BIRKHÄUSER

Progress in Mathematics
269

Algebra, Arithmetic, and Geometry

In Honor of Yu. I. Manin
Volume I

Yuri Tschinkel
Yuri Zarhin
Editors

BIRKHÄUSER

Progress in Mathematics
Volume 269

For other titles published in this series, go to
http://www.springer.com/series/4848

Algebra, Arithmetic, and Geometry

In Honor of Yu. I. Manin

Volume I

Yuri Tschinkel
Yuri Zarhin
Editors

Birkhäuser
Boston • Basel • Berlin

Editors
Yuri Tschinkel
New York University
Department of Mathematics
New York, NY 10012-1185
tschinkel@cims.nyu.edu

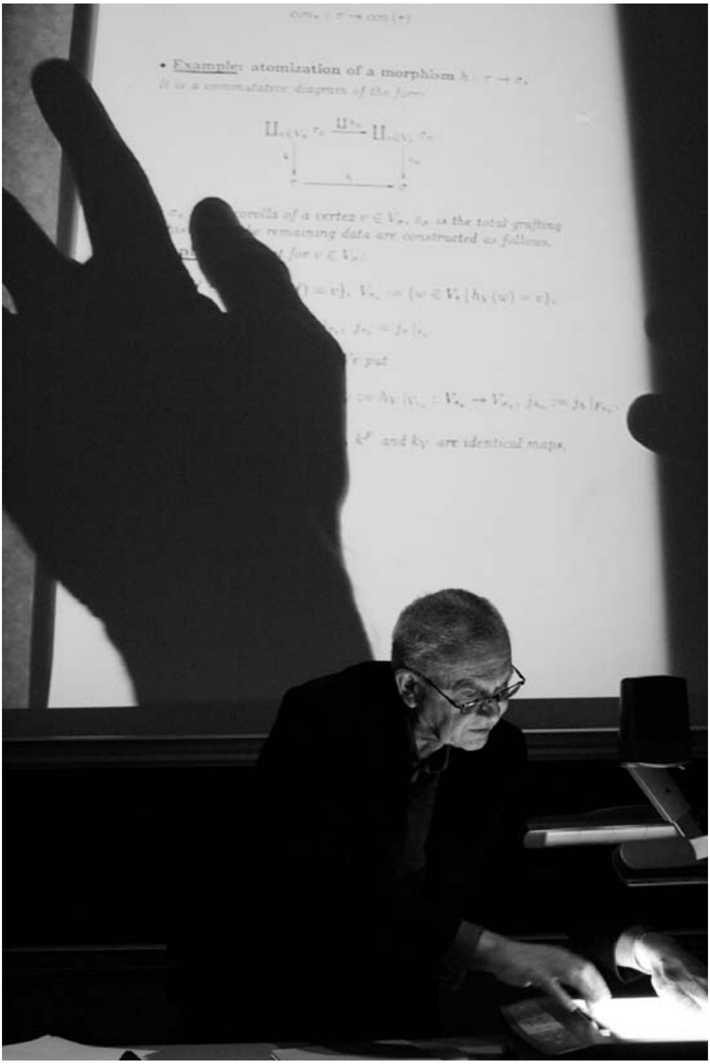
Yuri Zarhin
Pennsylvania State University
Eberly College of Science
Department of Mathematics
University Park, PA 16802
zarhin@math.psu.edu

ISBN 978-0-8176-4744-5 e-ISBN 978-0-8176-4745-2
DOI 10.1007/978-0-8176-4745-2
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2009939499

Printed on acid-free paper

Birkhäuser is part of Springer Science+Business Media (www.birkhauser.com)


• Example: atomization of a morphism
It is a commutative diagram of the form

Preface

Yuri Ivanovich Manin has made outstanding contributions to algebra, algebraic geometry, number theory, algorithmic complexity, noncommutative geometry and mathematical physics. His numerous achievements include the proof of the functional analogue of the Mordell Conjecture, the theory of the Gauss–Manin connection, proof with V. Iskovskikh of the nonrationality of smooth quartic threefolds, the theory of p-adic automorphic functions, construction of instantons (jointly with V. Drinfeld, M. Atiyah and N. Hitchin), and the theory of quantum computations.

We hope that the papers in this Festschrift, written in honor of Yu. I. Manin's seventieth birthday, will indicate the great respect and admiration that his students, friends and colleagues throughout the world all have for him.

June 2009

Courant Institute — Yuri Tschinkel
Penn State University — Yuri Zarhin

Contents

Curriculum Vitae
Yuri Ivanovich Manin

(updated May 30, 2009)

Personal

Born: February 16, 1937, Simferopol, Russia

Married to Ksenia Glebovna Semenova

Education

M.S. (summa cum laude): Moscow State University, 1958

Ph.D. (Candidate of Phys. Math. Sci.): Steklov Institute, 1960

Doctor of Phys.-Math. Sci. (Habilitation): Steklov Institute, 1963

Employment

Professor, Algebra Chair, Moscow University: 1965–1992

Principal Researcher, Steklov Institute, Academy of Sciences, Russia, Moscow: 1960–May 1993, since May 1993 Principal Researcher in absentia

Professor, Department of Mathematics, MIT: 1992–1993

Scientific Member and Collegium Member, Max-Planck-Institut für Math., Bonn: May 1993–February 2005; Director: November 1995–February 2005; Managing Director: November 1999–October 2001; Professor Emeritus: March 2005–now

Board of Trustees Professor, Northwestern University, Evanston: 2002–now

Selected visiting appointments

Pisa University, 1964

Montréal University, Aisenstadt Chair, 1988

IHES, Bures–sur–Yvette, France, 1967, 1989

Collège de France, Paris, 1989

Berkeley University, Miller Professor, 1989

Harvard University, 1991

Columbia University, Eilenberg Chair, 1992

University of Antwerpen, International Francqui Chair, 1996–7

Collège de France, Paris, 2006

Service to the mathematical community

Member, Shaw Prize in Mathematics Committee, 2008–2009

Member, Program Committee, European Congress of Mathematicians, Amsterdam, 2008

Member, Editorial Boards of two new Journals, founded in 2007: Non–Commutative Geometry; Algebra and Number Theory

Member, Assessment Committee of the Academy Professorships Programme, Royal Netherlands Academy of Arts and Sciences, 2004–2005, 2005–2006, 2006–2007

Member, Selection Committee, King Faisal International Prize for Mathematics, 2005

Chair, Program Committee, International Congress of Mathematicians, Beijing, 2002

Chair, Fields Prize Committee, International Congress of Mathematicians, Berlin, 1998

Member, Board of the Moscow Mathematical Society (various years)

Member, Program and Prize Committees of ICM and ECM (various years)

Member, Editorial and Advisory Boards of Mathematical Sbornik, Uspekhi Math. Nauk, Inventiones Mathematicae, Duke Mathematical Journal, Functional Analysis and its Applications, Crelle's Journal, Journal of Number Theory, Journal of Geometry and Physics, Advances of Mathematics, International Journal of Mathematics, American Journal of Mathematics et al. (various years)

Member, International Benchmarking of US Research Field Study, Math. Panel, 1997

Member, Senior Jury of Institut Universitaire de France, 1999

Co–organizer of international conferences and workshops

Mathematische Arbeitstagungen, Bonn, 1995, 1997 – 2007

Conformal Field Theory, Oberwolfach, 1996

Integrable systems, MPIM, Bonn, 1999

Non–Commutative Geometry, MPIM, Bonn, 1999

Frobenius Manifolds, MPIM, Bonn, 2002

Noncommutative Geometry and Number Theory, MPIM, Bonn, 2002, 2004

Dynamical Systems, MPIM, Bonn, 2004

Special session "Applications of Motives", AMS annual meeting 2004, Evanston

Geometry and Dynamics of Groups and Spaces, MPIM, 2006

Winter School and Workshop on Moduli Spaces, MPIM, 2008

Selected lectures

Invited speaker at the ICM congresses: Moscow, 1966; Nice, 1970; Helsinki, 1978 (plenary); Berkeley, 1986; Kyoto, 1990; Madrid 2006 (special activity)

Invited plenary speaker, European CM, Barcelona, 2000

Invited speaker at Bonn Arbeitstagungen: 1983, 1993

Rademacher Lecturer, University of Pennsylvania, 1989

Porter Lecturer, Rice University at Houston, 1989

Bowen Lecturer, Berkeley University, 1989

A. T. Brauer Lecturer, University of North Carolina, 1992

Leonardo da Vinci Lecturer, Milan University, 1992

Fermi Lecturer, Scuola Normale Superiore di Pisa, 1994

Schur Memorial Lecturer, Tel Aviv University, 1994

Regents' Lecturer, University of California at Riverside, 1997

Coble lecturer, University of Illinois at Urbana–Champaigne, 1999

Chern Lecturer, University of California at Berkeley, 1999

Blythe Lecturer, University of Toronto, 2004

Hardy Lecturer, London Mathematical Society, 2006

Honors and awards

Moscow Mathematical Society Award, 1963

Lenin Prize (Highest USSR National Prize), for work in Algebraic Geometry, 1967

Brouwer Gold Medal for work in Number Theory, Netherlands Royal Society and Mathematical Society, 1987

Frederic Esser Nemmers Prize in Mathematics, Northwestern University, Evanston, USA, 1994

Rolf Schock Prize in Mathematics of the Swedish Royal Academy of Sciences, Sweden, 1999

King Faisal International Prize for Mathematics, Saudi Arabia, 2002

Georg Cantor Medal of the German Mathematical Society, 2002

Order Pour le Mérite, Germany, 2007

Great Cross of Merit with Star, Germany, 2008

Elected Membership in Scientific Academies

Corresponding Member, Academy of Sciences, Russia, 1990

Foreign Member, Royal Academy of Sciences, the Netherlands, 1990

Member, Academia Europaea, 1993

Member, Max–Planck–Society for Scientific Research, Germany, 1993

Corresponding Member, Göttingen Academy of Sciences, Class of Physics and Mathematics, 1996

Member, Pontifical Academy of Sciences, Vatican, 1996

Member, German Academy of Sciences: Academia Leopoldina, 2000

Fellow, American Academy of Arts and Sciences, 2004

Foreign Member, Académie des sciences de l'Institut de France, 2005

Honorary degrees

Doctor Honoris Causa, Sorbonne (Université Pierre et Marie Curie, Paris VI), 1999

Abel Bicentennial Doctor Phil. Honoris Causa, University of Oslo, 2002

Doctor Honoris Causa, University of Warwick, 2006

Publications

Author and co-author of 11 monographs and about 240 papers in Algebraic Geometry, Number Theory, Mathematical Physics, History of Culture, Psycholinguistics

Teaching

Undergraduate teaching: courses of linear algebra and geometry, Moscow University, 1965–1991. Co-author (with A. I. Kostrikin) of the standard Course of Linear Algebra and Geometry (3rd printing in the USA).

Graduate teaching: Moscow University, 1965–1991, Massachusetts Institute of Technology 1992–1993; Northwestern University, 2002–2008; graduate courses and seminars on algebraic geometry, noncommutative geometry, homological algebra, number theory, theory of modular forms, non–linear differential equations, integrable systems, quantum field theory, applied algebra, error–correcting codes, mathematical logic et al.

Ph.D. Students

2007: D. Borisov

2006: A. Bayer

2002: M. Rosellen

1998: V. Kolmykov

1997: R. Kaufmann

1992: Yu. Tschinkel, E. Demidov

1990: A. Verevkin

1989: M. Kapranov

1988: A. Beilinson, A. Voronov, A. Levin

1987: I. Penkov, A. Vaintrob, A. Skorobogatov

1985: B. Tsygan

1984: M. Wodzicki, S. Vladut

1983: M. Tsfasman, G. Shabat, Hoang Le Minh

1981: D. Lebedev, V. Kolyvagin, D. Leites

1980: M. Vishik, A. Panchishkin

1979: V. Drinfeld

1978: G. Mustafin

1966–1977: Kha Huy Khoai, D. Kanevskii, P. Kurchanov, V. Berkovich, K. Kii, V. Shokurov, I. Cherednik, A. Roitman, M. Frumkin, Yu. Zarkhin, El Hushi, A. Geronimus, A. Shermenev, V. Danilov, V. Iskovskih, A. Belskii, Yu. Vainberg, B. Martynov.

List of Publications

1956

[1] On cubic congruences to a prime modulus. Russian: *Izv. AN SSSR, ser. mat.*, **20:6**, (1956), 673–678. English: *AMS Translations, ser. 2*, **13**, (1960), 1–7.

1958

[2] Algebraic curves over fields with differentiation. Russian: *Izv. AN SSSR, ser. mat.*, **22:6**, (1958), 737–756. English: *AMS Translations, ser. 2*, **37**, (1964), 59–78.

1959

[3] On the moduli of a field of algebraic functions. Russian: *Doklady AN SSSR*, **125:3**, (1959), 488–491.

1961

[4] The Hasse–Witt matrix of an algebraic curve. Russian: *Izv. AN SSSR, ser. mat.*, **25:1**, (1961), 153–172. English: *AMS Translations, ser. 2*, **45**, (1965), 245–264.

[5] On the Diophantine equations over functional fields. Russian: *Doklady AN SSSR*, **139:4**, (1961), 806–809.

[6] On the ramified coverings of algebraic curves. Russian: *Izv. AN SSSR, ser. mat.*, **25:6**, (1961), 789–796.

[7] On the theory of abelian varieties. Russian: *PhD Thesis*, Steklov Math. Institute, Moscow, (1961), 66 pp.

1962

[8] On the theory of abelian varieties over a field of finite characteristic. Russian: *Izv. AN SSSR, ser. mat*, **26:2**, (1962), 281–292.

[9] A remark on Lie p–algebras. Russian: *Sibirskii Mat. Journal*, **3:3**, (1962), 479–480.

[10] An elementary proof of Hasse's theorem. Russian: Chapter 10 in *Elementary Methods in Analytic Number Theory*, by A. O. Gelfond and Yu. V. Linnik, Moscow, Fizmatgiz, (1962). 13 pp.

[11] Two–dimensional formal abelian groups. Russian: *Doklady AN SSSR*, **143:1**, (1962), 35–37.

[12] On the classification of the formal abelian groups. Russian: *Doklady AN SSSR*, **144:3**, (1962), 490–492.

[13] On the geometric constructions with compasses and ruler. Russian: Chapter 6 in *Encyclopaedia of Elementary Mathematics*, **6**, Fizmatgiz, (1962), 14 pp.

1963

[14] Rational points of algebraic curves over functional fields. Russian: *Izv. AN SSSR, ser. ma.*, **27:6**, (1963), 1395–1440. English: *AMS Translations ser. 2*, **50**, (1966), 189–234.

[15] The theory of commutative formal groups over fields of finite characteristic. Russian: *Uspekhi Mat. Nauk*, **18:6**, (1963), 3–90. English: *Russian Math. Surveys*, **18:6**, (1963), 1–83.

[16] A proof of the analogue of the Mordell conjecture for curves over functional fields. Russian: *Doklady AN SSSR*, **152:5**, (1963), 1061–1063.

[17] On the arithmetic of rational surfaces. Russian: *Doklady AN SSSR*, **152:1**, (1963), 47–49. English: *Soviet Math. Dokl.*, **4**, (1963), 1243–1247.

1964

[18] The Tate height of points on an abelian variety, its variants and applications. Russian: *Izv. AN SSSR, ser. mat.*, **28:6**, (1964), 1363–1390. English: *AMS Translations, ser. 2*, **59**, (1966), 82–110.

[19] Diophantine equations and algebraic geometry. Russian: In: *Proc. of the 4th All–Union Math. Congress vol. 2*, Fizmatgiz, (1964), 15–21.

[20] Rational points on algebraic curves. Russian: *Uspekhi Mat. Nauk*, **19:6**, (1964), 83–87.

1965

[21] Moduli fuchsiani. *Annali Scuola Norm. Sup. di Pisa*, **19**, (1965), 113–126.

[22] Minimal models of ruled and rational surfaces (with Yu. R. Vainberg). Russian: In: *Algebraic surfaces*, ed. by I. R. Shafarevich, Trudy MIAN **75** (1965) 75–91.

[23] Algebraic topology of algebraic varieties. Russian: *Uspekhi Mat. Nauk*, **20:6**, (1965), 3–12.

1966

[24] Rational surfaces over perfect fields. Russian: *Publ. Math. IHES*, **30**, (1966), 415–457. English: *AMS Translations (ser. 2)*, **84**, (1969), 137–186.

[25] Differential forms and sections of elliptic pencils. Russian: In: *Contemporary Problems of the Theory of Analytic Functions*, Nauka, Moscow, (1966), 224–229.

[26] Two theorems on rational surfaces. In: *Simp. Int. di Geom. Algebrica*, Roma, (1966), 198–207.

1967

[27] Rational G–surfaces. Russian: *Doklady AN SSSR*, **1175:1**, (1967), 28–30.

[28] Rational surfaces over perfect fields. II. Russian: *Mat. Sbornik*, **72:2**, (1967), 161–192. English: *Math. USSR Sbornik*, **1**, (1967), 141–168.

1968

[29] Correspondences, motives, and monoidal transforms. Russian: *Mat. Sbornik*, **77:4**, (1968), 475–507. English: *Mathematics of the USSR Sbornik*, **6**, (1968), 439–470.

[30] Rational surfaces and Galois cohomology. In: *Proc. Moscow ICM*, MIR, Moscow (1968), 495–509.

[31] On some groups related to cubic surfaces. In: *Algebraic geometry.* Tata Press, Bombay, (1968), 255–263.

[32] Cubic hypersurfaces. I. Quasigroups of point classes. Russian: *Izv. AN SSSR, ser. mat.*, **32:6**, (1968), 1223–1244. English: *Math USSR – Izvestia*, **2**, (1968), 1171–1191.

[33] Lectures on Algebraic Geometry. *Comp. Center*, Moscow University, (1968), 185 pp.

1969

[34] The p–torsion of elliptic curves is uniformly bounded. Russian: *Izv. AN SSSR, ser. mat.*, **33:3**, (1969), 459–465. English: *Math. USSR Izvestia*, **3**, (1969), 433–438.

[35] Hypersurfaces cubiques. II. Automorphismes birationnelles en dimension deux. *Inv. Math.*, **6**, (1969), 334–352.

[36] Cubic hypersurfaces. III. Moufang loops and Brauer equivalence. Russian: *Mat. Sbornik*, **79:2**, (1969), 155–170. English: *Math. USSR–Sbornik*, **24:5**, (1969), 1–89.

[37] Lectures on the K–functor in algebraic geometry. Russian: *Uspekhi Mat. Nauk*, **24:5**, (1969), 3–86. English: *Russian Math. Surveys*, **24:5**, (1969), 1–89.

[38] Comments on the five Hilbert's problems. Russian: In: *Hilbert Problems*, Nauka, Moscow, (1969), 154–162, 171–181, 196–199.

[39] Regular elements in the Cremona group. Russian: *Mat. Zametki*, **5:2**, (1969), 145–148.

1970

[40] The refined structure of the Néron-Tate height. Russian: *Mat. Sbornik*, **83:3**, (1970), 331–348. English: *Math. USSR–Sbornik*, **12**, (1970), 325–342.

[41] Lectures on Algebraic Geometry I: Affine Schemes. Russian: Moscow University, (1970), 133 pp.

1971

[42] Le groupe de Brauer-Grothendieck en géométrie diophantienne. In: *Actes Congr. Int. Math. Nice*, Gauthier–Villars (1971), **1**, 401–411.

[43] Three–dimensional quartics and counterexamples to the Lüroth conjecture (with V. A. Iskovskih). Russian: *Mat. Sbornik*, **86:1**, (1971), 140–166. English: *Math. USSR–Sbornik*, **15**, (1972), 141–166.

[44] Cyclotomic fields and modular curves. Russian: *Uspekhi Mat. Nauk*, **26:6**, (1971), 7–71. English: *Russian Math. Surveys*, **26**, (1971), 7–78.

[45] Mordell–Weil theorem. Russian: Appendix to *Abelian Varieties* by D. Mumford, Mir, Moscow, (1971), 17 pp.

1972

[46] Parabolic points and zeta-functions of modular curves. Russian: *Izv. AN SSSR, ser. mat.*, **36:1**, (1972), 19–66. English: *Math. USSR Izvestija, publ. by AMS*, **6:1**, (1972), 19–64.

[47] Cubic Forms: Algebra, Geometry, Arithmetic. Russian: Nauka, Moscow, (1972), 304 pp.

[48] Height on families of abelian varieties (with Ju. G. Zarhin). Russian: *Mat. Sbornik*, **89:2**, (1972), 171–181. English: *Math. USSR Sbornik*, **18**, (1972), 169–179.

1973

[49] Periods of p–adic Schottky groups (with V. G. Drinfeld). *Journ. f. d. reine u. angew. Math.*, **262/263**, (1973), 239–247.

[50] Periods of parabolic forms and p–adic Hecke series. Russian: *Mat. Sbornik*, **92:3**, (1973), 378–401. English: *Math. USSR Sbornik*, **21:3**, (1973), 371–393.

[51] Explicit formulas for the eigenvalues of Hecke operators. *Acta Arithmetica*, **24**, (1973), 239–249.

[52] Hilbert's tenth problem. Russian: In: *Sovr. Probl. Mat.*, **1**, (1973), 5–37. English: *Journ of Soviet Math.*

1974

[53] The values of p–adic Hecke series at integer points of the critical strip. Russian: *Mat. Sbornik*, **93:4**, (1974), 621–626. English: *Math. USSR Sbornik*, **22:4**, (1974), 631–637.

[54] p–adic Hecke series of imaginary quadratic fields (with M. M. Vishik). Russian: *Mat. Sbornik*, **95:3**, (1974), 357–383. English: *Math. USSR Sbornik*, **24:3**, (1974), 345–371.

[55] p–adic automorphic functions. Russian: In: *Sovrem. Probl. Mat.*, **3**, (1974), 5–92. English: *Journ. of Soviet Math.*, **5**, (1976), 279–333.

[56] Lectures on Mathematical Logic. Russian: Moscow Institute of Electronic Engineering, (1974), Part 1, 133 pp., Part 2, 69 pp.

[57] Cubic Forms: Algebra, Geometry, Arithmetic. North Holland, Amsterdam, (1974), 292 pp.

1975

[58] Continuum problem. Russian: In: *Sovrem. Probl. Mat.*, **5**, (1975), 5–72.

[59] Gödel's Theorem. Russian: *Priroda*, **12**, (1975), 80–87.

1976

[60] Non–Archimedean integration and p–adic Jacquet–Langlands L–series. Russian: *Uspekhi Mat. Nauk*, **31:1**, (1976), 5–54. English: *Russian Math. Surveys*, **31:1**, (1976), 5–57.

1977

[61] Poisson brackets and the kernel of the variational derivation in the formal calculus of variations (with I. M. Gelfand and M. A. Shubin). Russian: *Funkc. Anal. i ego Prilozen.*, **10:4**, (1977), 31–42. English: *Func. Anal. Appl.*, **10:4**, (1977).

[62] Convolutions of Hecke series and their values at lattice points (with A. A. Panchishkin). Russian: *Mat. Sbornik*, **104:4**, (1977), 617–651. English: *Math. USSR Sbornik*, **33:4**, (1977), 539–571.

[63] Long wave equations with a free surface. I. Conservation laws and solutions (with B. Kupershmidt). Russian: *Funkc. Anal. i ego Prilozen.*, **11:3**, (1977), 31–42. English: *Func. Anal. Appl.*, **11:3**, (1977), 188–197.

[64] A Course in Mathematical Logic. Springer Verlag, (1977) XIII+286.

[65] Assioma/Postulato. In: *Enciclopaedia Einaudi*, **1**, (1977), 992–1010, Torino, Einaudi.

[66] Applicazioni. In: *Enciclopaedia Einaudi*, **1**, (1977), 701–743, Torino, Einaudi.

[67] Men and Signs. Russian: In: *Priroda*, **5**, (1977), 150–152.

1978

[68] Long wave equations with a free surface II. The Hamiltonian structure and the higher equations (with B. Kupershmidt). Russian: *Funkcional. Anal. i ego Prilozen.*, **12:1**, (1978), 25–37. English: *Func. Anal. Appl.*, **12:1**, (1978), 20–29.

[69] Matrix solitons and vector bundles over singular curves. Russian: *Func. Anal. i ego Prilozen.*, **12:4**, (1978), 53–63. English: *Func. Anal. Appl.*, **12:4**, (1978).

[70] Algebraic aspects of non–linear differential equations. Russian: *In: Sovrem. Probl. Mat.*, **11**, (1978), 5–152. English: *Journ. of Soviet Math.*, **11**, (1979), 1–122.

[71] Self–dual Yang–Mills fields over a sphere (with V. G. Drinfeld). Russian: *Func. Anal. i ego Prilozen.*, **12:2**, (1978), 78–79. English: *Func. Anal. Appl.*, **12:2**, (1978).

[72] On the locally free sheaves on CP^3 connected with Yang–Mills fields (with V. G. Drinfeld). Russian: *Uspekhi Mat. Nauk*, **33:3**, (1978), 241–242.

[73] Construction of instantons (with M. F. Atiyah, V. G. Drinfeld, and N. J. Hitchin). *Phys. Lett. A*, **65:3**, (1978), 185–187.

[74] Instantons and sheaves on CP^3 (with V. G. Drinfeld). In: *Springer Lecture Notes in Math.*, **732**, (1978), 60–81.

[75] A description of instantons (with V. G. Drinfeld). *Comm. Math. Phys.*, **63**, (1978), 177–192.

[76] A description of instantons II (with V. G. Drinfeld). Russian: In: *Proc. of Int. Seminar on the Physics of High Energy and Quantum Field Theory*, Serpoukhov, (1978), 71–92.

[77] Modular forms and number theory. In: *Proc. Int. Cogr. of Math.*, Helsinki (1978), **1**, 177–186.

[78] Continuo/Discreto. In: *Enciclopaedia Einaudi*, **3** (1978), 935–986, Torino, Einaudi.

[79] Dualitá (with I. M. Gelfand). In: *Enciclopaedia Einaudi*, **5**, (1978), 126–178, Torino, Einaudi.

[80] Divisibilitá. In: *Enciclopaedia Einaudi*, **5**, (1978), 14–37, Torino, Einaudi.

1979

[81] Conservation laws and Lax representation of Benney's long wave equations (with D. R. Lebedev). *Phys. Lett. A*, **74**, (1979), 154–156.

[82] Gelfand-Dikii Hamiltonian operator and the co-adjoint representation of the Volterra group (with D. R. Lebedev). Russian: *Func. Anal. i ego Prilozhen.*, **13:4**, (1979), 40–46. English: *Func. Anal. Appl.*, **13:4**, (1979), 268–273.

[83] Yang-Mills fields, instantons, tensor product of instantons (with V. G. Drinfeld). Russian: *Yad. Fiz.*, **29:6**, (1979), 1646–1653. English: *Soviet J. Nucl. Phys.*, **29:6**, (1979), 845–849.

[84] Modular forms and number theory. In: *Proc. Int. Math. Congr. 1978*, Helsinki, (1979), **1**, 177–186.

[85] Mathematics and Physics. Russian: Znaniye, Moscow, (1979), 63 pp.

[86] Provable and Unprovable. Russian: *Sov. Radio*, (1979), 166 pp.

[87] Insieme. In: *Enciclopaedia Einaudi*, **7**, (1979), 744–776, Torino, Einaudi.

[88] Razionale/Algebrico/Transcedente. In: *Enciclopaedia Einaudi*, **11**, (1979), 603–628, Torino, Einaudi.

[89] A new encounter with Alice. Russian: In: *Priroda*, **7**, (1979), 118–120.

1980

[90] An instanton is determined by its complex singularities (with A. A. Beilinson and S. I. Gelfand). Russian: *Funkc. Analiz i ego Priloz.*, **14:2**, (1980), 48–49. English: *Func. Anal. Appl*, **14:2**, (1980).

[91] Twistor description of classical Yang–Mills fields. *Phys. Lett. B*, **95:3–4**, (1980), 405–408.

[92] Benney's long wave equations II. The Lax representation and conservation laws (with D. R. Lebedev). Russian: *Zap. Nauchn. Sem. LOMI*, **96**, (1980), 169–178.

[93] Penrose transform and classical Yang–Mills fields. Russian: In: *Group Theoretical Methods in Physics*, Nauka, (1980), **2**, 133–144.

[94] Methods of algebraic geometry in modern mathematical physics (with V. G. Drinfeld, I. Krichever, and S. P. Novikov). *Math. Phys. Review 1*, (1980), Harwood Academic, Chur, 3–57.

[95] Computable and Uncomputable. Russian: *Sov. Radio*, (1980), 128 pp.

[96] Linear Algebra and Geometry (with A. I. Kostrikin). Russian: Moscow University, (1980), 319 pp.

1981

[97] Gauge fields and holomorphic geometry. Russian: In: *Sovr. Probl. Mat.*, **17**, (1981), 3–55. English: *Journal of Soviet Math.*

[98] Hidden symmetries of long waves. *Physica D*, **3:1–2**, (1981), 400–409.

[99] On the cohomology of twistor flag spaces (with G. M. Henkin). *Compositio Math.*, **44**, (1981), 103–111.

[100] Expanding constructive universes. In: *Springer Lect. Notes in Comp. Sci.*, **122**, (1981), 255–260.

[101] Simmetria (with I. M. Gelfand). In: *Enciclopaedia Einaudi*, **12**, (1981), 916–943, Torino, Einaudi.

[102] Strutture matematiche. In: *Enciclopaedia Einaudi*, **13**, (1981), 765–798, Torino, Einaudi.

[103] Natural language in scientific texts. Russian: In: *Structure of Text–81*, Moscow, Nauka, (1981), 25–27.
[104] Mathematics and Physics. Birkhäuser, Boston, (1981), 100 pp.
[105] Lo Dimostrable e Indemostrable. Mir, Moscow, (1981), 262 pp.

1982

[106] Null-geodesics of complex Einstein spaces (with I. B. Penkov). Russian: *Funk. Analiz i ego Priloz.*, **16:1**, (1982), 78–79. English: *Func. Anal. Appl.*, **16:1**, (1982).
[107] Yang–Mills–Dirac equations as Cauchy–Riemann equations on the twistor space (with G. M. Henkin). Russian: *Yad. Fiz.*, **35:6**, (1982), 1610–1626.
[108] Gauge fields and cohomology of spaces of null–geodesics. In: *Gauge Theories, Fundamental Interactions and Related Results*, Birkhäuser, (1982), 231–234.
[109] What is the maximum number of points on a curve over F_2? *Journ. Fac. Sci. Univ. Tokyo, ser. 1A*, **28:3**, (1982), 715–720.
[110] Remarks on algebraic supermanifolds. Russian: In: *Algebra (for the 90th anniversary of O. Yu. Schmidt)*, Moscow University, (1982), 95–101.
[111] Mathematics and Physics. Bulgarian: *Nauka i Izkusstvo*, Sofia, (1982), 56 pp.
[112] A Trilogy about Mathematics. Russian: In: *Znaniye–Sila*, **3** (1982), 32.

1983

[113] Flag superspaces and supersymmetric Yang–Mills equations. In: *Arithmetic and Geometry, papers to honor I. R. Shafarevich*, **2**, 175–198, Birkhäuser, (1983).
[114] Supersymmetry and supergravity in the superspace of null supergeodesics. Russian: In: *Group Theoretical Methods in Physics*, Nauka, Moscow, (1983), **1**, 203–208.
[115] Geometric ideas in the field theory. Russian: In: *Geometric ideas in physics*, Mir, (1983), 5–16.
[116] Tynyanov and Griboyedov. Russian: *Revue des Etudes Slaves*, t. LV, f. **3** (1983), 507–521, Paris.

1984

[117] Geometry of supergravity and Schubert supercells. Russian: *Zapiski Sem. LOMI*, **133**, (1984), 160–176.
[118] Grassmannians and flags in supergeometry. In: *Problems of Modern Analysis*, Moscow University, (1984), 83–101.
[119] Schubert supercells. Russian: *Func. Analiz i ego Priloz.*, **18:4**, (1984), 75–76. English: *Func. Anal. Appl.*, **18:4**, (1984), 329–330.

[120] New exact solutions and a cohomological analysis of the classical and supersymmetric Yang–Mills equations. Russian: *Trudy MIAN*, **165**, (1984), 98–114.

[121] Holomorphic supergeometry and Yang–Mills superfields. Russian: In: *Sovr. Probl. Mat.*, **24**, (1984), 3–80. English: *Journal of Soviet Math.*

[122] New directions in geometry. Russian: *Uspekhi Mat. Nauk*, **39:6**, (1984), 47–73. English: *Russian Math. Surveys*, **39:6**, (1984), 51–83, and *Springer Lecture Notes in Math.*, **1111**, (1985), 59–101.

[123] The inverse scattering transform and classical equations of motion. In: *Nonlinear and Turbulent Processes in Physics*, Harwood Academic, Chur, (1984), **3**, 1487–1502.

[124] Linear codes and modular curves (with S. G. Vladut). Russian: In: *Sovr. Probl. Mat.*, **25**, (1984), 209–257. English: *J. Soviet Math.*, **30**, (1985), 2611–2643.

[125] Some applications of algebraic geometry. Russian: *Trudy Mian*, **168**, (1984), 110–131.

[126] Gauge Fields and Complex Geometry. Russian: Nauka, Moscow, (1984). 335 pp.

1985

[127] A supersymmetric extension of the Kadomtsev-Petviashvili hierarchy (with A. O. Radul). *Comm. Math. Phys.*, **98**, (1985), 65–77.

[128] Sun, a Poor Totem. Russian: In: *Priroda*, **6**, (1985), 123–127.

[129] Geometry unbound. *Science*, **9**, (1985), 89–91.

1986

[130] The twistor transformation and algebraic–geometric constructions of solutions of the equations of field theory (with M. M. Kapranov). Russian: *Uspekhi Mat. Nauk*, **41:5**, (1986), 85–107. English: *Russian Math. Surveys*, **41:5**, (1986), 33–61.

[131] On the higher Bruhat orders related to the symmetric group. Russian: *Func. Anal. i ego Prilozhen.*, **20:2**, (1986), 74–75. English: *Func. Anal. Appl.*, **20:2**, (1986), 148–150.

[132] Critical dimensions of string theories and a dualizing sheaf on the moduli space of (super)curves. Russian: *Funkc. Anal. i ego Prilozhen.*, **20:3**, (1986), 88–89. English: *Func. Anal. Appl.*, **20:3**, (1986), 244–246.

[133] The partition function of the Polyakov string can be expressed in terms of theta functions. *Phys. Lett. B*, **172:2**, (1986), 184–185.

[134] The Mumford form and the Polyakov measure in string theory (with A. A. Beilinson). *Comm. Math. Phys.*, **107:3**, (1986), 359–376.

[135] The Radon-Penrose transformation for the group $SO(8)$ and instantons (with Hoang Le Minh). Russian: *Izv. AN SSSR, ser. mat.*, **50:1**, (1986), 195–206. English: *Math. USSR Izvestia*, **50 (1)**, (1986), 189–200.

[136] The multiloop contribution in the fermionic string (with M. A. Baranov, I. V. Frolov, and A. S. Schwartz). Russian: *Yad. Fiz.*, **43:4**, (1986), 1053–1055. English: *Sov. J. Nucl. Phys.*, **43**, (1986), 670–671.

[137] Arrangements of real hyperplanes and the Zamolodchikov equations (with V. V. Shekhtman). Russian: In: *Group Theoretical Methods in Physics*, **1**, (Jurmala, 1985), 316–325, Nauka, Moscow, (1986). English: *Group Theoretical Methods in Physics*, **1**, (Yurmala, 1985), 151–165, VNU Sci. Press, Utrecht, (1986).

[138] Rational varieties: algebra, geometry, and arithmetic (with M. Tsfasman). Russian: *Uspekhi mat. Nauk*, **41:2**, (1986), 43–94. English: *Russian Math. Surveys*, **41:2**, (1986), 51–116.

[139] Cubic Forms: Algebra, Geometry, Arithmetic. 2nd edition, revised and completed. North Holland, Amsterdam, (1986), 326 pp.

[140] Bevezetés a kiszámithatósag matematikai elméletébe (Hungarian translation of Computable and Uncomputable.) *Müszaki Könyvvkiadó*, Budapest, (1986), 144 pp.

1987

[141] Quantum strings and algebraic curves. In: *Proc. Int. Congr. of Math.*, Berkeley, (1986), AMS, Providence, RI, (1987), **2**, 1286–1295.

[142] Some remarks on Koszul algebras and quantum groups. *Ann. Inst. Fourier*, **37:4**, (1987), 191–205.

[143] Values of Selberg's zeta function at integer points (with A. A. Beilinson). Russian: *Func. Analiz i ego Prilozen.*, **21:1**, (1987), 68–69. English: *Func. Anal. Appl.*, **21:1**, (1987), 58–60.

[144] Sheaves of the Virasoro and Neveu-Schwarz algebras (with A. A. Beilinson and V. V. Shechtman). In: *Springer Lecture Notes in Math.* **1289**, (1987), 52–66.

[145] A superanalog of the Selberg trace formula and multiloop contribution for fermionic string (with M. A. Baranov, I. V. Frolov, and A. S. Schwartz). *Comm. Math. Phys.*, **111**, (1987), 373–392.

[146] On the problem of the early stages of speech and consciousness (phylogenesis). Russian: In: *Intellectual Processes and their Modeling*, Nauka, Moscow, (1987), 154–178.

[147] "It is love." Russian: In: *Priroda*, **4**, (1987), 118–120.

[148] The mythological trickster in psychology and history of culture. Russian: *Priroda*, **7**, (1987), 42–52.

1988

[149] Quantum groups and non-commutative geometry. Montreal University Press, (1988), 88 pp.

[150] Neveu–Schwarz sheaves and differential equations for Mumford superforms. *J. of Geometry and Physics*, **5:2**, (1988), 161–181.

[151] Brouwer Memorial Lecture 1987. *Nieuw Arch. Wisk.* (4) 6, (1988), no. 1–2, 1–6.

[152] Elements of supergeometry (with A. A. Voronov and I. B. Penkov). Russian: In: *Sovr. Probl. Mat.*, **32**, (1988), 3–25.

[153] Supercell partitions of flag superspaces (with A. A. Voronov). Russian: In: *Sovr. Probl. Mat.*, **32**, (1988), 27–70.

[154] Methods of the Homological Algebra I. Introduction to the Cohomology Theory and Derived Categories (with S. I. Gelfand). Russian: Nauka, Moscow, (1988), 416 pp.

[155] Gauge Field Theory and Complex Geometry. Springer Verlag, (1988), 295 pp.

[156] Homological Algebra (with S. I. Gelfand). Russian: *Sovr. Probl. Mat.*, **38**, (1988), 238 pp.

1989

[157] Arrangements of hyperplanes, higher braid groups, and higher Bruhat orders (with V. V. Schechtman). In: *Advanced Studies in Pure Math.*, **17**, *Algebraic Number Theory*, 289–308, Academic Press, (1989).

[158] Rational points of bounded height on Fano varieties (with J. Franke and Yu. Tschinkel). *Inv. Math.*, **95**, (1989), 421–435.

[159] Multiparametric quantum deformation of the general linear supergroup. *Comm. Math. Phys.*, **123**, (1989), 123–135.

[160] Determinants of Laplacians on Riemann surfaces. In: *Conformal Invariance and string theory* (Poiana Brasov, 1987), Academic Press, Boston, MA, (1989), 285–291.

[161] Reflections on arithmetical physics. In: *Conformal Invariance and string theory (Poiana Brasov, 1987)*, Academic Press, Boston, MA, (1989), 293–303.

[162] The formalism of left and right connections on supermanifolds (with I. B. Penkov). In: *Lectures on Supermanifolds, Geometrical Methods and Conformal Groups*, World Scientific, (1989), 3–13.

[163] Strings. *Math. Intelligencer*, **11:2**, (1989), 59–65.

[164] Multiparametric quantum deformation of the general linear supergroup. *Comm. Math. Phys.*, **123**, (1989), 123–135.

[165] Letter to the editor. Russian: *Izv. AN SSSR, ser. mat.*, **53:2**, (1989). English: *Math. USSR Izvestija*, **34:2**, (1990), 465–466.

[166] Linear Algebra and Geometry (with A. I. Kostrikin). Gordon and Breach, (1989), 309 pp.

[167] Elementary Particles: Mathematics, Physics and Philosophy (with I. Yu. Kobzarev). Reidel, Dordrecht, (1989), 227 pp.

1990

[168] Sur le nombre des points rationnels de hauteur bornée des variétés algébriques (with V. V. Batyrev). *Math. Annalen*, **286**, (1990), 27–43.

[169] Non-standard quantum deformations of $GL(n)$ and constant solutions of the Yang-Baxter equation (with E. E. Demidov, E. E. Mukhin and D. V. Zhdanovich). *Progress of Theor. Phys. Supplement*, **102**, (1990), 203–218.

[170] Quantized theta-functions. *Progress of Theor. Phys. Supplement*, **102**, (1990).

[171] Number Theory (with A. A. Panchishkin). Russian: *Sovr. Probl. Mat.*, **49**, (1990), 348 pp.

[172] Mathematics as metaphor. *Proc. of ICM, Kyoto 1990, vol. II*, 1665–1671, The AMS and Springer Verlag.

1991

[173] Three–dimensional hyperbolic geometry as ∞–adic Arakelov geometry. *Inv. Math.*, **104**, (1991), 223–244.

[174] Quantum Groups. *Nederl. Wetensk. Verslag, Afd. Naturkunde*, **100**, (1991), 55–68.

[175] Topics in Non–commutative Geometry. Princeton University Press, (1991), 163 pp.

1992

[176] Notes on quantum groups and quantum de Rham complexes. *Teoreticheskaya i Matematicheskaya Fizika*, **92:3**, (1992), 425–450.

[177] The Empty City archetype. Russian: In: *Arbor Mundi*, ed. by E. Meletinsky, **1**, (1992), 28–34.

1993

[178] Notes on the arithmetic of Fano threefolds. *Comp. Math.*, **85**, (1993), 37–55.

[179] Points of bounded height on del Pezzo surfaces (with Yu. Tschinkel). *Comp. Math*, **85**, (1993), 315–332.

1994

[180] Homological Algebra (with S. I. Gelfand). *Enc. of Math. Sci.*, **38**, Springer Verlag, (1994), 222 pp.

[181] Gromov–Witten classes, quantum cohomology, and enumerative geometry (with M. Kontsevich). *Comm. Math. Phys.*, **164:3**, (1994), 525–562.

1995

[182] Lectures on zeta functions and motives (according to Deninger and Kurokawa). In: *Columbia University Number Theory Seminar*, Astérisque **228**, (1995), 121–164.

[183] Number Theory I. Fundamental Problems, Ideas and Theories (with A. Panchishkin). *Enc. of Math. Sci.*, **49**, Springer Verlag, (1995), 303 pp.

[184] Generating functions in algebraic geometry and sums over trees. In: *The Moduli Space of Curves*, ed. by R. Dijkgraaf, C. Faber and G. van der Geer. *Progress in Math.*, **129**. Birkhäuser, (1995), 401–417.

[185] Problems on rational points and rational curves on algebraic varieties. *Surveys in Differential Geometry*, **2**, (1995), Int. Press, 214–245.

1996

[186] Quantum cohomology of a product (with M. Kontsevich and Appendix by R. Kaufmann). *Inv. Math.*, 124, f. 1–3 (1996), 313–339 (Remmert's Festschrift).

[187] Quantum groups and algebraic groups in non–commutative geometry. In: *Quantum Groups and their Applications in Physics*, ed. by L. Castellani and J. Wess, *Proc. of the Int. School of Physics "Enrico Fermi"*, IOS Press, (1996), 347–359.

[188] Distribution of rational points on Fano varieties. *Publ. RIMS Kokyuroku 958, Analytic Number Theory*, (1996), 98–104.

[189] Automorphic pseudodifferential operators (with P. B. Cohen and D. Zagier). In: *Algebraic Aspects of Integrable Systems, (in memory of Irene Dorfman)*, ed. by A. S. Fokas and I. M. Gelfand, Birkhäuser, Boston (1996), 17–47.

[190] Stacks of stable maps and Gromov–Witten invariants (with K. Behrend). *Duke Math. Journ.*, **85:1**, (1996), 1–60.

[191] Higher Weil–Petersson volumes of moduli spaces of stable n–pointed curves (with R. Kaufmann and D. Zagier). *Comm. Math. Phys.*, **181:3**, (1996), 763–787.

[192] Selected Papers. *World Scientific Series in 20th Century Mathematics*, **3**, World Sci., Singapore, (1996), xii+600 pp.

[193] Methods of homological algebra (with S. I. Gelfand, translation of 154). *Enc. of Math. Sci*, **38**, Algebra V, Springer Verlag, 1996, xv+372 pp.

1997

[194] Mordell-Weil Problem for Cubic Surfaces In: *Advances in the Mathematical Sciences—CRM's 25 Years (L. Vinet, ed.) CRM Proc. and Lecture Notes.*, **11**, Amer. Math. Soc., Providence, RI, (1997), 313–318.

[195] Semisimple Frobenius (super)manifolds and quantum cohomology of P^r (with S. A. Merkulov). *Topological Methods in Nonlinear Analysis*, **9:1**, (1997), 107–161 (Ladyzhenskaya's Festschrift).

[196] Vatican, Fall 1996. Russian: In: *Priroda*, **8**, (1997), 61–66.

[197] Gauge Field Theory and Complex Geometry, Second Edition (with Appendix by S. Merkulov). Springer Verlag, (1997), xii + 346 pp.

[198] Linear Algebra and Geometry (with A. I. Kostrikin), Paperback Edition. Gordon and Breach, (1997), ix + 308 pp.

[199] Elementary Particles (with I. Yu. Kobzarev). Russian version of 167, Phasis, Moscow (1997), 206 pp.

1998

[200] Sixth Painlevé equation, universal elliptic curve, and mirror of P^2. In: *Geometry of Differential Equations*, ed. by A. Khovanskii, A. Varchenko, and V. Vassiliev. Amer. Math. Soc. Transl. (2), **186**, 131–151.

[201] Relations between the correlators of the topological sigma–model coupled to gravity (with M. Kontsevich). *Comm. Math. Phys.*, **196**, (1998), 385–398.

[202] Stable maps of genus zero to flag spaces. *Topological Methods in Nonlinear Analysis*, **11:2**, (1998), 207–218 (Moser's Festschrift). Corrections: **15** (2000), 401.

[203] Interrelations between mathematics and physics. In: *Séminaires et Congrès, No 3*, Soc. Math. de France, (1998), 158–168.

[204] Truth, rigour, and common sense. In: *Truth in Mathematics*, ed. by H. G. Dales and G. Oliveri, Clarendon Press, Oxford, **1998**, 147–159.

1999

[205] Three constructions of Frobenius manifolds: a comparative study. *Asian J. Math*, **3:1**, (1999), 179–220. (Atiyah's Festschrift).

[206] Weak Frobenius manifolds (with C. Hertling.) *Int. Math. Res. Notices*, **6**, (1999), 277–286.

[207] The work of Maxim Kontsevich. *Notices of the AMS*, (Jan. 1999), 21–23.

[208] Homological algebra (with S. I. Gelfand.) (Paperback edition of 179). Springer Verlag, (1999), 222 pp.

[209] Frobenius manifolds, quantum cohomology, and moduli spaces. *AMS Colloquium Publications*, **47**, Providence, RI, (1999), xiii + 303 pp.

2000

[210] Classical computing, quantum computing, and Shor's factoring algorithm. *Séminaire Bourbaki, no. 862 (June 1999)*, Astérisque, **266**, (2000), 375–404.

[211] Invertible Cohomological Field Theories and Weil–Petersson volumes (with P. Zograf). *Ann. Ist. Fourier*, **50:2**, (2000), 519–535.

[212] New moduli spaces of pointed curves and pencils of flat connections (with A. Losev). *Fulton's Festschrift*, *Michigan Journ. of Math.*, **48**, (2000), 443–472.

[213] Mathematics as profession and vocation. In: *Mathematics: Frontiers and Perspectives*, ed. by V. Arnold et al., AMS, (2000), 153–159.

[214] Rational curves, elliptic curves and the Painlevé equation. Russian: In: *Studencheskie Chteniya MK NMU*, 1, McNMO, (2000), 27–36.

2001

[215] Mathematics: recent developments and cultural aspects. In: *Science and the Future of Mankind, Proceedings, Pontifical Ac. Sci., Vatican*, (2001), 89–94 and *Max Planck Research, Sci. Mag. of the Max Planck Society*, **2**, (2000), 58–59.

[216] Composition of points and Mordell–Weil problem for cubic surfaces (with D. Kanevsky). In: *Rational Points on Algebraic Varieties* ed. by E. Peyre and Yu. Tschinkel, *Progress in Mathematics*, **199**, Birkhäuser, Basel, (2001), 199–219.

[217] Theta functions, quantum tori and Heisenberg groups. *Letters in Math. Physics*, **56:3**, (2001), (special issue, Euroconference M. Flato, Part III), 295–320.

[218] Mirror symmetry and quantization of abelian varieties. In: *Moduli of Abelian Varieties*, ed. by C. Faber et al., *Progress in Math.*, **195**, Birkhäuser, (2001), 231–254.

[219] Modules and Morita theorem for operads (with M. Kapranov). *Am. J. of Math.*, **123:5**, (2001), 811–838.

[220] Moduli, Motives, Mirrors, 22 pp. In: *European Congress of Mathematicians, Barcelona 2000, vol. I. Progress in Math.*, **201**, Birkhäuser, (2001), 53–74.

[221] Holography principle and arithmetic of algebraic curves (with M. Marcolli). *Adv. Theor. Math. Phys.*, **5**, (2001), 617–650.

[222] A space of freedom Interview (in Russian). *Computerra*, 2[379], Jan. 23 (2001), 26–28.

2002

[223] Continued fractions, modular symbols, and non-commutative geometry (with M. Marcolli). *Selecta math., new ser. 8*, (2002), 475–521.

[224] Triangle of Thoughts (review of the book by A. Connes, A. Lichnerowicz, M. P. Schützenberger). *Notices of the AMS*, (March 2002), 325–327.

[225] Frobenius manifolds, quantum cohomology, and moduli spaces (Russian translation of 208). Factorial Press, Moscow, 343 pp.

2003

[226] Methods of homological algebra. Second edition. (with S. Gelfand). *Springer Monographs in Mathematics*. Springer-Verlag, Berlin, (2003). xx + 372 pp.

2004

[227] Unfoldings of meromorphic connections and a construction of Frobenius manifolds (with C. Hertling). In: *Frobenius Manifolds*, ed. by C. Hertling and M. Marcolli, Vieweg & Sohn Verlag, Wiesbaden, (2004), 113–144.

[228] Extended modular operad (with A. Losev). In: *Frobenius Manifolds*, ed. by C. Hertling and M. Marcolli, Vieweg & Sohn Verlag, Wiesbaden, (2004), 181–211.

[229] Functional equations for quantum theta functions. *Publ. Res. Inst. Mat. Sci. Kyoto*, **40:3**, (2004), 605–624.

[230] Multiple zeta–motives and moduli spaces $\overline{M}_{0,n}$ (with A. Goncharov). *Compos. Math.*, **140:1**, (2004), 1–14.

[231] Real multiplication and noncommutative geometry. In: *The legacy of Niels Henrik Abel*, ed. by O. A. Laudal and R. Piene, Springer Verlag, Berlin, (2004), 685–727.

[232] Moduli stacks $\overline{L}_{g,S}$. *Moscow Math. Journal*, **4:1**, (2004), 181–198.

[233] (Semi)simple exercises in quantum cohomology (with A. Bayer). In: *The Fano Conference Proceedings*, ed. by A. Collino, A. Conte, and M. Marchisio, Universitá di Torino, (2004), 143–173.

[234] Georg Cantor and his heritage. In: *Algebraic Geometry: Methods, Relations, and Applications: Collected papers dedicated to the memory of Andrei Nikolaevich Tyurin. Proc. V.* A. Steklov Inst. Math., Moscow, **246**, MAIK Nauka/Interperiodica, (2004), 195–203.

[235] Non–commutative geometry and quantum theta–functions. Russian: In: *Globus, Math. Seminar Notes of the Moscow Independent University*, **1**, (2004), 91–108.

[236] Mordell–Weil problem for cubic surfaces. Russian: In: *Globus, Math. Seminar Notes of the Moscow Independent University* **1**, (2004), 134–136.

2005

[237] F–manifolds with flat structure and Dubrovin's duality. *Advances in Math. (M. Artin's Fest)*, **198**, (2005), 5–26.

[238] Introduction to modern number theory (with A. A. Panchishkin). 2nd edition, revised and expanded. *Encyclopaedia of Mathematical Sciences*, **49**. Springer-Verlag, Berlin, (2005), xv + 514 pp.

[239] Von Zahlen und Figuren. In: *Géométrie au XXe siècle. Histoire et horizons.* Ed. by J. Kouneiher, D. Flament, Ph. Nabonnand, J.-J. Szczeciniarz. Hermann, Paris, (2005), 24–44.

[240] Iterated Shimura integrals. 16 pp. *Moscow Math. Journal*, **5:4**, (2005) 869–881.

2006

[241] Iterated integrals of modular forms and noncommutative modular symbols. In: *Algebraic Geometry and Number Theory. In honor of V. Drinfeld's 50th birthday.* Ed. by V. Ginzburg et al. *Progress in Math.*, **253**. Birkhäuser, Boston, 565–597.

[242] Manifolds with multiplication on the tangent sheaf. *Rendiconti Mat. Appl., Serie VII*, **26**, (2006), 69–85.

[243] The notion of dimension in geometry and algebra. *Bull. of the American Math. Soc.*, **43:2**, (2006), 139–161.

2007

[244] Generalized operads and their inner cohomomorhisms (with D. Borisov). In: *Geometry and Dynamics of Groups and spaces (In memory of Aleksader Reznikov).* Ed. by M. Kapranov et al. *Progress in Math.*, **265**. Birkhäuser, Boston, 247–308.

[245] Mathematics as Metaphor. *Selected Essays.* American Math. Society, (2007).

[246] Lectures on modular symbols. In *Arithmetic Geometry* CMI Summer school proceedings, Ed. by H. Darmon et al. 137–152, American Math. Society, (2009).

[247] Mathematical knowledge: internal, social and cultural aspects. Introductory Chapter to vol. 2 of *Mathematics and Culture*, ed. by C. Bartocci and P. Odifreddi. 27 pp.

2008

[248] Modular shadows (with M. Marcolli). 48 pp. math.NT/0703718.

[249] Matematika kak Metafora. (Expanded Russian version of 246). *Moscow Independent University* 400 pp.

[250] Matematica e conoscenza: aspetti interni, sociali e culturali.Italian version of 248. In: *La matematica II. Problemi i teoremi. A cura di C. Bartocci e P. Odifreddi. Einaudi.*

[251] Remarks on modular symbols for Maass wave forms. 25 pp. math.NT/0803.3270.

[252] Truth as value and duty: lessons of mathematics. 11 pp. math.NT/0805.4057.

[253] Cyclotomy and analytic geometry over F_1. 28 pp. math.NT/0809.2716.

2009

[254] Quantum theta functions and Gabor frames for modulation spaces (with F. Luef). *Lett. Math. Phys.*, **88:1–3**. 37 pp., 131–161.

[255] An update on on semisimple quantum cohomology and F–manifolds (with C. Hertling, C. Teleman). *Proc. Steklov Inst. Math.*, **264**. 12 pp., 62–69.

[256] Stability Conditions, wall-crossing and weighted Gromov-Witten Invariants (with Arend Bayer). *Moscow Math Journal*, **9:1**. 12 pp., 1–30.

[257] Introduction to modern number theory (with A. Panchishkin, in Russian, revised version of [238]). *Moscow, MCNMO*, 552 pp.

[258] Renormalization and computation I: motivation and background. arxiv/0904.4921.

[259] We Do Not Choose Mathematics as Our Profession, It Chooses Us: Interview with Yuri Manin (conducted by Mikhail Gelfand). *Notices of the AMS*, 56 no.10 (2009), 1268–1274.

Gerstenhaber and Batalin–Vilkovisky Structures on Lagrangian Intersections

Kai Behrend[1] and Barbara Fantechi[2]

[1] University of British Columbia, Vancouver, Canada. behrend@math.ubc.ca
[2] SISSA, Trieste, Italy. fantechi@sissa.it

To Yuri Ivanovich Manin on the occasion of his 70th birthday

Summary. Let M and N be Lagrangian submanifolds of a complex symplectic manifold S. We construct a Gerstenhaber algebra structure on $\mathcal{T}or_*^{\mathcal{O}_S}(\mathcal{O}_M, \mathcal{O}_N)$ and a compatible Batalin–Vilkovisky module structure on $\mathcal{E}xt^*_{\mathcal{O}_S}(\mathcal{O}_M, \mathcal{O}_N)$. This gives rise to a de Rham type cohomology theory for Lagrangian intersections.

Key words: symplectic manifolds, Lagrangian intersections, categorification, Batalin–Vilkovisky

2000 Mathematics Subject Classifications: 14C17, 16E45, 32G13, 53D12

Introduction

We are interested in intersections of Lagrangian submanifolds of holomorphic symplectic manifolds. Thus we work over the complex numbers in the analytic category.

There are two main aspects of this paper we would like to explain in the introduction: categorification of intersection numbers, and Gerstenhaber and Batalin–Vilkovisky structures on Lagrangian intersections.

Categorification of Lagrangian intersection numbers

This paper grew out of an attempt to categorify Lagrangian intersection numbers. We will explain what we mean by this, and how we propose a solution to the problem. Our construction looks very promising, but is still conjectural.

Lagrangian intersection numbers: smooth case

Let S be a (complex) symplectic manifold and L, M Lagrangian submanifolds. Since L and M are half-dimensional, the expected dimension of their intersection is zero. Intersection theory therefore gives us the intersection number

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*,
Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_1,

$$\#(L \cap M)$$

if the intersection is compact. In the general case, we get a class

$$[L \cap M]^{\mathrm{vir}} \in A_0(L \cap M)$$

in degree-zero Borel–Moore homology such that in the compact case,

$$\#(L \cap M) = \deg[L \cap M]^{\mathrm{vir}} .$$

If the intersection $X = L \cap M$ is smooth, then

$$[X]^{\mathrm{vir}} = c_{\mathrm{top}}(E) \cap [X] ,$$

where E is the excess bundle of the intersection, which fits into the exact sequence

$$0 \longrightarrow T_X \longrightarrow T_L|_X \oplus T_M|_X \longrightarrow T_S|_X \longrightarrow E \longrightarrow 0$$

of vector bundles on X. The symplectic form σ defines an isomorphism $T_S|_X = \Omega_S|_X$. Under this isomorphism, the subbundle $T_L|_X$ corresponds to the conormal bundle $N^\vee_{L/S}$. Thus we can rewrite our exact sequence as

$$0 \longrightarrow E^\vee \longrightarrow N^\vee_{L/S} \oplus N^\vee_{M/S} \longrightarrow \Omega_S|_X \longrightarrow \Omega_X \longrightarrow 0 ,$$

which shows that the excess bundle E is equal to the cotangent bundle Ω_X. Thus, in the smooth case,

$$[X]^{\mathrm{vir}} = c_{\mathrm{top}}(E) \cap [X] = c_{\mathrm{top}}(\Omega_X) \cap [X] = (-1)^n c_{\mathrm{top}}(T_X) \cap [X] ,$$

and in the smooth and compact case,

$$\#(L \cap M) = \deg[X]^{\mathrm{vir}} = (-1)^n \int_X c_{\mathrm{top}}(T_X) = (-1)^n \chi(X) ,$$

where $2n$ is the dimension of S and $\chi(X)$ is the topological Euler characteristic of X. This shows that we can make sense of the intersection number even if the intersection is not compact: define the intersection number to be the signed Euler characteristic.

Intersection numbers: singular case

In [1], it was shown how to make sense of the statement that Lagrangian intersection numbers are signed Euler characteristics in the case that the intersection X is singular. An integer invariant $\nu_X(P) \in \mathbb{Z}$ of the singularity of the analytic space X at the point $P \in X$ was introduced.

In the case of a Lagrangian intersection $X = L \cap M$, the number $\nu_X(P)$ can be described as follows. Locally around P, we can assume that S is equal to the cotangent bundle of M and $M \subset S$ is the zero section. Moreover, we can assume that L is the graph of a closed, even exact, 1-form ω on M. If $\omega = df$, for a holomorphic function $f : M \to \mathbb{C}$, defined near P, then

$$\nu_X(P) = (-1)^n \big(1 - \chi(F_P)\big)\,, \tag{1}$$

where $n = \dim M$ and F_P is the Milnor fiber of f at P.

The main theorem of [1] implies that if L and M are Lagrangian submanifolds of the symplectic manifold S, with compact intersection X, then

$$\#X = \deg[X]^{\mathrm{vir}} = \chi(X, \nu_X)\,,$$

the weighted Euler characteristic of X with respect to the constructible function ν_X, which is defined as

$$\chi(X, \nu_X) = \sum_{i \in \mathbb{Z}} i \cdot \chi(\{\nu_X = i\})\,.$$

In particular, arbitrary Lagrangian intersection numbers are always well-defined: the intersection need not be smooth or compact. The integer $\nu_X(P)$ may be considered as the contribution of the point P to the intersection $X = L \cap M$.

Categorifying intersection numbers: smooth case

To categorify the intersection number means to construct a cohomology theory such that the intersection number is equal to the alternating sum of Betti numbers. If X is smooth (not necessarily compact), a natural candidate is (shifted) holomorphic de Rham cohomology

$$\#X = (-1)^n \chi(X) = \sum (-1)^{i-n} \dim_{\mathbb{C}} \mathbb{H}^i\big(X, (\Omega_X^\bullet, d)\big)\,.$$

Here $(\Omega_X^\bullet, d)$ is the holomorphic de Rham complex of X and $\mathbb{H}^i$ its hypercohomology. Of course, by the holomorphic Poincaré lemma, hypercohomology reduces to cohomology.

Categorification: compact case

If the intersection $X = L \cap M$ is compact, but not necessarily smooth, we have

$$\begin{aligned}\#X &= \sum_i (-1)^{i-n} \dim_{\mathbb{C}} \mathrm{Ext}^i_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M) \\ &= \sum_{i,j} (-1)^i (-1)^{j-n} \dim_{\mathbb{C}} H^i\big(X, \mathcal{E}xt^j_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)\big)\,.\end{aligned}$$

For X smooth, $\mathcal{E}xt^j_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M) = \Omega^j_X$, so this reduces to Hodge cohomology

$$\#X = \sum_{i,j} (-1)^i (-1)^{j-n} \dim_{\mathbb{C}} H^i\left(X, \Omega^j_X\right).$$

This justifies using the sheaves $\mathcal{E}xt^j_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$ as replacements for the sheaves Ω^j_X if X is no longer smooth. To get finite-dimensional cohomology groups, we will construct de Rham type differentials

$$d : \mathcal{E}xt^j_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M) \longrightarrow \mathcal{E}xt^{j+1}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M),$$

so that the hypercohomology groups

$$\mathbb{H}^i\big(X, \big(\mathcal{E}xt^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M), d\big)\big)$$

are finite-dimensional, even if X is not compact. Returning to the compact case, for any such d, we necessarily have

$$\#X = \sum_{i} (-1)^{i-n} \dim_{\mathbb{C}} \mathbb{H}^i\big(X, \big(\mathcal{E}xt^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M), d\big)\big).$$

Categorification: local case

Every symplectic manifold S is locally isomorphic to the cotangent bundle Ω_N of a manifold N. The fibers of the induced vector bundle structure on S are Lagrangian submanifolds, and thus we have defined (locally on S) a foliation by Lagrangian submanifolds, i.e., a *Lagrangian foliation*. (Lagrangian foliations are also called *polarizations*.) We may assume that the leaves of our Lagrangian foliation of S are transverse to the two Lagrangians L and M whose intersection we wish to study. Then L and M turn into the graphs of 1-forms on N. The Lagrangian condition implies that these 1-forms on N are closed. Without loss of generality, we may assume that one of these 1-forms is the zero section of Ω_N and hence identify M with N. By making $M = N$ smaller if necessary, we may assume that the closed 1-form defined by L is exact. Then L is the graph of the 1-form df, for a holomorphic function f on M. Thus the intersection $L \cap M$ is now the zero locus of the 1-form df:

$$X = Z(df).$$

This is the *local case.*

Multiplying by df defines a differential

$$\begin{aligned} s : \Omega^j_M &\longrightarrow \Omega^{j+1}_M, \\ \omega &\longmapsto df \wedge \omega. \end{aligned}$$

Because df is closed, the differential s commutes with the de Rham differential $d : \Omega^j_M \to \Omega^{j+1}_M$. Thus the de Rham differential passes to cohomology with respect to s:

$$d : h^j\left(\Omega^\bullet_M, s\right) \longrightarrow h^{j+1}\left(\Omega^\bullet_M, s\right),$$

where h^j denotes the cohomology sheaves, which are coherent sheaves of $\mathcal{O}_X$-modules. Let us denote these cohomology sheaves by

$$\mathcal{E}^j = h^j\left(\Omega^\bullet_M, s\right).$$

We have thus defined a complex of sheaves on X,

$$(\mathcal{E}^\bullet, d), \tag{2}$$

where the $\mathcal{E}^i$ are coherent sheaves of $\mathcal{O}_X$-modules, and the differential d is $\mathbb{C}$-linear. It is a theorem of Kapranov [2] that the cohomology sheaves $h^i(\mathcal{E}^\bullet, d)$ are constructible sheaves on X and thus have finite-dimensional cohomology groups. It follows that the hypercohomology groups

$$\mathbb{H}^i\big(X, (\mathcal{E}^\bullet, d)\big)$$

are finite-dimensional as well.

We conjecture that the constructible function

$$P \mapsto \sum_i (-1)^{i-n} \dim_{\mathbb{C}} \mathbb{H}^i_{\{P\}}\big(X, (\mathcal{E}, d)\big),$$

of fiberwise Euler characteristic of $(\mathcal{E}, d)$ is equal to the function ν_X from (1) above. This would achieve the categorification in the local case. In particular, for the noncompact intersection numbers we would have

$$\chi(X, \nu_X) = \sum_i (-1)^{i-n} \dim_{\mathbb{C}} \mathbb{H}^i\big(X, (\mathcal{E}, d)\big).$$

We remark that if f is a homogeneous polynomial (in a suitable set of coordinates), then this conjecture is true.

To make the connection with the compact case (and because this construction is of central importance to the paper), let us explain why

$$\mathcal{E}^i = \mathcal{E}xt^i_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M).$$

Denote the projection $S = \Omega_M \to M$ by π. The 1-form on Ω_M that corresponds to the vector field generating the natural $\mathbb{C}^*$-action on the fibers we shall call α. Then $d\alpha = \sigma$ is the symplectic form on S. We consider the 1-form $s = \alpha - \pi^* df$ on S. Its zero locus in S is equal to the graph of df. Let us denote the subbundle of Ω_S annihilating vector fields tangent to the fibers of π by E. Then $s \in \Omega_S$ is a section of E and we obtain a resolution of the structure sheaf of $\mathcal{O}_L$ over $\mathcal{O}_S$:

$$\cdots \longrightarrow \Lambda^2 E^\vee \xrightarrow{\widetilde{s}} E^\vee \xrightarrow{\widetilde{s}} \mathcal{O}_S,$$

where $\widetilde{s}$ denotes the derivation of the differential graded $\mathcal{O}_S$-algebra $\Lambda^\bullet E^\vee$ given by contraction with s. Taking duals and tensoring with $\mathcal{O}_M$, we obtain a complex of vector bundles $(\Lambda E|_M, s|_M)$ that computes $\mathcal{E}xt^i_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$. One checks that $(\Lambda E|_M, s|_M) = (\Omega_M, s)$.

Categorification: global case

We now come to the contents of this paper. let S be a symplectic manifold and L, M Lagrangian submanifolds with intersection X. Let us use the abbreviation $\mathcal{E}^i = \mathcal{E}xt^i_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$. The $\mathcal{E}^i$ are coherent sheaves of $\mathcal{O}_X$-modules. The main theorem of this paper is that the locally defined de Rham differentials (2) do not depend on the way we write S as a cotangent bundle, or in other words, that d is independent of the chosen polarization of S. Thus, the locally defined d glue, and we obtain a globally defined canonical de Rham type differential

$$d : \mathcal{E}^i \to \mathcal{E}^{i+1}.$$

In the case that X is smooth, $\mathcal{E}^i = \Omega^i_X$, and d is the usual de Rham differential. We may call $(\mathcal{E}^\bullet, d)$ the *virtual de Rham complex* of the Lagrangian intersection X. Conjecturally, $(\mathcal{E}, d)$ categorifies Lagrangian intersection numbers in the sense that for the local contribution of the point $P \in X$ to the Lagrangian intersection we have

$$\nu_X(P) = \sum_i (-1)^{i-n} \dim_{\mathbb{C}} \mathbb{H}^i_{\{P\}}\big(X, (\mathcal{E}, d)\big).$$

Hence, for the noncompact intersection numbers we should have

$$\chi(X, \nu_X) = \sum_i (-1)^{i-n} \dim_{\mathbb{C}} \mathbb{H}^i\big(X, (\mathcal{E}, d)\big).$$

In particular, if the intersection is compact, $\#X = \chi(X, \nu_X)$ should be the alternating sum of the Betti numbers of the hypercohomology groups of the virtual de Rham complex.

Donaldson–Thomas invariants

Our original motivation for this research was a better understanding of Donaldson–Thomas invariants. It is to be hoped that the moduli spaces giving rise to Donaldson–Thomas invariants (spaces of stable sheaves of fixed determinant on Calabi–Yau threefolds) are Lagrangian intersections, at least locally. We have two reasons for believing this: First of all, the obstruction theory giving rise to the virtual fundamental class is *symmetric*, a property shared by the obstruction theories of Lagrangian intersections. Secondly, at least heuristically, these moduli spaces are equal to the critical set of the holomorphic Chern–Simons functional.

Our “exchange property” should be useful for gluing virtual de Rham complexes if the moduli spaces are only local Lagrangian intersections.

In this way we hope to construct a virtual de Rham complex on the Donaldson–Thomas moduli spaces and thus categorify Donaldson–Thomas invariants.

Gerstenhaber and Batalin–Vilkovisky structures on Lagrangian intersections

The virtual de Rham complex $(\mathcal{E}^\bullet, d)$ is just half of the story. There is also the graded sheaf of $\mathcal{O}_X$-algebras $\mathcal{A}^\bullet$ given by

$$\mathcal{A}^i = \mathcal{T}or_{-i}^{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)\,.$$

Locally, $\mathcal{A}^\bullet$ is given as the cohomology of $(\Lambda T_M, \widetilde{s})$, in the above notation. The Lie–Schouten–Nijenhuis bracket induces a $\mathbb{C}$-linear bracket operation

$$[\,,]: \mathcal{A}^\bullet \otimes_{\mathbb{C}} \mathcal{A}^\bullet \longrightarrow \mathcal{A}^\bullet$$

of degree $+1$. We show that these locally defined brackets glue to give a globally defined bracket making $(\mathcal{A}^\bullet, \wedge [\,,])$ a sheaf of Gerstenhaber algebras.

Then $\mathcal{E}^\bullet$ is a sheaf of modules over $\mathcal{A}^\bullet$. (The module structure is induced by contraction.) The bracket on $\mathcal{A}^\bullet$ and the differential on $\mathcal{E}^\bullet$ satisfy a compatibility condition; see (5). We say that $(\mathcal{E}, d)$ is a *Batalin–Vilkovisky module* over the Gerstenhaber algebra $(\mathcal{A}, \wedge[\,,])$. (This structure has been called a *calculus* by Tamarkin and Tsygan in [4].)

In the case that L and M are oriented submanifolds, i.e., the highest exterior powers of the normal bundles have been trivialized, we have an identification

$$\mathcal{A}^i = \mathcal{E}^{n+i}\,.$$

Transporting the differential from $\mathcal{E}^\bullet$ to $\mathcal{A}^\bullet$ via this identification turns $(\mathcal{A}, \wedge[\,,], d)$ into a Batalin–Vilkovisky algebra.

To prove these facts we have to study *differential* Gerstenhaber algebras and *differential* Batalin–Vilkovisky modules over them. We will prove that locally defined differential Gerstenhaber algebras and their differential Batalin–Vilkovisky modules are quasi-isomorphic, making their cohomologies isomorphic and hence yielding the well-definedness of the bracket and the differential.

First order truncation

In this paper we are interested only in the Gerstenhaber and Batalin–Vilkovisky structures on $\mathcal{A}$ and $\mathcal{E}$. In other words, we deal only with the structures induced on cohomology. This amounts to a truncation of the full derived Lagrangian intersection. Because of our modest goal, we need to study differential Gerstenhaber and Batalin–Vilkovisky structures only *up to first order*. In future research, we hope to address the complete derived structure on Lagrangian intersections.

This would certainly involve studying the Witten deformation of the de Rham complex in more detail. Related work along these lines has been done by Kashiwara and Schapira [3].

Overview

1. Algebra

In this introductory section, we discuss algebraic preliminaries. We review the definitions of *differential Gerstenhaber algebra* and *differential Batalin–Vilkovisky module*. This is mainly to fix our notation. There are quite a few definitions to keep track of; we apologize for the lengthiness of this section.

2. Symplectic geometry

Here we review a few basic facts about complex symplectic manifolds. In particular, the notions of *Lagrangian foliation*, *polarization*, and the *canonical partial connection* are introduced.

3. Derived Lagrangian intersections on polarized symplectic manifolds

On a *polarized* symplectic manifold, we define *derived intersections* of Lagrangian submanifolds. These are (sheaves of) Gerstenhaber algebras on the scheme-theoretic intersection of two Lagrangian submanifolds. The main theorem we prove about these derived intersections is a certain invariance property with respect to *symplectic correspondences.* We call it the *exchange property.*

We repeat this program for *derived homs* (the Batalin–Vilkovisky case), and *oriented* derived intersections (the oriented Batalin–Vilkovisky case).

4. The Gerstenhaber structure on $\mathcal{T}or$ and the Batalin–Vilkovisky structure on $\mathcal{E}xt$

In this section we use the exchange property to prove that after passing to cohomology, we no longer notice the polarization. The Gerstenhaber and Batalin–Vilkovisky structures are independent of the polarization chosen to define them.

This section closes with an example of a symplectic correspondence and the corresponding exchange property.

5. Further remarks

In this final section we define *virtual de Rham cohomology* of Lagrangian intersections. We speculate on what *virtual Hodge theory* might look like. We introduce a natural differential graded category associated to a complex symplectic manifold. (It looks like a kind of holomorphic, de Rham type analogue of the Fukaya category.) Finally, we mention the conjectures connecting the virtual de Rham complex to the perverse sheaf of vanishing cycles.

1 Algebra

Let M be a manifold. Regular functions, elements of $\mathcal{O}_M$, have degree 0. By ΛT_M we mean the graded sheaf of polyvector fields on M. We think of it as a sheaf of graded $\mathcal{O}_M$-algebras (the product being $\wedge$), concentrated in *nonpositive* degrees, the vector fields having degree -1. By $\Omega^\bullet_M$ we denote the graded sheaf of differential forms on M. This we think of as a sheaf of graded $\mathcal{O}_M$-modules, concentrated in *nonnegative* degrees, with 1-forms having degree $+1$. We will denote the natural pairing of T_M with Ω_M by $X \lrcorner\, \omega \in \mathcal{O}_S$, for $X \in T_M$ and $\omega \in \Omega_M$. The following is, of course, well known:

Lemma 1.1. *There exists a unique extension of $\lrcorner$ to an action of the sheaf of graded $\mathcal{O}_M$-algebras ΛT_M on the sheaf of graded $\mathcal{O}_S$-modules $\Omega^\bullet_M$ that satisfies*

(i) $f \lrcorner\, \omega = f\omega$, *for $f \in \mathcal{O}_S$ and $\omega \in \Omega^\bullet_M$ (linearity over $\mathcal{O}_M$),*

(ii) $X \lrcorner\,(\omega_1 \wedge \omega_2) = (X \lrcorner\, \omega_1) \wedge \omega_2 + (-1)^{\overline{\omega}_1}\omega_1 \wedge (X \lrcorner\, \omega_2)$, *for $X \in T_M$ and $\omega_1, \omega_2 \in \Omega^\bullet_M$, (the degree -1 part acts by derivations),*

(iii) $(X \wedge Y) \lrcorner\, \omega = X \lrcorner\,(Y \lrcorner\, \omega)$, *for $X, Y \in \Lambda T_M$, $\omega \in \Omega^\bullet_M$ (action property).*

Now turn things around and note that any section $s \in \Omega_M$ defines a derivation of degree $+1$ on ΛT_M, which we shall denote by $\widetilde{s}$. It is the unique derivation that extends the map $T_M \to \mathcal{O}_M$ given by $\widetilde{s}(X) = X \lrcorner\, s$, for all $X \in T_M$. (Note that this is not a violation of the universal sign convention; see Remark 1.3.)

Lemma 1.2. *The pair $(\Lambda T_M, \widetilde{s})$ is a sheaf of differential graded $\mathcal{O}_M$-algebras. Left multiplication by s defines a differential on $\Omega^\bullet_M$, and the pair $(\Omega^\bullet_M, s)$ is a sheaf of differential graded modules over $(\Lambda T_M, \widetilde{s})$.*

Proof. This amounts to the formula

$$s \wedge (X \lrcorner\, \omega) = \widetilde{s}(X) \lrcorner\, \omega + (-1)^{\overline{X}} X \lrcorner\,(s \wedge \omega) \tag{3}$$

for all $\omega \in \Omega^\bullet_M$ and $X \in \Lambda T_M$. □

Remark 1.3. Set $\langle X, \omega\rangle$ equal to the degree zero part of $X \lrcorner\, \omega$. This is a perfect pairing $\Lambda T_M \otimes_{\mathcal{O}_M} \Omega^\bullet_M \to \mathcal{O}_M$, expressing the fact that $\Omega^\bullet_M$ is the $\mathcal{O}_M$-dual of ΛT_M. According to formula (3), we have, if $\deg X + \deg \omega + 1 = 0$,

$$\langle \widetilde{s}(X), \omega\rangle + (-1)^{\overline{X}}\langle X, s \wedge \omega\rangle = 0\,.$$

This means that the derivation $\widetilde{s}$ and left multiplication by s are $\mathcal{O}_S$-duals of one another. To explain the signs, note that we think of $\widetilde{s}$ and s as differentials on the graded sheaves ΛT_M and $\Omega^\bullet_M$, and for differentials of degree $+1$ the sign convention is

$$0 = D\langle X, \omega\rangle = \langle DX, \omega\rangle + (-1)^{\overline{X}}\langle X, D\omega\rangle\,.$$

In particular, for $\deg X = 1$ and $\omega = 1$ we get $\widetilde{s}(X) = \langle X, s\rangle = X \lrcorner\, s$.

Remark 1.4. We can summarize formula (3) more succinctly as

$$[s, i_X] = i_{\widetilde{s}(X)}\,,$$

where $i_X : \Omega^\bullet_M \to \Omega^\bullet_M$ denotes the endomorphism $\omega \mapsto X \lrcorner\, \omega$.

1.1 Differential Gerstenhaber algebras

Let S be a manifold and A a graded sheaf of $\mathcal{O}_S$-modules.

Definition 1.5. A **bracket** on A of degree $+1$ is a homomorphism

$$[,] : A \otimes_{\mathbb{C}} A \longrightarrow A$$

of degree $+1$ satisfying:

(i) $[,]$ is a graded $\mathbb{C}$-linear derivation in each of its two arguments,

(ii) $[,]$ is graded commutative (not anticommutative).

If $[,]$ satisfies, in addition, the Jacobi identity, we shall call $[,]$ a **Lie bracket**.

The sign convention for brackets of degree $+1$ is that the comma is treated as carrying the degree $+1$, the opening and closing bracket as having degree 0. Thus, when passing an odd element past the comma, the sign changes. For example, the graded commutativity reads

$$[Y, X] = (-1)^{\overline{XY}+\overline{X}+\overline{Y}}[X, Y]\,.$$

Definition 1.6. A **Gerstenhaber algebra** over $\mathcal{O}_S$ is a sheaf of graded $\mathcal{O}_S$-modules A, concentrated in nonpositive degrees, endowed with

(i) a commutative (associative, of course) product $\wedge$ of degree 0 with unit, making A a sheaf of graded $\mathcal{O}_S$-algebras,

(ii) a Lie bracket $[,]$ of degree $+1$ (see Definition 1.5).

In our cases, the underlying $\mathcal{O}_S$-module of A will always be coherent and $\mathcal{O}_S \to A^0$ will be a surjection of coherent $\mathcal{O}_S$-algebras. The main example is the following:

Example 1.7. Let $M \subset S$ be a submanifold and $A = \Lambda_{\mathcal{O}_M} T_M$ the polyvector fields on M. The bracket is the Schouten–Nijenhuis bracket.

Definition 1.8. A **differential Gerstenhaber algebra** is a Gerstenhaber algebra A over $\mathcal{O}_S$ endowed with an additional $\mathbb{C}$-linear map $\widetilde{s} : A \to A$ of degree $+1$ that satisfies

(i) $[\widetilde{s}, \widetilde{s}] = \widetilde{s}^2 = 0$;

(ii) $\widetilde{s}$ is a derivation with respect to $\wedge$; in particular, it is $\mathcal{O}_S$-linear;

(iii) $\widetilde{s}$ is a derivation with respect to $[,]$.

Thus, neglecting the bracket, a differential Gerstenhaber algebra is a sheaf of differential graded algebras over $\mathcal{O}_S$.

Lemma 1.9. *Let $(A, \widetilde{s})$ be a differential Gerstenhaber algebra. Let $I \subset A^0$ be the image of $\widetilde{s} : A^{-1} \to A^0$. This is a sheaf of ideals in A^0. Then the cohomology $h^*(A, \widetilde{s})$ is a Gerstenhaber algebra with $h^0(A, \widetilde{s}) = A^0/I$.*

Proof. This is clear: the fact that $\widetilde{s}$ is a derivation with respect to both products on A implies that the two products pass to $h^*(A, \widetilde{s})$. Then all the properties of the products pass to cohomology. □

Example 1.10. Let $M \subset S$ and $A = \Lambda T_M$ be as in Example 1.7. In addition, let $s \in \Omega_M$ be a *closed* 1-form. Then $(\Lambda T_M, \widetilde{s})$ with $\wedge$ and Schouten–Nijenhuis bracket $[,]$ is a differential Gerstenhaber algebra. The closedness of s makes $\widetilde{s}$ a derivation with respect to $[,]$.

1.2 Morphisms of differential Gerstenhaber algebras

Definition 1.11. Let A and B be Gerstenhaber algebras over $\mathcal{O}_S$. A **morphism** of Gerstenhaber algebras is a homomorphism $\phi : A \to B$ of graded $\mathcal{O}_S$-modules (of degree zero) that is compatible with both $\wedge$ and $[,]$:

(i) $\phi(X \wedge Y) = \phi(X) \wedge \phi(Y)$,

(ii) $\phi([X, Y]) = [\phi(X), \phi(Y)]$.

Definition 1.12. Let $(A, \widetilde{s})$ and $(B, \widetilde{t})$ be differential Gerstenhaber algebras over $\mathcal{O}_S$. A **(first-order) morphism** of differential Gerstenhaber algebras is a pair $(\phi, \{,\})$, where $\phi : A \to B$ is a degree-zero homomorphism of graded $\mathcal{O}_S$-modules, and $\{,\} : A \otimes_{\mathbb{C}} A \to B$ is a degree-zero $\mathbb{C}$-bilinear map such that

(i) $\phi(X \wedge Y) = \phi(X) \wedge \phi(Y)$ and $\phi(\widetilde{s}X) = \widetilde{t}\phi(X)$, so that $\phi : A \to B$ is a morphism of differential graded $\mathcal{O}_S$-algebras;

(ii) $\{,\}$ is symmetric, i.e., $\{Y, X\} = (-1)^{\overline{XY}}\{X, Y\}$;

(iii) $\{,\}$ is a $\mathbb{C}$-linear derivation with respect to $\wedge$ in each of its arguments, where the A-module structure on B is given by ϕ, in other words,

$$\{X \wedge Y, Z\} = \phi(X) \wedge \{Y, Z\} + (-1)^{\overline{XY}}\phi(Y) \wedge \{X, Z\}$$

and

$$\{X, Y \wedge Z\} = \{X, Y\} \wedge \phi(Z) + (-1)^{\overline{YZ}}\{X, Z\} \wedge \phi(Y);$$

(iv) the failure of ϕ to commute with $[\,]$ is equal to the failure of the $\mathcal{O}_S$-linear differentials to behave as derivations with respect to $\{\,\}$,

$$\phi[X, Y] - [\phi(X), \phi(Y)] = (-1)^{\overline{X}}\widetilde{t}\{X, Y\} - (-1)^{\overline{X}}\{\widetilde{s}X, Y\} - \{X, \widetilde{s}Y\}\,. \quad (4)$$

Remark 1.13. We will always omit the qualifier "first order," since we will not consider any "higher-order" morphisms in this paper. This is because, in the end, we are interested only in the cohomology of our differential Gerstenhaber algebras. To keep track of the induced structure on cohomology, first-order morphisms suffice. We hope to return to "higher-order" questions in future research.

Remark 1.14. Suppose all conditions in Definition 1.12 except the last are satisfied. Then both sides of the equation in condition (iv) are symmetric of degree-one and $\mathbb{C}$-linear derivations with respect to $\wedge$ in each of the two arguments. Thus, to check condition (iv), it suffices to check on $\mathbb{C}$-algebra generators for A.

Lemma 1.15. *A morphism of differential Gerstenhaber algebras*

$$(\phi, \{\,\}) : (A, \widetilde{s}) \longrightarrow (B, \widetilde{t})$$

induces a morphism of Gerstenhaber algebras on cohomology. In other words,

$$h^*(\phi) : h^*(A, \widetilde{s}) \longrightarrow h^*(B, \widetilde{t})$$

respects both $\wedge$ and $[\,,]$.

Proof. Any morphism of differential graded $\mathcal{O}_S$-algebras induces a morphism of graded algebras upon passing to cohomology. Thus $h^*(\phi)$ respects $\wedge$. The fact that $h^*(\phi)$ respects the Lie brackets follows from property (iv) of Definition 1.12. All three terms on the right-hand side of said equation vanish in cohomology. □

Definition 1.16. A **quasi-isomorphism** of differential Gerstenhaber algebras is a morphism of differential Gerstenhaber algebras that induces an isomorphism of Gerstenhaber algebras on cohomology.

1.3 Differential Batalin–Vilkovisky modules

Definition 1.17. Let A be a Gerstenhaber algebra. A sheaf of graded $\mathcal{O}_S$-modules L with an action $\lrcorner$ of A making L a graded A-module is called a **Batalin–Vilkovisky module** over A if it is endowed with a $\mathbb{C}$-linear map $d : L \to L$ of degree $+1$ satisfying

(i) $[d, d] = d^2 = 0$;

(ii) For all $X, Y \in A$ and every $\omega \in L$ we have

$$\begin{aligned} d(X \wedge Y \lrcorner \omega) + (-1)^{\overline{X}+\overline{Y}} X \wedge Y \lrcorner d\omega + (-1)^{\overline{X}}[X, Y] \lrcorner \omega \\ = (-1)^{\overline{X}} X \lrcorner d(Y \lrcorner \omega) + (-1)^{\overline{XY}+\overline{Y}} Y \lrcorner d(X \lrcorner \omega)\,. \end{aligned} \tag{5}$$

Remark 1.18. Write i_X for the endomorphism $\omega \mapsto X \lrcorner \omega$ of L. Then formula (5) can be rewritten as

$$[X,Y] \lrcorner \omega = \big[[i_X, d], i_Y\big](\omega)$$

or simply

$$i_{[X,Y]} = \big[[i_X, d], i_Y\big]\,. \tag{6}$$

Note also that $\big[[i_X, d], i_Y\big] = \big[i_X, [d, i_Y]\big]$.

The action property $(X \wedge Y) \lrcorner \omega = X \lrcorner (Y \lrcorner \omega)$ translates into $i_{X \wedge Y} = i_X \circ i_Y$.

In our applications, Batalin–Vilkovisky modules will always be coherent over $\mathcal{O}_S$. Note that there is no multiplicative structure on L, so there is no requirement for the differential d to be a derivation.

Example 1.19. Let $M \subset S$ and $A = \Lambda T_M$ be the Gerstenhaber algebra of polyvector fields on M, as in Example 1.7. Then $\Omega_M^\bullet$ with exterior differentiation d is a Batalin–Vilkovisky module over ΛT_M.

Definition 1.20. A **differential Batalin–Vilkovisky module** over the differential Gerstenhaber algebra $(A, \widetilde{s})$ is a Batalin–Vilkovisky module L for the underlying Gerstenhaber algebra A, endowed with an additional $\mathbb{C}$-linear map $s : L \to L$ of degree $+1$ satisfying:

(i) $[s, s] = s^2 = 0$;

(ii) (M, s) is a differential graded module over the differential graded algebra $(A, \widetilde{s})$, i.e., we have

$$s(X \lrcorner \omega) = \widetilde{s}(X) \lrcorner \omega + (-1)^{\overline{X}} X \lrcorner s(\omega)$$

for all $X \in A$, $\omega \in L$. More succinctly: $[s, i_X] = i_{\widetilde{s}(X)}$;

(iii) $[d, s] = 0$.

Note that the differential s is necessarily $\mathcal{O}_S$-linear. This distinguishes it from d.

Lemma 1.21. *Let (L, s) be a differential Batalin–Vilkovisky module over the differential Gerstenhaber algebra $(A, \widetilde{s})$. Then $h^*(L, s)$ is a Batalin–Vilkovisky module for the Gerstenhaber algebra $h^*(A, \widetilde{s})$.*

Proof. First, $h^*(M, s)$ is a graded $h^*(A, \widetilde{s})$-module. The condition $[d, s] = 0$ implies that d passes to cohomology. Then the properties of d pass to cohomology as well. □

Example 1.22. Let $M \subset S$ be a submanifold and $s \in \Omega_M$ a closed 1-form. Then $(\Omega_M^\bullet, s)$ (see Lemma 1.2) is a differential Batalin–Vilkovisky module over the differential Gerstenhaber algebra $(\Lambda T_M, \widetilde{s})$ of Example 1.10.

1.4 Homomorphisms of differential Batalin–Vilkovisky modules

Definition 1.23. Let A and B be Gerstenhaber algebras and $\phi : A \to B$ a morphism of Gerstenhaber algebras. Let L be a Batalin–Vilkovisky module over A and M a Batalin–Vilkovisky module over B. A **homomorphism** of Batalin–Vilkovisky modules of degree n (covering ϕ) is a degree n homomorphism of graded A-modules $\psi : L \to M$ (where the A-module structure on M is defined via ϕ), which commutes with d:

(i) $\psi(X \lrcorner \omega) = (-1)^{n\overline{X}} \phi(X) \lrcorner \psi(\omega)$,

(ii) $\psi d_L(\omega) = (-1)^n d_M \psi(\omega)$.

We write the latter condition as $[\psi, d] = 0$.

Definition 1.24. Let $(A, \widetilde{s})$ and $(B, \widetilde{t})$ be differential Gerstenhaber algebras and $(\phi, \{,\}) : (A, \widetilde{s}) \to (B, \widetilde{t})$ a morphism of differential Gerstenhaber algebras. Let (L, s) be a differential Batalin–Vilkovisky module over $(A, \widetilde{s})$ and (M, t) a differential Batalin–Vilkovisky module over $(B, \widetilde{t})$. A **(first-order) homomorphism** of differential Batalin–Vilkovisky modules of degree n covering $(\phi, \{,\})$ is a pair (ψ, δ), where $\psi : (L, s) \to (M, t)$ is a degree n homomorphism of differential graded $(A, \widetilde{s})$-modules, where the $(A, \widetilde{s})$-module structure on (M, t) is through ϕ. Moreover, $\delta : L \to M$ is a $\mathbb{C}$-linear map, also of degree n, satisfying

(i) the commutator property

$$\psi \circ d - (-1)^n d \circ \psi = -2(-1)^n t \circ \delta + 2\delta \circ s \,, \tag{7}$$

(ii) compatibility with the bracket $\{,\}$ property

$$\begin{aligned}\delta(X \wedge Y \lrcorner \omega) + (-1)^{n(\overline{X}+\overline{Y})} \phi(X) \wedge \phi(Y) \lrcorner \delta\omega + (-1)^{n(\overline{X}+\overline{Y})} \{X, Y\} \lrcorner \psi(\omega) \\ = (-1)^{n\overline{X}} \phi(X) \lrcorner \delta(Y \lrcorner \omega) + (-1)^{\overline{XY}+n\overline{Y}} \phi(Y) \lrcorner \delta(X \lrcorner \omega) \,.\end{aligned} \tag{8}$$

Remark 1.25. The same comments as those in Remark 1.13 apply.

Remark 1.26. If we use the same letter s to denote the $\mathcal{O}_S$-linear differentials on L and M, we can rewrite the commutator conditions of Definition 1.24 more succinctly as

$$[\psi, s] = 0, \qquad [\psi, d] - 2[\delta, s] = 0 \,.$$

The compatibility with the bracket property can be rewritten as

$$\big[\iota_X, [\iota_Y, \delta]\big] = \iota_{\{X,Y\}} \circ \psi \,. \tag{9}$$

Note the absence of a condition on the commutator $[\delta, d]$. This would be a "higher-order" condition.

Remark 1.27. It is a formal consequence of properties of the commutator bracket that the left-hand side of (9) is a $\mathbb{C}$-linear derivation in each of its two arguments X, Y. The same is true of the right-hand side by assumption. Thus we have that if all properties of Definition 1.24 except for (i) and (ii) are satisfied, then to check that (ii) is satisfied, it suffices to do this for all X and Y belonging to a set of $\mathbb{C}$-algebra generators for A.

Remark 1.28. Suppose all properties of Definition 1.24 except for (i) are satisfied. Suppose also that L is free of rank one as an A-module on the basis $\omega^\circ \in L$. Then it suffices to prove equation (7) applied to elements of the form $X \lrcorner \omega^\circ$, where X runs over a set of generators of A as an A^0-module.

Lemma 1.29. *Let $(\psi, \delta) : (L, s) \to (M, t)$ be a homomorphism of differential Batalin–Vilkovisky modules over the morphism $(\phi, \{,\}) : (A, \widetilde{s}) \to (B, \widetilde{t})$ of differential Gerstenhaber algebras. Then $h^*(\psi) : h^*(L, s) \to h^*(M, t)$ is a homomorphism of Batalin–Vilkovisky modules over the morphism of Gerstenhaber algebras $h^*(\phi) : h^*(A, \widetilde{s}) \to h^*(B, \widetilde{t})$.*

Proof. Evaluating the right-hand side of equation (7) on s-cocycles in L yields t-boundaries in M. ☐

1.5 Invertible differential Batalin–Vilkovisky modules

Definition 1.30. We call the Batalin–Vilkovisky module L over the Gerstenhaber algebra A **invertible** if locally in S, there exists a section ω° of L such that the evaluation homomorphism

$$\begin{aligned} \Psi^\circ : A &\longrightarrow L, \\ X &\longmapsto (-1)^{\overline{X}\overline{\omega}^\circ} X \lrcorner \omega^\circ, \end{aligned}$$

is an isomorphism of sheaves of $\mathcal{O}_S$-modules. Any such ω° will be called a (local) **orientation** for L over A.

Note that if the degree of an orientation ω° is n, then $L^k = 0$ for all $k > n$, by our assumption on A. Thus orientations always live in the top degree of L. Moreover, if orientations exist everywhere locally, L^n is an invertible sheaf over A^0.

Lemma 1.31. *Let L be an invertible Batalin–Vilkovisky module over the Gerstenhaber algebra A and assume that ω° is a (global) orientation for L over A. Then, transporting the differential d via Ψ° to A yields a $\mathbb{C}$-linear map of degree $+1$, which we will call $d^\circ : A \to A$. It is characterized by the formula*

$$d^\circ(X) \lrcorner \omega^\circ = d(X \lrcorner \omega^\circ)\,.$$

It squares to 0 and it satisfies

$$(-1)^{\overline{X}}[X,Y] = d^\circ(X) \wedge Y + (-1)^{\overline{X}} X \wedge d^\circ(Y) - d^\circ(X \wedge Y)\,, \qquad (10)$$

for all $X, Y \in A$. In other words, d° is a **generator** *for the bracket $[,]$, making A a* **Batalin–Vilkovisky algebra**.

Proof. The assertion follows directly from formula (5) upon noticing that because ω° is top-dimensional, it is automatically d-closed: $d\omega^\circ = 0$. □

Corollary 1.32. *If the Gerstenhaber algebra admits an invertible Batalin–Vilkovisky module it is locally a Batalin–Vilkovisky algebra.*

Example 1.33. The Batalin–Vilkovisky module $\Omega^\bullet_M$ over the Gerstenhaber algebra ΛT_M of Example 1.10 is invertible. Any nonvanishing top-degree form $\omega^\circ \in \Omega^n_M$ is an orientation for $\Omega^\bullet_M$, where $n = \dim M$. Thus, the Schouten–Nijenhuis algebra ΛT_M is a Batalin–Vilkovisky algebra. For Calabi–Yau manifolds, i.e., $\Omega^n_M = \mathcal{O}_S$, a generator for the Batalin–Vilkovisky algebra is given.

Definition 1.34. Let (L, s) be a differential Batalin–Vilkovisky module over the differential Gerstenhaber algebra $(A, \widetilde{s})$. Then (L, s) is called **invertible** if the underlying Batalin–Vilkovisky module L is invertible over the underlying Gerstenhaber algebra A. An **orientation** for $(A, \widetilde{s})$ is an orientation of the underlying L.

Proposition 1.35. *Let (L, s) be an invertible differential Batalin–Vilkovisky module over the differential Gerstenhaber algebra $(A, \widetilde{s})$. Then under the isomorphism Ψ° defined by an orientation ω° of L over A, the differential $\widetilde{s}$ corresponds to the differential s. In particular, the induced differential d° on A has the property*

$$[d^\circ, \widetilde{s}] = 0\,,$$

besides satisfying (10). Hence $(A, d^\circ, \widetilde{s})$ is a **differential Batalin–Vilkovisky algebra**.

Moreover, the cohomology $h^*(L, s)$ is an invertible Batalin–Vilkovisky module over the Gerstenhaber algebra $h^*(A, \widetilde{s})$. We have $h^n(L, s) = L^n/I$, and the image of any orientation of L over A under the quotient map $L^n \to L^n/I$ gives an orientation for $h^*(L, s)$ over $h^*(A, \widetilde{s})$.

Proof. The equation $s \circ \Psi^\circ = (-1)^{\overline{\omega}^\circ} \Psi^\circ \circ \widetilde{s}$ follows immediately from $[s, i_X] = i_{\widetilde{s}(X)}$ upon noticing that $s(\omega) = 0$. Since Ψ° is therefore an isomorphism of differential graded $\mathcal{O}_S$-modules, the cohomology is an isomorphism: $h^*(A, \widetilde{s}) \xrightarrow{\sim} h^*(L, s)$. The rest follows from this. □

Example 1.36. For a closed 1-form s on M, the differential Batalin–Vilkovisky module $(\Omega^\bullet_M, s)$ over the differential Gerstenhaber algebra $(\Lambda T_M, \widetilde{s})$ of Example 1.22 is invertible. Any trivialization of Ω^n_M defines an orientation.

1.6 Oriented homomorphisms of invertible Batalin–Vilkovisky modules

Definition 1.37. Let $\phi : A \to B$ be a morphism of Gerstenhaber algebras and $\psi : L \to M$ a homomorphism of invertible Batalin–Vilkovisky modules covering ϕ. Let ω_L° and ω_M° be orientations for L and M, respectively. The homomorphism $\psi : L \to M$ is said to **preserve the orientations** (or be oriented) if $\psi(\omega_L^\circ) = \omega_M^\circ$.

Lemma 1.38. *Suppose we are given oriented invertible Batalin–Vilkovisky modules L and M over the Gerstenhaber algebras A and B, making A and B into Batalin–Vilkovisky algebras. Suppose $\psi : L \to M$ is an oriented homomorphism of Batalin–Vilkovisky modules. Then under the identifications of L and M with A and B given by ω_L° and ω_M°, the map $\psi : L \to M$ corresponds to $\phi : A \to B$. Hence $\phi : A \to B$ commutes with d°. Thus ϕ is a morphism of Batalin–Vilkovisky algebras: it respects $\wedge$, $[,]$, and d°.*

Definition 1.39. Let $(\psi, \delta) : (L, s) \to (M, t)$ be a homomorphism of invertible differentiable Batalin–Vilkovisky modules over $(\phi, \{,\}) : (A, \widetilde{s}) \to (B, \widetilde{t})$. Let ω_L° and ω_M° be orientations for L and M, respectively. We call (ψ, δ) **oriented** if $\psi(\omega_L^\circ) = \omega_M^\circ$ and $\delta(\omega_L^\circ) = 0$.

Proposition 1.40. *Suppose $(\psi, \delta) : (L, s, \omega_L^\circ) \to (M, t, \omega_M^\circ)$ is an oriented homomorphism of oriented invertible differential Batalin–Vilkovisky modules over $(\phi, \{,\}) : (A, \widetilde{s}) \to (B, \widetilde{t})$. Then $(A, \widetilde{s}, [,], d^\circ)$ and $(B, \widetilde{t}, [,], d^\circ)$ are differential Batalin–Vilkovisky algebras. Transporting $\delta : L \to M$ via the identifications of L and M with A and B to a map $\delta^\circ : A \to B$ satisfying*

$$\delta^\circ(X) \lrcorner\, \omega_M = (-1)^{\overline{\delta X}} \delta(X \lrcorner\, \omega_L)\,,$$

we get a triple

$$(\phi, \{,\}, \delta^\circ) : (A, \widetilde{s}, [,], d^\circ) \longrightarrow (B, \widetilde{t}, [,], d^\circ)\,,$$

which satisfies the following conditions:

(i) $\phi : (A, \widetilde{s}) \to (B, \widetilde{t})$ *is a morphism of differential graded algebras;*

(ii) we have the commutator property

$$\phi \circ d^\circ - d^\circ \circ \phi = -2\,\widetilde{t} \circ \delta^\circ + 2\,\delta^\circ \circ \widetilde{s}\,,$$

or, by abuse of notation, $[\phi, d^\circ] - 2[\delta^\circ, \widetilde{s}] = 0$*;*

(iii) the map δ° is a generator for the bracket $\{,\}$,

$$\{X, Y\} = \delta^\circ(X) \wedge \phi(Y) + \phi(X) \wedge \delta^\circ(Y) - \delta^\circ(X \wedge Y);$$

(iv) the failure of ϕ to preserve $[,]$ equals the failure of $\widetilde{s}$ to be a derivation with respect to $\{,\}$, equation (4).

Thus $(\phi, \{\,\}, \delta^\circ)$ *is a* **(first-order) morphism of differential Batalin–Vilkovisky algebras**.

The Lie bracket $[,]$ is determined by its generator d^0, and the bracket $\{,\}$ is determined by its generator δ^0. Thus, in a certain sense, the two brackets are redundant. Moreover, condition (iv) is implied by conditions (ii) and (iii).

Remark 1.41. A morphism of differential Batalin–Vilkovisky algebras

$$(\phi, \{,\}, \delta^\circ) : (A, \widetilde{s}, [,], d^\circ) \longrightarrow (B, \widetilde{t}, [,], d^\circ)$$

induces on cohomology

$$h^*(\phi) : \big(h^*(A, \widetilde{s}), [,], d^\circ\big) \to \big(h^*(B, \widetilde{t}), [,], d^\circ\big)$$

a morphism of Batalin–Vilkovisky algebras.

2 Symplectic geometry

Let (S, σ) be a *symplectic manifold,* i.e., a complex manifold S endowed with a closed holomorphic 2-form $\sigma \in \Omega^2_S$ that is everywhere nondegenerate, i.e., $X \to X \lrcorner\, \sigma$ defines an isomorphism of vector bundles $T_S \to \Omega_S$. The (complex) dimension of S is even, and we will denote it by $2n$.

A submanifold $M \subset S$ is *Lagrangian* if the restriction of this isomorphism $T_S|_M \to \Omega_S|_M$ identifies $T_M \subset T_S|_M$ with $T_M^\perp \subset \Omega_S|_M$. An equivalent condition is that the restriction of σ to a 2-form on M vanishes and that $\dim M = n$. More generally, we define an *immersed Lagrangian* to be an unramified morphism $i : M \to S$, where M is a manifold of dimension n, such that $i^*\sigma \in \Omega^2_M$ vanishes.

Holomorphic coordinates $x_1, \ldots, x_n, p_1, \ldots, p_n$ on S are called *Darboux coordinates* if

$$\sigma = \sum_{i=1}^n dp_i \wedge dx_i\,.$$

Let us introduce one further piece of notation. For a subbundle $E \subset \Omega_S$ we consider the associated bundles $E^\perp$, $E^\vee$, and $E^\dagger$ defined by the short exact sequences of vector bundles

$$0 \longrightarrow E^\perp \longrightarrow T_S \longrightarrow E^\vee \longrightarrow 0$$

and

$$0 \longrightarrow E \longrightarrow \Omega_S \longrightarrow E^\dagger \longrightarrow 0\,.$$

2.1 Lagrangian foliations

Definition 2.1. A **Lagrangian foliation** on S is an integrable distribution $F \subset T_S$, where $F \subset T_S$ is a Lagrangian subbundle, i.e., $X \to X \lrcorner \sigma$ defines an isomorphism of vector bundles $F \to F^\perp \subset \Omega_S$.

All leaves of the Lagrangian foliation F are Lagrangian submanifolds of S. The Lagrangian foliation $F \subset T_S$ may be equivalently defined in terms of the subbundle $E = F^\perp \subset \Omega_S$. Usually, we find it more convenient to specify $E \subset \Omega_S$, rather than $F \subset T_S$. In terms of E, we have the following isomorphism of short exact sequences of vector bundles:

$$\begin{tikzcd}
0 \arrow[r] & E^\perp \arrow[r] \arrow[d, "\cong"'] & T_S \arrow[r] \arrow[d, "\cong"', "\lrcorner\sigma"] & E^\vee \arrow[r] \arrow[d, "\cong"'] & 0 \\
0 \arrow[r] & E \arrow[r] & \Omega_S \arrow[r] & E^\dagger \arrow[r] & 0
\end{tikzcd}$$

Definition 2.2. A **polarized** symplectic manifold is a symplectic manifold endowed with a Lagrangian foliation.

The canonical partial connection

Any foliation $F \subset T_S$ defines a partial connection on the quotient bundle T_S/F:

$$\nabla : T_S/F \longrightarrow F^\vee \otimes T_S/F \,, \tag{11}$$

given by

$$\nabla_Y(X) = [Y, X] \,,$$

for $Y \in F$ and $X \in T_S/F$. This partial connection is flat. The dual bundle of T_S/F is $F^\perp \subset \Omega_S$. The dual connection

$$\nabla : F^\perp \longrightarrow F^\vee \otimes F^\perp$$

is given by

$$\nabla_Y(\omega) = Y \lrcorner d\omega$$

for $Y \in F$ and $\omega \in F^\perp \subset \Omega_S$.

Let us specialize to the case that F is Lagrangian. Then we can transport the partial connection from $F^\perp$ to F via the isomorphism $F \cong F^\perp$. We obtain the **canonical partial flat connection**

$$\nabla : F \longrightarrow F^\vee \otimes F$$

characterized by

$$\nabla_Y(X) \lrcorner \sigma = Y \lrcorner d(X \lrcorner \sigma) \,,$$

for $Y \in F$ and $X \in F$. The dual of this partial connection is

$$\nabla : F^\vee \longrightarrow F^\vee \otimes F^\vee ,$$

which is characterized by

$$\nabla_Y(X \lrcorner \sigma) = [Y, X] \lrcorner \sigma ,$$

for $Y \in F$ and $X \in T_S$.

Oriented Lagrangian foliations

Definition 2.3. Let $F \subset T_S$ be a Lagrangian foliation on S. An **orientation** of F is a nowhere vanishing global section

$$\theta \in \Gamma(S, \Lambda^n F)$$

that is *flat* with respect to the canonical partial connection on F.

A polarized symplectic manifold is called **oriented** if its Lagrangian foliation is endowed with an orientation.

Remark 2.4. If θ is an orientation of the Lagrangian foliation F, then we have $\nabla(\theta \lrcorner \sigma^n) = 0$. (Note that $\theta \lrcorner \sigma^n \in \Lambda^n F^\perp \subset \Lambda^n \Omega_S$.)

2.2 Polarizations and transverse Lagrangians

Let $E \subset \Omega_S$ define a Lagrangian foliation on S.

Lemma 2.5. *Let M be a Lagrangian submanifold of S that is everywhere transverse to E. Then there exists (locally near M) a unique section s of E such that $ds = \sigma$ and $M = Z(s)$, i.e., M is the zero locus of s (as a section of the vector bundle E).*

Definition 2.6. We call s the **Euler form** of M with respect to E, or the **Euler section** of M in E.

Remark 2.7. Conversely, if s is any section of E such that $ds = \sigma$, then $Z(s)$ is a Lagrangian submanifold. Thus we have a canonical one-to-one correspondence between sections s of E such that $ds = \sigma$ and Lagrangian submanifolds of S transverse to E.

Lemma 2.8. *Let (S, F, σ, θ) be an oriented polarized symplectic manifold and $E = F^\perp$. Let $M \subset S$ be a Lagrangian submanifold, everywhere transverse to F. Then near every point of M there exists a set of Darboux coordinates $x_1, \ldots, x_n, p_n \ldots, p_n$ such that*

(i) $M = Z(p_1, \ldots, p_n)$*;*

(ii) $F = \left\langle \frac{\partial}{\partial p_1} \ldots, \frac{\partial}{\partial p_n} \right\rangle$*;*

(iii) $\nabla\left(\frac{\partial}{\partial p_i}\right) = 0$*, for all* $i = 1, \dots, n$*;*

(iv) $\theta = \frac{\partial}{\partial p_i} \wedge \cdots \wedge \frac{\partial}{\partial p_n}$.

Moreover, in these coordinates we have

(i) $E = \langle dx_1, \dots, dx_n \rangle$*;*

(ii) the Euler form s *of* M *inside* E *is given by* $s = \sum p_i dx_i$.

3 Derived Lagrangian intersections on polarized symplectic manifolds

Definition 3.1. Let (S, E, σ) be a polarized symplectic manifold and L, M immersed Lagrangians of S that are both transverse to E. Then the **derived intersection**

$$L \Cap_S M$$

is the sheaf of differential Gerstenhaber algebras $(\Lambda T_M, \widetilde{t})$ on M, where $\widetilde{t}$ is the derivation on ΛT_M induced by the restriction to M of the Euler section $t \in E \subset \Omega_S$ of L.

Since $dt = \sigma$ and M is Lagrangian, the restriction of t to M is closed, and so $\widetilde{t}$ is a derivation with respect to the Schouten–Nijenhuis bracket on ΛT_M, making $(\Lambda T_M, \widetilde{t})$ a differential Gerstenhaber algebra.

Remark 3.2. After passing (locally in L) to suitable étale neighborhoods of L in S we can assume that L is embedded (not just immersed) in S and that L admits a globally defined Euler section t on S. This defines the derived intersection étale locally in M, and the global derived intersection is defined by gluing in the étale topology on M.

Remark 3.3. If we forget about the bracket, the underlying complex of $\mathcal{O}_S$-modules $(\Lambda T_M, \widetilde{t})$ represents the derived tensor product

$$\mathcal{O}_L \overset{L}{\otimes}_{\mathcal{O}_S} \mathcal{O}_M$$

in the derived category of sheaves of $\mathcal{O}_S$-modules.

Remark 3.4. The derived intersection $L \Cap_S M$ depends a priori on the polarization E. We will see later (see the proof of Theorem 4.2) that different polarizations lead to *locally* quasi-isomorphic derived intersections. (The quasi-isomorphism is not canonical, since it depends on the choice of a third polarization transverse to both of the polarizations being compared. It is not clear that such a third polarization can necessarily be found globally.)

Remark 3.5. The derived intersection does not seem to be symmetric. We will see below that $L \Cap_S M = M \Cap_{\overline{S}} L$, where $\overline{S} = (S, -\sigma)$, but only if $\overline{S}$ is endowed with a different polarization, transverse to E. Then the issue of change of polarization of Remark 3.4 arises.

Definition 3.6. Let S, L, M be as in Definition 3.1. Let M be *oriented*, i.e., endowed with a nowhere-vanishing top-degree differential form ω_M°. (Since M is Lagrangian, this amounts to the same as a trivialization of the determinant of the normal bundle $N_{M/C}$.) We call the differential Batalin–Vilkovisky algebra $(\Lambda T_M, \widetilde{t}, [,], d^\circ)$, where d° is induced by ω_M° as in Section 1.5, the **oriented derived intersection**, notation $L \Cap_S^\circ M$.

By a *local system* we mean a vector bundle (locally free sheaf of finite rank) endowed with a flat connection. Every local system P on a complex manifold M has an associated holomorphic de Rham complex $(P \otimes_{\mathcal{O}_M} \Omega_M^\bullet, d)$, where d denotes the covariant derivative.

Definition 3.7. Let (S, E, σ) be a polarized symplectic manifold and L, M immersed Lagrangians, both transverse to E. Let P be a local system on M and Q a local system on S. The **derived hom** from $Q|L$ to $P|M$ is the differential Batalin–Vilkovisky module

$$\mathbb{R}\mathcal{H}om_S\big(Q|L, P|M\big) = \big(\Omega_M^\bullet \otimes Q^\vee|_M \otimes P, t\big)$$

over the differential Gerstenhaber algebra

$$L \Cap_S M = (\Lambda T_M, \widetilde{t}).$$

The tensor products are taken over $\mathcal{O}_M$. The closed 1-form $t \in \Omega_M$ is the restriction to M of the Euler section of L inside E. The $\mathcal{O}_M$-linear differential t is multiplication by t and the $\mathbb{C}$-linear differential d is covariant derivative with respect to the induced flat connection on $Q^\vee|_M \otimes P$.

Remark 3.8. If we forget about the $\mathbb{C}$-linear differential d and the flat connections on P and Q, the underlying complex of $\mathcal{O}_S$-modules $\mathbb{R}\mathcal{H}om_S(Q|L, P|M)$ represents the derived sheaf of homomorphisms $\mathbb{R}\mathcal{H}om_{\mathcal{O}_S}(Q|_L, P)$ in the derived category of sheaves of $\mathcal{O}_S$-modules.

3.1 The exchange property: Gerstenhaber case

Given two symplectic manifolds S', S, of dimensions $2n'$ and $2n$, a **symplectic correspondence** between S' and S is a manifold C of dimension $n + n'$, together with morphisms $\pi' : C \to S'$ and $\pi : C \to S$, such that

(i) $\pi^*\sigma = \pi'^*\sigma'$ (as sections of Ω_C),

(ii) $C \to S' \times S$ is unramified.

Thus a symplectic correspondence is an immersed Lagrangian of

$$\overline{S}' \times S = (S' \times S, \sigma - \sigma').$$

Let $C \to S' \times S$ be a symplectic correspondence. We say that the immersed Lagrangian $L \to S$ is **transverse** to C if

(i) for every $(Q,P) \in C \times_S L$ we have that

$$T_C|_Q \oplus T_L|_P \longrightarrow T_S|_{\pi(Q)}$$

is surjective; hence the pullback $L' = C \times_S L$ is a manifold of dimension n';

(ii) the natural map $L' \to S'$ is unramified (and hence L' is an immersed Lagrangian of S').

By exchanging the roles of S and S' we also get the notion of transversality to C for immersed Lagrangians of S'.

Exchange property setup

Let (S, E, σ) and (S', E', σ') be polarized symplectic manifolds. Let $E^\perp \subset T_S$ and ${E'}^\perp \subset T_{S'}$ be the corresponding Lagrangian foliations. Consider a **transverse** symplectic correspondence $C \to S' \times S$. This means that $C \to S' \times S$ is transverse to the foliation ${E'}^\perp \times E^\perp$ of $S' \times S$. In particular, the composition

$$T_C \longrightarrow \pi^* T_S \longrightarrow \pi^* E^\vee$$

is surjective. Hence the foliation $E^\perp \subset T_S$ pulls back to a foliation $F \subset T_C$ of rank n'. We have the exact sequence of vector bundles

$$0 \longrightarrow F \longrightarrow T_C \longrightarrow \pi^* E^\vee \longrightarrow 0\,. \tag{12}$$

Similarly, the foliation ${E'}^\perp \subset T_{S'}$ pulls back to a foliation $F' \subset T_C$ of rank n with the exact sequence

$$0 \longrightarrow F' \longrightarrow T_C \longrightarrow {\pi'}^* {E'}^\vee \longrightarrow 0\,.$$

Moreover, F and F' are transverse foliations of C, and so we have

$$F \oplus F' = T_C = {\pi'}^* {E'}^\vee \oplus \pi^* E^\vee\,.$$

Even though it is not strictly necessary, we will make the assumption that $F \subset T_C$ descends to a Lagrangian foliation $\widetilde{F} \subset T_{S'}$ and $F' \subset T_C$ descends to a Lagrangian foliation $\widetilde{F'} \subset T_S$. This makes some of the arguments simpler.

Remark 3.9. The composition

$$F \longrightarrow T_C \longrightarrow {\pi'}^* T_{S'} \xrightarrow{\ \lrcorner\, {\pi'}^* \sigma'\ } {\pi'}^* \Omega_{S'} \longrightarrow {\pi'}^* {E'}^\dagger$$

defines an isomorphism of vector bundles $\beta : F \xrightarrow{\sim} {\pi'}^* {E'}^\dagger$ and its inverse $\eta : {\pi'}^* {E'}^\dagger \xrightarrow{\sim} F$. We can reinterpret these as perfect pairings $\beta : F \otimes_{\mathcal{O}_C} {\pi'}^* {E'}^\perp \to \mathcal{O}_C$ and $\eta : F^\vee \otimes_{\mathcal{O}_C} {\pi'}^* {E'}^\dagger \to \mathcal{O}_C$. These will be important in the proof below.

Now assume that we are given immersed Lagrangians L of S and M' of S'. Assume that both are transverse to C. Then we obtain manifolds L' and M by the pullback diagram

$$\begin{tikzcd} & L' \arrow[r] \arrow[d] \arrow[dr, phantom, "\square"] & L \arrow[d] \\ M \arrow[r] \arrow[d] \arrow[dr, phantom, "\square"] & C \arrow[r, "\pi"] \arrow[d, "\pi'"] & S \\ M' \arrow[r] & S' & \end{tikzcd} \tag{13}$$

Then L' is an immersed Lagrangian of S' and M an immersed Lagrangian of S.

Finally, we assume that L and M are transverse to E and that M' and L' are transverse to E'. As a consequence, L' is transverse to F' and M is transverse to F.

Remark 3.10. Since M is transverse to F, we have a canonical isomorphism $F|_M = N_{M/C}$. Also, since $\pi'^* N_{M'/S'} = N_{M/C}$, we have $\pi'^* E'^{\perp}|_M = N_{M/C}$. Thus, restricting the pairings β and η to M, we obtain

$$\beta|_M : N_{M/C} \otimes_{\mathcal{O}_M} N_{M/C} \to \mathcal{O}_M,$$

and

$$\eta|_M : N_{M/C}^\vee \otimes_{\mathcal{O}_M} N_{M/C}^\vee \to \mathcal{O}_M.$$

Lemma 3.11. *If $s \in E$ is the Euler section of M in E and s' the Euler section of M' in E', then the homomorphism $\beta|_M : N_{M/C} \to N_{M/C}^\vee$ fits into the commutative diagram*

$$\begin{tikzcd} T_C \arrow[rr, "\widetilde{s}-\widetilde{s}'"] \arrow[d] & & \mathcal{O}_C \arrow[r, "d"] & \Omega_C \arrow[d] \\ T_C|_M \arrow[r] & N_{M/C} \arrow[r, "\beta|_M"] & N_{M/C}^\vee \arrow[r] & \Omega_C|_M \end{tikzcd}$$

Proof. Let $P \in M \subset C$ be a point. It suffices to prove the claim locally near P. Let $\widetilde{F} \subset T_{S'}$ be the Lagrangian foliation on S', which pulls back to $F \subset T_C$. Then $\widetilde{F}$ is transverse to both E' and M'.

Choose holomorphic functions $x_1, \ldots, x_{n'}$ in a neighborhood of $\pi'(P)$ in S' such that $dx_1, \ldots, dx_{n'}$ is a basis for $E' \subset \Omega_{S'}$. Also, choose $y_1, \ldots, y_{n'}$, such that $dy_1, \ldots, dy_{n'}$ is a basis for $\widetilde{F}^\perp \subset \Omega_{S'}$. Then (x_i, y_j) is a set of coordinates for S' near $\pi'(P)$.

Let $\overline{s}$ be the Euler section of M' in $\widetilde{F}^\perp$ and let f be the unique holomorphic function on S', defined in a neighborhood of $\pi'(P)$, such that $f(\pi'(P)) = 0$

and $df = \overline{s} - s'$. Then we have $\overline{s} = \sum_j \frac{\partial f}{\partial y_j} dy_j$ and $s' = -\sum_i \frac{\partial f}{\partial x_i} dx_i$. Moreover, $\sigma' = \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial y_j} dx_i \wedge dy_j$.

We remark that the composition $T_{S'} \xrightarrow{df} \mathcal{O}_{S'} \xrightarrow{d} \Omega_{S'}$ factors through $N_{M'/S'} \to N^\vee_{M'/S'}$, because f vanishes on M'. The resulting map is, in fact, the Hessian of f. Via our identifications, this Hessian agrees with the map $\widetilde{F}|_{M'} \to {E'}^\dagger|_{M'}$ induced by σ', because, with our choice of coordinates, $\widetilde{F}|_{M'} = N_{M'/S'}$ has basis $\frac{\partial}{\partial x_i}$ and ${E'}^\dagger|_{M'} = N^\vee_{M'/S'}$ has basis dy_j.

To transfer this result from S' to C, we remark that the pullback of $\overline{s}$ to C is necessarily equal to the pullback of s to C. Thus the composition $d \circ (\widetilde{s} - \widetilde{s}')$ is equal to the Hessian of the pullback of f to C. This is, by what we proved above, equal to the pullback of the map induced by σ'. □

Theorem 3.12. *There are canonical quasi-isomorphisms of differential Gerstenhaber algebras*

$$(M' \times L) \Cap_{\overline{S}' \times S} C \longrightarrow L \Cap_S M$$

and

$$(M' \times L) \Cap_{\overline{S}' \times S} C \longrightarrow M' \Cap_{\overline{S}'} L' \,.$$

In particular, the derived intersections $L \Cap_S M$ and $M' \Cap_{\overline{S}'} L'$ are canonically quasi-isomorphic.

Proof. Passing to étale neighborhoods of L in S and M' in S' will not change anything about either derived intersection $L \Cap_S M$ or $M' \Cap_{\overline{S}'} L'$, so we may assume, without loss of generality, that

(i) L is embedded (not just immersed) in S (and the same for M' in S'),

(ii) L admits a global Euler section t with respect to E on S (and M' has the Euler section s' in E' on S').

Then the Euler section of M' with respect to E' on $\overline{S}'$ is $-s'$. Thus the derived intersection $L \Cap_S M$ is equal to $(\Lambda T_M, \widetilde{t})$ and the derived intersection $M' \Cap_{\overline{S}'} L'$ equals $(\Lambda T_{L'}, -\widetilde{s}')$.

Pulling back the 1-form t via π, we obtain a 1-form on C, which we shall, by abuse of notation, also denote by t. Similarly, pulling back s' via π' we get the 1-form s' on C. The difference $t - s'$ is closed on C, and thus we have the differential Gerstenhaber algebra $(\Lambda T_C, \widetilde{t} - \widetilde{s}')$. We remark that it is equal to $(M' \times L) \Cap_{\overline{S}' \times S} C$.

Recall that we have the identification $T_C = {\pi'}^* {E'}^\vee \oplus \pi^* E^\vee$. Under this direct sum decomposition $\widetilde{t} - \widetilde{s}'$ splits up into two components, $-\widetilde{s}'$ and $\widetilde{t}$. Hence we obtain the decomposition

$$(\Lambda T_C, \widetilde{t} - \widetilde{s}') = {\pi'}^*(\Lambda {E'}^\vee, -\widetilde{s}') \otimes \pi^*(\Lambda E^\vee, \widetilde{t})$$

of differential graded $\mathcal{O}_C$-algebras.

Recall that $\mathcal{O}_{S'} \to \mathcal{O}_{M'}$ induces a quasi-isomorphism of differential graded $\mathcal{O}_{S'}$-algebras $(\Lambda {E'}^\vee, -\widetilde{s}') \to \mathcal{O}_{M'}$. Because the pullback $M = M' \times_{S'} C$ is transverse, we get an induced quasi-isomorphism

$$\pi'^*(\Lambda {E'}^\vee, -\widetilde{s}') \longrightarrow \mathcal{O}_M$$

of differential graded $\mathcal{O}_C$-algebras. Tensoring with $\pi^*(\Lambda E^\vee, \widetilde{t})$, we obtain the quasi-isomorphism

$$(\Lambda T_C, \widetilde{t} - \widetilde{s}') \longrightarrow (\Lambda E^\vee, \widetilde{t})|_M \,.$$

Noting that $E^\vee|_M = T_M$, because M is an immersed submanifold in S transverse to E, we see that $(\Lambda E^\vee, \widetilde{t})|_M = (\Lambda T_M, \widetilde{t})$, and so we have a quasi-isomorphism of differential graded $\mathcal{O}_C$-algebras

$$\phi : (\Lambda T_C, \widetilde{t} - \widetilde{s}') \longrightarrow (\Lambda T_M, \widetilde{t}) \,. \tag{14}$$

For analogous reasons, we also have the quasi-isomorphism

$$\phi' : (\Lambda T_C, \widetilde{t} - \widetilde{s}') \longrightarrow (\Lambda T_{L'}, -\widetilde{s}') \,.$$

The proof will be finished if we can enhance ϕ and ϕ' by brackets, making them morphisms of differential Gerstenhaber algebras. We will concentrate on ϕ. The case of ϕ' follows by symmetry.

Thus we shall define a bracket

$$\{,\} : \Lambda T_C \otimes_{\mathbb{C}} \Lambda T_C \longrightarrow \Lambda T_M \tag{15}$$

such that $(\phi, \{,\})$ becomes a morphism of differential Gerstenhaber algebras.

We use the foliation $F \subset T_C$. It defines as in equation (11) a partial flat connection

$$\nabla : T_C/F \longrightarrow F^\vee \otimes_{\mathcal{O}_C} T_C/F \,.$$

By the usual formulas we can transport ∇ onto the exterior powers of T_C/F. In our context, we obtain

$$\nabla : \pi^* \Lambda E^\vee \longrightarrow F^\vee \otimes_{\mathcal{O}_C} \pi^* \Lambda E^\vee \,.$$

To get the signs right, we will consider the elements of the factor $F^\vee$ in this expression to have degree zero.

Let us write the projection $\Lambda T_C \to \pi^* \Lambda E^\vee$ as ρ. We identify $F^\vee|_M$ with $N^\vee_{M/C}$ and $(\pi^* \Lambda E^\vee)|_M$ with ΛT_M. Then ϕ is the composition of ρ with restriction to M. We now define for $X, Y \in \Lambda T_C$,

$$\{X, Y\} = \eta\big(\nabla(\rho X)|_M \wedge \nabla(\rho Y)|_M\big) \,. \tag{16}$$

In this formula, "$\wedge$" denotes the homomorphism (all tensors are over $\mathcal{O}_M$)

$$\begin{aligned}\Big(N^\vee_{M/C}\otimes\Lambda T_M\Big)\otimes\Big(N^\vee_{M/C}\otimes\Lambda T_M\Big)\longrightarrow\Big(N^\vee_{M/C}\otimes N^\vee_{M/C}\Big)\otimes\Lambda T_M,\\ v\otimes X\otimes w\otimes Y\longmapsto v\otimes w\otimes X\wedge Y.\end{aligned}$$

There is no sign correction in this definition, because the elements of $N^\vee_{M/C}$ are considered to have degree zero, by our sign convention. We have also extended the map η linearly to

$$\eta:\Big(N^\vee_{M/C}\otimes_{\mathcal{O}_M}N^\vee_{M/C}\Big)\otimes_{\mathcal{O}_M}\Lambda T_M\longrightarrow\Lambda T_M\,.$$

Claim. The conditions of Definition 1.12 are satisfied by $(\phi,\{\;\})$.

All but the last condition follow easily from the definitions. Let us check condition (iv). We use Remark 1.14. The $\mathbb{C}$-algebra ΛT_C is generated in degrees 0 and -1. As generators in degree -1, we may take the basic vector fields of a coordinate system for C. We choose this coordinate system such that M is cut out by a subset of the coordinates. Then, if we plug in generators of degree -1 for both X and Y in formula (4), every term vanishes. Also, if we plug in terms of degree 0 for both X and Y, both sides of (4) vanish for degree reasons. By symmetry, we thus reduce to considering the case in which X is of degree -1, i.e., a vector field on C, and Y is of degree 0, i.e., a regular function on C.

Hence we need to prove that for all $X\in T_C$ and $g\in\mathcal{O}_C$ we have

$$X(g)|_M-\rho(X)|_M\big(g|_M\big)=\{(\widetilde{t}-\widetilde{s}')X,g\}-\widetilde{t}\{X,g\}\,. \tag{17}$$

Let s denote the Euler section of M in $E\subset\Omega_S$, and its pullback to C. We will prove that

$$X(g)|_M-\rho(X)|_M\big(g|_M\big)=\{(\widetilde{s}-\widetilde{s}')X,g\} \tag{18}$$

and

$$\{(\widetilde{t}-\widetilde{s})X,g\}=\widetilde{t}\{X,g\}\,. \tag{19}$$

equation (18) involves only M, not L, and equation (19) involves only E, not E'. Together, they imply equation (17).

All terms in these three equations are $\mathcal{O}_S$-linear in X and derivations in g, and may hence be considered as $\mathcal{O}_C$-linear maps $T_C\to\mathcal{D}er(\mathcal{O}_C,\mathcal{O}_M)$. Since $\mathcal{D}er(\mathcal{O}_C,\mathcal{O}_M)=\mathcal{H}om_{\mathcal{O}_C}(\Omega_C,\mathcal{O}_M)=T_C|M$, we may also think of them as $\mathcal{O}_C$-linear maps $T_C\to T_C|_M$.

For example, the $\mathcal{O}_C$-linear map

$$\begin{aligned}T_C&\longrightarrow T_C|_M,\\ X&\longmapsto\{(\widetilde{s}-\widetilde{s}')X,\,\cdot\,\},\end{aligned} \tag{20}$$

is equal to the composition

$$T_C \xrightarrow{\widetilde{s}-\widetilde{s}'} \mathcal{O}_C \xrightarrow{d} \Omega_C \longrightarrow F^\vee|_M \xrightarrow{\eta} F|_M \longrightarrow T_C|_M \,.$$

Then the commutative diagram (Lemma 3.11)

$$\begin{tikzcd}
T_C \arrow[r, "\widetilde{s}-\widetilde{s}'"] \arrow[d] & \mathcal{O}_C \arrow[r, "d"] & \Omega_C \arrow[d] \arrow[dr] & & & \\
T_C|_M \arrow[d] & & \Omega_C|_M \arrow[r] & F^\vee|_M \arrow[r, "\eta"] & F|_M \arrow[r] & T_C|_M \\
N_{M/C} \arrow[rr, "\beta"] & & N^\vee_{M/C} \arrow[u] \arrow[ur, "\mathrm{id}"'] & & &
\end{tikzcd}$$

and the fact that η is the inverse of β proves that (20) is equal to the composition

$$T_C \longrightarrow T_C|_M \xrightarrow{p} T_C|_M \,,$$

where p is the projection onto the the second summand of the decomposition

$$T_C|_M = T_M \oplus N_{M/C}$$

given by the foliation F transverse to M in C. If we denote by q the projection onto the first summand, we see that the map

$$\begin{aligned} T_C &\longrightarrow T_C|_M, \\ X &\longmapsto \rho(X)|_M(\,\cdot\,|_M), \end{aligned} \tag{21}$$

is equal to

$$T_C \longrightarrow T_C|_M \xrightarrow{q} T_C|_M \ .$$

Thus (20) and (21) sum to the restriction map $T_C \to T_C|_M$, which is equal to the map given by $X \mapsto X(\,\cdot\,)|_M$. This proves (18).

Now let us remark that for any closed 1-form u on C we have

$$\widetilde{u}[Y,X] = Y\big(\widetilde{u}(X)\big) - X\big(\widetilde{u}(Y)\big)\,.$$

If $u \in \pi^*E \subset \Omega_C$, then $\widetilde{u}(Y) = 0$, for all $Y \in F$. So if $Y \in F$ we have

$$\widetilde{u}[Y,X] = Y\big(\widetilde{u}(X)\big)\,.$$

We have $\widetilde{u}[Y,X] = \widetilde{u}\big(\nabla(X)(Y)\big)$ by definition of the partial connection ∇ and we can write $Y\big(\widetilde{u}(X)\big) = \langle Y, d\big(\widetilde{u}(X)\big)\rangle$. In other words, the diagram

$$\begin{tikzcd}
T_C/F \arrow[rr, "\nabla"] \arrow[d, "\widetilde{u}"'] & & \mathcal{H}om(F, T_C/F) \arrow[d, "-\circ\widetilde{\mathrm{u}}"] \\
\mathcal{O}_C \arrow[r, "d"] & \Omega_C \arrow[r] & F^\vee
\end{tikzcd}$$

commutes. Thus, the larger diagram

$$\begin{tikzcd}
T_C \arrow[r, "\rho"] \arrow[dr, "\widetilde{u}"'] & \pi^* E^\vee \arrow[r, "\nabla"] \arrow[d, "\widetilde{u}"] & F^\vee \otimes \pi^* E^\vee \arrow[r] \arrow[d, "\mathrm{id} \otimes \widetilde{u}"] & F^\vee|_M \otimes T_M \arrow[r, "\eta\otimes\mathrm{id}"] \arrow[d, "\mathrm{id} \otimes \widetilde{u}|_M"] & F|_M \otimes T_M \arrow[d, "\mathrm{id} \otimes \widetilde{u}|_M"] \\
& \mathcal{O}_C \arrow[r, "d"] & F^\vee \arrow[r] & F^\vee|_M \arrow[r, "\eta"] & F|_M \arrow[d] \\
& & & & T_C|_M
\end{tikzcd}$$

commutes as well. We can apply these considerations to $u = t - s$. Then $\widetilde{u} = \widetilde{t} - \widetilde{s}$ and $\widetilde{u}|_M = \widetilde{t}|_M$. Thus the upper composition in this diagram represents the right-hand side of equation (19), and the lower composition represents the left-hand side of equation (19). This shows that (19) holds and finishes the proof of the theorem. □

3.2 The Batalin–Vilkovisky case

For the exchange property in the Batalin–Vilkovisky case, we require an *orientation* on the symplectic correspondence $C \to S' \times S$.

Definition 3.13. Let $\pi : C \to S$ be a morphism of complex manifolds, $F \subset T_C$ and $\widetilde{F} \subset T_S$ foliations. The foliations F, $\widetilde{F}$ are **compatible** (with respect to π), if $F \to T_C \to \pi^* T_S$ factors through $\pi^* \widetilde{F} \to \pi^* T_S$.

If F and $\widetilde{F}$ are compatible, then partial connections with respect to $\widetilde{F}$ pull back to partial connections with respect to F.

Now let (S, E, σ), (S', E', σ') and $C \to S' \times S$ be, as in Section 3.1, polarized symplectic manifolds with a transverse symplectic correspondence. Let F and F' be, as in 3.1, the inverse image foliations:

$$\begin{tikzcd}
& F \arrow[r] \arrow[d] \arrow[dr, phantom, "\square"] & E^\perp \arrow[d] \\
F' \arrow[r] \arrow[d] \arrow[dr, phantom, "\square"] & C \arrow[r, "\pi"] \arrow[d, "\pi'"] & S \\
E'^\perp \arrow[r] & S' &
\end{tikzcd} \tag{22}$$

Furthermore, we suppose that $\widetilde{F} \subset T_{S'}$ is a Lagrangian foliation on S' compatible with F via π' and that $\widetilde{F}' \subset T_S$ is a Lagrangian foliation on S, compatible with F' via π. Since the composition $F \to {\pi'}^* T_{S'} \to {\pi'}^* {E'}^\vee$ is an isomorphism, the map $F \to {\pi'}^* T_{S'}$ identifies F with a subbundle of ${\pi'}^* T_{S'}$.

Since F and $\widetilde{F}$ have the same rank, it follows that $F \to \pi'^* \widetilde{F}$ is an isomorphism of subbundles of $\pi'^* T_{S'}$. Similarly, we have an identification $F' \to \pi^* \widetilde{F}'$ of subbundles of $\pi^* T_S$.

Thus we have two Lagrangian foliations on $\overline{S}' \times S$, namely $E'^{\perp} \times E^{\perp}$ and $\widetilde{F} \times \widetilde{F}'$. Both Lagrangian foliations are transverse to C, and they are transverse to each other near C.

Definition 3.14. If $\theta \in \Gamma(S', \Lambda^{n'} \widetilde{F})$ and $\theta' \in \Gamma(S, \Lambda^n \widetilde{F}')$ are orientations of the Lagrangian foliations $\widetilde{F}$ on S' and $\widetilde{F}'$ on S, we call the data $(\widetilde{F}, \theta, \widetilde{F}', \theta')$ an **orientation** of the symplectic correspondence $C \to S' \times S$.

We call the transverse symplectic correspondence of polarized symplectic manifolds $C \to S' \times S$ **orientable** if it admits an orientation.

Exchange property setup

Let (S, E, σ) and (S', E', σ') be polarized symplectic manifolds and $C \to S' \times S$, $(\widetilde{F}, \theta, \widetilde{F}', \theta')$ an oriented transverse symplectic correspondence. Moreover, let $L \to S$ and $M' \to S'$ be, as in Section 3.1, immersed Lagrangians transverse to C such that the induced M and L' are transverse to E and E', respectively. (This latter condition is satisfied if M' and L are transverse to $\widetilde{F}$ and $\widetilde{F}'$, respectively.) Pulling back θ to C gives us a trivialization of $\Lambda^{n'} F$ and restricting further to M gives a trivialization of the determinant of the normal bundle $\Lambda^{n'} N_{M/C}$, because of the canonical identification $F|_M = N_{M/C}$. Similarly, θ' gives rise to a trivialization of the determinant of the normal bundle $\Lambda^n N_{L'/C}$.

Finally, let P' be a local system on S', and Q a local system on S. Let $P = \pi'^* P'$ and $Q' = \pi^* Q$ be the pullbacks of these local systems to C.

Theorem 3.15. *There exists a canonical quasi-isomorphism of differential Batalin–Vilkovisky modules*

$$\mathbb{R}\mathcal{H}om_{\overline{S}' \times S}\big(\mathcal{O}|(M' \times L), (P \otimes Q'^{\vee})|C\big) \longrightarrow \mathbb{R}\mathcal{H}om_S\big(Q|L, P|M\big)$$

of degree $-n'$, covering the corresponding canonical quasi-isomorphism of differential Gerstenhaber algebras of Theorem 3.12. Moreover, there is the quasi-isomorphism of differential Batalin–Vilkovisky modules

$$\mathbb{R}\mathcal{H}om_{\overline{S}' \times S}\big(\mathcal{O}|(M' \times L), (P \otimes Q'^{\vee})|C\big) \longrightarrow \mathbb{R}\mathcal{H}om_{\overline{S}'}\big(P'^{\vee}|M', Q'^{\vee}|L'\big)$$

of degree $-n$, covering the other canonical quasi-isomorphism of differential Gerstenhaber algebras of Theorem 3.12.

Thus, the derived homs $\mathbb{R}\mathcal{H}om_S(Q|L, P|M)$ and $\mathbb{R}\mathcal{H}om_{\overline{S}'}(P'^{\vee}|M', Q'^{\vee}|L')$ are canonically quasi-isomorphic, up to a degree shift $n' - n$.

Proof. Let t and s' be as in the proof of Theorem 3.12. We need to construct quasi-isomorphisms of Batalin–Vilkovisky modules

$$(\psi,\delta):\left(\Omega^\bullet_C\otimes P\otimes Q'^\vee,t-s'\right)\longrightarrow\left(\Omega^\bullet_M\otimes P\otimes Q'^\vee,t\right)$$

and

$$(\psi',\delta'):\left(\Omega^\bullet_C\otimes P\otimes Q'^\vee,t-s'\right)\longrightarrow\left(\Omega^\bullet_{L'}\otimes P\otimes Q'^\vee,-s'\right).$$

(Note that because the elements of P and $Q'^\vee$ have degree zero, it is immaterial in which order we write the two factors P and $Q'^\vee$.) The case of (ψ',δ') being analogous, we will discuss only (ψ,δ).

Let us start with ψ. Denote the pullback of the orientation $\theta\in\Lambda^{n'}\widetilde{F}$ to C by the same letter, thus giving us a trivialization $\theta\in\Lambda^{n'}F$. Note that contracting $\alpha\in\Omega^\bullet_C$ with θ gives a form $\theta\lrcorner\alpha$ in the subbundle $\pi^*\Lambda E\subset\Omega^\bullet_C$; see (12). Recall the nondegenerate symmetric bilinear form $\beta:N_{M/C}\otimes_{\mathcal{O}_M}N_{M/C}\to\mathcal{O}_M$ of Remark 3.10. Since $F|_M=N_{M/C}$, we may apply the discriminant of β to $\theta|_M\otimes\theta|_M$ to obtain the nowhere-vanishing regular function

$$g=\det\beta\left(\theta|_M\otimes\theta|_M\right)\in\mathcal{O}_M \tag{23}$$

on M. The homomorphism ψ is now defined as the composition

$$\psi:\Omega^\bullet_C\xrightarrow{\theta\lrcorner\cdot}\pi^*\Lambda E\xrightarrow{\operatorname{res}|_M}\Lambda E|_M=\Omega^\bullet_M\xrightarrow{\cdot g}\Omega^\bullet_M.$$

Tensoring with $P\otimes Q'^\vee$, we obtain the quasi-isomorphism of differential graded modules

$$\psi:\left(\Omega^\bullet_C\otimes P\otimes Q'^\vee,t-s\right)\longrightarrow\left(\Omega^\bullet_M\otimes P\otimes Q'^\vee,t\right)$$

covering the morphism of differential graded algebras ϕ of (14). The formula for ψ is

$$\psi(\alpha)=g\cdot\lceil\theta\lrcorner\alpha\rceil_M. \tag{24}$$

("Ceiling brackets" denote restriction.) Note that $\deg\psi=-n'$.

Let us next construct $\delta:\Omega^\bullet_C\to\Omega^\bullet_M$. Recall the canonical partial flat connection on the Lagrangian foliation $\widetilde{F}$ on S':

$$\widetilde{\nabla}:\widetilde{F}\longrightarrow\widetilde{F}^\vee\otimes_{\mathcal{O}_{S'}}\widetilde{F},$$

defined by the requirement

$$\widetilde{\nabla}_{\widetilde{Y}}(\widetilde{X})\lrcorner\sigma'=\widetilde{Y}\lrcorner d(\widetilde{X}\lrcorner\sigma'),$$

for $\widetilde{Y},\widetilde{X}\in\widetilde{F}$. Since F is compatible with $\widetilde{F}$ via π', we get the pullback partial flat connection

$$\nabla:\pi'^*\widetilde{F}\longrightarrow F^\vee\otimes_{\mathcal{O}_C}\pi'^*\widetilde{F}.$$

Making the identification $F=\pi'^*\widetilde{F}$ we rewrite this partial connection as

$$\nabla:F\longrightarrow F^\vee\otimes_{\mathcal{O}_C}F.$$

It is characterized by the formula

$$\nabla_Y(X) \lrcorner \sigma = Y \lrcorner d(X \lrcorner \sigma)\,,$$

for $Y, X \in F$. We have written σ for the restriction of the symplectic form σ' to C. The dual connection

$$\nabla : F^\vee \longrightarrow F^\vee \otimes_{\mathcal{O}_C} F^\vee \tag{25}$$

satisfies

$$\nabla_Y(X' \lrcorner \sigma) = [Y, X'] \lrcorner \sigma\,,$$

for $Y \in F$ and $X' \in T_C$.

Recall that we also have the partial connection

$$\nabla : \pi^* E \longrightarrow F^\vee \otimes_{\mathcal{O}_C} \pi^* E \tag{26}$$

defined by $\nabla_Y \omega = Y \lrcorner d\omega$, for $Y \in F$ and $\omega \in \pi^* E \subset \Omega_C$. We used the dual of this connection in the proof of Theorem 3.12.

Thus, we have partial flat connections on F and $\pi^* E$, in such a way that the canonical homomorphism $F \to \pi^* E$ given by $X \mapsto X \lrcorner \sigma$ is flat. We hope there will be no confusion from using the same symbol ∇ for both partial connections. As usual, we get induced partial connections on all tensor operations involving F and $\pi^* E$. We define

$$\nabla^2 : \pi^* \Lambda E \longrightarrow F^\vee \otimes F^\vee \otimes \pi^* \Lambda E$$

as the composition (all tensor products are over $\mathcal{O}_C$)

$$\pi^* \Lambda E \xrightarrow{\nabla} F^\vee \otimes \pi^* \Lambda E \xrightarrow{\nabla} F^\vee \otimes F^\vee \otimes \pi^* \Lambda E\,.$$

We will also need

$$\nabla^3 : \mathcal{O}_C \longrightarrow F^\vee \otimes F^\vee \otimes F^\vee\,.$$

To simplify notation, let us assume that the closed 1-form $s - s'$ on C is exact. Let I be the ideal of M in $\mathcal{O}_C$. Then there exists a unique regular function $f \in I^2$ such that $df = s - s'$. The fact that f is in I^2 follows because s and s' vanish in $\Omega_C|_M$, so df vanishes in $\Omega_C|_M$. Then the Hessian of f is a symmetric bilinear form $N_{M/C} \otimes_{\mathcal{O}_M} N_{M/C} \to \mathcal{O}_M$, and is equal to $\beta|_M$, by Lemma 3.11.

Finally, we define $\delta : \Omega^\bullet_C \to \Omega^\bullet_M$ as a certain $\mathbb{C}$-linear combination of the two compositions

$$\begin{tikzcd}
\Omega^\bullet_C \arrow[r, "\theta \lrcorner \cdot"] & \pi^* \Lambda E \arrow[r, "\nabla^2"] & F^\vee \otimes_{\mathcal{O}_C} F^\vee \otimes_{\mathcal{O}_C} \pi^* \Lambda E \arrow[d, "\mathrm{res}|_M"] & & \\
& & N^\vee_{M/C} \otimes_{\mathcal{O}_M} N^\vee_{M/C} \otimes_{\mathcal{O}_M} \Omega^\bullet_M \arrow[r, "\eta"] & \Omega^\bullet_M \arrow[r, "\cdot g"] & \Omega^\bullet_M
\end{tikzcd}$$

and

$$\begin{tikzcd}
\pi^*\Lambda E \arrow[r, "\nabla"] & F^\vee \otimes \pi^*\Lambda E \arrow[r, "\nabla^3(f)\otimes\mathrm{id}"] & (F^\vee)^{\otimes 4} \otimes \pi^*\Lambda E \arrow[d, "\mathrm{res}\,|_M"] & & \\
\Omega_C^\bullet \arrow[u, "\theta \,\lrcorner\, \cdot"] & & (N^\vee)^{\otimes 4} \otimes \Omega_M^\bullet \arrow[r, "\eta\otimes\eta"] & \Omega_M^\bullet \arrow[r, "\cdot\, g"] & \Omega_M^\bullet .
\end{tikzcd}$$

Here η and $\eta \otimes \eta$ are the linear extensions of the map η from Remark 3.10. In fact, we define

$$\delta(\alpha) = -\tfrac{1}{2} g \cdot \eta\big(\lceil \nabla^2(\theta \,\lrcorner\, \alpha)\rceil_M\big) + \tfrac{1}{2} g \cdot (\eta \otimes \eta)\big(\lceil \nabla^3(f) \otimes \nabla(\theta \,\lrcorner\, \alpha)\rceil_M\big).$$

Since P and $Q'^\vee$ have flat connections on them, their pullbacks to C do, too. In particular, we can *partially* differentiate. Thus, δ extends naturally to the map $\delta : \Omega_C^\bullet \otimes P \otimes Q'^\vee \to \Omega_M \otimes P \otimes Q'^\vee$.

We need to check properties (i) and (ii) of Definition 1.24. To simplify notation, we will spell out only the case in which $P = Q' = (\mathcal{O}_C, d)$, leaving the general case to the reader.

Proving (ii) is a straightforward but tedious calculation using the properties of the partial connections on $\pi^*\Lambda E$, $\pi^*\Lambda E^\vee$, F, and $F^\vee$, in particular, compatibility with contraction. One can simplify this calculation using Remark 1.27: choose $\mathbb{C}$-algebra generators for ΛT_C in such a way that the generators of degree -1 are flat for the partial connection (see below). This reduces to checking (8) for the case in which X and Y are of degree 0, i.e., regular functions x and y on C. The claim is that

$$\delta(xy\omega) + xy\delta(\omega) + \{x, y\}\psi(\omega) = x\delta(y\omega) + y\delta(x\omega),$$

for all $\omega \in \Omega_C$. We leave the details to the reader, and only write down the terms containing $dx \otimes dy$ and only after canceling $g \cdot (\theta \,\lrcorner\, \omega)|_M$. In fact, from the term $\delta(xy\omega)$ we get the contribution

$$-\frac{1}{2}\eta\lceil dx \otimes dy + dy \otimes dx\rceil_M,$$

and from the term $\{x, y\}\psi(\omega)$ we get the contribution

$$\eta\lceil dx \otimes dy\rceil_M,$$

and these two expressions do indeed add up to 0, because η is symmetric.

To prove (i), we shall use Remark 1.28. We will carefully choose a local trivialization of the vector bundle T_C, since this will give local generators for ΛT_C as an $\mathcal{O}_C$-algebra.

The equations we wish to prove can be checked locally. So we pick a point $P \in M$ and pass to a sufficiently small analytic neighborhood of P in C.

Choose holomorphic functions $p_1, \ldots, p_{n'}$ in a neighborhood of $\pi'(P)$ in S' satisfying

(i) $p_1, \dots, p_{n'}$ cut out the submanifold $M' \subset S'$,

(ii) $dp_1, \dots, dp_{n'}$ form a frame of $\widetilde{F}^\vee$,

(iii) $dp_1, \dots, dp_{n'}$ are flat for the partial connection on $\widetilde{F}^\vee$,

(iv) $\theta \lrcorner (dp_1 \wedge \dots \wedge dp_{n'}) = 1$.

Denote the pullbacks of these functions to C by the same letters. Then these functions on C cut out M; their differentials are flat for the partial connection (25) and form a frame for $F^\vee$. Also, the last property remains true as written. Such p_i exist by Lemma 2.8.

Similarly, we choose holomorphic functions $x_1, \dots, x_n$ in a neighborhood of $\pi(P)$ in S such that $dx_1, \dots, dx_n$ form a frame for the subbundle $E \subset \Omega_S$. Then the dx_i are automatically flat for the partial connection on E. Again, we denote the pullbacks to C of these functions by the same letters. For the functions $x_1, \dots, x_n$ on C we have that their differentials $dx_1, \dots, dx_n$ form a flat frame for the subbundle $\pi^* E$ of Ω_C. In particular, the restrictions of $x_1, \dots, x_n$ to $M \subset C$ form a set of coordinates for M near P.

Then the union of these two families $dp_1, \dots, dp_{n'}, dx_1, \dots, dx_n$ forms a basis for Ω_C. We denote the dual basis (as usual) by $\frac{\partial}{\partial p_1}, \dots, \frac{\partial}{\partial p_{n'}}, \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}$. We have

$$\theta = (-1)^{\frac{1}{2}n'(n'-1)} \frac{\partial}{\partial p_1} \wedge \dots \wedge \frac{\partial}{\partial p_{n'}} .$$

We define

$$\omega^\circ = dp_1 \wedge \dots \wedge dp_{n'} \wedge dx_1 \wedge \dots \wedge dx_n ,$$

which is a basis for $\Omega_C^\bullet$ as a ΛT_C-module. Note that

$$\theta \lrcorner \omega^\circ = dx_1 \wedge \dots \wedge dx_n .$$

We denote the restriction to M of $dx_1 \wedge \dots \wedge dx_n$ by τ°. This is a basis for $\Omega_M^\bullet$ as a ΛT_M-module. We have

$$\lceil \theta \lrcorner \omega^\circ \rceil_M = \tau^\circ .$$

We now have to prove that for $X = \frac{\partial}{\partial p_i}$ and $X = \frac{\partial}{\partial x_j}$ we have

$$\psi\big(d(X \lrcorner \omega^\circ)\big) - (-1)^{n'} d\big(\psi(X \lrcorner \omega^\circ)\big) \\ = -2(-1)^{n'} t \wedge \delta(X \lrcorner \omega^\circ) + 2\,\delta\big((t-s') \wedge (X \lrcorner \omega^\circ)\big) . \quad (27)$$

For any of our values for X we have $d(X \lrcorner \omega^\circ) = 0$. So the first term in (27) always vanishes. Similarly, the third term always vanishes, because all of our values for X, as well as θ and ω°, are flat for the partial connection ∇. Thus only the second and fourth terms of (27) contribute.

Consider the fourth term. We have

$$\delta\big((t-s') \wedge (X \lrcorner \omega^\circ)\big) = \delta\big((t-s) \wedge (X \lrcorner \omega^\circ)\big) + \delta\big(df \wedge (X \lrcorner \omega^\circ)\big) \\ = \delta\big(\big(X \lrcorner (t-s)\big)\omega^\circ\big)\big) + \delta\big(X(f)\,\omega^\circ\big) .$$

Since $t - s$ is a section of $\pi^* E$ and is also a closed 1-form, $t - s$ is flat with respect to the partial connection ∇. Since X is by assumption also flat with respect to ∇, it follows that $\nabla\big(X \lrcorner (t-s)\big) = 0$, and hence the fourth term of (27) is equal to

$$\delta\big((t-s') \wedge (X \lrcorner \omega^\circ)\big) = \delta\big(X(f)\, \omega^\circ\big).$$

Thus, (27) reduces to

$$(-1)^{n'} d\big(\psi(X \lrcorner \omega^\circ)\big) = -2\, \delta\big(X(f)\, \omega^\circ\big). \tag{28}$$

Let us first consider the case that $X = \frac{\partial}{\partial p_i}$. In this case $\theta \wedge X = 0$, so that the left-hand side of (28) vanishes. The claim is therefore that

$$\delta\big(\tfrac{\partial f}{\partial p_i}\, \omega^\circ\big) = 0\,,$$

for all $i = 1, \ldots, n'$. This is equivalent to

$$\eta\big(\lceil \nabla^2 \tfrac{\partial f}{\partial p_i} \rceil_M\big) = (\eta \otimes \eta)\big(\lceil \nabla^3 f \rceil_M \otimes \lceil \nabla \tfrac{\partial f}{\partial p_i} \rceil_M\big)\,. \tag{29}$$

To check (29), let us write it out in coordinates. The right-hand side is equal to

$$\begin{aligned}\sum_{k,l} \sum_{m,n} \eta_{kl}\eta_{mn}\, \frac{\partial^3 f}{\partial p_k \partial p_l \partial p_m}\bigg|_{p=0} \frac{\partial^2 f}{\partial p_n \partial p_i}\bigg|_{p=0} &= \sum_{k,l} \sum_{m} \eta_{kl}\delta^i_m\, \frac{\partial^3 f}{\partial p_k \partial p_l \partial p_m}\bigg|_{p=0} \\ &= \sum_{k,l} \eta_{kl}\, \frac{\partial^3 f}{\partial p_k \partial p_l \partial p_i}\bigg|_{p=0}\,,\end{aligned}$$

which is indeed equal to the left-hand side of (29).

Now let us consider the case $X = \frac{\partial}{\partial x_j}$. Recall that the ideal I defining M is given by $I = (p_1, \ldots, p_{n'})$. Since $f \in I^2$, we still have $\frac{\partial f}{\partial x_j} \in I^2$ and hence $\nabla\left(\frac{\partial f}{\partial x_j}\right)|_{p=0} = 0$. Thus, the right-hand side of (28) is equal to

$$\begin{aligned}-2\, \delta\left(\tfrac{\partial f}{\partial x_j} \omega^\circ\right) &= g \cdot \eta\big(\lceil \nabla^2 \tfrac{\partial f}{\partial x_j} \rceil_M\big)\, \tau^\circ + 0 \\ &= g \cdot \operatorname{tr}\big(\eta \cdot \tfrac{\partial}{\partial x_j} H(f)\big)\, \tau^\circ\,,\end{aligned}$$

because differentiating with respect to x_j commutes with restriction to $M = \{p = 0\}$. (We have written $H(f)$ for the Hessian of f.) On the other hand, the left-hand side of (28) is equal to

$$(-1)^{n'} d\big(\psi(X \lrcorner \omega^\circ)\big) = \tfrac{\partial g}{\partial x_j}\, \tau^\circ\,,$$

and thus our final claim is equivalent to

$$\tfrac{\partial g}{\partial x_j} = g \cdot \operatorname{tr}\big(\eta \cdot \tfrac{\partial}{\partial x_i} H(f)\big)\,.$$

Recalling that $g = \det H(f)$ and that η is the inverse of $H(f)$, this claim follows from the following:

Claim. Let A be an invertible square matrix of regular functions on the manifold M. Then for every vector field X on M we have

$$(\det A)^{-1} X(\det A) = \operatorname{tr}\left(A^{-1} X(A)\right).$$

This last claim is both well known and easy to check. □

3.3 The oriented Batalin–Vilkovisky case

The setup is exactly the same as in Section 3.2, with one additional ingredient, namely, an orientation of C, i.e., a nowhere-vanishing global section $\omega_C^\circ \in \Omega_C^{n+n'}$. We require ω_C° to be compatible with the orientation $(\widetilde{F}, \widetilde{F}', \theta, \theta')$ on the symplectic correspondence $C \to S' \times S$ in the following sense: we ask that

(i) $\nabla\left(\theta \lrcorner \omega_C^\circ\right) = 0$, where $\nabla : \pi^* E \to F^\vee \otimes \pi^* E$ is the partial connection (26) on $F^\perp = \pi^* E$ defined by the foliation F of C;

(ii) $\nabla'\left(\theta' \lrcorner \omega_C^\circ\right) = 0$, where $\nabla' : \pi'^* E' \to F'^\vee \otimes \pi'^* E'$ is the corresponding partial connection defined by the foliation F' of C.

Now, $\left(\theta \lrcorner \omega_C^\circ\right)|_M$ is an orientation of M (recalling that $(\pi^* E)|_M = \Omega_M$). We shall denote it by ω_M°. Similarly, $\left(\theta' \lrcorner \omega_C^\circ\right)|_{L'}$ is an orientation of L', which we shall denote by $\omega_{L'}^\circ$. This orients the three Lagrangian intersections in Theorem 3.12.

Theorem 3.16. *The quasi-isomorphisms of differential Gerstenhaber algebras of Theorem 3.12 are canonically enhanced to quasi-isomorphisms of differential Batalin–Vilkovisky algebras*

$$(M' \times L) \Cap^\circ_{S' \times S} C \longrightarrow L \Cap^\circ_S M$$

and

$$(M' \times L) \Cap^\circ_{S' \times S} C \longrightarrow M' \Cap^\circ_{S'} L'.$$

In particular, the oriented derived intersections $L \Cap^\circ_S M$ and $M' \Cap^\circ_{S'} L'$ are canonically quasi-isomorphic.

Proof. In view of Theorems 3.12 and 3.15 and the results of Section 1.6, we only need to check that

(i) $\psi\left(\omega_C^\circ\right) = \omega_M^\circ$,

(ii) $\psi'\left(\omega_C^\circ\right) = \omega_{L'}^\circ$,

(iii) $\delta\left(\omega_C^\circ\right) = 0$,

(iv) $\delta'\left(\omega_C^\circ\right) = 0$,

where (ψ, δ) and (ψ', δ') are the homomorphisms of differential Batalin–Vilkovisky modules constructed in the proof of Theorem 3.15. But the first two follow from the above definitions and the last two from the above assumptions. □

Remark 3.17. If $C = S$ and $C \to S \times S$ is the diagonal, a canonical choice for the orientation of C is $\omega_C^\circ = \sigma^n$, by Remark 2.4. In this case, we also have $\omega_M^\circ = \theta|_M$, via the identification $\Omega_M^n = \Lambda^n N_{M/S} = \Lambda^n F|_M$. Similarly, $\omega_{L'}^\circ = \theta'|_{L'}$.

4 The Gerstenhaber structure on $\mathcal{T}or$ and the Batalin–Vilkovisky structure on $\mathcal{E}xt$

4.1 The Gerstenhaber algebra structure on $\mathcal{T}or$

Let L and M be immersed Lagrangians in the symplectic manifold S. Write $\mathcal{T}or^i_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M) = \mathcal{T}or^{\mathcal{O}_S}_{-i}(\mathcal{O}_L, \mathcal{O}_M)$. The direct sum

$$\mathcal{T}or^\bullet_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M) = \bigoplus_i \mathcal{T}or^i_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$$

is a graded sheaf of $\mathcal{O}_S$-algebras, concentrated in nonpositive degrees.

Remark 4.1. To be precise, we have to use the analytic étale topology on S to be able to think of $\mathcal{T}or^\bullet_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$ as a sheaf of $\mathcal{O}_S$-algebras. If L and M are embedded, not just immersed, we can use the usual analytic topology. Alternatively, introduce the fibered product $Z = L \times_S M$ and think of $\mathcal{T}or^\bullet_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$ as a sheaf of graded $\mathcal{O}_Z$-algebras.

Theorem 4.2. *There exists a unique bracket of degree* $+1$ *on* $\mathcal{T}or^\bullet_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$ *such that*

(i) $\mathcal{T}or^\bullet_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$ *is a sheaf of Gerstenhaber algebras;*

(ii) whenever E *is a (local) polarization of* S *such that* L *and* M *are transverse to* E*, then this sheaf of Gerstenhaber algebras is obtained from the derived intersection* $L \Cap_S M$ *(defined with respect to* E*) by passing to cohomology.*

Proof. Without loss of generality, assume that L and M are submanifolds. For every point of S we can find an open neighborhood in S over which we can choose a polarization E that is transverse to L and M. This proves uniqueness.

For existence, we have to prove that any two polarizations E, E'' give rise to the same bracket on $\mathcal{T}or^\bullet_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$. This is a local question, so we may choose a third polarization E' that is transverse to both E and E'', and also to L and M.

We will apply the exchange property, Theorem 3.12, twice, first to the symplectic correspondence $\Delta : C = S \to S \times S$ between the polarized symplectic manifolds (S, E') and (S, E), then to the symplectic correspondence $\Delta : C = \overline{S} \to \overline{S} \times \overline{S}$ between $(\overline{S}, E'')$ and $(\overline{S}, E')$. We obtain the following diagram of quasi-isomorphisms of sheaves of differential Gerstenhaber algebras:

$$
\begin{tikzcd}
 & (M \times L) \Cap^{E' \times E}_{\overline{S} \times S} S \arrow[r] \arrow[d] & L \Cap^{E}_{S} M \\
(L \times M) \Cap^{E'' \times E'}_{S \times \overline{S}} \overline{S} \arrow[r] \arrow[d] & M \Cap^{E'}_{\overline{S}} L & \\
L \Cap^{E''}_{S} M & &
\end{tikzcd}
\tag{30}
$$

We have included the polarizations defining the derived intersections in the notation.

Passing to cohomology sheaves, we obtain the following diagram of isomorphisms of sheaves of Gerstenhaber algebras:

$$
\begin{tikzcd}
 & \mathcal{T}or^{\bullet}_{\mathcal{O}_{S \times S}}(\mathcal{O}_{M \times L}, \mathcal{O}_S) \arrow[r] \arrow[d] & \mathcal{T}or^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M) \\
\mathcal{T}or^{\bullet}_{\mathcal{O}_{S \times S}}(\mathcal{O}_{L \times M}, \mathcal{O}_S) \arrow[r] \arrow[d] & \mathcal{T}or^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_M, \mathcal{O}_L) & \\
\mathcal{T}or^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M) & &
\end{tikzcd}
$$

One checks that all these morphisms are the *canonical* ones, and hence that the composition of all four of them is the *identity* on $\mathcal{T}or^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$. If the identity preserves the two brackets on $\mathcal{T}or^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$ defined by E and E'', respectively, then the two brackets are equal. □

4.2 The Batalin–Vilkovisky structure on $\mathcal{E}xt$

Let L and M continue to denote immersed Lagrangians in the symplectic manifold S. Furthermore, let P be a local system on M, and Q a local system on L. The direct sum

$$\mathcal{E}xt^{\bullet}_{\mathcal{O}_S}(Q, P) = \bigoplus_i \mathcal{E}xt^{i}_{\mathcal{O}_S}(Q, P)$$

is a graded sheaf of $\mathcal{T}or^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$-modules.

Theorem 4.3. *There exists a unique* $\mathbb{C}$*-linear differential*

$$d : \mathcal{E}xt^{i}_{\mathcal{O}_S}(Q, P) \longrightarrow \mathcal{E}xt^{i+1}_{\mathcal{O}_S}(Q, P)$$

(for all i) such that

(i) $\mathcal{E}xt^{\bullet}_{\mathcal{O}_S}(Q, P)$ *is a sheaf of Batalin–Vilkovisky modules over the Gerstenhaber algebra* $\mathcal{T}or^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$*;*

(ii) whenever E is a (local) polarization of S such that L and M are transverse to E, and $\overline{Q}$ is a local system on S restricting to Q on L, this sheaf of Batalin–Vilkovisky modules is obtained from the derived hom $\mathbb{R}\mathcal{H}om_S(\overline{Q}|L, P|M)$ (defined with respect to E) by passing to cohomology.

Proof. Uniqueness is clear. Let us prove existence. For this, we assume that we are given two polarizations E, E'', transverse to L and M, and two extensions $\overline{Q}$ and $\widehat{Q}$ of Q to S. To compare the derived homs $\mathbb{R}\mathcal{H}om_S^E(\overline{Q}|L, P|M)$ and $\mathbb{R}\mathcal{H}om_S^{E''}(\widehat{Q}|L, P|M)$, we choose (locally) a third polarization E', transverse to L and M, and E and E'', and an extension $\overline{P}$ of P to S. These choices make the five derived homs in diagram (32), below, well-defined.

To define the homomorphisms of differential Batalin–Vilkovisky modules in (32), we orient the symplectic correspondence given by the diagonal of S in the canonical way, as in Remark 3.17, by σ^n. The corresponding symplectic correspondence given by the diagonal of $\overline{S}$ is hence oriented by $(-1)^n\sigma^n$. We also orient the three Lagrangian foliations F, F', F'' on S, by choosing $\theta \in \Lambda^n F$, $\theta' \in \Lambda^n F'$, and $\theta'' \in \Lambda^n F''$. (Note that $\widetilde{F} = F = E^\perp$, etc., in our case.) But the choice of θ, θ'' is not completely arbitrary. In fact, notice that both $F|_L$ and $F''|_L$ are complements to $T_L \subset T_S|_L$, so that we get a canonical identification $F|_L \xrightarrow{\sim} F''|_L$. We choose θ and θ'' in such a way that the composition

$$\mathcal{O}_L \xrightarrow{\theta|_L} \det F|_L \xrightarrow{\cong} \det F''|_L \xrightarrow{\theta''^\vee|_L} \mathcal{O}_L \tag{31}$$

is equal to the identity.

Now, by applying the exchange property, Theorem 3.15, twice, as in the proof of Theorem 4.2, we obtain the following diagram of quasi-isomorphisms of differential Batalin–Vilkovisky modules, covering diagram (30) of differential Gerstenhaber algebras:

$$\begin{tikzcd}
\mathbb{R}\mathcal{H}om^{E'\times E}_{\overline{S}\times S}\big(\mathcal{O}|(M\times L), (\overline{P}\otimes\overline{Q}^\vee)|S\big) \arrow[r] \arrow[dr] & \mathbb{R}\mathcal{H}om^E_S(\overline{Q}|L, P|M) \\
\mathbb{R}\mathcal{H}om^{E''\times E'}_{S\times\overline{S}}\big(\mathcal{O}|(L\times M), (\widehat{Q}^\vee, \otimes\overline{P})|S\big) \arrow[r] \arrow[dr] & \mathbb{R}\mathcal{H}om^{E'}_{\overline{S}}(\overline{P}^\vee|M, Q^\vee|L) \\
& \mathbb{R}\mathcal{H}om^{E''}_S(\widehat{Q}|L, P|M)
\end{tikzcd} \tag{32}$$

When passing to cohomology, the first and the last items in this diagram are both equal to $\mathcal{E}xt^\bullet_{\mathcal{O}_S}(Q, P)$. We claim that the induced isomorphism on cohomology is equal to the identity. For simplicity, we will prove this for the case that P and Q are the trivial rank-one local systems. Then the differential Batalin–Vilkovisky modules of diagram (32) are invertible. We orient them using σ^n, $(-1)^n\sigma^n$ and $(\theta \lrcorner \sigma^n)|_M$, $(\theta' \lrcorner \sigma^n)|_L$ and $(\theta'' \lrcorner \sigma^n)|_M$, respectively, as in Section 3.3. Then the homomorphisms in diagram (32) do not preserve orientations according to Definition 1.37, because of the presence of the functions g, defined in (23), entering into the definition of ψ, equation (24).

Let us call these functions, from the top to the bottom, g_1, g_2, g_3, g_4. We also need more detailed notation for the various maps β of Remark 3.9 and introduce

$$\beta_{ij} : F^{(i)} \to T_S \xrightarrow{\lrcorner\sigma} \Omega_S \to F^{(j)\vee},$$

where $i, j = 0, 1, 2$ denotes the number of primes on the letter F. Using similar notation, we introduce the functions

$$h_{ij} = (\theta^{(j)})^\vee \circ \det \beta_{ij} \circ \theta^{(i)}.$$

These are functions on S, invertible where they are defined.

On the submanifold M, we have canonical isomorphisms $\alpha_{ij} : F^{(i)}|_M \to F^{(j)}|_M$ and functions

$$a_{ij} = (\theta^{(j)})^{-1} \circ \det \alpha_{ij} \circ \theta^{(i)}.$$

Similarly, on L, we have canonical isomorphisms $\gamma_{ij} : F^{(i)}|_L \to F^{(j)}|_L$ and functions

$$c_{ij} = (\theta^{(j)})^{-1} \circ \det \alpha_{ij} \circ \theta^{(i)}.$$

With this notation we now have

$$\begin{aligned} g_1 &= h_{01}a_{01}, \\ g_2 &= h_{10}c_{10}, \\ g_3 &= h_{12}c_{12}, \\ g_4 &= h_{21}a_{21}. \end{aligned}$$

Hence the failure of the maps in (32) to preserve orientations is given by the product

$$\frac{g_2\big((-1)^n g_4\big)}{g_1\big((-1)^n g_3\big)} = \frac{h_{10}c_{10}h_{21}a_{21}}{h_{01}a_{01}h_{12}c_{12}} = \frac{c_{10}a_{21}}{a_{01}c_{12}},$$

noting that h_{ij} is dual, and hence equal, to h_{ji}.

Now note that we have two orientations on M, namely $(\theta \lrcorner \sigma^n)|_M$ and $(\theta'' \lrcorner \sigma^n)|_M$. On $\mathcal{E}xt^\bullet_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$, this difference induces a factor of a_{20}. Thus, to prove our claim, we need to show that

$$a_{20} = \frac{c_{10}a_{21}}{a_{01}c_{12}}.$$

Now, it is clear that $a_{ij}a_{jk} = a_{ik}$, and $c_{ij}c_{jk} = c_{ik}$. Thus, our claim is equivalent to

$$c_{20} = 1,$$

which is true, because c_{02} is the homomorphism of diagram (31), which is the identity, by assumption. $\square$

4.3 Oriented case

Theorem 4.4. *Let L and M be immersed Lagrangians of the symplectic manifold S of dimension $2n$. Then every orientation of M defines a generator for the bracket of Theorem 4.2. More precisely, every trivialization ω_M° of Ω_M^n defines a differential d° on $\mathcal{T}or^\bullet_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$ making the latter a sheaf of Batalin–Vilkovisky algebras.*

Proof. From Example 1.36 and Proposition 1.35, we get that ω_M° defines a differential d° on $\mathcal{T}or^\bullet_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$. We have to show that d° does not depend on the polarization. For this we repeat the proof of Theorem 4.3, making sure that all morphisms of differential Batalin–Vilkovisky modules preserve orientations. For this, we have to be more careful with our choices. Of course, F and F'', the two polarizations to be compared, are given. But we will choose F' in a special way, as follows.

First note that on L, both $F|_L$ and $F''|_L$ are complements to T_L inside $T_S|_L$. Thus we obtain an isomorphism $\phi : F|_L \to F''|_L$, characterized by $\phi(X) - X \in T_L$, for all $X \in F|_L$. There exists a canonical subbundle $H \subset T_S|_L$ such that H is complementary to $F|_L$, $F''|_L$, and T_L, and the isomorphism $\widetilde{\phi} : F|_L \to F''|_L$, characterized by $\widetilde{\phi}(X) - X \in H$ is equal to $-\phi$. (Essentially, H is obtained by negating the F''-components of the vectors in T_L, but preserving their F-components.) The subbundle $H \subset T_S|_L$ is isotropic, so we can extend it, at least locally, to a Lagrangian subbundle $F' \subset T_S$. With this choice we will have

$$\frac{h_{10}}{h_{12}} = \left(\frac{h_{10}}{h_{12}}\right)^\vee = (-1)^n c_{02},$$

and hence

$$g_2 = (-1)^n g_3.$$

Now, finally, we choose first an orientation ω_L° of L and then θ and θ' in such a way that

$$\begin{aligned} g_1(\theta \lrcorner \sigma^n)|_M &= \omega_M^\circ, \\ g_2(\theta' \lrcorner \sigma^n)|_M &= \omega_L^\circ. \end{aligned}$$

Then, by the choice of F', we have

$$g_3\big(\theta' \lrcorner (-1)^n \sigma^n\big)|_M = \omega_L^\circ\,.$$

We choose θ'' in such a way that $c_{02} = 1$, as above. Then $h_{10} = (-1)^n h_{12}$, and hence $h_{01} = (-1)^n h_{21}$ and $h_{01} a_{01} a_{20} = (-1)^n h_{21} a_{21}$. In other words, $g_1 a_{20} = (-1)^n g_4$, or

$$g_4\big(\theta'' \lrcorner (-1)^n \sigma^n\big)|_M = g_1(\theta \lrcorner \sigma^n)|_M = \omega_M^\circ\,.$$

Now all four homomorphisms of diagram (32) preserve orientations, and hence they are equal to the morphisms of diagram (30). This finishes the proof. □

Corollary 4.5. *In the nonoriented case, the sheaf of Gerstenhaber algebras* $\mathcal{T}or^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$ *is locally a Batalin–Vilkovisky algebra, albeit in a noncanonical way.*

4.4 The exchange property

Let S and S' be complex symplectic manifolds, of dimensions $2n$, $2n'$, respectively.

Definition 4.6. A symplectic correspondence $C \to S' \times S$ is called **regular** if for every point $P \in C$ one can find polarizations $E \subset \Omega_S$, defined in a neighborhood of $\pi(P) \in S$, and $E' \subset \Omega_{S'}$, defined in a neighborhood of $\pi'(P) \in S'$, such that

(i) C is transverse to ${E'}^{\perp} \times E^{\perp}$ inside $S' \times S$;

(ii) the induced foliations F, F' on C descend to foliations $\widetilde{F}$, $\widetilde{F}'$ on S' and S, respectively, as in Section 3.2.

Theorem 4.7. *Let* $C \to S' \times S$ *be a regular symplectic correspondence. Let* $L \to S$ *be an immersed Lagrangian transverse to* C *and* $M' \to S'$ *an immersed Lagrangian transverse to* C*. Then there is a canonical isomorphism of sheaves of Gerstenhaber algebras*

$$\mathcal{T}or^{\bullet}_{\mathcal{O}_{S'}}(\mathcal{O}_{L'}, \mathcal{O}_{M'}) = \mathcal{T}or^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)\,,$$

with notation as in (5.2).

If L*,* M' *and* C *are oriented, then this is an isomorphism of sheaves of Batalin–Vilkovisky algebras.*

Proof. We apply the exchange property, Theorem 3.12, twice, first to $C \to S' \times S$, then to $S' \to S' \times S'$. □

Theorem 4.8. *Let* $C \to S' \times S$ *be a regular symplectic correspondence. Let* $L \to S$ *be an immersed Lagrangian transverse to* C *and* $M' \to S'$ *an immersed Lagrangian transverse to* C*. Let* P' *be a local system on* M' *and* Q *a local system on* L*. Then there is a canonical isomorphism of sheaves of Batalin–Vilkovisky modules*

$$\mathcal{E}xt^{\bullet}_{\mathcal{O}_{S'}}(Q|_{L'}, P') = \mathcal{E}xt^{\bullet}_{\mathcal{O}_S}(Q, P'|_M)$$

covering the isomorphism of sheaves of Gerstenhaber algebras of Theorem 4.7.

Proof. We apply the exchange property, Theorem 3.15, twice, first to $C \to S' \times S$, then to $S' \to S' \times S'$. The details are similar to the proof of Theorem 4.3. □

An example

Let M be a complex manifold and $S = \Omega_M$ the cotangent bundle with its canonical symplectic structure. There are two typical examples of immersed Lagrangians:

(i) the graph of a closed 1-form $\omega \in \Gamma(M, \Omega_M)$, which we denote by $\Gamma_\omega \subset \Omega_M$. This is in fact embedded.

(ii) the conormal bundle $C_{Z/M} \to \Omega_M$, where $Z \to M$ is an immersion of complex manifolds, i.e., a holomorphic map, injective on tangent spaces.

Given one of each, we consider the Lagrangian intersection $\Gamma_\omega \cup C_{Z/M}$. Note that it is supported on $Z(\omega) \cap Z$. We use the notation

$$\mathcal{T}_M(\omega, Z) = \mathcal{T}or^\bullet_{\mathcal{O}_{\Omega_M}}(\mathcal{O}_{\Gamma_\omega}, \mathcal{O}_{C_{Z/M}})$$

and

$$\mathcal{E}_M(\omega, Z) = \mathcal{E}xt^\bullet_{\mathcal{O}_{\Omega_M}}(\mathcal{O}_{\Gamma_\omega}, \mathcal{O}_{C_{Z/M}}).$$

Let $f : M \to N$ be a holomorphic map between complex manifolds M, N. Consider the symplectic manifolds $S' = \Omega_M$ and $S = \Omega_N$. The pullback vector bundle $f^*\Omega_N$ is then a symplectic correspondence C:

$$\begin{tikzcd} f^*\Omega_N \arrow[r] \arrow[d] & \Omega_N \\ \Omega_M & \end{tikzcd}$$

Let us assume that $f^*\Omega_N \to \Omega_M$ fits into a short exact sequence of vector bundles

$$0 \longrightarrow K \longrightarrow f^*\Omega_N \longrightarrow \Omega_M \longrightarrow \Omega_{M/N} \longrightarrow 0\,.$$

Then the symplectic correspondence $C = f^*\Omega_N$ is regular.

If $Z \to M$ is an immersion (i.e., injective on tangent spaces) such that the composition $Z \to N$ is also an immersion, then the conormal bundle $C_{Z/M} \to \Omega_M$ is an immersed Lagrangian transverse to $f^*\Omega_N$. The corresponding immersed Lagrangian of Ω_N is the conormal bundle $C_{Z/N}$.

If $\omega \in \Gamma(N, \Omega_N)$ is a closed 1-form, then its graph is a Lagrangian submanifold of Ω_N, which is automatically transverse to $f^*\Omega_N$. The corresponding Lagrangian submanifold of Ω_M is the graph of the pullback form $f^*\omega$.

Corollary 4.9. *There is a canonical isomorphism of Gerstenhaber algebras with Batalin–Vilkovisky modules*

$$\mathcal{T}_N(\omega, Z) = \mathcal{T}_M(f^*\omega, Z)\,, \qquad \mathcal{E}_N(\omega, Z) = \mathcal{E}_M(f^*\omega, Z)\,.$$

5 Further remarks

5.1 Virtual de Rham cohomology

Let M and L be Lagrangian submanifolds of the complex symplectic manifold S. Let X be their scheme-theoretic intersection. Let $\mathcal{E} = \mathcal{E}xt^{\bullet}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M)$ be endowed with the differential d from Section 4.2. The sheaf $\mathcal{E}$ is a coherent $\mathcal{O}_X$-module; the differential d is $\mathbb{C}$-linear.

Definition 5.1. We call $(\mathcal{E}, d)$ the **virtual de Rham complex** of X.

Theorem 5.2. *The complex $(\mathcal{E}, d)$ is constructible.*

Proof. The claim is local in X, so we may assume that the symplectic manifold S is the cotangent bundle of the manifold M, that the first Lagrangian M is the zero section, and that the second Lagrangian L is the graph of an exact 1-form df, where $f : M \to \mathbb{C}$ is a holomorphic function. In this form, the theorem was proved by Kapranov; see the remarks toward the bottom of page 72 in [2]. □

Corollary 5.3. *The hypercohomology group $\mathbb{H}^p\big(X, (\mathcal{E}, d)\big)$ is finite-dimensional. Moreover, for $Z \subset X$ Zariski closed, $\mathbb{H}^p_Z\big(X, (\mathcal{E}, d)\big)$ is also finite-dimensional.*

By abuse of notation, we will write $\mathbb{H}^p(X, \mathcal{E})$ and $\mathbb{H}^p_Z(X, \mathcal{E})$, instead of $\mathbb{H}^p\big(X, (\mathcal{E}, d)\big)$ and $\mathbb{H}^p_Z\big(X, (\mathcal{E}, d)\big)$, respectively.

Definition 5.4. We call the hypercohomology group $\mathbb{H}^p(X, \mathcal{E})$ the pth **virtual de Rham cohomology** group of the Lagrangian intersection X.

Corollary 5.5. *The function*

$$P \longmapsto \sum_i (-1)^i \dim_{\mathbb{C}} \mathbb{H}^i_{\{P\}}(X, \mathcal{E})$$

is a constructible function $\chi : X \to \mathbb{Z}$.

We may think of $\chi : X \to \mathbb{Z}$ as the fiberwise Euler characteristic of the constructible complex $(\mathcal{E}, d)$.

5.2 A speculation in Hodge theory

Remark 5.6. There is the standard spectral sequence of hypercohomology

$$E_1^{pq} = H^q(X, \mathcal{E}^p) \Longrightarrow \mathbb{H}^{p+q}\big(X, (\mathcal{E}, d)\big). \tag{33}$$

This should be viewed as a generalization of the Hodge to de Rham spectral sequence.

There is also the usual local to global spectral sequence

$$E_2^{p,q} = H^p(X, \mathcal{E}^q) \Longrightarrow \operatorname{Ext}^{p+q}_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M). \tag{34}$$

One may speculate to what extent these spectral sequences degenerate.

Example 5.7. For example, if M is a manifold and $S = \Omega_M$ the cotangent bundle with its standard symplectic structure, and we consider the intersection of M (the zero section) with itself, we get

$$\mathcal{E}xt^p_{\mathcal{O}_S}(\mathcal{O}_M, \mathcal{O}_M) = \Omega^p_M .$$

Moreover, $(\mathcal{E}, d) = (\Omega^\bullet_M, d)$ is the de Rham complex of M, and Lagrangian intersection cohomology is equal to de Rham cohomology of M. Thus the spectral sequence (33) is the usual Hodge to de Rham spectral sequence:

$$E_1^{pq} = H^q(M, \Omega^p) \Longrightarrow \mathbb{H}^{p+q}\big(M, (\Omega^\bullet_M, d)\big).$$

On the other hand, we have

$$\mathrm{Ext}^i_{\mathcal{O}_S}(\mathcal{O}_M, \mathcal{O}_M) = \bigoplus_{p+q=i} H^i(M, \Omega^q),$$

in other words, the E_2-term of the spectral sequence (34) is equal to the abutment. Thus $\mathrm{Ext}^i_{\mathcal{O}_S}(\mathcal{O}_M, \mathcal{O}_M)$ is equal to Hodge cohomology of M.

We may, then, rewrite the Hodge to de Rham spectral sequence (33) as

$$\mathrm{Ext}^i_{\mathcal{O}_S}(\mathcal{O}_M, \mathcal{O}_M) \Longrightarrow \mathbb{H}^i\big(M, (\mathcal{E}, d)\big).$$

This, of course, degenerates if M is proper and gives the equality

$$\mathrm{Ext}^i_{\mathcal{O}_S}(\mathcal{O}_M, \mathcal{O}_M) = \mathbb{H}^i\big(M, (\mathcal{E}, d)\big), \tag{35}$$

if M is Kähler, by Hodge theory.

The following conjecture is thus a generalization of Hodge theory:

Conjecture 5.8. *Under sufficiently strong hypotheses, including certainly that the intersection $X = L \cap M$ is* complete *and some analogue of the Kähler condition, for example that X is projective, we have*

$$\mathbb{H}^p\big(X, (\mathcal{E}, d)\big) = \mathrm{Ext}^p_{\mathcal{O}_S}(\mathcal{O}_L, \mathcal{O}_M) .$$

5.3 A differential graded category

Let S be a symplectic variety and $\mathfrak{U} = \{U_i\}$ an affine open cover of S. Construct a category $\mathcal{A}$ as follows: objects of $\mathcal{A}$ are pairs (M, P), where M is a Lagrangian submanifold of S and P is a local system on M. We do not assume that $M \to S$ is a closed immersion, it suffices that this map be affine. We often omit the first component of such a pair (M, P) from the notation.

For objects (M, P) and (L, Q) of $\mathcal{A}$ we define $\mathrm{Hom}_{\mathcal{A}}(Q, P)$ to be the total complex associated to the double complex

$$C^\bullet\big(\mathfrak{U}, (\mathcal{E}^\bullet, d)\big)$$

given by Cech cochains with respect to the cover $\mathfrak{U}$ with values in the virtual de Rham complex $\mathcal{E} = \mathcal{E}xt^\bullet_{\mathcal{O}_S}(Q, P)$ (endowed with the differential from Theorem 4.3).

Theorem 5.9. *This defines a differential graded category.*

Proof. For simplicity of notation, we deal only with Lagrangian submanifolds M, N, and L, leaving the generalization to local systems to the reader. There are natural Yoneda pairings (all tensors and $\mathcal{E}xt$'s are over $\mathcal{O}_S$)

$$\mathcal{E}xt^i(\mathcal{O}_M, \mathcal{O}_N) \otimes \mathcal{E}xt^j(\mathcal{O}_N, \mathcal{O}_L) \longrightarrow \mathcal{E}xt^{i+j}(\mathcal{O}_M, \mathcal{O}_L).$$

We need to show that these are compatible with the canonical differential of Section 4.2. This is a local question, so we may assume that S has three Lagrangian foliations F, F', and F'', all transverse to each other, and all transverse to M, N, L. Let s, t', u'' be the Euler sections of M, N, L with respect to F, F', and F'', respectively. Then we can represent $\mathcal{E}xt^\bullet(\mathcal{O}_M, \mathcal{O}_N)$ as the cohomology of $(\Omega^\bullet_S, s - t')$, and $\mathcal{E}xt^\bullet(\mathcal{O}_N, \mathcal{O}_L)$ as the cohomology of $(\Omega^\bullet_S, t' - u'')$, and $\mathcal{E}xt^\bullet(\mathcal{O}_M, \mathcal{O}_L)$ as the cohomology of $(\Omega^\bullet_S, s - u'')$. So the claim will follow if we can produce a morphism of complexes

$$(\Omega^\bullet_S, s - t') \otimes (\Omega^\bullet_S, t' - u'') \longrightarrow (\Omega^\bullet_S, s - u'') \ .$$

But this is easy: just take the cup product. □

Remark 5.10. The cohomology groups of the hom-spaces in $\mathcal{A}$ are the virtual de Rham cohomology groups of Lagrangian intersections.

Remark 5.11. The category $\mathcal{A}$ does not depend on the affine cover $\mathfrak{U}$ in any essential way.

Remark 5.12. Of course, it is tempting to speculate on relations of $\mathcal{A}$ to the Fukaya category of S. We will leave this to future research.

5.4 Relation to vanishing cycles

Let S be a complex symplectic manifold of dimension $2n$. Let L, M be Lagrangian submanifolds, and $X = L \cap M$ their intersection.

In [1], we introduced for any scheme X a constructible function $\nu_X : X \to \mathbb{Z}$. The value $\nu_X(P)$ is an invariant of the singularity (X, P). In our context, the singularity (X, P) is the critical set of a holomorphic function $f : M \to \mathbb{C}$, locally defined near $P \in M$. Hence (see [1]), the invariant $\nu_X(P)$ is equal to the Milnor number of f at P, i.e., we have

$$\nu_X(P) = (-1)^n\big(1 - \chi(F_P)\big) \,,$$

where F_P is the Milnor fiber of f at the point P.

Conjecture 5.13. *We have $\chi(P) = \nu_X(P)$.*

This conjecture would follow from Remark 2.12 (b) of [2]. Note that Kapranov refers to this as a *fact*, which is *not obvious*, although *probably not very difficult*.

Conjecture 5.14. *If the intersection X is compact, so that the intersection number $\#^{\mathrm{vir}}(X)$ is well-defined, we have*

$$\#^{\mathrm{vir}}(X) = \sum_i (-1)^i \dim \mathbb{H}^i(X, \mathcal{E}),$$

i.e., the intersection number is equal to the **virtual Euler characteristic** *of X, defined in terms of virtual de Rham cohomology.*

To see that Conjecture 5.13 implies Conjecture 5.14, recall from [1] that the intersection X has a symmetric obstruction theory. The main result of [1] implies that $\#^{\mathrm{vir}}(X) = \chi(X, \nu_X)$. But if $\chi = \nu_X$, then $\chi(X, \nu_X) = \chi(X, \chi) = \sum_i (-1)^i \dim \mathbb{H}^i(X, \mathcal{E})$.

Remark 5.15. If S is the cotangent bundle of M, and L is the graph of df, where $f : M \to \mathbb{C}$ is a holomorphic function, then the Lagrangian intersection $X = L \cap M$ is the critical set of f. Thus X carries the perverse sheaf of vanishing cycles Φ_f. In [2], Kapranov constructs, at least conjecturally, a spectral sequence whose E_2-term is $(\mathcal{E}, d)$ and whose abutment is, in some sense, Φ_f.

Conjecture 5.16. *In the general case of a Lagrangian intersection $X = L \cap M$ inside a complex symplectic manifold S, we conjecture the existence of a natural perverse sheaf on X that locally coincides with the perverse sheaf of vanishing cycles of Remark 5.15. There should be a spectral sequence relating $(\mathcal{E}, d)$ to this perverse sheaf of vanishing cycles. We believe that [3] may be related to this question. This conjecture, in some sense, categorifies Conjecture 5.13.*

References

1. K. Behrend, *Donaldson-Thomas invariants via microlocal geometry*, `math.AG/0507523`, July 2005.
2. M. Kapranov, *On DG-modules over the de Rham complex and the vanishing cycles functor*, Algebraic Geometry (Chicago, 1989), Lecture Notes in Math., **1479**, Springer, Berlin, 1991, 57–86.
3. M. Kashiwara, P. Schapira, *Constructibility and duality for simple holonomic modules on complex symplectic manifolds*, Amer. J. Math., **130** (2008), no. 1, 207–237.
4. D. Tamarkin, B. Tsygan, *The ring of differential operators on forms in noncommutative calculus*: *Graphs and patterns in mathematics and theoretical physics*, Proc. Sympos. Pure Math. **73**, Amer. Math. Soc., Providence, RI, 2005, 105–131.

A Non-Archimedean Interpretation of the Weight Zero Subspaces of Limit Mixed Hodge Structures

Vladimir G. Berkovich

Department of Mathematics, Weizmann Institute, 76100 Rehovot, Israel
vladimir.berkovich@weizmann.ac.il

To Yuri Ivanovich Manin on the occasion of his 70th birthday

Summary. Let $\mathcal{X}$ be a proper scheme over the field F of functions meromorphic in an open neighborhood of zero in the complex plane. The scheme $\mathcal{X}$ gives rise to a proper morphism of complex analytic spaces $\mathcal{X}^h \to D^*$ and, if the radius of the open disc D is sufficiently small, the cohomology groups of the fibers $\mathcal{X}_t^h$ at points $t \in D^*$ form a variation of mixed Hodge structures on D^*, which admits a limit mixed Hodge structure. The purpose of the paper is to construct a canonical isomorphism between the weight zero subspace of this limit mixed Hodge structure and the rational cohomology group of the non-Archimedean analytic space $\mathcal{X}^{\mathrm{an}}$ associated to the scheme $\mathcal{X}$ over the completion of the field F.

Key words: complex and non-Archimedean analytic spaces, limit mixed Hodge structures.

2000 Mathematics Subject Classifications: 14D07, 14G22, 14C30, 32P05, 32G20, 32S35

Introduction

Let $\mathcal{X}$ be a proper scheme over the field F of functions meromorphic in an open neighborhood of zero in the complex plane $\mathbf{C}$. The scheme $\mathcal{X}$ gives rise to a proper morphism of complex analytic spaces $\mathcal{X}^h \to D^* = D\backslash\{0\}$, where D is an open disc with center at zero (see Section 3). It is well known that after shrinking the disc D (and replacing $\mathcal{X}^h$ by its preimage), the cohomology groups $H^i\left(\mathcal{X}_t^h, \mathbf{Z}\right)$ of the fiber $\mathcal{X}_t^h$ at a point $t \in D^*$ form a local system of finitely generated abelian groups, and that the corresponding action of the fundamental group $\pi_1(D^*) = \pi_1(D^*, t)$ on $H^i\left(\mathcal{X}_t^h, \mathbf{Z}\right)$ is quasi-unipotent. Furthermore, the mixed Hodge structures on the above groups define a variation of mixed Hodge structures on D^*. Let $\overline{D^*} \to D^*$ be a universal covering of

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*,
Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_2,

D^*, and $\overline{\mathcal{X}}^h = \mathcal{X}^h \times_{D^*} \overline{D^*}$. Then the cohomology group $H^i(\overline{\mathcal{X}}^h, \mathbf{Z})$ admits a mixed Hodge structure, which is the limit (in a certain sense) of the above variation of mixed Hodge structures on D^* (see [GNPP, Exp. IV, Theorem 7.4]). One of the purposes of this paper is to describe the weight zero subspace $W_0 H^i(\overline{\mathcal{X}}^h, \mathbf{Q})$ in terms of non-Archimedean analytic geometry.

Let K be the completion of the discrete valuation field F, and fix a corresponding multiplicative valuation on it. The scheme $\mathcal{X}$ gives rise to a proper K-analytic space $\mathcal{X}^{\mathrm{an}} = (\mathcal{X} \otimes_F K)^{\mathrm{an}}$ in the sense of [Ber1] and [Ber2]. Recall that, as a topological space, $\mathcal{X}^{\mathrm{an}}$ is compact and locally arcwise connected, and the topological dimension of $\mathcal{X}^{\mathrm{an}}$ is equal to the dimension of $\mathcal{X}$. If $\mathcal{X}$ is smooth, then $\mathcal{X}^{\mathrm{an}}$ is even locally contractible. Furthermore, let $\overline{\mathcal{X}}^{\mathrm{an}} = (\mathcal{X} \otimes_F \widehat{K}^{\mathrm{a}})^{\mathrm{an}}$, where $\widehat{K}^{\mathrm{a}}$ is the completion of the algebraic closure K^{a} of K, which corresponds to the universal covering $\overline{D^*} \to D^*$. Recall that the cohomology groups $H^i(\overline{\mathcal{X}}^{\mathrm{an}}, \mathbf{Z})$ of the underlying topological space of $\overline{\mathcal{X}}^{\mathrm{an}}$ are finitely generated, and there is a finite extension K'' of K in K^{a} such that they coincide with $H^i((\mathcal{X} \otimes_F K')^{\mathrm{an}}, \mathbf{Z})$ for any finite extension K' of K'' in K^{a} (see [Ber5, 10.1]).

In Section 3, we construct a topological space $\mathcal{X}^{\mathrm{An}}$ and a surjective continuous map $\lambda : \mathcal{X}^{\mathrm{An}} \to [0,1]$ for which there are an open embedding $\lambda^{-1}(]0,1]) \hookrightarrow \mathcal{X}^h \times]0,1]$, which is a homotopy equivalence, and a homeomorphism $\lambda^{-1}(0) \stackrel{\sim}{\to} \mathcal{X}^{\mathrm{an}} \times]0,r[$, where r is the radius of the disc D. We show that the induced maps $H^i(\mathcal{X}^{\mathrm{An}}, \mathbf{Z}) \to H^i(\mathcal{X}^{\mathrm{an}} \times]0,r[, \mathbf{Z}) = H^i(\mathcal{X}^{\mathrm{an}}, \mathbf{Z})$ are isomorphisms for all $i \geq 0$. In this way, we get a homomorphism $H^i(\mathcal{X}^{\mathrm{an}}, \mathbf{Z}) \to H^i(\mathcal{X}^h, \mathbf{Z})$, whose composition with the canonical map $H^i(\mathcal{X}^h, \mathbf{Z}) \to H^i(\overline{\mathcal{X}}^h, \mathbf{Z})$ gives a homomorphism $H^i(\mathcal{X}^{\mathrm{an}}, \mathbf{Z}) \to H^i(\overline{\mathcal{X}}^h, \mathbf{Z})$. The same construction applied to finite extensions of F in F^{a} gives rise to a homomorphism of $\pi_1(D^*)$-modules $H^i(\overline{\mathcal{X}}^{\mathrm{an}}, \mathbf{Z}) \to H^i(\overline{\mathcal{X}}^h, \mathbf{Z})$. Theorem 5.1 states that the latter gives rise to a functorial isomorphism of $\pi_1(D^*)$-modules

$$H^i(\overline{\mathcal{X}}^{\mathrm{an}}, \mathbf{Q}) \stackrel{\sim}{\to} W_0 H^i(\overline{\mathcal{X}}^h, \mathbf{Q}).$$

If $\mathcal{X}$ is projective and smooth, one can describe the group $W_0 H^i(\overline{\mathcal{X}}^h, \mathbf{Q})$ as follows. By the local monodromy theorem, the action of $(T^m - 1)^{i+1}$ on $H^i\left(\mathcal{X}^h_t, \mathbf{Z}\right)$ is zero (for some $m \geq 1$), where T is the canonical generator of $\pi_1(D^*)$. If we fix a point of $\overline{D^*}$ over t, there is an induced isomorphism $H^i(\overline{\mathcal{X}}^h, \mathbf{Z}) \stackrel{\sim}{\to} H^i\left(\mathcal{X}^h_t, \mathbf{Z}\right)$, which gives rise to an isomorphism between $W_0 H^i(\overline{\mathcal{X}}^h, \mathbf{Q})$ and the maximal unipotent monodromy subspace of $H^i\left(\mathcal{X}^h_t, \mathbf{Q}\right)$, i.e., $(T^m - 1)^i H^i\left(\mathcal{X}^h_t, \mathbf{Q}\right)$. Thus, in the case considered, there is a functorial isomorphism of $\pi_1(D^*)$-modules

$$H^i(\overline{\mathcal{X}}^{\mathrm{an}}, \mathbf{Q}) \stackrel{\sim}{\to} (T^m - 1)^i H^i\left(\mathcal{X}^h_t, \mathbf{Q}\right).$$

The mixed Hodge theory (see [St, p. 247], [Ill, p. 29]) provides an upper bound on the dimension of the space on the right-hand side, which implies the following bound on that of the left-hand side:

$$\dim_{\mathbf{Q}} H^i(\overline{\mathcal{X}}^{\mathrm{an}}, \mathbf{Q}) \leq \dim_F H^i(\mathcal{X}, \mathcal{O}_{\mathcal{X}}).$$

The equality is achieved for a totally degenerate family of abelian varieties (see [Ber1, §6]), and for a totally degenerate family of Calabi–Yau varieties (in the strong sense). In the latter example, $\mathcal{X}^{\mathrm{an}}$ has rational cohomology of the sphere of dimension $\dim(\mathcal{X})$, and is simply connected (see Remark 4.4(ii)).

In fact, $\mathcal{X}^{\mathrm{An}}$ is the underlying topological space of an analytic space over a commutative Banach ring. The idea of such an object was introduced in [Ber1, §1.5], and developed there in detail in the case when the Banach ring is a non-Archimedean field. The spaces considered in this paper are defined over the field of complex numbers $\mathbf{C}$ provided with the following Banach norm: $\|a\| = \max\{|a|_\infty, |a|_0\}$ for $a \in \mathbf{C}$, where $|\ |_\infty$ is the usual Archimedean valuation, and $|\ |_0$ is the trivial valuation (i.e., $|a|_0 = 1$ for $a \neq 0$). One has $[0,1] \xrightarrow{\sim} \mathcal{M}(\mathbf{C}, \|\ \|)$. Namely, a nonzero point $\rho \in]0,1]$ corresponds to the Archimedean valuation $|\ |_\infty^\rho$, and the zero point 0 corresponds to the trivial valuation $|\ |_0$. The above map λ is a canonical map $\mathcal{X}^{\mathrm{An}} \to \mathcal{M}(\mathbf{C}, \|\ \|) = [0,1]$. The preimage $\lambda^{-1}(\rho)$ of $\rho \in]0,1]$ is the restriction of the complex analytic space $\mathcal{X}^h$ to the smaller open disc $D(r^{\frac{1}{\rho}})$, and $\lambda^{-1}(0)$ is a non-Archimedean analytic space over the field $\mathbf{C}$ provided with the trivial valuation $|\ |_0$. Thus, the space $\mathcal{X}^{\mathrm{An}}$ incorporates both complex analytic and non-Archimedean analytic spaces, and the result on a non-Archimedean interpretation of the weight zero subspaces is evidence that analytic spaces over $(\mathbf{C}, \|\ \|)$ are worth studying.

In Section 1, we recall a construction from [Ber1, §1] that associates with an algebraic variety over a commutative Banach ring k the underlying topological space of a k-analytic space. We do not develop a theory of k-analytic spaces, but restrict ourselves to establishing basic properties necessary for this paper. In Section 2, we specify our study for the field $\mathbf{C}$ provided with the above Banach norm $\|\ \|$, and prove a particular case of the main result from Section 4. Let $\mathcal{O}_{\mathbf{C},0}$ be the local ring of functions analytic in an open neighborhood of zero in $\mathbf{C}$. In Section 3, we associate with an algebraic variety $\mathcal{X}$ over $\mathcal{O}_{\mathbf{C},0}$ analytic spaces of three types: a complex analytic space $\mathcal{X}^h$, a $(\mathbf{C}, \|\ \|)$-analytic space $\mathcal{X}^{\mathrm{An}}$, and a $(\mathbf{C}, |\ |_0)$-analytic space $\mathcal{X}_0^{\mathrm{An}}$. All three spaces are provided with a morphism to the corresponding open discs, and are closely interrelated. The construction gives rise to a commutative diagram of maps between topological spaces. In Section 4, we prove our main result (Theorem 4.1), which states that if $\mathcal{X}$ is proper over $\mathcal{O}_{\mathbf{C},0}$, the homomorphisms between integral cohomology groups induced by certain maps from that diagram are isomorphisms. Essential ingredients of the proof are C. H. Clemens's results from [Cle] and similar results from [Ber5]. If $\mathcal{X}$ is strictly semistable over $\mathcal{O}_{\mathbf{C},0}$, the former provide a strong deformation retraction of $\mathcal{X}^h$ to its fiber $\mathcal{X}_s^h$ at zero, and the latter provide a similar homotopy description of the non-Archimedean space $\mathcal{X}_0^{\mathrm{An}}$. In Section 5, we prove Theorem 5.1, which was already formulated.

We want to emphasize that the above result is an analogue of the description of the weight zero subspaces of l-adic étale cohomology groups of algebraic varieties defined over a local field in terms of cohomology groups of the associated non-Archimedean spaces (see [Ber6]). All of these results are evidence for the fact that the underlying topological space of the non-Archimedean analytic space associated with an algebraic variety somehow represents the weight zero part of the mixed motive of the variety.

I am very grateful to Johannes Nicaise for pointing out a mistake in a previous version of Theorem 4.1, and to the referee for many corrections and suggestions that significantly improved the paper.

1 Topological spaces associated with algebraic varieties over a commutative Banach ring

Let k be a commutative Banach ring, i.e., a commutative ring provided with a Banach norm $\| \ \|$ and complete with respect to it. For an affine scheme $\mathcal{X} = \mathrm{Spec}(A)$ of finite type over k, let $\mathcal{X}^{\mathrm{An}}$ denote the set of all nonzero multiplicative seminorms $| \ | : A \to \mathbf{R}_+$ on the ring A whose restriction to k is bounded with respect to the norm $\| \ \|$. The set $\mathcal{X}^{\mathrm{An}}$ is provided with the weakest topology with respect to which all real-valued functions of the form $| \ | \mapsto |f|$, $f \in A$, are continuous. For a point $x \in \mathcal{X}^{\mathrm{An}}$, the corresponding multiplicative seminorm $| \ |_x$ on A gives rise to a multiplicative norm on the integral domain $A/\mathrm{Ker}(| \ |_x)$ and therefore extends to a multiplicative norm on its field of fractions. The completion of the latter is denoted by $\mathcal{H}(x)$, and the image of an element $f \in A$ under the corresponding character $A \to \mathcal{H}(x)$ is denoted by $f(x)$. (In particular, $|f|_x = |f(x)|$ for all $f \in A$.) If $A = k \neq 0$, the space $\mathcal{X}^{\mathrm{An}}$ is the spectrum $\mathcal{M}(k)$ of k, which is a nonempty compact space, by [Ber1, 1.2.1]. If $A = k[T_1, \dots, T_n]$, the space $\mathcal{X}^{\mathrm{An}}$ is denoted by $\mathbf{A}^n$ (the n-dimensional affine space over k). Notice that the correspondence $\mathcal{X} \mapsto \mathcal{X}^{\mathrm{An}}$ is functorial in $\mathcal{X}$.

A continuous map of topological spaces $\varphi : Y \to X$ is said to be Hausdorff if, for any pair of distinct points $y_1, y_2 \in Y$ with $\varphi(y_1) = \varphi(y_2)$, there exist open neighborhoods $\mathcal{V}_1$ of y_1 and $\mathcal{V}_2$ of y_2 with $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$ (i.e., the image of Y in $Y \times_X Y$ is closed). Furthermore, let X be a topological space such that each point of it has a compact neighborhood. A continuous map $\varphi : Y \to X$ is said to be compact if the preimage of a compact subset of X is a compact subset of Y (i.e., φ is proper in the usual sense, but we use the terminology of [Ber2]). Such a map is Hausdorff, it takes closed subsets of Y to closed subsets of X, and each point of Y has a compact neighborhood.

Lemma 1.1. *(i) The space $\mathcal{X}^{\mathrm{An}}$ is locally compact and countable at infinity;*
(ii) given a closed (respectively an open) immersion $\varphi : \mathcal{Y} \to \mathcal{X}$, the map $\varphi^{\mathrm{an}} : \mathcal{Y}^{\mathrm{An}} \to \mathcal{X}^{\mathrm{An}}$ induces a homeomorphism of $\mathcal{Y}^{\mathrm{An}}$ with a closed (respectively open) subset of $\mathcal{X}^{\mathrm{An}}$;
(iii) given morphisms $\varphi : \mathcal{Y} \to \mathcal{X}$ and $\mathcal{Z} \to \mathcal{X}$, the canonical map

$$(\mathcal{Y} \times_{\mathcal{X}} \mathcal{Z})^{\mathrm{An}} \to \mathcal{Y}^{\mathrm{An}} \times_{\mathcal{X}^{\mathrm{An}}} \mathcal{Z}^{\mathrm{An}}$$

is compact.

Proof. (ii) If φ is a closed immersion, the required fact is trivial. If φ is an open immersion, it suffices to consider the case of a principal open subset $\mathcal{Y} = \mathrm{Spec}(A_f)$ for an element $f \in A$. It is clear that the map φ^{An} is injective and its image is an open subset of $\mathcal{X}^{\mathrm{An}}$. A fundamental system of open sets in $\mathcal{Y}^{\mathrm{An}}$ is formed by finite intersections of sets of the form $\mathcal{U} = \{y \in \mathcal{Y}^{\mathrm{An}} \big| |\frac{g}{f^n}(y)| < r\}$ and $\mathcal{V} = \{y \in \mathcal{Y}^{\mathrm{An}} \big| |\frac{g}{f^n}(y)| > r\}$, where $g \in A$, $n \geq 0$ and $r > 0$. It suffices therefore to verify that the sets $\mathcal{U}$ and $\mathcal{V}$ are open in $\mathcal{X}^{\mathrm{An}}$. Given a point $y \in \mathcal{U}$ (respectively $\mathcal{V}$), there exist $\varepsilon, \delta > 0$ such that $\frac{|g(y)|+\delta}{|f^n(y)|-\varepsilon} < r$ (respectively $\frac{|g(y)|-\delta}{|f^n(y)|+\varepsilon} > r$). Then the set $\{z \in \mathcal{X}^{\mathrm{An}} \big| |g(z)| < |g(y)|+\delta, |f^n(z)| > |f^n(y)|-\varepsilon\}$ (respectively $\{z \in \mathcal{X}^{\mathrm{An}} \big| |g(z)| > |g(y)| - \delta, |f^n(z)| < |f^n(y)| + \varepsilon\}$) is an open neighborhood of the point y in $\mathcal{X}^{\mathrm{An}}$, which is contained in $\mathcal{U}$ (respectively $\mathcal{V}$).

(i) By (ii), it suffices to consider the case of the affine space $\mathbf{A}^n$, which is associated with the ring of polynomials $k[T] = k[T_1, \ldots, T_n]$. One has $\mathbf{A}^n = \bigcup_r E(r)$, where the union is taken over tuples of positive numbers $r = (r_1, \ldots, r_n)$ and $E(r)$ is the closed polydisc of radius r with center at zero $\{x \in \mathbf{A}^n \big| |T_i(x)| \leq r_i \text{ for all } 1 \leq i \leq n\}$. The latter is a compact space. Indeed, let $k\langle r^{-1}T\rangle = k\left\langle r_1^{-1}T_1, \ldots, r_n^{-1}T_n\right\rangle$ denote the commutative Banach ring of all power series $f = \sum_\nu a_\nu T^\nu$ over k such that $\|f\| = \sum_\nu \|a_\nu\| r^\nu < \infty$. By [Ber1, Theorem 1.2.1], the spectrum $\mathcal{M}(k\langle r^{-1}T\rangle)$ is a nonempty compact space, and the canonical homomorphism $k[T] \to k\langle r^{-1}T\rangle$ induces a homeomorphism $\mathcal{M}(k\langle r^{-1}T\rangle) \xrightarrow{\sim} E(r)$.

(iii) Let $\mathcal{X} = \mathrm{Spec}(A)$, $\mathcal{Y} = \mathrm{Spec}(B)$ and $\mathcal{Z} = \mathrm{Spec}(C)$. By (ii), it suffices to consider the case $A = k[T_1, \ldots, T_n]$, $B = A[U_1, \ldots, U_p]$, and $C = A[V_1, \ldots, V_q]$, i.e., it suffices to verify that the corresponding map $\mathbf{A}^{n+p+q} \to \mathbf{A}^{n+p} \times_{\mathbf{A}^n} \mathbf{A}^{n+q}$ is compact. This is clear, since the preimage of $E(r') \times_{\mathbf{A}^n} E(r'')$ with $r' = (r_1, \ldots, r_n, s_1, \ldots, s_p)$ and $r'' = (r_1, \ldots, r_n, t_1, \ldots, t_q)$ is the polydisc $E(r)$ with $r = (r_1, \ldots, r_n, s_1, \ldots, s_p, t_1, \ldots, t_q)$. ☐

Now let $\mathcal{X}$ be a scheme of finite type over k. By Lemma 1.1(ii), one can glue the spaces $\mathcal{U}^{\mathrm{An}}$ for open affine subschemes $\mathcal{U} \subset \mathcal{X}$ to get a topological space $\mathcal{X}^{\mathrm{An}}$ in which all $\mathcal{U}^{\mathrm{An}}$ are open subspaces. Here is an equivalent description of the space $\mathcal{X}^{\mathrm{An}}$. For a bounded character $\chi : k \to K$ to a valuation field K (i.e., a field complete with respect to a valuation), let $\mathcal{X}(K)^\chi$ denote the set of all K-points of $\mathcal{X}$ that induce the character χ on k. Furthermore, let $\widetilde{\mathcal{X}}^{\mathrm{An}}$ be the disjoint union of the sets $\mathcal{X}(K)^\chi$ taken over bounded characters $\chi : k \to K$ to a valuation field K. Two points $x' \in \mathcal{X}(K')^{\chi'}$ and $x'' \in \mathcal{X}(K'')^{\chi''}$ are said to be equivalent if there exist a bounded character $\chi : k \to K$, a point $x \in \mathcal{X}(K)^\chi$, and isometric embeddings $K \to K'$ and $K \to K''$ that are compatible with the characters χ' and χ'', taking x to the points x' and x'', respectively. It is really an equivalence relation, and the space $\mathcal{X}^{\mathrm{An}}$ is the set of equivalence classes in $\widetilde{\mathcal{X}}^{\mathrm{An}}$.

The correspondence $\mathcal{X} \mapsto \mathcal{X}^{\mathrm{An}}$ is functorial in $\mathcal{X}$, and the properties (ii) and (iii) of Lemma 1.1 are straightforwardly extended to arbitrary schemes of finite type over k.

Lemma 1.2. *Let $\varphi : \mathcal{Y} \to \mathcal{X}$ be a morphism of schemes of finite type over k, and let φ^{An} be the induced map $\mathcal{Y}^{\mathrm{An}} \to \mathcal{X}^{\mathrm{An}}$. Then*

(i) if φ is separated, then the map φ^{An} is Hausdorff;
(ii) if φ is projective, then the map φ^{An} is compact;
(iii) if φ is proper and either the ring k is Noetherian or $\mathcal{Y}$ has a finite number of irreducible components, then the map φ^{An} is compact.

The assumptions in (iii) guarantee application of Chow's lemma (see [EGAII, 5.6.1]).

Proof. We may assume that the scheme $\mathcal{X} = \mathrm{Spec}(A)$ is affine.

(i) Since φ is separated, the diagonal map $\mathcal{Y}^{\mathrm{An}} \to (\mathcal{Y} \times_{\mathcal{X}} \mathcal{Y})^{\mathrm{An}}$ has a closed image. Since the image of the latter in $\mathcal{Y}^{\mathrm{An}} \times_{\mathcal{X}^{\mathrm{An}}} \mathcal{Y}^{\mathrm{An}}$ is closed, it follows that the map φ^{An} is Hausdorff.

(ii) It suffices to consider the case in which $\mathcal{Y} = \mathrm{Proj}(A[T_0, \dots, T_n])$ is the projective space over A. In this case, $\mathcal{Y} = \bigcup_{i=0}^n \mathcal{Y}_i$, where $\mathcal{Y}_i = \mathrm{Spec}\left(A\left[\frac{T_0}{T_i}, \dots, \frac{T_n}{T_i}\right]\right)$. If $E_i = \left\{y \in \mathcal{Y}_i^{\mathrm{An}} \middle| |\frac{T_j}{T_i}(y)| \le 1 \text{ for all } 0 \le j \le n\right\}$, then the map $E_i \to \mathcal{X}^{\mathrm{An}}$ is compact, and one has $\mathcal{Y}^{\mathrm{An}} = \bigcup_{i=0}^n E_i$. It follows that φ^{An} is a compact map.

(iii) Chow's lemma reduces the situation to the case considered in (ii). □

2 The case of the Banach ring $(\mathbf{C}, \| \ \|)$

We now consider the case when k is the field of complex numbers $\mathbf{C}$ provided with the following Banach norm: $\|a\| = \max\{|a|_\infty, |a|_0\}$ for all $a \in \mathbf{C}$. Notice that there is a homeomorphism $[0,1] \xrightarrow{\sim} \mathcal{M}(\mathbf{C}, \| \ \|) : \rho \mapsto p_\rho$, where the point p_0 corresponds to the trivial norm $| \ |_0$, and each point p_ρ with $\rho > 0$ corresponds to the Archimedean norm $| \ |_\infty^\rho$. Indeed, if $| \ |$ is a valuation that is different from those above, then it is nontrivial and not equivalent to $| \ |_\infty$. It follows that there exists a complex number $a \in \mathbf{C}$ with $|a|_\infty < 1$ and $|a| > 1$, i.e., $|a| > \|a\|$, and the valuation $| \ |$ is not bounded with respect to the Banach norm $\| \ \|$.

For every scheme $\mathcal{X}$ of finite type over $\mathbf{C}$, there is a canonical surjective map $\lambda = \lambda_{\mathcal{X}} : \mathcal{X}^{\mathrm{An}} \to \mathcal{M}(\mathbf{C}, \| \ \|) = [0,1]$. If $\rho \in]0,1]$, then $\mathcal{H}(p_\rho)$ is the field $\mathbf{C}$ provided with the Archimedean valuation $| \ |_\infty^\rho$. The fiber $\lambda^{-1}(1)$ is the complex analytic space $\mathcal{X}^h$ associated with $\mathcal{X}$, by complex GAGA [Serre]. The fiber $\lambda^{-1}(0)$ is the non-Archimedean $(\mathbf{C}, | \ |_0)$-analytic space $\mathcal{X}^{\mathrm{an}}$ associated with $\mathcal{X}$, by non-Archimedean GAGA [Ber1].

Lemma 2.1. *There is a functorial homeomorphism $\lambda^{-1}(]0,1]) \xrightarrow{\sim} \mathcal{X}^h \times]0,1] : x \mapsto (y, \rho)$, which commutes with the projections onto $]0,1]$ and, in the case of affine $\mathcal{X} = \mathrm{Spec}(A)$, is defined by $\rho = \lambda(x)$ and $|f(y)|_\infty = |f(x)|^{\frac{1}{\rho}}$, $f \in A$.*

Proof. Assume first that $\mathcal{X} = \mathrm{Spec}(A)$ is affine. The map considered is evidently continuous. It has an inverse map $\mathcal{X}^h \times]0,1] \to \lambda^{-1}(]0,1]) : (y,\rho) \mapsto y_\rho$, defined by $|f(y_\rho)| = |f(y)|_\infty^\rho$ for $f \in A$, and therefore it is bijective. The inverse map is continuous, since the topology on $\mathcal{X}^h \times]0,1]$ coincides with the weakest one with respect to which all functions of the form $\mathcal{X}^h \times]0,1] \to \mathbf{R}_+ : (y,\rho) \mapsto |f(y)|^\rho$ for $f \in A$ are continuous. It is trivial that the homeomorphisms are functorial in $\mathcal{X}$, and they extend to the class of all schemes of finite type over $\mathbf{C}$. □

Corollary 2.2. *If $\mathcal{X}$ is connected, then the topological space $\mathcal{X}^{\mathrm{An}}$ is also connected.*

Proof. Any $\mathbf{C}$-point of $\mathcal{X}$ defines a section $\mathcal{M}(\mathbf{C}, \|\ \|) = [0,1] \to \mathcal{X}^{\mathrm{An}}$ of the canonical map $\lambda : \mathcal{X}^{\mathrm{An}} \to [0,1]$, and so the required fact follows from the corresponding facts in complex GAGA [Serre] and non-Archimedean GAGA [Ber1, 3.5.3]. □

Proposition 2.3. *If $\mathcal{X}$ is proper, then $H^q(\mathcal{X}^{\mathrm{An}}, \mathbf{Z}) \xrightarrow{\sim} H^q(\mathcal{X}^{\mathrm{an}}, \mathbf{Z})$ for all $q \geq 0$.*

Proof. Since the space $\mathcal{X}^{\mathrm{An}}$ is compact, it suffices to show that the cohomology groups with compact support $H_c^q(\mathcal{X}^h \times]0,1], \mathbf{Z})$ are zero for all $q \geq 0$. For this we use the Leray spectral sequence

$$E_2^{p,q} = H_c^p(]0,1], R^q\lambda_*\mathbf{Z}) \Longrightarrow H_c^{p+q}(\mathcal{X}^h \times]0,1], \mathbf{Z}).$$

The sheaves $R^q\lambda_*\mathbf{Z}$ are constant, and therefore $E_2^{p,q} = 0$ for all $p, q \geq 0$, and the required fact follows. □

By Proposition 2.3, if $\mathcal{X}$ is proper, there is a homomorphism

$$H^q(\mathcal{X}^{\mathrm{an}}, \mathbf{Z}) \longrightarrow H^q(\mathcal{X}^h \times]0,1], \mathbf{Z}) = H^q(\mathcal{X}^h, \mathbf{Z}).$$

Corollary 2.4. *If $\mathcal{X}$ is proper, the above homomorphism gives rise to an isomorphism*

$$H^q(\mathcal{X}^{\mathrm{an}}, \mathbf{Q}) \xrightarrow{\sim} W_0 H^q(\mathcal{X}^h, \mathbf{Q}).$$

Proof. By the construction from [Del3, §6.2] and Hironaka's theorem on resolution of singularities, there exists a proper hypercovering $\mathcal{X}_\bullet \to \mathcal{X}$ such that each $\mathcal{X}_n$ is smooth. By [SGA4, Exp. V bis], it gives rise to a homomorphism of spectral sequences

$$\begin{array}{ccccc} {}'E_1^{p,q} & = H^q\left(\mathcal{X}_p^{\mathrm{an}}, \mathbf{Q}\right) & \Longrightarrow & H^{p+q}(\mathcal{X}^{\mathrm{an}}, \mathbf{Q}), \\ & \downarrow & & \downarrow \\ {}''E_1^{p,q} & = H^q\left(\mathcal{X}_p^h, \mathbf{Q}\right) & \Longrightarrow & H^{p+q}(\mathcal{X}^h, \mathbf{Q}). \end{array}$$

By [Ber5, §5], the connected components of each $\mathcal{X}_p^{\mathrm{an}}$ are contractible. This implies that ${}'E_1^{p,q} = 0$ for all $q \geq 1$, and therefore the first spectral sequence gives rise to isomorphisms ${}'E_2^{p,0} \xrightarrow{\sim} H^p(\mathcal{X}^{\mathrm{an}}, \mathbf{Q})$. On the other hand, by [Del2, 3.2.15(ii)], the mixed Hodge structure on $H^q\left(\mathcal{X}_p^h, \mathbf{Q}\right)$ has the property that $W_i = 0$ for $i < q$. Since the functor $H \mapsto W_0 H$ on the category of rational mixed Hodge structures H with $W_i H = 0$ for $i < 0$ is exact [Del2, 2.3.5(iv)], the latter implies that $W_0\left({}''E_1^{p,q}\right) = 0$ for all $q \geq 1$, and therefore the second spectral sequence gives rise to an isomorphism $W_0\left({}''E_2^{p,0}\right) \xrightarrow{\sim} W_0 H^p(\mathcal{X}^h, \mathbf{Q})$. The required fact now follows from Corollary 2.2. □

Remarks 2.5. (i) It would be interesting to know whether Proposition 2.3. is true for an arbitrary separated scheme $\mathcal{X}$ of finite type over F. If this is true, then the similar induced homomorphism $H^q(\mathcal{X}^{\mathrm{an}}, \mathbf{Z}) \to H^q(\mathcal{X}^h, \mathbf{Z})$ gives rise to an isomorphism analogous to that of Corollary 2.4. That such an isomorphism exists is shown in [Ber6, Theorem 1.1(c)] using the same reasoning as that used in the proof of Corollary 2.4 (see also Remark 5.3).

(ii) It is very likely that the map $\mathcal{X}^{\mathrm{an}} \to \mathcal{X}^{\mathrm{An}}$ is a homotopy equivalence at least in the case when $\mathcal{X}$ is a proper scheme over F with the property that for every $n \geq 1$, a nonempty intersection of n-irreducible components is smooth and of codimension $n-1$ (see also Remark 4.4(iii)).

3 Topological spaces associated with algebraic varieties over the ring $\mathcal{O}_{\mathbf{C},0}$

Let $\mathcal{X}$ be a scheme of finite type over the local ring $\mathcal{O}_{\mathbf{C},0}$. We are going to associate with $\mathcal{X}$ (the underlying topological spaces of) analytic spaces of three types. The first one is a classical object. This is a complex analytic space $\mathcal{X}^h$ over an open disc $D(r)$ in $\mathbf{C}$ of radius r (with center at zero). The second one is a $(\mathbf{C}, \| \ \|)$-analytic space $\mathcal{X}^{\mathrm{An}}$ over an open disc $\mathcal{D}(r)$ in $\mathbf{A}^1$. And the third one is a non-Archimedean $(\mathbf{C}, | \ |_0)$-analytic space $\mathcal{X}_0^{\mathrm{An}}$ over an open disc $D_0(r)$ in the $(\mathbf{C}, | \ |_0)$-analytic affine line $\mathbf{A}_0^1$. The first two objects are related to two representations of the ring $\mathcal{O}_{\mathbf{C},0}$ in the form of a filtered inductive limit of the same commutative rings, provided with two different commutative Banach ring structures. The third object is the analytic space associated with the base change of $\mathcal{X}$ under the homomorphism $\mathcal{O}_{\mathbf{C},0} \to \mathbf{C}[[T]] = \mathcal{O}_{\mathbf{A}_0^1,0}$, and is a particular case of an object introduced in [Ber3, §3].

For $r > 0$, let $\mathbf{C}\langle r^{-1}T\rangle$ denote the commutative Banach algebra of formal power series $f = \sum_{i=0}^\infty a_i T^i$ over $\mathbf{C}$ absolutely convergent at the closed disc $E(r) = \{x \in \mathbf{C} \big| |T(x)| \leq r\}$ and provided with the norm $\|f\| = \sum_{i=0}^\infty |a_i|_\infty r^i$. The canonical homomorphism $\mathbf{C}[T] \to \mathbf{C}\langle r^{-1}T\rangle$ induces a homeomorphism $\mathcal{M}(\mathbf{C}\langle r^{-1}T\rangle) \xrightarrow{\sim} E(r)$, and one has $\mathcal{O}_{\mathbf{C},0} = \varinjlim \mathbf{C}\langle r^{-1}T\rangle$ for r tending to zero. By [EGAIV, §8], for any scheme $\mathcal{X}$ of finite type over $\mathcal{O}_{\mathbf{C},0}$, there exist $r > 0$ and a scheme $\mathcal{X}'$ of finite type over $\mathbf{C}\langle r^{-1}T\rangle$ whose base change with respect to the canonical homomorphism $\mathbf{C}\langle r^{-1}T\rangle \to \mathcal{O}_{\mathbf{C},0}$ is $\mathcal{X}$. By the construction of Section 1, there is an associated topological space $\mathcal{X}'^h$, and we denote by

$\mathcal{X}^h$ the preimage of the open disc $D(r) = \{x \in \mathbf{C} \big| |T(x)| < r\}$ with respect to the canonical map $\mathcal{X}'^h \to E(r)$. The morphism $\mathcal{X}^h \to D(r)$ does not depend, up to a change of r, on the choice of $\mathcal{X}'$, and the construction is functorial in $\mathcal{X}$ (see Remark 3.3).

Furthermore, for $r > 0$, let $\mathbf{C}\langle\langle r^{-1}T\rangle\rangle$ denote the commutative Banach ring of formal power series $f = \sum_{i=0}^{\infty} a_i T^i$ over $\mathbf{C}$ such that $\|f\| = \sum_{i=0}^{\infty} \|a_i\| r^i < \infty$. The canonical homomorphism $\mathbf{C}[T] \to \mathbf{C}\langle\langle r^{-1}T\rangle\rangle$ gives rise to a homeomorphism $\mathcal{M}(\mathbf{C}\langle\langle r^{-1}T\rangle\rangle) \xrightarrow{\sim} \mathcal{E}(r) = \{x \in \mathbf{A}^1 \big| |T(x)| \le r\}$. If $r \ge 1$, then $\mathbf{C}[T] \xrightarrow{\sim} \mathbf{C}\langle\langle r^{-1}T\rangle\rangle$, and if $r < 1$, then $\mathbf{C}\langle\langle r^{-1}T\rangle\rangle \xrightarrow{\sim} \mathbf{C}\langle r^{-1}T\rangle$ (as $\mathbf{C}$-subalgebras of $\mathbf{C}[[T]]$). One has $\mathcal{O}_{\mathbf{C},0} = \varinjlim \mathbf{C}\langle\langle r^{-1}T\rangle\rangle$ for r tending to zero. By [EGAIV, §8] again, for any scheme $\mathcal{X}$ of finite type over $\mathcal{O}_{\mathbf{C},0}$, there exist $0 < r < 1$ and a scheme $\mathcal{X}'$ of finite type over $\mathbf{C}\langle\langle r^{-1}T\rangle\rangle$ whose base change with respect to the canonical homomorphism $\mathbf{C}\langle\langle r^{-1}T\rangle\rangle \to \mathcal{O}_{\mathbf{C},0}$ is $\mathcal{X}$. By the construction of Section 1, there is an associated topological space $\mathcal{X}'^{\mathrm{An}}$, and we denote by $\mathcal{X}^{\mathrm{An}}$ the preimage of $\mathcal{D}(r) = \{x \in \mathbf{A}^1 \big| |T(x)| < r\}$ with respect to the canonical map $\mathcal{X}'^{\mathrm{An}} \to \mathcal{E}(r)$. The map $\varphi : \mathcal{X}^{\mathrm{An}} \to \mathcal{D}(r)$ does not depend, up to a change of r, on the choice of $\mathcal{X}'$, and the construction is functorial in $\mathcal{X}$ (see again Remark 3.3).

Finally, for $r > 0$, let $\mathbf{C}\{r^{-1}T\}$ denote the commutative Banach ring of formal power series $f = \sum_{i=0}^{\infty} a_i T^i$ over $\mathbf{C}$ convergent at the closed disc $E_0(r) = \{x \in \mathbf{A}_0^1 \big| |T(x)| \le r\}$ and provided with the norm $\|f\| = \max_{i \ge 0}\{|a_i|_0 r^i\}$. The canonical homomorphism $\mathbf{C}[T] \to \mathbf{C}\{r^{-1}T\}$ gives rise to a homeomorphism $\mathcal{M}(\mathbf{C}\{r^{-1}T\}) \xrightarrow{\sim} E_0(r)$. If $r \ge 1$, then $\mathbf{C}[T] \xrightarrow{\sim} \mathbf{C}\{r^{-1}T\}$, and if $r < 1$, then $\mathbf{C}\{r^{-1}T\} \xrightarrow{\sim} \mathbf{C}[[T]]$. One has $\mathcal{O}_{\mathbf{A}_0^1,0} = \mathbf{C}[[T]] = \mathbf{C}\{r^{-1}T\}$ for every $0 < r < 1$. Thus, given a scheme $\mathcal{X}$ of finite type over $\mathcal{O}_{\mathbf{C},0}$, we set $\mathcal{X}_0 = \mathcal{X} \otimes_{\mathcal{O}_{\mathbf{C},0}} \mathbf{C}[[T]]$, and for $0 < r < 1$, we set $\mathcal{X}_0' = \mathcal{X}_0 \otimes_{\mathbf{C}[[T]]} \mathbf{C}\{r^{-1}T\}$. There is an associated topological space $\mathcal{X}_0'^{\mathrm{An}}$, and we denote by $\mathcal{X}_0^{\mathrm{An}}$ the preimage of $D_0(r) = \{x \in \mathbf{A}_0^1 \big| |T(x)| < r\}$ with respect to the canonical map $\mathcal{X}_0'^{\mathrm{An}} \to E_0(r)$.

Recall that F denotes the fraction field of $\mathcal{O}_{\mathbf{C},0}$, and K denotes the completion of F with respect to a fixed valuation, which is determined by its value at T. Let ε be the latter value. The K-analytic space associated with a scheme $\mathcal{X}$ of finite type over F is denoted by $\mathcal{X}^{\mathrm{an}}$ (instead of $(\mathcal{X} \otimes_F K)^{\mathrm{an}}$).

Lemma 3.1. *Let $\mathcal{X}$ be a scheme of finite type over $\mathcal{O}_{\mathbf{C},0}$, $\mathcal{X}_\eta = \mathcal{X} \otimes_{\mathcal{O}_{\mathbf{C},0}} F$ its generic fiber, and $\mathcal{X}_s = \mathcal{X} \otimes_{\mathcal{O}_{\mathbf{C},0}} \mathbf{C}$ its closed fiber. Let $\mathcal{X}^{\mathrm{An}}$ be the associated $(\mathbf{C}, \|\ \|)$-analytic space over $\mathcal{D}(r)$ with $0 < r < 1$, and let λ, λ_η, and λ_s be the canonical maps to $\mathcal{M}((\mathbf{C}, \|\ \|) = [0,1]$ from $\mathcal{X}^{\mathrm{An}}$, $\mathcal{X}_\eta^{\mathrm{An}}$ and $\mathcal{X}_s^{\mathrm{An}}$, respectively. Then*

(i) there is a homeomorphism

$$\lambda^{-1}(]0,1]) \xrightarrow{\sim} \left\{(x,\rho) \in \mathcal{X}^h \times]0,1] \big| |T(x)|_\infty < r^{\frac{1}{\rho}}\right\} : x' \mapsto (x,\rho),$$

which commutes with the projections onto $]0,1]$ *and, in the case of the affine* $\mathcal{X} = \mathrm{Spec}(A)$*, is defined by* $\rho = \lambda(x')$ *and* $|f(x)|_\infty = |f(x')|^{\frac{1}{\rho}}$*,* $f \in A$*;*

(ii) $\lambda^{-1}(0) \xrightarrow{\sim} \mathcal{X}_0^{\mathrm{An}}$*;*

(iii) there is a homeomorphism

$$\lambda_\eta^{-1}(0) \xrightarrow{\sim} \mathcal{X}_\eta^{\mathrm{an}} \times]0, r[: x' \mapsto (x, \rho),$$

which, in the case of the affine $\mathcal{X} = \mathrm{Spec}(A)$*, is defined by* $\rho = |T(x')|$ *and* $|f(x)| = |f(x')|^{\log_\rho(\varepsilon)}$*,* $f \in A$*;*

(iv) $\mathcal{X}^{\mathrm{An}} \backslash \mathcal{X}_\eta^{\mathrm{An}} = \mathcal{X}_s^{\mathrm{An}}$*, where the right-hand side is the* $(\mathbf{C}, \| \ \|)$*-analytic space associated with* $\mathcal{X}_s$ *in the sense of Section 2.*

Proof. In (i), the converse map $(x, \rho) \mapsto x_\rho$ is defined by $|f(x_\rho)| = |f(x)|_\infty^\rho$, $f \in A$, and in (iii), the converse map $\mathcal{X}_\eta^{\mathrm{an}} \times]0, r[\xrightarrow{\sim} \lambda^{-1}(0) \backslash \varphi^{-1}(0) : (x, \rho) \mapsto P_{x,\rho}$ is defined by $|f(P_{x,\rho})| = |f(x)|^{\log_\varepsilon(\rho)}$, $f \in A$. The statements (ii) and (iv) are trivial. □

Corollary 3.2. *The open embedding* $\lambda^{-1}(]0,1]) \hookrightarrow \mathcal{X}^h \times]0,1]$ *is a homotopy equivalence.*

Proof. The formula $((x,\rho),t) \mapsto (x, \max(\rho, t))$ defines a strong deformation retraction of $\lambda^{-1}(]0,1])$ and $\mathcal{X}^h \times]0,1]$ to $\mathcal{X}^h \times \{1\}$. □

Remarks 3.3. (i) The spaces $\mathcal{X}^h$, $\mathcal{X}^{\mathrm{An}}$, and $\mathcal{X}_0^{\mathrm{An}}$ are in fact pro-objects (i.e., filtered projective systems of objects) of the corresponding categories of analytic spaces (see [Ber3, §2]). The functoriality of their constructions means that they give rise to functors from the category of schemes of finite type over $\mathcal{O}_{\mathbf{C},0}$ to the corresponding categories of pro-objects.

(ii) Suppose that $\mathcal{X}$ is a scheme of finite type over $\mathcal{O}_{\mathbf{C},0}$ for which the canonical morphism to $\mathrm{Spec}(\mathcal{O}_{\mathbf{C},0})$ is a composition $\mathcal{X} \xrightarrow{\varphi} \mathrm{Spec}(\mathcal{O}_{\mathbf{C},0}) \xrightarrow{\psi} \mathrm{Spec}(\mathcal{O}_{\mathbf{C},0})$, where ψ is induced by the homomorphism $\mathcal{O}_{\mathbf{C},0} \to \mathcal{O}_{\mathbf{C},0} : T \mapsto T^n$ for $n \geq 1$. Let $\mathcal{Y}$ denote the same scheme $\mathcal{X}$ but considered over $\mathrm{Spec}(\mathcal{O}_{\mathbf{C},0})$ with respect to the morphism φ. Then there is a canonical homeomorphism of topological spaces $\mathcal{Y}_0^{\mathrm{An}} \xrightarrow{\sim} \mathcal{X}_0^{\mathrm{An}} : y \mapsto x$, which, in the case of affine $\mathcal{X} = \mathrm{Spec}(A)$, is defined by $|f(x)| = |f(y)|^n$ for $f \in A$. It induces homeomorphisms $\mathcal{Y}_\eta^{\mathrm{An}} \times]0, r^{\frac{1}{n}}[\xrightarrow{\sim} \mathcal{X}_\eta^{\mathrm{An}} \times]0, r[: (y, \rho) \mapsto (x, \rho^n)$ and $\mathcal{Y}_s^{\mathrm{An}} \xrightarrow{\sim} \mathcal{X}_s^{\mathrm{An}} : y \mapsto x$ (see Lemma 3.1(i) and (iii)), defined by $|f(x)| = |f(y)|^n$ for $f \in A$.

4 The main result

Let $\mathcal{X}$ be a scheme of finite type over $\mathcal{O}_{\mathbf{C},0}$. By the previous section, for some $0 < r < 1$ there is a commutative diagram in which hook and down arrows are open embeddings, left and up arrows are closed embeddings, and all squares are cartesian:

$$
\begin{tikzcd}
\mathcal{X}_\eta^h \times]0,1] \arrow[d, "*"] & \lambda_\eta^{-1}(]0,1]) \arrow[l, hook'] \arrow[r, hook, "*"] \arrow[d, "*"] & \mathcal{X}_\eta^{\mathrm{An}} \arrow[d] & \lambda_\eta^{-1}(0) \arrow[l] \arrow[r, "\sim"] \arrow[d] & \mathcal{X}_\eta^{\mathrm{an}} \times]0,r[\arrow[d] \\
\mathcal{X}^h \times]0,1] & \lambda^{-1}(]0,1]) \arrow[l, hook'] \arrow[r, hook, "*"] & \mathcal{X}^{\mathrm{An}} & \lambda^{-1}(0) \arrow[l] \arrow[r, "\sim"] & \mathcal{X}_0^{\mathrm{An}} \\
\mathcal{X}_s^h \times]0,1] \arrow[u] & \lambda_s^{-1}(]0,1]) \arrow[l, "\sim"'] \arrow[r, hook, "*"] \arrow[u] & \mathcal{X}_s^{\mathrm{An}} \arrow[u] & \lambda_s^{-1}(0) \arrow[l] \arrow[r, "\sim"] \arrow[u] & \mathcal{X}_s^{\mathrm{an}} \arrow[u]
\end{tikzcd}
$$

Theorem 4.1. *Assume that $\mathcal{X}$ is proper (respectively proper and strictly semistable) over $\mathcal{O}_{\mathbf{C},0}$. Then for a sufficiently small r, all of the horizontal (respectively vertical) arrows of the diagram, except those marked by $*$, induce an isomorphism between integral cohomology groups of the corresponding topological spaces.*

The following lemma is a version of Grothendieck's Proposition 3.10.2 from [Gro]. If F is a sheaf on a topological space X, Φ is a family of supports in X, and Y is a subspace of X, then $H^q_{\Phi\cap Y}(Y,F)$ denotes the cohomology groups with coefficients in the pullback of F at Y and with supports in $\Phi \cap Y = \{A\cap Y \big| A \in \Phi\}$.

Lemma 4.2. *Let X be a paracompact locally compact topological space, $X_1 \subset X_2 \subset \cdots$ an increasing sequence of closed subsets such that the union of their topological interiors in X coincides with X, and Φ a paracompactifying family of supports in X such that $A \in \Phi$ if and only if $A \cap X_i \in \Phi \cap X_i$ for all $i \geq 1$. Let F be an abelian sheaf on X, and let $q \geq 1$. Assume that for each $i \geq 1$ the image $H^{q-1}_{\Phi\cap X_{i+1}}(X_{i+1},F)$ in $H^{q-1}_{\Phi\cap X_i}(X_i,F)$ under the restriction homomorphism coincides with the image of $H^{q-1}_{\Phi\cap X_{i+2}}(X_{i+2},F)$. Then there is a canonical isomorphism*

$$H^q_\Phi(X,F) \xrightarrow{\sim} \varprojlim H^q_{\Phi\cap X_i}(X_i,F).$$

Remark 4.3. An analogue of [Gro, Proposition 3.10.2] for étale cohomology groups of non-Archimedean analytic spaces is [Ber2, Proposition 6.3.12]. In the formulation of the latter, only the assumption that X is a union of all X_i's was made. This is not enough, and one has to make the stronger assumption that X is a union of the topological interiors of all X_i's (this guarantees that $F(X) \xrightarrow{\sim} \varprojlim F(X_i)$ for any sheaf F on X).

Proof. First of all, by the above remark and the assumptions on the X_i's and Φ, for any abelian sheaf G on X one has $\Gamma_\Phi(X,G) \xrightarrow{\sim} \varprojlim \Gamma_{\Phi\cap X_i}(X_i,G)$. *We claim that, given an injective abelian sheaf J on X and a closed subset $Y \subset X$, the canonical map $\Gamma_\Phi(X,J) \to \Gamma_{\Phi\cap Y}(Y,J)$ is surjective.* Indeed, let g be an element from $\Gamma_{\Phi\cap Y}(Y,J)$ and let B be its support. By [God, Ch. II, Theorem 3.3.1], g is the restriction of a section g' of J over an open neighborhood $\mathcal{U}$ of B in X. Furthermore, let $A \in \Phi$ be such that $B = A \cap Y$, and let $A' \in \Phi$ be a neighborhood of A in X. Shrinking $\mathcal{U}$, we may assume that $\mathcal{U} \subset A'$. Since J is injective, the map $\Gamma(X,J) \to \Gamma(\mathcal{U} \coprod (X\backslash A'),J) =$

$\Gamma(\mathcal{U}, J) \oplus \Gamma(X \backslash A', J)$ is surjective. It follows that there exists an element $f \in \Gamma(X, J)$ whose restriction to $\mathcal{U}$ is g' and whose restriction to $X \backslash A'$ is zero. Since the support of f lies in A', one has $f \in \Gamma_\Phi(X, J)$, and the claim follows.

The claim implies that the pullback of J at any closed subset $Y \subset X$ is a $(\Phi \cap Y)$-soft sheaf on Y, i.e., for any $A \in \Phi \cap Y$, the canonical map $\Gamma_{\Phi \cap Y}(Y, J) \to \Gamma(A, J)$ is surjective (see [God, Ch. II, Section 3.5]). Thus, given an injective resolution $0 \to F \to J^0 \to J^1 \to \cdots$ of F, there is a commutative diagram

$$\begin{array}{ccccccc} 0 \to & \Gamma_\Phi(X, J^0) & \to & \Gamma_\Phi(X, J^1) & \to & \Gamma_\Phi(X, J^2) & \to \cdots \\ & \downarrow & & \downarrow & & \downarrow & \\ 0 \to & \Gamma_{\Phi \cap X_i}\left(X_i, J^0\right) & \to & \Gamma_{\Phi \cap X_i}\left(X_i, J^1\right) & \to & \Gamma_{\Phi \cap X_i}\left(X_i, J^2\right) & \to \cdots \end{array}$$

in which the first and second rows give rise to the groups $H^q_\Phi(X, F)$ and $H^q_{\Phi \cap X_i}(X_i, F)$, respectively, and the vertical arrows are surjections. The injectivity of the map considered is verified by a simple diagram search in the same way as in the proof of [Ber2, Proposition 6.3.12], and verification of its surjectivity is even easier (and because of that it was omitted in [Ber2]) and goes as follows.

Let $\overline{\beta}_i \in H^q_{\Phi \cap X_i}(X_i, F)$, $i \geq 1$, be a compatible system. Assume that for $i \geq 1$, we have constructed elements $\beta_j \in \Gamma_{\Phi \cap X_j}(X_j, J^q)$ each from the class of $\overline{\beta}_j$, $1 \leq j \leq i$, with $\beta_{j+1}\big|_{X_j} = \beta_j$ for $1 \leq j \leq i-1$, and let β'_{i+1} be an element from the class of $\overline{\beta}_{i+1}$. Then $\beta'_{i+1}\big|_{X_i} = \beta_i + d\gamma_i$ for some $\gamma_i \in \Gamma_{\Phi \cap X_i}\left(X_i, J^{q-1}\right)$. If $\alpha \in \Gamma_\Phi(X, J^{q-1})$ is such that $\alpha\big|_{X_i} = \gamma_i$, then for the element $\beta_{i+1} = \beta'_{i+1} - d\alpha\big|_{X_{i+1}}$, we have $\beta_{i+1}\big|_{X_i} = \beta_i$. By the remark at the beginning of the proof, there exists an element $\beta \in \Gamma_\Phi(X, J^q)$ such that $\beta\big|_{X_i} = \beta_i$ for all $i \geq 1$. Then $d\beta = 0$, and the surjectivity follows. □

Proof of Theorem 4.1. Step 1. First of all, the isomorphism $H^q(\mathcal{X}_s^{\mathrm{An}}, \mathbf{Z}) \xrightarrow{\sim} H^q(\lambda_s^{-1}(0), \mathbf{Z}) = H^q(\mathcal{X}_s^{\mathrm{an}}, \mathbf{Z})$ follows from Proposition 2.3. The isomorphisms

$$H^q(\mathcal{X}^h \times]0,1], \mathbf{Z}) \xrightarrow{\sim} H^q(\lambda^{-1}(]0,1]), \mathbf{Z})$$

and

$$H^q(\mathcal{X}_\eta^h \times]0,1], \mathbf{Z}) \xrightarrow{\sim} H^q(\lambda_\eta^{-1}(]0,1]), \mathbf{Z})$$

follow from Corollary 3.2.

Step 2. To get the isomorphisms

$$H^q(\mathcal{X}^{\mathrm{An}}, \mathbf{Z}) \xrightarrow{\sim} H^q(\lambda^{-1}(0), \mathbf{Z}) \quad \text{and} \quad H^q\left(\mathcal{X}_\eta^{\mathrm{An}}, \mathbf{Z}\right) \xrightarrow{\sim} H^q\left(\lambda_\eta^{-1}(0), \mathbf{Z}\right),$$

we assume that r is sufficiently small that the groups $H^q\left(\mathcal{X}_t^h, \mathbf{Z}\right)$, $t \in D^*(r)$, form a local system for all $q \geq 0$, and therefore, $R^q\psi_{\eta *}\left(\mathbf{Z}_{\mathcal{X}_\eta^h}\right)$ are locally constant quasi-unipotent sheaves of finitely generated abelian groups for all

$q \geq 0$, where ψ is the canonical morphism $\mathcal{X}^h \to D(r)$. Let $\widetilde{\mathcal{X}^h}$ and $\widetilde{\mathcal{X}^h_\eta}$ denote the images of $\lambda^{-1}(]0,1])$ and $\lambda_\eta^{-1}(]0,1])$ in $\mathcal{X}^h \times]0,1]$ and $\mathcal{X}^h_\eta \times]0,1]$, respectively. (If $\mathcal{X} = \mathrm{Spec}(\mathcal{O}_{\mathbf{C},0})$, they will be denoted by $\widetilde{D(r)}$ and $\widetilde{D^*(r)}$.) It suffices to show that $H^q_\Phi(\widetilde{\mathcal{X}^h}, \mathbf{Z}) = 0$ and $H^q_{\Phi_\eta}\left(\widetilde{\mathcal{X}^h_\eta}, \mathbf{Z}\right) = 0$ for all $q \geq 0$, where Φ and Φ_η are families of supports in $\widetilde{\mathcal{X}^h}$ and $\widetilde{\mathcal{X}^h_\eta}$ consisting of the closed subsets that are also closed in $\mathcal{X}^{\mathrm{An}}$ and $\mathcal{X}^{\mathrm{An}}_\eta$, respectively.

Consider the following commutative diagrams in which all squares are cartesian:

$$\begin{array}{ccccc} \widetilde{D(r)} & \hookrightarrow & \mathcal{D}(r) & \longleftarrow & [0,r[\\ \uparrow \widetilde{\varphi} & & \uparrow \varphi & & \uparrow \\ \widetilde{\mathcal{X}^h} & \hookrightarrow & \mathcal{X}^{\mathrm{An}} & \longleftarrow & \mathcal{X}^{\mathrm{An}}_0 \end{array} \qquad \begin{array}{ccccc} \widetilde{D^*(r)} & \hookrightarrow & \mathcal{D}^*(r) & \longleftarrow &]0,r[\\ \uparrow \widetilde{\varphi}_\eta & & \uparrow \varphi_\eta & & \uparrow \\ \widetilde{\mathcal{X}^h_\eta} & \hookrightarrow & \mathcal{X}^{\mathrm{An}}_\eta & \longleftarrow & \mathcal{X}^{\mathrm{an}}_\eta \times]0,r[\end{array}$$

Since all of the vertical maps are compact, there are spectral sequences (with initial terms $E_2^{p,q}$)

$$H^p_\Phi\left(\widetilde{D(r)}, R^q\widetilde{\varphi}_*\mathbf{Z}_{\widetilde{\mathcal{X}^h}}\right) \Longrightarrow H^{p+q}_\Phi(\widetilde{\mathcal{X}^h}, \mathbf{Z})$$

and

$$H^p_{\Phi_\eta}\left(\widetilde{D^*(r)}, R^q\widetilde{\varphi}_{\eta *}\mathbf{Z}_{\widetilde{\mathcal{X}^h_\eta}}\right) \Longrightarrow H^{p+q}_{\Phi_\eta}\left(\widetilde{\mathcal{X}^h_\eta}, \mathbf{Z}\right),$$

where Φ and Φ_η in the $E_2^{p,q}$ terms denote the similar families of supports in $\widetilde{D(r)}$ and $\widetilde{D^*(r)}$, respectively. Thus, it suffices to verify the following fact. Let L be an abelian sheaf on $D(r)$ whose restriction to $D^*(r)$ is locally constant and quasi-unipotent, and let π denote the canonical projection $\widetilde{D(r)} \to D(r)$. Then $(*)$ $H^p_\Phi(\widetilde{D(r)}, \pi^*L) = 0$ *and* $(*_\eta)$ $H^p_{\Phi_\eta}(\widetilde{D^*(r)}, \pi^*L) = 0$ *for all* $p \geq 0$. Both equalities are proved in the same way using Lemma 4.2 as follows.

The equality $(*)$. The space $\mathcal{D}(r)$ is a union of the closed discs $\mathcal{E}(\rho) = \{x \in \mathcal{D}(r) \big| |T(x)| \leq \rho\}$ with $\rho < r$. Let $\widetilde{\mathcal{E}(\rho)} = \mathcal{E}(\rho) \cap \widetilde{D(r)} = \{(y,t) \big| |T(y)|_\infty \leq \rho^{\frac{1}{t}}\}$. Then $\widetilde{D(r)}$ is a union of all $\widetilde{\mathcal{E}(\rho)}$ with $\rho < r$, and if $\rho < \rho'$, then $\mathcal{E}(\rho)$ and $\widetilde{\mathcal{E}(\rho)}$ are contained in the topological interiors of $\mathcal{E}(\rho')$ and $\widetilde{\mathcal{E}(\rho')}$ in $\mathcal{D}(r)$ and $\widetilde{D(r)}$, respectively. It follows easily that a closed subset $B \subset \widetilde{D(r)}$ is closed in $\mathcal{D}(r)$ if and only if $B \cap \widetilde{\mathcal{E}(\rho)}$ is closed in $\mathcal{E}(\rho)$ for all $\rho < r$. Since the spaces $\mathcal{E}(\rho)$ are compact, from Lemma 4.2 it follows that to prove the equality $(*)$, it suffices to show that $H^q_c\left(\widetilde{\mathcal{E}(\rho)}, \pi^*_\rho L\right) = 0$ for all $q \geq 0$, where π_ρ is the canonical projection $\widetilde{\mathcal{E}(\rho)} \to E(\rho)$.

One has $\pi_\rho^{-1}(0) \xrightarrow{\sim}]0,1]$ and, for $y \neq 0$, $\pi_\rho^{-1}(y) \xrightarrow{\sim} [t_y, 1]$, where $0 < t_y \leq 1$ is such that $|T(y)|_\infty = \rho^{\frac{1}{t_y}}$. It follows that $\left(R^q\pi_{\rho!}\left(\pi^*_\rho L\right)\right)_y$ is zero if $q \geq 1$ or $q = 0$ and $y = 0$, and coincides with L_y if $q = 0$ and $y \neq 0$. This means that $R^q\pi_{\rho!}\left(\pi^*_\rho L\right)$ is zero for $q \geq 1$, and coincides with $j_{\rho!}\left(j^*_\rho L\right)$ for $q = 0$, where j_ρ is the canonical open embedding $E^*(\rho) \hookrightarrow E(\rho)$. The Leray spectral sequence $E_2^{p,q} = H_c^p\left(E(\rho), R^q\pi_{\rho!}\left(\pi_\rho^* L\right)\right) \Longrightarrow H_c^{p+q}\left(\widetilde{\mathcal{E}(\rho)}, \pi_\rho^* L\right)$ implies that $H_c^q\left(\widetilde{\mathcal{E}(\rho)}, \pi_\rho^* L\right) = H_c^q(E^*(\rho), L)$ for all $q \geq 0$. Thus, the equality $(*)$ is a consequence of the following simple fact: $H_c^q(E^*(\rho), L) = 0$, $q \geq 0$, *for any abelian quasi-unipotent sheaf* L *on* $E^*(\rho)$.

If L is constant, the above fact follows from the long exact sequence of cohomology with compact supports associated with the maps

$$E^*(\rho) \xhookrightarrow{j_\rho} E(\rho) \longleftarrow \{0\}.$$

It follows easily that the same is true for any unipotent abelian sheaf L. Assume now that L is quasi-unipotent. Then there exists $n \geq 1$ such that the pullback of L under the n^{th}-power map $\varphi : E^*\left(\rho^{\frac{1}{n}}\right) \to E^*(\rho) : z \mapsto z^n$ is unipotent. By the previous case, $H_c^q\left(E^*\left(\rho^{\frac{1}{n}}\right), \varphi^* L\right) = 0$ for all $q \geq 0$. The spectral sequence $E_2^{p,q} = H^p\left(\mathbf{Z}/n\mathbf{Z}, H_c^q\left(E^*\left(\rho^{\frac{1}{n}}\right), \varphi^* L\right)\right) \Longrightarrow H_c^{p+q}(E^*(\rho), L)$ implies the required fact for such L.

The equality $(*_\eta)$ (see also Remark 4.4(i)). The space $\mathcal{D}^*(r)$ is a union of the closed annuli $\mathcal{A}_\rho = \{x \in \mathcal{D}(r) \big| \rho \leq |T(x)| \leq r - \rho\}$ with $0 < \rho < \frac{r}{2}$. Let $\widetilde{\mathcal{A}_\rho} = \mathcal{A}_\rho \cap \widetilde{D^*(r)} = \{(y,t) \big| \rho^{\frac{1}{t}} \leq |T(y)|_\infty \leq (r-\rho)^{\frac{1}{t}}\}$. Then $\widetilde{D^*(r)}$ is a union of $\widetilde{\mathcal{A}_\rho}$ with $0 < \rho < \frac{r}{2}$, and for $\rho < \rho'$, $\mathcal{A}_\rho$ and $\widetilde{\mathcal{A}_\rho}$ lie in the topological interiors of $\mathcal{A}_{\rho'}$ and $\widetilde{\mathcal{A}_{\rho'}}$ in $\mathcal{D}^*(r)$ and $\widetilde{D^*(r)}$, respectively. It follows that a closed subset $B \subset \widetilde{D^*(r)}$ is closed in $\mathcal{D}^*(r)$ if and only if $B \cap \widetilde{\mathcal{A}_\rho}$ is closed in $\mathcal{A}_\rho$ for all $0 < \rho < \frac{r}{2}$. Since the spaces $\mathcal{A}_\rho$ are compact, from Lemma 4.2 it follows that to prove the equality $(*_\eta)$ it suffices to show that $H_c^q\left(\widetilde{\mathcal{A}_\rho}, \pi_\rho^* L\right) = 0$ for all $q \geq 0$, where π_ρ is the canonical projection $\widetilde{\mathcal{A}_\rho} \to E^*(r-\rho)$.

Notice that in comparison with the previous case, the preimage of any point of $E^*(\rho)$ under the latter map is always a closed interval or a point. It follows that $R^q\pi_{\rho*}\left(\pi_\rho^* L\right)$ is zero if $q \geq 1$, and coincides with $L\big|_{E^*(r-\rho)}$ if $q = 0$. The Leray spectral sequence of the map π_ρ implies that $H_c^q\left(\widetilde{\mathcal{A}_\rho}, \pi_\rho^* L\right) = H_c^q(E^*(r-\rho), L)$ for all $q \geq 0$, and the equality $(*_\eta)$ follows from the fact we have already verified.

Step 3. It remains to show that if $\mathcal{X}$ is proper and strictly semistable over $\mathcal{O}_{\mathbf{C},0}$, then the unmarked vertical arrows in the extreme left and right columns induce isomorphisms of cohomology groups. In this case, $\mathcal{X}_s^h$ is even a strong deformation retract of $\mathcal{X}^h$, by the results of C. H. Clemens (see [Cle, §6]), and *both maps* $\mathcal{X}_\eta^{\text{an}} \times]0, r[\to \mathcal{X}_0^{\text{An}}$ *and* $\mathcal{X}_s^{\text{an}} \to \mathcal{X}_0^{\text{An}}$ *are homotopy equivalences*, by results from [Ber5], as we are going to explain.

Consider a more general situation. Let k be an arbitrary field (instead of $\mathbf{C}$) provided with the trivial valuation. The ring of formal power series $k[[z]]$ coincides with the ring $\mathcal{O}_{\mathbf{A}^1,0}$ of formal power series convergent in an open neighborhood of zero in the affine line $\mathbf{A}^1$ over k as well as with the ring $\mathcal{O}(D(1))$ of those power series that are convergent in the open disc $D(1)$ (of radius one with center at zero). The formal spectrum $\mathfrak{X} = \mathrm{Spf}(k[[z]])$ is a special formal scheme over $k^\circ = k$ in the sense of [Ber4], and its generic fiber $\mathfrak{X}_\eta$ coincides with $D(1)$. Notice that there is a canonical homeomorphism $[0,1[\stackrel{\sim}{\to} D(1)$: $\rho \mapsto P_\rho$, where P_ρ is defined by $|z(P_\rho)| = \rho$.

Let $\mathcal{X}$ be a scheme of finite type over $k[[z]]$. For any number $0 < r < 1$, the ring $k[[z]]$ coincides with the k-affinoid algebra $k\{r^{-1}z\}$, the algebra of analytic functions on the closed disc $E(r) \subset \mathbf{A}^1$ (which is canonically homeomorphic to $[0,r]$), and so there is an associated k-analytic space $\mathcal{Y}^{\mathrm{an}}(r)$. If $r < r'$, $\mathcal{X}^{\mathrm{an}}(r)$ is identified with a closed analytic subdomain of $\mathcal{X}^{\mathrm{an}}(r')$, and we set $\mathcal{X}^{\mathrm{an}} = \cup \mathcal{X}^{\mathrm{an}}(r)$. There is a canonical surjective morphism $\varphi : \mathcal{X}^{\mathrm{an}} \to D(1) \stackrel{\sim}{\to} [0,1[$. The fiber $\varphi^{-1}(\rho)$ at $\rho \in [0,1[$ is identified with the $\mathcal{H}(P_\rho)$-analytic space $\mathcal{X}^{\mathrm{an}}_\rho$ associated with the scheme $\mathcal{X} \otimes_{k[[z]]} \mathcal{H}(P_\rho)$. The formal completion $\widehat{\mathcal{X}}$ of $\mathcal{X}$ along its closed fiber $\mathcal{X}_s$ is a special formal scheme, and there is a canonical morphism of strictly k-analytic spaces $\widehat{\mathcal{X}}_\eta \to \mathcal{X}^{\mathrm{an}}$ whose composition with the above morphism φ is induced by the canonical morphism of formal schemes $\widehat{\mathcal{X}} \to \mathfrak{X}$. If $\mathcal{X}$ is separated and of finite type over $k[[z]]$, $\widehat{\mathcal{X}}_\eta$ is identified with a closed strictly analytic subdomain of $\mathcal{X}^{\mathrm{an}}$. If $\mathcal{X}$ is proper over $k[[z]]$, then $\widehat{\mathcal{X}}_\eta \stackrel{\sim}{\to} \mathcal{X}^{\mathrm{an}}$. If $\mathcal{X}$ is semistable over $k[[z]]$, then so is $\widehat{\mathcal{X}}$.

Assume now that $\mathfrak{Y}$ is a semistable formal scheme over $\mathfrak{X} = \mathrm{Spf}(k[[z]])$ (or, more generally, polystable in the sense of [Ber5]). For $\rho \in [0,1[$, we set $\mathfrak{Y}_\rho = \mathfrak{Y} \times_{\mathfrak{X}} \mathrm{Spf}(\mathcal{H}(P_\rho)^\circ)$. It is a semistable formal scheme of $\mathcal{H}(P_\rho)$, and there are canonical isomorphisms $\mathfrak{Y}_{\rho,\eta} \stackrel{\sim}{\to} \mathfrak{Y}_{\eta,\rho}$ and $\mathfrak{Y}_{\rho,s} \stackrel{\sim}{\to} \mathfrak{Y}_s$. In [Ber5, §5], we constructed a closed subset $S(\widehat{\mathfrak{Y}}_\rho)$ (the skeleton of $\mathfrak{Y}_\rho$) and a strong deformation retraction $\Phi_\rho : \mathfrak{Y}_{\rho,\eta} \times [0,1] \to \mathfrak{Y}_{\rho,\eta}$ of $\mathfrak{Y}_{\rho,\eta}$ to the skeleton $S(\widehat{\mathfrak{Y}}_\rho)$. We denote by $S(\mathfrak{Y}/\mathfrak{X})$ the union of $S(\widehat{\mathfrak{Y}}_\rho)$ over all $\rho \in [0,1[$, and by Φ the mapping $\mathfrak{Y}_\eta \times [0,1] \to \mathfrak{Y}_\eta$ that coincides with Φ_ρ at each fiber of φ. In [Ber5, §4], we also associated with the closed fiber $\mathfrak{Y}_s$ of $\mathfrak{Y}$ a simplicial set $\mathbf{C}(\mathfrak{Y}_s)$ that has a geometric realization $|\mathbf{C}(\mathfrak{Y}_s)|$. Thus, to prove the claim, it suffices to verify the following two facts:

(a) the mapping $\Phi : \mathfrak{Y}_\eta \times [0,1] \to \mathfrak{Y}_\eta$ is continuous and compact, and
(b) there is a homeomorphism $|\mathbf{C}(\mathfrak{Y}_s)| \times \mathfrak{X}_\eta \to S(\mathfrak{Y}/\mathfrak{X})$ that commutes with the canonical projections to $[0,1[$.

(a) The assertion follows from the proof of [Ber5, Theorem 7.1]. In the formulation of the latter, the formal scheme $\mathfrak{X}$ was in fact assumed to be locally finitely presented over the ring of integers of the ground field (in our case $k^\circ = k$), but its proof uses only the fact that the morphism $\mathfrak{Y} \to \mathfrak{X}$ is polystable and works in the case when $\mathfrak{X}$ is an arbitrary special formal scheme.

(b) By the properties of the skeleton established in [Ber5, §5], the situation is easily reduced to the case $\mathfrak{Y} = \mathrm{Spf}(B)$, where

$$B = k[[z]]\{T_0, \ldots, T_m\}/(T_0 \cdots T_n - z)\ ,\ 0 \le n \le m.$$

If $n = 0$, then $S(\mathfrak{Y}/\mathfrak{X}) \xrightarrow{\sim} \mathfrak{X}_\eta \xrightarrow{\sim} [0,1[$, $|\mathbf{C}(\mathfrak{Y}_s)|$ is a point, and (b) follows. Assume that $n \geq 1$. Then $S(\mathfrak{Y}/\mathfrak{X})$ is identified with the set $\{(P_\rho, r_0, \ldots, r_n) \in \mathfrak{X}_\eta \times [0,1]^{n+1} \big| r_0 \cdots r_n = \rho\}$, and $|\mathbf{C}(\mathfrak{Y}_s)|$ is identified with the set $\{(u_0, \ldots, u_n) \in [0,1]^{n+1} \big| u_0 + \cdots + u_n = 1\}$. The required map $|\mathbf{C}(\mathfrak{Y}_s)| \times \mathfrak{X}_\eta \to S(\mathfrak{Y}/\mathfrak{X})$ takes a point $((u_0, \ldots, u_n), \rho)$ to $(P_\rho, (r_0, \ldots, r_n))$, where $(r_0, \ldots, r_n)$ is the point of intersection of the line, connecting the points $(u_0, \ldots, u_n)$ and $(1, \ldots, 1)$, and the hypersurface defined by the equation $t_0 \cdot \cdots \cdot t_n = \rho$. $\square$

Remarks 4.4. (i) The equality $(*_\eta)$ can be established in a different way. Namely, *we claim that there is a strong deformation retraction of $\mathcal{D}^*(r)$ to the subset $\lambda_\eta^{-1}(0)$ (identified with $]0, r[$).* Indeed, let P_ρ denote the point of $\lambda_\eta^{-1}(0)$ that corresponds to $\rho \in]0, r[$ (it is a unique point from $\lambda_\eta^{-1}(0)$ with $|T(P_\rho)| = \rho$). Then the required strong deformation retraction $\Psi : \mathcal{D}^*(r) \times [0,1] \to \mathcal{D}^*(r)$ (with $\Psi(x, 1) = x$ and $\Psi(x, 0) \in \lambda_\eta^{-1}(0)$) is defined as follows:

(1) if $(\rho e^{i\varphi}, s) \in \widetilde{D^*(r)}$, then $\Psi((\rho e^{i\varphi}, s), t) = (\rho^{\frac{1}{t}} e^{i\varphi}, st) \in \widetilde{D^*(r)}$ for $t \in]0, 1]$;
(2) $\Psi((\rho e^{i\varphi}, s), 0) = P_{\rho^s}$;
(3) if $\rho \in]0, r[$, then $\Psi(P_\rho, t) = P_\rho$ for all $t \in [0, 1]$.

The claim implies that if the sheaf L is constant, then $H^p(\mathcal{D}^*(r), \pi^* L) \xrightarrow{\sim} H^p(]0, r[, \pi^* L) = 0$, and therefore, $H^p_{\Phi_\eta}(\widetilde{D^*(r)}, \pi^* L) = 0$ for all $p \geq 0$. Thus, the equality $(*_\eta)$ is true for constant L, and is easily extended to arbitrary quasi-unipotent sheaves L.

(ii) Let $\mathcal{X}$ be a connected projective scheme over F that admits a strictly semistable reduction over $\mathcal{O}_{\mathbf{C},0}$. *Then the fundamental group of $\mathcal{X}^{\mathrm{an}}$ is isomorphic to a quotient of the fundamental group of the fiber $\mathcal{X}_t^h$, $t \in D^*(r)$,* and in particular, if the latter is simply connected, then so is $\mathcal{X}^{\mathrm{an}}$. Indeed, let $\mathcal{Y}$ be a projective strictly semistable scheme over $\mathcal{O}_{\mathbf{C},0}$ with $\mathcal{Y}_\eta = \mathcal{X}$. The canonical morphism $\mathcal{Y} \to \mathrm{Spec}(\mathcal{O}_{\mathbf{C},0})$ has a section $\mathrm{Spec}(\mathcal{O}_{\mathbf{C},0}) \to \mathcal{Y}$ (with the image in the smooth locus of that morphism), and therefore, for some $0 < r < 1$, the canonical morphism $\mathcal{Y}^h \to D(r)$ has a section $D(r) \to \mathcal{Y}^h$. It follows that the canonical surjection $\pi_1(\mathcal{X}^h) \to \pi_1(D^*)$ has a section whose image lies in the kernel of the canonical homomorphism $\pi_1(\mathcal{X}^h) \to \pi_1(\mathcal{Y}^h)$, and therefore, the image of $\pi_1\left(\mathcal{X}_t^h\right)$ in $\pi_1(\mathcal{Y}^h)$ coincides with that of $\pi_1(\mathcal{X}^h)$. But the canonical homomorphism $\pi_1(\mathcal{X}^h) \to \pi_1(\mathcal{Y}^h)$ is surjective, since the preimage of $\mathcal{X}^h$ in a universal covering of $\mathcal{Y}^h$ is connected (it is the complement of a Zariski closed subset of a connected smooth complex analytic space). Thus, $\pi_1(\mathcal{Y}^h)$ is a quotient of $\pi_1\left(\mathcal{X}_t^h\right)$. Furthermore, by the result of Clemens [Cle] used in the proof of Theorem 4.1, $\mathcal{Y}_s^h$ is a strong deformation retraction of $\mathcal{Y}^h$, i.e., $\pi_1\left(\mathcal{Y}_s^h\right)$ is a quotient of $\pi_1\left(\mathcal{X}_t^h\right)$. If C is the simplicial set associated with the scheme $\mathcal{Y}_s$, there is a surjective homomorphism from $\pi_1\left(\mathcal{Y}_s^h\right)$ to the fundamental group of the geometric realization $|C|$ of C. It remains to notice that by [Ber5, Theorem 5.2], $\mathcal{X}^{\mathrm{an}}$ is homotopy equivalent to $|C|$. (I am grateful to O. Gabber for the above reasoning.)

(iii) Assume that $\mathcal{X}$ is proper and strictly semistable over $\mathcal{O}_{\mathbf{C},0}$. It is very likely that all of the maps in the diagram from the beginning of this section, except those marked by $*$, are in fact homotopy equivalences.

(iv) It would be interesting to know whether Theorem 4.1 is true for not necessarily proper schemes.

5 An interpretation of the weight zero subspaces

Let $\mathcal{X}$ be a proper scheme over F, and let $0 < r < 1$ be small enough that the isomorphisms $H^q(\mathcal{X}^{\mathrm{an}} \times]0, r[, \mathbf{Z}) = H^q(\mathcal{X}^{\mathrm{an}}, \mathbf{Z}) \xrightarrow{\sim} H^q(\mathcal{X}^{\mathrm{An}}, \mathbf{Z})$ from Theorem 4.1 take place. They give rise to homomorphisms $H^q(\mathcal{X}^{\mathrm{an}}, \mathbf{Z}) \to H^q(\mathcal{X}^h, \mathbf{Z})$, $q \geq 0$. Let $\overline{D^*}(r) \to D^*(r)$ be a universal covering of $D^*(r)$. The fundamental group $\pi_1(D^*) = \pi_1(D^*(r), t)$ (which does not depend on the choice of r and a point $t \in D^*(r)$) acts on $\overline{D^*}(r)$, and therefore, it acts on $\overline{\mathcal{X}}^h = \mathcal{X}^h \times_{D^*(r)} \overline{D^*}(r)$. Furthermore, let F^{a} be the field of functions meromorphic in the preimage of an open neighborhood of zero in $\overline{D^*}(r)$ that are algebraic over F. It is an algebraic closure of F, and in particular, $\pi_1(D^*)$ acts on F^{a}. Let K^{a} be the corresponding algebraic closure of K, $\widehat{K}^{\mathrm{a}}$ the completion of K^{a}, and $\overline{\mathcal{X}}^{\mathrm{an}} = (\mathcal{X}^{\mathrm{an}} \otimes_F \widehat{K}^{\mathrm{a}})^{\mathrm{an}}$. As was mentioned in the introduction, the constructed homomorphisms of cohomology groups induce $\pi_1(D^*)$-equivariant homomorphisms $H^q(\overline{\mathcal{X}}^{\mathrm{an}}, \mathbf{Z}) \to H^q(\overline{\mathcal{X}}^h, \mathbf{Z})$, $q \geq 0$.

Theorem 5.1. *The above homomorphisms give rise to $\pi_1(D^*)$-equivariant isomorphisms*

$$H^q(\overline{\mathcal{X}}^{\mathrm{an}}, \mathbf{Q}) \xrightarrow{\sim} W_0 H^q(\overline{\mathcal{X}}^h, \mathbf{Q}), \ q \geq 0.$$

Proof. Consider first the case $\mathcal{X} = \mathcal{Y}_\eta$, where $\mathcal{Y}$ is a projective strictly semistable scheme over $\mathcal{O}_{\mathbf{C},0}$. By Corollary 2.4, in the commutative diagram of Theorem 4.1, for such $\mathcal{Y}$ the maps from the lower row give rise to an isomorphism $H^q(\mathcal{Y}^{\mathrm{an}}, \mathbf{Q}) \xrightarrow{\sim} W_0 H^q\left(\mathcal{Y}_s^h, \mathbf{Q}\right)$, and by Steenbrink's work [St], the homomorphisms $H^q\left(\mathcal{Y}_s^h, \mathbf{Z}\right) \xrightarrow{\sim} H^q\left(\mathcal{Y}^h, \mathbf{Z}\right) \to H^q(\mathcal{X}^h, \mathbf{Z}) \to H^q(\overline{\mathcal{X}}^h, \mathbf{Z})$ give rise to an isomorphism $W_0 H^q\left(\mathcal{Y}_s^h, \mathbf{Q}\right) \xrightarrow{\sim} W_0 H^q(\overline{\mathcal{X}}^h, \mathbf{Q})$, $q \geq 0$. Since the residue field of K is algebraically closed, the canonical map $\overline{\mathcal{X}}^{\mathrm{an}} \to \mathcal{X}^{\mathrm{an}}$ is a homotopy equivalence (see [Ber5, §5]) and, in particular, $H^q(\mathcal{X}^{\mathrm{an}}, \mathbf{Z}) \xrightarrow{\sim} H^q(\overline{\mathcal{X}}^{\mathrm{an}}, \mathbf{Z})$. Thus, the required isomorphism follows from Theorem 4.1. Consider now the case when $\mathcal{X}$ is projective and smooth over F. One can find an integer $n \geq 1$ such that, if F' is the cyclic extension of F of degree n in F^{a}, then the scheme $\mathcal{X}' = \mathcal{X} \otimes_F F'$ is of the previous type over F'. The extensions $F^{\mathrm{a}} \supset F' \supset F$ correspond to morphisms $\overline{D^*}(r) \to D^*\left(r^{\frac{1}{n}}\right) \xrightarrow{z \mapsto z^n} D^*(r)$, and so there is a canonical isomorphism of complex analytic spaces $\overline{\mathcal{X}}'^h \xrightarrow{\sim} \overline{\mathcal{X}}^h$. The latter gives rise to isomorphisms $H^q(\overline{\mathcal{X}}^h, \mathbf{Z}) \xrightarrow{\sim} H^q(\overline{\mathcal{X}}'^h, \mathbf{Z})$ of cohomology groups provided with the limit mixed Hodge structures (see [GNPP, p. 126]). Similarly, one has canonical isomorphisms $H^q(\overline{\mathcal{X}}^{\mathrm{an}}, \mathbf{Z}) \xrightarrow{\sim} H^q(\overline{\mathcal{X}}'^{\mathrm{an}}, \mathbf{Z})$, and the required isomorphism follows from the previous case. Finally, if $\mathcal{X}$ is an arbitrary proper scheme over F, one gets the required isomorphism using the same reasoning as in the proof of Corollary 2.4, i.e., using a proper hypercovering $\mathcal{X}_\bullet \to \mathcal{X}$ with projective and smooth $\mathcal{X}_n$'s and the fact that the functor $H \mapsto W_0 H$ on the category of rational mixed Hodge structures H with $W_i H = 0$ for $i < 0$ is exact. □

Corollary 5.2. *In the above situation, the following is true:*

(i) $H^q(\mathcal{X}^{\mathrm{an}}, \mathbf{Q}) \xrightarrow{\sim} (W_0 H^q(\overline{\mathcal{X}}^h, \mathbf{Q}))^{T=1}$*;*
(ii) if $\mathcal{X}$ *is projective and smooth, then*

$$H^q(\mathcal{X}^{\mathrm{an}}, \mathbf{Q}) \xrightarrow{\sim} ((T^m - 1)^i H^q\left(\mathcal{X}_t^h, \mathbf{Q}\right))^{T=1}.$$

Here T is the canonical generator of $\pi_1(D^*)$, and m is a positive integer for which the action of $(T^m - 1)^{i+1}$ on $H^q\left(\mathcal{X}_t^h, \mathbf{Q}\right)$ is zero (see the introduction).

Proof. It suffices to show that the canonical map

$$H^q(\mathcal{X}^{\mathrm{an}}, \mathbf{Q}) \to H^q(\overline{\mathcal{X}}^{\mathrm{an}}, \mathbf{Q})^{T=1}$$

is an isomorphism. For this we recall that by [Ber5, Theorem 10.1], one has $H^q\left((\mathcal{X} \otimes_F K')^{\mathrm{an}}, \mathbf{Q}\right) \xrightarrow{\sim} H^q(\overline{\mathcal{X}}^{\mathrm{an}}, \mathbf{Q})$ for some finite Galois extension K' of K in K^{a}. Since the topological space $\mathcal{X}^{\mathrm{an}}$ is the quotient of $(\mathcal{X} \otimes_F K')^{\mathrm{an}}$ by the action of the Galois group of K' over K, the required fact follows from [Gro, Corollary 5.2.3]. □

Remark 5.3. As was mentioned at the end of the introduction, Theorem 5.1 is an analogue of a similar description of the weight zero subspaces in the l-adic cohomology groups of algebraic varieties over a local field, which holds for arbitrary separated schemes of finite type (see [Ber6]). And so it is very likely that the isomorphism of Theorem 5.1 also takes place for arbitrary separated schemes of finite type over F. The latter would follow from the validity of Theorem 4.1 for that class of schemes (see Remark 4.4(iv)), and is easily extended to separated smooth schemes. (Recall that the theory of limit mixed Hodge structures on the cohomology groups $H^q(\overline{\mathcal{X}}^h, \mathbf{Q})$ for separated schemes $\mathcal{X}$ of finite type over F is developed in [EZ].)

References

[Ber1] Berkovich, V. G., *Spectral theory and analytic geometry over non-Archimedean fields*, Math. Surveys Monogr. 33, Amer. Math. Soc., Providence, R.I., 1990.

[Ber2] Berkovich, V. G., *Étale cohomology for non-Archimedean analytic spaces*, Publ. Math. IHES **78** (1993), 5–161.

[Ber3] Berkovich, V. G., *Vanishing cycles for non-Archimedean analytic spaces*, J. Amer. Math. Soc. **9** (1996), no. 4, 1187–1209.

[Ber4] Berkovich, V. G., *Vanishing cycles for formal schemes. II*, Invent. Math. **125** (1996), no. 2, 367–390.

[Ber5] Berkovich, V. G., *Smooth p-adic analytic spaces are locally contractible*, Invent. Math. **137** (1999), no. 1, 1–84.

[Ber6] Berkovich, V. G., *An analog of Tate's conjecture over local and finitely generated fields*, Internat. Math. Res. Notices **2000**, no. 13, 665–680.

[Cle] Clemens, C. H., *Degeneration of Kähler manifolds*, Duke Math. J. **44** (1977), no. 2, 215–290.

[Del1] Deligne, P., *Théorie de Hodge. I*, in Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 1, pp. 425–430. Gauthier-Villars, Paris, 1971.

[Del2] Deligne, P., *Théorie de Hodge. II*, Publ. Math. IHES **40** (1971), 5–57.

[Del3] Deligne, P., *Théorie de Hodge. III*, Publ. Math. IHES **44** (1974), 5–77.

[EZ] El Zein, F., *Théorie de Hodge des cycles évanescents*, Ann. Sci. École Norm. Sup. **19** (1986), no. 1, 107–184.

[GNPP] Guillén, F.; Navarro Aznar, V.; Pascual Gainza, P.; Puerta, F., *Hyperrésolutions cubiques et descente cohomologique*, Papers from the Seminar on Hodge–Deligne Theory held in Barcelona, 1982. Lecture Notes in Math. **1335**, Springer, Berlin–Heidelberg–New York, 1988.

[God] Godement, R., *Topologie algébrique et théorie des faisceaux*, Paris, Hermann, 1958.

[Gro] Grothendieck, A., *Sur quelques points d'algèbre homologique*, Tôhoku Math. J. **9** (1957), 119–221.

[EGAII] Grothendieck, A.; Dieudonné, J., *Éléments de géométrie algébrique. II. Étude globale élémentaire de quelques classes de morphismes*, Publ. Math. IHES **8** (1961).

[EGAIV] Grothendieck, A.; Dieudonné, J., *Eléments de géométrie algébrique. IV. Étude locale des schémas et des morphismes de schémas*, Publ. Math. IHES **20** (1964), **25** (1965), **28** (1966), **32** (1967).

[Ill] Illusie, L., *Autour du théorème de monodromie locale*, in Périodes p-adiques, Astérisque **223** (1994), 9–57.

[Serre] Serre J.-P., *Géométrie algébrique et géométrie analytique*, Ann. Inst. Fourier, Grenoble **6** (1955-1956), 1–42.

[SGA4] Artin, M; Grothendieck, A; Verdier, J.-L., *Théorie des Topos et Cohomologie Étale des Schémas*, Lecture Notes in Math. **269, 270, 305**, Springer, Berlin–Heidelberg–New York, 1972–1973.

[St] Steenbrink, J., *Limits of Hodge structures*, Invent. Math. **31** (1976), no. 3, 229–257.

Analytic Curves in Algebraic Varieties over Number Fields

Jean-Benoît Bost[1] and Antoine Chambert-Loir[2]

[1] Département de mathématiques, Université de Paris-Sud (Orsay), Bâtiment 425, F-91405 Orsay Cedex, France. Jean-Benoit.Bost@math.u-psud.fr

[2] Université de Rennes 1, Institut de recherche mathématique de Rennes, Campus de Beaulieu, F-35042 Rennes Cedex, France. antoine.chambert-loir@univ-rennes1.fr

To Yuri Ivanovich Manin on the occasion of his 70th birthday

Summary. We establish algebraicity criteria for formal germs of curves in algebraic varieties over number fields and apply them to derive a rationality criterion for formal germs of functions on algebraic curves that extends the classical rationality theorems of Borel–ŰDwork and Polya–ŰBertrandias, valid over the projective line, to arbitrary algebraic curves over a number field. The formulation and the proof of these criteria involve some basic notions in Arakelov geometry, combined with complex and rigid analytic geometry (notably, potential theory over complex and p-adic curves). We also discuss geometric analogues, pertaining to the algebraic geometry of projective surfaces of these arithmetic criteria.

Key words: Arakelov geometry, capacity theory, BorelŰ–Dwork rationality criterion, algebricity of formal subschemes, rigid analytic geometry, slope inequality.

2000 Mathematics Subject Classifications: 14G40 (Primary); 14G22, 31A15, 14B20 (Secondary)

1 Introduction

The purpose of this article is to establish *algebraicity criteria* for *formal germs of curves* in algebraic varieties over number fields and to apply them to derive a *rationality criterion* for *formal germs of functions*, which extends the classical rationality theorems of Borel–Dwork [6], [22] and Pólya–Bertrandias [1, Chapter 5], [43], (see also [16]), valid over the projective line, to arbitrary algebraic curves over a number field.

Our algebraicity criteria improve on those in [12] and [13], which themselves were inspired by the papers [19] and [20] of D. V. and G. V. Chudnovsky

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*, Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_3,

and by the subsequent works by André [2] and Graftieaux [26, 27]. As in [12] and [13], our results will be proved by means of a geometric version of "transcendence techniques," which avoids the traditional constructions of "auxiliary polynomials" and the explicit use of Siegel's lemma, replacing them by a few basic concepts of Arakelov geometry. In the proofs, our main objects of interest will be some evaluation maps, defined on the spaces of global sections of powers of an ample line bundle on a projective variety by restricting these sections to formal subschemes or to subschemes of finite lengths. Arakelov geometry enters through the estimates satisfied by the heights of the evaluation maps, and the slopes and Arakelov degrees of the hermitian vector bundles defined by spaces of sections (see [17] and [14] for more details and references on this approach).

Our main motivation in investigating the algebraicity and rationality criteria presented in this article has been the desire to obtain theorems respecting the classical principle of number theory that "all places of number fields should appear on an equal footing"—which actually is not the case in "classical" Arakelov geometry and in its applications in [12]. A closely related aim has been to establish arithmetic theorems whose geometric counterparts (obtained through the analogy between number fields and function fields) have simple formulations and proofs. These concerns led us to two technical developments in this paper: the use of (rigid) analytic geometry over p-adic fields to define and estimate local invariants of formal curves over number fields,[3] and the derivation of a rationality criterion from an algebraicity criterion by means of the Hodge index theorem on (algebraic or arithmetic) surfaces.

Let us describe the contents of this article in more detail.

In Section 2, we discuss geometric analogues of our arithmetic theorems. Actually, these are classical results in algebraic geometry, going back to Hartshorne [32] and Hironaka–Matsumura [35]. For instance, our algebraicity result admits as analogue the following fact. *Let X be a quasiprojective variety over a field k, and let Y be a smooth projective integral curve in X; let $\widehat{S}$ be a smooth formal germ of surface through Y* (that is, a smooth formal subscheme of dimension 2, containing Y, of the completion $\widehat{X}_Y$). *If the degree* $\deg_Y N_Y\widehat{S}$ *of the normal bundle to Y in $\widehat{S}$ is positive, then $\widehat{S}$ is algebraic.*

Our point is that, transposed to a geometric setting, the arguments leading to our algebraicity and rationality criteria in the arithmetic setting—which rely on the consideration of suitable evaluation maps and on the Hodge index theorem—provide simple proofs of these nontrivial algebro-geometric results, in which the geometric punch line of the arguments appears more clearly.

In Section 3, we introduce the notion of *A-analytic curve* in an algebraic variety X over a number field K. By definition, this will be a smooth for-

[3]Since the first version of this paper was written, the relevance of rigid analytic geometry à la Berkovich to develop a non-archimedean potential theory on p-adic curves, and consequently a "modern" version of Arakelov geometry of arithmetic surfaces satisfying the above principle of "equality of places," has been largely demonstrated by A. Thuillier in his thesis [51].

mal curve $\widehat{C}$ through a rational point P in $X(K)$—that is, a smooth formal subscheme of dimension 1 in the completion $\widehat{X}_P$—which, firstly, is analytic at every place of K, finite or infinite. Namely, if v denotes any such place and K_v the corresponding completion of K, the formal curve $\widehat{C}_{K_v}$ in X_{K_v} deduced from $\widehat{C}$ by the extension of scalars $K \hookrightarrow K_v$ is the formal curve defined by a K_v-analytic curve in $X(K_v)$. Moreover the v-adic radius r_v (in $]0,1]$) of the open ball in $X(K_v)$ in which $\widehat{C}_{K_v}$ "analytically extends" is required to "stay close to 1 when v varies," in the sense that the series $\sum_v \log r_v^{-1}$ has to be convergent. The precise formulation of this condition relies on the notion of *size* of a smooth analytic germ in an algebraic variety over a p-adic field. This notion was introduced in [12, 3.1]; we review it in Section 3.A, adding some complements.

With the above notation, if $\mathscr{X}$ is a model of X over the ring of integers $\mathscr{O}_K$ of K, and if P extends to an integral point $\mathscr{P}$ in $\mathscr{X}(\mathscr{O}_K)$, then a formal curve $\widehat{C}$ through P is A-analytic if it is analytic at each archimedean place of K and extends to a smooth formal surface $\widehat{\mathscr{C}}$ in $\widehat{\mathscr{X}_{\mathscr{P}}}$. For a general formal curve $\widehat{C}$ that is analytic at archimedean places, being an A-analytic germ may be seen as a weakened form of the existence of such a smooth extension $\widehat{\mathscr{C}}$ of $\widehat{C}$ along $\mathscr{P}$. In this way, an A-analytic curve through the point P appears as an arithmetic counterpart of the smooth formal surface $\widehat{S}$ along the curve Y in the geometric algebraicity criterion above.

The tools needed to formulate the arithmetic counterpart of the positivity condition $\deg_Y N_Y \widehat{S} > 0$ are developed in Sections 4 and 5. We first show in Section 4 how, for any germ of analytic curve $\widehat{C}$ through a rational point P in some algebraic variety X over a local field K, one is led to introduce the so-called *canonical seminorm* $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C}}$ on the K-line $T_P\widehat{C}$ through the consideration of the metric properties of the evaluation maps involved in our geometric version of the method of auxiliary polynomials. This extends a definition introduced in [13] when $K = \mathbf{C}$. In Section 5, we discuss the construction of Green functions and capacities on rigid analytic curves over p-adic fields. We then extend the comparison of "canonical seminorms" and "capacitary metrics" in [13], 3.4, to the non-archimedean setting.

In Section 6, we apply these notions to formulate and establish our algebraicity results. If $\widehat{C}$ is an A-analytic curve through a rational point P in an algebraic variety X over some number field K, then the K-line $T_P\widehat{C}$ may be equipped with a "K_v-adic semi-norm" for every place v by the above construction—namely, the seminorm $\|\cdot\|^{\mathrm{can}}_{X_{K_v},\widehat{C}_{K_v}}$ on

$$T_P\widehat{C}_{K_v} \simeq T_P\widehat{C} \otimes_K K_v.$$

The so-defined metrized K-line $\overline{T_P\widehat{C}}$ has a well-defined Arakelov degree in $]-\infty,+\infty]$, and our main algebraicity criterion asserts that $\widehat{C}$ *is algebraic if the Arakelov degree* $\widehat{\deg}\,\overline{T_P\widehat{C}}$ *is positive.* Actually, the converse implication also holds: when $\widehat{C}$ is algebraic, the canonical seminorms $\|.\|^{\mathrm{can}}_{X_{K_v},\widehat{C}_{K_v}}$ all vanish, and $\widehat{\deg}\,\overline{T_P\widehat{C}} = +\infty$.

Finally, in Section 7, we derive an extension of the classical theorems of Borel, Dwork, Pólya, and Bertrandias, which gives a criterion for the rationality of a formal germ of function φ on some algebraic curve Y over a number field. By considering the graph of φ—a formal curve $\widehat{C}$ in the surface $X := Y \times \mathbf{A}^1$—we easily obtain the algebraicity of φ as a corollary of our previous algebraicity criterion. In this way, we are reduced to establishing a rationality criterion for an algebraic formal germ. Actually, rationality results for algebraic functions on the projective line have been investigated by Harbater [30], and used by Ihara [36] to study the fundamental group of some arithmetic surfaces. Ihara's results have been extended in [11] using Arakelov geometry on arithmetic surfaces. Our rationality argument in Section 7, based on the Hodge index theorem on arithmetic surfaces of Faltings–Hriljac, is a variation on the proof of the Lefschetz theorem on arithmetic surfaces in [11].

It is a pleasure to thank A. Ducros for his helpful advice on rigid analytic geometry during the preparation of this article.

Some of the results below were presented, in a preliminary form, during the "Arithmetic Geometry and Number Theory" conference in honor of N. Katz, in Princeton, December 2003, and were announced in [14].

During the preparation of this article, the authors benefitted from the support of the Institut Universitaire de France.

It would be difficult to acknowledge fairly the multifaceted influences of Yuri Ivanovich Manin on our work. We hope that this article will appear as a tribute, not only to his multiple contributions to algebraic geometry and number theory, but also to his global vision of mathematics, emphasizing geometric insights and analogies. The presentation of this vision in his 25th-Arbeitstagung report *New Dimensions in Geometry* [38] has been, since it was written, a source of wonder and inspiration to one of the authors, and we allowed ourselves to borrow the terminology "A-analytic" from the "A-geometry" programmatically discussed in [38]. It is an honor for us to dedicate this article to Yuri Ivanovich Manin.

2 Preliminary: the geometric case

The theorems we want to prove in this paper are analogues in arithmetic geometry of classical algebro-geometric results going back—at least in an implicit form—to Hartshorne, Hironaka, and Matsumura ([31],[34],[35]). Conversely, in this section we give short proofs of algebraic analogues of our main arithmetic theorems.

Proposition 2.1. *Let $\mathscr{X}$ be a quasi-projective scheme over a field k and let $\mathscr{P}$ be a projective connected subscheme of dimension 1 in $\mathscr{X}$. Let $\widehat{\mathscr{C}}$ be a formal* *subscheme of dimension* 2 *in* $\widehat{\mathscr{X}_{\mathscr{P}}}$ *admitting* $\mathscr{P}$ *as a scheme of definition. Assume that* $\widehat{\mathscr{C}}$ *is (formally) smooth over* k*, and that* $\mathscr{P}$ *has no embedded component (of dimension* 0*), or equivalently, that* $\mathscr{P}$ *defines a Cartier divisor in* $\widehat{\mathscr{C}}$*, and let* $\mathscr{N}$ *be the normal bundle of the immersion* $\iota : \mathscr{P} \hookrightarrow \widehat{\mathscr{C}}$*, that is, the invertible sheaf* $\iota^* \mathscr{O}_{\widehat{\mathscr{C}}}(\mathscr{P})$ *on* $\mathscr{P}$.

If the divisor $[\mathscr{P}]$ *on the formal surface* $\widehat{\mathscr{C}}$ *is* nef *and has* positive self-intersection, *then the formal surface* $\widehat{\mathscr{C}}$ *is algebraic; namely, the Zariski-closure of* $\widehat{\mathscr{C}}$ *in* $\mathscr{X}$ *is an algebraic subvariety of dimension* 2.

Let $(\mathscr{P}_i)_{i\in I}$ be the family of irreducible components of $\mathscr{P}$, and $(n_i)_{i\in I}$ their multiplicities in $\mathscr{P}$. Recall that $[\mathscr{P}]$ is said to be nef on $\widehat{\mathscr{C}}$ when

$$[\mathscr{P}_i] \cdot [\mathscr{P}] := \deg_{\mathscr{P}_i} \mathscr{N} \geqslant 0 \ \text{ for any } i \in I,$$

and to have positive self-intersection if

$$[\mathscr{P}] \cdot [\mathscr{P}] := \sum_{i\in I} n_i . \deg_{\mathscr{P}_i} \mathscr{N} > 0,$$

or equivalently, when $[\mathscr{P}]$ is nef, if one the nonnegative integers $\deg_{\mathscr{P}_i} \mathscr{N}$ is positive. Observe that these conditions are satisfied if $\mathscr{N}$ is ample on $\mathscr{P}$.

More general versions of the algebraicity criterion in Proposition 2.1 and of its proof below, without restriction on the dimensions of $\widehat{\mathscr{C}}$ and $\mathscr{P}$, can be found in [12, §3.3], [5], [13, Theorem 2.5], (see also [17, 18]). Besides, it will be clear from the proof that, suitably reformulated, Proposition 2.1 still holds with the smoothness assumption on $\widehat{\mathscr{C}}$ omitted; we leave this to the interested reader.

Such algebraicity criteria may also be deduced from the works of Hironaka, Matsumura, and Hartshorne on the condition G_2 [34], [35], [31]. We refer the reader to the monographs [33] and [3] for extensive discussions and references about related results concerning formal functions and projective algebraic varieties.

Note that Proposition 2.1 has consequences for the study of algebraic varieties over function fields. Indeed, let S be a smooth, projective, and geometrically connected curve over a field k and let $K = k(S)$. Let $f\colon \mathscr{X} \to S$ be a surjective map of k-schemes and assume that $\mathscr{P}$ is the image of a section of f. Let $X = \mathscr{X}_K$, $P = \mathscr{P}_K$, and $\widehat{C} = \widehat{\mathscr{C}}_K$ be the generic fibers of $\mathscr{X}$, $\mathscr{P}$, and $\mathscr{C}$. Then P is a K-rational point of X and $\widehat{C}$ is a germ of curve in X at P. Observe that $\widehat{\mathscr{C}}$ is algebraic if and only if $\widehat{C}$ is algebraic. Consequently, in this situation, Proposition 2.1 appears as an algebraicity criterion for a formal germ of curves $\widehat{C}$ in X. In particular, it shows that such *a smooth formal curve* $\widehat{C}$ *in* X *is algebraic if it extends to a smooth formal scheme* $\widehat{\mathscr{C}}$ *through* $\mathscr{P}$ *in* $\mathscr{X}$ *such that the normal bundle of* $\mathscr{P}$ *in* $\widehat{\mathscr{C}}$ *has positive degree.*

Proof (of Proposition 2.1). We may assume that $\mathscr{X}$ is projective and that $\widehat{\mathscr{C}}$ is Zariski-dense in $\mathscr{X}$. We let $d = \dim \mathscr{X}$. One has obviously $d \geqslant 2$ and our goal is to prove the equality.

Let $\mathscr{O}(1)$ be any very ample line bundle on $\mathscr{X}$. The method of "auxiliary polynomials," borrowed from transcendence theory, suggests the introduction of the "evaluation maps"

$$\varphi_D \colon \Gamma(\mathscr{X}, \mathscr{O}(D)) \to \Gamma(\widehat{\mathscr{C}}, \mathscr{O}(D)), \quad s \mapsto s|_{\widehat{\mathscr{C}}},$$

for positive integers D.

Let us write $E_D = \Gamma(\mathscr{X}, \mathscr{O}(D))$, and for any integer $i \geqslant 0$, let E_D^i be the set of all $s \in E_D$ such that $\varphi_D(s) = s|_{\widehat{\mathscr{C}}}$ vanishes to order at least i along $\mathscr{P}$, i.e., such that the restriction of $\varphi_D(s)$ to $i\mathscr{P}$ vanishes. Since $\widehat{\mathscr{C}}$ is Zariski-dense in $\mathscr{X}$, no nonzero section of $\mathscr{O}(D)$ has a restriction to $\widehat{\mathscr{C}}$ that vanishes at infinite order along $\mathscr{P}$, and we have

$$\bigcap_{i=0}^{\infty} E_D^i = 0.$$

Consequently,

$$\operatorname{rank} E_D = \sum_{i=0}^{\infty} \operatorname{rank}\left(E_D^i / E_D^{i+1}\right).$$

Moreover, there is a canonical injective map of k-vector spaces

$$E_D^i / E_D^{i+1} \hookrightarrow \Gamma(\mathscr{P}, \mathscr{O}(D) \otimes \mathscr{N}^{\vee \otimes i}),$$

which amounts to taking the ith jet along $\mathscr{P}$—that is, the restriction to $(i+1)\mathscr{P}$—of a section that vanishes to order at least i. Indeed, the quotient sheaf

$$\left(\mathscr{O}(D) \otimes \mathscr{O}_{\widehat{\mathscr{C}}}(-i\mathscr{P})\right) / \left(\mathscr{O}(D) \otimes \mathscr{O}_{\widehat{\mathscr{C}}}(-(i+1)\mathscr{P})\right)$$

over $\widehat{\mathscr{C}}$ may be identified with $\mathscr{O}(D) \otimes \iota_* \mathscr{N}^{\vee \otimes i}$. Observe also that the dimension of the range of this injection satisfies an upper bound of the form

$$\dim \Gamma(\mathscr{P}, \mathscr{O}(D) \otimes \mathscr{N}^{\vee \otimes i}) \leqslant c(D+i),$$

valid for any nonnegative integers D and i.

Assume that $E_D^i \neq 0$ and let $s \in E_D^i$ be any nonzero element. By assumption, $\varphi_D(s)$ vanishes to order i along $\mathscr{P}$; hence $\operatorname{div} \varphi_D(s) - i[\mathscr{P}]$ is an effective divisor on $\widehat{\mathscr{C}}$ and its intersection number with $[\mathscr{P}]$ is nonnegative, for $[\mathscr{P}]$ is nef. Consequently

$$\operatorname{div} \varphi_D(s) \cdot [\mathscr{P}] \geqslant i[\mathscr{P}] \cdot [\mathscr{P}].$$

Since

$$\operatorname{div} \varphi_D(s) \cdot [\mathscr{P}] = \deg_{\mathscr{P}}(\mathscr{O}(D)) = D \deg_{\mathscr{P}}(\mathscr{O}(1))$$

and $[\mathscr{P}] \cdot [\mathscr{P}] > 0$ by the assumption of positive self-intersection, this implies $i \leqslant aD$, where $a := \deg_{\mathscr{P}} \mathscr{O}(1) / [\mathscr{P}] \cdot [\mathscr{P}]$. Consequently E_D^i is reduced to 0 if $i > aD$.

Finally, we obtain

$$\operatorname{rank} E_D = \sum_{i=0}^{\infty} \operatorname{rank}\left(E_D^i/E_D^{i+1}\right) = \sum_{i=0}^{\lfloor aD \rfloor} \operatorname{rank}\left(E_D^i/E_D^{i+1}\right) \leqslant \sum_{i=0}^{\lfloor aD \rfloor} c(D+i).$$

This proves that as D goes to $+\infty$,

$$\operatorname{rank} E_D \ll D^2.$$

On the other hand,

$$\operatorname{rank} E_D = \operatorname{rank} \Gamma(\mathscr{X}, \mathscr{O}(D)) \asymp D^d,$$

by Hilbert–Samuel's theorem. This establishes that the integer d, which is at least 2, actually equals 2. □

Proposition 2.2. *Let $f\colon S' \to S$ be a dominant morphism between two normal projective surfaces over a field k. Let $D \subset S$ and $D' \subset S'$ be effective divisors such that $f(D') = D$.*

Assume that $f|_{D'}\colon D' \to D$ is an isomorphism and that f induces an isomorphism $\widehat{f}\colon \widehat{S'_{D'}} \to \widehat{S_D}$ between formal completions. If, moreover, D is nef and $D \cdot D > 0$, then f is birational.

Recall that D is said to be nef if, for any effective divisor E on S, the (rational) intersection number $D \cdot E$ is nonnegative.

Proof. By hypothesis, f is étale in a neighborhood of D'. If $\deg(f) > 1$, one can therefore write $f^*D = D' + D''$, where D'' is a nonzero effective Cartier divisor on S' that is disjoint from D'. Now, f^*D is a nef divisor on S' such that $f^*D \cdot f^*D = \deg(f) D \cdot D > 0$. As a classical consequence of the Hodge index theorem (see [24], [45] and also [11, Proposition 2.2]), the effective divisor f^*D is numerically connected, hence connected. This contradicts the decomposition $f^*D = D' \sqcup D''$. □

Proposition 2.3. *Let $\mathscr{S}$ be a smooth projective connected surface over a perfect field k. Let $\mathscr{P}$ be a smooth projective connected curve in $\mathscr{S}$. If the divisor $[\mathscr{P}]$ on $\mathscr{S}$ is big and nef, then any formal rational function along $\mathscr{P}$ is defined by a (unique) rational function on $\mathscr{S}$. In other words, one has an isomorphism of fields*

$$k(\mathscr{S}) \xrightarrow{\sim} \Gamma(\mathscr{P}, \operatorname{Frac} \mathscr{O}_{\widehat{\mathscr{S}_{\mathscr{P}}}}).$$

Proof. Let φ be any formal rational function along $\mathscr{P}$. We may introduce a sequence of blowups of closed points $\nu\colon \mathscr{S}' \to \mathscr{S}$ such that $\varphi' = \nu^*\varphi$ has no point of indeterminacy and may be seen as a map (of formal k-schemes) $\widehat{\mathscr{S}'_{\mathscr{P}'}} \to \mathbf{P}^1_k$, where $\mathscr{P}' = \nu^*\mathscr{P}$.

Let us consider the graph $\operatorname{Gr}\varphi'$ of φ' in $\mathscr{S}' \times \mathbf{P}^1_k$. This is a formally smooth 2-dimensional formal scheme, admitting the graph of $\varphi'_{|\mathscr{P}'}\colon \mathscr{P}' \to \mathbf{P}^1_k$

as a scheme of definition, and the morphism φ' defines an isomorphism of formal schemes

$$\psi' := (\mathrm{Id}, \varphi) \colon \widehat{\mathscr{S}'}_{\mathscr{P}'} \to \mathrm{Gr}\,\varphi'.$$

Like the divisor $\mathscr{P}$ in $\mathscr{S}$, its inverse image $\mathscr{P}'$ in $\mathscr{S}'$ is nef and has positive self-intersection. Proposition 2.1 therefore implies that $\mathrm{Gr}\,\varphi'$ is algebraic in $\mathscr{S}' \times \mathbf{P}^1_k$. In other words, φ' is an algebraic function.

To establish its rationality, let us introduce the Zariski closure Γ of the graph of $\mathrm{Gr}\,\varphi'$ in $\mathscr{S}' \times \mathbf{P}^1_k$, the projections $\mathrm{pr}_1 \colon \Gamma \to \mathscr{S}'$ and $\mathrm{pr}_2 \colon \Gamma \to \mathbf{P}^1_k$, and the normalization $n \colon \tilde{\Gamma} \to \Gamma$ of Γ. Consider also the Cartier divisor $\mathscr{P}'_\Gamma$ (respectively $\mathscr{P}'_{\tilde{\Gamma}}$) defined as the inverse image $\mathrm{pr}_1^* \mathscr{P}'$ (respectively $n^* \mathscr{P}'_\Gamma$) of $\mathscr{P}'$ in Γ (respectively $\tilde{\Gamma}$). The morphisms n and pr_1 induce morphisms of formal completions:

$$\widehat{\tilde{\Gamma}}_{\mathscr{P}'_{\tilde{\Gamma}}} \xrightarrow{\widehat{n}} \widehat{\Gamma}_{\mathscr{P}'_\Gamma} \xrightarrow{\widehat{\mathrm{pr}_1}} \widehat{\mathscr{S}'}_{\mathscr{P}'}.$$

The morphism ψ' may be viewed as a section of $\widehat{\mathrm{pr}_1}$; by normality of $\widehat{\mathscr{S}'_{\mathscr{P}'}}$, it admits a factorization through $\widehat{n}$ of the form $\psi' = \widehat{n} \circ \tilde{\psi}$, for some uniquely determined morphism of k-formal schemes $\tilde{\psi} \colon \widehat{\mathscr{S}'}_{\mathscr{P}'} \to \widehat{\tilde{\Gamma}}_{\mathscr{P}'_{\tilde{\Gamma}}}$. This morphism $\tilde{\psi}$ is a section of $\widehat{\mathrm{pr}_1} \circ \widehat{n}$. Therefore the (scheme-theoretic) image $\tilde{\psi}(\mathscr{P}')$ defines a (Cartier) divisor in $\tilde{\Gamma}$ such that

$$(f \colon S' \to S, D', D) = (\mathrm{pr}_1 \circ n \colon \tilde{\Gamma} \to \mathscr{S}', \tilde{\mathscr{P}}, \mathscr{P}')$$

satisfy the hypotheses of Proposition 2.2. Consequently the morphism $\mathrm{pr}_1 \circ n$ is birational. Therefore, pr_1 is birational too and φ' is the restriction of a rational function on $\mathscr{S}'$, namely $\mathrm{pr}_2 \circ \mathrm{pr}_1^{-1}$. This implies that φ is the restriction of a rational function on $\mathscr{S}$. The uniqueness of this rational function follows from the Zariski density of the formal neighborhood of $\mathscr{P}$ in $\mathscr{S}$. □

Remark 2.4. In the terminology of Hironaka and Matsumura [35], the last proposition asserts that $\mathscr{P}$ is G_3 in $\mathscr{S}$, and has been established by Hironaka in [34]. Hartshorne observes in [32, Proposition 4.3, and Remark p. 123] that Proposition 2.2 holds more generally under the assumption that D and D' are G_3 in $\mathscr{S}$ and $\mathscr{S}'$. Our approach to Propositions 2.2 and 2.3 follows an order opposite to that in [34] and [32], and actually provides a simple proof of [32, Proposition 4.3].

3 *A*-analyticity of formal curves

3.A Size of smooth formal curves over *p*-adic fields

In this section, we briefly recall some definitions and results from [12].

Let K be field equipped with some complete ultrametric absolute value $|\cdot|$ and assume that its valuation ring R is a discrete valuation ring. Let also $\overline{K}$

be an algebraic closure of K. We shall still denote by $|.|$ the non-archimedean absolute value on $\overline{K}$ that extends the absolute value $|.|$ on K.

For any positive real number r, we define the norm $\|g\|_r$ of a formal power series $g = \sum_{I \in \mathbf{N}^N} a_I X^I \in K[[X_1, \ldots, X_n]]$ by the formula

$$\|g\|_r = \sup_I |a_I| r^{|I|} ;$$

it belongs to $\mathbf{R}_+ \cup \{\infty\}$. The power series g such that $\|g\|_r < \infty$ are precisely those that are convergent and bounded on the open N-ball of radius r in $\overline{K}^N$.

The group $G_{\mathrm{for},K} := \mathrm{Aut}\left(\widehat{\mathbf{A}^N_{K,0}}\right)$ of automorphisms of the formal completion of $\mathbf{A}^N_K$ at 0 may be identified with the set of all N-tuples $f = (f_1, \ldots, f_n)$ of power series in $K[[X_1, \ldots, X_N]]$ such that $f(0) = 0$ and $Df(0) := \left(\frac{\partial f_j}{\partial X_i}(0)\right)$ belongs to $\mathrm{GL}_N(K)$. We consider its following subgroups:

- the subgroups G_{for} consisting of all elements $f \in G_{\mathrm{for},K}$ such that $Df(0) \in \mathrm{GL}_N(R)$;
- the subgroup $G_{\mathrm{an},K}$ consisting of those $f = (f_1, \ldots, f_N)$ in $G_{\mathrm{for},K}$ such that for each j, f_j has a positive radius of convergence;
- $G_{\mathrm{an}} := G_{\mathrm{an},K} \cap G_{\mathrm{for}}$;
- for any positive real number r, the subgroup $G_{\mathrm{an},r}$ of G_{an} consisting of all N-tuples $f = (f_1, \ldots, f_N)$ such that $\|f_j\|_r \leqslant r$ for each j. This subgroup may be identified with the group of all analytic automorphisms, preserving the origin, of the open N-dimensional ball of radius r.

One has the inclusion $G_{\mathrm{an},r'} \subset G_{\mathrm{an},r}$ for any $r' > r > 0$, and the equalities

$$\bigcup_{r>0} G_{\mathrm{an},r} = G_{\mathrm{an}} \quad \text{and} \quad G_{\mathrm{an},1} = \mathrm{Aut}\left(\widehat{\mathbf{A}^N_{R,0}}\right).$$

It is straightforward that a formal subscheme $\widehat{V}$ of $\widehat{\mathbf{A}^N_{K,0}}$ is (formally) *smooth* of dimension d iff there exists $\varphi \in G_{\mathrm{for},K}$ such that $\varphi^* \widehat{V}$ is the formal subscheme $\widehat{\mathbf{A}^d_{K,0}} \times \{0\}$ of $\widehat{\mathbf{A}^N_{K,0}}$; when this holds, one can even find such a φ in G_{for}. Moreover such a smooth formal subscheme $\widehat{V}$ is K-*analytic* iff one can find φ as above in $G_{\mathrm{an},K}$, or equivalently in G_{an}.

Let $\mathscr{X}$ be a flat quasiprojective R-scheme, and $X = \mathscr{X} \otimes_R K$ its generic fiber. Let $\mathscr{P} \in \mathscr{X}(R)$ be an R-point, and let $P \in X(K)$ be its restriction to $\mathrm{Spec}\, K$. In [12, §3.1.1], we associated to any smooth formal scheme $\widehat{V}$ of dimension d in $\widehat{X}_P$, its *size* $S_{\mathscr{X}}(\widehat{V})$ *with respect to the model* $\mathscr{X}$ of X. It is a number in $[0, 1]$ whose definition and basic properties may be summarized in the following statement:

Theorem 3.1. *There is a unique way to attach a number* $S_{\mathscr{X}}(\widehat{V})$ *in* $[0, 1]$ *to any such data* $(\mathscr{X}, \mathscr{P}, \widehat{V})$ *so that the following properties hold:*

(a) if $\mathscr{X} \to \mathscr{X}'$ is an immersion, then $S_{\mathscr{X}'}(\widehat{V}) = S_{\mathscr{X}}(\widehat{V})$ (invariance under immersions)*;*

(b) for any two triples $(\mathscr{X}, \mathscr{P}, \widehat{V})$ and $(\mathscr{X}', \mathscr{P}', \widehat{V}')$ as above, if there exists an R-morphism $\varphi\colon \mathscr{X} \to \mathscr{X}'$ mapping $\mathscr{P}$ to $\mathscr{P}'$, étale along $\mathscr{P}$, and inducing an isomorphism $\widehat{V} \simeq \widehat{V}'$, then $S_{\mathscr{X}'}(\widehat{V}') = S_{\mathscr{X}}(\widehat{V})$ (invariance by étale localization)*;*

(c) if $\mathscr{X} = \mathbf{A}^N_R$ is the affine space over R and $\mathscr{P} = (0, \dots, 0)$, then $S_{\mathscr{X}}(\widehat{V})$ is the supremum in $[0,1]$ of the real numbers $r \in (0,1]$ for which there exists $f \in G_{\mathrm{an},r}$ such that $f^*\widehat{V} = \widehat{\mathbf{A}^d_0} \times \{0\}$ (normalization)*.*

As a straightforward consequence of these properties of the size, we obtain the following:

Proposition 3.2. *A smooth formal subscheme $\widehat{V}$ in $\widehat{X}_P$ is K-analytic if and only if its size $S_{\mathscr{X}}(\widehat{V})$ is a positive real number.*

Proposition 3.3. *Let $\mathscr{X}$, $\mathscr{P}$, and $\widehat{V}$ be as above and assume that there exists a smooth formal R-subscheme $\mathscr{V} \subset \widehat{\mathscr{X}_{\mathscr{P}}}$ such that $\widehat{V} = \mathscr{V}_K$. Then $S_{\mathscr{X}}(\widehat{V}) = 1$.*

The remainder of this section is devoted to further properties of the size.

Proposition 3.4. *The size is invariant under isometric extensions of valued fields (complete with respect to a discrete valuation).*

Proof. It suffices to check this assertion in the case of a smooth formal subscheme $\widehat{V}$ through the origin of the affine space $\mathbf{A}^N$. By its very definition, the size cannot decrease under extensions of the base field.

To show that it cannot increase either, let us fix an isomorphism of K-formal schemes

$$\xi = (\xi_1, \dots, \xi_N) : \widehat{\mathbf{A}}^d_0 \xrightarrow{\sim} \widehat{V} \hookrightarrow \widehat{\mathbf{A}}^N_0$$

given by N power series $\xi_i \in K[[T_1, \dots, T_d]]$ such that $\xi_1(0) = \dots = \xi_N(0) = 0$. We then observe that for any N-tuple $g = (g_1, \dots, g_N)$ of series in $K[[X_1, \dots, X_N]]$, the following two conditions are equivalent:

(i) g belongs to $G_{\mathrm{for},K}$ and $(g^{-1})^*\widehat{V} = \widehat{\mathbf{A}^d_0} \times \{0\}$;

(ii) $g_1(0) = \dots = g_N(0) = 0$, $g_{d+1}(\xi_1, \dots, \xi_N) = \dots = g_N(\xi_1, \dots, \xi_N) = 0$, and $\left(\frac{\partial g_i}{\partial X_j}(0)\right)$ belongs to $\mathrm{GL}_N(K)$.

Let K' be a valued field, satisfying the same condition as K, that contains K and whose absolute value restricts to the given one on K. Let R' be its valuation ring. Let $G'_{\mathrm{for}}, G'_{\mathrm{an}}, G'_{\mathrm{an},r}, \dots$ denote the analogues of $G_{\mathrm{for}}, G_{\mathrm{an}}, G_{\mathrm{an},r}, \dots$ defined by replacing the valued field K by K'. Recall that there exists an "orthogonal projection" from K' to K, namely a K-linear map $\lambda\colon K' \to K$ such that $|\lambda(a)| \leqslant |a|$ for any $a \in K'$ and $\lambda(a) = a$ for any $a \in K$; see for instance [28, p. 58, Corollary (2.3)].

Let $\widehat{V}' = \widehat{V}_{K'}$ be the formal subscheme of $\widehat{\mathbf{A}_{K'}^N}$ deduced from $\widehat{V}$ by the extension of scalars $K \hookrightarrow K'$, and let r be an element in $]0, S_{\mathbf{A}_{R'}^N}(\widehat{V}')[$. By the very definition of the size, there exists some $g' = (g_1', \ldots, g_N')$ in $G'_{\mathrm{an},r}$ such that $(g'^{-1})^* \widehat{V} = \widehat{\mathbf{A}_0^d} \times \{0\}$. Since the tangent space at the origin of V' is defined over K, by composing g' with a suitable element in $\mathrm{GL}_N(R')$, we may even find g' such that $Dg'(0)$ belongs to $\mathrm{GL}_N(R)$. Then the series $g_i := \lambda \circ g_i'$, deduced from the series g_i' by applying the linear map λ to their coefficients, satisfy $g_i(0) = 0$, $(\partial g_i / \partial X_j)(0) = (\partial g_i' / \partial X_j)(0)$, and $\|g_i\|_r \leqslant \|g_i'\|_r$. Therefore $g := (g_1, \ldots, g_N)$ is an element of $G_{\mathrm{an},r}$. Moreover, from the equivalence of conditions (i) and (ii) above and its analogue with K' instead of K, we derive that g satisfies $(g^{-1})^* \widehat{V} = \widehat{\mathbf{A}_0^d} \times \{0\}$. This shows that $S_{\mathbf{A}_R^N}(\widehat{V}) \geqslant r$ and establishes the required inequality $S_{\mathbf{A}_R^N}(\widehat{V}) \geqslant S_{\mathbf{A}_{R'}^N}(\widehat{V})$. □

The next proposition relates sizes, radii of convergence, and Newton polygons.

Proposition 3.5. *Let $\varphi \in K[[X]]$ be a power series such that $\varphi(0) = 0$ and $\varphi'(0) \in R$, and let $\widehat{C}$ be its graph, namely the formal subscheme of $\widehat{\mathbf{A}}_0^2$ defined by the equation $x_2 = \varphi(x_1)$.*

(1) The radius of convergence ρ of φ satisfies

$$\rho \geqslant S_{\mathbf{A}_R^2}(\widehat{C}).$$

(2) Suppose that ρ is positive and that $\varphi'(0)$ is a unit in R. Then

$$S_{\mathbf{A}_R^2}(\widehat{C}) = \min(1, \exp \lambda_1),$$

where λ_1 denotes the first slope of the Newton polygon of the power series $\varphi(x)/x$.

Recall that if $\varphi = \sum_{i \geqslant 1} c_i T^i$, under the hypothesis in (2), we have

$$\lambda_1 := \inf_{i \geqslant 1} -\frac{\log|c_{i+1}|}{i} \leqslant \lim \inf_{i \to +\infty} -\frac{\log|c_{i+1}|}{i} = \log \rho.$$

Moreover, $\exp \lambda_1$ is the supremum of the numbers $r \in]0, \rho[$ such that for any t in $\overline{K}$ satisfying $|t| < r$, we have $|\varphi(t)| = |t|$.

Proof. Let r be a positive real number such that $r < S_{\mathbf{A}_0^2}(\widehat{C})$. By assumption, there are power series f_1 and $f_2 \in K[[X_1, X_2]]$ such that $f = (f_1, f_2)$ belongs to $G_{\mathrm{an},r}$ and such that $f^* \widehat{C} = \widehat{\mathbf{A}^1} \times \{0\}$. This last condition implies (actually is equivalent to) the identity

$$f_2(T, 0) = \varphi(f_1(T, 0))$$

in $K[[T]]$. Let us write $f_1(T, 0) = \sum_{i \geqslant 1} a_i T^i$, $f_2(T, 0) = \sum_{i \geqslant 1} b_i T^i$, and $\varphi(X) = \sum_{i \geqslant 1} c_i X^i$.

One has $b_1 = c_1 a_1$, and $c_1 = \varphi'(0)$ belongs to R by hypothesis. Moreover, the first column of the matrix $Df(0)$ is $\binom{a_1}{b_1} = a_1 \binom{1}{c_1}$. Since $Df(0)$ belongs to $\mathrm{GL}_2(R)$ and c_1 to R, this implies that a_1 is a unit in R. Then, looking at the expansion of $f_1(T,0)$ (which satisfies $\|f_1(T,0)\|_r \leqslant r$), we see that $|f_1(t,0)| = |t|$ for any $t \in \overline{K}$ such that $|t| < r$. Consequently, if $g \in K[[T]]$ denotes the reciprocal power series of $f_1(T,0)$, then g converges in the open disc of radius r and satisfies $|g(t)| = |t|$ for any $t \in \overline{K}$ such that $|t| < r$.

The identity

$$\varphi(T) = \varphi(f_1(g(T),0)) = f_2(g(T),0)$$

in $K[[T]]$ then shows that the radius of convergence of φ is at least r. This establishes (1).

Let us now assume that ρ is positive and that $\varphi'(0)(= c_1)$ is a unit of R. Then $b_1 = a_1 c_1$ is also a unit, and similarly, we have $|f_2(t,0)| = |t|$ for any $t \in \overline{K}$ such that $|t| < r$. This implies that $|\varphi(t)| = |t|$ for any such t. This shows that $\exp \lambda_1 \geqslant S_{\mathbf{A}^2_R}(\widehat{C})$.

To complete the proof of (2), observe that the element f of G_{an} defined as $f(T_1,T_2) = (T_1 + T_2, \varphi(T_1))$ satisfies $f^*\widehat{C} = \widehat{\mathbf{A}^1} \times \{0\}$ and belongs to $G_{\mathrm{an},r}$ for any r in $]0, \min(1, \exp \lambda_1)[$. $\square$

Observe that for any nonzero $a \in R$, the series $\varphi(T) = T/(a-T)$ has radius of convergence $\rho = |a|$, while the size of its graph $\widehat{C}$ is 1 (observe that $f(T_1,T_2) := (aT_1 + T_2, T_1/(1-T_1))$ satisfies $f^*\widehat{C} = \widehat{\mathbf{A}^1} \times \{0\}$). Taking $|a| < 1$, this shows that the size of the graph of a power series φ can be larger than its radius of convergence when the assumption $\varphi'(0) \in R$ is omitted.

As an application of the second assertion in Proposition 3.5, we obtain that when K is a p-adic field, the size of the graph of $\log(1+x)$ is equal to $|p|^{1/(p-1)}$. Considering this graph as the graph of the exponential power series with axes exchanged, this also follows from the first assertions of Propositions 3.5 and 3.6 below.

Finally, let us indicate that by analyzing the construction à la Cauchy of local solutions of analytic ordinary differential equations, one may establish the following lower bounds on the size of a formal curve obtained by integrating an algebraic one-dimensional foliation over a p-adic field (cf. [13, Proposition 4.1]):

Proposition 3.6. *Assume that K is a field of characteristic 0, and that its residue field k has positive characteristic p. Assume also that $\mathscr{X}$ is smooth over R in a neighborhood of $\mathscr{P}$. Let $\mathscr{F} \subset T_{\mathscr{X}/R}$ be a rank 1 subbundle and let $\widehat{C}$ be the formal integral curve through P of the one-dimensional foliation $F = \mathscr{F}_K$. Then*

$$S_{\mathscr{X}}(\widehat{C}) \geqslant |p|^{1/(p-1)}.$$

If, moreover, K is absolutely unramified (that is, if the maximal ideal of R is pR) *and if the one-dimensional subbundle $\mathscr{F}_k \subset T_{\mathscr{X}_k}$ is closed under p-th powers, then*

$$S_{\mathscr{X}}(\widehat{C}) \geqslant |p|^{1/p(p-1)}.$$

3.B A-analyticity of formal curves in algebraic varieties over number fields

Let K be a number field and let R denote its ring of integers. For any maximal ideal $\mathfrak{p}$ of R, let $|\cdot|_{\mathfrak{p}}$ denote the $\mathfrak{p}$-adic absolute value, normalized by the condition $|\pi|_{\mathfrak{p}} = (\#(R/\mathfrak{p}))^{-1}$ for any uniformizing element π at $\mathfrak{p}$. Let $K_{\mathfrak{p}}$ and $R_{\mathfrak{p}}$ be the $\mathfrak{p}$-adic completions of K and R, and $\mathbf{F}_{\mathfrak{p}} := R/\mathfrak{p}$ the residue field of $\mathfrak{p}$.

In this section, we consider a quasiprojective algebraic variety X over K, a rational point P in $X(K)$, and a smooth formal curve $\widehat{C}$ in $\widehat{X}_P$.

It is straightforward that if N denotes a sufficiently divisible positive integer, there exists a model $\mathscr{X}$ of X, quasiprojective over $R[1/N]$, such that P extends to a point $\mathscr{P}$ in $\mathscr{X}(R[1/N])$. Then, for any maximal ideal $\mathfrak{p}$ not dividing N, the size $S_{\mathscr{X}_{R_{\mathfrak{p}}}}(\widehat{C}_{K_{\mathfrak{p}}})$ is a well-defined real number in $[0,1]$.

Definition 3.7. *We will say that the formal curve $\widehat{C}$ in X is A-*analytic *if the following conditions are satisfied:*

(i) for any place v of K, the formal curve $\widehat{C}_{K_v}$ is K_v-analytic;
(ii) the infinite product $\prod_{\mathfrak{p}\nmid N} S_{\mathscr{X}_{R_{\mathfrak{p}}}}(\widehat{C}_{K_{\mathfrak{p}}})$ converges to a positive real number.

Condition (ii) asserts precisely that the series with nonnegative terms

$$\sum_{\mathfrak{p}\nmid N} \log S_{\mathscr{X}_{R_{\mathfrak{p}}}}(\widehat{C}_{K_{\mathfrak{p}}})^{-1}$$

is convergent.

Observe that the above definition does not depend on the choices required to formulate it. Indeed, condition (i) does not involve any choice. Moreover, if condition (i) holds and if N' is any positive multiple of N, condition (ii) holds for $(N, \mathscr{X}, \mathscr{P})$ if and only if it holds for $\big(N', \mathscr{X}_{R[1/N']}, \mathscr{P}_{R[1/N']}\big)$. Moreover, for any two such triples $(N_1, \mathscr{X}_1, \mathscr{P}_1)$ and $(N_2, \mathscr{X}_2, \mathscr{P}_2)$, there is a positive integer M, a multiple of both N_1 and N_2, such that the models $(\mathscr{X}_1, \mathscr{P}_1)$ and $(\mathscr{X}_2, \mathscr{P}_2)$ of (X, P) become isomorphic over $R[1/M]$. This shows that when (i) is satisfied, conditions (ii) for any two triples $(N, \mathscr{X}, \mathscr{P})$ are indeed equivalent.

It follows from the properties of the size recalled in Proposition 3.1 that A-analyticity is invariant under immersions and compatible to étale localization.

As a consequence of Propositions 3.2 and 3.3, we also have the following:

Proposition 3.8. *Let $\widehat{C}$ be a smooth formal curve which is K_v-analytic for any place v of K. Assume that $\widehat{C}$ extends to a smooth formal curve $\mathscr{C} \hookrightarrow \mathscr{X}$ over $R[1/N]$, for some $N \geqslant 1$. Then $\widehat{C}$ is A-analytic.*

Indeed, these conditions imply that the size of $\widehat{C}$ at almost every finite place of K is equal to 1, while being positive at every place.

As observed in essence by Eisenstein [23], any algebraic smooth formal curve satisfies the hypothesis of Proposition 3.8. Therefore, we have the following corollary:

Corollary 3.9. *If the smooth formal curve $\widehat{C}$ is algebraic, then it is A-analytic.*

The invariance of size under extensions of valued fields established in Proposition 3.4 easily implies that *for any number field K' containing K, the smooth formal curve $\widehat{C}' := \widehat{C}_{K'}$ in $X_{K'}$ deduced from $\widehat{C}$ by the extension of scalars $K \hookrightarrow K'$ is A-analytic iff $\widehat{C}$ is A-analytic.*

Let $\varphi \in K[[X]]$ be any formal power series, and let $P := (0, \varphi(0))$. From the inequality in Proposition 3.5(1), between the convergence radius of a power series and the size of its graph, it follows that *the A-analyticity of the graph $\widehat{C}$ of φ in $\widehat{\mathbf{A}^2_P}$ implies that the convergence radii R_v of φ at the places v of K satisfy the so-called Bombieri condition*

$$\prod_v \min(1, R_v) > 0,$$

or equivalently

$$\sum_v \log^+ R_v^{-1} < +\infty.$$

However, the converse does not hold, as can be seen by considering the power series $\varphi(X) = \log(1 + X)$, which satisfies Bombieri's condition (since all the R_v equal 1) but is not A-analytic (its p-adic size is $|p|^{1/(p-1)}$ and the infinite series $\sum \frac{1}{p-1} \log p$ diverges).

Let us conclude this section with a brief discussion of the relevance of A-analyticity in the arithmetic theory of differential equations (we refer to [12, 13, 17] for more details).

Assume that X is smooth over K, that F is a sub-vector bundle of rank one in the tangent bundle T_X (defined over K), and that $\widehat{C}$ is the formal leaf at P of the one-dimensional algebraic foliation on X defined by F. By a model of (X, F) over $R[1/N]$, we mean the data of a scheme $\mathscr{X}$ quasi-projective and smooth over $\operatorname{Spec} R$, of a coherent subsheaf $\mathscr{F}$ of $T_{\mathscr{X}/R}$, and of an isomorphism $X \simeq \mathscr{X} \otimes K$ inducing an isomorphism $F \simeq \mathscr{F} \otimes K$. Such models clearly exist if N is sufficiently divisible. Let us choose one of them $(\mathscr{X}, \mathscr{F})$. We say that the foliation F satisfies the *Grothendieck–Katz condition* if for almost every maximal ideal $\mathfrak{p} \subset R$, the subsheaf $\mathscr{F}_{\mathbf{F}_\mathfrak{p}}$ of $T_{\mathscr{X}_{\mathbf{F}_\mathfrak{p}}/\mathbf{F}_\mathfrak{p}}$ is closed under p-th powers, where p denotes the characteristic of $\mathbf{F}_\mathfrak{p}$. As above, this condition does not depend on the choice of the model $(\mathscr{X}, \mathscr{F})$.

Proposition 3.10. *With the above notation, if F satisfies the Grothendieck–Katz condition, then its formal integral curve $\widehat{C}$ through any rational point P in $X(K)$ is A-analytic.*

Proof. It follows from Cauchy's theory of analytic ordinary differential equations over local fields that the formal curve $\widehat{C}$ is K_v-analytic for any place v of K.

After possibly increasing N, we may assume that P extends to a section $\mathscr{P}$ in $\mathscr{X}(R[1/N])$. For any maximal ideal $\mathfrak{p} \subset R$ that is unramified over a prime number p and such that $\mathscr{F}_{\mathbf{F}_\mathfrak{p}}$ is closed under p-th power, Proposition 3.6 shows that the $\mathfrak{p}$-adic size of $\widehat{C}$ is at least $|p|^{1/p(p-1)}$. When F satisfies the Grothendieck–Katz condition, this inequality holds for almost all maximal ideals of R. Since the series over primes $\sum_p \frac{1}{p(p-1)} \log p$ converges, this implies the convergence of the series $\sum_{\mathfrak{p} \nmid N} \log S_{\mathscr{X}_{R_\mathfrak{p}}}(\widehat{C}_{K_\mathfrak{p}})^{-1}$ and consequently the A-analyticity of $\widehat{C}$. □

4 Analytic curves in algebraic varieties over local fields and canonical seminorms

4.A Consistent sequences of norms

Let K be a local field, X a projective scheme over K, and L a line bundle over X.

We may consider the following natural constructions of sequences of norms on the spaces of sections $\Gamma(X, L^{\otimes n})$:

(1) When $K = \mathbf{C}$ and X is reduced, we may choose an arbitrary continuous norm $\|\cdot\|_L$ on the $\mathbf{C}$-analytic line bundle L_{an} defined by L on the compact and reduced complex analytic space $X(\mathbf{C})$. Then, for any integer n, the space of algebraic regular sections $\Gamma(X, L^{\otimes n})$ may be identified with a subspace of the space of continuous sections of $L_{\mathrm{an}}^{\otimes n}$ over $X(\mathbf{C})$. It may therefore be equipped with the restriction of the L^∞-norm, defined by

$$\|s\|_{\mathrm{L}^\infty,n} := \sup_{x \in X(\mathbf{C})} \|s(x)\|_{L^{\otimes n}} \quad \text{for any } s \in \Gamma(X, L^{\otimes n}), \tag{4.1}$$

where $\|\cdot\|_{L^{\otimes n}}$ denotes the continuous norm on $L_{\mathrm{an}}^{\otimes n}$ deduced from $\|\cdot\|_L$ by taking the n-th tensor power.

This construction admits a variant in which instead of the sup-norms (4.1), one considers the L^p-norms defined by using some "Lebesgue measure" (cf. [12, 4.1.3], and [46, Théorème 3.10]).

(2) When $K = \mathbf{R}$ and X is reduced, we may choose a continuous norm on $L_\mathbf{C}$ that is invariant under complex conjugation. The previous constructions define complex norms on the complex vector spaces

$$\Gamma(X, L^{\otimes n}) \otimes_\mathbf{R} \mathbf{C} \simeq \Gamma\left(X_\mathbf{C}, L_\mathbf{C}^{\otimes n}\right),$$

which are invariant under complex conjugation, and by restriction, real norms on the real vector spaces $\Gamma(X, L^{\otimes n})$.

(3) When K is a p-adic field, with ring of integers $\mathscr{O}$, we may choose a pair $(\mathscr{X}, \mathscr{L})$, where $\mathscr{X}$ is a projective flat model of X over $\mathscr{O}$, and $\mathscr{L}$ a line bundle over $\mathscr{X}$ extending L. Then, for any integer n, the $\mathscr{O}$-module $\Gamma(\mathscr{X}, \mathscr{L}^{\otimes n})$ is free of finite rank and may be identified with an $\mathscr{O}$-lattice in the K-vector space $\Gamma(X, L^{\otimes n})$, and consequently defines a norm on the latter, namely, the norm $\|\cdot\|_n$ such that a section $s \in \Gamma(X, L^{\otimes n})$ satisfies $\|s\|_n \leqslant 1$ iff s extends to a section of $\mathscr{L}^{\otimes n}$ over $\mathscr{X}$.

(4) A variant of construction (1) can be used when K is a p-adic field and X is reduced. Let $\|\cdot\|$ be a metric on L (see Appendix A for basic definitions concerning metrics in the p-adic setting). For any integer n, the space $\Gamma(X, L^{\otimes n})$ admits an L^∞-norm, defined for any $s \in \Gamma(X, L^{\otimes n})$ by $\|s\|_{\mathrm{L}^\infty, n} := \sup_{x \in X(C)} \|s(x)\|$, where C denotes the completion of an algebraic closure of K. When the metric of L is defined by a model $\mathscr{L}$ of L on a *normal* projective model $\mathscr{X}$ of X on R, then this norm coincides with that defined by construction (3) (see, e.g., [48, Proposition 1.2]).

For any given K, X, and L as above, we shall say that two sequences $(\|.\|_n)_{n\in\mathbf{N}}$ and $\left(\|.\|'_n\right)_{n\in\mathbf{N}}$ of norms on the finite-dimensional K-vector spaces $(\Gamma(X, L^{\otimes n}))_{n\in\mathbf{N}}$ are *equivalent* when for some positive constant C and any positive integer n,

$$C^{-n}\,\|\cdot\|'_n \leqslant \|.\|_n \leqslant C^n\,\|\cdot\|'_n\,.$$

One easily checks that for any given K, X, and L, the above constructions provide sequences of norms $(\|.\|_n)_{n\in\mathbf{N}}$ on the sequence of spaces $(\Gamma(X, L^{\otimes n}))_{n\in\mathbf{N}}$ that are all equivalent. In particular, their equivalence class does not depend on the auxiliary data (models, norms on L, ...) involved. (For the comparison of the L^2 and L^∞ norms in the archimedean case, see notably [46, Théorème 3.10].)

A sequence of norms on the spaces $\Gamma(X, L^{\otimes n})$ that is equivalent to one (or, equivalently, to any) of the sequences thus constructed will be called *consistent*. This notion immediately extends to sequences $(\|\cdot\|_n)_{n\geqslant n_0}$ of norms on the spaces $\Gamma(X, L^{\otimes n})$, defined only for n large enough.

When the line bundle L is ample, consistent sequences of norms are also provided by additional constructions. Indeed we have the following result.

Proposition 4.2. *Let K be a local field, X a projective scheme over K, and L an* ample *line bundle over X. Let, moreover, Y be a closed subscheme of X, and assume X and Y reduced when K is archimedean.*

For any consistent sequence of norms $(\|\cdot\|_n)_{n\in\mathbf{N}}$ on $(\Gamma(X, L^{\otimes n}))_{n\in\mathbf{N}}$, the quotient norms $(\|\cdot\|'_n)_{n\geqslant n_0}$ on the spaces $\left(\Gamma\left(Y, L^{\otimes n}_{|Y}\right)\right)_{n\geqslant n_0}$, deduced from the norms $\|\cdot\|_n$ by means of the restriction maps $\Gamma(X, L^{\otimes n}) \longrightarrow \Gamma\left(Y, L^{\otimes n}_{|Y}\right)$ — which are surjective for $n \geqslant n_0$ large enough since L is ample—*constitute a consistent sequence.*

When K is archimedean, this is proved in [13, Appendix], by introducing a positive metric on L, as a consequence of Grauert's finiteness theorem for pseudoconvex domains applied to the unit disk bundle of $L^\vee$ (see also [46]).

When K is a p-adic field with ring of integers $\mathscr{O}$, Proposition 4.2 follows from the basic properties of ample line bundles over projective $\mathscr{O}$-schemes. Indeed, let $\mathscr{X}$ be a projective flat model of X over $\mathscr{O}$, $\mathscr{L}$ an ample line bundle on $\mathscr{X}$, $\mathscr{Y}$ the closure of Y in $\mathscr{X}$, and $\mathscr{I}_{\mathscr{Y}}$ the ideal sheaf of $\mathscr{Y}$. If the positive integer n is large enough, then the cohomology group $H^1(\mathscr{Y}, \mathscr{I}_{\mathscr{Y}} \cdot \mathscr{L}^{\otimes n})$ vanishes, and the restriction morphism $\Gamma(\mathscr{X}, \mathscr{L}^{\otimes n}) \to \Gamma\left(\mathscr{Y}, \mathscr{L}^{\otimes n}_{|\mathscr{Y}}\right)$ is therefore surjective. Consequently, the norm on $\Gamma\left(Y, L^{\otimes n}_{|Y}\right)$ attached to the lattice $\Gamma\left(\mathscr{Y}, \mathscr{L}^{\otimes n}_{|\mathscr{Y}}\right)$ is the quotient of the one on $\Gamma(X, L^{\otimes n})$ attached to $\Gamma(\mathscr{X}, \mathscr{L}^{\otimes n})$.

Let E be a finite-dimensional vector space over the local field K, equipped with some norm, assumed to be euclidean or hermitian in the archimedean case. This norm induces similar norms on the tensor powers $E^{\otimes n}$, $n \in \mathbf{N}$, hence—by taking the quotient norms—on the symmetric powers $\mathrm{Sym}^n E$. If X is the projective space $\mathbf{P}(E) := \mathbf{Proj}\,\mathrm{Sym}^{\cdot}(E)$ and L the line bundle $\mathscr{O}(1)$ over $\mathbf{P}(E)$, then the canonical isomorphisms $\mathrm{Sym}^n E \simeq \Gamma(X, L^{\otimes n})$ allow one to see these norms as a sequence of norms on $(\Gamma(X, L^{\otimes n}))_{n\in\mathbf{N}}$. One easily checks that this sequence is consistent. (This is straightforward in the p-adic case. When K is archimedean, this follows, for instance, from [15, Lemma 4.3.6].)

For any closed subvariety Y in $\mathbf{P}(E)$ and any $n \in \mathbf{N}$, we may consider the following commutative diagram of K-linear maps:

$$\begin{tikzcd}
\mathrm{Sym}^n E \arrow[r, "\sim"] & \mathrm{Sym}^n \Gamma(\mathbf{P}(E), \mathscr{O}(1)) \arrow[r, "\sim"] \arrow[d] & \Gamma(\mathbf{P}(E), \mathscr{O}(n)) \arrow[d, "\alpha_n"] \\
& \mathrm{Sym}^n \Gamma(Y, \mathscr{O}(1)) \arrow[r, "\beta_n"] & \Gamma(Y, \mathscr{O}(n))
\end{tikzcd}$$

where the vertical maps are the obvious restriction morphisms. The maps α_n, and consequently β_n, are surjective if n is large enough.

Together with Proposition 4.2, these observations yield the following corollary:

Corollary 4.3. *Let K, E, and Y, a closed subscheme of $\mathbf{P}(E)$, be as above. Assume that Y is reduced if K is archimedean. Let us choose a norm on E (respectively on $\Gamma(Y, \mathscr{O}(1))$) and let us equip $\mathrm{Sym}^n E$ (respectively $\mathrm{Sym}^n \Gamma(Y, \mathscr{O}(1))$) with the induced norm, for any $n \in \mathbf{N}$.*

Then the sequence of quotient norms on $\Gamma(Y, \mathscr{O}(n))$ defined for n large enough by means of the surjective morphisms $\alpha_n \colon \mathrm{Sym}^n E \to \Gamma(Y, \mathscr{O}(n))$ (respectively by means of $\beta_n \colon \mathrm{Sym}^n \Gamma(Y, \mathscr{O}(1)) \to \Gamma(Y, \mathscr{O}(n))$) is consistent.

4.B Canonical seminorms

Let K be a local field. Let X be a *projective* variety over K, P a rational point in $X(K)$, and $\widehat{C}$ a smooth K-*analytic* formal curve in $\widehat{X}_P$. To these data, we are going to attach a canonical seminorm $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C}}$ on the tangent line $T_P\widehat{C}$ of $\widehat{C}$ at P. It will be defined by considering an analogue of the evaluation map

$$E^i_D/E^{i+1}_D \hookrightarrow \Gamma(\mathscr{P}, \mathscr{O}(D)\otimes \mathscr{N}^{\vee\otimes i}),$$

which played a prominent role in our proof of Proposition 2.1.

The construction of $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C}}$ will require auxiliary data, on which it will eventually not depend.

Let us choose a line bundle L on X and a consistent sequence of norms on the K-vector spaces $E_D = \Gamma(X, L^{\otimes D})$, for $D\in\mathbf{N}$. Let us also fix norms $\|\cdot\|_0$ on the K-lines $T_P\widehat{C}$ and $L_{|P}$.

Let us denote by C_i the ith neighborhood of P in $\widehat{C}$. Thus we have $C_{-1}=\emptyset$, $C_0=\{P\}$, and C_i is a K-scheme isomorphic to $\operatorname{Spec} K[t]/(t^{i+1})$; moreover, $\widehat{C}=\varinjlim C_i$. Let us denote by E^i_D the K-vector subspace of the $s\in E_D$ such that $s_{|C_{i-1}}=0$. The restriction map $E_D\to\Gamma(C_i, L^{\otimes D})$ induces a linear map of finite-dimensional K-vector spaces

$$\varphi^i_D\colon E^i_D \to \Gamma\left(C_i, \mathscr{I}_{C_{i-1}}\otimes L^{\otimes D}\right) \simeq \left(T^\vee_P\widehat{C}\right)^{\otimes i}\otimes L^{\otimes D}_{|P}.$$

We may consider the norm $\left\|\varphi^i_D\right\|$ of this map, computed by using the chosen norms on E_D, $T_P\widehat{C}$, and $L_{|P}$, and the ones they induce by restriction, duality, and tensor product on E^i_D and on $\left(T^\vee_P\widehat{C}\right)^{\otimes i}\otimes L^{\otimes D}_{|P}$.

Let us now define $\rho(L)$ by the following formula:

$$\rho(L) = \limsup_{i/D\to+\infty} \frac{1}{i}\log\left\|\varphi^i_D\right\|.$$

The analyticity of $\widehat{C}$ implies that $\rho(L)$ belongs to $[-\infty,+\infty[$. Indeed, when K is $\mathbf{C}$ or $\mathbf{R}$, as observed in [13, §3.1], from Cauchy's inequality we easily derive the existence of positive real numbers r and C such that

$$\left\|\varphi^i_D\right\| \leqslant C^{D+1} r^{-i}. \tag{4.4}$$

When K is ultrametric, we may actually bound $\rho(L)$ in terms of the size of $\widehat{C}$:

Lemma 4.5. *Assume that K is ultrametric and let R be its ring of integers. Let $\mathscr{X}$ be a projective flat R-model of X and let $\mathscr{P}\colon \operatorname{Spec} R\to\mathscr{X}$ be the section extending P. Assume, moreover, that the metric of L is given by a line bundle $\mathscr{L}$ on $\mathscr{X}$ extending L and the consistent sequence of norms on (E_D)*

by the construction (3) in Section 4.A, and fix the norm $\|\cdot\|_0$ *on* $T_P\widehat{C}$ *so that its unit ball is equal to* $N_{\mathscr{P}}\mathscr{X} \cap T_P\widehat{C}$*.*

Then, one has

$$\rho(L) \leqslant -\log S_{\mathscr{X},\mathscr{P}}(\widehat{C}).$$

Proof. Let r be an element of $]0, S_{\mathscr{X}}(\widehat{C})[$. We claim that with the notation above, we have

$$\left\|\varphi_i^D\right\| \leqslant r^{-i}.$$

This will establish that $\rho(L) = \limsup_{i/D\to+\infty} \frac{1}{i}\log\left\|\varphi_i^D\right\| \leqslant -\log r$, hence the required inequality by letting r go to $S_{\mathscr{X}}(\widehat{C})$.

To establish the above estimate on $\left\|\varphi_i^D\right\|$, let us choose an affine open neighborhood U of $\mathscr{P}$ in $\mathscr{X}$ such that $\mathscr{L}_U$ admits a nonvanishing section l, and a closed embedding $i\colon U \hookrightarrow \mathbf{A}_R^N$ such that $i(\mathscr{P}) = (0,\ldots,0)$. Let $\widehat{C}'$ denote the image of $\widehat{C}$ by the embedding of formal schemes $\widehat{i_{KP}} : \widehat{X}_P \hookrightarrow \widehat{\mathbf{A}_{K,0}^N}$. By the very definition of the size, we may find Φ in $G_{\mathrm{an},r}$ such that $\Phi^*\widehat{C}' = \widehat{\mathbf{A}_0^1} \times \{0\}^{N-1}$. Let s be an element of $\Gamma(\mathscr{X},\mathscr{L}^{\otimes D})$. We may write $s_{|U} = i^*Q\cdot l^{\otimes D}$ for some Q in $R[X_1,\ldots,X_N]$. Then, Φ^*Q is given by a formal series $g = \sum b_I X^I$ that satisfies $\|g\|_r \leqslant 1$, or equivalently, $|b_I| r^{|I|} \leqslant 1$ for any multi-index I. If s belongs to E_D^i, with the chosen normalizations of norms, we have $\left\|\varphi_i^D(s)\right\| = |b_{i,0,\ldots,0}| \leqslant r^{-i}$. □

The exponential of $\rho(L)$ is a well-defined element in $[0,+\infty[$, and we may introduce the following definition:

Definition 4.6. *The* canonical seminorm *on* $T_P\widehat{C}$ *attached to* $(X,\widehat{C},L)$ *is*

$$\|\cdot\|^{\mathrm{can}}_{X,\widehat{C},L} := e^{\rho(L)}\,\|\cdot\|_0\,.$$

Observe that if $\widehat{C}$ is algebraic, then there exists a real number λ such that the filtration $\left(E_D^i\right)_{i\in\mathbf{N}}$ becomes stationary—or equivalently φ_D^i vanishes—for $i/D > \lambda$ (for instance, we may take the degree of the Zariski closure of $\widehat{C}$ for λ). Consequently, in this case, $\rho(L) = -\infty$ and the canonical seminorm $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C},L}$ vanishes.

The notation $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C},L}$ for the canonical seminorm—which makes reference to X, $\widehat{C}$, and L only—is justified by the first part of the next proposition:

Proposition 4.7. *(a) The seminorm* $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C},L}$ *is independent of the choices of norms on* $T_P\widehat{C}$ *and* $L_{|P}$*, and of the consistent sequence of norms on the spaces* $E_D := \Gamma(X, L^{\otimes D})$*.*

(b) For any positive integer k*, the seminorm* $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C},L}$ *is unchanged if* L *is replaced by* $L^{\otimes k}$*.*

(c) Let L_1 and L_2 be two line bundles and assume that $L_2 \otimes L_1^{-1}$ has a regular section σ over X that does not vanish at P. Then

$$\|\cdot\|^{\mathrm{can}}_{X,\widehat{C},L_1} \leqslant \|\cdot\|^{\mathrm{can}}_{X,\widehat{C},L_2}.$$

Proof. (a) Let us denote by primes another family of norms on the spaces $T_P\widehat{C}$, $L_{|P}$, and E_D, and by $\rho'(L)$ and $\left(\|\cdot\|^{\mathrm{can}}_{X,L,\widehat{C}}\right)'$ the attached "rho-invariant" and canonical seminorm. There are positive real numbers a, b, c such that $\|t\|'_0 = a\,\|t\|_0$ for any $t \in T_P\widehat{C}$, $\|s(P)\|' = b\|s(P)\|$ for any local section s of L at P, and

$$c^{-D}\|s\| \leqslant \|s\|' \leqslant c^D\|s\|$$

for any positive integer D and any global section $s \in E_D$. Consequently, for $(i,D) \in \mathbf{N}^2$ and $s \in E^i_D$,

$$\|\varphi^i_D(s)\|' = a^{-i}b^D\|\varphi^i_D(s)\| \leqslant a^{-i}b^D\|\varphi^i_D\|\|s\| \leqslant a^{-i}b^D\|\varphi^i_D\|c^D\|s\|',$$

hence

$$\|\varphi^i_D\|' \leqslant a^{-i}c^Db^D\|\varphi^i_D\|$$

and

$$\frac{1}{i}\log\|\varphi^i_D\|' \leqslant -\log a + \frac{D}{i}\log(bc) + \frac{1}{i}\log\|\varphi^i_D\|.$$

When i/D goes to infinity, this implies

$$\rho'(L) \leqslant -\log a + \rho(L),$$

from which follows that

$$\left(\|\cdot\|^{\mathrm{can}}_{X,L,\widehat{C}}\right)' \leqslant \|\cdot\|^{\mathrm{can}}_{X,L,\widehat{C}},$$

by definition of the canonical seminorm. The opposite inequality also holds by symmetry, hence the desired equality.

(b) To define $\rho(L)$ and $\rho(L^{\otimes k})$, let us use the same norm $\|\cdot\|_0$ on $T_P\widehat{C}$, and assume that the consistent sequence of norms chosen on $(\Gamma(X, L^{\otimes D}))$ is defined by one of the constructions (1–4) in the Section 4.A above, and finally that the one on $(\Gamma(X,(L^{\otimes k})^{\otimes D})) = (\Gamma(X, L^{\otimes kD}))$ is extracted from the one on $(\Gamma(X, L^{\otimes D}))$.

Specifying the line bundle with a supplementary index, one has

$$\varphi^i_{D,L^{\otimes k}} = \varphi^i_{kD,L}.$$

The definition of an upper limit therefore implies that $\rho(L^k) \leqslant \rho(L)$.

To establish the opposite inequality, observe that for any section s in $E^i_{D,L}$ and any positive integer k, the k - tensor power $s^{\otimes k}$ belongs to $E^{ki}_{D,L^{\otimes k}}$ and

$$\varphi^{ki}_{D,L^{\otimes k}}(s^{\otimes k}) = \left(\varphi^i_{D,L}(s)\right)^{\otimes k}.$$

Let ρ be any real number such that $\rho < \rho(L)$, and choose i, D, and $s \in E^i_{D,L}$ such that $\|\varphi^i_{D,L}(s)\| \geqslant e^{\rho i}\|s\|$. Then, for any positive integer k, we have

$$\|\varphi^{ki}_{D,L^{\otimes k}}(s^{\otimes k})\| = \|\varphi^i_{D,L}(s)\|^k \geqslant e^{\rho k i}\|s\|^k = e^{\rho k i}\|s^{\otimes k}\|,$$

so that $\|\varphi^{ki}_{D,L^{\otimes k}}\|^{1/ki} \geqslant e^\rho$. Consequently, $\rho(L^k) \geqslant \rho$.

(c) Here again, we may use the same norm $\|\cdot\|_0$ on $T_P\widehat{C}$ to define $\rho(L_1)$ and $\rho(L_2)$, and assume that the consistent sequences of norms chosen on $\left(\Gamma\left(X, L_1^{\otimes D}\right)\right)$ and $\left(\Gamma\left(X, L_2^{\otimes D}\right)\right)$ are defined by one of the constructions (1–4) above.

If s is a global section of $L_1^{\otimes D}$, then $s \otimes \sigma^{\otimes D}$ is a global section of $L_2^{\otimes D}$; if s vanishes to order i along $\widehat{C}$, so does $s \otimes \sigma^{\otimes D}$, and

$$\varphi^i_{D,L_2}(s \otimes \sigma^{\otimes D}) = \varphi^i_{D,L_1}(s) \otimes \sigma(P)^{\otimes D}.$$

Consequently,

$$\|\varphi^i_{D,L_1}(s)\| \leqslant \|\varphi^i_{D,L_2}\|.\|s \otimes \sigma^{\otimes D}\|.\|\sigma(P)\|^{-D} \leqslant (\|\sigma(P)\|^{-1}\|\sigma\|)^D.\|\varphi^i_{D,L_2}\|.\|s\|,$$

and $\rho(L_1) \leqslant \rho(L_2)$, as was to be shown. □

Corollary 4.8. *The set of seminorms on $T_P\widehat{C}$ described by $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C},L}$ when L varies in the class of line bundles on X possesses a maximal element, namely the canonical seminorm $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C},L}$ attached to any ample line bundle L on X.*

We shall denote by $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C}}$ this maximal element. The formation of $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C}}$ satisfies the following compatibility properties with respect to rational morphisms.

Proposition 4.9. *Let X' be another projective algebraic variety over K, and let $f\colon X \dashrightarrow X'$ be a rational map that is defined near P. Let $P' := f(P)$, and assume that f defines an (analytic, or equivalently, formal) isomorphism from $\widehat{C}$ onto a smooth K-analytic formal curve $\widehat{C'}$ in $\widehat{X'}_{P'}$.*

Then for any $v \in T_P\widehat{C}$,

$$\|Df(P)v\|_{X',f(\widehat{C})} \leqslant \|v\|_{X,\widehat{C}}.$$

If, moreover, f is an immersion in a neighborhood of P, then the equality holds.

When K is archimedean, this summarizes the results established in [13, Sections 3.2 and 3.3]. The arguments in that work may be immediately transposed to the ultrametric case using consistent norms as defined above instead of L^∞ norms on the spaces of sections E_D. We leave the details to the reader.

Observe finally that this proposition allows us to define the canonical seminorm $\|\cdot\|^{\mathrm{can}}_{X,\widehat{C}}$ when the algebraic variety X over K is assumed to be only quasiprojective. Indeed, if $\overline{X}$ denotes some projective variety containing X as an open subvariety, the seminorm $\|\cdot\|^{\mathrm{can}}_{\overline{X},\widehat{C}}$ is independent of the choice of $\overline{X}$, and we let

$$\|\cdot\|^{\mathrm{can}}_{X,\widehat{C}} := \|\cdot\|^{\mathrm{can}}_{\overline{X},\widehat{C}}.$$

5 Capacitary metrics on p-adic curves

5.A Review of the complex case

Let M be a compact Riemann surface and let Ω be an open subset of M. We assume that the compact subset complementary to Ω in any connected component of M is not polar. Let D be an effective divisor on M whose support is contained in Ω. Potential theory on Riemann surfaces (see [11, 3.1.3-4]) shows the existence of a unique subharmonic function $g_{D,\Omega}$ on M satisfying the following assumptions:

(1) $g_{D,\Omega}$ is harmonic on $\Omega \setminus |D|$;
(2) the set of points $z \in M \setminus \Omega$ such that $g_{D,\Omega}(z) \neq 0$ is a polar subset of $\partial\Omega$;
(3) for any open subset V of Ω and any holomorphic function f on V such that $\operatorname{div}(f) = D$, the function $g_{D,\Omega} - \log|f|^{-2}$ on $V \setminus |D|$ is the restriction of a harmonic function on V.

Moreover, $g_{D,\Omega}$ takes nonnegative values, is locally integrable on M, and defines an L^2_1-Green current for D in the sense of [11]. It is the so-called *equilibrium potential* attached to the divisor D in Ω.

If E is another effective divisor on M supported in Ω, one has $g_{D+E,\Omega} = g_{D,\Omega} + g_{E,\Omega}$. We can therefore extend by linearity the definition of the equilibrium potential $g_{D,\Omega}$ to arbitrary divisors D on M that are supported on Ω. Recall also that if Ω_0 denotes the union of the connected components of Ω that meet $|D|$, then $g_{D,\Omega_0} = g_{D,\Omega}$ [11, p. 258].

The function $g_{D,\Omega}$ allows one to define a generalized metric on the line bundle $\mathscr{O}_M(D)$ by the formula

$$\|\mathbf{1}_D\|^2(z) = \exp(-g_{D,\Omega}(z)),$$

where $\mathbf{1}_D$ denotes the canonical global section of $\mathscr{O}_M(D)$. We will call this metric the *capacitary metric*[4] on $\mathscr{O}_M(D)$ attached to Ω and denote by $\|f\|^{\mathrm{cap}}_{\Omega}$ the norm of a local section f of $\mathscr{O}_M(D)$.

[4]Our terminology differs slightly from that in [11]. In the present article, the term *capacitary metric* will be used for two distinct notions: for the metrics on line bundles defined using equilibrium potentials just defined, and for some metrics on the tangent line to M at a point; see Section 5.C. In [11], it was used for the latter notion only.

5.B Equilibrium potential and capacity on *p*-adic curves

Let R be a complete discrete valuation ring, and let K be its field of fractions and k its residue field. Let X be a smooth projective curve over K and let U be an affinoid subspace of the associated rigid K-analytic curve X^{an}. We shall always require that *U meets every connected component of X^{an}*; this hypothesis is analogous to the nonpolarity assumption in the complex case. We also let $\Omega = X^{\mathrm{an}} \setminus U$, which we view as a (non-quasicompact) rigid K-analytic curve; its affinoid subspaces are just affinoid subspaces of X^{an} disjoint from U. See Appendix B for a detailed proof that this endowes Ω with the structure of a rigid K-analytic space in the sense of Tate.

The aim of this subsection is to endow the line bundle $\mathscr{O}(D)$, where D is a divisor that does not meet U, with a metric (in the sense of Appendix A) canonically attached to Ω, in a way that parallels the construction over Riemann surfaces recalled in the previous subsection.

Related constructions of equilibrium potentials over p-adic curves have been developed by various authors, notably Rumely [49] and Thuillier [51] (see also [37]). Our approach will be self-contained, and formulated in the framework of classical rigid analytic geometry. Our main tool will be intersection theory on a model $\mathscr{X}$ of X over R. This point of view will allow us to combine potential theory on p-adic curves and Arakelov intersection theory on arithmetic surfaces in a straightforward way.

We want to indicate that by using an adequate potential theory on analytic curves in the sense of Berkovich [4] such as the one developed by Thuillier [51], one could give a treatment of equilibrium potential on p-adic curves and their relations to canonical seminorms that would more closely parallel the one in the complex case. For instance, in the Berkovich setting, the affinoid subspace U is a compact subset of the analytic curve attached to X, and Ω is an open subset. We leave the transposition and the extension of our results in the framework of Berkovich and Thuillier to the interested reader.

By Raynaud's general results on formal/rigid geometry, see for instance [8, 9], there exists a normal projective flat model $\mathscr{X}$ of X over R such that U is the set of rigid points of X^{an} reducing to some open subset U of the special fiber X. We shall write $U =]\mathsf{U}[_{\mathscr{X}}$ and say that U is the tube of U in $\mathscr{X}$; similarly, we write $\Omega =]\mathscr{X} \setminus \mathsf{U}[_{\mathscr{X}}$. (We remove the index $\mathscr{X}$ from the notation when it is clear from the context.) The reduction map identifies the connected components of U with those of U, and the connected components of Ω with those of $\mathsf{X} \setminus \mathsf{U}$. Since we assumed that U meets every connected component of X, this shows that U meets every connected component of X.

Recall that to any two Weil divisors Z_1 and Z_2 on $\mathscr{X}$ such that $Z_{1,K}$ and $Z_{2,K}$ have disjoint supports is attached their the *intersection number* (Z_1, Z_2). It is a rational number, which depends linearly on Z_1 and Z_2. It may be defined à la Mumford (see [42, II.(b)]), and it coincides with the degree over the residue field k of the intersection class $Z_1.Z_2$ in $\mathrm{CH}_0(\mathsf{X})$ when Z_1 or Z_2 is Cartier. Actually, when the residue field k is an algebraic extension of a finite field—for instance when K is a p-adic field, the case in which we are interested in the sequel—any Weil divisor on $\mathscr{X}$ has a multiple that is Cartier (see [40, Théorème 2.8]), and this last property, together with their bilinearity, completely determines the intersection numbers.

The definition of intersection numbers immediately extends by bilinearity to pairs of Weil divisors with coefficients in $\mathbf{Q}$ ($\mathbf{Q}$-*divisors*, for short) in $\mathscr{X}$ whose supports do not meet in X.

Proposition 5.1. *For any divisor D on X, there is a unique $\mathbf{Q}$-divisor $\mathscr{D}$ on $\mathscr{X}$ extending D and satisfying the following two conditions:*

(1) *For any irreducible component v of codimension 1 of $\mathsf{X} \setminus \mathsf{U}$, $\mathscr{D} \cdot v = 0$.*
(2) *The vertical components of $\mathscr{D}$ do not meet U.*

Moreover, the map $D \mapsto \mathscr{D}$ so defined is linear and sends effective divisors to effective divisors.

Proof. Let S denote the set of irreducible components of X and let $T \subset S$ be the subset consisting of components that do not meet U. Let $\mathscr{D}_0$ be the schematic closure of D in $\mathscr{X}$. Since U meets every connected component of X, T does not contain all of the irreducible components of some connected component of X, so that the restriction of the intersection pairing of $\mathrm{Div}_{\mathbf{Q}}(\mathscr{X})$ to the subspace generated by the components of X that belong to T is negative definite (see, for instance, [21, Corollaire 1.8], when $\mathscr{X}$ is regular; one reduces to this case by considering a resolution of $\mathscr{X}$, as in [42, II.(b)]). Therefore, there is a unique vertical divisor V, a linear combination of components in T, such that $(\mathscr{D}_0 + V, s) = 0$ for any $s \in T$. (In the analogy with the theory of electric networks, the linear system one has to solve corresponds to that of a Dirichlet problem on a graph, with at least one electric source per connected component.) Set $\mathscr{D} = \mathscr{D}_0 + V$; it satisfies assumptions (1) and (2). The linearity of the map $D \mapsto \mathscr{D}$ follows immediately from the uniqueness of V.

Let us assume that D is effective and show that so is V. (In graph-theoretic language, this is a consequence of the maximum principle for the discrete Laplacian.) Denote by m_s the multiplicity of the component s in the special fiber of $\mathscr{X}$, so that $\sum_{s \in S} m_s s$ belongs to the kernel of the intersection pairing. Write $V = \sum_{s \in S} c_s s$, where $c_s = 0$ if $s \notin T$.

Let S' be the set of elements $s \in S$ where c_s/m_s achieves its minimal value. Then, for any element τ of $S' \cap T$,

$$\begin{aligned} 0 = (c_\tau/m_\tau)\Big(\sum_{s \in S} m_s s, \tau\Big) &= c_\tau(\tau,\tau) + \sum_{s \neq \tau} (c_\tau/m_\tau) m_s (s,\tau) \\ &\leqslant c_\tau(\tau,\tau) + \sum_{s \neq \tau} c_s(s,\tau) = \sum_{s \in S} (c_s s, \tau) \\ &\leqslant (\mathscr{D},\tau) - (\mathscr{D}_0,\tau) = -(\mathscr{D}_0,\tau). \end{aligned}$$

Since $\mathscr{D}_0$ is effective and horizontal, $(\mathscr{D}_0,\tau) \geqslant 0$; hence all previous inequalities are in fact equalities. In particular, $(\mathscr{D}_0,\tau) = 0$ and $c_s/m_s = c_\tau/m_\tau$ for any $s \in S$ such that $(s,\tau) \neq 0$.

Assume by contradiction that V is not effective, i.e., that there is some s with c_s negative. Then S' is contained in T (for $c_s = 0$ we have $s \notin T$) and the preceding argument implies that S' is a union of connected components of X. (In the graph-theoretic analogue, all neighbours of a vertex in S' belong to S'.) This contradicts the assumption that U meets every connected component of X^{an} and concludes the proof that V is effective. $\square$

In order to describe the functoriality properties of the assignment $D \mapsto \mathscr{D}$ constructed in Proposition 5.1, we consider two smooth projective curves X and X' over K, some normal projective flat models $\mathscr{X}$ and $\mathscr{X}'$ over R of these curves, and $\pi\colon \mathscr{X}' \to \mathscr{X}$ an R-morphism such that the K-morphism $\pi_K\colon X' \to X$ is finite.

Recall that the direct image of 1-dimensional cycles defines a $\mathbf{Q}$-linear map between spaces of $\mathbf{Q}$-divisors:

$$\pi_*\colon \operatorname{Div}_{\mathbf{Q}}(\mathscr{X}') \longrightarrow \operatorname{Div}_{\mathbf{Q}}(\mathscr{X}),$$

and that the inverse image of Cartier divisors defines a $\mathbf{Q}$-linear map between spaces of $\mathbf{Q}$-Cartier divisors,

$$\pi^*\colon \operatorname{Div}_{\mathbf{Q}}^{\mathrm{Cartier}}(\mathscr{X}) \longrightarrow \operatorname{Div}_{\mathbf{Q}}^{\mathrm{Cartier}}(\mathscr{X}').$$

These two maps satisfy the following adjunction formula, valid for any Z in $\operatorname{Div}_{\mathbf{Q}}^{\mathrm{Cartier}}(\mathscr{X})$ and any Z' in $\operatorname{Div}_{\mathbf{Q}}(\mathscr{X}')$:

$$(\pi^* Z, Z') = (Z, \pi_* Z')\,. \tag{5.2}$$

When k is an algebraic extension of a finite field, as recalled above, $\mathbf{Q}$-divisors and $\mathbf{Q}$-Cartier divisors on $\mathscr{X}$ or $\mathscr{X}'$ coincide, and π^* may be seen as a linear map from $\operatorname{Div}_{\mathbf{Q}}(\mathscr{X})$ to $\operatorname{Div}_{\mathbf{Q}}(\mathscr{X}')$ adjoint to π_*.

In general, the map π^* above admits a unique extension to a $\mathbf{Q}$-linear map

$$\pi^*\colon \operatorname{Div}_{\mathbf{Q}}(\mathscr{X}) \longrightarrow \operatorname{Div}_{\mathbf{Q}}(\mathscr{X}'),$$

compatible with the pullback of divisors on the generic fiber

$$\pi_K^*\colon \operatorname{Div}_{\mathbf{Q}}(X) \longrightarrow \operatorname{Div}_{\mathbf{Q}}(X'),$$

such that the adjunction formula (2.3) holds for any (Z, Z') in $\operatorname{Div}_{\mathbf{Q}}(\mathscr{X}) \times \operatorname{Div}_{\mathbf{Q}}(\mathscr{X}')$. The uniqueness of such a map map follows from the nondegeneracy properties of the intersection pairing, which show that if a divisor Z_1' supported by the closed fiber X of $\mathscr{X}$ satisfies $Z_1' \cdot Z_2' = 0$ for every Z_2' in $\operatorname{Div}_{\mathbf{Q}}(\mathscr{X}')$, then $Z_1' = 0$. The existence of π^* is known when $\mathscr{X}'$ is regular (then $\operatorname{Div}_{\mathbf{Q}}(\mathscr{X})$ and $\operatorname{Div}_{\mathbf{Q}}^{\mathrm{Cartier}}(\mathscr{X})$ coincide), and when π is birational—i.e., when π_K is an isomorphism—and $\mathscr{X}$ is regular, according to Mumford's construction in [42, II.(b)]. To deal with the general case, observe that there exist two projective flat regular curves $\tilde{\mathscr{X}}$ and $\tilde{\mathscr{X}}'$ equipped with birational R-morphisms $\nu\colon \tilde{\mathscr{X}} \to \mathscr{X}$ and $\nu'\colon \tilde{\mathscr{X}}' \to \mathscr{X}'$, and an R-morphism $\tilde{\pi}\colon \tilde{\mathscr{X}}' \to \tilde{\mathscr{X}}$ such that $\pi \circ \tilde{\nu} = \tilde{\pi} \circ \nu$. Then it is straightforward that $\pi^* := \tilde{\nu}_* \tilde{\pi}^* \nu^*$ satisfies the required properties.

Observe also that the assignment $\pi \mapsto \pi^*$ so defined is functorial, as follows easily from its definition.

Proposition 5.3. *Let* U *be a Zariski open subset of the special fiber* X *and let* $\mathsf{U}' = \pi^{-1}(\mathsf{U})$*. Assume that* $]\mathsf{U}[_{\mathscr{X}}$ *meets every connected component of* X^{an}*; then* $]\mathsf{U}'[_{\mathscr{X}'}$ *meets every connected component of* $(X')^{\mathrm{an}}$*.*

Let D *and* D' *be divisors on* X *and* X' *respectively, and let* $\mathscr{D}$ *and* $\mathscr{D}'$ *be the extensions to* $\mathscr{X}$ *and* $\mathscr{X}'$*, relative to the open subsets* U *and* U' *respectively, given by Proposition 5.1.*

(a) Assume that $D' = \pi^* D$*. If* $|D|$ *does not meet* $]\mathsf{U}[$*, then* $|D'|$ *is disjoint from* $]\mathsf{U}'[$ *and* $\mathscr{D}' = \pi^*\mathscr{D}$*.*
(b) Assume that $D = \pi_* D'$*. If* $|D'|$ *does not meet* $]\mathsf{U}'[$*, then* $|D| \cap]\mathsf{U}[= \emptyset$ *and* $\mathscr{D} = \pi_*\mathscr{D}'$*.*

Proof. Let us denote by S the set of irreducible components of the closed fiber X of $\mathscr{X}$, and by T its subset of the components that do not meet U. Define similarly S' and T' to be the set of irreducible components of X' and its subset corresponding to the components that do not meet U'. Also let N denote the set of all irreducible components of X' that are contracted to a point by π.

By construction of π^*, the divisor $\pi^*(\mathscr{D})$ satisfies $(\pi^*(\mathscr{D}), n) = 0$ for any $n \in N$ and has no multiplicity along the components of N that are not contained in $\pi^{-1}(|\mathscr{D}|)$.

Since $\mathsf{U}' = \pi^{-1}(\mathsf{U})$, T' is the union of all components of X' that are mapped by π, either to a point outside U or to a component in T.

(a) Let $t' \in T'$. One has $(\pi^*\mathscr{D}, t') = (\mathscr{D}, \pi_* t') = 0$, since t' maps to a component in T, or to a point. Moreover, by the construction of π^*, the vertical components of $\pi^*\mathscr{D}$ are elements $s' \in S'$ such $\pi(s')$ meets the support of $\mathscr{D}$. By assumption, the Zariski closure of D in X is disjoint from U; in other words, the vertical components of $\pi^*\mathscr{D}$ all belong to T'. This shows that the divisor $\pi^*\mathscr{D}$ on $\mathscr{X}'$ satisfies the conditions of Proposition 5.1; since it extends $D' = \pi^* D$, one has $\pi^*\mathscr{D} = \mathscr{D}'$.

(b) Let s be a vertical component appearing in $\pi_*(\mathscr{D}')$; necessarily, there is a vertical component s' of $\mathscr{D}'$ such that $s = \pi(s')$. This implies that $s' \in T'$, hence $s \in T$. For any $t \in T$, $\pi^*(t)$ is a linear combination of vertical components of X' contained in $\pi^{-1}(t)$. Consequently, they all belong to T' and one has $(\pi_*(\mathscr{D}'), t) = (\mathscr{D}', \pi^*(t)) = 0$. By uniqueness, $\pi_*(\mathscr{D}) = \mathscr{E}$. □

Corollary 5.4. *Let* X *be a projective smooth algebraic curve over* K*, let* U *be an affinoid subspace of* X^{an} *that meets any connected component of* X^{an}*. Let* D *be a divisor on* X *whose support is disjoint from* U*.*

Then the metrics on the line bundle $\mathscr{O}_X(D)$ *induced by the line bundle* $\mathscr{O}_{\mathscr{X}}(\mathscr{D})$ *defined by Proposition 5.1 does not depend on the choice of the projective flat model* $\mathscr{X}$ *of* X *such that* U *is the tube of a Zariski open subset of the special fiber of* $\mathscr{X}$*.*

Proof. For $i = 1, 2$, let $(\mathscr{X}_i, \mathsf{U}_i)$ be a pair as above, consisting of a normal flat, projective model $\mathscr{X}_i$ of X over R, and an open subset U_i of its special fiber X_i such that $]\mathsf{U}_i[_{\mathscr{X}_i} = U$. Let $\mathscr{D}_i$ denote the extension of D on $\mathscr{X}_i$ relative to U_i.

There exists a third model $(\mathscr{X}', \mathsf{U}')$ that admits maps $\pi_i \colon \mathscr{X}' \to \mathscr{X}_i$, for $i = 1, 2$, extending the identity on the generic fiber. Let $\mathscr{D}'$ denote the extension of D on $\mathscr{X}'$. For $i = 1, 2$, one has $\pi_i^{-1}(\mathsf{U}_i) = \mathsf{U}'$. By Proposition 5.3, one thus has the equalities $\pi^* \mathscr{D}_i = \mathscr{D}'$; hence the line bundles $\mathscr{O}_{\mathscr{X}'}(\mathscr{D}')$ on $\mathscr{X}'$ and $\mathscr{O}_{\mathscr{X}_i}(\mathscr{D}_i)$ on $\mathscr{X}$ induce the same metric on $\mathscr{O}_X(D)$. □

We shall call this metric the *capacitary metric* and denote by $\|f\|_{\Omega}^{\mathrm{cap}}$ the norm of a local section f of $\mathscr{O}_X(D)$ for this metric.

Proposition 5.5. *Let X be a projective smooth algebraic curve over K, and let U be an affinoid subspace of X^{an} that meets any connected component of X^{an}. Let D be a divisor on X whose support is disjoint from U and let $\Omega = X^{\mathrm{an}} \setminus U$.*

If Ω' denotes the union of the connected components of Ω that meet $|D|$, then the capacitary metrics of $\mathscr{O}(D)$ relative to Ω and to Ω' coincide.

Proof. Let us fix a normal projective flat model $\mathscr{X}$ of X over R and a Zariski open subset U of its special fiber X such that $U =]\mathsf{U}[_{\mathscr{X}}$. Let $\mathsf{Z} = \mathsf{X} \setminus \mathsf{U}$ and let Z' denote the union of those connected components of Z that meet the specialization of $|D|$. Then $\Omega' =]\mathsf{Z}'[$ is the complementary subset to the affinoid $]\mathsf{U}'[$, where $\mathsf{U}' = \mathsf{X} \setminus \mathsf{Z}'$; in particular, $]\mathsf{U}'[$ meets every connected component of X^{an}.

Let $\mathscr{D}_0$ denote the horizontal divisor on $\mathscr{X}$ that extends D. The divisor $\mathscr{D}' := \mathscr{D}_{\Omega'}$ is the unique $\mathbf{Q}$-divisor of the form $\mathscr{D}_0 + V$ on $\mathscr{X}$ where V is a vertical divisor supported by Z' such that $(\mathscr{D}', t) = 0$ for any irreducible component of Z'. By the definition of Z', an irreducible component of Z that is not contained in Z' meets neither Z' nor $\mathscr{D}_0$. It follows that for any such component t, $(\mathscr{D}', t) = (\mathscr{D}_0, t) + (V, t) = 0$. By uniqueness, $\mathscr{D}'$ is the extension of D on $\mathscr{X}$ relative to U, so that $\mathscr{D}_{\Omega'} = \mathscr{D}_{\Omega}$. This implies the proposition. □

As an application of the capacitary metric, in the next proposition we establish a variant of a classical theorem by Fresnel and Matignon [25, Théorème 1] asserting that affinoids of a curve can be defined by one equation. (While these authors make no hypothesis on the residue field of k, or on the complementary subset of the affinoid U, we are able to impose the polar divisor of f.) Using the terminology of Rumely [49, §4.2, p. 220], this proposition means that affinoid subsets of a curve are RL-domains ("rational lemniscates"), and that RL-domains with connected complement are PL-domains ("polynomial lemniscates"). It is thus essentially equivalent to Rumely's theorem [49, Theorem 4.2.12, p. 244] asserting that *island domains* coincide with PL-domains. Rumely's proof relies on his non-archimedean potential theory, which we replace here by Proposition 5.1.

This proposition will also be used to derive further properties of the capacitary metric.

Proposition 5.6. *Assume that the residue field k of K is algebraic over a finite field. Let (X, U, Ω) be as above, and let D be an effective divisor that does not meet U but meets every connected component of Ω. There is a rational function $f \in K(X)$ with polar divisor a multiple of D such that $U = \{x \in X\,;\, |f(x)| \leqslant 1\}$.*

Proof. Keep notation as in the proof of Proposition 5.1; in particular, S denotes the set of irreducible components of X. The closed subset $\mathsf{X} \setminus \mathsf{U}$ has only finitely many connected components, say $\mathsf{V}_1, \dots, \mathsf{V}_r$. Moreover, we may assume that for each i, V_i is the union of a family $T_i \subset S$ of components of X. For any i, the tube $]\mathsf{V}_i[$ in X^{an} consisting of the rigid points of X that reduce to points of V_i is a connected analytic subset of X^{an}, albeit not quasicompact, and X^{an} is the disjoint union of $U =]\mathsf{U}[$ and of the $]\mathsf{V}_i[$. (See [47] for more details.) We let m_s denote the multiplicity of the component s in the special fiber, and $F = \sum_{s\in S} m_s s$.

Let $\mathscr{D} = \mathscr{D}_0 + V$ be the extension of D to a $\mathbf{Q}$-divisor of $\mathscr{X}$ given by Proposition 5.1, where $\mathscr{D}_0$ is horizontal and $V = \sum_{s\in S} c_s s$ is a vertical divisor supported by the special fiber X. One has $c_s = 0$ for $s \notin T$ and $c_s \geqslant 0$ if $s \in T$. For any $s \notin T$, we define $a_s = (V, s)$. This is a nonnegative rational number and we have

$$\sum_{s\in S\setminus T} a_s m_s = (V, F) - \sum_{t\in T} m_t (V, t) = \sum_{t\in T} m_t(\mathscr{D}_0, t) = \sum_{s\in S} m_s(\mathscr{D}_0, s),$$

since D does not meet U; hence

$$\sum_{s\in S\setminus T} a_s m_s = (\mathscr{D}_0, F) = \deg(D). \tag{5.7}$$

For any $s \in S \setminus T$, let us fix a point x_s of X that is contained on the component s as well as on the smooth locus of $\mathscr{X}$. Using either a theorem of Rumely [49, Th. 1.3.1, p. 48], or van der Put's description of the Picard group of any one-dimensional K-affinoid, cf. [44, Prop. 3.1],[5] there is a rational function $f_s \in K(X)$ with polar divisor a multiple of D and of which all zeros specialize to x_s. We may write its divisor as a sum

$$\operatorname{div}(f_s) = -n_s\mathscr{D} + E_s + W_s,$$

where n_s is a positive integer, E_s is a horizontal effective divisor having no common component with $\mathscr{D}$ and W_s is a vertical divisor. Since E_s is the closure of the divisor of zeroes of f_s, it meets only the component labeled s. One thus has $(E_s, s') = 0$ for $s' \in S \setminus \{s\}$, while

$$(E_s, s) = \frac{1}{m_s}(E_s, m_s s) = \frac{1}{m_s}(E_s, F) = \frac{n_s}{m_s}\deg D.$$

[5]The proofs in both references are similar and rely on the Abel–Jacobi map, together with the fact that K is the union of its locally compact subfields.

Let $t \in T$. One has $(\operatorname{div}(f_s), t) = 0$, hence

$$(W_s, t) = n_s(\mathscr{D}, t) - (E_s, t) = 0.$$

Similarly, if $s' \in S \setminus T$, then $(\operatorname{div}(f_s), s') = 0$ and

$$\begin{aligned}(W_s, s') &= n_s(\mathscr{D}, s') - (E_s, s') \\ &= n_s\left(\mathscr{D}_0, s'\right) + n_s(V, s') - (E_s, s') \\ &= 0 + n_s a_{s'} - (E_s, s').\end{aligned}$$

If $s' \neq s$, it follows that

$$(W_s, s') = n_s a_{s'},$$

while

$$(W_s, s) = n_s a_s - \frac{n_s}{m_s} \deg(D).$$

We now define a vertical divisor

$$W = \sum_{s \notin T} \frac{a_s m_s}{n_s} W_s.$$

For any $t \in T$, $(W, t) = 0$. Moreover, for any $s' \in S \setminus T$,

$$\begin{aligned}(W, s') &= \sum_{s \notin T} \frac{a_s m_s}{n_s} (W_s, s') \\ &= \sum_{s \notin T} a_s m_s a_{s'} - \frac{a_{s'}}{n_{s'}} n_{s'} \deg(D) \\ &= a_{s'} \left(\sum_{s \notin T} a_s m_s \right) - a_{s'} \deg(D),\end{aligned}$$

hence $(W, s') = 0$ by (5.7). Therefore, the vertical $\mathbf{Q}$-divisor W is a multiple of the special fiber and there is $\lambda \in \mathbf{Q}$ such that $W = \lambda F$. Finally,

$$\sum_{s \notin T} \frac{a_s m_s}{n_s} \operatorname{div}(f_s) - \lambda F = -\deg(D)\mathscr{D} + \sum_{s \notin T} \frac{a_s m_s}{n_s} E_s$$

is a principal $\mathbf{Q}$-divisor. It follows that there are positive integers μ and λ_s, for $s \notin T$, such that

$$\mathscr{P} = \sum_{s \notin T} \lambda_s E_s - \mu \mathscr{D}$$

is the divisor of a rational function $f \in K(X)$.

By construction, the polar divisor of f on X is a multiple of D. Moreover, the reduction of any $x \notin U$ belongs to a component labeled by T at which the multiplicity of $\mathscr{P}$ is positive. Consequently, $|f(x)| > 1$. In contrast, if $x \in U$, it reduces to a component outside T, and $|f(x)| \leqslant 1$. More precisely, $|f(x)| < 1$ if and only if x reduces to one of the points x_s, $s \notin T$. □

The definition of an algebraic metric now implies the following explicit description of the capacitary metric.

Corollary 5.8. *Let (X, U, Ω) be as above, let D be any divisor that does not meet U, and let f be a rational function defining U, as in the preceding proposition, and whose polar divisor is equal to mD for some positive integer m. Then the capacitary metric on $\mathscr{O}_X(D)$ can be computed as*

$$-\log\|1_{\mathsf{D}}\|_{\Omega}^{\mathrm{cap}}(x) = \frac{1}{m}\log^{+}|f(x)| = \max(0, \log|f(x)|^{1/m}).$$

Proposition 5.9. *Let (X, U, Ω) and (X', U', Ω') be as above and let $\varphi\colon \Omega' \to \Omega$ be any rigid analytic isomorphism. Let D' be any divisor in X' whose support does not meet U' and let $D = \varphi(D')$.*

Then for any $x \in \Omega'$,

$$\|1_{D'}\|_{\Omega'}^{\mathrm{cap}}(x) = \|1_D\|_{\Omega}^{\mathrm{cap}}(\varphi(x)).$$

Proof. By linearity, we may assume that D is effective. Let $f \in K(X)$ and $f' \in K(X')$ be rational functions as in Proposition 5.6. Let m and m' be positive integers such that the polar divisors of f and f' are mD and $m'D'$ respectively. The function $f\circ\varphi$ is a meromorphic function on Ω' whose divisor is mD'. Consequently, the meromorphic function

$$g = (f \circ \varphi)^{m'}/(f')^{m}$$

on Ω' is in fact invertible. We have to prove that $|g|(x) = 1$ for any $x \in \Omega'$.

Let (ε_n) be any decreasing sequence of elements of $\sqrt{|K^*|}$ converging to 1. The sets $V'_n = \{x \in X' \,;\, |f'(x)| \geqslant \varepsilon_n\}$ are affinoid subspaces of Ω' and exhaust it. By the maximum principle (see Proposition B.1 below), one has

$$\sup_{x\in V'_n} |g(x)| = \sup_{|f'(x)|=\varepsilon_n} |g(x)| \leqslant 1/(\varepsilon_n)^m \leqslant 1.$$

Consequently, $\sup_{x\in\Omega'}|g(x)| \leqslant 1$. The opposite inequality is shown similarly by considering the isomorphism $\varphi^{-1}\colon \Omega \to \Omega'$. This proves the proposition. □

5.C Capacitary norms on tangent spaces

Definition 5.10. *Let (X, U, Ω) be as above and let $P \in X(K)$ be a rational point lying in Ω. Let us endow the line bundle $\mathscr{O}_X(P)$ with its capacitary metric relative to Ω. The capacitary norm $\|\cdot\|_{P,\Omega}^{\mathrm{cap}}$ on the K-line T_PX is then defined as the restriction of $(\mathscr{O}_X(P), \|\cdot\|_{\Omega}^{\mathrm{cap}})$ to the point P, composed with the adjunction isomorphism $\mathscr{O}_X(P)|_P \simeq T_PX$.*

Example 5.11. Let us fix a normal projective flat model $\mathscr{X}$, let $\mathscr{P}$ be the divisor extending P, meeting the special fiber X in a smooth point P. Let $\mathsf{U} = \mathsf{X} \setminus \{\mathsf{P}\}$ and define $U =]\mathsf{U}[$, $\Omega =]\mathsf{P}[$. In other words, Ω is the set of rig-points of X^{an} that have the same reduction P as P. Then $\Omega =]\mathsf{P}[$ is isomorphic to an open unit ball, the divisor $\mathscr{P}$ is simply the image of the section that extends the point P, and the capacitary metric on $T_P S$ is simply the metric induced by the integral model.

Example 5.12 (Comparison with other definitions). Let us show how this norm fits with Rumely's definition in [49] of the capacity of U with respect to the point P. Let f be a rational function on X, without pole except P, such that $U = \{x \in X\,;\, |f(x)| \leqslant 1\}$. Let m be the order of f at P and let us define $c_P \in K^*$ so that $f(x) = c_P t(x)^{-m} + \cdots$ around P, where t is a fixed local parameter at P. By definition of the adjunction map, the local section $\frac{1}{t} 1_P$ of $\mathscr{O}_X(P)$ maps to the tangent vector $\frac{\partial}{\partial t}$. Consequently,

$$\left\| \frac{\partial}{\partial t} \right\|_{P,\Omega}^{\mathrm{cap}} = \left\| \frac{1}{t} 1_P \right\| (P) = \lim_{x \to P} |t(x)|^{-1} \min(1, |f(x)|^{-1/m}) = |c_P|^{-1/m}. \tag{5.13}$$

As an example, and to make explicit the relation of our rationality criterion below with the classical theorem of Borel–Dwork later on, let us consider the classical case in which $X = \mathbf{P}^1$ (containing the affine line with t coordinate), and U is the affinoid subspace of $\mathbf{P}^1$ defined by the inequality $|t| \geqslant r$ (to which we add the point at infinity), where $r \in \sqrt{|K^*|}$. Let us note that $\Omega = \complement U$ and choose for the point $P \in \Omega$ the point with coordinate $t = 0$. Let m be a positive integer and $a \in K^*$ such that $r^m = |a|$; let $f = a/t^m$; this is a rational function on $\mathbf{P}^1$ with a single pole at P, and U is defined by the inequality $|f| \leqslant 1$. It follows that

$$\left\| \frac{\partial}{\partial t} \right\|_{P,\Omega}^{\mathrm{cap}} = |a|^{-1/m} = 1/r.$$

Similarly, assume that U is an affinoid subset of $\mathbf{P}^1$ that does not contain the point $P = \infty$. Then U is bounded and $\|t^2 \frac{\partial}{\partial t}\|_{P,\Omega}$ is nothing but its transfinite diameter in the sense of Fekete. (See [1]; the equivalence of both notions follows from [49, Theorem 4.1.19, p. 204]; see also [49, Theorem 3.1.18, p. 151] for its archimedean counterpart.)

Remark 5.14. (a) Let (X, U) be as above, let $P \in X(K)$ be a rational point such that $P \notin U$. Let $\Omega = X^{\mathrm{an}} \setminus U$ and define Ω_0 to be the connected component of Ω that contains P. It follows from Proposition 5.5 that the norms $\|\cdot\|_{P,\Omega_0}^{\mathrm{cap}}$ and $\|\cdot\|_{P,\Omega}^{\mathrm{cap}}$ on $T_P X$ coincide.

(b) Let U' be another affinoid subspace of X^{an} such that $U' \subset U$; the complementary subset Ω' to U' satisfies $\Omega \subset \Omega'$. If, moreover, Ω and Ω' are connected, then for any $P \in \Omega$ and any vector $v \in T_P X$, one has

$$\|v\|_{P,\Omega'}^{\mathrm{cap}} \leqslant \|v\|_{P,\Omega}^{\mathrm{cap}}.$$

Indeed, since Ω and Ω' are connected and contain P, Proposition 5.6 implies that there exist rational functions f and f' on X, without pole except P, such that the affinoids U and U' are defined by the inequalities $|f| \leqslant 1$ and $|f'| \leqslant 1$ respectively. Replacing f and f' by some positive powers, we may also assume that $\operatorname{ord}_P(f) = \operatorname{ord}_P(f')$; let us denote it by $-d$. Let t be a local parameter at P; it is enough to prove the desired inequality for $v = \frac{\partial}{\partial t}$.

We may expand f and f' around P as Laurent series in $t - t(P)$, writing

$$f = \frac{c}{(t-t(P))^d} + \cdots, \quad f' = \frac{c'}{(t-t(P))^d} + \cdots \quad .$$

The rational function $g = f/f'$ on X defines a holomorphic function on the affinoid subspace defined by the inequality $\{|f'| \geqslant 1\}$, since the poles at P in the numerator and denominator cancel each other; moreover, $g(P) = c/c'$. Using the maximum principle twice (Proposition B.1) we have

$$\begin{aligned} |g(P)| &\leqslant \sup_{|f'(x)|\geqslant 1} |g(x)| = \sup_{|f'(x)|=1} |g(x)| = \sup_{|f'(x)|=1} |f(x)| \\ &= \sup_{|f'(x)|\leqslant 1} |f(x)| \leqslant 1, \end{aligned}$$

since $\Omega \subset \Omega'$. This implies that $|g(P)| \leqslant 1$, so that $|c| \leqslant |c'|$. Therefore,

$$\left\| \frac{\partial}{\partial t} \right\|_{P,\Omega'}^{\mathrm{cap}} = |c'|^{-1/d} \leqslant |c|^{-1/d} = \left\| \frac{\partial}{\partial t} \right\|_{P,\Omega}^{\mathrm{cap}},$$

as was to be shown.

5.D Canonical seminorms and capacities

Let K be a local field.

In the case that K is archimedean, we assume moreover that $K = \mathbf{C}$; let M be a connected Riemann surface, and let Ω be an open subset in M, relatively compact. In the case that K is ultrametric, let M be a smooth projective curve over K, let U be an affinoid in M^{an}, let us set $\Omega = M^{\mathrm{an}} \setminus U$.

In both cases, let O be a point in Ω.

We endow the K-line $T_O M$ with its capacitary seminorm, as defined by the first author in [13] when $K = \mathbf{C}$, or in the previous section in the p-adic case.

Let X be a projective variety over K, let $P \in X(K)$ be a rational point, and let $\widehat{C}$ be a smooth formal curve in $\widehat{X}_P$. Assume that $\widehat{C}$ is K-analytic and let $\varphi\colon \Omega \to X^{\mathrm{an}}$ be an analytic map such that $\varphi(O) = P$ that maps the germ of Ω at O to $\widehat{C}$. (Consequently, if $D\varphi(O) \neq 0$, then φ defines an analytic isomorphism from the formal germ of Ω at O to $\widehat{C}$.) We endow $T_P X$ with its canonical seminorm $\|\cdot\|_{X,\widehat{C}}^{\mathrm{can}}$.

Proposition 5.15. *For any $v \in T_O\Omega$, one has*

$$\|D\varphi(O)(v)\|^{\mathrm{can}}_{X,\widehat{C}} \leqslant \|v\|^{\mathrm{cap}}_{P,\Omega}.$$

Proof. The case $K = \mathbf{C}$ is treated in [13, Proposition 3.6]. It therefore remains to treat the ultrametric case.

In view of Remark 5.14 (a), we may assume that Ω is connected. By Proposition 5.6, there exists a rational function $f \in K(M)$ without pole except O such that $U = \{x \in M\,;\, |f(x)| \leqslant 1\}$. Let $m > 0$ denote the order of the pole of f at the point O. For any real number $r > 1$ belonging to $\sqrt{|K|^*}$, let us denote by U_r and ∂U_r the affinoids $\{|f(x)| \geqslant r\}$ and $\{|f(x)| = r\}$ in M. One has $\bigcup_{r>1} U_r = \Omega$. We shall denote by φ_r the restriction of φ to the affinoid U_r. Let us also fix a local parameter t at O and let us define $c_P = \lim_{x\to O} t(x)^m f(x)$. One has $\|\frac{\partial}{\partial t}\|^{\mathrm{cap}}_{O,\Omega} = |c_P|^{-1/m}$.

Let L be an ample line bundle on X For the proof of the proposition, we may assume that $D\varphi(O)$ is nonzero; then φ is a formal isomorphism and we may consider the formal parameter $\tau = t\circ\varphi^{-1}$ on $\widehat{C}$ at P. We have $dt = \varphi^* d\tau$, hence $D\varphi(O)(\frac{\partial}{\partial t}) = \frac{\partial}{\partial \tau}$. Let us also fix a norm $\|\cdot\|_0$ on the K-line $T_P\widehat{C}$, and let us still denote by $\|\cdot\|_0$ the associated norm on its dual $T_P^\vee\widehat{C}$.

Let us choose a real number $r > 1$ such that $r \in \sqrt{|K^*|}$, fixed for the moment. Since the residue field of K is finite, the line bundle φ_r^*L on U_r is torsion (see [44, Proposition 3.1]); we may therefore consider a positive integer n and a nonvanishing section ε of $\varphi_r^*L^{\otimes n}$. For any integer D and any section $s \in \Gamma(X, L^{\otimes nD})$, let us write $\varphi_r^* s = \sigma\varepsilon^{\otimes nD}$, where σ is an analytic function on U_r. Since we assumed that $D\varphi(O) \neq 0$, the condition that s vanishes to order i along $\widehat{C}$ means exactly that σ vanishes to order i at O. Consequently, the i-th jet of $\varphi_r^* s$ at O is given by

$$\mathrm{j}^i_O\,(\varphi_r^* s) = (\sigma t^{-i})(O)\varepsilon^{nD}(O) \otimes d\tau^{\otimes i}.$$

Writing $(\sigma t^{-i})^m = (\sigma^m f^i)(f t^m)^{-i}$, it follows that

$$\|\mathrm{j}^i_O\,(\varphi_r^* s)\|^m = |\sigma^m f^i|(O)|c_P|^{-i}\|\varepsilon(O)\|^{nmD}\|d\tau\|_0^{im}.$$

Notice that $\sigma^m f^i$ is an analytic function on U_r. By the maximum principle (Proposition B.1),

$$|\sigma^m f^i|(O) \leqslant \sup_{U_r}|\sigma^m f^i| = \sup_{x\in\partial U_r}|\sigma^m f^i(x)| = \|\sigma\|^m_{\partial U_r} r^i.$$

Consequently,

$$\begin{aligned}\|\mathrm{j}^i_O(s)\| &\leqslant \|\sigma\|_{\partial U_r}|c_P|^{-i/m} r^i \|\varepsilon(O)\|^{nD}\|d\tau\|_0^i \\ &\leqslant \|s\|_{\partial U_r}|c_P|^{-i/m} r^i \left(\frac{\|\varepsilon(O)\|}{\inf_{x\in\partial U_r}\|\varepsilon(x)\|}\right)^{nD} \|d\tau\|_0^i.\end{aligned}$$

With the notation of Section 4.B, it follows that the norm of the evaluation morphism
$$\varphi^i_{nD} \colon E^i_{nD} \to L^{\otimes nD}_{|P} \otimes \left(T^\vee_P \widehat{C}\right)^{\otimes i}$$
satisfies the inequality
$$\|\varphi^i_{nD}\|^{1/i} \leqslant r^{1/m} |c_P|^{-1/m} \left(\|\varepsilon(O)\| / \inf_{\partial U_r} \|\varepsilon\|\right)^{nD/i} \|d\tau\|_0,$$
hence
$$\limsup_{i/D\to\infty} \frac{1}{i} \log\|\varphi^i_{nD}\| \leqslant \frac{1}{m} \log \frac{r}{|c_P|} = \log\|d\tau\|_0.$$
Using the notation introduced for defining the canonical semi-norm, we thus have $\rho(L) = \rho(L^{\otimes n}) \leqslant \log\|d\tau\|_0$ and
$$\begin{aligned}\left\| D\varphi(O)\left(\frac{\partial}{\partial t}\right)\right\|^{\mathrm{can}}_{X,\widehat{C},P} &= \left\|\frac{\partial}{\partial \tau}\right\|^{\mathrm{can}}_{X,\widehat{C},P} = e^{\rho(L)} \left\|\frac{\partial}{\partial\tau}\right\|_0 \\ &\leqslant \left(\frac{r}{|c_P|}\right)^{1/m} = r^{1/m} \left\|\frac{\partial}{\partial t}\right\|^{\mathrm{cap}}_{\Omega,P}.\end{aligned}$$
Letting r go to 1, we obtain the desired inequality. □

5.E Global capacities

Let K be a number field, and let R denote the ring of integers in K. Let X be a projective smooth algebraic curve over K. For any ultrametric place v of R, let us denote by $\mathbf{F}_v$ the residue field of R at v, by K_v the completion of K at v, and by X_v the rigid K_v-analytic variety attached to X_{K_v}. For any archimedean place v of X, corresponding to an embedding $\sigma\colon K \hookrightarrow \mathbf{C}$, we let X_v be the compact Riemann surface $X_\sigma(\mathbf{C})$. When v is real, by an open subset of X_v we shall mean an open subset of $X_\sigma(\mathbf{C})$ invariant under complex conjugation.

Our goal in this section is to show how capacitary metrics at all places fit within the framework of the Arakelov intersection theory (with L^2_1-regularity) introduced in [11]. Let us briefly recall here the main notation and properties of this arithmetic intersection theory, referring to this article for more details.

For any normal projective flat model $\mathscr{X}$ of X over R, the Arakelov Chow group $\widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X})$ consists of equivalence classes of pairs $(\mathscr{D}, g) \in \widehat{\mathrm{Z}}^1_{\mathbf{R}}(\mathscr{X})$, where $\mathscr{D}$ is an $\mathbf{R}$-divisor on $\mathscr{X}$ and g is a Green current with L^2_1-regularity on $\mathscr{X}(\mathbf{C})$ for the real divisor $\mathscr{D}_K$, stable under complex conjugation. For any class α of an Arakelov divisor $(\mathscr{D}, g)$, we shall denote, as usual, $\omega(\alpha) = dd^c g + \delta_{\mathscr{D}_{\mathbf{C}}}$.

Arithmetic intersection theory endowes the space $\widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X})$ with a symmetric $\mathbf{R}$-valued bilinear form. Any morphism $\pi\colon \mathscr{X}' \to \mathscr{X}$ between normal projective flat models of curves X' and X induces morphisms of abelian groups

$\pi_* \colon \widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X}') \to \widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X})$ and $\pi^* \colon \widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X}) \to \widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X}')$. For any classes α and $\beta \in \widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X})$, $\gamma \in \widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X}')$, one has $\pi^*\alpha \cdot \pi^*\beta = \alpha \cdot \beta$ and a projection formula $\pi_*(\pi^*\alpha \cdot \gamma) = \deg(\pi)\alpha \cdot \pi_*(\gamma)$, when π has constant generic degree $\deg(\pi)$.

Any class $\alpha \in \widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X})$ defines a *height* function h_α that is a linear function on the subspace of $Z^1_{\mathbf{R}}(\mathscr{X})$ consisting of real 1-cycles Z on $\mathscr{X}$ such that $\omega(\alpha)$ is locally L^∞ on a neighbourhood of $|Z|(\mathbf{C})$. If D is a real divisor on X such that $\omega(\alpha)$ is locally L^∞ in a neighbourhood of $|D|(\mathbf{C})$, we shall still denote by $h_\alpha(D)$ the height of the unique horizontal 1-cycle on $\mathscr{X}$ that extends D. Moreover, for any effective divisor D on X such that $\omega(\alpha)$ is locally L^∞ in a neighborhood of $|\pi_*(D)|(\mathbf{C})$, then $\omega(\pi^*\alpha)$ is locally L^∞ in a neighborhood of $|D|(\mathbf{C})$, and one has the equality $h_{\pi^*\alpha}(D) = h_\alpha(\pi_*(D))$.

Definition 5.16. *Let D be a divisor on X. For each place v of K, let Ω_v be an open subset of X_v (stable under complex conjugation if v is archimedean). One says that the collection (Ω_v) is an* adelic tube adapted to D *if the following conditions are satisfied:*

(1) *for any ultrametric place v, the complement of Ω_v in any connected component of X_v is a nonempty affinoid subset;*
(2) *for any archimedean place v, the complement of Ω_v in any connected component of X_v is nonpolar;*
(3) *there exist an effective reduced divisor E containing $|D|$, a finite set of places F of K, and a normal projective flat model $\mathscr{X}$ of X over R such that for any ultrametric place v of K such that $v \notin F$, $\Omega_v =]\overline{E}[_v$ is the tube in X_v around the specialization of E in the special fiber $\mathscr{X}_{\mathbf{F}_v}$.*

Let $\Omega = (\Omega_v)$ be a family where, for each place v of K, Ω_v is an open subset of the analytic curve X_v satisfying conditions (1) and (2). Let D be a divisor on X whose support is contained in Ω_v for any place v of K. By the considerations of this section, the line bundle $\mathscr{O}_X(D)$ is then endowed, for each place v of K, with a v-adic metric $\|\cdot\|^{\mathrm{cap}}_{\Omega_v}$. If Ω is an adelic tube adapted to D, then for almost all places of K, this metric is in fact induced by the horizontal extension of the divisor D in an adequate model $\mathscr{X}$ of X. Actually, one has the following proposition:

Proposition 5.17. *Assume that Ω is an adelic tube adapted to $|D|$. There is a normal, flat, projective model $\mathscr{X}$ of X over R and a (unique) Arakelov* $\mathbf{Q}$*-divisor extending D, inducing at any place v of K the v-adic capacitary metric on $\mathscr{O}_X(D)$.*

Such an arithmetic surface $\mathscr{X}$ will be said to be *adapted to Ω*. Then the Arakelov $\mathbf{Q}$-divisor on $\mathscr{X}$ whose existence is asserted by the proposition will be denoted by $\widehat{D}_\Omega$. Observe, moreover, that the current $\omega(\widehat{D}_\Omega)$ is locally L^∞ on Ω, since it vanishes there. Consequently, the height $h_{\widehat{D}_\Omega}(E)$ is defined when E is any 0-cycle on X that is supported by Ω.

Proof. It has already been recalled that archimedean Green functions defined by potential theory have the required L^2_1-regularity. It thus remains to show that the metrics at finite places can be defined using a single model $(\mathscr{X}, \mathscr{D})$ of (X, D) over R.

Lemma 5.18. *There exists a normal, flat projective model $\mathscr{X}$ of X over R, and, for any ultrametric place v of K, a Zariski closed subset Z_v of the special fiber $\mathsf{X}_{\mathbf{F}_v}$ at v such that $\Omega_v =]\mathsf{Z}_v[$. We may, moreover, assume that for almost all ultrametric places v of K, $\mathsf{Z}_v = \mathscr{E} \cap \mathsf{X}_{\mathbf{F}_v}$, where $\mathscr{E}$ is an effective reduced horizontal divisor on $\mathscr{X}$.*

Proof. Let $\mathscr{X}_1$ be a projective flat model of X over R, E an effective reduced divisor on $\mathscr{X}$, and F a finite set of places satisfying condition (3) of the definition of an adelic tube. Up to enlarging F, we may assume that the fiber product $\mathscr{X}_1 \otimes_R R_1$ is normal, where R_1 denotes the subring of K obtained from R by localizing outside places in F.

By Raynaud's formal/rigid geometry comparison theorem, there are, for each finite place $v \in F$, a normal projective and flat model $\mathscr{X}_v$ of X over the completion $\widehat{R_v}$, and a Zariski closed subset Z_v of the special fiber of $\mathscr{X}_v$, such that $\Omega_v =]\mathsf{Z}_v[$.

By a general descent theorem of Moret-Bailly ([41, Th. 1.1]; see also [10, 6.2, Lemma D]), there exists a projective and flat R-scheme $\mathscr{X}$ that coincides with $\mathscr{X}_1$ over $\operatorname{Spec} R_1$ and such that its completion at any finite place $v \in F$ is isomorphic to $\mathscr{X}_v$. By faithfully flat descent, such a scheme is normal (see [39, 21.E, Corollary]).

For any ultrametric place v over $\operatorname{Spec} R_1$, we just let Z_v be the specialization of E in $\mathscr{X}_{\mathbf{F}_v} = (\mathscr{X}_0)_{\mathbf{F}_v}$; one has $\Omega_v =]\mathsf{Z}_v[$ by assumption, since v does not belong to the finite set F of excluded places. For any ultrametric place $v \in F$, Z_v is identified with a Zariski closed subset of the special fiber $\mathscr{X}_{\mathbf{F}_v}$ and its tube is equal to Ω_v by construction. This concludes the proof of the lemma. □

Fix such a model $\mathscr{X}$ and let $\mathscr{D}_0$ be the Zariski closure of D in $\mathscr{X}$. For any ultrametric place v of F, let V_v be the unique divisor on the special fiber $\mathscr{X}_{\mathbf{F}_v}$ such that $\mathscr{D}_0 + V_v$ satisfies the assumptions of Proposition 5.1. One has $V_v = 0$ for any ultrametric place v such that Z_v has no component of dimension 1, hence for all but finitely places v. We thus may consider the $\mathbf{Q}$-divisor $\mathscr{D} = \mathscr{D}_0 + \sum_v V_v$ on $\mathscr{X}$ and observe that it induces the capacitary metric at all ultrametric places. □

Proposition 5.19. *Let D be a divisor on X and let Ω be an adelic tube adapted to $|D|$. One has the equality*

$$\widehat{D}_\Omega \cdot \widehat{D}_\Omega = h_{\widehat{D}_\Omega}(D).$$

Proof. Let us consider a model $\mathscr{X}$ of X and an Arakelov $\mathbf{Q}$-divisor $\mathscr{D}$ on $\mathscr{X}$ defining the capacitary metric $\|\cdot\|^{\mathrm{cap}}_{\Omega_v}$ at all ultrametric places v of K.

Let $\mathscr{D}_0$ denote the Zariski closure of D in $\mathscr{X}$. For any ultrametric place v of K, let V_v be the vertical part of $\mathscr{D}$ lying above v, so that $\mathscr{D} = \mathscr{D}_0 + \sum_v V_v$. By [11, Cor. 5.4], one has

$$\widehat{D}_\Omega \cdot \widehat{D}_\Omega = h_{\widehat{D}_\Omega}(\mathscr{D}).$$

By the definition of the capacitary metric at ultrametric places, the geometric intersection number of $\mathscr{D}$ with any vertical component of $\mathscr{D}$ is zero. Consequently,

$$\widehat{D}_\Omega \cdot \widehat{D}_\Omega = h_{\widehat{D}_\Omega}(\mathscr{D}_0) + \sum_v h_{\widehat{D}_\Omega}(V_v) = h_{\widehat{D}_\Omega}(\mathscr{D}_0),$$

as was to be shown. □

Corollary 5.20. *Let $P \in X(K)$ be a rational point of X and let Ω be an adelic tube adapted to P. One has*

$$\widehat{P}_\Omega \cdot \widehat{P}_\Omega = \widehat{\deg}\,(T_P X, \|\cdot\|^{\mathrm{cap}}_\Omega)\,.$$

6 An algebraicity criterion for *A*-analytic curves

Let K be a number field, R its ring of integers, X a quasiprojective algebraic variety over K, and let P be a point in $X(K)$. Let $\widehat{C} \hookrightarrow \widehat{X}_P$ be a smooth formal curve that is A-analytic.

For any place v of K, the formal curve $\widehat{C}$ is K_v-analytic, and we may equip the K-line $T_P\widehat{C}$ with the canonical v-adic seminorm $\|\cdot\|^{\mathrm{can}}_v = \|\cdot\|^{\mathrm{can}}_{X,\widehat{C},P,v}$ constructed in Section 4.B. We claim that equipped with these semi-norms, $T_P\widehat{C}$ defines a seminormed K-line $(T_P\widehat{C}, \|\cdot\|^{\mathrm{can}})$ with a well-defined Arakelov degree in $]-\infty, +\infty]$, in the sense of [13, 4.2]. Recall that this means that, for any (or equivalently, for some) nonzero element in $T_P\widehat{C}$, the series $\sum_v \log^+\|t\|^{\mathrm{can}}_v$ is convergent. To see this, consider a quasiprojective flat R-scheme $\mathscr{X}$ with generic fiber X, together with a section $\mathscr{P}\colon \operatorname{Spec} R \to \mathscr{X}$ that extends P. According to Lemma 4.5 (applied to projective compactifications of X and $\mathscr{X}$, and an ample line bundle $\mathscr{L}$), the inequality

$$\log\|t\|^{\mathrm{can}}_v \leqslant -\log S_{\mathscr{X},v}(\widehat{C})$$

holds for almost all finite places v, where $S_{\mathscr{X},v}$ denotes the size of $\widehat{C}$ with respect to the R_v model $\mathscr{X} \otimes R_v$. Since by definition of A-analyticity the series with nonnegative terms $\sum_v \log S_{\mathscr{X},v}(\widehat{C})^{-1}$ has a finite sum, this establishes the required convergence.

The Arakelov degree of $\left(T_P\widehat{C}, \|\cdot\|^{\mathrm{can}}\right)$ is defined as the sum

$$\widehat{\deg}\left(T_P\widehat{C}, \|\cdot\|^{\mathrm{can}}\right) := \sum_v \left(-\log\|t\|^{\mathrm{can}}_v\right).$$

It is a well-defined element in $]-\infty, +\infty]$, independent of the choice of t by the product formula (we follow the usual convention $-\log 0 = +\infty$.)

The following criterion extends Theorem 4.2 of [13], where instead of canonical seminorms, larger norms constructed by means of the sizes were used at finite places.

Theorem 6.1. *Let $\widehat{C}$ be, as above, an A-analytic curve through a rational point P in some algebraic variety X over K.*

If $\widehat{\deg}\left(T_P\widehat{C}, \|\cdot\|^{\mathrm{can}}\right) > 0$, then $\widehat{C}$ is algebraic.

Proof. We keep the above notation, and we assume, as we may, X (respectively $\mathscr{X}$) to be projective over K (respectively over R). We choose an ample line bundle $\mathscr{L}$ over $\mathscr{X}$ and we let $L := \mathscr{L}_K$.

We let $\mathscr{E}_D := \Gamma(\mathscr{X}, \mathscr{L}^{\otimes D})$, and for any embedding $\sigma : K \hookrightarrow \mathbf{C}$, we choose a consistent sequence of hermitian norms $(\|\cdot\|_{D,\sigma})$ on the $\mathbf{C}$-vector spaces $\mathscr{E}_{D,\sigma} \simeq \Gamma\left(X_\sigma, L_\sigma^{\otimes D}\right)$, in a way compatible with complex conjugation. Using these norms, we define hermitian vector bundles $\overline{\mathscr{E}}_D := (\mathscr{E}_D, (\|\cdot\|_{D,\sigma})_\sigma)$ over $\operatorname{Spec} R$.

We also choose a hermitian structure on $\mathscr{P}^*\mathscr{L}$, and we denote by $\overline{\mathscr{P}^*\mathscr{L}}$ the so-defined hermitian line bundle over $\operatorname{Spec} R$. Finally, we equip $T_P\widehat{C}$ with the R-structure defined by $N_{\mathscr{P}}\mathscr{X} \cap T_P\widehat{C}$ and with an arbitrary hermitian structure, and in this way we define a hermitian line bundle $\overline{T}_0$ over $\operatorname{Spec} R$ such that $(T_0)_K = T_P\widehat{C}$.

We define the K-vector spaces $E_D := \mathscr{E}_{D,K} \simeq \Gamma(X, L^{\otimes D})$, their subspaces E_D^i, and the evaluation maps

$$\varphi_D^i \colon E_D^i \to \left(T_P^\vee\widehat{C}\right)^{\otimes i} \otimes L_{|P}^{\otimes D}$$

as in the "local" situation considered in Section 4.B. According to the basic algebraicity criteria in [13, 2.2], to prove that $\widehat{C}$ is algebraic, it suffices to prove that the ratio

$$\frac{\sum\limits_{i\geqslant 0}(i/D)\operatorname{rank}\left(E_D^i/E_D^{i+1}\right)}{\sum\limits_{i\geqslant 0}\operatorname{rank}\left(E_D^i/E_D^{i+1}\right)} \tag{6.2}$$

stays bounded as D goes to $+\infty$.

For any place v of K, the morphism φ_D^i has a v-adic norm, defined by means of the integral and hermitian structures introduced above. If $\varphi_D^i \neq 0$, the height of φ_D^i is the real number defined as the (finite) sum

$$h\left(\varphi_D^i\right) = \sum_v \log\|\varphi_D^i\|_v.$$

When φ_D^i vanishes, we define $h\left(\varphi_D^i\right) = -\infty$; observe that in this case $E_D^{i+1} = E_D^i$.

As established in the proof of Lemma 4.5 above (see also [12, Lemma 3.3]), the following inequality holds for any finite place v and any two nonnegative integers i and D:

$$\log\|\varphi_D^i\|_v \leqslant -i \log S_{\mathscr{X},v}(\widehat{C}). \tag{6.3}$$

Since $\widehat{C}$ is A-analytic, the upper bounds (4.4) and (6.3) show the existence of some positive real number c such that

$$h\left(\varphi_D^i\right) \leqslant c(i+D). \tag{6.4}$$

For any place v of K, we let

$$\rho_v(L) = \limsup_{i/D\to\infty} \frac{1}{i} \log\|\varphi_D^i\|_v.$$

This is an element in $[-\infty, +\infty[$, which, according to (6.3), satisfies

$$\rho_v(L) \leqslant -\log S_{\mathscr{X},v}(\widehat{C})$$

for any finite place v. Moreover, by its very definition, the Arakelov degree of $\left(T_P\widehat{C}, \|\cdot\|^{\mathrm{can}}\right)$ is given by

$$\begin{aligned}\widehat{\deg}\left(T_P\widehat{C}, \|\cdot\|^{\mathrm{can}}\right) &= \sum_v (-\rho_v(L)) + \widehat{\deg}\,\overline{T}_0 \\ &= \sum_{v \text{ finite}} \left(-\rho_v(L) - \log S_{\mathscr{X},v}(\widehat{C})\right) \\ &\qquad + \sum_{v \text{ finite}} \log S_{\mathscr{X},v}(\widehat{C}) + \sum_{v|\infty} (-\rho_v(L)) + \widehat{\deg}\,\overline{T}_0.\end{aligned}$$

In the last expression, the terms of the first sum belong to $[0, +\infty]$ — and the sum itself is therefore well-defined in $[0, +\infty]$ — and the second sum is convergent by A-analyticity of $\widehat{C}$.

Observe also that since the sums

$$\sum_{v \text{ finite}} \left(-\frac{1}{i} \log\|\varphi_D^i\|_v + \log S_{\mathscr{X},v}(\widehat{C})\right)$$

have nonnegative terms, we get, as a special instance of Fatou's lemma:

$$\begin{aligned}&\sum_{v \text{ finite}} \liminf_{i/D\to\infty} \left(-\frac{1}{i} \log\|\varphi_D^i\|_v + \log S_{\mathscr{X},v}(\widehat{C})\right) \\ &\qquad\qquad \leqslant \liminf_{i/D\to\infty} \sum_{v \text{ finite}} \left(-\frac{1}{i} \log\|\varphi_D^i\|_v + \log S_{\mathscr{X},v}(\widehat{C})\right).\end{aligned}$$

Consequently
$$\limsup_{i/D\to\infty} \frac{1}{i} h\left(\varphi_D^i\right) \leqslant \sum_v \rho_v(L)$$
and
$$\widehat{\deg}\left(T_P\widehat{C}, \|\cdot\|^{\mathrm{can}}\right) \leqslant -\limsup_{i/D\to\infty} \frac{1}{i} h\left(\varphi_D^i\right) + \widehat{\deg}\,\overline{T}_0. \tag{6.5}$$

When $\widehat{\deg}\left(T_P\widehat{C}, \|\cdot\|^{\mathrm{can}}\right)$ is positive, the inequality (6.5) implies the existence of positive real numbers ε and λ such that, for any two positive integers i and D,
$$\widehat{\deg}\,\overline{T}_0 - \frac{1}{i} h\left(\varphi_D^i\right) \geqslant \varepsilon \quad \text{if } i \geqslant \lambda D. \tag{6.6}$$

Let $\mathscr{E}_D^i := \mathscr{E}_D \cap E_D^i$ and let $\overline{\mathscr{E}_D^i/\mathscr{E}_D^{i+1}}$ be the hermitian vector bundle on $\operatorname{Spec} R$ defined by the quotient $\mathscr{E}_D^i/\mathscr{E}_D^{i+1}$ equipped with the hermitian structure induced by that of $\overline{\mathscr{E}}_D$. The evaluation map φ_D^i induces an injection $E_D^i/E_D^{i+1} \hookrightarrow \left(T_P^\vee\widehat{C}\right)^{\otimes i} \otimes L_{|P}^{\otimes D}$. Actually, either $\varphi_D^i = 0$ and then $E_D^i = E_D^{i+1}$, or $\varphi_D^i \neq 0$, and this inclusion is an isomorphism of K-lines. In either case, we have
$$\widehat{\deg}\,\overline{\mathscr{E}_D^i/\mathscr{E}_D^{i+1}} = \operatorname{rank}\left(E_D^i/E_D^{i+1}\right)\left(\widehat{\deg}\left(\overline{\mathscr{P}^*\mathscr{L}}^{\otimes D} \otimes \overline{T}_0^{\vee\otimes i}\right) + h\left(\varphi_D^i\right)\right).$$

Indeed, if $\varphi_D^i = 0$, both sides vanish (we follow the usual convention $0\cdot(-\infty) = 0$). If $\varphi_D^i \neq 0$, the equality is a straightforward consequence of the definitions of the Arakelov degree of a hermitian line bundle over $\operatorname{Spec} R$ and of the heights $h\left(\varphi_D^i\right)$.

The above equality may also be written
$$\widehat{\deg}\,\overline{\mathscr{E}_D^i/\mathscr{E}_D^{i+1}} = \operatorname{rank}\left(E_D^i/E_D^{i+1}\right)\left(D\,\widehat{\deg}\,\overline{\mathscr{P}^*\mathscr{L}} - i\,\widehat{\deg}\,\overline{T}_0 + h\left(\varphi_D^i\right)\right). \tag{6.7}$$

Moreover, by [12, Proposition 4.4], there is a constant c' such that for any $D \geqslant 0$ and any saturated submodule $\mathscr{F}$ of $\mathscr{E}_D$,
$$\widehat{\deg}\,\overline{\mathscr{E}_D/\mathscr{F}} \geqslant -c' D \operatorname{rank}(\mathscr{E}_D/\mathscr{F}).$$

(This is an easy consequence of the fact that the K-algebra $\bigoplus_{D\geqslant 0} \mathscr{E}_{D,K}$ is finitely generated.) Applied to $\mathscr{F} := \bigcap_{i\geqslant 0} \mathscr{E}_D^i$, this estimate becomes
$$\sum_{i\geqslant 0} \widehat{\deg}\,\overline{\mathscr{E}_D^i/\mathscr{E}_D^{i+1}} \geqslant -c' D \sum_{i\geqslant 0} \operatorname{rank}\left(E_D^i/E_D^{i+1}\right). \tag{6.8}$$

Using (6.7) and (6.8), we derive the inequality

$$- (c' + \widehat{\deg}\,\overline{\mathscr{P}^*\mathscr{L}}) D \sum_{i \geqslant 0} \operatorname{rank}\left(E_D^i / E_D^{i+1}\right) \leqslant \sum_{i \geqslant 0} \operatorname{rank}\left(E_D^i / E_D^{i+1}\right) \left(-i\,\widehat{\deg}\,\overline{T}_0 + h\left(\varphi_D^i\right)\right). \quad (6.9)$$

Finally, using (6.9), (6.4), and (6.6), we obtain

$$\sum_{i < \lambda D} \operatorname{rank}\left(E_D^i / E_D^{i+1}\right) \left(\frac{i}{D}\,\widehat{\deg}\,\overline{T}_0 - c\frac{i+D}{D}\right) + \sum_{i \geqslant \lambda D} \operatorname{rank}\left(E_D^i / E_D^{i+1}\right) \varepsilon \frac{i}{D} \leqslant (c' + \widehat{\deg}\,\overline{\mathscr{P}^*\mathscr{L}}) \sum_{i \geqslant 0} \operatorname{rank}\left(E_D^i / E_D^{i+1}\right).$$

This implies that the ratio (6.2) is bounded by

$$\lambda + \frac{1}{\varepsilon}\left(c' + \widehat{\deg}\,\overline{\mathscr{P}^*\mathscr{L}} + c + \lambda \max(0, c - \widehat{\deg}\,\overline{T}_0)\right),$$

and completes the proof. □

7 Rationality criteria

7.A Numerical equivalence and numerical effectivity on arithmetic surfaces

The following results are variations on a classical theme in Arakelov geometry of arithmetic surfaces. The first theorem characterizes numerically trivial Arakelov divisors with real coefficients. It is used in the next proposition to describe effective Arakelov divisors whose sum is numerically effective. We allow ourselves to use freely the notation of [11].

Theorem 7.1. (Compare [11, Thm. 5.5]) *Let $\mathscr{X}$ be a normal flat projective scheme over the ring of integers of a number field K whose generic fiber is a smooth and geomerically connected curve. Let (D,g) be any element in $\widehat{\mathrm{Z}}^1_{\mathbf{R}}(\mathscr{X})$ that is numerically trivial. Then there exist an integer n, real numbers λ_i and rational functions $f_i \in K(\mathscr{X})^*$, for $1 \leqslant i \leqslant n$, and a family $(c_\sigma)_{\sigma\colon K \hookrightarrow \mathbf{C}}$ of real numbers such that $c_{\overline{\sigma}} = c_\sigma$, $\sum c_\sigma = 0$, and $(D,g) = (0,(c_\sigma)) + \sum_{i=1}^n \lambda_i \widehat{\operatorname{div}}(f_i)$.*

Proof. There are real numbers λ_i and Arakelov divisors $(D_i, g_i) \in \widehat{\mathrm{Z}}^1(\mathscr{X})$ such that $(D,g) = \sum \lambda_i (D_i, g_i)$. We may assume that the λ_i are linearly independent over $\mathbf{Q}$. By assumption, the degree of D on any vertical component of $\mathscr{X}$ is zero; the linear independence of the λ_i implies that the same holds for any D_i. Let us then denote by g'_i any Green current for D_i such that $\omega(D_i, g'_i) = 0$. One has

$$0 = \omega(D, g) = \sum \lambda_i \omega(D_i, g_i) = \sum \lambda_i \omega(D_i, g'_i),$$

so that the difference $g - \sum \lambda_i g'_i$ is harmonic, and therefore constant on any connected component of $\mathscr{X}(\mathbf{C})$. By adding a locally constant function to some g'_i, we may assume that $g = \sum \lambda_i g'_i$. Then $(D, g) = \sum \lambda_i (D_i, g'_i)$. This shows that we may assume that one has $\omega(D_i, g_i) = 0$ for any i. By Faltings–Hriljac's formula, the Néron–Tate quadratic form on $\operatorname{Pic}^0(\mathscr{X}_K) \otimes \mathbf{R}$ takes the value 0 on the class of the real divisor $\sum \lambda_i (D_i)_K$. Since this quadratic form is positive definite (see [50, 3.8, p. 42]), this class is zero. Using that the λ_i are linearly independent over $\mathbf{Q}$, we deduce that the class of each divisor $(D_i)_K$ in $\operatorname{Pic}^0(\mathscr{X}_K)$ is torsion. Since D_i has degree zero on any vertical component of $\mathscr{X}$ and the Picard group of the ring of integers of K is finite, the class in $\operatorname{Pic}(\mathscr{X})$ of the divisor D_i is torsion too. Let us then choose positive integers n_i and rational functions f_i on $\mathscr{X}$ such that $\operatorname{div}(f_i) = n_i D_i$. The Arakelov divisors $\widehat{\operatorname{div}}(f_i) - n_i(D_i, g_i)$ are of the form $(0, c_i)$, where $c_i = (c_{i,\sigma})_{\sigma \colon K \hookrightarrow \mathbf{C}}$ is a family of real numbers such that $c_{i,\overline{\sigma}} = c_{i,\sigma}$ and $\sum_\sigma c_{i,\sigma} = 0$. Then, letting $c_\sigma = \sum_i (\lambda_i / n_i) c_{i,\sigma}$, one has

$$(D, g) = (0, (c_\sigma)) + \sum \frac{\lambda_i}{n_i} \widehat{\operatorname{div}}(f_i)$$

as requested. □

Let $f_1, \dots, f_n$ be meromorphic functions on some Riemann surface M, let $\lambda_1, \dots, \lambda_n$ be real numbers, and let $f \in \mathbf{C}(M)^* \otimes_{\mathbf{Z}} \mathbf{R}$ be defined as $f = \sum_{i=1}^n f_i \otimes \lambda_i$. We shall denote by $|f|$ the real function on M given by $\prod |f_i|^{\lambda_i}$, and by $\operatorname{div} f$ the $\mathbf{R}$-divisor $\sum \lambda_i \operatorname{div}(f_i)$; they don't depend on the decomposition of f as a sum of tensors. One has $dd^c \log|f|^{-2} + \delta_{\operatorname{div}(f)} = 0$.

We shall say that a pair (D, g) formed of a divisor D on M and of a Green current g with L^2_1 regularity for D is *effective*[6] if the divisor D is effective and if the Green current g of degree 0 for D may be represented by a nonnegative summable function (see [11, Def. 6.1]).

Similarly, we say that an Arakelov divisor $(D, g) \in \widehat{\mathrm{Z}}^1_{\mathbf{R}}(\mathscr{X})$ on the arithmetic surface $\mathscr{X}$ is effective if D is effective on $\mathscr{X}$ and if $(D_{\mathbf{C}}, g)$ is effective on $\mathscr{X}(\mathbf{C})$.

We say that an Arakelov divisor, or the class α of an Arakelov divisor, is *numerically effective* (for short, *nef*) if $[(D, g)] \cdot \alpha \geqslant 0$ for any effective Arakelov divisor $(D, g) \in \widehat{\mathrm{Z}}^1_{\mathbf{R}}(\mathscr{X})$ (according to [11, Lemma 6.6], it is sufficient to consider Arakelov divisors (D, g) with $\mathscr{C}^\infty$-regularity). If (D, g) is an effective and numerically effective Arakelov divisor, then the current $\omega(g) := dd^c g + \delta_D$ is a positive measure (see [11, proof of Proposition 6.9]).

[6] In the terminology of [11], *nonnegative*.

Proposition 7.2. *Let $\mathscr{X}$ be a normal, flat projective scheme over the ring of integers of a number field K whose generic fiber is a smooth geometrically connected algebraic curve.*

Let (D, g) and (E, h) be nonzero elements of $\widehat{\mathrm{Z}}^1_{\mathbf{R}}(\mathscr{X})$; let α and β denote their classes in $\widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X})$. Let us assume that the following conditions are satisfied:

(1) *the Arakelov divisors (D, g) and (E, h) are effective;*
(2) *the supports of D and E do not meet and $\int_{\mathscr{X}(\mathbf{C})} g * h = 0$.*

If the class $\alpha + \beta$ is numerically effective, then there exist a positive real number λ, an element $f \in K(\mathscr{X})^* \otimes_{\mathbf{Z}} \mathbf{R}$, and a family $(c_\sigma)_{\sigma\colon K \hookrightarrow \mathbf{C}}$ of real numbers that is invariant under conjugation and satisfies $\sum_\sigma c_\sigma = 0$ such that for any embedding $\sigma\colon K \hookrightarrow \mathbf{C}$,

$$g_\sigma = (c_\sigma + \log|f|^{-2})^+ \quad \text{and} \quad h_\sigma = \lambda(c_\sigma + \log|f|^{-2})^-,$$

where for any real-valued function φ, we define $\varphi^+ = \max(0, \varphi)$ and $\varphi^- = \max(0, -\varphi)$, so that $\varphi^+ - \varphi^- = \varphi$.

Moreover, $\alpha^2 = \alpha\beta = \beta^2 = 0$.

Proof. Since (D, g) and (E, h) are effective and nonzero, the classes α and β are not equal to zero ([11, Proposition 6.10]). Moreover, the assumptions of the proposition imply that

$$\alpha \cdot \beta = \deg \pi_*(D, E) + \frac{1}{2} \int_{\mathscr{X}(\mathbf{C})} g * h = 0.$$

Since $\alpha + \beta$ is numerically effective, it follows from Lemma 6.11 of [11] (which in turn is an application of the Hodge index theorem in Arakelov geometry) that there exists $\lambda \in \mathbf{R}^*_+$ such that $\beta = \lambda\alpha$ in $\widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X})$. In particular, α and β are nef, and $\alpha^2 = \beta^2 = \alpha \cdot \beta = 0$.

Replacing (E, h) by $(\lambda E, \lambda h)$, we may assume $\lambda = 1$. Then, $(D - E, g - h)$ belongs to the kernel of the canonical map $\rho\colon \widehat{\mathrm{Z}}^1_{\mathbf{R}}(\mathscr{X}) \to \widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X})$, so is numerically trivial. By Theorem 7.1, there exist real numbers λ_i, rational functions $f_i \in K(\mathscr{X})^*$, and a family $c = (c_\sigma)_{\sigma\colon K \hookrightarrow \mathbf{C}}$ of real numbers, invariant under conjugation, such that $\sum_\sigma c_\sigma = 0$ and $(D - E, g - h) = (0, c) + \sum \lambda_i \widehat{\mathrm{div}}(f_i)$ in $\widehat{\mathrm{Z}}^1_{\mathbf{R}}(\mathscr{X})$. Let us denote by f the element $\sum f_i \otimes \lambda_i$ of $K(\mathscr{X})^* \otimes_{\mathbf{Z}} \mathbf{R}$. The proposition now follows by applying Lemma 7.3 below to the connected Riemann surface $\mathscr{X}_\sigma(\mathbf{C})$, the pairs (D, g_σ), (E, h_σ), and the "meromorphic function" $e^{-2c_\sigma} f_{\mathscr{X}_\sigma(\mathbf{C})}$, for each embedding $\sigma\colon K \hookrightarrow \mathbf{C}$. $\square$

Lemma 7.3. *Let M be a compact connected Riemann surface, let D and D' be two nonzero $\mathbf{R}$-divisors on M, and let g and g' be two Green functions with L^2_1 regularity for D and D'. We make the following assumptions: $|D| \cap |D'| = \emptyset$, the pairs (D, g) and (D', g') are effective, the currents $\omega(g) = dd^c g + \delta_D$ and $\omega(g') = dd^c g' + \delta_{D'}$ are positive measures, $\int_M g * g' = 0$. If there exists an* *element $f \in \mathbf{C}(M)^* \otimes \mathbf{R}$ such that $g - g' = \log|f|^{-2}$, then $g = \max(0, \log|f|^{-2})$ and $g' = \max(0, \log|f|^2)$.*

Proof. First observe that

$$\omega(g) - \omega(g') = dd^c(g - g') + \delta_D - \delta_{D'} = dd^c \log|f|^{-2} + \delta_D - \delta_{D'} = \delta_{D - D' - \mathrm{div}(f)},$$

by the Poincaré–Lelong formula. By assumption, the current $\omega(g) - \omega(g')$ belongs to the Sobolev space L^2_{-1}; it is therefore nonatomic (see [11, Appendix, A.3.1]), so that $D - D' = \mathrm{div}(f)$ and $\omega(g) = \omega(g')$.

Observe also that $g_{|M \setminus |D|}$ (respectively $g'_{|M \setminus |D|'}$) is a subharmonic current. In the sequel, we denote by g (respectively g') the unique subharmonic function on $M \subset |D|$ (respectively on $M' \subset |D|'$) that represents this current.

Let F be the set of points $x \in M$ where $|f(x)| = 1$ and let $\Omega = M \setminus F$ be its complementary subset. The functions $h = \max(0, \log|f|^{-2})$ and $h' = \max(0, \log|f|^2)$ are continuous Green functions with L^2_1 regularity for D and D' respectively. The currents $dd^c\, h + \delta_D$, $dd^c\, h' + \delta_{D'}$ are equal to a common positive measure, which we denote by ν. Since h (respectively h') is harmonic on $M \setminus (|D| \cup F)$ (respectively on $M \setminus (|D|' \cup F)$), this measure is supported by F.

Let S be the support of the positive measure $\omega(g)$. It follows from [11, Remark 6.5] that g and g' vanish $\omega(g)$-almost everywhere on M. Consequently, the equality $\log|f|^{-2} = g - g' = 0$ holds $\omega(g)$-almost everywhere; in particular, $S \subset F$.

Let us pose $u = h - g = h' - g'$; this is a current with L^2_1 regularity on M and $dd^c\, u = dd^c\, h - dd^c\, g = \nu - \omega(g)$. In particular, $dd^c(u|_\Omega) = 0$: u is harmonic on Ω. Since g is nonnegative, one has $u \leqslant 0$ on $F = \complement\Omega$. By the maximum principle, this implies that $u \leqslant 0$ on Ω (cf. [11, Theorem A.6.1]; observe that u is finely continuous on M).

Finally, one has

$$0 = \int_M g * g' = \int_M h * h' - \int_M u\nu - \int_M u\omega(g) \geqslant \int_M h * h'.$$

By [11, Corollary 6.4], this last term is nonnegative, so that all terms of the formula vanish. In particular, $\int u\nu = 0$; hence $u = 0$ (ν-a.e.). Using again that u is harmonic on Ω, it follows that its Dirichlet norm vanishes, and finally that $u \equiv 0$. □

Remark 7.4. The Green currents g and h appearing in the conclusion of Proposition 7.2 are very special. Assume, for example, that the Arakelov divisors $\widehat{D}$ and $\widehat{E}$ are defined using capacity theory at the place σ, with respect to an open subset Ω_σ of X_σ. Then, g_σ and h_σ vanish nearly everywhere on $\complement\Omega_\sigma$. In other words, $\complement\Omega_\sigma$ is contained in the set of $x \in X_\sigma$ such that $|f(x)|^2 = \exp(-c_\sigma)$, which is a real semialgebraic curve in X_σ, viewed as a real algebraic surface. In particular, it contradicts any of the following hypotheses on Ω_σ, respectively denoted by $(4.2)_{\mathscr{X},\Omega_\sigma}$ and $(4.3)_{\mathscr{X},\Omega_\sigma}$ in [11]:

(1) the interior of $\mathscr{X}_\sigma(\mathbf{C}) \setminus \Omega_\sigma$ is not empty;
(2) there exists an open subset U of $\mathscr{X}_\sigma(\mathbf{C}) \setminus |D|(\mathbf{C})$ not contained in Ω such that any harmonic function on U that vanishes nearly everywhere on $U \setminus \Omega$ vanishes on U.

7.B Rationality criteria for algebraic and analytic functions on curves over number fields

Let K be a number field and X a smooth projective geometrically connected curve over K. For any place v of K, we denote by X_v the associated rigid analytic curve over K_v if v is ultrametric, respectively the corresponding Riemann surface $X_\sigma(\mathbf{C})$ if v is induced by an embedding of K in $\mathbf{C}$.

Let D be an effective divisor in X and $\Omega = (\Omega_v)_v$ an adelic tube adapted to $|D|$. We choose a normal projective flat model of X over the ring of integers $\mathscr{O}_K$ of K, say $\mathscr{X}$, and an Arakelov $\mathbf{Q}$-divisor $\widehat{D}_\Omega$ on $\mathscr{X}$ inducing the capacitary metrics $\|\cdot\|^{\mathrm{cap}}_{\Omega_v}$ at all places v of K. In particular, we assume that for any ultrametric place v, Ω_v is the tube $]\mathsf{Z}_v[$ around a closed Zariski subset Z_v of its special fiber $\mathscr{X}_{\mathbf{F}_v}$, and $\mathsf{Z}_v = \overline{D} \cap \mathscr{X}_{\mathbf{F}_v}$ for almost all places v.

Our first statement in this section is the following arithmetic analogue of Proposition 2.2.

Proposition 7.5. *Let X' be another geometrically connected smooth projective curve over K and let $f\colon X' \to X$ be a nonconstant morphism. Let D' be an effective divisor in X'. We make the following assumptions:*

(1) *by restriction, f defines an isomorphism from the subscheme D' of X' to the subscheme D of X and is étale in a neighbourhood of $|D'|$;*
(2) *for any place v of K, the morphism f admits an analytic section $\varphi_v\colon \Omega_v \to X'_v$ defined over Ω_v whose formal germ is equal to $\widehat{f_{D K_v}}^{-1}$;*
(3) *the class of the Arakelov $\mathbf{Q}$-divisor $\widehat{D}_\Omega$ is numerically effective.*

Assume moreover

($4'$) *either that $\widehat{D}_\Omega \cdot \widehat{D}_\Omega > 0$;*
($4''$) *or that there is an archimedean place v such that the complementary subset to Ω_v in X_v is not contained in a real semialgebraic curve of X_v.*

Then f is an isomorphism.

Proof. Let us denote by E the divisor f^*D on X'; we will prove that $E = D'$. Observe that according to assumption (1), this divisor may be written

$$E := f^*D = D' + R,$$

where R denotes an effective or zero divisor on X whose support is disjoint from that of D'.

Let $\mathscr{X}'$ denote the normalization of $\mathscr{X}$ in the function field of X' and let us still denote by f the natural map from $\mathscr{X}'$ to $\mathscr{X}$ that extends f. Then $\mathscr{X}'$

is a normal projective flat model of X' over $\mathscr{O}_K$. For any place v of K, let Ω'_v denote the preimage $f^{-1}(\Omega_v)$ of Ω_v by f. The complementary subset of Ω_v is a nonempty affinoid subspace of X'_v if v is ultrametric, and a nonpolar compact subset of X'_v if v is archimedean. Moreover, for almost all ultrametric places v, Ω'_v is the tube around the specialization in $\mathscr{X}'_{\mathbf{F}_v}$ of $f^{-1}(D)$. In particular, the collection $\Omega' = (\Omega'_v)$ is an adelic tube adapted to $|E|$.

We thus may assume that the capacitary metrics on $\mathscr{O}_{X'}(D')$ and $\mathscr{O}_{X'}(E)$ relative to the open subsets Ω'_v are induced by Arakelov $\mathbf{Q}$-divisors on $\mathscr{X}'$. Let us denote them by $\widehat{D'}_{\Omega'}$ and $\widehat{E}_{\Omega'}$ respectively.

Since X and X' are normal, and the associated rigid analytic spaces as well, the image $\varphi_v(\Omega_v)$ of Ω_v by the analytic section φ_v is a closed and open subset Ω^1_v of Ω'_v containing $|D'|$, and the collection $\Omega^1 = \big(\Omega^1_v\big)$ is an adelic tube adapted to $|D'|$. Consequently, by Proposition 5.5, one has $\widehat{D'}_{\Omega'} = \widehat{D'}_{\Omega^1}$. Similarly, writing $\Omega^2_v = \Omega'_v \setminus \Omega^1_v$, the collection $\Omega^2 = (\Omega^2_v)$ is an adelic tube adapted to $|R|$ and $\widehat{R}_{\Omega'} = \widehat{R}_{\Omega^2}$. One has $\widehat{E}_{\Omega'} = f^*\widehat{D}_{\Omega} = \widehat{D'}_{\Omega^1} + \widehat{R}_{\Omega^2}$. Since $\Omega^1_v \cap \Omega^2_v = \emptyset$ for any place v, Lemma 7.6 below implies that $[\widehat{R}_{\Omega^2}]\cdot[\widehat{D'}_{\Omega^1}] = 0$.

Since $\widehat{D}$ is nonzero and its class is numerically effective, the class in $\widehat{\mathrm{CH}}^1_{\mathbf{Q}}(\mathscr{X}')$ of the Arakelov divisor $f^*\widehat{D} = \widehat{D} + \widehat{R}$ is numerically effective too. Proposition 7.2 and Remark 7.4 show that, when either of the hypotheses (4′) (4″) is satisfied, necessarily $\widehat{R}_{\Omega^2} = 0$. In particular, $R = 0$ and $E = D'$. It follows that f has degree one, hence is an isomorphism. □

Lemma 7.6. *Let X be a geometrically connected smooth projective curve over a number field K, let D_1 and D_2 be divisors on X, and let Ω_1 and Ω_2 be adelic tubes adapted to $|D_1|$ and $|D_2|$. Let us consider a normal projective and flat model $\mathscr{X}$ of X over the ring of integers of K as well as Arakelov divisors $\widehat{D_1}_{\Omega_1}$ and $\widehat{D_2}_{\Omega_2}$ inducing the capacitary metrics on $\mathscr{O}_X(D_1)$ and $\mathscr{O}_X(D_2)$ relative to the adelic tubes Ω_1 and Ω_2.*

If $\Omega_{1,v} \cap \Omega_{2,v} = \emptyset$ for any place v of K, then

$$\widehat{D_1}_{\Omega_1} \cdot \widehat{D_2}_{\Omega_2} = 0.$$

Proof. Observe that D_1 and D_2 have no common component, since any point P common to D_1 and D_2 would belong to $\Omega_{1,v} \cap \Omega_{2,v}$.

Let $\mathscr{X}$ be a normal projective flat model of X adapted to Ω_1 and Ω_2, so that the classes $\widehat{D_i}_{\Omega_i}$ live in $\widehat{\mathrm{CH}}^1_{\mathbf{R}}(\mathscr{X})$. Namely $\widehat{D_i}_{\Omega_i} = (\mathscr{D}_i, g_i)$, where $\mathscr{D}_i$ is the $\mathbf{Q}$-divisor on $\mathscr{X}$ extending D_i defined by Proposition 5.1 and $g_i = (g_{D_i,\Omega_{i,v}})$ is the family of capacitary Green currents at archimedean places. The vertical components of $\mathscr{D}_1$ and $\mathscr{D}_2$ lying over any finite place v are distinct one from one another, since $\Omega_{1,v} \cap \Omega_{2,v} = \emptyset$. Consequently, the geometric part of the Arakelov intersection product is zero. In view of [11, Lemma 5.1] the contribution of any archimedean place v is zero too, since $\Omega_{1,v}$ and $\Omega_{2,v}$ are disjoint. This concludes the proof. □

The following proposition makes more explicit the numerical effectivity hypothesis in Proposition 7.5.

Proposition 7.7. *Let X, Ω, D, $\mathscr{X}$, $\widehat{D}_\Omega$ be as in the beginning of this subsection.*

(a) If D is effective, then the Arakelov divisor $\widehat{D}_\Omega$ on $\mathscr{X}$, attached to the effective divisor D and to the adelic tube Ω, is effective.
(b) Write $D = \sum_i n_i P_i$, for some closed points P_i of X and positive integers n_i. Then $\widehat{D}_\Omega$ is numerically effective if and only if $h_{\widehat{D}_\Omega}(P_i) \geqslant 0$ for each i.
(c) If D is a rational point P, then $\widehat{D}_\Omega$ is numerically effective (respectively $\widehat{D}_\Omega \cdot \widehat{D}_\Omega > 0$) if and only if the Arakelov degree $\widehat{\deg}(T_P X, \|\cdot\|_\Omega^{\mathrm{cap}})$ is nonnegative (respectively positive).

Proof. (a) Let us assume that D is an effective divisor. For each archimedean place v of K, the capacitary Green function g_{D,Ω_v} is therefore nonnegative [11, 3.1.4]. Moreover, we have proved in Proposition 5.1 that the $\mathbf{Q}$-divisor $\mathscr{D}$ in $Z^1_{\mathbf{Q}}(\mathscr{X})$ is effective. These two facts together imply that $\widehat{D}_\Omega$ is an effective Arakelov divisor.

(b) For any archimedean place v, the definition of the archimedean capacitary Green currents involved in $\widehat{D}_\Omega$ implies that $\omega(\widehat{D}_{\Omega_v})$ is a positive measure on X_v, zero near $|D|$ [11, Theorem 3.1, (iii)]. By [11, Proposition 6.9], in order for $\widehat{D}_\Omega$ to be numerically effective, it is necessary and sufficient that $h_{\widehat{D}_\Omega}(E) \geqslant 0$ for any irreducible component E of $\mathscr{D}$. This holds by construction if E is a vertical component of $\mathscr{X}$: according to the conditions of Proposition 5.1, one has $\mathscr{D} \cdot V = 0$ for any vertical component V of the support of $\mathscr{D}$; for any other vertical component V, one has $\mathscr{D} \cdot V \geqslant 0$ because the divisor $\widehat{D}_\Omega$ is effective. Consequently, $\widehat{D}_\Omega$ is nef if and only if $h_{\widehat{D}_\Omega}(P_i) \geqslant 0$ for all i.

(c) This follows from (b) and from the equality (Corollary 5.20)

$$h_{\widehat{D}_\Omega}(P) = \widehat{D}_\Omega.\widehat{D}_\Omega = \widehat{\deg}\,(T_P X, \|\cdot\|_\Omega^{\mathrm{cap}})\,.$$

□

Theorem 7.8. *Let X be a geometrically connected smooth projective curve over K. Let P be a rational point in $X(K)$, and $\Omega := (\Omega_v)$ an adelic tube adapted to P.*

Let $\varphi \in \widehat{\mathscr{O}_{X,P}}$ be any formal function around P satisfying the following assumptions:

(1) *for any $v \in F$, φ extends to an analytic meromorphic function on Ω_v;*
(2) *φ is algebraic over $\mathscr{O}_{X,P}$;*
(3) $\widehat{\deg}\,(T_P X, \|\cdot\|_\Omega^{\mathrm{cap}}) \geqslant 0$.

If equality holds in the last inequality, assume moreover that there is an archimedean place v of F such that $X_v \setminus \Omega_v$ is not contained in a real semi-algebraic curve of X_v.

Then φ is the formal germ at P of a rational function in $K(X)$.

Proof. Let X' be the normalization of X in the field extension of $K(X)$ generated by φ. This is a geometrically connected smooth projective curve over K, which may be identified with the normalization of the Zariski closure Z in $X \times \mathbf{P}^1_K$ of the graph of φ. It is endowed with a finite morphism $f\colon X' \to X$, namely the composite morphism $X' \to Z \overset{\mathrm{pr}_1}{\to} X$. Moreover, the formal function φ may be identified with the composition of the formal section σ of f at P that lifts the formal section (Id_X, φ) of $Z \overset{\mathrm{pr}_1}{\to} X$ and the rational function $\tilde{\varphi}$ in the local ring $\mathscr{O}_{X',\sigma(P)}$ defined as the composition $X' \to Z \overset{\mathrm{pr}_2}{\to} \mathbf{P}^1_K$.

To show that φ is the germ at P of a rational function, we want to show that f is an isomorphism.

For any place v, Ω_v is a smooth analytic curve in X_v, and σ extends to an analytic section $\sigma_v\colon \Omega_v \to X'_v$ of f. Indeed, according to (1), the formal morphism (Id_X, φ) extends to an analytic section of $Z \overset{\mathrm{pr}_1}{\to} X$ over Ω_v, which in turn lifts to an analytic section of f by normality.

By Corollary 5.20, the Arakelov $\mathbf{Q}$-divisor $\widehat{P}_\Omega$ attached to the point P and the adelic tube Ω is nef. When $\widehat{\deg}\,(T_P X, \|\cdot\|^{\mathrm{cap}}_\Omega)$ is positive, Proposition 7.5 implies that f is an isomorphism; hence φ is the formal germ to a rational function on X. This still holds when $\widehat{\deg}\,(T_P X, \|\cdot\|^{\mathrm{cap}}_\Omega) = 0$, thanks to the supplementary assumption at archimedean places in that case. □

As an example, this theorem applies when X is the projective line, P is the origin, and when, for each place v in F, Ω_v is the disk of center 0 and radius $R_v \in \sqrt{|K_v^*|}$ in the affine line. Then (Ω_v) is an adelic tube adapted to P iff almost every R_v equals 1, and $\widehat{\deg}\,(T_P X, \|\cdot\|^{\mathrm{cap}}_\Omega)$ is nonnegative iff $\prod_v R_v \geqslant 1$. In this special case, Theorem 7.8 becomes Harbater's rationality criterion [30, Proposition 2.1].

Actually Harbater's result is stated without the assumption $R_v \in \sqrt{|K_v^*|}$ on the non-archimedean radii. The reader will easily check that his rationality criterion may be derived in full generality from Theorem 7.8, by shrinking the disks Ω_v for v non-archimedean, and replacing them by larger simply connected domains for v archimedean.

When $\widehat{\deg}\,(T_P X, \|\cdot\|^{\mathrm{cap}}_\Omega) = 0$, some hypothesis on the sets $X_v \setminus \Omega_v$ is really necessary for a rationality criterion to hold. As an example, let us consider the Taylor series of the algebraic function $\varphi(x) = 1/\sqrt{1-4x} - 1$, viewed as a formal function around the origin of the projective line $\mathbf{P}^1_{\mathbf{Q}}$. As shown by the explicit expansion

$$\frac{1}{\sqrt{1-4x}} - 1 = \sum_{n=1}^{\infty} (-4)^n \binom{-1/2}{n} x^n = \sum_{n=1}^{\infty} \binom{2n}{n} x^n,$$

the coefficients of this series are rational integers. Moreover, the complementary subset Ω of the real interval $[1/4, \infty]$ in $\mathbf{P}^1(\mathbf{C})$ is a simply connected open Riemann surface on which the algebraic function has no ramification. Consequently, there is a meromorphic function φ_∞ on Ω such that $\varphi_\infty(x) =$

$(1-4x)^{-1/2}-1$ around 0. One has $\mathrm{cap}_0(\Omega) = 1$, hence $\widehat{\deg}(T_0\mathbf{P}^1, \|\cdot\|^{\mathrm{cap}}_{\Omega})) = 0$. However, φ is obviously not a rational function.

By combining the algebraicity criterion of Theorem 6.1 and the previous corollary, we deduce the following result, a generalization to curves of any genus of Borel–Dwork's criterion.

Theorem 7.9. *Let X be a geometrically connected smooth projective curve over K, P a rational point in $X(K)$, and $\Omega := (\Omega_v)$ an adelic tube adapted to P.*

Let $\varphi \in \widehat{\mathscr{O}_{X,P}}$ be any formal function around P satisfying the following assumptions:

(1) *for any $v \in F$, φ extends to an analytic meromorphic function on Ω_v;*
(2) *the formal graph of φ in $\widehat{X \times \mathbf{A}^1}_{(P,\varphi(P))}$ is A-analytic.*

If, moreover, $\widehat{\deg}\,(T_PX, \|\cdot\|^{\mathrm{cap}}_{\Omega}) > 0$, then φ is the formal germ at P of a rational function on X (in other words, φ belongs to $\mathscr{O}_{X,P}$).

Proof. In view of Corollary 7.8, it suffices to prove that φ is algebraic. Let $V = X \times \mathbf{P}^1$ and let $\widehat{C} \subset \widehat{V}_{(P,\varphi(P))}$ be the formal graph of φ. We need to prove that $\widehat{C}$ is algebraic. Indeed, since at each place v of K, the canonical v-adic seminorm on $T_P\widehat{C}$ is smaller than the capacitary one, $\widehat{\deg}\left(T_PX, \|\cdot\|^{\mathrm{can}}_{X,\widehat{C}}\right) \geqslant \widehat{\deg}\,(T_PX, \|\cdot\|^{\mathrm{cap}}_{\Omega}) > 0$. By Theorem 6.1, $\widehat{C}$ is then algebraic, and φ is algebraic over $K(X)$. □

Observe that when condition (1) is satisfied in Theorem 7.9, the A-analyticity condition (2) is implied by the following one:

(2′) *there exist a positive integer N and a smooth model $\mathscr{X}$ of X over* $\mathrm{Spec}\,\mathscr{O}_K[1/N]$ *such that P extends to an integral point $\mathscr{P}$ in $\mathscr{X}(\mathscr{O}_K[1/N])$, and φ extends to a regular formal function on the formal completion $\widehat{\mathscr{X}_{\mathscr{P}}}$.*

This follows from Proposition 3.8, since then the formal graph of φ extends to a smooth formal curve in $\mathscr{X} \times \mathbf{A}^1$ over $\mathrm{Spec}\,\mathscr{O}_K[1/N]$.

Example 7.10. Theorem 7.9 may be applied when X is $\mathbf{P}^1_K$, P is the origin 0 in $\mathbf{A}^1(K) \hookrightarrow \mathbf{P}^1(K)$, and when, for each place v, $\Omega_v \subset F_v$ is a disk of center 0 and positive radius R_v in the affine line, provided these radii are almost all equal to 1 and satisfy $\prod R_v > 1$. In this case, the rationality of any φ in $\widehat{\mathscr{O}_{X,P}} \simeq K[[X]]$ under the assumptions (1) and (2′) is precisely Borel–Dwork's rationality criterion [6, 22].

More generally, the expression of capacitary norms in terms of transfinite diameters and a straightforward approximation argument[7] allows one to recover the criterion of Pólya–Bertrandias ([1, 43]) from our Theorem 7.9 with $X = \mathbf{P}^1_K$.

[7]Using the fact that bounded subsets of $\mathbf{C}_p$ are contained in affinoids (actually, lemniscates) with arbitrarily close transfinite diameters.

Appendix

A Metrics on line bundles

Let K be a field that is complete with respect to the topology defined by a discrete absolute value $|\cdot|$ on K. Let R be its valuation ring and let π be a uniformizing element of R. We denote by $v = \log|\cdot|/\log|\pi|$ the corresponding normalized valuation on K.

Let X be an algebraic variety over K and let L be a line bundle on X. In this appendix, we set out some basic facts concerning the definition of a metric on the fibers of L.

Let $\overline{K}$ be an algebraic closure of K; endow it with the unique absolute value that extends the given one on K. It might not be complete, however, its completion, denoted by C, is a complete field containing $\overline{K}$ as a dense subset on which the absolute value extends uniquely, endowing it with the structure of a complete valued field.

A metric on the fibers of L is the datum, for any $x \in X(C)$, of a norm $\|\cdot\|$ on the one-dimensional C-vector space $L(x)$. Namely, $\|\cdot\|$ is a map $L(x) \to \mathbf{R}_+$ satisfying the following properties:

- $\|s_1 + s_2\| \leqslant \max(\|s_1\|, \|s_2\|)$ for all s_1, $s_2 \in L(x)$;
- $\|as\| = |a|\|s\|$ for all $a \in C$ and $s \in L(x)$;
- $\|s\| = 0$ implies $s = 0$.

We also assume that these norms are stable under the natural action of the Galois group $\mathrm{Gal}(C/K)$, namely that for any $x \in X(C)$, $s \in L(x)$ and $\sigma \in \mathrm{Gal}(C/K)$, $\|\sigma(s)\| = \|s\|$.

We say that a metric is continuous if for any open subset $U \subset X$ (for the Zariski topology) and any section $s \in \Gamma(U, L)$, the function $x \mapsto \|s(x)\|$ on $U(C)$ is continuous. This definition corresponds to the classical notion of a Weil function attached to a Cartier divisor on X and will be sufficient for our purposes; a better one would be to impose that this function extend to a continuous function on the analytic space attached to U by Berkovich [4]; see, e.g., [29] for this point of view.

Assume that X is projective and let $\mathscr{X}$ be any projective and flat R-scheme with generic fiber X, together with a line bundle $\mathscr{L}$ on $\mathscr{X}$ extending L. Let $x \in X(C)$; if C^0 denotes the valuation ring of C, there is a unique morphism $\varepsilon_x \colon \operatorname{Spec} C^0 \to \mathscr{X}$ by which the generic point of $\operatorname{Spec} C^0$ maps to x. Then, $\varepsilon_x^*\mathscr{L}$ is a sub-C^0-module of $L(x)$. For any section $s \in L(x)$, there exists $a \in C^0$ such that $as \in \varepsilon_x^*\mathscr{L}$. Define, for any $s \in L(x)$,

$$\|s\| = \inf\left\{|a|^{-1}, \ as \in \varepsilon_x^*\mathscr{L}, \quad a \in C \setminus \ \{0\}\right\}.$$

This is a continuous metric on the fibers of L, which we call an algebraic metric.

Algebraic metrics are in fact the only metrics that we use in this article, where the language of metrics is just a convenient way of comparing various extensions of X and L over R. In that respect, we make the following two remarks:

(1) Let Y be another projective algebraic variety over K and let $f\colon Y \to X$ be a morphism. Let $(L, \|\cdot\|_L)$ be a metrized line bundle on X. Then the line bundle f^*L on Y admits a metric $\|\cdot\|_{f^*L}$, defined by the formula $\|f^*s(y)\|_{f^*L} = \|s(f(y))\|_L$, where $y \in Y(C)$ and s is a section of L in a neighborhood of $f(y)$. Assume that the metric of L is algebraic, defined by a model $(\mathscr{X}, \mathscr{L})$. Let $\mathscr{Y}$ be any projective flat model of Y over R such that f extends to a morphism $\varphi\colon \mathscr{Y} \to \mathscr{X}$. Then the metric $\|\cdot\|_{f^*L}$ is algebraic, defined by the pair $(\mathscr{Y}, \varphi^*\mathscr{L})$.

(2) Let $\mathscr{X}$ be a projective and flat model of X on R and let $\mathscr{L}$ and $\mathscr{L}'$ be two line bundles on $\mathscr{X}$ that induce the same (algebraic) metric on L. If $\mathscr{X}$ is *normal*, then the identity map $\mathscr{L}_K = \mathscr{L}'_K$ on the generic fiber extends uniquely to an isomorphism $\mathscr{L} \simeq \mathscr{L}'$.

B Background on rigid analytic geometry

The results of this appendix are basic facts of rigid analytic geometry: the first one is a version of the maximum principle, while the second proposition states that the complementary subsets to an affinoid subspace in a rigid analytic space has a canonical structure of a rigid space. They are well known to specialists, but having been unable to find a convenient reference, we decided to write them down here.

Let K be a field, endowed with an ultrametric absolute value for which it is complete.

Proposition B.1. *Let C be a smooth projective connected curve over K, let $f \in K(C)$ be a nonconstant rational function, and let X denote the Weierstrass domain $C(f) = \{x \in X\,;\, |f(x)| \leqslant 1\}$ in X. Then any affinoid function g on X is bounded; moreover, there exists $x \in U$ such that*

$$|g(x)| = \sup_X |g| \quad \textit{and} \quad |f(x)| = 1.$$

The fact that g is bounded and attains its maximum is the classical maximum principle; we just want to ensure that the maximum is attained on the "boundary" of U.

Proof. The analytic map $f\colon C^{\mathrm{an}} \to (\mathbf{P}^1)^{\mathrm{an}}$ induced by f is finite, hence restricts to a finite map $f_X\colon X \to \mathbf{B}$ of rigid analytic spaces, where $\mathbf{B} = \operatorname{Sp} K\langle t\rangle$ is the unit ball. It corresponds to f_X a morphism of affinoid algebras $K\langle t\rangle \hookrightarrow \mathscr{O}(X)$ that makes $\mathscr{O}(X)$ a $K\langle t\rangle$-module of finite type. Let $g \in \mathscr{O}(X)$ be an analytic function. Then g is integral over $K\langle t\rangle$, hence there is a smallest positive integer n, as well as analytic functions $a_i \in K\langle t\rangle$, for $1 \leqslant i \leqslant n$, such that

$$g(x)^n + a_1(f(x))g(x)^{n-1} + \cdots + a_n(f(x)) = 0$$

for any $x \in X$. Then, (see [7, p. 239, Proposition 6.2.2/4])

$$\sup_{x\in X}|g(x)| = \max_{1\leqslant i\leqslant n}|a_i(t)|^{1/i}.$$

The usual proof of the maximum principle on $\mathbf{B}$ shows that there is for each integer $i \in \{1,\ldots,n\}$ a point $t_i \in \mathbf{B}$ satisfying $|t_i| = 1$ and $|a_i(t_i)| = \|a_i\|$. (After having reduced to the case $\|a_i\| = 1$, it suffices to lift any nonzero element of the residue field at which the reduced polynomial $\overline{t_i}$ does not vanish.) Consequently, there is therefore a point $t \in \mathbf{B}$ such that $|t| = 1$ and

$$\max_i|a_i(t)|^{1/i} = \max_i\|a_i\|^{1/i}.$$

Applying Proposition 3.2.1/2, p. 129, of [7] to the polynomial

$$Y^n + a_1(t)Y^{n-1} + \cdots + a_n(t)$$

there is a point $y \in \mathbf{P}^1$ and $|y| = \max_i\|a_i\|^{1/i}$. Since the morphism $K\langle t\rangle[g] \subset \mathscr{O}(X)$ is integral, there is a point $x \in X$ such that $f(x) = t$ and $g(x) = y$. For such a point, one has $|f(x)| = 1$ and $|g|(x) = \|g\|$. □

Proposition B.2. *Let X be a rigid analytic variety over K and let $A \subset X$ be the union of finitely many affinoid subsets.*

Then $X \setminus A$, endowed with the induced G-topology, is a rigid analytic variety.

Proof. By [7, p. 357, Proposition 9.3.1/5], and the remark that follows that proposition, it suffices to prove that $X \setminus A$ is an admissible open subset.

Let (X_i) be an admissible affinoid covering of X; then, for each i, $A_i = A \cap X_i$ is a finite union of affinoid subsets of X_i. Assume that the Proposition holds when X is affinoid; then, each $X_i \setminus A_i$ is an admissible open subset of X_i, hence of X. Then $X \setminus A = \bigcup_i (X_i \setminus A_i)$ is an admissible open subset of X, by the property (G_1) satisfied by the G-topology of rigid analytic varieties.

We thus may assume that X is an affinoid variety. By Gerritzen–Grauert's theorem [7, p. 309, Cor. 7.3.5/3], A is a finite union of rational subdomains $(A_i)_{1\leqslant i\leqslant m}$ in X. For each i, let us consider affinoid functions $(f_{i,1},\ldots,f_{i,n_i},g_i)$ on X generating the unit ideal such that

$$\begin{aligned} A_i &= X\left(\frac{f_{i,1}}{g_i},\ldots,\frac{f_{i,n_i}}{g_i}\right) \\ &= \left\{x \in X\,;\, |f_{i,1}(x)| \leqslant |g_i(x)|,\ldots,|f_{i,n_i}(x)| \leqslant |g_i(x)|\right\}. \end{aligned}$$

We have

$$X \setminus A = \bigcap_{i=1}^m (X \setminus A_i) = \bigcap_{i=1}^m \bigcup_{j=1}^{n_i} \left\{x \in X\,;\, |f_{i,j}(x)| > |g_i(x)|\right\}.$$

Since any finite intersection of admissible open subsets is itself admissible open, it suffices to treat the case $m = 1$, i.e., when A is a rational subdomain $X(f_1, \dots, f_n; g)$ of X, which we now assume.

By assumption, $f_1, \dots, f_n, g$ have no common zero. By the maximum principle [7, p. 307, Lemma 7.3.4/7], there is $\delta \in \sqrt{|K^*|}$ such that for any $x \in X$,

$$\max(|f_1(x)|, \dots, |f_n(x)|, |g(x)|) \geqslant \delta.$$

For any $\alpha \in \sqrt{|K^*|}$ with $\alpha > 1$, and any $j \in \{1, \dots, n\}$, define

$$X_{j,\alpha} = X\left(\delta \frac{1}{f_j}, \alpha^{-1}\frac{g}{f_j}\right) = \left\{x \in X \,;\, \delta \leqslant |f_j(x)|, \quad \alpha|g(x)| \leqslant |f_j(x)|\right\}.$$

This is a rational domain in X. For any $x \in X_{j,\alpha}$, one has $f_j(x) \neq 0$, and $|g(x)| < |f_j(x)|$, hence $x \in X \setminus A$. Conversely, if $x \in X \setminus A$, there exists $j \in \{1, \dots, n\}$ such that $\max(|f_1(x)|, \dots, |f_n(x)|, |g(x)|) = |f_j(x)| > |g(x)|$; it follows that there is $\alpha \in \sqrt{|K^*|}$, $\alpha > 1$, such that $x \in X_{j,\alpha}$. This shows that the affinoid domains $X_{j,\alpha}$ of X, for $1 \leqslant j \leqslant n$ and $\alpha \in \sqrt{|K^*|}$, $\alpha > 1$, form a covering of $X \setminus A$. Let us show that this covering is admissible. Let Y be an affinoid space and let $\varphi\colon Y \to X$ be an affinoid map such that $\varphi(Y) \subset X \setminus A$. By [7, p. 342, Proposition 9.1.4/2], we need to show that the covering $(\varphi^{-1}(X_{j,\alpha}))_{j,\alpha}$ of Y has a (finite) affinoid covering that refines it. For that, it is sufficient to prove that there are real numbers $\alpha_1, \dots, \alpha_n$ in $\sqrt{|K^*|}$, greater than 1, such that $\varphi(Y) \subset \bigcup_{j=1}^n X_{j,\alpha_j}$.

For $j \in \{1, \dots, n\}$, define an affinoid subspace Y_j of Y by

$$Y_j = \left\{y \in Y \,;\, |f_i(\varphi(y))| \leqslant |f_j(\varphi(y))| \text{ for } 1 \leqslant i \leqslant n\right\}.$$

One has $Y = \bigcup_{j=1}^n Y_j$. Fix some $j \in \{1, \dots, n\}$. Since $\varphi(Y_j) \subset X \setminus A$, $|g(x)| < |f_j(x)|$ on Y_j. It follows that $f_j \circ \varphi$ does not vanish on Y_j; hence $g \circ \varphi / f_j \circ \varphi$ is an affinoid function on Y_j such that

$$\left|\frac{g \circ \varphi}{f_j \circ \varphi}(y)\right| < 1$$

for any $y \in Y_j$. By the maximum principle, there is $\alpha_j \in \sqrt{|K^*|}$ such that $\alpha_j > 1$ and $\left|\frac{g\circ\varphi}{f_j\circ\varphi}\right| < \frac{1}{\alpha_j}$ on Y_j. One then has $\varphi(Y) \subset \bigcup_{j=1}^n X_{j,\alpha_j}$, which concludes the proof of the proposition. □

References

1. Y. Amice, *Les nombres p-adiques*, Collection SUP: Le Mathématicien, vol. 14, Presses Universitaires de France, Paris, 1975.
2. Y. André, *G-functions and geometry*, Vieweg, Braunschweig, 1989.

3. L. Bădescu, *Projective geometry and formal geometry*, Mathematics Institute of the Polish Academy of Sciences. Mathematical Monographs (New Series), vol. 65, Birkhäuser Verlag, Basel, 2004.
4. V. G. Berkovich, *Spectral theory and analytic geometry over non-archimedean fields*, Mathematical Surveys and Monographs, vol. 33, American Mathematical Society, Providence, RI, 1990.
5. F. A. Bogomolov, M. L. McQuillan, *Rational curves on foliated varieties*, prépublication M/01/07, I.H.É.S., 2001.
6. É. Borel, *Sur une application d'un théorème de M. Hadamard*, Bulletin des sciences mathématiques **18** (1894), 22–25.
7. S. Bosch, U. Güntzer, R. Remmert, *Non-archimedean analysis*, Grundlehren der Mathematischen Wissenschaften, vol. 261, Springer-Verlag, Berlin, 1984.
8. S. Bosch, W. Lütkebohmert, *Formal and rigid geometry. I. Rigid spaces*, Math. Ann. **295** (1993), no. 2, 291–317.
9. S. Bosch, W. Lütkebohmert, *Formal and rigid geometry. II. Flattening techniques*, Math. Ann. **296** (1993), no. 3, 403–429.
10. S. Bosch, W. Lütkebohmert, M. Raynaud, *Néron models*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 21, Springer-Verlag, 1990.
11. J.-B. Bost, *Potential theory and Lefschetz theorems for arithmetic surfaces*, Ann. Sci. École Norm. Sup. **32** (1999), no. 2, p. 241–312.
12. J.-B. Bost, *Algebraic leaves of algebraic foliations over number fields*, Publ. Math. Inst. Hautes Études Sci. **93** (2001), 161–221.
13. J.-B. Bost, *Germs of analytic varieties in algebraic varieties: canonical metrics and arithmetic algebraization theorems*, (A. Adolphson, F. Baldassarri, P. Berthelot, N. Katz, F. Loeser, eds.), vol. I, Walter de Gruyter GmbH & Co. KG, Berlin, 2004, 371–418.
14. J.-B. Bost, *Evaluation maps, slopes, and algebraicity criteria*, Proceedings of the International Congress of Mathematicians (Madrid 2006) (M. Sanz-Solé, J. Soria, J. L. Varona, J. Verdera, eds.), vol. II, European Mathematical Society, 2007, 371–418.
15. J.-B. Bost, H. Gillet, C. Soulé, *Heights of projective varieties and positive Green forms*, J. Amer. Math. Soc. **7** (1994), 903–1027.
16. D. G. Cantor, *On an extension of the definition of transfinite diameter and some applications*, J. Reine Angew. Math. **316** (1980), 160–207.
17. A. Chambert-Loir, *Théorèmes d'algébricité en géométrie diophantienne (d'après J.-B. Bost, Y. André, D. & G. Chudnovsky)*, Astérisque (2002), no. 282, 175–209, Exp. No. 886.
18. H. Chen, *Positivité en géométrie algébrique et en géométrie d'Arakelov : application à l'algébrisation et à l'étude asymptotique des polygones de Harder-Narasimhan*, Thèse, École polytechnique, 2006.
19. D. V. Chudnovsky, G. V. Chudnovsky, *Applications of Padé approximations to the Grothendieck conjecture on linear differential equations*, Number theory (New York, 1983–84), Lecture Notes in Math., vol. 1135, 1985, 52–100.
20. D. V. Chudnovsky, G. V. Chudnovsky, *Padé approximations and Diophantine geometry*, Proc. Nat. Acad. Sci. U.S.A. **82** (1985), no. 8, 2212–2216.
21. P. Deligne, *Intersections sur les surfaces régulières*, Groupes de monodromie en géométrie algébrique (SGA 7 II), Lecture Notes in Math., vol. 340, Springer-Verlag, 1973, 1–38.

22. B. Dwork, *On the rationality of the zeta function of an algebraic variety*, Amer. J. Math. **82** (1960), 631–648.
23. G. Eisenstein, *Über eine allgemeine Eigenschaft der Reihen-Entwicklungen aller algebraischen Funktionen (1852)*, Mathematische Gesammelte Werke, Band II, Chelsea Publishing Co., New York, 1975, 765–767.
24. A. Franchetta, *Sulle curve riducibili appartenenti ad una superficie algebrica*, Univ. Roma. Ist. Naz. Alta. Mat. Rend. Mat. e Appl. (5) **8** (1949), 378–398.
25. J. Fresnel, M. Matignon, *Sur les espaces analytiques quasi-compacts de dimension 1 sur un corps valué complet ultramétrique*, Ann. Mat. Pura Appl. (4) **145** (1986), 159–210.
26. P. Graftieaux, *Formal groups and the isogeny theorem*, Duke Math. J. **106** (2001), no. 1, 81–121.
27. P. Graftieaux, *Formal subgroups of abelian varieties*, Invent. Math. **145** (2001), no. 1, 1–17.
28. L. Gruson, M. van der Put, *Banach spaces*, Mém. Soc. Math. France (1974), no. 39–40, 55–100.
29. W. Gubler, *Local heights of subvarieties over non-archimedean fields*, J. Reine Angew. Math. **498** (1998), 61–113.
30. D. Harbater, *Galois covers of an arithmetic surface*, Amer. J. Math. **110** (1988), no. 5, 849–885.
31. R. Hartshorne, *Cohomological dimension of algebraic varieties*, Ann. of Math. (2) **88** (1968), 403–450.
32. R. Hartshorne, *Curves with high self-intersection on algebraic surfaces*, Publ. Math. Inst. Hautes Études Sci. (1969), no. 36, 111–125.
33. R. Hartshorne, Ample subvarieties of algebraic varieties, Lecture Notes in Mathematics, vol. 156, Springer-Verlag, Berlin, 1970.
34. H. Hironaka, *On some formal imbeddings*, Illinois J. Math. **12** (1968), 587–602.
35. H. Hironaka, H. Matsumura, *Formal functions and formal embeddings*, J. Math. Soc. Japan **20** (1968), 52–82.
36. Y. Ihara, *Horizontal divisors on arithmetic surfaces associated with Belyĭ uniformizations*, The Grothendieck theory of dessins d'enfants (Luminy 1993) (L. Schneps, ed.), London Math. Soc. Lecture Note Ser., vol. 200, Cambridge Univ. Press, Cambridge, 1994, 245–254.
37. E. Kani, *Potential theory on curves*, Théorie des nombres (Quebec, PQ, 1987), de Gruyter, Berlin, 1989, 475–543.
38. Yu. I. Manin, *New dimensions in geometry*, Workshop Bonn 1984 (Bonn, 1984), Lecture Notes in Math., vol. 1111, Springer, Berlin, 1985, 59–101.
39. H. Matsumura, *Commutative Algebra*, Mathematics Lecture Notes Series, Benjamin/Cummings, 1980.
40. L. Moret-Bailly, *Groupes de Picard et problèmes de Skolem. I.*, Ann. Sci. École Norm. Sup. **22** (1989), no. 2, 161–179.
41. L. Moret-Bailly, *Un problème de descente*, Bull. Soc. Math. France **124** (1996), 559–585.
42. D. Mumford, *The topology of normal singularities of an algebraic surface and a criterion for simplicity*, Publ. Math. Inst. Hautes Études Sci. **9** (1961), 5–22.
43. G. Pólya, *Über gewisse notwendige Determinantenkriterien für die Fortsetzbarkeit einer Potenzreihe*, Math. Ann. **99** (1928), 687–706.
44. M. van der Put, *The class group of a one-dimensional affinoid space*, Ann. Inst. Fourier (Grenoble) **30** (1980), no. 4, 155–164.

45. C. P. Ramanujam, *Remarks on the Kodaira vanishing theorem*, J. Indian Math. Soc. (N.S.) **36** (1972), 41–51.
46. H. Randriambololona, *Métriques de sous-quotient et théorème de Hilbert–Samuel arithmétique pour les faisceaux cohérents*, J. Reine Angew. Math. **590** (2006), 67–88.
47. M. Raynaud, *Revêtements de la droite affine en caractéristique $p > 0$ et conjecture d'Abhyankar*, Invent. Math. **116** (1994), 425–462.
48. R. Rumely, C. F. Lau, R. Varley, *Existence of the sectional capacity*, Mem. Amer. Math. Soc. **145** (2000), no. 690, 1–130.
49. R. S. Rumely, *Capacity theory on algebraic curves*, Lecture Notes in Math., vol. 1378, Springer-Verlag, Berlin, 1989.
50. J.-P. Serre, Lectures on the Mordell–Weil theorem, third ed., Aspects of Mathematics, Friedr. Vieweg & Sohn, Braunschweig, 1997.
51. A. Thuillier, *Théorie du potentiel sur les courbes en géométrie non archimédienne. Applications à la théorie d'Arakelov*, Thèse, Université de Rennes 1, 2005. http://tel.archives-ouvertes.fr/tel-00010990/

Riemann–Roch for Real Varieties

Paul Bressler,[1] Mikhail Kapranov,[2] Boris Tsygan,[3] and Eric Vasserot[4]

[1] I.A.S., Princeton bressler@math.ias.edu
[2] Yale University mikhail.kapranov@yale.edu
[3] Northwestern University tsygan@math.northwestern.edu
[4] Université Paris 7 vasserot@math.jussieu.fr

To Yuri Ivanovich Manin on his 70th birthday.

Summary. We prove a Riemann–Roch type result for any smooth family of smooth oriented compact manifolds. It describes the class of the conjectural higher determinantal gerbe associated to the fibers of the family.

Key words: Riemann–Roch, determinantal gerbe, Lie algebroid, cyclic homology.

2000 Mathematics Subject Classifications: 57R20 58J (Primary); 17B65 (Secondary)

1 Introduction

1.1

Let Σ be an oriented real analytic manifold of dimension d and let X be a complex envelope of Σ, i.e., a complex manifold of the same dimension containing Σ as a totally real submanifold. Then, (real) geometric objects on Σ can be viewed as (complex) geometric objects on X involving cohomology classes of degree d. For example, a C^∞-function f on Σ can be considered as a section of $\mathcal{B}_\Sigma$, the sheaf of hyperfunctions on Σ, which, according to Sato, can be defined as

$$\mathcal{B}_\Sigma = \underline{H}^d_\Sigma(\mathcal{O}_X \otimes \mathrm{or}_{\Sigma/X}) \ , \tag{1}$$

where or Σ/X is the relative orientation sheaf. So f can be viewed as a class in dth local cohomology.

More generally, the equality (1) suggests that various results of holomorphic geometry on X should have consequences for the purely real geometry on Σ, consequences that involve raising the cohomological degree by d. The goal of this paper is to investigate the consequences of one such result, the Grothendieck–Riemann–Roch theorem (GRR).

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*, 125
Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_4,

1.2

Let $p : X \to B$ be a smooth proper morphism of complex algebraic manifolds. We denote the fibers of p by $X_b = p^{-1}(b)$ and assume them to be of dimension d. If $\mathcal{E}$ is an algebraic vector bundle on X, the GRR theorem says that

$$ch_m(Rp_*(\mathcal{E})) = \int_{X/B} \Big[ch(\mathcal{E}) \cdot \mathrm{Td}(\mathcal{T}_{X/B}) \Big]_{2m+2d} \quad \in \quad H^{2m}(B, \mathbb{C}). \tag{2}$$

Here $\int_{X/B} : H^{2m+2d}(X, \mathbb{C}) \to H^{2m}(B, \mathbb{C})$ is the cohomological direct image (integration over the fibers of p).

In the case $m = 1$, the class on the left comes from the class, in the Picard group of B, of the determinantal line bundle $\det(Rp_*\mathcal{E})$ whose fiber, at a generic point $b \in B$, is

$$\det H^\bullet(X_b, \mathcal{E}) \quad = \quad \bigotimes_i \big(\Lambda^{max} H^i(X_b, \mathcal{E})\big)^{\otimes (-1)^i}. \tag{3}$$

Deligne [9] posed the problem of describing $\det(Rp_*\mathcal{E})$ in a functorial way as a refinement of GRR for $m = 1$. This problem makes sense already for the case $B = pt$ when we have to describe the 1-dimensional vector space (3) as a functor of $\mathcal{E}$. Deligne solved this problem for a family of curves, and further results have been obtained in [11].

1.3

To understand the real counterpart of (2), assume first that $B = pt$, so $X = X_{pt}$ and let $\Sigma \subset X$ be as in Section 1.1. Denote by E the restriction of $\mathcal{E}$ to Σ and by $C^\infty_\Sigma(E)$ the sheaf of its C^∞ sections. Then, similarly to (1), we have the embedding

$$C^\infty_\Sigma(E) \subset \underline{H}^d_\Sigma(\mathcal{E} \otimes \Omega^d_X).$$

Assume further that $d = 1$, so X is an algebraic curve, and that Σ is a small circle in X cutting it into two pieces: X_+ (a small disk) and X_-. Let $\mathcal{E}_\pm = \mathcal{E}|_{X_\pm}$. We are then in the situation of the Krichever correspondence [26]. Namely, the space $\Gamma(E)$ of L^2-sections has a canonical polarization in the sense of Pressley and Segal [26] and therefore possesses a determinantal gerbe $\mathrm{Det}\,\Gamma(E)$. The latter is a category with every Hom-set made into a $\mathbb{C}^*$-torsor (a 1-dimensional vector space with zero deleted). The extensions $\mathcal{E}_\pm$ of E to $X_\pm$ define two objects $[\mathcal{E}_\pm]$ of this gerbe, and

$$\det H^\bullet(X, \mathcal{E}) = \mathrm{Hom}_{\mathrm{Det}\,\Gamma(E)}([\mathcal{E}_+], [\mathcal{E}_-]).$$

The real counterpart of the problem of describing the $\mathbb{C}^*$-torsor $\det H^\bullet(X, \mathcal{E})$ is the problem of describing the gerbe $\mathrm{Det}\,\Gamma(E)$. If we now have a family $p : X \to B$ as before (with $d = 1$), equipped with a subfamily of circles

$q : \Sigma \to B$, $\Sigma \subset X$, then we have an $\mathcal{O}_B^*$-gerbe $\operatorname{Det} q_*(E)$, which, according to the the classification of gerbes [5], has a class in $H^2(B, \mathcal{O}_B^*)$. The latter group maps naturally to $H^3(B, \mathbb{Z})$ and in fact can be identified with the Deligne cohomology group $H^3(B, \mathbb{Z}_D(1))$, see [6]. The Real Riemann–Roch for a circle fibration describes the above class (modulo 2-torsion) as

$$[\operatorname{Det} q_*(E)] = \int_{\Sigma/B} ch_2(E) \quad \in \quad H^3(B, \mathbb{Z}_{\mathcal{D}}(2)) \otimes \mathbb{Z}\left[\frac{1}{2}\right]. \tag{4}$$

Here $\int_{\Sigma/B} : H^4(\Sigma, \mathbb{Z}_D(2)) \to H^3(B, \mathbb{Z}_D(1))$ is the direct image in Deligne cohomology. Note the absence of the characteristic classes of $\mathcal{T}_{\Sigma/B}$ (they are 2-torsion for a real rank one bundle). If one is interested in the image of the determinantal class in $H^3(B, \mathbb{Z})$, then one can understand the RHS of the above formula in the purely topological sense.

Both sides of (4) do not involve anything other than $q : \Sigma \to B$ and a vector bundle E on Σ (equipped with CR-structures coming from the embeddings into $X, \mathcal{E}$). One has a similar result for any C^∞ circle fibration (no CR structure) and any C^∞ complex bundle E on Σ. In this case we get a gerbe with lien $C_B^{\infty *}$, the sheaf of invertible complex-valued C^∞-functions on B, and its class lies in $H^2(B, C_B^{\infty *}) = H^3(B, \mathbb{Z})$. It is this purely C^∞ setting that we adopt and generalize in the present paper.

1.4

Let Σ be a compact oriented C^∞-manifold of arbitrary dimension d and E a C^∞ complex vector bundle on Σ. One expects that the space $\Gamma(E)$ should have some kind of d-fold polarization, giving rise to a "determinantal d-gerbe", $\operatorname{Det} \Gamma(E)$. This structure is rather clear when Σ is a 2-torus, but in general, the theory of higher gerbes is not fully developed. In any case, one expects a C^∞ family of such gerbes over a base B to give a class in $H^{d+1}(B, C_B^{\infty *}) = H^{d+2}(B, \mathbb{Z})$. In this paper we consider a C^∞ family $q : \Sigma \to B$ of relative dimension d and a C^∞ bundle E on Σ. We then define by means of the Chern–Weil approach what should be the characteristic class of the would-be d-gerbe $\operatorname{Det}(q_*(E))$:

$$C_1(q_*(E)) \quad \in \quad H^{d+2}(B, \mathbb{C}).$$

We denote it by C_1, since it is a kind of d-fold delooping of the usual first Chern (determinantal) class. We then show the compatibility of this class with the gerbe approach whenever the latter can be carried out rigorously. Our main result is the Real Riemann–Roch theorem (RRR):

$$C_1(q_* E) = \int_{\Sigma/B} \Big[ch(E) \cdot \operatorname{Td}(\mathcal{T}_{\Sigma/B}) \Big]_{2d+2} \quad \in \quad H^{d+2}(B, \mathbb{C}).$$

Here, $\mathcal{T}_{\Sigma/B}$ is the complexified relative tangent bundle, and $\int_{\Sigma/B}$, the integration along the fibers of q, lowers the degree by d.

Note that the above theorem is a statement of purely real geometry and is quite different from the “Riemann–Roch theorem for differentiable manifolds” proved by Atiyah and Hirzebruch [1]. The latter expresses properties of a Dirac operator on a real manifold Σ, while our RRR deals with the $\overline{\partial}$-operator on a complex envelope X of Σ. The $d = 1$ case above can be deduced from a result of Lott [22] on “higher” index forms for Dirac operators (because the polarization in the circle case can be described in terms of the signs of eigenvalues of the Dirac operator). In general, however, our results proceed in a different direction.

1.5

Our definition of $C_1(q_*E)$ uses the description of the cyclic homology of differential operators [7] [29], which provides a construction of a natural Lie algebra cohomology class γ of the Atiyah algebra, i.e., of the Lie algebra of infinitesimal automorphisms of a pair (Σ, E), where Σ is a compact oriented d-dimensional C^∞-manifold and E is a vector bundle on Σ. The intuition with higher gerbes suggests that this class comes in fact from a group cohomology class of the infinite-dimensional group of all the automorphisms of (Σ, E), see Proposition 40, and moreover, that there are similar classes coming from the higher Chern classes in formula (39). This provides a new point of view on the rather classical subject of “cocycles on gauge groups and Lie algebras”, i.e., on groups of diffeomorphisms of manifolds and automorphisms of vector bundles as well as their Lie algebra analogues.

There have been two sources of interest in this subject. The first one was the study of the cohomology of the Lie algebras of vector fields following the work of Gelfand–Fuks; see [13] for a systematic account. In particular, Bott [3] produced a series of cohomology classes of the Lie algebra of vector fields on a compact manifold and integrated them to group cohomology classes of the group of diffeomorphisms. Later, group cocycles were studied with connections with various anomalies in physics, see [27].

From our point of view, the approach of [27] can be seen as producing “integrals of products of Chern classes” in families over a base B, (cf. [9] [11]), in other words, as producing the ingredients for the right-hand side of a group-theoretical RRR. This is the same approach that leads to the construction of the Morita–Miller characteristic classes for surface fibrations [24]. The anomalies themselves, however, should be seen as the hypothetical classes from Conjectures 39, 41 and whose description through integrals of products of Chern classes constitutes the RRR.

1.6

As far as the proof of the RRR goes, we use two techniques. The first is that of differential graded Lie algebroids (which can be seen as infinitesimal analogues of higher groupoids appearing in the heuristic discussion above).

The second technique is that of "formal geometry" of Gelfand and Kazhdan, i.e., reduction of global problems in geometry of manifolds and vector bundles to problems related to cohomology of Lie algebras of formal vector fields and currents. The first work relating Riemann–Roch to Lie algebra cohomology was [12], and this approach was further developed in [4]. To prove the RRR we use results of [25] and [4] on the Lie algebra cohomology of formal Atiyah algebras.

1.7

The second author would like to acknowledge support from NSF, Université Paris-7, and Max-Planck Institut für Mathematik. The fourth author is partially supported by the NSF.

2 Background on Lie algebroids, groupoids and gerbes

2.1 Conventions

All manifolds will be understood to be C^∞ unless otherwise specified. For a manifold Σ we denote by C^∞_Σ the sheaf of $\mathbb{C}$-valued C^∞-functions. By a vector bundle over Σ we mean a locally trivial, C^∞ complex vector bundle, possibly infinite-dimensional. For such a bundle E we denote by $C^\infty(E) = C^\infty_\Sigma(E)$ the sheaf of smooth sections, which is a locally free sheaf of C^∞_Σ-modules. By $\mathcal{T}_\Sigma$ we denote the *complexified* tangent bundle of Σ, so its sections are derivations of C^∞_Σ. We denote by $\mathcal{D}_\Sigma$ the sheaf of differential operators acting on C^∞_Σ, and by $\mathcal{D}_{\Sigma,E}$ the sheaf of differential operators acting from sections of E to sections of E. The notations $\mathcal{D}(\Sigma)$ and $\mathcal{D}(\Sigma, E)$ will be used for the spaces of global sections of $\mathcal{D}_\Sigma$ and $\mathcal{D}_{\Sigma,E}$.

2.2 Lie algebroids

Recall [23] that a Lie algebroid on Σ consists of a vector bundle $\mathcal{G}$, a morphism of vector bundles $\alpha : \mathcal{G} \to \mathcal{T}_\Sigma$ (the anchor map), and a Lie algebra structure in $C^\infty(\mathcal{G})$ satisfying the following properties:

1. α takes the Lie bracket on sections of $\mathcal{G}$ to the standard Lie bracket on vector fields.
2. For any smooth function f on Σ and sections x, y of $\mathcal{G}$ we have

$$[fx, y] - f \cdot [x, y] = \mathrm{Lie}_{\alpha(y)}(f) \cdot x.$$

A Lie algebroid is called transitive if α is surjective.

Example 1. When $\Sigma = \mathrm{pt}$, a Lie algebroid is the same as a Lie algebra.

Example 2. $\mathcal{T}_X$ with the standard Lie bracket and $\alpha = \mathrm{id}$ is a Lie algebroid.

Example 3. If $\alpha = 0$, then the bracket in $\mathcal{G}$ is C^∞_Σ-linear. In this case we say that $\mathcal{G}$ is a bundle of Lie algebras: every fiber of $\mathcal{G}$ is a Lie algebra.

Morphisms of Lie algebroids are defined in an obvious way. Note that for any transitive Lie algebroid $\mathcal{G}$ the kernel $\mathrm{Ker}(\alpha) \subset \mathcal{G}$ is a bundle of Lie algebras, i.e., a Lie algebroid with trivial anchor map, and the maps in the short exact sequence

$$0 \to \mathrm{Ker}(\alpha) \to \mathcal{G} \xrightarrow{\alpha} \mathcal{T}_X \to 0$$

are morphisms of Lie algebroids.

2.3 The de Rham complex of a Lie algebroid

Let $\mathcal{G}$ be a Lie algebroid on Σ. Let

$$\mathrm{DR}^i(\mathcal{G}) := \mathrm{Hom}(\Lambda^i\mathcal{G}, C^\infty_\Sigma).$$

The differential $\mathrm{d} : \mathrm{DR}^i(\mathcal{G}) \to \mathrm{DR}^{i+1}(\mathcal{G})$ is defined by the standard formula of Cartan: for an antisymmetric i-linear function $l : \mathcal{G}^i \to C^\infty_\Sigma$ we set

$$\begin{aligned} \mathrm{d}l(x_1, \ldots, x_{i+1}) = &\sum_{j=1}^{i+1} (-1)^j \mathrm{Lie}_{\alpha(x_j)} l(x_1, \ldots, \widehat{x_j}, \ldots, x_{i+1}) \\ &+ \sum_{j<k} (-1)^{j+k} l([x_j, x_k], x_1, \ldots, \widehat{x_j}, \ldots, \widehat{x_k}, \ldots, x_{i+1}). \end{aligned} \tag{5}$$

We get a complex $\mathrm{DR}^\bullet(\mathcal{G})$ called the de Rham complex of $\mathcal{G}$. A morphism of Lie algebroids $\phi : \mathcal{G} \to \mathcal{H}$ gives rise to the morphism of de Rham complexes $\phi^* : \mathrm{DR}^\bullet(\mathcal{H}) \to \mathrm{DR}^\bullet(\mathcal{G})$.

Example 4. If $\Sigma = \mathrm{pt}$, so $\mathcal{G}$ is a Lie algebra, then $\mathrm{DR}^\bullet(\mathcal{G}) = C^\bullet(\mathcal{G})$ is the cochain complex of $\mathcal{G}$ with trivial coefficients.

Example 5. If $\mathcal{G} = \mathcal{T}_\Sigma$, then $\mathrm{DR}^\bullet(\mathcal{G}) = \Omega^\bullet_\Sigma$ is the C^∞ de Rham complex of Σ.

2.4 The enveloping algebra of a Lie algebroid

Let $\mathcal{G}$ be a Lie algebroid on Σ, as before. The enveloping algebra $U(\mathcal{G})$ is the sheaf of associative algebras on Σ defined by generators $x \in \mathcal{G}$ (local sections) and $f \in C^\infty_\Sigma$ (local functions) subject to the relations

$$xy - yx = [x, y] \ ,$$
$$f \cdot x - x \cdot f = \mathrm{Lie}_{\alpha(x)}(f).$$

Example 6. If $\Sigma = pt$, so $\mathcal{G}$ is a Lie algebra, then $U(\mathcal{G})$ is the usual enveloping algebra of $\mathcal{G}$.

Example 7. If $\mathcal{G} = \mathcal{T}_\Sigma$, then $U(\mathcal{G}) = \mathcal{D}_\Sigma$ is the sheaf of differential operators $C^\infty_\Sigma \to C^\infty_\Sigma$.

Example 8. If $\mathcal{G}$ is any Lie algebroid, then the anchor map α induces a morphism

$$U(\alpha) : U(\mathcal{G}) \to U(\mathcal{T}_\Sigma) = \mathcal{D}_\Sigma$$

of sheaves of associative algebras. In particular, C^∞_Σ is a left $U(\mathcal{G})$-module.

The sheaf $U(\mathcal{G})$ has an increasing ring filtration $\{U^m(\mathcal{G})\}$ with $U^m(\mathcal{G})$ generated by products involving at most m sections of $\mathcal{G}$. The following is then standard.

Proposition 9. *The associated graded sheaf of algebras* $\operatorname{gr} U(\mathcal{G})$ *is identified with the symmetric algebra* $S^\bullet(\mathcal{G})$*.*

2.5 The Koszul resolution

Let $\mathcal{G}$ be a Lie algebroid on Σ. We have then the complex

$$\cdots \to U(\mathcal{G}) \otimes \Lambda^2\mathcal{G} \to U(\mathcal{G}) \otimes \mathcal{G} \to U(\mathcal{G}) \to C^\infty_\Sigma \to 0. \tag{6}$$

with the differential defined by:

$$\begin{aligned} \mathrm{d}(u \otimes (\gamma_1 \wedge \cdots \wedge \gamma_n)) = &\sum_{j=1}^{n} (-1)^j (u\gamma_j) \otimes \big(\gamma_1 \wedge \cdots \wedge \widehat{\gamma_j} \wedge \cdots \wedge \gamma_n\big) \\ &+ \sum_{j<k} (-1)^{j+k} u \otimes \big([\gamma_i, \gamma_j] \wedge \cdots \wedge \widehat{\gamma_i} \wedge \cdots \wedge \widehat{\gamma_j} \wedge \cdots \wedge \gamma_n\big). \end{aligned}$$

Proposition 10. *The complex (6) is exact. Thus, it provides a locally free resolution of* C^∞_Σ *as a* $U(\mathcal{G})$*-module.*

Corollary 11. *We have*

$$\mathrm{DR}^\bullet(\mathcal{G}) \simeq R\underline{\mathrm{Hom}}_{U(\mathcal{G})}(C^\infty_\Sigma, C^\infty_\Sigma).$$

2.6 The Atiyah algebra

Let G be a Lie group, $\mathfrak{g}$ its Lie algebra, and $\rho : P \to \Sigma$ a principal G-bundle on Σ. The Atiyah algebra $\mathcal{A}_P$ is the sheaf of Lie algebras on Σ whose sections are G-invariant vector fields on P:

$$\mathcal{A}_P = (\rho_* \mathcal{T}_P)^G.$$

The map $\alpha = d\rho$ makes $\mathcal{A}_P$ into a transitive Lie algebroid of the form

$$0 \longrightarrow \mathrm{Ad}(P) \longrightarrow \mathcal{A}_P \xrightarrow{\alpha} \mathcal{T}_\Sigma \longrightarrow 0. \tag{7}$$

Here, $\mathrm{Ad}(P)$ is the bundle of Lie algebras on Σ associated to P via the adjoint representation.

If $\Sigma = \bigcup U_i$ is a covering in which P is trivialized, $P|_{U_i} = U_i \times G$, and $g_{ij} : U_i \cap U_j \to \mathrm{Aut}(\mathfrak{g})$ are the transition functions, then $\mathcal{A}_P$ is glued out of $\mathcal{A}_P|_{U_i} = \mathcal{T}_{U_i} \times \mathfrak{g}$ via the transition functions

$$(v, x) \mapsto (v, i_v(dg_{ij} \cdot g_{ij}^{-1}) + \mathrm{Ad}_{g_{ij}}(x)). \tag{8}$$

Example 12. Let $G = \mathrm{GL}_r(\mathbb{C})$, so $\mathfrak{g} = \mathfrak{gl}_r(\mathbb{C})$. A principal G-bundle P corresponds then to a rank r vector bundle E on Σ. In this case, $\mathcal{A}_P$ will also be denoted by $\mathcal{A}_E$, and it has a well-known alternative description. It consists of differential operators $L : E \to E$ such that:

1. L has order ≤ 1.
2. The first order symbol of L (which is a priori a section of $\mathcal{T}_\Sigma \otimes \mathrm{End}(E)$) lies in the subsheaf $\mathcal{T}_\Sigma = \mathcal{T}_\Sigma \otimes 1$.

2.7 Modules over Lie algebroids

We follow [17, §3] but use a more geometric language. Let $\mathcal{G}$ be a Lie algebroid on Σ. A $\mathcal{G}$-module is a vector bundle $\mathcal{M}$ on Σ equipped with a Lie algebra action $(x, m) \mapsto xm$ of $\mathcal{G}$ on the sections that satisfies

1. the Leibniz rule

$$x(f \cdot m) - f \cdot (xm) = (\mathrm{Lie}_{\alpha(x)} f) \cdot m, \quad f \in C^\infty_\Sigma, x \in \mathcal{G}, m \in \mathcal{M} \ ;$$

   in particular, the assignment $x \mapsto (m \mapsto x \cdot m)$ defines a map $\mathcal{G} \to \mathcal{A}_\mathcal{M}$ that commutes with respective anchor maps
2. the map $\mathcal{G} \to \mathcal{A}_\mathcal{M}$ is C^∞_Σ-linear.

Example 13. For any $\mathcal{G}$ the trivial bundle (whose sheaf of sections is) C^∞_Σ is a $\mathcal{G}$-module with the $\mathcal{G}$ action given via the anchor map and the Lie derivations of functions.

Example 14. An ideal in $\mathcal{G}$ is a sub-Lie algebroid $\mathcal{G}'$ such that $[\mathcal{G}, \mathcal{G}'] \subset \mathcal{G}'$. Suppose that $\mathcal{G}'$ is an ideal in $\mathcal{G}$ such that the restriction of the anchor map to $\mathcal{G}'$ is trivial. Then, $\mathcal{G}'$ is a $\mathcal{G}$-module via the adjoint action.

Any $\mathcal{G}$-module has a structure of a sheaf of modules over the sheaf of rings $U(\mathcal{G})$.

2.8 Cohomology of Lie algebroids

Let $\mathcal{M}$ be a $\mathcal{G}$-module. The de Rham complex $\mathrm{DR}^\bullet(\mathcal{G},\mathcal{M})$ with coefficients in $\mathcal{M}$ is defined by

$$\mathrm{DR}^i(\mathcal{G},\mathcal{M}) = \underline{\mathrm{Hom}}(\Lambda^i\mathcal{G},\mathcal{M})$$

with the differential of $l : \mathcal{G}^i \to \mathcal{M}$ defined by the following modification of (5):

$$\begin{aligned}\mathrm{d}l(x_1,\dots,x_{i+1}) = &\sum_{j=1}^{i+1}(-1)^j x_j(l(x_1,\dots,\widehat{x_j},\dots,x_{i+1})) \\ &+\sum_{j<k}(-1)^{j+k}l([x_j,x_k],x_1,\dots,\widehat{x_j},\dots,\widehat{x_k},\dots,x_{i+1}).\end{aligned}$$

Its cohomology sheaves will be denoted by $\underline{H}^i_{\mathrm{Lie}}(\mathcal{G},\mathcal{M})$ and the corresponding cohomology groups of the complex of global smooth sections of $\mathrm{DR}^\bullet(\mathcal{G},\mathcal{M})$ by $H^i_{\mathrm{Lie}}(\mathcal{G},\mathcal{M})$. See [23, §7.1]. As before, it is easy to see that

$$\mathrm{DR}^\bullet(\mathcal{G},\mathcal{M}) \simeq \underline{R\mathrm{Hom}}_{U(\mathcal{G})}(C^\infty_\Sigma,\mathcal{M}).$$

Therefore,

$$\underline{H^i}_{\mathrm{Lie}}(\mathcal{G},\mathcal{M}) = \underline{\mathrm{Ext}}^i_{U(\mathcal{G})}(C^\infty_\Sigma,\mathcal{M}), \quad H^i_{\mathrm{Lie}}(\mathcal{G},\mathcal{M}) = \mathrm{Ext}^i_{U(\mathcal{G})}(C^\infty_\Sigma,\mathcal{M}).$$

Example 15. The trivial bundle C^∞_Σ is always a $\mathcal{G}$-module, and for $\mathcal{G} = \mathcal{T}_\Sigma$ we have $H^i_{\mathrm{Lie}}(\mathcal{T}_\Sigma, C^\infty_\Sigma) = H^i(\Sigma,\mathbb{C})$.

2.9 The Hochschild–Serre spectral sequence and the transgression

Let

$$0 \to \mathcal{G}' \to \mathcal{G} \to \mathcal{G}'' \to 0 \tag{9}$$

be an extension of Lie algebroids on Σ, so $\mathcal{G}'$ is an ideal with zero anchor in $\mathcal{G}$. Note that $\mathcal{G}'$ is then a bundle of Lie algebras. Let $\mathcal{M}$ be a $\mathcal{G}$-module. Then for every point $x \in \Sigma$ the fiber $\mathcal{M}_x$ is a module over the Lie algebra $\mathcal{G}'_x$. Assume that for any $i \geq 0$ the Lie algebra cohomology spaces $H^i_{\mathrm{Lie}}(\mathcal{G}'_x,\mu_x)$ have finite dimension which is independent of i. Then the sheaves $\underline{H}^i_{\mathrm{Lie}}(\mathcal{G}',\mathcal{M})$ are vector bundles on Σ with fiber $H^i_{\mathrm{Lie}}(\mathcal{G}'_x,\mathcal{M}_x)$ at $x \in \Sigma$. These vector bundles have natural structures of $\mathcal{G}''$-modules. In this case we have (a Lie algebroid generalization of) the Hochshild–Serre spectral sequence with

$$E_2^{pq} = H^p_{\mathrm{Lie}}(\mathcal{G}'', \underline{H}^q_{\mathrm{Lie}}(\mathcal{G}',\mathcal{M})) \Rightarrow H^{p+q}_{\mathrm{Lie}}(\mathcal{G},\mathcal{M}). \tag{10}$$

The construction is parallel to the classical (Lie algebra) case as in [13]. One uses the short exact sequence (9) to produce, in a standard way, a filtration on $\mathrm{DR}^\bullet(\mathcal{G},\mathcal{M})$. See [23, Section 7.4] for the treatment of the case $\mathcal{G}'' = \mathcal{T}_\Sigma$, which is the only case we will use in this paper.

Example 16. Similarly to the classical case, one can use (10) (or elementary considerations) to identify $H^2_{\mathrm{Lie}}(\mathcal{G},\mathcal{M})$ with the set of isomorphism classes of central extensions of Lie algebroids

$$0 \to \mathcal{M} \to \widetilde{\mathcal{G}} \to \mathcal{G} \to 0.$$

Central extensions of this type with $\mathcal{G} = \mathcal{T}_\Sigma$, $\mathcal{M} = C^\infty_\Sigma$, and the $\mathcal{G}$-action on $\mathcal{M}$ being the standard one (by Lie derivations), were called in [17] Picard Lie algebroids. The set of their isomorphism classes is thus identified with $H^2_{\mathrm{Lie}}(\mathcal{T}_\Sigma, C^\infty_\Sigma)$, which is the same as the topological (de Rham) cohomology $H^2(\Sigma,\mathbb{C})$.

Fix $n > 0$ and assume that

$$H^j(\mathcal{G}',\mathcal{M}) = 0, \quad 0 < j < n. \tag{11}$$

In this case $E_2^{0,n} = E_{n+1}^{0,n}$ as well as $E_2^{0,n+1} = E_{n+1}^{0,n+1}$. We obtain therefore the *transgression map*

$$d_{n+1} : E_{n+1}^{0,n} = E_2^{0,n} = H^n_{\mathrm{Lie}}(\mathcal{G}',\mathcal{M})^{\mathcal{G}''} \longrightarrow \\ H^{n+1}_{\mathrm{Lie}}(\mathcal{G}'',\mathcal{M}^{\mathcal{G}'}) = E_2^{n+1,0} = E_{n+1}^{n+1,0}. \tag{12}$$

We will use this map later in the paper. Without the assumption (11) we have that $E_{n+1}^{0,n}$ is a subspace of $E_2^{0,n} = H^n_{\mathrm{Lie}}(\mathcal{G}',\mathcal{M})^{\mathcal{G}''}$, namely the intersection of the kernels of $d_2,\dots,d_n$. For convenience we will call elements of this space *transgressive* elements of $E_2^{0,n}$. Similarly, $E_{n+1}^{n+1,0}$ is a quotient space of $E_2^{n+1,0} = H^{n+1}_{\mathrm{Lie}}(\mathcal{G}'',\mathcal{M}^{\mathcal{G}'})$ by the union of images of $d_2,\dots,d_n$.

Example 17. Suppose that $n = 2$ and $\Sigma = \mathrm{pt}$, so (9) is a central extension of Lie algebras and $\mathcal{M}$ is a $\mathcal{G}$-module in the usual sense. Let $\gamma \in E_2^{0,2} = H^2_{\mathrm{Lie}}(\mathcal{G}',\mathcal{M})^{\mathcal{G}''}$ be a $\mathcal{G}''$-invariant class in H^2 and let

$$0 \to \mathcal{M} \to \widetilde{\mathcal{G}}' \to \mathcal{G}' \to 0$$

be a central extension representing γ. The class γ is transgressive (i.e., annihilated by d_2) if and only if $\widetilde{\mathcal{G}}'$ can be made into a $\mathcal{G}$-equivariant central extension (as opposed to the fact that the class of the extension remains unchanged under the $\mathcal{G}$-action or, what is the same, under $\mathcal{G}''$-action). Given such an equivariant extension, one obtains a crossed module of Lie algebras (i.e., a dg-Lie algebra situated in degrees (-1) and 0)

$$\widetilde{\mathcal{G}}'' \xrightarrow{\partial} \mathcal{G},$$

with $\mathrm{Ker}(\partial) = \mathcal{M}$ and $\mathrm{Coker}(\partial) = \mathcal{G}''$. As is well known (see, e.g., [21, Example E.10.3]), such a crossed module represents an element in $H^3(\mathcal{G}'',\mathcal{M})$, and this element is the lifting of $d_3(\beta)$. Different choices of equivariant structure on $\widetilde{\mathcal{G}}'$ correspond to the ambiguity of the values of d_3 modulo the image of d_2. One can generalize this picture easily to the case of an arbitrary Σ.

2.10 Reminder on gerbes

We follow the same conventions as in [19] and use [5] as the background reference.

If B is a topological space and $\mathcal{F}$ is a sheaf of abelian groups on B, then we can speak of $\mathcal{F}$-gerbes (= gerbes with *band* $\mathcal{F}$). Recall that such a gerbe $\mathfrak{G}$ is the following:

1. A category $\mathfrak{G}(U)$ given for all open $U \subset B$, the restriction functors $r_{UV} : \mathfrak{G}(U) \to \mathfrak{G}(V)$ given for any morphism $V \subset U$ and natural isomorphisms of functors $s_{UVW} : r_{VW} \circ r_{UV} \Rightarrow r_{UW}$ given for each $W \subset V \subset U$ and satisfying the transitivity conditions.
2. The structure of $\mathcal{F}|_U$-torsor (possibly empty) on each sheaf $\underline{\mathrm{Hom}}_{\mathfrak{G}(U)}(x, y)$ compatible with the r_{UV} and such that the composition of morphisms is bi-additive.

These data have to satisfy the local uniqueness and gluing properties, for which we refer to [5].

By a sheaf of $\mathcal{F}$-groupoids we will mean a sheaf of categories $\mathfrak{C}$ on B (so both $\mathrm{Ob}\,\mathfrak{C}$ and $\mathrm{Mor}\,\mathfrak{C}$ are sheaves of sets) in which each sheaf $\underline{\mathrm{Hom}}_{\mathfrak{C}(U)}(x, y)$ is either empty or is made into a sheaf of $\mathcal{F}|_U$-torsors so that the composition is bi-additive. A sheaf $\mathfrak{C}$ of $\mathcal{F}$-groupoids is called locally connected if locally on B all the $\mathrm{Ob}\,\mathfrak{C}(U)$ and $\mathrm{Hom}_{\mathfrak{C}(U)}(x, y)$ are nonempty.

Each sheaf of $\mathcal{F}$-groupoids can be seen as a fibered category over B; in fact, it is a prestack, see, e.g., [20]. Recall (see, e.g., [20] Lemma 2.2) that for any prestack $\mathfrak{C}$ there is an associated stack $\mathfrak{C}^{\sim}$. If $\mathfrak{C}$ is a locally connected sheaf of $\mathcal{F}$-groupoids, then $\mathfrak{C}^{\sim}$ is an $\mathcal{F}$-gerbe.

As is well known (see, e.g., [5]), the set formed by $\mathcal{F}$-gerbes up to equivalence is identified with $H^2(B, \mathcal{F})$. The identification of the set of isomorphism classes of Picard Lie algebroids in Example 16 can be seen as an infinitesimal analogue of this fact. Given an $\mathcal{F}$-gerbe $\mathfrak{G}$, we denote by $[\mathfrak{G}] \in H^2(B, \mathcal{F})$ its class. Given a sheaf $\mathfrak{C}$ of $\mathcal{F}$-groupoids, we denote by $[\mathfrak{C}]$ the class of the corresponding gerbe.

Let B be a C^∞-manifold. We will be particularly interested in $C_B^{\infty *}$-gerbes on B. Recall that we have the following exponential sequence of sheaves on B:

$$0 \to \underline{\mathbb{Z}}_B \to C_B^\infty \xrightarrow{e^{2\pi i x}} C_B^{\infty *} \to 0. \tag{13}$$

The corresponding coboundary map

$$\delta_n : H^n(B, C_B^{\infty *}) \to H^{n+1}(B, \mathbb{Z}) \tag{14}$$

is an isomoprhism for $n \geq 1$, since C_B^∞ is a soft sheaf. Thus $[\mathfrak{G}]$ gives rise to a class in $H^3(B, \mathbb{Z})$.

Let $\mathfrak{G}$ be a $C_B^{\infty *}$-gerbe. Recall [6] that a *connective structure* Δ on $\mathfrak{G}$ is a set of data that associates to each open $U \subset B$ and each object $x \in \mathrm{Ob}\,\mathfrak{G}(U)$ an Ω_U^1-torsor $\Delta(x)$ (whose sections can be thought of as “formal connections” on x) and for any local (iso)morphism $g : x \to y$ over U an identification of torsors $g_* : \Delta(x) \to \Delta(y)$, satisfying the compatibility property plus the following gauge condition: if $x = y$ so $g \in C^{\infty *}(U)$ is an invertible function, then $g_*(\nabla) = \nabla - g^{-1}d(g)$.

A *curving* of a connective structure Δ is a rule K associating to any x as above and any global object $\nabla \in \Delta(x)$ a 2-form $K(\nabla) \in \Omega^2(U)$ satisfying the compatibility with pullbacks, invariance under isomorphisms, as well as the gauge condition $K(\nabla + \alpha) = K(\nabla) + d\alpha$, $\alpha \in \Omega^1(U)$. In this situation J.-L. Brylinski defined the 3-curvature of the connective structure and curving, which is a closed 3-form $S = S_{\Delta,K} \in \Omega^3(B)$.

Example 18. Let G be a Lie group and

$$1 \to \mathbb{C}^* \to \widetilde{G} \to G \to 1$$

a central extension of Lie groups. Let $\rho : P \to B$ be a principal G-bundle. We then have the $C_B^{\infty *}$-gerbe $\mathrm{Lift}_G^{\widetilde{G}}(P)$, whose objects over $U \subset B$ are liftings of $P|_U$ to a principal $\widetilde{G}$-bundle over U; compare [2]. Let ∇_P be a connection on P. Then $\mathrm{Lift}_G^{\widetilde{G}}(P)$ has a connective structure Δ that to every lifting $\widetilde{P}$ of P to a $\widetilde{G}$-bundle associates the space of all connections on $\widetilde{P}$ extending ∇_P. Further, let $R_\nabla \in \Omega^2(B) \otimes \mathrm{Ad}(P)$ be the curvature of ∇. A choice of a lifting of R_∇ to a form $\widetilde{R}_\nabla \in \Omega^2(B) \otimes \mathrm{Ad}(\widetilde{P})$ gives a curving K on Δ. This curving associates to any section $\widetilde{\nabla}$ of $\Delta(\widetilde{P})$, i.e., to a connection on $\widetilde{P}$ extending ∇, the 2-form $R_{\widetilde{\nabla}} - \widetilde{R}_\nabla$, where $R_{\widetilde{\nabla}}$ is the curvature of $\widetilde{\nabla}$.

We will need the following result [6, Thm. 5.3.12].

Theorem 19. *If $\mathfrak{G}$ is a $C_B^{\infty *}$-gerbe with a connective structure Δ and a curving K, then the class of $S_{\Delta,K}$ in $H^3(B, \mathbb{C})$ is integral and is equal to the image of $\delta_2[\mathfrak{G}]$ under the natural map from $H^3(B, \mathbb{Z})$ to $H^3(B, \mathbb{C})$.*

3 Background on homology of differential operators

3.1 Conventions

Let A be an associative algebra over $\mathbb{C}$. We denote by $\mathrm{Hoch}_\bullet(A)$ the Hochschild complex of A with coefficients in A:

$$\cdots \to A \otimes A \otimes A \to A \otimes A \to A$$

with the differential given by the formula

$$b(a_0 \otimes \cdots \otimes a_p) = \sum_{i=0}^{p-1} (-1)^i a_0 \otimes \cdots \otimes a_i a_{i+1} \otimes \cdots \otimes a_p + (-1)^p a_p a_0 \otimes a_1 \otimes \cdots \otimes a_{p-1}.$$

By $HH_\bullet(A)$ we denote the homology of $\mathrm{Hoch}_\bullet(A)$. As is well known,

$$HH_\bullet(A) = \mathrm{Tor}_\bullet^{A\otimes A^{op}}(A, A). \tag{15}$$

Put

$$\tau(a_0 \otimes \cdots \otimes a_p) = (-1)^p a_1 \otimes \cdots \otimes a_p \otimes a_0.$$

Let $N = 1+\tau+\tau^2+\cdots+\tau^n$ on $\mathrm{Hoch}_n(A)$. The cyclic complex of A is defined as the total complex

$$\begin{aligned} &CC_\bullet(A) \\ &\quad = \mathrm{Tot}_\bullet\left\{\cdots \to \mathrm{Hoch}_\bullet(A) \overset{1-\tau}{\to} \mathrm{Hoch}_\bullet(A) \overset{N}{\to} \mathrm{Hoch}_\bullet(A) \overset{1-\tau}{\to} \mathrm{Hoch}_\bullet(A)\right\}. \end{aligned} \tag{16}$$

The cyclic homology $HC_\bullet(A)$ is the homology of the complex $CC_\bullet(A)$. We recall the theorem relating the cyclic homology with the Lie algebra homology of the algebra of matrices; see [21].

Theorem 20. $H_\bullet^{\mathrm{Lie}}(\mathfrak{gl}(A)) = S^\bullet(HC_{\bullet-1}(A))$

Corollary 21. *If $HC_j(A) = 0$ for $j = 0, \ldots, p-1$, then $H_j^{\mathrm{Lie}}(\mathfrak{gl}(A)) = 0$ for $j = 1, \ldots, p$, and $H_{p+1}^{\mathrm{Lie}}(\mathfrak{gl}(A)) = HC_p(A)$.*

3.2 Homology of differential operators: algebro-geometric version

Let X be a smooth affine algebraic variety over $\mathbb{C}$ of dimension d, and let $\mathcal{E}$ be an algebraic vector bundle on X. Then the Hochschild–Kostant–Rosenberg theorem (together with Morita invariance of $HH_\bullet$) gives an identification

$$HH_m(\mathrm{End}(\mathcal{E})) = \Omega^m(X),$$

where on the right we have the space of global regular m-forms on X. Furthermore,

$$HC_m(\mathrm{End}(\mathcal{E})) = \Omega^m(X)/d\Omega^{m-1}(X) \oplus H^{m-2}(X, \mathbb{C}) \oplus H^{m-4}(X, \mathbb{C}) \oplus \cdots,$$

where on the right we have the usual topological (de Rham) cohomology; see [21, Th. 3.4.12]. Let $\mathcal{D}(\mathcal{E})$ be the ring of global differential operators from $\mathcal{E}$ to $\mathcal{E}$. Then the results of [7], [29] imply

$$HH_m(\mathcal{D}(\mathcal{E})) = H^{2d-m}(X, \mathbb{C}).$$

Furthermore,

$$HC_m(\mathcal{D}(\mathcal{E})) = \bigoplus_{i\geq 0} H^{2d-m+2i}(X, \mathbb{C}).$$

We recall that the approach of [7], [29] is to use the filtration by the order of differential operators and realize the E_1-term of the corresponding spectral sequence for HH as the complex of forms on the cotangent bundle with the differential adjoint to the de Rham differential by means of the symplectic form. The spectral sequence is then seen to degenerate at E_2.

Let us note the particular case in which $X = \mathbb{A}^d$ and $E = \mathcal{O}_{\mathbb{A}^d}$ is the trivial bundle of rank 1. Then $\mathcal{D}(\mathcal{E}) = W_d$ is the Weyl algebra with generators x_i, ∂_i, $i = 1, \ldots, d$, and relations

$$[x_i, x_j] = [\partial_i, \partial_j] = 0, \quad [\partial_i, x_j] = \delta_{ij} \cdot 1.$$

The above results imply that

$$HH_i(W_d) = 0 \quad \text{if} \quad i \neq 2d, \qquad HH_{2d}(W_d) = \mathbb{C}, \tag{17}$$

and

$$HC_i(W_d) = \mathbb{C},\, i - 2d \in 2\mathbb{Z}_{\geq 0}, \quad HC_i(W_d) = 0,\, i - 2d \notin 2\mathbb{Z}_{\geq 0}.$$

3.3 The C^∞ version

Let Σ be an oriented C^∞-manifold of dimension d and let E be a smooth complex vector bundle on Σ. We have then the algebras $\mathrm{End}(E)$, $\mathcal{D}(E)$ of smooth endomorphisms and differential operators on E. Following [29] we present the analogues of the results cited in Section 3.2 for these algebras. These rings have natural Fréchet topologies. As pointed out in [29], to get reasonable results, all tensor products occurring in the Hochschild and cyclic complexes of the above algebras should be taken in the category of topological vector spaces, i.e., be completed. In plain terms, this means that $\mathrm{End}(E)^{\otimes p}$ should be understood as the ring of endomorphisms of the vector bundle $E^{\boxtimes p}$ on the p-fold Cartesian product Σ^p and similarly for differential operators. Under these conventions, we have

$$HH_m(\mathcal{D}(E)) = H^{2d-m}(\Sigma, \mathbb{C}), \tag{18}$$

$$HC_m(\mathcal{D}(E)) = \bigoplus_{i \geq 0} H^{2d-m+2i}(\Sigma, \mathbb{C}), \tag{19}$$

where on the right we have the topological cohomology.

Remark 22. The Lie algebra cochain complexes of $\mathcal{D}(E)$ and of $\mathfrak{gl}_N \mathcal{D}(E) = \mathcal{D}(E \otimes \mathbb{C}^r)$ involve exterior products of these algebras over $\mathbb{C}$. If we understand these products in the completed sense as above (compare also with Fuks [13]), then the analogue of Theorem 20 holds, and we have the following.

Corollary 23. *Let Σ be a compact, oriented C^∞ manifold of dimension d. Then, for $N \gg 0$ we have*

$$H_i^{\mathrm{Lie}} \mathfrak{gl}_N \mathcal{D}(E) = 0, \quad 0 < i < d+1,$$
$$H_{d+1}^{\mathrm{Lie}} \mathfrak{gl}_N \mathcal{D}(E) = \mathbb{C}.$$

3.4 The formal series version

Let

$$\widehat{W}_d = W_d \otimes_{\mathbb{C}[x_1,\dots,x_d]} \mathbb{C}[[x_1,\dots,x_d]]$$

be the algebra of differential operators whose coefficients are formal power series. Similarly to the above, we consider the Hochschild and cyclic complexes of $\widehat{W}_d$ using the adic topology on $\mathbb{C}[[x_1,\dots,x_d]]$ and taking completions. Thus $\widehat{W}_d^{\otimes p}$ is understood as the ring of differential operators whose coefficients are power series in p groups of d variables. With this understanding, we have the following analogue of (17):

$$HH_{2d}(\widehat{W}_d) = \mathbb{C}, \quad HH_i(\widehat{W}_d) = 0, \quad i \neq 2d.$$

For the proof, see [12]. One can also apply the spectral sequence argument of [7] and [29] and then use the Poincaré lemma on the (contangent bundle to the) formal disk.

Our next step is to consider such formal completions simultaneously at all points of a given C^∞-manifold Σ. So, let Σ, E be as above. Let $\widehat{\mathrm{Hoch}}_p(\mathcal{D}(E))$ be the completion of $\mathcal{D}(E^{\boxtimes(p+1)})$ (differential operators in the bundle $E^{\boxtimes(p+1)}$ on Σ^{p+1}) along the diagonal $\Sigma \subset \Sigma^{p+1}$. This is a sheaf on Σ.

Then the Hochschild differential extends to $\widehat{\mathrm{Hoch}}_\bullet(\mathcal{D}(E))$, making it into a complex, and we denote by $\widehat{HH}_\bullet(\mathcal{D}(E))$ its homology. Similarly, we define the completed cyclic complex $\widehat{CC}_\bullet(\mathcal{D}(E))$ by the procedure identical to (16) and denote its homology by $\widehat{HC}_\bullet(\mathcal{D}(E))$. Thus, $\widehat{HH}_\bullet(\mathcal{D}(E))$ and $\widehat{CC}_\bullet(\mathcal{D}(E))$ are sheaves on Σ.

Proposition 24. *We have $\widehat{HH}_p(\mathcal{D}(E)) = \underline{\mathbb{C}}_\Sigma$ (constant sheaf) for $p = 2d$ and $\widehat{HH}_p(\mathcal{D}(E)) = 0$ for $p \neq 2d$.*

Proof. Consider the case in which Σ is an open contractible domain in $\mathbb{R}^d$ and E is trivial. Let us prove that in this case the complex of global sections of $\widehat{HH}_\bullet(\mathcal{D}(E))$ is exact everywhere except degree $2d$, where the cohomology is isomorphic to $\mathbb{C}$. (This is the standard Hochschild–Kostant–Rosenberg theorem in the context of completed Hochschild complexes).

We start with the case of $\widehat{\mathrm{Hoch}}_\bullet(C^\infty_\Sigma)$ defined, as before, using the completion of the functions on $\Sigma^{\bullet+1}$ along the diagonal. Recall the interpretation of HH as Tor, see (15). Assume for a moment that Σ is the affine space viewed as an affine algebraic variety. Choose the standard Koszul resolution of $\mathbb{C}[\Sigma]$ over $\mathbb{C}[\Sigma \times \Sigma]$. We see that

$$HH_\bullet(\mathbb{C}[\Sigma]) = \Omega^\bullet(\Sigma),$$

and the same will hold if we replace $\mathbb{C}[\Sigma]$ by a matrix algebra (i.e., take E of higher rank).

Now let us get back to the C^∞ case. There is a small difference, namely that we are using completed tensor products, and therefore the standard argument of comparing two projective resolutions is not quite applicable. But if we follow this standard argument in the algebraic case, we see that it gives the embedding of complexes $i : \Omega^\bullet(\Sigma) \to \mathrm{Hoch}_\bullet(\mathbb{C}[\Sigma])$, a projection $j : \mathrm{Hoch}_\bullet(\mathbb{C}[\Sigma]) \to \Omega^\bullet(\Sigma)$, and a homotopy $s : \mathrm{Hoch}_\bullet(\mathbb{C}[\Sigma]) \to \mathrm{Hoch}_{\bullet+1}(\mathbb{C}[\Sigma])$ such that $ji = 1$, $ij - 1 = sd + ds$. It is easy to see that the maps i, j, and s extend from $\mathbb{C}[\Sigma]$ to $C^\infty(\Sigma)$ and from the algebraic Hochschild complex to the completed one. We conclude that

$$\widehat{HH}_\bullet(C^\infty(\Sigma)) = \Omega^\bullet(\Sigma),$$

and the same will hold if we replace $\mathbb{C}[\Sigma]$ by a matrix algebra.

Next, we replace C^∞_Σ by the sheaf of commutative algebras

$$\mathcal{A} = S^\bullet(\mathcal{T}_\Sigma)$$

(polynomial functions on the cotangent bundle) and define $\widehat{\mathrm{Hoch}}_\bullet(\mathcal{A})$ using the completions of sheaves of sections of $\mathcal{A}^{\boxtimes(p+1)}$ on Σ^{p+1} along the diagonals. The same argument will apply, so we conclude that

$$\widehat{HH}_\bullet(\mathcal{A}) = p_*(\Omega^\bullet_{T^*\Sigma}), \tag{20}$$

where $p : T^*\Sigma \to \Sigma$ is the projection. Again, a similar statement will hold for matrices.

Finally, we use the approach of [7], [29] and consider the spectral sequence for $\widehat{HH}_\bullet(\mathcal{D}(E))$ associated to the filtration by degree of operators. We get the E_1-term to be (20) with the differential being the adjoint of the de Rham differential on $T^*\Sigma$. Since we assumed Σ to be a contractible domain in the flat space, we conclude that the E_2-term reduces to one space $\mathbb{C}$. Moreover, we see that the class of the cycle

$$1 \otimes \mathrm{Alt}_{S_{2d}}(\partial_{x_1} \otimes \cdots \otimes \partial_{x_d} \otimes x_1 \otimes \cdots \otimes x_d)$$

is a generator of $\widehat{HH}_{2d}(\mathcal{D}(E))$. We will call it the canonical generator. Note also that the above argument works not only for the ring of algebraic or smooth (or holomorphic) differential operators but also for formal differential operators, i.e., differential operators whose coefficients are formal power series.

Now consider a diffeomorphism from one contractible domain in the flat space to another. It induces an isomorphism of the rings of differential operators. It is enough to show that this isomorphism sends the canonical generator to the canonical generator. Take a point of Σ. We have seen that the homomorphism that associates to a function its jet at this point induces an isomorphism on the Hochschild homology. Furthermore, any shift in the affine space sends the canonical generator to itself. We are reduced to proving that any formal coordinate change induces an automorphism of the ring of formal differential operators that sends the canonical generator to itself. Since a reflection preserves the canonical generator, we may assume that our formal coordinate change is oriented. Therefore it may be included into a one-parameter group of formal coordinate changes. We are reduced to proving that if X is a formal vector field then the corresponding derivation of the ring of formal differential operators is trivial on the Hochschild homology. But such a derivation is inner, and any inner derivation acts on the Hochschild homology trivially (the operator ι_X from (41) is a contracting homotopy).

More generally, any change of the trivialization of the vector bundle E induces an automorphism of $\widehat{HH}^\bullet(\mathcal{D}_(E))$ that sends the fundamental generator to itself.

We have proven that the only sheaf of cohomology of $\widehat{HH}_\bullet(\mathcal{D}(E))$ in the case that Σ is a contractible domain in a flat space (and thus in the general case) is $\underline{\mathbb{C}}_\Sigma$. □

Furthermore, we need a relative version of the above statements. Let

$$q : \Sigma \to B$$

be a submersion (smooth fibration) of C^∞-manifolds, whose fibers are of dimension d and are oriented. Let E be a C^∞-bundle on Σ, as above. We then have the subring

$$\mathcal{D}_{\Sigma/B}(E) \subset \mathcal{D}(E),$$

consisting of differential operators that are $q^{-1}C^\infty_B$-linear, i.e., act along the fibers only.

Let $\Sigma^{p+1}_B \subset \Sigma^{p+1}$ be the $(p+1)$-fold fiber product of Σ over B. We denote by $E_B^{\boxtimes(p+1)}$ the restriction of $E^{\boxtimes(p+1)}$ to Σ^{p+1}_B.

Let $\widehat{\mathrm{Hoch}}_p(\mathcal{D}_{\Sigma/B}(E))$ denote the completion of $\mathcal{D}_{\Sigma^{p+1}_B/B}(E_B^{\boxtimes(p+1)})$ along the diagonal. Then the Hochschild differential extends to $\widehat{\mathrm{Hoch}}_p(\mathcal{D}_{\Sigma/B}(E))$. We also define the completed cyclic complex $\widehat{CC}_\bullet(\mathcal{D}_{\Sigma/B}(E))$ by implementing (16).

Theorem 25.

1. *The complex $\widehat{\mathrm{Hoch}}_p(\mathcal{D}_{\Sigma/B}(E))$ is acyclic in degrees other than $2d$, and its $2d$-th cohomology sheaf is isomorphic to $q^{-1}C^\infty_B$. In other words, we have an isomorphism in the derived category of sheaves of $q^{-1}C^\infty_B$-modules on Σ:*

$$\mu_{\mathcal{D}} : \widehat{\mathrm{Hoch}}_p(\mathcal{D}_{\Sigma/B}(E)) \to q^{-1}C^\infty_B[2d].$$

2. *We have $H^i(\widehat{CC}_\bullet(\mathcal{D}_{\Sigma/B}(E)) = 0$ unless $i = -2d + k$, $k \in \mathbb{Z}_+$, and*

$$H^{-2d+k}(\widehat{CC}_\bullet(\mathcal{D}_{\Sigma/B}(E)) = q^{-1}C^\infty_B.$$

Proof. Similar to that of Proposition 24. □

Corollary 26. *We have a morphism (no longer an isomorphism) in the derived category*

$$\nu_{\mathcal{D}} : \widehat{CC}_\bullet(\mathcal{D}_{\Sigma/B}(E)) \to q^{-1}C^\infty_B[2d].$$

4 Characteristic classes from Lie algebra cohomology

4.1 The finite-dimensional case

Let G be a Lie group with Lie algebra $\mathfrak{g}$. We denote by $C^\bullet(\mathfrak{g})$ the cochain complex of $\mathfrak{g}$ with trivial coefficients $\mathbb{C}$ and by $H^n(\mathfrak{g})$ its nth cohomology space.

Let $\gamma \in H^n(\mathfrak{g})$ be a cohomology class. We want to associate (under certain conditions) to γ a characteristic class of principal G-bundles. In other words, we want to produce, for each C^∞-manifold B and each smooth principal G-bundle P on B, a topological (de Rham) cohomology class

$$c_\gamma(P) \in H^{n+1}(B) = H^{n+1}(B, \mathbb{C})$$

(note the shift of degree by 1).

Indeed, let a principal G-bundle $\rho : P \to B$ be given and let $\mathcal{A}_P$ be its Atiyah algebra. We then have the extension of Lie algebroids (7) on B and the corresponding Hochschild–Serre spectral sequence (10), which in our case has the form

$$E_2^{pq} = H^p_{\mathrm{Lie}}(\mathcal{T}_B, \underline{H}^q_{\mathrm{Lie}}(\mathrm{Ad}(P), C^\infty_B)) \Rightarrow H^{p+q}_{\mathrm{Lie}}(\mathcal{A}_P, C^\infty_B). \tag{21}$$

This sequence was considered in [23, Thm. 7.4.19]. Note that $\underline{H}^q_{\mathrm{Lie}}(\mathrm{Ad}(P), C^\infty_B)$ is the cohomology of the cochain complex of $\mathrm{Ad}(P)$ as a Lie algebra over C^∞_B, i.e., of the complex of bundles formed by the duals of the fiberwise exterior products of fibers of $\mathrm{Ad}(P)$. We will also use the notation $C^\bullet(\mathrm{Ad}(P)_{/B})$ for this complex.

Lemma 27. *For any $q \geq 0$ the bundle $H^q_{\mathrm{Lie}}(\mathrm{Ad}(P), C^\infty_B) = H^q(\mathrm{Ad}(P)_{/B})$ on B formed by the Lie algebra cohomology spaces of the fibers of* $\mathrm{Ad}(P)$ *is canonically identified with the trivial bundle with fiber $H^q(\mathfrak{g})$.*

Proof. This follows from the fact adjoint action of G on $\mathfrak{g}$ induces the trivial action on $H^q(\mathfrak{g})$. □

Corollary 28. *The E_2-term of the spectral sequence (21) is given by $E_2^{pq} = H^p(B) \otimes H^q(\mathfrak{g})$. In particular, the assignment $\gamma \mapsto 1 \otimes \gamma$ defines a map $H^n(\mathfrak{g}) \to E_2^{0n}$.*

Assume now that there exists $n > 0$ such that the Lie algebra $\mathfrak{g}$ satisfies the acyclicity condition

$$H^i(\mathfrak{g}) = 0, \quad 0 < i < n. \tag{22}$$

Then we are in the situation of (11), so we have the transgression map (12), which in our case has the form

$$d_{n+1} : H^n(\mathfrak{g}) \to H^{n+1}(B), \tag{23}$$

and we define

$$c_\gamma(P) = d_{n+1}(1 \otimes \gamma). \tag{24}$$

Without the assumption (22) we have that $c_\gamma(P)$ is defined only if $1 \otimes \gamma$ is transgressive (i.e., annihilated by $d_2, \dots, d_n$ and takes values not in $H^{n+1}(B)$ but in the quotient of $H^{n+1}(B)$ by the images of $d_2, \dots, d_n$).

If the latter is true for a cohomology class γ, we say that γ is *transgressive.*

Example 29. Let $n = 1$. Then the condition (22) is trivially satisfied. A class γ is just a trace functional $\gamma : \mathfrak{g} \to \mathbb{C}$. The class $c_\gamma(P) \in H^2(B)$ can be obtained by choosing a connection ∇ in P with curvature $R \in \Omega_B^2 \otimes \mathfrak{g}$ and taking the class of the closed 2-form $\gamma(R) \in \Omega_B^2$. Alternatively, one can use γ to produce a trace functional $\gamma_P : \mathrm{Ad}(P) \to C_B^\infty$ and then use γ_P to push forward the extension (7) to a central extension of Lie algebroids

$$0 \to C_B^\infty \to \mathcal{G} \to \mathcal{T}_B \to 0.$$

As is well known (see Section 2.7) the set of isomorphism classes of such central extensions is identified with $H^2_{\mathrm{Lie}}(\mathcal{T}_B, C_B^\infty) = H^2(B, \mathbb{C})$.

Example 30. Let $n = 2$, so γ is represented by a central extension

$$0 \to \mathbb{C} \to \widetilde{\mathfrak{g}} \to \mathfrak{g} \to 0. \tag{25}$$

A sufficient condition for γ to be basic for any P is that $\widetilde{\mathfrak{g}}$ can be made into a G-equivariant central extension; compare Example 17. Suppose that such an equivariant structure has been chosen. Then the class $c_\gamma(P) \in H^3(B, \mathbb{C})$ can be constructed as follows. We have the representation $\widetilde{\mathrm{Ad}}$ of G on $\widetilde{\mathfrak{g}}$, and therefore an extension of associated vector bundles on B:

$$0 \to C_B^\infty \to \widetilde{\mathrm{Ad}}(P) \to \mathrm{Ad}(P) \to 0.$$

Choose a connection ∇ in P. Then we have associated linear connections ∇_{Ad} in $\mathrm{Ad}(P)$ and $\nabla_{\widetilde{\mathrm{Ad}}}$ in $\widetilde{\mathrm{Ad}}(P)$. We also have the curvature $R_\nabla \in \Omega^2(B) \otimes \mathrm{Ad}(P)$. Choose a lifting $\widetilde{R_\nabla}$ of R_∇ to $\Omega^2(B) \otimes \widetilde{\mathrm{Ad}}(P)$, and take

$$S = \nabla_{\widetilde{\mathrm{Ad}}}(\widetilde{R}_\nabla) \in \Omega^3(B) \otimes \widetilde{\mathrm{Ad}}(P).$$

By the Bianchi identity, $\nabla(R_\nabla) = 0$, and so S lies in the tensor product of $\Omega^3(B)$ and the subbundle $C_B^\infty \subset \widetilde{\mathrm{Ad}}(P)$), i.e., it is a scalar differential form $S \in \Omega^3(B)$. Furthermore, it is clear that S is a closed 3-form. The class $c_\gamma(P)$ is then the class of the form S. A different choice of an equivariant structure on $\widetilde{\mathfrak{g}}$ leads to a change of the class of S by an element from the image of d_2.

Example 31. Let $G = \mathrm{GL}_N(\mathbb{C})$, so $\mathfrak{g} = \mathfrak{gl}_N(\mathbb{C})$. Then $H^\bullet(\mathfrak{g})$ is the exterior algebra on generators $\gamma_1, \dots, \gamma_N$ with $\gamma_i \in H^{2i-1}(\mathfrak{g})$. A principal G-bundle P on B is the same as a rank-N vector bundle E. In this case, each $1 \otimes \gamma_i$ is transgressive, and $c_{\gamma_i}(P)$ is the image of $c_i(E) \in H^{2i}(B)$ under the natural projection $H^{2i}(B) \to E_{n+1}^{0,n+1}$. Here $c_i(E)$ is the usual ith Chern class of E.

4.2 Other interpretations

Here we collect, for future use, some more or less straightforward reformulations of the construction of $c_\gamma(P)$.

4.2.1 The Chern–Weil picture

If we choose a connection ∇ in P, then the sequence (7) splits (such splitting is in fact the definition of a connection following Atiyah). So we can identify

$$\Omega^\bullet(P)^G = \mathrm{DR}^\bullet(\mathcal{A}_P) = \Omega^\bullet_B \otimes C^\bullet(\mathrm{Ad}(P)_{/B}). \tag{26}$$

Let R be the curvature of ∇. Then the differential in the RHS of (26) has the form $\partial + \nabla + i_R$, where ∂ is the differential in $C^\bullet(\mathfrak{g})$ and

$$i_R : \Omega^\bullet_B \otimes C^\bullet(\mathfrak{g}) \to \Omega^{\bullet+2}_B \otimes C^{\bullet-1}(\mathfrak{g})$$

is the contraction with R. This leads to a definition of $c_\gamma(P)$ in terms of differential forms. Namely, we have an injective map of complexes followed by a surjective one:

$$\Omega^\bullet_B = \Omega^\bullet_B \otimes C^0(\mathfrak{g}) \overset{\phi}{\hookrightarrow} \Omega^\bullet_B \otimes C^\bullet(\mathfrak{g}) \xrightarrow{\psi} \Omega^0_B \otimes C^\bullet(\mathfrak{g}).$$

Here, ψ is identified with the projection to gr^0_F, where F is the filtration from (21). If our class γ is basic, then it lifts uniquely to a class in $H^n(\mathrm{Coker}(\phi))$, so $c_\gamma(P)$ is the image of that lifted class under the coboundary map corresponding to the short exact sequence

$$0 \to \Omega^\bullet_B \overset{\phi}{\hookrightarrow} \Omega^\bullet_B \otimes C^\bullet(\mathfrak{g}) \to \mathrm{Coker}(\phi) \to 0.$$

4.2.2 The differential graded picture

Let $\mathfrak{A}$ denote the cone of the map $i : \mathrm{Ad}(P) \to \mathcal{A}_P$ viewed as a differential graded Lie algebroid. Thus $\mathcal{A}_P$ is put in degree 0, and $\mathrm{Ad}(P)$ in degree (-1). The anchor map α induces the quasi-isomorphism of Lie algebroids $\mathfrak{A} \to \mathcal{T}_B$, hence the map of respective universal enveloping (differential graded) algebras $U(\mathfrak{A}) \to U(\mathcal{T}_B) = \mathcal{D}_B$ (the latter concentrated in degree zero), which is a quasi-isomorphism. Define the map

$$\mathrm{DR}^\bullet(\mathcal{A}_P)/\mathrm{DR}^\bullet(\mathcal{T}_B) \xrightarrow{\delta} \mathrm{DR}^{\bullet+1}(\mathfrak{A})$$

as follows. For $X \in \mathrm{Ad}(P)$, denote by $\underline{X}$ the element $(X, 0)$ in the cone $\mathfrak{A}$ of i; for $Y \in \mathcal{A}_P$, denote the element $(0, Y)$ simply by Y. Given a p-cochain ω from $\mathrm{DR}^\bullet(\mathcal{A}_P)$, define the cochain $\delta\omega$ by

$$\delta\omega(\underline{X}_1, \ldots, \underline{X}_q, Y_1, \ldots, Y_r) = \omega(\underline{X}_1, Y_1, \ldots, Y_r)$$

for $q = 1$ and zero for $q \neq 1$.

It is easy to see that the sequence

$$\mathrm{DR}^\bullet(\mathcal{A}_P)/\mathrm{DR}^\bullet(\mathcal{T}_B) \xrightarrow{\delta} \mathrm{DR}^{\bullet+1}(\mathfrak{A}) \leftarrow \mathrm{DR}^{\bullet+1}(\mathcal{T}_B) = \Omega_B^{\bullet+1}$$

represents the boundary map

$$H^\bullet(\mathrm{DR}^\bullet(\mathcal{A}_P)/\mathrm{DR}^\bullet(\mathcal{T}_B)) \to H^{\bullet+1}(\mathrm{DR}^\bullet(\mathcal{T}_B)) = H^{\bullet+1}(B). \tag{27}$$

A basic class γ as above defines an n-dimensional cohomology class $\tilde{\gamma}$ of $\mathrm{DR}^\bullet(\mathcal{A}_P)/\mathrm{DR}^\bullet(\mathcal{T}_B)$, and $c_\gamma(P)$ is the image of $\tilde{\gamma}$ under (27).

4.2.3 The $\mathcal{D}$-module picture

Consider the short exact sequence

$$0 \to C^{\geq 1}(\mathrm{Ad}(P)_{/B}) \to C^\bullet(\mathrm{Ad}(P)_{/B}) \to C_B^\infty \to 0 \tag{28}$$

coming from the fact that $C_B^\infty = C^0(\mathrm{Ad}(P)_{/B})$ is the 0-th term of the relative cochain complex. If $\mathfrak{A}$ is as in (b), then all three complexes in (28) are graded $U(\mathfrak{A})$-modules in the following way. The elements $Y = (0, Y), Y \in \mathcal{A}$, act via the adjoint action. The element $\underline{X} = (X, 0), X \in \mathrm{Ad}(P)$, acts by contraction, i.e., by substitution of X into a cochain. The action of $U(\mathfrak{A})$ on C_B^∞ is via the quasi-isomorphism with $\mathcal{D}_B$.

Note that (28) splits as a short exact sequence of complexes of vector bundles but not of $U(\mathfrak{A})$-modules. We will use the corresponding connecting morphism

$$\delta : C_B^\infty \to C^{\geq 1}(\mathrm{Ad}(P)_{/B})[1]$$

in $D(U(\mathfrak{A}))$, the derived category of differential graded $U(\mathfrak{A})$-modules.

Since $\mathfrak{A}$ is quasi-isomorphic to $\mathcal{T}_B$, the DG algebra $U(\mathfrak{A})$ is quasi-isomorphic to $\mathcal{D}_B$, and the category $D(U(\mathfrak{A}))$ is equivalent to $D(\mathcal{D}_B)$. Now recall (Corollary 11) that

$$H^m(B; \mathbb{C}) = \mathrm{Hom}_{D(\mathcal{D}_B)}(C_B^\infty, C_B^\infty[m]).$$

On the other hand, suppose that $\mathfrak{g}$ is such that $H^i(\mathfrak{g}) = 0$ for $0 < i < n$. Then $H^i(\mathrm{Ad}(P)_{/B}) = H^i(\mathfrak{g}) \otimes C_B^\infty = 0$ for $0 < i < n$ as well. In other words, the complex $C^{\geq 1}(\mathrm{Ad}(P)_{/B})$ is acyclic in degrees $< n$, and therefore each class ξ in its n-th cohomology (which is isomorphic to $H^n(\mathfrak{g}) \otimes C_B^\infty$) defines a morphism in the derived category of complexes of vector bundles

$$\tilde{\xi} : C^{\geq 1}(\mathrm{Ad}(P)_{/B}) \to C_B^\infty[n].$$

Furthermore, a "constant" class ξ, i.e., a class of the form $\gamma \otimes 1$, $\gamma \in H^n(\mathfrak{g})$, defines in fact a morphism in the category $D(U(\mathfrak{A})) \sim D(\mathcal{D}_B)$. Composing $\widetilde{\gamma \otimes 1}$ with δ, we get a morphism

$$C_B^\infty \to C_B^\infty[n+1] \;, \tag{29}$$

i.e., a class in $H^{n+1}(B; \mathbb{C})$.

Proposition 32. *The class in $H^{n+1}(B;\mathbb{C})$ corresponding to (29) is equal to $c_\gamma(P)$.*

Proof. This follows directly from the definitions (in fact, we could take (29) as the definition of $c_\gamma(P)$). Indeed, the morphism in the derived category from the cohomology of a quotient complex such as C_B^∞ to the homology of a subcomplex such as $C^{\geq 1}(\mathrm{Ad}(P)_{/B})$ acyclic up to degree n is precisely the differential d_{n+1} in the corresponding spectral sequence. □

4.3 Infinite-dimensional groups

Slightly reformulating the approach of K.-T. Chen [8], we introduce the following definition.

Definition 33. *A differentiable space is an* ind-*object in the category of C^∞-manifolds.*

For background on ind-objects, see [10]. Thus a differentiable space M is a formal limit "$\varinjlim$"$_{\alpha\in A} M_\alpha$ of (finite-dimensional) C^∞-manifolds. In particular, M defines a functor

$$S \mapsto M(S) = C^\infty(S, M) = \varinjlim C^\infty(S, M_\alpha) \tag{30}$$

on such manifolds and can in fact be identified with this functor. In practice, however, we will identify M with the set $M(\mathrm{pt}) = \varinjlim M_\alpha$ with (30) providing an additional structure on this set (description of what it means for an element of this set to vary in a smooth family).

For a differential space M we define (compare [8]) the space of p-forms (in particular, of C^∞-functions) on M by

$$\Omega^p(M) = \varprojlim \Omega^p(M_\alpha).$$

For a point $m \in M(\mathrm{pt})$ the tangent space $T_m M$ is defined by

$$T_m M = \varinjlim T_s S \ ,$$

where the limit is taken over C^∞-maps $(S, s) \to (M, m)$.

A differentiable group G is a group object in the category of differentiable spaces. For such a group the space $\mathfrak{g} = T_e G$ is a Lie algebra in the standard way.

Example 34 (Groups of diffeomorphisms). Let Σ_0 be a compact oriented C^∞-manifold of dimension d. Then we have a differentiable group $G = \mathrm{Diffeo}(\Sigma_0)$ of orientation-preserving diffeomorphisms. The corresponding functor (30) is as follows. A smooth map $S \to \mathrm{Diffeo}(\Sigma_0)$ is a diffeomorphism of $S \times \Sigma_0$ preserving the projection to S. The Lie algebra of this group is $\mathrm{Vect}(\Sigma_0)$, the algebra of C^∞ vector fields.

Example 35 (Gauge groups). Let Σ_0 be as before and let E_0 be a C^∞ complex vector bundle on Σ_0. Then we have the differentiable group $\mathrm{Aut}(E_0)$ of C^∞-automorphisms of E_0 (the differentiable structure defined as in Example 34). Its Lie algebra is $\mathrm{End}(E_0)$.

Example 36 (Atiyah groups). Let Σ_0, E_0 be as before. The Atiyah group $AT(\Sigma_0, E_0)$ consists of pairs (ϕ, f), where ϕ is an orientation-preserving diffeomorphism of Σ_0, and $f : \phi^* E_0 \to E_0$ is an isomorphism of vector bundles. Thus we have an extension of differentiable groups:

$$1 \to \mathrm{Aut}(E_0) \to AT(\Sigma_0, E_0) \to \mathrm{Diffeo}(\Sigma_0) \to 1.$$

The Lie algebra of $AT(\Sigma_0, E_0)$ is $\mathcal{A}_{E_0}(\Sigma_0)$, the algebra of global C^∞-sections of the Atiyah Lie algebroid.

More generally, one can replace the vector bundle in Examples 35, 36 by a principal bundle with a Lie group of arbitrary structure. In this paper we will be interested in the vector bundle case and will concentrate on Example 36 as the most general.

Let us now describe a class of principal bundles with structure groups as in Example 36. Suppose that $q : \Sigma \to B$ is a smooth fibration with compact oriented fibers of dimension d. Suppose that B is connected. Then all the fibers $\Sigma_b = q^{-1}(b), b \in B$, are diffeomorphic to each other. Let Σ_0 be one such fiber. Futher, let E be a smooth $\mathbb{C}$-vector bundle on Σ and $E_b = E|_{\Sigma_b}$. Then, for different b the pairs (Σ_b, E_b) are isomorphic, in particular, isomorphic to (Σ_0, E_0). Let $G = AT(\Sigma_0, E_0)$. We have the principal G-bundle

$$\rho : P = P(\Sigma/B, E) \to B \tag{31}$$

whose fiber $P_b = \rho^{-1}(b)$, $b \in B$, consists of isomorphisms of pairs $(\Sigma_0, E_0) \to (\Sigma_b, E_b)$.

For any differentiable G-bundle P over a finite-dimensional base B the Atiyah algebra $\mathcal{A}_P$ can be defined by (8). In the example where $G = AT(\Sigma_0, E_0)$ and $P = P(\Sigma/B, E)$, this gives

$$\mathcal{A}_{P(\Sigma/B,E)} = q_* \mathcal{A}_E$$

(the sheaf-theoretic direct image of the Atiyah algebra of E).

4.4 The first Chern class

Let $q : \Sigma \to B$ and E be as before, so that we have a principal bundle $P = P(\Sigma/B, E) \to B$ with structure group $G = AT(\Sigma_0, E_0)$. Since the corresponding Lie algebra $\mathfrak{g} = \mathcal{A}_{E_0}(\Sigma_0)$ consists of global sections of the Atiyah Lie algebroid of Σ_0, we have the embeddings

$$\mathfrak{g} \hookrightarrow \mathcal{D}(E_0) \hookrightarrow \mathfrak{gl}(\mathcal{D}(E_0)).$$

By Corollary 23, $\mathfrak{gl}(\mathcal{D}(E_0))$ has a unique continuous (in the Fréchet topology) cohomology class c in degree $d + 1$. We denote by γ the restriction of c to $\mathfrak{g}$.

Proposition 37.

1. *There exists a Lie algebroid*

$$0 \to q_*(\mathfrak{gl}(\mathcal{D}_{\Sigma/B}(E))) \to \mathcal{A}_{\Sigma/B,E} \xrightarrow{\alpha} \mathcal{T}_B \to 0$$

and a morphism (embedding) of Lie algebroids $\mathcal{A}_P \to \mathcal{A}_{\Sigma/B,E}$ *that restricts to the embedding* $\mathfrak{g} \hookrightarrow \mathfrak{gl}(\mathcal{D}(E_0))$*.*

2. *The class* $1 \otimes \gamma$ *is transgressive, so* $d_{d+2}(1 \otimes \gamma)$ *is defined.*

Proof. The construction of $\mathcal{A}_{\Sigma/B,E}$ is given in Section 4.5 below.

The fibers of $\mathrm{Ker}(\alpha)$ are Lie algebras isomorphic to $\mathfrak{gl}(\mathcal{D}_{\Sigma_0,E_0})$ via an isomorphism defined uniquely up to an inner automorphism and thus satisfy the acyclicity condition (22) with $n = d+1$. Therefore the class c is transgressive.

The Hochschild–Serre spectral sequence for $\mathcal{A}_{\Sigma/B,E}$ maps into the analogous spectral sequence for $\mathcal{A}_P$. Since γ is the restriction of c, the naturality of the Hochschild–Serre spectral sequence implies that γ is transgressive. □

Definition 38. The first Chern class $\mathrm{C}_1(\mathrm{q}_*\mathrm{E})$ is defined by

$$C_1(q_*E) := d_{d+2}(1 \otimes \gamma) \in H^{d+2}(B, \mathbb{C}).$$

The class $C_1(q_*E)$ will be the main object of study in the rest of the paper.

4.5 Construction of $\mathcal{A}_{\Sigma/B,E}$

We start with the Atiyah Lie algebroid on Σ:

$$0 \to \mathrm{End}(E) \xrightarrow{i} \mathcal{A}_E \xrightarrow{\alpha} \mathcal{T}_\Sigma \to 0.$$

Let $U(\mathcal{A}_E)_{/B}$ denote the centralizer of $q^{-1}C^\infty_B$ in $U(\mathcal{A}_E)$. Let $F_1U(\mathcal{A}_E) = \{a | [a, q^{-1}C^\infty_B] \subseteq q^{-1}C^\infty_B\}$. Then, $F_1U(\mathcal{A}_E)$ is a Lie algebra under the commutator, $U(\mathcal{A}_E)_{/B}$ is a Lie ideal in $F_1U(\mathcal{A}_E)$, and there is an exact sequence

$$0 \to U(\mathcal{A}_E)_{/B} \to F_1U(\mathcal{A}_E) \to q^{-1}\mathcal{T}_B \to 0 \tag{32}$$

exhibiting $F_1U(\mathcal{A}_E)$ as a transitive $q^{-1}C^\infty_B$-algebroid.

The inclusion $\mathcal{A}_E \to \mathcal{D}_\Sigma(E)$ induces the surjective map $U(\mathcal{A}_E)_{/B} \to \mathcal{D}_{\Sigma/B,E}$ with kernel being the ideal generated by the relation that identifies $1 \in C^\infty_\Sigma \subset U(\mathcal{A}_E)_{/B}$ with $1 \in \underline{\mathrm{End}}_{C^\infty_\Sigma}(E) \subset U(\mathcal{A}_E)_{/B}$. The pushout of the exact sequence (32) by the map $U(\mathcal{A}_E) \to \mathcal{D}_\Sigma(E)$ gives the transitive Lie algebroid (the middle term in the exact sequence)

$$0 \to \mathcal{D}_{\Sigma/B,E} \to F_1\mathcal{D}_{\Sigma,E} \to q^{-1}\mathcal{T}_B \to 0. \tag{33}$$

Replacing E by its tensor product by the trivial bundle of rank r in the above example, (33) can be rewritten as

$$0 \to \mathfrak{gl}_r(\mathcal{D}_{\Sigma/B,E}) \to F_1\mathfrak{gl}_r(\mathcal{D}_{\Sigma,E}) \to q^{-1}\mathcal{T}_B \to 0.$$

Taking the limit over inclusions $\mathfrak{gl}_r \to \mathfrak{gl}_{r+1}$, we obtain a $q^{-1}C^\infty_B$-algebroid

$$0 \to \mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}) \to \mathcal{A}_{q,E} \to q^{-1}\mathcal{T}_B \to 0. \tag{34}$$

Let $\mathfrak{A}_{q,E}$ denote the cone of the inclusion $\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}) \to \mathcal{A}_{q,E}$. There are quasi-isomorphisms

$$\mathfrak{A}_{q,E} \to q^{-1}\mathcal{T}_B, \quad U_{q^{-1}C^\infty_B}(\mathfrak{A}_{q,E}) \to q^{-1}\mathcal{D}_B.$$

Taking the direct image of (34) under q and pulling back by the canonical map $\mathcal{T}_B \to q_*q^{-1}\mathcal{T}_B$, we get the following transitivity (since $R^1\pi_*\mathfrak{gl}_r(\mathcal{D}_{\Sigma/B,E}) = 0$) Lie algebroid on B:

$$0 \to \mathcal{G} \to \mathcal{A}_{\Sigma/B,E} \to \mathcal{T}_B \to 0\ ,$$

where $\mathcal{G} = q_*\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E})$, as we wanted. Let $\mathfrak{A}_{\Sigma/B,E}$ denote the differential graded Lie algebroid on B equal to the cone of the inclusion $\mathcal{G} \to \mathcal{A}_{\Sigma/B,E}$.

For any Lie algebra $\mathfrak{h}$, we denote by $C_+(\mathfrak{h})$ the positive part of the Chevalley–Eilenberg complex, i.e. $\oplus_{p>0}\Lambda^p\mathfrak{h}$ with the Chevalley–Eilenberg differential. There is an exact sequence of complexes

$$0 \to C_+(\mathfrak{h}) \to C_\bullet(\mathfrak{h}) \to C_0(\mathfrak{h}) \to 0.$$

The exact sequence

$$0 \to C_+(\mathcal{G}) \to C_\bullet(\mathcal{G}) \to C_0(\mathcal{G}) \to 0$$

is, in fact, an exact sequence of differential graded $U(\mathfrak{A}_{\Sigma/B})$-modules (this is a construction dual to (28)). Note that $C_0(\mathcal{G}) = C^\infty_B$. Let

$$\delta_{\Sigma/B} : C^\infty_B \to C_+(\mathcal{G})[1] \tag{35}$$

denote the correponding morphism in the derived category of differential graded modules over the universal enveloping (differential graded) algebra $U(\mathfrak{A}_{\Sigma/B})$.

4.6 Smooth cohomology and characteristic classes

A more traditional way of getting characteristic classes of principal G-bundles is by using group cohomology classes of G. Let us present a framework that we will then compare with the Lie algebra framework above.

Let S be a topological space and $\mathcal{F}$ a sheaf of abelian groups on S. We denote by $\Phi^\bullet(\mathcal{F})$ the standard Godement resolution of $\mathcal{F}$ by flabby sheaves. Thus $\Phi^0(\mathcal{F}) = DS(\mathcal{F})$ is the sheaf of (possibly discontinuous) sections of the (étale space associated to) $\mathcal{F}$, and $\Phi^{n+1}(\mathcal{F}) = DS(\Phi^n(\mathcal{F}))$. In this and the next sections we write $R\Gamma(S, \mathcal{F})$ for the complex of global sections $\Gamma(S, \Phi^\bullet(\mathcal{F}))$.

Let G be a differentiable group and $B_\bullet G$ its classifying space. Thus $B_\bullet G = (B_nG)_{n\geq 0}$ is a simplicial object in the category of differentiable spaces with

$B_nG = G^n$, and the face and degeneracy maps are given by the standard formulas. We define the smooth cohomology of G with coefficients in $\mathbb{C}^*$ to be

$$H^n_{sm}(G, \mathbb{C}^*) = \mathbb{H}^n(B_\bullet G, C^{\infty *}).$$

Here the hypercohomology on the right is defined as the cohomology of the double complex whose rows are the complexes $R\Gamma(B_nG, C^{\infty *}_{B_nG})$ and the differential between the neighboring slices coming from the simplicial structure on $B_\bullet G$. This is a version of the Segal cohomology theory for topological groups [13, p. 305]. In particular, we have a spectral sequence

$$H^i(B_nG, \mathbb{C}^*) \Rightarrow H^{i+n}_{sm}(G, \mathbb{C}^*).$$

We will use some other natural (complexes of) sheaves on $B_\bullet G$ to get natural cohomology theories for G, for example, the Deligne cohomology

$$H^n_{sm}(G, \mathbb{Z}_D(p)) = \mathbb{H}^n(B_\bullet G, \mathbb{Z}_D(p)) \ ,$$

where for any differentiable space M we set

$$\mathbb{Z}_D(p) = \left\{ \underline{\mathbb{Z}}_M \to \Omega^0_M \to \Omega^1_m \to \cdots \to \Omega^{p-1}_M \right\} ,$$

with $\underline{\mathbb{Z}}_M$ placed in degree zero; compare [6].

Let B be a C^∞-manifold and $\mathcal{U} = \{U_i\}_{i \in I}$ an open covering of B. We denote by $N_\bullet \mathcal{U}$ the simplicial nerve of $\mathcal{U}$, i.e., the simplicial manifold with

$$N_n \mathcal{U} = \coprod_{i_0, \ldots, i_n} U_{i_0} \cap \cdots \cap U_{i_n}.$$

For any sheaf $\mathcal{F}$ on B there is a natural isomorphism

$$\mathbb{H}^i(N_\bullet \mathcal{U}, \mathcal{F}_\bullet) = H^i(B, \mathcal{F}) \ ,$$

where $\mathcal{F}_\bullet$ is the natural sheaf on $N_\bullet \mathcal{U}$ whose n-th component is the sheaf on $N_n \mathcal{U}$ formed by the restrictions of $\mathcal{F}$.

Let $\rho : P \to B$ be a principal G-bundle and suppose that P is trivial on each U_i. Then a collection of trivializations (i.e., sections) $\tau = (\tau_i : U_i \to P)$ gives a morphism of simplicial differentiable spaces

$$u_\tau : N_\bullet \mathcal{U} \to B_\bullet G.$$

Given a class $\beta \in H^n_{sm}(G, \mathbb{C}^*)$, we define the characteristic class

$$\mathfrak{c}_\beta(P) = u^*_\phi(\beta) \in H^n(B, C^{\infty *}_B). \tag{36}$$

Similarly, one can define characteristic classes corresponding to group cohomology classes with values in the Deligne cohomology.

4.7 Integrality and integrability

Let G be as in Section 4.6, and let $\mathfrak{g}$ be the Lie algebra of G. We construct the "derivative" map

$$\partial : H^n_{sm}(G, \mathbb{C}^*) \to H^n_{\mathrm{Lie}}(\mathfrak{g}, \mathbb{C}). \tag{37}$$

To do this, we first observe that for any topological space S, any sheaf of abelian groups $\mathcal{F}$ on S, and any point $s_0 \in S$ we have a natural morphism of complexes

$$\epsilon_{s_0} : R\Gamma(S, \mathcal{F}) \to \mathcal{F}_{s_0},$$

where $\mathcal{F}_{s_0}$ is the stalk of $\mathcal{F}$ at s_0. To construct ϵ_{s_0}, we first project $R\Gamma(S, \mathcal{F}) = \Gamma(S, \Phi^\bullet(\mathcal{F}))$ to its 0-th term $\Gamma(S, \Phi^0(\mathcal{F}))$, which, by definition, is the space of all sections $\phi = (s \mapsto \phi_s)$ of the étale space of $\mathcal{F}$. Thus any such ϕ is a rule that to any point $s \in S$ associates an element of $\mathcal{F}_s$. We define ϵ_{s_0} by further mapping any ϕ as above to $\phi_{s_0} \in \mathcal{F}_{s_0}$.

We now specialize to $S = B_m G = G^m$, to $s_0 = e_m := (1, \ldots, 1)$, and to $\mathcal{F} = C^{\infty *}_S$. We get a morphism from the double complex

$$\{R\Gamma(B_m G, C^{\infty *}_{B_m G})\}_{m \geq 0} \tag{38}$$

to the complex of stalks

$$\mathbb{C}^* \to C^{\infty *}_{G, e_1} \to C^{\infty *}_{G \times G, e_2} \to \cdots .$$

Thus, an n-cocycle in (38) gives a germ of a smooth function

$$\xi = \xi(g_1, \ldots, g_n) : G^n \to \mathbb{C}^*$$

satisfying the group cocycle equation (on a neighborhood of e_{n+1} in G^{n+1}). Similarly to [13, p. 293], one associates to ξ a Lie algebra cocycle $\partial(\xi) \in C^n(\mathfrak{g})$ by

$$\partial(\xi)(x_1, \ldots, x_n) = \frac{d}{dt} \mathrm{Alt} \log \xi(\exp(tx_1), \ldots, \exp(tx_n))\bigg|_{t=0}.$$

A Lie algebra cohomology class $\gamma \in H^n(\mathfrak{g}, \mathbb{C})$ will be called integrable if it lies in the image of the map ∂ from (37). Consider the exponential exact sequence (13) of sheaves on B and its coboundary map δ_n from (14). The intuition with determinantal d-gerbes (1.4) suggests the following.

Conjecture 39.

1. The class $\gamma \in H^{d+1}(\mathcal{A}_{E_0}(\Sigma_0))$ constructed in Section 4.4 is integrable and comes from a natural class $\beta \in H^{d+1}_{sm}(\mathrm{AT}(\Sigma_0, E_0), \mathbb{C}^*)$ (the "higher determinantal class").

2. Furthermore, for any $q : \Sigma \to B$ and E as above, the class $C_1(q_*E) = c_\gamma(P) \in H^{d+2}(B, \mathbb{C})$ is integral and is the image of the following class in the integral cohomology:

$$\delta_{d+1}(\mathfrak{c}_\beta(P)) \in H^{d+2}(B, \mathbb{Z}).$$

This conjecture holds for $d = 1$ (i.e., for the case of a circle fibration). We will verify this in Section 6. In general, the second statement seems to follow from the first by virtue of some compatibility result between group cohomology classes with coefficients in $\mathbb{C}^*$ and Lie algebra cohomology classes with coefficients in $\mathbb{C}$. Here we present a $d = 1$ version of such a result.

Let G be a differentiable group with Lie algebra $\mathfrak{g}$. Let $\beta \in H^2_{sm}(G, \mathbb{C}^*)$ and let $\gamma = \partial(\beta) \in H^2_{\mathrm{Lie}}(\mathfrak{g}, \mathbb{C})$ be the derivative of β. Suppose β is represented by an extension of differentiable groups

$$1 \to \mathbb{C}^* \to \widetilde{G} \to G \to 1,$$

whose Lie algebra is the extension (25) representing γ. Let $\rho : P \to B$ be a principal G-bundle over a C^∞-manifold B. Then we have the characteristic class $c_\gamma(P) \in H^3(B, \mathbb{C})$ (the lifting to H^3 is well-defined because $\widetilde{\mathfrak{g}}$ is a G-module via the adjoint representation of $\widetilde{G}$, see Example 30). On the other hand, β gives rise to a class $\mathfrak{c}_\beta(P) \in H^2(B, C^{\infty *}_B)$; see (36).

Proposition 40. *In the above situation, $c_\gamma(P) \in H^3(B, \mathbb{C})$ is the image of $\delta_2(\mathfrak{c}_\beta(P)) \in H^3(B, \mathbb{Z})$ under the natural homomorphism from integral to complex cohomology.*

Proof. This follows from Theorem 19 using Example 30 and an obvious generalization of Example 18 to differentiable groups. □

Conjecture 41. We further conjecture the existence of the natural "deloopings" of the higher Chern classes as well, i.e., the existence of classes

$$\beta_m \in H^{d+2m}_{sm}(\mathrm{AT}(\Sigma_0, E_0), \mathbb{Z}_D(m)), \quad m \geq 1, \tag{39}$$

which then give characteristic classes in families:

$$C_m(q_*E) \in H^{d+2m}(B, \mathbb{Z}_D(m)). \tag{40}$$

5 The real Riemann–Roch

Here is the main result of the present paper.

Theorem 42. *Let $q : \Sigma \to B$ be a C^∞ fibration with compact oriented fibers of dimension d. Let E be a complex C^∞ vector bundle on Σ. Then*

$$C_1(q_*E) = \int_{\Sigma/B} \Big[\mathrm{ch}(E) \cdot \mathrm{Td}(\mathcal{T}_{\Sigma/B})\Big]_{2d+2} \in H^{d+2}(B, \mathbb{C}).$$

The proof consists of several steps.

5.1 A $\mathcal{D}$-module interpretation of C_1 using $\mathcal{A}_{\Sigma/B,E}$

We use the notation of Section 4.4 and introduce the following abbreviations:

$$\mathcal{G} = q_*(\mathfrak{gl}(\mathcal{D}_{\Sigma/B}(E))),$$

which is a bundle of infinite-dimensional Lie algebras on B,

$$\mathfrak{A} = \mathfrak{A}_{q,E},$$

which is a DG Lie algebroid on Σ quasi-isomorphic to $q^{-1}\mathcal{T}_B$,

$$U\mathfrak{A} = U_{q^{-1}C_B^\infty}(\mathfrak{A}_{q,E}) \ ,$$

which is a sheaf of DG-algebras on Σ quasi-isomorphic to $q^{-1}\mathcal{D}_B$.

Now, $U\mathfrak{A}$ acts on $C_+(\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}))_B$. Furthermore, a similar algebra acts on the Hochschild and cyclic complexes of $\mathcal{D}_{\Sigma/B,E}$. Let $\mathfrak{A}_0$ be the Lie algebroid defined exactly in the same way as $\mathfrak{A}$ but without tensoring by $\mathfrak{gl}$. In the same spirit as in Section 4.2.3, elements $Y = (0,Y), Y \in \mathcal{A}_{q,E}$, act via the adjoint action. Elements of the form $\underline{X} = (X,0)$ act via the shuffle multiplication

$$\iota_X(a_0 \otimes \cdots \otimes a_p) = \sum_{i=0}^{p}(-1)^i a_0 \otimes \cdots \otimes a_i \otimes X \otimes a_{i+1} \otimes \cdots \otimes a_p. \qquad (41)$$

Denoting by b, B the standard operators on Hochschild chains, see [21], we have

$$[b,\iota_X] = \mathrm{ad}(X), \quad [B,\iota_X] = 0.$$

Therefore $U\mathfrak{A}_0$ acts on both the Hochschild and the cyclic complexes. This action extends to the completions described in Section 3.4. Furthermore, the morphisms $\mu_{\mathcal{D}}, \nu_{\mathcal{D}}$ from Theorem 25 and Corollary 26 are in fact morphisms in $D(U\mathfrak{A})$. Indeed, there is a spectral sequence

$$E_2^{pq} = \mathrm{Ext}^p_{q^{-1}\mathcal{D}_B}\big(\underline{H}^q(\widehat{\mathrm{Hoch}}_\bullet(\mathcal{D}_{\Sigma/B}(E))), C_B^\infty\big) \Rightarrow \mathrm{Ext}^{p+q}_{U\mathfrak{A}_0}(\big(\widehat{\mathrm{Hoch}}_\bullet(\mathcal{D}_{\Sigma/B}(E)), C_B^\infty), \qquad (42)$$

and similarly for the cyclic complex.

The action of $q^{-1}\mathcal{D}_B$ on $\underline{H}^q(\widehat{\mathrm{Hoch}}_\bullet(\mathcal{D}_{\Sigma/B}(E)))$ is induced on the cohomology by the action of $U\mathfrak{A}$ on $\widehat{\mathrm{Hoch}}_\bullet(\mathcal{D}_{\Sigma/B}(E))$. The map $\mu_{\mathcal{D}}$ defines an element of E_2^{0d}, and $E_2^{pq} = 0$ for $q < d$, so $\mu_{\mathcal{D}}$ gives rise to a well-defined class in Ext^d on the RHS of (42). Similarly for $\nu_{\mathcal{D}}$.

We would like to compare the Lie algebra chain complex to the cyclic complex as modules over the algebras above. Roughly speaking, this comparison involves the embedding of $\mathfrak{A}_0$ into $\mathfrak{A}$ induced by the embedding of differential operators into matrix-valued differential operators as diagonal matrices all of whose diagonal entries are the same. Unfortunately, these operators are not finite and therefore do not lie in $\mathfrak{gl}$. This causes a minor technical difficulty that we are going to address next.

Let

$$C_+(\mathfrak{gl}(\mathcal{D}_{\Sigma/B}(E)))_B \xrightarrow{\beta} CC_\bullet(\mathcal{D}_{\Sigma/B}(E))_B[1]$$

be the standard map from the Lie algebra chain complex to the cyclic complex; see [21, (10.2.3)]. Observe that this map factors into the composition

$$C_+(\mathfrak{gl}((\mathcal{D}_{\Sigma/B}(E)))_B \xrightarrow{\text{proj}} (C_+(\mathfrak{gl}(\mathcal{D}_{\Sigma/B}(E)))_B)_{\mathfrak{gl}(\mathbb{C})} \to CC_\bullet(\mathcal{D}_{\Sigma/B}(E))_B[1]$$

(the complex in the middle is the complex of coinvariants). For each p the coinvariants stabilize: the projection

$$(C_p(\mathfrak{gl}_N(\mathcal{D}_{\Sigma/B}(E)))_B)_{\mathfrak{gl}_N(\mathbb{C})} \xrightarrow{\text{proj}_\text{N}} (C_p(\mathfrak{gl}(\mathcal{D}_{\Sigma/B}(E)))_B)_{\mathfrak{gl}(\mathbb{C})}$$

is an isomorphism for $N > p$. The DG Lie algebroid $\mathfrak{A}_0$ acts on the complex of $\mathfrak{gl}_N$-coinvariants via the diagonal embedding of $\mathcal{D}_{\Sigma/B}(E)$ into $\mathfrak{gl}_N(\mathcal{D}_{\Sigma/B}(E))$ for N big enough; this action is independent of N.

Let

$$\alpha : (C_+(\mathfrak{gl}(\mathcal{D}_{\Sigma/B}(E)))_B)_{\mathfrak{gl}(\mathbb{C})} \to q^{-1}C^\infty_B[2d]$$

denote the composition

$$\begin{aligned}(C_+(\mathfrak{gl}(\mathcal{D}_{\Sigma/B}(E)))_B)_{\mathfrak{gl}(\mathbb{C})} \xrightarrow{\beta} CC_\bullet(\mathcal{D}_{\Sigma/B}(E))_B[1] \to \\ \widehat{CC}_\bullet(\mathcal{D}_{\Sigma/B}(E))_B[1] \xrightarrow{\nu_{\mathcal{D}}[1]} q^{-1}C^\infty_B[2d+1]. \quad (43)\end{aligned}$$

It is checked directly that β commutes with the operators ι_X, so it is $U\mathfrak{A}_0$-invariant. Therefore, all maps in (43) and the map α are morphisms in $D(U\mathfrak{A}_0)$.

Let us now take the direct image and define the morphism

$$\int_{\Sigma/B} \alpha : (C_+(\mathcal{G})_B)_{\mathfrak{gl}(\mathbb{C})} \to C^\infty_B[d]$$

as the composition

$$\begin{aligned}(C_+(\mathcal{G})_B)_{\mathfrak{gl}(\mathbb{C})} \to q_*(C_+\mathfrak{gl}(\mathcal{D}_{\Sigma/B}(E))_B)_{\mathfrak{gl}(\mathbb{C})} \xrightarrow{\sim} \\ Rq_*(C_+\mathfrak{gl}(\mathcal{D}_{\Sigma/B}(E))_B)_{\mathfrak{gl}(\mathbb{C})} \xrightarrow{\alpha} Rq_*q^{-1}C^\infty_B[2d+1] \xrightarrow{\int_{\Sigma/B}} C^\infty_B[d+1].\end{aligned}$$

Here the last map is the integration over the relative (topological) fundamental class of Σ/B. Consider the composition

$$C^\infty_B \xrightarrow{\delta_{\Sigma/B}} (C_+(\mathcal{G})_B)_{\mathfrak{gl}(\mathbb{C})}[1] \xrightarrow{\int_{\Sigma/B}\alpha} C^\infty_B[d+2]\ , \quad (44)$$

where $\delta_{\Sigma/B}$ is as in (35). Since both maps in (44) are morphisms in $D(\mathcal{D}_B)$, the composition (denote it by C) is an element

$$C \in \operatorname{Ext}^{d+2}_{\mathcal{D}_B}(C^\infty_B, C^\infty_B) = H^{d+2}(B,\mathbb{C}).$$

Proposition 43. *We have $C = C_1(q_*E)$.*

Proof. This follows from the interpretation of $C_1(q_*E) = c_\gamma(P(\Sigma/B,E))$ given in Sections 4.2.2 and 4.2.3, and from the compatibility of the Atiyah algebroid of $P(\Sigma/B)$ with $\mathcal{A}_{\Sigma/B,E}$. □

5.2 A local RRR in the total space

Proposition 43 reduces the RRR to the following "local" statement taking place in the total space Σ.

Theorem 44. *Let ξ be the morphism in $D(q^{-1}\mathcal{D}_B)$ defined as the composition*

$$q^{-1}C_B^\infty \to C_+(\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}))_{\mathfrak{gl}(\mathbb{C})}[1] \to q^{-1}C_B^\infty[2d+2].$$

Then the class in

$$\mathrm{Ext}^{2d+2}_{q^{-1}\mathcal{D}_B}(q^{-1}C_B^\infty, q^{-1}C_B^\infty) = H^{2d+2}(\Sigma,\mathbb{C})$$

corresponding to ξ is equal to

$$\Big[\mathrm{ch}(E)\cdot \mathrm{Td}(\mathcal{T}_{\Sigma/B})\Big]_{2d+2}.$$

We now concentrate on the proof of Theorem 44. First, we recall the definition of periodic cyclic homology [21]. Let A be an associative algebra. The "negative" cyclic complex of A is defined, similarly to (16), as

$$CC_\bullet^-(A) = \mathrm{Tot}\Big\{\mathrm{Hoch}_\bullet(A) \xrightarrow{N} \mathrm{Hoch}_\bullet(A) \xrightarrow{1-\tau} \mathrm{Hoch}_\bullet(A) \to \cdots\Big\}$$

Here, the grading of the copies of $\mathrm{Hoch}_\bullet(A)$ in the horizontal direction goes in increasing integers $0,1,2$ etc. So $CC_\bullet^-(A)$ is a module over the formal Taylor series ring $\mathbb{C}[[u]]$, where u has degree (-2). The original cyclic complex is a module over the polynomial ring $\mathbb{C}[u^{-1}]$. Finally, the periodic cyclic complex $CC_\bullet^{\mathrm{per}}(A)$ is obtained by merging together $CC_\bullet(A)$ and $CC_\bullet^-(A)$ into one double complex that is repeated 2-periodically both in the positive and negative horizontal directions. In other words,

$$CC_\bullet^{\mathrm{per}}(A) = CC_\bullet^-(A)\otimes_{\mathbb{C}[[u]]}\mathbb{C}((u)).$$

We extend this construction to other situations (see Section 3) in which the tensor products are understood in the sense of various completions. In particular, the morphism ν_D of Corollary 26 extends to morphisms

$$\nu_D^- : CC_\bullet^-(\mathcal{D}_{\Sigma/B,E}) \to q^{-1}C_B^\infty[2d][[u]],$$
$$\nu_D^{\mathrm{per}} : CC_\bullet^{\mathrm{per}}(\mathcal{D}_{\Sigma/B,E}) \to q^{-1}C_B^\infty[2d]((u)).$$

These morphisms can be included in the commutative diagram

$$\begin{tikzcd}
CC_\bullet^-(\mathcal{D}_{\Sigma/B,E}) \arrow[r] \arrow[d, "\nu_{\mathcal{D}}^-"'] & CC_\bullet^{\mathrm{per}}(\mathcal{D}_{\Sigma/B,E}) \arrow[r] \arrow[d, "\nu_{\mathcal{D}}^{\mathrm{per}}"'] & CC(\mathcal{D}_{\Sigma/S,E})[2] \arrow[d, "\nu_{\mathcal{D}}"] \\
C_B^\infty[2d][[u]] \arrow[r] & C_B^\infty[2d]((u)) \arrow[r, "\mathrm{Res}_{u=0}"] & C_B^\infty[2d+2]
\end{tikzcd}$$

We now want to reduce Theorem 44 to the following statement.

Theorem 45. *The composition*

$$C_B^\infty \xrightarrow{1} CC_\bullet^{\mathrm{per}}(\mathcal{D}_{\Sigma/B,E}) \xrightarrow{\nu_{\mathcal{D}}^{\mathrm{per}}} C_B^\infty[2d]((u))$$

defines an element of $\mathrm{Ext}^\bullet_{q^{-1}\mathcal{D}_B}(q^{-1}C_B^\infty, q^{-1}C_B^\infty[2d])((u))$ *that is equal to*

$$\sum_{i=0}^{\infty} u^i \cdot \left[\mathrm{ch}(E)\mathrm{Td}(\mathcal{T}_{\Sigma/B})\right]_{2(d-i)}.$$

Proof (Theorem 44). Assuming Theorem 45, it is sufficient to prove that the composition

$$q^{-1}C_B^\infty \to C_+(\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}))_{\mathfrak{gl}(\mathbb{C})}[1] \to CC_\bullet(\mathcal{D}_{\Sigma/B,E})[2]$$

is equal to the composition

$$q^{-1}C_B^\infty \xrightarrow{1} CC_\bullet^{\mathrm{per}}(\mathcal{D}_{\Sigma/B,E}) \to CC_\bullet(\mathcal{D}_{\Sigma/B,E})[2]\ ,$$

since the latter one is related to Chern and Todd via Theorem 45. In order to perform the comparison, let K be the cone of the inclusion $C_+(\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}) \to C_\bullet(\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}))$, so that we have a quasi-isomorphism $K \to q^{-1}C_B^\infty$ as well as an isomorphism of distinguished triangles

$$\begin{tikzcd}
C_\bullet(\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}))_{\mathfrak{gl}(\mathbb{C})} \arrow[r] \arrow[d] & K \arrow[r] \arrow[d] & C_+(\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}))_{\mathfrak{gl}(\mathbb{C})}[1] \arrow[d] \\
C(\mathfrak{gl}(\mathcal{D}_{\Sigma/B,\mathcal{E}}))_{\mathfrak{gl}(\mathbb{C})} \arrow[r] & q^{-1}C_B^\infty \arrow[r] & C_+(\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}))[1]_{\mathfrak{gl}(\mathbb{C})}
\end{tikzcd}$$

(with the top row a short exact sequence of complexes). Note that there is a morphism of distinguished triangles

$$\begin{tikzcd}
C_\bullet(\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}))_{\mathfrak{gl}(\mathbb{C})} \arrow[r] \arrow[d] & K \arrow[r] \arrow[d] & C_+(\mathfrak{gl}(\mathcal{D}_{\Sigma/B,E}))_{\mathfrak{gl}(\mathbb{C})}[1] \arrow[d] \\
CC_\bullet^-(\mathcal{D}_{\Sigma/B,E}) \arrow[r] & CC_\bullet^{\mathrm{per}}(\mathcal{D}_{\Sigma/B,E}) \arrow[r] & CC_\bullet(\mathcal{D}_{\Sigma/B,E})[2]
\end{tikzcd}$$

It remains to notice further that the diagram

$$q^{-1}C_B^\infty \longleftarrow K \longrightarrow CC_\bullet^{\mathrm{per}}(\mathcal{D}_{\Sigma/B,E})$$

represents the morphism $C_B^\infty \xrightarrow{1} CC_\bullet^{\mathrm{per}}(\mathcal{D}_{\Sigma/B,E})$ in the derived category. □

5.3 Proof of Theorem 45

This statement can be deduced from the results of [25] on the cohomology of the Lie algebras of formal vector fields and formal matrix functions. We recall the setting of [25], which extends that of the Chern–Weil definition of characteristic classes. Recall that the latter provides a map

$$S^\bullet[[\mathfrak{h}_0]]^{H_0} \to H^{2\bullet}(\Sigma, \mathbb{C}) \;, \tag{45}$$

where $H_0 = \mathrm{GL}_d(\mathbb{C}) \times \mathrm{GL}_r(\mathbb{C})$ with $r = \mathrm{rk}(E)$, while $\mathfrak{h}_0$ is the Lie algebra of H_0, i.e., $\mathfrak{gl}_d(\mathbb{C}) \oplus \mathfrak{gl}_r(\mathbb{C})$. To be precise, the elementary symmetric functions of the two copies of $\mathfrak{gl}$ are mapped to the Chern classes of $\mathcal{T}_{\Sigma/B}$ and E.

In [25], this construction was generalized in the following way. Let $k = \dim(B)$, and let $\widehat{\mathfrak{g}}$ be the Lie algebra of formal differential operators of the form

$$\sum_{i=1}^{k} P_i(y_1, \ldots, y_k) \frac{\partial}{\partial y_i} + \sum_{j=1}^{d} Q_j(x_1, \ldots, x_d, y_1, \ldots, y_k) \frac{\partial}{\partial x_i} + R(x_1, \ldots, x_d, y_1, \ldots, y_k),$$

where P_i, Q_j are formal power series, and $R(x)$ is an $r \times r$ matrix whose entries are power series. Thus $\widehat{\mathfrak{g}}$ is the formal version of the relative Atiyah algebra. Consider the Lie subalgebra $\mathfrak{h}$ of fields such that all P_i and Q_j are of degree-one and all entries of R are of degree zero. We can identify this subalgebra with

$$\mathfrak{h} = \mathfrak{gl}_d(\mathbb{C}) \oplus \mathfrak{gl}_k(\mathbb{C}) \oplus \mathfrak{gl}_r(\mathbb{C}).$$

Let

$$H = \mathrm{GL}_d(\mathbb{C}) \times \mathrm{GL}_k(\mathbb{C}) \times \mathrm{GL}_r(\mathbb{C})$$

be the corresponding Lie group. Thus $(\widehat{\mathfrak{g}}, H)$ form a Harish–Chandra pair. Following the ideas of "formal geometry" (or "localization") of Gelfand and Kazhdan, one sees that every $(\widehat{\mathfrak{g}}, H)$-module L induces a sheaf $\mathcal{L}$ on Σ. Similarly, a complex $L^\bullet$ of modules gives rise to a complex of sheaves $\mathcal{L}^\bullet$. A complex $L^\bullet$ of modules is called *homotopy constant* if the action of $\widehat{\mathfrak{g}}$ extends to an action of the differential graded Lie algebra $(\widehat{\mathfrak{g}}[\epsilon], \frac{\partial}{\partial \epsilon})$. Here ϵ is a formal variable of degree -1 and square zero. In this case, there is a generalization of the Chern–Weil map constructed in [25]:

$$\mathrm{CW} : \mathbb{H}^\bullet(\mathfrak{h}_0[\epsilon], \mathfrak{h}_0; L^\bullet) \to \mathbb{H}^\bullet(\Sigma, \mathcal{L}^\bullet),$$

which gives (45) when $L = \mathbb{C}$ with the trivial action. Consider the following $(\widehat{\mathfrak{g}}, H)$-modules:

$$\mathcal{D} = \left\{ \sum P_\alpha(x_1, \ldots, x_d, y_1, \ldots, y_k) \partial_x^\alpha \right\},$$

where P_α are $r \times r$ matrices whose entries are power series, and

$$\Omega^\bullet = \left\{ \sum_I P_I(x_1, \ldots, x_d, y_1, \ldots, y_k) d^I x \right\},$$

which is the space of differential forms whose coefficients are formal power series. The latter is a complex with the (fiberwise) de Rham differential. Moreover, $\Omega^\bullet$ is homotopy constant ($\epsilon\widehat{\mathfrak{g}}$ acts on it by exterior multiplication). The Hochschild, cyclic, etc. complexes of $\mathcal{D}$ inherit the $(\widehat{\mathfrak{g}}, H)$-module structure; moreover, they also become homotopy constant (an element $\epsilon X \in \epsilon\widehat{\mathfrak{g}}$ acts by the operator ι_X from equation (41). One constructs [4, pt. II, Lemma 3.2.4] a class

$$\nu \in \mathbb{H}^0(\mathfrak{h}_0[\epsilon], \mathfrak{h}_0; \underline{\mathrm{Hom}}(\mathrm{CC}^{\mathrm{per}}_{-\bullet}(\mathcal{D}), \Omega^{2d+\bullet}))$$

such that $\mathrm{CW}(\nu)$ coincides with

$$\nu_{\mathcal{D}} \in \mathbb{H}^0(\Sigma; \underline{\mathrm{Hom}}(\mathrm{CC}^{\mathrm{per}}_{-\bullet}(\mathcal{D}_{\Sigma/B}), \Omega^{2d+\bullet}_{\Sigma/B})).$$

To be precise, the cited lemma concerns the Weyl algebra of power series in both coordinates and derivations with the Moyal product (clearly, differential operators of finite order form a subalgebra). Second, the construction there is for the relative cohomology of the pair $(\mathfrak{g}, \mathfrak{h})$, but it extends to the case of the pair $(\mathfrak{g}[\epsilon], \mathfrak{h})$, of which $(\mathfrak{h}_0[\epsilon], \mathfrak{h}_0)$ is a subpair.

The cochain ν is actually independent of y. There is the canonical class 1 in ${\mathrm{HC}_0}^{\mathrm{per}}(\mathcal{D})$; it is $\mathfrak{h}_0$-invariant, and it is shown in [25] how to extend it to a class in $\mathbb{H}^0(\mathfrak{h}_0[\epsilon], \mathfrak{h}_0; \mathrm{CC}^{\mathrm{per}}_{-\bullet}(\mathcal{D}))$. On the other hand,

$$\mathbb{H}^0(\mathfrak{h}_0[\epsilon], \mathfrak{h}_0; \Omega^\bullet)$$

can be naturally identified with

$$\mathbb{H}^0(\mathfrak{h}_0[\epsilon], \mathfrak{h}_0; \mathbb{C}).$$

It remains to show that

$$\nu(1) = \sum [\mathrm{ch} \cdot \mathrm{Td}]_{2(d+i)} \cdot u^i,$$

where ch is the corresponding invariant power series in $H^\bullet(\mathfrak{gl}_r[\epsilon], \mathfrak{gl}_r; \mathbb{C})$ and Td is the corresponding invariant power series in $H^\bullet(\mathfrak{gl}_d[\epsilon], \mathfrak{gl}_d; \mathbb{C})$. This was carried out in [4, Lemma 5.3.2] □

6 Comparison with the gerbe picture

6.1 L^2-sections of a vector bundle on a circle

Let Σ be an oriented C^∞-manifold diffeomorphic to the circle S^1 with the standard orientation, and let E be a complex C^∞-vector bundle on Σ. Choose a smooth Riemannian metric g on Σ and a smooth Hermitian metric h on E. Let $\Gamma(\Sigma, E)$ be the space of C^∞-sections of E. The choice of g, h defines a positive definite scalar product on this space, and we denote by $L^2_{g,h}(\Sigma, E)$ the Hilbert space obtained by completion with respect to this scalar product.

Lemma 46. *For a different choice g', h' of metrics on Σ, E we have a canonical identification of topological vector spaces*

$$L^2_{g,h}(\Sigma, E) \to L^2_{g',h'}(\Sigma, E).$$

Proof. The Hilbert norms on $\Gamma(\Sigma, E)$ associated to (g, h) and (g', h') are equivalent, since Σ is compact. □

We will denote the completion simply by $L^2(\Sigma, E)$.

Consider now the case in which $\Sigma = S^1$ is the standard circle and $E = \mathbb{C}^r$ is the trivial bundle of rank r. In this case, $L^2(\Sigma, E) = L^2(S^1)^{\oplus r}$. Let us denote this Hilbert space by H. It comes with a polarization in the sense of Pressley and Segal [26]. In other words, H is decomposed as $H_+ \oplus H_-$, where H_+, H_- are infinite-dimensional orthogonal closed subspaces defined as follows.

The space H_+ consists of vector functions extending holomorphically into the unit disk $D_+ = \{|z| < 1\}$. The space H_- consists of vector functions extending holomorphically into the opposite annulus $D_- = \{|z| > 1\}$ and vanishing at ∞.

The decomposition $H = H_+ \oplus H_-$ yields the groups $\mathrm{GL}_{\mathrm{res}}(H) \subset \mathrm{GL}(H)$, see [26, (6.2.1)], as well as the Sato Grassmannian $\mathrm{Gr}(H)$ on which $\mathrm{GL}_{\mathrm{res}}(H)$ acts transitively. We recall that $\mathrm{Gr}(H)$ consists of closed subspaces $W \subset H$ whose projection to H_+ is a Fredholm operator and whose projection to H_- is a Hilbert–Schmidt operator; see [26, (7.1.1)].

Given arbitrary Σ, E as before, we can choose an orientation-preserving diffeomorphism $\phi : S^1 \to \Sigma$ and a trivialization $\psi : \phi^* E \to \mathbb{C}^r$. This gives an identification

$$u_{\phi,\psi} : L^2(\Sigma, E) \to H = L^2(S^1)^{\oplus r}.$$

In particular, we get a distinguished set of subspaces in $L^2(\Sigma, E)$, namely

$$\mathrm{Gr}_{\phi,\psi}(\Sigma, E) = u^{-1}_{\phi,\psi}(\mathrm{Gr}(H)),$$

and a distinguished subgroup of its automorphisms, namely

$$\mathrm{GL}^{\phi,\psi}_{\mathrm{res}}(L^2(\Sigma, E)) = u^{-1}_{\phi,\psi} \mathrm{GL}_{\mathrm{res}}(H) u_{\phi,\psi}.$$

Lemma 47. *The subgroup $\mathrm{GL}^{\phi,\psi}_{\mathrm{res}}(L^2(\Sigma, E))$ and the set $\mathrm{Gr}_{\phi,\psi}(L^2(\Sigma, E))$ are independent of the choice of ϕ and ψ.*

Proof. Any two choices of ϕ, ψ differ by an element of the Atiyah group $\mathrm{AT}(S^1, \mathbb{C}^r)$; see Example 36. This group being a semidirect product of $\mathrm{Diffeo}(S^1)$ and $\mathrm{GL}_r C^\infty(S^1)$, our statement follows from the known fact that both of these groups are subgroups of $\mathrm{GL}_{\mathrm{res}}(H)$; see [26]. □

From now on, we drop ϕ, ψ from the notation, writing $\mathrm{Gr}(L^2(\Sigma, E))$ and $\mathrm{GL}_{\mathrm{res}}(L^2(\Sigma, E))$. Recall further that $\mathrm{Gr}(H) \times \mathrm{Gr}(H)$ is equipped with a line bundle Δ (the relative determinantal bundle) which has the following additional structures:

1. Equivariance with respect to $\mathrm{GL}_{\mathrm{res}}(H)$.
2. A multiplicative structure, i.e., an identification
$$p_{12}^*\Delta \otimes p_{23}^*\Delta \to p_{13}^*\Delta \tag{46}$$
of vector bundles on $\mathrm{Gr}(H) \times \mathrm{Gr}(H) \times \mathrm{Gr}(H)$, which is equivariant under $\mathrm{GL}_{\mathrm{res}}(H)$ and satisfies the associativity, unit and inversion properties.

It follows from the above that we have a canonically defined line bundle (still denoted by Δ) on $\mathrm{Gr}(L^2(\Sigma, E)) \times \mathrm{Gr}(L^2(\Sigma, E))$ equivariant under $\mathrm{GL}_{\mathrm{res}}(L^2(\Sigma, E))$ and equipped with a multiplicative structure. For $W, W' \in \mathrm{Gr}(L^2(\Sigma, E))$ we denote by $\Delta_{W,W'}$ the fiber of Δ at (W, W').

As is well known, the multiplicative bundle Δ gives rise to a category ($\mathbb{C}^*$-gerbe) $\mathcal{D}et\, L^2(\Sigma, E)$ whose set of objects is $\mathrm{Gr}(L^2(\Sigma, E))$, while

$$\mathrm{Hom}_{\mathcal{D}et\, L^2(\Sigma,E)}(W, W') = \Delta_{W,W'} - \{0\}.$$

The composition of morphisms comes from the identification

$$\Delta_{W,W'} \otimes \Delta_{W',W''} \to \Delta_{W,W''}$$

given by (46).

6.2 L^2-direct image in a circle fibration

Now let $q : \Sigma \to B$ be a fibration in oriented circles and E a vector bundle on Σ. We then have a bundle of Hilbert spaces $q_*^{L^2}(E)$ whose fiber at $b \in B$ is $L^2(\Sigma_b, E_b)$. Furthermore, by Lemma 47 this bundle has a $\mathrm{GL}_{\mathrm{res}}(H)$-structure, where $H = L^2(S^1)^{\oplus r}$. Therefore we have the associated bundle of Sato Grassmannians $\mathrm{Gr}(q_*^{L^2}(E))$ on B and the (fiberwise) multiplicative line bundle Δ on
$$\mathrm{Gr}(q_*^{L^2}(E)) \times_B \mathrm{Gr}(q_*^{L^2}(E)).$$

We define a sheaf of $C_B^{\infty *}$-groupoids on B whose local objects are local sections of $\mathrm{Gr}(q_*^{L^2}(E))$, and for any two such sections defined on $U \subset B$,

$$\underline{\mathrm{Hom}}(s_1, s_2) = (s_1, s_2)^*\Delta - 0_U,$$

where 0_U stands for the zero section of the induced line bundle. This sheaf of groupoids is locally connected and so gives rise to a $C_B^{\infty *}$-gerbe, which we denote by $\mathcal{D}et(q_*E)$. So we have the class

$$\left[\mathcal{D}et(q_*e)\right] \in H^2(B, C_B^{\infty *}).$$

Alternatively, consider the Atiyah group $G = \mathrm{AT}(S^1, \mathbb{C}^r)$; see Example 36. By the above, $G \subset \mathrm{GL}_{\mathrm{res}}(H)$. The determinantal $\mathbb{C}^*$-gerbe $\mathcal{D}et(H)$ (over a point) with G-action gives a central extension $\widetilde{G}$ of G by $\mathbb{C}^*$. A circle fibration $q : \Sigma \to B$ gives a principal G-bundlle $P(\Sigma/B)$, as in (31), and the following is clear.

Proposition 48. *The gerbe $\mathcal{D}et(q_*E)$ is equivalent to $\mathrm{Lift}_G^{\widetilde{G}}(P(\Sigma/B, E))$; see Example 18.*

Consider the exponential sequence (13) of sheaves on B and the corresponding coboundary map δ_2, see (14). Then we have the class

$$\delta_2\big[\mathcal{D}et(q_*E)\big] \in H^3(B, \mathbb{Z}).$$

Theorem 49. *The image of $\delta\big[\mathcal{D}et(q_*E)\big]$ in $H^3(B, \mathbb{C})$ coincides with the negative of the class $C_1(q_*E)$ (see Definition 38).*

Proof. We apply Proposition 40 to $G = \mathrm{AT}(S^1, \mathbb{C}^r)$ and β being the class of the central extension $\widetilde{G}$. Then $\mathfrak{g} = \mathcal{A}_{\mathbb{C}^r}(S^1)$ is the Atiyah algebra of the trivial bundle on S^1 and γ is the class of the "trace" central extension induced from the Lie algebra $\mathfrak{gl}_{\mathrm{res}}(H)$ of $\mathrm{GL}_{\mathrm{res}}(H)$. We have the embeddings

$$\mathfrak{g} \subset \mathfrak{gl}_r(\mathcal{D}(S^1)) \subset \mathfrak{gl}_{\mathrm{res}}(H),$$

and the trace central extension is represented by an explicit cocycle Ψ of $\mathfrak{gl}_{\mathrm{res}}(H)$ (going back to [28]). Let z be the standard complex coordinate on S^1 such that $|z| = 1$. Then the formula for the restriction of Ψ to $\mathfrak{gl}_r(\mathcal{D}(S^1))$ was given in [15], see also [16, formula (1.5.2)]:

$$\Psi(f(z)\partial_z^m, g(z)\partial_z^n) = \frac{m!n!}{(m+n+1)!}\mathrm{Res}_{z=0} dz \cdot \mathrm{Tr}(f^{(n+1}(z) g^{(m)}(z)),$$

where $f^{(n)}$ denotes the n-th derivative with respect to z. Our statement now reduces to the following lemma. □

Lemma 50. *The second Lie cohomology class of $\mathfrak{gl}_r\mathcal{D}(S^1)$ given by the cocycle Ψ is equal to the negative of the class corresponding to the fundamental class of S^1 via the identification of Corollary 23.*

Proof. Since the space of (continuous) Lie algebra homology in question is 1-dimensional, it is enough to evaluate the cocycle Ψ on the Lie algebra 2-homology class σ from Corollary 23 and to show that this value is precisely equal to 1. For this it is enough to consider $r = 1$. Let $\mathcal{D} = \mathcal{D}(S^1)$ for simplicity.

We need to recall the explicit form of the identification (18) for the case $n = 1$ (first Hochschild homology maps to the second Lie algebra homology). In other words, we need to recall the definition of the map

$$\epsilon : HH_1(\mathcal{D}) \to H_2^{\mathrm{Lie}}(\mathfrak{gl}(\mathcal{D})) \to \mathbb{C}.$$

As explained in [7] and [29], this map is defined via the order filtration F on the ring $\mathcal{D}$ and uses the corresponding spectral sequence. This means that we need to start with a Hochschild 1-cycle $\sigma = \sum P_i \otimes Q_i \in \mathcal{D} \otimes \mathcal{D}$ and form its highest symbol cycle

$$\mathrm{Smbl}(\sigma) = \sum \mathrm{Smbl}(P_i) \otimes \mathrm{Smbl}(Q_i) \in \mathrm{gr}(\mathcal{D}) \otimes \mathrm{gr}(\mathcal{D}),$$

which gives an element in $\mathrm{Hoch}_1(\mathrm{gr}(\mathcal{D}))$. Since $\mathrm{gr}(\mathcal{D})$ is the ring of polynomial functions on T^*S^1, Hochschild–Kostant–Rosenberg gives $HH_1(\mathrm{gr}(\mathcal{D})) = \Omega^1(T^*S^1)$, the space of 1-forms on T^*S^1 polynomial along the fibers. So the class of $\mathrm{Smbl}(\sigma)$ is a 1-form $\omega = \omega(\sigma)$ on T^*S^1. This is an element of the E_1-term of the spectral sequence for the Hochschild homology of the filtered ring $\mathcal{D}$.

Furthermore, one denotes by $*$ the symplectic Hodge operator on forms on T^*S^1. The results of [7], [29] imply that the differential in the E_1-term is $*d*$, where d is the de Rham differential on T^*S^1 while higher differentials vanish. This means that under our assumptions, $*\omega(\sigma)$ is a closed 1-form and

$$\epsilon(\sigma) = \int_{S^1} *\omega(\sigma).$$

To finish the proof we need to exhibit just one σ as above such that

$$0 \neq \epsilon(\sigma) = \Psi(\sigma) := \sum \Psi(P_i, Q_i).$$

We take

$$\sigma = z^2 \otimes z^{-1}\partial_z - 2z \otimes \partial_z.$$

Then one sees that σ is a Hochschild 1-cycle and $\Psi(\sigma) = 1$. On the other hand, let θ be the real coordinate on S^1 such that $z = \exp(2\pi i\theta)$. Then the real coordinates on T^*S^1 are θ, ξ with $\xi = \mathrm{Smbl}(\partial/\partial\theta)$, so the Poisson bracket $\{\theta, \xi\}$ is equal to 1. In terms of the coordinate z it means that $\xi = \mathrm{Smbl}(z\partial/\partial z)$ and $\{z, \xi\} = z$. Therefore

$$\mathrm{Smbl}(\sigma) = z^2 \otimes z^{-2}\xi - 2z \otimes z^{-1}\xi,$$

and hence

$$\omega(\sigma) = z^2 d(z^{-2}\xi) = 2zd(z^{-1}\xi) = -dz - z^{-1}\xi\ ,$$

see [21, p.11]. The symplectic (volume) form on T^*S^1 is $(dz/z) \wedge d\xi$, so the symplectic Hodge operator is given by

$$*d\xi = dz/z, \quad *dz/z = d\xi, \quad *^2 = 1.$$

Therefore,

$$*\omega(\sigma) = -dz/z - \xi d\xi, \qquad \int_{S^1} *\omega(\sigma) = -1,$$

and we are done. $\square$

References

1. M.F. Atiyah, F. Hirzebruch, *Riemann–Roch theorems for differentiable manifolds*, Bull. Amer. Math. Soc. **65** (1959), 276–281.
2. S. Bloch, *K_2 and algebraic cycles*, Ann. of Math. (2) **99** (1974), 349–379.
3. R. Bott, *On the characteristic classes of groups of diffeomorphisms*, L'Enseignment Math. **23** (1977), 209–220.
4. P. Bressler, R. Nest, B. Tsygan, *Riemann–Roch theorems via deformation quantization.* I, II, Adv. Math. **167** (2002), no. 1, 1–25, 26–73.
5. L. Breen, *On the classification of 2-gerbes and 2-stacks, Astérisque*, **225** (1994).
6. J.-L. Brylinski, *Loop Spaces, Characteristic Classes and Geometric Quantization*, Birkhäuser, Boston, 1993.
7. J.-L. Brylinski, E. Getzler, *The homology of algebras of pseudodifferential symbols and the noncommutative residue*, K-Theory **1** (1987), no. 4, 385–403.
8. K.T. Chen, *Iterated path integrals*, Bull. Amer. Math. Soc. **83** (1977), no. 5, 831–879.
9. P. Deligne, *Le déterminant de la cohomologie*, Current trends in arithmetical algebraic geometry (Arcata, Calif., 1985), 93–177, Contemp. Math., 67, Amer. Math. Soc., Providence, RI, 1987.
10. P. Deligne, *Le groupe fondamental de la droite projective moins trois points*, Galois groups over $\mathbb{Q}$ (Berkeley, CA, 1987), 79–297, Math. Sci. Res. Inst. Publ., **16**, Springer, New York, 1989.
11. R. Elkik, *Fibrés d'intersections et intégrales de classes de Chern*, Ann. Sci. École Norm. Sup. (4) **22** (1989), no. 2, 195–226.
12. B. Feigin, B. Tsygan, *Riemann–Roch theorem and Lie algebra cohomology I*, Proceedings of the Winter School on Geometry and Physics (Srní, 1988). Rend. Circ. Mat. Palermo (2) Suppl. **21** (1989), 15–52.
13. D.B. Fuks, *Cohomology of Infinite-Dimensional Lie Algebras*, Consultants Bureau, New York and London, 1986.
14. V. Hinich, V. Schechtman, *Deformation theory and Lie algebra homology*, I,II, Algebra Colloq. **4** (1997), no. 2, 213–240, 291–316.
15. V. Kac, D. Peterson, *Spin and wedge representations of infinite-dimensional Lie algebras and groups*, Proc. Natl. Acad. Sci. USA, **78** (1981), 3308–3312.
16. V. Kac, A. Radul, *Quasifinite highest weight modules over the Lie algebra of differential operators on the circle*, Comm. Math. Phys. **157** (1993), no. 3, 429–457.
17. R. Kallstrom, *Smooth Modules over Lie Algebroids I*, preprint math. AG/9808108.
18. M. Kapranov, E. Vasserot, *Vertex algebras and the formal loop space*, Publ. Math. Inst. Hautes Études Sci. **100** (2004), 209–269.
19. M. Kapranov, E. Vasserot, *Formal loops II: a local Riemann–Roch theorem for determinantal gerbes*, Ann. Sci. École Norm. Sup. (4) 40 (2007), 113–133.
20. G. Laumon, L. Moret-Bailly, *Champs Algébriques*, Springer-Verlag, Berlin, 2000.
21. J. L. Loday, *Cyclic Homology*, Second Edition, Springer-Verlag, Berlin, 1998.
22. J. Lott, Higher-degree analogs of the determinant line bundle, Comm. Math. Phys. **230** (2002), no. 1, 41–69.
23. K. McKenzie, *General Theory of Lie Groupoids and Lie Algebroids*, London Mathematical Society Lecture Note Series, **213**. Cambridge University Press, Cambridge, 2005.

24. S. Morita, *Geometry of Characteristic Classes*, Translations of Mathematical Monographs, **199**. Iwanami Series in Modern Mathematics, American Mathematical Society, Providence, RI, 2001.
25. R. Nest, B. Tsygan, *Algebraic index theorem for families*, Adv. Math. **113** (1995), no. 2, 151–205.
26. A. Pressley, G.B. Segal, *Loop Groups*, Cambridge University Press, 1986.
27. A.G. Reiman, M.A. Semenov-Tyan-Shanskii, L.D. Faddeev, *Quantum anomalies and cocycles on gauge groups*, Funkt. Anal. Appl. **18** (1984), No. 4, 64–72.
28. J. Tate, *Residues of differentials on curves*, Ann. Sci. École Norm. Sup. (4) **1** (1968) 149–159.
29. M. Wodzicki, *Cyclic homology of differential operators*, Duke Math. J. **54** (1987), no. 2, 641–647.

Universal KZB Equations: The Elliptic Case

Damien Calaque,[1] Benjamin Enriquez,[2] and Pavel Etingof[3]

[1] Université Lyon 1, Institut Camille Jordan (CNRS), F-69622 Villeurbanne, France calaque@math.univ-lyon1.fr
[2] ULP Strasbourg 1 et IRMA (CNRS), 7 rue René Descartes, F-67084 Strasbourg, France enriquez@math.u-strasbg.fr
[3] Department of Mathematics, MIT, Cambridge, MA 02139, USA etingof@math.mit.edu

To Yuri Ivanovich Manin on his 70th birthday

Summary. We define a universal version of the Knizhnik–Zamolodchikov–Bernard (KZB) connection in genus 1. This is a flat connection over a principal bundle on the moduli space of elliptic curves with marked points. It restricts to a flat connection on configuration spaces of points on elliptic curves, which can be used for proving the formality of the pure braid groups on genus 1 surfaces. We study the monodromy of this connection and show that it gives rise to a relation between the KZ associator and a generating series for iterated integrals of Eisenstein forms. We show that the universal KZB connection is realized as the usual KZB connection for simple Lie algebras, and that in the $\mathfrak{sl}_n$ case this realization factors through the Cherednik algebras. This leads us to define a functor from the category of equivariant D-modules on $\mathfrak{sl}_n$ to that of modules over the Cherednik algebra, and to compute the character of irreducible equivariant D-modules over $\mathfrak{sl}_n$ that are supported on the nilpotent cone.

Key words: Knizhnik–Zamolodchikov–Bernard equations, elliptic curve, monodromy.

2000 Mathematics Subject Classifications: 17B37, 32G34, 11F55, 20C08, 52C38

Introduction

The KZ system was introduced in [KZ84] as a system of equations satisfied by correlation functions in conformal field theory. It was then realized that this system has a universal version [Dri91]. The monodromy of this system leads to representations of the braid groups, which can be used for proving that the pure braid groups, which are the fundamental groups of the configuration

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*,
Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_5,

spaces of $\mathbb{C}$, are formal (i.e., their Lie algebras are isomorphic to their associated graded Lie algebras, which is a holonomy Lie algebra and thus has an explicit presentation). This fact was first proved in the framework of minimal model theory [Sul77, Koh83]. These results gave rise to Drinfeld's theory of associators and quasi-Hopf algebras [Dri90b, Dri91]; one of the purposes of this work was to give an algebraic construction of the formality isomorphisms, and indeed, one of its by-products is the fact that these isomorphisms can be defined over $\mathbb{Q}$.

In the case of configuration spaces over surfaces of genus ≥ 1, similar Lie algebra isomorphisms were constructed by Bezrukavnikov [Bez94], using results of Kriz [Kri94]. In this series of papers, we will show that this result can be re-proved using a suitable flat connection over configuration spaces. This connection is a universal version of the KZB connection [Ber98a, Ber98b], which is the higher-genus analogue of the KZ connection.

In this paper, we focus on the case of genus 1. We define the universal KZB connection (Section 1), and rederive from there the formality result (Section 2). As in the integrable case of the KZB connection, the universal KZB connection extends from the configuration spaces $\bar{C}(E_\tau, n)/S_n$ to the moduli space $\mathcal{M}_{1,[n]}$ of elliptic curves with n unordered marked points (Section 3). This means that (a) the connection can be extended to the directions of variation of moduli, and (b) it is modular invariant.

This connection then gives rise to a monodromy morphism $\gamma_n : \Gamma_{1,[n]} \to \mathbf{G}_n \rtimes S_n$, which we analyze in Section 4. The images of most generators can be expressed using the KZ associator, but the image $\tilde{\Theta}$ of the S-transformation can be expressed using iterated integrals of Eisenstein series. The relations between generators give rise to relations between $\tilde{\Theta}$ and the KZ associator, identities (28). This identity may be viewed as an elliptic analogue of the pentagon identity, since it is a "de Rham" analogue of the relation 6AS in [HLS00] (in [Man05], the question was asked of the existence of this kind of identity).

In Section 5, we investigate how to algebraically construct a morphism $\Gamma_{1,[n]} \to \mathbf{G}_n \rtimes S_n$. We show that a morphism $\overline{\mathrm{B}}_{1,n} \to \exp(\hat{\mathfrak{t}}_{1,n}) \rtimes S_n$ can be constructed using an associator only (here $\overline{\mathrm{B}}_{1,n}$ is the reduced braid group of n points on the torus). [Dri91] then implies that the formality isomorphism can be defined over $\mathbb{Q}$. In the last part of Section 5, we develop the analogue of the theory of quasitriangular quasibialgebras (QTQBAs), namely elliptic structures over QTQBAs. These structures give rise to representations of $\overline{\mathrm{B}}_{1,n}$, and they can be modified by twist. We hope that in the case of a simple Lie algebra, and using suitable twists, the elliptic structure given in Section 5.4 will give rise to elliptic structures over the quantum group $U_q(\mathfrak{g})$ (where $q \in \mathbb{C}^\times$) or over the Lusztig quantum group (when q is a root of unity), recovering the representations of $\overline{\mathrm{B}}_{1,n}$ from conformal field theory.

In Section 6, we show that the universal KZB connection indeed specializes to the ordinary KZB connection.

Sections 7–9 are dedicated to applications of the ideas of the preceding sections (in particular, Section 6) to the representation theory of Cherednik algebras.

More precisely, in Section 7, we construct a homomorphism from the Lie algebra $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}$ to the rational Cherednik algebra $H_n(k)$ of type A_{n-1} (here $\mathfrak{d}$ is a Lie algebra introduced in Section 3, which acts on $\bar{\mathfrak{t}}_n$ by derivations). This allows us to consider the elliptic KZB connection with values in representations of the rational Cherednik algebra. The monodromy of this connection then gives representations of the true Cherednik algebra (i.e., the double affine Hecke algebra). In particular, this gives a simple way of constructing an isomorphism between the rational Cherednik algebra and the double affine Hecke algebra, with formal deformation parameters.

In Section 8, we consider the special representation V_N of the rational Cherednik algebra $H_n(k)$, $k = N/n$, for which the elliptic KZB connection is the KZB connection for (holomorphic) n-point correlation functions of the WZW model for $\mathrm{SL}_N(\mathbb{C})$ on the elliptic curve, when the marked points are labeled by the vector representation $\mathbb{C}^N$. This representation is realized in the space of equivariant polynomial functions on $\mathfrak{sl}_N$ with values in $(\mathbb{C}^N)^{\otimes n}$, and we show that it is irreducible, and calculate its character.

In Section 9, we generalize the construction of Section 8 by replacing, in the construction of V_N, the space of polynomial functions on $\mathfrak{sl}_N$ with an arbitrary D-module on $\mathfrak{sl}_N$. This gives rise to an exact functor from the category of (equivariant) D-modules on $\mathfrak{sl}_N$ to the category of representations of $H_n(N/n)$. We study this functor in detail. In particular, we show that this functor maps D-modules concentrated on the nilpotent cone to modules from the category $\mathcal{O}_-$ of highest weight modules over the Cherednik algebra, and is closely related to the Gan–Ginzburg functor [GG04]. Using these facts, we show that it maps irreducible D-modules on the nilpotent cone to irreducible representations of the Cherednik algebra, and determine their highest weights. As an application, we compute the decomposition of cuspidal D-modules into irreducible representations of $\mathrm{SL}_N(\mathbb{C})$. Finally, we describe the generalization of the above result to the trigonometric case (which involves D-modules on the group and trigonometric Cherednik algebras) and point out several directions for generalization.

1 Bundles with flat connections on (reduced) configuration spaces

1.1 The Lie algebras $\mathfrak{t}_{1,n}$ and $\bar{\mathfrak{t}}_{1,n}$

Let $n \geq 1$ be an integer and $\mathbf{k}$ a field of characteristic zero. We define $\mathfrak{t}^{\mathbf{k}}_{1,n}$ as the Lie algebra with generators x_i, y_i $(i = 1, \ldots, n)$ and t_{ij} $(i \neq j \in \{1, \ldots, n\})$ and relations

$$t_{ij} = t_{ji}, \quad [t_{ij}, t_{ik} + t_{jk}] = 0, \quad [t_{ij}, t_{kl}] = 0, \tag{1}$$

$$[x_i, y_j] = t_{ij}, \quad [x_i, x_j] = [y_i, y_j] = 0, \quad [x_i, y_i] = -\sum_{j|j\neq i} t_{ij},$$

$$[x_i, t_{jk}] = [y_i, t_{jk}] = 0, \quad [x_i + x_j, t_{ij}] = [y_i + y_j, t_{ij}] = 0$$

(i, j, k, l are distinct). In this Lie algebra, $\sum_i x_i$ and $\sum_i y_i$ are central; we then define $\bar{\mathfrak{t}}^{\mathbf{k}}_{1,n} := \mathfrak{t}^{\mathbf{k}}_{1,n} / (\sum_i x_i, \sum_i y_i)$. Both $\mathfrak{t}^{\mathbf{k}}_{1,n}$ and $\bar{\mathfrak{t}}^{\mathbf{k}}_{1,n}$ are positively graded, where $\deg(x_i) = \deg(y_i) = 1$.

The symmetric group S_n acts by automorphisms of $\mathfrak{t}^{\mathbf{k}}_{1,n}$ by $\sigma(x_i) := x_{\sigma(i)}$, $\sigma(y_i) := y_{\sigma(i)}$, $\sigma(t_{ij}) := t_{\sigma(i)\sigma(j)}$; this induces an action of S_n by automorphisms of $\bar{\mathfrak{t}}^{\mathbf{k}}_{1,n}$.

We will set $\mathfrak{t}_{1,n} := \mathfrak{t}^{\mathbb{C}}_{1,n}$, $\bar{\mathfrak{t}}_{1,n} := \bar{\mathfrak{t}}^{\mathbb{C}}_{1,n}$ in Sections 1 to 4.

1.2 Bundles with flat connections over $C(E, n)$ and $\bar{C}(E, n)$

Let E be an elliptic curve, $C(E, n)$ the configuration space $E^n - \{\text{diagonals}\}$ ($n \geq 1$) and $\bar{C}(E, n) := C(E, n)/E$ the reduced configuration space. We will define an[4] $\exp(\hat{\bar{\mathfrak{t}}}_{1,n})$-principal bundle with a flat (holomorphic) connection $(\bar{P}_{E,n}, \bar{\nabla}_{E,n}) \to \bar{C}(E, n)$. For this, we define an $\exp(\hat{\mathfrak{t}}_{1,n})$-principal bundle with a flat connection $(P_{E,n}, \nabla_{E,n}) \to C(E, n)$. Its image under the natural morphism $\exp(\hat{\mathfrak{t}}_n) \to \exp(\hat{\bar{\mathfrak{t}}}_n)$ is an $\exp(\hat{\bar{\mathfrak{t}}}_{1,n})$-bundle with connection $(\tilde{P}_{E,n}, \tilde{\nabla}_{E,n}) \to C(E, n)$, and we then prove that $(\tilde{P}_{E,n}, \tilde{\nabla}_{E,n})$ is the pullback of a pair $(\bar{P}_{E,n}, \bar{\nabla}_{E,n})$ under the canonical projection $C(E, n) \to \bar{C}(E, n)$.

For this, we fix a uniformization $E \simeq E_\tau$, where for $\tau \in \mathfrak{H}$, $\mathfrak{H} := \{\tau \in \mathbb{C} | \Im(\tau) > 0\}$, $E_\tau := \mathbb{C}/\Lambda_\tau$, and $\Lambda_\tau := \mathbb{Z} + \mathbb{Z}\tau$. We then have $C(E_\tau, n) = (\mathbb{C}^n - \mathrm{Diag}_{n,\tau})/\Lambda_\tau^n$, where

$$\mathrm{Diag}_{n,\tau} := \{\mathbf{z} = (z_1, \ldots, z_n) \in \mathbb{C}^n \,|\, z_{ij} := z_i - z_j \in \Lambda_\tau \text{ for some } i \neq j\}.$$

We define $P_{\tau,n}$ as the restriction to $C(E_\tau, n)$ of the bundle over $\mathbb{C}^n/\Lambda_\tau^n$ for which a section on $U \subset \mathbb{C}^n/\Lambda_\tau^n$ is a regular map $f : \pi^{-1}(U) \to \exp(\hat{\mathfrak{t}}_{1,n})$, such that[5] $f(\mathbf{z} + \delta_i) = f(\mathbf{z})$, $f(\mathbf{z} + \tau\delta_i) = e^{-2\pi \mathrm{i} x_i} f(\mathbf{z})$ (here $\pi : \mathbb{C}^n \to \mathbb{C}^n/\Lambda_\tau^n$ is the canonical projection and δ_i is the ith vector of the canonical basis of $\mathbb{C}^n$).

The bundle $\tilde{P}_{\tau,n} \to C(E_\tau, n)$ derived from $P_{\tau,n}$ is the pullback of a bundle $\bar{P}_{\tau,n} \to \bar{C}(E_\tau, n)$ since the $e^{-2\pi \mathrm{i} \bar{x}_i} \in \exp(\hat{\bar{\mathfrak{t}}}_{1,n})$ commute pairwise and their product is 1. Here $x \mapsto \bar{x}$ is the map $\hat{\mathfrak{t}}_{1,n} \to \hat{\bar{\mathfrak{t}}}_{1,n}$.

A flat connection $\nabla_{\tau,n}$ on $P_{\tau,n}$ is then the same as an equivariant flat connection over the trivial bundle over $\mathbb{C}^n - \mathrm{Diag}_{n,\tau}$, i.e., a connection of the form

$$\nabla_{\tau,n} := d - \sum_{i=1}^{n} K_i(\mathbf{z}|\tau) \mathrm{d}z_i,$$

[4] We will denote by $\hat{\mathfrak{g}}$ or $\mathfrak{g}^\wedge$ the degree completion of a positively graded Lie algebra $\mathfrak{g}$.

[5] We set $\mathrm{i} := \sqrt{-1}$, leaving i for indices.

where $K_i(-|\tau) : \mathbb{C}^n \to \hat{\mathfrak{t}}_{1,n}$ is holomorphic on $\mathbb{C}^n - \mathrm{Diag}_{n,\tau}$ such that:

(a) $K_i(\mathbf{z} + \delta_j|\tau) = K_i(\mathbf{z}|\tau)$, $K_i(\mathbf{z} + \tau\delta_j|\tau) = e^{-2\pi\mathrm{i}\,\mathrm{ad}(x_j)}(K_i(\mathbf{z}|\tau))$,

(b) $[\partial/\partial z_i - K_i(\mathbf{z}|\tau), \partial/\partial z_j - K_j(\mathbf{z}|\tau)] = 0$ for any i, j.

Then $\nabla_{\tau,n}$ induces a flat connection $\tilde{\nabla}_{\tau,n}$ on $\tilde{P}_{\tau,n}$. Then $\tilde{\nabla}_{\tau,n}$ is the pull-back of a (necessarily flat) connection on $\bar{P}_{\tau,n}$ iff

(c) $\bar{K}_i(\mathbf{z}|\tau) = \bar{K}_i(\mathbf{z}+u(\sum_i \delta_i)|\tau)$ and $\sum_i \bar{K}_i(\mathbf{z}|\tau) = 0$ for $\mathbf{z} \in \mathbb{C}^n - \mathrm{Diag}_{n,\tau}$, $u \in \mathbb{C}$.

In order to define the $K_i(\mathbf{z}|\tau)$, we first recall some facts on theta functions. There is a unique holomorphic function $\mathbb{C} \times \mathfrak{H} \to \mathbb{C}$, $(z, \tau) \mapsto \theta(z|\tau)$, such that

- $\{z|\theta(z|\tau) = 0\} = \Lambda_\tau$,
- $\theta(z + 1|\tau) = -\theta(z|\tau) = \theta(-z|\tau)$,
- $\theta(z + \tau|\tau) = -e^{-\pi\mathrm{i}\tau}e^{-2\pi\mathrm{i}z}\theta(z|\tau)$, and
- $\theta_z(0|\tau) = 1$.

We have $\theta(z|\tau + 1) = \theta(z|\tau)$, while $\theta(-z/\tau| - 1/\tau) = -(1/\tau)e^{(\pi\mathrm{i}/\tau)z^2}\theta(z|\tau)$. If $\eta(\tau) = q^{1/24}\prod_{n\geq 1}(1 - q^n)$, where $q = e^{2\pi\mathrm{i}\tau}$, and if we set $\vartheta(z|\tau) := \eta(\tau)^3\theta(z|\tau)$, then $\partial_\tau\vartheta = (1/4\pi\mathrm{i})\partial_z^2\vartheta$.

Let us set

$$k(z, x|\tau) := \frac{\theta(z + x|\tau)}{\theta(z|\tau)\theta(x|\tau)} - \frac{1}{x}.$$

When τ is fixed, $k(z, x|\tau)$ belongs to $\mathrm{Hol}(\mathbb{C} - \Lambda_\tau)[[x]]$. Substituting $x = \mathrm{ad}\, x_i$, we get a linear map $\mathfrak{t}_{1,n} \to (\mathfrak{t}_{1,n} \otimes \mathrm{Hol}(\mathbb{C} - \Lambda_\tau))^\wedge$, and taking the image of t_{ij}, we define

$$K_{ij}(z|\tau) := k(z, \mathrm{ad}\, x_i|\tau)(t_{ij}) = \left(\frac{\theta(z + \mathrm{ad}(x_i)|\tau)}{\theta(z|\tau)}\frac{\mathrm{ad}(x_i)}{\theta(\mathrm{ad}(x_i)|\tau)} - 1\right)(y_j);$$

it is a holomorphic function on $\mathbb{C} - \Lambda_\tau$ with values in $\hat{\mathfrak{t}}_{1,n}$.

Now set $\mathbf{z} := (z_1, \ldots, z_n)$, $z_{ij} := z_i - z_j$ and define

$$K_i(\mathbf{z}|\tau) := -y_i + \sum_{j|j\neq i} K_{ij}(z_{ij}|\tau).$$

Let us check that the $K_i(\mathbf{z}|\tau)$ satisfy condition (c). We have clearly $K_i(\mathbf{z}+u(\sum_i \delta_i)) = K_i(\mathbf{z})$. We have $k(z, x|\tau)+k(-z, -x|\tau) = 0$, so $K_{ij}(z|\tau)+K_{ji}(-z|\tau) = 0$, so that $\sum_i K_i(\mathbf{z}|\tau) = -\sum_i y_i$, which implies $\sum_i \bar{K}_i(\mathbf{z}|\tau) = 0$.

Lemma 1. *$K_i(\mathbf{z} + \delta_j|\tau) = K_i(\mathbf{z}|\tau)$ and $K_i(\mathbf{z} + \tau\delta_j|\tau) = e^{-2\pi\mathrm{i}\,\mathrm{ad}\, x_j}(K_i(\mathbf{z}|\tau))$, i.e., the $K_i(\mathbf{z}|\tau)$ satisfy condition (a).*

Proof. We have $k(z \pm 1, x|\tau) = k(z, x|\tau)$ so for any j, $K_i(\mathbf{z} + \delta_j|\tau) = K_i(\mathbf{z}|\tau)$. We have $k(z \pm \tau, x|\tau) = e^{\mp 2\pi\mathrm{i}x}k(z, x|\tau) + (e^{\mp 2\pi\mathrm{i}x} - 1)/x$, so if $j \neq i$,

$$K_i(\mathbf{z}+\tau\delta_j|\tau) = \sum_{j'\neq i,j} K_{ij'}(z_{ij'}|\tau) + e^{2\pi\mathrm{i}\,\mathrm{ad}\, x_i}K_{ij}(z_{ij}|\tau) + \frac{e^{2\pi\mathrm{i}\,\mathrm{ad}\, x_i} - 1}{\mathrm{ad}\, x_i}(t_{ij}) - y_i.$$

Then

$$\frac{e^{2\pi \mathrm{i} \operatorname{ad} x_i}-1}{\operatorname{ad} x_i}(t_{ij}) = \frac{1-e^{-2\pi \mathrm{i} \operatorname{ad} x_j}}{\operatorname{ad} x_j}(t_{ij}) = (1-e^{-2\pi \mathrm{i} \operatorname{ad} x_j})(y_i),$$

$e^{2\pi \mathrm{i} \operatorname{ad} x_i}(K_{ij}(z_{ij}|\tau)) = e^{-2\pi \mathrm{i} \operatorname{ad} x_j}(K_{ij}(z_{ij}|\tau))$ and for $j' \neq i,j$, $K_{ij'}(z_{ij'}|\tau) = e^{-2\pi \mathrm{i} \operatorname{ad} x_j}(K_{ij'}(z_{ij'}|\tau))$, so $K_i(\mathbf{z}+\tau\delta_j|\tau) = e^{-2\pi \mathrm{i} \operatorname{ad} x_j}(K_i(\mathbf{z}|\tau))$. Now

$$\begin{aligned} K_i(\mathbf{z}+\tau\delta_i|\tau) &= -\sum_i y_i - \sum_{j|j\neq i} K_j(\mathbf{z}+\tau\delta_i|\tau) \\ &= -\sum_i y_i - e^{-2\pi \mathrm{i} \operatorname{ad} x_i}\left(\sum_{j|j\neq i} K_j(\mathbf{z}|\tau)\right) \\ &= e^{-2\pi \mathrm{i} \operatorname{ad} x_i}\left(-\sum_i y_i - \sum_{j|j\neq i} K_j(\mathbf{z}|\tau)\right) \\ &= e^{-2\pi \mathrm{i} \operatorname{ad} x_i} K_i(\mathbf{z}|\tau) \end{aligned}$$

(the first and last equalities follow from the proof of (c), the second equality has just been proved, and the third equality follows from the centrality of $\sum_i y_i$). □

Proposition 2. $[\partial/\partial z_i - K_i(\mathbf{z}|\tau), \partial/\partial z_j - K_j(\mathbf{z}|\tau)] = 0$, *i.e., the* $K_i(\mathbf{z}|\tau)$ *satisfy condition (b).*

Proof. For $i \neq j$, let us set $K_{ij} := K_{ij}(z_{ij}|\tau)$. Recall that $K_{ij} + K_{ji} = 0$, and therefore if $\partial_i := \partial/\partial z_i$, then

$$\partial_i K_{ij} - \partial_j K_{ji} = 0, \quad [y_i - K_{ij}, y_j - K_{ji}] = -[K_{ij}, y_i + y_j].$$

Moreover, if i,j,k,l are distinct, then $[K_{ik}, K_{jl}] = 0$. It follows that if $i \neq j$, then $[\partial_i - K_i(\mathbf{z}|\tau), \partial_j - K_j(\mathbf{z}|\tau)]$ equals

$$[y_i+y_j, K_{ij}] + \sum_{k|k\neq i,j} \big([K_{ik}, K_{jk}] + [K_{ij}, K_{jk}] + [K_{ij}, K_{ik}] + [y_j, K_{ik}] - [y_i, K_{jk}]\big).$$

Let us assume for a while that if $k \notin \{i,j\}$, then

$$-[y_i, K_{jk}] - [y_j, K_{ki}] - [y_k, K_{ij}] + [K_{ji}, K_{ki}] + [K_{kj}, K_{ij}] + [K_{ik}, K_{jk}] = 0 \quad (2)$$

(this is the universal version of the classical dynamical Yang–Baxter equation).

Then (2) implies that

$$[\partial_i - K_i(\mathbf{z}|\tau), \partial_j - K_j(\mathbf{z}|\tau)] = [y_i + y_j, K_{ij}] + \sum_{k|k\neq i,j} [y_k, K_{ij}] = [\sum_k y_k, K_{ij}] = 0$$

(since $\sum_k y_k$ is central), which proves the proposition.

Let us now prove (2). If $f(x) \in \mathbb{C}[[x]]$, then

$$
\begin{aligned}
[y_k, f(\mathrm{ad}\, x_i)(t_{ij})] &= \frac{f(\mathrm{ad}\, x_i) - f(-\mathrm{ad}\, x_j)}{\mathrm{ad}\, x_i + \mathrm{ad}\, x_j}[-t_{ki}, t_{ij}], \\
[y_i, f(\mathrm{ad}\, x_j)(t_{jk})] &= \frac{f(\mathrm{ad}\, x_j) - f(-\mathrm{ad}\, x_k)}{\mathrm{ad}\, x_j + \mathrm{ad}\, x_k}[-t_{ij}, t_{jk}] \\
&= \frac{f(\mathrm{ad}\, x_j) - f(\mathrm{ad}\, x_i + \mathrm{ad}\, x_j)}{-\mathrm{ad}\, x_i}[-t_{ij}, t_{jk}], \\
[y_j, f(\mathrm{ad}\, x_k)(t_{ki})] &= \frac{f(\mathrm{ad}\, x_k) - f(-\mathrm{ad}\, x_i)}{\mathrm{ad}\, x_k + \mathrm{ad}\, x_i}[-t_{jk}, t_{ki}] \\
&= \frac{f(-\mathrm{ad}\, x_i - \mathrm{ad}\, x_j) - f(-\mathrm{ad}\, x_i)}{-\mathrm{ad}\, x_j}[-t_{jk}, t_{ki}].
\end{aligned}
$$

The first identity is proved as follows:

$$
\begin{aligned}
[y_k, (\mathrm{ad}\, x_i)^n(t_{ij})] &= -\sum_{s=0}^{n-1} (\mathrm{ad}\, x_i)^s (\mathrm{ad}\, t_{ki})(\mathrm{ad}\, x_i)^{n-1-s}(t_{ij}) \\
&= -\sum_{s=0}^{n-1} (\mathrm{ad}\, x_i)^s (\mathrm{ad}\, t_{ki})(-\mathrm{ad}\, x_j)^{n-1-s}(t_{ij}) \\
&= -\sum_{s=0}^{n-1} (\mathrm{ad}\, x_i)^s (-\mathrm{ad}\, x_j)^{n-1-s}(\mathrm{ad}\, t_{ki})(t_{ij}) \\
&= f(\mathrm{ad}\, x_i, -\mathrm{ad}\, x_j)([-t_{ki}, t_{ij}]),
\end{aligned}
$$

where $f(u,v) = (u^n - v^n)/(u-v)$. The two next identities follow from this one and from the fact that $x_i + x_j + x_k$ commutes with t_{ij}, t_{ik}, t_{jk}.

Then, if we write $k(z,x)$ instead of $k(z,x|\tau)$, the l.h.s. of (2) is equal to

$$
\begin{aligned}
&\Bigg(k(z_{ij}, -\mathrm{ad}\, x_j)k(z_{ik}, \mathrm{ad}\, x_i + \mathrm{ad}\, x_j) - k(z_{ij}, \mathrm{ad}\, x_i)k(z_{jk}, \mathrm{ad}\, x_i + \mathrm{ad}\, x_j) \\
&\quad + k(z_{ik}, \mathrm{ad}\, x_i)k(z_{jk}, \mathrm{ad}\, x_j) + \frac{k(z_{jk}, \mathrm{ad}\, x_j) - k(z_{jk}, \mathrm{ad}\, x_i + \mathrm{ad}\, x_j)}{\mathrm{ad}\, x_i} \\
&\quad + \frac{k(z_{ik}, \mathrm{ad}\, x_i) - k(z_{ij}, \mathrm{ad}\, x_i + \mathrm{ad}\, x_j)}{\mathrm{ad}\, x_j} - \frac{k(z_{ij}, \mathrm{ad}\, x_i) - k(z_{ij}, -\mathrm{ad}\, x_j)}{\mathrm{ad}\, x_i + \mathrm{ad}\, x_j} \Bigg) \\
&\quad [t_{ij}, t_{ik}].
\end{aligned}
$$

So (2) follows from the identity

$$
\begin{aligned}
&k(z,-v)k(z',u+v) - k(z,u)k(z'-z,u+v) + k(z',u)k(z'-z,v) \\
&\quad + \frac{k(z'-z,v) - k(z'-z,u+v)}{u} + \frac{k(z',u) - k(z',u+v)}{v} \\
&\quad - \frac{k(z,u) - k(z,-v)}{u+v} = 0,
\end{aligned}
$$

where u, v are formal variables, which is a consequence of the theta-functions identity

$$\begin{aligned}\left(k(z,-v)-\frac{1}{v}\right)\left(k(z',u+v)+\frac{1}{u+v}\right)\\ -\left(k(z,u)+\frac{1}{u}\right)\left(k(z'-z,u+v)+\frac{1}{u+v}\right)\\ +\left(k(z',u)+\frac{1}{u}\right)\left(k(z'-z,v)+\frac{1}{v}\right)=0.\end{aligned} \tag{3}$$

□

We have therefore proved:

Theorem 3. $(P_{\tau,n},\nabla_{\tau,n})$ *is a flat connection on* $C(E_\tau,n)$*, and the induced flat connection* $(\tilde{P}_{\tau,n},\tilde{\nabla}_{\tau,n})$ *is the pullback of a unique flat connection* $(\bar{P}_{\tau,n},\bar{\nabla}_{\tau,n})$ *on* $\bar{C}(E_\tau,n)$*.*

1.3 Bundles with flat connections on $C(E,n)/S_n$ and $\bar{C}(E,n)/S_n$

The group S_n acts freely by automorphisms of $C(E,n)$ by $\sigma(z_1,\dots,z_n) := (z_{\sigma^{-1}(1)},\dots,z_{\sigma^{-1}(n)})$. This descends to a free action of S_n on $\bar{C}(E,n)$. We set $C(E,[n]) := C(E,n)/S_n$, $\bar{C}(E,[n]) := \bar{C}(E,n)/S_n$.

We will show that $(P_{\tau,n},\nabla_{\tau,n})$ induces a bundle with flat connection $(P_{\tau,[n]},\nabla_{\tau,[n]})$ on $C(E_\tau,[n])$ with group $\exp(\hat{\mathfrak{t}}_{1,n}) \rtimes S_n$, and similarly $(\bar{P}_{\tau,n},\bar{\nabla}_{\tau,n})$ induces $(\bar{P}_{\tau,[n]},\bar{\nabla}_{\tau,[n]})$ on $\bar{C}(E_\tau,[n])$ with group $\exp(\hat{\bar{\mathfrak{t}}}_{1,n}) \rtimes S_n$.

We define $P_{\tau,[n]} \to C(E_\tau,[n])$ by the condition that a section of $U \subset C(E_\tau,[n])$ is a regular map $\pi^{-1}(U) \to \exp(\hat{\mathfrak{t}}_{1,n}) \rtimes S_n$, satisfying again

$f(\mathbf{z}+\delta_i) = f(\mathbf{z})$, $f(\mathbf{z}+\tau\delta_i) = e^{-2\pi \mathrm{i} x_i} f(\mathbf{z})$ and the additional requirement $f(\sigma\mathbf{z}) = \sigma f(\mathbf{z})$ (where $\tilde{\pi} : \mathbb{C}^n - \mathrm{Diag}_{\tau,n} \to C(E_\tau,[n])$ is the canonical projection). It is clear that $\nabla_{\tau,n}$ is S_n-invariant, which implies that it defines a flat connection $\nabla_{\tau,[n]}$ on $C(E_\tau,[n])$.

The bundle $\bar{P}(E_\tau,[n]) \to \bar{C}(E_\tau,[n])$ is defined by the additional requirement $f(\mathbf{z}+u(\sum_i \delta_i)) = f(\mathbf{z})$ and $\bar{\nabla}_{\tau,n}$ then induces a flat connection $\bar{\nabla}_{\tau,[n]}$ on $\bar{C}(E_\tau,[n])$.

2 Formality of pure braid groups on the torus

2.1 Reminders on Malcev Lie algebras

Let $\mathbf{k}$ be a field of characteristic 0 and let $\mathfrak{g}$ be a pronilpotent $\mathbf{k}$-Lie algebra. Set $\mathfrak{g}^1 = \mathfrak{g}$, $\mathfrak{g}^{k+1} = [\mathfrak{g},\mathfrak{g}^k]$; then $\mathfrak{g} = \mathfrak{g}^1 \supset \mathfrak{g}^2 \cdots$ is a decreasing filtration of $\mathfrak{g}$. The associated graded Lie algebra is $\mathrm{gr}(\mathfrak{g}) := \oplus_{k\geq 1}\mathfrak{g}^k/\mathfrak{g}^{k+1}$; we also consider its completion $\hat{\mathrm{gr}}(\mathfrak{g}) := \hat{\oplus}_{k\geq 1}\mathfrak{g}^k/\mathfrak{g}^{k+1}$ (here $\hat{\oplus}$ is the direct product).

We say that $\mathfrak{g}$ is formal if there exists an isomorphism of filtered Lie algebras $\mathfrak{g} \simeq \hat{\mathrm{gr}}(\mathfrak{g})$, whose associated graded morphism is the identity. We will use the following fact: if $\mathfrak{g}$ is a pronilpotent Lie algebra, $\mathfrak{t}$ is a positively graded Lie algebra, and there exists an isomorphism $\mathfrak{g} \simeq \hat{\mathfrak{t}}$ of filtered Lie algebras, then $\mathfrak{g}$ is formal and the associated graded morphism $\mathrm{gr}(\mathfrak{g}) \to \mathfrak{t}$ is an isomorphism of graded Lie algebras.

If Γ is a finitely generated group, there exists a unique pair $(\Gamma(\mathbf{k}), i_\Gamma)$ of a prounipotent algebraic group $\Gamma(\mathbf{k})$ and a group morphism $i_\Gamma : \Gamma \to \Gamma(\mathbf{k})$, which is initial in the category of all pairs (U, j), where U is a prounipotent $\mathbf{k}$-algebraic group and $j : \Gamma \to U$ is a group morphism.

We denote by $\mathrm{Lie}(\Gamma)_\mathbf{k}$ the Lie algebra of $\Gamma(\mathbf{k})$. Then we have $\Gamma(\mathbf{k}) = \exp(\mathrm{Lie}(\Gamma)_\mathbf{k})$; $\mathrm{Lie}(\Gamma)_\mathbf{k}$ is a pronilpotent Lie algebra. We have $\mathrm{Lie}(\Gamma)_\mathbf{k} = \mathrm{Lie}(\Gamma)_\mathbb{Q} \otimes \mathbf{k}$. We say that Γ is formal iff $\mathrm{Lie}(\Gamma)_\mathbb{C}$ is formal (one can show that this implies that $\mathrm{Lie}(\Gamma)_\mathbb{Q}$ is formal).

When Γ is presented by generators $g_1, \ldots, g_n$ and relations $R_i(g_1, \ldots, g_n)$ $(i = 1, \ldots, p)$, $\mathrm{Lie}(\Gamma)_\mathbb{Q}$ is the quotient of the topologically free Lie algebra $\hat{\mathfrak{f}}_n$ generated by $\gamma_1, \ldots, \gamma_n$ by the topological ideal generated by $\log(R_i(e^{\gamma_1}, \ldots, e^{\gamma_n}))$ $(i = 1, \ldots, p)$.

The decreasing filtration of $\hat{\mathfrak{f}}_n$ is $\hat{\mathfrak{f}}_n = (\hat{\mathfrak{f}}_n)^1 \supset (\hat{\mathfrak{f}}_n)^2 \supset \cdots$, where $(\hat{\mathfrak{f}}_n)^k$ is the part of $\hat{\mathfrak{f}}_n$ of degree $\geq k$ in the generators $\gamma_1, \ldots, \gamma_n$. The image of this filtration by the projection map is the decreasing filtration $\mathrm{Lie}(\Gamma)_\mathbb{Q} = \mathrm{Lie}(\Gamma)_\mathbb{Q}^1 \supset \mathrm{Lie}(\Gamma)_\mathbb{Q}^2 \supset \cdots$ of $\mathrm{Lie}(\Gamma)_\mathbb{Q}$.

2.2 Presentation of $\mathbf{PB}_{1,n}$

For $\tau \in \mathfrak{H}$, let $U_\tau \subset \mathbb{C}^n - \mathrm{Diag}_{n,\tau}$ be the open subset of all $\mathbf{z} = (z_1, \ldots, z_n)$ of the form $z_i = a_i + \tau b_i$, where $0 < a_1 < \cdots < a_n < 1$ and $0 < b_1 < \cdots < b_n < 1$. If $\mathbf{z}_0 = \left(z_1^0, \ldots, z_n^0\right) \in U_\tau$, its image $\bar{\mathbf{z}}_0$ in E_τ^n actually belongs to the configuration space $C(E_\tau, n)$.

The pure braid group of n points on the torus $\mathrm{PB}_{1,n}$ may be viewed as $\mathrm{PB}_{1,n} = \pi_1(C(E_\tau, n), \bar{\mathbf{z}}_0)$. Denote by $X_i, Y_i \in \mathrm{PB}_{1,n}$ the classes of the projection of the paths $[0,1] \ni t \mapsto \mathbf{z}_0 - t\delta_i$ and $[0,1] \ni t \mapsto \mathbf{z}_0 - t\tau\delta_i$.

Set $A_i := X_i \cdots X_n$, $B_i := Y_i \cdots Y_n$ for $i = 1, \ldots, n$. According to [Bir69a], A_i, B_i $(i = 1, \ldots, n)$ generate $\mathrm{PB}_{1,n}$ and a presentation of $\mathrm{PB}_{1,n}$ is, in terms of these generators:

$$(A_i, A_j) = (B_i, B_j) = 1 \text{ (any } i, j), \quad (A_1, B_j) = (B_1, A_j) = 1 \text{ (any } j),$$
$$(B_k, A_k A_j^{-1}) = (B_k B_j^{-1}, A_k) = C_{jk} \ (j \leq k);$$
$$(A_i, C_{jk}) = (B_i, C_{jk}) = 1 \ (i \leq j \leq k),$$

where $(g, h) = ghg^{-1}h^{-1}$.

2.3 Alternative presentations of $\mathfrak{t}_{1,n}$

We now give two variants of the defining presentation of $\mathfrak{t}_{1,n}$. Presentation (A) below is the original presentation in [Bez94], and presentation (B) will be suited to the comparison with the above presentation of $\mathrm{PB}_{1,n}$.

Lemma 4. $\mathfrak{t}_{1,n}$ *admits the following presentations:*

(A) generators are x_i, y_i $(i = 1, \dots, n)$, *relations are* $[x_i, y_j] = [x_j, y_i]$ $(i \neq j)$, $[x_i, x_j] = [y_i, y_j] = 0$ *(any* i, j*)*, $[\sum_j x_j, y_i] = [\sum_j y_j, x_i] = 0$ *(any* i*)*, $[x_i, [x_j, y_k]] = [y_i, [y_j, x_k]] = 0$ *(*i, j, k *are distinct);*

(B) generators are a_i, b_i $(i = 1, \dots, n)$, *relations are* $[a_i, a_j] = [b_i, b_j] = 0$ *(any* i, j*)*, $[a_1, b_j] = [b_1, a_j] = 0$ *(any* j*)*, $[a_j, b_k] = [a_k, b_j]$ *(any* i, j*)*, $[a_i, c_{jk}] = [b_i, c_{jk}] = 0$ $(i \leq j \leq k)$, *where* $c_{jk} = [b_k, a_k - a_j]$.

The isomorphism of presentations (A) and (B) is $a_i = \sum_{j=i}^n x_j$, $b_i = \sum_{j=i}^n y_j$.

Proof. Let us prove that the initial relations for x_i, y_i, t_{ij} imply the relations (A) for x_i, y_i. Let us assume the initial relations. If $i \neq j$, since $[x_i, y_j] = t_{ij}$ and $t_{ij} = t_{ji}$, we get $[x_i, y_j] = [x_j, y_i]$. The relations $[x_i, x_j] = [y_i, y_j] = 0$ (any i, j) are contained in the initial relations. For any i, since $[x_i, y_i] = -\sum_{j|j\neq i} t_{ij}$ and $[x_j, y_i] = t_{ji} = t_{ij}$ $(j \neq i)$, we get $[\sum_j x_j, y_i] = 0$. Similarly, $[\sum_j y_j, x_i] = 0$ (for any i). If i, j, k are distinct, since $[x_j, y_k] = t_{jk}$ and $[x_i, t_{jk}] = 0$, we get $[x_i, [x_j, y_k]] = 0$, and similarly we prove $[x_i, [y_j, x_k]] = 0$.

Let us now prove that the relations (A) for x_i, y_i imply the initial relations for x_i, y_i and $t_{ij} := [x_i, y_j]$ $(i \neq j)$. Assume the relations (A). If $i \neq j$, since $[x_i, y_j] = [x_j, y_i]$, we have $t_{ij} = t_{ji}$. The relation $t_{ij} = [x_i, y_j]$ $(i \neq j)$ is clear and $[x_i, x_j] = [y_i, y_j] = 0$ (any i, j) are already in relations (A). Since for any i, $[\sum_j x_j, y_i] = 0$, we get $[x_i, y_i] = -\sum_{j|j\neq i}[x_j, y_i] = -\sum_{j|j\neq i} t_{ji} = -\sum_{j|j\neq i} t_{ij}$. If i, j, k are distinct, the relations $[x_i, [x_j, y_k]] = [y_i, [y_j, x_k]] = 0$ imply $[x_i, t_{jk}] = [y_i, t_{jk}] = 0$. If $i \neq j$, since $[\sum_k x_k, x_i] = [\sum_k x_k, y_j] = 0$, we get $[\sum_k x_k, t_{ij}] = 0$, and $[x_k, t_{ij}] = 0$ for $k \notin \{i, j\}$ then implies $[x_i + x_j, t_{ij}] = 0$. One proves similarly $[y_i + y_j, t_{ij}] = 0$. We have already shown that $[x_i, t_{kl}] = [y_j, t_{kl}] = 0$ for i, j, k, l distinct, which implies $[[x_i, y_j], t_{kl}] = 0$, i.e., $[t_{ij}, t_{kl}] = 0$. If i, j, k are distinct, we have shown that $[t_{ij}, y_k] = 0$ and $[t_{ij}, x_i + x_j] = 0$, which implies $[t_{ij}, [x_i + x_j, y_k]] = 0$, i.e., $[t_{ij}, t_{ik} + t_{jk}] = 0$.

Let us prove that the relations (A) for x_i, y_i imply relations (B) for $a_i := \sum_{j=i}^n x_j$, $b_i := \sum_{j=i}^n y_j$. Summing up the relations $[x_{i'}, x_{j'}] = [y_{i'}, y_{j'}] = 0$ and $[x_{i'}, y_{j'}] = [x_{j'}, y_{i'}]$ for $i' = i, \dots, n$ and $j' = j, \dots, n$, we get $[a_i, a_j] = [b_i, b_j] = 0$ and $[a_i, b_j] = [a_j, b_i]$ (for any i, j). Summing up $[\sum_j x_j, y_{i'}] = [\sum_j y_j, x_{i'}] = 0$ for $i' = i, \dots, n$, we get $[a_1, b_i] = [a_i, b_1] = 0$ (for any i). Finally, $c_{jk} = \sum_{\alpha=j}^{k-1} \sum_{\beta=k}^n t_{\alpha\beta}$ (in terms of the initial presentation), so the relations $[x_{i'}, t_{\alpha\beta}] = 0$ for $i' \neq \alpha, \beta$ and $[x_\alpha + x_\beta, t_{\alpha\beta}] = 0$ imply $[a_i, c_{jk}] = 0$ for $i \leq j \leq k$. Similarly, one shows that $[b_i, c_{jk}] = 0$ for $i \leq j \leq k$.

Let us prove that the relations (B) for a_i, b_i imply relations (A) for $x_i := a_i - a_{i+1}$, $y_i := b_i - b_{i+1}$ (with the convention $a_{n+1} = b_{n+1} = 0$). As before,

$[a_i, a_j] = [b_i, b_j] = 0$, $[a_i, b_j] = [a_j, b_i]$ imply $[x_i, x_j] = [y_i, y_j] = 0$, $[x_i, y_j] = [x_j, y_i]$ (for any i, j). We set $t_{ij} := [x_i, y_j]$ for $i \neq j$, then we have $t_{ij} = t_{ji}$. We have for $j < k$, $t_{jk} = c_{jk} - c_{j,k+1} - c_{j+1,k} + c_{j+1,k+1}$ (we set $c_{i,n+1} := 0$), so $[a_i, c_{jk}] = 0$ implies $[\sum_{i'=i}^{n} x_{i'}, t_{jk}] = 0$ for $i \le j < k$. When $i < j < k$, the difference between this relation and its analogue for $(i+1, j, k)$ gives $[x_i, t_{jk}] = 0$ for $i < j < k$. This can be rewritten $[x_i, [x_j, y_k]] = 0$, and since $[x_i, x_j] = 0$, we get $[x_j, [x_i, y_k]] = 0$, so $[x_j, t_{ik}] = 0$, and by changing indices, $[x_i, t_{jk}] = 0$ for $j < i < k$. Rewriting again $[x_i, t_{jk}] = 0$ for $i < j < k$ as $[x_i, [y_j, x_k]] = 0$ and using $[x_i, x_k] = 0$, we get $[x_k, [x_i, y_j]] = 0$, i.e., $[x_k, t_{ij}] = 0$, which we rewrite $[x_i, t_{jk}] = 0$ for $j < k < i$. Finally, $[x_i, t_{jk}] = 0$ for $j < k$ and $i \notin \{j, k\}$, which implies $[x_i, t_{jk}] = 0$ for i, j, k distinct. One proves similarly $[y_i, t_{jk}] = 0$ for i, j, k distinct. □

2.4 The formality of $\mathbf{PB}_{1,n}$

The flat connection $d - \sum_{i=1}^{n} K_i(\mathbf{z}|\tau) dz_i$ gives rise to a monodromy representation

$$\mu_{\mathbf{z}_0,\tau} : \mathrm{PB}_{1,n} = \pi_1(C, \overline{\mathbf{z}}_0) \to \exp(\hat{\mathfrak{t}}_{1,n}),$$

which factors through a morphism $\mu_{\mathbf{z}_0,\tau}(\mathbb{C}) : \mathrm{PB}_{1,n}(\mathbb{C}) \to \exp(\hat{\mathfrak{t}}_{1,n})$. Let $\mathrm{Lie}(\mu_{\mathbf{z}_0,\tau}) : \mathrm{Lie}(\mathrm{PB}_{1,n})_{\mathbb{C}} \to \hat{\mathfrak{t}}_{1,n}$ be the corresponding morphism between pronilpotent Lie algebras.

Proposition 5. $\mathrm{Lie}(\mu_{\mathbf{z}_0,\tau})$ *is an isomorphism of filtered Lie algebras, so that* $\mathrm{PB}_{1,n}$ *is formal.*

Proof. As we have seen, $\mathrm{Lie}(\mathrm{PB}_{1,n})_{\mathbb{C}}$ (denoted by $\mathrm{Lie}(\mathrm{PB}_{1,n})$ in this proof) is the quotient of the topologically free Lie algebra generated by α_i, β_i $(i = 1, \ldots, n)$ by the topological ideal generated by $[\alpha_i, \alpha_j]$, $[\beta_i, \beta_j]$, $[\alpha_1, \beta_j]$, $[\beta_1, \alpha_j]$, $\log(e^{\beta_k}, e^{\alpha_k - \alpha_j}) - \log(e^{\beta_k - \beta_j}, e^{\alpha_k})$, $[\alpha_i, \gamma_{jk}]$, $[\beta_i, \gamma_{jk}]$ where $\gamma_{jk} = \log(e^{\beta_k}, e^{\alpha_k - \alpha_j})$.

This presentation and the above presentation (B) of $\mathfrak{t}_{1,n}$ imply that there is a morphism of graded Lie algebras $p_n : \mathfrak{t}_{1,n} \to \mathrm{grLie}(\mathrm{PB}_{1,n})$ defined by $a_i \mapsto [\alpha_i]$, $b_i \mapsto [\beta_i]$, where $\alpha \mapsto [\alpha]$ is the projection map $\mathrm{Lie}(\mathrm{PB}_{1,n}) \to \mathrm{gr}_1\mathrm{Lie}(\mathrm{PB}_{1,n})$.

The morphism p_n is surjective because $\mathrm{grLie}\Gamma$ is generated in degree 1 (as the associated graded of any quotient of a topologically free Lie algebra).

There is a unique derivation $\tilde{\Delta}_0 \in \mathrm{Der}(\mathfrak{t}_{1,n})$, such that $\tilde{\Delta}_0(x_i) = y_i$ and $\tilde{\Delta}_0(y_i) = 0$. This derivation gives rise to a one-parameter group of automorphisms of $\mathrm{Der}(\mathfrak{t}_{1,n})$, defined by $\exp(s\tilde{\Delta}_0)(x_i) := x_i + s y_i$, $\exp(s\tilde{\Delta}_0)(y_i) = y_i$.

$\mathrm{Lie}(\mu_{\mathbf{z}_0,\tau})$ induces a morphism $\mathrm{grLie}(\mu_{\mathbf{z}_0,\tau}) : \mathrm{grLie}(\mathrm{PB}_{1,n}) \to \mathfrak{t}_{1,n}$. We will now prove that

$$\mathrm{grLie}(\mu_{\mathbf{z}_0,\tau}) \circ p_n = \exp\left(-\frac{\tau}{2\pi \mathrm{i}} \tilde{\Delta}_0\right) \circ w, \tag{4}$$

where w is the automorphism of $\mathfrak{t}_{1,n}$ defined by $w(a_i) = -b_i$, $w(b_i) = 2\pi\mathrm{i} a_i$.

Then $\mu_{\mathbf{z}_0,\tau}$ is defined as follows. Let $F_{\mathbf{z}_0}(\mathbf{z})$ be the solution of

$$(\partial/\partial z_i)F_{\mathbf{z}_0}(\mathbf{z}) = K_i(\mathbf{z}|\tau)F_{\mathbf{z}_0}(\mathbf{z}), \ \ F_{\mathbf{z}_0}(\mathbf{z}_0) = 1$$

on U_τ; let

$$H_\tau := \{\mathbf{z} = (z_1, \dots, z_n) \,|\, z_i = a_i + \tau b_i, 0 < a_1 < \cdots < a_n < 1\}$$

and

$$V_\tau := \{\mathbf{z} = (z_1, \dots, z_n) \,|\, z_i = a_i + \tau b_i, 0 < b_1 < \cdots < b_n < 1\};$$

let $F^H_{\mathbf{z}_0}$ and $F^V_{\mathbf{z}_0}$ be the analytic prolongations of $F_{\mathbf{z}_0}$ to H_τ and V_τ; then

$$F^H_{\mathbf{z}_0}(\mathbf{z}+\delta_i) = F^H_{\mathbf{z}_0}(\mathbf{z})\mu_{\mathbf{z}_0,\tau}(X_i), \quad e^{2\pi \mathrm{i} x_i}F^V_{\mathbf{z}_0}(\mathbf{z}+\tau\delta_i) = F^V_{\mathbf{z}_0}(\mathbf{z})\mu_{\mathbf{z}_0,\tau}(Y_i).$$

We have $\log F_{\mathbf{z}_0}(\mathbf{z}) = -\sum_i \left(z_i - z_i^0\right) y_i$ + terms of degree ≥ 2, where $\mathfrak{t}_{1,n}$ is graded by $\deg(x_i) = \deg(y_i) = 1$, which implies that $\log \mu_{\mathbf{z}_0,\tau}(X_i) = -y_i +$ terms of degree ≥ 2, $\log \mu_{\mathbf{z}_0,\tau}(Y_i) = 2\pi\mathrm{i}x_i - \tau y_i$ + terms of degree ≥ 2. Therefore $\mathrm{Lie}(\mu_{\mathbf{z}_0,\tau})(\alpha_i) = \log \mu_{\mathbf{z}_0,\tau}(A_i) = -b_i$ + terms of degree ≥ 2, $\mathrm{Lie}(\mu_{\mathbf{z}_0,\tau})(\beta_i) = \log \mu_{\mathbf{z}_0,\tau}(B_i) = 2\pi\mathrm{i}a_i - \tau b_i$ + terms of degree ≥ 2. So $\mathrm{grLie}(\mu_{\mathbf{z}_0,\tau})([\alpha_i]) = -b_i$, $\mathrm{grLie}(\mu_{\mathbf{z}_0,\tau})([\beta_i]) = 2\pi\mathrm{i}a_i - \tau b_i$.

It follows that $\mathrm{grLie}(\mu_{\mathbf{z}_0,\tau})\circ p_n$ is the endomorphism $a_i \mapsto -b_i$, $b_i \mapsto 2\pi\mathrm{i}a_i - \tau b_i$ of $\mathfrak{t}_{1,n}$, which is the automorphism $\exp\left(-\frac{\tau}{2\pi\mathrm{i}}\tilde{\Delta}_0\right) \circ w$; this proves (4).

Since we have already proved that p_n is surjective, it follows that $\mathrm{grLie}(\mu_{\mathbf{z}_0,\tau})$ and p_n are both isomorphisms. Since $\mathrm{Lie}(\mathrm{PB}_{1,n})$ and $\hat{\mathfrak{t}}_{1,n}$ are both complete and separated, $\mathrm{Lie}(\mu_{\mathbf{z}_0,\tau})$ is bijective, and since it is a morphism, it is an isomorphism of filtered Lie algebras. □

2.5 The formality of $\overline{\mathrm{PB}}_{1,n}$

Let $\mathbf{z}_0 \in U_\tau$ and let $[\mathbf{z}_0] \in \bar{C}(E_\tau, n)$ be its image. We set

$$\overline{\mathrm{PB}}_{1,n} := \pi_1(\bar{C}(E_\tau, n), [\mathbf{z}_0]).$$

Then $\overline{\mathrm{PB}}_{1,n}$ is the quotient of $\mathrm{PB}_{1,n}$ by its central subgroup (isomorphic to $\mathbb{Z}^2$) generated by A_1 and B_1. We have $\mu_{\mathbf{z}_0,\tau}(A_1) = e^{-\sum_i y_i}$ and $\mu_{\mathbf{z}_0,\tau}(B_1) = e^{2\pi\mathrm{i}\sum_i x_i - \tau\sum_i y_i}$, so $\mathrm{Lie}(\mu_{\mathbf{z}_0,\tau})(\alpha_1) = -a_1$, $\mathrm{Lie}(\mu_{\mathbf{z}_0,\tau})(\beta_1) = 2\pi\mathrm{i}a_1 - \tau b_1$, which implies that $\mathrm{Lie}(\mu_{\mathbf{z}_0,\tau})$ induces an isomorphism between $\mathrm{Lie}(\overline{\mathrm{PB}}_{1,n})_{\mathbb{C}}$ and $\bar{\mathfrak{t}}_{1,n}$. In particular, $\overline{\mathrm{PB}}_{1,n}$ is formal.

Remark 6. Let $\mathrm{Diag}_n := \{(\mathbf{z}, \tau) \in \mathbb{C}^n \times \mathfrak{H} | \mathbf{z} \in \mathrm{Diag}_{n,\tau}\}$ and let $U \subset (\mathbb{C}^n \times \mathfrak{H}) - \mathrm{Diag}_n$ be the set of all $(\mathbf{z}, \tau)$ such that $\mathbf{z} \in U_\tau$. Each element of U gives rise to a Lie algebra isomorphism $\mu_{\mathbf{z},\tau} : \mathrm{Lie}(\mathrm{PB}_{1,n}) \simeq \hat{\mathfrak{t}}_{1,n}$. For an infinitesimal $(\mathrm{d}\mathbf{z}, \mathrm{d}\tau)$, the composition $\mu_{\mathbf{z}+\mathrm{d}\mathbf{z},\tau+\mathrm{d}\tau} \circ \mu_{\mathbf{z},\tau}^{-1}$ is then an infinitesimal automorphism of $\hat{\mathfrak{t}}_{1,n}$. This defines a flat connection over U with values in the trivial Lie algebra bundle with Lie algebra $\mathrm{Der}(\hat{\mathfrak{t}}_{1,n})$. When $d\tau = 0$, the infinitesimal automorphism has the form $\exp(\sum_i K_i(\mathbf{z}|\tau)dz_i)$, so the connection has the form $d - \sum_i \mathrm{ad}(K_i(\mathbf{z}|\tau))dz_i - \tilde{\Delta}(\mathbf{z}|\tau)d\tau$, where $\tilde{\Delta} : U \to \mathrm{Der}(\hat{\mathfrak{t}}_{1,n})$ is a meromorphic map with poles at Diag_n. In the next section, we determine a map $\Delta : (\mathbb{C}^n \times \mathfrak{H}) - \mathrm{Diag}_n \to \mathrm{Der}(\hat{\mathfrak{t}}_{1,n})$ with the same flatness properties as $\tilde{\Delta}(\mathbf{z}|\tau)$.

2.6 The isomorphisms $\mathbf{B}_{1,n}(\mathbb{C}) \simeq \exp(\hat{\mathfrak{t}}_{1,n}) \rtimes S_n$, $\overline{\mathbf{B}}_{1,n}(\mathbb{C}) \simeq \exp(\hat{\bar{\mathfrak{t}}}_{1,n}) \rtimes S_n$

Let $\mathbf{z}_0$ be as above; we define $\mathrm{B}_{1,n} := \pi_1(C(E_\tau,[n]),[\mathbf{z}_0])$ and $\overline{\mathrm{B}}_{1,n} := \pi_1(\bar{C}(E_\tau,[n]),[\overline{\mathbf{z}}_0])$, where $x \mapsto [x]$ is the canonical projection $C(E_\tau,n) \to C(E_\tau,[n])$ or $\bar{C}(E_\tau,n) \to \bar{C}(E_\tau,[n])$.

We have an exact sequence $1 \to \mathrm{PB}_{1,n} \to \mathrm{B}_{1,n} \to S_n \to 1$, We then define groups $\mathrm{B}_{1,n}(\mathbb{C})$ fitting in an exact sequence $1 \to \mathrm{PB}_{1,n}(\mathbb{C}) \to \mathrm{B}_{1,n}(\mathbb{C}) \to S_n \to 1$ as follows: the morphism $\mathrm{B}_{1,n} \to \mathrm{Aut}(\mathrm{PB}_{1,n})$ extends to $\mathrm{B}_{1,n} \to \mathrm{Aut}(\mathrm{PB}_{1,n}(\mathbb{C}))$; we then construct the semidirect product $\mathrm{PB}_{1,n}(\mathbb{C}) \rtimes \mathrm{B}_{1,n}$; then $\mathrm{PB}_{1,n}$ embeds diagonally as a normal subgroup of this semidirect product, and $\mathrm{B}_{1,n}(\mathbb{C})$ is defined as the quotient $(\mathrm{PB}_{1,n}(\mathbb{C}) \rtimes \mathrm{B}_{1,n})/\mathrm{PB}_{1,n}$.

The monodromy of $\nabla_{\tau,[n]}$ then gives rise to a group morphism $\mathrm{B}_{1,n} \to \exp(\hat{\mathfrak{t}}_{1,n}) \rtimes S_n$, which factors through $\mathrm{B}_{1,n}(\mathbb{C}) \to \exp(\hat{\mathfrak{t}}_{1,n}) \rtimes S_n$. Since this map commutes with the natural morphisms to S_n, using the isomorphism $\mathrm{PB}_{1,n}(\mathbb{C}) \simeq \exp(\hat{\mathfrak{t}}_{1,n})$, we obtain that $\mathrm{B}_{1,n}(\mathbb{C}) \to \exp(\hat{\mathfrak{t}}_{1,n}) \rtimes S_n$ is an isomorphism.

Similarly, from the exact sequence $1 \to \overline{\mathrm{PB}}_{1,n} \to \overline{\mathrm{B}}_{1,n} \to S_n \to 1$ one defines a group $\overline{\mathrm{B}}_{1,n}(\mathbb{C})$ fitting in an exact sequence $1 \to \overline{\mathrm{PB}}_{1,n} \to \overline{\mathrm{B}}_{1,n}(\mathbb{C}) \to S_n \to 1$ together with an isomorphism $\overline{\mathrm{B}}_{1,n}(\mathbb{C}) \to \exp(\hat{\bar{\mathfrak{t}}}_{1,n}) \rtimes S_n$.

3 Bundles with flat connection on $\mathcal{M}_{1,n}$ and $\mathcal{M}_{1,[n]}$

We first define Lie algebras of derivations of $\bar{\mathfrak{t}}_{1,n}$ and a related group $\mathbf{G}_n$. We then define a principal $\mathbf{G}_n$-bundle with flat connection of $\mathcal{M}_{1,n}$ and a principal $\mathbf{G}_n \rtimes S_n$-bundle with flat connection on the moduli space $\mathcal{M}_{1,[n]}$ of elliptic curves with n unordered marked points.

3.1 Derivations of the Lie algebras $\mathfrak{t}_{1,n}$ and $\bar{\mathfrak{t}}_{1,n}$ and associated groups

Let $\mathfrak{d}$ be the Lie algebra with generators Δ_0, d, X, and δ_{2m} $(m \geq 1)$, and relations

$$[d,X] = 2X, \quad [d,\Delta_0] = -2\Delta_0, \quad [X,\Delta_0] = d,$$

$$[\delta_{2m},X] = 0, \quad [d,\delta_{2m}] = 2m\delta_{2m}, \quad \mathrm{ad}(\Delta_0)^{2m+1}(\delta_{2m}) = 0.$$

Proposition 7. *We have a Lie algebra morphism* $\mathfrak{d} \to \mathrm{Der}(\mathfrak{t}_{1,n})$, *denoted by* $\xi \mapsto \tilde{\xi}$, *such that* $\tilde{d}(x_i) = x_i$, $\tilde{d}(y_i) = -y_i$, $\tilde{d}(t_{ij}) = 0$, $\tilde{X}(x_i) = 0$, $\tilde{X}(y_i) = x_i$, $\tilde{X}(t_{ij}) = 0$, $\tilde{\Delta}_0(x_i) = y_i$, $\tilde{\Delta}_0(y_i) = 0$, $\tilde{\Delta}_0(t_{ij}) = 0$, $\tilde{\delta}_{2m}(x_i) = 0$,

$$\tilde{\delta}_{2m}(t_{ij}) = [t_{ij}, (\mathrm{ad}\, x_i)^{2m}(t_{ij})], \quad \textit{and}$$

$$\tilde{\delta}_{2m}(y_i) = \sum_{j|j\neq i} \frac{1}{2} \sum_{p+q=2m-1} [(\mathrm{ad}\, x_i)^p(t_{ij}), (-\mathrm{ad}\, x_i)^q(t_{ij})].$$

This induces a Lie algebra morphism $\mathfrak{d} \to \mathrm{Der}(\bar{\mathfrak{t}}_{1,n})$.

Proof. The fact that $\tilde{\Delta}_0, \tilde{d}, \tilde{X}$ are derivations and commute according to the Lie bracket of $\mathfrak{sl}_2$ is clear.

Let us prove that $\tilde{\delta}_{2m}$ is a derivation. We have

$$\tilde{\delta}_{2m}(t_{ij}) = \left[t_{ij}, \sum_{i<j} (\mathrm{ad}\, x_i)^{2m}(t_{ij}) \right],$$

which implies that $\tilde{\delta}_{2m}$ preserves the infinitesimal pure braid identities. It clearly preserves the relations

$$[x_i, x_j] = 0,\ [x_i, y_j] = t_{ij},\ [x_k, t_{ij}] = 0,\ [x_i + x_j, t_{ij}] = 0.$$

Let us prove that $\tilde{\delta}_{2m}$ preserves the relation $[y_k, t_{ij}] = 0$. On the one hand,

$$\begin{aligned}
[\tilde{\delta}_{2m}(y_k), t_{ij}] &= \frac{1}{2} \sum_{p+q=2m-1} (-1)^q [[(\mathrm{ad}\, x_k)^p(t_{ki}), (\mathrm{ad}\, x_k)^q(t_{ki})] \\
&\quad + [(\mathrm{ad}\, x_k)^p(t_{kj}), (\mathrm{ad}\, x_k)^q(t_{kj})], t_{ij}] \\
&= \frac{1}{2} \sum_{p+q=2m-1} (-1)^{q+1} [[(\mathrm{ad}\, x_k)^p(t_{ki}), (\mathrm{ad}\, x_k)^q(t_{kj})] \\
&\quad + [(\mathrm{ad}\, x_k)^p(t_{kj}), (\mathrm{ad}\, x_k)^q(t_{ki})], t_{ij}] \\
&= \sum_{p+q=2m-1} (-1)^{q+1} [[(\mathrm{ad}\, x_k)^p(t_{ki}), (\mathrm{ad}\, x_k)^q(t_{kj})], t_{ij}] \\
&= \left[t_{ij}, \sum_{p+q=2m-1} (-1)^p (\mathrm{ad}\, x_i)^p (\mathrm{ad}\, x_j)^q ([t_{ki}, t_{kj}]) \right].
\end{aligned}$$

On the other hand

$$[y_k, \tilde{\delta}_{2m}(t_{ij})] = [y_k, [t_{ij}, (\mathrm{ad}\, x_i)^{2m}(t_{ij})]] = [t_{ij}, [y_k, (\mathrm{ad}\, x_i)^{2m}(t_{ij})]].$$

Now

$$
\begin{aligned}
[y_k, (\operatorname{ad} x_i)^{2m}(t_{ij})] &= -\sum_{\alpha+\beta=2m-1} (\operatorname{ad} x_i)^\alpha \big([t_{ki}, (\operatorname{ad} x_i)^\beta(t_{ij})]\big) \\
&= -\sum_{\alpha+\beta=2m-1} (\operatorname{ad} x_i)^\alpha [t_{ki}, (-\operatorname{ad} x_j)^\beta(t_{ij})] \\
&= -\sum_{\alpha+\beta=2m-1} (\operatorname{ad} x_i)^\alpha (-\operatorname{ad} x_j)^\beta([t_{ki}, t_{ij}]) \\
&= \sum_{p+q=2m-1} (-1)^{p+1} (\operatorname{ad} x_i)^p (\operatorname{ad} x_j)^q([t_{ki}, t_{kj}]).
\end{aligned}
$$

Hence $[\tilde{\delta}_{2m}(y_k), t_{ij}] + [y_k, \tilde{\delta}_{2m}(t_{ij})] = 0$.

Let us prove that $\tilde{\delta}_{2m}$ preserves the relation $[y_i, y_j] = 0$, i.e., that $[\tilde{\delta}_{2m}(y_i), y_j] + [y_i, \tilde{\delta}_{2m}(y_j)] = 0$. We have

$$
\begin{aligned}
[y_i, \tilde{\delta}_{2m}(y_j)] = \frac{1}{2}&\left[y_i, \sum_{p+q=2m-1} (-1)^q [(\operatorname{ad} x_j)^p(t_{ji}), (\operatorname{ad} x_j)^q(t_{ji})]\right] \\
&+ \frac{1}{2}\sum_{k\neq i,j}\left[y_i, \sum_{p+q=2m-1} (-1)^q [(\operatorname{ad} x_j)^p(t_{jk}), (\operatorname{ad} x_j)^q(t_{jk})]\right].
\end{aligned}
$$

Now

$$
\frac{1}{2}\left[y_i, \sum_{p+q=2m-1} (-1)^q [(\operatorname{ad} x_j)^p(t_{ji}), (\operatorname{ad} x_j)^q(t_{ji})]\right] - (i \leftrightarrow j) \tag{5}
$$

$$
\begin{aligned}
&= -\frac{1}{2}\left[y_i + y_j, \sum_{p+q=2m-1} (-1)^q [(\operatorname{ad} x_i)^p(t_{ij}), (\operatorname{ad} x_i)^q(t_{ij})]\right] \\
&= \sum_{p+q=2m-1} (-1)^{q+1} \left[[y_i + y_j, (\operatorname{ad} x_i)^p(t_{ij})], (\operatorname{ad} x_i)^q(t_{ij})\right].
\end{aligned}
$$

A computation similar to the above computation of $[y_k, (\operatorname{ad} x_i)^{2m}(t_{ij})]$ yields

$$
[y_i + y_j, (\operatorname{ad} x_i)^p(t_{ij})] = (-1)^p \sum_{\alpha+\beta=p-1} [(\operatorname{ad} x_k)^\alpha(t_{ik}), (\operatorname{ad} x_j)^\beta(t_{jk})],
$$

so

$$
(5) = \sum_{\alpha+\beta+\gamma=2m-2} [(\operatorname{ad} x_i)^\alpha(t_{ij}), [(\operatorname{ad} x_k)^\beta(t_{ik}), (\operatorname{ad} x_j)^\gamma(t_{jk})]].
$$

If now $k \neq i, j$, then

$$\left[y_i, \frac{1}{2} \sum_{p+q=2m-1} (-1)^q [(\mathrm{ad}\, x_j)^p(t_{jk}), (\mathrm{ad}\, x_j)^q(t_{jk})]\right] \\ = \sum_{p+q=2m-1} (-1)^q [[y_i, (\mathrm{ad}\, x_j)^p(t_{jk})], (\mathrm{ad}\, x_j)^q(t_{jk})].$$

As we have seen,

$$\begin{aligned}[y_j, (\mathrm{ad}\, x_i)^p(t_{ik})] &= (-1)^p \sum_{\alpha+\beta=p-1} (-\mathrm{ad}\, x_i)^\alpha (\mathrm{ad}\, x_k)^\beta [t_{ij}, t_{ik}] \\ &= (-1)^{p+1} \sum_{\alpha+\beta=p-1} [(-\mathrm{ad}\, x_i)^\alpha(t_{ij}), (\mathrm{ad}\, x_k)^\beta(t_{jk})].\end{aligned}$$

So we get that $\left[y_i, \frac{1}{2} \sum_{p+q=2m-1} (-1)^q [(\mathrm{ad}\, x_j)^p(t_{jk}), (\mathrm{ad}\, x_j)^q(t_{jk})]\right]$ equals

$$\sum_{\alpha+\beta+\gamma=2m-2} [[(\mathrm{ad}\, x_i)^\alpha(t_{ij}), (\mathrm{ad}\, x_k)^\beta(t_{ik})], (\mathrm{ad}\, x_j)^\gamma(t_{jk})]$$

and thus $\left[y_i, \frac{1}{2} \sum_{p+q=2m-1} (-1)^q [(\mathrm{ad}\, x_j)^p(t_{jk}), (\mathrm{ad}\, x_j)^q(t_{jk})]\right] - (i \leftrightarrow j)$ equals

$$\sum_{\alpha+\beta+\gamma=2m-2} [(\mathrm{ad}\, x_i)^\alpha(t_{ij}), [(\mathrm{ad}\, x_k)^\beta(t_{ik}), (\mathrm{ad}\, x_j)^\gamma(t_{jk})]].$$

Therefore $[y_i, \tilde{\delta}_{2m}(y_j)] + [\tilde{\delta}_{2m}(y_i), y_j] = 0$.

Since $\tilde{\delta}_{2m}(\sum_i x_i) = \tilde{\delta}_{2m}(\sum_i y_i) = 0$ and $\sum_i x_i$ and $\sum_i y_i$ are central, $\tilde{\delta}_{2m}$ preserves the relations $[\sum_i x_i, y_j] = 0$ and $[\sum_k x_k, t_{ij}] = [\sum_k y_k, t_{ij}] = 0$. It follows that $\tilde{\delta}_{2m}$ preserves the relations $[x_i + x_j, t_{ij}] = [y_i + y_j, t_{ij}] = 0$ and $[x_i, y_i] = -\sum_{j|j\neq i} t_{ij}$. All this proves that $\tilde{\delta}_{2m}$ is a derivation.

Let us show that $\mathrm{ad}(\tilde{\Delta}_0)^{2m+1}(\tilde{\delta}_{2m}) = 0$ for $m \geq 1$. We have

$$\begin{aligned}&\mathrm{ad}(\tilde{\Delta}_0)^{2m+1}(\tilde{\delta}_{2m})(x_i) \\ &\quad = -(2m+1)\tilde{\Delta}_0^{2m} \circ \tilde{\delta}_{2m} \circ \tilde{\Delta}_0(x_i) \\ &\quad = -(2m+1)\tilde{\Delta}_0^{2m} \circ \tilde{\delta}_{2m}(y_i) \\ &\quad = -(2m+1)\tilde{\Delta}_0^{2m} \left(\frac{1}{2} \sum_{\substack{j|j \neq i, \\ p+q=2m-1}} [(\mathrm{ad}\, x_i)^p(t_{ij}), (-\mathrm{ad}\, x_i)^q(t_{ij})] \right) \\ &\quad = 0;\end{aligned}$$

the last part of this computation implies that $\mathrm{ad}(\tilde{\Delta}_0)^{2m+1}(\tilde{\delta}_{2m})(y_i) = 0$; therefore $\mathrm{ad}(\tilde{\Delta}_0)^{2m+1}(\tilde{\delta}_{2m}) = 0$.

We have clearly $[\tilde{X}, \tilde{\delta}_{2m}] = 0$ and $[\tilde{d}, \tilde{\delta}_{2m}] = 2m\tilde{\delta}_{2m}$. It follows that we have a Lie algebra morphism $\mathfrak{d} \to \mathrm{Der}(\mathfrak{t}_{1,n})$. Since $\tilde{d}, \tilde{\Delta}_0, \tilde{X}$, and $\tilde{\delta}_{2m}$ all map $\mathbb{C}(\sum_i x_i) \oplus \mathbb{C}(\sum_i y_i)$ to itself, this induces a Lie algebra morphism $\mathfrak{d} \to \mathrm{Der}(\bar{\mathfrak{t}}_{1,n})$. □

Let e, f, h be the standard basis of $\mathfrak{sl}_2$. Then we have a Lie algebra morphism $\mathfrak{d} \to \mathfrak{sl}_2$, defined by $\delta_{2n} \mapsto 0$, $d \mapsto h$, $X \mapsto e$, $\Delta_0 \mapsto f$. We denote by $\mathfrak{d}_+ \subset \mathfrak{d}$ its kernel.

Since the morphism $\mathfrak{d} \to \mathfrak{sl}_2$ has a section (given by $e, f, h \mapsto X, \Delta_0, d$), we have a semidirect product decomposition $\mathfrak{d} = \mathfrak{d}_+ \rtimes \mathfrak{sl}_2$.

We then have

$$\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d} = (\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}_+) \rtimes \mathfrak{sl}_2.$$

Lemma 8. $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}_+$ *is positively graded.*

Proof. We define compatible $\mathbb{Z}^2$-gradings of $\mathfrak{d}$ and $\bar{\mathfrak{t}}_{1,n}$ by $\deg(\Delta_0) = (-1, 1)$, $\deg(d) = (0, 0)$, $\deg(X) = (1, -1)$, $\deg(\delta_{2m}) = (2m + 1, 1)$, $\deg(x_i) = (1, 0)$, $\deg(y_i) = (0, 1)$, $\deg(t_{ij}) = (1, 1)$.

We define the support of $\mathfrak{d}$ (respectively, $\bar{\mathfrak{t}}_{1,n}$) as the subset of $\mathbb{Z}^2$ of indices for which the corresponding component of $\mathfrak{d}$ (respectively, $\bar{\mathfrak{t}}_{1,n}$) is nonzero.

Since the $\bar{x}_i$ on the one hand and the $\bar{y}_i$ on the other hand generate abelian Lie subalgebras of $\bar{\mathfrak{t}}_{1,n}$, the support of $\bar{\mathfrak{t}}_{1,n}$ is contained in $\mathbb{N}^2_{>0} \cup \{(1, 0), (0, 1)\}$.

On the other hand, $\mathfrak{d}_+$ is generated by the $\mathrm{ad}(\Delta_0)^p(\delta_{2m})$, which all have degrees in $\mathbb{N}^2_{>0}$. It follows that the support of $\mathfrak{d}_+$ is contained in $\mathbb{N}^2_{>0}$.

Therefore the support of $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}_+$ is contained in $\mathbb{N}^2_{>0} \cup \{(1, 0), (0, 1)\}$, so this Lie algebra is positively graded. □

Lemma 9. $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}_+$ *is a sum of finite-dimensional* $\mathfrak{sl}_2$*-modules;* $\mathfrak{d}_+$ *is a sum of irreducible odd-dimensional* $\mathfrak{sl}_2$*-modules.*

Proof. A generating space for $\bar{\mathfrak{t}}_{1,n}$ is $\sum_i (\mathbb{C}\bar{x}_i \oplus \mathbb{C}\bar{y}_i)$, which is a sum of finite-dimensional $\mathfrak{sl}_2$-modules, so $\bar{\mathfrak{t}}_{1,n}$ is a sum of finite-dimensional $\mathfrak{sl}_2$-modules.

A generating space for $\mathfrak{d}_+$ is the sum over $m \geq 1$ of its $\mathfrak{sl}_2$-submodules generated by the δ_{2m}, which are zero or irreducible odd-dimensional; therefore $\mathfrak{d}_+$ is a sum of odd-dimensional $\mathfrak{sl}_2$-modules. (In fact, the $\mathfrak{sl}_2$-submodule generated by δ_{2m} is nonzero, since it follows from the construction of the above morphism $\mathfrak{d}_+ \to \mathrm{Der}(\bar{\mathfrak{t}}_{1,n})$ that $\delta_{2m} \neq 0$.) □

It follows that $\bar{\mathfrak{t}}_{1,n}$, $\bar{\mathfrak{d}}_+$, and $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}_+$ integrate to $\mathrm{SL}_2(\mathbb{C})$-modules (while $\bar{\mathfrak{d}}_+$ even integrates to a $\mathrm{PSL}_2(\mathbb{C})$-module).

We can form in particular the semidirect products

$$\mathbf{G}_n := \exp((\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}_+)^\wedge) \rtimes \mathrm{SL}_2(\mathbb{C})$$

and $\exp(\hat{\mathfrak{d}}_+) \rtimes \mathrm{PSL}_2(\mathbb{C})$; we have morphisms $\mathbf{G}_n \to \exp(\hat{\mathfrak{d}}_+) \rtimes \mathrm{PSL}_2(\mathbb{C})$ (this is a 2-covering if $n = 1$, since $\bar{\mathfrak{t}}_{1,1} = 0$).

Observe that the action of S_n by automorphisms of $\bar{\mathfrak{t}}_{1,n}$ extends to an action on $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}$, where the action on $\mathfrak{d}$ is trivial. This gives rise to an action of S_n by automorphisms of $\mathbf{G}_n$.

3.2 Bundle with flat connection on $\mathcal{M}_{1,n}$

The semidirect product $((\mathbb{Z}^n)^2 \times \mathbb{C}) \rtimes \mathrm{SL}_2(\mathbb{Z})$ acts on $(\mathbb{C}^n \times \mathfrak{H}) - \mathrm{Diag}_n$ by

$$(\mathbf{n}, \mathbf{m}, u) * (\mathbf{z}, \tau) := \left(\mathbf{n} + \tau\mathbf{m} + u\left(\sum_i \delta_i\right), \tau\right) \text{ for } (\mathbf{n}, \mathbf{m}, u) \in (\mathbb{Z}^n)^2 \times \mathbb{C}$$

and

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} * (\mathbf{z}, \tau) := \left(\frac{\mathbf{z}}{\gamma\tau + \delta}, \frac{\alpha\tau + \beta}{\gamma\tau + \delta}\right) \text{ for } \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$$

(here $\mathrm{Diag}_n := \{(\mathbf{z}, \tau) \in \mathbb{C}^n \times \mathfrak{H} |$ for some $i \neq j, z_{ij} \in \Lambda_\tau\}$). The quotient is then identified with the moduli space $\mathcal{M}_{1,n}$ of elliptic curves with n marked points.

Set $\mathbf{G}_n := \exp((\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}_+)^\wedge) \rtimes \mathrm{SL}_2(\mathbb{C})$. We will define a principal $\mathbf{G}_n$-bundle with flat connection $(\mathcal{P}_n, \nabla_{\mathcal{P}_n})$ over $\mathcal{M}_{1,n}$.

For $u \in \mathbb{C}^\times$, $u^d := \left(\begin{smallmatrix} u & 0 \\ 0 & u^{-1} \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{C}) \subset \mathbf{G}_n$ and for $v \in \mathbb{C}$, $e^{vX} := \left(\begin{smallmatrix} 1 & v \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{C}) \subset \mathbf{G}_n$. Since $[X, \bar{x}_i] = 0$, we consistently set

$$\exp\left(aX + \sum_i b_i \bar{x}_i\right) := \exp(aX)\exp\left(\sum_i b_i \bar{x}_i\right).$$

Proposition 10. *There exists a unique principal $\mathbf{G}_n$-bundle $\mathcal{P}_n$ over $\mathcal{M}_{1,n}$ such that a section of $U \subset \mathcal{M}_{1,n}$ is a function $f : \pi^{-1}(U) \to \mathbf{G}_n$ (where*

$$\pi : (\mathbb{C}^n \times \mathfrak{H}) - \mathrm{Diag}_n \to \mathcal{M}_{1,n}$$

is the canonical projection) such that

- $f(\mathbf{z} + \delta_i|\tau) = f(\mathbf{z} + u(\sum_i \delta_i)|\tau) = f(\mathbf{z}|\tau)$,
- $f(\mathbf{z} + \tau\delta_i|\tau) = e^{-2\pi \mathrm{i} \bar{x}_i} f(\mathbf{z}|\tau)$,
- $f(\mathbf{z}|\tau + 1) = f(\mathbf{z}|\tau)$, *and*
- $f\left(\frac{\mathbf{z}}{\tau}| - \frac{1}{\tau}\right) = \tau^d \exp\left(\frac{2\pi \mathrm{i}}{\tau}\left(\sum_i z_i \bar{x}_i + X\right)\right) f(\mathbf{z}|\tau)$.

Proof. Let $c_{\tilde{g}} : \mathbb{C}^n \times \mathfrak{H} \to \mathbf{G}_n$ be a family of holomorphic functions (where $\tilde{g} \in ((\mathbb{Z}^n)^2 \times \mathbb{C}) \rtimes \mathrm{SL}_2(\mathbb{Z})$) satisfying the cocycle condition

$$c_{\tilde{g}\tilde{g}'}(\mathbf{z}|\tau) = c_{\tilde{g}}(\tilde{g}' * (\mathbf{z}|\tau)) c_{\tilde{g}'}(\mathbf{z}|\tau).$$

Then there exists a unique principal $\mathbf{G}_n$-bundle over $\mathcal{M}_{1,n}$ such that a section of $U \subset \mathcal{M}_{1,n}$ is a function $f : \pi^{-1}(U) \to \mathbf{G}_n$ such that $f(\tilde{g} * (\mathbf{z}|\tau)) = c_{\tilde{g}}(\mathbf{z}|\tau) f(\mathbf{z}|\tau)$.

We will now prove that there is a unique cocycle such that $c_{(u,0,0)} = c_{(0,\delta_i,0)} = 1$, $c_{(0,0,\delta_i)} = e^{-2\pi \mathrm{i} \bar{x}_i}$, $c_S = 1$ and $c_T(\mathbf{z}|\tau) = \tau^d \exp\left(\frac{2\pi \mathrm{i}}{\tau}\left(\sum_i z_i \bar{x}_i + X\right)\right)$, where $S = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, $T = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

Such a cocycle is the same as a family of functions $c_g : \mathbb{C}^n \times \mathfrak{H} \to \mathbf{G}_n$ (where $g \in \mathrm{SL}_2(\mathbb{Z})$), satisfying the cocycle conditions $c_{gg'}(\mathbf{z}|\tau) = c_g(g' * (\mathbf{z}|\tau)) c_{g'}(\mathbf{z}|\tau)$ for $g, g' \in \mathrm{SL}_2(\mathbb{Z})$, and $c_g(\mathbf{z} + \delta_i|\tau) = e^{2\pi \mathrm{i} \gamma \bar{x}_i} c_g(\mathbf{z}|\tau)$, $c_g(\mathbf{z} + \tau\delta_i|\tau) = e^{-2\pi \mathrm{i} \delta \bar{x}_i} c_g(\mathbf{z}|\tau) e^{2\pi \mathrm{i} \bar{x}_i}$, and $c_g(\mathbf{z} + u(\sum_i \delta_i)|\tau) = c_g(\mathbf{z}|\tau)$ for $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$.

Lemma 11. *There exists a unique family of functions* $c_g : \mathbb{C}^n \times \mathfrak{H} \to \mathbf{G}_n$ *such that* $c_{gg'}(\mathbf{z}|\tau) = c_g(g' * (\mathbf{z}|\tau))c_{g'}(\mathbf{z}|\tau)$ *for* $g, g' \in \mathrm{SL}_2(\mathbb{Z})$*, with*

$$c_S(\mathbf{z}|\tau) = 1, \quad c_T(\mathbf{z}|\tau) = \tau^d e^{(2\pi \mathrm{i}/\tau)(\sum_j z_j \bar{x}_j + X)}.$$

Proof. $\mathrm{SL}_2(\mathbb{Z})$ is the group generated by $\tilde{S}$, $\tilde{T}$, and relations $\tilde{T}^4 = 1$, $(\tilde{S}\tilde{T})^3 = \tilde{T}^2$, $\tilde{S}\tilde{T}^2 = \tilde{T}^2\tilde{S}$. Let $\langle \tilde{S}, \tilde{T} \rangle$ be the free group with generators $\tilde{S}$, $\tilde{T}$; then there is a unique family of maps $c_{\tilde{g}} : \mathbb{C}^n \times \mathfrak{H} \to \mathbf{G}_n$, $\tilde{g} \in \langle \tilde{S}, \tilde{T} \rangle$ satisfying the cocycle conditions (with respect to the action of $\langle \tilde{S}, \tilde{T} \rangle$ on $\mathbb{C}^n \times \mathfrak{H}$ through its quotient $\mathrm{SL}_2(\mathbb{Z})$) and $c_{\tilde{S}} = c_S$, $c_{\tilde{T}} = c_T$. It remains to show that $c_{\tilde{T}^4} = 1$, $c_{(\tilde{S}\tilde{T})^3} = c_{\tilde{T}^2}$, and $c_{\tilde{S}\tilde{T}^2} = c_{\tilde{T}^2\tilde{S}}$.

For this, we show that $c_{\tilde{T}^2}(\mathbf{z}|\tau) = (-1)^d$. We have

$$\begin{aligned} c_{\tilde{T}^2}(\mathbf{z}|\tau) &= c_T(\mathbf{z}/\tau| -1/\tau)c_T(\mathbf{z}|\tau) \\ &= (-\tau)^{-d}\exp\left(-2\pi \mathrm{i}\tau\left(\sum_j (z_j/\tau)\bar{x}_j + X\right)\right)\tau^d\exp\left(\frac{2\pi\mathrm{i}}{\tau}\left(\sum_j z_j\bar{x}_j + X\right)\right) \\ &= (-1)^d, \end{aligned}$$

since $\tau^d X \tau^{-d} = \tau^2 X$, $\tau^d \bar{x}_i \tau^{-d} = \tau \bar{x}_i$.

Since $((-1)^d)^2 = 1^d = 1$, we get $c_{\tilde{T}^4} = 1$. Since $c_{\tilde{S}}$ and $c_{\tilde{T}^2}$ are both constant and commute, we also get $c_{\tilde{S}\tilde{T}^2} = c_{\tilde{T}^2\tilde{S}}$.

We finally have $c_{\tilde{S}\tilde{T}}(\mathbf{z}|\tau) = c_T(\mathbf{z}|\tau)$, while $\tilde{S}\tilde{T} = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $(\tilde{S}\tilde{T})^2 = \left(\begin{smallmatrix} 0 & -1 \\ 1 & -1 \end{smallmatrix}\right)$, so

$$\begin{aligned} & c_{(\tilde{S}\tilde{T})^3}(\mathbf{z}|\tau) \\ &= c_T\left(\frac{\mathbf{z}}{\tau-1}|\frac{1}{1-\tau}\right)c_T\left(\frac{\mathbf{z}}{\tau}|\frac{\tau-1}{\tau}\right)c_T(\mathbf{z}|\tau) \\ &= \left(\frac{1}{1-\tau}\right)^d \exp\left(-2\pi\mathrm{i}\sum z_j\bar{x}_j + 2\pi\mathrm{i}(1-\tau)X\right)\left(\frac{\tau-1}{\tau}\right)^d \\ &\quad \exp\left(\frac{2\pi\mathrm{i}}{\tau-1}\sum_j z_j\bar{x}_j + 2\pi\mathrm{i}\frac{\tau}{\tau-1}X\right)\tau^d\exp\left(\frac{2\pi\mathrm{i}}{\tau}\left(\sum_j z_j\bar{x}_j + X\right)\right) \\ &= (-1)^d\exp\left(\frac{2\pi\mathrm{i}}{1-\tau}\left(\sum_j z_j\bar{x}_j + X\right)\right)\exp\left(\frac{2\pi\mathrm{i}}{\tau(\tau-1)}\left(\sum_j z_j\bar{x}_j + X\right)\right) \\ &\quad \exp\left(\frac{2\pi\mathrm{i}}{\tau}\left(\sum_j z_j\bar{x}_j + X\right)\right) \\ &= (-1)^d, \end{aligned}$$

so $c_{(\tilde{S}\tilde{T})^3} = c_{\tilde{T}^2}$. $\square$

End of proof of Proposition 10. We now check that the maps c_g satisfy the remaining conditions, i.e., $c(\mathbf{z}+u(\sum_i \delta_i)|\tau) = c_g(\mathbf{z}|\tau)$, $c_g(\mathbf{z}+\delta_i|\tau) = e^{2\pi \mathrm{i}\gamma\bar{x}_i}c_g(\mathbf{z}|\tau)$, $c_g(\mathbf{z}+\tau\delta_i|\tau) = e^{-2\pi\mathrm{i}\delta\bar{x}_i}c_g(\mathbf{z}|\tau)e^{2\pi\mathrm{i}\bar{x}_i}$. The cocycle identity $c_{gg'}(\mathbf{z}|\tau) = c_g(g' * (\mathbf{z}|\tau))c_{g'}(\mathbf{z}|\tau)$ implies that it suffices to prove these identities for $g = S$ and $g = T$. They are trivially satisfied if $g = S$. When $g = T$, the first identity follows from $\sum_i \bar{x}_i = 0$, the third identity follows from the fact that $(X, \bar{x}_1, \ldots, \bar{x}_n)$ is a commutative family, and the second identity follows from the same fact together with $\tau^d \bar{x}_i \tau^{-d} = \tau\bar{x}_i$. □

Set

$$g(z,x|\tau) := \frac{\theta(z+x|\tau)}{\theta(z|\tau)\theta(x|\tau)}\left(\frac{\theta'}{\theta}(z+x|\tau) - \frac{\theta'}{\theta}(x|\tau)\right) + \frac{1}{x^2} = k_x(z,x|\tau)$$

(we set $f'(z|\tau) := (\partial/\partial z)f(z|\tau)$).

We have $g(z,x|\tau) \in \mathrm{Hol}((\mathbb{C}\times\mathfrak{H}) - \mathrm{Diag}_1)[[x]]$, therefore $g(z, \mathrm{ad}\,\bar{x}_i|\tau)$ is a linear map $\bar{\mathfrak{t}}_{1,n} \to (\mathrm{Hol}((\mathbb{C}\times\mathfrak{H}) - \mathrm{Diag}_1)\otimes\bar{\mathfrak{t}}_{1,n})^\wedge$, so $g(z,\mathrm{ad}\,\bar{x}_i|\tau)(\bar{t}_{ij}) \in (\mathrm{Hol}((\mathbb{C}\times\mathfrak{H}) - \mathrm{Diag}_1)\otimes\bar{\mathfrak{t}}_{1,n})^\wedge$. Therefore

$$g(\mathbf{z}|\tau) := \sum_{i<j} g(z_{ij}, \mathrm{ad}\,\bar{x}_i|\tau)(\bar{t}_{ij})$$

is a meromorphic function $\mathbb{C}^n \times \mathfrak{H} \to \hat{\bar{\mathfrak{t}}}_{1,n}$ with poles only at Diag_n.

We set

$$\bar{\Delta}(\mathbf{z}|\tau) := -\frac{1}{2\pi\mathrm{i}}\Delta_0 - \frac{1}{2\pi\mathrm{i}}\sum_{n\geq 1} a_{2n}E_{2n+2}(\tau)\delta_{2n} + \frac{1}{2\pi\mathrm{i}}g(\mathbf{z}|\tau),$$

where $a_{2n} = -(2n+1)B_{2n+2}(2\mathrm{i}\pi)^{2n+2}/(2n+2)!$ and B_n are the Bernoulli numbers given by $x/(e^x-1) = \sum_{r\geq 0}(B_r/r!)x^r$. This is a meromorphic function $\mathbb{C}^n\times\mathfrak{H} \to (\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}_+)^\wedge \rtimes \mathfrak{n}_+ \subset \mathrm{Lie}(\mathbf{G}_n)$ (where $\mathfrak{n}_+ = \mathbb{C}\Delta_0 \subset \mathfrak{sl}_2$) with poles only at Diag_n.

For $\psi(x) = \sum_{n\geq 1} b_{2n}x^{2n}$, we set $\delta_\psi := \sum_{n\geq 1} b_{2n}\delta_{2n}$, $\Delta_\psi := \Delta_0 + \sum_{n\geq 1} b_{2n}\delta_{2n}$. If we set

$$\varphi(x|\tau) = -x^{-2} - (\theta'/\theta)'(x|\tau) + (x^{-2} + (\theta'/\theta)'(x|\tau))_{|x=0} = g(0,0|\tau) - g(0,x|\tau),$$

then $\varphi(x|\tau) = \sum_{n\geq 1} a_{2n}E_{2n+2}(\tau)x^{2n}$, so that

$$\bar{\Delta}(\mathbf{z}|\tau) = -\frac{1}{2\pi\mathrm{i}}\Delta_{\varphi(*|\tau)} + \frac{1}{2\pi\mathrm{i}}g(\mathbf{z}|\tau).$$

Theorem 12. *There is a unique flat connection $\nabla_{\mathcal{P}_n}$ on $\mathcal{P}_n$ whose pullback to $(\mathbb{C}^n\times\mathfrak{H}) - \mathrm{Diag}_n$ is the connection*

$$d - \bar{\Delta}(\mathbf{z}|\tau)d\tau - \sum_i \bar{K}_i(\mathbf{z}|\tau)dz_i$$

on the trivial $\mathbf{G}_n$-bundle.

Proof. We should check that the connection $d - \bar{\Delta}(\mathbf{z}|\tau)d\tau - \sum_i \bar{K}_i(\mathbf{z}|\tau)dz_i$ is equivariant and flat, which is expressed as follows (taking into account that we already checked the equivariance and flatness of $d - \sum_i \bar{K}_i(\mathbf{z}|\tau)dz_i$ for any τ):

(equivariance) for $g = \left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$

$$\frac{1}{\gamma\tau+\delta}\bar{K}_i\left(\frac{\mathbf{z}}{\gamma\tau+\delta}|\frac{\alpha\tau+\beta}{\gamma\tau+\delta}\right) = \mathrm{Ad}(c_g(\mathbf{z}|\tau))(\bar{K}_i(\mathbf{z}|\tau)) + [(\partial/\partial z_i)c_g(\mathbf{z}|\tau)]c_g(\mathbf{z}|\tau)^{-1} \tag{6}$$

$$\bar{\Delta}(\mathbf{z}+\delta_i|\tau) = \bar{\Delta}(\mathbf{z}+u(\sum_i \delta_i)|\tau) = \bar{\Delta}(\mathbf{z}|\tau)$$

$$\text{and} \quad \bar{\Delta}(\mathbf{z}+\tau\delta_i|\tau) = e^{-2\pi \mathrm{i}\,\mathrm{ad}\, x_i}(\bar{\Delta}(\mathbf{z}|\tau) - \bar{K}_i(\mathbf{z}|\tau)), \tag{7}$$

$$\frac{1}{(\gamma\tau+\delta)^2}\bar{\Delta}\left(\frac{\mathbf{z}}{\gamma\tau+\delta}|\frac{\alpha\tau+\beta}{\gamma\tau+\delta}\right) = \mathrm{Ad}(c_g(\mathbf{z}|\tau))(\bar{\Delta}(\mathbf{z}|\tau)) + \frac{\gamma}{\gamma z+\delta}\sum_{i=1}^n z_i \mathrm{Ad}(c_g(\mathbf{z}|\tau))(\bar{K}_i(\mathbf{z}|\tau)) + \left[\left(\frac{\partial}{\partial\tau} + \frac{\gamma}{\gamma\tau+\delta}\sum_{i=1}^n z_i \frac{\partial}{\partial z_i}\right)c_g(\mathbf{z}|\tau)\right]c_g(\mathbf{z}|\tau)^{-1}; \tag{8}$$

(flatness) $[\partial/\partial\tau - \bar{\Delta}(\mathbf{z}|\tau), \partial/\partial z_i - \bar{K}_i(\mathbf{z}|\tau)] = 0$.

Let us now check the equivariance identity (6) for $\bar{K}_i(\mathbf{z}|\tau)$. The cocycle identity $c_{gg'}(\mathbf{z}|\tau) = c_g(g'*(\mathbf{z}|\tau))c_{g'}(\mathbf{z}|\tau)$ implies that it suffices to check it when $g = S$ and $g = T$. When $g = S$, this is the identity $\bar{K}_i(\mathbf{z}|\tau+1) = \bar{K}_i(\mathbf{z}|\tau)$, which follows from the identity $\theta(z|\tau+1) = \theta(z|\tau)$. When $g = T$, we have to check the identity

$$\frac{1}{\tau}\bar{K}_i\left(\frac{\mathbf{z}}{\tau}|-\frac{1}{\tau}\right) = \mathrm{Ad}\left(\tau^d e^{\frac{2\pi\mathrm{i}}{\tau}(\sum_i z_i\bar{x}_i + X)}\right)(\bar{K}_i(\mathbf{z}|\tau)) + 2\pi\mathrm{i}\bar{x}_i. \tag{9}$$

We have

$$\begin{aligned}
&2\pi\mathrm{i}\bar{x}_i - \mathrm{Ad}(e^{2\pi\mathrm{i}(\sum_i z_i\bar{x}_i + X)})(\bar{y}_i/\tau)\\
&\quad= -\mathrm{Ad}(e^{2\pi\mathrm{i}(\sum_i z_i\bar{x}_i)})(\bar{y}_i/\tau) \quad (\text{since } \mathrm{Ad}(e^{2\pi\mathrm{i}\tau X})(\bar{y}_i/\tau) = \bar{y}_i/\tau + 2\pi\mathrm{i}\bar{x}_i)\\
&\quad= -\frac{\bar{y}_i}{\tau} - \frac{e^{2\pi\mathrm{i}\,\mathrm{ad}(\sum_k z_k\bar{x}_k)} - 1}{\mathrm{ad}(\sum_k z_k\bar{x}_k)}\left(\left[\sum_j z_j\bar{x}_j, \frac{\bar{y}_i}{\tau}\right]\right)\\
&\quad= -\frac{\bar{y}_i}{\tau} - \frac{e^{2\pi\mathrm{i}\,\mathrm{ad}(\sum_k z_k\bar{x}_k)} - 1}{\mathrm{ad}(\sum_k z_k\bar{x}_k)}\left(\sum_{j|j\neq i}\frac{z_{ji}}{\tau}\bar{t}_{ij}\right)\\
&\quad= -\frac{\bar{y}_i}{\tau} - \sum_{j|j\neq i}\frac{e^{2\pi\mathrm{i}\,\mathrm{ad}(\sum_k z_k\bar{x}_k)} - 1}{\mathrm{ad}(\sum_k z_k\bar{x}_k)}\left(\frac{z_{ji}}{\tau}\bar{t}_{ij}\right)
\end{aligned}$$

$$= -\frac{\bar{y}_i}{\tau} - \sum_{j|j\neq i} \frac{e^{2\pi\mathrm{i}\mathrm{ad}(z_{ij}\bar{x}_i)}-1}{\mathrm{ad}(z_{ij}\bar{x}_i)}\left(\frac{z_{ji}}{\tau}\bar{t}_{ij}\right)$$
$$= -\frac{\bar{y}_i}{\tau} + \sum_{j|j\neq i} \frac{e^{2\pi\mathrm{i}\mathrm{ad}(z_{ij}\bar{x}_i)}-1}{\mathrm{ad}(\bar{x}_i)}\left(\frac{\bar{t}_{ij}}{\tau}\right);$$

therefore

$$\frac{1}{\tau}\left(\sum_j \frac{e^{2\pi\mathrm{i}z_{ij}\mathrm{ad}\,\bar{x}_i}-1}{\mathrm{ad}\,\bar{x}_i}(\bar{t}_{ij}) - \bar{y}_i\right) = -\mathrm{Ad}\left(\tau^d e^{\frac{2\pi\mathrm{i}}{\tau}(\sum_i z_i\bar{x}_i+X)}\right)(\bar{y}_i) + 2\pi\mathrm{i}\bar{x}_i. \tag{10}$$

We have $\theta(z/\tau|-1/\tau) = (1/\tau)e^{(\pi\mathrm{i}/\tau)z^2}\theta(z|\tau)$; therefore

$$\frac{1}{\tau}k\left(\frac{z}{\tau},x|-\frac{1}{\tau}\right) = e^{2\pi\mathrm{i}zx}k(z,\tau x|\tau) + \frac{e^{2\pi\mathrm{i}zx}-1}{x\tau}. \tag{11}$$

Substituting $(z,x) = (z_{ij},\mathrm{ad}\,\bar{x}_i)$ $(j\neq i)$, applying to $\bar{t}_{ij}$, summing over j and adding up identity (10), we get

$$\frac{1}{\tau}\left(\sum_{j|j\neq i} k\left(\frac{z_{ij}}{\tau},\mathrm{ad}\,\bar{x}_i|-\frac{1}{\tau}\right)(\bar{t}_{ij}) - \bar{y}_i\right)$$
$$= \sum_{j|j\neq i} e^{2\pi\mathrm{i}z_{ij}\mathrm{ad}\,\bar{x}_i}k(z_{ij},\tau\mathrm{ad}\,\bar{x}_i|\tau)(\bar{t}_{ij})$$
$$-\mathrm{Ad}\left(\tau^d e^{\frac{2\pi\mathrm{i}}{\tau}(\sum_i z_i\bar{x}_i+X)}\right)(\bar{y}_i) + 2\pi\mathrm{i}\bar{x}_i.$$

Since

$$e^{2\pi\mathrm{i}z_{ij}\mathrm{ad}\,\bar{x}_i}k(z_{ij},\tau\mathrm{ad}\,\bar{x}_i|\tau)(\bar{t}_{ij})$$
$$= \mathrm{Ad}(\tau^d e^{(2\pi\mathrm{i}/\tau)(\sum_i z_i\bar{x}_i+X)})\big(k(z_{ij},\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})\big),$$

this implies (9). This ends the proof of (6).

Let us now check the shift identities (7) in $\bar{\Delta}(\mathbf{z}|\tau)$. The first part is immediate; let us check the last identity. We have $k(z+\tau,x|\tau) = e^{-2\pi\mathrm{i}x}g(z,x|\tau) + (e^{-2\pi\mathrm{i}x}-1)/x$, therefore $g(z+\tau,x|\tau) = e^{-2\pi\mathrm{i}x}g(z,x|\tau) - 2\pi\mathrm{i}e^{-2\pi\mathrm{i}x}k(z,x|\tau) + \frac{1}{x}\left(\frac{1-e^{-2\pi\mathrm{i}x}}{x} - 2\pi\mathrm{i}e^{-2\pi\mathrm{i}x}\right)$. Substituting $(z,x) = (z_{ij},\mathrm{ad}\,\bar{x}_i)$ $(j\neq i)$, applying to $\bar{t}_{ij}$, summing up and adding up $\sum_{k,l|k,l\neq j} g(z_{kl},\mathrm{ad}\,\bar{x}_k|\tau)(\bar{t}_{kl})$, we get that $g(\mathbf{z}+\tau\delta_i|\tau)$ equals

$$e^{-2\pi\mathrm{i}\mathrm{ad}\,\bar{x}_i}(g(\mathbf{z}|\tau)) - 2\pi\mathrm{i}e^{-2\pi\mathrm{i}\mathrm{ad}\,\bar{x}_i}(\bar{K}_i(\mathbf{z}|\tau)+\bar{y}_i)$$
$$+\sum_{j|j\neq i}\frac{1}{\mathrm{ad}\,\bar{x}_i}\left(\frac{1-e^{-2\pi\mathrm{i}\mathrm{ad}\,\bar{x}_i}}{\mathrm{ad}\,\bar{x}_i} - 2\pi\mathrm{i}e^{-2\pi\mathrm{i}\mathrm{ad}\,\bar{x}_i}\right)(\bar{t}_{ij})$$

$$= e^{-2\pi \mathrm{i} \operatorname{ad} \bar{x}_i}(g(\mathbf{z}|\tau)) - 2\pi\mathrm{i} e^{-2\pi\mathrm{i}\operatorname{ad}\bar{x}_i}(\bar{K}_i(\mathbf{z}|\tau) + \bar{y}_i)$$
$$- \left(\frac{1 - e^{-2\pi\mathrm{i}\operatorname{ad}\bar{x}_i}}{\operatorname{ad}\bar{x}_i} - 2\pi\mathrm{i} e^{-2\pi\mathrm{i}\operatorname{ad}\bar{x}_i}\right)(\bar{y}_i)$$
$$= e^{-2\pi\mathrm{i}\operatorname{ad}\bar{x}_i}(g(\mathbf{z}|\tau)) - 2\pi\mathrm{i} e^{-2\pi\mathrm{i}\operatorname{ad}\bar{x}_i}(\bar{K}_i(\mathbf{z}|\tau)) - \frac{1 - e^{-2\pi\mathrm{i}\operatorname{ad}\bar{x}_i}}{\operatorname{ad}\bar{x}_i}(\bar{y}_i);$$

on the other hand, we have $e^{-2\pi\mathrm{i}\operatorname{ad}\bar{x}_i}(\Delta_0) = \Delta_0 + \frac{1-e^{-2\pi\mathrm{i}\operatorname{ad}\bar{x}_i}}{\operatorname{ad}\bar{x}_i}(\bar{y}_i)$ (since $[\Delta_0, \bar{x}_i] = \bar{y}_i$); therefore $g(\mathbf{z}+\delta_i|\tau) - \Delta_0 = e^{-2\pi\mathrm{i}\operatorname{ad}\bar{x}_i}(g(\mathbf{z}|\tau) - \Delta_0 - 2\pi\mathrm{i}\bar{K}_i(\mathbf{z}|\tau))$. Since the δ_{2n} commute with $\bar{x}_i$, we get $\bar{\Delta}(\mathbf{z} + \tau\delta_i|\tau) = e^{-2\pi\mathrm{i}\operatorname{ad}\bar{x}_i}(\bar{\Delta}(\mathbf{z}|\tau) - \bar{K}_i(\mathbf{z}|\tau))$, as desired.

Let us now check the equivariance identities (8) for $\bar{\Delta}(\mathbf{z}|\tau)$. As above, the cocycle identities imply that it suffices to check (8) for $g = S, T$. When $g = S$, this identity follows from $\sum_i \bar{K}_i(\mathbf{z}|\tau) = 0$. When $g = T$, it is written

$$\frac{1}{\tau^2}\bar{\Delta}\left(\frac{\mathbf{z}}{\tau}| - \frac{1}{\tau}\right) = \operatorname{Ad}(c_T(\mathbf{z}|\tau))\Big(\bar{\Delta}(\mathbf{z}|\tau) + \frac{1}{\tau}\sum_i z_i\bar{K}_i(\mathbf{z}|\tau)\Big) + \frac{d}{\tau} - 2\pi\mathrm{i}X. \quad (12)$$

The modularity identity (11) for $k(z, x|\tau)$ implies that

$$\frac{1}{\tau^2} g\left(\frac{z}{\tau}, x| - \frac{1}{\tau}\right) = e^{2\pi\mathrm{i}zx} g(z, \tau x|\tau) + \frac{2\pi\mathrm{i}z}{\tau} e^{2\pi\mathrm{i}zx} k(z, \tau x|\tau)$$
$$+ \frac{1 - e^{2\pi\mathrm{i}zx}}{\tau^2 x^2} + \frac{2\pi\mathrm{i}z}{\tau^2}\frac{e^{2\pi\mathrm{i}zx}}{x}.$$

This implies that

$$\frac{1}{\tau^2}\sum_{i<j} g\left(\frac{z_{ij}}{\tau}, \operatorname{ad}\bar{x}_i| - \frac{1}{\tau}\right)(\bar{t}_{ij}) =$$
$$\sum_{i<j} e^{2\pi\mathrm{i}z_{ij}\operatorname{ad}\bar{x}_i} g(z_{ij}, \tau\operatorname{ad}\bar{x}_i|\tau)(\bar{t}_{ij})$$
$$+ \sum_{i<j}\frac{2\pi\mathrm{i}}{\tau} z_{ij} e^{2\pi\mathrm{i}z_{ij}\operatorname{ad}\bar{x}_i} k(z_{ij}, \tau\operatorname{ad}\bar{x}_i|\tau)(\bar{t}_{ij})$$
$$+ \sum_{i<j}\left(\frac{1 - e^{2\pi\mathrm{i}z_{ij}\operatorname{ad}\bar{x}_i}}{\tau^2(\operatorname{ad}\bar{x}_i)^2} + \frac{2\pi\mathrm{i}z_{ij}}{\tau^2}\frac{e^{2\pi\mathrm{i}z_{ij}\operatorname{ad}\bar{x}_i}}{\operatorname{ad}\bar{x}_i}\right)(\bar{t}_{ij}).$$

We compute as above

$$\sum_{i<j} e^{2\pi\mathrm{i}z_{ij}\operatorname{ad}\bar{x}_i} g(z_{ij}, \tau\operatorname{ad}\bar{x}_i|\tau)(\bar{t}_{ij})$$
$$= \operatorname{Ad}\left(\tau^d e^{\frac{2\pi i}{\tau}(\sum_i z_i\bar{x}_i + X)}\right)(g(\mathbf{z}|\tau)),$$

$$\sum_{i<j} \frac{2\pi\mathrm{i}}{\tau} z_{ij} e^{2\pi\mathrm{i} z_{ij} \mathrm{ad}\, \bar{x}_i} k(z_{ij}, \tau \mathrm{ad}\, \bar{x}_i | \tau)(\bar{t}_{ij})$$
$$= \sum_i \frac{2\pi\mathrm{i}}{\tau} z_i \left(\sum_{j|j\neq i} e^{2\pi\mathrm{i} z_{ij} \mathrm{ad}\, \bar{x}_i} k(z_{ij}, \tau \mathrm{ad}\, \bar{x}_i|\tau)(\bar{t}_{ij}) \right)$$

(using $k(z, x|\tau) + k(-z, -x|\tau) = 0$), and

$$\sum_{i<j} e^{2\pi\mathrm{i} z_{ij} \mathrm{ad}\, \bar{x}_i} k(z_{ij}, \tau \mathrm{ad}\, \bar{x}_i|\tau)(\bar{t}_{ij}) = \mathrm{Ad}\left(\tau^d e^{\frac{2\pi i}{\tau}(\sum_i z_i \bar{x}_i + X)}\right)(\bar{K}_i(\mathbf{z}|\tau) + \bar{y}_i).$$

Therefore

$$\frac{1}{\tau^2} g\left(\frac{\mathbf{z}}{\tau}| - \frac{1}{\tau}\right) = \mathrm{Ad}(c_T(\mathbf{z}|\tau))\Big(g(\mathbf{z}|\tau) + \frac{2\pi\mathrm{i}}{\tau}\sum_i z_i \bar{K}_i(\mathbf{z}|\tau) + \frac{2\pi\mathrm{i}}{\tau}\sum_i z_i \bar{y}_i\Big)$$
$$+ \sum_{i<j}\left(\frac{1 - e^{2\pi\mathrm{i} z_{ij} \mathrm{ad}\, \bar{x}_i}}{\tau^2 (\mathrm{ad}\, \bar{x}_i)^2} + \frac{2\pi\mathrm{i} z_{ij}}{\tau^2} \frac{e^{2\pi\mathrm{i} z_{ij} \mathrm{ad}\, \bar{x}_i}}{\mathrm{ad}\, \bar{x}_i}\right)(\bar{t}_{ij}),$$

which implies

$$\frac{1}{\tau^2}\bar{\Delta}\left(\frac{\mathbf{z}}{\tau}| - \frac{1}{\tau}\right) = \mathrm{Ad}(c_T(\mathbf{z}|\tau))\left(\bar{\Delta}(\mathbf{z}|\tau) + \frac{1}{\tau}\sum_i \bar{K}_i(\mathbf{z}|\tau)\right)$$
$$+ \mathrm{Ad}(c_T(\mathbf{z}|\tau))\left(\frac{1}{\tau}\sum_i z_i \bar{y}_i\right)$$
$$+ \frac{1}{2\pi\mathrm{i}}\sum_{i<j}\left(\frac{1 - e^{2\pi\mathrm{i} z_{ij} \mathrm{ad}\, \bar{x}_i}}{\tau^2 (\mathrm{ad}\, \bar{x}_i)^2} + \frac{2\pi\mathrm{i} z_{ij}}{\tau^2} \frac{e^{2\pi\mathrm{i} z_{ij} \mathrm{ad}\, \bar{x}_i}}{\mathrm{ad}\, \bar{x}_i}\right)(\bar{t}_{ij})$$
$$+ \frac{1}{2\pi\mathrm{i}}\big(\mathrm{Ad}(c_T(\mathbf{z}|\tau))(\Delta_{\varphi(*|\tau)}) - \frac{1}{\tau^2}\Delta_{\varphi(*|-1/\tau)}\big).$$

To prove (12), it then suffices to prove

$$\mathrm{Ad}(c_T(\mathbf{z}|\tau))\left(\frac{1}{\tau}\sum_i z_i \bar{y}_i\right) + \frac{1}{2\pi\mathrm{i}}\sum_{i<j}\left(\frac{1 - e^{2\pi\mathrm{i} z_{ij} \mathrm{ad}\, \bar{x}_i}}{\tau^2 (\mathrm{ad}\, \bar{x}_i)^2} + \frac{2\pi\mathrm{i} z_{ij}}{\tau^2} \frac{e^{2\pi\mathrm{i} z_{ij} \mathrm{ad}\, \bar{x}_i}}{\mathrm{ad}\, \bar{x}_i}\right)(\bar{t}_{ij})$$
$$+ \frac{1}{2\pi\mathrm{i}}\big(\mathrm{Ad}(c_T(\mathbf{z}|\tau))(\Delta_{\varphi(*|\tau)}) - \frac{1}{\tau^2}\Delta_{\varphi(*|-1/\tau)}\big) = \frac{d}{\tau} - 2\pi\mathrm{i}X. \tag{13}$$

We compute

$$\mathrm{Ad}(c_T(\mathbf{z}|\tau))\left(\frac{1}{\tau}\sum_i z_i \bar{y}_i\right) = \frac{1}{\tau^2}\sum_i z_i \bar{y}_i + \frac{2\pi\mathrm{i}}{\tau}\sum_i z_i \bar{x}_i$$
$$+ \sum_{i<j}\left(-\frac{1}{\tau^2}\right) z_{ij} \frac{e^{2\pi\mathrm{i} z_{ij} \mathrm{ad}\, \bar{x}_i} - 1}{\mathrm{ad}\, \bar{x}_i}(\bar{t}_{ij}).$$

We also have $\mathrm{Ad}(c_T(\mathbf{z}|\tau))(E_{2n+2}(\tau)\delta_{2n}) = \frac{1}{\tau^2}E_{2n+2}\left(-\frac{1}{\tau}\right)\delta_{2n}$ since $[\delta_{2n}, \bar{x}_i] = [\delta_{2n}, X] = 0$ and $[d, \delta_{2n}] = 2n\delta_{2n}$, and since $E_{2n+2}(-1/\tau) = \tau^{2n+2}E_{2n+2}(\tau)$, this implies

$$\mathrm{Ad}(c_T(\mathbf{z}|\tau))(\delta_{\varphi(*|\tau)}) = \delta_{\varphi(*|-1/\tau)}.$$

We now compute $\mathrm{Ad}(c_T(\mathbf{z}|\tau))(\Delta_0) - (1/\tau^2)\Delta_0$. We have

$$\mathrm{Ad}(c_T(\mathbf{z}|\tau))(\Delta_0) = \mathrm{Ad}(e^{2\pi\mathrm{i}\sum_i z_i\bar{x}_i}) \circ \mathrm{Ad}(\tau^d e^{(2\pi\mathrm{i}/\tau)X})(\Delta_0)$$

and

$$\mathrm{Ad}(\tau^d e^{(2\pi\mathrm{i}/\tau)X})(\Delta_0) = (1/\tau^2)\Delta_0 + (2\pi\mathrm{i}/\tau)d - (2\pi\mathrm{i})^2 X.$$

Now $\mathrm{Ad}(e^{2\pi\mathrm{i}\sum_i z_i\bar{x}_i})(X) = X$, $\mathrm{Ad}(e^{2\pi\mathrm{i}\sum_i z_i\bar{x}_i})(d) = d - 2\pi\mathrm{i}\sum_i z_i\bar{x}_i$. We now compute

$$\begin{aligned}
&\mathrm{Ad}(e^{2\pi\mathrm{i}\sum_i z_i\bar{x}_i})(\Delta_0)\\
&\quad= \Delta_0 + \frac{e^{2\pi\mathrm{i}\sum_i z_i \,\mathrm{ad}\,\bar{x}_i} - 1}{2\pi\mathrm{i}\,\mathrm{ad}(\sum_i z_i\bar{x}_i)}\left(\left[2\pi\mathrm{i}\sum_i z_i\bar{x}_i, \Delta_0\right]\right)\\
&\quad= \Delta_0 - \frac{e^{2\pi\mathrm{i}\sum_i z_i \,\mathrm{ad}\,\bar{x}_i} - 1}{\mathrm{ad}(\sum_i z_i\bar{x}_i)}\left(\sum_i z_i\bar{y}_i\right)\\
&\quad= \Delta_0 - \sum_i \frac{e^{2\pi\mathrm{i}\sum_{j|j\neq i} z_{ji}\,\mathrm{ad}\,\bar{x}_j} - 1}{\mathrm{ad}(\sum_{j|j\neq i} z_{ji}\bar{x}_j)}(z_i\bar{y}_i)\\
&\quad= \Delta_0 - \sum_i \left(2\pi\mathrm{i} z_i\bar{y}_i + \frac{1}{\mathrm{ad}(\sum_{j|j\neq i} z_{ji}\bar{x}_j)}\left(\frac{e^{2\pi\mathrm{i}\sum_{j|j\neq i} z_{ji}\,\mathrm{ad}\,\bar{x}_j} - 1}{\mathrm{ad}(\sum_{j|j\neq i} z_{ji}\bar{x}_j)} - 2\pi\mathrm{i}\right)\right.\\
&\qquad\quad \left.\left(\left[\sum_{j|j\neq i} z_{ji}\bar{x}_j, z_i\bar{y}_i\right]\right)\right)\\
&\quad= \Delta_0 - \sum_i 2\pi\mathrm{i} z_i\bar{y}_i - \sum_{i\neq j}\left(\frac{1}{\mathrm{ad}(\bar{x}_j)}\left(\frac{e^{2\pi\mathrm{i} z_{ji}\,\mathrm{ad}\,\bar{x}_j} - 1}{\mathrm{ad}(z_{ji}\bar{x}_j)} - 2\pi\mathrm{i}\right)(z_i\bar{t}_{ij})\right);
\end{aligned}$$

the last sum decomposes as

$$\begin{aligned}
&\sum_{i<j}\frac{1}{\mathrm{ad}(\bar{x}_j)}\left(\frac{e^{2\pi\mathrm{i} z_{ji}\,\mathrm{ad}\,\bar{x}_j} - 1}{\mathrm{ad}(z_{ji}\bar{x}_j)} - 2\pi\mathrm{i}\right)(z_i\bar{t}_{ij})\\
&\qquad + \sum_{i>j}\frac{1}{\mathrm{ad}(\bar{x}_j)}\left(\frac{e^{2\pi\mathrm{i} z_{ji}\,\mathrm{ad}\,\bar{x}_j} - 1}{\mathrm{ad}(z_{ji}\bar{x}_j)} - 2\pi\mathrm{i}\right)(z_i\bar{t}_{ij})\\
&\quad= \sum_{i<j}\frac{1}{\mathrm{ad}(\bar{x}_j)}\left(\frac{e^{2\pi\mathrm{i} z_{ji}\,\mathrm{ad}\,\bar{x}_j} - 1}{\mathrm{ad}(z_{ji}\bar{x}_j)} - 2\pi\mathrm{i}\right)(z_i\bar{t}_{ij})\\
&\qquad + \frac{1}{\mathrm{ad}(\bar{x}_i)}\left(\frac{e^{2\pi\mathrm{i} z_{ij}\,\mathrm{ad}\,\bar{x}_i} - 1}{\mathrm{ad}(z_{ij}\bar{x}_i)} - 2\pi\mathrm{i}\right)(z_j\bar{t}_{ij})\\
&\quad= \sum_{i<j}\frac{1}{\mathrm{ad}(\bar{x}_i)}\left(\frac{e^{2\pi\mathrm{i} z_{ij}\,\mathrm{ad}\,\bar{x}_i} - 1}{\mathrm{ad}(z_{ij}\bar{x}_i)} - 2\pi\mathrm{i}\right)(z_{ji}\bar{t}_{ij}),
\end{aligned}$$

so

$$\mathrm{Ad}(e^{2\pi\mathrm{i}\sum_i z_i\bar{x}_i})(\Delta_0) = \Delta_0 - 2\pi\mathrm{i}\sum_i z_i\bar{y}_i - \sum_{i<j}\frac{1}{\mathrm{ad}(\bar{x}_i)}\left(\frac{e^{2\pi\mathrm{i}z_{ij}\,\mathrm{ad}\,\bar{x}_i}-1}{\mathrm{ad}(z_{ij}\bar{x}_i)} - 2\pi\mathrm{i}\right)(z_{ji}\bar{t}_{ij});$$

and finally

$$\begin{aligned}&\mathrm{Ad}(c_T(\mathbf{z}|\tau))(\Delta_{\varphi(*|\tau)}) - \frac{1}{\tau^2}\Delta_{\varphi(*|-1/\tau)}\\ &\quad= -\frac{2\pi\mathrm{i}}{\tau^2}\sum_i z_i\bar{y}_i - \frac{1}{\tau^2}\sum_{i<j}\frac{1}{\mathrm{ad}(\bar{x}_i)}\left(\frac{e^{2\pi\mathrm{i}z_{ij}\,\mathrm{ad}\,\bar{x}_i}-1}{\mathrm{ad}(z_{ij}\bar{x}_i)} - 2\pi\mathrm{i}\right)(z_{ji}\bar{t}_{ij})\\ &\qquad+\frac{2\pi\mathrm{i}}{\tau}\left(d - 2\pi\mathrm{i}\sum_i z_i\bar{x}_i\right) - (2\pi\mathrm{i})^2 X,\end{aligned}$$

which implies (13). This proves (12) and therefore (8).

We prove the flatness identity $[\partial/\partial\tau - \bar{\Delta}(\mathbf{z}|\tau), \partial/\partial z_i - \bar{K}_i(\mathbf{z}|\tau)] = 0$. For this, we prove that $(\partial/\partial\tau)\bar{K}_i(\mathbf{z}|\tau) = (\partial/\partial\tau)\bar{\Delta}(\mathbf{z}|\tau)$ and $[\bar{\Delta}(\mathbf{z}|\tau), \bar{K}_i(\mathbf{z}|\tau)] = 0$.

Let us first prove that

$$(\partial/\partial\tau)\bar{K}_i(\mathbf{z}|\tau) = (\partial/\partial z_i)\bar{\Delta}(\mathbf{z}|\tau). \tag{14}$$

We have

$$(\partial/\partial\tau)\bar{K}_i(\mathbf{z}|\tau) = \sum_{j|j\neq i}(\partial_\tau k)(z_{ij}, \mathrm{ad}\,\bar{x}_i|\tau)(\bar{t}_{ij})$$

and $(\partial/\partial z_i)\bar{\Delta}(\mathbf{z}|\tau) = (2\pi\mathrm{i})^{-1}\sum_{j|j\neq i}(\partial_z g)(z_{ij}, \mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})$ (where $\partial_\tau := \partial/\partial\tau$, $\partial_z = \partial/\partial z$), so it suffices to prove the identity

$$(\partial_\tau k)(z,x|\tau) = (2\pi\mathrm{i})^{-1}(\partial_z g)(z,x|\tau),$$

i.e., $(\partial_\tau k)(z,x|\tau) = (2\pi\mathrm{i})^{-1}(\partial_z\partial_x k)(z,x|\tau)$. In this identity, $k(z,x|\tau)$ may be replaced by $\tilde{k}(z,x|\tau) := k(z,x|\tau) + 1/x = \theta(z+x|\tau)/(\theta(z|\tau)\theta(x|\tau))$. Dividing by $\tilde{k}(z,x|\tau)$, the desired identity is rewritten as

$$\begin{aligned}&2\pi\mathrm{i}\left(\frac{\partial_\tau\theta}{\theta}(z+x|\tau) - \frac{\partial_\tau\theta}{\theta}(z|\tau) - \frac{\partial_\tau\theta}{\theta}(x|\tau)\right)\\ &\quad= \left(\frac{\theta'}{\theta}\right)'(z+x|\tau) + \left(\frac{\theta'}{\theta}(z+x|\tau) - \frac{\theta'}{\theta}(z|\tau)\right)\left(\frac{\theta'}{\theta}(z+x|\tau) - \frac{\theta'}{\theta}(x|\tau)\right)\end{aligned}$$

(recall that $f'(z|\tau) = \partial_z f(z|\tau)$), or taking into account the heat equation $4\pi\mathrm{i}(\partial_\tau\theta/\theta)(z|\tau) = (\theta''/\theta)(z|\tau) - 12\pi\mathrm{i}(\partial_\tau\eta/\eta)(\tau)$, as follows:

$$2\Big(\tfrac{\theta'}{\theta}(z|\tau)\tfrac{\theta'}{\theta}(x|\tau)-\tfrac{\theta'}{\theta}(x|\tau)\tfrac{\theta'}{\theta}(z+x|\tau)-\tfrac{\theta'}{\theta}(z|\tau)\tfrac{\theta'}{\theta}(z+x|\tau)\Big) \tag{15}$$
$$+\tfrac{\theta''}{\theta}(z|\tau)+\tfrac{\theta''}{\theta}(x|\tau)+\tfrac{\theta''}{\theta}(z+x|\tau)-12\pi\mathrm{i}\tfrac{\partial_\tau\eta}{\eta}(\tau)=0.$$

Let us prove (15). Denote its l.h.s. by $F(z,x|\tau)$. Since $\theta(z|\tau)$ is odd w.r.t. z, $F(z,x|\tau)$ is invariant under the permutation of $z,x,-z-x$. The identities

$$(\theta'/\theta)(z+\tau|\tau)=(\theta'/\theta)(z|\tau)-2\pi\mathrm{i}$$

and

$$(\theta''/\theta)(z+\tau|\tau)=(\theta''/\theta)(z|\tau)-4\pi\mathrm{i}(\theta'/\theta)(z|\tau)+(2\pi\mathrm{i})^2$$

imply that $F(z,x|\tau)$ is elliptic in z,x (w.r.t. the lattice Λ_τ). The possible poles of $F(z,x|\tau)$ as a function of z are simple at $z=0$ and $z=-x$ (mod Λ_τ), but one checks that $F(z,x|\tau)$ is regular at these points, so it is constant in z. By the $\mathfrak{S}_3$-symmetry, it is also constant in x; hence it is a function of τ only: $F(z,x|\tau)=F(\tau)$.

To compute this function, we compute

$$F(z,0|\tau)=[-2(\theta'/\theta)'-2(\theta'/\theta)^2+2\theta''/\theta](z|\tau)+(\theta''/\theta)(0|\tau)-12\pi\mathrm{i}(\partial_\tau\eta/\theta)(\tau);$$

hence

$$F(\tau)=(\theta''/\theta)(0|\tau)-12\pi\mathrm{i}(\partial_\tau\eta/\eta)(\tau);$$

The above heat equation then implies that $F(\tau)=4\pi\mathrm{i}(\partial_\tau\theta/\theta)(0|\tau)$. Now $\theta'(0|\tau)=1$ implies that $\theta(z|\tau)$ has the expansion $\theta(z|\tau)=z+\sum_{n\geq 2}a_n(\tau)z^n$ as $z\to 0$, which implies $(\partial_\tau\theta/\theta)(0|\tau)=0$. So $F(\tau)=0$, which implies (15) and therefore (14).

We now prove

$$[\bar{\Delta}(\mathbf{z}|\tau),\bar{K}_i(\mathbf{z}|\tau)]=0. \tag{16}$$

Since τ is constant in what follows, we will write $k(z,x)$, $g(z,x)$, φ instead of $k(z,x|\tau)$, $g(z,x|\tau)$, $\varphi(*|\tau)$. For $i\neq j$, let us set $g_{ij}:=g(z_{ij},\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})$. Since $g(z,x|\tau)=g(-z,-x|\tau)$, we have $g_{ij}=g_{ji}$. Recall that $\bar{K}_{ij}=k(z_{ij},\mathrm{ad}\,\bar{x}_i)(t_{ij})$.

We have

$$\begin{aligned}
&2\pi\mathrm{i}[\bar{\Delta}(\mathbf{z}|\tau),\bar{K}_i(\mathbf{z}|\tau)] \qquad (17)\\
&=\Big[-\Delta_\varphi+\sum_{i,j|i<j}g_{ij},-\bar{y}_i+\sum_{j|j\neq i}\bar{K}_{ij}\Big]\\
&=[\Delta_\varphi,\bar{y}_i]+\sum_{j|j\neq i}\Big(-[\Delta_\varphi,\bar{K}_{ij}]+[\bar{y}_i,g_{ij}]+[g_{ij},\bar{K}_{ij}]\Big)\\
&\quad+\sum_{j,k|j\neq i,k\neq i,j<k}\big([\bar{y}_i,g_{jk}]+[g_{ik}+g_{jk},\bar{K}_{ij}]+[g_{ij}+g_{jk},\bar{K}_{ik}]\big).
\end{aligned}$$

One computes

$$[\Delta_\varphi, \bar{y}_i] = \sum_\alpha [f_\alpha(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij}), g_\alpha(-\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})], \tag{18}$$

where $\sum_\alpha f_\alpha(u)g_\alpha(v) = \frac{1}{2}\frac{\varphi(u)-\varphi(v)}{u-v}$. If $f(x) \in \mathbb{C}[[x]]$, then

$$[\Delta_0, f(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})] - [\bar{y}_i, f'(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})] = \sum_\alpha [h_\alpha(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij}), k_\alpha(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})]$$
$$+ \sum_{k|k\neq i,j} \frac{f(\mathrm{ad}\,\bar{x}_i) - f(-\mathrm{ad}\,\bar{x}_j) - f'(-\mathrm{ad}\,\bar{x}_j)(\mathrm{ad}\,\bar{x}_i + \mathrm{ad}\,\bar{x}_j)}{(\mathrm{ad}\,\bar{x}_i + \mathrm{ad}\,\bar{x}_j)^2}([\bar{t}_{ij}, \bar{t}_{ik}]),$$

where

$$\sum_\alpha h_\alpha(u)k_\alpha(v) = \frac{1}{2}\Big(\frac{1}{v^2}\big(f(u+v)-f(u)-vf'(u)\big) - \frac{1}{u^2}\big(f(u+v)-f(v)-uf'(v)\big)\Big).$$

Since $g(z,x) = k_x(z,x)$, we get

$$-[\Delta_0, \bar{K}_{ij}] + [\bar{y}_i, g_{ij}] = -\sum_\alpha \left[f_\alpha^{ij}(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij}), g_\alpha^{ij}(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})\right] \tag{19}$$
$$+ \sum_{k|k\neq i,j} \frac{k(z_{ij}, \mathrm{ad}\,\bar{x}_i) - k(z_{ij}, -\mathrm{ad}\,\bar{x}_j) - (\mathrm{ad}\,\bar{x}_i + \mathrm{ad}\,\bar{x}_j)k_x(z_{ij}, -\mathrm{ad}\,\bar{x}_j)}{(\mathrm{ad}\,\bar{x}_i + \mathrm{ad}\,\bar{x}_j)^2}([\bar{t}_{ij}, \bar{t}_{jk}]),$$

where $\sum_\alpha f_\alpha^{ij}(u)g_\alpha^{ij}(v)$ equals

$$\tfrac{1}{2}\Big(\tfrac{1}{v^2}\big(k(z_{ij}, u+v) - k(z_{ij}, u) - vk_x(z_{ij}, u)\big)$$
$$- \tfrac{1}{u^2}\big(k(z_{ij}, u+v) - k(z_{ij}, v) - uk_x(z_{ij}, v)\big)\Big).$$

For $f(x) \in \mathbb{C}[[x]]$, we have

$$[\delta_\varphi, f(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})] = \sum_\alpha [l_\alpha(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij}), m_\alpha(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})],$$

where $\sum_\alpha l_\alpha(u)m_\alpha(v) = f(u+v)\varphi(v)$; therefore

$$-[\delta_\varphi, \bar{K}_{ij}] = -\sum_\alpha \left[l_\alpha^{ij}(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij}), m_\alpha^{ij}(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})\right], \tag{20}$$

where $\sum_\alpha l_\alpha^{ij}(u)m_\alpha^{ij}(v) = k(z_{ij}, u+v)\varphi(v)$.

For $j, k \neq i$ and $j < k$, we have

$$[\bar{y}_i, g_{jk}] + [g_{ik} + g_{jk}, \bar{K}_{ij}] + [g_{ij} + g_{jk}, \bar{K}_{ik}]$$
$$= [\bar{y}_i, g_{jk}] - [g_{ki}, \bar{K}_{ji}] - [g_{ji}, \bar{K}_{ki}] + [g_{jk}, \bar{K}_{ij}] + [g_{jk}, \bar{K}_{ik}],$$

and since for any $f(x) \in \mathbb{C}[[x]]$,

$$[\bar{y}_i, f(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{jk})] = -\frac{f(\mathrm{ad}\,\bar{x}_j) - f(-\mathrm{ad}\,\bar{x}_k)}{\mathrm{ad}\,\bar{x}_j + \mathrm{ad}\,\bar{x}_k}([\bar{t}_{ij}, \bar{t}_{jk}]),$$

we get

$$\begin{aligned}
&[\bar{y}_i, g_{jk}] + [g_{ik} + g_{jk}, \bar{K}_{ij}] + [g_{ij} + g_{jk}, \bar{K}_{ik}] \\
&\quad = \Big(- \frac{g(z_{jk}, \mathrm{ad}\,\bar{x}_j) - g(z_{jk}, -\mathrm{ad}\,\bar{x}_k)}{\mathrm{ad}\,\bar{x}_j + \mathrm{ad}\,\bar{x}_k} \\
&\qquad -g(z_{ki}, \mathrm{ad}\,\bar{x}_k)k(z_{ji}, \mathrm{ad}\,\bar{x}_j) + g(z_{ji}, \mathrm{ad}\,\bar{x}_j)k(z_{ki}, \mathrm{ad}\,\bar{x}_k) \\
&\qquad -g(z_{kj}, \mathrm{ad}\,\bar{x}_k)k(z_{ij}, \mathrm{ad}\,\bar{x}_i) + g(z_{jk}, \mathrm{ad}\,\bar{x}_j)k(z_{ik}, \mathrm{ad}\,\bar{x}_i)\Big)([\bar{t}_{ij}, \bar{t}_{jk}]).
\end{aligned} \tag{21}$$

Summing up (18), (19), (20), and (21), (17) gives

$$\begin{aligned}
2\pi\mathrm{i}[\bar{\Delta}(\mathbf{z}|\tau), \bar{K}_i(\mathbf{z}|\tau)] = &\sum_{j|j\neq i} \sum_{\alpha} \big[F_\alpha^{ij}(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij}), G_\alpha^{ij}(\mathrm{ad}\,\bar{x}_i)(\bar{t}_{ij})\big] \\
&+ \sum_{j,k|j\neq i,k\neq i} H(z_{ij}, z_{ik}, -\mathrm{ad}\,\bar{x}_j, -\mathrm{ad}\,\bar{x}_k)([t_{ij}, t_{jk}]),
\end{aligned}$$

where $\sum_\alpha F_\alpha^{ij}(u)G_\alpha^{ij}(v) = L(z_{ij}, u, v)$,

$$\begin{aligned}
L(z, u, v) = &\frac{1}{2}\frac{\varphi(u) - \varphi(v)}{u + v} + \frac{1}{2}k(z, u + v)(\varphi(u) - \varphi(v)) \\
&+\frac{1}{2}(g(z, u)k(z, v) - k(z, u)g(z, v)) \\
&-\frac{1}{2}\Big(\frac{1}{v^2}\big(k(z, u + v) - k(z, u) - vk_x(z, u)\big) \\
&-\frac{1}{u^2}\big(k(z, u + v) - k(z, v) - uk_x(z, v)\big)\Big)
\end{aligned}$$

and

$$\begin{aligned}
H(z, z', u, v) = &\frac{1}{v^2}\big(k(z, u + v) - k(z, u) - vk_x(z, u)\big) \\
&-\frac{1}{u^2}\big(k(z', u + v) - k(z', v) - uk_x(z', v)\big) \\
&+\frac{1}{u + v}\big(g(z' - z, -u) - g(z' - z, v)\big) \\
&-g(-z', -v)k(-z, -u) + g(-z, -u)k(-z', -v) \\
&-g(z - z', -v)k(z, u + v) + g(z' - z, -u)k(z', u + v).
\end{aligned}$$

Explicit computation shows that $H(z, z', u, v) = 0$, which implies that $L(z, u, v) = 0$ since $L(z, u, v) = -\frac{1}{2}H(z, z, u, v)$. This proves (16). □

Remark 13. Define $\Delta(\mathbf{z}|\tau)$ by the same formula as $\bar{\Delta}(\mathbf{z}|\tau)$, replacing $\bar{x}_i, \bar{y}_i$ by x_i, y_i. Then $d - \Delta(\mathbf{z}|\tau)d\tau - \sum_i K_i(\mathbf{z}|\tau)dz_i$ is flat. This can be interpreted as follows.

Let $N_+ \subset \mathrm{SL}_2(\mathbb{C})$ be the connected subgroup with Lie algebra $\mathbb{C}\Delta_0$. Set $\tilde{\mathbf{N}}_n := \exp((\mathfrak{t}_{1,n} \rtimes \mathfrak{d}_+)^\wedge) \rtimes N_+$, $\mathbf{N}_n := \exp((\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}_+)^\wedge) \rtimes N_+$ and $\tilde{\mathbf{G}}_n := \exp((\mathfrak{t}_{1,n} \rtimes \mathfrak{d}_+)^\wedge) \rtimes \mathrm{SL}_2(\mathbb{C})$. Then we have a diagram of groups

$$\begin{pmatrix} \tilde{\mathbf{N}}_n & \to & \mathbf{N}_n \\ \downarrow & & \downarrow \\ \tilde{\mathbf{G}}_n & \to & \mathbf{G}_n \end{pmatrix}.$$

The trivial $\mathbf{N}_n$-bundle on $(\mathfrak{H}\times\mathbb{C}^n)-\mathrm{Diag}_n$ with flat connection $d-\bar{\Delta}(\mathbf{z}|\tau)d\tau-\sum_i \bar{K}_i(\mathbf{z}|\tau)dz_i$ admits a reduction to $\tilde{\mathbf{N}}_n$, where the bundle is again trivial and the connection is $d-\Delta(\mathbf{z}|\tau)d\tau-\sum_i K_i(\mathbf{z}|\tau)dz_i$.

$((\mathbb{Z}^2)^2 \times \mathbb{C}) \rtimes \mathrm{SL}_2(\mathbb{Z})$ contains the subgroups $(\mathbb{Z}^n)^2$, $(\mathbb{Z}^n)^2 \times \mathbb{C}$, $(\mathbb{Z}^n)^2 \rtimes \mathrm{SL}_2(\mathbb{Z})$. We denote the corresponding quotients of $(\mathbb{C}^n \times \mathfrak{H}) - \mathrm{Diag}_n$ by $C(n)$, $\bar{C}(n)$, $\tilde{\mathcal{M}}_{1,n}$. These fit in the diagram

$$\begin{pmatrix} \tilde{C}(n) & \to & C(n) \\ \downarrow & & \downarrow \\ \tilde{\mathcal{M}}_{1,n} & \to & \mathcal{M}_{1,n} \end{pmatrix}.$$

The pair $(\mathcal{P}_n, \nabla_{\mathcal{P}_n})$ can be pulled back to $\mathbf{G}_n$-bundles over these covers of $\mathcal{M}_{1,n}$. These pullbacks admit G-structures, where G is the corresponding group in the above diagram of groups.

We have natural projections $C(n) \to \mathfrak{H}$, $\bar{C}(n) \to \mathfrak{H}$. The fibers of $\tau \in \mathfrak{H}$ are respectively $C(E_\tau, n)$ and $\bar{C}(E_\tau, n)$. The pair $(\mathcal{P}_n, \nabla_n)$ can be pulled back to $C(E_\tau, n)$ and $\bar{C}(E_\tau, n)$; these pullbacks admit G-structures, where $G = \exp(\hat{\mathfrak{t}}_{1,n})$ and $\exp(\hat{\bar{\mathfrak{t}}}_{1,n})$, which coincide with $(P_{n,\tau}, \nabla_{n,\tau})$ and $(\bar{P}_{n,\tau}, \bar{\nabla}_{n,\tau})$.

3.3 Bundle with flat connection over $\mathcal{M}_{1,[n]}$

The semidirect product $((\mathbb{Z}^n)^2 \times \mathbb{C}) \rtimes (\mathrm{SL}_2(\mathbb{Z}) \times S_n)$ acts on $(\mathbb{C}^n \times \mathfrak{H}) - \mathrm{Diag}_n$ as follows: the action of $((\mathbb{Z}^n)^2 \times \mathbb{C}) \rtimes \mathrm{SL}_2(\mathbb{C})$ is as above and the action of S_n is $\sigma * (z_1, \ldots, z_n, \tau) := (z_{\sigma^{-1}(1)}, \ldots, z_{\sigma^{-1}(n)}, \tau)$. The quotient then is identified with $\mathcal{M}_{1,[n]}$.

We will define a principal $\mathbf{G}_n \rtimes S_n$-bundle with a flat connection $(\mathcal{P}_{[n]}, \nabla_{\mathcal{P}_{[n]}})$ over $\mathcal{M}_{1,[n]}$.

Proposition 14. *There exists a unique principal $\mathbf{G}_n \rtimes S_n$-bundle $\mathcal{P}_{[n]}$ over $\mathcal{M}_{1,[n]}$ such that a section of $U \subset \mathcal{M}_{1,[n]}$ is a function $f : \tilde{\pi}^{-1}(U) \to \mathbf{G}_n \rtimes S_n$, satisfying the conditions of Proposition 10 as well as $f(\sigma\mathbf{z}|\tau) = \sigma f(\mathbf{z}|\tau)$ for $\sigma \in S_n$ (here $\tilde{\pi} : (\mathbb{C}^n \times \mathfrak{H}) - \mathrm{Diag}_n \to \mathcal{M}_{1,[n]}$ is the canonical projection).*

Proof. One checks that $\sigma c_{\tilde{g}}(\mathbf{z}|\tau)\sigma^{-1} = c_{\sigma\tilde{g}\sigma^{-1}}(\sigma^{-1}\mathbf{z})$, where $\tilde{g} \in ((\mathbb{Z}^n)^2 \times \mathbb{C}) \rtimes \mathrm{SL}_2(\mathbb{Z})$, $\sigma \in S_n$. It follows that there is a unique cocycle $c_{(\tilde{g},\sigma)} : \mathbb{C}^n \times \mathfrak{H} \to \mathbf{G}_n \rtimes S_n$ such that $c_{(\tilde{g},1)} = c_{\tilde{g}}$ and $c_{(1,\sigma)}(\mathbf{z}|\tau) = \sigma$. □

Theorem 15. *There is a unique flat connection* $\nabla_{\mathcal{P}_{[n]}}$ *on* $\mathcal{P}_{[n]}$ *whose pullback to* $(\mathbb{C}^n \times \mathfrak{H}) - \mathrm{Diag}_n$ *is the connection* $d - \bar{\Delta}(\mathbf{z}|\tau)d\tau - \sum_i \bar{K}_i(\mathbf{z}|\tau)dz_i$ *on the trivial* $\mathbf{G}_n \rtimes S_n$*-bundle.*

Proof. Taking into account Theorem 12, it remains to show that this connection is S_n-equivariant. We have already mentioned that $\sum_i \bar{K}_i(\mathbf{z}|\tau)dz_i$ is equivariant; $\bar{\Delta}(\mathbf{z}|\tau)$ is also checked to be equivariant. □

4 The monodromy morphisms $\boldsymbol{\Gamma_{1,[n]}} \longrightarrow \mathbf{G}_{\boldsymbol{n}} \rtimes \boldsymbol{S_n}$

Let $\Gamma_{1,[n]}$ be the mapping class group of genus 1 surfaces with n unordered marked points. It can be viewed as the fundamental group $\pi_1(\mathcal{M}_{1,[n]}, *)$, where $*$ is a base point at infinity that will be specified later. The flat connection on $\mathcal{M}_{1,[n]}$ introduced above gives rise to morphisms $\gamma_n : \Gamma_{1,[n]} \to \mathbf{G}_n \rtimes S_n$, which we now study. This study in divided in two parts: In the first, analytic, part, we show that γ_n can be obtained from γ_1 and γ_2, and show that the restriction of γ_n to $\overline{\mathrm{B}}_{1,n}$ can be expressed in terms of the KZ associator only. In the second part, we show that morphisms $\overline{\mathrm{B}}_{1,n} \to \exp(\hat{\bar{\mathfrak{t}}}_{1,n}) \rtimes S_n$ can be constructed algebraically using an arbitrary associator. Finally, we introduce the notion of an elliptic structure over a quasi-bialgebra.

4.1 The solution $\boldsymbol{F^{(n)}(\mathbf{z}|\tau)}$

The elliptic KZB system is now

$$(\partial/\partial z_i)F(\mathbf{z}|\tau) = \bar{K}_i(\mathbf{z}|\tau)F(\mathbf{z}|\tau), \quad (\partial/\partial\tau)F(\mathbf{z}|\tau) = \bar{\Delta}(\mathbf{z}|\tau)F(\mathbf{z}|\tau),$$

where $F(\mathbf{z}|\tau)$ is a function $(\mathbb{C}^n \times \mathfrak{H}) - \mathrm{Diag}_n \supset U \to \mathbf{G}_n \rtimes S_n$ invariant under translation by $\mathbb{C}(\sum_i \delta_i)$. Let $D_n := \{(\mathbf{z},\tau) \in \mathbb{C}^n \times \mathfrak{H} | z_i = a_i + b_i\tau, a_i, b_i \in \mathbb{R}, a_1 < a_2 < \cdots < a_n < a_1 + 1, b_1 < b_2 < \cdots < b_n < b_1 + 1\}$. Then $D_n \subset (\mathbb{C}^n \times \mathfrak{H}) - \mathrm{Diag}_n$ is simply connected and invariant under $\mathbb{C}(\sum_i \delta_i)$. A solution of the elliptic KZB system on this domain is then unique, up to right multiplication by a constant. We now determine a particular solution $F^{(n)}(\mathbf{z}|\tau)$.

Let us study the elliptic KZB system in the region $z_{ij} \ll 1$, $\tau \to \mathrm{i}\infty$. Then $\bar{K}_i(\mathbf{z}|\tau) = \sum_{j|j\neq i} \bar{t}_{ij}/(z_i - z_j) + O(1)$.

We now compute the expansion of $\bar{\Delta}(\mathbf{z}|\tau)$. The heat equation for ϑ implies the expansion $\vartheta(x|\tau) = \eta(\tau)^3\big(x + 2\pi\mathrm{i}\partial_\tau \mathrm{log}\eta(\tau)x^3 + O(x^5)\big)$, so $\theta(x|\tau) = x + 2\pi\mathrm{i}\partial_\tau \mathrm{log}\,\eta(\tau)x^3 + O(x^5)$, hence

$$g(0,x|\tau) = \left(\frac{\theta'}{\theta}\right)'(x|\tau) + \frac{1}{x^2} = 4\pi\mathrm{i}\partial_\tau \mathrm{log}\,\eta(\tau) + O(x) = -(\pi^2/3)E_2(\tau) + O(x),$$

since $E_2(\tau) = \frac{24}{2\pi \mathrm{i}}\partial_\tau \log\,\eta(\tau)$. We have $g(0,x|\tau) = g(0,0|\tau) - \varphi(x|\tau)$, so

$$g(0,x|\tau) = -\sum_{k\geq 0} a_{2k}x^{2k}E_{2k+2}(\tau),$$

where $a_0 = \pi^2/3$. Then

$$\bar{\Delta}(\mathbf{z}|\tau) = -\frac{1}{2\pi\mathrm{i}}\Big(\Delta_0 + \sum_{k\geq 0} a_{2k}E_{2k+2}(\tau)\big(\delta_{2k} + \sum_{i,j|i<j}(\mathrm{ad}\,\bar{x}_i)^{2k}(\bar{t}_{ij})\big)\Big) + o(1)$$

for $z_{ij} \ll 1$ and any $\tau \in \mathfrak{H}$. Since we have an expansion $E_{2k}(\tau) = 1 + \sum_{l>0} a_{kl}e^{2\pi\mathrm{i}l\tau}$ as $\tau \to \mathrm{i}\infty$, then using Proposition 85 with $u_n = z_{n1}$, $u_{n-1} = z_{n-1,1}/z_{n1}, \ldots$, $u_2 = z_{21}/z_{31}$, and $u_1 = q = e^{2\pi\mathrm{i}\tau}$, there is a unique solution $F^{(n)}(\mathbf{z}|\tau)$ with the expansion

$$\begin{aligned} F^{(n)}(\mathbf{z}|\tau) \simeq\ & z_{21}^{\bar{t}_{12}} z_{31}^{\bar{t}_{13}+\bar{t}_{23}} \ldots z_{n1}^{\bar{t}_{1n}+\ldots+\bar{t}_{n-1,n}} \\ & \times \exp\Big(-\frac{\tau}{2\pi\mathrm{i}}\Big(\Delta_0 + \sum_{k\geq 0} a_{2k}\big(\delta_{2k} + \sum_{i<j}(\mathrm{ad}\,\bar{x}_i)^{2k}(\bar{t}_{ij})\big)\Big)\Big) \end{aligned}$$

in the region $z_{21} \ll z_{31} \ll \cdots \ll z_{n1} \ll 1$, $\tau \to \mathrm{i}\infty$, $(\mathbf{z},\tau) \in D_n$ (here $z_{ij} = z_i - z_j$); here the sign $\simeq$ means that any of the ratios of both sides has the form $1 + \sum_{k>0}\sum_{i,a_1,\ldots,a_n} r_k^{i,a_1,\ldots,a_n}(u_1,\ldots,u_n)$, where the second sum is finite with $a_i \geq 0$, $i \in \{1,\ldots,n\}$, $r_k^{i,a_1,\ldots,a_n}(u_1,\ldots,u_n)$ has degree k, and is $O(u_i(\log u_1)^{a_1}\cdots(\log u_n)^{a_n})$.

4.2 Presentation of $\Gamma_{1,[n]}$

According to [Bir69b], $\Gamma_{1,[n]} = \{\overline{\mathrm{B}}_{1,n} \rtimes \widetilde{\mathrm{SL}_2(\mathbb{Z})}\}/\mathbb{Z}$, where $\widetilde{\mathrm{SL}_2(\mathbb{Z})}$ is a central extension $1 \to \mathbb{Z} \to \widetilde{\mathrm{SL}_2(\mathbb{Z})} \to \mathrm{SL}_2(\mathbb{Z}) \to 1$; the action $\alpha : \widetilde{\mathrm{SL}_2(\mathbb{Z})} \to \mathrm{Aut}(\overline{\mathrm{B}}_{1,n})$ is such that for Z the central element $1 \in \mathbb{Z} \subset \widetilde{\mathrm{SL}_2(\mathbb{Z})}$, $\alpha_Z(x) = Z'x(Z')^{-1}$, where Z' is the image of a generator of the center of PB_n (the pure braid group of n points on the plane) under the natural morphism $\mathrm{PB}_n \to \overline{\mathrm{B}}_{1,n}$; $\overline{\mathrm{B}}_{1,n} \rtimes \widetilde{\mathrm{SL}_2(\mathbb{Z})}$ is then $\overline{\mathrm{B}}_{1,n} \times \widetilde{\mathrm{SL}_2(\mathbb{Z})}$ with the product $(p,A)(p',A') = (p\alpha_A(p'),AA')$; this semidirect product is then factored by its central subgroup (isomorphic to $\mathbb{Z}$) generated by $((Z')^{-1},Z)$.

The group $\Gamma_{1,[n]}$ is presented explicitly as follows. Generators are σ_i ($i = 1,\ldots,n-1$), A_i, B_i ($i = 1,\ldots,n$), C_{jk} ($1 \leq j < k \leq n$), Θ and Ψ, and relations are

$$\begin{gathered} \sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}\ (i = 1,\ldots,n-2), \quad \sigma_i\sigma_j = \sigma_j\sigma_i\ (1 \leq i < j \leq n), \\ \sigma_i^{-1}X_i\sigma_i^{-1} = X_{i+1}, \quad \sigma_iY_i\sigma_i = Y_{i+1}\ (i = 1,\ldots,n-1), \\ (\sigma_i, X_j) = (\sigma_i, Y_j) = 1\ (i \in \{1,\ldots,n-1\}, j \in \{1,\ldots,n\}, j \neq i, i+1), \end{gathered}$$

$$\begin{aligned}
\sigma_i^2 &= C_{i,i+1}C_{i+1,i+2}C_{i,i+2}^{-1}\ (i=1,\ldots,n-1),\\
(A_i,A_j) &= (B_i,B_j)=1(\text{any } i,j),\quad A_1=B_1=1,\\
\left(B_k,A_kA_j^{-1}\right) &= \left(B_kB_j^{-1},A_k\right)=C_{jk}\ (1\le j<k\le n),\\
(A_i,C_{jk}) &= (B_i,C_{jk})=1\ (1\le i\le j<k\le n),\\
\Theta A_i\Theta^{-1} &= B_i^{-1},\quad \Theta B_i\Theta^{-1}=B_iA_iB_i^{-1},\\
\Psi A_i\Psi^{-1} &= A_i,\quad \Psi B_i\Psi^{-1}=B_iA_i,\quad (\Theta,\sigma_i)=(\Psi,\sigma_i)=1,\\
(\Psi,\Theta^2) &= 1,\quad (\Theta\Psi)^3=\Theta^4=C_{12}\cdots C_{n-1,n}.
\end{aligned}$$

Here $X_i = A_iA_{i+1}^{-1}$, $Y_i = B_iB_{i+1}^{-1}$ for $i=1,\ldots,n$ (with the convention $A_{n+1}=B_{n+1}=C_{i,n+1}=1$). The relations imply

$$C_{jk}=\sigma_{j,j+1,\cdots,k}\cdots\sigma_{j+n-k,j+n-k+1,\ldots,n}\sigma_{j,j+1,\ldots,n-k+j+1}\cdots\sigma_{k-1,k,\ldots,n},$$

where $\sigma_{i,i+1,\ldots,j}=\sigma_{j-1}\cdots\sigma_i$. Observe that $C_{12},\ldots,C_{n-1,n}$ commute with each other.

The group $\widetilde{\mathrm{SL}_2(\mathbb{Z})}$ is presented by generators Θ,Ψ, and Z, and relations Z is central, $\Theta^4=(\Theta\Psi)^3=Z$ and $(\Psi,\Theta^2)=1$. The morphism $\widetilde{\mathrm{SL}_2(\mathbb{Z})}\to\mathrm{SL}_2(\mathbb{Z})$ is $\Theta\mapsto\left(\left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right)\right)$, $\Psi\mapsto\left(\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)\right)$, and the morphism $\Gamma_{1,[n]}\to\mathrm{SL}_2(\mathbb{Z})$ is given by the same formulas and $A_i,B_i,\sigma_i\mapsto 1$.

The elliptic braid group $\overline{\mathrm{B}}_{1,n}$ is the kernel of $\Gamma_{1,[n]}\to\mathrm{SL}_2(\mathbb{Z})$; it has the same presentation as $\Gamma_{1,[n]}$, except for the omission of the generators Θ,Ψ and the relations involving them. The “pure” mapping class group $\Gamma_{1,n}$ is the kernel of $\Gamma_{1,[n]}\to S_n$, $A_i,B_i,C_{jk}\mapsto 1$, $\sigma_i\mapsto\sigma_i$; it has the same presentation as $\Gamma_{1,[n]}$, except for the omission of the σ_i. Finally, recall that $\overline{\mathrm{PB}}_{1,n}$ is the kernel of $\Gamma_{1,[n]}\to\mathrm{SL}_2(\mathbb{Z})\times S_n$.

Remark 16. The extended mapping class group $\tilde{\Gamma}_{1,n}$ of classes of not necessarily orientation-preserving self-homeomorphisms of a surface of type $(1,n)$ fits in a split exact sequence $1\to\Gamma_{1,n}\to\tilde{\Gamma}_{1,n}\to\mathbb{Z}/2\mathbb{Z}\to 1$; it may be viewed as $\left\{\overline{\mathrm{PB}}_{1,n}\rtimes\widetilde{\mathrm{GL}_2(\mathbb{Z})}\right\}/\mathbb{Z}$; it has the same presentation as $\Gamma_{1,n}$ with the additional generator Σ subject to

$$\Sigma^2=1,\quad \Sigma\Theta\Sigma^{-1}=\Theta^{-1},\quad \Sigma\Psi\Sigma^{-1}=\Psi^{-1},$$

$$\Sigma A_i\Sigma^{-1}=A_i^{-1},\quad \Sigma B_i\Sigma^{-1}=A_iB_iA_i^{-1}.$$

4.3 The monodromy morphisms $\gamma_n:\Gamma_{1,[n]}\to\mathbf{G}_n\rtimes S_n$

Let $F(\mathbf{z}|\tau)$ be a solution of the elliptic KZB system defined on D_n.

Recall that $D_n=\{(\mathbf{z},\tau)\in\mathbb{C}^n\times\mathfrak{H}|z_i=a_i+b_i\tau, a_i,b_i\in\mathbb{R}, a_1<a_2<\cdots<a_n<a_1+1, b_1<b_2<\cdots<b_n<b_1+1\}$. The domains $H_n:=\{(\mathbf{z},\tau)\in\mathbb{C}^n\times\mathfrak{H}|z_i=a_i+b_i\tau, a_i,b_i\in\mathbb{R},\ a_1<a_2<\cdots<a_n<a_1+1\}$ and $D_n:=\{(\mathbf{z},\tau)\in\mathbb{C}^n\times\mathfrak{H}|z_i=a_i+b_i\tau, a_i,b_i\in\mathbb{R}, b_1<b_2<\cdots<b_n<b_1+1\}$ are

also simply connected and invariant, and we denote by $F^H(\mathbf{z}|\tau)$ and $F^V(\mathbf{z}|\tau)$ the prolongations of $F(\mathbf{z}|\tau)$ to these domains.

Then $(\mathbf{z},\tau) \mapsto F^H\left(\mathbf{z}+\sum_{j=1}^n \delta_i|\tau\right)$ and $(\mathbf{z},\tau) \mapsto e^{2\pi\mathrm{i}(\bar{x}_i+\cdots+\bar{x}_n)}F^V\Big(\mathbf{z}+\tau\left(\sum_{j=1}^n \delta_i\right)|\tau\Big)$ are solutions of the elliptic KZB system on H_n and D_n respectively. We define $A_i^F, B_i^F \in \mathbf{G}_n$ by

$$F^H\Big(\mathbf{z}+\sum_{j=1}^n \delta_i|\tau\Big) = F^H(\mathbf{z}|\tau)A_i^F,$$

$$e^{2\pi\mathrm{i}(\bar{x}_i+\cdots+\bar{x}_n)}F^V\Big(\mathbf{z}+\tau\Big(\sum_{j=1}^n \delta_i\Big)|\tau\Big) = F^V(\mathbf{z}|\tau)B_i^F.$$

The action of $T^{-1} = \left(\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)\right)$ is $(\mathbf{z},\tau) \mapsto (-\mathbf{z}/\tau, -1/\tau)$; this transformation takes H_n to V_n. Then $(\mathbf{z},\tau) \mapsto c_{T^{-1}}(\mathbf{z}|\tau)^{-1}F^V(-\mathbf{z}/\tau|-1/\tau)$ is a solution of the elliptic KZB system on H_n (recall that $c_{T^{-1}}(\mathbf{z}|\tau)^{-1} = e^{2\pi\mathrm{i}(-\sum_i z_i\bar{x}_i+\tau X)}(-\tau)^d = (-\tau)^d e^{(2\pi\mathrm{i}/\tau)(\sum_i z_i\bar{x}_i+X)}$). We define Θ^F by

$$c_{T^{-1}}(\mathbf{z}|\tau)^{-1}F^V(-\mathbf{z}/\tau|-1/\tau) = F^H(\mathbf{z}|\tau)\Theta^F.$$

The action of $S = \left(\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)\right)$ is $(\mathbf{z},\tau) \mapsto (\mathbf{z},\tau+1)$. This transformation takes H_n to itself. Since $c_S(\mathbf{z}|\tau)=1$, the function $(\mathbf{z},\tau) \mapsto F^H(\mathbf{z},\tau+1)$ is a solution of the elliptic KZB system on H_n. We define Ψ^F by

$$F^H(\mathbf{z}|\tau+1) = F^H(\mathbf{z}|\tau)\Psi^F.$$

Finally, define σ_i^F by

$$\sigma_i F\left(\sigma_i^{-1}\mathbf{z}|\tau\right) = F(\mathbf{z}|\tau)\sigma_i^F,$$

where on the l.h.s. F is extended to the universal cover of $(\mathbb{C}^n \times \mathfrak{H}) - \mathrm{Diag}_n$ (σ_i exchanges z_i and z_{i+1}, z_{i+1} passing to the right of z_i).

Lemma 17. *There is a unique morphism $\Gamma_{1,[n]} \to \mathbf{G}_n \rtimes S_n$, taking X to X^F, where $X = A_i, B_i, \Theta$, or Ψ.*

Proof. This follows from the geometric description of generators of $\Gamma_{1,[n]}$: if $(\mathbf{z}_0,\tau_0) \in D_n$, then A_i is the class of the projection of the path $[0,1] \ni t \mapsto \left(\mathbf{z}_0 + t\sum_{j=i}^n \delta_j, \tau_0\right)$, B_i is the class of the projection of $[0,1] \ni t \mapsto \left(\mathbf{z}_0 + t\tau\sum_{j=i}^n \delta_j, \tau_0\right)$, Θ is the class of the projection of any path connecting $(\mathbf{z}_0,\tau_0)$ to $(-\mathbf{z}_0/\tau_0, -1/\tau_0)$ contained in H_n, and Ψ is the class of the projection of any path connecting $(\mathbf{z}_0,\tau_0)$ to $(\mathbf{z}_0,\tau_0+1)$ contained in H_n. □

We will denote by $\gamma_n : \Gamma_{1,[n]} \to \mathbf{G}_n \rtimes S_n$ the morphism induced by the solution $F^{(n)}(\mathbf{z}|\tau)$.

4.4 Expression of $\gamma_n : \Gamma_{1,[n]} \to \mathbf{G}_n \rtimes S_n$ using γ_1 and γ_2

Lemma 18. *There exists a unique Lie algebra morphism* $\mathfrak{d} \to \bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}$, $x \mapsto [x]$, *such that* $[\delta_{2n}] = \delta_{2n} + \sum_{i<j} (\mathrm{ad}\, \bar{x}_i)^{2n}(\bar{t}_{ij})$, $[X] = X$, $[\Delta_0] = \Delta_0$, $[d] = d$.

It induces a group morphism $\mathbf{G}_1 \to \mathbf{G}_n$, *also denoted by* $g \mapsto [g]$.

Lemma 19. *For each map* $\phi : \{1, \dots, m\} \to \{1, \dots, n\}$, *there exists a* Lie *algebra morphism* $\bar{\mathfrak{t}}_{1,n} \to \bar{\mathfrak{t}}_{1,m}$, $x \mapsto x^\phi$, *defined by* $(\bar{x}_i)^\phi := \sum_{i' \in \phi^{-1}(i)} \bar{x}_{i'}$, $(\bar{y}_i)^\phi := \sum_{i' \in \phi^{-1}(i)} \bar{y}_{i'}$, $(\bar{t}_{ij})^\phi := \sum_{i' \in \phi^{-1}(i), j' \in \phi^{-1}(j)} \bar{t}_{i'j'}$.

It induces a group morphism $\exp(\hat{\bar{\mathfrak{t}}}_{1,n}) \to \exp(\hat{\bar{\mathfrak{t}}}_{1,m})$, *also denoted by* $g \mapsto g^\phi$.

The proofs are immediate. We now recall the definition and properties of the KZ associator [Dri91].

If $\mathbf{k}$ is a field with $\mathrm{char}(\mathbf{k}) = 0$, we let $\mathfrak{t}_n^{\mathbf{k}}$ be the $\mathbf{k}$-Lie algebra generated by t_{ij}, where $i \neq j \in \{1, \dots, n\}$, with relations

$$t_{ji} = t_{ij}, \quad [t_{ij} + t_{ik}, t_{jk}] = 0, \quad [t_{ij}, t_{kl}] = 0$$

for i, j, k, l distinct (in this section, we set $\mathfrak{t}_n := \mathfrak{t}_n^{\mathbb{C}}$). For each partially defined map $\{1, \dots, m\} \supset D_\phi \xrightarrow{\phi} \{1, \dots, n\}$, we have a Lie algebra morphism $\mathfrak{t}_n \to \mathfrak{t}_m$, $x \mapsto x^\phi$, defined by[6] $(t_{ij})^\phi := \sum_{i' \in \phi^{-1}(i), j' \in \phi^{-1}(j)} t_{i'j'}$. We also have morphisms $\mathfrak{t}_n \to \mathfrak{t}_{1,n}$, $t_{ij} \mapsto \bar{t}_{ij}$, compatible with the maps $x \mapsto x^\phi$ on both sides.

The KZ associator $\Phi = \Phi(t_{12}, t_{23}) \in \exp(\hat{\mathfrak{t}}_3)$ is defined by $G_0(z) = G_1(z)\Phi$, where $G_i :]0, 1[\to \exp(\hat{\mathfrak{t}}_3)$ are the solutions of $G'(z)G(z)^{-1} = t_{12}/z + t_{23}/(z-1)$ with $G_0(z) \sim z^{t_{12}}$ as $z \to 0$ and $G_1(z) \sim (1-z)^{t_{23}}$ as $z \to 1$. The KZ associator satisfies the duality, hexagon, and pentagon equations (37), (38) below (where $\lambda = 2\pi \mathrm{i}$).

Lemma 20. $\gamma_2(A_2)$ *and* $\gamma_2(B_2)$ *belong to* $\exp(\hat{\bar{\mathfrak{t}}}_{1,2}) \subset \mathbf{G}_2$.

Proof. If $F(\mathbf{z}|\tau) : H_2 \to \mathbf{G}_2$ is a solution of the KZB equation for $n = 2$, then $A_2^F = F^H(\mathbf{z} + \delta_2|\tau) F^H(\mathbf{z}|\tau)^{-1}$ is expressed as the iterated integral, from $\mathbf{z}_0 \in D_n$ to $\mathbf{z}_0 + \delta_2$, of $\bar{K}_2(\mathbf{z}|\tau) \in \hat{\bar{\mathfrak{t}}}_{1,2}$; hence $A_2^F \in \exp(\hat{\bar{\mathfrak{t}}}_{1,2})$. Since $\gamma_2(A_2)$ is a conjugate of A_2^F, it belongs to $\exp(\hat{\bar{\mathfrak{t}}}_{1,2})$, since $\exp(\hat{\bar{\mathfrak{t}}}_{1,2}) \subset \mathbf{G}_2 \rtimes S_2$ is normal. One proves similarly that $\gamma_2(B_2) \in \exp(\hat{\bar{\mathfrak{t}}}_{1,2})$. □

Set

$$\Phi_i := \Phi^{1,\dots,i-1,i,i+1,\dots,n} \cdots \Phi^{1,\dots,n-2,n-1,n} \in \exp(\hat{\mathfrak{t}}_n).$$

We denote by $x \mapsto \{x\}$ the morphism $\exp(\hat{\mathfrak{t}}_n) \to \exp(\hat{\bar{\mathfrak{t}}}_{1,n})$ induced by $t_{ij} \mapsto \bar{t}_{ij}$.

[6] We will also use the notation $x^{I_1,\dots,I_n}$ for x^ϕ, where $I_i = \phi^{-1}(i)$.

Proposition 21. *If $n \geq 2$, then*

$$\gamma_n(\Theta) = [\gamma_1(\Theta)]e^{\mathrm{i}\frac{\pi}{2}\sum_{i<j}\bar{t}_{ij}}, \quad \gamma_n(\Psi) = [\gamma_1(\Psi)]e^{\mathrm{i}\frac{\pi}{6}\sum_{i<j}\bar{t}_{ij}},$$

and if $n \geq 3$, then

$$\begin{aligned}
\gamma_n(A_i) &= \{\Phi_i\}^{-1}\gamma_2(A_2)^{1,\dots,i-1,i,\dots,n}\{\Phi_i\}, \ (i = 1,\dots,n),\\
\gamma_n(B_i) &= \{\Phi_i\}^{-1}\gamma_2(B_2)^{1,\dots,i-1,i,\dots,n}\{\Phi_i\}, \ (i = 1,\dots,n),\\
\gamma_n(\sigma_i) &= \{\Phi^{1,\dots,i-1,i,i+1}\}^{-1}e^{\mathrm{i}\pi\bar{t}_{i,i+1}}\{\Phi^{1,\dots,i-1,i,i+1}\}, \ (i = 1,\dots,n-1).
\end{aligned}$$

Proof. In the region $z_{21} \ll z_{31} \ll \cdots \ll z_{n1} \ll 1$, $(\mathbf{z},\tau) \in D_n$, we have

$$F^{(n)}(\mathbf{z}|\tau) \simeq z_{21}^{\bar{t}_{12}} \cdots z_{n1}^{\bar{t}_{1n}+\cdots+\bar{t}_{n-1,n}} \exp\left(-\frac{a_0}{2\pi\mathrm{i}}\left(\int_{\mathrm{i}}^{\tau} E_2 + C\right)\left(\sum_{i<j}\bar{t}_{ij}\right)\right)[F(\tau)],$$

where $F(\tau) = F^{(1)}(z|\tau)$ for any z. Here C is the constant such that $\int_{\mathrm{i}}^{\tau} E_2 + C = \tau + o(1)$ as $\tau \to \mathrm{i}\infty$.

We have $F(\tau+1) = F(\tau)\gamma_1(\Psi)$, $F(-1/\tau) = F(\tau)\gamma_1(\Theta)$. Since $\sum_{i<j}\bar{t}_{ij}$ commutes with the image of $x \mapsto [x]$, we get

$$F^{(n)}(\mathbf{z}|\tau+1) = F^{(n)}(\mathbf{z}|\tau)\exp\left(-\frac{a_0}{2\pi\mathrm{i}}\left(\sum_{i<j}\bar{t}_{ij}\right)\right)[\gamma_1(\Psi)],$$

so

$$\gamma_n(\Psi) = \exp\left(\mathrm{i}\frac{\pi}{6}\sum_{i<j}\bar{t}_{ij}\right)[\gamma_1(\Psi)].$$

In the same region,

$$\begin{aligned}
&c_{T^{-1}}(\mathbf{z}|\tau)^{-1}F^{(n)V}\left(-\frac{\mathbf{z}}{\tau}| - \frac{1}{\tau}\right)\\
&\simeq (-\tau)^d e^{\frac{2\pi\mathrm{i}}{\tau}(\sum_i z_i\bar{x}_i + X)}(-z_{21}/\tau)^{\bar{t}_{12}} \cdots (-z_{n1}/\tau)^{\bar{t}_{1n}+\cdots+\bar{t}_{n-1,n}}\\
&\exp\left(-\frac{a_0}{2\pi\mathrm{i}}\left(\int_{\mathrm{i}}^{-1/\tau} E_2 + C\right)\left(\sum_{i<j}\bar{t}_{ij}\right)\right)[F(-1/\tau)].
\end{aligned}$$

Now $E_2(-1/\tau) = \tau^2 E_2(\tau) + (6\mathrm{i}/\pi)\tau$, so

$$\int_{\mathrm{i}}^{-1/\tau} E_2 - \int_{\mathrm{i}}^{\tau} E_2 = (6\mathrm{i}/\pi)[\log(-1/\tau) - \log \mathrm{i}]$$

(where $\log(re^{i\theta}) = \log r + \mathrm{i}\theta$ for $\theta \in]-\pi,\pi[$).

It follows that

$$
\begin{aligned}
& c_{T^{-1}}(\mathbf{z}|\tau)^{-1} F^{(n)V}\left(-\frac{\mathbf{z}}{\tau}| -\frac{1}{\tau}\right) \simeq e^{2\pi\mathrm{i}(\sum_i z_i\bar{x}_i)} z_{21}^{\bar{t}_{12}} \cdots z_{n1}^{\bar{t}_{1n}+\cdots+\bar{t}_{n-1,n}} \\
& \quad \exp\left(-\frac{a_0}{2\pi\mathrm{i}}\left(\int_{\mathrm{i}}^{\tau} E_2 + C\right)\left(\sum_{i<j}\bar{t}_{ij}\right)\right) \\
& \quad \exp\left(-\frac{a_0}{2\pi\mathrm{i}}\frac{-6\mathrm{i}}{\pi}(\log \mathrm{i})\left(\sum_{i<j}\bar{t}_{ij}\right)\right)[(-\tau)^d e^{(2\pi\mathrm{i}/\tau)X} F(-1/\tau)] \\
& \simeq z_{21}^{\bar{t}_{12}} \cdots z_{n1}^{\bar{t}_{1n}+\cdots+\bar{t}_{n-1,n}} \exp\left(-\frac{a_0}{2\pi\mathrm{i}}\left(\int_{\mathrm{i}}^{\tau} E_2 + C\right)\left(\sum_{i<j}\bar{t}_{ij}\right)\right) \\
& \quad [F(\tau)\gamma_1(\Theta)]\exp\left(\frac{\mathrm{i}\pi}{2}\sum_{i<j}\bar{t}_{ij}\right) \\
& \simeq F^{(n)H}(\mathbf{z}|\tau)[\gamma_1(\Theta)]\exp\left(\frac{\mathrm{i}\pi}{2}\sum_{i<j}\bar{t}_{ij}\right)
\end{aligned}
$$

(the second $\simeq$ follows from $\sum_i z_i\bar{x}_i = \sum_{i>1} z_{i1}\bar{x}_i$ and $z_{i1} \to 0$), so

$$\gamma_n(\Theta) = [\gamma_1(\Theta)]\exp\left(\mathrm{i}\frac{\pi}{2}\sum_{i<j}\bar{t}_{ij}\right).$$

Let $G_i(\mathbf{z}|\tau)$ be the solution of the elliptic KZB system such that

$$
\begin{aligned}
G_i(\mathbf{z}|\tau) \simeq\ & z_{21}^{\bar{t}_{12}} \cdots z_{i-1,1}^{\bar{t}_{12}+\cdots+\bar{t}_{1,i-1}} z_{n,i}^{\bar{t}_{i,n}+\cdots+\bar{t}_{n-1,n}} \cdots z_{n,n-1}^{\bar{t}_{n-1,n}} \\
& \times \exp\Big(-\frac{\tau}{2\pi\mathrm{i}}\Big(\Delta_0 + \sum_{n\geq 0} a_{2n}\big(\delta_{2n} + \sum_{i<j}(\mathrm{ad}\,\bar{x}_i)^{2n}(\bar{t}_{ij})\big)\Big)\Big)
\end{aligned}
$$

when $z_{21} \ll \cdots \ll z_{i-1,1} \ll 1$, $z_{n,n-1} \ll \cdots \ll z_{n,i} \ll 1$, $\tau \to \mathrm{i}\infty$, and $(\mathbf{z},\tau) \in D_n$. Then $G_i\left(\mathbf{z} + \sum_{j=i}^n \delta_i|\tau\right) = G_i(\mathbf{z}|\tau)\gamma_2(A_2)^{1,\ldots,i-1,i,\ldots,n}$, because in the domain considered, $\bar{K}_i(\mathbf{z}|\tau)$ is close to $\bar{K}_2(z_1,z_n|\tau)^{1,\ldots,i-1,i,\ldots,n}$ (where $\bar{K}_2(\cdots)$ corresponds to the 2-point system); on the other hand, $F(\mathbf{z}|\tau) = G_i(\mathbf{z}|\tau)\{\Phi_i\}$, which implies the formula for $\gamma_n(A_i)$. The formula for $\gamma_n(B_i)$ is proved in the same way. The behavior of $F^{(n)}(\mathbf{z}|\tau)$ for $z_{21} \ll \cdots \ll z_{n1} \ll 1$ is similar to that of a solution of the KZ equations, which implies the formula for $\gamma_n(\sigma_i)$. □

Remark 22. One checks that the composition $\mathrm{SL}_2(\mathbb{Z}) \simeq \Gamma_{1,1} \to \mathbf{G}_1 \to \mathrm{SL}_2(\mathbb{C})$ is a conjugation of the canonical inclusion. It follows that the composition $\widetilde{\mathrm{SL}_2(\mathbb{Z})} \subset \Gamma_{1,n} \to \mathbf{G}_1 \to \mathrm{SL}_2(\mathbb{C})$ is a conjugation of the canonical projection for any $n \geq 1$.

Let us set $\tilde{A} := \gamma_2(A_2)$, $\tilde{B} := \gamma_2(B_2)$. The image of $A_2A_3^{-1} = \sigma_1^{-1}A_2^{-1}\sigma_1^{-1}$ by γ_3 yields

$$\tilde{A}^{12,3} = e^{\mathrm{i}\pi\bar{t}_{12}}\{\Phi\}^{3,1,2}\tilde{A}^{2,13}\{\Phi\}^{2,1,3}e^{\mathrm{i}\pi\bar{t}_{12}} \cdot \{\Phi\}^{3,2,1}\tilde{A}^{1,23}\{\Phi\}^{1,2,3}, \tag{22}$$

and the image of $B_2B_3^{-1} = \sigma_1 B_2^{-1}\sigma_1$ yields

$$\tilde{B}^{12,3} = e^{-\mathrm{i}\pi\bar{t}_{12}}\{\Phi\}^{3,1,2}\tilde{B}^{2,13}\{\Phi\}^{2,1,3}e^{-\mathrm{i}\pi\bar{t}_{12}} \cdot \{\Phi\}^{3,2,1}\tilde{B}^{1,23}\{\Phi\}^{1,2,3}. \tag{23}$$

Since $(\gamma_3(A_2), \gamma_3(A_3)) = (\gamma_3(B_2), \gamma_3(B_3)) = 1$, we get

$$(\{\Phi\}^{3,2,1}\tilde{A}^{1,23}\{\Phi\}, \tilde{A}^{12,3}) = (\{\Phi\}^{3,2,1}\tilde{B}^{1,23}\{\Phi\}, \tilde{B}^{12,3}) = 1 \tag{24}$$

(this equation can also be directly derived from (22) and (23) by noting that the l.h.s. is invariant under $x \mapsto x^{2,1,3}$ and commutes with $e^{\pm\mathrm{i}\pi\bar{t}_{12}}$). We have for $n = 2$, $C_{12} = (B_2, A_2)$, so $(\tilde{A}, \tilde{B}) = \gamma_2(C_{12})^{-1}$. Also $\gamma_1(\Theta)^4 = 1$, so $\gamma_2(C_{12}) = \gamma_2(\Theta)^4 = (e^{\mathrm{i}\pi\bar{t}_{12}/2}[\gamma_1(\Theta)])^4 = e^{2\pi\mathrm{i}\bar{t}_{12}}[\gamma_1(\Theta)^4] = e^{2\pi\mathrm{i}\bar{t}_{12}}$, so

$$(\tilde{A}, \tilde{B}) = e^{-2\pi\mathrm{i}\bar{t}_{12}}. \tag{25}$$

For $n = 3$, we have $\gamma_3(\Theta)^4 = e^{2\pi\mathrm{i}(\bar{t}_{12}+\bar{t}_{13}+\bar{t}_{23})} = \gamma_3(C_{12}C_{23})$; since $\gamma_3(C_{12}) = (\gamma_3(B_2), \gamma_3(A_2)) = \{\Phi\}^{-1}(\tilde{B}, \tilde{A})^{1,23}\{\Phi\} = \{\Phi\}^{-1}e^{2\pi\mathrm{i}(\bar{t}_{12}+\bar{t}_{13})}\{\Phi\}$, we get $\gamma_3(C_{23}) = \{\Phi\}^{-1}e^{2\pi\mathrm{i}\bar{t}_{23}}\{\Phi\}$. The image by γ_3 of $\big(B_3, A_3A_2^{-1}\big) = \big(B_3B_2^{-1}, A_3\big) = C_{23}$ then gives

$$\begin{aligned}(\tilde{B}^{12,3}, \tilde{A}^{12,3}\{\Phi\}^{-1}(\tilde{A}^{1,23})^{-1}\{\Phi\}) &= (\tilde{B}^{12,3}\{\Phi\}^{-1}(\tilde{B}^{1,23})^{-1}\{\Phi\}, \tilde{A}^{12,3}) \\ &= \{\Phi\}^{-1}e^{2\pi\mathrm{i}\bar{t}_{23}}\{\Phi\}\end{aligned} \tag{26}$$

(applying $x \mapsto x^{\emptyset,1,2}$, this identity implies (25)).

Let us set $\tilde{\Theta} := \gamma_1(\Theta)$, $\tilde{\Psi} := \gamma_1(\Theta)$. Since γ_1, γ_2 are group morphisms, we have

$$\tilde{\Theta}^4 = (\tilde{\Theta}\tilde{\Psi})^3 = (\tilde{\Theta}^2, \tilde{\Psi}) = 1, \tag{27}$$

$$[\tilde{\Theta}]e^{\mathrm{i}\frac{\pi}{2}\bar{t}_{12}}\tilde{A}\left([\tilde{\Theta}]e^{\mathrm{i}\frac{\pi}{2}\bar{t}_{12}}\right)^{-1} = \tilde{B}^{-1}, \quad [\tilde{\Theta}]e^{\mathrm{i}\frac{\pi}{2}\bar{t}_{12}}\tilde{B}\left([\tilde{\Theta}]e^{\mathrm{i}\frac{\pi}{2}\bar{t}_{12}}\right)^{-1} = \tilde{B}\tilde{A}\tilde{B}^{-1}, \tag{28}$$

$$[\tilde{\Psi}]e^{\mathrm{i}\frac{\pi}{6}\bar{t}_{12}}\tilde{A}\left([\tilde{\Psi}]e^{\mathrm{i}\frac{\pi}{6}\bar{t}_{12}}\right)^{-1} = \tilde{A}, \quad [\tilde{\Psi}]e^{\mathrm{i}\frac{\pi}{6}\bar{t}_{12}}\tilde{B}\left([\tilde{\Psi}]e^{\mathrm{i}\frac{\pi}{6}\bar{t}_{12}}\right)^{-1} = \tilde{B}\tilde{A}. \tag{29}$$

We note that, (27) (respectively, (28), (29)) are identities in $\mathbf{G}_1$ (respectively, $\mathbf{G}_2$); in (28), (29), $x \mapsto [x]$ is induced by the map $\mathfrak{d} \to \mathfrak{d} \rtimes \bar{\mathfrak{t}}_{1,2}$ defined above.

4.5 Expression of $\tilde{\Psi}$ and of $\tilde{A}$ and $\tilde{B}$ in terms of Φ

In this section, we compute $\tilde{A}$ and $\tilde{B}$ in terms of the KZ associator Φ. We also compute $\tilde{\Psi}$.

Recall the definition of $\tilde{\Psi}$. The elliptic KZB system for $n = 1$ is

$$2\pi\mathrm{i}\partial_\tau F(\tau) + \left(\Delta_0 + \sum_{k\geq 1} a_{2k}E_{2k+2}(\tau)\delta_{2k}\right)F(\tau) = 0.$$

The solution $F(\tau) := F^{(1)}(z|\tau)$ (for any z) is determined by $F(\tau) \simeq \exp\left(-\frac{\tau}{2\pi\mathrm{i}}(\Delta_0 + \sum_{k\geq 1} a_{2k}\delta_{2k})\right)$. Then $\tilde{\Psi}$ is determined by $F(\tau+1) = F(\tau)\tilde{\Psi}$. We have therefore the following:

Lemma 23. $\tilde{\Psi} = \exp\left(-\frac{1}{2\pi\mathrm{i}}(\Delta_0 + \sum_{k\geq 1} a_{2k}\delta_{2k})\right)$.

Recall the definition of $\tilde{A}$ and $\tilde{B}$. The elliptic KZB system for $n = 2$ is

$$\partial_z F(z|\tau) = -\left(\frac{\theta(z + \mathrm{ad}\,x|\tau)\mathrm{ad}\,x}{\theta(z|\tau)\theta(\mathrm{ad}\,x|\tau)}\right)(y) \cdot F(z|\tau), \tag{30}$$

$$2\pi\mathrm{i}\partial_\tau F(z|\tau) + \left(\Delta_0 + \sum_{k\geq 1} a_{2k}E_{2k+2}(\tau)\delta_{2k} - g(z, \mathrm{ad}\,x|\tau)(t)\right) F(z|\tau) = 0, \tag{31}$$

where $z = z_{21}$, $x = \bar{x}_2 = -\bar{x}_1$, $y = \bar{y}_2 = -\bar{y}_1$, $t = \bar{t}_{12} = -[x, y]$.

The solution $F(z|\tau) := F^{(2)}(z_1, z_2|\tau)$ is determined by its behavior $F(z|\tau) \simeq z^t \exp\left(-\frac{\tau}{2\pi\mathrm{i}}(\Delta_0 + \sum_{k\geq 0} a_{2k}(\delta_{2k} + (\mathrm{ad}\,x)^{2k})(t))\right)$ as $z \to 0^+$, $\tau \to \mathrm{i}\infty$. We then have $F^H(z+1|\tau) = F^H(z|\tau)\tilde{A}$, $e^{2\pi\mathrm{i}x}F^V(z+\tau|\tau) = F^V(z|\tau)\tilde{B}$.

Proposition 24. *We have*[7]

$$\begin{aligned}\tilde{A} &= (2\pi/\mathrm{i})^t \Phi(\tilde{y}, t) e^{2\pi\mathrm{i}\tilde{y}} \Phi(\tilde{y}, t)^{-1} (\mathrm{i}/2\pi)^t \\ &= (2\pi)^t \mathrm{i}^{-3t} \Phi(-\tilde{y} - t, t) e^{2\pi\mathrm{i}(\tilde{y}+t)} \Phi(-\tilde{y} - t, t)^{-1} (2\pi\mathrm{i})^{-t},\end{aligned}$$

where $\tilde{y} = -\frac{\mathrm{ad}\,x}{e^{2\pi\mathrm{i}\,\mathrm{ad}\,x} - 1}(y)$.

Proof. $\tilde{A} = F^H(z|\tau)^{-1}F^H(z+1|\tau)$, which we will compute in the limit $\tau \to \mathrm{i}\infty$. For this, we will compute $F(z|\tau)$ in the limit $\tau \to \mathrm{i}\infty$. In this limit, $\theta(z|\tau) = (1/\pi)\sin(\pi z)[1 + O(e^{2\pi\mathrm{i}\tau})]$, so the system becomes

$$\partial_z F(z|\tau) = \left(\pi\cot\mathrm{g}(\pi z)t - \pi\cot\mathrm{g}(\pi\,\mathrm{ad}\,x)\mathrm{ad}\,x(y) + O(e^{2\pi\mathrm{i}\tau})\right) F(z|\tau), \tag{32}$$

$$\begin{aligned}2\pi\mathrm{i}\partial_\tau F(z|\tau) + \Big(\Delta_0 + \textstyle\sum_{k\geq 1} a_{2k}\delta_{2k} + &\left(\frac{\pi^2}{\sin^2(\pi\,\mathrm{ad}\,x)} - \frac{1}{(\mathrm{ad}\,x)^2}\right) \\ &\times (t) + O(e^{2\pi\mathrm{i}\tau})\Big) F(z|\tau) = 0,\end{aligned}$$

where the last equation is

$$\begin{aligned}&2\pi\mathrm{i}\partial_\tau F(z|\tau) \\ &\quad + \left(\Delta_0 + a_0 t + \sum_{k\geq 1} a_{2k}\left(\delta_{2k} + (\mathrm{ad}\,x)^{2k}(t)\right) + O(e^{2\pi\mathrm{i}\tau})\right) F(z|\tau) = 0.\end{aligned}$$

[7] By convention, if $z \in \mathbb{C} \setminus \mathbb{R}_-$ and $x \in \mathfrak{n}$, where $\mathfrak{n}$ is a pronilpotent Lie algebra, then z^x is $\exp(x \log z) \in \exp(\mathfrak{n})$, where $\log z$ is chosen with imaginary part in $]-\pi, \pi[$.

We set

$$\Delta := \Delta_0 + \sum_{k\geq 1} a_{2k}\delta_{2k}, \quad \text{so} \quad \Delta_0 + a_0 t + \sum_{k\geq 1} a_{2k}(\delta_{2k} + (\mathrm{ad}\, x)^{2k}(t)) = [\Delta] + a_0 t.$$

The compatibility of this system implies that $[\Delta] + a_0 t$ commutes with t and $(\pi \mathrm{ad}\, x) \cot \mathrm{g}(\pi \mathrm{ad}\, x)(y) = \mathrm{i}\pi(-t - 2\tilde{y})$, hence with t and $\tilde{y}$; actually t commutes with each $[\delta_{2k}] = \delta_{2k} + (\mathrm{ad}\, x)^{2k}(t)$.

Equation (30) can be written $\partial_z F(z|\tau) = (t/z + O(1))F(z|\tau)$. We then let $F_0(z|\tau)$ be the solution of (30) in $V := \{(z,\tau)|\tau \in \mathfrak{H}, z = a + b\tau, a \in]0,1[, b \in \mathbb{R}\}$ such that $F_0(z|\tau) \simeq z^t$ when $z \to 0^+$, for any τ. This means that the left (equivalently, right) ratio of these quantities has the form $1 + \sum_{k>0}(\text{degree } k)O(z(\log z)^{f(k)})$ where $f(k) \geq 0$.

We now relate $F(z|\tau)$ and $F_0(z|\tau)$. Let $F(\tau) = F^{(1)}(z|\tau)$ for any z be the solution of the KZB system for $n = 1$, such that $F(\tau) \simeq \exp\left(-\frac{\tau}{2\pi \mathrm{i}}\Delta\right)$ as $\tau \to \mathrm{i}\infty$ (meaning that the left, or equivalently right, ratio of these quantities has the form $1 + \sum_{k>0}(\text{degree } k)O(\tau^{f(k)} e^{2\pi \mathrm{i}\tau})$, where $f(k) \geq 0$).

Lemma 25. *We have* $F(z|\tau) = F_0(z|\tau)\exp\left(-\frac{a_0}{2\pi \mathrm{i}}\left(\int_{\mathrm{i}}^{\tau} E_2 + C\right) t\right) [F(\tau)]$, *where C is such that* $\int_{\mathrm{i}}^{\tau} E_2 + C = \tau + O(e^{2\pi \mathrm{i}\tau})$.

Proof of Lemma. $F(z|\tau) = F_0(z|\tau)X(\tau)$, where $X : \mathfrak{H} \to \mathbf{G}_2$ is a map. We have $g(z, \mathrm{ad}\, x|\tau)(t) = a_0 E_2(\tau)t + \sum_{k>0} a_{2k}E_{2k+2}(\tau)(\mathrm{ad}\, x)^{2k}(t) + O(z)$ when $z \to 0^+$ and for any τ, so (31) is written as

$$2\pi \mathrm{i}\partial_\tau F(z|\tau) + \left(\Delta_0 + a_0 E_2(\tau)t + \sum_{k>0} a_{2k}E_{2k+2}(\tau)[\delta_{2k}] + O(z)\right) F(z|\tau) = 0,$$

where $O(z)$ has degree > 0. Since Δ_0, t and the $[\delta_{2k}]$ all commute with t, the ratio $F_0(z|\tau)^{-1}F(z|\tau)$ satisfies

$$2\pi \mathrm{i}\partial_\tau \left(F_0^{-1}F(z|\tau)\right) + \Big(\Delta_0 + a_0 E_2(\tau)t + \sum_{k>0} a_{2k}E_{2k+2}(\tau)[\delta_{2k}] + \sum_{k>0}(\text{degree } k)O(z(\log z)^{h(k)})\Big) \left(F_0^{-1}F(z|\tau)\right) = 0$$

where $h(k) \geq 0$. Since $F_0(z|\tau)^{-1}F(z|\tau) = X(\tau)$ is in fact independent of z, we have

$$2\pi \mathrm{i}\partial_\tau (X(\tau)) + \left(\Delta_0 + a_0 E_2(\tau)t + \sum_{k>0} a_{2k}E_{2k+2}(\tau)[\delta_{2k}]\right) (X(\tau)) = 0,$$

which implies that $X(\tau) = \exp\left(-\frac{a_0}{2\pi \mathrm{i}}\left(\int_{\mathrm{i}}^{\tau} E_2 + C\right) t\right) [F(\tau)]X_0$, where X_0 is a suitable element in $\mathbf{G}_2$. The asymptotic behavior of $F(z|\tau)$ as $\tau \to \mathrm{i}\infty$ and $z \to 0^+$ then implies $X_0 = 1$. □

End of proof of Proposition. We then have $F(z|\tau) = F_0(z|\tau)X(\tau)$, where $X(\tau) \simeq \exp\big(-\frac{\tau}{2\pi \mathrm{i}}([\Delta] + a_0 t)\big)$ as $\tau \to \mathrm{i}\infty$, where this means that the left ratio (equivalently, the right ratio) of these quantities has the form $1 + \sum_{k>0}(\text{degree } k)O(\tau^{x(k)}e^{2\pi \mathrm{i}\tau})$, where $x(k) \geq 0$.

If we set $u := e^{2\pi \mathrm{i} z}$, then (30) is rewritten as

$$\partial_u \bar{F}(u|\tau) = (\tilde{y}/u + t/(u-1) + O(e^{2\pi \mathrm{i}\tau}))\bar{F}(u|\tau), \tag{33}$$

where $\bar{F}(u|\tau) = F(z|\tau)$.

Let $D' := \{u||u| \leq 1\} - [0,1]$ be the complement of the unit interval in the unit disc. Then we have a bijection $\big\{(z,\tau)|\tau \in \mathrm{i}\mathbb{R}_+^\times, z = a + \tau b, a \in [0,1], b \geq 0\big\} \to D' \times \mathrm{i}\mathbb{R}_+^\times$, given by $(z,\tau) \mapsto (u,\tau) := (e^{2\pi \mathrm{i} z}, \tau)$.

Let $\bar{F}_a, \bar{F}_f$ be the solutions of (33) in $D' \times \mathrm{i}\mathbb{R}_+$ such that $\bar{F}_a(u|\tau) \simeq ((u-1)/(2\pi \mathrm{i}))^t$ when $u = 1 + \mathrm{i}0^+$ and for any τ, and $\bar{F}_f(u|\tau) \simeq e^{\mathrm{i}\pi t}((1-u)/(2\pi \mathrm{i}))^t$ when $u = 1 - \mathrm{i}0^+$, for any τ.

Then one checks that $F_0(z|\tau) = \bar{F}_a(e^{2\pi \mathrm{i} z}|\tau)$, $F_0(z-1|\tau) = \bar{F}_f(e^{2\pi \mathrm{i} z}|\tau)$ when $(z,\tau) \in \big\{(z,\tau)|\tau \in \mathrm{i}\mathbb{R}_+^\times, z = a + \tau b | a \in [0,1], b \geq 0\big\}$.

We then define $\bar{F}_b, \ldots, \bar{F}_e$ as the solutions of (33) in $D' \times \mathrm{i}\mathbb{R}_+^\times$, such that $\bar{F}_b(u|\tau) \simeq (1-u)^t$ as $u = 1 - 0^+$, $\Im(u) > 0$ for any τ, $\bar{F}_c(u|\tau) \simeq u^{\tilde{y}}$ as $u \to 0^+$, $\Im(u) > 0$ for any τ, $\bar{F}_d(u|\tau) \simeq u^{\tilde{y}}$ as $u \to 0^+$, $\Im(u) < 0$ for any τ, $\bar{F}_e(u|\tau) \simeq (1-u)^t$ as $u = 1 - 0^+$, $\Im(u) < 0$ for any τ.

Then $\bar{F}_b = \bar{F}_a(-2\pi \mathrm{i})^t$, $\bar{F}_c(-|\tau) = \bar{F}_b(-|\tau)[\Phi(\tilde{y},t) + O(e^{2\pi \mathrm{i}\tau})]$, $\bar{F}_d(-|\tau) = \bar{F}_c(-|\tau)e^{-2\pi \mathrm{i}\tilde{y}}$, $\bar{F}_e(-|\tau) = \bar{F}_d(-|\tau)[\Phi(\tilde{y},t)^{-1} + O(e^{2\pi \mathrm{i}\tau})]$, $\bar{F}_f = \bar{F}_e(\mathrm{i}/2\pi)^t$.

So $\bar{F}_f(-|\tau) = \bar{F}_a(-|\tau)\big((-2\pi \mathrm{i})^t\Phi(\tilde{y},t)e^{-2\pi \mathrm{i}\tilde{y}}\Phi(\tilde{y},t)^{-1}(\mathrm{i}/2\pi)^t + O(e^{2\pi \mathrm{i}\tau})\big)$. It follows that $F_0(z+1|\tau) = F_0(z|\tau)A(\tau)$, where

$$A(\tau) = (-2\pi \mathrm{i})^t\Phi(\tilde{y},t)e^{2\pi \mathrm{i}\tilde{y}}\Phi(\tilde{y},t)^{-1}(\mathrm{i}/2\pi)^t + O(e^{2\pi \mathrm{i}\tau}).$$

Now

$$\begin{aligned}
\tilde{A} &= F(z|\tau)^{-1}F(z+1|\tau) = X(\tau)^{-1}A(\tau)X(\tau) \\
&= \left(1 + \sum_{k>0}(\text{degree } k)O(\tau^{x(k)}e^{2\pi \mathrm{i}\tau})\right)^{-1} \exp\left(\frac{\tau}{2\pi \mathrm{i}}([\Delta] + a_0 t)\right) \\
&\quad \big((-2\pi \mathrm{i})^t\Phi(\tilde{y},t)e^{2\pi \mathrm{i}\tilde{y}}\Phi(\tilde{y},t)^{-1}(\mathrm{i}/2\pi)^t + O(e^{2\pi \mathrm{i}\tau})\big)\exp\left(-\frac{\tau}{2\pi \mathrm{i}}([\Delta] + a_0 t)\right) \\
&\quad \left(1 + \sum_{k>0}(\text{degree } k)O(\tau^{x(k)}e^{2\pi \mathrm{i}\tau})\right).
\end{aligned}$$

As we have seen, $[\Delta] + a_0 t$ commutes with $\tilde{y}$ and t; on the other hand,

$$\begin{aligned}
&\exp\left(\frac{\tau}{2\pi \mathrm{i}}([\Delta] + a_0 t)\right) O(e^{2\pi \mathrm{i}\tau})\exp\left(-\frac{\tau}{2\pi \mathrm{i}}([\Delta] + a_0 t)\right) \\
&\quad = \exp\left(\tau \mathrm{ad}\left(\frac{[\Delta] + a_0 t}{2\pi \mathrm{i}}\right)\right)(O(e^{2\pi \mathrm{i}\tau})) = \sum_{k\geq 0}(\text{degree } k)O(\tau^{n_1(k)}e^{2\pi \mathrm{i}\tau}),
\end{aligned}$$

where $n_1(k) \geq 0$, since $[\Delta] + a_0 t$ is a sum of terms of positive degree and of Δ_0, which is locally ad-nilpotent.

Then

$$\tilde{A} = \left(1 + \sum_{k>0} (\text{degree } k) O(\tau^{x(k)} e^{2\pi \mathrm{i} \tau})\right)^{-1}$$
$$\left((-2\pi\mathrm{i})^t \Phi(\tilde{y}, t) e^{2\pi\mathrm{i}\tilde{y}} \Phi(\tilde{y}, t)^{-1} (\mathrm{i}/2\pi)^t + \sum_{k\geq 0} (\text{degree } k) O(\tau^{n_1(k)} e^{2\pi\mathrm{i}\tau})\right)$$
$$\left(1 + \sum_{k>0} (\text{degree } k) O(\tau^{x(k)} e^{2\pi\mathrm{i}\tau})\right).$$

It follows that

$$\tilde{A} = (-2\pi\mathrm{i})^t \Phi(\tilde{y}, t) e^{2\pi\mathrm{i}\tilde{y}} \Phi(\tilde{y}, t)^{-1} (\mathrm{i}/2\pi)^t + \sum_{k\geq 0} (\text{degree } k) O(\tau^{n_2(k)} e^{2\pi\mathrm{i}\tau}),$$

where $n_2(k) \geq 0$, which implies the first formula for $\tilde{A}$. The second formula either follows from the first one by using the hexagon identity, or can be obtained by repeating the above argument using a path $1 \to +\infty \to 1$, winding around 1 and ∞. $\square$

Theorem 26.

$$\tilde{B} = (2\pi\mathrm{i})^t \Phi(-\tilde{y} - t, t) e^{2\pi\mathrm{i}x} \Phi(\tilde{y}, t)^{-1} (2\pi/\mathrm{i})^{-t}.$$

Proof. We first define $F_0(z|\tau)$ as the solution in $V := \{a + b\tau | a \in]0, 1[, b \in \mathbb{R}\}$ of (30) such that $F_0(z|\tau) \sim z^t$ as $z \to 0^+$. Then there exists $B(\tau)$ such that $e^{2\pi\mathrm{i}x} F_0(z + \tau|\tau) = F_0(z|\tau) B(\tau)$. We compute the asymptotics of $B(\tau)$ as $\tau \to \mathrm{i}\infty$.

We define four asymptotic zones (z is assumed to remain on the segment $[0, \tau]$, and τ on the line $\mathrm{i}\mathbb{R}_+$): (1) $z \ll 1 \ll \tau$, (2) $1 \ll z \ll \tau$, (3) $1 \ll \tau - z \ll \tau$, (4) $\tau - z \ll 1 \ll \tau$.

In the transition (1)–(2), the system takes the form (32), or if we set $u := e^{2\pi\mathrm{i}z}$, (33).

In the transition (3)–(4), $G(z'|\tau) := e^{2\pi\mathrm{i}x} F(\tau + z'|\tau)$ satisfies (30), so $\bar{G}(u'|\tau) = e^{2\pi\mathrm{i}x} F(\tau + z'|\tau)$ satisfies (33), where $u' = e^{2\pi\mathrm{i}z'}$.

We now compute the form of the system in the transition (2)–(3). We first prove:

Lemma 27. *Set $u := e^{2\pi\mathrm{i}z}$, $v := e^{2\pi\mathrm{i}(\tau - z)}$. When $0 < \Im(z) < \Im(\tau)$, we have $|u| < 1$, $|v| < 1$. When $k \geq 0$, $(\theta^{(k)}/\theta)(z|\tau) = (-\mathrm{i}\pi)^k + \sum_{s,t\geq 0, s+t>0} a_{st}^{(k)} u^s v^t$, where the sum in the r.h.s. is convergent in the domain $|u| < 1$, $|v| < 1$.*

Proof. This is clear if $k=0$. Set $q=uv=e^{2\pi \mathrm{i}\tau}$. We have $\theta(z|\tau)=u^{1/2}\prod_{s>0}(1-q^s u)\prod_{s\geq 0}(1-q^s u^{-1})\cdot(2\pi\mathrm{i})^{-1}\prod_{s>0}(1-q^s)^{-2}$, so

$$\begin{aligned}(\theta'/\theta)(z|\tau) &= \mathrm{i}\pi - 2\pi\mathrm{i}\sum_{s>0} q^s u/(1-q^s u) + 2\pi\mathrm{i}\sum_{s\geq 0} q^s u^{-1}/(1-q^s u^{-1}) \\ &= -\mathrm{i}\pi - 2\pi\mathrm{i}\sum_{s\geq 0}\frac{u^{s+1}v^s}{1-u^{s+1}v^s} + 2\pi\mathrm{i}\sum_{s\geq 0}\frac{u^s v^{s+1}}{1-u^s v^{s+1}} \\ &= -\mathrm{i}\pi + \sum_{s+t>0} a_{st}u^s v^t,\end{aligned}$$

where $a_{st}=2\pi\mathrm{i}$ if $(s,t)=k(r,r+1)$, $k>0$, $r\geq 0$, and $a_{st}=-2\pi\mathrm{i}$ if $(s,t)=k(r+1,r)$, $k>0$, $r\geq 0$. One checks that this series is convergent in the domain $|u|<1$, $|v|<1$. This proves the lemma for $k=1$.

We then prove the remaining cases by induction, using

$$\frac{\theta^{(k+1)}}{\theta}(z|\tau) = \frac{\theta^{(k)}}{\theta}(z|\tau)\frac{\theta'}{\theta}(z|\tau) + \frac{\partial}{\partial z}\frac{\theta^{(k)}}{\theta}(z|\tau). \qquad \square$$

Using the expansion

$$\begin{aligned}\frac{\theta(z+x|\tau)x}{\theta(z|\tau)\theta(x|\tau)} &= \frac{x}{\theta(x|\tau)}\sum_{k\geq 0}(\theta^{(k)}/\theta)(z|\tau)\frac{x^k}{k!} \\ &= \frac{\pi x}{\sin(\pi x)}\left(1+\sum_{n>0}q^n P_n(x)\right)\left(\sum_{k\geq 0}\left((-\mathrm{i}\pi)^k + \sum_{s+t>0}a_{st}^{(k)}u^s v^t\right)\frac{x^k}{k!}\right) \\ &= \frac{\pi x}{\sin(\pi x)}e^{-\mathrm{i}\pi x} + \sum_{s+t>0}a_{st}(x)u^s v^t = \frac{2\mathrm{i}\pi x}{e^{2\mathrm{i}\pi x}-1} + \sum_{s+t>0}a_{st}(x)u^s v^t,\end{aligned}$$

the form of the system in the transition (2)–(3) is

$$\begin{aligned}\partial_z F(z|\tau) &= \left(-\frac{2\mathrm{i}\pi \operatorname{ad} x}{e^{2\mathrm{i}\pi \operatorname{ad} x}-1}(y) + \sum_{s,t|s+t>0} a_{st}u^s v^t\right)F(z|\tau) \\ &= \left(2\mathrm{i}\pi\tilde{y} + \sum_{s,t|s+t>0} a_{st}u^s v^t\right)F(z|\tau), \end{aligned} \tag{34}$$

where each homogeneous part of $\sum_{s,t} a_{st}u^s v^t$ converges for $|u|<1$, $|v|<1$.

Lemma 28. *There exists a solution $F_c(z|\tau)$ of (34) defined for $0<\Im(z)<\Im(\tau)$ such that*

$$F_c(z|\tau) = u^{\tilde{y}}(1+\sum_{k>0}\sum_{s\leq s(k)}\log(u)^s f_{ks}(u,v))$$

($\log u = i\pi z$, $u^{\tilde y} = e^{2\pi \mathrm{i} z\tilde y}$), where $f_{ks}(u,v)$ is an analytic function taking its values in the homogeneous part of the algebra of degree k, convergent for $|u|<1$ and $|v|<1$, and vanishing at $(0,0)$. This function is uniquely defined up to right multiplication by an analytic function of the form $1+\sum_{k>0} a_k(q)$ (recall that $q=uv$), where $a_k(q)$ is an analytic function on $\{q||q|<1\}$, vanishing at $q=0$, with values in the degree k part of the algebra.

Proof of Lemma. We set $G(z|\tau) := u^{-\tilde y}F(z|\tau)$, so $G(z|\tau)$ should satisfy

$$\partial_z G(z|\tau) = \exp(-\mathrm{ad}(\tilde y)\log u)\left\{\sum_{s+t>0} a_{st}u^s v^t\right\} G(z|\tau),$$

which has the general form

$$\partial_z G(z|\tau) = \Big(\sum_{k>0}\sum_{s\le a(k)} \log(u)^s a_{ks}(u,v)\Big) G(z|\tau),$$

where $a_{ks}(u,v)$ is analytic in $|u|<1$, $|v|<1$ and vanishes at $(0,0)$. We show that this system admits a solution of the form

$$1+\sum_{k>0}\sum_{s\le s(k)} \log(u)^s f_{ks}(u,v),$$

with $f_{ks}(u,v)$ analytic in $|u|<1$, $|v|<1$, in the degree k part of the algebra, vanishing at $(0,0)$ for $s\neq 0$. For this, we solve inductively (in k) the system of equations

$$\partial_z\left(\sum_s (\log u)^s f_{ks}(u,v)\right) = \sum_{s',s'',k',k''|k'+k''=k} (\log u)^{s'+s''} a_{k's'}(u,v) f_{k''s''}(u,v). \tag{35}$$

Let $\mathcal{O}$ be the ring of analytic functions on $\{(u,v)||u|<1,|v|<1\}$ (with values in a finite-dimensional vector space) and let $\mathfrak{m}\subset\mathcal{O}$ be the subset of functions vanishing at $(0,0)$. We have an injection $\mathcal{O}[X]\to\{$analytic functions in (u,v), $|u|<1$, $|v|<1$, $u\notin\mathbb{R}_-\}$, given by $f(u,v)X^k\mapsto(\log u)^k f(u,v)$. The endomorphism $\frac{\partial}{\partial z} = 2\pi\mathrm{i}\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right)$ then corresponds to the endomorphism of $\mathcal{O}[X]$ given by $2\pi\mathrm{i}\left(\frac{\partial}{\partial X} + u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right)$. It is surjective, and restricts to a surjective endomorphism of $\mathfrak{m}[X]$. The latter surjectivity implies that equation (35) can be solved.

Let us show that the solution $G(z|\tau)$ is unique up to right multiplication by functions of q as in the lemma. The ratio of two solutions is of the form $1+\sum_{k>0}\sum_{s\le s(k)}\log(u)^s f_{ks}(u,v)$ and is killed by ∂_z. Now the kernel of the endomorphism of $\mathfrak{m}[X]$ given by $2\pi\mathrm{i}\left(\frac{\partial}{\partial X} + u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right)$ is $m^*(\mathfrak{m}_1)$, where

$m^*(\mathfrak{m}_1) \subset \mathfrak{m}$ is the set of all functions of the form $a(uv)$, where a is an analytic function on $\{q||q| < 1\}$ vanishing at 0. This implies that the ratio of two solutions is as above. $\square$

End of proof of Theorem. Similarly, there exists a solution $F_d(z|\tau)$ of (34) defined in the same domain, such that

$$F_d(z|\tau) = v^{-\tilde{y}} \left(1 + \sum_{k>0} \sum_{s \le t(k)} \log(v)^t g_{ks}(u,v) \right),$$

where $b_{ks}(u,v)$ is as above (and $\log v = \mathrm{i}\pi(\tau - z)$, $v^{-\tilde{y}} = \exp(2\pi\mathrm{i}(z-\tau)\tilde{y})$). The solution $F_d(z|\tau)$ is defined up to right multiplication by a function of q as above.

We now study the ratio $F_c(z|\tau)^{-1}F_d(z|\tau)$. This is a function of τ only, and it has the form

$$q^{-\tilde{y}}\Big(1 + \sum_{k>0} \sum_{s \le s(k), t \le t(k)} (\log u)^s (\log v)^t a_{kst}(u,v)\Big),$$

where $a_{kst}(u,v) \in \mathfrak{m}$ (as $v^{-\tilde{y}}(1 + \sum_{k>0}\sum_{s\le s(k)}(\log u)^s c_{ks}(u,v))v^{\tilde{y}}$ has the form $1 + \sum_{k>0}\sum_{s,t\le t(k)}$ $(\log u)^s(\log v)^t d_{ks}(u,v)$, where $d_{ks}(u,v) \in \mathfrak{m}$ if $c_{ks}(u,v) \in \mathfrak{m}$). Set $\log q := \log u + \log v = 2\pi\mathrm{i}\tau$, then this ratio can be rewritten $q^{-\tilde{y}}\{1 + \sum_{k>0}\sum_{s\le s(k),t\le t(k)}(\log u)^s(\log q)^t b_{kst}(u,v)\}$, where $b_{kst}(u,v) \in \mathfrak{m}$, and since the product of this ratio with $q^{\tilde{y}}$ is killed by ∂_z (which identifies with the endomorphism $2\pi\mathrm{i}\left(\frac{\partial}{\partial X} + u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right)$ of $\mathcal{O}[X]$), the ratio is in fact of the form

$$F_c^{-1}F_d(z|\tau) = q^{\tilde{y}} \left(1 + \sum_{k>0} \sum_{s \le s(k)} (\log q)^s a_{ks}(q) \right),$$

where a_{ks} is analytic in $\{q||q| < 1\}$, vanishing at $q = 0$.

It follows that

$$F_c^{-1}F_d(z|\tau) = e^{-2\pi\mathrm{i}\tau\tilde{y}} \left(1 + \sum_{k>0} (\text{degree } k) O(\tau^k e^{-2\pi\mathrm{i}\tau}) \right). \tag{36}$$

In addition to F_c and F_d, which have prescribed behaviors in zones (2) and (3), we define solutions of (30) in V by prescribing behaviors in the remaining asymptotic zones: $F_a(z|\tau) \simeq z^t$ as $z \to 0^+$ for any τ; $F_b(z|\tau) \simeq (2\pi z/\mathrm{i})^t$ as $z \to \mathrm{i}0^+$ for any τ (in particular in zone (1)); $e^{2\pi\mathrm{i}x}F_e(z|\tau) \simeq (2\pi(\tau - z)/\mathrm{i})^t$ as $z = \tau - \mathrm{i}0^+$ for any τ; $e^{2\pi\mathrm{i}x}F_f(z|\tau) \simeq (z-\tau)^t$ when $z = \tau + 0^+$ for any τ (in particular in zone (4)).

Then $F_0(z|\tau) = F_a(z|\tau)$, and $e^{-2\pi\mathrm{i}x}F_0(z-\tau|\tau) = F_f(z|\tau)$. We have $F_b = F_a(2\pi/\mathrm{i})^t$, $F_f = F_e(2\pi\mathrm{i})^{-t}$.

Let us now compute the ratio between F_b and F_c. Recall that $u = e^{2\pi \mathrm{i} z}$, $v = e^{2\pi\mathrm{i}(\tau - z)}$. Set $\bar{F}(u,v) := F(z|\tau)$. Using the expansion of $\theta(z|\tau)$, one shows that (30) has the form

$$\partial_u \bar{F}(u,v) = \left(\frac{A(u,v)}{u} + \frac{B(u,v)}{u-1}\right)\bar{F}(u,v),$$

where $A(u,v)$ is holomorphic in the region $|v| < 1/2$, $|u| < 2$, and $A(u,0) = \tilde{y}$, $B(u,0) = t$. We have $\bar{F}_b(u,v) = (1-u)^t(1 + \sum_k \sum_{s\le s(k)} \log(1-u)^k b_{ks}(u,v))$ and $\bar{F}_b(u,v) = u^{\tilde{t}}(1+\sum_k \sum_{s\le s(k)} \log(u)^k a_{ks}(u,v))$, with a_{ks}, b_{ks} analytic, and $a_{ks}(0,v) = b_{ks}(1,v) = 0$. The ratio $\bar{F}_b^{-1}\bar{F}_c$ is an analytic function of q only, which coincides with $\Phi(\tilde{y},t)$ for $q=0$, so it has the form $\Phi(\tilde{y},t) + \sum_{k>0} a_k(q)$, where $a_k(q)$ has degree k, is analytic in a neighborhood of $q=0$, and vanishes at $q=0$. Therefore

$$F_c(z|\tau) = F_b(z|\tau)\big(\Phi(\tilde{y},t) + O(e^{2\pi\mathrm{i}\tau})\big).$$

In the same way, one proves that

$$F_e(z|\tau) = F_d\big(e^{-2\pi\mathrm{i}x}\Phi(-\tilde{y}-t,t)^{-1} + O(e^{2\pi\mathrm{i}\tau})\big).$$

Let us set $\bar{G}_d(u',v') := e^{2\pi\mathrm{i}x}F_d(\tau+z'|\tau)$, $\bar{G}_e(u',v') := e^{2\pi\mathrm{i}x}F_e(\tau+z'|\tau)$, where $u' = e^{2\pi\mathrm{i}(\tau+z')}$, $v' = e^{-2\pi\mathrm{i}z'}$, then $\bar{G}_d(u',v') \simeq (v')^{-\tilde{y}-t}e^{2\pi\mathrm{i}x}$ as $(u',v') \to (0^+,0^+)$ and $\bar{G}_e(u',v') \simeq (1-v')^t$ as $v' \to 1^-$ for any u', and both $\bar{G}_d$ and $\bar{G}_e$ are solutions of $\partial_{v'}\bar{G}(u',v') = [-(\tilde{y}+t)/v' + t/(v'-1) + O(u')]\bar{G}(v')$. Therefore $\bar{G}_d = \bar{G}_e[\Phi(-\tilde{y}-t,t)e^{2\pi\mathrm{i}x} + O(u')]$.

Combining these results, we get the following:

Lemma 29.

$$B(\tau) \simeq (2\pi\mathrm{i})^t\Phi(-\tilde{y}-t,t)e^{2\pi\mathrm{i}x}e^{2\mathrm{i}\pi\tau\tilde{y}}\Phi(\tilde{y},t)^{-1}(2\pi/\mathrm{i})^{-t},$$

in the sense that the left (equivalently, right) ratio of these quantities has the form $1 + \sum_{k>0}(\text{degree } k)O(\tau^{n(k)}e^{2\pi\mathrm{i}\tau})$ *for* $n(k) \ge 0$.

Recall that we have proved:

$$F(z|\tau) = F_0(z|\tau)\exp\left(-\frac{a_0}{2\pi\mathrm{i}}\left(\int_{\mathrm{i}}^{\tau} E_2 + C\right)t\right)[F(\tau)],$$

where C is such that $\int_{\mathrm{i}}^{\tau} E_2 + C = \tau + O(e^{2\pi\mathrm{i}\tau})$.

Set $X(\tau) := \exp\left(-\frac{a_0}{2\pi\mathrm{i}}\left(\int_{\mathrm{i}}^{\tau} E_2 + C\right)t\right)[F(\tau)]$. As $\tau \to \mathrm{i}\infty$,

$$X(\tau) = \exp\left(-\frac{\tau}{2\pi\mathrm{i}}([\Delta] + a_0 t)\right)\left(1 + \sum_{k>0}(\text{degree } k)O(\tau^{f(k)}e^{2\pi\mathrm{i}\tau})\right).$$

Then

$$\begin{aligned}\tilde{B} &= F(z|\tau)^{-1}e^{2\pi \mathrm{i}x}F(z+\tau|\tau) = X(\tau)^{-1}B(\tau)X(\tau)\\ &= \mathrm{Ad}\left(\left(1+\sum_{k>0}(\text{degree } k)O(\tau^{f(k)}e^{2\pi\mathrm{i}\tau})\right)^{-1}\exp\left(\frac{\tau}{2\pi \mathrm{i}}([\varDelta]+a_0 t)\right)\right)\\ &\quad\times\Big(\big((2\pi\mathrm{i})^t\varPhi(-\tilde{y}-t,t)e^{2\pi\mathrm{i}x}e^{2\pi\mathrm{i}\tau\tilde{y}}\varPhi(\tilde{y},t)^{-1}(2\pi/\mathrm{i})^{-t}\big)\\ &\quad\times\big(1+\sum_{k>0}(\text{degree } k)O(\tau^{n(k)}e^{2\pi\mathrm{i}\tau})\big)\Big),\end{aligned}$$

where $\mathrm{Ad}(u)(x) = uxu^{-1}$.

Now, $[\varDelta]+a_0t$ commutes with $\tilde{y}$ and t; assume for a moment that

$$\mathrm{Ad}\left(\exp\left(\frac{\tau}{2\pi\mathrm{i}}([\varDelta]+a_0t)\right)\right)(e^{2\pi\mathrm{i}x}e^{2\pi\mathrm{i}\tau\tilde{y}}) = e^{2\pi\mathrm{i}x}$$

(Lemma 30 below), then

$$\begin{aligned}&\mathrm{Ad}\left(\exp\left(\frac{\tau}{2\pi\mathrm{i}}([\varDelta]+a_0t)\right)\right)\left((2\pi\mathrm{i})^t\varPhi(-\tilde{y}-t,t)e^{2\pi\mathrm{i}x}e^{2\pi\mathrm{i}\tau\tilde{y}}\varPhi(\tilde{y},t)^{-1}(2\pi/\mathrm{i})^{-t}\right)\\ &\qquad= (2\pi\mathrm{i})^t\varPhi(-\tilde{y}-t,t)e^{2\pi\mathrm{i}x}\varPhi(\tilde{y},t)^{-1}(2\pi/\mathrm{i})^{-t}.\end{aligned}$$

On the other hand, $\mathrm{Ad}\left(\exp\left(\frac{\tau}{2\pi\mathrm{i}}([\varDelta]+a_0t)\right)\right)(1+\sum_{k>0}(\text{degree } k)O(\tau^{n(k)} e^{2\pi\mathrm{i}\tau}))$ has the form $1+\sum_{k>0}(\text{degree } k)O(\tau^{n'(k)}e^{2\pi\mathrm{i}\tau})$, where $n'(k)\geq 0$. It follows that

$$\begin{aligned}\tilde{B} = \mathrm{Ad}\big(1&+\sum_{k>0}(\text{degree } k)O(\tau^{f(k)}e^{2\pi\mathrm{i}\tau})\big)\\ &\times\Big(\big((2\pi\mathrm{i})^t\varPhi(-\tilde{y}-t,t)e^{2\pi\mathrm{i}x}\varPhi(\tilde{y},t)^{-1}(2\pi/\mathrm{i})^{-t}\big)\\ &\times\left(1+\sum_{k>0}(\text{degree } k)O(\tau^{n'(k)}e^{2\pi\mathrm{i}\tau})\right)\Big);\end{aligned}$$

now

$$\begin{aligned}\mathrm{Ad}\Big((2\pi\mathrm{i})^t&\varPhi(-\tilde{y}-t,t)e^{2\pi\mathrm{i}x}\varPhi(\tilde{y},t)^{-1}(2\pi/\mathrm{i})^{-t}\Big)^{-1}\\ &\times\left(1+\sum_{k>0}(\text{degree } k)O(\tau^{f(k)}e^{2\pi\mathrm{i}\tau})\right)\\ &= 1+\sum_{k>0}(\text{degree } k)O(\tau^{f(k)}e^{2\pi\mathrm{i}\tau}),\end{aligned}$$

so

$$\begin{aligned}\tilde{B} &= \left((2\pi\mathrm{i})^t \Phi(-\tilde{y}-t,t)e^{2\pi\mathrm{i}x}\Phi(\tilde{y},t)^{-1}(2\pi/\mathrm{i})^{-t}\right)\\ &\quad\times\left(1+\sum_{k>0}(\text{degree } k)O(\tau^{f(k)}e^{2\pi\mathrm{i}\tau})\right)\\ &\quad\times\left(1+\sum_{k>0}(\text{degree } k)O(\tau^{n'(k)}e^{2\pi\mathrm{i}\tau})\right)\\ &= \left((2\pi\mathrm{i})^t \Phi(-\tilde{y}-t,t)e^{2\pi\mathrm{i}x}\Phi(\tilde{y},t)^{-1}(2\pi/\mathrm{i})^{-t}\right)\\ &\quad\times\left(1+\sum_{k>0}(\text{degree } k)O(\tau^{n''(k)}e^{2\pi\mathrm{i}\tau})\right)\end{aligned}$$

for $n''(k) \geq 0$. Since $\tilde{B}$ is constant w.r.t. τ, this implies

$$\tilde{B} = (2\pi\mathrm{i})^t \Phi(-\tilde{y}-t,t)e^{2\pi\mathrm{i}x}\Phi(\tilde{y},t)^{-1}(2\pi/\mathrm{i})^{-t},$$

as claimed.

We now prove the conjugation used above.

Lemma 30. *For any $\tau \in \mathbb{C}$, we have*

$$e^{\frac{\tau}{2\pi\mathrm{i}}([\Delta]+a_0t)}e^{2\pi\mathrm{i}x}e^{-\frac{\tau}{2\pi\mathrm{i}}([\Delta]+a_0t)}e^{2\mathrm{i}\pi\tau\tilde{y}} = e^{2\pi\mathrm{i}x}.$$

Proof. We have $[\Delta]+a_0t = \Delta_0 + \sum_{k\geq 0} a_{2k}(\delta_{2k} + (\mathrm{ad}\,x)^{2k}(t))$ (where $\delta_0 = 0$), so $[[\Delta]+a_0t, x] = y - \sum_{k\geq 0} a_{2k}(\mathrm{ad}\,x)^{2k+1}(t)$. Recall that

$$\sum_{k\geq 0} a_{2k}u^{2k} = \frac{\pi^2}{\sin^2(\pi u)} - \frac{1}{u^2},$$

then $[[\Delta]+a_0t, x] = y - (\mathrm{ad}\,x)\left(\frac{\pi^2}{\sin^2(\pi \mathrm{ad}\,x)} - \frac{1}{(\mathrm{ad}\,x)^2}\right)(t)$. So

$$\begin{aligned}&e^{-2\pi\mathrm{i}x}\left(\frac{1}{2\pi\mathrm{i}}([\Delta]+a_0t)\right)e^{2\pi\mathrm{i}x}\\ &\quad= \frac{1}{2\pi\mathrm{i}}([\Delta]+a_0t) + \frac{e^{-2\pi\mathrm{i}\,\mathrm{ad}\,x}-1}{\mathrm{ad}\,x}\left(\left[x, \frac{1}{2\pi\mathrm{i}}([\Delta]+a_0t)\right]\right)\\ &\quad= \frac{1}{2\pi\mathrm{i}}([\Delta]+a_0t) - \frac{1}{2\pi\mathrm{i}}\frac{e^{-2\pi\mathrm{i}\,\mathrm{ad}\,x}-1}{\mathrm{ad}\,x}\\ &\qquad\left(y - (\mathrm{ad}\,x)\left(\frac{\pi^2}{\sin^2(\pi\mathrm{ad}\,x)} - \frac{1}{(\mathrm{ad}\,x)^2}\right)(t)\right).\end{aligned}$$

We have

$$-\frac{1}{2\pi \mathrm{i}} \frac{e^{-2\pi \mathrm{i} \mathrm{ad}\, x}-1}{\mathrm{ad}\, x}\left(y-(\mathrm{ad}\, x)\left(\frac{\pi^2}{\sin^2(\pi \mathrm{ad}\, x)}-\frac{1}{(\mathrm{ad}\, x)^2}\right)(t)\right)=-2\pi \mathrm{i} \tilde{y};$$

therefore we get

$$e^{-2\pi \mathrm{i} x}\left(\frac{1}{2\pi \mathrm{i}}([\Delta]+a_0 t)\right) e^{2\pi \mathrm{i} x}=\frac{1}{2\pi \mathrm{i}}([\Delta]+a_0 t)-2\pi \mathrm{i} \tilde{y}.$$

Multiplying by τ, taking the exponential, and using the fact that $[\Delta]+a_0 t$ commutes with $\tilde{y}$, we get

$$e^{-2\pi \mathrm{i} x} e^{\frac{\tau}{2\pi \mathrm{i}}([\Delta]+a_0 t)} e^{2\pi \mathrm{i} x}=e^{\frac{\tau}{2\pi \mathrm{i}}([\Delta]+a_0 t)} e^{-2\pi \mathrm{i} \tau \tilde{y}},$$

which proves the lemma. □

This ends the proof of Theorem 26.

5 Construction of morphisms $\Gamma_{1,[n]} \to \mathbf{G}_n \rtimes S_n$

In this section, we fix a field $\mathbf{k}$ of characteristic zero. We denote the algebras $\bar{\mathfrak{t}}^{\mathbf{k}}_{1,n}$, $\mathfrak{t}^{\mathbf{k}}_n$ simply by $\bar{\mathfrak{t}}_{1,n}$, $\mathfrak{t}_n$. The above group $\mathbf{G}_n$ is the set of $\mathbb{C}$-points of a group scheme defined over $\mathbb{Q}$, and we now again denote by $\mathbf{G}_n$ the set of its $\mathbf{k}$-points.

5.1 Construction of morphisms $\Gamma_{1,[n]} \to \mathbf{G}_n \rtimes S_n$ from a 5-tuple $(\Phi_\lambda, \tilde{A}, \tilde{B}, \tilde{\Theta}, \tilde{\Psi})$

Let Φ_λ be a λ-associator defined over $\mathbf{k}$. This means that $\Phi_\lambda \in \exp(\hat{\mathfrak{t}}_3)$ (the Lie algebras are now over $\mathbf{k}$),

$$\Phi_\lambda^{3,2,1}=\Phi_\lambda^{-1}, \quad \Phi_\lambda^{2,3,4} \Phi_\lambda^{1,23,4} \Phi_\lambda^{1,2,3}=\Phi_\lambda^{1,2,34} \Phi_\lambda^{12,3,4}, \tag{37}$$

$$e^{\lambda t_{31}/2} \Phi_\lambda^{2,3,1} e^{\lambda t_{23}/2} \Phi_\lambda e^{\lambda t_{12}/2} \Phi_\lambda^{3,1,2}=e^{\lambda(t_{12}+t_{23}+t_{13})/2}. \tag{38}$$

For example, the KZ associator is a $2\pi\mathrm{i}$-associator over $\mathbb{C}$.

Proposition 31. *If $\tilde{\Theta}, \tilde{\Psi} \in \mathbf{G}_1$ and $\tilde{A}, \tilde{B} \in \exp(\hat{\bar{\mathfrak{t}}}_{1,2})$ satisfy the "$\Gamma_{1,1}$ identities" (27), the "$\Gamma_{1,2}$ identities" (28), (29), and the "$\Gamma_{1,[3]}$ identities" (23), (22), (26) (with $2\pi\mathrm{i}$ replaced by λ), as well as $\tilde{A}^{\emptyset,1}=\tilde{A}^{1,\emptyset}=\tilde{B}^{\emptyset,1}=\tilde{B}^{1,\emptyset}=1$, then one defines a morphism $\Gamma_{1,[n]} \to \mathbf{G}_n \rtimes S_n$ by*

$$\Theta \mapsto [\tilde{\Theta}] e^{\mathrm{i}\frac{\pi}{2} \sum_{i<j} \bar{t}_{ij}},$$
$$\Psi \mapsto [\tilde{\Psi}] e^{\mathrm{i}\frac{\pi}{6} \sum_{i<j} \bar{t}_{ij}},$$

$$\sigma_i \mapsto \left\{\Phi_\lambda^{1\ldots i-1,i,i+1}\right\}^{-1} e^{\lambda \bar t_{i,i+1}/2}(i,i+1)\left\{\Phi_\lambda^{1\ldots i-1,i,i+1}\right\},$$
$$C_{jk} \mapsto \Big\{\Phi_{\lambda,j}^{-1}\Phi_\lambda^{j,j+1,\ldots,n}\ldots \Phi_\lambda^{j\ldots,k-1,\ldots,n}(e^{\lambda t_{12}})^{j\ldots k-1,k\ldots n} \times \left(\Phi_\lambda^{j,j+1,\ldots,n}\ldots\Phi_\lambda^{j\ldots,k-1,\ldots,n}\right)^{-1}\Phi_{\lambda,j}\Big\},$$
$$A_i \mapsto \{\Phi_{\lambda,i}\}^{-1}\tilde A^{1\ldots i-1,i\ldots n}\{\Phi_{\lambda,i}\},$$
$$B_i \mapsto \{\Phi_{\lambda,i}\}^{-1}\tilde B^{1\ldots i-1,i\ldots n}\{\Phi_{\lambda,i}\},$$

where $\Phi_{\lambda,i} = \Phi_\lambda^{1\ldots i-1,i,i+1\ldots n}\ldots\Phi_\lambda^{1\ldots n-2,n-1,n}$.

According to Section 4.4, the representations γ_n are obtained by the procedure described in this proposition from the KZ associator, $\tilde\Theta, \tilde\Psi$ arising from γ_1, and $\tilde A, \tilde B$ arising from γ_2.

Note also that the analogue of (22) is equivalent to the pair of equations

$$e^{\lambda\bar t_{12}/2}\tilde A^{2,1}e^{\lambda\bar t_{12}/2}\tilde A = 1,$$
$$(e^{\lambda\bar t_{12}/2}\tilde A)^{3,12}\Phi_\lambda^{3,1,2}(e^{\lambda\bar t_{12}/2}\tilde A)^{2,31}\Phi_\lambda^{2,3,1}(e^{\lambda\bar t_{12}/2}\tilde A)^{1,23}\Phi_\lambda^{1,2,3} = 1,$$

and similarly, (23) is equivalent to the same equations, with $\tilde A, \lambda$ replaced by $\tilde B, -\lambda$.

Remark 32. One can prove that if Φ_λ satisfies only the pentagon equation and $\tilde\Theta, \tilde\Psi, \tilde A, \tilde B$ satisfy the the "$\Gamma_{1,1}$ identities" (27), the "$\Gamma_{1,2}$ identities" (28), (29), and the "$\Gamma_{1,3}$ identities" (24), (26), then the above formulas (removing σ_i) define a morphism $\Gamma_{1,n} \to \mathbf{G}_n$. In the same way, if Φ_λ satisfies all the associator conditions and $\tilde A, \tilde B$ satisfy the $\Gamma_{1,[3]}$ identities (22), (23), (26), then the above formulas (removing Θ, Ψ) define a morphism $\overline{\mathrm{B}}_{1,n} \to \exp(\hat{\bar{\mathfrak{t}}}_{1,n}) \rtimes S_n$.

Proof. Let us prove that the identity $(A_i, A_j) = 1$ $(i < j)$ is preserved. Applying $x \mapsto x^{1,\ldots,i-1,i\cdots j-1,j\cdots n}$ to the first identity of (24), we get

$$\left(\tilde A^{1\ldots i-1,i\ldots n}, \Phi_\lambda^{1\ldots,i\ldots j-1,\ldots n}\tilde A^{1\ldots j-1,j\ldots n}\left(\Phi_\lambda^{-1}\right)^{1\ldots,i\ldots j-1,\ldots n}\right) = 1.$$

The pentagon identity implies

$$\Phi_\lambda^{1\ldots,i,\ldots n}\cdots\Phi_\lambda^{1\ldots,j-1,\ldots n} = \left(\Phi_\lambda^{i,i+1,\ldots n}\cdots\Phi_\lambda^{i\ldots,j-1,\ldots,n}\right)\Phi_\lambda^{1\ldots,i\ldots j-1,\ldots n}\left(\Phi_\lambda^{1\ldots,i,\ldots j-1}\ldots\Phi_\lambda^{1\ldots,j-2,j-1}\right), \tag{39}$$

so the above identity is rewritten

$$(\Phi_\lambda^{i,i+1,\ldots n}\cdots\Phi_\lambda^{i\ldots,j-1,\ldots,n}\tilde A^{1\ldots i-1,i\ldots n}\left(\Phi_\lambda^{i,i+1,\ldots n}\cdots\Phi_\lambda^{i\ldots,j-1,\ldots,n}\right)^{-1},$$

$$\Phi_\lambda^{1\dots,i,\dots n}\cdots\Phi_\lambda^{1\dots,j-1,\dots n}\left(\Phi_\lambda^{1\dots,i,\dots j-1}\cdots\Phi_\lambda^{1\dots,j-2,\dots j-1}\right)^{-1}\tilde{A}^{1\dots j-1,j\dots n}$$
$$\Phi_\lambda^{1\dots,i,\dots j-1}\cdots\Phi_\lambda^{1\dots,j-2,\dots j-1}\left(\Phi_\lambda^{1\dots,i,\dots n}\cdots\Phi_\lambda^{1\dots,j-1,\dots n}\right)^{-1}\Big) = 1.$$

Now $\Phi_\lambda^{i,i+1,\dots n}, \dots, \Phi_\lambda^{i\dots,j-1,\dots,n}$ commute with $\tilde{A}^{1\dots i-1,i\dots n}$, and $\Phi_\lambda^{1\dots,i,\dots j-1}$, $\dots, \Phi_\lambda^{1\dots,j-2,\dots j-1}$ commute with $\Phi_\lambda^{1\dots,i,\dots j-1}\dots\Phi_\lambda^{1\dots,j-2,\dots j-1}$, which implies

$$\Big(\tilde{A}^{1\dots i-1,i\dots n},\Phi_\lambda^{1\dots,i,\dots n}\cdots$$
$$\Phi_\lambda^{1\dots,j-1,\dots n}\tilde{A}^{1\dots j-1,j\dots n}\left(\Phi_\lambda^{1\dots,i,\dots n}\dots\Phi_\lambda^{1\dots,j-1,\dots n}\right)^{-1}\Big) = 1,$$

so that $(A_i, A_j) = 1$ is preserved. In the same way, one shows that $(B_i, B_j) = 1$ is preserved.

Let us show that $\left(B_k, A_k A_j^{-1}\right) = C_{jk}$ is preserved (if $j \le k$).

$$\begin{aligned}
&\left(\Phi_{\lambda,k}^{-1}\tilde{B}^{1\dots k-1,k\dots n}\Phi_{\lambda,k},\Phi_{\lambda,k}^{-1}\tilde{A}^{1\dots k-1,k\dots n}\Phi_{\lambda,k}\Phi_{\lambda,j}^{-1}(\tilde{A}^{1\dots j-1,j\dots n})^{-1}\Phi_{\lambda,j}\right)\\
&= \Phi_{\lambda,j}^{-1}\Big(\left(\Phi_\lambda^{1\dots,j,\dots n}\cdots\Phi_\lambda^{1\dots,k-1,\dots n}\right)\tilde{B}^{1\dots k-1,k\dots n}\left(\Phi_\lambda^{1\dots,j,\dots n}\cdots\Phi_\lambda^{1\dots,k-1,\dots n}\right)^{-1},\\
&\quad\left(\Phi_\lambda^{1\dots,j,\dots n}\cdots\Phi_\lambda^{1\dots,k-1,\dots n}\right)\tilde{A}^{1\dots k-1,k\dots n}\left(\Phi_\lambda^{1\dots,j,\dots n}\cdots\Phi_\lambda^{1\dots,k-1,\dots n}\right)^{-1}\\
&\quad(\tilde{A}^{1\dots j-1,j\dots n})^{-1}\Big)\Phi_{\lambda,j}\\
&= \Phi_{\lambda,j}^{-1}\big(\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\Phi_\lambda^{1\dots,j\dots k-1,\dots n}\tilde{B}^{1\dots k-1,k\dots n}\\
&\quad\left(\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\Phi_\lambda^{1\dots,j\dots k-1,\dots n}\right)^{-1},\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\\
&\quad\Phi_\lambda^{1\dots,j\dots k-1,\dots n}\tilde{A}^{1\dots k-1,k\dots n}\left(\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\Phi_\lambda^{1\dots,j\dots k-1,\dots n}\right)^{-1}\\
&\quad(\tilde{A}^{1\dots j-1,j\dots n})^{-1}\big)\Phi_{\lambda,j}\\
&= \Phi_{\lambda,j}^{-1}\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\big(\Phi_\lambda^{1\dots,j\dots k-1,\dots n}\tilde{B}^{1\dots k-1,k\dots n}\left(\Phi_\lambda^{1\dots,j\dots k-1,\dots n}\right)^{-1},\\
&\quad\Phi_\lambda^{1\dots,j\dots k-1,\dots n}\tilde{A}^{1\dots k-1,k\dots n}\left(\Phi_\lambda^{1\dots,j\dots k-1,\dots n}\right)^{-1}(\tilde{A}^{1\dots j-1,j\dots n})^{-1}\big)\\
&\quad\left(\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\right)^{-1}\Phi_{\lambda,j}\\
&= \Phi_{\lambda,j}^{-1}\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\Big\{\Phi\left(\tilde{B}^{12,3},\tilde{A}^{12,3}\Phi_\lambda^{-1}(\tilde{A}^{1,23})^{-1}\Phi_\lambda\right)\Phi_\lambda^{-1}\Big\}^{1\dots,j\dots k-1,\dots n}\\
&\quad\left(\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\right)^{-1}\Phi_{\lambda,j}\\
&= \Phi_{\lambda,j}^{-1}\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}(e^{2\pi i\bar{t}_{12}})^{j\dots k-1,k\dots n}\left(\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\right)^{-1}\Phi_{\lambda,j},
\end{aligned}$$

where the second identity uses (39) and the invariance of Φ_λ, the third identity uses the fact that $\Phi_\lambda^{j,j+1,\dots n}, \dots, \Phi_\lambda^{j\dots,k-1,\dots n}$ commute with $\tilde{A}^{1,\dots,j-1,j\dots n}$ (again by the invariance of Φ_λ), and the last identity uses (26). So $\left(B_k, A_k A_j^{-1}\right) = C_{jk}$ is preserved. One shows similarly that

$$\begin{aligned}&\big(\Phi_{\lambda,k}^{-1}\tilde{B}^{1\dots k-1,k\dots n}\Phi_{\lambda,k}\Phi_{\lambda,j}^{-1}(\tilde{B}^{1\dots j-1,j\dots n})^{-1}\Phi_{\lambda,j},\Phi_{\lambda,k}^{-1}\tilde{A}^{1\dots k-1,k\dots n}\Phi_{\lambda,k}\big)\\ &\quad=\Phi_j^{-1}\Phi^{j,j+1,\dots n}\cdots\Phi^{j\dots,k-1,\dots n}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{j\dots k-1,k\dots n}\\ &\qquad\left(\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\right)^{-1}\Phi_{\lambda,j},\end{aligned}$$

so that $\left(B_kB_j^{-1},A_k\right)=C_{jk}$ is preserved.

Let us show that $(A_i,C_{jk})=1$ $(i\le j\le k)$ is preserved. We have

$$\begin{aligned}&\big(\Phi_{\lambda,i}^{-1}\tilde{A}^{1\dots i-1,i\dots n}\Phi_{\lambda,i},\Phi_{\lambda,j}^{-1}\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{j\dots k-1,k\dots n}\\ &\left(\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\right)^{-1}\Phi_{\lambda,j}\big)\\ &\ =\Phi_{\lambda,i}^{-1}\big(\tilde{A}^{1\dots i-1,i\dots n},\Phi_\lambda^{1\dots,i,\dots n}\cdots\Phi_\lambda^{1\dots,j-1,\dots n}\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\\ &\quad(e^{2\pi\mathrm{i}\bar{t}_{12}})^{j\dots k-1,k\dots n}\left(\Phi_\lambda^{1\dots,i,\dots n}\cdots\Phi_\lambda^{1\dots,j-1,\dots n}\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\right)^{-1}\big)\Phi_{\lambda,i}\\ &\ =\Phi_{\lambda,i}^{-1}\big(\tilde{A}^{1\dots i-1,i\dots n},\Phi_\lambda^{i,i+1,\dots n}\cdots\Phi_\lambda^{i\dots,j-1,\dots n}\Phi_\lambda^{1\dots,i\dots j-1,\dots n}\Phi_\lambda^{1\dots,i,\dots j-1}\cdots\\ &\quad\Phi_\lambda^{1\dots,j-2,j-1}\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{j\dots k-1,k\dots n}\\ &\quad\Big(\Phi_\lambda^{i,i+1,\dots n}\cdots\Phi_\lambda^{i\dots,j-1,\dots n}\Phi_\lambda^{1\dots,i\dots j-1,\dots n}\Phi_\lambda^{1\dots,i,\dots j-1}\cdots\\ &\quad\Phi_\lambda^{1\dots,j-2,j-1}\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\Big)^{-1}\big)\Phi_{\lambda,i}\\ &\ =\Phi_{\lambda,i}^{-1}\big(\tilde{A}^{1\dots i-1,i\dots n},\Phi_\lambda^{i,i+1,\dots n}\cdots\Phi_\lambda^{i\dots,j-1,\dots n}\Phi_\lambda^{1\dots,i\dots j-1,\dots n}\Phi_\lambda^{j,j+1,\dots n}\cdots\\ &\quad\Phi_\lambda^{j\dots,k-1,\dots n}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{j\dots k-1,k\dots n}\Big(\Phi_\lambda^{i,i+1,\dots n}\cdots\Phi_\lambda^{i\dots,j-1,\dots n}\Phi_\lambda^{1\dots,i\dots j-1,\dots n}\Phi_\lambda^{j,j+1,\dots n}\\ &\quad\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\Big)^{-1}\big)\Phi_{\lambda,i}\\ &\ =\Phi_{\lambda,i}^{-1}\Phi_\lambda^{i,i+1,\dots n}\cdots\Phi_\lambda^{i\dots,j-1,\dots n}\big(\tilde{A}^{1\dots i-1,i\dots n},\Phi_\lambda^{1\dots,i\dots j-1,\dots n}\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\\ &\quad(e^{2\pi\mathrm{i}\bar{t}_{12}})^{j\dots k-1,k\dots n}\left(\Phi_\lambda^{1\dots,i\dots j-1,\dots n}\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\right)^{-1}\big)\\ &\quad\left(\Phi_\lambda^{i,i+1,\dots n}\cdots\Phi_\lambda^{i\dots,j-1,\dots n}\right)^{-1}\Phi_{\lambda,i}\\ &\ =\Phi_{\lambda,i}^{-1}\Phi_\lambda^{i,i+1,\dots n}\cdots\Phi_\lambda^{i\dots,j-1,\dots n}\Big(\tilde{A}^{1\dots i-1,i\dots n},\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\Phi_\lambda^{1\dots,i\dots j-1,\dots n}\\ &\quad(e^{2\pi\mathrm{i}\bar{t}_{12}})^{j\dots k-1,k\dots n}\left(\Phi_\lambda^{j,j+1,\dots n}\cdots\Phi_\lambda^{j\dots,k-1,\dots n}\Phi_\lambda^{1\dots,i\dots j-1,\dots n}\right)^{-1}\Big)\\ &\quad\left(\Phi_\lambda^{i,i+1,\dots n}\cdots\Phi_\lambda^{i\dots,j-1,\dots n}\right)^{-1}\Phi_{\lambda,i}\\ &\ =1,\end{aligned}$$

where the second equality follows from the generalized pentagon identity (39), the third equality follows from the fact that $\Phi_\lambda^{1\dots,i,\dots j-1}$, $\cdots$, $\Phi_\lambda^{1\dots,j-2,j-1}$ commute with $(e^{2\pi\mathrm{i}\bar{t}_{12}})^{j\dots k-1,k\dots n}$, $\Phi_\lambda^{j,j+1,\dots n}$, $\cdots$, $\Phi_\lambda^{j\dots,k-1,\dots n}$, the fourth equality follows from the fact that $\Phi_\lambda^{i,i+1,\dots n}$, $\cdots$, $\Phi_\lambda^{i\dots,j-1,\dots n}$ commute with $\tilde{A}^{1\dots i-1,i\dots n}$ (since Φ_λ is invariant), the last equality follows from the fact that $\Phi_\lambda^{1\dots,i\dots j-1,j\dots n}$ commutes with $\Phi_\lambda^{j,j+1,\dots n}$, $\dots$, $\Phi_\lambda^{j\dots,k-1,\dots n}$ (again as Φ_λ is invariant) and with $(e^{2\pi\mathrm{i}\bar{t}_{12}})^{j\dots k-1,k\dots n}$ (since t_{34} commutes with the image of

$\mathfrak{t}_3 \to \mathfrak{t}_4$, $x \mapsto x^{1,2,34}$). Therefore $(A_i, C_{jk}) = 1$ is preserved. One shows similarly that $(B_i, C_{jk}) = 1$ $(i \le j \le k)$, $X_{i+1} = \sigma_i X_i \sigma_i$ and $Y_{i+1} = \sigma_i^{-1} Y_i \sigma_i^{-1}$ are preserved.

The fact that the relations $\Theta A_i \Theta^{-1} = B_i^{-1}$, $\Theta B_i \Theta^{-1} = B_i A_i B_i^{-1}$, $\Psi A_i \Psi^{-1} = A_i$, $\Psi B_i \Psi^{-1} = B_i A_i$, are preserved follows from the identities (28), (29) and that if we denote by $x \mapsto [x]_n$ the morphism $\mathfrak{d} \to \mathfrak{d} \rtimes \bar{\mathfrak{t}}_{1,n}$ defined above, then (a) Φ_i commutes with $\sum_{i,j|i<j} \bar{t}_{ij}$ and with the image of $\mathfrak{d} \to \mathfrak{d} \rtimes \bar{\mathfrak{t}}_{1,n}$, $x \mapsto [x]_n$; (b) for $x \in \mathfrak{d}$, $y \in \bar{\mathfrak{t}}_{1,2}$, we have $[[x]_n, y^{1\ldots i-1,i\ldots n}] = [[x]_2, y]^{1\ldots i-1,i\ldots n}$. Let us prove (a): the first part follows from the fact that Φ commutes with $t_{12}+t_{13}+t_{23}$; the second part follows from the fact that X, d, Δ_0 and $\delta_{2n} + \sum_{k<l} (\mathrm{ad}\, \bar{x}_k)^{2n}(\bar{t}_{kl})$ commute with $\bar{t}_{ij}$ for any $i<j$. Let us prove (b): the identity holds for $[x,x']$ whenever it holds for x and for x', so it suffices to check it for x a generator of $\mathfrak{d}$; x being such a generator, both sides are (as functions of y) derivations $\bar{\mathfrak{t}}_{1,2} \to \bar{\mathfrak{t}}_{1,n}$ w.r.t. the morphism $\bar{\mathfrak{t}}_{1,2} \to \bar{\mathfrak{t}}_{1,n}$, $y \mapsto y^{1\ldots i-1,i\ldots n}$, so it suffices to check the identity for y a generator of $\bar{\mathfrak{t}}_{1,2}$. The identity is obvious if $x \in \{\Delta_0, d, X\}$ and $y \in \{\bar{x}_1, \bar{y}_1, \bar{x}_2, \bar{y}_2\}$. If $x = \delta_{2s}$ and $y = \bar{x}_1$, then the identity holds because we have

$$[\delta_{2s} + (\mathrm{ad}\, \bar{x}_1)^{2s}(\bar{t}_{12}), \bar{x}_1]^{1\ldots i-1,i\ldots n} = -\big((\mathrm{ad}\, \bar{x}_1)^{2s+1}(\bar{t}_{12})\big)^{1\ldots i-1,i\ldots n}$$
$$= -\left(\mathrm{ad}\left(\sum_{u'=1}^{i-1} \bar{x}_{u'}\right)\right)^{2s+1} \left(\sum_{1\le u<i\le v\le n} \bar{t}_{uv}\right) = -\sum_{1\le u<i\le v\le n} (\mathrm{ad}\, \bar{x}_u)^{2s+1}(\bar{t}_{uv}),$$

while

$$\left[\delta_{2s} + \sum_{1\le u<v\le n} (\mathrm{ad}\, \bar{x}_u)^{2s}(\bar{t}_{uv}), \sum_{u'=1}^{i-1} \bar{x}_{u'}\right] = \left[\sum_{1\le u<i\le v\le n} (\mathrm{ad}\, \bar{x}_u)^{2s}(\bar{t}_{uv}), \sum_{u'=1}^{i-1} \bar{x}_{u'}\right]$$
$$= -\sum_{1\le u<i\le v\le n} (\mathrm{ad}\, \bar{x}_u)^{2s+1}(\bar{t}_{uv}),$$

where the first equality follows from the fact that $(\mathrm{ad}\, \bar{x}_u)^{2s}(\bar{t}_{uv})$ commutes with $\sum_{u'=1}^{i-1} \bar{x}_{u'}$ whenever $u<v<i$ or $i \le u < v$. If $x = \delta_{2s}$ and $y = \bar{x}_2$, then the identity follows because $[\delta_{2s} + (\mathrm{ad}\, \bar{x}_1)^{2s}(\bar{t}_{12}), \bar{x}_1 + \bar{x}_2] = 0$ and $\left[\delta_{2s} + \sum_{1\le u<v\le n} (\mathrm{ad}\, \bar{x}_u)^{2s}(\bar{t}_{uv}), \sum_{u'=1}^{n} \bar{x}_{u'}\right] = 0$.

If $x = \delta_{2s}$ and $y = \bar{y}_1$, then

$$[\delta_{2s} + (\mathrm{ad}\, \bar{x}_1)^{2s}(\bar{t}_{12}), \bar{y}_1]^{1\ldots i-1,i\ldots n}$$
$$= \left\{\frac{1}{2} \sum_{p+q=2s-1} [(\mathrm{ad}\, \bar{x}_1)^p(\bar{t}_{12}), (-\mathrm{ad}\, \bar{x}_1)^q(\bar{t}_{12})] + [(\mathrm{ad}\, \bar{x}_1)^{2s}(\bar{t}_{12}), \bar{y}_1]\right\}^{1\ldots i-1,i\ldots n}$$

$$= \frac{1}{2} \sum_{p+q=2s-1} \left[\sum_{1 \le u < i \le v \le n} (\mathrm{ad}\, \bar{x}_u)^p (\bar{t}_{uv}), \sum_{1 \le u' < i \le v' \le n} (\mathrm{ad}\, \bar{x}_{u'})^q (\bar{t}_{u'v'}) \right]$$
$$+ \left[\sum_{1 \le u < i \le v \le n} (\mathrm{ad}\, \bar{x}_u)^{2s} (\bar{t}_{uv}), \bar{y}_1 + \cdots + \bar{y}_{i-1} \right];$$

on the other hand,

$$\left[\delta_{2s} + \sum_{1 \le u < v \le n} (\mathrm{ad}\, \bar{x}_u)^{2s} (\bar{t}_{uv}), \bar{y}_1 + \cdots + \bar{y}_{i-1} \right]$$
$$= \sum_{1 \le u < v \le n} [(\mathrm{ad}\, \bar{x}_u)^{2s} (\bar{t}_{uv}), \bar{y}_1 + \cdots + \bar{y}_{i-1}]$$
$$+ \sum_{u=1}^{i-1} \sum_{v | v \ne u} \sum_{p+q=2s-1} \frac{1}{2} [(\mathrm{ad}\, \bar{x}_u)^p (\bar{t}_{uv}), (-\mathrm{ad}\, \bar{x}_u)^q (\bar{t}_{uv})]$$
$$= \sum_{1 \le u < v \le n} [(\mathrm{ad}\, \bar{x}_u)^{2s} (\bar{t}_{uv}), \bar{y}_1 + \cdots + \bar{y}_{i-1}]$$
$$+ \sum_{1 \le u < i \le v \le n} \sum_{p+q=2s-1} \frac{1}{2} [(\mathrm{ad}\, \bar{x}_u)^p (\bar{t}_{uv}), (-\mathrm{ad}\, \bar{x}_u)^q (\bar{t}_{uv})],$$

where the second equality follows from the fact that

$$[(\mathrm{ad}\, \bar{x}_u)^p (\bar{t}_{uv}), (-\bar{x}_u)^q (\bar{t}_{uv})] + [(\mathrm{ad}\, \bar{x}_v)^p (\bar{t}_{uv}), (-\mathrm{ad}\, \bar{x}_v)^q (\bar{t}_{uv})] = 0$$

as $p+q$ is odd.

Then

$$[\delta_{2s} + (\mathrm{ad}\, \bar{x}_1)^{2s} (\bar{t}_{12}), \bar{y}_1]^{1 \ldots i-1, i \ldots n} - \left[\delta_{2s} + \sum_{1 \le u < v \le n} (\mathrm{ad}\, \bar{x}_u)^{2s} (\bar{t}_{uv}), \bar{y}_1 + \cdots + \bar{y}_{i-1} \right]$$
$$= - \sum_{1 \le u < v < i} [(\mathrm{ad}\, \bar{x}_u)^{2s} (\bar{t}_{uv}), \bar{y}_1 + \cdots + \bar{y}_{i-1}]$$
$$- \sum_{i \le u < v \le n} [(\mathrm{ad}\, \bar{x}_u)^{2s} (\bar{t}_{uv}), \bar{y}_1 + \cdots + \bar{y}_{i-1}]$$
$$+ \frac{1}{2} \sum_{p+q=2s-1} \sum_{\substack{1 \le u < i \le v \le n \\ 1 \le u' < i \le v' \le n, (u,v) \ne (u',v')}} [(\mathrm{ad}\, \bar{x}_u)^p (\bar{t}_{uv}), (-\mathrm{ad}\, \bar{x}_{u'})^q (\bar{t}_{u'v'})]$$
$$= \sum_{1 \le u < v < i} [(\mathrm{ad}\, \bar{x}_u)^{2s} (\bar{t}_{uv}), \bar{y}_i + \cdots + \bar{y}_n]$$
$$- \sum_{i \le u < v \le n} [(\mathrm{ad}\, \bar{x}_u)^{2s} (\bar{t}_{uv}), \bar{y}_1 + \cdots + \bar{y}_{i-1}]$$
$$+ \frac{1}{2} \sum_{p+q=2s-1} \sum_{\substack{1 \le u < i \le v \le n \\ 1 \le u < i \le v' \le n, v \ne v'}} [(\mathrm{ad}\, \bar{x}_u)^p (\bar{t}_{uv}), (-\mathrm{ad}\, \bar{x}_u)^q (\bar{t}_{uv'})]$$

$$+\frac{1}{2}\sum_{p+q=2s-1}\sum_{\substack{1\le u<i\le v\le n\\ 1\le u'<i\le v\le n,u\neq u'}}[(\mathrm{ad}\,\bar{x}_u)^p(\bar{t}_{uv}),(-\mathrm{ad}\,\bar{x}_{u'})^q(\bar{t}_{u'v})],$$

where the second equality follows from the centrality of $\bar{y}_1+\cdots+\bar{y}_n$, the last equality follows for the fact that $(\mathrm{ad}\,\bar{x}_u)^p(\bar{t}_{uv})$ and $(-\mathrm{ad}\,\bar{x}_{u'})^q(\bar{t}_{u'v'})$ commute for u,v,u',v' all distinct. Since $p+q$ is odd, it follows that

$$\begin{aligned}
&[\delta_{2s}+(\mathrm{ad}\,\bar{x}_1)^{2s}(\bar{t}_{12}),\bar{y}_1]^{1\ldots i-1,i\ldots n}-\left[\delta_{2s}+\sum_{1\le u<v\le n}(\mathrm{ad}\,\bar{x}_u)^{2s}(\bar{t}_{uv}),\bar{y}_1+\cdots+\bar{y}_{i-1}\right]\\
&\quad=\sum_{1\le u<v<i}[(\mathrm{ad}\,\bar{x}_u)^{2s}(\bar{t}_{uv}),\bar{y}_i+\cdots+\bar{y}_n]\\
&\quad\quad-\sum_{i\le u<v\le n}[(\mathrm{ad}\,\bar{x}_u)^{2s}(\bar{t}_{uv}),\bar{y}_1+\cdots+\bar{y}_{i-1}]\\
&\quad\quad+\sum_{p+q=2s-1}\sum_{1\le u<i\le v<v'\le n}[(\mathrm{ad}\,\bar{x}_u)^p(\bar{t}_{uv}),(-\mathrm{ad}\,\bar{x}_u)^q(\bar{t}_{uv'})]\\
&\quad\quad+\sum_{p+q=2s-1}\sum_{1\le u<u'<i\le v\le n}[(\mathrm{ad}\,\bar{x}_u)^p(\bar{t}_{uv}),(-\mathrm{ad}\,\bar{x}_{u'})^q(\bar{t}_{u'v})].
\end{aligned}$$

Now if $1\le u<v<i$, we have

$$\begin{aligned}
&[(\mathrm{ad}\,\bar{x}_u)^{2s}(\bar{t}_{uv}),\bar{y}_i+\cdots+\bar{y}_n]\\
&\quad=\sum_{p+q=2s-1}(\mathrm{ad}\,\bar{x}_u)^p\mathrm{ad}(\bar{t}_{ui}+\cdots+\bar{t}_{un})(\mathrm{ad}\,\bar{x}_u)^q(\bar{t}_{uv})\\
&\quad=\sum_{w=i}^{n}\sum_{p+q=2s-1}(\mathrm{ad}\,\bar{x}_u)^p[\bar{t}_{uw},(-\mathrm{ad}\,\bar{x}_v)^q(\bar{t}_{uv})]\\
&\quad=\sum_{w=i}^{n}\sum_{p+q=2s-1}(\mathrm{ad}\,\bar{x}_u)^p(-\mathrm{ad}\,\bar{x}_v)^q([\bar{t}_{uw},\bar{t}_{uv}])\\
&\quad=-\sum_{w=i}^{n}\sum_{p+q=2s-1}(\mathrm{ad}\,\bar{x}_u)^p(-\mathrm{ad}\,\bar{x}_v)^q([\bar{t}_{uw},\bar{t}_{vw}])\\
&\quad=-\sum_{w=i}^{n}\sum_{p+q=2s-1}[(\mathrm{ad}\,\bar{x}_u)^p(\bar{t}_{uw}),(-\mathrm{ad}\,\bar{x}_v)^q(\bar{t}_{vw})];
\end{aligned}$$

one shows in the same way that if $i\le u<v\le n$, then

$$[(\mathrm{ad}\,\bar{x}_u)^{2s}(\bar{t}_{uv}),\bar{y}_1+\cdots+\bar{y}_{i-1}]=\Sigma_{w=1}^{i-1}\Sigma_{p+q=2s-1}[(\mathrm{ad}\,\bar{x}_u)^p(\bar{t}_{uw}),(-\mathrm{ad}\,\bar{x}_v)^q(\bar{t}_{vw})];$$

all this implies that

$$[\delta_{2s}+(\mathrm{ad}\,\bar{x}_1)^{2s}(\bar{t}_{12}),\bar{y}_1]^{1\dots i-1,i\dots n}-\left[\delta_{2s}+\sum_{1\le u<v\le n}(\mathrm{ad}\,\bar{x}_u)^{2s}(\bar{t}_{uv}),(\bar{y}_1)^{1\dots i-1}\right].$$

Since $[\delta_{2s}+(\mathrm{ad}\,\bar{x}_1)^{2s}(\bar{t}_{12}),\bar{y}_1+\bar{y}_2]=0$ and

$$\left[\delta_{2s}+\sum_{1\le u<v\le n}(\mathrm{ad}\,\bar{x}_u)^{2s}(\bar{t}_{uv}),\bar{y}_1+\dots+\bar{y}_n\right]=0,$$

this equality implies

$$[\delta_{2s}+(\mathrm{ad}\,\bar{x}_1)^{2s}(\bar{t}_{12}),\bar{y}_2]^{1\dots i-1,i\dots n}-\left[\delta_{2s}+\sum_{1\le u<v\le n}(\mathrm{ad}\,\bar{x}_u)^{2s}(\bar{t}_{uv}),(\bar{y}_2)^{1\dots i-1}\right],$$

which ends the proof of (b) above, and therefore of the fact that the identities $\Theta A_i\Theta^{-1}=B_i^{-1}$, $\dots$, $\Psi B_i\Psi^{-1}=B_iA_i$ are preserved.

The relation $(\Theta,\Psi^2)=1$ is preserved because

$$\begin{aligned}\left([\tilde{\Theta}]e^{\mathrm{i}\frac{\pi}{2}\sum_{i<j}\bar{t}_{ij}},\left([\tilde{\Psi}]e^{\mathrm{i}\frac{\pi}{6}\sum_{i<j}\bar{t}_{ij}}\right)^2\right)&=\left([\tilde{\Theta}]e^{\mathrm{i}\frac{\pi}{2}\sum_{i<j}\bar{t}_{ij}},[\tilde{\Psi}]^2e^{\mathrm{i}\frac{\pi}{3}\sum_{i<j}\bar{t}_{ij}}\right)\\&=([\tilde{\Theta}],[\tilde{\Psi}]^2)=[(\tilde{\Theta},\tilde{\Psi}^2)]=1,\end{aligned}$$

where the first two identities follow from the fact that $\sum_{i<j}\bar{t}_{ij}$ commutes with the image of $\mathfrak{d}\to\mathfrak{d}\ltimes\bar{\mathfrak{t}}_{1,n}$, $x\mapsto[x]$, the third identity follows from the fact that $\mathbf{G}_1\to\mathbf{G}_n$, $g\mapsto[g]$ is a group morphism, and the last identity follows from (27).

The image of $C_{i,i+1}$ is $\Phi_{\lambda,i}^{-1}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{i,i+1\dots n}\Phi_{\lambda,i}$, to the product of the images of $C_{12},\dots,C_{n-1,n}$ is

$$\begin{aligned}&\Phi_{\lambda,1}^{-1}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{1,2\dots n}\left(\Phi_{\lambda,1}\Phi_{\lambda,2}^{-1}\right)(e^{2\pi\mathrm{i}\bar{t}_{12}})^{2,3\dots n}\left(\Phi_{\lambda,2}\Phi_{\lambda,3}^{-1}\right)\\&\times(e^{2\pi\mathrm{i}\bar{t}_{12}})^{3,4\dots n}\cdots\left(\Phi_{\lambda,n-1}\Phi_{\lambda,n}^{-1}\right)e^{2\pi\mathrm{i}\bar{t}_{n-1,n}}\Phi_{\lambda,n}\\&\quad=\Phi_{\lambda,1}^{-1}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{1,2\dots n}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{2,3\dots n}\Phi_\lambda^{1,2,3\dots n}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{3,4\dots n}\cdots\Phi_\lambda^{1\dots,i-1,\dots n}\\&\qquad(e^{2\pi\mathrm{i}\bar{t}_{12}})^{i,i+1\dots n}\cdots\Phi_\lambda^{1\dots,n-2,n-1\ n}e^{2\pi\mathrm{i}\bar{t}_{n-1,n}}\\&\quad=\Phi_{\lambda,1}^{-1}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{1,2\dots n}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{2,3\dots n}(e^{2\pi\mathrm{i}\bar{t}_{12}})^{3,4\dots n}\cdots(e^{2\pi\mathrm{i}\bar{t}_{12}})^{i,i+1\dots n}\cdots\\&\qquad e^{2\pi\mathrm{i}\bar{t}_{n-1,n}}\Phi_\lambda^{1,2,3\dots n}\cdots\Phi_\lambda^{1\dots,i-1,\dots n}\cdots\Phi_\lambda^{1\dots,n-2,n-1\ n}\\&\quad=\Phi_{\lambda,1}^{-1}e^{2\pi\mathrm{i}\sum_{i<j}\bar{t}_{ij}}\Phi_{\lambda,1}=e^{2\pi\mathrm{i}\sum_{i<j}\bar{t}_{ij}},\end{aligned}$$

where the second equality follows from the fact that $\Phi^{1\dots,i,\dots n}$ commutes with $(e^{2\pi\mathrm{i}\bar{t}_{12}})^{j,j+1\dots n}$ whenever $j>i$, and the last equality follows from the fact that $\sum_{i<j}t_{ij}$ is central is $\mathfrak{t}_n$.

So the product of the images of $C_{12}\cdots C_{n-1,n}$ is $e^{2\pi\mathrm{i}\sum_{i<j}\bar t_{ij}}$.

The relation $(\Theta\Psi)^3 = C_{12}\cdots C_{n-1,n}$ is then preserved because

$$\left([\tilde\Theta]e^{\mathrm{i}\frac{\pi}{2}\sum_{i<j}\bar t_{ij}}[\tilde\Psi]e^{\mathrm{i}\frac{\pi}{6}\sum_{i<j}\bar t_{ij}}\right)^3 = ([\tilde\Theta][\tilde\Psi])^3 e^{2\pi\mathrm{i}\sum_{i<j}\bar t_{ij}} = [(\tilde\Theta\tilde\Psi)^3]e^{2\pi\mathrm{i}\sum_{i<j}\bar t_{ij}}$$
$$= e^{2\pi\mathrm{i}\sum_{i<j}\bar t_{ij}},$$

where the first equality follows from the fact that $\sum_{i<j}\bar t_{ij}$ commutes with the image of $\mathbf{G}_1\to\mathbf{G}_n$, $g\mapsto[g]$, and the second equality follows from the fact that $g\mapsto[g]$ is a group morphism and the last equality follows from (27). In the same way, one proves that $\Theta^4 = C_{12}\cdots C_{n-1,n}$, $\sigma_i^2 = C_{i,i+1}C_{i+1,i+2}C_{i,i+1}^{-1}$, and $(\Theta,\sigma_i) = (\Psi,\sigma_i) = 1$ are preserved. □

5.2 Construction of morphisms $\overline{\mathbf{B}}_{1,n}\to\exp\left(\widehat{\bar{\mathfrak{t}}^{\mathbf{k}}_{1,n}}\right)\rtimes S_n$ using an associator Φ_λ

Let us keep the notation of the previous section. Set

$$\begin{aligned}
a_{2n}(\lambda) &:= -(2n+1)B_{2n+2}\lambda^{2n+2}/(2n+2)!,\quad \tilde y_\lambda := -\frac{\mathrm{ad}\,x}{e^{\lambda\mathrm{ad}\,x}-1}(y),\\
\tilde A_\lambda &:= \Phi_\lambda(\tilde y_\lambda,t)e^{\lambda\tilde y_\lambda}\Phi_\lambda(\tilde y_\lambda,t)^{-1}\\
&= e^{-\lambda t/2}\Phi_\lambda(-\tilde y_\lambda-t,t)e^{\lambda(\tilde y_\lambda+t)}\Phi_\lambda(-\tilde y_\lambda-t,t)^{-1}e^{-\lambda t/2},\\
\tilde B_\lambda &:= e^{\lambda t/2}\Phi_\lambda(-\tilde y_\lambda-t,t)e^{\lambda x}\Phi_\lambda(\tilde y_\lambda,t)^{-1}
\end{aligned}$$

(the identity in the definition of A_λ follows from the hexagon relation).

Proposition 33. *We have*

$$\begin{aligned}
&\tilde A_\lambda^{12,3} = e^{\lambda\bar t_{12}/2}\{\Phi_\lambda\}^{3,1,2}\tilde A_\lambda^{2,13}\{\Phi_\lambda\}^{2,1,3}e^{\lambda\bar t_{12}/2}\cdot\{\Phi_\lambda\}^{3,2,1}\tilde A_\lambda^{1,23}\{\Phi_\lambda\}^{1,2,3},\\
&\tilde B_\lambda^{12,3} = e^{-\lambda\bar t_{12}/2}\{\Phi_\lambda\}^{3,1,2}\tilde B_\lambda^{2,13}\{\Phi_\lambda\}^{2,1,3}e^{-\lambda\bar t_{12}/2}\cdot\{\Phi_\lambda\}^{3,2,1}\tilde B_\lambda^{1,23}\{\Phi_\lambda\}^{1,2,3},\\
&\left(\tilde B_\lambda^{12,3}, e^{\lambda\bar t_{12}/2}\{\Phi_\lambda\}^{3,1,2}\tilde A_\lambda^{2,13}\{\Phi_\lambda\}^{2,1,3}e^{\lambda\bar t_{12}/2}\right)\\
&\qquad = \left(e^{-\lambda\bar t_{12}/2}\{\Phi_\lambda\}^{3,1,2}\tilde B_\lambda^{2,13}\{\Phi_\lambda\}^{2,1,3}e^{-\lambda\bar t_{12}/2},\tilde A_\lambda^{12,3}\right)\\
&\qquad = \{\Phi_\lambda\}^{3,2,1}e^{\lambda\bar t_{23}}\{\Phi_\lambda\}^{1,2,3},
\end{aligned}$$

so the formulas of Proposition 31 (restricted to the generators $A_i, B_i, \sigma_i, C_{jk}$) induce a morphism $\overline{\mathrm{B}}_{1,n}\to\exp\left(\widehat{\bar{\mathfrak{t}}^{\mathbf{k}}_{1,n}}\right)\rtimes S_n$ (here $\widehat{\bar{\mathfrak{t}}^{\mathbf{k}}_{1,n}}$ is the degree completion of $\bar{\mathfrak{t}}^{\mathbf{k}}_{1,n}$).

Proof. In this proof, we shift the indices of the generators of $\mathfrak{t}_{n+1}$ by 1, so these generators are now t_{ij}, $i\neq j\in\{0,\dots,n\}$ (recall that $\mathfrak{t}_{n+1}=\mathfrak{t}^{\mathbf{k}}_{n+1}$, $\bar t_{1,n}=\bar{\mathfrak{t}}^{\mathbf{k}}_{1,n}$).

We have a morphism $\alpha_n : \mathfrak{t}_{n+1} \to \bar{\mathfrak{t}}_{1,n}$, defined by $t_{ij} \mapsto \bar{t}_{ij}$ if $1 \le i < j \le n$ and $t_{0i} \mapsto \tilde{y}_i := -\frac{\mathrm{ad}\,\bar{x}_i}{e^{\lambda \mathrm{ad}\,\bar{x}_i}-1}(\bar{y}_i)$ if $1 \le i \le n$ (it takes the central element $\sum_{0\le i<j\le n} t_{ij}$ to 0).

Let $\phi : \{1,\ldots,m\} \to \{1,\ldots,n\}$ be a map and let $\phi' : \{0,\ldots,m\} \to \{0,\ldots,n\}$ be given by $\phi'(1) = 1$, $\phi'(i) = \phi(i)$ for $i = 1,\ldots,m$. The diagram

$$\begin{array}{ccc} \mathfrak{t}_{n+1} & \stackrel{x\mapsto x^{\phi'}}{\to} & \mathfrak{t}_{m+1} \\ \alpha_n \downarrow & & \downarrow \alpha_m \\ \bar{\mathfrak{t}}_{1,n} & \stackrel{x\mapsto x^{\phi}}{\to} & \bar{\mathfrak{t}}_{1,m} \end{array}$$

is not commutative, we have instead the identity

$$\alpha_m(x^{\phi'}) = \alpha_n(x)^{\phi} - \sum_{i=1}^{n} \xi_i(x) \left(\sum_{i',j' \in \phi^{-1}(i) | i'<j'} \bar{t}_{i'j'} \right),$$

where $\xi_i : \bar{\mathfrak{t}}_{1,n} \to \mathbf{k}$ is the linear form defined by $\xi_i(t_{0i}) = 1$, ξ_i (any other homogeneous Lie polynomial in the $t_{kl}) = 0$.

Since the various $\sum_{i',j'\in\phi^{-1}(i)|i'<j'} \bar{t}_{i'j'}$ commute with each other and with the image of $x \mapsto x^{\phi}$, this implies

$$\alpha_m(g^{\phi'}) = \alpha_n(g)^{\phi} \prod_{i=1}^{n} e^{-\xi_i(\log g)(\sum_{i',j'\in\phi^{-1}(i), i'<j'} \bar{t}_{i'j'})}$$

for $g \in \exp(\hat{\mathfrak{t}}_{n+1})$.

Set $\bar{A}_\lambda := \Phi_\lambda^{0,1,2} e^{\lambda t_{01}} \left(\Phi_\lambda^{0,1,2}\right)^{-1} \in \exp(\hat{\mathfrak{t}}_3)$. One proves that

$$\bar{A}_\lambda^{0,12,3} e^{\lambda t_{12}} = e^{\lambda t_{12}/2} \Phi_\lambda^{3,1,2} \bar{A}_\lambda^{0,2,13} \Phi_\lambda^{2,1,3} e^{\lambda t_{12}/2} \cdot \Phi_\lambda^{3,2,1} \bar{A}_\lambda^{0,1,23} \Phi_\lambda^{1,2,3}$$

(relation in $\exp(\hat{\mathfrak{t}}_4)$). We then have $\alpha_2(\bar{A}_\lambda) = \tilde{A}_\lambda$, $\alpha_3\left(\Phi_\lambda^{1,2,3}\right) = \Phi_\lambda^{1,2,3}$, and the relation between the α_i and coproducts implies $\alpha_3\left(\bar{A}_\lambda^{0,1,23}\right) = \tilde{A}_\lambda^{1,23}$ and $\alpha_3\left(\bar{A}_\lambda^{0,12,3} e^{\lambda t_{12}}\right) = \tilde{A}_\lambda^{12,3}$. Taking the image by α_3, we get the first identity. As we have already mentioned, this identity implies $\left(\Phi_\lambda^{-1} \tilde{A}_\lambda^{1,23} \Phi_\lambda, \tilde{A}_\lambda^{12,3}\right) = 1$.

Let $\exp(\hat{\mathfrak{t}}_{n+1}) * \mathbb{Z}^n / I_n$ be the quotient of the free product of $\exp(\hat{\mathfrak{t}}_{n+1})$ with $\mathbb{Z}^n = \oplus_{i=1}^n \mathbb{Z} X_i$ by the normal subgroup generated by the ratios of the exponentials of the sides of each of the equations

$$X_i t_{0i} X_i^{-1} = \sum_{0\le\alpha\le n, \alpha\ne i} t_{\alpha i}, \quad X_i(t_{0j} + t_{ij}) X_i^{-1} = t_{0j},$$
$$X_i t_{jk} X_i^{-1} = t_{jk}, \; X_j X_k t_{jk} (X_j X_k)^{-1} = t_{jk}$$

where i, j, k are distinct in $\{1,\ldots,n\}$. Then the morphism $\alpha_n : \mathfrak{t}_{n+1} \to \bar{\mathfrak{t}}_{1,n}$ extends to $\tilde{\alpha}_n : \exp(\hat{\mathfrak{t}}_{n+1}) * \mathbb{Z}^n / I_n \to \exp(\hat{\bar{\mathfrak{t}}}_{1,n})$ by $X_i \mapsto e^{\lambda x_i}$.

If $\phi : \{1,\dots,m\} \to \{1,\dots,n\}$ is a map, then the Lie algebra morphism $\mathfrak{t}_{n+1} \to \mathfrak{t}_{m+1}$, $x \mapsto x^{\phi'}$ extends to a group morphism $\exp(\hat{\mathfrak{t}}_n) * \mathbb{Z}^n/I_n \to \exp(\hat{\mathfrak{t}}_m) * \mathbb{Z}^m/I_m$ by $X_i \mapsto \prod_{i'\in\phi^{-1}(i)} X_{i'}$.

Let

$$\bar{B}_\lambda := e^{\lambda t_{12}/2}\Phi_\lambda^{0,2,1} X_1 \Phi_\lambda^{2,1,0} \in \exp(\hat{\mathfrak{t}}_3) * \mathbb{Z}^2/I_2,$$

Then $\alpha_2(\bar{B}_\lambda) = \tilde{B}_\lambda$.

We will prove that

$$\bar{B}_\lambda^{0,12,3} = e^{-\lambda t_{12}/2}\Phi_\lambda^{3,1,2}\bar{B}_\lambda^{0,2,13}\Phi_\lambda^{2,1,3}e^{-\lambda t_{12}/2}\cdot\Phi^{3,2,1}\bar{B}_\lambda^{0,1,23}\Phi_\lambda^{1,2,3}. \tag{40}$$

The l.h.s. is

$$\bar{B}_\lambda^{0,12,3} = e^{\lambda t_{3,12}/2}\Phi_\lambda^{0,3,12}X_1X_2\Phi_\lambda^{3,21,0}$$

and the r.h.s. is

$$e^{-\lambda t_{12}/2}\Phi_\lambda^{3,1,2}e^{\lambda t_{31,2}/2}\Phi_\lambda^{0,13,2}X_2\Phi_\lambda^{13,2,0}\Phi_\lambda^{2,1,3}e^{-\lambda t_{12}/2}\Phi_\lambda^{3,2,1}e^{\lambda t_{23,1}/2}$$
$$\Phi_\lambda^{0,23,1}X_1\Phi_\lambda^{32,1,0}\Phi_\lambda^{1,2,3}.$$

The equality between these terms is rewritten as

$$X_1X_2 = \Phi_\lambda^{03,1,2}\Phi_\lambda^{1,3,0}e^{-\lambda t_{13}/2}X_2\Phi_\lambda^{13,2,0}e^{\lambda t_{13}/2}\Phi_\lambda^{2,3,1}\Phi_\lambda^{0,23,1}X_1\Phi_\lambda^{01,2,3}\Phi_\lambda^{2,1,0},$$

or, using the fact that X_i commutes with t_{jk} (i,j,k distinct), as

$$X_1X_2 = \Phi_\lambda^{03,1,2}\Phi_\lambda^{1,3,0}X_2\Phi_\lambda^{02,3,1}\Phi_\lambda^{3,2,0}X_1\Phi_\lambda^{01,2,3}\Phi_\lambda^{2,1,0}.$$

Now

$$\begin{aligned} X_2\Phi_\lambda^{02,3,1} &= \Phi_\lambda^{0,3,1}X_2, \\ X_1\Phi_\lambda^{01,2,3} &= \Phi_\lambda^{0,2,3}X_1, \text{ and} \\ X_1X_2\Phi_\lambda^{2,1,0} &= \Phi_\lambda^{2,1,03}X_1X_2, \end{aligned}$$

so the r.h.s. is rewritten as

$$\Phi_\lambda^{03,1,2}\Phi_\lambda^{1,3,0}\Phi_\lambda^{0,3,1}X_2\Phi_\lambda^{3,2,0}\Phi_\lambda^{0,2,3}X_1\Phi_\lambda^{2,1,0} = X_1X_2.$$

This ends the proof of (40). Taking the image by α_4, we then get the second identity of the proposition.

Let us prove the next identity. We have

$$\begin{aligned} &\left(\bar{B}_\lambda^{0,12,3}, e^{\lambda\bar{t}_{12}/2}\Phi_\lambda^{3,1,2}\bar{A}_\lambda^{0,2,13}\Phi_\lambda^{2,1,3}e^{\lambda\bar{t}_{12}/2}\right) \\ &\quad = e^{\lambda t_{12,3}/2}\Phi_\lambda^{0,3,12}X_1X_2\Phi_\lambda^{3,12,0}e^{\lambda\bar{t}_{12}/2}\Phi_\lambda^{3,1,2}\Phi_\lambda^{0,2,13}e^{\lambda t_{0,2}}\Phi_\lambda^{13,2,0}\Phi_\lambda^{2,1,3} \\ &\quad\quad e^{\lambda\bar{t}_{12}/2}\Phi_\lambda^{0,12,3}(X_1X_2)^{-1}\Phi_\lambda^{12,3,0}e^{-\lambda t_{12,3}/2}e^{-\lambda\bar{t}_{12}/2}\Phi_\lambda^{3,1,2}\Phi_\lambda^{0,2,13}e^{-\lambda t_{0,2}} \\ &\quad\quad \times\Phi_\lambda^{13,2,0}\Phi_\lambda^{2,1,3}e^{-\lambda\bar{t}_{12}/2}. \end{aligned}$$

Now

$$\begin{aligned}
&X_1X_2\Phi_\lambda^{3,12,0}e^{\lambda\bar t_{12}/2}\Phi_\lambda^{3,1,2}\Phi_\lambda^{0,2,13}e^{\lambda t_{0,2}}\Phi_\lambda^{13,2,0}\Phi_\lambda^{2,1,3}e^{\lambda\bar t_{12}/2}\Phi_\lambda^{0,12,3}(X_1X_2)^{-1}\\
&\quad= e^{\lambda\bar t_{12}/2}X_1X_2\Phi_\lambda^{3,12,0}\Phi_\lambda^{3,1,2}\Phi_\lambda^{0,2,13}e^{\lambda t_{0,2}}\Phi_\lambda^{13,2,0}\Phi_\lambda^{2,1,3}\Phi_\lambda^{0,12,3}(X_1X_2)^{-1}e^{\lambda\bar t_{12}/2}\\
&\quad= e^{\lambda\bar t_{12}/2}X_1X_2\Phi_\lambda^{0,2,1}\Phi_\lambda^{3,1,02}e^{\lambda t_{0,2}}\Phi_\lambda^{02,1,3}\Phi_\lambda^{0,2,1}(X_1X_2)^{-1}e^{\lambda\bar t_{12}/2}\\
&\quad= e^{\lambda\bar t_{12}/2}X_1X_2\Phi_\lambda^{0,2,1}e^{\lambda t_{0,2}}\Phi_\lambda^{0,2,1}(X_1X_2)^{-1}e^{\lambda\bar t_{12}/2}\\
&\quad= e^{\lambda\bar t_{12}/2}\Phi_\lambda^{03,2,1}X_1X_2e^{\lambda t_{0,2}}(X_1X_2)^{-1}\Phi_\lambda^{03,2,1}e^{\lambda\bar t_{12}/2}\\
&\quad= e^{\lambda\bar t_{12}/2}\Phi_\lambda^{03,2,1}e^{\lambda t_{03,2}}\Phi_\lambda^{03,2,1}e^{\lambda\bar t_{12}/2}.
\end{aligned}$$

Plugging this in the above expression for

$$\left(\bar B_\lambda^{0,12,3}, e^{\lambda\bar t_{12}/2}\Phi_\lambda^{3,1,2}\bar A_\lambda^{0,2,13}\Phi_\lambda^{2,1,3}e^{\lambda t_{12}/2}\right),$$

one then obtains

$$\left(\bar B_\lambda^{0,12,3}, e^{\lambda\bar t_{12}/2}\Phi_\lambda^{3,1,2}\bar A_\lambda^{0,2,13}\Phi_\lambda^{2,1,3}e^{\lambda\bar t_{12}/2}\right) = \Phi_\lambda^{3,2,1}e^{\lambda t_{23}}\Phi_\lambda^{1,2,3}.$$

Taking the image by α_4, we then obtain

$$\left(\tilde B_\lambda^{12,3}, e^{\lambda\bar t_{12}/2}\Phi_\lambda^{3,1,2}\tilde A_\lambda^{2,13}\Phi_\lambda^{2,1,3}e^{\lambda\bar t_{12}/2}\right) = \Phi_\lambda^{3,2,1}e^{\lambda\bar t_{23}}\Phi_\lambda^{1,2,3}.$$

Let us prove this last identity. For this, we will show that

$$\left(e^{-\lambda t_{12}/2}\Phi_\lambda^{3,1,2}\bar B_\lambda^{0,2,13}\Phi_\lambda^{2,1,3}e^{-\lambda t_{12}/2}, \bar A_\lambda^{0,12,3}e^{\lambda t_{12}}\right) = \Phi_\lambda^{3,2,1}e^{\lambda t_{23}}\Phi_\lambda^{1,2,3}$$

and take the image by α_4.

We have

$$\begin{aligned}
&\left(e^{-\lambda t_{12}/2}\Phi_\lambda^{3,1,2}\bar B_\lambda^{0,2,13}\Phi_\lambda^{2,1,3}e^{-\lambda t_{12}/2}, \bar A_\lambda^{0,12,3}e^{\lambda t_{12}}\right)\\
&= e^{-\lambda t_{12}/2}\Phi_\lambda^{3,1,2}e^{\lambda t_{2,13}/2}\Phi_\lambda^{0,13,2}X_2\Phi_\lambda^{13,2,0}\Phi_\lambda^{2,1,3}e^{-\lambda t_{12}/2}\Phi_\lambda^{0,12,3}e^{\lambda t_{0,12}}\\
&\quad\times\Phi_\lambda^{3,12,0}e^{\lambda t_{12}}e^{\lambda t_{12}/2}\Phi_\lambda^{3,1,2}\Phi_\lambda^{0,2,13}X_2^{-1}\Phi_\lambda^{2,13,0}e^{-\lambda t_{2,13}/2}\Phi_\lambda^{2,1,3}e^{\lambda t_{12}/2}\\
&\quad\times\Phi_\lambda^{0,12,3}e^{-\lambda t_{0,12}}\Phi_\lambda^{3,12,0}e^{-\lambda t_{12}}\\
&= e^{-\lambda t_{12}/2}\Phi_\lambda^{3,1,2}e^{\lambda t_{2,13}/2}\Phi_\lambda^{0,13,2}X_2\Phi_\lambda^{13,2,0}\Phi_\lambda^{2,1,3}\Phi_\lambda^{0,12,3}e^{\lambda t_{0,12}+\lambda t_{12}}\Phi_\lambda^{3,12,0}\Phi_\lambda^{3,1,2}\\
&\quad\times\Phi_\lambda^{0,2,13}\Phi_\lambda^{2,13,0}e^{-\lambda t_{2,13}/2}X_2^{-1}\Phi_\lambda^{2,1,3}e^{-\lambda t_{12}/2}\Phi_\lambda^{0,12,3}e^{-\lambda t_{0,12}}\Phi_\lambda^{3,12,0}.
\end{aligned}$$

Now

$$\begin{aligned}
X_2\Phi_\lambda^{13,2,0}&\Phi_\lambda^{2,1,3}\Phi_\lambda^{0,12,3}e^{\lambda t_{0,12}+\lambda t_{12}}\Phi_\lambda^{3,12,0}\Phi_\lambda^{3,1,2}\Phi_\lambda^{0,2,13}X_2^{-1}\\
&= X_2\Phi_\lambda^{02,1,3}\Phi_\lambda^{1,2,0}e^{\lambda t_{0,12}+\lambda t_{12}}\Phi_\lambda^{0,2,1}\Phi_\lambda^{3,1,02}X_2^{-1}\\
&= \Phi_\lambda^{0,1,3}X_2\Phi_\lambda^{1,2,0}e^{\lambda t_{0,12}+\lambda t_{12}}\Phi_\lambda^{0,2,1}X_2^{-1}\Phi_\lambda^{3,1,0}\\
&= \Phi_\lambda^{0,1,3}X_2e^{\lambda(t_{01}+t_{02}+t_{12})}X_2^{-1}\Phi_\lambda^{3,1,0}\\
&= \Phi_\lambda^{0,1,3}e^{\lambda(t_{01}+t_{02}+t_{12}+t_{23})}\Phi_\lambda^{3,1,0}.
\end{aligned}$$

So

$$\left(e^{-\lambda t_{12}/2}\Phi_\lambda^{3,1,2}\bar{B}_\lambda^{0,2,13}\Phi_\lambda^{2,1,3}e^{-\lambda t_{12}/2}, \bar{A}_\lambda^{0,12,3}e^{\lambda t_{12}}\right)$$
$$= e^{-\lambda t_{12}/2}\Phi_\lambda^{3,1,2}e^{\lambda t_{2,13}/2}\Phi_\lambda^{0,13,2}\Phi_\lambda^{0,1,3}e^{\lambda(t_{01}+t_{02}+t_{12}+t_{23})}$$
$$\times \Phi_\lambda^{3,1,0}\Phi_\lambda^{2,13,0}e^{-\lambda t_{2,13}/2}\Phi_\lambda^{2,1,3}e^{-\lambda t_{12}/2}\Phi_\lambda^{0,12,3}e^{-\lambda t_{0,12}}\Phi_\lambda^{3,12,0};$$

after some computation, we find that this equals $\Phi_\lambda^{3,2,1}e^{\lambda t_{23}}\Phi_\lambda^{1,2,3}$. □

In particular, $(\Phi_\lambda, \tilde{A}_\lambda, \tilde{B}_\lambda)$ give rise to a morphism $\overline{\mathrm{B}}_{1,n} \to \exp\left(\widehat{\bar{\mathfrak{t}}^{\mathbf{k}}_{1,n}}\right) \rtimes S_n$; one proves as in Section 2 that it induces an isomorphism of filtered Lie algebras $\mathrm{Lie}(\overline{\mathrm{PB}}_{1,n})_{\mathbf{k}} \simeq \widehat{\bar{\mathfrak{t}}^{\mathbf{k}}_{1,n}}$. Taking Φ_λ to be a rational associator [Dri91], we then obtain the following:

Corollary 34. *We have a filtered isomorphism* $\mathrm{Lie}(\overline{\mathrm{PB}}_{1,n})_{\mathbb{Q}} \simeq \widehat{\bar{\mathfrak{t}}^{\mathbb{Q}}_{1,n}}$, *which can be extended to an isomorphism* $\overline{\mathrm{B}}_{1,n}(\mathbb{Q}) \simeq \exp\left(\widehat{\bar{\mathfrak{t}}^{\mathbb{Q}}_{1,n}}\right) \rtimes S_n$.

5.3 Construction of morphisms $\boldsymbol{\Gamma_{1,[n]}} \to \mathbf{G}_{\boldsymbol{n}} \rtimes \boldsymbol{S_n}$ using a pair $(\boldsymbol{\Phi_\lambda}, \boldsymbol{\tilde{\Theta}_\lambda})$

Keep the notation of the previous section and set

$$\tilde{\Psi}_\lambda := \exp\left(-\frac{1}{\lambda}\left(\Delta_0 + \sum_{k\geq 1} a_{2k}(\lambda)\delta_{2k}\right)\right).$$

Proposition 35. *We have*

$$[\tilde{\Psi}_\lambda]e^{\lambda\bar{t}_{12}/12}\tilde{A}_\lambda([\tilde{\Psi}_\lambda]e^{\lambda\bar{t}_{12}/12})^{-1} = \tilde{A}_\lambda,$$
$$[\tilde{\Psi}_\lambda]e^{\lambda\bar{t}_{12}/12}\tilde{B}_\lambda([\tilde{\Psi}_\lambda]e^{\lambda\bar{t}_{12}/12})^{-1} = \tilde{B}_\lambda\tilde{A}_\lambda.$$

Proof. The first identity follows from the fact that $\Delta_0 + \sum_{k\geq 1} a_{2k}(\lambda)[\delta_{2k}] - \lambda^2 t/12$ commutes with t and $\tilde{y}_\lambda$; the second identity follows from these facts and the analogue of Lemma 30, where $2\pi\mathrm{i}$ is replaced by λ. □

Assume that $\tilde{\Theta}_\lambda \in \mathbf{G}_1$ satisfies

$$\tilde{\Theta}_\lambda^4 = (\tilde{\Theta}_\lambda\tilde{\Psi}_\lambda)^3 = \left(\tilde{\Theta}_\lambda^2, \tilde{\Psi}_\lambda\right) = 1,$$
$$[\tilde{\Theta}_\lambda]e^{\lambda\bar{t}_{12}/4}\tilde{A}_\lambda([\tilde{\Theta}_\lambda]e^{\lambda\bar{t}_{12}/4})^{-1} = \tilde{B}_\lambda^{-1},$$
$$[\tilde{\Theta}_\lambda]e^{\lambda\bar{t}_{12}/4}\tilde{B}_\lambda([\tilde{\Theta}_\lambda]e^{\lambda\bar{t}_{12}/4})^{-1} = \tilde{B}_\lambda\tilde{A}_\lambda\tilde{B}_\lambda^{-1}$$

(one can show that the last two equations are equivalent), then $\Theta \mapsto [\tilde{\Theta}_\lambda]e^{\lambda(\sum_{i<j}\bar{t}_{ij})/4}$, $\Psi \mapsto [\tilde{\Psi}_\lambda]e^{\lambda(\sum_{i<j}\bar{t}_{ij})/12}$ extends the morphism defined in Proposition 33 to a morphism $\Gamma_{1,[n]} \to \mathbf{G}_n \rtimes S_n$.

We do not know whether for each Φ_λ defined over $\mathbf{k}$ there exists a $\tilde{\Theta}_\lambda$ defined over $\mathbf{k}$, satisfying the above conditions.

5.4 Elliptic structures over QTQBAs

Let $(H, \Delta_H, R_H, \Phi_H)$ be a quasitriangular quasi-bialgebra (QTQBA). Recall that this means that [Dri90b] (H, m_H) is an algebra, $\Delta_H : H \to H^{\otimes 2}$ is an algebra morphism, $R_H \in H^{\otimes 2}$ and $\Phi_H \in H^{\otimes 3}$ are invertible, and

$$\begin{aligned} \Delta_H(x)^{2,1} &= R_H \Delta_H(x) R_H^{-1}, \\ (\mathrm{id} \otimes \Delta_H) \circ \Delta_H(x) &= \Phi_H (\Delta_H \otimes \mathrm{id}) \circ \Delta_H(x) \Phi_H^{-1}, \\ R_H^{12,3} &= \Phi_H^{3,1,2} R_H^{1,3} \left(\Phi_H^{1,3,2}\right)^{-1} R_H^{2,3} \Phi_H^{1,2,3}, \\ R_H^{1,23} &= \left(\Phi_H^{2,3,1}\right)^{-1} R_H^{1,3} \Phi_H^{2,1,3} R_H^{1,2} \left(\Phi_H^{1,2,3}\right)^{-1}, \\ \Phi_H^{1,2,34} \Phi_H^{12,3,4} &= \Phi_H^{2,3,4} \Phi_H^{1,23,4} \Phi_H^{1,2,3}. \end{aligned}$$

One also assumes the existence of a unit 1_H and a counit ε_H.

If $\mathbf{A}$ is an algebra and $J_1, J_2 \subset \mathbf{A}$ are left ideals, define the Hecke bimodule $\mathcal{H}(\mathbf{A}|J_1, J_2)$ or $\mathcal{H}(J_1, J_2)$ as $\mathrm{Hom}_{\mathbf{A}}(\mathbf{A}/J_1, \mathbf{A}/J_2) = (\mathbf{A}/J_2)^{J_1}$ where J_1 acts on the quotient from the left; we thus have $\mathcal{H}(J_1, J_2) = \{x \in \mathbf{A} | J_1 x \subset J_2\}/J_2$. The product of $\mathbf{A}$ induces a product $\mathcal{H}(J_1, J_2) \otimes \mathcal{H}(J_2, J_3) \to \mathcal{H}(J_1, J_3)$. When $J_1 = J_2 = J$, $\mathcal{H}(J) := \mathcal{H}(J, J)$ is the usual Hecke algebra, and $\mathcal{H}(J_1, J_2)$ is a $(\mathcal{H}(J_1), \mathcal{H}(J_2))$-bimodule. Recall that we have a functor $\mathbf{A}-\mathrm{mod} \to \mathcal{H}(J)-\mathrm{mod}$, $V \mapsto V^J := \{v \in V | Jv = 0\}$.

If H is an algebra with unit equipped with a morphism $\Delta_H : H \to H^{\otimes 2}$ and $a : H \to D$ is a morphism of algebras with unit, we define for each $n \geq 1$ and each pair of words w, w' in the free magma generated by $1, \ldots, n$ containing $1, \ldots, n$ exactly once (recall that a magma is a set with a not necessarily associative binary operation) the Hecke bimodule

$$\mathcal{H}^{w,w'}(D, H) := \mathcal{H}(D \otimes H^{\otimes n} | J_w, J_{w'}),$$

(or simply $\mathcal{H}^{w,w'}$) where $J_w \subset D \otimes H^{\otimes n}$ is the left ideal generated by the image of $(a \otimes \Delta_H^w) \circ \Delta_H : H_+ \to D \otimes H^{\otimes n}$. Here $H_+ = \mathrm{Ker}(H \xrightarrow{\varepsilon_H} \mathbf{k})$ and for example $\Delta_H^{(21)3} = (213) \circ (\Delta_H \otimes \mathrm{id}_H) \circ \Delta_H$, etc. We have products $\mathcal{H}^{w,w'} \otimes \mathcal{H}^{w',w''} \to \mathcal{H}^{w,w''}$. We denote the Hecke algebra $\mathcal{H}^{w,w}$ by $\mathcal{H}^w(D, H)$ or $\mathcal{H}^w$; we denote by 1_w its unit. We denote by $(\mathcal{H}^{w,w'})^\times$ the set of invertible elements of $\mathcal{H}^{w,w'}$, i.e., the set of elements X such that for some $X' \in \mathcal{H}^{w',w}$, $X'X = 1_{w'}$, $XX' = 1_w$. The symmetric group S_n acts on the system of bimodules $\mathcal{H}^{w,w'}$ by permuting the factors, so we get maps $\mathrm{Ad}(\sigma) : \mathcal{H}^{w,w'} \to \mathcal{H}^{\sigma(w),\sigma(w')}$ (where $\sigma(w)$ is the word w, and where i is replaced by $\sigma(i)$). If $w_0 = ((12)\ldots)n$, we define an algebra structure on $\oplus_{\sigma \in S_n} \mathcal{H}^{w_0,\sigma(w_0)}\sigma$ by $\left(\sum_{\sigma \in S_n} h_\sigma \sigma\right)\left(\sum_{\tau \in S_n} h'_\tau \tau\right) := \sum_{\sigma,\tau \in S_n} h_\sigma \mathrm{Ad}(\sigma)\left(h'_\tau\right) \sigma\tau$. Then $\sqcup_{\sigma \in S_n} (\mathcal{H}^{w_0,\sigma(w_0)})^\times \sigma \subset \oplus_{\sigma \in S_n} \mathcal{H}^{w_0,\sigma(w_0)}\sigma$ is a group with unit 1_{w_0}. We have an exact sequence $1 \to (\mathcal{H}^{w_0})^\times \to \sqcup_{\sigma \in S_n} (\mathcal{H}^{w_0,\sigma(w_0)})^\times \sigma \to S_n$, but the last map is not necessarily surjective (and if it is, does not necessarily split).

If H is a quasi-bialgebra, then Φ_H gives rise to an element of $\mathcal{H}^{1(23),(12)3}(D,H)$, which we also denote by Φ_H; similarly, Φ_H^{-1} gives rise to the inverse (w.r.t. composition of Hecke bimodules) element $\Phi_H^{-1} \in \mathcal{H}^{(12)3,1(23)}(D,H)$. We have algebra morphisms $\mathcal{H}^{12}(D,H) \to \mathcal{H}^{(12)3}(D,H)$ induced by $X \mapsto X^{0,12,3} := (\mathrm{id}_D \otimes \Delta_H \otimes \mathrm{id}_H)(X)$ (0 is the index of D) and similarly morphisms $\mathcal{H}^{12}(D,H) \to \mathcal{H}^{2(13)}(D,H)$, $X \mapsto X^{0,2,13}$, $\mathcal{H}^{12}(D,H) \to \mathcal{H}^{1}(D,H)$, $X \mapsto X^{0,1,\emptyset}$ and $X^{0,\emptyset,1}$, etc. If, moreover, H is quasi-triangular, then $R_H \in \mathcal{H}^{21,12}(D,H)$, $R_H^{-1} \in \mathcal{H}^{12,21}(D,H)$, so in that case $\sqcup_{\sigma\in S_n}\mathcal{H}^{w_0,\sigma(w_0)}\sigma \to S_n$ is surjective, and we have a morphism $\mathrm{B}_n \to \sqcup_{\sigma\in S_n}\mathcal{H}^{w_0,\sigma(w_0)}\sigma$ such that the composition $\mathrm{B}_n \to \sqcup_{\sigma\in S_n}\mathcal{H}^{w_0,\sigma(w_0)}\sigma \to S_n$ is the canonical projection.

Definition 36. *If H is a QTQBA, an elliptic structure on H is a triple (D,A,B), where D is an algebra with unit, equipped with an algebra morphism $a : H \to D$, and $A, B \in \mathcal{H}^{12}(D,H)$ are invertible such that $A^{0,1,\emptyset} = A^{0,\emptyset,1} = B^{0,1,\emptyset} = B^{0,\emptyset,1} = 1_D \otimes 1_H$,*

$$A^{0,12,3} = R_H^{2,1}\left(\Phi_H^{2,1,3}\right)^{-1} A^{0,2,13}\Phi_H^{2,1,3}R_H^{1,2}\left(\Phi_H^{1,2,3}\right)^{-1} A^{0,1,23}\Phi_H^{1,2,3}, \quad (41)$$

$$B^{0,12,3} = \left(R_H^{1,2}\right)^{-1}\left(\Phi_H^{2,1,3}\right)^{-1} B^{0,2,13}\Phi_H^{2,1,3} \left(R_H^{2,1}\right)^{-1}\left(\Phi_H^{1,2,3}\right)^{-1} B^{0,1,23}\Phi_H^{1,2,3} \quad (42)$$

and

$$\begin{aligned}&\left(B^{0,12,3}, R_H^{2,1}\left(\Phi_H^{2,1,3}\right)^{-1} A^{0,2,13}\Phi_H^{2,1,3}R_H^{1,2}\right)\\ &\quad = \left(\left(R_H^{1,2}\right)^{-1}\left(\Phi_H^{2,1,3}\right)^{-1} B^{0,2,13}\Phi_H^{2,1,3}\left(R_H^{2,1}\right)^{-1}, A^{0,12,3}\right)\\ &\quad = \left(\Phi_H^{1,2,3}\right)^{-1} R_H^{3,2}R_H^{2,3}\Phi_H^{1,2,3}\end{aligned}$$

(identities in $\mathcal{H}^{(12)3}(D,H)$).

The pair of identities (41), (42) is equivalent to

$$\begin{cases} R_H^{2,1}A^{0,2,1}R_H^{1,2}A^{0,1,2} = 1 \\ R_H^{3,12}A^{0,3,12}\Phi_H^{3,1,2}R_H^{2,31}A^{0,2,31}\Phi_H^{2,3,1}R_H^{1,23}A^{0,1,23}\Phi_H^{1,2,3} = 1, \end{cases}$$

and

$$\begin{cases} \left(R_H^{1,2}\right)^{-1} B^{0,2,1}\left(R_H^{2,1}\right)^{-1} B^{0,1,2} = 1 \\ \left(R_H^{-1}\right)^{12,3} B^{0,3,12}\Phi_H^{3,1,2}\left(R_H^{-1}\right)^{31,2} B^{0,2,31}\Phi_H^{2,3,1}\left(R_H^{-1}\right)^{23,1} B^{0,1,23}\Phi_H^{1,2,3} = 1, \end{cases}$$

so the invertibility conditions on A, B follow from (41), (42).

If $F \in H^{\otimes 2}$ is invertible with $(\varepsilon_H \otimes \mathrm{id}_H)(F) = (\mathrm{id}_H \otimes \varepsilon_H)(F) = 1_H$, then the twist of H by F is the quasi-Hopf algebra ${}^F H$ with product

m_H, coproduct $\tilde{\Delta}_H(x) = F\Delta_H(x)F^{-1}$, R-matrix $\tilde{R}_H = F^{2,1}R_HF^{-1}$, and associator $\tilde{\Phi}_H = F^{2,3}F^{1,23}\Phi_H(F^{1,2}F^{12,3})^{-1}$. If $a : H \to D$ is an algebra morphism, it can be viewed as a morphism ${}^FH \to D$, and we have an algebra isomorphism $\mathcal{H}^{(12)3}(D,H) \to \mathcal{H}^{(12)3}(D,{}^FH)$, induced by $X \mapsto F^{1,2}F^{0,12}X(F^{1,2}F^{0,12})^{-1}$ (more generally, we have an isomorphism of the systems of bimodules $\mathcal{H}^{w,w'}(D,H) \to \mathcal{H}^{w,w'}(D,{}^FH)$ induced by $X \mapsto F_wXF_{w'}^{-1}$ for suitable F_w).

If (D,A,B) is an elliptic structure on H, then an elliptic structure on ${}^FH$ is $(D,\tilde{A},\tilde{B})$, where $\tilde{A} = F^{1,2}F^{0,12}A(F^{1,2}F^{0,12})^{-1}$ and $\tilde{B} = F^{1,2}F^{0,12}B$ $(F^{1,2}F^{0,12})^{-1}$.

An elliptic structure (D,A,B) over H gives rise to a unique group morphism

$$\overline{\mathrm{B}}_{1,n} \to \sqcup_{\sigma\in S_n}\mathcal{H}^{w_0,\sigma(w_0)}(D,H)^\times\sigma$$

such that

$$\begin{aligned}
\sigma_i &\mapsto \left(\Phi_H^{(((12)3)\dots i-1),i,i+1}\right)^{-1} R_H^{i,i+1}\,(i,i+1)\,\Phi_H^{(((12)3)\dots i-1),i,i+1},\\
A_i &\mapsto \Phi_{H,i}^{-1}A^{0,(((12)3)\dots i-1),(i\dots(n-1,n))}\Phi_{H,i},\\
B_i &\mapsto \Phi_{H,i}^{-1}B^{0,(((12)3)\dots i-1),(i\dots(n-1,n))}\Phi_{H,i},
\end{aligned}$$

where

$$\Phi_{H,i} = \Phi_H^{((12)\dots i-1),i,(i+1(\dots(n-1,n)))}\cdots\Phi_H^{((12)\dots n-2),n-1,n};$$

here we have, for example, $x^{((12)3)} = (\Delta_H\otimes\mathrm{id}_H)\circ\Delta_H(x)$ for $x\in H$.

If $\mathfrak{g}$ is a Lie algebra and $t_\mathfrak{g}\in S^2(\mathfrak{g})^\mathfrak{g}$ is nondegenerate, then $H = U(\mathfrak{g})[[\hbar]]$ is a QTQBA, with m_H, Δ_H are the undeformed product and coproduct, $R_H = e^{\hbar t_\mathfrak{g}/2}$, and $\Phi_H = \Phi\left(\hbar t_\mathfrak{g}^{1,2}, \hbar t_\mathfrak{g}^{2,3}\right)$, where Φ is a 1-associator. The results of the next section then imply that (D,A,B) is an elliptic structure over H, where $D = D_{(\mathfrak{g})}[[\hbar]]$ ($D_{(\mathfrak{g})}$ is the algebra of differential operators on the formal neighborhood of the origin in $\mathfrak{g}$) and A,B are given by the formulas for $\tilde{A}_\lambda, \tilde{B}_\lambda$ with t replaced by $\hbar t_\mathfrak{g}^{1,2}$, x replaced by $\sum_\alpha \mathrm{x}_\alpha\otimes\left(e_\alpha^1\right)$, y replaced by $-\hbar\sum_\alpha\partial_\alpha\otimes$ $\left(e_\alpha^1\right)$.

Remark 37. If H is a Hopf algebra, we have an isomorphism

$$\mathcal{H}^{w_0}(D,H) \simeq (D\otimes H^{\otimes n-1})^H,$$

where the right side is the commutant of the diagonal map $H \to D\otimes H^{\otimes n-1}$, $h \mapsto \left(a\otimes\mathrm{id}_H^{\otimes n-1}\right)\circ\Delta_H^{(n)}(h)$. This map takes the class of $d\otimes h_1\otimes\cdots\otimes h_n$ to $da(S_H\left(h_n^{(n)}\right))\otimes h_1S_H\left(h_n^{(n-1)}\right)\otimes\cdots\otimes h_{n-1}S_H\left(h_n^{(1)}\right)$ (S_H is the antipode of H). So A,B identify with elements $\mathcal{A},\mathcal{B}\in(D\otimes H)^H$; the conditions are then

$$\mathcal{A}^{0,12} = R_H^{2,1}\mathcal{A}^{0,2}R_H^{1,2}\mathcal{A}^{0,1}, \quad \mathcal{B}^{0,12} = \left(R_H^{1,2}\right)^{-1}\mathcal{B}^{0,2}\left(R_H^{2,1}\right)^{-1}\mathcal{B}^{0,1},$$

$$\left(\mathcal{B}^{0,12}, R_H^{2,1}\mathcal{A}^{0,2}R_H^{1,2}\right) = \left(\left(R_H^{1,2}\right)^{-1}\mathcal{B}^{0,2}\left(R_H^{2,1}\right)^{-1}, \mathcal{A}^{0,12}\right)$$
$$= \left(R_H^{3,2}R_H^{1,2}R_H^{0,2}R_H^{2,0}R_H^{2,1}R_H^{2,3}\right)^{\tilde{0},\tilde{1},2\cdot\tilde{3}}$$

(conditions in $(D\otimes H^{\otimes 2})^H$), where the superscript $\mathrm{B}'_n \rtimes \mathbb{Z}^{n-1} \to \mathrm{B}_{n-1} \rtimes \mathbb{Z}^{n-1}$ is the map $x_0 \otimes \cdots \otimes x_3 \mapsto S_H(x_0) \otimes S_H(x_1) \otimes x_2 S_H(x_3)$.

Moreover, the morphism $\mathrm{PB}_n \to (\mathcal{H}^{w_0})^\times \simeq (D\otimes H^{\otimes n-1})^H$ factors through $\mathrm{PB}_n \to \mathrm{PB}_{n-1} \times \mathbb{Z}^{n-1} \to (D \otimes H^{\otimes n-1})^H$, where (a) the first morphism is induced by $\mathbb{Z}^{n-1} \rtimes \mathrm{B}'_n \to \mathbb{Z}^{n-1} \rtimes \mathrm{B}_{n-1}$ (where $\mathrm{B}'_n = \mathrm{B}_n \times_{S_n} S_{n-1}$ is the group of braids leaving the last strand fixed), constructed as follows: we have a composition $\mathrm{B}'_{n+1} \to \pi_1((\mathbb{P}^1)^{n+1} - \text{diagonals}/S_n) \to \pi_1(\mathbb{C}^n - \text{diagonals}/S_n) = \mathrm{B}_n$, where the first map is induced by $\mathbb{C} \subset \mathbb{P}^1$, and the middle map comes from the fibration $\mathbb{C}^n - \text{diagonals} \to (\mathbb{P}^1)^{n+1} - \text{diagonals} \to \mathbb{P}^1$, $(z_1, \ldots, z_n) \to (z_1, \ldots, z_n, \infty)$ and $(z_1, \ldots, z_{n+1}) \to z_{n+1}$ [the second projection has a section so the map between π_1's is an isomorphism]; viewing $\mathbb{Z}^{n-1} \rtimes \mathrm{B}'_n$, $\mathbb{Z}^{n-1} \rtimes \mathrm{B}_{n-1}$ as fundamental groups of configuration spaces of points equipped with a nonzero tangent vector, we then get the morphism $\mathbb{Z}^{n-1} \rtimes \mathrm{B}'_n \to \mathbb{Z}^{n-1} \rtimes \mathrm{B}_{n-1}$ (which does not restrict to a morphism $\mathrm{B}'_n \to \mathrm{B}_{n-1}$); (b) the second map is induced by the standard map $\mathrm{PB}_{n-1} \times \mathbb{Z}^{n-1} \to (H^{\otimes n-1})^\times$ induced by $R_H = \sum_\alpha r'_\alpha \otimes r''_\alpha$ and the map taking the i-th generator of $\mathbb{Z}^{n-1}$ to $1 \otimes \cdots \otimes uS_H(u) \otimes \cdots \otimes 1$, where $u = \sum_i S_H(r''_\alpha)\, r'_\alpha$ (see [Dri90a]). The morphism $\mathrm{B}_n \to \mathrm{Aut}((\mathcal{H}^{w_0})^\times) = \mathrm{Aut}((D \otimes H^{\otimes n-1})^H)$ extends the inner action of PB_n by

$$\sigma_{n-1} \cdot X := \left\{R_H^{n-1,n\ldots 2n-1} X^{0,1,\ldots,n-2,n\ldots 2n-1} R_H^{n\ldots 2n-1,n-1}\right\}^{0\cdot\widetilde{2n-1},\ldots,n-1\cdot\tilde{n}}$$

(where the superscript means that $x_0 \otimes \cdots \otimes x_{2n-1}$ maps to

$$x_0 S_H(x_{2n-1}) \otimes \cdots \otimes x_{n-1} S_H(x_n)).$$

We have then $\sqcup_{\sigma\in S_n} (\mathcal{H}^{w_0,\sigma(w_0)})^\times \sigma \simeq ((D \otimes H^{\otimes n-1})^\times)^H \rtimes_{\mathrm{PB}_n} \mathrm{B}_n$ (the index means that $\mathrm{PB}_n \subset \mathrm{B}_n$ is identified with its image in $((D \otimes H^{\otimes n-1})^\times)^H$).

Then if $(\mathcal{A}, \mathcal{B})$ is an elliptic structure over $a : H \to D$, the morphism $\mathrm{B}_n \to ((D \otimes H^{\otimes n-1})^\times)^H \rtimes_{\mathrm{PB}_n} \mathrm{B}_n$ extends to a morphism

$$\overline{\mathrm{B}}_{1,n} \to ((D \otimes H^{\otimes n-1})^\times)^H \rtimes_{\mathrm{PB}_n} \mathrm{B}_n$$

via $A_i \mapsto \mathcal{A}^{0,1\ldots i-1}$, $B_i \mapsto \mathcal{B}^{0,1\ldots i-1}$.

This interpretation of $\mathcal{H}^{w_0}$ and of the relations between $\mathcal{A}, \mathcal{B}$ can be extended to the case when H is a quasi-Hopf algebra.

Remark 38. Let $\mathcal{C}$ be a rigid braided monoidal category. We define an elliptic structure on $\mathcal{C}$ as a quadruple $(\mathcal{E}, A, B, F)$, where $\mathcal{E}$ is a category, $F : \mathcal{E} \to \mathcal{C}$ is a functor, and A, B are functorial automorphisms of $F(?)\otimes ?$, which reduce to the identity if the second factor is the neutral object $\mathbf{1}$, and such that the following equalities of automorphisms of $F(M)\otimes(X\otimes Y)$ hold (we write them omitting associativity maps, since they can be put in automatically):

$$A_{M,X\otimes Y}=\beta_{Y,X}A_{M,Y}\beta_{X,Y}A_{M,X},$$

$$B_{M,X\otimes Y}=\beta_{X,Y}^{-1}B_{M,Y}\beta_{Y,X}^{-1}B_{M,X},$$

$$\begin{aligned}(B_{M,X\otimes Y},\beta_{Y,X}A_{M,Y}\beta_{X,Y}) &= \left(\beta_{Y,X}^{-1}B_{M,Y}\beta_{X,Y}^{-1},A_{M,X\otimes Y}\right)\\ &= \beta_{(M\otimes X\otimes Y)^*,Y}\beta_{Y,(M\otimes X\otimes Y)^*}\circ\mathrm{can}_{M\otimes X\otimes Y},\end{aligned}$$

where $\mathrm{can}_X\in\mathrm{Hom}_{\mathcal{C}}(\mathbf{1},X\otimes X^*)$ is the canonical map and the r.h.s. of the last identity is viewed as an element of $\mathrm{End}_{\mathcal{C}}(M\otimes X\otimes Y)$ using its identification with $\mathrm{Hom}_{\mathcal{C}}(\mathbf{1},(M\otimes X\otimes Y)\otimes(M\otimes X\otimes Y)^*)$. An elliptic structure on a quasitriangular quasi-Hopf algebra H gives rise to an elliptic structure on H-mod. An elliptic structure over a rigid braided monoidal category $\mathcal{C}$ gives rise to representations of $\overline{\mathrm{B}}_{1,n}$ by $\mathcal{C}$-automorphisms of $F(M)\otimes X^{\otimes n-1}$.

6 The KZB connection as a realization of the universal KZB connection

6.1 Realizations of $\bar{\mathfrak{t}}_{1,n}$

Let $\mathfrak{g}$ be a Lie algebra and let $t_{\mathfrak{g}}\in S^2(\mathfrak{g})^{\mathfrak{g}}$ be nondegenerate. We denote by $(a,b)\mapsto\langle a,b\rangle$ the corresponding invariant pairing.

Let $D(\mathfrak{g})$ be the algebra of algebraic differential operators on $\mathfrak{g}$. It has generators x_a, ∂_a, $a\in\mathfrak{g}$, and relations $a\mapsto\mathrm{x}_a$, $a\mapsto\partial_a$ are linear, $[\mathrm{x}_a,\mathrm{x}_b]=[\partial_a,\partial_b]=0$, $[\partial_a,\mathrm{x}_b]=\langle a,b\rangle$.

There is a unique Lie algebra morphism $\mathfrak{g}\to D(\mathfrak{g})$, $a\mapsto X_a$, where $X_a:=\sum_\alpha \mathrm{x}_{[a,e_\alpha]}\partial_{e_\alpha}$, and $t_{\mathfrak{g}}=\sum_\alpha e_\alpha\otimes e_\alpha$ (it is the infinitesimal of the adjoint action). We also have a Lie algebra morphism $\mathfrak{g}\to A_n:=D(\mathfrak{g})\otimes U(\mathfrak{g})^{\otimes n}$, $a\mapsto Y_a:=X_a\otimes 1+1\otimes\left(\sum_{i=1}^n a^{(i)}\right)$. We denote by $\mathfrak{g}^{\mathrm{diag}}$ the image of this morphism. We denote by $\mathcal{H}_n(\mathfrak{g})$ the Hecke algebra of $(A_n,\mathfrak{g}^{\mathrm{diag}})$. It is defined as the quotient $\{x\in A_n|\forall a\in\mathfrak{g},Y_ax\in A_n\mathfrak{g}^{\mathrm{diag}}\}/A_n\mathfrak{g}^{\mathrm{diag}}$. We have a natural action of S_n on A_n, which induces an action of S_n on $\mathcal{H}_n(\mathfrak{g})$.

If $(V_i)_{i=1,\dots,n}$ are $\mathfrak{g}$-modules, then $(S(\mathfrak{g})\otimes(\otimes_{i=1}^n V_i))^{\mathfrak{g}}$ is a module over $\mathcal{H}_n(\mathfrak{g})$. If, moreover, $V_1=\cdots=V_n$, this is a module over $\mathcal{H}_n(\mathfrak{g})\rtimes S_n$.

Proposition 39. *There is a unique Lie algebra morphism* $\rho_{\mathfrak{g}}:\bar{\mathfrak{t}}_{1,n}\to\mathcal{H}_n(\mathfrak{g})$, $\bar{x}_i\mapsto\sum_\alpha\mathrm{x}_\alpha\otimes e_\alpha^{(i)}$, $\bar{y}_i\mapsto-\sum_\alpha\partial_\alpha\otimes e_\alpha^{(i)}$, $\bar{t}_{ij}\mapsto 1\otimes t_{\mathfrak{g}}^{(ij)}$ *(we set* $\mathrm{x}_\alpha:=\mathrm{x}_{e_\alpha}$, $\partial_\alpha:=\partial_{e_\alpha}$*).*

Proof. The images of all the generators of $\bar{\mathfrak{t}}_{1,n}$ are contained in the commutant of $\mathfrak{g}^{\mathrm{diag}}$ in A_n, therefore also in its normalizer. According to Lemma 4, we will use the following presentation of $\bar{\mathfrak{t}}_{1,n}$. Generators are $\bar{x}_i,\bar{y}_i,\bar{t}_{ij}$, relations are

$[\bar{x}_i, \bar{x}_j] = [\bar{y}_i, \bar{y}_j] = 0$, $[\bar{x}_i, \bar{y}_j] = \bar{t}_{ij}$ $(i \neq j)$, $\bar{t}_{ij} = \bar{t}_{ji}$, $\sum_i \bar{x}_i = \sum_i \bar{y}_i = 0$, $[\bar{x}_i, \bar{t}_{jk}] = [\bar{y}_i, \bar{t}_{jk}] = 0$ $(i, j, k$ distinct$)$.

The relations $[\bar{x}_i, \bar{x}_j] = [\bar{y}_i, \bar{y}_j] = 0$, $[\bar{x}_i, \bar{y}_j] = \bar{t}_{ij}$ $(i \neq j)$, $\bar{t}_{ij} = \bar{t}_{ji}$ and $[\bar{x}_i, \bar{t}_{jk}] = [\bar{y}_i, \bar{t}_{jk}] = 0$ are obviously preserved. Let us check that $\sum_i \bar{x}_i = \sum_i \bar{y}_i = 0$ are preserved.

We have

$$\sum_i \rho_{\mathfrak{g}}(\bar{x}_i) = \sum_\alpha \mathrm{x}_\alpha \otimes \left(\sum_i e_\alpha^{(i)} \right) = \sum_\alpha (\mathrm{x}_\alpha \otimes 1)(Y_\alpha - X_\alpha \otimes 1)$$
$$\equiv -\sum_\alpha \mathrm{x}_\alpha X_\alpha \otimes 1 = \sum_{\alpha,\beta} \mathrm{x}_{e_\alpha} \mathrm{x}_{[e_\alpha, e_\beta]} \partial_{e_\beta} \otimes 1 = 0,$$

since x_α commutes with $\mathrm{x}_{[e_\alpha, e_\beta]}$ and $\sum_\beta e_\beta \otimes e_\beta = t_{\mathfrak{g}}$ is invariant. We also have

$$\sum_i \rho_{\mathfrak{g}}(\bar{y}_i) = -\sum_\alpha \partial_\alpha \otimes \left(\sum_i e_\alpha^{(i)} \right) = -\sum_\alpha (\partial_\alpha \otimes 1)(Y_\alpha - X_\alpha \otimes 1)$$
$$\equiv \sum_\alpha \partial_\alpha X_\alpha \otimes 1 = -\sum_{\alpha,\beta} \partial_{e_\alpha} \mathrm{x}_{[e_\alpha, e_\beta]} \partial_{e_\beta}$$
$$= -\sum_{\alpha,\beta} \langle e_\alpha, [e_\alpha, e_\beta] \rangle \partial_{e_\beta} - \sum_{\alpha,\beta} \mathrm{x}_{[e_\alpha, e_\beta]} \partial_{e_\alpha} \partial_{e_\beta}.$$

Since $t_{\mathfrak{g}}$ is invariant and $\langle -, - \rangle$ is symmetric, we have $\sum_\alpha \langle e_\alpha, [e_\alpha, e_\beta] \rangle = 0$ for any β, and since $[\partial_{e_\alpha}, \partial_{e_\beta}] = 0$, we have $\sum_{\alpha,\beta} \mathrm{x}_{[e_\alpha, e_\beta]} \partial_{e_\alpha} \partial_{e_\beta}$, so $\sum_i \rho_{\mathfrak{g}}(\bar{y}_i) = 0$. □

6.2 Realizations of $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}$

Let $(\mathfrak{g}, t_{\mathfrak{g}})$ be as in Section 6.1. We keep the same notation.

Proposition 40. *The Lie algebra morphism $\rho_{\mathfrak{g}} : \bar{\mathfrak{t}}_{1,n} \to \mathcal{H}_n(\mathfrak{g})$ of Proposition 39 extends to a Lie algebra morphism $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d} \to \mathcal{H}_n(\mathfrak{g})$, defined by $\Delta_0 \mapsto -\frac{1}{2} \left(\sum_\alpha \partial_\alpha^2 \right) \otimes 1$, $X \mapsto \frac{1}{2} \left(\sum_\alpha \mathrm{x}_\alpha^2 \right) \otimes 1$, $d \mapsto \frac{1}{2}(\sum_\alpha \mathrm{x}_\alpha \partial_\alpha + \partial_\alpha \mathrm{x}_\alpha) \otimes 1$, and*

$$\delta_{2m} \to \frac{1}{2} \sum_{\alpha_1, \ldots, \alpha_{2m}, \alpha} \mathrm{x}_{\alpha_1} \cdots \mathrm{x}_{\alpha_{2m}} \otimes \left(\sum_{i=1}^n (ad(e_{\alpha_1}) \cdots ad(e_{\alpha_{2m}})(e_\alpha) \cdot e_\alpha)^{(i)} \right)$$

for $m \geq 1$. This morphism further extends to a morphism $U(\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}) \rtimes S_n \to \mathcal{H}_n(\mathfrak{g}) \rtimes S_n$ by $\sigma \mapsto \sigma$.

Proof. First of all $[\rho_{\mathfrak{g}}(\delta_{2m}), \rho_{\mathfrak{g}}(\bar{x}_i)]$ equals

$$\frac{1}{2} \sum_{\alpha_1, \ldots, \alpha_{2m}, \alpha, \beta} \mathrm{x}_{\alpha_1} \cdots \mathrm{x}_{\alpha_{2m}} \mathrm{x}_\beta \otimes [e_\beta, \mathrm{ad}(e_{\alpha_1}) \cdots \mathrm{ad}(e_{\alpha_{2m}})(e_\alpha) e_\alpha]^{(i)}$$

$$= \frac{1}{2} \sum_{\alpha_1,\dots,\alpha_{2m},\alpha,\beta} \mathrm{x}_{\alpha_1} \cdots \mathrm{x}_{\alpha_{2m}} \mathrm{x}_\beta$$
$$\otimes \sum_{\ell=1}^{2m} \big(\mathrm{ad}(e_{\alpha_1}) \cdots \mathrm{ad}([e_\beta, e_{\alpha_\ell}]) \cdots \mathrm{ad}(e_{\alpha_{2m}})(e_\alpha) e_\alpha\big)^{(i)}$$

(the equality follows from the invariance of $t_{\mathfrak{g}}$), which equals zero since the first factor is symmetric in (β, α_ℓ), while the second is antisymmetric in (β, α_l).

We note that $\rho_{\mathfrak{g}}$ preserves the relation $[\delta_{2m}, \bar{t}_{ij}] = [\bar{t}_{ij}, \mathrm{ad}(\bar{x}_i)^{2m}(\bar{t}_{ij})]$, because $\rho_{\mathfrak{g}}(\delta_{2m} + \sum_{i<j} \mathrm{ad}(\bar{x}_i)^{2m}(\bar{t}_{ij}))$ belongs to $D(\mathfrak{g}) \otimes \mathrm{Im}(\Delta^{(n)} : U(\mathfrak{g}) \to U(\mathfrak{g})^{\otimes n})$, where $\Delta^{(n)}$ is the n-fold coproduct and $U(\mathfrak{g})$ is equipped with its standard bialgebra structure.

Now $[\rho_{\mathfrak{g}}(\delta_{2m}), \rho_{\mathfrak{g}}(\bar{y}_i)]$ yields

$$\frac{1}{2} \sum_{\alpha_1,\dots,\alpha_{2m},\alpha,\beta} \Bigg(\sum_j [\partial_\beta, x_{\alpha_1} \cdots x_{\alpha_{2m}}] \otimes e_\beta^{(i)} \mathrm{ad}(e_{\alpha_1}) \cdots \mathrm{ad}(e_{\alpha_{2m}})(e_\alpha)^{(j)} e_\alpha^{(j)}$$
$$+ x_{\alpha_1} \cdots x_{\alpha_{2m}} \partial_\beta \otimes [e_\beta, \mathrm{ad}(e_{\alpha_1}) \cdots \mathrm{ad}(e_{\alpha_{2m}})(e_\alpha) \cdot e_\alpha]^{(i)} \Bigg)$$
$$= \frac{1}{2} \sum_{l=1}^{2m} \sum_{\alpha_1,\dots,\alpha_{2m},\alpha} \Bigg(\sum_j x_{\alpha_1} \cdots \check{x}_{\alpha_l} \cdots x_{\alpha_{2m}} \otimes e_{\alpha_l}^{(i)} \mathrm{ad}(e_{\alpha_1}) \cdots \mathrm{ad}(e_{\alpha_{2m}})(e_\alpha)^{(j)} e_\alpha^{(j)}$$
$$+ x_{\alpha_1} \cdots x_{\alpha_{2m}} \partial_\beta \otimes \mathrm{ad}(e_{\alpha_1}) \cdots \mathrm{ad}([e_\beta, e_{\alpha_l}]) \cdots \mathrm{ad}(e_{\alpha_{2m}})(e_\alpha)^{(i)} e_\alpha^{(i)} \Bigg)$$
$$\equiv \frac{1}{2} \sum_{l=1}^{2m} \sum_{\alpha_1,\dots,\alpha_{2m},\alpha} \sum_j \Big(x_{\alpha_1} \cdots \check{x}_{\alpha_l} \cdots x_{\alpha_{2m}} \otimes e_{\alpha_l}^{(i)} \mathrm{ad}(e_{\alpha_1}) \cdots \mathrm{ad}(e_{\alpha_{2m}})(e_\alpha)^{(j)} e_\alpha^{(j)}$$
$$- x_{\alpha_1} \cdots \check{x}_{\alpha_l} \cdots x_{\alpha_{2m}} \otimes \mathrm{ad}(e_{\alpha_1}) \cdots \mathrm{ad}(e_{\alpha_{2m}})(e_\alpha)^{(i)} e_\alpha^{(i)} e_{\alpha_l}^{(j)} \Big) .$$

The term corresponding to $j = i$ is

$$\frac{1}{2} \sum_{l=1}^{2m} \sum_{\alpha_1,\dots,\alpha_{2m},\alpha} x_{\alpha_1} \cdots \check{x}_{\alpha_l} \cdots x_{\alpha_{2m}} \otimes [e_{\alpha_l}, \mathrm{ad}(e_{\alpha_1}) \cdots \mathrm{ad}(e_{\alpha_{2m}})(e_\alpha) \cdot e_\alpha]^{(i)}.$$

It corresponds to the linear map $S^{2m-1}(\mathfrak{g}) \to U(\mathfrak{g})$ such that for $x \in \mathfrak{g}$,

$$x^{2m-1} \mapsto \frac{1}{2} \sum_{p+q=2m-1} \sum_{\alpha,\beta} [e_\beta, \mathrm{ad}(x)^p \mathrm{ad}(e_\beta) \mathrm{ad}(x)^q (e_\alpha) \cdot e_\alpha]$$
$$= \frac{1}{2} \sum_{\alpha,\beta} \sum_{p+q+r=2m-2} \mathrm{ad}(x)^p \mathrm{ad}([e_\beta, x]) \mathrm{ad}(x)^q \mathrm{ad}(e_\beta) \mathrm{ad}(x)^r (e_\alpha) \cdot e_\alpha$$
$$+ \mathrm{ad}(x)^p \mathrm{ad}(e_\beta) \mathrm{ad}(x)^q \mathrm{ad}([e_\beta, x]) \mathrm{ad}(x)^r (e_\alpha) \cdot e_\alpha,$$

since $\mu(t_{\mathfrak{g}}) = 0$ ($\mu : \mathfrak{g}^{\otimes 2} \to \mathfrak{g}$ is the Lie bracket) and $t_{\mathfrak{g}}$ is $\mathfrak{g}$-invariant. Now this is zero, since $t_{\mathfrak{g}} = \sum_\beta e_\beta \otimes e_\beta$ is invariant.

The term corresponding to $j \neq i$ corresponds to the map $S^{2m-1}(\mathfrak{g}) \to U(\mathfrak{g})^{\otimes n}$ such that for $x \in \mathfrak{g}$,

$$x^{2m-1} \mapsto -\frac{1}{2}\sum_{l=1}^{2m}\sum_{\alpha,\beta}\big((\mathrm{ad}\,x)^{l-1}(\mathrm{ad}\,e_\beta)(\mathrm{ad}\,x)^{2m-l}(e_\alpha)\cdot e_\alpha\big)^{(i)}e_\beta^{(j)} - (i \leftrightarrow j)$$

$$= \frac{1}{2}\sum_{l=1}^{2m}(-1)^{l+1}\sum_{\alpha,\beta}\big((\mathrm{ad}\,x)^{l-1}([e_\beta,e_\alpha])\cdot(\mathrm{ad}\,x)^{2m-l}(e_\alpha)\big)^{(i)}e_\beta^{(j)} - (i \leftrightarrow j)$$

$$= \frac{1}{2}\sum_{l=1}^{2m}(-1)^{l-1}\sum_{\alpha,\beta}\big((\mathrm{ad}\,x)^{l-1}(e_\beta)\cdot(\mathrm{ad}\,x)^{2m-l}(e_\alpha)\big)^{(i)}[e_\alpha,e_\beta]^{(j)} - (i \leftrightarrow j)$$

$$= \frac{1}{2}\sum_{l=1}^{2m}(-1)^{l}\left[\sum_{\alpha}\big((\mathrm{ad}\,x)^{l-1}(e_\alpha)\big)^{(i)}e_\alpha^{(j)}, \sum_{\beta}\big((\mathrm{ad}\,x)^{2m-l}(e_\beta)\big)^{(i)}e_\beta^{(j)}\right],$$

which coincides with the image of

$$\frac{1}{2}\sum_{p+q=2m-1}(-1)^q[(\mathrm{ad}\ \bar{x}_i)^p(\bar{t}_{ij}),(\mathrm{ad}\ \bar{x}_i)^q(\bar{t}_{ij})].$$

It is then clear that $\rho_{\mathfrak{g}}$ preserves the commutation relations of $\varDelta_0, X$, and d with δ_{2m}. □

6.3 Reductions

Assume that $\mathfrak{g}$ is finite-dimensional and we have a reductive decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{n}$, i.e., $\mathfrak{h} \subset \mathfrak{g}$ is a Lie subalgebra and $\mathfrak{n} \subset \mathfrak{g}$ is a vector subspace such that $[\mathfrak{h},\mathfrak{n}] \subset \mathfrak{n}$; assume also that $t_{\mathfrak{g}} = t_{\mathfrak{h}} + t_{\mathfrak{n}}$, where $t_{\mathfrak{h}} \in S^2(\mathfrak{h})^{\mathfrak{h}}$ and $t_{\mathfrak{n}} \in S^2(\mathfrak{n})^{\mathfrak{h}}$.

We assume that for a generic $h \in \mathfrak{h}$, $\mathrm{ad}(h)_{|\mathfrak{n}} \in \mathrm{End}(\mathfrak{n})$ is invertible. This condition is equivalent to the nonvanishing of $P(\lambda) := \det(\mathrm{ad}(\lambda^\vee)_{|\mathfrak{n}}) \in S^{\dim\mathfrak{n}}(\mathfrak{h})$, where $\lambda \mapsto \lambda^\vee$ is the map $\mathfrak{h}^* \to \mathfrak{h}$, with $\lambda^\vee := (\lambda \otimes \mathrm{id})(t_{\mathfrak{h}})$. If G is a Lie group with Lie algebra $\mathfrak{g}$, an equivalent condition is that a generic element of $\mathfrak{g}^*$ is conjugate to some element in $\mathfrak{h}^*$ (see [EE05]).

Let us set, for $\lambda \in \mathfrak{h}^*$,

$$r(\lambda) := (\mathrm{id} \otimes (\mathrm{ad}\,\lambda^\vee)^{-1}_{|\mathfrak{n}})(t_{\mathfrak{n}}).$$

Then $r : \mathfrak{h}^*_{\mathrm{reg}} \to \wedge^2(\mathfrak{n})$ is an $\mathfrak{h}$-equivariant map (here $\mathfrak{h}^*_{\mathrm{reg}} = \{\lambda \in \mathfrak{h}^* | P(\lambda) \neq 0\}$), satisfying the classical dynamical Yang–Baxter (CDYB) equation

$$\mathrm{CYB}(r) - \mathrm{Alt}(\mathrm{d}r) = 0$$

(see [EE05]). Here for $r = \sum_\alpha a_\alpha \otimes b_\alpha \otimes \ell_\alpha \in (\mathfrak{n}^{\otimes 2} \otimes S(\mathfrak{h})[1/P])^{\mathfrak{h}}$, we set $\mathrm{CYB}(r) = \sum_{\alpha,\alpha'}([a_\alpha,a_{\alpha'}]\otimes b_\alpha \otimes b_{\alpha'} + a_\alpha \otimes [b_\alpha, a_{\alpha'}] \otimes b_{\alpha'} + a_\alpha \otimes a_{\alpha'} \otimes [b_\alpha, b_{\alpha'}])\otimes$

$\ell_\alpha \ell_{\alpha'}$, $dr := \sum_\alpha a_\alpha \otimes b_\alpha \otimes d\ell_\alpha$, where d extends $S(\mathfrak{h}) \to \mathfrak{h} \otimes S(\mathfrak{h})$, $x^k \mapsto kx \otimes x^{k-1}$, and $\mathrm{Alt}(X \otimes \ell) = (X + X^{2,3,1} + X^{3,1,2}) \otimes \ell$.

We also set

$$\psi(\lambda) := \left(\mathrm{id} \otimes (\mathrm{ad}\,\lambda^\vee)^{-2}_{|\mathfrak{n}}\right)(t_\mathfrak{n}).$$

We write $\psi(\lambda) = \sum_\alpha A_\alpha \otimes B_\alpha \otimes L_\alpha$.

Let $D(\mathfrak{h})[1/P]$ be the localization at P of the algebra $D(\mathfrak{h})$ of differential operators on $\mathfrak{h}$; the latter algebra is generated by $\bar{\mathrm{x}}_h$, $\bar{\partial}_h$, $h \in \mathfrak{h}$, with relations $h \mapsto \bar{\mathrm{x}}_h$, $h \mapsto \bar{\partial}_h$ linear, $[\bar{\mathrm{x}}_h, \bar{\mathrm{x}}_{h'}] = [\bar{\partial}_h, \bar{\partial}_{h'}] = 0$, and $[\bar{\partial}_h, \bar{\mathrm{x}}_{h'}] = \langle h, h' \rangle$.

Set $B_n := D(\mathfrak{h})[1/P] \otimes U(\mathfrak{g})^{\otimes n}$. For $h \in \mathfrak{h}$, we define $\bar{X}_h := \sum_\nu \bar{\mathrm{x}}_{[h,h_\nu]} \bar{\partial}_{h_\nu} \in D(\mathfrak{h})$, where $t_\mathfrak{h} = \sum_\nu h_\nu \otimes h_\nu$. We then set $\bar{Y}_h := \bar{X}_h + \sum_{i=1}^n h^{(i)}$. The map $\mathfrak{h} \to B_n$ is a Lie algebra morphism; we denote by $\mathfrak{h}^{\mathrm{diag}}$ its image.

We denote by $\mathcal{H}_n(\mathfrak{g}, \mathfrak{h})$ the Hecke algebra of B_n relative to $\mathfrak{h}^{\mathrm{diag}}$. Explicitly, $\mathcal{H}_n(\mathfrak{g}, \mathfrak{h}) = \{x \in B_n | \forall h \in \mathfrak{h}, \bar{Y}_h x \in B_n \mathfrak{h}^{\mathrm{diag}}\} / B_n \mathfrak{h}^{\mathrm{diag}}$.

Proposition 41. *There is a unique Lie algebra morphism*

$$\rho_{\mathfrak{g},\mathfrak{h}} : \bar{\mathfrak{t}}_{1,n} \to \mathcal{H}_n(\mathfrak{g}, \mathfrak{h})$$

such that $\bar{x}_i \mapsto \sum_\nu \bar{\mathrm{x}}_\nu \otimes h_\nu^{(i)}$, $\bar{y}_i \mapsto -\sum_\nu \bar{\partial}_\nu \otimes h_\nu^{(i)} + \sum_j \sum_\alpha \ell_\alpha \otimes a_\alpha^{(i)} b_\alpha^{(j)}$, $\bar{t}_{ij} \mapsto t_\mathfrak{g}^{(ij)}$. *Here* $r(\lambda) = \sum_\alpha \ell_\alpha(\lambda)(a_\alpha \otimes b_\alpha)$.

If $V_1, \ldots, V_n$ are $\mathfrak{g}$-modules, then $S(\mathfrak{h})[1/P] \otimes (\otimes_i V_i)$ is a module over $D(\mathfrak{h})[1/P] \otimes U(\mathfrak{g})^{\otimes n}$, and $(S(\mathfrak{h})[1/P] \otimes (\otimes_i V_i))^\mathfrak{h}$ is a module over $H_n(\mathfrak{g}, \mathfrak{h})$. Moreover, we have a restriction morphism $(S(\mathfrak{g}) \otimes (\otimes_i V_i))^\mathfrak{g} \to (S(\mathfrak{h})[1/P] \otimes (\otimes V_i))^\mathfrak{h}$. Note that $(S(\mathfrak{g}) \otimes (\otimes_i V_i))^\mathfrak{g}$ is a $\bar{\mathfrak{t}}_{1,n}$-module using the morphism $\bar{\mathfrak{t}}_{1,n} \to \mathcal{H}_n(\mathfrak{g})$, while $(S(\mathfrak{h})[1/P] \otimes (\otimes V_i))^\mathfrak{h}$ is a $\bar{\mathfrak{t}}_{1,n}$-module using the morphism $\bar{\mathfrak{t}}_{1,n} \to \mathcal{H}_n(\mathfrak{g}, \mathfrak{h})$. Then one checks that the restriction morphism $(S(\mathfrak{g}) \otimes (\otimes_i V_i))^\mathfrak{g} \to (S(\mathfrak{h})[1/P] \otimes (\otimes V_i))^\mathfrak{h}$ is a $\bar{\mathfrak{t}}_{1,n}$-module morphism.

Proof. The images of the above elements are all $\mathfrak{h}$-invariant. To lighten the notation, we will imply summation over repeated indices and denote elements of B_n as follows: $\bar{\partial}_\nu \otimes 1$ by $\bar{\partial}_\nu$, $\bar{\mathrm{x}}_\nu \otimes 1$ by $\langle \lambda, h_\nu \rangle$, $1 \otimes x^{(i)}$ by x^i. Then $\rho_{\mathfrak{g},\mathfrak{h}}(\bar{x}_i) = (\lambda^\vee)^i$, $\rho_{\mathfrak{g},\mathfrak{h}}(\bar{y}_i) = -h_\nu^i \bar{\partial}_\nu + \sum_{j=1}^n r(\lambda)^{ij}$ (here for $x \otimes y \in \mathfrak{g}^{\otimes 2}$, $(x \otimes y)^{ii} := x^i y^i$).

We will use the same presentation of $\bar{\mathfrak{t}}_{1,n}$ as in Proposition 39. The relations $[\bar{x}_i, \bar{x}_j] = 0$ and $\bar{t}_{ij} = \bar{t}_{ji}$ are obviously preserved.

Let us check that $[\bar{x}_i, \bar{y}_j] = \bar{t}_{ij}$ is preserved ($i \neq j$):

$$[\rho_{\mathfrak{g},\mathfrak{h}}(\bar{x}_i), \rho_{\mathfrak{g},\mathfrak{h}}(\bar{y}_j)] = \left[\bar{x}_\nu h_\nu^i, -h_\nu^j \bar{\partial}_\nu + \sum_k r(\lambda)^{jk}\right] = t_\mathfrak{h}^{ij} + [\lambda^i, r(\lambda)^{ji}]$$
$$= t_\mathfrak{h}^{ij} + t_\mathfrak{n}^{ij} = t_\mathfrak{g}^{ij} = \rho_{\mathfrak{g},\mathfrak{h}}(\bar{t}_{ij}).$$

Let us check that $\sum_i \bar{x}_i = \sum_i \bar{y}_i = 0$ are preserved. We have $\sum_i \rho_{\mathfrak{g},\mathfrak{h}}(\bar{x}_i) = 0$ by the same argument as above and $\sum_i \rho_{\mathfrak{g},\mathfrak{h}}(\bar{y}_i) = \sum_i (\lambda^\vee)^i$ (by the antisymmetry of $r(\lambda)$), which vanishes by the same argument as above.

Let us check that $[\bar{y}_i, \bar{y}_j] = 0$ is preserved, for $i \neq j$. We have

$$\begin{aligned}
&[\rho_{\mathfrak{g},\mathfrak{h}}(\bar{y}_i), \rho_{\mathfrak{g},\mathfrak{h}}(\bar{y}_j)] \\
&\quad = \sum_{k|k\neq i,j} \big(-h^i_\nu (\partial_\nu r(\lambda))^{jk} + h^j_\nu (\partial_\nu r(\lambda))^{ik} + [r(\lambda)^{ij}, r(\lambda)^{jk}] \\
&\qquad + [r(\lambda)^{ik}, r(\lambda)^{jk}] + [r(\lambda)^{ik}, r(\lambda)^{ji}]\big) + \left[\left(h^i_\nu + h^j_\nu\right)\bar{\partial}_\nu, r(\lambda)^{ij}\right] \\
&\qquad - \left[h^i_\nu \bar{\partial}_\nu, r(\lambda)^{jj}\right] + \left[h^j_\nu \bar{\partial}_\nu, r(\lambda)^{ii}\right] + [r(\lambda)^{ij}, r(\lambda)^{ii} + r(\lambda)^{jj}] \\
&\quad = \sum_{k|k\neq i,j} h^k_\nu (\partial_\nu r(\lambda))^{ij} + \left[\left(h^i_\nu + h^j_\nu\right)\bar{\partial}_\nu, r(\lambda)^{ij}\right] - \left[h^i_\nu \bar{\partial}_\nu, r(\lambda)^{jj}\right] \\
&\qquad + \left[h^j_\nu \bar{\partial}_\nu, r(\lambda)^{ii}\right] + [r(\lambda)^{ij}, r(\lambda)^{ii} + r(\lambda)^{jj}] \\
&\quad \equiv (\partial_\nu r(\lambda))^{ij}\left(-h^i_\nu - h^j_\nu - \bar{X}_\nu\right) + \left[\left(h^i_\nu + h^j_\nu\right)\bar{\partial}_\nu, r(\lambda)^{ij}\right] \\
&\qquad - h^i_\nu (\partial_\nu r(\lambda))^{jj} + h^j_\nu (\partial_\nu r(\lambda))^{ii} + [r(\lambda)^{ij}, r(\lambda)^{ii} + r(\lambda)^{jj}] \\
&\quad = \left[h^i_\nu + h^j_\nu, r(\lambda)^{ij}\right]\bar{\partial}_\nu - (\partial_\nu r^{ij}(\lambda))\bar{X}_\nu + \left[h^i_\nu + h^j_\nu, \partial_\nu r(\lambda)^{ij}\right] \\
&\qquad - h^i_\nu (\partial_\nu r(\lambda))^{jj} + h^j_\nu (\partial_\nu r(\lambda))^{ii} + [r(\lambda)^{ij}, r(\lambda)^{ii} + r(\lambda)^{jj}].
\end{aligned}$$

The second equality follows from the CDYBE and the antisymmetry on $r(\lambda)$. Then

$$\left[h^i_\nu + h^j_\nu, r(\lambda)^{ij}\right]\bar{\partial}_\nu - (\partial_\nu r^{ij}(\lambda))\bar{X}_\nu = \left(\left[h^i_{\nu'} + h^j_{\nu'}, r(\lambda)^{ij}\right] - \partial_\nu r^{ij}(\lambda)\langle\lambda, [h_\nu, h_{\nu'}]\rangle\right)\bar{\partial}_{\nu'}$$

is zero thanks to the $\mathfrak{h}$-invariance of $r(\lambda)$. Applying $x^i y^j z^k \mapsto x^i (yz)^j$ to the CDYB identity

$$\begin{aligned}
&\left[r(\lambda)^{ij}, r(\lambda)^{ik}\right] + \left[r(\lambda)^{ij}, r(\lambda)^{jk}\right] + \left[r(\lambda)^{ik}, r(\lambda)^{jk}\right] - h^i_\nu \partial_\nu r(\lambda)^{jk} + h^j_\nu \partial_\nu r(\lambda)^{ik} \\
&\qquad - h^j_\nu \partial_\nu r(\lambda)^{ij} = 0,
\end{aligned}$$

we get

$$\begin{aligned}
&\frac{1}{2} \sum_{\alpha,\beta} \ell_\alpha \ell'_\beta(\lambda)[a_\alpha, a_\beta]^i [b_\alpha, b_\beta]^j + [r(\lambda)^{ij}, r(\lambda)^{ii}] - h^i_\nu (\partial_\nu r(\lambda))^{jj} \\
&\qquad + \left[h^j_\nu, \partial_\nu r(\lambda)^{ij}\right] = 0.
\end{aligned}$$

Since $r(\lambda)$ is antisymmetric, the sum $(1/2)\sum_{\alpha,\beta}\dots$ is symmetric in (i,j); antisymmetrizing in (i,j), we get

$$\left[h^i_\nu + h^j_\nu, \partial_\nu r(\lambda)^{ij}\right] - h^i_\nu (\partial_\nu r(\lambda))^{jj} + h^j_\nu (\partial_\nu r(\lambda))^{ii} + [r(\lambda)^{ij}, r(\lambda)^{ii} + r(\lambda)^{jj}] = 0.$$

All this implies that $[\rho_{\mathfrak{g},\mathfrak{h}}(\bar{y}_i), \rho_{\mathfrak{g},\mathfrak{h}}(\bar{y}_j)] = 0$.

Let us check that $[\bar{x}_i, \bar{t}_{jk}] = 0$ is preserved (i,j,k distinct). We have $[\rho_{\mathfrak{g},\mathfrak{h}}(\bar{x}_i), \rho_{\mathfrak{g},\mathfrak{h}}(\bar{t}_{jk})] = \left[(\lambda^\vee)^i, t^{jk}_{\mathfrak{g}}\right] = 0$.

Let us prove that $[\bar{y}_i, \bar{t}_{jk}] = 0$ is preserved (i,j,k distinct). We have $[\rho_{\mathfrak{g},\mathfrak{h}}(\bar{y}_i), \rho_{\mathfrak{g},\mathfrak{h}}(\bar{t}_{jk})] = \left[-h^i_\nu \bar{\partial}_\nu + \sum_l r(\lambda)^{il}, t^{jk}_{\mathfrak{g}}\right] = \left[r(\lambda)^{ij} + r(\lambda)^{ik}, t^{jk}_{\mathfrak{g}}\right] = 0$ because $t_{\mathfrak{g}}$ is $\mathfrak{g}$-invariant. $\square$

Proposition 42. *If $V_1, \dots, V_n$ are $\mathfrak{g}$-modules, then $(S(\mathfrak{h})[1/P] \otimes (\otimes_i V_i))^{\mathfrak{h}}$ is a $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}$-module. The $\bar{\mathfrak{t}}_{1,n}$-module structure is induced by the morphism $\bar{\mathfrak{t}}_{1,n} \to \mathcal{H}_n(\mathfrak{g}, \mathfrak{h})$ of Proposition 41, so*

$$\rho_{(V_i)}(\bar{x}_i)(f(\lambda) \otimes (\otimes_i v_i)) = (\lambda^\vee)^i (f(\lambda) \otimes (\otimes_i v_i)),$$
$$\rho_{(V_i)}(\bar{y}_i)(f(\lambda) \otimes (\otimes_i v_i)) = \left(-h_\nu^i \partial_\nu + \sum_j r(\lambda)^{ij}\right)(f(\lambda) \otimes (\otimes_i v_i)),$$
$$\rho_{(V_i)}(\bar{t}_{ij})(f(\lambda) \otimes (\otimes_i v_i)) = t_{\mathfrak{g}}^{ij}(f(\lambda) \otimes (\otimes_i v_i)),$$

and the $\mathfrak{d}$-module structure is given by

$$\rho_{(V_i)}(\delta_{2m})(f(\lambda) \otimes (\otimes_i v_i)) = \frac{1}{2}\left(\sum_i \{(\mathrm{ad}\,\lambda^\vee)^{2m}(e_\alpha) \cdot e_\alpha\}^i\right)(f(\lambda) \otimes (\otimes_i v_i)),$$
$$\rho_{(V_i)}(\Delta_0)(f(\lambda) \otimes (\otimes_i v_i)) = \Bigg(-\frac{1}{2}\partial_\nu^2 + \frac{1}{2}\langle \mu(r(\lambda)), h_\nu \rangle \partial_\nu$$
$$+ \left\{\frac{1}{2}\psi(\lambda)^{11} - \frac{1}{2}(\mathrm{ad}\,\lambda^\vee)_{|\mathfrak{n}}^{-1}(\mu(r(\lambda))_{\mathfrak{n}})\right\}^{12\dots n}\Bigg)$$
$$(f(\lambda) \otimes (\otimes_i v_i)),$$
$$\rho_{(V_i)}(d)(f(\lambda) \otimes (\otimes_i v_i)) = \frac{1}{2}(\langle \lambda, h_\nu \rangle \partial_\nu + \partial_\nu \langle \lambda, h_\nu \rangle + \langle \mu(r(\lambda)), \lambda^\vee \rangle)$$
$$(f(\lambda) \otimes (\otimes_i v_i)),$$
$$\rho_{(V_i)}(X)(f(\lambda) \otimes (\otimes_i v_i)) = \frac{1}{2}\langle \lambda^\vee, \lambda^\vee \rangle (f(\lambda) \otimes (\otimes_i v_i)).$$

Here $(-)_{\mathfrak{n}}$ denotes the projection of $\mathfrak{g}$ on $\mathfrak{n}$ along $\mathfrak{h}$.

To summarize, we have a diagram

$$\begin{tikzcd} \bar{\mathfrak{t}}_{1,n} \arrow[r] \arrow[dr, hook] & \mathcal{H}_n(\mathfrak{g}, \mathfrak{h}) \arrow[r] & \mathrm{End}((S(\mathfrak{h})[1/P] \otimes (\otimes_i V_i))^{\mathfrak{h}}) \\ & \bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d} \arrow[u, "(1)"] \arrow[ur] & \end{tikzcd}$$

As before, the restriction morphism $(S(\mathfrak{g}) \otimes (\otimes_i V_i))^{\mathfrak{g}} \to (S(\mathfrak{h})[1/P] \otimes (\otimes_i V_i))^{\mathfrak{h}}$ extends to a $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}$-module morphism.

The action of $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}$ factors through a morphism $\tilde{\rho}_{\mathfrak{g},\mathfrak{h}} : \bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d} \to \mathcal{H}_n(\mathfrak{g}, \mathfrak{h})$ extending $\rho_{\mathfrak{g},\mathfrak{h}} : \bar{\mathfrak{t}}_{1,n} \to \mathcal{H}_n(\mathfrak{g}, \mathfrak{h})$ (denoted by (1) in the diagram).

Proof. Let $\lambda \in \mathfrak{h}^*_{\mathrm{reg}}$. Then if V is a $\mathfrak{g}$-module, we have $(\hat{\mathcal{O}}_{\mathfrak{g}^*,\lambda} \otimes V)^{\mathfrak{g}} = (\hat{\mathcal{O}}_{\mathfrak{h}^*,\lambda} \otimes V)^{\mathfrak{h}}$ (where $\hat{\mathcal{O}}_{X,x}$ is the completed local ring of a variety X at the point x). We then have a morphism $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d} \to \mathcal{H}_n(\mathfrak{g}) \to \mathrm{End}((\hat{\mathcal{O}}_{\mathfrak{g}^*,\lambda} \otimes (\otimes_i V_i))^{\mathfrak{g}})$ for any $\lambda \in \mathfrak{g}^*$, so when $\lambda \in \mathfrak{h}^*_{\mathrm{reg}}$ we get a morphism $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d} \to \mathrm{End}((\hat{\mathcal{O}}_{\mathfrak{h}^*,\lambda} \otimes (\otimes_i V_i))^{\mathfrak{h}})$.

Let show that the images of the generators of $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}$ under this morphism are given by the above formulas.

Since the actions of $\bar{x}_i$, $\bar{t}_{ij}$, and X on $(\hat{\mathcal{O}}_{\mathfrak{g}^*,\lambda} \otimes (\otimes_i V_i))^{\mathfrak{g}}$ are given by multiplication by elements of $(\hat{\mathcal{O}}_{\mathfrak{g}^*,\lambda} \otimes U(\mathfrak{g})^{\otimes n})^{\mathfrak{g}}$, their actions on $(\hat{\mathcal{O}}_{\mathfrak{h}^*,\lambda} \otimes (\otimes_i V_i))^{\mathfrak{h}}$ are given by multiplication by restrictions of these elements to $\mathfrak{h}^*$.

Let us compute the action of $\bar{y}_i$. Let $\tilde{f}(\lambda) \in (\hat{\mathcal{O}}_{\mathfrak{h}^*,\lambda} \otimes (\otimes_i V_i))^{\mathfrak{h}}$ and $\tilde{F}(\lambda) \in (\hat{\mathcal{O}}_{\mathfrak{g}^*,\lambda} \otimes (\otimes_i V_i))^{\mathfrak{g}}$ be its equivariant extension to a formal map $\mathfrak{g}^* \to \otimes_i V_i$. Then for $x \in \mathfrak{n}$, we have $(\partial_{x^\wedge} + \sum_i (\mathrm{ad}\,\lambda^\vee)^{-1}(x)^i)(\tilde{F}(\lambda))_{|\mathfrak{h}^*} = 0$ (the map $x \mapsto x^\wedge$ is the inverse of $\mathfrak{g}^* \to \mathfrak{g}$, $\lambda \mapsto \lambda^\vee$). Then $\rho_{(V_i)}(\bar{y}_i)(\tilde{f}(\lambda)) = \Big(- h_\nu^i \partial_\nu + \sum_j e_\beta^i \big((\mathrm{ad}\,\lambda^\vee)^{-1}(e_\beta) \big)^j \Big) \tilde{f}(\lambda) = \Big(-h_\nu^i \partial_\nu + \sum_j r(\lambda)^{ij} \Big) (\tilde{f}(\lambda))$.

Let us now compute the action of Δ_0. Let $\lambda_0 \in \mathfrak{h}^*$ be such that $\lambda_0^\vee \in U$ and $\lambda \in \mathfrak{g}^*$ be close to λ_0. We set $\delta\lambda := \lambda - \lambda_0$. We then have $\lambda = e^{\mathrm{ad}\,x}(\lambda_0 + h^\wedge)$, where $x \in \mathfrak{n}$ and $h \in \mathfrak{h}$ are close to 0. We have the expansions

$$h = (\delta\lambda)_{\mathfrak{h}}^\vee + \frac{1}{2} \Big[(\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} \left((\delta\lambda)_{\mathfrak{n}}^\vee \right), (\delta\lambda)_{\mathfrak{n}}^\vee \Big]_{\mathfrak{h}},$$

$$\begin{aligned} x = &- (\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} \left((\delta\lambda)_{\mathfrak{n}}^\vee + \left[(\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} \left((\delta\lambda)_{\mathfrak{n}}^\vee \right), (\delta\lambda)_{\mathfrak{h}}^\vee \right] \right. \\ &\left. + \frac{1}{2} \Big[(\mathrm{ad}\,\lambda_0^\vee)_{|\mathfrak{n}}^{-1} \left((\delta\lambda)_{\mathfrak{n}}^\vee \right), (\delta\lambda)_{\mathfrak{n}}^\vee \Big]_{\mathfrak{n}} \right) \end{aligned}$$

up to terms of order > 2; here the indices $u_{\mathfrak{n}}$ and $u_{\mathfrak{h}}$ mean the projections of $u \in \mathfrak{g}$ to $\mathfrak{n}$ and $\mathfrak{h}$. If now $\tilde{f}(\lambda) : \mathfrak{h}^* \supset V(\lambda_0, \mathfrak{h}^*) \to \otimes_i V_i$ is an $\mathfrak{h}$-equivariant function defined in the vicinity of λ_0 and $\tilde{F}(\lambda) : \mathfrak{g}^* \supset V(\lambda_0, \mathfrak{g}^*) \to \otimes_i V_i$ is its $\mathfrak{g}$-equivariant extension to a neighborhood of λ_0 in $\mathfrak{g}^*$, then $\tilde{F}(\lambda) = (e^x)^{1\ldots n} \tilde{f}(\lambda_0 + h)$, which implies the expansion

$$\begin{aligned} \tilde{F}(\lambda) = \tilde{f}(\lambda_0) &+ \Big((\delta\lambda)_\nu + \frac{1}{2} \Big\langle \Big[(\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} (e_\beta), e_{\beta'} \Big], h_\nu \Big\rangle (\delta\lambda)_\beta (\delta\lambda)_{\beta'} \Big) \partial_\nu \tilde{f}(\lambda_0) \\ &+ \frac{1}{2} (\delta\lambda)_\nu (\delta\lambda)_{\nu'} \partial^2_{\nu\nu'} \tilde{f}(\lambda_0) + \Big(- (\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} (e_\beta)(\delta\lambda)_\beta \\ &- (\mathrm{ad}\ \lambda_0^\vee)^{-1} \Big(\Big[(\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} (e_\beta), h_\nu \Big] \Big) (\delta\lambda)_\nu (\delta\lambda)_\beta \\ &- \frac{1}{2} (\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} \Big(\Big[(\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} (e_\beta), e_{\beta'} \Big]_{\mathfrak{n}} \Big) (\delta\lambda)_\beta (\delta\lambda)_{\beta'} \\ &+ \frac{1}{2} (\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} (e_\beta) (\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} (e_{\beta'}) (\delta\lambda)_\beta (\delta\lambda)_{\beta'} \Big)^{1\ldots n} \tilde{f}(\lambda_0) \\ &- (\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} (e_\beta)^{1\ldots n} (\delta\lambda)_\beta (\delta\lambda)_\nu \partial_\nu \tilde{f}(\lambda_0) \end{aligned}$$

up to terms of order > 2.

Then

$$\begin{aligned} \left(\partial_\alpha^2 F \right)(\lambda_0) = \left(\partial_\nu^2 \tilde{f} \right)(\lambda_0) &+ \Big\langle \Big[(\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} (e_\beta), e_\beta \Big], h_\nu \Big\rangle \partial_\nu \tilde{f}(\lambda_0) \\ &+ \Big(- (\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} \Big(\Big[(\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} (e_\beta), e_\beta \Big]_{\mathfrak{n}} \Big) \\ &+ \Big((\mathrm{ad}\ \lambda_0^\vee)_{|\mathfrak{n}}^{-1} (e_\beta) \Big)^2 \Big)^{1\ldots n} \tilde{f}(\lambda_0), \end{aligned}$$

which implies the formula for the action of Δ_0.

Then $(S(\mathfrak{h})[1/P] \otimes (\otimes_i V_i))^{\mathfrak{h}} \subset \prod_{\lambda \in \mathfrak{h}^*_{\mathrm{reg}}} (\hat{\mathcal{O}}_{\mathfrak{h}^*,\lambda} \otimes (\otimes_i V_i))^{\mathfrak{h}}$ is preserved by the action of the generators of $\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}$-module, hence it is a sub-$(\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d})$-module, with action given by the above formulas. □

6.4 Realization of the universal KZB system

The realization of the flat connection $d - \sum_i \bar{K}_i(\mathbf{z}|\tau)dz_i - \bar{\Delta}(\mathbf{z}|\tau)d\tau$ on $(\mathfrak{H} \times \mathbb{C}^n) - \mathrm{Diag}_n$ is a flat connection on the trivial bundle with fiber $\left(\mathcal{O}_{\mathfrak{h}^*_{\mathrm{reg}}} \otimes (\otimes_i V_i)\right)^{\mathfrak{h}}$.

We now compute this realization, under the assumption that $\mathfrak{h} \subset \mathfrak{g}$ is a maximal abelian subalgebra. In this case, two simplifications occur:

(a) $(\mathrm{ad}\,\lambda^\vee)(h_\nu) = 0$ since $\mathfrak{h}$ is abelian,

(b) $\left[(\mathrm{ad}\,\lambda^\vee)^{-1}_{|\mathfrak{n}}(e_\beta), e_\beta\right]_{\mathfrak{n}} = 0$, since $\left[(\mathrm{ad}\,\lambda^\vee)^{-1}_{|\mathfrak{n}}(e_\beta), e_\beta\right]$ commutes with any element in $\mathfrak{h}$, so that it belongs to $\mathfrak{h}$.

The image of $\bar{K}_i(\mathbf{z}|\tau)$ is then the operator

$$\begin{aligned} &K_i^{(V_i)}(\mathbf{z}|\tau) \\ &\quad = h^i_\nu \partial_\nu - \sum_j r(\lambda)^{ij} + \sum_{j|j\neq i} k(z_{ij}, (\mathrm{ad}\,\lambda^\vee)^i|\tau)\left(t^{ij}_{\mathfrak{n}} + t^{ij}_{\mathfrak{h}}\right) \\ &\quad = h^i_\nu \partial_\nu - r(\lambda)^{ii} + \sum_{j|j\neq i} \frac{\theta(z_{ij} + (\mathrm{ad}\,\lambda^\vee)^i|\tau)}{\theta(z_{ij}|\tau)\theta((\mathrm{ad}\,\lambda^\vee)^i|\tau)}\left(t^{ij}_{\mathfrak{n}}\right) + \sum_{j|j\neq i} \frac{\theta'}{\theta}(z_{ij}|\tau) t^{ij}_{\mathfrak{h}}. \end{aligned}$$

The image of $2\pi \mathrm{i}\bar{\Delta}(\mathbf{z}|\tau)$ is the operator

$$\begin{aligned} 2\pi\mathrm{i}\Delta^{(V_i)}(\mathbf{z}|\tau) &= \frac{1}{2}\partial_\nu^2 + \frac{1}{2}\left\langle \left[(\mathrm{ad}\,\lambda^\vee)^{-1}(e_\beta), e_\beta\right], h_\nu \right\rangle \partial_\nu - g(0,0|\tau)\sum_i \frac{1}{2} t^{ii}_{\mathfrak{g}} \\ &+ \sum_{i,j} \frac{1}{2}\left(\left[g(z_{ij}, \mathrm{ad}\,\lambda^\vee|\tau) - (\mathrm{ad}\,\lambda^\vee)^{-2}\right](e_\beta)\right)^i e^j_\beta + \sum_{i,j} \frac{1}{2} g(z_{ij}, 0|\tau) h^i_\nu h^j_\nu, \end{aligned}$$

and the connection is now

$$\nabla^{(V_i)} = d - \sum_i K_i^{(V_i)}(\mathbf{z}|\tau)dz_i - \Delta^{(V_i)}(\mathbf{z}|\tau)d\tau.$$

Recalling $P(\lambda) = \det((\mathrm{ad}\,\lambda^\vee)_{|\mathfrak{n}})$, we compute the conjugations $P^{1/2}\nabla^{(V_i)} P^{-1/2}$, where $P^{\pm 1/2}$ is the operator of multiplication by (inverse branches of) $P^{\pm 1/2}$ on $\mathcal{O}_{\mathfrak{h}^*_{\mathrm{reg}}} \otimes (\otimes_i V_i)^{\mathfrak{h}}$.

Lemma 43. $\partial_\nu \log P(\lambda) = -\langle h_\nu, \mu(r(\lambda))\rangle$, $P^{1/2}\left[h^i_\nu \partial_\nu - r(\lambda)^{ii}\right] P^{-1/2} = h^i_\nu \partial_\nu$, $P^{1/2}\left[\partial_\nu^2 + \left\langle \left[(\mathrm{ad}\,\lambda^\vee)^{-1}_{|\mathfrak{n}}(e_\beta), e_\beta\right], h_\nu \right\rangle \partial_\nu\right] P^{-1/2} = \partial_\nu^2 + \partial_\nu\left(\langle h_\nu, \frac{1}{2}\mu(r(\lambda))\rangle\right) - \left\langle h_\nu, \frac{1}{2}\mu(r(\lambda))\right\rangle^2$.

Proof. $\partial_\nu \log P(\lambda) = (d/dt)_{|t=0}\det\left[(\mathrm{ad}(\lambda^\vee + th_\nu)_{|\mathfrak{n}})(\mathrm{ad}\,\lambda^\vee)^{-1}_{|\mathfrak{n}}\right] = \mathrm{tr}\left[(\mathrm{ad}h_\nu)_{|\mathfrak{n}}\circ (\mathrm{ad}\,\lambda^\vee)^{-1}_{|\mathfrak{n}}\right] = \left\langle e_\beta, (\mathrm{ad}h_\nu)\circ(\mathrm{ad}\,\lambda^\vee)^{-1}_{|\mathfrak{n}}(e_\beta)\right\rangle = \left\langle \left[(\mathrm{ad}\,\lambda^\vee)^{-1}_{|\mathfrak{n}}(e_\beta), e_\beta\right], h_\nu\right\rangle = -\langle h_\nu, \mu(r(\lambda))\rangle$. The next equality follows from $\mu(r(\lambda))^i = 2r(\lambda)^{ii}$. The last equality is a direct consequence. □

Proposition 44. *$P^{1/2}\nabla^{(V_i)}P^{-1/2} = d - \sum_i \tilde{K}_i(\mathbf{z}|\tau)dz_i - \tilde{\Delta}(\mathbf{z}|\tau)d\tau$, where*

$$\tilde{K}_i(\mathbf{z}|\tau) = h_\nu^i\partial_\nu + \sum_{j|j\neq i}\frac{\theta(z_{ij} + (\mathrm{ad}\,\lambda^\vee)^i|\tau)}{\theta(z_{ij}|\tau)\theta((\mathrm{ad}\,\lambda^\vee)^i|\tau)}\left(t^{ij}_{\mathfrak{n}}\right) + \sum_{j|j\neq i}\frac{\theta'}{\theta}(z_{ij}|\tau)t^{ij}_{\mathfrak{h}}$$

$$\begin{aligned}2\pi\mathrm{i}\tilde{\Delta}(\mathbf{z}|\tau) = {} & \frac{1}{2}\partial_\nu^2 + \partial_\nu\left(\left\langle h_\nu, \frac{1}{2}\mu(r(\lambda))\right\rangle\right) - \left\langle h_\nu, \frac{1}{2}\mu(r(\lambda))\right\rangle^2 - g(0,0|\tau)\sum_i \frac{1}{2}t^{ii}_{\mathfrak{g}}\\ & + \sum_{i,j}\frac{1}{2}\Big(\big(g(z_{ij}, \mathrm{ad}\ \lambda^\vee|\tau) - (\mathrm{ad}\,\lambda^\vee)^{-2}\big)(e_\beta)\Big)^i e^j_\beta\\ & + \sum_{i,j}\frac{1}{2}g(z_{ij},0|\tau)h^i_\nu h^j_\nu,\end{aligned}$$

where

$$g(z,0|\tau) = \frac{1}{2}\frac{\theta''}{\theta}(z|\tau) - 2\pi\mathrm{i}\frac{\partial_\tau\eta}{\eta}(\tau)$$

and

$$g(z,\alpha|\tau) - \alpha^{-2} = \frac{1}{2}\frac{\theta(z+\alpha|\tau)}{\theta(x|\tau)\theta(\alpha|\tau)}\left(\frac{\theta'}{\theta}(z+\alpha|\tau) - \frac{\theta'}{\theta}(\alpha|\tau)\right).$$

The term in $\sum_i(1/2)t^{ii}_{\mathfrak{g}}$ is central and can be absorbed by a suitable further conjugation. Rescaling $t_{\mathfrak{g}}$ into $\kappa^{-1}t_{\mathfrak{g}}$, where $\kappa \in \mathbb{C}^\times$, $\tilde{K}_i(\mathbf{z}|\tau)$ and $\tilde{\Delta}(\mathbf{z}|\tau)$ get multiplied by κ. Moreover, we have:

Lemma 45. *When $\mathfrak{g}$ is simple and $\mathfrak{h}\subset\mathfrak{g}$ is the Cartan subalgebra,*

$$\partial_\nu\left\{\left\langle h_\nu, \frac{1}{2}\mu(r(\lambda))\right\rangle\right\} = \left\langle h_\nu, \frac{1}{2}\mu(r(\lambda))\right\rangle^2.$$

Proof. Let $D(\lambda) := \prod_{\alpha\in\Delta^+}(\alpha,\lambda)$, where Δ^+ is the set of positive roots of $\mathfrak{g}$. Then $D(\lambda)$ is W-anti-invariant, where W is the Weyl group. Therefore $\partial_\nu^2 D(\lambda)$ is also W-anti-invariant, so it is divisible (as a polynomial on $\mathfrak{h}^*$) by all the (α,λ), where $\alpha\in\Delta^+$, so it is divisible by $D(\lambda)$; since $\partial^2_\nu D(\lambda)$ has degree strictly lower than $D(\lambda)$, we get $\partial^2_\nu D(\lambda) = 0$.

Now if $(e_\alpha, f_\alpha, h_\alpha)$ is a basis of the $\mathfrak{sl}_2$-triple associated with α, we have $r(\lambda) = \sum_{\alpha\in\Delta^+} -(e_\alpha\otimes f_\alpha - f_\alpha\otimes e_\alpha)/(\alpha,\lambda)$, so $\frac{1}{2}\mu(r(\lambda)) = -\sum_{\alpha\in\Delta^+} h_\alpha/(\alpha,\lambda)$. Therefore $\frac{1}{2}\mu(r(\lambda)) = -\partial_\nu \log D(\lambda) h_\nu$. Then $\partial^2_\nu D(\lambda) = 0$ implies that $\partial^2_\nu \log D + (\partial_\nu \log D)^2 = 0$, which implies the lemma. □

The resulting flat connection then coincides with that of [Ber98a, FW96].

7 The universal KZB connection and representations of Cherednik algebras

7.1 The rational Cherednik algebra of type A_{n-1}

Let k be a complex number, and $n \geq 1$ an integer. The rational Cherednik algebra $H_n(k)$ of type A_{n-1} is the quotient of the algebra $\mathbb{C}[S_n] \ltimes \mathbb{C}\langle \mathrm{x}_1, \ldots, \mathrm{x}_n, \mathrm{y}_1, \ldots, \mathrm{y}_n \rangle$ by the relations

$$\sum_i \mathrm{x}_i = 0, \ \sum_i \mathrm{y}_i = 0, \ [\mathrm{x}_i, \mathrm{x}_j] = 0 = [\mathrm{y}_i, \mathrm{y}_j],$$
$$[\mathrm{x}_i, \mathrm{y}_j] = \frac{1}{n} - k s_{ij}, \ i \neq j,$$

where $s_{ij} \in S_n$ is the permutation of i and j (see, e.g., [EG02]).[8]

Let $e := \frac{1}{n!} \sum_{\sigma \in S_n} \sigma \in \mathbb{C}[S_n]$ be the Young symmetrizer. The spherical subalgebra $B_n(k)$ (often called the spherical Cherednik algebra) is defined to be the algebra $eH_n(k)e$.

We define an important element

$$\mathbf{h} := \frac{1}{2} \sum_i (\mathrm{x}_i \mathrm{y}_i + \mathrm{y}_i \mathrm{x}_i).$$

We recall that category $\mathcal{O}$ is the category of $H_n(k)$-modules that are locally nilpotent under the action of the operators y_i and decompose into a direct sum of finite dimensional generalized eigenspaces of $\mathbf{h}$. Similarly, one defines the category $\mathcal{O}$ over $B_n(k)$ to be the category of $B_n(k)$-modules that are locally nilpotent under the action of $\mathbb{C}[\mathrm{y}_1, \ldots, \mathrm{y}_n]^{S_n}$ and decompose into a direct sum of finite dimensional generalized eigenspaces of $\mathbf{h}$.

7.2 The homomorphism from $\bar{\mathfrak{t}}_{1,n}$ to the rational Cherednik algebra

Proposition 46. *For each $k, a, b \in \mathbb{C}$, we have a homomorphism of Lie algebras $\xi_{a,b} : \bar{\mathfrak{t}}_{1,n} \to H_n(k)$, defined by the formula*

$$\bar{x}_i \mapsto a\mathrm{x}_i, \quad \bar{y}_i \mapsto b\mathrm{y}_i, \quad \bar{t}_{ij} \mapsto ab\left(\frac{1}{n} - k s_{ij}\right).$$

Proof. Straightforward. □

Remark 47. Obviously, a, b can be rescaled independently, by rescaling the generators $\bar{x}_i$ and $\bar{y}_i$ of the source algebra $\bar{\mathfrak{t}}_{1,n}$. On the other hand, if we are allowed only to apply automorphisms of the target algebra $H_n(k)$, then a, b can be rescaled only in such a way that the product ab is preserved.

[8] The generators $\mathrm{x}_\alpha, \partial_\alpha$ of Section 6.1 will be henceforth renamed q_α, p_α.

This shows that any representation V of the rational Cherednik algebra $H_n(k)$ yields a family of realizations for $\bar{\mathfrak{t}}_{1,n}$ parametrized by $a, b \in \mathbb{C}$, and gives rise to a family of flat connections $\nabla_{a,b}$ over the configuration space $\bar{C}(E_\tau, n)$.

7.3 Monodromy representations of double affine Hecke algebras

Let $\mathcal{H}_n(q,t)$ be Cherednik's double affine Hecke algebra of type A_{n-1}. By definition, $\mathcal{H}_n(q,t)$ is the quotient of the group algebra of the orbifold fundamental group $\overline{\mathrm{B}}_{1,n}$ of $\bar{C}(E_\tau, n)/S_n$ by the additional relations

$$(T - q^{-1}t)(T + q^{-1}t^{-1}) = 0,$$

where T is any element of $\overline{\mathrm{B}}_{1,n}$ homotopic (as a free loop) to a small loop around the divisor of diagonals in the counterclockwise direction.

Let V be a representation of $H_n(k)$, and let $\nabla_{a,b}(V)$ be the universal connection $\nabla_{a,b}$ evaluated in V. In some cases, for example if a, b are formal or if V is finite-dimensional, we can consider the monodromy of this connection, which obviously gives a representation of $\mathcal{H}_n(q,t)$ on V, with

$$q = e^{-2\pi \mathrm{i} ab/n}, \quad t = e^{-2\pi \mathrm{i} kab}.$$

In particular, taking $a = b$, $V = H_n(k)$, this monodromy representation defines a homomorphism $\theta_a : \mathcal{H}_n(q,t) \to H_n(k)[[a]]$, where

$$q = e^{-2\pi \mathrm{i} a^2/n}, \quad t = e^{-2\pi \mathrm{i} ka^2}.$$

It is easy to check that this homomorphism becomes an isomorphism upon inverting a. The existence of such an isomorphism was pointed out by Cherednik (see [Che03, end of Section 6], and the end of [Che97]), but his proof is different.

Example 48. Let $k = r/n$, where r is an integer relatively prime to n. In this case, it is known (see, e.g., [BEG03a]) that the algebra $H_n(k)$ admits an irreducible finite dimensional representation $Y(r,n)$ of dimension r^{n-1}. By virtue of the above construction, the space $Y(r,n)$ carries an action of $\mathcal{H}_n(q,t)$ with any nonzero q, t such that $q^r = t$. This finite-dimensional representation of $\mathcal{H}_n(q,t)$ is irreducible for generic q, and is called a perfect representation; it was first constructed in [Eti94, p. 500], and later in [Che03, Theorem 6.5], in a greater generality.

7.4 The modular extension of $\xi_{a,b}$

Assume that $a, b \neq 0$.

Proposition 49. *The homomorphism $\xi_{a,b}$ can be extended to the algebra $U(\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}) \rtimes S_n$ by the formulas*

$$\xi_{a,b}(s_{ij}) = s_{ij},$$
$$\xi_{a,b}(d) = \mathbf{h} = \frac{1}{2}\sum_i (\mathrm{x}_i\mathrm{y}_i + \mathrm{y}_i\mathrm{x}_i), \quad \xi_{a,b}(X) = -\frac{1}{2}ab^{-1}\sum_i \mathrm{x}_i^2,$$
$$\xi_{a,b}(\Delta_0) = \frac{1}{2}ba^{-1}\sum_i \mathrm{y}_i^2, \quad \xi_{a,b}(\delta_{2m}) = -\frac{1}{2}a^{2m-1}b^{-1}\sum_{i<j}(\mathrm{x}_i - \mathrm{x}_j)^{2m}.$$

Proof. Direct computation. □

Thus, the flat connections $\nabla_{a,b}$ extend to flat connections on $\mathcal{M}_{1,[n]}$.

This shows that the monodromy representation of the connection $\nabla_{a,b}(V)$, when it can be defined, is a representation of the double affine Hecke algebra $\mathcal{H}_n(q,t)$ with a compatible action of the extended modular group $\widetilde{\mathrm{SL}_2(\mathbb{Z})}$. In particular, this is the case if $V = Y(r,n)$. Such representations of $\widetilde{\mathrm{SL}_2(\mathbb{Z})}$ were considered by Cherednik [Che03]. The element T of $\widetilde{\mathrm{SL}_2(\mathbb{Z})}$ acts in this representation by "the Gaussian," and the element S by the "Fourier–Cherednik transform." They are generalizations of the $\widetilde{\mathrm{SL}_2(\mathbb{Z})}$-action on Verlinde algebras.

8 Explicit realizations of certain highest weight representations of the rational Cherednik algebra of type A_{n-1}

8.1 The representation V_N

Let N be a divisor of n, and $\mathfrak{g} = \mathfrak{sl}_N(\mathbb{C})$, $G = \mathrm{SL}_N(\mathbb{C})$. Let $V_N = (\mathbb{C}[\mathfrak{g}] \otimes (\mathbb{C}^N)^{\otimes n})^{\mathfrak{g}}$ (the divisor condition is needed for this space to be nonzero). It turns out that V_N has a natural structure of a representation of $H_n(k)$ for $k = N/n$.

Proposition 50. *We have a homomorphism $\zeta_N : H_n(N/n) \to \mathrm{End}(V_N)$, defined by the formulas*

$$\zeta_N(s_{ij}) = s_{ij}, \quad \zeta_N(\mathrm{x}_i) = X_i, \quad \zeta_N(\mathrm{y}_i) = Y_i \quad (i = 1, \ldots, n),$$

where for $f \in V_N, A \in \mathfrak{g}$ we have

$$(X_i f)(A) = A_i f(A),$$
$$(Y_i f)(A) = \frac{N}{n}\sum_p (b_p)_i \frac{\partial f}{\partial b_p}(A),$$

where $\{b_p\}$ is an orthonormal basis of $\mathfrak{g}$ with respect to the trace form.

Proof. Straightforward verification. □

The relationship of the representation V_N to other results in this paper is described by the following proposition.

Proposition 51. *The connection $\nabla_{a,1}(V_N)$ corresponding to the representation V_N is the usual KZB connection for the n-point correlation functions on the elliptic curve for the Lie algebra $\mathfrak{sl}_N$ and n copies of the vector representation $\mathbb{C}^N$, at level $K = -\frac{n}{aN} - N$.*

Proof. We have a sequence of maps

$$U(\bar{\mathfrak{t}}_{1,n} \rtimes \mathfrak{d}) \rtimes S_n \to H_n(N/n) \to \mathcal{H}_n(\mathfrak{g}) \rtimes S_n \to \mathrm{End}(V_N),$$

where the first map is $\xi_{a,b}$, the second map sends s_{ij} to s_{ij}, x_i to the class of $\sum_\alpha q_\alpha \otimes e_\alpha^i$, and y_i to the class of $\sum_\alpha p_\alpha \otimes e_\alpha^i$ (recall that the x_a, ∂_a of Section 6.1 have been renamed q_a, p_a), and the last map is explained in Section 6.1. The composition of the first two maps is then that of Proposition 40, and the composition of the last two maps is the map ζ_N of Proposition 50. This implies the statement. □

Remark 52. Suppose that K is a nonnegative integer, i.e., $a = -\frac{n}{N(K+N)}$, where $K \in \mathbb{Z}_+$. Then the connection $\nabla_{a,1}$ on the infinite-dimensional vector bundle with fiber V_N preserves a finite-dimensional subbundle of conformal blocks for the WZW model at level K. The subbundle gives rise to a finite dimensional monodromy representation V_N^K of the Cherednik algebra $\mathcal{H}_n(q,t)$ with

$$q = e^{\frac{2\pi \mathrm{i}}{N(K+N)}}, \quad t = q^N$$

(so both parameters are roots of unity). The dimension of V_N^K is given by the Verlinde formula, and it carries a compatible action of $\mathrm{SL}_2(\mathbb{Z})$ to the action of the Cherednik algebra. Representations of this type were studied by Cherednik in [Che03].

8.2 The spherical part of V_N.

Note that

$$\left(\left(\sum_{i=1}^n X_i^p\right) f\right)(A) = \frac{n}{N}(\mathrm{tr}\, A^p) f(A), \tag{43}$$

$$\left(\left(\sum_{i=1}^n Y_i^p\right) f\right)(A) = \left(\frac{N}{n}\right)^{p-1} (\mathrm{tr}\, \partial_A^p)\, f(A). \tag{44}$$

Consider the space $U_N = eV_N = (\mathbb{C}[\mathfrak{g}] \otimes S^n\mathbb{C}^N)^{\mathfrak{g}}$ as a module over the spherical subalgebra $B_n(k)$. It is known (see e.g. [BEG03b]) that the spherical subalgebra is generated by the elements $(\sum \mathrm{x}_i^p)\, e$ and $(\sum \mathrm{y}_i^p)\, e$. Thus formulas (43), (44) determine the action of $B_n(k)$ on U_N.

We note that by restriction to the set $\mathfrak{h}$ of diagonal matrices $\mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, and dividing by $\Delta^{n/N}$, where $\Delta = \prod_{i<j}(\lambda_i - \lambda_j)$, one identifies U_N with $\mathbb{C}[\mathfrak{h}]^{S_N}$. Moreover, it follows from [EG02] that formulas (43), (44) can be viewed as defining an action of another spherical Cherednik algebra, namely $B_N(1/k)$, on $\mathbb{C}[\mathfrak{h}]^{S_N}$. Moreover, this representation is the symmetric part W of the standard polynomial representation of $H_N(1/k)$, which is faithful and irreducible, since $1/k = n/N$ is an integer [GGOR03]. In other words, we have the following proposition.

Proposition 53. *There exists a surjective homomorphism $\phi : B_n(N/n) \to B_N(n/N)$, such that $\phi^* W = U_N$. In particular, U_N is an irreducible representation of $B_n(N/n)$.*

Proposition 53 can be generalized as follows. Let $0 \le p \le n/N$ be an integer. Consider the partition $\mu(p) = (n - p(N-1), p, \ldots, p)$ of n. The representation of $\mathfrak{g}$ attached to $\mu(p)$ is $S^{n-pN}\mathbb{C}^N$.

Let $e(p)$ be a primitive idempotent of the representation of S_n attached to $\mu(p)$. Let $U_N^p = e(p)V_N = (\mathbb{C}[\mathfrak{g}] \otimes S^{n-pN}\mathbb{C}^N)^{\mathfrak{g}}$. Then the algebra $e(p)H_n(N/n)e(p)$ acts on U_N^p, and the above situation of U_N is the special case $p = 0$.

Proposition 54. *There exists a surjective momorphism*

$$\phi_p : e(p)H_n(N/n)e(p) \to B_N(n/N - p)$$

such that $\phi_p^* W = U_N^p$. In particular, U_N^p is an irreducible representation of $B_n(N/n - p)$.

Proof. Similar to the proof of Proposition 53. □

Example 55. $p = 1$, $n = N$. In this case $e(p) = e_- = \frac{1}{n!}\sum_{\sigma \in S_n} \varepsilon(\sigma)\sigma$, the antisymmetrizer, and the map ϕ_p is the shift isomorphism $e_- H_N(1) e_- \to eH_N(0)e$.

8.3 Coincidence of the two $\mathfrak{sl}_2$ actions

As before, let $\{b_p\}$ be an orthonormal basis of $\mathfrak{g}$ (under some invariant inner product). Consider the $\mathfrak{sl}_2$-triple

$$H = \sum b_p \frac{\partial}{\partial b_p} + \frac{\dim \mathfrak{g}}{2} \tag{45}$$

(the shifted Euler field),

$$E = \frac{1}{2}\sum_p b_p^2, \quad F = -\frac{1}{2}\Delta_{\mathfrak{g}}, \tag{46}$$

where $\Delta_{\mathfrak{g}}$ is the Laplace operator on $\mathfrak{g}$. Recall also (see, e.g., [BEG03b]) that the rational Cherednik algebra contains the $\mathfrak{sl}_2$-triple $\mathbf{h} = \frac{1}{2}\sum_i (\mathrm{x}_i \mathrm{y}_i + \mathrm{y}_i \mathrm{x}_i)$, $\mathbf{f} = -\frac{1}{2}\sum_i \mathrm{y}_i^2$, $\mathbf{e} = \frac{1}{2}\sum_i \mathrm{x}_i^2$.

The following proposition shows that the actions of these two $\mathfrak{sl}_2$ algebras on V_N essentially coincide.

Proposition 56. *On V_N, one has*

$$\mathbf{h} = H, \quad \mathbf{e} = \frac{n}{N}E, \quad \mathbf{f} = \frac{N}{n}F.$$

Proof. The last two equations follow from formulas (43), (44), and the first one follows from the last two by taking commutators. □

8.4 The irreducibility of V_N.

Let $\Delta(n,N)$ be the representation of the symmetric group S_n corresponding to the rectangular Young diagram with N rows (and correspondingly n/N columns), i.e., to the partition $\left(\frac{n}{N}, \ldots, \frac{n}{N}\right)$; e.g., $\Delta(n,1)$ is the trivial representation.

For a representation π of S_n, let $L(\pi)$ denote the irreducible lowest weight representation of $H_n(k)$ with lowest weight π.

Theorem 57. *The representation V_N is isomorphic to $L(\Delta(n,N))$.*

Proof. The representation V_N is graded by the degree of polynomials, and in degree-zero we have $V_N[0] = ((\mathbb{C}^N)^{\otimes n})^{\mathfrak{g}} = \Delta(n,N)$ by the Weyl duality.

Let us show that the module V_N is semisimple. It is sufficient to show that V_N is a unitary representation, i.e., admits a positive definite contravariant Hermitian form. Such a form can be defined by the formula

$$(f,g) = \langle f(\partial_A), g(A)\rangle|_{A=0},$$

where $\langle -,-\rangle$ is the Hermitian form on $(\mathbb{C}^N)^{\otimes n}$ obtained by tensoring the standard forms on the factors. This form is obviously positive definite, and satisfies the contravariance properties

$$(Y_i f, g) = \frac{N}{n}(f, X_i g), \ (f, Y_i g) = \frac{N}{n}(X_i f, g).$$

The existence of the form $(-,-)$ implies the semisimplicity of V_N. In particular, we have a natural inclusion $L(\Delta(n,N)) \subset V_N$.

Next, formula (43) implies that V_N is a torsion-free module over $R := \mathbb{C}[\mathrm{x}_1, \ldots, \mathrm{x}_N]^{S_N} = \mathbb{C}\left[\sum_{i=1}^N \mathrm{x}_i^p, 2 \le p \le N\right]$. Since V_N is semisimple, this implies that $V_N/L(\Delta(n,N))$ is torsion-free as well.

On the other hand, we will now show that the quotient $V_N/L(\Delta(n,N))$ is a torsion module over R. This will imply that the quotient is zero, as desired.

Let $v_1, \ldots, v_N$ be the standard basis of $\mathbb{C}^N$, and for each sequence $J = (j_1, \ldots, j_n)$, $j_i \in \{1, \ldots, N\}$, let $v_J := v_{j_1} \otimes \cdots \otimes v_{j_n}$. Let us say that a sequence J is balanced if it contains each of its members exactly n/N times. Let B be the set of balanced sequences. The set B has commuting left and right actions S_N and S_n, $\sigma * (j_1, \ldots, j_n) * \tau = (\sigma(j_{\tau(1)}), \ldots, \sigma(j_{\tau(n)}))$. Let $J_0 = (1 \ldots 1, 2 \ldots 2, \ldots, N \ldots N)$, then any $J \in B$ has the form $J = J_0 * \tau$ for some $\tau \in S_n$.

Let $f \in V_N$. Then f is a function $\mathfrak{h} \to ((\mathbb{C}^N)^{\otimes n})^{\mathfrak{h}}$, equivariant under the action of S_N (here $\mathfrak{h} \subset \mathfrak{g}$ is the Cartan subalgebra, so $\mathfrak{h} = \{(\lambda_1, \ldots, \lambda_N) | \sum_i \lambda_i = 0\}$), so

$$f(\lambda) = \sum_{J \in B} f_J(\lambda) v_J, \tag{47}$$

where $\lambda = (\lambda_1, \ldots, \lambda_N)$ and f_J are scalar functions (the summation is over B, since $f(\lambda)$ must have zero weight). By the S_N-invariance, we have $f_{\sigma * J}(\sigma(\lambda)) = f_J(\lambda)$. We then decompose $f(\lambda) = \sum_{o \in S_N \backslash B} f_o(\lambda)$, where $f_o(\lambda) = \sum_{J \in o} f_J(\lambda) v_J$.

For each $o \in S_N \setminus B$, we construct a nonzero $\phi_o \in \mathbb{C}[\mathrm{x}_1, \ldots, \mathrm{x}_n]$ such that $\phi_o \cdot f_o(\lambda) \in L(\Delta(n, N))$. Then $\phi := \prod_{o \in S_N \backslash B} \prod_{\sigma \in S_N} \sigma(\phi_o) \in R$ is nonzero and such that $\phi \cdot f(\lambda) \in L(\Delta(n, N))$.

We first construct ϕ_o when $o = o_0$, the class of J_0. By S_N-invariance, $f_{o_0}(\lambda)$ has the form

$$f_{o_0}(\lambda) = \sum_{\sigma \in S_N} g(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(N)}) v_{\sigma(1)}^{\otimes n/N} \otimes \cdots \otimes v_{\sigma(N)}^{\otimes n/N}$$

where $g(\lambda, \ldots, \lambda_N) \in \mathbb{C}[\lambda_1, \ldots, \lambda_N]$. For $\phi_{o_0} \in \mathbb{C}[\mathrm{x}_1, \ldots, \mathrm{x}_N]$, we have

$$\phi_{o_0} \cdot f_{o_0}(\lambda) = \sum_{\sigma \in S_N} (\phi_{o_0} g)(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(N)}) v_{\sigma(1)}^{\otimes n/N} \otimes \cdots \otimes v_{\sigma(N)}^{\otimes n/N}. \tag{48}$$

On the other hand, let $v \in \Delta(n, N)$; expand $v = \sum_{J \in B} c_J v_J$. One checks that v can be chosen such that $c_{J_0} \neq 0$ (one starts with a nonzero vector v' and $J' \in B$ such that the coordinate of v' along J' is nonzero, and then acts on v' by an element of S_n bringing J' to J_0). Then since v is $\mathfrak{g}$-invariant (and therefore S_N-invariant), we have

$$c_{\sigma(1) \cdots \sigma(1) \cdots \sigma(N) \cdots \sigma(N)} = c_{J_0} \tag{49}$$

for any $\sigma \in S_N$.

If $Q \in \mathbb{C}[\mathrm{x}_1, \ldots, \mathrm{x}_n]$, then

$$(Q \cdot v)(\lambda) = \sum_{(j_1, \ldots, j_n) \in B} c_{j_1 \cdots j_n} Q(\lambda_{j_1}, \ldots, \lambda_{j_n}) v_{j_1} \otimes \cdots \otimes v_{j_n} \in L(\Delta(n, N)). \tag{50}$$

Set

$$Q_0(\lambda_1,\ldots,\lambda_n) := \prod_{1\le a<b\le n, j_a^0\neq j_b^0} (\lambda_a-\lambda_b),$$

where

$$\begin{aligned}\left(j_1^0,\ldots,j_n^0\right) &= J_0,\\ q_0(\lambda_1,\ldots,\lambda_N) &:= Q_0(\lambda_1\cdots\lambda_1,\ldots,\lambda_N\cdots\lambda_N),\end{aligned}$$

so

$$q_0(\lambda_1,\ldots,\lambda_N) = \Big(\prod_{1\le i<j\le N}(\lambda_i-\lambda_j)\Big)^{(n/N)^2}.$$

Set $\phi_{o_0}(\lambda_1,\ldots,\lambda_N) := q_0(\lambda_1,\ldots,\lambda_N)$ and

$$Q(\lambda_1,\ldots,\lambda_n) := Q_0(\lambda_1,\ldots,\lambda_n)q(\lambda_1,\lambda_{(n/N)+1},\ldots,\lambda_{(N-1)\frac{n}{N}+1}).$$

Then (48) and (50) coincide, since: (a) for $J\notin o_0$, $Q_0(\lambda_{j_1},\ldots,\lambda_{j_n})=0$, so the coefficient of v_J in both expressions is zero, (b) the coefficients of v_{J_0} in both expressions coincide, (c) for $J\in o_0$, the coefficients of v_J coincide because of (b) and (49). The functions ϕ_o are constructed in the same way for a general $o\in S_N\setminus B$. This ends the proof of the theorem. □

Remark 58. Theorem 57 is a special case of the much more general (but much less elementary) Theorem 68, which is proved below.

8.5 The character formula for V_N

For each partition μ of n, let $V(\mu)$ be the representation of $\mathfrak{g}$, and $\pi(\mu)$ the representation of S_n corresponding to μ.

Let $P_\mu(q)$ be the q-analogue of the weight multiplicity of the zero weight in $V(\mu)$. Namely, we have a filtration $F^\bullet$ on $V(\mu)[0]$ such that F^i is the space of vectors in $V(\mu)[0]$ killed by the $(i+1)$-power of the principal nilpotent element $\sum e_i$ of $\mathfrak{g}$. Then $P_\mu(q)=\sum_{j\ge0}\dim(F^j/F^{j-1})q^j$. The coefficients of $P_\mu(q)$ are called the generalized exponents of $V(\mu)$ (see [Kos63, He, Lus81] for more details).

We have $V_N=\oplus_\mu\pi(\mu)\otimes(\mathbb{C}[\mathfrak{g}]\otimes V(\mu))^{\mathfrak{g}}$. This together with Theorem 57 implies the following.

Corollary 59. *The character of $L(\Delta(n,N))$ is given by the formula*

$$\mathrm{Tr}|_{L(\Delta(n,N))}(w\cdot q^{\mathbf{h}}) = q^{(N^2-1)/2}\frac{\sum_\mu \chi_{\pi(\mu)}(w)P_\mu(q)}{(1-q^2)\ldots(1-q^N)},$$

where $w\in S_n$, and $\chi_{\pi(\mu)}$ is the character of $\pi(\mu)$. Here the summation is over partitions μ of n with at most N parts.

Proof. The formula follows, using Proposition 56, from Kostant's result [Kos63] that $(\mathbb{C}[\mathfrak{g}] \otimes V(\mu))^{\mathfrak{g}}$ is a free module over $\mathbb{C}[\mathfrak{g}]^{\mathfrak{g}}$, and the fact that the Hilbert polynomial of the space of generators for this module is the q-weight multiplicity of the zero weight, $P_\mu(q)$ [Kos63, Lus81, He]. □

Remark 60. It would be interesting to compare this formula with the character formula of [Rou05] for the same module.

9 Equivariant D-modules and representations of the rational Cherednik algebra

9.1 The category of equivariant D-modules on the nilpotent cone

The theory of equivariant D-modules on the nilpotent cone arose from Harish-Chandra's work on invariant distributions on nilpotent orbits of real groups, and was developed further in many papers; see, e.g., [HK84, LS97, Lev98, Mir04] and references therein. Let us recall some of the basics of this theory.

Let G be a simply connected simple algebraic group over $\mathbb{C}$, and $\mathfrak{g}$ its Lie algebra. Let $\mathcal{N} \subset \mathfrak{g}$ be the nilpotent cone of $\mathfrak{g}$. We denote by $\mathcal{D}(\mathfrak{g})$ the category of finitely generated D-modules on $\mathfrak{g}$, by $\mathcal{D}_G(\mathfrak{g})$ the subcategory of G-equivariant D-modules, and by $\mathcal{D}_G(\mathcal{N})$ the category of G-equivariant D-modules that are set-theoretically supported on $\mathcal{N}$ (here we do not make a distinction between a D-module on an affine space and the space of its global sections). Since G acts on $\mathcal{N}$ with finitely many orbits, it is well known that any object in $\mathcal{D}_G(\mathcal{N})$ is regular and holonomic.

Moreover, the category $\mathcal{D}_G(\mathcal{N})$ has finitely many simple objects, and every object of this category has finite length (so this category is equivalent to the category of modules over a finite-dimensional algebra).

9.2 Simple objects in $\mathcal{D}_G(\mathcal{N})$

Recall (see e.g. [Mir04] and references) that irreducible objects in the category $\mathcal{D}_G(\mathcal{N})$ are parametrized by pairs (O, χ), where O is a nilpotent orbit of G in $\mathfrak{g}$, and χ is an irreducible representation of the fundamental group $\pi_1(O)$, which is clearly isomorphic to the component group $A(O)$ of the centralizer G_x of a point $x \in O$. Namely, χ defines a local system L_χ on O, and the simple object $M(O, \chi) \in \mathcal{D}_G(\mathcal{N})$ is the direct image of the Goresky–MacPherson extension of L_χ to the closure $\bar{O}$ of O, under the inclusion of $\bar{O}$ into $\mathfrak{g}$.

9.3 Semisimplicity of $\mathcal{D}_G(\mathcal{N})$

The proof of the following theorem was explained to us by G. Lusztig.

Theorem 61. *The category $\mathcal{D}_G(\mathcal{N})$ is semisimple.*

Proof. We may replace the category $\mathcal{D}_G(\mathcal{N})$ by the category of G-equivariant perverse sheaves (of complex vector spaces) on $\mathfrak{g}$ supported on $\mathcal{N}$, $\mathrm{Perv}_G(\mathcal{N})$, since these two categories are known to be equivalent. We must show that $\mathrm{Ext}^1(P,Q) = 0$ for every two simple objects $P,Q \in \mathrm{Perv}_G(\mathcal{N})$.

Let P', Q' be the Fourier transforms of P,Q. Then P', Q' are character sheaves on $\mathfrak{g}$, and it suffices to show that $\mathrm{Ext}^1(P',Q') = 0$.

Recall that to each character sheaf S one can naturally attach a conjugacy class of pairs (L,θ), where L is a Levi subgroup of G, and θ is a cuspidal local system on a nilpotent orbit for L. It is shown by arguments parallel to those in [Lus85] (which treats the more difficult case of character sheaves on the group) that if (L_i,θ_i) corresponds to S_i, $i=1,2$, and (L_1,θ_1) is not conjugate to (L_2,θ_2) then $\mathrm{Ext}^*(S_1,S_2) = 0$. Thus it is sufficient to assume that the pair (L,θ) attached to P' and Q' is the same.

Using standard properties of constructible sheaves (in particular, Poincaré duality), we have

$$\begin{aligned}\mathrm{Ext}^1(P',Q') &= H^1(\mathfrak{g}, \underline{\mathrm{Hom}}(P',Q'))\\ &= H_c^{2\dim\mathfrak{g}-1}(\mathfrak{g}, \underline{\mathrm{Hom}}(P',Q')^*)^* = H_c^{2\dim\mathfrak{g}-1}(\mathfrak{g}, (Q')^*\otimes P')^*,\end{aligned}$$

where $*$ for sheaves denotes the Verdier duality functor.

Recall that to each character sheaf one can attach an irreducible representation of a certain Weyl group, via the generalized Springer correspondence. Let R be the direct sum of all character sheaves corresponding to a given pair (L,θ) with multiplicities given by the dimensions of the corresponding representations. Then it is sufficient to show that $H_c^{2\dim\mathfrak{g}-1}(\mathfrak{g},(R')^*\otimes R') = 0$.

This fact is essentially proved in [Lus88]. Namely, it follows from the computations of [Lus88] that $H_c^i(\mathfrak{g},(R')^*\otimes R')$ is the cohomology with compact support of a certain generalized Steinberg variety with twisted coefficients, and it is shown that this cohomology is concentrated in even degrees.[9] The theorem is proved. □

9.4 Monodromicity

We will need the following lemma.

Lemma 62. *Let $Q \in \mathcal{D}_G(\mathcal{N})$. Then for any finite-dimensional representation U of $\mathfrak{g}$, the action of the shifted Euler operator H defined by (45) on $(Q\otimes U)^{\mathfrak{g}}$ is locally finite (so Q is a monodromic D-module), and has finite-dimensional generalized eigenspaces. Moreover, the eigenvalues of H on $(Q\otimes U)^{\mathfrak{g}}$ are bounded from above. In particular, $(Q\otimes U)^{\mathfrak{g}}$ belongs to category $\mathcal{O}$ for the $\mathfrak{sl}_2$-algebra spanned by H and the elements E, F given by (46).*

[9] More precisely, in the arguments of [Lus88] the vanishing of odd cohomology is proved for G-equivariant cohomology with compact supports, and in the nonequivariant case one should use parallel arguments, rather than exactly the same arguments.

Proof. Since Q has finite length, it is sufficient to assume that Q is irreducible. We may further assume that Q is generated by an irreducible G-submodule V, annihilated by multiplication by any invariant polynomial on $\mathfrak{g}$ of positive degree. Indeed, let V_0 be an irreductible G-submodule of Q, let $J_{V_0} := \{f \in \mathbb{C}[\mathfrak{g}]^{\mathfrak{g}} | fV_0 = 0\}$ and for any $v \in V_0$, let $J_v := \{f \in \mathbb{C}[\mathfrak{g}]^{\mathfrak{g}} | fv = 0\}$. Then if $v \in V_0$ is nonzero, $J_v = J_{V_0}$, since $Gv = V_0$. Moreover, the support condition implies that $J_v \subset \mathfrak{m}^k$ for some $k \geq 0$, where $\mathfrak{m} = \mathbb{C}[\mathfrak{g}]^{\mathfrak{g}}_+$. So $J_{V_0} \subset \mathfrak{m}^k$ and is an ideal of $\mathbb{C}[\mathfrak{g}]^{\mathfrak{g}}$. Let $f \in \mathbb{C}[\mathfrak{g}]^{\mathfrak{g}}$ be such that $f \notin J_{V_0}$ and $f\mathfrak{m} \subset J_{V_0}$; we set $V := fV_0$.

Then Q is a quotient of the D-module $\tilde{Q} \otimes V$ by a G-stable submodule, where

$$\tilde{Q} := D(\mathfrak{g})/(D(\mathfrak{g})\mathrm{ad}(\mathrm{Ann}(V)) + D(\mathfrak{g})I),$$

$\mathrm{Ann}(V)$ is the annihilator of V in $U(\mathfrak{g})$, and I is the ideal in $\mathbb{C}[\mathfrak{g}]$ generated by invariant polynomials on $\mathfrak{g}$ of positive degree. Thus, it suffices to show that the lemma holds for the module $\tilde{Q}$ (which is only weakly G-equivariant, i.e., the group action and the Lie algebra action coming from differential operators do not agree, in general).

The algebra $D(\mathfrak{g})$ has a grading in which $\deg(\mathfrak{g}^*) = -1$, $\deg(\mathfrak{g}) = 1$. This grading descends to a grading on $\tilde{Q}$. We will show that for each U, this grading on $(\tilde{Q} \otimes U)^{\mathfrak{g}}$ has finite-dimensional pieces, and is bounded from above. This implies the lemma, since the Euler operator preserves the grading.

Consider the associated graded module $\tilde{Q}_0$ of $\tilde{Q}$ under the Bernstein filtration. This is a bigraded module over $\mathbb{C}[\mathfrak{g} \oplus \mathfrak{g}]$ (where we identify $\mathfrak{g}$ and $\mathfrak{g}^*$ using the trace form). We have to show that the homogeneous subspaces of $(\tilde{Q}_0 \otimes U)^{\mathfrak{g}}$ under the grading defined by $\deg(\mathfrak{g} \oplus 0) = -1$, $\deg(0 \oplus \mathfrak{g}) = 1$ are finite-dimensional.

The associated graded of the ideal $\mathrm{Ann}(V) \subset U(\mathfrak{g})$ is such that $\mathbb{C}[\mathfrak{g}]^k_+ \subset \mathrm{grAnn}(V) \subset \mathbb{C}[\mathfrak{g}]_+$ for some $k \geq 1$; therefore

$$\tilde{Q}_0 = \mathbb{C}[\mathfrak{g} \oplus \mathfrak{g}]/J,$$

where J is a (not necessarily radical) ideal whose zero set is the variety $\mathcal{Z}$ of pairs $(u, v) \in \mathcal{N} \times \mathfrak{g}$ such that $[u, v] = 0$. Let

$$Q'_0 = \mathbb{C}[\mathfrak{g} \oplus \mathfrak{g}]/\sqrt{J}.$$

Because of the Hilbert basis theorem, it suffices to prove that the homogeneous subspaces of $(Q'_0 \otimes U)^{\mathfrak{g}}$ are finite-dimensional and that the degree is bounded above. But Q'_0 is the algebra of regular functions on $\mathcal{Z}$. By the result of [Jos97], one has $\mathbb{C}[\mathcal{Z}]^{\mathfrak{g}} = \mathbb{C}[\mathfrak{g}]^{\mathfrak{g}}$, the algebra of invariant polynomials of Y. But it follows from Hilbert's theorem on invariants that every isotypic component of $\mathbb{C}[\mathcal{Z}]$ is a finitely generated module over $\mathbb{C}[\mathcal{Z}]^{\mathfrak{g}}$. This implies the result. □

9.5 Characters

Lemma 62 allows one to define the character of an object $M \in \mathcal{D}_G(\mathcal{N})$. Namely, let $\mu = (\mu_1, \ldots, \mu_N)$ be a dominant integral weight for $\mathfrak{g}$, and $V(\mu)$

the irreducible representation of $\mathfrak{g}$ with highest weight μ. Let $K_M(\mu) = (M \otimes V(\mu))^{\mathfrak{g}}$. Then the character of M is defined by the formula

$$\mathrm{Ch}_M(t,g) = \mathrm{Tr}_{|M}(gt^{-H}) = \sum_{\mu} \mathrm{Tr}_{|K_M(\mu)}(t^{-H})\chi_\mu(g), \quad g \in G,$$

where χ_μ denotes the character of μ. It can be viewed as a linear functional from $\mathbb{C}[G]^G$ to $\mathbb{F} := \oplus_{\beta \in \mathbb{C}} t^\beta \mathbb{C}[[t]]$, via the integration pairing.

In other words, the multiplicity spaces $K_M(\mu)$ are representations from the category $\mathcal{O}$ of the Lie algebra $\mathfrak{sl}_2$ spanned by E, F, H, and the character of M carries the information about the characters of these representations.

The problem of computing characters of simple objects in $\mathcal{D}_G(\mathcal{N})$ is interesting and, to our knowledge, open. Below we will show how these characters for $G = \mathrm{SL}_N(\mathbb{C})$ can be expressed via characters of irreducible representations of the rational Cherednik algebra.

Example 63. Recall (see, e.g., [Mir04]) that an object $M \in \mathcal{D}_G(\mathcal{N})$ is cuspidal iff $\mathcal{F}(M) \in \mathcal{D}_G(\mathcal{N})$, where $\mathcal{F}$ is the Fourier transform (Lusztig's criterion). If follows that in the case of cuspidal objects M, the spaces $K_M(\mu)$ are also in the category $\mathcal{O}$ for the opposite Borel subalgebra of $\mathfrak{sl}_2$, hence are finite dimensional representations of $\mathfrak{sl}_2$, and in particular, their dimensions are of interest.

9.6 The functors F_n, F_n^*

The representation V_N is a special case of representations of the rational Cherednik algebra that can be constructed via a functor similar to the one defined in [GG04]. Namely, the construction of V_N can be generalized as follows.

Let n and N be positive integers (we no longer assume that N is a divisor of n), and $k = N/n$. We again consider the special case $G = \mathrm{SL}_N(\mathbb{C})$, $\mathfrak{g} = \mathfrak{sl}_N(\mathbb{C})$. Then we have a functor $F_n : \mathcal{D}(\mathfrak{g}) \to H_n(k)$-mod defined by the formula

$$F_n(M) = (M \otimes (\mathbb{C}^N)^{\otimes n})^{\mathfrak{g}},$$

where $\mathfrak{g}$ acts on M by adjoint vector fields. The action of $H_n(k)$ on $F_n(M)$ is defined by the same formulas as in Proposition 50, and Proposition 56 remains valid.

Note that $F_n(M) = F_n(M_{\mathrm{fin}})$, where M_{fin} is the set of $\mathfrak{g}$-finite vectors in M. Clearly M_{fin} is a G-equivariant D-module. Thus, it is sufficient to consider the restriction of F_n to the subcategory $\mathcal{D}_G(\mathfrak{g})$, which we will do from now on.

In general, $F_n(M)$ does not belong to the category $\mathcal{O}$. However, we have the following lemma.

Lemma 64. *If the Fourier transform $\mathcal{F}(M)$ of M is set-theoretically supported on the nilpotent cone $\mathcal{N}$ of $\mathfrak{g}$, then $F_n(M)$ belongs to the category $\mathcal{O}$.*

Proof. Since $\mathcal{F}(M)$ is supported on $\mathcal{N}$, invariant polynomials on $\mathfrak{g}$ act locally nilpotently on $\mathcal{F}(M)$. Hence invariant differential operators on $\mathfrak{g}$ with constant coefficients act locally nilpotently on M. Thus, it follows from formula (44) that the algebra $\mathbb{C}[\mathrm{y}_1, \ldots, \mathrm{y}_n]^{S_n}$ acts locally nilpotently on $F_n(M)$. Also, by Lemma 62, the operator $\mathbf{h}$ acts with finite-dimensional generalized eigenspaces on $F_n(M)$. This implies the statement. □

Thus we obtain an exact functor $F_n^* = F_n \circ \mathcal{F} : \mathcal{D}_G(\mathcal{N}) \to \mathcal{O}(H_n(k))$.

9.7 The symmetric part of F_n

Consider the symmetric part $eF_n(M)$ of $F_n(M)$. We have $eF_n(M) = (M \otimes S^n\mathbb{C}^N)^{\mathfrak{g}}$, and we have an action of the spherical subalgebra $B_n(k)$ on $eF_n(M)$, given by formulas (43), (44).

This allows us to relate the functor F_n with the functor defined in [GG04]. Namely, recall from [GG04] that for any $c \in \mathbb{Z}$, one may define the category $\mathcal{D}_c(\mathfrak{g} \times \mathbb{P}^{N-1})$ of coherent D-modules on $\mathfrak{g} \times \mathbb{P}^{N-1}$ that are twisted by the cth power of the tautological line bundle on the second factor (this makes sense for all complex c even though the cth power is defined only for integer c). Then the paper [GG04][10] defines a functor

$$\mathbb{H} : \mathcal{D}_c(\mathfrak{g} \times \mathbb{P}^{N-1}) \to B_N(c/N)\text{-mod},$$

given by $\mathfrak{H}(M) = M^{\mathfrak{g}}$.

Proposition 65. *(i) If n is divisible by N then one has a functorial isomorphism $eF_n(M) \simeq \phi^*\mathbb{H}(M \otimes S^n\mathbb{C}^N)$, where $S^n\mathbb{C}^N$ is regarded as a twisted D-module on $\mathbb{P}^{N-1}$ (with $c = n$).*

(ii) For any n, the actions of $B_n(N/n)$ and $B_N(n/N)$ on the space $eF_n(M) = \mathbb{H}(M \otimes S^n\mathbb{C}^N)$ have the same image in the algebra of endomorphisms of this space.

Proof. This follows from the definition of $\mathbb{H}$ and formulas (43), (44). □

Corollary 66. *The functor eF_n^* on the category $\mathcal{D}_G(\mathcal{N})$ maps irreducible objects into irreducible ones.*

Proof. This follows from Proposition 65, (ii), and Proposition 7.4.3 of [GG04], which states that the functor $\mathbb{H}$ maps irreducible objects to irreducible ones. □

Formulas (43), (44) can also be used to study the support of $F_n^*(M)$ for $M \in \mathcal{D}_G(\mathcal{N})$, as a $\mathbb{C}[\mathrm{x}_1, \ldots, \mathrm{x}_n]$-module. Namely, we have the following proposition.

[10]There seems to be a misprint in [GG04]: in the definition of $\mathbb{H}$, c should be replaced by c/N.

Proposition 67. *Let $q = \gcd(n, N)$ be the greatest common divisor of n and N. Then the support S of $F_n^*(M)$ is contained in the union of the S_n-translates of the subspace E_q of $\mathbb{C}^n$ defined by the equations $\sum_{i=1}^n x_i = 0$ and $x_i = x_j$ if $\frac{n}{q}(l-1)+1 \le i, j \le \frac{nl}{q}$ for some $1 \le l \le q$.*

Proof. It follows from equation (44) that for any $(x_1, \ldots, x_n) \in S$ there exists a point $(z_1, \ldots, z_N) \in \mathbb{C}^N$ such that one has

$$\frac{1}{n}\sum_{i=1}^{n} x_i^p = \frac{1}{N}\sum_{j=1}^{N} z_j^p$$

for all positive integers p. In particular, writing generating functions, we find that

$$N\sum_{i=1}^{n}\frac{1}{1-tx_i} = n\sum_{j=1}^{N}\frac{1}{1-tz_j}.$$

In particular, every fraction occurs on both sides at least $\mathrm{lcm}(n, N)$ times, and hence the numbers x_i fall into n/q-tuples of equal numbers (and the numbers z_j into N/q-tuples of equal numbers). The proposition is proved. □

9.8 Irreducible equivariant D-modules on the nilpotent cone for $G = \mathrm{SL}_N(\mathbb{C})$

Nilpotent orbits for $\mathrm{SL}_N(\mathbb{C})$ are labeled by Young diagrams, or partitions. Namely, if $x \in \mathfrak{sl}_N(\mathbb{C})$ is a nilpotent element, then we let μ_i be the sizes of its Jordan blocks enumerated in decreasing order. The partition $\mu = (\mu_1, \ldots, \mu_m)$ and the corresponding Young diagram whose rows have lengths μ_i are attached to x. If O is the orbit of x then we will denote μ by $\mu(O)$. For instance, if $O = \{0\}$ then $\mu = (1^N)$, and if O is the open orbit, then $\mu = (N)$.

It is known (and easy to show) that the group $A(O)$ is naturally isomorphic to $\mathbb{Z}/d\mathbb{Z}$, where d is the greatest common divisor of the μ_i. Namely, let $Z = \mathbb{Z}/N\mathbb{Z}$ be the center of G (we identify $\mathbb{Z}/N\mathbb{Z}$ with Z by $p \to e^{2\pi \mathrm{i} p/N}\mathrm{Id}$). Then we have a natural surjective homomorphism $\theta : Z \to A(O)$ induced by the inclusion $Z \to G_x$, $x \in O$. This homomorphism sends d to 0, and thus $A(O)$ gets identified with $\mathbb{Z}/d\mathbb{Z}$.

Thus, any character $\chi : A(O) \to \mathbb{C}^*$ is defined by the formula $\chi(p) = e^{-2\pi \mathrm{i} ps/d}$, where $0 \le s < d$. We will denote this character by χ_s.

9.9 The action of F_n^* on irreducible objects

Obviously, the center Z of G acts on $F_n^*(M)$ by $z \to z^{-sN/d}$. Thus, a necessary condition for $F_n^*(M(O, \chi_s))$ to be nonzero is

$$n = N\left(p + \frac{s}{d}\right), \tag{51}$$

where p is a nonnegative integer.

Our main result in this section is the following theorem.

Theorem 68. *The functor F_n^* maps irreducible objects into irreducible ones or zero. Specifically, if condition (51) holds, then we have*

$$F_n^*(M(O,\chi_s)) = L(\pi(n\mu(O)/N)),$$

the irreducible representation of $H_n(k)$ whose lowest weight is the representation of S_n corresponding to the partition $n\mu(O)/N$.

Remark 69. Here if μ is a partition and $c \in \mathbb{Q}$ is a rational number, then we denote by $c\mu$ the partition whose parts are $c\mu_i$, provided that these numbers are all integers. In our case, this integrality condition holds, since all parts of $\mu(O)$ are divisible by d.

Corollary 70. *Let λ be a partition of n into at most N parts. Let $M = M(O_\mu, \chi_s)$, and assume that condition (51) is satisfied. Then*

$$(M \otimes V(\lambda))^{\mathfrak{g}} = \mathrm{Hom}_{S_n}(\pi(\lambda), L(\pi(n\mu/N)))$$

as graded vector spaces.

This corollary allows us to express the characters of the irreducible D-modules $M(O,\chi)$ in terms of characters of certain special lowest weight irreducible representations of $H_n(k)$. We note that characters of lowest weight irreducible representations of rational Cherednik algebras of type A have been computed by Rouquier, [Rou05].

Remark 71. Note that Theorem 57 is the special case of Theorem 68 for $O = \{0\}$.

9.10 Proof of Theorem 68

Our proof of Theorem 68 is based on the following result of [BE09], which is proved in [GS05] under the assumption that k is not a half-integer.

Theorem 72. *Let $k > 0$. Then the functor $V \mapsto eV$ is an equivalence of categories between $H_n(k)$-modules and $B_n(k)$-modules.*

Theorem 72 implies the first statement of the theorem, i.e., that if (51) holds then $F_n^*(M(O_\mu,\chi_s))$ is irreducible. Indeed, it follows from Corollary 66 that $eF_n^*(M(O_\mu,\chi_s))$ is irreducible over $B_n(k)$. Thus, it remains to find the lowest weight of $F_n^*(M(O_\mu,\chi_s))$.

Let $\mu = (\mu_1,\dots,\mu_N)$ be a partition of N ($\mu_i \geq 0$). Let O_μ be the nilpotent orbit of $\mathfrak{g}$ corresponding to the partition μ. Denote by d the greatest common divisor of μ_i, and by m a divisor of d. Define the following function f on O_μ with values in $\otimes_{i=1}^N S^{\mu_i}\mathbb{C}^N$

$$f(X,\xi_1,\dots,\xi_N) = \bigwedge_{i=1}^{N} \bigwedge_{j=0}^{\mu_i-1} \xi_i X^j,$$

$\xi_i \in (\mathbb{C}^N)^*$ (here $X^j \in \mathrm{M}_N(\mathbb{C})$ is the jth power of X, so $\xi_i X^j \in \mathbb{C}^N$).

Lemma 73. *(i) For any $X \in O_\mu$, $f(X,\dots)^{1/m}$ is a polynomial in $\xi_1,\dots,\xi_N$. Thus, $f^{1/m}$ is a regular function on the universal cover $\tilde{O}_\mu$ of O_μ with values in $\otimes_{i=1}^N S^{\mu_i/m}\mathbb{C}^N$.*

(ii) For any $X \in O_\mu$, the function $f(X,\dots)^{1/m}$ generates a copy of the representation $V(\mu/m)$ inside $\otimes_{i=1}^N S^{\mu_i/m}\mathbb{C}^N$.

(iii) Specifically, let the standard basis $u_1,\dots,u_N$ of $(\mathbb{C}^N)^*$ be filled into the squares of the Young diagram of μ (filling the first column top to bottom, then the second one, etc.), and let X be the matrix J acting by the horizontal shift to the right on this basis. Then $f(J,\dots)^{1/m}$ is a highest weight vector of the representation $V(\mu/m)$.

Proof. It is sufficient to prove (iii). Let $\mu^* = (\mu_1^*,\dots,\mu_N^*)$ be the conjugate partition. Let p_j be the number of times the part j occurs in this partition. Clearly, p_j is divisible by m. By looking at the matrix whose determinant is f, we see that we have, up to sign,

$$f(J,\xi_1,\dots,\xi_N) = \prod_j \Delta_j(\xi_1,\dots,\xi_N)^{p_j},$$

where Δ_j is the left upper j-by-j minor of the matrix $(\xi_1,\dots,\xi_N)$. Thus $f^{1/m} = \prod_j \Delta_j^{p_j/m}$ is clearly a highest weight vector of weight $\sum_j p_j\varpi_j/m$, where ϖ_j are the fundamental weights, but $\sum p_j\varpi_j = \mu$, so we are done. □

Corollary 74. *The function f gives rise to a G-equivariant regular map $f : \tilde{O}_\mu \to V(\mu/d)$, whose image is the orbit of the highest weight vector. In particular, we have a G-equivariant inclusion of commutative algebras,*

$$f^* : \oplus_{\ell\geq 0} V(\ell\mu/d)^* \to \mathbb{C}[\tilde{O}_\mu].$$

Now let $0 \leq s \leq d-1$, and denote by $\mathbb{C}[\tilde{O}_\mu]_s$ the subspace of $\mathbb{C}[\tilde{O}_\mu]$ on which central elements $z \in G$ act by $z \to z^{-s}$. Then we have an inclusion

$$f^* : \oplus_{\ell : d^{-1}(\ell-s)\in\mathbb{Z}} V(\ell\mu/d)^* \to \mathbb{C}[\tilde{O}_\mu]_s.$$

Now recall that by construction, $\mathbb{C}[\tilde{O}_\mu]_s$ sits inside $M = M(O_\mu,\chi_s)$ as a $\mathbb{C}[O_\mu]$-submodule. In particular, the operators X_i act on the space $(\mathbb{C}[\tilde{O}_\mu]_s \otimes (\mathbb{C}^N)^{\otimes n})^{\mathfrak{g}}$.

Let $\pi(\mu)$ be the representation of S_n corresponding to μ, and regard $V(\lambda)\otimes\pi(\lambda)$, for any partition λ of n, as a subspace of $(\mathbb{C}^N)^{\otimes n}$ using the Weyl duality. Then for any $u \in \pi(n\mu/N)$, we can define the element $a(u) \in F_n^*(M)$ by $a(u) = f_n^* \otimes u$, where $f_n^* \in \mathbb{C}[\tilde{O}_\mu]_s \otimes V(n\mu/N)$ is the homogeneous part of f^* of degree n.

Lemma 75. *$a(u)$ is annihilated by the elements y_i of $H_n(k)$.*

Proof. We need to show that the operators X_i (or, equivalently, the elements $\mathrm{x}_i \in H_n(k)$) annihilate $a(u) \in F_n(M)$. Since $a(u)$ is G-invariant, it is sufficient to prove the statement at the point $X = J$. This boils down to showing that, for any j not exceeding the number of parts of μ (i.e. $j \le \mu_1^*$), the application of J, in any component, annihilates the element $\Delta_j(\xi_1, \dots, \xi_N) \in \wedge^j \mathbb{C}^N \subset (\mathbb{C}^N)^{\otimes j}$. This is clear, since the first μ_1^* columns of J are zero. □

This implies that the lowest weight of $F_n^*(M(O_\mu, \chi_s))$ is $\pi(n\mu/N)$, as desired. The theorem is proved.

Remark 76. Here is another, short, proof of Theorem 68 for $n = N$. We have

$$e_- F_N^*(M(O, 1)) = \mathcal{F}(M(O, 1))^G.$$

According to [Lev98, LS97],

$$\mathcal{F}(M(O, 1))^G = (\mathbb{C}[\mathfrak{h}] \otimes \pi(\mu(O)))^{S_N}$$

as a module over $D(\mathfrak{h})^W = e_- H_N(1) e_-$. Thus,

$$e_- F_N^*(M(O, 1)) = e_- L(\pi(\mu(O)))$$

as $e_- H_N(1) e_-$-modules. But the functor $V \to e_- V$ is an equivalence of categories $H_N(1)$-mod $\to e_- H_N(1) e_-$-mod (see [BEG03b]). Thus, $F_N^*(M(O, 1)) = L(\pi(\mu(O)))$ as $H_N(1)$-modules, as desired.

9.11 The support of $L(\pi(n\mu/N))$

Corollary 77. *Let μ be a partition of N such that $n\mu_i/N$ are integers. Then the support of the representation $L(\pi(n\mu/N))$ of $H_n(N/n)$ as a module over $\mathbb{C}[\mathrm{x}_1, \dots, \mathrm{x}_n]$ is contained in the union of S_n-translates of E_q, $q = \gcd(n, N)$.*

Proof. This follows from Theorem 68 and Proposition 67. □

We note that in the case $\mu = (N)$, Corollary 77 follows from Theorem 3.2 from [CE03].

9.12 The cuspidal case

An interesting special case of Theorem 68 is the cuspidal case. In this case N and n are relatively prime, $d = N$ (i.e., O is the open orbit), and s is relatively prime to N.

Here is a short proof of Theorem 68 in the cuspidal case.

Since the Fourier transform of $M(O, \chi_s)$ in the cuspidal case is supported on the nilpotent cone, $F_n^*(M(O, \chi_s))$ belongs not only to the category $\mathcal{O}$ generated by lowest-weight modules, but also to the "dual" category $\mathcal{O}_-$ generated by highest weight modules over $H_n(k)$. Thus, by the results of [BEG03a], $F_n^*(M(O, \chi_s))$ is a multiple of the unique finite-dimensional irreducible $H_n(k)$-module $L(\mathbb{C}) = Y(N, n)$, of dimension N^{n-1}. But this multiple must be a single copy by Corollary 66, so the theorem is proved.

Theorem 68 implies the following formula for the characters of the cuspidal D-modules $M(O, \chi_s)$.

Let μ be a dominant integral weight for $\mathfrak{g}$ such that the center Z of G acts on $V(\mu)$ via $z \to z^s = z^n$. Let ρ be the half-sum of positive roots of $\mathfrak{g}$. Let $K_s(\mu) = (M(O, \chi_s) \otimes V(\mu))^{\mathfrak{g}}$ be the isotypic components of $M(O, \chi_s)$.

Theorem 78. *We have*

$$\mathrm{Tr}_{|K_s(\mu)}(q^{2H}) = \frac{q - q^{-1}}{q^N - q^{-N}} \varphi_\mu(q),$$

where

$$\varphi_\mu(q) := \prod_{1 \le p < r \le N} \frac{q^{\mu_r - \mu_p + r - p} - q^{\mu_p - \mu_r + p - r}}{q^{r-p} - q^{p-r}} = \chi_{V(\mu)}(q^{2\rho}),$$

where $\chi_{V(\mu)}$ is the character of $V(\mu)$. In particular,

$$\dim K_s(\mu) = \frac{1}{N} \prod_{1 \le p < r \le N} \frac{\mu_r - \mu_p + r - p}{r - p} = \frac{1}{N} \dim V(\mu).$$

Proof. We extend the representation $V(\mu)$ to $\mathrm{GL}_N(\mathbb{C})$ by setting $z \to z^n$ for all scalar matrices z, so that its $\mathrm{GL}_N(\mathbb{C})$-highest weight is

$$\tilde{\mu} := (\mu_1 + n/N, \ldots, \mu_N + n/N).$$

Note that we automatically have $\mu_i + n/N \in \mathbb{Z}$. Assume that n is so big that $\tilde{\mu}$ is a partition of n (i.e., $\mu_i + n/N \ge 0$).

It follows from the results of [BEG03a] that the character of the irreducible representation $L(\mathbb{C})$ of the rational Cherednik algebra $H_n(k)$, $k = N/n$, is given by the formula

$$\mathrm{Tr}_{|L(\mathbb{C})}(g q^{2\mathbf{h}}) = \frac{q - q^{-1}}{q^N - q^{-N}} \frac{\det(q^{-N} - q^N g)}{\det(q^{-1} - qg)}, \quad g \in S_n, \tag{52}$$

where the determinants are taken in $\mathbb{C}^n$.

Let us equip $\mathbb{C}^N$ with the structure of an irreducible representation of $\mathfrak{sl}_2$ with basis e, f, h. Let $g \in S_n$. Then

$$\mathrm{Tr}_{|\mathrm{Hom}_{S_n}(\pi(\tilde{\mu}), (\mathbb{C}^N)^{\otimes n})}(q^h) = \mathrm{Tr}_{|V(\mu)}(q^{2\rho}) = \varphi_\mu(q),$$

by the Weyl character formula. On the other hand, it is easy to show that

$$\mathrm{Tr}_{|(\mathbb{C}^N)^{\otimes n}}(g q^h) = \frac{\det(q^{-N} - q^N g)}{\det(q^{-1} - qg)}.$$

Thus,

$$\begin{aligned}\mathrm{Tr}_{|\mathrm{Hom}_{S_n}(\pi(\tilde{\mu}), L(\mathbb{C}))}(q^{2\mathbf{h}}) &= \frac{q - q^{-1}}{q^N - q^{-N}} \mathrm{Tr}_{|\mathrm{Hom}_{S_n}(\pi(\tilde{\mu}), (\mathbb{C}^N)^{\otimes n})}(q^h) \\ &= \frac{q - q^{-1}}{q^N - q^{-N}} \varphi_\mu(q).\end{aligned}$$

By Theorem 68 and Weyl duality, this implies that

$$\mathrm{Tr}_{|(M(O,\chi_s)\otimes V(\mu))^{\mathfrak{g}}}(q^{2H}) = \frac{q-q^{-1}}{q^N - q^{-N}}\varphi_\mu(q),$$

as desired. □

Example 79. Let $N=2$, $s=1$. In this case Theorem iii) gives us the following decomposition of $M(O,\chi_s)$:

$$M(O,\chi_s) = \oplus_{j\geq 1} N_j \otimes V_{2j-1},$$

where V_j is the irreducible representation of $\mathfrak{sl}_2$ of dimension $j+1$, and the spaces N_j satisfy the equation

$$\mathrm{Tr}_{|N_j}(q^{2H}) = \frac{q^{2j}-q^{-2j}}{q^2-q^{-2}}.$$

This shows that $N_j = V_{j-1}$ as a representation of the $\mathfrak{sl}_2$-subalgebra spanned by E, F, H, which commutes with $\mathfrak{g}$.

9.13 The case of general orbits

Let $W = S_N$, the Weyl group of G, $\lambda \in \mathfrak{h}/W$, and let $\mathcal{N}_\lambda$ be the closure in $\mathfrak{g}$ of the adjoint orbit of a regular element of $\mathfrak{g}$ whose semisimple part is λ. Denote by $\mathcal{D}_G(\mathcal{N}_\lambda)$ the category of G-equivariant D-modules on G that are concentrated on $\mathcal{N}_\lambda$. We also let $\mathcal{O}_\lambda$ be the category of finitely generated $H_n(k)$-modules in which the subalgebra $\mathbb{C}[\mathrm{y}_1,\ldots,\mathrm{y}_n]^{S_n}$ acts through the character λ. Then one can show, similarly to the above, that the functor F_n^* restricts to a functor $F_{n,\lambda}^* : \mathcal{D}_G(\mathcal{N}_\lambda) \to \mathcal{O}_\lambda$. The functor considered above is $F_{n,0}^*$. We plan to study the functor $F_{n,\lambda}^*$ for general λ in a future work.

9.14 The trigonometric case

Our results about rational Cherednik algebras can be extended to the trigonometric case. For this purpose, D-modules on the Lie algebra $\mathfrak{g}$ should be replaced with D-modules on the group G. Let us describe this generalization.

First, let us introduce some notation. As above, we let $G = \mathrm{SL}_N(\mathbb{C})$. For $b \in \mathfrak{g}$, let L_b be the right-invariant vector field on G equal to b at the identity element; that is, L_b generates the group of left translations by e^{tb}. As before, we let $k = N/n$.

Now let M be a D-module on G. Similarly to the above, we define $F_n(M)$ to be the space

$$F_n(M) = (M \otimes (\mathbb{C}^N)^{\otimes n})^G,$$

where G acts on itself by conjugation.

Consider the operators X_i, Y_i, $i = 1, \ldots, n$, on $F_n(M)$, defined by formulas similar to the rational case:

$$X_i = \sum_{j,l} A_{jl} \otimes (E_{lj})_i, \ Y_i = \frac{N}{n} \sum_p L_{b_p} \otimes (b_p)_i,$$

where A_{jl} is the jl-th matrix element of $A \in G$ regarded as the multiplication operator in M by a regular function on G.

Proposition 80. *The operators X_i, Y_i satisfy the following relations:*

$$\prod_i X_i = 1, \ \sum_i Y_i + k \sum_{i<j} s_{ij} = 0,$$

$$s_{ij} X_i = X_j s_{ij}, \ s_{ij} Y_i = Y_j s_{ij}, \ [s_{ij}, X_l] = [s_{ij}, Y_l] = 0,$$
$$[X_i, X_j] = 0, \ [Y_i, Y_j] = k s_{ij} (Y_i - Y_j),$$
$$[Y_i, X_j] = \left(k s_{ij} - \frac{1}{n} \right) X_j,$$

where i, j, l denote distinct indices.

Proof. Straightforward computation. □

Corollary 81. *The operators $\bar{Y}_i = Y_i + k \sum_{j<i} s_{ij}$ pairwise commute.*

The relations of Proposition 80 are nothing but the defining relations of the *degenerate double affine Hecke algebra* of type A_{n-1}, which we will denote by $H_n^{\mathrm{tr}}(k)$ (where "tr" stands for trigonometric, to illustrate the fact that this algebra is a trigonometric deformation of the rational Cherednik algebra $H_n(k)$). Thus we have defined an exact functor $F_n : \mathcal{D}(G) \to H_n^{\mathrm{tr}}(k)$-mod. As before, it is sufficient to consider the restriction of this functor to the category of equivariant finitely generated D-modules, $\mathcal{D}_G(G)$.

This allows us to generalize much of our story for rational Cherednik algebras to the trigonometric case. In particular, let $\mathcal{U}$ be the unipotent variety on G, and let $\mathcal{D}_G(\mathcal{U})$ be the category of finitely generated G-equivariant D-modules on G concentrated on $\mathcal{U}$. If we restrict the functor F_n to this category, we get a situation identical to that in the rational case. Indeed, one can show that for any M in this category, $F_n(M)$ belongs to the category $\mathcal{O}_-^{\mathrm{tr}}$ of finitely generated modules over $H_n^{\mathrm{tr}}(k)$ that are locally unipotent with respect to the action of X_i. The latter category is equivalent to the category $\mathcal{O}_-$ over the rational Cherednik algebra $H_n(k)$, because the completion of $H_n^{\mathrm{tr}}(k)$ with respect to the ideal generated by $X_i - 1$ is isomorphic to the completion of $H_n(k)$ with respect to the ideal generated by x_i. On the other hand, the exponential map identifies the categories $\mathcal{D}_G(\mathcal{U})$ and $\mathcal{D}_G(\mathcal{N})$. It is clear that after we make these two identifications, the functor F_n becomes the functor F_n in the rational case that we considered above.

On the other hand, because of the absence of a Fourier transform on the group (as opposed to Lie algebra), the trigonometric story is richer than the rational one. Namely, we can consider another subcategory of $\mathcal{D}_G(G)$, the category of character sheaves. By definition, a character sheaf on G is an object M in $\mathcal{D}_G(G)$ that is locally finite with respect to the action of the algebra of bi-invariant differential operators, $U(\mathfrak{g})^G$. This category is denoted by $\mathrm{Char}(G)$. It is known that one has a decomposition

$$\mathrm{Char}(G) = \oplus_{\lambda \in T^\vee/W} \mathrm{Char}_\lambda(G),$$

where $T^\vee$ is the dual torus, and $\mathrm{Char}_\lambda(G)$ the category of those $M \in \mathcal{D}_G(G)$ for which the generalized eigenvalues of $U(\mathfrak{g})^G$ (which we identify with $U(\mathfrak{h})^W$ via the Harish-Chandra homomorphism) project to λ under the natural projection $\mathfrak{h}^* \to T^\vee$.

On the other hand, one can define the category $\mathrm{Rep}_{Y-\mathrm{fin}}\,(H_n^{\mathrm{tr}}(k))$ of modules over $H_n^{\mathrm{tr}}(k)$ on which the commuting elements $\bar{Y}_i$ act in a locally finite manner. We have a similar decomposition

$$\mathrm{Rep}_{Y-\mathrm{fin}}\left(H_n^{\mathrm{tr}}(k)\right) = \oplus_{\lambda \in T^\vee/W} \mathrm{Rep}_{Y-\mathrm{fin}}\left(H_n^{\mathrm{tr}}(k)\right)_\lambda,$$

where $\mathrm{Rep}_{Y-\mathrm{fin}}\,(H_n^{\mathrm{tr}}(k))_\lambda$ is the subcategory of all objects where the generalized eigenvalues of $\bar{Y}_i$ project to $\lambda \in T^\vee/W$. Then one can show, similarly to the rational case, that the functor F_n gives rise to the functors

$$F_{n,\lambda} : \mathrm{Char}_\lambda(G) \to \mathrm{Rep}_{Y-fin}\left(H_n^{\mathrm{tr}}(k)\right)_\lambda$$

for each $\lambda \in T^\vee/W$. The most interesting case is $\lambda = 0$ (unipotent character sheaves). We plan to study these functors in subsequent works.

9.15 Relation with the Arakawa–Suzuki functor

Note that the elements Y_i and s_{ij} generate the degenerate affine Hecke algebra $\mathcal{H}_n$ of Drinfeld and Lusztig (of type A_{n-1}). To define the action of this algebra on $F_n(M) = (M \otimes (\mathbb{C}^N)^{\otimes n})^{\mathfrak{g}}$ by the formula of Proposition 80, we need only the action of the operators L_b, $b \in \mathfrak{g}$ in M. So M can be taken to be an arbitrary $\mathfrak{g}$-bimodule that is locally finite with respect to the diagonal action of $\mathfrak{g}$ (in this case, $\sum_i Y_i + \sum_{i<j} s_{ij}$ is a central element that does not necessarily act by zero, so we get a representation of a central extension $\widetilde{\mathcal{H}}_n$ of $\mathcal{H}_n$). In particular, we have an exact functor $F_n : \mathrm{HC}(\mathfrak{g}) \to \widetilde{\mathcal{H}}_n$-mod from the category of Harish-Chandra bimodules over $\mathfrak{g}$ to the category of finite dimensional representations of the degenerate affine Hecke algebra $\widetilde{\mathcal{H}}_n$. This functor was essentially considered in [AS98] (where it was applied to the Harish-Chandra modules of the form $M = \mathrm{Hom}_{\mathfrak{g}-\mathrm{finite}}(M_1, M_2)$, where M_1 and M_2 are modules from the category $\mathcal{O}$ over $\mathfrak{g}$). We note that the paper [AST96] describes the extension of this construction to affine Lie algebras, which yields representations of degenerate double affine Hecke algebras.

9.16 Directions of further study

In conclusion we would like to discuss (in a fairly speculative manner) several directions of further study and generalizations (we note that these generalizations can be combined with each other).

1. The q-case: the group G is replaced with the corresponding quantum group, D-modules with q-D-modules, and degenerate double affine Hecke algebras with the usual double affine Hecke algebras (defined by Cherednik). It is especially interesting to consider this generalization if q is a root of unity.

2. The quiver case. This generalization was suggested by Ginzburg, and will be studied in his subsequent work with the third author. In this case, one has a finite subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{C})$, and one should consider equivariant D-modules on the representation space of the affine quiver attached to Γ (with some orientation). Then there should exist an analogue of the functor F_n, which takes values in the category of representations of an appropriate symplectic reflection algebra for the wreath product $S_n \ltimes \Gamma^n$ [EG02] (or, equivalently, the Gan–Ginzburg algebra [GG05]). This generalization should be especially nice in the case when Γ is a cyclic group, when the symplectic reflection algebra is a Cherednik algebra for a complex reflection group, and one has the notion of category $\mathcal{O}$ for it.

3. The symmetric space case. This is the trigonometric version of the previous generalization for $\Gamma = \mathbb{Z}/2$. In this generalization one considers (monodromic) equivariant D-modules on the symmetric space $\mathrm{GL}_{p+q}(\mathbb{C})/(\mathrm{GL}_p \times \mathrm{GL}_q)(\mathbb{C})$ (see [Gin89]), and one expects a functor from this category to the category of representations of an appropriate degenerate double affine Hecke algebra of type $C^\vee C_n$. This functor should be related, similarly to the previous subsection, to an analogue of the Arakawa–Suzuki functor, which would attach to a Harish-Chandra module for the pair $(\mathrm{GL}_{p+q}(\mathbb{C}), \mathrm{GL}_p(\mathbb{C}) \times \mathrm{GL}_q(\mathbb{C}))$, a finite dimensional representation of the degenerate double affine Hecke algebra of type BC_n.

A

Let $\mathcal{O}$ be the ring $\mathbb{C}[[u_1, \dots, u_n]][\ell_1, \dots, \ell_n]$. Define commuting derivations D_i of $\mathcal{O}$ by $D_i(u_j) = \delta_{ij} u_i$, $D_i(\ell_j) = \delta_{ij}$ (we will later think of ℓ_i and D_i as $\log u_i$ and $u_i \frac{\partial}{\partial u_i}$).

We set $\mathcal{O}_+ := \mathfrak{m}[\ell_1, \dots, \ell_n]$, where $\mathfrak{m} = \mathrm{Ker}(\mathbb{C}[[u_1, \dots, u_n]] \to \mathbb{C})$ is the augmentation ideal. Let $A = \oplus_{k \geq 0} A_k$ be a graded ring with finite-dimensional homogeneous components.

Proposition 83. *Let $X_i(u_1, \dots, \ell_n) \in \hat{\oplus}_{k>0}(A_k \otimes \mathcal{O}_+)$ be such that $D_i(X_j) = D_j(X_i)$. Then there exists a unique $F(u_1, \dots, \ell_n) \in \hat{\oplus}_{k>0}(A_k \otimes \mathcal{O}_+)$ such that $D_i(F) = X_i$ for $i = 1, \dots, n$.*

Let us say that $f \in \mathcal{O}$ has radius of convergence $R > 0$ if $f = \sum_{k_1, \dots, k_n \geq 0} f_{k_1, \dots, k_n}(u_1, \dots, u_n) \ell_1^{k_1} \cdots \ell_n^{k_n}$, where each $f_{k_1, \dots, k_n}(u_1, \dots, u_n)$

converges for $|u_1|, \ldots, |u_n| \leq R$. Then if $X_1, \ldots, X_n$ have radius of convergence R, so does F.

Proof. For each i, D_i restricts to an endomorphism of $\mathcal{O}_+$; one checks that $\cap_{i=1}^n \mathrm{Ker}(D_i : \mathcal{O}_+ \to \mathcal{O}_+) = 0$ which implies the uniqueness. To prove the existence, we work by induction. One proves that $D_n : \mathcal{O}_+ \to \mathcal{O}_+$ is surjective, and its kernel is $\mathfrak{m}_{n-1}[\ell_1, \ldots, \ell_{n-1}]$, where $\mathfrak{m}_{n-1} = \mathrm{Ker}(\mathbb{C}[[u_1, \ldots, u_{n-1}]] \to \mathbb{C})$. Let G be a solution of $D_n(G) = X_n$. Then the system $D_i(F') = X_i - D_i(G)$ $(i = 1, \ldots, n)$ is compatible, which implies $D_n(X_i') = 0$, where $X_i' := X_i - D_i(G)$, so $X_i' \in \oplus_{k>0} \left(A_k \otimes \mathcal{O}_+^{(n-1)}\right)$, where $\mathcal{O}_+^{(n-1)}$ is the analogue of $\mathcal{O}_+$ at order $n-1$. Hence the system $D_i(F') = X_i - D_i(G)$ $(i = 1, \ldots, n-1)$ is compatible and we may apply to it the result at order $n-1$ to obtain a solution F'. Then a solution of $D_i(F) = X_i$ is $F' + G$.

Let $D : u\mathbb{C}[[u]] \to u\mathbb{C}[[u]]$ be the map $u\frac{\partial}{\partial u}$ and let $I := D^{-1}$. The map $D_1 : u\mathbb{C}[[u]][\ell] \to u\mathbb{C}[[u]][\ell]$ is bijective and its inverse is given by $D_1^{-1}(F(u)\ell^a) = \sum_{k=0}^{a} (-1)^k a(a-1)\cdots(a-k+1)(I^{k+1}(F))(u)\ell^{a-k}$.

We have $\mathcal{O}_+ = \mathcal{O}^{(n-1)} \hat{\otimes} u_n \mathbb{C}[[u_n]][\ell_n] \oplus \mathfrak{m}^{(n-1)} \hat{\otimes} \mathbb{C}[\ell_n]$ (where $\mathcal{O}^{(n-1)}$, $\mathfrak{m}^{(n-1)}$ are the analogues of $\mathcal{O}, \mathfrak{m}$ at order $n-1$, $\hat{\otimes}$ is the completed tensor product). The endomorphism D_n preserves this decomposition and a section of D_n is given by $\left(\mathrm{id} \otimes D_1^{-1}\right) \oplus (\mathrm{id} \otimes J)$, where $J \in \mathrm{End}(\mathbb{C}[\ell])$ is a section of $\partial/\partial\ell$.

It follows from the fact that I preserves the radius of convergence of a series that the same holds for the section of D_n defined above. One then follows the above construction of a solution X of $D_i(X) = X_i$ and uses the fact that D_i also preserves the radius of convergence to show by induction that X has radius R if the X_i do. ◻

Proposition 84. *Let $X_i(u_1, \ldots, \ell_n) \in \hat{\oplus}_{k>0}(A_k \otimes \mathcal{O}_+)$ be such that $D_i(X_j) - D_j(X_i) = [X_i, X_j]$. Then there exists a unique $F(u_1, \ldots, \ell_n) \in 1 + \hat{\oplus}_{k>0}(A_k \otimes \mathcal{O}_+)$ such that $D_i(F) = X_i F$ for $i = 1, \ldots, n$. If the X_i have radius R, then so does F.*

Proof. Let us prove the uniqueness. If F, F' are two solutions, then $F^{-1}F'$ is a constant (since $\cap_{i=0}^n \mathrm{Ker}(D_i : \mathcal{O} \to \mathcal{O}) = 0$), and it also belongs to $1 + \hat{\oplus}_{k>0}(A_k \otimes \mathcal{O}_+)$, which implies that $F = F'$. To prove the existence, one sets $F = 1 + f_1 + f_2 + \cdots$, $X_i = x_1^{(i)} + \cdots$, where $f_k, x_k^{(i)} \in A_k \otimes \mathcal{O}_+$ and solves by induction the system $D_i(f_k) = x_1^{(i)} f_{k-1} + \cdots + x_k^{(i)}$ using Proposition 83. ◻

Proposition 85. *Let $C_i(u_1, \ldots, u_n) \in \hat{\oplus}_{k>0} A_k[[u_1, \ldots, u_n]]$ $(i = 1, \ldots, n)$ be such that $u_i \partial_{u_i}(C_j) - u_j \partial_{u_j}(C_i) = [C_i, C_j]$ for any i, j. Assume that the series C_i have radius R. Then there exists a unique solution of the system $u_i \partial_{u_i}(X) = C_i X$, analytic in the domain $\{u | |u| \leq R, u \notin \mathbb{R}_-\}^n$, such that the ratio $\left(u_1^{C_0^1} \cdots u_n^{C_0^n}\right)^{-1} X(u_1, \ldots, u_n)$ (we set $C_0^i := C_i(0, \ldots, 0)$) has the form $1 + \sum_{k>0} \sum_{a_1, \ldots, a_n, i} r_k^{a_1, \ldots, a_n, i}(u_1, \ldots, u_n)$ (the second sum is finite for any* *k), $r_k^{a_1,\dots,a_n,i}$ has degree k, $a_i \geq 0$, $i \in \{1,\dots,n\}$, and $r_k^{a_1,\dots,a_n,i}(u_1,\dots,u_n) = O(u_i(\log u_1)^{a_1}\cdots(\log u_n)^{a_n})$.*

The same is then true of the ratio $X(u_1,\dots,u_n)\left(u_1^{C_0^1}\cdots u_n^{C_0^n}\right)^{-1}$; we write $X(u_1,\dots,u_n) \simeq u_1^{C_0^1}\cdots u_n^{C_0^n}$.

Proof. Let us show the existence of X. The compatibility condition implies that $\left[C_0^i, C_0^j\right] = 0$. If we set $Y(u_1,\dots,u_n) := \left(u_1^{C_0^1}\cdots u_n^{C_0^n}\right)^{-1} X(u_1,\dots,u_n)$, then X is a solution iff Y is a solution of $u_i\partial_{u_i}(Y) = \exp\left(-\sum_{j=1}^n(\log u_j)C_j^0\right)\left(C_i - C_i^0\right)\cdot Y$.

Let us set

$$X_i(u_1,\dots,\ell_n) := \exp\left(-\sum_{j=1}^n \ell_j C_j^0\right)(C_i(u_1,\dots,u_n) - C_i(0,\dots,0));$$

then $X_i(u_1,\dots,\ell_n) \in \hat{\oplus}_{k>0}(A_k \otimes \mathcal{O}_+)$. We then apply Proposition 84 and find a solution $Y \in 1 + \hat{\oplus}_{k>0}A_k \otimes \mathcal{O}_+$ of $D_i(Y) = X_iY$. Let Y_k be the component of Y of degree k. Since Y has radius R, the replacement $\ell_i = \log u_i$ in Y_k for $u_i \in \{u||u| \leq R, u \notin \mathbb{R}_-\}$ gives an analytic function on $\{u||u| \leq R, u \notin \mathbb{R}_-\}^n$. Moreover, $\mathcal{O}_+ = \sum_{i=1}^n u_i\mathbb{C}[[u_1,\dots,u_n]][\ell_1,\dots,\ell_n]$, which gives a decomposition $Y_k = \sum_{i,a_1,\dots,a_n} u_i\ell_1^{a_1}\cdots\ell_n^{a_n}y^k_{i,a_1,\dots,a_n}(u_1,\dots,u_n)$ and leads (after substitution $\ell_i = \log u_i$) to the above estimates.

The ratio $X(u_1,\dots,u_n)\left(u_1^{C_0^1}\cdots u_n^{C_0^n}\right)^{-1}$ is then $1 + \exp\left(\sum_j C_0^j \log u_j\right)(Y(u_1,\dots,u_n) - 1)$; the term of degree k has finitely many contributions to which we apply the above estimates.

Let us prove the uniqueness of X. Any other solution has the form $X = X(1 + c_k + \cdots)$ where $c_j \in A_j$, and $c_k \neq 0$. Then the degree k term is transformed by the addition of c_k, which cannot be split as a sum of terms in the various $O(u_i(\log u_1)^{a_1}\cdots(\log u_n)^{a_n})$. □

Acknowledgments. This project started in June 2005, while the three authors were visiting ETH; they would like to thank Giovanni Felder for his kind invitations.

P.E. is deeply grateful to Victor Ginzburg for a great deal of help with proofs of the main results in the part about equivariant D-modules, and for explanations on the Gan–Ginzburg functors. Without this help, this part could not have been written. P.E. is also very grateful to G. Lusztig for explaining to him the proof of Theorem 61, and to V. Ostrik and D. Vogan for useful discussions.

The work of D.C. has been partially supported by the European Union through the FP6 Marie Curie RTN ENIGMA (contract number MRTN-CT-2004-5652).

References

[AS98] Arakawa, T., Suzuki, T., *Duality between $\mathfrak{sl}_n(\mathbb{C})$ and the degenerate affine Hecke algebra*. J. Algebra **209**, no. 1, 288–304, (1998).

[AST96] Arakawa, T., Suzuki, T., Tsuchiya, A., *Degenerate double affine Hecke algebra and conformal field theory*. In: Kashiwara, Masaki (ed.) et al., Topological field theory, primitive forms and related topics. Birkhäuser, Boston. Prog. Math. **160**, 1–34, (1998).

[BK00] Bakalov, B., Kirillov (Jr.), A., *On the Lego-Teichmüller game*. Transform. Groups **5**, no. 3, 207–244, (2000).

[BEG03a] Berest, Y., Etingof, P., Ginzburg, V., *Finite-dimensional representations of rational Cherednik algebras*. Int. Math. Res. Not. **19**, 1053–1088, (2003).

[BEG03b] Berest, Y., Etingof, P., Ginzburg, V., *Cherednik algebras and differential operators on quasi-invariants*. Duke Math. J. **118**, no. 2, 279–337, (2003).

[Ber98a] Bernard, D., *On the Wess-Zumino-Witten model on the torus*. Nucl. Phys. **B 303**, 77–93, (1998).

[Ber98b] Bernard, D., *On the Wess-Zumino-Witten model on Riemann surfaces*. Nucl. Phys. **B 309**, 145–174, (1998).

[Bez94] Bezrukavnikov, R., *Koszul* DG-*algebras arising from configuration spaces*. Geom. and Funct. Analysis **4**, no. 2, 119–135, (1994).

[BE09] Bezrukavnikov, R., P. Etingof, *Parabolic induction and restriction functors for rational Cherednik algebras*, math/0803.3639, Selecta Math., **14**, nos. 3–4, 397–425, May (2009).

[Bir69a] Birman, J., *On braid groups*. Commun. Pure Appl. Math. **12**, 41–79, (1969).

[Bir69b] Birman, J., *Mapping class groups and their relationship to braid groups*. Commun. Pure Appl. Math. **12**, 213–38 (1969).

[CE03] Chmutova, T., Etingof, P., *On some representations of the rational Cherednik algebra*. Represent. Theory **7**, 641–650, (2003).

[Che97] Cherednik, I., *Intertwining operators of double affine Hecke algebras*. Selecta Math. (N.S.) **3**, no. 4, 459–495, (1997).

[Che03] Cherednik, I., *Double affine Hecke algebras and difference Fourier transforms*. Invent. Math. **152**, no. 2, 213–303, (2003).

[Dri90a] Drinfeld, V., *On almost cocommutative Hopf algebras*. Leningrad Math. J. **1**, no. 2, 321–342, (1990).

[Dri90b] Drinfeld, V., *Quasi-Hopf algebras*. Leningrad Math. J. **1**, no. 6, 1419–1457, (1990).

[Dri91] Drinfeld, V., *On quasitriangular quasi-Hopf algebras and a group closely connected with* $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$. Leningrad Math. J. **2**, 829–860, (1991).

[EE05] Enriquez, B., Etingof, P., *Quantization of classical dynamical r-matrices with nonabelian base*. Commun. Math. Phys. **254**, no. 3, 603–650, (2005).

[Eti94] Etingof, P., *Representations of affine Lie algebras, elliptic r-matrix systems, and special functions*. Commun. Math. Phys. **159**, no. 3, 471–502, (1994).

[EG02] Etingof, P., Ginzburg, V., *Symplectic reflection algebras, Calogero-Moser space, and deformed Harish-Chandra homomorphism*. Invent. Math. **147**, no. 2, 243–348, (2002).

[FW96] Felder, G., Wieczerkowski, C., *Conformal blocks on elliptic curves and the Knizhnik-Zamolodchikov-Bernard equations.* Commun. Math. Phys. **176**, no. 1, 133–161, (1996).

[GG04] Gan, W.L., Ginzburg, V., *Almost-commuting variety, D-modules, and Cherednik Algebras.* IMRP Int. Math. Res. Pap. **26439**, 1–54, (2006).

[GG05] Gan, W.L., Ginzburg, V., *Deformed preprojective algebras and symplectic reflection algebras for wreath products.* J. Algebra **283**, no. 1, 350–363, (2005).

[Gin89] Ginzburg, V., *Admissible modules on a symmetric space.* In: Orbites unipotentes et reprÃĹsentations. III. Orbites et faisceaux pervers. Astérisque **173-174**, 199–255, (1989).

[GGOR03] Ginzburg, V., Guay, N., Opdam, E., Rouquier, R., *On the category $\mathcal{O}$ for rational Cherednik algebras.* Invent. Math. **154**, no. 3, 617–651, (2003).

[GS05] Gordon, I., Stafford, J.T., *Rational Cherednik algebras and Hilbert schemes.* Adv. Math. **198**, no. 1, 222–274, (2005).

[HLS00] Hatcher, A., Lochak, P., Schneps, L., *On the Teichmüller tower of mapping class groups.* J. Reine Angew. Math. **521**, 1–24, (2000).

[Hes80] Hesselink, W.H., *Characters of the nullcone*; Math. Ann. **252**, no. 3, 179–182, (1980).

[HK84] Hotta, R., Kashiwara, M., *The invariant holonomic system on a semisimple Lie algebra.* Invent. Math. **75**, no. 2, 327–358, (1984).

[Jos97] Joseph, A., *On a Harish-Chandra homomorphism.* C. R. Acad. Sci. Paris **324**, 759–764, (1997).

[KZ84] Knizhnik V., Zamolodchikov, A., *Current algebras and the Wess-Zumino model in two dimensions.* Nuclear Phys. **B 247**, 83–103, (1984).

[Koh83] Kohno, T., *On the holonomy Lie algebra and the nilpotent completion of the fundamental group of the complement of hypersurfaces.* Nagoya Math. J. **92**, 21–37, (1983).

[Kos63] Kostant, B., *Lie group representations on polynomial rings.* Amer. J. Math. **85**, 327–404, (1963).

[Kri94] Kriz, I., *On the rational homotopy type of configuration spaces.* Annals Math. **139**, 227–237, (1994).

[Lus81] Lusztig, G., *Singularities, character formulas, and a q-analog of weight multiplicities.* In: Analysis and topology on singular spaces, II, III (Luminy, 1981), 208–229.

[Lus88] Lusztig, G., *Cuspidal local systems and graded Hecke algebras. I.* Publ. Math. IHÉS **67**, 145–202, (1988).

[Lus85] Lusztig, G., *Character sheaves. II.* Adv. in Math. **57**, no. 3, 226–265, (1985).

[Lev98] Levasseur, T., *Equivariant D-modules attached to nilpotent orbits in a semisimple Lie algebra.* Transformation Groups **3**, 337–353, (1998).

[LS97] Levasseur, T., Stafford, J.T., *Semi-simplicity of invariant holonomic systems on a reductive Lie algebra.* Amer. J. Math. **119**, 1095–1117, (1997).

[Man05] Manin, Y.I., *Iterated integrals of modular forms and noncommutative modular symbols.* In: Algebraic geometry and number theory. Birkhäuser, Boston. Progr. Math. **253**, 565–597, (2006).

[Mir04] Mirkovic, I., *Character sheaves on reductive Lie algebras.* Mosc. Math. J. **4**, no. 4, 897–910, (2004).

[Rou05] Rouquier, R., *q-Schur algebras and complex reflection groups.* Mosc. Math. J. **8**, no. 1, 119–158 (2008).

[Sul77] Sullivan, D., *Infinitesimal computations in topology.* Publ. Math. IHÉS, **47**, 260–331 (1977).

Exact Category of Modules of Constant Jordan Type

Jon F. Carlson[1,2] and Eric M. Friedlander[3,4]

[1] Department of Mathematics, University of Georgia
jfc@math.uga.edu
[3] Department of Mathematics, University of Southern California, Los Angeles, CA 90089, ericmf@usc.edu

To Yuri Manin with admiration

Summary. For a finite group scheme G, we continue our investigation of those finite-dimensional kG-modules that are of constant Jordan type. We introduce a Quillen exact category structure $\mathcal{C}(kG)$ on these modules and investigate $K_0(\mathcal{C}(kG))$. We study which Jordan types can be realized as the Jordan types of (virtual) modules of constant Jordan type. We also briefly consider thickenings of $\mathcal{C}(kG)$ inside the triangulated category $\mathrm{stmod}(kG)$.

Key words: finite group scheme, Jordan type, exact category.

2000 Mathematics Subject Classifications: 16G10, 20C20, 20G10, 16D90, 14D15, 19A31

1 Introduction

Together with Julia Pevtsova, the authors introduced in [6] an intriguing class of modules for a finite group G (or, more generally, for an arbitrary finite group scheme), the kG-modules of constant Jordan type. This class includes projective modules and endotrivial modules. It is closed under taking direct sums, direct summands, k-linear duals, and tensor products. We have several methods for constructing modules of constant Jordan type, typically using cohomological techniques. In the very special case that $G = \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, the authors and Andrei Suslin have recently introduced several interesting constructions that associate modules of constant Jordan type to an arbitrary finite-dimensional kG-module and have identified cyclic kG-modules of constant Jordan type [7].

[2]Partially supported by the NSF.
[4]Partially supported by NSF Grant #03000525.

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*,
Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_6,

What strikes us as remarkable is how challenging the problem of classifying modules of constant Jordan type is even for relatively simple finite group schemes. In this paper we address two other aspects of the theory that also present formidable challenges. The first is the realization problem of determining which Jordan types can actually occur for modules of constant Jordan type. The second question concerns stratification of the entire module category by modules of constant Jordan type.

To consider realization, we give the class of kG-modules of constant Jordan type the structure of a Quillen exact category $\mathcal{C}(kG)$ using "locally split short exact sequences." This structure suggests itself naturally once kG-modules are treated from the point of view of π-points as in [11], a point of view necessary to even define modules of constant Jordan type. With respect to this exact category structure, the Grothendieck group $K_0(\mathcal{C}(kG))$ arises as a natural invariant. There are natural Jordan type functions JType, $\overline{\text{JType}}$ defined on $K_0(\mathcal{C}(kG))$ that are useful for formulating questions of realizability of (virtual) modules of constant Jordan type. The reader will find several results concerning the surjectivity of these functions.

A seemingly very difficult goal is the classification of kG-modules of constant Jordan type, or at least the determination of $K_0(\mathcal{C}(kG))$. In this paper, we provide a calculation of $K_0(\mathcal{C}(kG))$ for two very simple examples: the Klein four group and the first infinitesimal kernel of SL_2.

The category $\mathcal{C}(kG)$ possesses many closure properties. However, the complexity of this category is reflected in the observation that an extension of modules of constant Jordan type need not be of constant Jordan type. We conclude this paper by a brief consideration of a stratification of the stable module category stmod(kG) by "thickenings" of $\mathcal{C}(kG)$.

We are very grateful to Julia Pevtsova and Andrei Suslin for many discussions. This paper is part of a longer-term project that will reflect their ideas and constructions.

2 The exact category $\mathcal{C}(kG)$

As shown in [6], there is a surprising array of kG-modules of constant Jordan type. One evident way to construct new examples out of old is to use locally split extensions (see, for example, Proposition 2.4). In order to focus on examples that seem more essential, we introduce in Definition 2.3 the Quillen exact category $\mathcal{C}(kG)$ of modules of constant Jordan type whose admissible short exact sequences are those that are locally split.

We begin by recalling the definition of a π-point of a finite group scheme over k. This is a construction that is necessary to formulate the concept of modules of constant Jordan type.

Definition 2.1. *Let G be a finite group scheme with group algebra kG (the linear dual of the coordinate algebra $k[G]$). A π-point of G is a map of K-algebras*

$\alpha_K : K[t]/t^p \to KG_K$ that is left flat and that factors through some abelian unipotent subgroup scheme $U_K \subset G_K$; here K is an arbitrary field extension of k and G_K is the base extension of G along K/k.

Two π-points α_K, β_L of G are said to be equivalent (denoted by $\alpha_K \sim \beta_L$) provided that for every finite-dimensional kG-module M the $K[t]/t^p$-module $\alpha_K^*(M_K)$ is projective if and only if the $L[t]/t^p$-module $\beta_L^*(M_L)$ is projective.

In [11], it is shown that the set of equivalence classes of π-points of G admits a scheme structure that is defined in terms of the representation theory of G and that is denoted by $\Pi(G)$. Moreover, it is verified that this scheme is isomorphic to the projectivization of the affine scheme of $\mathrm{H}^\bullet(G,k)$,

$$\Pi(G) \;\cong\; \mathrm{Proj}\ \mathrm{H}^\bullet(G,k).$$

Here, $\mathrm{H}^\bullet(G,k)$ is the finitely generated commutative k-algebra defined to be the cohomology algebra $\mathrm{H}^*(G,k)$ if $p = \mathrm{char}(k)$ equals 2 and to be the subalgebra of $\mathrm{H}^*(G,k)$ generated by homogeneous classes of even degree if $p > 2$.

We next introduce admissible monomorphisms and admissible epimorphisms, formulated in terms of π-points.

Definition 2.2. *Let G be a finite group scheme. A short exact sequence of kG-modules $0 \to M_1 \to M_2 \to M_3 \to 0$ is said to be locally split if its pull-back via any π-point $\alpha_K : K[t]/t^p \to KG$ is split as a short exact sequence of $K[t]/t^p$-modules. We shall frequently refer to such a locally split short exact sequence as an admissible sequence.*

Moreover, we say that a monomorphism $f : M_1 \to M_2$ of kG-modules is an admissible monomorphism if it can be completed to a locally split short exact sequence. Similarly, an epimorphism $g : M_2 \to M_3$ is said to be an admissible epimorphism if it can be completed to a locally split short exact sequence.

We typically identify an admissible monomorphism with the inclusion of the image of the injective map $f : M \to N$.

The objects of our study are kG-modules of constant Jordan type as defined below (and introduced in [6]). We recall that a finite-dimensional $K[t]/t^p$ module M of dimension N is isomorphic to

$$a_p[p] + \cdots + a_i[i] + \cdots + a_1[1], \quad \sum_i a_i \cdot i = N,$$

where $[i] = K[t]/t^i$ is the indecomposable $K[t]/t^p$-module of dimension i. We refer to the p-tuple $(a_p, \ldots, a_1)$ as the Jordan type of M and designate this Jordan type by $\mathrm{JType}(M)$; the $(p-1)$-tuple $(a_{p-1}, \ldots, a_1)$ will be called the stable Jordan type of M.

Definition 2.3. *A finite-dimensional module M for a finite group scheme G is said to be of* **constant Jordan type** *if the Jordan type of $\alpha_K^*(M_K)$ is independent of the choice of the π-point α_K of G.*

In the following proposition, we see that the class of modules of constant Jordan type is closed under locally split extensions.

Proposition 2.4. *Let G be a finite group scheme and let $\mathcal{E}$ denote a short exact sequence $0 \to M_1 \to M_2 \to M_3 \to 0$ of finite-dimensional kG-modules.*

1. *Assume that $\mathcal{E}$ is locally split. Then M_1 and M_3 are of constant Jordan type if and only if M_2 is of constant Jordan type.*
2. *If M_1 and M_3 are modules of constant Jordan type, then $\mathcal{E}$ is locally split if and only if $\alpha_K^*(M_{2,K}) \simeq \alpha_K^*(M_{1,K}) \oplus \alpha_K^*(M_{3,K})$ for some representative α_K of each generic point of $\Pi(G)$.*

Proof. Observe that a short exact sequence of finite-dimensional $K[t]/t^p$-modules $0 \to N_1 \to N_2 \to N_3 \to 0$ is split if and only if $\mathrm{JType}(N_2) = \mathrm{JType}(N_1) + \mathrm{JType}(N_3)$. Thus, if $\mathcal{E}$ is locally split and if $M_1,\ M_3$ have constant Jordan type, then M_2 does as well.

If $\mathcal{E}$ is locally split and M_2 has constant Jordan type, then the proof that both M_1 and M_3 have constant Jordan type is verified using the same argument as that of [6, 3.7] using [6, 3.5]; the point is that the Jordan type of M_1 at some representative of a generic point of $\Pi(G)$ must be greater than or equal to the Jordan type of M_1 at some representative of any specialization. The same applies to M_2 and M_3. As shown in [12, 4.2], the Jordan type of any finite-dimensional kG-module at a generic point of $\Pi(G)$ is independent of the choice of π-point representing that generic point.

If $0 \to V_1 \to V_2 \to V_3 \to 0$ is a short exact sequence of $k[t]/t^p$-modules, then the Jordan type of V_2 must be greater than or equal to the sum of the Jordan types of V_1 and V_3. Thus, the observation of the preceding paragraph verifies the following: assume that $\mathcal{E}$ is a short exact sequence of kG-modules with M_1, M_3 of constant Jordan type; if M_2 has the minimal possible Jordan type (namely the sum of the Jordan types of M_1 and M_3) at each generic π-point of G, then M_2 must have constant Jordan type. □

As the following proposition asserts, admissible monomorphisms and admissible epimorphisms as in Definition 2.2 are associated to structures of exact categories (in the sense of Quillen [18]) on the category $\mathrm{mod}(kG)$ of finite-dimensional kG-modules and the full subcategory $\mathcal{C}(kG)$ of modules of constant Jordan type.

Proposition 2.5. *The collection $\underline{E}$ of locally split short exact sequences of finite-dimensional kG-modules constitutes a class of admissible sequences providing $\mathrm{mod}(kG)$ with the structure of an exact category in the sense of Quillen (cf. [18]).*

Similarly, the class $\underline{E}_{\mathcal{C}}$ of locally split short exact sequences of kG-modules of constant Jordan type also constitutes a class of admissible sequences, thereby providing $\mathcal{C}(kG)$ with the structure of an exact subcategory of $\mathrm{mod}(kG)$.

Proof. According to Quillen, to verify that $\underline{E}$ provides $\mathrm{mod}(kG)$ with the structure of an exact category we must verify three properties. The first property consists of the conditions that any short exact sequence isomorphic to one in $\underline{E}$ is itself in $\underline{E}$; that $\underline{E}$ contains all split short exact sequences; and that if $0 \to M' \xrightarrow{i} M \xrightarrow{j} M'' \to 0$ is a short exact sequence in $\underline{E}$, then $i : M' \to M$ is a kernel for $j : M \to M''$ and $j : M \to M''$ is a cokernel for $i : M' \to M$. These conditions are essentially immediate.

The second property consists of the conditions that the class of admissible monomorphisms (i.e., monomorphisms occurring in an exact sequence of $\underline{E}$) is closed under composition and closed under push-out with respect to any map of $\mathrm{mod}(kG)$; similarly that the class of admissible epimorphisms is closed under composition and pullback. This follows from the observation that these properties hold for split exact sequences, thus also for those exact sequences split at every π-point.

The third property asserts that any map $f : M' \to M$ of kG-modules with the property that there exists some map $g : M \to Q$ of kG-modules such that the composition $g \circ f : M' \to M \to Q$ is an admissible monomorphism is itself an admissible monomorphism; and the analogous statement for admissible epimorphisms. This is clear, for any splitting of the composition $\alpha_K^*(g \circ f) : \alpha_K^*(M'_K) \to \alpha_K^*(M_K) \to \alpha_K^*(Q_K)$ gives a splitting of $\alpha_K^*(f) : \alpha_K^*(M'_K) \to \alpha_K^*(M_K)$.

In view of Proposition 2.4, the preceding discussion for $\underline{E}$ applies equally to the class $\underline{E}_{\mathcal{C}}$ of locally split short exact sequences of $\mathcal{C}(kG)$. □

The following proposition makes the evident points that not every short exact sequence of modules of constant Jordan type is locally split, that some nonsplit short exact sequences are locally split, and that a non-locally split extension of modules of constant Jordan type might not have a middle term that is of constant Jordan type.

Proposition 2.6. *Let E be an elementary abelian p-group and let $I \subset kE$ be the augmentation ideal. Then the short exact sequence of kE-modules of constant Jordan type*

$$0 \longrightarrow I^i/I^j \longrightarrow I^i/I^\ell \longrightarrow I^j/I^\ell \longrightarrow 0$$

is not locally split for any $i < j < \ell \leq p$.

If the rank of E is at least 2, then a nontrivial negative Tate cohomology class $\xi \in \hat{\mathrm{H}}^n(G,k)$ determines a locally split (but not split) short exact sequence of the form

$$0 \to k \to E \to \Omega^{n-1}(k) \to 0.$$

In contrast, if the rank of E is at least 2 and if $0 \neq \zeta \in \mathrm{H}^1(E,k)$, then the associated short exact sequence $0 \to k \to M \to k \to 0$ of kE-modules is such that M does not have constant Jordan type.

Proof. By [6, 2.1], each I^i/I^ℓ is of constant Jordan type. The fact that the sequence $0 \to I^i/I^j \to I^i/I^\ell \to I^j/I^\ell \to 0$ is not locally split follows from Proposition 2.4 and an easy computation of Jordan types. The second assertion is a special case of [6, 6.3].

Finally, the extension of k by k determined by $\zeta \in \mathrm{H}^1(E,k)$ is split when pulled back via the π-point α_K if and only if $\alpha_K^*(\zeta_K) = 0 \in \mathrm{H}^1(K[t]/t^p, K)$. The subset of $\Pi(E)$ consisting of those $[\alpha_K]$ satisfying $\alpha_K^*(\zeta_K) = 0$ is a hyperplane of $\Pi(E) \cong \mathbb{P}^{r-1}$, where r is the rank of E. Consequently, the Jordan type of $\alpha_K^*(M_K)$ is $2[1]$ on a hyperplane of $\Pi(E)$ and is $[2]$ otherwise. $\square$

3 The Grothendieck group $K_0(\mathcal{C}(kG))$

For any exact category $\mathcal{E}$ specified by a class $\underline{E}$ of admissible exact sequences, we denote by $K_0(\mathcal{E})$ the Grothendieck group given as the quotient of the free abelian group of isomorphism classes of objects of $\mathcal{E}$ modulo the relations generated by $[M_1] - [M_2] + [M_3]$ whenever $0 \to M_1 \to M_2 \to M_3 \to 0$ is an admissible sequence in $\underline{E}$. In this section, we begin a consideration of $K_0(\mathcal{C}(kG))$, the Grothendieck group of modules of constant Jordan type (with respect to locally split short exact sequences).

Following Quillen [18], we could further consider $K_i(\mathcal{C}(kG))$ for $i > 0$. Granted the current state of our understanding, the challenge of investigating $K_0(\mathcal{C}(kG))$ is sufficiently daunting that we postpone any consideration of $K_i(\mathcal{C}(kG)), i > 0$.

Proposition 3.1. *For any finite group scheme G, there are natural embeddings of exact categories*

$$\mathcal{P}(G) \longrightarrow \mathcal{C}(kG), \qquad\qquad \mathcal{C}(kG) \longrightarrow \mathrm{mod}(kG).$$

The first is from the exact category $\mathcal{P}(G)$ of finitely generated projective kG-modules into the category $\mathcal{C}(kG)$ of modules of constant Jordan type; the second is from $\mathcal{C}(kG)$ into the category mod(kG) *of all finitely generated kG-modules as in Proposition 2.5. These embeddings induce homomorphisms*

$$K_0(kG) \equiv K_0(\mathcal{P}(G)) \longrightarrow K_0(\mathcal{C}(kG)) \longrightarrow K_0(\mathrm{mod}(kG)).$$

Moreover, these homomorphisms are contravariantly functorial with respect to a closed immersion $i : H \to G$ of finite group schemes.

Proof. Since every short exact sequence of projective modules is split, we conclude that the full embedding $\mathcal{P}(G) \to \mathcal{C}(kG)$ of the category of finite-dimensional projective kG-modules into the category of kG-modules of constant Jordan type (with admissible short exact sequences being the locally split short exact sequences) is an embedding of exact categories. The embedding $\mathcal{C}(kG) \subseteq \mathrm{mod}(kG)$ is clearly an embedding of exact categories.

If $i : H \to G$ is a closed embedding, then kG is projective as a kH-module (cf. [16, 8.16]), so that any projective kG-module restricts to a projective kH-module. By [6, 1.9], the restriction of a kG-module of constant Jordan type to kH is again of constant Jordan type. Since restriction is exact and preserves locally split sequences (because every π-point of H when composed with i becomes a π-point of G), we obtain the asserted naturality with respect to $i : H \to G$. □

Remark 3.2. If a module $M \in \mathcal{C}(kG)$ admits an admissible filtration (meaning that the inclusion maps are admissible monomorphisms)

$$M_0 \subset M_1 \subset \cdots \subset M_n = M,$$

then $[M] = \sum_{i=1}^{n} [M_i/M_{i-1}] \in K_0(\mathcal{C}(kG))$.

As an immediate corollary of Proposition 3.1, we have the following.

Corollary 3.3. *If G is a finite group, then $K_0(\mathcal{P}(G)) \to K_0(\mathcal{C}(kG))$ is injective. More generally, if the Cartan matrix for the finite group scheme G is nondegenerate, then $K_0(\mathcal{P}(G)) \to K_0(\mathcal{C}(kG))$ is injective.*

Proof. A basis for $K_0(\mathcal{P}(G))$ is given by the classes of the indecomposable projective modules, while a basis $K_0(\mathrm{mod}(kG))$ is given by the classes of the irreducible modules. The Cartan matrix for kG represents the natural map $K_0(\mathcal{P}(G))$ to $K_0(\mathrm{mod}(kG))$ with respect to these bases. By Proposition 3.1, this map factors through $K_0(\mathcal{C}(kG))$. This proves the second statement. A theorem of Brauer (cf. [3, I.5.7.2]) says that if G is a finite group then the Cartan matrix for kG is nonsingular. □

We consider a few elementary examples.

Examples 3.4. Let G be the cyclic group $\mathbb{Z}/p$. Then

$$K_0(\mathcal{C}(k\mathbb{Z}/p)) \quad \simeq \quad \mathbb{Z}^p.$$

The map $K_0(\mathcal{P}(G)) \to K_0(\mathcal{C}(k\mathbb{Z}/p))$ is identified with the map $\mathbb{Z} \to \mathbb{Z}^p, \quad a \mapsto (a, 0, \ldots, 0)$. The map $K_0(\mathcal{C}(k\mathbb{Z}/p)) \to K_0(\mathrm{mod}(k\mathbb{Z}/p))$ is identified with the map $\mathbb{Z}^p \to \mathbb{Z}, \quad (a_p, \ldots, a_1) \mapsto \sum_i i a_i$.

Proposition 2.4 and the universal property of the Grothendieck group $K_0(\mathcal{C}(kG))$ immediately imply the following proposition.

Proposition 3.5. *Sending a kG-module M of constant Jordan type to the Jordan type of $\alpha_K^*(M_K)$ and to the stable Jordan type of $\alpha_K^*(M_K)$ for any π-point α_K of G determines homomorphisms*

$$\mathrm{JType} : K_0(\mathcal{C}(kG)) \longrightarrow \mathbb{Z}^p, \qquad \overline{\mathrm{JType}} : K_0(\mathcal{C}(kG)) \longrightarrow \mathbb{Z}^{p-1}. \tag{1}$$

We view an element of $K_0(\mathcal{C}(kG))$ as the class of a *virtual* kG-module of constant Jordan type. In the next section, we shall investigate to what extent the homomorphisms JType, $\overline{\text{JType}}$ are surjective; in other words, the realizability of Jordan types by virtual kG-modules of constant Jordan type.

The function JType is not injective: for example, a module of constant Jordan type and its k-linear dual have the same Jordan type. The example of $\mathbb{Z}/2 \times \mathbb{Z}/2$ given in Proposition 3.6 provides a more explicit example of noninjectivity of JType.

We can achieve the next example because we know exactly the indecomposable kE-modules for the Klein four-group $E = \mathbb{Z}/2 \times \mathbb{Z}/2$ (cf. [2], [14]).

Proposition 3.6. *For any field k of characteristic 2, the group algebra kE of the Klein four-group $E = \mathbb{Z}/2 \times \mathbb{Z}/2$ satisfies*

$$K_0(\mathcal{C}(kE)) \simeq \mathbb{Z}^3.$$

Proof. We recall from [6, 6.2] that any kE-module of constant Jordan type is of the form $kE^e \bigoplus (\oplus_i \Omega^{n_i}(k))$. One easily checks there is an admissible (i.e., locally split, short) exact sequence of kE-modules of the following form:

$$0 \longrightarrow \Omega^2(k) \longrightarrow \Omega^1(k) \oplus \Omega^1(k) \longrightarrow k \longrightarrow 0. \tag{2}$$

Hence, $[\Omega^2(k)] = 2[\Omega^1(k)] - [k]$ in $K_0(\mathcal{C}(kE))$. Consecutive applications of the Heller shift to the sequence (2) thus imply that $K_0(\mathcal{C}(kE))$ is generated by the classes of the three kE-modules: $kE,\ k,\ \Omega^1(k)$.

We define a function σ on the class of modules of constant Jordan type by sending $M \simeq kE^e \bigoplus (\oplus_i \Omega^{n_i}(k))$ to $\sigma(M) = \sum_i n_i$. We proceed to show that σ is additive on admissible sequences, and hence induces a homomorphism $\sigma : K_0(\mathcal{C}(kE)) \to \mathbb{Z}$.

Let $\xi : 0 \to M \to L \to N \to 0$ be an admissible sequence of kE-modules of constant Jordan type, and assume $N = N_1 \oplus N_2$. Since $\operatorname{Ext}^1_E(N, M) = \operatorname{Ext}^1_E(N_1, M) \oplus \operatorname{Ext}^1_E(N_2, M)$, we have $\xi = \xi_1 + \xi_2$, where $\xi_i \in \operatorname{Ext}^1_E(N_i, M)$. Moreover, both ξ_1, ξ_2 are admissible, and the additivity of σ on ξ_1, ξ_2 implies additivity of σ on ξ. Hence, we may assume that N is an indecomposable nonprojective module of constant Jordan type, that is, $N \simeq \Omega^n(k)$. Similarly, we may assume $M \simeq \Omega^m(k)$.

Thus, we may assume that ξ has the form

$$0 \longrightarrow \Omega^m(k) \longrightarrow kE^e \bigoplus (\oplus_i \Omega^{n_i}(k)) \longrightarrow \Omega^n(k) \longrightarrow 0.$$

Because the Jordan type of the middle term is the sum of the Jordan types of the ends, we conclude that the middle term has Jordan type with exactly the non-projective summands and thus is of the form $kE^e \oplus \Omega^a(k) \oplus \Omega^b(k)$. We immediately conclude that σ is additive.

We consider the map

$$\Psi = (\text{JType}, \sigma) : K_0(\mathcal{C}(kE)) \longrightarrow \mathbb{Z}^2 \oplus \mathbb{Z}.$$

This is well-defined by Proposition 3.5 and the observation that σ is additive as shown above. To prove the proposition, it suffices to show that Ψ has image a subgroup of finite index inside $\mathbb{Z}^2 \oplus \mathbb{Z}$. This follows from the observation that the vectors $\Psi(k) = (0,1,0)$, $\Psi(\Omega^1(k)) = (1,1,1)$, and $\Psi(kE) = (2,0,0)$ are linearly independent over $\mathbb{Q}$. $\square$

The preceding proof shows that the admissible sequences for the Klein four-group have a generating set consisting of sequences of the form

$$0 \longrightarrow \Omega^{m+n+a}(k) \longrightarrow \Omega^{m+a}(k) \oplus \Omega^{n+a}(k) \oplus kE^e \longrightarrow \Omega^a(k) \longrightarrow 0. \tag{3}$$

We proceed to verify that $m, n > 0$ and

$$2e = |m+n+a| + |a| - |m+a| - |n-a|;$$

hence, $e = 0$ unless a is negative and $m+n+a$ is positive. This fact is of use in the discussion to follow.

By [6, 6.9], $\dim \Omega^{2a}(k) = 4a+1$; hence, $\dim \Omega^{2a+1}(k) = 4(a+1)+1$ (using the short exact sequence $0 \to \Omega^{2a+1}(k) \to P_{2a} \to \Omega^{2a}(k) \to 0$). By applying Ω^i to ζ for some i, we may further assume that ζ has the form

$$0 \longrightarrow \Omega^m(k) \longrightarrow kE^e \bigoplus (\oplus_i \Omega^{a_i}(k)) \longrightarrow k \longrightarrow 0.$$

Since ξ is split on restriction to any π-point, there exists i such that the restriction $\Omega^{a_i}(k) \to k$ is not the zero map. Hence, the map $\Omega^{a_i}(k) \to k$ is surjective. Since the left end of the sequence ξ is projective free, we conclude that $e = 0$. Because the Jordan type of the middle term is the sum of the Jordan type of the ends, the middle term must have exactly two summands. Hence, the sequence has the form

$$0 \longrightarrow \Omega^m(k) \longrightarrow \Omega^a(k) \oplus \Omega^b(k) \longrightarrow k \longrightarrow 0.$$

Because ξ represents an element in $\mathrm{H}^1(G, \Omega^m(k)) \simeq \hat{\mathrm{H}}^{1-m}(G,k)$ that vanishes on restriction along any π-point, $1 - m < 0$ by [6, 6.3]; in other words, m is positive. Likewise, a and b are not negative. That is, the map $\Omega^a(k) \to k$ represents an element in $\hat{\mathrm{H}}^a(G,k)$ that does not vanish on restriction to some π-point. This can happen only if $a \geq 0$. The same argument shows that $b \geq 0$.

There are examples of group schemes for which the Cartan matrix is singular, so that we cannot apply Corollary 3.3 in these examples. Perhaps the simplest is the first Frobenius kernel $G = \mathfrak{G}_1$ of the algebraic group $\mathfrak{G} = \mathrm{SL}_2$ with $p > 2$. In this case, kG is isomorphic to the restricted enveloping algebra of the restricted p-Lie algebra sl_2. It is known (cf. [13, 2.4]) that kG has $(p+1)/2$ blocks one of which has only a single projective irreducible module. Each of the other $(p-1)/2$ blocks has two nonisomorphic irreducible modules. Now suppose that B is one of these blocks. It has irreducible modules S and T. The projective covers Q_S of S and Q_T of T have the forms

$$Q_S = \begin{tikzcd} & S \arrow[dl] \arrow[dr] & \\ T \arrow[dr] & & T \arrow[dl] \\ & S & \end{tikzcd}, \qquad Q_T = \begin{tikzcd} & T \arrow[dl] \arrow[dr] & \\ S \arrow[dr] & & S \arrow[dl] \\ & T & \end{tikzcd}.$$

In particular, for any module M in B, $\operatorname{Rad}^3(M) = 0$, and moreover, if $\operatorname{Rad}^2(M) \neq \{0\}$, then M contains a projective direct summand. That is, if M is an indecomposable nonprojective B-module then $\operatorname{Rad}^2(M) = \{0\}$.

We see from the above that the Cartan matrix of kG is a $p \times p$ matrix consisting of $(p+1)/2$ block matrices along the diagonal, $(p-1)/2$ of which are the Cartan matrices of blocks B of the group algebra as above. The Cartan matrix of such a block is a 2×2 matrix with 2 in every entry. The other block matrix of the Cartan matrix of kG is a 1×1 identity matrix. Hence the natural map from $K_0(\mathcal{P}(G))$ to $K_0(\operatorname{mod}(kG))$ is not injective. However, as we see below, $K_0(\mathcal{P}(G))$ still injects into $K_0(\mathcal{C}(kG))$.

Proposition 3.7. *Let $G = \mathfrak{G}_1$, where $\mathfrak{G} = SL_2$ and assume that $p \geq 3$. Then*

$$K_0(\mathcal{C}(kG)) \simeq \mathbb{Z}^{3p-2}.$$

Moreover, the map $K_0(\mathcal{P}(G)) \to K_0(\mathcal{C}(kG))$ is injective.

Proof. We begin by recalling that rational SL_2-modules are kG-modules of constant Jordan type by [6, 2.5]; in particular, every simple kG-module is a module of constant Jordan type.

As stated above, there are $(p-1)/2$ blocks $B_1, \ldots, B_{(p-1)/2}$ as above. In addition there is another block B' containing only a single indecomposable module, which is projective. Therefore,

$$K_0(\mathcal{C}(kG)) \simeq K_0(\mathcal{C}(B')) \oplus \sum_{i=1}^{(p-1)/2} K_0(\mathcal{C}(B_i)).$$

We know that $K_0(\mathcal{C}(kB')) \simeq \mathbb{Z}$. Consequently, it suffices to prove that $K_0(\mathcal{C}(B_i)) \simeq \mathbb{Z}^6$ for each i.

We consider $B = B_i$ with simple modules S and T. We may assume that S and T have stable constant Jordan types $1[i]$ and $1[p-i]$ respectively. The indecomposable B-modules can be classified using standard methods similar to those of [19] or the diagrammatic methods of [5]. Every nonsimple, nonprojective indecomposable B-module M has the property that $\operatorname{Soc}(M) = \operatorname{Rad}(M)$ is a direct sum of t copies of one of the simple modules S or T. The quotient $M/\operatorname{Rad}(M)$ is a direct sum of r copies of the other simple T or S. Moreover, we must have that r is one of $t-1$, t, and $t+1$, that is, $|t-r| \leq 1$. In the case that $r = t$, it is evident that the module M has periodic cohomology. Or at least, it is clear that the dimensions of $\Omega^n(M)$ are bounded for all n. This means that the module M must have proper nontrivial support variety in $\Pi(G)$, since the annihilator of its cohomology must be nontrivial. Consequently, such a module M cannot have constant Jordan type. Thus, we conclude that $|t - r| = 1$, and that every indecomposable module of constant Jordan type is a syzygy of an irreducible module (cf. (7), (8), and (9), below). This last fact can also be deduced from recent results of Benson [4] on algebras with radical cube zero.

Using the diagrammatic methods sketched below (as can also be done in the situation of Proposition 3.6), we first verify that the nontrivial admissible sequences are generated by sequences of the form

$$0 \to \Omega^{2(m+n)+a}(\mathfrak{X}) \to \Omega^{2m+a}(\mathfrak{X}) \oplus \Omega^{2n+a}(\mathfrak{X}) \oplus P \to \Omega^{a}(\mathfrak{X}) \to 0 \quad (4)$$

and

$$0 \to \Omega^{2(m+n-1)+a}(\mathfrak{X}) \to \Omega^{2m-1+a}(\mathfrak{Y}) \oplus \Omega^{2n-1+a}(\mathfrak{Y}) \oplus Q \to \Omega^{a}(\mathfrak{X}) \to 0 \quad (5)$$

and

$$0 \to \Omega^{2(m+n)-1+a}(\mathfrak{Y}) \to \Omega^{2m+a}(\mathfrak{X}) \oplus \Omega^{2n-1+a}(\mathfrak{Y}) \oplus R \to \Omega^{a}(\mathfrak{X}) \to 0 \quad (6)$$

where $\mathfrak{X}$ and $\mathfrak{Y}$ are either S or T and $\mathfrak{Y}$ is not the same as $\mathfrak{X}$. Here a can be any integer and $m, n > 0$. The modules P, Q, and R are projective modules which are required for the exactness. In any of these sequences, the projective module P, Q or R is a number of copies of the projective cover of a simple module in the socle of the leftmost term of the sequence, or in the top of the rightmost term of the sequence. In what follows it might be helpful to note that, for dimensional reasons, the projective module P in sequence (4) is zero except in the cases that $2(m + n) + a$ is positive and a is negative. Likewise, Q in sequence (5) is zero except when $2(m + n - 1) + a$ is positive and a is negative. And a similar thing happens for sequence (6). To be very specific, P in (4) is a sum of copies of Q_T if $\mathfrak{X} = S$ and a is even or if $\mathfrak{X} = T$ and a is odd. Otherwise, it is a sum of copies of Q_S. The module Q in (5) and R in (6) are sums of copies of Q_T in the cases that $\mathfrak{X} = S$, $\mathfrak{Y} = T$, and a is even and $\mathfrak{X} = T$, $\mathfrak{Y} = S$, and a is odd. Otherwise, they are sums of copies of Q_S. The verification that these sequences, (4), (5), and (6) generate the collection of admissible sequences prodeeds as follows.

As in the proof of Proposition 3.6, we are looking for sequences representing elements of $\mathrm{Ext}^1_{kG}(-,-)$ that vanish on restriction along any π-point. Because Ext^1_{kG} distributes over direct sums, and any such distribution still satisfies the vanishing condition, we can restrict our consideration to sequences that have indecomposable end terms. Consequently, the right end term can be considered to be $\Omega^a(\mathfrak{X})$ for $\mathfrak{X}$ either S or T. Suppose that $\mathfrak{X}$ is S. We can translate the sequence by Ω^{-a} so that the right term is isomorphic to S. Note also that because the two end terms each have only one single nonprojective Jordan block on restriction to any Π-point, there are at exactly two nonprojective direct summands in the middle term of the sequence.

Now we consider the diagrams for the syzygies of S amd T. For example,

$$\Omega^{-1}(S) = \begin{tikzcd}[column sep=small] & S \arrow[dl] \arrow[dr] & \\ T & & T \end{tikzcd} \qquad \Omega^{-1}(T) = \begin{tikzcd}[column sep=small] & T \arrow[dl] \arrow[dr] & \\ S & & S \end{tikzcd} \tag{7}$$

$$\Omega(S) = \begin{tikzcd}[column sep=small] T \arrow[dr] & & T \arrow[dl] \\ & S & \end{tikzcd} \qquad \Omega(T) = \begin{tikzcd}[column sep=small] S \arrow[dr] & & S \arrow[dl] \\ & T & \end{tikzcd} \tag{8}$$

$$\Omega^2(S) = \begin{tikzcd}[column sep=small] S \arrow[dr] & & S \arrow[dl] \arrow[dr] & & S \arrow[dl] \\ & T & & T & \end{tikzcd} \quad \Omega^2(T) = \begin{tikzcd}[column sep=small] T \arrow[dr] & & T \arrow[dl] \arrow[dr] & & T \arrow[dl] \\ & S & & S & \end{tikzcd} \tag{9}$$

The diagram for an arbitrary syzygy of either S or T is merely an elongation of these diagrams.

If we have a locally split sequence whose right-hand term is S (or T) and if $\Omega^m(\mathfrak{X})$ occurs in the middle term (with $\mathfrak{X}$ either S or T), then m must be nonnegative. The reason is that otherwise ($m < 0$), the Jordan type of the kernel of such a map has too many nonprojective blocks at any Π-point. In addition, such a map could not be right split at *any* π-point, because the socle of the kernel would have more irreducible constituents than the head.

To finish the verification that the nontrivial admissible sequences are generated by sequences of the form (4), (5), or (6), we suppose that there is a locally split sequence whose right-hand term is S. The middle term must consist of two terms of the form $\Omega^m(\mathfrak{X})$ and $\Omega^n(\mathfrak{Y})$, with $m, n > 0$ (if either m or n is 0, then the sequence splits). Because both of these terms must map surjectively to S, we must have that $\mathfrak{X} \simeq S$ if m is even, and $\mathfrak{X} \simeq T$ if m is odd. The same happens for $\mathfrak{Y}$ and n. Finally, we notice that the left term of the sequence is determined entirely by its composition factors. A complete analysis, which we leave to the reader, reveals that the sequence must look like one of (4), (5), or (6).

Consider the following two collections of indecomposable B-modules:

$$\mathcal{U}_1 = \{\Omega^m(S),\ m \text{ even}\} \cup \{\Omega^n(T),\ n \text{ odd}\} \cup \{Q_T\}$$

and

$$\mathcal{U}_2 = \{\Omega^m(S),\ m \text{ odd}\} \cup \{\Omega^n(T),\ n \text{ even}\} \cup \{Q_S\}.$$

Observe that each of the sequences (4), (5), (6) for any values of a, m, n, involves either modules all of which are isomorphic to elements of $\mathcal{U}_1$ or modules isomorphic to elements of $\mathcal{U}_2$.

Let $\mathcal{E}(B)$ be the exact category of all B-modules of constant Jordan type with the admissible sequences being the split exact sequences. Thus, $K_0(\mathcal{E}(B))$ is the Green ring $\mathbb{Z}[B]$ of B and is a free $\mathbb{Z}$-module on the classes of elements in the union of $\mathcal{U}_1$ and $\mathcal{U}_2$. Moreover, factoring out the relators coming from the sequences (4) and (5) defines $K_0(\mathcal{C}(B))$ as a quotient of $\mathbb{Z}[B]$. As in the proof of Proposition 3.6, we conclude that the images of classes in $\mathcal{U}_1$ (respectively, $\mathcal{U}_2$) in $K_0(\mathcal{C}(B))$ are generated by the images of the classes $[S]$, $[\Omega^1(T)]$ and $[Q_T]$ (respectively, $[T]$, $[\Omega^1(S)]$, and $[Q_S]$).

Now let $\mathcal{M}_1$ be the subgroup of $K_0(\mathcal{E}(B))$ generated by the classes of modules in $\mathcal{U}_1$, and let $\mathcal{M}_2$ be the subgroup generated by the classes of modules in $\mathcal{U}_2$. Then $K_0(\mathcal{E}(B)) \simeq \mathcal{M}_1 \oplus \mathcal{M}_2$. Let $\mathcal{K}$ be the subgroup generated by the relators determined by sequences (4), (5), and (6). That is, $\mathcal{K}$ is the kernel of the homomorphism of $K_0(\mathcal{E}(B))$ onto $K_0(\mathcal{C}(B))$. Notice that $\mathcal{K} = \mathcal{K}_1 \oplus \mathcal{K}_2$, where each $\mathcal{K}_i = \mathcal{M}_i \cap \mathcal{K}$ is generated by the relators determined by those sequences of the form (4), (5), and (6) that involve only modules from $\mathcal{U}_i$. Consequently, we have that

$$K_0(\mathcal{C}(B)) \simeq K_0(\mathcal{E}(B))/\mathcal{K} \simeq \mathcal{M}_1/\mathcal{K}_1 \oplus \mathcal{M}_2/\mathcal{K}_2.$$

Therefore, to complete the proof of the proposition, it suffices to exhibit homomorphisms $\mathcal{M}_1/\mathcal{K}_1 \to \mathbb{Z}^3$ and $\mathcal{M}_2/\mathcal{K}_2 \to \mathbb{Z}^3$ whose images are cofinite. We do this by showing that there are isomorphisms $\theta_i : \mathcal{M}_i/\mathcal{K}_i \simeq K_0(\mathcal{C}(k'E))$ for $i = 1, 2$ and appealing to Proposition 3.6; here, $E = \mathbb{Z}/2 \times \mathbb{Z}/2$ is the Klein four group and k' is a field of characteristic 2.

For this purpose we define maps $\gamma_1 : \mathcal{M}_1 \to K_0(\mathcal{E}(k'E))$ and $\gamma_2 : \mathcal{M}_2 \to K_0(\mathcal{E}(k'E))$ as follows:

$$\gamma_1([\Omega^n(S)]) = [\Omega^n(k')] \text{ (for } n \text{ even)}, \qquad \gamma_2([\Omega^n(S)]) = [\Omega^n(k')] \text{ (for } n \text{ odd)},$$

$$\gamma_1([\Omega^n(T)]) = [\Omega^n(k')] \text{ (for } n \text{ odd)}, \qquad \gamma_2([\Omega^n(T)]) = [\Omega^n(k')] \text{ (for } n \text{ even)},$$

$$\gamma_1([Q_T]) = [k'E], \qquad \gamma_2([Q_S]) = [k'E].$$

The reader should remember that γ_i is defined only on the classes of modules in $\mathcal{U}_i$. Moreover, each γ_i is clearly surjective.

The important thing to note is that γ_1 takes the relators coming from those sequences (4), (5), and (6) involving only the modules of $\mathcal{U}_1$ bijectively to the relators coming from the sequences (3), which are then zero in $K_0(\mathcal{E}(k'E))$. To see this we need only replace the module $\mathfrak{X}$, $\mathfrak{Y}$ in sequences (4), (5), and (6) by k' and the projective modules P and Q by the appropriate sum of copies of $k'E$.

Moreover, relators coming from the sequences (3) generate the kernel of the natural quotient map from $K_0(\mathcal{E}(k'E))$ to $K_0(\mathcal{C}(k'E))$. Consequently, γ_1 induces an isomorphism θ_1 from $\mathcal{M}_1/\mathcal{K}_1$ to $K_0(\mathcal{C}(k'E))$. Likewise, γ_2 induces an isomorphism $\theta_2 : \mathcal{M}_2/\mathcal{K}_2 \simeq K_0(\mathcal{C}(k'E))$. □

4 Realization of Jordan types

In this section, we initiate the investigation of the images of the Jordan type functions, JType, $\overline{\text{JType}}$, introduced in Proposition 3.5. Our first proposition establishes the surjectivity of JType for an elementary abelian p-group of arbitrary rank. The reader is cautioned that this establishes realizability of Jordan types by virtual modules of constant Jordan type. Even for the rank 2 elementary abelian p-group $E = \mathbb{Z}/p \times \mathbb{Z}/p$ with $p > 3$, we do not know[4] whether there is a kE-module M of constant Jordan type with $\overline{\text{JType}}(M) = [2]$, whereas Proposition 4.1 shows that we can realize theJordan type [2] in a virtual module.

Proposition 4.1. *The map*

$$\text{JType} : K_0(\mathcal{C}(kE)) \longrightarrow \mathbb{Z}^p$$

of Proposition 3.5 is surjective, provided that E is an elementary abelian p-group.

Proof. It suffices to verify that $M = kE/I^i$ has constant Jordan type of the form $1[i] + a_{i-1}[i-1] + \cdots + a_1[1]$ for each $i, 1 \leq i \leq p$, where I is the augmentation ideal of kE. The fact that M has constant Jordan type is verified in [6, 2.1]. The fact that $a_j = 0$ for $j > i$ follows from the fact that $I^i \cdot M = 0$. The fact that $a_i = 1$ follows from the observation that the generator, of M is not annihilated by $(g-1)^i \in I$ for any generator g of E whereas any element of I/I^i is annihilated by $(g-1)^i$ for every generator g of E. □

The Jordan type of a direct sum of $k[t]/t^p$-modules is the sum of the Jordan types. The stable Jordan type of the Heller shift of the $k[t]/t^p$-module $\sum_{i=1}^{p-1} a_i[i]$ equals $\sum_{i=1}^{p-1} a_{p-i}[i]$. The Jordan type of a tensor product is given by the following proposition.

Proposition 4.2. *(cf. [6, 10.2]) Let $[i]$ be an indecomposable $k[t]/t^p$-module of dimension i for $1 \leq i \leq p$. Then if $j \geq i$, we have that*

$$[i] \otimes [j] = \begin{cases} [j-i+1] + [j-i+3] + \cdots + [j+i-3] + [j+i-1] & \textit{if } j+i \leq p \\ [j-i+1] + \cdots + [2p-1-i-j] + (j+i-p)[p] & \textit{if } j+i > p. \end{cases}$$

The following proposition strongly restricts the possible images of the stable Jordan type function $\overline{\text{JType}} : K_0(\mathcal{C}(kG)) \to \mathbb{Z}^{p-1}$ of Proposition 3.5. We say that a subset $S \subset \mathbb{Z}^{p-1}$ is closed under Heller shifts if $\{a_1, \ldots, a_{p-1}\}$ is in S whenever $\sigma = \{a_{p-1}, \ldots, a_1\} \in S$. Similarly, we say that $S \subset \mathbb{Z}^{p-1}$ is closed under direct sums (respectively, tensor products) if $\sigma + \tau \in S$ (respectively $\sigma \otimes \tau \in S$) whenever $\sigma,\ \tau \in S$. Here, $\sigma \otimes \tau$ is defined by the formula of Proposition 4.2 if both σ, τ have a single nonzero entry and is defined more generally by imposing biadditivity.

[4] D. Benson has recently shown that such a module does not exist.

Proposition 4.3. *Let $S \subset \mathbb{Z}^{p-1}$ be a set of stable Jordan types.*

1. *If S has the form $\overline{\mathrm{JType}}(\mathcal{C}(kG))$ for some finite group scheme G, then S is closed under Heller shifts, direct sums, and tensor products.*
2. *If $S = \{m[1] + n[p-1]; m, n \in \mathbb{Z}\}$, then S is closed under Heller shifts, direct sums, and tensor products.*
3. *If $[1],\ [2] \in S$ and S is closed under Heller shifts, direct sums, and tensor products, then $S = \mathbb{Z}^{p-1}$.*
4. *Similarly, if $[1],\ [3] \in S$ with $p > 2$ and S is closed under Heller shifts, direct sums, and tensor products, then $S = \mathbb{Z}^{p-1}$.*
5. *On the other hand, the assumption that $[1],\ [4] \in S$ and S is closed under Heller shifts, direct sums,and tensor products does not imply that $S = \mathbb{Z}^{p-1}$ for $p = 11$.*

Proof. Statement (1) is a consequence of [6, 1.8]. Statement (2) easily follows from the fact that $[p-1] \otimes [p-1] = [1] + (p-2)[p]$.

To prove (3), we observe that $[i] \otimes [2] = [i+1] + [i-1]$ for $i, 2 \leq i < p$. Thus, $[1], [2] \otimes [2] \in S$ implies that $[3] \in S$. Proceeding by induction, we see that $[i], [i-1], [i] \otimes [2] \in S$ implies that $[i+1] \in S$.

For (4), we may assume that $p > 5$. Then we obtain $[3] \otimes [3] = [1] + [3] + [5]$ which implies that $[5] \in S$; $[5] \otimes [3] = [3] + [5] + [7]$, so that $[7] \in S$. Continuing, we conclude that $[2i-1] \in S$, $1 \leq [2i-1] \leq p-2$. Applying Heller shifts to $[2i-1]$, we conclude the stable type $[p-2i+1] \in S$ for $1 \leq [2i-1] \leq p-2$, so that $[j] \in S$ for all $j, 1 \leq j \leq [p-1]$.

Finally, let $p = 11$ so that $[4] \otimes [4] = [1] + [3] + [5] + [7]$ and $p - 4 = 7$. Moreover, $[7] \otimes [4] = [7] + [9] + [11]$. We conclude that S contains the span of $\{[1], [4], [7], [10], [3] + [5], [8] + [6], [7] + [9]\}$. On the other hand, further tensor products with these classes never yield a Jordan type $\sum a_i[i]$ with $a_3 \neq a_5,\ a_6 \neq a_8$ or $a_7 \neq a_9$. □

The applicability of Proposition 4.3 to the question of realizability of stable Jordan types is reflected in the following proposition and its corollary.

Proposition 4.4. *Suppose that G is a finite group that has a normal abelian Sylow p-subgroup. There exists a kG-module with constant stable Jordan type $[2] + n[1]$ for some n. Moreover, the stable Jordan type function*

$$\overline{\mathrm{JType}}:\ K_0(\mathcal{C}(kG)) \longrightarrow \mathbb{Z}^{p-1}$$

is surjective.

Proof. Let $E \subseteq G$ be the subgroup consisting of all elements of order dividing p. This is the unique maximal elementary abelian subgroup of G. Let $M = k_E^{\uparrow G}$ be the induced module from the trivial module on E. Notice that because E is normal in G, E acts trivially on M and hence M is a module of constant Jordan type $s[1]$, where s is the index of E in G.

For any $t > 0$ we have that

$$\mathrm{Ext}^t_{kG}(k, M) \;\cong\; \mathrm{H}^t\left(G, k_E^{\uparrow G}\right) \;\cong\; \mathrm{H}^t(E, k)$$

by the Eckmann–Shapiro lemma. In particular, $\mathrm{Ext}^1_{kG}(k, M) \cong \mathrm{H}^1(E, k)$ has a k-basis consisting of elements $\gamma_1, \ldots, \gamma_r$, where r is the rank of E. The elements have the property that for any π-point $\alpha_K : K[t]/(t^p) \to KG$ the restriction of some $\alpha_K^*(\gamma_i)$ is not zero for some i. It follows that the tuple $\zeta = (\gamma_1, \ldots, \gamma_r)$ is an element of $\mathrm{Ext}^1_{kG}\left(k, \left(k_E^{\uparrow G}\right)^r\right)$ that does not vanish when restricted along any π-point. Then ζ represents a sequence

$$\zeta : \qquad 0 \longrightarrow \left(k_E^{\uparrow G}\right)^r \longrightarrow B \longrightarrow k \longrightarrow 0$$

that is not split on restriction along any π-point. Since the first term in the sequence has constant Jordan type $rs[1]$, the middle term of the sequence must have constant Jordan type $[2] + (rs - 1)[1]$.

The surjectivity of $\overline{\mathrm{JType}}$ now follows from Proposition 4.3(2). $\square$

Corollary 4.5. *Let G be a finite unipotent abelian group scheme. Then the stable Jordan type function*

$$\overline{\mathrm{JType}} : \; K_0(\mathcal{C}(kG)) \longrightarrow \mathbb{Z}^{p-1}$$

is surjective.

Proof. Because G is a unipotent abelian group scheme, its group algebra kG is isomorphic as an algebra (though not as a Hopf algebra) to the group algebra of a finite abelian p-group. Hence, the previous proposition applies, since kG has a module of constant Jordan type $[2]+n[1]$ as constructed in the proof. $\square$

We next briefly consider the question of realizability of stable Jordan types by indecomposable kG-modules by utilizing the Auslander–Reiten theory of almost split sequences (cf. [1]). In the following proposition, $\tau : \mathrm{stmod}(kG) \to \mathrm{stmod}(kG)$ denotes the Auslander–Reiten translation. This is a functor on the stable category, and if G is a finite group or if kG is a symmetric algebra, then $\tau(M) = \Omega^2(M)$ for any finite-dimensional kG-module M.

Proposition 4.6. *Let Θ be a connected component of the Auslander–Reiten quiver that has tree class A_∞. Let $\hat{K}$ be the subgroup of $K_0(\mathcal{C}(kG))$ generated by the classes of the modules in Θ. Let X_0 denote a kG-module in Θ that becomes an initial node of the tree of Θ once projectives are deleted. Then, $\hat{K}$ is generated by the classes $\{[\tau^n(X_0)] \mid n \in \mathbb{Z}\}$, and if Θ contains a projective module P, by the class $[P]$ of that projective module.*

Proof. It suffices to show that the class $[M]$ of any indecomposable module M in Θ can be written as a linear combination of the classes of elements on the bottom row of the stable part of Θ together with the class $[P]$. This is obvious if M lies on the bottom row of Θ. So assume that M is not on the bottom row. Moreover, to prove this for M it suffices to do so for $\tau^n(M)$ for some $n \in \mathbb{Z}$. This is because the functor τ is additive.

The stable part of the component Θ has the form

$$\begin{tikzcd}
 & & & & \cdots & & \\
 & \cdots \arrow[dr] & & X_2 \arrow[ur] \arrow[dr] & & \tau^{-1}(X_2) & \\
 {} \arrow[dr] & & X_1 \arrow[ur] \arrow[dr] & & \tau^{-1}(X_1) \arrow[ur] \arrow[dr] & & {} \\
 & X_0 \arrow[ur] & & \tau^{-1}(X_0) \arrow[ur] & & \tau^{-2}(X_0) \arrow[ur] &
\end{tikzcd}$$

Without loss of generality, we can assume that $M = X_{n+1}$ for some n. In the case that $n = 0$, the almost split sequence for M has the form

$$0 \longrightarrow X_0 \longrightarrow X_1 \oplus \epsilon \longrightarrow \tau^{-1}(X_0) \longrightarrow 0,$$

where ϵ is either the zero module or the projective module P. This is an admissible sequence by [6]. Hence we have that $[M] = [X_1] = [X_0] + [\tau^{-1}(X_0)] - [\epsilon]$ in $K_0(\mathcal{C}(kG))$.

If $n > 0$, then there is an almost split sequence having the form

$$0 \longrightarrow X_n \longrightarrow X_{n+1} \oplus \tau^{-1}(X_{n-1}) \longrightarrow \tau^{-1}(X_n) \longrightarrow 0.$$

Again this sequence is admissible, and we have that $[X_n] + [\tau^{-1}(X_n)] = [X_{n+1}] + [\tau^{-1}(X_{n-1})]$. The proposition now follows by induction. $\square$

As an example of the applicability of the following corollary, recall that a finite p-group G has a single block, and this block has wild representation type, provided $\Pi(G)$ has dimension at least 1 and G is not a dihedral, quaternion or semi-dihedral 2-group. The proof of this corollary is essentially a verbatim repetition of the proof of [6, 8.8] granted Proposition 4.6.

Corollary 4.7. *Let G be a finite group and assume that k is algebraically closed. Assume that there exists an indecomposable kG-module of constant Jordan type with stable Jordan type $\underline{a} = \sum_{i=1}^{p-1} a_i[i]$ that lies in a block of wild representation type. Then there exists an indecomposable kG-module of constant Jordan type with stable Jordan type $n\underline{a}$ for any $n > 0$.*

Proof. By Erdmann's theorem [8], the connected component of the Auslander–Reiten quiver of such a module has tree class A_∞. In the notation of the proof of Proposition 4.6, if X_0 has stable constant Jordan type $\underline{a}$, then because τ commutes with restrictions along π-points, so does $\tau^n(X_0)$. Thus, by the relations developed in the proof, X_1 has stable constant Jordan type $2\underline{a}$, and inductively, X_n has stable constant Jordan type $(n+1)\underline{a}$. □

For an arbitrary finite group scheme, Corollary 4.7 would appear to remain valid in view of work of R. Farnsteiner [9], [10].

We next give a cohomological criterion for the realizability of all stable Jordan types by virtual modules of constant Jordan type.

Theorem 4.8. *Let G be a finite group scheme defined over $\mathbb{Z}$ with the property that there exist odd-dimensional classes $\zeta_1 \in \mathrm{H}^{2d_1-1}(G,k), \ldots, \zeta_m \in \mathrm{H}^{2d_m-1}(G,k)$ such that $\cap_{i=1}^m V(\beta(\zeta_i)) = 0$, where $\beta : \mathrm{H}^{odd}(G,k) \to \mathrm{H}^{ev}(G,k)$ is the Bockstein cohomological operation of degree $+1$. Let*

$$L_{\zeta_1,\ldots,\zeta_m} \;=\; \ker\{\, \textstyle\sum \tilde{\zeta}_i : \sum_{i=1}^m \Omega^{2d_i-1}(k) \longrightarrow k \,\}.$$

If $p > 2$, then $L_{\zeta_1,\ldots,\zeta_m}$ is a module of constant Jordan type whose stable Jordan type has the form $(m-1)[p-1]+1[p-2]$. Consequently, for such G,

$$\overline{\mathrm{JType}} : K_0(\mathcal{C}(kG)) \longrightarrow \mathbb{Z}^{p-1}$$

is surjective.

Proof. Because $\cap_{i=1}^m V(\beta(\zeta_i)) = 0$, there does not exist an equivalence class of π-points $[\alpha_K] \in \Pi(G)$ with $\alpha_K^*(\beta(\zeta_{i,K})) = 0 \in \mathrm{H}^{2d_i}(K[t]/t^p, K)$ for all $i, 1 \le i \le m$. Hence, for each α_K of G there exists some i such that $\alpha_K^*(\zeta_{i,K}) \neq 0 \in \mathrm{H}^{2d_i-1}(K[t]/t^p, K)$. Hence, Proposition [6, 6.7] implies that $L_{\zeta_1,\ldots,\zeta_m}$ is of constant Jordan type with stable Jordan type $(m-1)[p-1]+1[p-2]$.

The second assertion follows from the first by Proposition 4.3(2). □

We proceed to verify (in Proposition 4.11) below that the condition of Theorem 4.8 is satisfied by many finite groups. To do so, we use the following theorem of D. Quillen, which asserts that $\mathrm{H}^*(GL(n,\mathbb{F}_\ell),k)$ maps isomorphically onto the invariants of the cohomology of a direct product of cyclic groups. Specifically, we have the following.

Theorem 4.9. *(D. Quillen [17, §8]) Assume $p > 2$, let $\mathbb{F}_\ell$ be a finite field with $(\ell,p) = 1$, and let r be the least integer such that p divides the order of the units $\mathbb{F}_{\ell^r}^*$ of $\mathbb{F}_{\ell^r}$. Let $\pi = \mathrm{Gal}(\mathbb{F}_{\ell^r}/\mathbb{F}_\ell)$. Then the restriction map*

$$\mathrm{H}^*(GL(n,\mathbb{F}_\ell),k) \;\to\; \mathrm{H}^*\left((\mathbb{F}_{\ell^r}^*)^{\times m}, k\right)^{\pi^{\times m} \rtimes \Sigma_m}$$

is an isomorphism, where $n = mr + e,\ 0 \le e < r$.

In particular, $\mathrm{H}^*(GL(n,\mathbb{F}_\ell),k)_{red}$ is a polynomial algebra on generators $x_i \in \mathrm{H}^{2ri}(GL(n,\mathbb{F}_\ell),k)), 1 \le i \le m$.

The following is an easy consequence of Quillen's theorem.

Proposition 4.10. *We retain the hypotheses and notation of Theorem 4.9. Assume furthermore that p^2 does not divide the order of the units of $\mathbb{F}_{\ell^r}$. Then the cohomology $\mathrm{H}^*\left((\mathbb{F}_{\ell^r}^*)^{\times m}, k\right)^{\pi^{\times m}}$ is an exterior algebra on generators $t_1, \ldots, t_m$ in degrees $2r-1$ tensor a symmetric algebra on the Bocksteins of the t_i's, $u_1 = \beta(t_1), \ldots, u_m = \beta(t_m)$ in degrees $2r$.*

Set ω_i^s to be the class

$$\sum_{j_1 < \ldots < j_i} t_{j_s} \cdot u_{j_1} \cdots u_{j_{s-1}} \cdot u_{j_{s+1}} \cdots u_{j_i} \;\in\; \mathrm{H}^{2ri-1}\left((\mathbb{F}_{\ell^r}^*)^{\times m}, k\right)^{\pi^{\times m}}$$

for any $s, 1 \le s \le i$, and set ζ_i to be the class

$$\sum_{1 \le s \le i} \omega_i^s \in \left(\mathrm{H}^{2ri-1}\left((\mathbb{F}_{\ell^r}^*)^{\times m}, k\right)^{\pi^{\times m}}\right)^{\Sigma_m} \;\cong\; \mathrm{H}^{2ri-1}(GL(n, \mathbb{F}_\ell), k).$$

Then the Bockstein applied to ζ_i equals i times x_i,

$$\beta(\zeta_i) \;=\; i \cdot x_i \in \mathrm{H}^{2ri}(GL(n, \mathbb{F}_\ell), k).$$

Proof. The Bockstein of each ω_i^s equals

$$\sum_{j_1 < \cdots < j_i} u_{j_1} \cdots u_{j_r} \in \mathrm{H}^{2ri}(T(n, \mathbb{F}_\ell), k),$$

for any s, where $T(n, \mathbb{F}_\ell)$ is the torus, which we can take to consist of the diagonal matrices. This element is equal to the restriction of $x_i \in \mathrm{H}^{2ri}(GL(n, \mathbb{F}_\ell), k)$. Thus, $i \cdot x_i$ and $\beta(\zeta_i) \in \mathrm{H}^{2ri}(GL(n, \mathbb{F}_\ell), k)$ both restrict to $i \cdot \sum_{j_1 < \cdots j_i} u_{j_1} \cdots u_{j_i} \in \mathrm{H}^{2ri}(T(n, \mathbb{F}_\ell), k)$. Since the restriction map $\mathrm{H}^{2ri}(GL(n, \mathbb{F}_\ell), k) \to \mathrm{H}^{2ri}(T(n, \mathbb{F}_\ell), k)$ is injective, the proposition follows. $\square$

Proposition 4.11. *Assume $p > 2$, let $\mathbb{F}_\ell$ be a finite field with $(\ell, p) = 1$ and let r be the least integer such that p divides the order of the units $\mathbb{F}_{\ell^r}^*$ of $\mathbb{F}_{\ell^r}$. Let $\pi = \mathrm{Gal}(\mathbb{F}_{\ell^r}/\mathbb{F}_\ell)$. Assume further that p^2 does not divide the order of the units $\mathbb{F}_{\ell^r}^*$ of $\mathbb{F}_{\ell^r}$. If $n = mr + e$, then $0 \le e < r$, and if $m < p$, then any finite group G admitting an embedding $G \subset \mathrm{GL}(n, \mathbb{F}_\ell)$ satisfies the hypothesis of Theorem 4.8.*

In particular, for any such finite group G,

$$\overline{\mathrm{JType}} : K_0(\mathcal{C}(kG)) \longrightarrow \mathbb{Z}^{p-1}$$

is surjective.

Proof. By Theorem 4.8, we may construct a $kGL(n, \mathbb{F}_\ell)$-module $L_{\zeta_1, \ldots, \zeta_m}$ of constant Jordan type with stable Jordan type $m[p-1]+1[p-2]$. The restriction of $L_{\zeta_1, \ldots, \zeta_m}$ to kG remains a module of constant Jordan type with the same stable Jordan type. Thus, the corollary follows by applying Proposition 4.3(2). $\square$

5 Stratification by $\mathcal{C}(kG)$

The modules of constant Jordan type do not form a triangulated subcategory of the stable module category $\mathrm{stmod}(kG)$. On the other hand, $\mathcal{C}(kG)$ enjoys many good properties: as recalled earlier (from [6]), $\mathcal{C}(kG)$ is closed under direct sums, tensor products, k-linear duals, retracts, and Heller shifts. In $\mathrm{stmod}(kG)$, $M \mapsto \Omega^{-1}(M)$ is the translation functor on the triangulated category $\mathrm{stmod}(kG)$. Hence, $\mathcal{C}(kG)$ is a full subcategory of $\mathrm{stmod}(kG)$ closed under translations.

In this section, we briefly consider "thickenings" of $\mathcal{C}(kG) \subset \mathrm{stmod}(kG)$, adopting terminology of [15]. The question here is what modules can be assembled by successive extensions of modules of constant Jordan type. This is motivated partly by the observation that if a kG-module M is an extension of two modules of constant Jordan type, then there is a lower bound on the Jordan type (and perhaps even a minimal Jordan type) that can occur at any π-point, namely, the sum of Jordan types of the two extending modules. A well-known class of examples is that of the modules L_ζ for $\zeta \in \mathrm{H}^{2m}(G,k)$ for some $m > 0$. Such a module is defined by a sequence

$$0 \longrightarrow L_\zeta \longrightarrow \Omega^{2m}(k) \xrightarrow{\tilde{\zeta}} k \longrightarrow 0,$$

where the map $\tilde{\zeta}$ represents the cohomology class ζ. For any π-point α_K of G, $\alpha_K^*(L_{\zeta,K})$ has Jordan type $s[p]$ for some $s > 0$ if $\alpha_K^*(\zeta_K) \neq 0$ and Jordan type $s[p] + [p-1] + [1]$ if $\alpha_K^*(\zeta_K) = 0$.

Definition 5.1. *Set* $\mathrm{thick}^1(\mathcal{C})$ *equal to* $\mathcal{C}(kG)$. *For* $n > 1$, *set* $\mathrm{thick}^n(\mathcal{C})$ *equal to the smallest full subcategory of* $\mathrm{stmod}(kG)$ *that is closed under retracts and that contains all finite-dimensional kG-modules M fitting in a distinguished triangle* $M' \to M \to N \to \Omega^{-1}(M')$ *with* $M' \in \mathrm{thick}^{n-1}(\mathcal{C})$ *and N a kG-module of constant Jordan type (i.e.,* $M \in \mathrm{thick}^1(\mathcal{C})$*).*

Furthermore, define

$$\mathrm{Thick}(\mathcal{C}) \;\equiv\; \cup_n \mathrm{thick}^n(\mathcal{C}) \;\subset\; \mathrm{stmod}(kG),$$

the smallest thick subcategory of $\mathcal{C}(kG)$ *containing* $\mathcal{C}(kG)$.

Any finite-dimensional kG-module has a finite filtration with associated graded module a direct sum of irreducible modules (obtained by successively considering socles of quotient modules). Since the only isomorphism class of irreducible modules for a finite p-group is that of the trivial module and since every irreducible $kSL_{2(1)}$-module has constant Jordan type by [6], we immediately conclude the following proposition.

Proposition 5.2. *Let G be a finite group scheme with the property that every irreducible kG-module has constant Jordan type. Then*

$$\mathrm{Thick}(\mathcal{C}) \;=\; \mathrm{stmod}(kG).$$

In particular, if G is either a finite p-group or the first infinitesimal kernel $(SL_2)_1$ of SL_2, then $\mathrm{Thick}(\mathcal{C}) = \mathrm{stmod}(kG)$.

The socle filtration employed in the proof of the above filtration might not be a good measure of the "level" of a kG-module as we shall see in Proposition 5.9.

Recall that a block B of kG is said to have defect group a given p-subgroup $P \subset G$ if P is the smallest subgroup of G such that every module in B is a direct summand of a kG-module obtained as the induced module of a kP-module. As shown in [12, 4.12], this implies that any kG-module in B has support in $i_*(\Pi(P)) \subset \Pi(G)$, where $i : \Pi(P) \to \Pi(G)$ is induced by the inclusion $P \subset G$.

Proposition 5.3. *Let G be a finite group and let B be a block of kG with defect group $P \neq \{1\}$. If $i_* : \Pi(P) \to \Pi(G)$ is not surjective, then no non-projective module in B has constant Jordan type.*

Moreover, if B is such a block, then no nonprojective module in B is in $\mathrm{Thick}(\mathcal{C})$.

Proof. If M is a kG-module in B, then the support variety $\Pi(G)_M$ is contained in $i_*(\Pi(P)) \subset \Pi(G)$, a proper subvariety of $\Pi(G)$. Consequently, M does not have constant Jordan type. Thus the only modules of constant Jordan type in B are the projective modules. If a B-module were in $\mathrm{Thick}(\mathcal{C})$, then it would have to be an extension of projective B-modules and hence projective. □

Remark 5.4. The hypothesis of Proposition 5.3 is satisfied in numerous examples. For example, the alternating group A_9 on nine letters and the first Janko group J_1 have such blocks in characteristic 2. The Mathieu group M_{12} has such a block in characteristic 3.

We conclude this investigation of modules of constant Jordan type by returning to the special case $G = E$ an elementary abelian p-group. We utilize a class of modules of constant Jordan type discovered by Andrei Suslin, those with the "constant image property." The widespread prevalence of such modules reveals that $\mathcal{C}(kE)$ must necessarily be large and complicated.

We consider an elementary abelian p-group E of rank r, and we consistently use the notation $kE = k[x_1, \dots, x_r]/(x_1^p, \dots, x_r^p)$. We denote by I_E the augmentation ideal of kE, and by $\overline{k}$ some choice of algebraic closure of k.

The following definition is similar to those definitions found in [12, §1].

Definition 5.5. *A finite-dimensional kE-module M is said to have the constant image property if for any $0 \neq w_\alpha = \alpha_1 x_1 + \cdots + \alpha_r x_r \in \overline{k}E$,*

$$w_\alpha M_{\overline{k}} = \mathrm{Rad}\, M_{\overline{k}}.$$

Examples 5.6. The module

$$M \;=\; \mathrm{Rad}^{(r-1)(p-1)}(kE) \equiv I_E^{(r-1)(p-1)}$$

has the constant image property.

To verify this, first observe that $M_{\overline{k}} \;=\; \mathrm{Rad}^{(r-1)(p-1)}(\overline{k}E)$, so that we may assume that k is algebraically closed. Every nonzero monomial of degree $(r-1)(p-1)+1$ in $k[x_1,\ldots,x_r]$ has the form $x_1^{a_1}\cdots x_r^{a_r}$ for $0 \le a_i < p$ for all i. But since $a_1+\cdots+a_r = (r-1)(p-1)+1$, it must be that $a_1 > 0$; i.e., $I_E^{(r-1)(p-1)+1} = x_1 I_E^{(r-1)(p-1)}$. Hence,

$$\mathrm{Rad}\, M \;\equiv\; I_E M \;= \mathrm{Rad}^{(r-1)(p-1)+1}(kE) = x_1 M,$$

and we have verified the condition of Definition 5.5 for $w_\alpha \;=\; x_1$. This is sufficient, for given any $w_\alpha \neq 0$ there is an automorphism of kE that takes w_α to x_1 and takes I_E^n to I_E^n for all n.

Remark 5.7. As the reader can easily verify, the direct sum of modules with the constant image property and any quotient of a module with constant image property again have the constant image property. Thus, starting with Example 5.6, we obtain many additional modules with constant image property. Moreover, a simple induction argument implies that if the kE-module M has the constant image property, then for any $n > 0$ and any $0 \neq w_\alpha = \alpha_1 x_1 + \cdots + \alpha_r x_r \in \overline{k}E$,

$$w_\alpha^n M_{\overline{k}} \;=\; \mathrm{Rad}^n\, M_{\overline{k}}.$$

Proposition 5.8. *Suppose that M is a kE-module with the constant image property. Then M has constant Jordan type.*

Proof. For any $\alpha = (\alpha_1,\ldots,\alpha_r) \neq \underline{0} \in \overline{k}^r$, the Jordan type of w_α on $M_{\overline{k}}$ is equivalent to the data consisting of the sequence of dimensions

$$\mathrm{Dim}\, M_{\overline{k}}, \quad \mathrm{Dim}\, w_\alpha M_{\overline{k}}, \quad \mathrm{Dim}\, w_\alpha^2 M_{\overline{k}}, \quad \ldots, \quad \mathrm{Dim}\, w_\alpha^{p-1} M_{\overline{k}}.$$

Thus, Remark 5.7 implies that all of the w_α have the same Jordan type on $M_{\overline{k}}$. According to the original definition of maximal Jordan type [12, 1.4], some w_α has maximal Jordan type for M. Thus, the complement of the nonmaximal variety of M contains all $\overline{k}$-rational points of $\mathbb{A}^r$, and hence the nonmaximal variety of M is empty. This is equivalent to the assertion that M has constant Jordan type. □

Proposition 5.9. *Suppose that E is an elementary abelian p-group of rank r. Then*

$$\mathrm{thick}^{2r}(\mathcal{C}) = \mathrm{stmod}(kE).$$

Proof. We first assume that $\mathrm{Rad}^p(M) = I_E^p M = 0$ and proceed to verify that M is in $\mathrm{thick}^2(\mathcal{C})$. Namely, let $Q = kE^t$ be the injective hull of M. Then we have an injection $\varphi : M \longrightarrow kE^t$. Because $I_E^p M = 0$, we must have that $\varphi(M) \subseteq I_E^{(r-1)(p-1)} kE^t = \left(I_E^{(r-1)(p-1)}\right)^t$. This yields the exact sequence

$$0 \longrightarrow M \longrightarrow \left(I_E^{(r-1)(p-1)}\right)^t \longrightarrow N \longrightarrow 0,$$

where N is the quotient. Hence in $\mathrm{stmod}(kG)$ there is a distinguished triangle

$$M \longrightarrow \left(I_E^{(r-1)(p-1)}\right)^t \longrightarrow N \longrightarrow \Omega^{-1}(M) \longrightarrow \cdots.$$

Consequently, applying Proposition 5.8, we conclude that $M \in \mathrm{thick}^2(\mathcal{C})$.

In general, the modules

$$M/\mathrm{Rad}^p(M),\ \mathrm{Rad}^p(M)/\mathrm{Rad}^{2p}(M), \ldots, \mathrm{Rad}^{(r-1)p}(M)/\mathrm{Rad}^{rp}(M)$$

each belong to $\mathrm{thick}^2(\mathcal{C})$, because they are all annihilated by I_E^p. Because $I_E^{rp}(M) = 0$, this readily implies that M is in $\mathrm{thick}^{2r}(\mathcal{C})$ as asserted. □

References

1. M. Auslander, I. Reiten, and S. Smalo, *Representation Theory of Artin Algebras*, Cambridge Studies in Advanced Mathematics 36, 1995.
2. V. Basev, *Representations of the group $\mathbb{Z}_2 \times \mathbb{Z}_2$ in a field of characteristic 2*, (Russian) Dokl. Akad. Nauk. USSR **141** (1961), 1015–1018.
3. D. Benson, *Representations and Cohomology, Vols. I and II*, Cambridge University Press, 1991.
4. D. Benson, *Resolutions over symmetric algebras with radical cube zero*, Journal of Algebra, **320** (2008), 48–56.
5. D. J. Benson and J. F. Carlson, *Diagrammatic methods for group representations and cohomology*, Comm. in Algebra, **15** (1987), 53–121.
6. J. Carlson, E. Friedlander, and J. Pevtsova, *Modules of constant Jordan type*, J. für die reine und ang. Math. **614** (2008), 191–234.
7. J. Carlson, E. Friedlander, and A. Suslin, *Modules for $\mathbb{Z}/p \times \mathbb{Z}/p$*. To appear in Commentarii Mathematici Helvetici.
8. K. Erdmann, *On Auslander-Reiten components of group algebras*, J. Pure Appl. Algebra **104** (1995), 149–160.
9. R. Farnsteiner, *On the Auslander-Reiten quiver of an infinitesimal group*, Nagoya Math. J. **160** (2000), 103–121.
10. R. Farnsteiner, *Support spaces and Auslander-Reiten components*, *Lie algebras, vertex operator algebras and their applications*, 61–87, Contemp. Math. **442**, Amer. Math. Soc. 2007.
11. E. Friedlander and J. Pevtsova, *Π-supports for modules for finite group schemes*, Duke Math. J. **139** (2007), no. 2, 317–368.

12. E. Friedlander, J. Pevtsova, and A. Suslin, *Maximal and generic Jordan types*, Invent. Math. **168** (2007), no. 3, 485–522.
13. E. Friedlander and B. Parshall, *Modular representation theory of Lie algebras*, American J. Math **110** (1988), 1055–1094.
14. A. Heller, I. Reiner, *Indecomposable representations*, Illinois J. Math 5 (1961), 314–323.
15. S. Iyengar, *Levels in triangulated categories*, Lecture given at Leeds.
16. J. C. Jantzen, Representations of Algebraic Groups, Academic Press (1987).
17. D. Quillen, *On the cohomology and K-theory of the general linear groups over a finite field*, Annals of Math **96** (1972), 552–586.
18. D. Quillen, Higher K-theory, I; in Algebraic K-theory I, SLN 341 (1973), 77–139.
19. C. M. Ringel, *The indecomposable representations of the dihedral* 2*-groups*, Math. Ann. 214 (1975), 19–34.

Del Pezzo Moduli via Root Systems

Elisabetta Colombo,[1] Bert van Geemen,[1] and Eduard Looijenga[2]

[1] Dipartimento di Matematica, Università di Milano, via Saldini 50, I-20133 Milano, Italia, elisabetta.colombo@mat.unimi.it, geemen@mat.unimi.it
[2] Mathematisch Instituut, Universiteit Utrecht, P.O. Box 80.010, NL-3508 TA Utrecht, E.J.N.Looijenga@uu.nl

To Professor Yuri Manin, for his 70th birthday

Summary. Coble defined in his 1929 treatise invariants for cubic surfaces and quartic curves. We reinterpret these in terms of the root systems of type E_6 and E_7 that are naturally associated to these varieties, thereby giving some of his results a more intrinsic treatment. Our discussion is uniform for all Del Pezzo surfaces of degree 2, 3, 4, and 5.

Key words: Del Pezzo surfaces, moduli spaces, root systems.

2000 Mathematics Subject Classifications: 14J10, 14J26, 14NB5, 17B20

Introduction

A *Del Pezzo surface of degree d* is a smooth projective surface with semi-ample anticanonical bundle whose class has self-intersection d. The degree is always between 1 and 9 and the surface is either a quadric ($d = 8$ in that case) or is obtained from blowing up $9-d$ points in the projective plane that satisfy a mild genericity condition. So moduli occur only for $1 \leq d \leq 4$. The anticanonical system is d-dimensional, and when $d \neq 1$, it is also base-point-free. For $d = 4$, the resulting morphism is birational onto a complete intersection of two quadrics in $\mathbb{P}^4$, for $d = 3$ it is birational onto a cubic surface in $\mathbb{P}^3$ and for $d = 2$ we get a degree two map onto $\mathbb{P}^2$ whose discriminant curve is a quartic (we will ignore the case $d = 1$ here). This image surface (respectively discriminant) is smooth in case the anticanonical bundle is ample; we then call the Del Pezzo surface a *Fano surface*. Otherwise, it might have simple singularities in the sense of Arnol'd (that have a root system label A, D, or E). Conversely, every complete intersection of two quadrics in $\mathbb{P}^4$, cubic surface in $\mathbb{P}^2$, or quartic curve in $\mathbb{P}^2$ with such singularities thus arises.

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*, Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_7,

We mentioned that a degree-d Del Pezzo surface, $d \neq 8$, is obtained from blowing up $9-d$ points in $\mathbb{P}^2$ in general position. A more precise statement is that if we are given as many disjoint exceptional curves $E_1, \dots, E_{9-d}$ on the Del Pezzo surface S, then these can be simultaneously contracted to produce a projective plane. So the images of these curves yield $9-d$ numbered points $p_1, \dots, p_{9-d}$ in $\mathbb{P}^2$ given up to projective equivalence. Hence every polynomial expression in terms of the projective coordinates of these points that is invariant under $\mathrm{SL}(3, \mathbb{C})$ is a "covariant" for the tuple $(S; E_1, \dots, E_{9-d})$. Coble exhibited such covariants for the important cases $d = 2$ and $d = 3$. These are in general not covariants of S itself, since the surface may have many exceptional systems $(E_1, \dots, E_{9-d})$. Indeed, if we assume that S is Fano, then, as Manin observed, these systems are simply transitively permuted by a Weyl group W, which acts here as a group of Cremona transformations. Therefore, this group will act on the space of such covariants. Coble's covariants span a W-invariant subspace, and Coble was able to identify the W-action as a Cremona group (although the Weyl group interpretation was not available to him). For $d = 3$ he found an irreducible representation of degree 10 of an E_6-Weyl group, and for $d = 2$ he obtained an irreducible representation of degree 15 of an E_7-Weyl group.

The present paper purports to couch Coble's results in terms of a moduli space of tuples $(S; E_1, \dots, E_{9-d})$ as above, for which S is semistable in the sense of geometric invariant theory. This moduli space comes with an action of the Weyl group W. It also carries a natural line bundle, called the *determinant bundle*, to which the W-action lifts: this line bundle assigns to a Del Pezzo surface the line that is the dual of the top exterior power of the space of sections of its anticanonical bundle. It turns out that this bundle is proportional to the one that we use to do geometric invariant theory (and from which our notion of semistability originates). We show that the Coble covariants can be quite naturally understood as sections of this bundle, and we re-prove the fact known to Coble that these sections span an irreducible representation of W. We also show that these sections separate the points of the above moduli space, so that one might say that Coble's covariants of a stable tuple $(S; E_1, \dots, E_{9-d})$ make up a complete set of invariants. This approach not only covers the cases Coble considered (degree 2 and 3), but also the degree 4 case and, somewhat amusingly, even the degree 5 case, for which there are no moduli at all. For the case of degree 3 we also make the connection with earlier work of Naruki and Yoshida. This allows us to conclude that the Coble covariants define a complete linear system and define a closed immersion of the GIT-compactification of the moduli space of marked cubic surfaces in a 9-dimensional projective space. Our results are less complete when the degree is 2; for instance, we did not manage to establish that the Coble covariants define a complete linear system.

We end up with a description of the GIT moduli space that is entirely in terms of the corresponding root system. Our results lead us to some remarkable integrability properties of the module of W-invariant vector fields on the vector space that underlies the defining (reflection) representation of W, and we raise the question whether this is a special case of a general phenomenon.

Since the appearance of Coble's book a great deal of work on Del Pezzo moduli has seen the light of day. Since its sheer volume makes it impossible to give our predecessors their fair due, any singling out of contributions will be biased. While keeping that in mind we nevertheless wish to mention the influential book by Manin [18], the *Astérisque* volume by Dolgachev–Ortland [10], Naruki's construction of a smooth compactification of the moduli space of marked cubic surfaces [16], the determination of its Chow groups in [5], the Lecture Notes by Hunt [15], and Yoshida's revisit of the Coble covariants [19]. The ball quotient description of the moduli space of cubic surfaces by Allcock, Carlson, and Toledo [2], combined with Borcherds' theory of modular forms, led Allcock and Freitag [1] to construct an embedding of the moduli space of marked cubic surfaces, which coincides with the map given by the Coble invariants [13], [14].

We now briefly review the organization of the paper. The first section introduces a moduli space for marked Fano surfaces of degree $d \geq 2$ as well as the line bundle over that space that is central to this paper, the determinant line bundle. This assigns to a Fano surface the determinant line of the dual of the space of sections of its anticanonical sheaf (this is also the determinant of the cohomology of its structure sheaf). We show that this line bundle can be used to obtain in a uniform manner a compactification (by means of GIT), so the determinant bundle extends over this compactification as an ample bundle. In Section 2 we introduce the Coble covariants and show that they can be identified with sections of the determinant bundle. The next section expresses these covariants purely in terms of the associated root system. In Section 4 we identify (and discuss) the Weyl group representation spanned by the Coble covariants. The final section investigates the separating properties of the Coble covariants, where the emphasis is on the degree 3 case.

As we indicated, Manin's work on Del Pezzo surfaces has steered this beautiful subject in a new direction. Although this represents only a small part of his many influential contributions to mathematics, we find it therefore quite appropriate to dedicate this paper to him on the occasion of his 70th birthday.

1 Moduli spaces for marked Del Pezzo surfaces

We call a smooth complete surface S a *Del Pezzo surface of degree* d if its anticanonical bundle ω_S^{-1} is semi-ample and $\omega_S \cdot \omega_S = d$. It is known that then $1 \leq d \leq 9$ and that S is isomorphic either to a smooth quadric or to a surface obtained from successively blowing up $9 - d$ points of $\mathbb{P}^2$. In order that a successive blowing up of $9-d$ points of $\mathbb{P}^2$ yields a Del Pezzo surface, it is necessary and sufficient that we blow up on (i.e., over the strict transform of) a smooth cubic curve (which is an anticanonical divisor of $\mathbb{P}^2$). This is equivalent to the apparently weaker condition that we blow up at most three times on a line and at most six times on a conic. It is also equivalent to the apparently stronger condition that the anticanonical system on this surface be nonempty and have dimension d.

Let S be a Del Pezzo surface of degree d. The vector space $V(S) := H^0\left(\omega_S^{-1}\right)^*$ (which we will usually abbreviate by V) has dimension $d+1$. The anticanonical system defines the (rational) *anticanonical map* $S \dashrightarrow \mathbb{P}(V)$. When $d \geq 2$, it has no base points, so that the anticanonical map is a morphism. For $d = 1$, it has a single base point; if we blow up this point, then the anticanonical map lifts to a morphism $\tilde{S} \to \mathbb{P}(V)$ that makes $\tilde{S}$ a rational elliptic surface (with a section defined by the exceptional curve of the blowup).

Let $S \to \bar{S}$ contract the (-2)-curves on S (we recall that a curve on a smooth surface is called a (-2)*-curve* if it is a smooth rational curve with self-intersection -2). Then $\bar{S}$ has rational double-point singularities only and its dualizing sheaf $\omega_{\bar{S}}$ is invertible and anti-ample. We shall call such a surface an *anticanonical surface.* If $d \geq 2$, then the anticanonical morphism factors through $\bar{S}$. For $d \geq 3$ the second factor is an embedding of $\bar{S}$ in a projective space of dimension d; for $d = 3$ this yields a cubic surface and for $d = 4$ a complete intersection of two quadrics. When $d = 2$, the second factor realizes $\bar{S}$ as a double cover of a projective plane ramified along a quartic curve with only simple singularities in the sense of Arnol'd (accounting for the rational double points on $\bar{S}$).

Adopting the terminology in [11], we say that S is a *Fano surface* of degree d if ω_S^{-1} is ample (but beware that other authors call this a Del Pezzo surface). If S is given as a projective plane blown up in $9 - d$ points, then it is Fano precisely when the points in question are distinct, no three lie on a line, no six lie on a conic, and no eight lie on a cubic that has a singular point at one of them. This is equivalent to requiring that S contain no (-2)-curves.

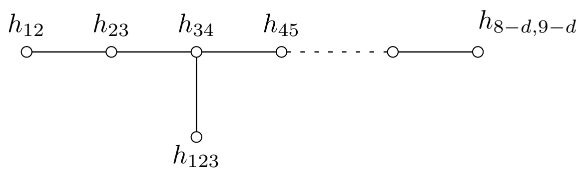
From now on we assume that S is not isomorphic to a smooth quadric. We denote the canonical class of S by $k \in \mathrm{Pic}(S)$ and its orthogonal complement in $\mathrm{Pic}(S)$ by $\mathrm{Pic}_0(S)$. An element $e \in \mathrm{Pic}(S)$ is called an *exceptional class* of S if $e \cdot e = e \cdot k = -1$. Every exceptional class is representable by a unique effective divisor. A *marking* of S is an ordered $(9-d)$-tuple of exceptional classes $(e_1, \dots, e_{9-d})$ on $\mathrm{Pic}(S)$ with $e_i \cdot e_j = -\delta_{ij}$. Given a marking, there is a unique class $\ell \in \mathrm{Pic}(S)$ characterized by the property that $3\ell = -k + e_1 + \cdots + e_{9-d}$ and $(\ell, e_1, \dots, e_{9-d})$ will be a basis of $\mathrm{Pic}(S)$. The marking is said to be *geometric* if S can be obtained by $(9-d)$-successive blowups of a projective plane in such a manner that e_i is the class of the total transform E_i of the exceptional curve of the ith blowup. An ℓ-marking of S consists of merely giving the class ℓ. So if $\mathcal{L}$ is a representative line bundle, then $\mathcal{L}$ is base-point-free and defines a birational morphism from S to a projective plane, and the anticanonical system on S projects onto a d-dimensional linear system of cubic curves on this plane.

Since we are interested here in the moduli of Fano surfaces, we usually restrict to the case $d \leq 4$: if S is given as a blown-up projective plane, then four of the $9-d$ points to be blown up can be used to fix a coordinate system, from which it follows that we have a fine moduli space $\mathcal{M}^\circ_{m,d}$ of marked Fano surfaces of degree d that is isomorphic to an affine open subset of $(\mathbb{P}^2)^{5-d}$.

From now on we assume that $d \le 6$. With Manin we observe that then the classes $e_i - e_{i+1}$, $i = 1, \dots, 8-d$, and $\ell - e_1 - e_2 - e_3$ make up a basis of $\mathrm{Pic}_0(S)$ and can be thought of as a system of simple roots of a root system R_{9-d}. This root system is of type E_8, E_7, E_6, D_5, A_4, and $A_2 + A_1$ respectively. The roots that have fixed inner product with ℓ make up a single $\mathcal{S}_{9-d}$-orbit and we label them accordingly:

(0) $h_{ij} := e_i - e_j$, $(i \neq j)$.
(1) $h_{ijk} := \ell - e_i - e_j - e_k$ with i, j, k pairwise distinct.
(2) $(2\ell - e_1 - e_2 - e_3 - e_4 - e_5 - e_6 - e_7) + e_i$, denoted by h_i when $d = 2$; for $d = 3$, this makes sense only for $i = 7$ and we then may write h instead.
(3) $-k - e_i$ $(d = 1$ only).

Notice that $h_{ij} = -h_{ji}$, but that in h_{ijk} the order of the subscripts is irrelevant.



The marking defined by $(e_1, \dots, e_{9-d})$ is geometric if and only if for every (-2)-curve C its intersection product with each of the simple roots is not positive.

The Weyl group $W(R_{9-d})$ is precisely the group of orthogonal transformations of $\mathrm{Pic}(S)$ that fix k. It acts simply transitively on the markings. In particular it acts on $\mathcal{M}^\circ_{m,d}$, and the quotient variety $\mathcal{M}^\circ_d := W(R_{9-d})\backslash\mathcal{M}^\circ_{m,d}$ can be interpreted as the coarse moduli space of Fano surfaces of degree d. The orbit space of $\mathcal{M}^\circ_{m,d}$ relative to the permutation group of the $e_1, \dots, e_{9-d}$ (a Weyl subgroup of type A_{8-d}) is the moduli space $\mathcal{M}^\circ_{\ell,d}$ of ℓ-marked Fano surfaces of degree d.

Completion of the moduli spaces by means of GIT

Fix a 3-dimensional complex vector space A and a generator $\alpha \in \det(A)$. We think of α as a translation-invariant 3-vector field on A. If $f \in \mathrm{Sym}^3 A^*$ is a cubic form on A, then the contraction of α with df, $\iota_{df}\alpha$, is a 2-vector field on A that is invariant under scalar multiplication and hence defines a 2-vector field on $\mathbb{P}(A)$. We thus obtain an isomorphism between $\mathrm{Sym}^3 A^*$ and $H^0\left(\omega^{-1}_{\mathbb{P}(A)}\right)$.

Let $d \in \{2, 3, 4\}$. A $(d+1)$-dimensional linear quotient V of $\mathrm{Sym}^3 A$ defines a linear subspace $V^* \subset \mathrm{Sym}^3 A^*$, i.e., a linear system of cubics on $\mathbb{P}(A)$ of dimension d. If we assume that this system does not have a fixed component and that its base locus consists of $9-d$ points (multiplicities counted), then blowing up this base locus produces an ℓ-marked Del Pezzo surface S with the property that $H^0\left(\omega_S^{-1}\right)^*$ can be identified with V (we excluded $d=1$ here because then the base locus has 9 points and we get a rational elliptic surface). If we specify an order for the blowing up, then S is even geometrically marked. The quotient surface $\bar{S}$ obtained from contracting (-2)-curves is more canonically defined since it can be described in terms of the rational map $\mathbb{P}(A) \dashrightarrow \mathbb{P}(V)$: for $d=3,4$ it is the image of this map, and for $d=2$ the Stein factorization realizes $\bar{S}$ as a double cover $\mathbb{P}(V)$ ramified over a quartic curve.

The condition that the linear system has no fixed component and has $9-d$ base points defines a subset $\Omega_d \subset G_{d+1}(\mathrm{Sym}^3 A^*)$. Over Ω_d we have a well-defined ℓ-marked family $\bar{\mathcal{S}}_d/\Omega_d$ to which the $\mathrm{SL}(A)$-action lifts. Any ℓ-marked anticanonical surface is thus obtained, so that we have a bijection between the *set* of isomorphism classes of ℓ-marked Del Pezzo surfaces and the *set* of $\mathrm{SL}(A)$-orbits in Ω_d. It is unlikely that this can be lifted to the level of varieties, and we therefore invoke geometric invariant theory. We begin by defining the line bundle that is central to this paper.

Definition 1.1. *If $f : \mathcal{S} \to B$ is a family of anticanonical surfaces of degree d, then its* determinant bundle $\mathrm{Det}(\mathcal{S}/B)$ *is the line bundle over B that is the dual of the determinant of the rank $9-d$ vector bundle $R^1 f_* \omega_{\mathcal{S}/B}^{-1}$ (so this assigns to a Del Pezzo surface S the line $\det H^0\left(\omega_S^{-1}\right)^*$).*

Thus we have a line bundle $\mathrm{Det}(\mathcal{S}_d/\Omega_d)$. Its fiber over the $(d+1)$-dimensional subspace $V^* \subset \mathrm{Sym}^3 A^*$ is the line $\det(V)$ and hence the fiber of the ample bundle $\mathcal{O}_{G_{d+1}(\mathrm{Sym}^3 A^*)}(1)$. A section of $\mathcal{O}_{G_{d+1}(\mathrm{Sym}^3 A^*)}(k)$ determines a section of $\mathrm{Det}^{\otimes k}(\mathcal{S}_d/\Omega_d)$. Since the action of $\mathrm{SL}(A)$ on $\mathrm{Sym}^3 A^*$ is via $\mathrm{PGL}(A)$ (the center μ_3 of $\mathrm{SL}(A)$ acts trivially), we shall regard this as a representation of the latter. Consider the subalgebra of $\mathrm{PGL}(A)$-invariants in the homogeneous coordinate ring of $\overline{\Omega_d}$,

$$R_d^\bullet := \left(\oplus_{k=0}^\infty H^0(\mathcal{O}_{\overline{\Omega_d}}(k))\right)^{\mathrm{PGL}(A)}.$$

The affine cone $\mathrm{Spec}\,(R_d^\bullet)$ has the interpretation as the categorical $\mathrm{PGL}(A)$-quotient of the (affine) cone over $\overline{\Omega_d}$. It may be thought of as the affine hull of the moduli space of triples (S,ℓ,δ) with (S,ℓ) an ℓ-marked Del Pezzo surface of degree d and δ a generator of $\det H^0\left(\omega_S^{-1}\right)$. Since we shall find that the base of this cone, $\mathrm{Proj}\,(R_d^\bullet)$, defines a projective completion of $\mathcal{M}_{\ell,d}^\circ$, we denote it by $\mathcal{M}_{\ell,d}^*$. The asserted interpretation of $\mathcal{M}_{\ell,d}^*$ of course requires that we verify that the orbits defined by Fano surfaces are stable. We will do that in a case-by-case discussion that relates this to GIT completions that are obtained in a different manner. In fact, for each of the three cases $d=2,3,4$ we shall construct a GIT completion $\mathcal{M}_d^*$ of $\mathcal{M}_d^\circ$ in such a way that the forgetful morphism $\mathcal{M}_{\ell,d}^\circ \to \mathcal{M}_d^\circ$ extends to a finite morphism of GIT completions $\mathcal{M}_{\ell,d}^* \to \mathcal{M}_d^*$. This description will also help us to identify (and interpret) the boundary strata.

We recall that the proj construction endows $\mathcal{M}^*_{\ell,d}$ for every $k \geq 0$ with a coherent sheaf $\mathcal{O}_{\mathcal{M}^*_{\ell,d}}(k)$ of rank one whose space of sections is R^k_d. We call $\mathcal{O}_{\mathcal{M}^*_{\ell,d}}(1)$ the *determinant sheaf*; it is a line bundle in the orbifold setting.

In what follows, V_{d+1} is a fixed complex vector space of dimension $d+1$ endowed with a generator μ of $\det(V_{d+1})$. We often regard μ as a translation-invariant $(d+1)$-polyvector field on V_{d+1}.

Degree 4 surfaces in projective 4-space

If a pencil of quadrics in $\mathbb{P}(V_5)$ contains a smooth quadric, then the number of singular members of this pencil (counted with multiplicity) is 5. According to Wall [20], the geometric invariant theory for intersections of quadrics is as follows: for a plane $P \subset \mathrm{Sym}^2 V_5^*$, $[\wedge^2 P] \in \mathbb{P}\left(\wedge^2\left(\mathrm{Sym}^2 V_5^*\right)\right)$ is $\mathrm{SL}(V_5)$-stable (respectively $\mathrm{SL}(V_5)$-semistable) if and only if the divisor on $\mathbb{P}(P)$ parameterizing singular members is reduced (respectively has all its multiplicities ≤ 2). A semistable pencil belongs to a minimal orbit if and only if its members can be simultaneously diagonalized. So a stable pencil is represented by a pair $\left\langle Z_0^2+Z_1^2+Z_2^2+Z_3^2+Z_4^2, a_0Z_0^2+a_1Z_1^2+a_2Z_2^2+a_3Z_3^2+a_4Z_4^2\right\rangle$ with $a_0,\dots,a_4$ distinct. This is equivalent to the corresponding surface S_P in $\mathbb{P}(V_5)$ being smooth. The minimal strictly semistable orbits allow at most two pairs of coefficients to be equal. In case we have only one pair of equal coefficients, S_P has two A_1-singularities and in case we have two such pairs, four. The fact that these singularities come in pairs can be "explained" in terms of the D_5-root system in the Picard group of a Del Pezzo surface of degree 4: an A_1-singularity is resolved by a single blowup with a (-2)-curve as exceptional curve whose class is a root in the Picard root system. The roots perpendicular to this root make up a root system of type $D_4 + A_1$, and the class of the companion (-2)-curve will sit in the A_1-summand. Also, a minimal strictly semistable orbit with 2 (respectively 4) A_1-singularities is adjacent to a semistable orbit without such A_1-pairs and represented by a pair of quadrics, one of which is defined by $Z_0^2+Z_1^2+Z_2^2+Z_3^2+Z_4^2$ and the other by $Z_0Z_1+a_1Z_1^2+a_2Z_2^2+a_3Z_3^2+a_4Z_4^2$ (respectively $Z_0Z_1+Z_2Z_3+a_3Z_3^2+a_4Z_4^2$).

The center μ_5 of $\mathrm{SL}(V_5)$ acts acts faithfully by scalars on $\wedge^2\left(\mathrm{Sym}^2 V_5^*\right)$, and for that reason the $\mathrm{SL}(V_5)$-invariant part of the homogeneous coordinate ring of $\mathrm{Gr}_2\left(\mathrm{Sym}^2 V_5^*\right)$ lives in degrees that are multiples of 5:

$$S_4^\bullet := \bigoplus_{k=0}^{\infty} S_4^k, \quad S_4^k := H^0\left(\mathcal{O}_{\mathrm{Gr}_2\left(\mathrm{Sym}^2 V_5^*\right)}(5k)\right)^{\mathrm{SL}(V_5)}.$$

We obtain a projective completion $\mathcal{M}_4^* := \mathrm{Proj}\, S_4^\bullet$ of $\mathcal{M}_4^\circ$ with twisting sheaves $\mathcal{O}_{\mathcal{M}_4^*}(k)$ such that $S_4^k = H^0\left(\mathcal{O}_{\mathcal{M}_4^*}(k)\right)$. The singular complete intersections are parameterized by a hypersurface in $\mathrm{Gr}_2\left(\mathrm{Sym}^2 V_5^*\right)$. Since the Picard group of this Grassmannian is generated by $\mathcal{O}_{\mathrm{Gr}_2\left(\mathrm{Sym}^2 V_5^*\right)}(1)$, this discriminant is defined by a section of $\mathcal{O}_{\mathrm{Gr}_2\left(\mathrm{Sym}^2 V_5^*\right)}(20)$ and so $B_4 := \mathcal{M}_4^* - \mathcal{M}_4^\circ$ is defined by a section of $\mathcal{O}_{\mathcal{M}_4^*}(4)$.

Suppose we are given a surface $S \subset \mathbb{P}(V_5)$ defined by a pencil of quadrics. So S determines a line Φ_S in $\wedge^2 \left(\mathrm{Sym}^2 V_5^*\right)$. Any generator $F_1 \wedge F_2 \in \Phi_S$ and $u \in V_5^*$ determine a 2-vector field on V by $\iota_{du\wedge dF_1 \wedge dF_2}\mu$. This 2-vector field is invariant under scalar multiplication and tangent to the cone over S. Hence it defines a 2-vector field on S, or equivalently, an element of $H^0\left(\omega_S^{-1}\right)$. The map thus defined is an isomorphism

$$V_5^* \otimes \Phi_S \cong H^0\left(\omega_S^{-1}\right).$$

By taking determinants we get an identification of $\Phi_S^5 \cong \det H^0\left(\omega_S^{-1}\right)$. We may think of Φ_S^5 as the quotient of the line Φ_S by the center μ_5 of $\mathrm{SL}(V_5)$. Thus $\mathrm{Spec}\left(S_4^\bullet\right)$ may be regarded as the affine hull of the moduli space of pairs (S,δ) with S a Del Pezzo surface of degree 4 and δ a generator of $\det\left(H^0\left(\omega_S^{-1}\right)\right)$.

Proposition 1.2. *We have a natural finite embedding $S_4^\bullet \subset R_4^\bullet$ of graded $\mathbb{C}$-algebras such that the forgetful morphism $\mathcal{M}_{\ell,4}^\circ \to \mathcal{M}_4^\circ$ extends to a finite morphism of GIT completions $\mathcal{M}_{\ell,4}^* \to \mathcal{M}_4^*$ (and the notions of semistability coincide in the two cases) and $\mathcal{O}_{\mathcal{M}_4^*}(1)$ is the determinant sheaf.*

Cubic surfaces

Following Hilbert, the cubic surfaces in $\mathbb{P}(V_4)$ that are stable (respectively semistable) relative to the $\mathrm{SL}(V_4)$-action are those that have an A_1-singularity (respectively A_2-singularity) at worst. There is only one strictly semistable minimal orbit and that is the one that has three A_2-singularities.

The center μ_4 of $\mathrm{SL}(V_4)$ acts faithfully by scalars on $\mathrm{Sym}^3 V_4^*$, and so the $\mathrm{SL}(V_4)$-invariant part of the homogeneous coordinate ring of $\mathrm{Sym}^3 V_4^*$ lives in degrees that are multiples of 4:

$$S_3^\bullet := \bigoplus_{k=0}^{\infty} S_3^k, \quad S_3^k := H^0\left(\mathcal{O}_{\mathbb{P}\left(\mathrm{Sym}^3 V_4^*\right)}(4k)\right)^{\mathrm{SL}(V_4)}.$$

We thus find the projective completion $\mathcal{M}_3^* := \mathrm{Proj} S_3^\bullet$ of $\mathcal{M}_3^\circ$ with twisting sheaves $\mathcal{O}_{\mathcal{M}_3^*}(k)$ such that $S_3^k = H^0\left(\mathcal{O}_{\mathcal{M}_3^*}(k)\right)$. The discriminant hypersurface in the linear system of degree d hypersurfaces in $\mathbb{P}^n$ has degree $(n+1)(d-1)^n$. So the singular cubic surfaces are parameterized by a hypersurface in $\mathbb{P}\left(\mathrm{Sym}^3 V_4^*\right)$ of degree 32. The stable locus $\mathcal{M}_3^\circ \subset \mathcal{M}_3 \subset \mathcal{M}_3^*$ is the complement of a single point. Furthermore, $B_3 := \mathcal{M}_3^* - \mathcal{M}_3^\circ$ is defined by a section of $\mathcal{O}_{\mathcal{M}_3^*}(8)$.

Let $S \subset \mathbb{P}(V_4)$ be a cubic surface defined by a line Φ_S in $\mathrm{Sym}^3 V_4^*$. Proceeding as in the degree 4 case we find that for a generator $F \in \Phi_S$ and $u \in V^*$, the expression $\iota_{du\wedge dF}\mu$ defines a 2-vector field on S and that we thus get an isomorphism

$$V_4^* \otimes \Phi_S \cong H^0\left(\omega_S^{-1}\right).$$

By taking determinants we get an identification $\Phi_S^4 \cong \det H^0\left(\omega_S^{-1}\right)$. We think of Φ_S^4 as the quotient of the line Φ_S by the center μ_4 of $\mathrm{SL}(V_4)$ and conclude as before:

Proposition 1.3. *We have a natural finite embedding $S_3^\bullet \subset R_3^\bullet$ of graded $\mathbb{C}$-algebras such that the forgetful morphism $\mathcal{M}_{\ell,3}^\circ \to \mathcal{M}_3^\circ$ extends to a finite morphism of GIT completions $\mathcal{M}_{\ell,3}^* \to \mathcal{M}_3^*$ (and the notions of semistability coincide in the two cases) and $\mathcal{O}_{\mathcal{M}_3^*}(1)$ is the determinant sheaf.*

Quartic curves

The case of degree 2 is a bit special because $W(E_7)$ has a nontrivial center (of order two). The center leaves invariant the (isomorphism type of the) surface: it acts as an involution and changes only the marking. The latter even disappears if we remember only the fixed point set of this involution, the quartic curve. Van Geemen [10] observed that the marking of the Del Pezzo surface then amounts to a principal level-two structure on the quartic curve (this is based on the fact that $W(E_7)$ modulo its center is isomorphic to the symplectic group $\mathrm{Sp}(6, \mathbb{Z}/2)$). Since a smooth quartic curve is a canonically embedded genus three curve, $\mathcal{M}_2^\circ$ can also be interpreted as the moduli space of nonhyperelliptic genus three curves with principal level-two structure (here we ignore the orbifold structure).

The projective space $\mathbb{P}\left(\mathrm{Sym}^4 V_3^*\right)$ parameterizes the quartic curves in the projective plane $\mathbb{P}(V_3)$. The geometric invariant theory relative to its $\mathrm{SL}(V_3)$-action is as follows: a quartic curve is stable if and only if it has singularities no worse than of type A_2. A quartic is unstable if and only if it has a point of multiplicity ≥ 3 (or equivalently, a D_4-singularity or worse) or consists of a cubic plus an inflectional tangent. The latter gives generically an A_5-singularity, but such a singularity may also appear on a semistable quartic, for instance on the union of two conics having a point in common where they intersect with multiplicity 3. Let us, in order to understand the incidence relations, review (and redo) the classification of nonstable quartics.

A plane quartic curve C that is not stable has a singularity of type A_3 or worse. So it has an equation of the form $cy^2z^2 + yzf_2(x,y) + f_4(x,y)$ with f_2 and f_4 homogeneous. Consider its orbit under the $\mathbb{C}^\times$-copy in $\mathrm{SL}(V_3)$ for which $t \in \mathbb{C}^\times$ sends (x,y,z) to $(x, ty, t^{-1}z)$. If we let $t \to 0$, then the equation tends to $cy^2z^2 + ax^2yz + bx^4$, where $f_2(x,0) = ax^2$ and $f_4(x,0) = bx^4$. We go through the possibilities.

If $c = 0$, then C has a triple point and the equation $ax^2yz + bx^4$ is easily seen to be unstable. We therefore assume that $c = 1$ and we denote the limit curve by C_0.

If $a^2 - 4b \neq 0 \neq b$, then C_0 is made up of two nonsingular conics meeting in two distinct points with a common tangent (having therefore an A_3-singularity at each) and the original singularity was of type A_3.

If $a \neq 0 = b$, then we have the same situation except that one of the conics has now degenerated into a union of two lines.

The most interesting case is $a^2 - 4b = 0 \neq b$. Then C_0 is a double nonsingular conic, and in case $C \neq C_0$, C has a singularity of type A_k for some $4 \leq k \leq 7$. The case of an A_7-singularity occurs for the curve C_1 given by $(yz+x^2)(yz+x^2+y^2)$: it consists of two nonsingular conics meeting in a single point with multiplicity 4. This is also the most degenerate case after C_0: any $\mathrm{SL}(V_3)$-orbit that has C_0 in its closure is either the orbit of C_0 or has C_1 in its closure. So although a double conic does not yield a Del Pezzo surface, the corresponding point of $\mathcal{M}_2^*$ is uniquely represented by a geometrically marked Del Pezzo surface with an A_7-singularity.

On the other hand, the condition $a = b = 0$ (which means that f_2 and f_4 are divisible by y, so that we have a cubic plus an inflectional tangent or worse), gives the limiting curve defined by $y^2z^2 = 0$, which is clearly unstable.

We shall later find that the ambiguous behavior of an A_5-singularity reflects a feature of the E_7-root system: this system contains two Weyl group equivalence classes of subsystems of type A_5: one type is always contained in an A_7-subsystem (the semistable case) and the other is not (the unstable case). Since the center μ_3 of $\mathrm{SL}(V_3)$ acts faithfully by scalars on $\mathrm{Sym}^4 V_3^*$, we have as algebra of invariants

$$S_2^\bullet := \bigoplus_{k=0}^{\infty} S_2^k, \quad S_2^k := H^0\left(\mathcal{O}_{\mathbb{P}\left(\mathrm{Sym}^4 V_3^*\right)}(3k)\right)^{\mathrm{SL}(V_3)}.$$

Thus $\mathcal{M}_2^* := \mathrm{Proj}\,\mathbb{C}\left[\mathrm{Sym}^4 V_3^*\right]^{\mathrm{SL}(V_3)}$ is a projective completion of $\mathcal{M}_2^\circ$. It comes with twisting sheaves $\mathcal{O}_{\mathcal{M}_2^*}(k)$ such that $S_2^k = H^0\left(\mathcal{O}_{\mathcal{M}_2^*}(k)\right)$. Let us write $\mathcal{M}_2^\circ \subset \mathcal{M}_2 \subset \mathcal{M}_2^*$ for the stable locus; this can be interpreted as the moduli space of marked Del Pezzo surfaces of degree 2 with A_2-singularities at worst. Its complement in $\mathcal{M}_2^*$ is of dimension one. Since the singular quartics make up a hypersurface of degree 27 in $\mathbb{P}\left(\mathrm{Sym}^4 V_3^*\right)$, $B_2 := \mathcal{M}_2^* - \mathcal{M}_2^\circ$ is defined by a section of $\mathcal{O}_{\mathcal{M}_2^*}(9)$. In particular, B_2 is a Cartier divisor.

Let C be a quartic curve in $\mathbb{P}(V_3)$ defined by the line $\varPhi_C \subset \mathrm{Sym}^4 V_3^*$. If $F \in \varPhi_C$ is a generator, then a double cover S of $\mathbb{P}(V_3)$ totally ramified along C is defined by $w^2 = F$ in $V_3 \times \mathbb{C}$ (more precisely, it is Proj of the graded algebra obtained from $\mathbb{C}[V_3]$ by adjoining to it a root of F). Then for every $u \in V_3^*$, the 2-vector field $w^{-1}\iota_{du\wedge dF}\mu$ defines a section of ω_S^{-1}. We thus get an isomorphism $V_3^* \otimes w^{-1}dF \cong H^0\left(\omega_S^{-1}\right)$. If we take the determinants of both sides, we find that $(w^{-1}dF)^3$ determines a generator of $\det H^0\left(\omega_S^{-1}\right)$. So $F^{-3}(dF)^6$ gives one of $\left(\det H^0\left(\omega_S^{-1}\right)\right)^2$, in other words, we have a natural isomorphism $\varPhi_C^3 \cong \left(\det H^0\left(\omega_S^{-1}\right)\right)^2$. That the square of the determinant appears here reflects the fact that the central element $-\mathbf{1}$ of $W(E_7)$ induces an involution in S that acts as the scalar -1 on $\det H^0\left(\omega_S^{-1}\right)$. We obtain the following.

Proposition 1.4. *We have a natural finite embedding $S_2^\bullet \subset R_2^\bullet$ of graded $\mathbb{C}$-algebras such that the forgetful morphism $\mathcal{M}_{\ell,2}^\circ \to \mathcal{M}_2^\circ$ extends to a finite morphism of GIT completions $\mathcal{M}_{\ell,2}^* \to \mathcal{M}_2^*$ (and the notions of semistability coincide in the two cases) and $\mathcal{O}_{\mathcal{M}_2^*}(1)$ is the square of the determinant sheaf.*

Completion of the moduli space of marked Fano surfaces

We have produced for $d = 4, 3, 2$, a GIT completion $\mathcal{M}_d^*$ of $\mathcal{M}_d^\circ$ that we were able to identify with a finite quotient of $\mathcal{M}_{\ell,d}^*$. This implies that $\mathcal{M}_{\ell,d}^*$ contains $\mathcal{M}_{\ell,d}^\circ$ as an open dense subset and proves that every point of $\mathcal{M}_{\ell,d}^*$ can be represented by an ℓ-marked Fano surface.

We define a completion $\mathcal{M}_{m,d}^*$ of the moduli space $\mathcal{M}_{m,d}^\circ$ of marked Fano surfaces of degree d simply as the normalization of $\mathcal{M}_d^*$ in $\mathcal{M}_{m,d}^\circ$. This comes with an action of $W(R_{9-d})$. and the preceding discussion shows that $\mathcal{M}_{\ell,d}^*$ can be identified with the orbit space of $\mathcal{M}_{m,d}^\circ$ by the permutation group of the $e_1, \dots, e_{9-d}$ (a Weyl subgroup of type A_{8-d}).

2 Coble's covariants

In this section we assume that the degree d of a Del Pezzo surface is at most 6 (we later make further restrictions).

Let $(S; e_1, \dots, e_{9-d})$ be a geometrically marked Del Pezzo surface. Recall that we have a class $\ell \in \mathrm{Pic}(S)$ characterized by the property that $-3\ell + e_1 + \cdots + e_{9-d}$ equals the canonical class k. Let us choose a line bundle $\mathcal{L}$ on S that represents ℓ: $H^0(\mathcal{L})$ is then of dimension 3, and if we denote its dual by A, then the associated linear system defines a birational morphism $S \to \mathbb{P}(A)$ that has $E = E_1 + \cdots + E_{9-d}$ as its exceptional divisor. The direct image of $\mathcal{L}$ on $\mathbb{P}(A)$ is still a line bundle (namely $\mathcal{O}_{\mathbb{P}(A)}(1)$, but we continue to denote this bundle by $\mathcal{L}$).

We claim that there is a natural identification

$$\omega_S^{-1} \cong \mathcal{L}^3(-E) \otimes \det A.$$

To see this, we note that if $p \in \mathbb{P}(A)$ and $\lambda \subset A$ is the line defined by p, then the tangent space of $\mathbb{P}(A)$ at p appears in the familiar exact sequence

$$0 \to \mathbb{C} \to \mathrm{Hom}(\lambda, A) \to T_p\mathbb{P}(A) \to 0,$$

from which it follows that $\det T_p\mathbb{P}(A) = \lambda^{-3} \det(A)$ (we often omit the $\otimes$-symbol when lines or line bundles are involved). So the anticanonical bundle $\omega_{\mathbb{P}(A)}^{-1}$ of $\mathbb{P}(A)$ is naturally identified with $\mathcal{L}^3 \otimes \det(A)$. Since $S \to \mathbb{P}(A)$ is the blowup with exceptional divisor E, we see that the above identification makes ω_S^{-1} correspond to $\mathcal{L}^3(-E) \otimes \det(A)$.

The following simple lemma will help us to understand Coble's covariants.

Lemma 2.1. *For a Del Pezzo surface S, the determinant lines of the vector spaces $H^0(\mathcal{O}_E \otimes \mathcal{L}^3) \otimes \det(A)$ and $V(S) := H^0\left(\omega_S^{-1}\right)^*$ are canonically isomorphic.*

Proof. The identification $\omega_S^{-1} \cong \mathcal{L}^3(-E) \otimes \det A$ above gives rise to the short exact sequence

$$0 \to \omega_S^{-1} \to \mathcal{L}^3 \otimes \det A \to \mathcal{O}_E \otimes \mathcal{L}^3 \otimes \det A \to 0.$$

This yields an exact sequence on H^0 because $H^1\left(\omega_S^{-1}\right) = 0$. If we take into account that $H^0(\mathcal{L}^3) = \mathrm{Sym}^3 H^0(\mathcal{L}) = \mathrm{Sym}^3 A^*$, then we obtain the exact sequence

$$0 \to V^* \to \mathrm{Sym}^3 A^* \otimes \det A \to H^0(\mathcal{O}_E \otimes \mathcal{L}^3) \otimes \det A \to 0.$$

Since $\dim(\mathrm{Sym}^3 A^*) = 10$ and $\det(\mathrm{Sym}^3 A^*) = (\det A)^{-10}$, the determinant of the middle term has a canonical generator. This identifies the determinant of V with that of the right-hand side. □

It will be convenient to have a notation for the one-dimensional vector space appearing in the preceding lemma: we define

$$L(S,E) := \det(H^0(\mathcal{O}_E \otimes \mathcal{L}^3 \otimes \det A)) = \det(H^0(\mathcal{O}_E \otimes \mathcal{L}^3)) \otimes (\det A)^{9-d},$$

so that the lemma asserts that $L(S,E)$ may be identified with $\det V(S)$.

We continue with $(S; e_1, \dots, e_{9-d})$ and $\mathcal{L}$. If p_i denotes the image point of E_i, then the geometric fiber of $\mathcal{L}$ over p_i is $\lambda_i := H^0(\mathcal{L} \otimes \mathcal{O}_{E_i})^*$ (a one-dimensional vector space). So $L(S,E) = (\lambda_1 \cdots \lambda_{9-d})^{-3} \otimes \det(A)^{9-d}$. For e_i, e_j, e_k distinct, the map defined by componentwise inclusion

$$\lambda_i \oplus \lambda_j \oplus \lambda_k \to A$$

is a linear map between 3-dimensional vector spaces. It is an isomorphism if p_i, p_j, p_k are not collinear. Hence the corresponding map on the third exterior powers yields an element

$$|ijk| \in \lambda_i^{-1} \lambda_j^{-1} \lambda_k^{-1} \det A$$

that is nonzero in the Fano case (recall that we usually omit the $\otimes$-sign when lines are involved). Notice that the line $\lambda_i^{-1} \lambda_j^{-1} \lambda_k^{-1} \det A$ attached to $\mathcal{L}$ depends only on the marked surface $(S; e_1, \dots, e_{9-d})$ and not on the choice of the $\mathcal{L}$: as said above, $\mathcal{L}$ is unique up to isomorphism and the only possible ambiguity therefore originates from the action of $\mathbb{C}^\times$ in the fibers of $\mathcal{L}$. But it is clear that this $\mathbb{C}^\times$-action is trivial on this line. For $e_{i_1}, \dots, e_{i_6}$ distinct, we also have a linear map between 6-dimensional vector spaces

$$\lambda_{i_1}^2 \oplus \cdots \oplus \lambda_{i_6}^2 \to \mathrm{Sym}^2 A.$$

Since $\det(\mathrm{Sym}^2 A) = (\det A)^4$, this defines a determinant

$$|i_1 \cdots i_6| \in \lambda_{i_1}^{-2} \cdots \lambda_{i_6}^{-2} (\det A)^4.$$

It is nonzero if and only if $p_1, \ldots, p_6$ do not lie on a conic, which is the case when S is Fano. The elements $|ijk|$ and $|i_1 \cdots i_6|$ just introduced will be referred to as *Coble factors*.

Action of the Weyl group on the Coble factors

We now assume that S is a Fano surface of degree ≤ 6. Another marking of S yields another ℓ' and hence another line $L(S, E')$. Nevertheless they are canonically isomorphic to each other since both have been identified with $\det V$. For what follows it is important to make this isomorphism concrete. We will do this for the case that the new marking is the image of the former under the reflection in h_{123}. So $\ell' = 2\ell - e_1 - e_2 - e_3$, $e_1' = \ell - e_2 - e_3$ (E_1' is the strict transform of $\overline{p_2 p_3}$), e_2', and e_3' are expressed in a likewise manner and $e_i' = e_i$ for $i > 3$. We represent ℓ' by

$$\mathcal{L}' := \mathcal{L}^2(-E_1 - E_2 - E_3),$$

so that $A' = H^0(\mathcal{L}')^*$. Before proceeding, let us see what happens if we do this twice, that is, if we apply h_{123} once more:

$$\mathcal{L}'' = \mathcal{L}'^2\left(-E_1' - E_2' - E_3'\right) = \mathcal{L}^4\left(-2E_1 - 2E_2 - 2E_3 - E_1' - E_2' - E_3'\right).$$

The line bundle $\mathcal{L}^3\left(-2E_1 - 2E_2 - 2E_3 - E_1' - E_2' - E_3'\right)$ is trivial (a generator is given by a section of $\mathcal{L}^3$ whose divisor is the triangle spanned by p_1, p_2, p_3), and so if I denotes its (one-dimensional) space of sections, then $\mathcal{L}''$ is identified with $\mathcal{L} \otimes I$. We note that for $i > 3$, the restriction map $I \to \lambda_i^{-3}$ is an isomorphism of lines and that we also have a natural isomorphism $I \to (\lambda_1\lambda_2\lambda_3)^{-1}$, which, after composition with the inverse of $|123|$ yields an isomorphism $I \to \det(A)^{-1}$.

The space of sections of $\mathcal{L}'$ is the space of quadratic forms on A that are zero on λ_1, λ_2, and λ_3. This leads to an exact sequence

$$0 \to \lambda_1^2 \oplus \lambda_2^2 \oplus \lambda_3^2 \to \operatorname{Sym}^2 A \to A' \to 0.$$

The exactness implies that

$$\det A' = (\det A)^4 (\lambda_1\lambda_2\lambda_3)^{-2}.$$

We have $(\lambda_i')^{-1} = H^0(\mathcal{L}^2(-E_1 - E_2 - E_3) \otimes \mathcal{O}_{E_i'})$ by definition. For $i > 3$, this is just the space of quadratic forms on the line λ_i, i.e., λ_i^{-2}, and so $\lambda_i' = \lambda_i^2$ in that case. For $i = 1$,

$$(\lambda_1')^{-1} = H^0(\mathcal{L}^2(-E_1 - E_2 - E_3) \otimes \mathcal{O}_{E_1'}) = H^0(\mathcal{L}^2 \otimes \mathcal{O}_{\overline{p_2 p_3}})(-(p_2) - (p_3)))$$

is the space of quadratic forms on $\lambda_2 + \lambda_3$ that vanish on each summand, i.e., $\lambda_2^{-1}\lambda_3^{-1}$. Thus $\lambda_1' = \lambda_2\lambda_3$ and likewise $\lambda_2' = \lambda_3\lambda_1$, $\lambda_3' = \lambda_1\lambda_3$. Notice that λ_i'' is naturally identified with $\lambda_i \otimes I^{-1}$. Thus

$$\begin{aligned} L(S,E') &= \left(\lambda'_1 \cdots \lambda'_{9-d}\right)^{-3} (\det A')^{9-d} \\ &= (\lambda_1 \cdots \lambda_{9-d})^{-6} (\lambda_1\lambda_2\lambda_3)^{-2(9-d)} (\det A)^{4(9-d)} \\ &= L(S,E) \left(\lambda_1^{-1}\lambda_2^{-1}\lambda_3^{-1} \det A\right)^{21-2d} \textstyle\prod_{i=4}^{9-d} \left(\lambda_i^{-3} \det A\right). \end{aligned}$$

The identifications above of $\lambda_1\lambda_2\lambda_3$, λ_i^3 $(i>3)$, and $\det(A)$ with I^{-1} show that the twisting line $\left(\lambda_1^{-1}\lambda_2^{-1}\lambda_3^{-1} \det A\right)^{21-2d} \prod_{i=4}^{9-d} \left(\lambda_i^{-3} \det A\right)$ has a canonical generator δ. This generator can be expressed in terms of our Coble factors as

$$\delta := |123|^9 \prod_{i=4}^{9-d} (|12i||23i||31i|)$$

(which indeed lies in $\left(\lambda_1^{-1}\lambda_2^{-1}\lambda_3^{-1} \det A\right)^{21-2d} \prod_{i=4}^{9-d} \left(\lambda_i^{-3} \det A\right)$).

Proposition 2.2. *The isomorphism $L(S;E) \cong L(S',E')$ defined above coincides with the isomorphism that we obtain from the identification of domain and range with* $\det V$.

Proof. Choose generators $a_i \in \lambda_i$ and write x_1, x_2, x_3 for the basis of A^* dual to a_1, a_2, a_3. The basis (a_1, a_2, a_3) of A defines a generator $a_1 \wedge a_2 \wedge a_3$ of $\det A$. This determines an isomorphism $\phi : \omega_S^{-1} \cong \mathcal{L}^3(-E)$. That isomorphism fits in the exact sequence

$$0 \to V^* \to \mathrm{Sym}^3 A^* \to \bigoplus_{i=1}^{9-d} \lambda_i^{-3} \to 0.$$

The middle space has the cubic monomials in x_1, x_2, x_3 as a basis. The triple $\left(x_1^3, x_2^3, x_3^3\right)$ defines a basis dual to $\left(a_1^3, a_2^3, a_3^3\right) \in \lambda_1^3 \oplus \lambda_2^3 \oplus \lambda_3^3$. It follows that we have an exact subsequence

$$0 \to V^* \to K \to \bigoplus_{i=4}^{9-d} \lambda_i^{-3} \to 0, \tag{1}$$

where $K \subset \mathrm{Sym}^3 A^*$ is the span of the cubic monomials that are not a third power. This yields an identification

$$(\lambda_4 \cdots \lambda_{9-d})^{-3} \cong \det V \det\left(\langle x_1^2 x_2, \ldots, x_2 x_3^2, x_1 x_2 x_3 \rangle\right).$$

We now do the same for $\mathcal{L}' = \mathcal{L}^2(-E_1 - E_2 - E_3)$. The space A' comes with a basis $(a'_1 = a_2a_3, a'_2 = a_3a_1, a'_3 = a_1a_2)$ that is dual to the basis (x_2x_3, x_3x_1, x_1x_2) of $H^0(\mathcal{L}^2(-E_1 - E_2 - E_3))$. The monomial $x_1x_2x_3$ is the obvious generator of $I = H^0(\mathcal{L}^3(-E'_1 - E'_2 - E'_3 - 2E_1 - 2E_2 - 2E_3)$, so that we have an associated isomorphism

$$\phi' := \phi \otimes x_1x_2x_3 : \omega_S^{-1} \to (\mathcal{L}')^3(-E').$$

We fit this in the exact sequence

$$0 \to H^0\left(\omega_S^{-1}\right) \to H^0((\mathcal{L}')^3) \to \bigoplus_{i=1}^{9-d} {\lambda'}_i^{-3} \to 0.$$

The vector space of sections of the middle term has the cubic monomials in x_2x_3, x_3x_1, x_1x_2 as a basis. The lines $\lambda'_1, \lambda'_2, \lambda'_3$ are spanned by a_2a_3, a_3a_1 and a_1a_2 respectively and λ'_i is for $i = 4, 5, \ldots, 9-d$ spanned by a_i^2. So $(x_2x_3)^3$ spans ${\lambda'}_1^{-3}$ and similarly for ${\lambda'}_2^{-3}$ and ${\lambda'}_3^{-3}$. The space spanned by cubic monomials in x_2x_3, x_3x_1, x_1x_2 that are not pure powers is just $K' := x_1x_2x_3K$. It follows that we get an exact subsequence analogous to the sequence (1):

$$0 \to V^* \to K' \to \bigoplus_{i=4}^{9-d} {\lambda'}_i^{-3} \to 0, \tag{2}$$

where the embedding $V^* \to K'$ is the composite of $V^* \to K$ and the isomorphism $K \cong K'$ given by multiplication by $x_1x_2x_3$. This identifies $(a'_i)^{-3} \in {\lambda'}_i^{-3}$ with $(x_1x_2x_3)(a_i).a_i^{-3} \in \lambda_i^{-3}$. From this we deduce that the generator $\left(a'_1 \cdots a'_{9-d}\right)^{-3}(a'_1 \wedge a'_2 \wedge a'_3)^{9-d}$ of $\left(\lambda'_1 \cdots \lambda'_{9-d}\right)^{-3}(\det A')^{9-d}$ corresponds to $\prod_{i=4}^{9-d}(x_1x_2x_3)(a_i)$ times the generator $(a_1 \cdots a_{9-d})^{-3}(a_1 \wedge a_2 \wedge a_3)^{9-d}$ of $(\lambda_1 \cdots \lambda_{9-d})^{-3}(\det A)^{9-d}$. Since $\delta = \prod_{i=4}^{9-d}(x_1x_2x_3)(a_i)$, this proves that the isomorphism $L(S; E) \cong \det V \cong L(S, E')$ sends the generator of the former to δ times the generator of the latter. □

We next determine how the Coble factors for ℓ' are expressed in terms of those of ℓ. We retain our $0 \neq a_i \in \lambda_i$ and write x_1, x_2, x_3 for the basis of A^* dual to a_1, a_2, a_3 as before. We identify $|ijk|$ (respectively $|i_1 \cdots i_6|$) with the $a_1 \wedge a_2 \wedge a_3$-coefficient of $a_i \wedge a_j \wedge a_k$ (respectively the $a_1^2 \wedge a_2^2 \wedge a_3^2 \wedge a_2a_3 \wedge a_3a_1 \wedge a_1a_2$-coefficient of $a_{i_1}^2 \wedge \cdots \wedge a_{i_6}^2$). Here are some typical cases, where it is assumed that the free indices are distinct and > 3:

$$\begin{aligned}
|123| &= 1,\\
|12k| &= \langle x_3 \,|\, a_k\rangle,\\
|1jk| &= \langle x_2 \wedge x_3 \,|\, a_j \wedge a_k\rangle,\\
|ijk| &= \langle x_1 \wedge x_2 \wedge x_3 \,|\, a_i \wedge a_j \wedge a_k\rangle,\\
|123ijk| &= \left\langle x_2x_3 \wedge x_3x_1 \wedge x_1x_2 \,|\, a_i^2 \wedge a_j^2 \wedge a_k^2\right\rangle,\\
|12ijkl| &= \left\langle x_3^2 \wedge x_2x_3 \wedge x_3x_1 \wedge x_1x_2 \,|\, a_i^2 \wedge a_j^2 \wedge a_k^2 \wedge a_l^2\right\rangle,\\
|1ijklm| &= \left\langle x_2^2 \wedge x_3^2 \wedge x_2x_3 \wedge x_3x_1 \wedge x_1x_2 \,|\, a_i^2 \wedge a_j^2 \wedge a_k^2 \wedge a_l^2 \wedge a_m^2\right\rangle.
\end{aligned}$$

The corresponding expressions for the new marking are converted into the old marking by the substitutions

$$a'_i = \begin{cases} a_2a_3 \text{ when } i = 1,\\ a_3a_1 \text{ when } i = 2,\\ a_1a_2 \text{ when } i = 3,\\ a_i^2 \quad \text{ when } i > 3, \end{cases} \qquad x'_i = \begin{cases} x_2x_3 \text{ when } i = 1,\\ x_3x_1 \text{ when } i = 2,\\ x_1x_2 \text{ when } i = 3,\\ x_i^2 \quad \text{ when } i > 3. \end{cases}$$

We thus obtain:

$$\begin{aligned}
|123|' &= 1 = |123|,\\
|12k|' &= x_1x_2(a_k) = |23k||31k|,\\
|1jk|' &= -x_1(a_j)x_1(a_k)|1jk|,\\
|ijk|' &= |123ijk|,\\
|123ijk|' &= x_1x_2x_3(a_i)\cdot x_1x_2x_3(a_j)\cdot x_1x_2x_3(a_k)\cdot |ijk|,\\
|12ijkl|' &= x_1x_2(a_i)\cdot x_1x_2(a_j)\cdot x_1x_2(a_k)\cdot x_1x_2(a_l)\cdot |12ijkl|,
\end{aligned}$$

The expression for $|1ijklm|'$ does not appear to have a pleasant form: we find that

$$\begin{aligned}
|1ijklm|' = {} & x_1(a_i)x_1(a_j)x_1(a_k)x_1(a_l)x_1(a_m)\\
& \cdot \left\langle x_3^2x_1 \wedge x_1x_3^2 \wedge x_1x_2x_3 \wedge x_2^2x_3 \wedge x_2x_3^2 \mid a_i^3 \wedge a_j^3 \wedge a_k^3 \wedge a_l^3 \wedge a_m^3 \right\rangle.
\end{aligned}$$

The covariants

Here is the definition.

Definition 2.3. *Let $(S; e_1, \dots, e_{9-d})$ be a marked Fano surface of degree $d \le 6$. A* Coble covariant *is an element of $L(S, E)$ that is a product of Coble factors $|ijk|$ and $|i_1 \cdots i_6|$ in such a manner that every unordered pair in $\{1, 2, \dots, 9-d\}$ appears in one of these factors.*

This notion also makes sense for a marked Del Pezzo surface, and indeed, in case the $E_1, \dots, E_{9-d}$ are irreducible (or equivalently, $p_1, \dots, p_{9-d}$ are distinct), then we adopt this as a definition. But when this is not the case, this is not the "right" definition (see Remark 2.5).

It is easily verified that Coble covariants exist only when $2 \le d \le 5$. In these cases they are as follows:

- $(d = 5)$ There is only one Coble covariant, namely $|123||234||341||412|$. It is nonzero if and only if no three points are collinear, that is, if S is a Fano surface.
- $(d = 4)$ A typical Coble covariant is $|123||234||345||451||512|$. It depends on a cyclic ordering of $\{1, 2, \dots, 5\}$, with the opposite cycle giving the same element. So the number of Coble covariants up to sign is equal to $4!/2 = 12$. A Coble covariant can be nonzero even if S has (-2)-curves. For instance, if (p_1, p_2, p_4) and (p_2, p_3, p_5) are collinear but are otherwise generic then the given Coble covariant is nonzero and S has a $2A_1$-configuration (i.e., two disjoint (-2)-curves).
- $(d = 3)$ We have two typical cases: one is $|134||234||356||456||512||612|$ and another is $|123||456||123456|$. The former type amounts to dividing the 6-element set $\{e_1, \dots, e_6\}$ into three equal parts (of two) and cyclically order the three parts (there are 30 such) and the latter to splitting of $\{e_1, \dots, e_6\}$ into two equal parts (there are 10 of these). So there are 40 Coble covariants up to sign.

- $(d = 2)$ We have two typical cases: $|351||461||342||562||547||217||367|$ (of which there are 30) and $|123456||127||347||567|$ (105 in number). So up to sign we obtain 135 cases.

Proposition 2.4. *If S is Fano, then the collection of Coble covariants, when considered as elements of* $\det V(S)$*, is independent of the marking.*

Proof. It is enough to show that the collection is invariant under the reflection in h_{123}. So in view of Proposition 2.2 we need to verify that if we make the above substitutions for a Coble covariant relative to $\left(e'_1, \ldots, e'_{9-d}\right)$, then we get δ times a Coble covariant relative to $(e_1, \ldots, e_{9-d})$. This is a straightforward check. We do a few examples. For $d = 3$, we find that

$$\begin{aligned}
&|134|'|234|'|356|'|456|'|512|'|612|' \\
&\quad = -x_1x_3(a_4) \cdot x_2x_3(a_4) \cdot x_3(a_5)x_3(a_6)|356| \cdot |123456| \cdot x_1x_2(a_5) \cdot x_1x_2(a_6) \\
&\quad = -\delta|124||356||123456|,
\end{aligned}$$

which is indeed the image of $|124||356||123456|$. Similarly,

$$|123|'|456|'|123456|' = |123| \cdot |123456| \cdot \delta|456|,$$

which is the image of $|123||456||123456|$.

The other cases are similar and are left to the reader to verify. □

Remark 2.5. The notion of a Coble covariant extends to the case of a geometrically marked Del Pezzo surface. There is not much of an issue here as long as the points $p_1, \ldots, p_{9-d}$ remain distinct, but when two coalesce the situation becomes a bit delicate, since we wish to land in $\det V(S)$. It is clear that if p_2 approaches p_1, then any Coble covariant involving these points such as $|123| \in \lambda_1^{-1}\lambda_2^{-1}\lambda_3^{-1} \det A$ tends to 0. But we should regard $|123|$ as an element of $\det(H^0(\mathcal{O}_{E_1+E_2+E_3} \otimes \mathcal{L}^3)) \otimes \det A$, and when p_2 tends to p_1, then E_1 becomes decomposable and of the form $F + E_2$. The component F is the strict transform of the exceptional curve of the first blowup and hence a (-2)-curve. Let q be the point where F and E_2 meet. This corresponds to tangent direction at p_1, or equivalently, to a plane $P_q \subset A$ that contains λ_1. If λ_2 moves in W towards λ_1, then $H^0(\mathcal{O}_{E_1+E_2})$ becomes $H^0(\mathcal{O}_{F+2E_2})$. The exact sequence

$$0 \to \mathcal{I}_{F+E_2}/\mathcal{I}_{F+2E_2} \to \mathcal{O}_{F+2E_2} \to \mathcal{O}_{F+E_2} \to 0$$

induces an exact sequence on sections. Notice that the first term is a constant sheaf on E_2. Its fiber over q is $T_q^*F \otimes T_q^*E_2$. This fiber is apparently also the determinant of $H^0(\mathcal{O}_{F+2E_2})$. Since $T_qF = \mathrm{Hom}(\lambda, W_q/\lambda)$ and $T_qE_2 = \mathrm{Hom}(W_q/\lambda, A/W_q)$, we have $T_q^*F \otimes T_q^*E_2 = (\det A)^{-1} \det P_q \otimes \lambda_1$. It follows that

$$\det(H^0(\mathcal{O}_{F+2E_2+E_3} \otimes \mathcal{L}^3)) \otimes \det A = \lambda_1^{-1}\lambda_3^{-1} \det P_q.$$

It is in this line where $|123|$ should take its value. (This also explains why in the next section we need to divide by a discriminant $\Delta(t_q, \ldots, t_{9-d})$.)

Since every point of $\mathcal{M}^*_{m,d}$ is representable by a marked Del Pezzo surface, a Coble covariant can be regarded as a section of the determinant sheaf $\mathcal{O}_{\mathcal{M}^*_{m,d}}(1)$. It follows from Section 2 that $W(R_{9-d})$ permutes these sections transitively.

Definition 2.6. *The* Coble space $\mathcal{C}_d$ *is the subspace of* $H^0(\mathcal{O}_{\mathcal{M}^*_{m,d}}(1))$ *spanned by the Coble covariants.*

We shall prove that for $d = 3, 4$, $\mathcal{C}_d$ is complete, i.e., all of $H^0(\mathcal{O}_{\mathcal{M}^*_{m,d}}(1))$, but we do not know whether that is true when $d = 2$.

Remark 2.7. We shall see that $\mathcal{C}_d$ is an irreducible representation of $W(R_{9-d})$ of dimension $6, 10, 15$ for respectively $d = 4, 3, 2$. It follows from the discussion in Section 1 that any $W(R_{9-d})$-invariant polynomial of degree k in the Coble covariants has an interpretation in terms of classical invariant theory: for $d = 4$ we get an SL(5)-invariant of degree $5k$ for pencils of quadrics, for $d = 3$, an SL(4)-invariant of degree $4k$ for cubic forms, and for $d = 2$, an SL(3)-invariant of degree $3k/2$ for quartic forms.

Here is an example that illustrates this. There is only one irreducible representation of $W(E_6)$ of degree 10. This representation is real and has therefore a nonzero $W(E_6)$-invariant quadratic form. According to the preceding this produces an SL(4)-invariant of degree 8 for cubic forms. This is indeed the lowest degree of such an invariant.

3 Anticanonical divisors with a cusp

Anticanonical cuspidal cubics on Del Pezzo surfaces

Let S be a Del Pezzo surface of degree d that is not isomorphic to a smooth quadric. Assume that there is also given a reduced anticanonical curve K on S isomorphic to a cuspidal cubic (notice that if $d = 1$ such a curve will not always exist). The curve K is given by a hyperplane $V_K \subset V = H^0\left(\omega_S^{-1}\right)^*$. We write l_K for the line V/V_K, so that l_K^* is a line in $V^* = H^0\left(\omega_S^{-1}\right) = \mathrm{Hom}(\omega_S, \mathcal{O}_S)$. The image of l_K^* is $\mathrm{Hom}(\omega_S, \mathcal{O}_S(-K)) \subset \mathrm{Hom}(\omega_S, \mathcal{O}_S)$, and hence l_K may be identified with $H^0(\omega_S(K))$. So a nonzero $\kappa \in l_K$ can be understood as a rational 2-form κ on S whose divisor is K. The residue $\mathrm{Res}_K(\kappa)$ of κ on the smooth part of K identifies $\mathrm{Pic}^0(K)$ with $\mathbb{C}$ as an algebraic group: we may represent an element of $\mathrm{Pic}^0(K)$ by a difference $(q) - (p)$, and then $\mathrm{Res}_K(\kappa)$ assigns to this element the integral of $\mathrm{Res}_K(\kappa)$ along any arc in K from p to q. This identifies l_K with $\mathrm{Hom}(\mathrm{Pic}^0(K), \mathbb{C})$ or equivalently, $\mathrm{Pic}^0(K)$ with l_K^*.

Recall that $\mathrm{Pic}_0(S) \subset \mathrm{Pic}(S)$ denotes the orthogonal complement of the class of ω_S^{-1}. It is then clear that restriction defines a homomorphism $r : \mathrm{Pic}_0(S) \to \mathrm{Pic}^0(K) \to l_K^*(\subset V^*)$. We extend r to an algebra homomorphism

$$r : \mathrm{Sym}^\bullet \mathrm{Pic}_0(S) \to \mathrm{Sym}^\bullet l_K^*(\subset \mathrm{Sym}^\bullet V^*).$$

For $\kappa \in l_K$, we compose this map with the evaluation in κ and obtain an algebra homomorphism

$$r_\kappa : \mathrm{Sym}^\bullet \mathrm{Pic}_0(S) \to \mathbb{C}.$$

Suppose now that S is geometrically marked by $(e_1, \dots, e_{9-d})$ as before. With the notation of the previous section, we have a line bundle $\mathcal{L}$ on S and an associated contraction morphism $S \to \mathbb{P}(A)$, where $A = H^0(\mathcal{L})^*$, with E_i mapping to a singleton. The cuspidal curve K meets E_i in a single point p_i (with intersection number one). It is mapped isomorphically to its image in $\mathbb{P}(A)$.

The Zariski tangent space of K at its cusp is a line (with multiplicity two, but that will be irrelevant here). Let $u \in A^*$ be such that $u = 0$ defines the corresponding line in $\mathbb{P}(A)$. We may then extend u to a coordinate system (u, v, w) for A such that K is given by the equation $u^2w - v^3$ (this makes $[1 : 0 : 0]$ the unique flex point of K and $w = 0$ its tangent line). This coordinate system is for a given u almost unique: if (u, v', w') is any other such coordinate system, then $v' = cv$ and $w' = c^3 w$ for some $c \in \mathbb{C}^\times$.

However, a choice of a generator $\kappa \in l_K$ singles out a natural choice (v, w) by requiring that the residue of κ on K be the restriction of $d(v/u)$. We put $t := v/u$ (we should write $v_\kappa, w_\kappa, t_\kappa$, but we do not want to overburden the reader with the notation; let us just remember that v, w, t are homogeneous of degree $1, 3, 1$ in κ). The dependence of v and w on u is clearly homogeneous of degree 1. The smooth part K is then parameterized by $t \in \mathbb{C} \mapsto p(t) := [1 : t : t^3]$ such that dt corresponds to $\mathrm{Res}_K \kappa$ and r_κ sends $(p(t)) - (p(t')) \in \mathrm{Pic}^o(K)$ to $t - t' \in \mathbb{C}$.

Assume for the moment that the p_i's are distinct (so that the E_i's are irreducible and the λ_i's are distinct). Let us denote the restriction of $u \in A^*$ to λ_i by u_i. This is clearly a coordinate for λ_i, and hence a generator of λ_i^{-1}. We thus obtain the generator

$$\begin{aligned} \epsilon'_\kappa &:= (u_1 \cdots u_{9-d})^3 (du \wedge dv \wedge dw)^{-(9-d)} \\ &\in \lambda_1^{-3} \cdots \lambda_{9-d}^{-3} (\det A)^{9-d} = L(S, E). \end{aligned}$$

Since v and w are homogeneous of degree one in u, it follows that ϵ'_κ is independent of the choice of u. But they are homogeneous of degree 1 and 3 respectively in κ, and so ϵ'_κ is homogeneous of degree $-4(9-d)$ in κ.

Let us now see which linear forms we get on $\lambda_1^3 \oplus \cdots \oplus \lambda_{9-d}^3$ by restriction of cubic monomials in u, v, w. They will be of the form $\left(t_1^k u_1^3, \dots, t_{9-d}^k u_{9-d}^3\right)$ for some $k \geq 0$: for the monomial $u^a v^b w^c$ we have $k = b + 3c$. We thus get all integers $0 \leq k \leq 9$ except 8 (and 3 occurs twice since $u^2 v$ and w^3 yield the same restriction):

$$\begin{array}{ccccc} u^3 \mapsto 1 & u^2 v \mapsto t & uv^2 \mapsto t^2 & u^2 w, v^3 \mapsto t^3 & uvw \mapsto t^4 \\ v^2 w \mapsto t^5 & uw^2 \mapsto t^6 & vw^2 \mapsto t^7 & w^3 \mapsto t^9. & \end{array}$$

If we select $9-d$ such monomials and compute the determinant of their restrictions to $\lambda_1^3 \oplus \cdots \oplus \lambda_{9-d}^3$, we see that that it is either zero or equal to $(u_1 \cdots u_{9-d})^3 \det((t_i)^{k_j})_{1\le i,j\le 9-d}$ for some $0 \le k_1 < \cdots < k_{9-d}$. The latter expression lies in $\mathbb{Z}[t_1,\dots,t_{9-d}]$ and is divisible by the discriminant that we get by taking $9-d$ monomials corresponding to $1, t, \dots, t^{8-d}$, namely

$$\Delta(t_1,\dots,t_{9-d}) := \prod_{1\le i<j\le 9-d} (t_i - t_j).$$

In other words, if we regard $t_1,\dots,t_{9-d}$ as variables, then the $(9-d)$th exterior power over $\mathbb{Z}[t_1,\dots,t_{9-d}]$ of the homomorphism

$$\mathbb{Z}[t_1,\dots,t_{9-d}] \otimes \operatorname{Sym}^3 A^* \to \oplus_{i=1}^{9-d} \mathbb{Z}[t_1,\dots,t_{9-d}]u_i^3$$

has its image generated by $\Delta(t_1,\dots,t_{9-d})u_1^3 \cdots u_{9-d}^3$. That a division by $\Delta(t_1,\dots,t_{9-d})$ is appropriate is suggested by Remark 2.5 and so we use

$$\epsilon_\kappa := \Delta(t_1,\dots,t_{9-d})\epsilon'_\kappa$$

as a generator of $\lambda_1^{-3} \cdots \lambda_{9-d}^{-3} (\det A)^{9-d}$ instead. We may rephrase this more sensibly in terms of our exact sequence

$$0 \to V^* \to \operatorname{Sym}^3(A^*)(\det A)^{10} \to H^0(\mathcal{L}^3 \otimes \mathcal{O}_E \otimes \det A) \to 0.$$

We see that our coordinates define a basis of the middle term in such a manner that they make the sequence split: $9-d$ cubic monomials that yield $1, t, \dots, t^{8-d}$ define a partial basis of $\operatorname{Sym}^3(A^*)(\det A)^{10}$, whose $(9-d)$th exterior power maps onto ϵ_κ. Notice that this remains true if some of the points p_i coalesce, that is, if S is just a geometrically marked Del Pezzo surface—this is in contrast to ϵ'_κ. Since ϵ_κ is homogeneous in κ of degree $-4(9-d) + \binom{9-d}{2} = \frac{1}{2}d(9-d)$, we have constructed an isomorphism

$$\epsilon : l_K^{d(9-d)/2} \cong L(S,E).$$

Anticanonical cuspidal cubics on Fano surfaces

Let us return to the Fano case. If $p_i = p(t_i) = \left[1 : t_i : t_i^3\right]$, then the generator of λ_i dual to u_i is clearly $\tilde{p}_i = \left(1, t_i, t_i^3\right)$. It is not hard to verify that for i, j, k distinct in $\{1,\dots,9-d\}$, we have the following identity in $(\lambda_i\lambda_j\lambda_k)^{-1} \det A$:

$$\begin{aligned} \tilde{p}_i \wedge \tilde{p}_j \wedge \tilde{p}_k &= \Delta(t_i,t_j,t_k)(-t_i - t_j - t_k)(du \wedge dv \wedge dw)^{-1} \\ &= \Delta(t_i,t_j,t_k) r_\kappa(h_{ijk})(du \wedge dv \wedge dw)^{-1}. \end{aligned}$$

Similarly we find for $i_1,\dots,i_6$ distinct in $\{1,\dots,9-d\}$ the following identity in the line $\lambda_{i_1}^{-2} \cdots \lambda_{i_6}^{-2} \det(\operatorname{Sym}^2 A)$:

$$\begin{aligned} \tilde{p}_{i_1}^2 \wedge \cdots \wedge \tilde{p}_{i_6}^2 &= \mp(t_{i_1} + \cdots + t_{i_6})\Delta(t_{i_1},\dots,t_{i_6})(du \wedge dv \wedge dw)^{-4} \\ &= \pm\Delta(t_{i_1},\dots,t_{i_6}) r_\kappa(2\ell - e_{i_1} - \cdots - e_{i_6})(du \wedge dv \wedge dw)^{-4}. \end{aligned}$$

In this manner we get, for example, when $d = 3$,

$$\begin{aligned}\pm|123456||123||456| &= \varDelta(t_1, \ldots, t_6) r_\kappa(\varDelta(R')) u_1^3 \cdots u_6^3 (du \wedge dv \wedge dw)^{-6} \\ &= r_\kappa(\varDelta') \epsilon_\kappa,\end{aligned}$$

where $\varDelta'$ is the discriminant of the root subsystem of type $3A_2$ given by $\langle h_{12}, h_{23}, h_{45}, h_{56}, h_{123}, h\rangle$, and similarly

$$\pm|134||234||356||456||512||612| = r_\kappa(\varDelta'')\epsilon_\kappa,$$

where $\varDelta''$ is the $3A_2$-discriminant of $\langle h_{12}, h_{134}, h_{34}, h_{356}, h_{56}, h_{125}\rangle$. In this manner to each Coble covariant there is associated the discriminant of an A_2^3-subsystem of the E_6 root system and vice versa (there are indeed 40 such subsystems). In similar fashion we find that for $d = 2$ a Coble covariant is the discriminant of an A_1^7-subsystem of the E_7 root system and vice versa (there are 135 such); the two typical cases yield

$$\begin{aligned}\pm|135||146||234||256||457||127||367| &= r_\kappa(h_{135}h_{146}h_{234}h_{256}h_{457}h_{127}h_{367})\epsilon_\kappa, \\ \pm|123456||127||347||567| &= r_\kappa(h_{12}h_{34}h_{56}h_{127}h_{347}h_{567}h_7)\epsilon_\kappa.\end{aligned}$$

For $d = 4$, we do not get the discriminant of a root subsystem. The best way to describe this case is perhaps by just giving a typical case in terms of the standard representation of D_5 in $\mathbb{R}^5$ as in Bourbaki [3], where the roots are $\pm\epsilon_i \pm \epsilon_j$ with $1 \le i < j \le 5$. One such case is

$$\prod_{i\in\mathbb{Z}/5} (\epsilon_i - \epsilon_{i+1})(\epsilon_i + \epsilon_{i+1}) = \prod_{i\in\mathbb{Z}/5} \left(\epsilon_i^2 - \epsilon_{i+1}^2\right).$$

There are indeed 12 such elements up to sign.

Finally, we observe that for $d = 5$, we get

$$\pm|123||234||341||412| = r_\kappa(\varDelta)\epsilon_\kappa,$$

where $\varDelta$ is the discriminant of the full A_4-system $\langle h_{12}, h_{23}, h_{34}, h_{123}\rangle$.

This makes it clear that in all these cases $\epsilon : l_K^{d(9-d)/2} \cong L(S,E)$ is independent of the marking once we identify $L(S,E)$ with $\det V$. Notice that the polynomials defining Coble covariants indeed have the predicted degree $\frac{1}{2}d(9-d)$.

We can restate this as follows. Consider the Lobachevski lattice $\varLambda_{1,9-d}$, whose basis elements are denoted by $(\ell, e_1, \ldots, e_{9-d})$ (the inner product matrix is in diagonal form with $(+1, -1, \ldots, -1)$ on the diagonal). Put $k := -3\ell + e_1 + \cdots + e_{9-d}$ and let $(h_{123} = 3\ell - e_1 - e_2 - e_3, h_{12} = e_1 - e_2, \ldots, h_{8-d,9-d} = e_{8-d} - e_{9-d})$. This is a basis of $k^\perp$ that is at the same time a root basis of a root system R_{9-d}.

If R is a root system, then let us denote its root lattice by $Q(R)$, by $W(R)$ its Weyl group, and by $\mathfrak{h}(R) := \mathrm{Hom}(Q(R), \mathbb{C})$ the Cartan algebra on which R is defined. Let $\mathfrak{h}(R)^\circ \subset \mathfrak{h}(R)$ stand for the reflection hyperplane complement (which, in the parlance of Lie theory, is the set of its *regular elements*). We abbreviate the projectivizations of these last two spaces by $\mathbb{P}(R)$ and $\mathbb{P}(R)^\circ$. In the presence of a nondegenerate $W(R)$-invariant symmetric bilinear form on $Q(R)$ we tacitly identify $\mathfrak{h}(R)$ with its dual.

So $Q(R_{9-d}) = k^\perp$. It is clear that a marking of a Del Pezzo surface amounts to an isomorphism $\mathrm{Pic}(S) \cong \Lambda_{1,9-d}$ that sends the canonical class to k (and hence $\mathrm{Pic}_0(S)$ to $Q(R_{9-d})$). These isomorphisms are simply transitively permuted by the Weyl group $W(R_{9-d})$. If we are given a marked Fano surface $(S; e_1, \dots, e_{9-d})$ of degree d and a rational 2-form κ on S whose divisor is a cuspidal curve K, then we can associate to these data an element of $\mathfrak{h}(R_{9-d})$ by

$$Q(R_{9-d}) \cong \mathrm{Pic}_0(S) \to \mathrm{Pic}^o(K) \xrightarrow{r_\kappa} \mathbb{C}.$$

It is known that we land in $\mathfrak{h}(R_{9-d})^\circ$ and that we thus obtain a bijection between the set of isomorphism classes of systems $(S; e_1, \dots, e_{9-d}; \kappa)$ and the points of $\mathfrak{h}(R_{9-d})^\circ$. This isomorphism is evidently homogeneous of degree one: replacing κ by $c\kappa$ multiplies the image by a factor c. In other words, the set of isomorphism classes of systems $(S; e_1, \dots, e_{9-d}; K)$, where $(S; e_1, \dots, e_{9-d})$ is a marked Fano surface of degree d and K is a cuspidal anticanonical curve on K, can be identified with $\mathbb{P}(\mathfrak{h}_{9-d})^\circ$ (where we have abbreviated $\mathfrak{h}(R_{9-d})$ by $\mathfrak{h}_{9-d}$) in a such a manner that if l is a line in $\mathfrak{h}_{9-d}$ associated to $(S; e_1, \dots, e_{9-d}; K)$, then l gets identified with the line $H^0(\omega_S(K))$.

We sum up the preceding in terms of the forgetful morphism $\mathbb{P}(\mathfrak{h}_{9-d})^\circ \to \mathcal{M}^\circ_{m,d}$:

Theorem 3.1. *Assume that $d \in \{2, 3, 4, 5\}$. Then the forgetful morphism from $\mathbb{P}(\mathfrak{h}_{9-d})^\circ$ to $\mathcal{M}^\circ_{m,d}$ is surjective and flat. It is covered by a natural isomorphism between the pullback of the determinant bundle and $\mathcal{O}_{\mathbb{P}(\mathfrak{h}_{9-d})^\circ}\left(\frac{1}{2}d(9-d)\right)$ and under this isomorphism the Coble covariants form a single $W(R_{9-d})$-orbit, which up to a constant common scalar factor is as follows:*

- *(d=5) the discriminant of the A_4-system (a polynomial of degree 10),*
- *(d=4) the $W(D_5)$-orbit of the degree 10 polynomial $\prod_{i\in\mathbb{Z}/5}\left(\epsilon_i^2 - \epsilon_{i+1}^2\right)$,*
- *(d=3) the discriminants of subsystems of type $3A_2$ (a $W(E_6)$-orbit of polynomials of degree 9),*
- *(d=2) the discriminants of subsystems of type $7A_1$ (a $W(E_7)$-orbit of polynomials of degree 7).*

In particular, the Coble space $\mathcal{C}_d$ can be identified with the linear span of the above orbit of polynomials.

The Weyl group representation $\mathcal{C}_d$ was, at least for $d = 2$ and $d = 3$, already considered by Coble [4], although the notion of a Coxeter group was not available to him. We shall consider these representations in more detail in Section 4. In that section we also investigate the rational map $\mathbb{P}(\mathfrak{h}_{9-d}) \dashrightarrow \mathcal{M}^*_{m,d}$ defined by the morphism $\mathbb{P}(\mathfrak{h}_{9-d})^\circ \to \mathcal{M}^\circ_{m,d}$.

Theorem 3.1 presents (for $d = 2, 3, 4, 5$) the moduli space $\mathcal{M}^\circ_{m,d}$ as a flat $W(R_{9-d})$-equivariant quotient of $\mathbb{P}(R_{9-d})^\circ$ with $(d-2)$-dimensional fibers. Admittedly that description is somewhat indirect from a geometric point of view. We will here offer in the next two subsections a somewhat more concrete characterization when fiber and base are positive-dimensional (so for $d = 3$ and $d = 4$). A fiber is then irreducible, and we show that the $(d-1)$ $W(R_{9-d})$-invariant vector fields of lowest degree span in $\mathbb{P}(R_{9-d})^\circ$ a $(d-2)$-dimensional foliation whose leaves are the fibers of $\mathbb{P}(R_{9-d})^\circ \to \mathcal{M}^\circ_{m,d}$. We first do the case $d = 3$.

The universal parabolic curve of a cubic surface

We begin by recalling a classical definition.

Definition 3.2. *Let S be a Fano surface of degree* 3*. A point of S is said to be* parabolic *if there is a cuspidal anticanonical curve on S that has its cusp singularity at that point.*

We may think of the surface S as anticanonically embedded in $\mathbb{P}^3$ as a smooth cubic surface. If $F(Z_0, Z_1, Z_2, Z_3)$ is a defining equation, then the locus of parabolic points is precisely the part of S where S meets its *Hessian surface* (defined by $\det(\partial^2 F/\partial Z_i \partial Z_j)$) transversally. So this is a nonsingular curve on S that need not be closed (in fact, it isn't: a point of S where its tangent plane intersects S in a union of a conic and a line tangent lies in the Zariski boundary of the parabolic curve). If we fix a marking for S, so that a point of $p \in \mathcal{M}^\circ_{m,3}$ is determined, then the parabolic locus can be identified with the fiber of $\mathbb{P}(R_{9-d})^\circ$ over p.

We now return to the situation of the beginning of this section, where we essentially have a fixed cuspidal cubic curve K in the projective plane $\mathbb{P}^2$ whose smooth part in terms of affine coordinates has the parameter form $(v, w) = (t, t^3)$ (this puts the cusp at infinity and the unique flex point at the origin). The points $p_1, \dots, p_6 \in \mathbb{P}^2_\ell$ that we blow up in order to produce S lie on K_{reg}; we denote their $t(= v)$-coordinates by $t_1, \dots, t_6$. The system $(S; e_1, \dots, e_6; K)$ defines a point $\tilde{p} \in \mathbb{P}(E_6)^\circ$, and $(t_1, \dots, t_6)$ describes a point of $\mathfrak{h}(E_6)^\circ$ that lies over $\tilde{p}$. The vector fields X on $\mathbb{P}^2$ with the property that X is tangent to K at the points $p_1, \dots, p_6$ make up a vector space of dimension two. It contains the field $X_E = v\partial/\partial v + 3w\partial/\partial w$, which is tangent to K_{reg} everywhere (it generates a $\mathbb{C}^\times$-action on $\mathbb{P}^2$ that preserves K). If X is in this vector space, then

$$\hat{X} := \sum_i X(p_i) \frac{\partial}{\partial t_i}$$

is a tangent vector of $\mathfrak{h}(E_6)$ at $(t_1, \dots, t_6) \in \mathfrak{h}(E_6)^\circ$. For X_E this yields $\hat{X}_E = \sum_{i=1}^6 t_i \frac{\partial}{\partial t_i}$, in other words, we get the Euler field of $\mathfrak{h}(E_6)$ at $(t_1, \dots, t_6)$.

Lemma 3.3. *If X is not proportional to X_E, then the line in $\mathbb{P}(E_6)$ through $\tilde{p}$ that is defined by $\hat{X}$ is tangent to the fiber of $\mathbb{P}(E_6)^\circ \to \mathcal{M}^\circ_{m,3}$ at $\tilde{p}$.*

Proof. This is mostly a matter of geometric interpretation. If we view $\hat{X} = (X(p_1), \dots, X(p_6))$ as an infinitesimal displacement of the point configuration $(p_1, \dots, p_6)$ in $\mathbb{P}^2$, then $\hat{X}$ does not effectively deform the corresponding Fano surface, because X is an infinitesimal automorphism of $\mathbb{P}^2$. But if we view $\hat{X}$ as an infinitesimal displacement $(p_1, \dots, p_6)$, then it will induce a nontrivial line field (a priori with singularities) unless X is tangent to K. This last condition is equivalent to X being proportional to X_E. □

We now calculate the resulting field on $\mathbb{P}(E_6)^\circ$. A vector field X on $\mathbb{P}^2$ has in our affine coordinates (v, w) the form

$$X = (a_0 + a_1 v + a_2 w + c_1 v^2 + c_2 vw)\frac{\partial}{\partial v} + (b_0 + b_1 v + b_2 w + c_1 vw + c_2 w^2)\frac{\partial}{\partial w}.$$

Since we may calculate modulo X_E, we assume $b_2 = 0$. The condition that X be tangent to C at p_i amounts to

$$3t^2(a_0 + a_1 t + a_2 t^3 + c_1 t^2 + c_2 t^4) = b_0 + b_1 t + c_1 t^4 + c_2 t^6$$

for $t = t_i$, or equivalently, that $2c_2 t^6 + 3a_2 t^5 + 2c_1 t^4 + 3a_1 t^3 + 3a_0 t^2 - b_1 t - b_0$ have (the distinct) zeros $t = t_i$ for $i = 1, \dots, 6$. This means that

$$\frac{a_2}{c_2} = -\tfrac{2}{3}\sigma_1, \ \frac{c_1}{c_2} = \sigma_2, \ \frac{a_1}{c_2} = -\tfrac{2}{3}\sigma_3, \ \frac{a_0}{c_2} = \tfrac{2}{3}\sigma_4, \ \frac{b_1}{c_2} = 2\sigma_5, \ \frac{b_0}{c_2} = -2\sigma_6,$$

where σ_i stands for the ith symmetric function of $t_1, \dots, t_6$. So we may normalize X by taking $c_2 = 1$. The value of X in p_i is in terms of the t-coordinate its x-component, and hence equal to

$$\left(a_0 + a_1 t_i + c_1 t_i^2 + a_2 t_i^3 + c_2 t_i^4\right)\frac{\partial}{\partial t} = \left(\tfrac{2}{3}\sigma_4 - \tfrac{2}{3}\sigma_3 t_i + \sigma_2 t_i^2 - \tfrac{2}{3}\sigma_1 t_i^3 + t_i^4\right)\frac{\partial}{\partial t}.$$

It follows that

$$\hat{X} = \sum_{i=1}^{6}\left(\tfrac{2}{3}\sigma_4 - \tfrac{2}{3}\sigma_3 t_i + \sigma_2 t_i^2 - \tfrac{2}{3}\sigma_1 t_i^3 + t_i^4\right)\frac{\partial}{\partial t_i}. \qquad (*)$$

We see in particular that if we regard $\hat{X}$ as a vector field ($(t_1, \dots, t_6)$ varies), then it is homogeneous of degree 3. The space of homogeneous vector fields of degree 3 on $\mathfrak{h}(E_6)$ is as a W_6-representation space isomorphic to $\mathrm{Sym}^4(\mathfrak{h}^*(E_6)) \otimes \mathfrak{h}(E_6)$; this has a one-dimensional space of invariants and contains no other $W(E_6)$-invariant one-dimensional subspace. It follows that $\hat{X}$ is $W(E_6)$-invariant and is characterized by this property up to a constant factor.

Corollary 3.4. *The fibration $\mathbb{P}(E_6)^\circ \to \mathcal{M}_{m,3}^\circ$ integrates the one-dimensional foliation defined by a $W(E_6)$-invariant vector field that is homogeneous of degree three.*

A natural way to produce such an invariant vector field is to take the nonzero $W(E_6)$ invariant polynomials f_2, f_5 on $\mathfrak{h}(E_6)$ of degrees 2 and 5 (these are unique up to a constant factor); since f_2 is nondegenerate we can choose coordinates $z_1, \dots, z_6$ such that $f_2 = \sum_i z_i^2$. The gradient vector field relative to f_2,

$$\nabla f_5 = \sum_{i=1}^{6} \frac{\partial f_5}{\partial z_i} \frac{\partial}{\partial z_i},$$

is a $W(E_6)$-invariant homogeneous vector field of degree 3. So we can restate the preceding corollary as the following theorem:

Theorem 3.5. *Let $\mathfrak{h}$ denote the natural representation space of a Coxeter group W of type E_6. The natural W-invariant rational dimension-one foliation on $\mathbb{P}(\mathfrak{h})^\circ$ of degree 3 (i.e., the one defined by the gradient of a nonzero invariant quintic form with respect to a nonzero (hence nondegenerate) invariant quadratic form on $\mathfrak{h}$) is algebraically integrable and has a leaf space that is in a W-equivariant manner isomorphic to the moduli space of marked smooth cubic surfaces.*

Moduli of degree 4 Del Pezzo surfaces

This is a slight modification of the argument for the degree 3 case. We have one point fewer and so by letting t_6 move over the affine line we may regard formula $(*)$ as defining a one-parameter family of vector fields on $\mathfrak{h}(D_5)$. For $i \geq 1$, we have $\sigma_i(t_1, \dots, t_6) = \sigma_i(t_1, \dots, t_5) + t_6 \sigma_{i-1}(t_1, \dots, t_5)$, and so we immediately see that this is a linear family spanned by the two vector fields

$$\hat{X}_3 = \textstyle\sum_{i=1}^{5} \left(\frac{2}{3}\sigma_4 - \frac{2}{3}\sigma_3 t_i + \sigma_2 t_i^2 - \frac{2}{3}\sigma_1 t_i^3 + t_i^4 \right) \frac{\partial}{\partial t_i},$$

$$\hat{X}_2 = \textstyle\sum_{i=1}^{5} \left(\frac{2}{3}\sigma_3 - \frac{2}{3}\sigma_2 t_i + \sigma_1 t_i^2 - \frac{2}{3} t_i^3 \right) \frac{\partial}{\partial t_i}.$$

The subscript indicates of course the degree. Since $\hat{X}$ is $W(E_6)$-invariant, $\hat{X}_3$ and $\hat{X}_2$ will be invariant under the $W(E_6)$-stabilizer of e_6, that is, $W(D_5)$. The $W(D_5)$-invariant vector fields on $\mathfrak{h}(D_5)$ form a free module on the polynomial algebra of $W(D_5)$-invariant functions. The latter algebra has its generators f_2, f_4, f_5, f_6, f_8 in degrees indicated by the subscript. The generator f_2 is a nondegenerate quadratic form, and the module of invariant vector fields is freely generated by the gradients of the f_i relative to f_2, $X_{i-2} := \nabla f_i$ (these have the degree indicated by the subscript; X_0 is the Euler field). We conclude that the plane distribution on $\mathbb{P}(D_5)^\circ$ spanned by the vector fields X_2 and X_3 is also defined by $\hat{X}_2$ and $\hat{X}_3$. We therefore have the following:

Theorem 3.6. *Let $\mathfrak{h}$ denote the natural representation space of a Coxeter group W of type D_5. The natural W-invariant rational dimension-two foliation on $\mathbb{P}(\mathfrak{h})^\circ$ defined by the gradients of nonzero invariant forms of degree*

4 and 5 with respect to a nonzero (hence nondegenerate) invariant quadratic form on $\mathfrak{h}$ is algebraically integrable and has a leaf space that is in a W-equivariant manner isomorphic to the moduli space of marked Fano surfaces of degree 4.

Remark 3.7. The Frobenius integrability is remarkable, because it tells us that the degree four vector field $[X_2, X_3]$ does not involve the degree four generator X_4. It is of course even more remarkable that it is algebraically so (in the sense that its leaves are the fibers of a morphism). It makes one wonder how often this happens. For instance, one can ask, given a Coxeter arrangement complement $\mathfrak{h}^\circ$ and a positive integer k, when is the distribution on $\mathfrak{h}^\circ$ spanned by the subset of homogeneous invariant generating vector fields whose degree is $\leq k$ algebraically integrable?

4 Coble's representations

This section discusses the main properties of the representations of a Weyl group of type D_5, E_6, or E_7 that we encountered in Theorem 3.1.

Macdonald's irreducibility theorem

We will use the following beautiful (and easily proved!) theorem of MacDonald [17], which states that the type of representation under consideration is irreducible.

Proposition 4.1 (Macdonald). *Let R be a root system, $\mathfrak{h}$ the complex vector space it spans and $S \subset R$ a reduced root subsystem. Then the $W(R)$-subrepresentation of $\mathrm{Sym}^{|S|/2}\mathfrak{h}$ generated by the discriminant of S is irreducible. In particular, the Coble representations of type E_6 and E_7 are irreducible.* □

Proof. Since the proof is short, we reproduce it here. If $L \subset \mathrm{Sym}^{|S|/2}\mathfrak{h}$ denotes the line spanned by the discriminant of S, then $W(S)$ acts on L with the sign character. In fact, L is the entire eigensubspace of $\mathrm{Sym}^{|S|/2}\mathfrak{h}$ defined by that character, for if $G \in \mathrm{Sym}^{|S|/2}\mathfrak{h}$ is such that $s(G) = -G$ for every reflection s in $W(S)$, then G is zero on each reflection hyperplane of $W(S)$ and hence divisible by the discriminant of S. Since G and the discriminant have the same degree, G must be proportional to it.

Let $V = \mathbb{C}[W(R)]L$ be the $W(R)$-subrepresentation of $\mathrm{Sym}^{|S|/2}\mathfrak{h}$ generated by L. We must prove that every W-equivariant map $\phi : V \to V$ is given by a scalar. From the preceding it follows that ϕ preserves L and so is given on L as multiplication by a scalar, λ say. Then $\phi - \lambda\mathbf{1}_V$ is zero on L and hence zero on V. □

Unfortunately, Macdonald's theorem does not come with an effective way to compute the degree of such representations, and that is one of several good reasons to have a closer look at them. (The Coble representations had indeed been considered by Coble and presumably by others before him. Their irreducibility and their degrees were known at the time.)

It will be convenient (and of course quite relevant for the application we have in mind) to work for $d = 2, 3, 4$ with the Manin model of the R_{9-d} root system as sitting in the Lobachevski lattice $\Lambda_{1,9-d}$, so that $\mathfrak{h}_{9-d} := \mathfrak{h}(R_{9-d})$ is the orthogonal element of $k = -3\ell + e_1 + \cdots + e_{9-d}$. For $d = 2, 3, 4$ we have a corresponding (Coble) representation $\mathcal{C}_d = \mathcal{C}(R_{9-d})$, which in case $d = 3$ (respectively $d = 2$) is spanned by the discriminants subsystems of type $3A_2$ (respectively $7A_1$). We may regard $\mathcal{C}_d$ as a linear system of hypersurfaces of degree 10 ($d = 4$), 9 ($d = 3$), or 7 ($d = 2$) in $\mathbb{P}(\mathfrak{h}_{9-d})$. Among our goals is to compute the dimension of this system and to investigate its separating properties.

The Manin basis recognizes one particular weight, namely the orthogonal projection of ℓ in $\mathfrak{h}_{9-d}$. Its $W(R_{9-d})$-stabilizer is the symmetric group $\mathcal{S}_{9-d}$ of $e_1, \dots, e_{9-d}$ (a Weyl subgroup of type A_{8-d}). We shall denote by $\pi_{9-d} : \mathbb{C} \otimes \Lambda_{1,9-d} \to \mathfrak{h}_{9-d}$ the orthogonal projection. So $\pi_{9-d}(e_i) = e_i + \frac{1}{3}k$.

The Coble representation of a Weyl group of type E_6

Here $R = R_6$ and the representation of $W(E_6)$ in question is the subspace $\mathcal{C}_3 \subset \mathrm{Sym}^9 \mathfrak{h}_6$ spanned by the discriminants of subsystems of type $3A_2$ of R. Following Proposition 4.1, this representation is irreducible. We shall prove that $\dim \mathcal{C}_3 = 10$ and that quotients of elements of $\mathcal{C}_3$ separate the isomorphism types of cubic surfaces.

Lemma 4.2. *The Weyl group $W(R)$ acts transitively on the collection of ordered triples of mutually orthogonal roots. If $(\alpha_1, \alpha_2, \alpha_3)$ is such a triple, then*

(i) there is a root $\alpha \in R$ perpendicular to each α_i and this root is unique up to sign,
(ii) the roots α_1, α_2, α_3, α belong to the (unique) subsystem of type D_4,
(iii) there are precisely two subsystems of type $3A_2$ containing $\{\alpha_1, \alpha_2, \alpha_3\}$ and these two subsystems are interchanged by s_α.

Proof. The transitivity assertion and the properties (i) and (ii) are known and a proof goes like this: The orthogonal complement of a root $\alpha_1 \in R$ is a subsystem $R' \subset R$ of type A_5, the orthogonal complement of a root $\alpha_2 \in R'$ in R' is a subsystem $R'' \subset R'$ of type A_3, and the orthogonal complement of a root $\alpha_3 \in R''$ is a subsystem $R''' \subset R''$ of type A_1. Since all the root systems encountered have the property that their Weyl group acts transitively on the roots, the first assertion follows. Notice that we proved (i) at the same time. The remaining properties now need to be verified only for a particular choice of $(\alpha_1, \alpha_2, \alpha_3)$.

We take this triple to be (h_{12}, h_{34}, h_{56}). Then we may take $\alpha = h$, and we see that these are roots of the D_4-system spanned by h_{12}, h_{34}, h_{56}, h_{135}. The two $3A_2$-subsystems containing $\{h_{12}, h_{34}, h_{56}\}$ are then easily seen to be $\langle h_{12}, h_{134}\rangle \perp \langle h_{34}, h_{356}\rangle \perp \langle h_{56}, h_{125}\rangle$ and $\langle h_{12}, h_{156}\rangle \perp \langle h_{34}, h_{123}\rangle \perp \langle h_{56}, h_{345}\rangle$. We observe that s_h interchanges them. □

The following notion is the root-system analogue of its namesake introduced by Allcock and Freitag [1].

Lemma-Definition 4.3. Let R be a root system of type E_6 and $S \subset R$ a subsystem of type $3A_2$. If $\alpha \in R$ is not orthogonal to any summand of S, then the roots in S orthogonal to α make up a subsystem of type $3A_1$ (which then must meet every summand of S). This sets up a bijection between the antipodal pairs $\{\pm\alpha\}$ that are not orthogonal to any summand of S and $3A_1$-subsystems of S.

For (S, α) as above and S^+ a set of positive roots for S, the degree-nine polynomial $(1 - s_\alpha)\Delta(S^+)$ is called a *cross* of R.

Proof. If we are given a A_2-subsystem of R, then any root not in that subsystem is orthogonal to some root in that subsystem. This implies that in the above definition we can find in each of the three A_2-summands a root that is orthogonal to α. Since α is not orthogonal to any summand, this root is unique up to sign and so the roots in S fixed by s form a $3A_1$-subsystem as asserted.

Conversely, if $R' \subset S$ is a subsystem of type $3A_1$, and $\alpha \in R - R'$ is as in Lemma 4.2, then $s = s_\alpha$ has the desired property. □

Lemma 4.4. *Let $S^+ \subset R$ and $\alpha \in R - S$ be as in Lemma 4.3. If $\alpha_1, \alpha_2, \alpha_3$ are the roots in S^+ perpendicular to α, then the cross $(1 - s_\alpha)\Delta(S^+)$ is divisible by $\alpha_1\alpha_2\alpha_3\alpha$,*

$$(1 - s_\alpha)\Delta(S^+) = \alpha_1\alpha_2\alpha_3\alpha F_1,$$

and the quotient $F_1 \in \mathrm{Sym}^5\mathfrak{h}_6$ is invariant under the Weyl group of the D_4-subsystem that contains $\alpha_1, \alpha_2, \alpha_3, \alpha$.

Proof. It is clear that both $\Delta(S^+)$ and $s_\alpha\Delta(S^+) = \Delta(s_\alpha S^+)$ are divisible by $\alpha_1\alpha_2\alpha_3$. It is also clear that $(1 - s_\alpha)\Delta(S^+)$ is divisible by α. So F_1 is defined as an element of $\mathrm{Sym}^5\mathfrak{h}$.

We will now prove that there exists a $g \in \mathrm{GL}(\mathbb{Q} \otimes \Lambda_{1,6})$ that centralizes the Weyl group in question and is such that the transform of F_1 under g^{-1} is a $W(R)$-invariant in $\mathrm{Sym}^5\mathfrak{h}$. This will clearly suffice.

We may, in view of Lemma 4.2, assume without loss of generality that $(\alpha_1, \alpha_2, \alpha_3, \alpha) = (h_{12}, h_{34}, h_{56}, h)$, so that the D_4-subsystem containing these roots is $\langle h_{12}, h_{34}, h_{56}, h_{135}\rangle$. Denote by $\mathfrak{h}'$ the subspace of $\mathfrak{h}$ spanned by these roots.

We first recall a remarkable result due to Naruki. The set of exceptional classes that have inner product 1 with α is $\{e_1, \ldots, e_6\}$ (this set and its s_α-transform make up what is classically known as a *double six*). Consider the element $(1-s_\alpha)\prod_{i=1}^6 \pi_*(e_i) \in \mathrm{Sym}^6\mathfrak{h}$. It is clearly divisible by α and the quotient $F \in \mathrm{Sym}^5\mathfrak{h}$ will evidently be invariant under a Weyl subgroup of $W(R)$ of type A_5+A_1. But according to Naruki [15, p. 235], F is even invariant under all of $W(R)$.

The orthogonal complement of $\mathfrak{h}'$ in $\mathbb{C}\otimes\Lambda_{1,6}$ is spanned by the members of the "anticanonical triangle"

$$(\epsilon_0 := \ell - e_1 - e_2, \epsilon_1 := \ell - e_3 - e_4, \epsilon_2 := \ell - e_5 - e_6),$$

and the intersection $\mathfrak{h}\cap\mathfrak{h}'^{\perp}$ is spanned by the differences $\epsilon_0-\epsilon_1$ and $\epsilon_1-\epsilon_2$. Let $g \in \mathrm{GL}(\mathbb{Q}\otimes\Lambda_{1,6})$ be the transformation that is the identity on $\mathfrak{h}'$ and takes ϵ_i to $\epsilon_i + 2\epsilon_{i+1}$ for $i \in \mathbb{Z}/3$. This transformation preserves $\mathfrak{h}'^{\perp}$ and hence commutes with all the transformations that preserve $\mathfrak{h}'$ and act as the identity on $\mathfrak{h}'^{\perp}$. We also note that $g(k) = 3k$, and that g preserves the orthogonal complement of k, and hence g commutes with π.

We claim that $g\pi(e_1) = -h_{134} \in R$. One easily checks that $2e_1+\epsilon_0+k \in \mathfrak{h}'$ and so

$$\begin{aligned} g(e_1) &= g\left(\frac{1}{2}(2e_1+\epsilon_0+k)\right) - \frac{1}{2}g(\epsilon_0) - \frac{1}{2}g(k) \\ &= \frac{1}{2}(2e_1+\epsilon_0+k) - \frac{1}{2}(\epsilon_0+2\epsilon_1) - \frac{3}{2}k \\ &= e_1 - \epsilon_1 - k \\ &= -\ell + e_1 + e_3 + e_4 - k \\ &= -h_{134} + k. \end{aligned}$$

Applying π to this identity yields $g\pi(e_1) = -h_{134}$.

We get similar formulas for the $g\pi(e_i)$ and thus find that

$$g\pi\{e_1,\dots,e_6\} = \langle -h_{134}, -h_{234}\rangle \perp \langle -h_{356}, -h_{456}\rangle \perp \langle -h_{125}, -h_{126}\rangle.$$

Notice that the union of this set with $\{h_{12}, h_{34}, h_{56}\}$ is a system of positive roots of a $3A_2$-system. This union will be our S^+. So g_* takes the polynomial $h_{12}h_{34}h_{56}\prod_{i=1}^6 \pi(e_i)$ to $\Delta(S^+)$. Since s_α commutes with g we have

$$\begin{aligned} (1-s_\alpha)\Delta(S^+) &= h_{12}h_{34}h_{56}\,g\Big((1-s_\alpha)\textstyle\prod_{i=1}^6 \pi(e_i)\Big) \\ &= h_{12}h_{34}h_{56}\,g(\alpha F) \\ &= h_{12}h_{34}h_{56}\,\alpha g(F) \end{aligned}$$

so that $F_1 = g(F)$. This proves the lemma. □

Corollary 4.5. *For any three pairwise perpendicular roots in R there exists a cross that is divisible by their product. This cross is unique up to sign and is also divisible by a root perpendicular to these three. This yields a bijection between $4A_1$-subsystems of R and antipodal pairs of crosses.* □

Corollary 4.6. *Let $R' \subset R$ be a subsystem of type D_4. The $4A_1$-subsystems of R' define three crosses up to sign whose sum is zero. These crosses span a $W(R')$-invariant plane in $\mathcal{C}_3$. A quotient of the discriminants of two $4A_1$-subsystems of R' is a quotient of two crosses.* □

We now fix a D_5-subsystem $R_o \subset R$. It has precisely five subsystems of type D_4. As we just observed, each of these defines a plane in $\mathcal{C}_3$. Therefore, the crosses associated to the $4A_1$-subsystems of R_o span a subspace of $\mathcal{C}_3$ of dimension at most 10.

Lemma 4.7. *For every subsystem $S \subset R$ of type $3A_2$, $S \cap R_o$ is of type $2A_1 + A_2$ and hence contains three subsystems of type $3A_1$; for every such $3A_1$-subsystem the associated $4A_1$-subsystem of R is in fact contained in R_o. Moreover, $S \mapsto S \cap R_o$ defines a bijection between the $3A_2$-subsystems of R and the $2A_1 + A_2$-subsystems of R_o, and $W(R_o)$ acts transitively on both sets.*

Proof. It is easy (and left to the reader) to find one subsystem $S \subset R$ of type $3A_2$ such that $R_o \cap S$ has the stated properties. It therefore suffices to prove the transitivity property. This involves a simple count: The $W(R_o)$-stabilizer $R_o \cap S$ of R_o contains $W(R_o \cap S)$ as a subgroup of index two (there is an element in the stabilizer that interchanges the A_1-summands and is minus the identity on the A_2-summand), and so the number of systems $W(R)$-equivalent to $R_o \cap S$ is $|W(D_5)|/2|W(A_2 + 2A_1)| = 40$. That is just as many as there are $3A_2$-subsystems of R. ◻

We continue with the D_5-subsystem $R_o \subset R$ that we fixed above. Let $S \subset R$ be any subsystem of type $3A_2$. By Lemma 4.7, $R_o \cap S$ is of type $2A_1 + A_2$. Let be s_1, s_2, s_3 be the three reflections in the Weyl group of the A_2-summand. The two A_1-summands and the antipodal root pair attached to s_i make up a $3A_1$-subsystem $R_1^{(i)}$ of $R_o \cap S$. Each of these subsystems is contained in a unique $4A_1$-subsystem. Let $s^{(i)}$ denote the reflection in the extra A_1-summand. According to Lemma 4.7, $s^{(i)} \in W(R_o)$, so that $R_2^{(i)} := s^{(i)}S$ has the property that $R^{(i)} \cap R_o = R \cap R_o$.

Lemma 4.8. *The discriminant $\Delta(S^+)$ is fixed under $s' + s'' + s'''$, in other words,*

$$2\Delta(S^+) = (1 - s')\Delta(S^+) + (1 - s'')\Delta(S^+) + (1 - s''')\Delta(S^+),$$

where we note that the right-hand side is a sum of three crosses attached to subsystems of R_o of type $4A_1$. In particular, $\mathcal{C}_3$ is generated by the crosses.

Proof. It is clear that $s_i\Delta(S^+) = -\Delta(S^+)$. Since $s_3 = s_1 s_2 s_1^{-1}$, we have $R_1''' = s_1(R_1'')$. This implies that $s''' = s_1 s'' s_1$ and so

$$(s'' + s''')\Delta(S^+) = (s'' + s_1 s'' s_1)\Delta(S^+) = (1 - s_1)s''\Delta(S^+) = (1 - s_1)\Delta(s''S^+).$$

Since $s_1 \notin W(s''S)$, the right-hand side is a cross. We claim that this cross equals the cross $(1 - s')\Delta(S^+)$ up to sign. For this it suffices to show that there exist four perpendicular roots such that each is divisible by three of them. It is clear that $(1 - s_1)s''\Delta(S^+)$ is divisible by a root attached to s_1 and by the roots in the two A_1-summands of S^+ (for these are unaffected by s'' and s_1).

On the other hand, $(1-s')\Delta(S^+)$ is divisible by a root attached to s' and the roots in the two A_1-summands of S. It remains to observe that the roots attached to s_1 and s' are perpendicular.

Thus $(s''+s''')\Delta(S^+) = \pm(1-s')\Delta(S^+)$. Suppose the minus sign holds, so that $1-s'+s''+s'''$ kills $\Delta(S^+)$. The cyclic permutation $1+s'+s''-s'''$ then also kills $\Delta(S^+)$ and hence so will $1+s''$. In other words, $\Delta(S^+)$ will be anti-invariant under s''. Since $s''\notin W(S)$, this is a contradiction. Hence the plus sign holds and the lemma follows. □

Theorem 4.9. *The planes defined by the five subsystems of R_o of type D_4 make up a direct sum decomposition of $\mathcal{C}_3$. In particular, $\mathcal{C}_3$ is the irreducible representation of the E_6-Weyl group of degree 10.*

Proof. Lemma 4.8 shows that $\mathcal{C}_3$ is spanned by the crosses attached to $4A_1$-subsystems of R_o. So the five planes in question span $\mathcal{C}_3$, and $\dim \mathcal{C}_3 \le 10$. The irreducible representations of $W(R)$ of degree < 10 are the trivial representation, the sign representation (which are both of degree 1), and the defining representation (of degree 6), and $\mathcal{C}_3$ is clearly none of these. The theorem follows. □

We now determine the common zero set of the Coble covariants. We first make some general remarks that also apply to the D_5 and the E_7-cases. The zero set of a Coble invariant is a union of reflection hyperplanes, and hence each irreducible component of their common intersection, $Z_r \subset \mathfrak{h}_r$, is an intersection of reflection hyperplanes. Since Z_r is invariant under the Weyl group, so is the collection of its irreducible components. So an irreducible component is always the translate of common zero set of a subset of the given root basis of R_r. (This subset need not be unique.)

Proposition 4.10. *The common zero set $Z_6 \subset \mathfrak{h}$ of the members of $\mathcal{C}_3$ is the union of the linear subspaces that are pointwise fixed by a Weyl subgroup of type A_3.*

Proof. We first verify that for any A_3-subsystem of R, the subspace of $\mathfrak{h}$ perpendicular to it is in the common zero set of the members of $\mathcal{C}_3$. Since an A_3-subsystem is contained in a D_5-subsystem, it is in view of Lemma 4.7 enough to show that an (A_2+2A_1)-subsystem and an A_3-subsystem in a D_5-system always meet. This is easily verified.

We next show that Z_6 is not larger. Any subsystem generated by fundamental roots that does not contain an A_3-system is contained in a subsystem of type $2A_2+A_1$. There is a single Weyl group equivalence class of such subsystems, and so it suffices to give two subsystems of R, of type $3A_2$ and of type $2A_2+A_1$ that are disjoint. We take $\langle h_{12}, h_{23}, h_{45}, h_{56}, h_{123}, h\rangle$ and $\langle h_{16}, h_{125}, h_{34}, h_{136}, h_{25}\rangle$. □

Question 4.11. Is $\mathcal{C}_3$ the space of degree 9 polynomials on $\mathfrak{h}$ that vanish on Z_6? This is probably equivalent to the completeness of $\mathcal{C}_3$ as a linear system on $\mathcal{M}^*_{m,3}$ (which is known, though in a rather indirect manner; see Remark 5.10).

The Coble representation of a Weyl group of type E_7

The Weyl group $W(E_7)$ decomposes as $W_+(E_7) \times \{1, c\}$, where $W_+(E_7) \subset W(E_7)$ is the subgroup of elements that have determinant one in the Coxeter representation and $c \in W(E_7)$ is minus the identity in the Coxeter representation. This implies that every irreducible representation of $W(E_7)$ is obtained as an irreducible representation of $W_+(E_7)$ plus a decree as to whether c acts as 1 or as -1.

We know that the representation of $W(E_7)$ defined by $\mathcal{C}_2$ (which we recall, is spanned by products of seven pairwise perpendicular roots of the E_7 root system) is irreducible, and we want to prove the following:

Proposition 4.12. *The representation $\mathcal{C}_2$ of $W(E_7)$ is of degree* 15*, and the nontrivial central element of $W(E_7)$ acts as -1.*

It is known that there is just one isomorphism type of irreducible representations of $W_+(E_7)$ in degree 15, and so Proposition 4.12 identifies the isomorphism type of the representation.

In what follows, R stands for the root system R_7 of type E_7.

Lemma 4.13. *The Weyl group $W(R)$ acts transitively on the collection of $7A_1$-subsystems of R. If we are given a subsystem R' of type $2A_1$, then the roots perpendicular to R' make up a subsystem of type $A_1 + D_4$. In particular, there is a unique subsystem of type $3A_1$ that contains R' and is orthogonal to a subsystem of type D_4. Conversely, the roots perpendicular to a given subsystem of R of type D_4 make up a system of type $3A_1$.*

Proof. This lemma is known and the proof is standard. The first assertion follows from the fact that the roots orthogonal to a given root of R form a subsystem of type D_6 and the roots orthogonal to a root of a root system of type D_6 form a subsystem of type $D_4 + A_1$.

Any root subsystem of R of type D_4 is saturated, and so a root basis of this subsystem extends to a root basis of R. Since the group $W(R)$ acts transitively on the set of root bases, it also acts transitively on the set of subsystems of type D_4. $\square$

So if we have a subsystem $R_1 \subset R$ of type $7A_1$, then any two summands of R_1 (making up a subsystem $R' \subset R_1$ of type $2A_1$) determine a third summand, and the remaining four summands will lie in a D_4-subsystem. In this way we can construct a 2-dimensional simplicial complex with seven vertices indexed by the summands of R_1: three vertices span a 2-simplex if and only if the orthogonal complement of the sum of their associated A_1-summands is of type D_4. The preceding lemma tells us that every edge is in exactly one 2-simplex. Probably the $W(E_7)$-stabilizer of R_1 is the full automorphism group of this complex.

The subsystems of type $7A_1$ make up two $\mathcal{S}_7$-orbits, represented by

(A) $\langle h_7, h_{12}, h_{34}, h_{56}, h_{127}, h_{347}, h_{567}\rangle$, 105 in number, and
(B) $\langle h_{123}, h_{145}, h_{167}, h_{256}, h_{247}, h_{357}, h_{346}\rangle$, of which there are 30.

We designate by the same letters (A) and (B) the type of the corresponding product of roots.

Lemma 4.14. *Let F be a product of roots of type* (B)*. Then the $\mathcal{S}_7$-stabilizer of F acts transitively on its factors and has order* $7.3.2^3$*. The subgroup that stabilizes a given factor is isomorphic to $\mathcal{S}_4$.*

Proof. Let F be of type (B) and let α be a factor of F. Write $\alpha = h_{abc}$. Then the other factors are of the form h_{axy}, h_{azw}, h_{bxz}, h_{byw}, h_{cxw}, h_{cyz}, where x, y, z, w are the distinct elements of $\{1, 2, \ldots, 7\} - \{a, b, c\}$. So these factors are given by an indexing by a, b, c of the three ways we can split $\{1, 2, \ldots, 7\} - \{a, b, c\}$ into two pairs. This description proves that $\mathcal{S}_7$ is transitive on the collection of pairs (F, α) with stabilizer mapping isomorphically onto the permutation group of $\{x, y, z, w\}$. The lemma follows. □

Lemma 4.15. *The space $\mathcal{C}_2$ is spanned by the* 30 *root products of type (B) and is annihilated by $\sum_{w\in\mathcal{S}(i,j,k)} \mathrm{sign}(w)w$ for any* 3*-element subset $\{i, j, k\}$ of $\{1, \ldots, 7\}$.*

For the proof we need the following lemma:

Lemma 4.16. *A root system S of type D_4 contains exactly three subsystems of type $4A_1$, and the discriminants of these three subsystems (relative to some choice of positive roots) are such that a signed sum is zero. More precisely, if $S_o \subset S$ is a subsystem of type $4A_1$, then the eight reflections in $W(S) - W(S_o)$ decompose into two equivalence classes with the property that two reflections s, s' belong to different classes if and only if they do not commute. In that case S_o, sS_o, $s'S_o$ are the distinct $4A_1$-subsystems of S, and if f is the product of four pairwise perpendicular roots in S_o, then $f = s(f) + s'(f)$. The plane in the fourth symmetric power of the complex span of the root system generated by these discriminants affords an irreducible representation of the Weyl group of the root system.*

Proof. In terms of the standard model for the D_4-system, the set of vectors $\pm\epsilon_i \pm \epsilon_j$, $1 \le i < j \le 4$, in Euclidean 4-space, the $4A_1$-subsystems correspond to the three ways of partitioning $\{1, 2, 3, 4\}$ into parts of size 2. For instance, the partition $\{\{1, 2\}, \{3, 4\}\}$ yields $\{\pm\epsilon_1 \pm \epsilon_2, \pm\epsilon_3 \pm \epsilon_4\}$, whose discriminant is (up to sign) equal to $\epsilon_1^2\epsilon_3^2 + \epsilon_2^2\epsilon_4^2 - \epsilon_1^2\epsilon_4^2 + \epsilon_2^2\epsilon_3^2$. We can verify the lemma for $s = s_{\epsilon_1-\epsilon_3}$, $s' = s_{\epsilon_1-\epsilon_4}$ and deduce the general case from that.

The last clause is easily verified. □

Proof (of Lemma 4.15). Consider $F = h_7h_{12}h_{34}h_{56}h_{127}h_{347}h_{567}$ (a typical root product of type (A)). The four factors that are not of type (1),

$h_7, h_{12}, h_{34}, h_{56}$, lie in a subsystem of type D_4. If we let s (respectively s') be the reflection in h_{135} (respectively h_{246}), then

$$s(h_7h_{12}h_{34}h_{56}) = h_{246}h_{235}h_{145}h_{136},$$
$$s'(h_7h_{12}h_{34}h_{56}) = h_{135}(-h_{146})(-h_{236})(-h_{245}) = -h_{135}h_{146}h_{236}h_{245}.$$

Notice that the second product is obtained from the first by applying minus the transposition (34). According to Lemma 4.16 we then have

$$h_7h_{12}h_{34}h_{56} = (1-(34))\, h_{246}h_{235}h_{145}h_{136}. \tag{3}$$

After multiplying both sides by $h_{127}h_{347}h_{567}$, we see that F has been written as a difference of two products of type (B): $f = (1-(34))\, G$ with

$$G := h_{127}h_{347}h_{567}h_{246}h_{235}h_{145}h_{136}.$$

In particular, the type (B) products generate $\mathcal{C}_2$. It follows from Lemma 4.14 that the $\mathcal{S}_7$-stabilizer of G has two orbits in the collection of 3-element subsets $\{i,j,k\} \subset \{1,\dots,7\}$: those for which h_{ijk} is a factor of G and those for which there exists a factor h_{abc} of G with $\{a,b,c\} \cap \{i,j,k\} = \emptyset$. So there are only two cases to verify.

We first do the case $I = \{3,4,5\}$. We are then in the second case because $I \cap \{1,2,7\} = \emptyset$ and h_{127} is a factor of G. To this end we look at the D_4-system defined by the pair h_7, h_{127}: it is the system that contains the four roots h_{34}, h_{56}, h_{347}, h_{567}. The reflections s (respectively s') perpendicular to h_{367} (respectively h_{467}) lie in this D_4 summand and do not commute. We have

$$s(h_{34}h_{56}h_{347}h_{567}) = h_{467}(-h_{357})h_{64}h_{35},$$
$$s'(h_{34}h_{56}h_{347}h_{567}) = (-h_{367})(-h_{457})h_{63}h_{45},$$

so that

$$h_{34}h_{56}h_{347}h_{567} - h_{467}h_{357}h_{46}h_{35} + h_{367}h_{457}h_{36}h_{45} = 0.$$

The second (respectively third) term is obtained from the first by applying to it minus the transposition (45) (respectively minus the transposition (35)), so that

$$(1-(45)-(35))\, h_7h_{12}h_{127}h_{34}h_{56}h_{347}h_{567} = 0.$$

If we combine this with equation (3) and observe that

$$(1-(45)-(35))\,(1-(34)) = \sum_{w\in\mathcal{S}(3,4,5)} \operatorname{sign}(w)w,$$

then we find that the latter kills $G = h_{127}h_{347}h_{567}h_{246}h_{235}h_{145}h_{136}$.

An instance of the first case, namely the assertion that G is also killed by $\sum_{w\in\mathcal{S}(1,2,7)} \operatorname{sign}(w)wG$, follows by exploiting the symmetry properties of G: the transpositions (34) and (35) have the same effect on G as respectively (12) and (17). This implies that

$$\sum_{w\in\mathcal{S}(1,2,7)} \operatorname{sign}(w)wG = \sum_{w\in\mathcal{S}(3,4,5)} \operatorname{sign}(w)wG = 0.$$

□

Corollary 4.17. *We have* $\dim \mathcal{C}_2 \le 15$.

Proof. Let $f \in \mathcal{C}_2$. Since a monomial of type (B) has a unique factor of the form $(12a)$, $a \in \{3,\dots,7\}$, we can write f accordingly: $f = (124)f_3 + \cdots + (127)f_7$. For every a, we have 6 type-(B) monomials corresponding to the ways we index the splittings of the complement of a in $\{3,4,5,6,7\}$ by $\{1,2,a\}$. The symmetric group $\mathcal{S}(1,2,a)$ permutes these 6 root products simply transitively. These root products satisfy the corresponding alternating sum relation, and so we can arrange that each f_a is a linear combination of 6 monomials whose alternating sum of coefficients is zero. If we take as our guiding idea to make a as small as possible, then it turns out that in half of the cases we can do better.

Let us first assume $a \in \{5,6,7\}$. We then invoke the relation defined by $\{3,4,a\}$:

$$\sum_{w\in\mathcal{S}(3,4,a)} \operatorname{sign}(w)wf_a = 0.$$

Four of the six terms have a factor (123) or (124), whereas the other two have a factor $(12a)$ and combine to $(1-(34))f_a$. So this relation allows us to arrange that f_a and $(34)f_a$ have the same coefficient. We thus make f_a vary in a space of dimension ≤ 3. If $a \in \{6,7\}$, then we can repeat this game with $\{4,5,a\}$. This allows us to assume in addition that f_a and $(45)f_a$ have the same coefficient. But then f_a must have all its coefficients equal. So f_a varies in a space of dimension ≤ 1 for $a = 6,7$, of dimension ≤ 3 for $a = 5$, and of dimension ≤ 5 for $a = 3,4$. This proves that $\dim \mathcal{C}_2 \le 15$. □

Proof (of Proposition 4.12). If we combine Proposition 4.1 and Corollary 4.17, we see that $\mathcal{C}_2$ is an irreducible representation of $W(R)$ of dimension ≤ 15. Since $W(R) = W(R)_+ \times \{1,c\}$, it will then also be an irreducible representation of $W(R)_+$. The symmetric bilinear form on the root lattice induces a nondegenerate form on the root lattice modulo two times the weight lattice (this is an $\mathbb{F}_2$-vector space of dimension 6). This identifies $W(R)_+$ with the symplectic group $\mathrm{Sp}(6,\mathbb{Z}/2)$, and it is well known (see, for instance, [7], where this group is denoted by $S_6(2)$) that the irreducible representations of dimension < 15 are the trivial representation, the sign representation, and the standard representation of degree 7. It is easy to see that $\mathcal{C}_2$ is none of these. Since c acts as -1 in $\mathcal{C}_2$, the proposition follows. □

The roots orthogonal to an A_5-subsystem of an E_6-system make up a system of type A_1 or A_2. In terms of our root basis, they are represented by $\langle h_{23}, h_{34}, h_{45}, h_{56}, h_{67}\rangle$ (with $\langle h_1\rangle$ as the perpendicular system) and the system $\langle h_{123}, h_{34}, h_{45}, h_{56}, h_{67}\rangle$ (with $\langle h_1, h_{12}\rangle$ as perpendicular system). We shall call an A_5-subsystem of the second type *special.* Conversely, the roots perpendicular to an A_2-subsystem form a special A_5-system. Since the A_2-subsystems make up a single Weyl group equivalence class, the same is true for the special A_5-subsystems.

Proposition 4.18. *If we regard $\mathcal{C}_2$ as a vector space of degree 7 polynomials on $\mathfrak{h}_7$, then their common zero set Z_7 is the union of the linear subspaces perpendicular to a system of type D_4 or to a special system of type A_5.*

Proof. The D_4-subsystems constitute a single Weyl group equivalence class, and so we may take as our system the one spanned by the fundamental roots $\langle h_{23}, h_{34}, h_{45}, h_{123}\rangle$. We must show that every $7A_1$-subsystem of R meets this D_4-system. The positive roots of the D_4-system are $\{h_{ij}\}_{2\leq i<j\leq 5}$ and $\{h_{1ij}\}_{2\leq i<j\leq 5}$. It is easy to see from our description that every $7A_1$-system of type (A) contains a root h_{ij} with $2 \leq i < j \leq 5$. Similarly, we see that every $7A_1$-system of type (B) contains a root h_{1ij} with $2 \leq i < j \leq 5$.

We argue for the special A_5-system $\langle h_{123}, h_{34}, h_{45}, h_{56}, h_{67}\rangle$ in a similar fashion. Its positive roots are $\{h_{ij}\}_{3\leq i<j\leq 7}$ and $\{h_{12i}\}_{3\leq i\leq 7}$. Every $7A_1$-system of type (A) contains a root h_{ij} with $3 \leq i < j \leq 7$, and every $7A_1$-system of type (B) contains a root h_{12i} with $3 \leq i \leq 7$.

It remains to show that this exhausts Z_7. Every subsystem of R that does not contain a D_4-subsystem has only components of type A. If in addition it does not contain a special A_5-system, then any such a subsystem is Weyl group-equivalent to a proper subsystem in the (nonsaturated) A_7-system spanned by the fundamental roots $h_{12}, h_{34}, \dots, h_{67}$ and the highest root h_1. The latter has as its positive roots $\{h_{ij}\}_{1\leq i<j\leq 7} \cup \{h_i\}_{i=1}^{7}$ and is therefore disjoint from the $7A_1$-subsystem $\langle h_{123}, h_{145}, h_{167}, h_{256}, h_{247}, h_{357}, h_{346}\rangle$. This implies that Z_7 is as asserted. $\square$

Question 4.19. Is $\mathcal{C}_2$ the space of degree 7 polynomials on $\mathfrak{h}_7$ that vanish on Z_7? We expect this to be equivalent to the question whether $\mathcal{C}_2$ is complete as a linear system on $\mathcal{M}^*_{m,2}$.

5 The Coble linear system

Let A be a vector space of dimension three, so that $\mathbb{P}(A)$ is a projective plane. Given a numbered set $(p_1, \dots, p_N)$ of $N \geq 5$ points in $\mathbb{P}(A)$ that are in generic position, then for any 5-tuple $(i_0, \dots, i_4)$ with $i_0, \dots, i_4$ distinct and taken from $\{1, \dots, N\}$, the four ordered lines $p_{i_0}p_{i_1}$, $p_{i_0}p_{i_2}$, $p_{i_0}p_{i_3}$, $p_{i_0}p_{i_4}$ through p_{i_0} have a cross ratio. The collection of cross ratios thus obtained forms a complete projective invariant of $(p_1, \dots, p_N)$: we may choose coordinates such that $p_1 = [1 : 0 : 0]$, $p_2 = [0 : 1 : 0]$, $p_3 = [0 : 0 : 1]$, $p_4 = [1 : 1 : 1]$ and the coordinates $[z_0 : z_1 : z_2]$ for p_i, $i > 4$, are then given by cross ratios. For instance, $z_1 : z_2 = (p_1p_2 : p_1p_3 : p_1p_4 : p_1p_i)$. If $a_i \in A$ represents p_i, then we can write this as a cross ratio of four lines in the plane $a_1 \wedge A \subset \wedge^2 A$: $(a_1 \wedge a_2 : a_1 \wedge a_3 : a_1 \wedge a_4 : a_1 \wedge a_i)$.

Now let us observe that if (v_1, v_2, v_3, v_4) is a generic ordered 4-tuple in a vector space T of dimension two, then the corresponding points in $\mathbb{P}(T)$ have a cross ratio that can be written as a ratio of two elements of $\det(T)^2$, namely $(v_1 \wedge v_4)(v_2 \wedge v_3) : (v_2 \wedge v_4)(v_1 \wedge v_3)$. If we apply this to to the present case, then we get

$$z_1 : z_2 = (p_1p_2 : p_1p_3 : p_1p_4 : p_1p_i) = |12i||134| : |13i||124|,$$

where we used the Coble notation. Thus the cross ratios formed in this manner allow us to reconstruct $(p_1, \dots, p_5)$ up to projective equivalence. We can express this in terms of roots as follows.

Lemma 5.1. *Let $(S; e_1, \dots, e_{9-d})$ be a marked Del Pezzo surface of degree $d \leq 4$, $S \to \check{\mathbb{P}}^2(H^0(S, \ell))$ the contraction morphism defined by the linear system $|\ell|$ (as usual), and p_i the image of E_i. Then $p_1, \dots, p_4$ are in general position if and only if none of the roots in the A_4-subsystem generated by $(h_{123}, h_{12}, h_{23}, h_{34})$ is effective in* $\mathrm{Pic}(S)$.

If that is the case and K is a cuspidal anticanonical curve on S, then for $i > 4$ the cross ratio $(p_1p_2 : p_1p_3 : p_1p_4 : p_1p_i)$ equals the ratio of the two elements of the line $\mathrm{Pic}(K)^o$ *given by $r_K(h_{2i}h_{12i}h_{34}h_{134})$ and $r_K(h_{3i}h_{13i}h_{24}h_{124})$.*

Proof. The first part is left to the reader as an exercise. As to the second part, choose affine coordinates (x, y) in $\check{\mathbb{P}}^2(H^0(S, \ell))$ such that K is given by $y^3 = x^2$. So $p_i = \left(t_i, t_i^3\right)$ for some t_i. For $i \neq 1$, the line p_1p_i has tangent $\left[t_i - t_1 : t_i^3 - t_1^3\right] = \left[1 : t_i^2 + t_it_1 + t_1^2\right]$. So the cross ratio of the lines p_1p_i involves factors of the form

$$\left(t_j^2 + t_jt_1 + t_1^2\right) - \left(t_i^2 + t_it_1 + t_1^2\right) = (t_j - t_i)(t_j + t_i + t_1), \quad 2 \leq i < j \leq 5.$$

If we use the x-coordinate to identify $\mathrm{Pic}_0(K)$ with $\mathbb{C}$, then such a factor can be written $r_K(h_{ij}h_{1ij})$. The last assertion follows. □

Remark 5.2. Notice that the roots that appear in the numerator (respectively the denominator) of $h_{25}h_{12i}h_{34}h_{134} : h_{3i}h_{13i}h_{24}h_{124}$ are four pairwise perpendicular roots all of which lie in a single D_4-subsystem.

The Coble system in the degree four case

We first consider a Fano surface of degree 5. We recall that such a surface S can be obtained by blowing up four points of a projective plane in general position, and so is unique up to isomorphism. Any automorphism of this surface that acts trivially on its Picard group preserves every exceptional curve (= line) and hence is the identity. It follows that the automorphism group of S is the Weyl group $W(A_4)$. There are 10 lines on S. If five of them make up a pentagon, then their sum is an anticanonical divisor. There are 12 such pentagons and they generate the anticanonical system. Let us now fix a marking (e_1, e_2, e_3, e_4) for S. To every $p \in S$ we associate a marked Del Pezzo surface $(S_p; e_1, \ldots, e_5)$ of degree 4 by letting S_p be the blowup of S in p and letting e_5 be the class of the exceptional divisor. This defines a rational map $S \dashrightarrow \mathcal{M}^*_{m,4}$. We will see that this is in fact an isomorphism.

Proposition 5.3. *The* 12 *Coble covariants for* D_5 *span a complete linear system of dimension* 6 *and define an embedding of* $\mathcal{M}^*_{m,4}$ *in a projective space of dimension* 5*. The image is "the" anticanonically embedded Fano surface of degree* 5 *(so that the Coble system is anticanonical) with* $\mathcal{M}^*_{m,4} - \mathcal{M}^\circ_{m,4}$ *mapping onto the union of its ten lines. The divisor of every Coble covariant is a pentagon on this Fano surface, and every pentagon thus occurs.*

Proof. Let A be a complex vector space of dimension 3 and let $p_1, \ldots, p_5 \in \mathbb{P}(A)$. We first assume that $p_1, \ldots, p_4$ are in general position (i.e., no three collinear). We then adapt our coordinate system accordingly: $p_i = [a_i]$, with $a_1 = (1,0,0)$, $a_2 = (0,1,0)$, $a_3 = (0,0,1)$, and $a_4 = (1,1,1)$. If $a_5 = (z_0, z_1, z_2)$, then typical determinants involving a_5 are

$$|125| = z_2, \quad |145| = z_2 - z_1.$$

So for instance,

$$|415||152||523||234||341| = (z_1 - z_2).(-z_2).z_0.1. - 1 = z_0z_1z_2 - z_1^2z_2.$$

We obtain in this manner the six polynomials $z_0z_1z_2 - z_i^2z_j$, $i \neq j$, and it is easily verified that any other Coble covariant is a linear combination of these. They are visibly linearly independent and hence form a basis for $\mathcal{C}_5$. It is precisely the linear system of cubic curves that pass through $p_1, \ldots, p_4$. So the Coble system is anticanonical and defines an embedding of the blowup S of $\mathbb{P}(A)$ in $p_1, \ldots, p_4$ to $\mathbb{P}^5$. The remaining assertions are verified in a straightforward manner. $\square$

Remark 5.4. This proposition and its proof show that the moduli space $\mathcal{M}^*_{m,4}$ is as a variety simply obtained from $\mathbb{P}^2$ by blowing up the vertices $p_1, \ldots, p_4$ of the coordinate simplex. This argument then also shows that there is universal semistable marked Del Pezzo surface of degree 4, $\mathcal{S}_{m,4} \to \mathcal{M}^*_{m,4}$: over $p_5 \in \mathcal{M}^*_{m,4}$ we put the blowup of the surface $\mathcal{M}^*_{m,4}$ in p_5 so that $\mathcal{S}_{m,4}$ is simply $\mathcal{M}^*_{m,4} \times \mathcal{M}^*_{m,4}$ blown up along the diagonal with one of the projections serving as the structural morphism.

The Coble system in the degree three case

Our discussion starts off with the following lemma.

Lemma 5.5. *The Coble system* $\mathcal{C}_3$ *has no base points.*

Proof. The Coble linear system pulled back to $\mathbb{P}(\mathfrak{h}_6)$ has according to Proposition 4.10 as its base-point locus the projective arrangement $\mathbb{P}(Z_6)$, the union of the fixed point hyperplanes of Weyl subgroups of type A_3. Since $\mathcal{M}_3^*$ is a quotient of $\mathbb{P}(\mathfrak{h}_6) - \mathbb{P}(Z_6)$, it follows that $\mathcal{C}_3$ has no base points and hence defines a morphism to a $\mathbb{P}^9$. □

We use what we shall call the *Naruki model* of $\mathcal{M}_{m,3}^*$. This is based on a particular way of getting a degree 3 Fano surface as a blown-up projective plane: we suppose the points in question to be labeled p_i, q_i with $i \in \mathbb{Z}/3$ and to lie on the coordinate lines of $\mathbb{P}^2$ as follows:

$$p_0 = [0:1:a_0],\, p_1 = [a_1:0:1],\, p_2 = [1:a_2:0],$$
$$q_0 = [0:1:b_0],\ q_1 = [b_1:0:1],\ q_2 = [1:b_2:0].$$

For the moment we assume that blowing up these points gives rise to a Fano surface S, so that in particular, none of the a_i, b_i is zero and $a_i \neq b_i$. If we blow up these points, the strict transform of the coordinate triangle is an anti-canonical curve K (it is a tritangent of the corresponding cubic surface). This "partial rigidification" reduces the projective linear group $\mathrm{PGL}(3,\mathbb{C})$ to its maximal subtorus that leaves the coordinate triangle invariant. The following expressions are invariant under that torus

$$\alpha_i := a_i/b_i\, (i \in \mathbb{Z}/3), \quad \delta := -b_0b_1b_2,$$

and together they form a complete projective invariant of the configuration. Notice the formulas

$$a_0b_1b_2 = -\alpha_0\delta, \quad a_0a_1b_2 = -\alpha_0\alpha_1\delta, \quad a_0a_1a_2 = -\alpha_0\alpha_1\alpha_2\delta.$$

As explained in the appendix of [16], α_i and δ have a simple interpretation in terms of the of pair (S,K): if we denote the exceptional curves by A_i, B_i, and E is the strict transform of the line through q_1 and q_2, then the cycles $A_i - B_i$, $i \in \mathbb{Z}/3$, and $B_0 - E$ span in $\mathrm{Pic}(S)$ the orthogonal complement to the components of K, and the numbers in question can be interpreted as their images in $\mathrm{Pic}^0(K) \cong \mathbb{C}^\times$. The classes themselves make up the basis of a D_4 root system with the last one representing the central node. We denote the 4-torus for which $\alpha_0, \alpha_1, \alpha_2, \delta$ is a basis of characters by T. This torus comes with an action of $W(D_4)$, and this makes it an adjoint torus of type D_4. We denote by T° the open set of its regular elements. This is the complement of the union of reflection hypertori, i.e., the locus where none of the D_4-roots is 1. It has the interpretation as the moduli space of marked nonsingular cubic surfaces with the property that a particular tritangent (which is entirely given by the marking) does not have its three lines collinear. If we want to include that case too, we must first blow up the identity element of T, $\mathrm{Bl}_1(T) \to T$, and then remove the strict transforms of the reflection hypertori. This open subset, $\mathrm{Bl}_1(T)^\circ \subset \mathrm{Bl}_1(T)$, is a model for the moduli space of marked nonsingular cubic surfaces; in other words, it can be identified with $\mathcal{M}_{m,3}^\circ$. The modular interpretation implies that this variety has a $W(E_6)$-action, although only the action of a Weyl subgroup of type D_4 is manifest. The action of the two missing fundamental reflections was written down by Naruki and Sekiguchi: the one that in the Dynkin diagram is attached to α_i is given by

$$\alpha_i \mapsto -\alpha_0\alpha_1\alpha_2\delta^2 \frac{1-\alpha_i}{1-\alpha_0\alpha_1\alpha_2\delta^2},$$

$$\alpha_{i\pm 1} \mapsto \frac{(1-\alpha_0\alpha_{i\pm 1}\delta)(1-\alpha_0\alpha_1\alpha_2\delta)}{(1-\alpha_0\delta)(\alpha_{i\pm 1}-\alpha_0\alpha_1\alpha_2\delta)},$$

$$\delta \mapsto \delta^{-1}\frac{(1-\alpha_i\delta)(1-\alpha_0\alpha_1\alpha_2\delta^2)}{(1-\alpha_i)(1-\alpha_0\alpha_1\alpha_2\delta)}.$$

(We give this formula only because of its remarkable form—we shall not use it.) Subsequently, Naruki [16] found a nice $W(E_6)$-equivariant smooth projective completion of this space with a normal crossing divisor as boundary. What is more relevant here is a (projective) blowdown of his completion that was also introduced by him. We shall take Naruki's construction of the latter as our guide and re-prove some of his results in the process.

We identify the complexification of $\mathrm{Hom}(\mathbb{C}^\times, T)$ with the Lie algebra $\mathfrak{t}$ of T, so that the latter has a natural $\mathbb{Q}$-structure. The decomposition Σ of $\mathfrak{t}(\mathbb{R})$ into its $W(D_4)$-chambers has its rays spanned by the coweights that lie in the $W(D_4)$-orbit of a fundamental coweight. One of these is the orbit of coroots and has 24 elements; the other three consist of minuscule weights and are a single orbit under the full automorphism group of the D_4-system; it also has 24 elements. If we remove the faces that contain a coroot we obtain a coarser decomposition of $\mathfrak{t}(\mathbb{R})$ that we denote by Σ'; a maximal face of Σ' is now an orbit of a Weyl chamber under the stabilizer of a coroot (a type $3A_1$-Weyl group).

Let $T \subset T_\Sigma$ be the associated torus embedding. It is smooth with normal crossing boundary. The boundary divisors are in bijective correspondence with the above coweights. We shall refer to those that correspond to coroots (respectively minuscule weights) as *toric coroot divisors* (respectively *toric minuscule weight divisors*). So there are 24 of each.

Now blow up successively in T_Σ: the identity element (in other words, the fixed point set of $W(D_4)$), the fixed-point sets of the Weyl subgroups of type A_3, the fixed-point sets of the Weyl subgroups of type A_2. We denote the resulting blowup by $\hat{T}_\Sigma$. In this blowup the exceptional divisors of type A_3 have been separated, and each is naturally a product. To be precise, a $W(A_3)$-Weyl subgroup $G \subset W(D_4)$ has as its fixed-point locus in $\hat{T}_\Sigma$ a copy of a $\mathbb{P}^1$ whose tangent line at the identity is the G-fixed point set in $\mathfrak{t}$, and the divisor associated to G is then naturally the product of the projective line $\hat{T}_\Sigma^G$ and the projective plane $\mathbb{P}(\mathfrak{t}/\mathfrak{t}^G)$ blown up in the fixed points of the A_2-Weyl subgroups of G (there are four such and they are in general position).

The Coble system on Naruki's completion, together with the $W(E_6)$-action on it, was identified in [14, (5.9 and 4.5)]. It is the pullback of

$$H^0\left(T_\Sigma, \mathcal{O}(D_S + 2D_R) \otimes \mathfrak{m}_e^3\right)$$

to $\hat{T}_\Sigma$; here D_S, D_R are the sums of the 24 divisors corresponding to the rays spanned by the minuscule weight and the coroots respectively, and $\mathfrak{m}_e$ is the ideal sheaf of the identify element $e \in T_\Sigma$.

We now give an explicit description of the Coble covariants in terms of $(\alpha_0, \alpha_1, \alpha_2, \delta)$. For this we begin with observing the following simple identities:

$$|p_0p_1p_2| = a_0a_1a_2 + 1, \quad |p_iq_ip_{i+1}| = (b_i - a_i)a_{i+1}, \quad |p_iq_ip_{i-1}| = b_i - a_i.$$

A straightforward computation yields

$$|p_0p_1p_2q_0q_1q_2| = \pm(b_0 - a_0)(b_1 - a_1)(b_2 - a_2)(1 - a_0a_1a_2b_0b_1b_2).$$

We substitute these values in the formulas for the Coble covariants, but for reasons similar to those in Section 3, we divide these by $(b_0 - a_0)(b_1 - a_1)(b_2 - a_2)$. For example,

$$\begin{aligned}
&|p_0p_1p_2q_0q_1q_2|.|p_0q_0p_1|.|q_1p_2q_2| \\
&\qquad\qquad = \pm(1 - a_0a_1a_2b_0b_1b_2).(b_0 - a_0)a_1.(b_2 - a_2) \\
&\qquad\qquad = \pm\alpha_1\delta(1 - \alpha_0\alpha_1\alpha_2\delta)(1 - \alpha_0)(1 - \alpha_2)
\end{aligned}$$

and

$$\begin{aligned}
&|p_0q_0p_1|.|p_0q_0q_2|.|p_1q_1q_2|.|p_1p_2q_2|.|p_0q_1p_2|.|q_0q_1p_2| \\
&\quad = \pm\frac{(b_0 - a_0)a_1.(b_0 - a_0).(b_1 - a_1)b_2.(b_2 - a_2)(a_0b_1a_2 + 1)(b_0b_1a_2 + 1)}{(b_0 - a_0)(b_1 - a_1)(b_2 - a_2)} \\
&\quad = \pm\alpha_0\alpha_1\delta(1 - \alpha_0)(1 - \alpha_0\alpha_2\delta)(1 - \alpha_2\delta).
\end{aligned}$$

We thus find for the Coble covariants (40 up to sign) the following expressions:

1	$\delta(1 - \alpha_0)(1 - \alpha_1)(1 - \alpha_2)$
2	$\alpha_0\alpha_1\alpha_2\delta^2(1 - \alpha_0)(1 - \alpha_1)(1 - \alpha_2)$
3	$(1 - \alpha_0\delta)(1 - \alpha_1\delta)(1 - \alpha_2\delta)$
4	$\alpha_0\alpha_1\alpha_2\delta(1 - \alpha_0\delta)(1 - \alpha_1\delta)(1 - \alpha_2\delta)$
5	$(1 - \alpha_0\alpha_1\delta)(1 - \alpha_1\alpha_2\delta)(1 - \alpha_2\alpha_0\delta)$
6	$\delta(1 - \alpha_0\alpha_1\delta)(1 - \alpha_1\alpha_2\delta)(1 - \alpha_2\alpha_0\delta)$
7_i	$(1 - \alpha_{i-1}\delta)(1 - \alpha_{i+1}\delta)(1 - \alpha_0\alpha_1\alpha_2\delta)$
8_i	$\alpha_i\delta(1 - \alpha_{i-1})(1 - \alpha_{i+1})(1 - \alpha_0\alpha_1\alpha_2\delta^2)$
9_i	$\delta(1 - \alpha_{i-1})(1 - \alpha_{i+1})(1 - \alpha_0\alpha_1\alpha_2\delta^2)$
10_i	$\alpha_{i-1}\alpha_{i+1}\delta(1 - \delta)(1 - \alpha_i\alpha_{i-1}\delta)(1 - \alpha_i\alpha_{i+1}\delta)$

11_i	$(1-\delta)(1-\alpha_i\alpha_{i+1}\delta)(1-\alpha_i\alpha_{i-1}\delta)$
12_i	$\alpha_i\delta(1-\alpha_{i+1}\delta)(1-\alpha_{i-1}\delta)(1-\alpha_0\alpha_1\alpha_2\delta)$
13	$(1-\delta)(1-\alpha_0\alpha_1\alpha_2\delta)(1-\alpha_0\alpha_1\alpha_2\delta^2)$
14_i	$\alpha_{i-1}\alpha_{i+1}\delta(1-\delta)(1-\alpha_i)(1-\alpha_i\delta)$
15_i	$(1-\alpha_i\delta)(1-\alpha_{i-1}\alpha_{i+1}\delta)(1-\alpha_0\alpha_1\alpha_2\delta^2)$
16_i	$\delta(1-\alpha_i)(1-\alpha_{i-1}\alpha_{i+1}\delta)(1-\alpha_0\alpha_1\alpha_2\delta)$
17_{ijk}	$\alpha_i\delta(1-\alpha_j)(1-\alpha_k\delta)(1-\alpha_j\alpha_k\delta)$

Notice that the zero set in T of any of these expressions is the union of the reflection hypertori of a subsystem of type $3A_1$ (in the first 12 cases) or the A_2 (for the last 5). We now quote from [16, Proposition 11.3]:

Theorem 5.6 (Naruki). *There is a projective contraction*

$$\hat{T}_\Sigma \to \check{T}_\Sigma$$

that contracts each A_3-divisor along the projection on its 2-dimensional factor. The contracted variety $\check{T}_\Sigma$ is nonsingular, and the action of $W(E_6)$ on $(\mathrm{Bl}_e T)^\circ$ extends regularly to it. This action is transitive on the collection of 40 divisors that are of toric coroot type or of A_2-type.

The 40 divisors in question are easily seen to be pairwise disjoint. Naruki also shows (Section 12 of [16]) that each of these can be contracted to a point. We can see this quickly using the theory of torus embeddings: the 24 toric coroot divisors get contracted if we replace in the above discussion the decomposition Σ of $\mathfrak{t}(\mathbb{R})$ by the coarser one, Σ', that we obtain by removing the faces that contain a coroot. The $W(E_6)$-action and the theorem above imply that this contraction is then also possible for the remaining 16 divisors. The singularities thus created, for instance the one defined by the ray spanned by the coroot $\delta^\vee$, can be understood as follows: the natural affine T-invariant neighborhood of the point of $T_{\Sigma'}$ defined by the ray spanned by the coroot $\delta^\vee$ is Spec of the algebra generated by the elements of the orbit of δ under the Weyl group of the $3A_1$-subsystem $\langle\alpha_0,\alpha_1,\alpha_2\rangle$: $\mathrm{Spec}\mathbb{C}[\delta,\alpha_0\delta,,\alpha_1\delta,\dots,\alpha_0\alpha_1\alpha_2\delta]$. This is a cone over the Veronese embedding of $(\mathbb{P}^1)^3$.

The following more precise result is in [14] (Theorem 5.7) and follows also from [19].

Theorem 5.7. *The Coble covariants generate on $\check{T}_\Sigma$ a linear system without base points that has the property that its restriction to each of the 40 divisors of toric coroot type or of A_2-type is trivial. The resulting morphism to a nine-dimensional projective space realizes Naruki's contraction.*

Proof (Outline of proof). We have a natural decomposition of T_Σ into strata by type: D_4 (yielding the identity element); A_3, A_2, and $\{1\}$ (being open in T_Σ). There is a corresponding decomposition of $\hat{T}_\Sigma$ (and of $\check{T}_\Sigma$, but we find it more convenient to work on the former), albeit those strata are then indexed by chains of strata in T_Σ that are totally ordered for incidence. We first check that along every stratum the Coble covariants define, modulo the stated contractions property, an embedding. For this, the $W(D_4)$-equivariance allows us to concentrate on the open subset $U = \mathrm{Spec}(\mathbb{C}[\alpha_0, \alpha_1, \alpha_2, \delta])$ of T_Σ and its preimage $\hat{U}$ in $\hat{T}_\Sigma$.

It is clear from the expressions we found that the Coble covariants generate $\mathbb{C}[\delta, \alpha_0\delta, \dots, \alpha_0\alpha_1\alpha_2\delta]$ after we localize away from the kernels of the roots (that is, we make for each root alpha, the expression $\alpha - 1$ invertible). So we have an embedding on the corresponding open subset of $\check{T}_\Sigma$ (this contains the singular point defined by the coroot $\delta^\vee$). A closer look at the equations shows that this is in fact true even if we allow some of the roots to be 1, provided that they are mutually perpendicular. In other words, the linear system defines an embedding on the intersection of the open stratum with U.

Now let Z be the A_3-stratum that is open in $\alpha_0 = \alpha_1 = \delta = 1$. All Coble covariants vanish on Z, and we readily verify that Coble covariants generate the ideal defining Z. Thus the system has no base points on the blowup of Z, and the system contracts this exceptional divisor along the Z-direction.

Now let us look at an A_2-stratum Z', say the one that is open in $\alpha_0 = \delta = 1$. We observe that the restriction of every covariant to Z' is proportional to $(1 - \alpha_1)(1 - \alpha_2)(1 - \alpha_1\alpha_2)$ (and can be nonzero). So the linear system will define the constant map on Z' (or the exceptional divisor over Z').

We turn to the situation at the identity of T. Every covariant vanishes there to order three and has for initial part a product of three roots, viewed as linear forms on $\mathfrak{t}$. Up to sign, the roots in such a product are the positive roots of a $3A_1$-subsystem (in the first 12 cases) or of an A_2-system (the last 5 cases). The preceding implies that the linear system restricted to an A_2-line in $\mathbb{P}(\mathfrak{t})$ is constant. With some work we find that the Coble linear system is without base points and generates the ideal defining the identity away from the union of the A_2-loci.

The remaining strata on $\hat{T}_\Sigma$ are defined by "flags": chains of strata above totally ordered by incidence, with $\{1\} < Z' < Z$ as a typical degenerate case. In that situation, one checks that the covariants generate on that stratum the ideal $\mathcal{I}_{\{e\}}\mathcal{I}_{Z'}\mathcal{I}_Z$. We thus see that we have a local embedding along this stratum. The other strata are dealt with in the same way.

This shows that the linear system defines local embeddings modulo the contraction property. So $\check{T}_\Sigma$ is defined as a projective quotient of $\hat{T}_\Sigma$, and the linear system maps it as a local embedding to a nine-dimensional projective space. It remains to see that the images of the strata are disjoint. This is left to the reader. □

Remark 5.8. The construction of the Naruki quotient comes with a stratification, and as one may expect, each of its members has a modular interpretation. We here give that interpretation without proof. As mentioned, the open stratum T° is the moduli space of systems $(S; e_1, \dots, e_6; K)$ with $(S; e_1, \dots, e_6)$ a marked Fano surface of degree 3 (equivalently, a cubic surface) and K an anticanonical divisor made up of three nonconcurrent exceptional curves (so that the isomorphism $S \to \bar{S}$ maps K onto a tritangent $\bar{K} \subset \bar{S}$). Suppose now that S is merely a Del Pezzo surface whose configuration of (-2)-curves is nonempty, but disjoint from K. Then that configuration is of type rA_1, and we are on a stratum contained in T of type rA_1 $(1 \le r \le 4)$ or it is of type $rA_1 + sA_1$ with $r \ge 1$ and we are on one of the 24 points that are images of toric coroot divisors (these are the punctual strata of $T_{\Sigma'}$). In these cases $\bar{K}$ is a genuine tritangent of $\bar{S}$ (which lies in the smooth part of $\bar{S}$). The other strata are loci for which $\bar{K}$ is no longer a tritangent: if $\bar{K}$ defines an Eckardt point (so that K consists of three distinct concurrent exceptional curves), then we find ourselves in the stratum that is open in the preimage of the unit element of T. If $\bar{K}$ becomes a union of a double line and another line, then it contains two distinct A_1-singularities of $\bar{S}$, and K is of the form $2E + E' + C + C'$, where E, E' are exceptional curves and C, C' are (-2)-curves with E', C, C' pairwise disjoint and meeting E normally. We are then on a stratum that is open in the image of an A_3-locus in $\check{T}$. If $\bar{K}$ becomes a triple line, then it contains two distinct A_2-singularities of $\bar{S}$ and K is of the form $3E + C + C'$, where E is an exceptional curve and and C, C' are disjoint A_2-curves meeting E normally. We are then representing one of the 16 punctual strata that are images of an A_2-locus in T.

Corollary 5.9. *The GIT completion of the moduli space of marked cubic surfaces, $\mathcal{M}^*_{m,3}$, is $W(E_6)$-equivariantly isomorphic to the Naruki contraction of $\hat{T}_\Sigma$. The Coble linear system embeds $\mathcal{M}^*_{m,3}$ in projective nine-space.*

Proof. The Coble linear system on $\hat{T}_\Sigma$ is without base points, and so the resulting morphism $f : \hat{T}_\Sigma \to \mathbb{P}^9$ realizes the Naruki contraction. Recall that we have an identification of $\mathcal{M}^\circ_{m,3}$ with $(\mathrm{Bl}_e T)^\circ$. This isomorphism clearly extends to a morphism $\mathcal{M}^*_{m,3} \to f(\hat{T}_\Sigma)$. This morphism is birational, and since $f(\hat{T}_\Sigma)$ is normal, it must be a contraction. □

Remark 5.10. It is known [2] that the moduli space of stable cubic surfaces is Galois covered by the complex 4-ball with an arithmetic group Γ as Galois group. The group Γ has a single cusp, and this cusp represents the minimal strictly stable orbit of cubic surfaces (i.e., those having three A_2-singularities). This gives $\mathcal{M}_3$ the structure of an arithmetic ball quotient for which $\mathcal{M}^*_3$ is its Baily–Borel compactification. Allcock and Freitag [1] have used Γ-modular forms to construct an embedding of this Baily–Borel compactification in a 9-dimensional projective space. This is precisely the embedding that appears here (see also Freitag [13] and van Geemen [14]). Via this interpretation it also follows that the Coble system is complete [12].

Remark 5.11. The linear system $\mathcal{C}_3$ can also be studied by restricting it, as was done in [6], to the exceptional divisor $\mathbb{P}(\mathfrak{h}_5)$ of the blowup of the e_6-point in $\mathbb{P}(\mathfrak{h}_6)$. The generic point of $\mathbb{P}(\mathfrak{h}_5)$ has a modular interpretation: it parameterizes marked cubic surfaces with a point where the tangent space meets the surface in the union of a conic and a line tangent to that conic. The marking determines the line, but not the conic, for the system of conics on a cubic surface that lie in plane that contains a given line on that surface has two members that are tangent to the line. So we have a natural involution ι on that space. The projective space $\mathbb{P}(\mathfrak{h}_5)$ can be seen as the projective span of the D_5-subsystem R_5, spanned by the roots not involving e_6.

In order to be explicit we also use the standard model for the D_5 root system, i.e., the model for which $\epsilon_i - \epsilon_{i+1} = h_{i,i+1}$ $(i = 1, \ldots, 4)$ and $\epsilon_4 + \epsilon_5 = h_{123}$. This makes $W(R_5)$ the semidirect product of the group of permutations of the basis elements $\epsilon_1, \ldots, \epsilon_5$ and the group of sign changes in the basis elements $(\epsilon_1, \ldots, \epsilon_5) \mapsto (\pm\epsilon_1, \ldots, \pm\epsilon_5)$ with an even number of minus signs. We denote the basis dual to $(\epsilon_1, \ldots, \epsilon_5)$ by $(x_1, \ldots, x_5)$. The Coble covariant we attached to the $3A_2$-system $\langle h_{ij}, h_{jk}, h_{lm}, h_{m6}, h_{ijk}, h\rangle$ gives, after dividing by a common degree 4 factor, the quintic form on $\mathfrak{h}_5$ defined by

$$h_{ij}h_{jk}h_{ik}h_{lm}h_{ijk} = (x_i - x_j)(x_j - x_k)(x_i - x_k)\left(x_l^2 - x_m^2\right),$$

whereas the Coble covariant attached to $\langle h_{ij}, h_{ilm}, h_{lm}, h_{km6}, h_{k6}, h_{ijk}\rangle$ gives

$$h_{ij}h_{jlm}h_{ilm}h_{lm}h_{ijk} = (x_i - x_j)(x_j + x_k)(x_i + x_k)\left(x_l^2 - x_m^2\right).$$

These all lie in a single $W(R_5)$-orbit as predicted by Lemma 4.7. It can be easily checked that the base locus of the system on $\mathbb{P}(\mathfrak{h}_5)$ is the set of points fixed by a Weyl subgroup of type A_3. There are two orbits of subroot systems of type A_3: one has 40 elements and is represented by $\langle \epsilon_1 - \epsilon_2, \epsilon_2 - \epsilon_3, \epsilon_3 - \epsilon_4\rangle$, and the other has 10 elements and is represented by $\langle \epsilon_1 - \epsilon_2, \epsilon_2 - \epsilon_3, \epsilon_2 + \epsilon_3\rangle$. So the base locus is a union of 50 lines. The locus where two such lines meet is the (16) fixed points of a Weyl subgroup of type A_4 and the (5) fixed points of a Weyl subgroup of type D_4. Blowing up first the 21 points and then the strict transforms of the 50 lines, we obtain a smooth fourfold $\widetilde{\mathbb{P}(\mathfrak{h}_5)}$ in which the strict transforms of the planes defined by the (40) root subsystems of type A_2 have become disjoint. The Coble system defines a morphism

$$\Psi : \widetilde{\mathbb{P}(\mathfrak{h}_5)} \longrightarrow \mathbb{P}^9,$$

which is generically two to one: it identifies the orbits of the involution ι, which on $\mathbb{P}(\mathfrak{h}_5)$ is given as the rational map

$$[x_1 : \cdots : x_5] \longmapsto \left[x_1^{-1} : \cdots : x_5^{-1}\right].$$

The morphism Ψ is ramified along the exceptional divisors over the A_4-points and contracts the exceptional divisors over the A_3-lines to planes and the 40 planes of type A_2 to points.

The Coble system in the degree two case

An analogue of Lemma 5.5 holds:

Proposition 5.12. *The linear system $\mathcal{C}_2$ is without base points on $\mathcal{M}^*_{m,2}$. Its restriction to $\mathcal{M}^\circ_{m,2}$ is an embedding.*

Proof. The proof of the first assertion differs from that of Lemma 5.12 essentially only by replacing the reference to Proposition 4.10 by a reference to Proposition 4.18: the pullback of $\mathcal{C}_2$ to $\mathbb{P}(\mathfrak{h}_7)$ has, according to Proposition 4.18, as base locus $\mathbb{P}(Z_7)$ the projective arrangement defined by the subsystems of type D_4 and the special subsystems of type A_5. Since the map from $\mathbb{P}(\mathfrak{h}_7) - \mathbb{P}(Z_7)$ to $\mathcal{M}^*_2$ is a surjective morphism, $\mathcal{C}_2$ is without base points.

The second assertion follows from Lemma 4.13: if we are given a D_4-subsystem of a root system of type E_6, then orthogonal to it we have a $3A_1$-system. Thus two disjoint $4A_1$-subsystems of the given D_4-system have discriminants whose quotient is a quotient of Coble covariants. According to Lemma 5.1 the quotient of two such $4A_1$-subsystems is a cross ratio. Hence all (generalized) cross ratios are recovered from the Coble covariants. If the seven points are in general position, then these cross ratios determine the points up to projective equivalence. $\square$

Recall from Proposition 1.4 that we identified $\mathcal{C}_2$ with a space of sections of a square root of $\mathcal{O}_{\mathcal{M}^*_2}(1)$.

Conjecture 5.13. *The linear system $\mathcal{C}_2$ is without base points and hence defines an injective morphism from the moduli space $\mathcal{M}^*_2$ of semistable quartic curves with level-two structure to a* 14*-dimensional projective space.*

We also expect that there is an analogue of the results of Naruki and Yoshida with the role of D_4 taken by E_6.

References

1. D. Allcock, E. Freitag, *Cubic surfaces and Borcherds products*, Comment. Math. Helv. **77** (2002), 270–296.
2. D. Allcock, J.A. Carlson, D. Toledo, *The complex hyperbolic geometry of the moduli space of cubic surfaces*, J. Algebraic Geom. **11** (2002), 659–724.
3. N. Bourbaki, *Groupes et algèbres de Lie*, Ch. 4,5 et 6, 2ième éd., Masson, Paris 1981.
4. A.B. Coble, *Algebraic Geometry and Theta Functions*, AMS New York, 1929.
5. E. Colombo, B. van Geemen, *The Chow group of the moduli space of marked cubic surfaces,* Ann. Mat. Pura Appl. (4) **183** (2004), 291–316.
6. E. Colombo, B. van Geemen, *A family of marked cubic surfaces and the root system D_5,* Internat. J. Math. **18** (2007), no. 5, 505–525.
7. J.H. Conway, R.T. Curtis, S.P. Norton, R.A. Parker, R.A. Wilson, *Atlas of finite groups*, Oxford University Press, Eynsham, 1985.
8. F.R. Cossec, I.V. Dolgachev, *Enriques Surfaces I*, Birkhäuser, PM **76** (1989).
9. M. Demazure, *Surfaces de Del Pezzo*, in Séminaire sur les Singularités des Surfaces, Lecture Notes in Math. **777**, Springer, Berlin, New York, 1980.

10. I. Dolgachev, D. Ortland, *Point sets in projective spaces and theta functions*, Astérisque **165** (1988).
11. I. Dolgachev, B. van Geemen, S. Kondō, *A complex ball uniformization of the moduli space of cubic surfaces via periods of $K3$ surfaces*, J. Reine Angew. Math. **588** (2005), 99–148.
12. E. Freitag, *A graded algebra related to cubic surfaces*, Kyushu J. Math. **56** (2002), no. 2, 299–312.
13. E. Freitag, *Comparison of different models of the moduli space of marked cubic surfaces*, Proc. Japanese-German seminar Riushi-do, T. Ibukyama, W. Kohnen eds., 74–79 (2002), also available from `www.rzuser.uni-heidelberg.de/~t91/index4.html`.
14. B. van Geemen, *A linear system on Naruki's moduli space of marked cubic surfaces,* Internat. J. Math. **13** (2002), 183–208.
15. B. Hunt, *The geometry of some special arithmetic quotients*. Lecture Notes in Mathematics **1637**, Springer-Verlag, Berlin, 1996.
16. I. Naruki, *Cross ratio variety as a moduli space of cubic surfaces*, Proc. London Math. Soc. **45** (1982), 1–30.
17. I.G. Macdonald, *Some irreducible representations of Weyl groups*, Bull. London Math. Soc. **4** (1972), 148–150.
18. Y. Manin, *Cubic Forms*, 2^{nd} edition, North Holland Math. Library 4, North Holland, 1986.
19. M. Yoshida, *A $W(E_6)$-equivariant projective embedding of the moduli space of cubic surfaces*, available at `arXiv:math.AG/0002102`.
20. C.T.C. Wall, *Geometric invariant theory of linear systems*, Math. Proc. Cambridge Philos. Soc. **93** (1983), 57–62.
21. C.T.C. Wall, *Nets of quadrics, and theta-characteristics of singular curves*, Philos. Trans. Roy. Soc. London Ser. A **289** (1978), no. 1357, 229–269.

The Weil Proof and the Geometry of the Adèles Class Space

Alain Connes,[1] Caterina Consani,[2] and Matilde Marcolli[3]

[1] Collège de France, 3, rue d'Ulm, Paris, F-75005 France
alain@connes.org
[2] Mathematics Department, Johns Hopkins University, Baltimore, MD 21218 USA
kc@math.jhu.edu
[3] Department of Mathematics, California Institute of Technology 1200 E California Blvd, Pasadena, CA 91101, USA
matilde@caltech.edu

Dedicated to Yuri Manin on the occasion of his 70th birthday

O simili o dissimili che sieno questi mondi
non con minor raggione sarebe bene a l'uno
l'essere che a l'altro

Giordano Bruno – *De l'infinito, universo e mondi*

Summary. This paper explores analogies between the Weil proof of the Riemann hypothesis for function fields and the geometry of the adèles class space, which is the noncommutative space underlying Connes' spectral realization of the zeros of the Riemann zeta function. We consider the cyclic homology of the cokernel (in the abelian category of cyclic modules) of the "restriction map" defined by the inclusion of the idèles class group of a global field in the noncommutative adèles class space. Weil's explicit formula can then be formulated as a Lefschetz trace formula for the induced action of the idèles class group on this cohomology. In this formulation the Riemann hypothesis becomes equivalent to the positivity of the relevant trace pairing. This result suggests a possible dictionary between the steps in the Weil proof and corresponding notions involving the noncommutative geometry of the adèles class space, with good working notions of correspondences, degree, and codegree etc. In particular, we construct an analog for number fields of the algebraic points of the curve for function fields, realized here as classical points (low temperature KMS states) of quantum statistical mechanical systems naturally associated to the periodic orbits of the action of the idèles class group, that is, to the noncommutative spaces on which the geometric side of the trace formula is supported.

Key words: noncommutative geometry, cyclic homology, thermodynamics, factors, Riemann zeta function, correspondences, Lefschetz trace formula, adèles, idèles, Frobenius.

2000 Mathematics Subject Classifications: 11M36, 58B34, 11M26, 46L55

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*,
Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_8,

1 Introduction

This paper explores analogies between the Weil proof of the Riemann hypothesis for function fields and the geometry of the adèles class space, which is the noncommutative space underlying the spectral realization of the zeros of the Riemann zeta function constructed in [10]. Our purpose is to build a dictionary between the algebro-geometric setting of algebraic curves, divisors, the Riemann–Roch formula, and the Frobenius map, around which the Weil proof is built, and the world of noncommutative spaces, cyclic cohomology and KK-theory, index formulas, and the thermodynamic notions of quantum statistical mechanics, which, as we already argued in [11], provide an analogue of the Frobenius in characteristic zero via the scaling action on the dual system.

The present work builds on several previous results. The first input is the spectral realization of [10], where the adèles class space was first identified as the natural geometric space underlying the Riemann zeta function, where the Weil explicit formula acquires an interpretation as a trace formula. In [10] the analytic setting is that of Hilbert spaces, which provide the required positivity, but the spectral realization involves only the critical zeros. In [11], we provided a cohomological interpretation of the trace formula, using cyclic homology. In the setting of [11], the analysis is as developed by Ralf Meyer in [33] and uses spaces of rapidly decaying functions instead of Hilbert spaces. In this case, all zeros contribute to the trace formula, and the Riemann hypothesis becomes equivalent to a positivity question. This mirrors more closely the structure of the two main steps in the Weil proof, namely the explicit formula and the positivity $\mathrm{Tr}(Z * Z') > 0$ for correspondences (see below). The second main building block we need to use is the theory of endomotives and their quantum statistical mechanical properties we studied in [11]. Endomotives are a pseudo-abelian category of noncommutative spaces that naturally generalize the category of Artin motives. They are built from semigroup actions on projective limits of Artin motives. The morphisms in the category of endomotives generalize the notion of correspondence given by algebraic cycles in the product used in the theory of motives to the setting of étale groupoids, to account naturally for the presence of the semigroup actions. Endomotives carry a Galois action inherited from Artin motives and they have both an algebraic and an analytic manifestation. The latter provides the data for a quantum statistical mechanical system, via the natural time evolution associated by Tomita's theory to a probability measure carried by the analytic endomotive.

The main example that is of relevance to the Riemann zeta function is the endomotive underlying the Bost–Connes quantum statistical mechanical system of [5]. One can pass from a quantum statistical mechanical system to the "dual system" (in the sense of the duality of type III and type II factors in [6], [37]), which comes endowed with a scaling action induced by the time evolution. A general procedure described in [11] shows that there is a "restriction map" (defined as a morphism in the abelian category of modules over the cyclic category) from the dual system to a line bundle over the space of low-temperature KMS states of the quantum statistical mechanical system. The cokernel of this map is not defined at the level of algebras, but it makes sense in the abelian category and carries a corresponding scaling action. We argued in [11] that the induced scaling action on the cyclic homology of this cokernel may be thought of as an analogue of the action of Frobenius on étale cohomology. This claim is justified by the role that this scaling action of $\mathbb{R}_+^*$, combined with the action of $\hat{\mathbb{Z}}^*$ carried by the Bost–Connes endomotive, has in the trace formula; see [10], [11], and §4 of [13]. Further evidence for the role of the scaling action as Frobenius is given in [20], where it is shown that in the case of function fields, for a natural quantum statistical mechanical system that generalizes the Bost–Connes system to rank one Drinfeld modules, the scaling action on the dual system can be described in terms of the Frobenius and inertia groups.

In the present paper we continue along this line of thought. We begin by reviewing the main steps in the Weil proof for function fields, where we highlight the main conceptual steps and the main notions that will need an analogue in the noncommutative geometry setting. We conclude this part by introducing the main entries in our still tentative dictionary. The rest of the paper discusses in detail some parts of the dictionary and provides evidence in support of the proposed comparison. We begin this part by recalling briefly the properties of the Bost–Connes endomotive from [11] followed by the description of the "restriction map" corresponding to the inclusion of the idèles class group $C_{\mathbb{K}} = \mathbb{A}_{\mathbb{K}}^*/\mathbb{K}^*$ in the noncommutative adèles class space $X_{\mathbb{K}} = \mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$. We discuss its relation to the exact sequence of Hilbert spaces of [10], which plays a crucial role in obtaining the spectral realization as an "absorption spectrum."

We then concentrate on the geometry of the adèles class space over an arbitrary global field and the restriction map in this general setting, viewed as a map of cyclic modules. We introduce the actions ϑ_a and ϑ_m (with a and m respectively for additive and multiplicative) of $\mathbb{A}_{\mathbb{K}}^*$ on suitable function spaces on $\mathbb{A}_{\mathbb{K}}$ and on $C_{\mathbb{K}}$ and the induced action on the cokernel of the restriction map in the category of cyclic modules. We prove the corresponding general form of the associated Lefschetz trace formula, as a cohomological reformulation of the trace formula of [10] using the analytical setting of [33].

The form of the trace formula and the positivity property that is equivalent, in this setting, to the Riemann hypothesis for the corresponding L-functions with Grössencharakter, suggest by comparison with the analogous notions in the Weil proof a natural candidate for the analogue of the Frobenius correspondence on the curve. This is given by the graph of the scaling action. We can also identify the analogue of the degree and co-degree of a correspondence, and the analogue of the self-intersection of the diagonal on the curve, by looking at the explicit form of our Lefschetz trace formula. We also have a clear analogue of the first step in the Weil proof of positivity, which consists in adjusting the degree by multiples of the trivial correspondences.

This step is possible, with our notion of correspondences, due to a subtle failure of Fubini's theorem that allows us to modify the degree by adding elements in the range of the "restriction map", which play in this way the role of the trivial correspondences. This leaves open the more difficult question of identifying the correct analogue of the principal divisors, which is needed in order to continue the dictionary.

We then describe how to obtain a good analogue of the algebraic points of the curve in the number field case (in particular in the case of $\mathbb{K} = \mathbb{Q}$), in terms of the thermodynamic properties of the system. This refines the general procedure described in [11]. In fact, after passing to the dual system, one can consider the periodic orbits. We explain how, by the result of [10], these are the noncommutative spaces where the geometric side of the Lefschetz trace formula concentrates. We show that, in turn, these periodic orbits carry a time evolution and give rise to quantum statistical mechanical systems, of which one can consider the low-temperature KMS states. To each periodic orbit one can associate a set of "classical points" and we show that these arise as extremal low temperature KMS states of the corresponding system. We show that in the function field case, the space obtained in this way indeed can be identified, compatibly with the Frobenius action, with the algebraic points of the curve, albeit by a noncanonical identification. Passing to the dual system is the analogue in characteristic zero of the transition from $\mathbb{F}_q$ to its algebraic closure $\bar{\mathbb{F}}_q$. Thus, the procedure of considering periodic orbits in the dual system and classical points of these periodic orbits can be seen as an analogue, for our noncommutative space, of considering points defined over the extensions $\mathbb{F}_{q^n}$ of $\mathbb{F}_q$ in the case of varieties defined over finite fields (cf. [11] and §4 of [13]).

We analyze the behavior of the adèles class space under field extensions and the functoriality question. We then finish the paper by sketching an analogy between some aspects of the geometry of the adèles class spaces and the theory of singularities, which may be useful in adapting to this context some of the techniques of vanishing and nearby cycles.

2 A look at the Weil proof

In this preliminary section, we briefly review some aspects of the Weil proof of the Riemann hypothesis for function fields, with an eye on extending some of the basic steps and concepts to a noncommutative framework, which is what we will be doing in the rest of the paper.

In this section we let $\mathbb{K}$ be a global field of positive characteristic $p > 0$. One knows that in this case, there exists a smooth projective curve over a finite field $\mathbb{F}_q$, with $q = p^r$ for some $r \in \mathbb{N}$, such that

$$\mathbb{K} = \mathbb{F}_q(C) \tag{2.1}$$

is the field of functions of C. For this reason, a global field of positive characteristic is called a function field.

We denote by $\Sigma_{\mathbb{K}}$ the set of places of $\mathbb{K}$. A place $v \in \Sigma_{\mathbb{K}}$ is a Galois orbit of points of $C(\bar{\mathbb{F}}_q)$. The degree $n_v = \deg(v)$ is its cardinality, namely the number of points in the orbit of the Frobenius acting on the fiber of the natural map from points to places:

$$C(\bar{\mathbb{F}}_q) \to \Sigma_{\mathbb{K}}. \tag{2.2}$$

This means that the fiber over v consists of n_v conjugate points defined over $\mathbb{F}_{q^{n_v}}$, the residue field of the local field $\mathbb{K}_v$.

The curve C over $\mathbb{F}_q$ has a zeta function of the form

$$Z_C(T) = \exp\left(\sum_{n=1}^{\infty} \frac{\#C(\mathbb{F}_{q^n})}{n} T^n\right), \tag{2.3}$$

with $\log Z_C(T)$ the generating function for the number of points of C over the fields $\mathbb{F}_{q^n}$. It is customary to use the notation

$$\zeta_{\mathbb{K}}(s) = \zeta_C(s) = Z_C(q^{-s}). \tag{2.4}$$

It converges for $\Re(s) > 1$. In terms of Euler product expansions one writes

$$\zeta_{\mathbb{K}}(s) = \prod_{v \in \Sigma_{\mathbb{K}}} (1 - q^{-n_v s})^{-1}. \tag{2.5}$$

In terms of divisors of C, one has equivalently

$$\zeta_{\mathbb{K}}(s) = \zeta_C(s) = \sum_{D \geq 0} N(D)^{-s}, \tag{2.6}$$

where the norm of the divisor D is $N(D) = q^{\deg(D)}$.

The Riemann–Roch formula for the curve C states that

$$\ell(D) - \ell(\kappa_C - D) = \deg(D) - g + 1, \tag{2.7}$$

where κ_C is the canonical divisor on C, with degree $\deg(\kappa_C) = 2g - 2$ and $h^0(\kappa_C) = g$, and $\ell(D)$ the dimension of $H^0(D)$. Both $\deg(D)$ and $N(D)$ are well defined on the equivalence classes obtained by adding principal divisors, that is,

$$D \sim D' \iff D - D' = (f), \tag{2.8}$$

for some $f \in \mathbb{K}^*$. The Riemann–Roch formula (2.7) also implies that the zeta function $\zeta_{\mathbb{K}}(s)$ satisfies the functional equation

$$\zeta_{\mathbb{K}}(1 - s) = q^{(1-g)(1-2s)} \zeta_{\mathbb{K}}(s). \tag{2.9}$$

The zeta function $\zeta_{\mathbb{K}}(s)$ can also be written as a rational function

$$Z_C(T) = \frac{P(T)}{(1-T)(1-qT)}\,, \quad T = q^{-s}\,, \tag{2.10}$$

where $P(T)$ is a polynomial of degree $2g$ and integer coefficients

$$P(T) = \prod_{j=1}^{2g} (1 - \lambda_j T). \tag{2.11}$$

In particular, one has

$$\# C(\mathbb{F}_{q^n}) = q^n + 1 - \sum_{j=1}^{2g} \lambda_j. \tag{2.12}$$

Another important reformulation of the zeta function can be given in terms of étale cohomology. Namely, the coefficients $\# C(\mathbb{F}_{q^n})$ that appear in the zeta function can be rewritten as

$$\# C(\mathbb{F}_{q^n}) = \#\mathrm{Fix}(\mathrm{Fr}^n : \bar{C} \to \bar{C}) \tag{2.13}$$

with $\bar{C} = C \otimes_{\mathbb{F}_q} \bar{\mathbb{F}}_q$. The *Lefschetz fixed-point formula* for étale cohomology then shows that

$$\# C(\mathbb{F}_{q^n}) = \sum_{i=0}^{2} (-1)^i \mathrm{Tr}\left(\mathrm{Fr}^n |\, H^i_{\text{ét}}(\bar{C}, \mathbb{Q}_\ell)\right). \tag{2.14}$$

Thus, the zeta function can be written in the form

$$\zeta_{\mathbb{K}}(s) = \prod_{i=0}^{2} \left(\exp\left(\sum_{n=1}^{\infty} \mathrm{Tr}\left(\mathrm{Fr}^n |\, H^i_{\text{ét}}(\bar{C}, \mathbb{Q}_\ell)\right) \frac{q^{-sn}}{n} \right)\right)^{(-1)^i}. \tag{2.15}$$

The analogue of the Riemann hypothesis for the zeta functions $\zeta_{\mathbb{K}}(s)$ of function fields was stated in 1924 by E. Artin as the property that the zeros lie on the line $\Re(s) = 1/2$. Equivalently, it states that the complex numbers λ_j of (2.11), which are the eigenvalues of the Frobenius acting on $H^1_{\text{ét}}(\bar{C}, \mathbb{Q}_\ell)$, are algebraic numbers satisfying

$$|\lambda_j| = \sqrt{q}. \tag{2.16}$$

The Weil proof can be formulated either using étale cohomology, or purely in terms of the Jacobian of the curve, or again (equivalently) in terms of divisors on $C \times C$. We follow this last viewpoint and we recall in detail some of the main steps in the proof.

2.1 Correspondences and divisors

Correspondences Z, given by (nonvertical) divisors on $C \times C$, form a ring under composition,

$$Z_1 \star Z_2 = (p_{13})_* \left(p_{12}^* Z_1 \bullet p_{23}^* Z_2\right), \tag{2.17}$$

with $p_{ij} : C \times C \times C \to C \times C$ the projections, and $\bullet$ the intersection product. The ring has an involution

$$Z' = \sigma(Z), \tag{2.18}$$

where σ is the involution that exchanges the two copies of C in the product $C \times C$.

The degree $d(Z)$ and the codegree $d'(Z)$ are defined as the intersection numbers

$$d(Z) = Z \bullet (P \times C) \quad \text{and} \quad d'(Z) = Z \bullet (C \times P), \quad \forall P \in C. \tag{2.19}$$

They satisfy the relations

$$d(Z') = d'(Z) \quad \text{and} \quad d(Z_1 \star Z_2) = d(Z_1)d(Z_2). \tag{2.20}$$

The correspondences $P \times C$ and $C \times P$ are called *trivial correspondences*. One can consider the abelian group $\mathrm{Div}_{tr}(C \times C)$ generated by these trivial correspondences and take the quotient

$$\mathcal{C}(C) := \mathrm{Div}(C \times C)/\mathrm{Div}_{\mathrm{tr}}(C \times C). \tag{2.21}$$

It is always possible to change the degree and codegree of a correspondence Z by adding a multiple of the trivial correspondences $P \times C$ and $C \times P$, so that for any element in $\mathcal{C}$ we find a representative $Z \in \mathrm{Corr}$ with

$$d(Z) = d'(Z) = 0. \tag{2.22}$$

One also wants to consider correspondences up to linear equivalence,

$$Z_1 \sim Z_2 \Longleftrightarrow Z_1 - Z_2 = (f), \tag{2.23}$$

where (f) is a principal divisor on $C \times C$. Thus, one can consider

$$\mathrm{Pic}(C \times C) = \mathrm{Div}(C \times C)/ \sim$$

and its quotient $\mathcal{P}(C)$ modulo the classes of the trivial correspondences.

A correspondence Z is effective if it is given by an effective divisor on $C \times C$, namely if it is a combination $Z = \sum_i n_i Z_i$ of curves $Z_i \subset C \times C$ with coefficients $n_i \geq 0$. We write $Z \geq 0$ to mean its effectiveness. An effective correspondence $Z \geq 0$ that is nonempty can be viewed as a multivalued map

$$Z : C \to C, \qquad P \mapsto Z(P), \tag{2.24}$$

with $Z(P) = \mathrm{proj}_C(Z \bullet (P \times C))$, of which the divisor is the graph and with the product (2.17) given by the composition.

The trace of a correspondence Z on $C \times C$ is the expression

$$\mathrm{Tr}(Z) = d(Z) + d'(Z) - Z \bullet \Delta, \tag{2.25}$$

with Δ the diagonal (identity correspondence) and $\bullet$ the intersection product. This is well defined on $\mathcal{P}(C)$ since $\mathrm{Tr}(Z) = 0$ for principal divisors and trivial correspondences.

Consider a correspondence of degree g of the form $Z = \sum a_n \mathrm{Fr}^n$, given by a combination of powers of the Frobenius. Then Z can be made effective by adding a principal correspondence that is defined over $\mathbb{F}_q$ and that commutes with Fr.

This can be seen as follows. The Riemann–Roch theorem ensures that one can make Z effective by adding a principal correspondence, over the field $k(P)$, where k is the common field of definition of the correspondence Z and of the curve (cf. [35]) and P is a generic point. A correspondence of the form $Z = \sum a_n \mathrm{Fr}^n$ is in fact defined over $\mathbb{F}_q$, hence the principal correspondence is also defined over $\mathbb{F}_q$. As such, it automatically commutes with Fr (cf. [38, p. 287]).

Notice, however, that in general, it is not possible to modify a divisor D of degree one on C to an effective divisor in such a way that the added principal divisor has support on the same Frobenius orbit. An illustrative example is given in Chapter 4 of [13].

2.2 The explicit formula

The main steps in the Weil proof of the Riemann hypothesis for function fields are

1. The explicit formula
2. Positivity

Let $\mathbb{K}$ be a global field and let $\mathbb{A}_{\mathbb{K}}$ denote its ring of adèles. Let $\Sigma_{\mathbb{K}}$ denote the set of places of $\mathbb{K}$. Let α be a nontrivial character of $\mathbb{A}_{\mathbb{K}}$ that is trivial on $\mathbb{K} \subset \mathbb{A}_{\mathbb{K}}$. We write

$$\alpha = \prod_{v \in \Sigma_{\mathbb{K}}} \alpha_v \tag{2.26}$$

for the decomposition of α as a product of its restrictions to the local fields $\alpha_v = \alpha|_{\mathbb{K}_v}$.

Consider the bicharacter

$$\langle z, \lambda \rangle := \lambda^z, \qquad \text{for } (z, \lambda) \in \mathbb{C} \times \mathbb{R}_+^*. \tag{2.27}$$

Let N denote the range of the norm $|\cdot| : C_{\mathbb{K}} \to \mathbb{R}_+^*$. Then $N^\perp \subset \mathbb{C}$ denotes the subgroup

$$N^\perp := \{z \in \mathbb{C} |\, \lambda^z = 1,\ \forall \lambda \in N\}. \tag{2.28}$$

Consider then the expression

$$\sum_{\rho\in\mathbb{C}/N^{\perp}\,|\,L(\tilde{\chi},\rho)=0} \hat{f}(\tilde{\chi},\rho), \tag{2.29}$$

with $L(\tilde{\chi},\rho)$ the L-function with Grössencharakter χ, where $\tilde{\chi}$ denotes the extension to $C_{\mathbb{K}}$ of the character $\chi \in \widehat{C_{\mathbb{K},1}}$, the Pontryagin dual of $C_{\mathbb{K},1}$. Here $\hat{f}(\tilde{\chi},\rho)$ denotes the Fourier transform

$$\hat{f}(\tilde{\chi},\rho) = \int_{C_{\mathbb{K}}} f(u)\tilde{\chi}(u)\,|u|^{\rho}\,d^*u \tag{2.30}$$

of a test function f in the Schwartz space $\mathcal{S}(C_{\mathbb{K}})$.

In the case $N = q^{\mathbb{Z}}$ (function fields), the subgroup $N^{\perp}$ is nontrivial and given by

$$N^{\perp} = \frac{2\pi i}{\log q}\mathbb{Z}. \tag{2.31}$$

Since in the function field case the L-fuctions are functions of q^{-s}, there is a periodicity by $N^{\perp}$, hence we need to consider only $\rho \in \mathbb{C}/N^{\perp}$. In the number field case this does not matter, since $N = \mathbb{R}_+^*$ and $N^{\perp}$ is trivial.

The Weil explicit formula is the remarkable identity [43]

$$\hat{h}(0) + \hat{h}(1) - \sum_{\rho\in\mathbb{C}/N^{\perp}\,|\,L(\tilde{\chi},\rho)=0} \hat{h}(\tilde{\chi},\rho) = \sum_{v\in\Sigma_{\mathbb{K}}} \int'_{(\mathbb{K}_v^*,\alpha_v)} \frac{h(u^{-1})}{|1-u|}\,d^*u. \tag{2.32}$$

Here the Fourier transform $\hat{h}$ is as in (2.30). The test function h has compact support and belongs to the Schwartz space $\mathcal{S}(C_{\mathbb{K}})$. As soon as $h(1) \neq 0$ the integrals on the right-hand side are singular, so that one needs to specify how to take their principal value. This was done in [43] and it was shown in [10] that the same principal value can in fact be defined in the following unified way.

Definition 2.1. *For a local field K and a given (nontrivial) additive character β of K, one lets ϱ_β denote the unique distribution extending d^*u at $u = 0$, whose Fourier transform*

$$\hat{\varrho}(y) = \int_K \varrho(x)\beta(xy)\,dx \tag{2.33}$$

satisfies the vanishing condition $\hat{\varrho}(1) = 0$.

Then by definition the principal value $\int'$ is given by

$$\int'_{(K,\beta)} \frac{f(u^{-1})}{|1-u|}\,d^*u = \langle \varrho_\beta, g\rangle, \quad \text{with } g(\lambda) = \frac{f((\lambda+1)^{-1})}{|\lambda+1|}, \tag{2.34}$$

where $\langle \varrho_\beta, g\rangle$ denotes the pairing of the distribution ϱ_β and the function $g(\lambda)$. This makes sense provided the support of f is compact, which implies that $g(\lambda)$ vanishes identically in a neighborhood of $\lambda = -1$.

The Weil explicit formula is a far-reaching generalization of the relation between primes and zeros of the Riemann zeta function, originally due to Riemann [34].

2.3 Riemann–Roch and positivity

Weil positivity is the statement that if Z is a nontrivial correspondence in $\mathcal{P}(C)$ (i.e., as above, a correspondence on $C \times C$ modulo trivial ones and up to linear equivalence), then

$$\mathrm{Tr}(Z \star Z') > 0. \tag{2.35}$$

This is proved using the Riemann–Roch formula on C to show that one can achieve effectivity. In fact, using trivial correspondences to adjust the degree one can assume that $d(Z) = g$. Then the Riemann–Roch formula (2.7) shows that if D is a divisor on C of degree $\deg(D) = g$, then there are effective representatives in the linear equivalence class of D. The intersection of $Z \subset C \times C$ with $P \times C$ defines a divisor $Z(P)$ on C with

$$\deg(Z(P)) = d(Z) = g.$$

Thus, the argument above shows that there exists $f_P \in \mathbb{K}^*$ such that $Z(P) + (f_P)$ is effective. This determines an effective divisor $Z + (f)$ on $C \times C$. Thus, we can assume that Z is effective; hence we can write it as a multivalued function

$$P \mapsto Z(P) = Q_1 + \cdots + Q_g. \tag{2.36}$$

The product $Z \star Z'$ is of the form

$$Z \star Z' = d'(Z)\Delta + Y, \tag{2.37}$$

where Δ is the diagonal in $C \times C$ and Y is the effective correspondence such that $Y(P)$ is the divisor on C given by the sum of points in

$$\{Q \in C |\, Q = Q_i(P) = Q_j(P),\, i \neq j\}.$$

One sees this is what is intended from the description in terms of intersection product that it is given by the multivalued function

$$(Z \star Z')(Q) = \sum_{i,j} \sum_{P \in \mathcal{U}_{ij}(Q)} P, \tag{2.38}$$

where

$$\mathcal{U}_{ij}(Q) = \{P \in C |\, Q_i(P) = Q_j(P) = Q\}.$$

One can separate this out in the contribution of the locus where $Q_i = Q_j$ for $i \neq j$ and the part where $i = j$,

$$(Z \star Z')(Q) = U(Q) + Y(Q).$$

Notice that

$$\#\{P \in C |\, Q = Q_i(P),\ \text{for some } i = 1, \dots, g\} = d'(Z). \tag{2.39}$$

Thus, for $i = j$ we obtain that the divisor $U(Q) = \sum_i \sum_{P \in \mathcal{U}_{ii}(Q)} P$ is just $d'(Z)\Delta(Q)$, while for $i \neq j$ one obtains the remaining term Y of (2.37).

In the case $g = 1$, the effective correspondence $Z(P) = Q(P)$ is single-valued and the divisor $(Z \star Z')(P)$ of (2.38) reduces to the sum of points in

$$\mathcal{U}(Q) = \{P \in C |\, Q(P) = Q\}.$$

There are $d'(Z)$ such points, so one obtains

$$Z \star Z' = d'(Z)\Delta, \quad \text{with} \quad \mathrm{Tr}(Z \star Z') = 2d'(Z) \geq 0, \tag{2.40}$$

since for $g = 1$ one has $\Delta \bullet \Delta = 0$ and $d'(Z) \geq 0$ since Z is effective.

In the case of genus $g > 1$, the Weil proof proceeds as follows. Let κ_C be a choice of an effective canonical divisor for C without multiple points, and let $\{f_1, \dots, f_g\}$ be a basis of the space $H^0(\kappa_C)$. One then considers the function $C \to M_{g\times g}(\mathbb{F}_q)$ to $g \times g$ matrices

$$P \mapsto M(P), \quad \text{with} \quad M_{ij}(P) = f_i(Q_j(P)), \tag{2.41}$$

and the function $K : C \to \mathbb{F}_q$ given by

$$K(P) = \det(M(P))^2. \tag{2.42}$$

The function $P \mapsto K(P)$ of (2.42) is a rational function with $(2g-2)d'(Z)$ double poles. In fact, $K(P)$ is a symmetric function of the $Q_j(P)$, because of the squaring of the determinant. The composition $P \mapsto (Q_j(P)) \mapsto K(P)$ is then a rational function of $P \in C$. The poles occur (as double poles) at those points $P \in C$ for which some $Q_i(P)$ is a component of κ_C. The canonical divisor κ_C has degree $2g - 2$. This means that there are $(2g - 2)d'(Z)$ such double poles.

For $Z \star Z' = d'(Z)\Delta + Y$ as above, the intersection number $Y \bullet \Delta$ satisfies the estimate

$$Y \bullet \Delta \leq (4g - 4)\, d'(Z). \tag{2.43}$$

In fact, the rational function $K(P)$ of (2.42) has a number of zeros equal to $(4g - 4)\, d'(Z)$. On the other hand, $Y \bullet \Delta$ counts the number of times that $Q_i = Q_j$ for $i \neq j$. Since each point P with $Q_i(P) = Q_j(P)$ for $i \neq j$ produces a zero of $K(P)$, one sees that $Y \bullet \Delta$ satisfies the estimate (2.43). Notice that for genus $g > 1$, the self-intersection of the diagonal is the Euler characteristic

$$\Delta \bullet \Delta = 2 - 2g = \chi(C). \tag{2.44}$$

Moreover, we have

$$d(Z \star Z') = d(Z)d'(Z) = g\, d'(Z) = d'(Z \star Z'). \tag{2.45}$$

Thus, using again the decomposition (2.37) and the definition of the trace of a correspondence (2.25), together with (2.44) and (2.45) one obtains

$$\begin{aligned} \mathrm{Tr}(Z \star Z') &= 2g\, d'(Z) + (2g-2)\, d'(Z) - Y \bullet \Delta \\ &\geq (4g-2)\, d'(Z) - (4g-4)\, d'(Z) = 2d'(Z) \geq 0. \end{aligned} \tag{2.46}$$

This gives the positivity (2.35).

In the Weil proof of the Riemann hypothesis for function fields, one concentrates on a particular type of correspondences, namely those that are of the form

$$Z_{n,m} = m\Delta + n\mathrm{Fr}, \tag{2.47}$$

for $n, m \in \mathbb{Z}$, with Fr the Frobenius correspondence.

Notice that while the correspondence depends linearly on $n, m \in \mathbb{Z}$, the expression for the trace gives

$$\mathrm{Tr}\left(Z_{n,m} \star Z'_{n,m}\right) = 2gm^2 + 2(1 + q - N)mn + 2gqn^2, \tag{2.48}$$

where $N = \#C(\mathbb{F}_q)$. In particular, (2.48) depends quadratically on (n, m). In the process of passing from a correspondence of degree g to an effective correspondence, this quadratic dependence on (n, m) is contained in the multiplicity $d'(Z)$. Notice, moreover, that the argument does not depend on the torsion part of the ring of correspondences.

2.4 A tentative dictionary

In the rest of the paper we illustrate some steps toward the creation of a dictionary relating the main steps in the Weil proof described above to the noncommutative geometry of the adèles class space of a global field. The noncommutative geometry approach has the advantage that it provides (see [10], [33], [11]) a Lefschetz trace formula interpretation for the Weil explicit formula and that it gives a parallel formulation for both function fields and number fields. Parts of the dictionary sketched below are very tentative at this stage, so we mostly concentrate, in the rest of the paper, on illustrating what we put in the first few lines of the dictionary, and on the role of the scaling correspondence as Frobenius and its relation to the explicit formula.

Frobenius correspondence	$Z(f) = \int_{C_{\mathbb{K}}} f(g)\, Z_g\, d^*g$
Trivial correspondences	Elements of the range $\mathcal{V}$
Adjusting the degree by trivial correspondences	Fubini step on the test functions
Correspondences	Bivariant elements $Z(f) \Rightarrow \Gamma(f)$
Degree of a correspondence	Pointwise index
Riemann–Roch	Index theorem
Effective correspondences	Epimorphism of C^*-modules
$\deg Z(P) \geq g \Rightarrow Z + (f)$ effective	$d(\Gamma) > 0 \Rightarrow \Gamma + K$ onto
Lefschetz formula	Bivariant Chern of $\Gamma(f)$ (by localization on the graph $Z(f)$)

3 Quantum statistical mechanics and arithmetic

The work of Bost–Connes [5] first revealed the presence of an interesting interplay between quantum statistical mechanics and Galois theory. More recently, several generalizations [12], [14], [15], [11], [20], [25], [26] have confirmed and expanded this viewpoint. The general framework of interactions between noncommutative geometry and number theory described in [30], [31], [13], [32] recast these phenomena into a broader picture, of which we explore in this paper but one of many facets.

The basic framework that combines quantum statistical mechanics and Galois theory can be seen as an extension, involving noncommutative spaces, of the category of Artin motives. In the setting of pure motives (see [28]), Artin motives correspond to the subcategory generated by zero-dimensional objects, with morphisms given by algebraic cycles in the product (in this case without the need to specify with respect to which equivalence relation). Endomotives were introduced in [11] as noncommutative spaces of the form

$$\mathcal{A}_{\mathbb{K}} = A \rtimes S, \tag{3.1}$$

where A is an inductive limit of reduced finite-dimensional commutative algebras over the field $\mathbb{K}$, i.e. a projective limit of Artin motives, and S is a unital abelian semigroup of algebra endomorphisms $\rho : A \to A$. The crossed product (3.1) is obtained by adjoining to A new generators U_ρ and U_ρ^*, for $\rho \in S$, satisfying the relations

$$\begin{aligned} &U_\rho^* U_\rho = 1, && U_\rho U_\rho^* = \rho(1), && \forall \rho \in S, \\ &U_{\rho_1 \rho_2} = U_{\rho_1} U_{\rho_2}, && U^*_{\rho_2 \rho_1} = U^*_{\rho_1} U^*_{\rho_2}, && \forall \rho_1, \rho_2 \in S, \\ &U_\rho\, a = \rho(a)\, U_\rho, && a\, U_\rho^* = U_\rho^*\, \rho(a), && \forall \rho \in S, \forall a \in A. \end{aligned} \tag{3.2}$$

The algebras (3.1) have the following properties: the algebra A is unital; the image $e = \rho(1) \in A$ is an idempotent, for all $\rho \in S$; each $\rho \in S$ is an isomorphism of A with the compressed algebra eAe. A general construction given in [11] based on self maps of algebraic varieties provides a large class of examples over different fields $\mathbb{K}$. We are mostly interested here in the case in which $\mathbb{K}$ is a number field, and for part of our discussion below we will concentrate on a special case (the Bost–Connes endomotive) over the field $\mathbb{K} = \mathbb{Q}$.

Endomotives form a pseudo-abelian category in which morphisms are correspondences given by $\mathcal{A}_{\mathbb{K}}$–$\mathcal{B}_{\mathbb{K}}$-bimodules that are finite and projective as right modules. These define morphisms in the additive KK-category and in the abelian category of cyclic modules. In fact, in addition to the algebraic form described above, endomotives also have an analytic structure given by considering, instead of the $\mathbb{K}$-algebra (3.1), the C^*-algebra

$$C(X) \rtimes S, \tag{3.3}$$

where X denotes the totally disconnected Hausdorff space $X = X(\bar{\mathbb{K}})$ of algebraic points of the projective limit of Artin motives. This follows by first showing that the semigroup S acts by endomorphisms of the C^*-algebra of continuous functions $C(X)$ and then passing to the norm completion of the algebraic crossed product as in the general theory of semigroups crossed product C^*-algebras (see [27]; see also [13, Chapter 4, §2.2]). The C^*-completion is taken with respect to the norm

$$\|f\| = \sup_{x \in X} \|\pi_x(f)\|, \tag{3.4}$$

where, for each $x \in X$, one lets π_x be the representation by left convolution on the Hilbert space $\ell^2(\mathcal{G}_x)$ of the countable fiber $\mathcal{G}_x$ over $x \in X$ of the source map of the groupoid $\mathcal{G}$ associated to the action of S on X.

There is a canonical action of the Galois group $G = \mathrm{Gal}(\bar{\mathbb{K}}/\mathbb{K})$ by automorphisms of the C^*-algebra (3.3) globally preserving $C(X)$, coming from the action by composition of G on $X(\bar{\mathbb{K}}) = \mathrm{Hom}(A, \bar{\mathbb{K}})$. We refer the reader to [11] for a more detailed discussion of algebraic and analytic endomotives and the properties of morphisms in the corresponding categories.

If the endomotive is "uniform" in the sense specified in [11], the space X comes endowed with a probability measure μ that induces a state φ on the C^*-algebra (3.3). The general Tomita theory of modular automorphism groups in the context of von Neumann algebras [36] shows that there is a natural time evolution for which the state φ is KMS_1. We refer the reader to [13, Chap. 4, §4.1] for a more detailed discussion of this step and the necessary von Neumann algebra background. Here the KMS_1 condition means the following.

Given a system of an algebra of observables with a time evolution, there is a good notion of thermodynamic equilibrium states given by the KMS condition. A state on a unital C^*-algebra is a continuous linear functional $\varphi : \mathcal{A} \to \mathbb{C}$ with $\varphi(1) = 1$ and $\varphi(a^*a) \geq 0$ for all $a \in \mathcal{A}$. The KMS condition at inverse temperature β for a state φ is the property that for all $a, b \in \mathcal{A}$ there exists a function $F_{a,b}(z)$ that is holomorphic on the strip $I_\beta = \{z \in \mathbb{C} \,|\, 0 < \Im(z) < \beta\} \subset \mathbb{C}$ and continuous on $I_\beta \cup \partial I_\beta$, satisfying $F_{a,b}(t) = \varphi(a\sigma_t(b))$ and $F_{a,b}(t + i\beta) = \varphi(\sigma_t(b)a)$, for all $t \in \mathbb{R}$. For a more detailed account of the quantum statistical mechanical formalism, the notion of KMS state, and the properties of such states, we refer the reader to Chaper 3 of [13], Section 2.

One can then consider the set Ω_β of low-temperature (large β) KMS states for the same quantum statistical mechanical system obtained from a uniform endomotive in the way described above.

One also associates to the system $(\mathcal{A}, \sigma)$ of the C^*-algebra with the time evolution its dual system $(\hat{\mathcal{A}}, \theta)$, where the algebra $\hat{\mathcal{A}} = \mathcal{A} \rtimes_\sigma \mathbb{R}$ is obtained by taking the crossed product with the time evolution and θ is the scaling action of $\mathbb{R}_+^*$:

$$\theta_\lambda \left(\int x(t)\, U_t\, dt \right) = \int \lambda^{it}\, x(t)\, U_t\, dt. \tag{3.5}$$

One then constructs an $\mathbb{R}_+^*$-equivariant map

$$\pi : \hat{\mathcal{A}}_\beta \to C(\tilde{\Omega}_\beta, \mathcal{L}^1), \tag{3.6}$$

from a suitable subalgebra $\hat{\mathcal{A}}_\beta \subset \hat{\mathcal{A}}$ of the dual system to functions on a principal $\mathbb{R}_+^*$-bundle $\tilde{\Omega}_\beta$ over the low-temperature KMS states of the system, with values in trace class operators. Since traces define morphisms in the cyclic category, the map (3.6) can be used to construct a morphism $\delta = (\mathrm{Tr} \circ \pi)^\natural$ at the level of cyclic modules

$$\hat{\mathcal{A}}_\beta^\natural \stackrel{(\mathrm{Tr}\circ\pi)^\natural}{\longrightarrow} C(\tilde{\Omega}_\beta)^\natural. \tag{3.7}$$

This map can be loosely thought of as a "restriction map" corresponding to the inclusion of the "classical points" in the noncommutative space. One can then consider the cokernel of this map in the abelian category of cyclic modules. In [11] we called the procedure described above "cooling and distillation" of endomotives. We refer the reader to [11] for the precise technical hypotheses under which this procedure can be performed. Here we have given only an impressionistic sketch aimed at recalling briefly the main steps involved.

3.1 The Bost–Connes endomotive

The main example of endomotive we will consider here in relation to the geometry of the adèles class space is the Bost–Connes system. This can be constructed as an endomotive over $\mathbb{K} = \mathbb{Q}$, starting from the projective system $X_n = \mathrm{Spec}(A_n)$, with $A_n = \mathbb{Q}[\mathbb{Z}/n\mathbb{Z}]$ the group ring of $\mathbb{Z}/n\mathbb{Z}$. The inductive limit is the group ring $A = \mathbb{Q}[\mathbb{Q}/\mathbb{Z}]$ of $\mathbb{Q}/\mathbb{Z}$. The endomorphism ρ_n associated to an element $n \in S$ of the (multiplicative) semigroup $S = \mathbb{N} = \mathbb{Z}_{>0}$ is given on the canonical basis $e_r \in \mathbb{Q}[\mathbb{Q}/\mathbb{Z}]$, $r \in \mathbb{Q}/\mathbb{Z}$, by

$$\rho_n(e_r) = \frac{1}{n} \sum_{ns=r} e_s. \tag{3.8}$$

The corresponding analytic endomotive is the crossed product C^*-algebra

$$\mathcal{A} = C^*(\mathbb{Q}/\mathbb{Z}) \rtimes \mathbb{N}.$$

The Galois action is given by composing a character $\chi : A_n \to \bar{\mathbb{Q}}$ with an element g of the Galois group $G = \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$. Since χ is determined by the nth root of unity $\chi(e_{1/n})$, one obtains the cyclotomic action.

In the case of the Bost–Connes endomotive, the state φ on $\mathcal{A}$ induced by the measure μ on $X = \hat{\mathbb{Z}}$ is of the form

$$\varphi(f) = \int_{\hat{\mathbb{Z}}} f(1,\rho)\, d\mu(\rho), \tag{3.9}$$

and the modular automorphism group restricts to the C^*-algebra as the time evolution of the BC system; cf. [5], [11] and [13, §4].

The dual system of the Bost–Connes system is best described in terms of commensurability classes of $\mathbb{Q}$-lattices. In [12] the Bost–Connes system is reinterpreted as the noncommutative space describing the relation of commensurability for 1-dimensional $\mathbb{Q}$-lattices up to scaling. One can also consider the same equivalence relation without dividing out by the scaling action. If we let $\mathcal{G}_1$ denote the groupoid of the commensurability relation on 1-dimensional $\mathbb{Q}$-lattices, and $\mathcal{G}_1/\mathbb{R}_+^*$ the one obtained after modding out by scaling, we identify the C^*-algebra of the Bost–Connes system with $C^*\left(\mathcal{G}_1/\mathbb{R}_+^*\right)$ (cf. [12]). The algebra $\hat{\mathcal{A}}$ of the dual system is then obtained in the following way (cf. [11]). There is a C^*-algebra isomorphism $\iota : \hat{\mathcal{A}} \to C^*(\mathcal{G}_1)$ of the form

$$\iota(X)(k,\rho,\lambda) = \int_{\mathbb{R}} x(t)(k,\rho)\, \lambda^{it}\, dt \tag{3.10}$$

for $(k,\rho,\lambda) \in \mathcal{G}_1$ and $X = \int x(t)\, U_t\, dt \in \hat{\mathcal{A}}$.

3.2 Scaling as Frobenius in characteristic zero

In the general setting described in [11] one denotes by $D(\mathcal{A},\varphi)$ the cokernel of the morphism (3.7), viewed as a module in the cyclic category. The notation is meant to recall the dependence of the construction on the initial data of an analytic endomotive $\mathcal{A}$ and a state φ. The cyclic module $D(\mathcal{A},\varphi)$ inherits a scaling action of $\mathbb{R}_+^*$, and one can consider the induced action on the cyclic homology $HC_0(D(\mathcal{A},\varphi))$. We argued in [11] that this cyclic homology with the induced scaling action plays a role analogous to the role played by the Frobenius action on étale cohomology in the algebro-geometric context. Our main supporting evidence is the Lefschetz trace formula for this action which gives a cohomological interpretation of the spectral realization of the zeros of the Riemann zeta function of [10]. We return to discuss the Lefschetz trace formula for the more general case of global fields in Section 6 below.

The main results of [10] show that we have the following setup. There is an exact sequence of Hilbert spaces

$$0 \to L^2_\delta(\mathbb{A}_\mathbb{Q}/\mathbb{Q}^*)_0 \to L^2_\delta(\mathbb{A}_\mathbb{Q}/\mathbb{Q}^*) \to \mathbb{C}^2 \to 0, \tag{3.11}$$

which defines the subspace $L^2_\delta(\mathbb{A}_\mathbb{Q}/\mathbb{Q}^*)_0$ by imposing the conditions $f(0)=0$ and $\hat{f}(0)=0$ and a suitable decay condition imposed by the weight δ. The space $L^2_\delta(\mathbb{A}_\mathbb{Q}/\mathbb{Q}^*)_0$ fits into another exact sequence of Hilbert spaces of the form

$$0 \to L^2_\delta(\mathbb{A}_\mathbb{Q}/\mathbb{Q}^*)_0 \xrightarrow{\mathfrak{E}} L^2_\delta(C_\mathbb{Q}) \to \mathcal{H} \to 0, \tag{3.12}$$

where the map $\mathfrak{E}$ is defined by

$$\mathfrak{E}(f)(g) = |g|^{1/2} \sum_{q\in\mathbb{Q}^*} f(qg), \qquad \forall g \in C_\mathbb{Q} = \mathbb{A}^*_\mathbb{Q}/\mathbb{Q}^*. \tag{3.13}$$

The map is equivariant with respect to the actions of $C_\mathbb{Q}$ i.e.,

$$\mathfrak{E} \circ \vartheta_a(\gamma) = |\gamma|^{1/2} \vartheta_m(\gamma) \circ \mathfrak{E}, \tag{3.14}$$

where $(\vartheta_a(\gamma)\xi)(x) = \xi(\gamma^{-1}x)$ for $\xi \in L^2_\delta(\mathbb{A}_\mathbb{Q}/\mathbb{Q}^*)_0$, and similarly $\vartheta_m(\gamma)$ is the regular representation of $C_\mathbb{K}$.

We showed in [11] that the map $\mathfrak{E}$, translated from the context of Hilbert spaces to that of nuclear spaces as in [33], has a natural interpretation in terms of the "cooling and distillation process" for the Bost–Connes (BC) endomotive. In fact, we showed in [11] that if $(\mathcal{A},\sigma)$ denotes the BC system, then the following properties hold:

1. For $\beta > 1$ there is a canonical isomorphism

$$\tilde{\Omega}_\beta \simeq \hat{\mathbb{Z}}^* \times \mathbb{R}_+^* \simeq C_\mathbb{Q} \tag{3.15}$$

   of $\tilde{\Omega}_\beta$ with the space of invertible 1-dimensional $\mathbb{Q}$-lattices.

2. For $X \in \hat{\mathcal{A}}$ and $f = \iota(X) \in C^*(\mathcal{G}_1)$, the cooling map (3.7) takes the form

$$\delta(X)(u,\lambda) = \sum_{n \in \mathbb{N} = \mathbb{Z}_{>0}} f(1, nu, n\lambda), \quad \forall (u,\lambda) \in C_{\mathbb{Q}} \simeq \tilde{\Omega}_\beta. \tag{3.16}$$

One can compare directly the right-hand side of (3.16) with the map $\mathfrak{E}$ (up to the normalization by $|j|^{1/2}$) written as in (3.13) by considering a function $f(\rho, v) = f(1, \rho, v)$ and its unique extension $\tilde{f}$ to adèles, where f is extended by 0 outside $\hat{\mathbb{Z}} \times \mathbb{R}^*$ and one requires the parity

$$\tilde{f}(-u, -\lambda) = f(u, \lambda)\,. \tag{3.17}$$

This then gives

$$\sum_{n \in \mathbb{N}} f(1, nu, n\lambda) = \frac{1}{2} \sum_{q \in \mathbb{Q}^*} \tilde{f}(q\,j), \qquad \text{where } j = (u,\lambda) \in C_{\mathbb{Q}}. \tag{3.18}$$

4 The adèles class space

Let $\mathbb{K}$ be a global field, with $\mathbb{A}_{\mathbb{K}}$ its ring of adèles.

Definition 4.1. *The adèles class space of a global field $\mathbb{K}$ is the quotient $\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$.*

When viewed from the classical standpoint this is a "bad quotient" due to the ergodic nature of the action, which makes the quotient ill-behaved topologically. Thus, following the general philosophy of noncommutative geometry, we describe it by a noncommutative algebra of coordinates, which allows one to continue to treat the quotient as a "nice quotient" in the context of noncommutative geometry.

A natural choice of the algebra is the crossed product

$$C_0(\mathbb{A}_{\mathbb{K}}) \rtimes \mathbb{K}^* \quad \text{with the smooth subalgebra} \quad \mathcal{S}(\mathbb{A}_{\mathbb{K}}) \rtimes \mathbb{K}^*. \tag{4.1}$$

A better description can be given in terms of groupoids.

Consider the groupoid law $\mathcal{G}_{\mathbb{K}} = \mathbb{K}^* \ltimes \mathbb{A}_{\mathbb{K}}$ given by

$$(k,x) \circ (k', y) = (kk', y), \quad \forall k, k' \in \mathbb{K}^*, \quad \text{and } \forall x, y \in \mathbb{A}_{\mathbb{K}} \text{ with } x = k'y, \tag{4.2}$$

with the composition (4.2) defined whenever the source $s(k,x) = x$ agrees with the range $r(k', y) = k'y$.

Lemma 4.2. *The algebras (4.1) are, respectively, the groupoid C^*-algebra $C^*(\mathcal{G}_{\mathbb{K}})$ and its dense subalgebra $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$.*

Proof. The product in the groupoid algebra is given by the associative convolution product

$$(f_1 * f_2)\,(k,x) = \sum_{s\in\mathbb{K}^*} f_1(k\,s^{-1}, s\,x) f_2(s,x), \tag{4.3}$$

and the adjoint is given by $f^*(k,x) = \overline{f(k^{-1}, k\,x)}$.

The functions (on the groupoid) associated to $f \in \mathcal{S}(\mathbb{A}_K)$ and U_k are given, respectively, by

$$\begin{aligned} f(1,x) &= f(x) \text{ and } \ f(k,x) &= 0 \ \forall k \neq 1, \\ U_k(k,x) &= \ \ 1 \quad \text{and } U_g(k,x) &= 0 \ \forall g \neq k. \end{aligned} \tag{4.4}$$

The product $f\,U_k$ is then the convolution product of the groupoid.

The algebra $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$ is obtained by considering finite sums of the form

$$\sum_{k\in\mathbb{K}^*} f_k\,U_k, \quad \text{for} \quad f_k \in \mathcal{S}(\mathbb{A}_{\mathbb{K}}). \tag{4.5}$$

The product is given by the convolution product

$$(U_k\,f\,U_k^*)\,(x) = f(k^{-1}x), \tag{4.6}$$

for $f \in \mathcal{S}(\mathbb{A}_{\mathbb{K}})$, $k \in \mathbb{K}^*$, and $x \in \mathbb{A}_{\mathbb{K}}$. □

4.1 Cyclic module

We can associate to the algebra $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$ of the adèles class space an object in the category of Λ-modules. This means that we consider the cyclic module $\mathcal{S}(\mathcal{G}_{\mathbb{K}})^\natural$ and the two cyclic morphisms

$$\varepsilon_j : \mathcal{S}(\mathcal{G}_{\mathbb{K}})^\natural \to \mathbb{C} \tag{4.7}$$

given by

$$\varepsilon_0(\sum f_k\,U_k) = f_1(0) \quad \text{and} \quad \varepsilon_1(\sum f_k\,U_k) = \int_{\mathbb{A}_{\mathbb{K}}} f_1(x)\,dx \tag{4.8}$$

and in higher degree by

$$\varepsilon_j^\natural(a^0 \otimes \cdots \otimes a^n) = \varepsilon_j(a^0 \, \cdots \, a^n). \tag{4.9}$$

The morphism ε_1 is given by integration on $\mathbb{A}_{\mathbb{K}}$ with respect to the additive Haar measure. This is $\mathbb{K}^*$-invariant; hence it defines a trace on $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$. In the case of $\mathbb{K} = \mathbb{Q}$, this corresponds to the dual trace τ_φ for the KMS$_1$-state φ associated to the time evolution of the BC system. The morphism ε_0 here takes into account the fact that we are imposing a vanishing condition at $0 \in \mathbb{A}_{\mathbb{K}}$ (cf. [11] and [13, Chapter 4]). In fact, the Λ-module we associate to $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$ is given by

$$\mathcal{S}(\mathcal{G}_{\mathbb{K}})_0^\natural := \operatorname{Ker} \varepsilon_0^\natural \cap \operatorname{Ker} \varepsilon_1^\natural. \tag{4.10}$$

Note that since $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$ is nonunital, the cyclic module $\mathcal{S}(\mathcal{G}_{\mathbb{K}})^\natural$ is obtained using the adjunction of a unit to $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$.

4.2 The restriction map

Consider the idèles $\mathbb{A}_{\mathbb{K}}^* = \mathrm{GL}_1(\mathbb{A}_{\mathbb{K}})$ of $\mathbb{K}$ with their natural locally compact topology induced by the map

$$\mathbb{A}_{\mathbb{K}}^* \ni g \mapsto (g, g^{-1}). \tag{4.11}$$

We can see the idèles class group $C_{\mathbb{K}} = \mathbb{A}_{\mathbb{K}}^*/\mathbb{K}^*$ as a subspace of the adèles class space $X_{\mathbb{K}} = \mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$ in the following way.

Lemma 4.3. *The pairs $((k,x),(k',y)) \in \mathcal{G}_{\mathbb{K}}$ such that both x and y are in $\mathbb{A}_{\mathbb{K}}^*$ form a full subgroupoid of $\mathcal{G}_{\mathbb{K}}$ that is isomorphic to $\mathbb{K}^* \ltimes \mathbb{A}_{\mathbb{K}}^*$.* □

Proof. Elements of $\mathbb{A}_{\mathbb{K}}$ whose orbit under the $\mathbb{K}^*$ action contains an idelè are also idèles. Thus, we obtain a groupoid that is a full subcategory of $\mathcal{G}_{\mathbb{K}}$.

This implies the existence of a restriction map. Consider the map

$$\rho : \mathcal{S}(\mathbb{A}_K) \ni f \mapsto f|_{\mathbb{A}_{\mathbb{K}}^*}. \tag{4.12}$$

We denote by $C_\rho\left(\mathbb{A}_{\mathbb{K}}^*\right) \subset C\left(\mathbb{A}_{\mathbb{K}}^*\right)$ the range of ρ.

Corollary 4.4. *The restriction map ρ of (4.12) extends to an algebra homomorphism*

$$\rho : \mathcal{S}(\mathcal{G}_{\mathbb{K}}) \to C_\rho\left(\mathbb{A}_{\mathbb{K}}^*\right) \rtimes \mathbb{K}^*. \tag{4.13}$$

Proof. The map (4.12) induced by the inclusion $\mathbb{A}_{\mathbb{K}}^* \subset \mathbb{A}_{\mathbb{K}}$ is continuous and $\mathbb{K}^*$ equivariant; hence the map

$$\rho\left(\sum_{k\in\mathbb{K}^*} f_k\, U_k\right) = \sum_{k\in\mathbb{K}^*} \rho(f_k)\, U_k$$

is an algebra homomorphism. □

The action of $\mathbb{K}^*$ on $\mathbb{A}_{\mathbb{K}}^*$ is free and proper, so that we have an equivalence of the locally compact groupoids $\mathbb{K}^* \ltimes \mathbb{A}_{\mathbb{K}}^*$ and $\mathbb{A}_{\mathbb{K}}^*/\mathbb{K}^* = C_{\mathbb{K}}$. We use the exact sequence of locally compact groups

$$1 \to \mathbb{K}^* \to \mathbb{A}_{\mathbb{K}}^* \xrightarrow{p} C_{\mathbb{K}} \to 1 \tag{4.14}$$

to parameterize the orbits of $\mathbb{K}^*$ as the fibers $p^{-1}(x)$ for $x \in C_{\mathbb{K}}$. By construction the Hilbert spaces

$$\mathcal{H}_x = \ell^2(p^{-1}(x)), \quad \forall x \in C_K, \tag{4.15}$$

form a continuous field of Hilbert spaces over $C_{\mathbb{K}}$. We let $\mathcal{L}^1(\mathcal{H}_x)$ be the Banach algebra of trace class operators in $\mathcal{H}_x$; these form a continuous field over $C_{\mathbb{K}}$.

Proposition 4.5. *The restriction map* ρ *of* (4.12) *extends to an algebra homomorphism*

$$\rho : \mathcal{S}(\mathcal{G}_{\mathbb{K}}) \to C(C_{\mathbb{K}}, \mathcal{L}^1(\mathcal{H}_x)). \tag{4.16}$$

Proof. Each $p^{-1}(x)$ is globally invariant under the action of $\mathbb{K}^*$, so the crossed product rules in $C_\rho\left(\mathbb{A}_{\mathbb{K}}^*\right) \rtimes \mathbb{K}^*$ are just multiplication of operators in $\mathcal{H}_x$. To show that the obtained operators are in $\mathcal{L}^1$ we just need to consider monomials $f_k\, U_k$. In that case the only nonzero matrix elements correspond to $k = xy^{-1}$. It is enough to show that for any $f \in \mathcal{S}(\mathbb{A}_{\mathbb{K}})$, the function $k \mapsto f(k\, b)$ is summable. This follows from the discreteness of $b\,\mathbb{K} \subset \mathbb{A}_{\mathbb{K}}$ and the construction of the Bruhat–Schwartz space $\mathcal{S}(\mathbb{A}_{\mathbb{K}})$, cf. [10]. In fact, the associated operator is of finite rank when f has compact support. In general, what happens is that the sum will look like the sum over $\mathbb{Z}$ of the values $f(nb)$ of a Schwartz function f on $\mathbb{R}$. □

In general, the exact sequence (4.14) does not split, and one does not have a natural $C_{\mathbb{K}}$-equivariant trivialization of the continuous field $\mathcal{H}_x$. Thus it is important in the general case to retain the nuance between the algebras $C(C_{\mathbb{K}}, \mathcal{L}^1(\mathcal{H}_x))$ and $C(C_{\mathbb{K}})$. We shall first deal with the special case $\mathbb{K} = \mathbb{Q}$, in which this issue does not arise.

4.3 The Morita equivalence and cokernel for $\mathbb{K} = \mathbb{Q}$

The exact sequence (4.14) splits for $\mathbb{K} = \mathbb{Q}$ and admits a natural continuous section that corresponds to the open and closed fundamental domain $\Delta_{\mathbb{Q}} = \hat{\mathbb{Z}}^* \times \mathbb{R}_+^* \subset \mathbb{A}_{\mathbb{Q}}^*$ for the action of $\mathbb{Q}^*$ on idèles. This allows us to construct a cyclic morphism between the cyclic module associated, respectively, to the algebra $C_\rho\left(\mathbb{A}_{\mathbb{Q}}^*\right) \rtimes \mathbb{Q}^*$ and to a suitable algebra $C_\rho(C_{\mathbb{Q}})$ of functions on $C_{\mathbb{Q}}$.

Lemma 4.6. *The composition* $d_{\mathbb{Q}} \circ e_{\mathbb{Q}}$ *of the maps*

$$e_{\mathbb{Q}} : (k, hb) \mapsto (b, (k, h)) \quad \text{and} \quad d_{\mathbb{Q}}(k, h) = (kh, h) \tag{4.17}$$

with $b \in \Delta_{\mathbb{Q}}$ *and* $k, h \in \mathbb{Q}^*$ *gives an isomorphism of the locally compact groupoids*

$$\mathbb{Q}^* \ltimes \mathbb{A}_{\mathbb{Q}}^* \simeq \Delta_{\mathbb{Q}} \times \mathbb{Q}^* \times \mathbb{Q}^*. \tag{4.18}$$

Proof. The map $e_{\mathbb{Q}}$ realizes an isomorphism between the locally compact groupoids

$$\mathbb{Q}^* \ltimes \mathbb{A}_{\mathbb{Q}}^* \simeq \Delta_{\mathbb{Q}} \times (\mathbb{Q}^* \ltimes \mathbb{Q}^*),$$

where $\mathbb{Q}^* \ltimes \mathbb{Q}^*$ is the groupoid of the action of $\mathbb{Q}^*$ on itself by multiplication. The latter is isomorphic to the trivial groupoid $\mathbb{Q}^* \times \mathbb{Q}^*$ via the map $d_{\mathbb{Q}}$. □

We then have the following result.

Proposition 4.7. *The map*

$$\sum_{k\in\mathbb{Q}^*} f_k\, U_k \mapsto M_b(x,y) = f_{xy^{-1}}(x\,b), \tag{4.19}$$

for $x, y \in \mathbb{Q}^*$ with $k = xy^{-1}$ and $b \in \Delta_{\mathbb{Q}}$, defines an algebra homomorphism

$$C_\rho\left(\mathbb{A}^*_{\mathbb{Q}}\right) \rtimes \mathbb{Q}^* \to C(\Delta_{\mathbb{Q}}, M_\infty(\mathbb{C}))$$

to the algebra of matrix-valued functions on $\Delta_{\mathbb{Q}}$. For any $f \in \mathcal{S}(\mathcal{G}_{\mathbb{Q}})$ the element M_b obtained in this way is of trace class.

Proof. We use the groupoid isomorphism (4.17) to write $k = xy^{-1}$ and $khb = xb$, for $x = kh$ and $y = h$. The second statement follows from Proposition 4.5. $\square$

Let $\pi = M \circ \rho : \mathcal{S}(\mathcal{G}_{\mathbb{K}}) \to C(\Delta_{\mathbb{Q}}, M_\infty(\mathbb{C}))$ be the composition of the restriction map ρ of (4.13) with the algebra morphism (4.19). Since the trace Tr on $M_\infty(\mathbb{C})$ gives a cyclic morphism, one can use this to obtain a morphism of cyclic modules $(\mathrm{Tr}\circ\pi)^\natural$, which we now describe explicitly. We let, in the number field case,

$$\mathbf{S}\,(C_{\mathbb{K}}) = \bigcap\nolimits_{\beta\in\mathbb{R}} \mu^\beta \mathcal{S}(C_{\mathbb{K}}), \tag{4.20}$$

where $\mu \in C(C_{\mathbb{K}})$ is the module morphism from $C_{\mathbb{K}}$ to $\mathbb{R}^*_+$. In the function field case one can simply use for $\mathbf{S}\,(C_{\mathbb{K}})$ the Schwartz functions with compact support.

Proposition 4.8. *The map* $\mathrm{Tr}\circ\pi$ *defines a morphism* $(\mathrm{Tr}\circ\pi)^\natural$ *of cyclic modules from $\mathcal{S}(\mathcal{G}_{\mathbb{Q}})_0^\natural$ to the cyclic submodule $\mathbf{S}^\natural(C_{\mathbb{Q}}) \subset C(C_{\mathbb{Q}})^\natural$ whose elements are continuous functions whose restriction to the main diagonal belongs to $\mathbf{S}\,(C_{\mathbb{Q}})$.*

Proof. By Proposition 4.7 the map π is an algebra homomorphism from $\mathcal{S}(\mathcal{G}_{\mathbb{Q}})$ to $C(\Delta_{\mathbb{Q}}, \mathcal{L}^1) \sim C(C_{\mathbb{Q}}, \mathcal{L}^1)$. We need to show that the corresponding cyclic morphism using $\mathrm{Tr}^\natural$ lands in the cyclic submodule $S^\natural(C_{\mathbb{Q}})$.

For simplicity we can just restrict to the case of monomials, where we consider elements of the form

$$Z = f_{k_0}\, U_{k_0} \otimes f_{k_1}\, U_{k_1} \otimes \cdots \otimes f_{k_n}\, U_{k_n}. \tag{4.21}$$

The matrix-valued functions associated to the monomials $f_{k_j}\, U_{k_j}$ as in Proposition 4.7 have matrix elements at a point $b \in \Delta_{\mathbb{Q}}$ that are nonzero only for $x_{j+1} = x_j k_j^{-1}$ and are of the form

$$f_{k_j}\, U_{k_j} \mapsto M_b(x_j, x_{j+1}) = f_{k_j}(x_j b). \tag{4.22}$$

Composing with the cyclic morphism $\mathrm{Tr}^\natural$ gives

$$(\mathrm{Tr} \circ \pi)^\natural(Z)(b_0, b_1, \ldots, b_n) = \sum \prod M_{b_j}(x_j, x_{j+1}), \tag{4.23}$$

where the $x_j \in \mathbb{K}^*$ and $x_{n+1} = x_0$. Let $\gamma_0 = 1$ and $\gamma_{j+1} = k_j \gamma_j$. Then we find that $(\mathrm{Tr} \circ \pi)^\natural(Z) = 0$, unless $\prod_j k_j = 1$, i.e., $\gamma_{n+1} = 1$. In this case we obtain

$$\mathrm{Tr} \circ \pi(Z)(b_0, b_1, \ldots, b_n) = \sum_{k \in \mathbb{Q}^*} \prod_{j=0}^{n} f_{k_j}\left(\gamma_j^{-1} k b_j\right), \quad \forall b_j \in \Delta_{\mathbb{Q}}. \tag{4.24}$$

For $n = 0$ the formula (4.24) reduces to

$$\mathrm{Tr} \circ \pi(f)(b) = \sum_{k \in \mathbb{Q}^*} f(kb), \quad \forall b \in \Delta_{\mathbb{Q}}, \quad \forall f \in \mathcal{S}(\mathbb{A}_{\mathbb{Q}})_0, \tag{4.25}$$

where $\mathcal{S}(\mathbb{A}_{\mathbb{Q}})_0 = \mathrm{Ker}\, \varepsilon_0 \cap \mathrm{Ker}\, \varepsilon_1 \subset \mathcal{S}(\mathbb{A}_{\mathbb{Q}})$. This gives an element of $\mathbf{S}\,(C_{\mathbb{Q}})$, by [10, Lemma 2 Appendix 1]. In general, (4.24) gives a continuous function of $n+1$ variables on $C_{\mathbb{Q}}$, and its restriction to the main diagonal belongs to $\mathbf{S}\,(C_{\mathbb{Q}})$. ☐

Since the category of cyclic modules is an abelian category, we can consider the cokernel in the category of Λ-modules of the cyclic morphism $(\mathrm{Tr} \circ \pi)^\natural$, with π the composite of (4.13) and (4.19). This works nicely for $\mathbb{K} = \mathbb{Q}$ but makes use of the splitting of the exact sequence (4.14).

4.4 The cokernel of ρ for general global fields

To handle the general case in a canonical manner one just needs to work directly with $C(C_{\mathbb{K}}, \mathcal{L}^1(\mathcal{H}_x))$ instead of $C(C_{\mathbb{K}})$ and express at that level the decay condition of the restrictions to the diagonal in the cyclic submodule $\mathbf{S}^{\,\natural}(C_{\mathbb{Q}})$ of Proposition 4.8.

Definition 4.9. *We define $\mathbf{S}^{\,\natural}(C_{\mathbb{K}}, \mathcal{L}^1(\mathcal{H}_x))$ to be the cyclic submodule of the cyclic module $C(C_{\mathbb{K}}, \mathcal{L}^1(\mathcal{H}_x))^\natural$ whose elements are continuous functions such that the trace of the restriction to the main diagonal belongs to $\mathbf{S}\,(C_{\mathbb{K}})$.*

Note that for $T \in C(C_{\mathbb{K}}, \mathcal{L}^1(\mathcal{H}_x))^\natural$ of degree n, $T(x_0, \ldots, x_n)$ is an operator in $\mathcal{H}_{x_0} \otimes \cdots \otimes \mathcal{H}_{x_n}$. On the diagonal, $x_j = x$ for all j, the trace map corresponding to $\mathrm{Tr}^\natural$ is given by

$$\mathrm{Tr}^\natural(T_0 \otimes T_1 \otimes \cdots \otimes T_n) = \mathrm{Tr}(T_0\, T_1\, \cdots\, T_n)\,. \tag{4.26}$$

This makes sense, since on the diagonal all the Hilbert spaces $\mathcal{H}_{x_j}$ are the same.

The argument of Proposition 4.8 extends to the general case and shows that the cyclic morphism $\rho^\natural$ of the restriction map ρ lands in $\mathbf{S}^{\,\natural}(C_{\mathbb{K}}, \mathcal{L}^1(\mathcal{H}_x))$.

Definition 4.10. *We define $\mathcal{H}^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$ to be the cokernel of the cyclic morphism*

$$\rho^\natural \;:\; \mathcal{S}(\mathcal{G}_{\mathbb{K}})_0^\natural \to \mathbf{S}^{\,\natural}(C_{\mathbb{K}}, \mathcal{L}^1(\mathcal{H}_x)).$$

Moreover, an important issue arises, since the ranges of continuous linear maps are not necessarily closed subspaces. In order to preserve the duality between cyclic homology and cyclic cohomology we shall define the cokernel of a cyclic map $T : \mathcal{A}^\natural \to \mathcal{B}^\natural$ as the quotient of $\mathcal{B}^\natural$ by the closure of the range of T. In a dual manner, the kernel of the transposed map $T^t : \mathcal{B}^\sharp \to \mathcal{A}^\sharp$ is automatically closed and is the dual of the above.

The choice of the notation $\mathcal{H}^1(\mathbb{A}_\mathbb{K}/\mathbb{K}^*, C_\mathbb{K})$ is explained by the fact that we consider this a first cohomology group, in the sense that it is a cokernel in a sequence of cyclic homology groups for the inclusion of the idèles class group in the adèles class space (dually for the restriction map of algebras); hence we can think of it as giving rise to an H^1 in the relative cohomology sequence of an inclusion of $C_\mathbb{K}$ in the noncommutative space $\mathbb{A}_\mathbb{K}/\mathbb{K}^*$. We can use the result of [7], describing the cyclic (co)homology in terms of derived functors in the category of cyclic modules, to write the cyclic homology as

$$HC_n(\mathcal{A}) = \mathrm{Tor}_n(\mathbb{C}^\natural, \mathcal{A}^\natural). \tag{4.27}$$

Thus, we obtain a cohomological realization of the cyclic module $\mathcal{H}^1(\mathbb{A}_\mathbb{K}/\mathbb{K}^*, C_\mathbb{K})$ by setting

$$H^1(\mathbb{A}_\mathbb{K}/\mathbb{K}^*, C_\mathbb{K}) := \mathrm{Tor}(\mathbb{C}^\natural, \mathcal{H}^1(\mathbb{A}_\mathbb{K}/\mathbb{K}^*, C_\mathbb{K})). \tag{4.28}$$

We think of this as an H^1 because of its role as a relative term in a cohomology exact sequence of the pair $(\mathbb{A}_\mathbb{K}/\mathbb{K}^*, C_\mathbb{K})$.

We now show that $H^1(\mathbb{A}_\mathbb{K}/\mathbb{K}^*, C_\mathbb{K})$ carries an action of $C_\mathbb{K}$, which we can view as the abelianization $W_\mathbb{K}^{\mathrm{ab}} \sim C_\mathbb{K}$ of the Weil group. This action is induced by the multiplicative action of $C_\mathbb{K}$ on $\mathbb{A}_\mathbb{K}/\mathbb{K}^*$ and on itself. This generalizes to global fields the action of $C_\mathbb{Q} = \hat{\mathbb{Z}}^* \times \mathbb{R}_+^*$ on $HC_0(D(\mathcal{A}, \varphi))$ for the Bost–Connes endomotive (cf. [11]).

Proposition 4.11. *The cyclic modules $\mathcal{S}(\mathcal{G}_\mathbb{K})_0^\natural$ and $\mathbf{S}^\natural(C_\mathbb{K}, \mathcal{L}^1(\mathcal{H}_x))$ are endowed with an action of $\mathbb{A}_\mathbb{K}^*$, and the morphism $\rho^\natural$ is $\mathbb{A}_\mathbb{K}^*$-equivariant. This induces an action of $C_\mathbb{K}$ on $H^1(\mathbb{A}_\mathbb{K}/\mathbb{K}^*, C_\mathbb{K})$.*

Proof. For $\gamma \in \mathbb{A}_\mathbb{K}^*$ one defines an action by automorphisms of the algebra $\mathcal{A} = \mathcal{S}(\mathcal{G}_\mathbb{K})$ by setting

$$\vartheta_a(\gamma)(f)(x) := f(\gamma^{-1}x), \quad \text{for } f \in \mathcal{S}(\mathbb{A}_\mathbb{K}), \tag{4.29}$$

$$\vartheta_a(\gamma)(\sum_{k\in\mathbb{K}^*} f_k\, U_k) := \sum_{k\in\mathbb{K}^*} \vartheta_a(\gamma)(f_k)\, U_k\,. \tag{4.30}$$

This action is inner for $\gamma \in \mathbb{K}^*$ and induces an outer action

$$C_\mathbb{K} \to \mathrm{Out}(\mathcal{S}(\mathcal{G}_\mathbb{K}))\,. \tag{4.31}$$

Similarly, the continuous field $\mathcal{H}_x = \ell^2(p^{-1}(x))$ over $C_\mathbb{K}$ is $\mathbb{A}_\mathbb{K}^*$-equivariant for the action of $\mathbb{A}_\mathbb{K}^*$ on $C_\mathbb{K}$ by translations, and the equality

$$(V(\gamma)\xi)(y) := \xi(\gamma^{-1}\, y)\,, \quad \forall y \in p^{-1}(\gamma x)\,,\ \xi \in \ell^2(p^{-1}(x))\,, \tag{4.32}$$

defines an isomorphism $\mathcal{H}_x \xrightarrow{V(\gamma)} \mathcal{H}_{\gamma x}$. One then obtains an action of $\mathbb{A}_{\mathbb{K}}^*$ on $C(C_{\mathbb{K}}, \mathcal{L}^1(\mathcal{H}_x))$ by setting

$$\vartheta_m(\gamma)(f)(x) := V(\gamma)\, f(\gamma^{-1}\, x)\, V(\gamma^{-1}), \quad \forall f \in C(C_{\mathbb{K}}, \mathcal{L}^1(\mathcal{H}_x))\,. \tag{4.33}$$

The morphism ρ is $\mathbb{A}_{\mathbb{K}}^*$-equivariant, so that one obtains an induced action on the cokernel $\mathcal{H}^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$. This action is inner for $\gamma \in \mathbb{K}^*$ and thus induces an action of $C_{\mathbb{K}}$ on $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$. $\square$

We denote by

$$C_{\mathbb{K}} \ni \gamma \mapsto \underline{\vartheta}_m(\gamma) \tag{4.34}$$

the induced action on $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$.

We have a noncanonical isomorphism

$$C_{\mathbb{K}} \simeq C_{\mathbb{K},1} \times N, \tag{4.35}$$

where $N \subset \mathbb{R}_+^*$ is the range of the norm $|\cdot| : C_{\mathbb{K}} \to \mathbb{R}_+^*$. For number fields this is $N = \mathbb{R}_+^*$, while for function fields in positive characteristic $N \simeq \mathbb{Z}$ is the subgroup $q^{\mathbb{Z}} \subset \mathbb{R}_+^*$ with $q = p^\ell$ the cardinality of the field of constants. We denote by $\widehat{C_{\mathbb{K},1}}$ the group of characters of the compact subgroup $C_{\mathbb{K},1} \subset C_{\mathbb{K}}$, i.e., the Pontryagin dual of $C_{\mathbb{K},1}$. Given a character χ of $C_{\mathbb{K},1}$, we let $\tilde{\chi}$ denote the unique extension of χ to $C_{\mathbb{K}}$ that is equal to one on N.

One obtains a decomposition of $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$ according to projectors associated to characters of $C_{\mathbb{K},1}$.

Proposition 4.12. *Characters $\chi \in \widehat{C_{\mathbb{K},1}}$ determine a canonical direct-sum decomposition*

$$\begin{aligned} H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}}) &= \bigoplus\nolimits_{\chi \in \widehat{C_{\mathbb{K},1}}} H^1_\chi(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}}), \\ H^1_\chi(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}}) &= \{\xi |\, \underline{\vartheta}_m(\gamma)\, \xi = \chi(\gamma)\, \xi,\, \forall \gamma \in C_{\mathbb{K},1}\}, \end{aligned} \tag{4.36}$$

where $\underline{\vartheta}_m(\gamma)$ denotes the induced action (4.34) on $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$.

Proof. The action of $\mathbb{A}_{\mathbb{K}}^*$ on $\mathcal{H}^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$ induces a corresponding action of $C_{\mathbb{K}}$ on $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$. $\square$

We can then reformulate the result of [11] based on the trace formula of [10] in the formulation of [33] in terms of the cohomology $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$ in the following way.

Proposition 4.13. *The induced representation of $C_{\mathbb{K}}$ on $H^1_\chi(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$ gives the spectral realization of the zeros of the L-function with Grössencharakter χ.*

This result is a variant of [10, Corollary 2]; the proof is similar and essentially reduces to the result of [33]. There is a crucial difference from [10] in that all zeros (including those not located on the critical line) now appear due to the choice of the function spaces. To see what happens it is simpler to deal with the dual spaces, i.e., to compute the cyclic cohomology HC^0. Its elements are cyclic morphisms T from $\mathcal{H}^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$ to $\mathbb{C}^\natural$ and they are determined by the map T^0 in degree 0. The cyclic morphism property then shows that T^0 defines a trace on $\mathbf{S}^\natural(C_{\mathbb{K}}, \mathcal{L}^1(\mathcal{H}_x))$ which vanishes on the range of $\rho^\natural$. The freeness of the action of $\mathbb{K}^*$ on $\mathbb{A}_{\mathbb{K}}^*$ then ensures that these traces are given by continuous linear forms on $\mathbf{S}(C_{\mathbb{K}})$ that vanish on the following subspace of $\mathbf{S}(C_{\mathbb{K}})$, which is the range of the restriction map, defined as follows.

Definition 4.14. *Let $\mathcal{V} \subset \mathbf{S}(C_{\mathbb{K}})$ denote the range of the map $\mathrm{Tr} \circ \rho$, that is,*

$$\mathcal{V} = \{h \in \mathbf{S}(C_{\mathbb{K}}) | \, h(x) = \sum_{k \in \mathbb{K}^*} \xi(kx), \ \text{with } \xi \in \mathcal{S}(\mathbb{A}_{\mathbb{K}})_0\}, \tag{4.37}$$

where $\mathcal{S}(\mathbb{A}_{\mathbb{K}})_0 = \mathrm{Ker}\ \varepsilon_0 \cap \mathrm{Ker}\ \varepsilon_1 \subset \mathcal{S}(\mathbb{A}_{\mathbb{K}})$.

We have seen above in the case $\mathbb{K} = \mathbb{Q}$ (cf. [10]) that the range of $\mathrm{Tr} \circ \rho$ is indeed contained in $\mathbf{S}(C_{\mathbb{K}})$.

Moreover, we have the following results about the action $\underline{\vartheta}_m(\gamma)$, for $\gamma \in C_{\mathbb{K}}$, on $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$. Suppose we are given $f \in \mathbf{S}(C_{\mathbb{K}})$. We define a corresponding operator

$$\underline{\vartheta}_m(f) = \int_{C_{\mathbb{K}}} f(\gamma)\, \underline{\vartheta}_m(\gamma)\, d^*\gamma, \tag{4.38}$$

acting on the complex vector space $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$. Here $d^*\gamma$ is the multiplicative Haar measure on $C_{\mathbb{K}}$. We have the following description of the action of $\underline{\vartheta}_m(f)$.

Lemma 4.15. *For $f \in \mathbf{S}(C_{\mathbb{K}})$, the action of the operator $\underline{\vartheta}_m(f)$ of (4.38) on $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$ is the action induced on the quotient of $\mathbf{S}(C_{\mathbb{K}})$ by $\mathcal{V} \subset \mathbf{S}(C_{\mathbb{K}})$ of the action of $\vartheta_m(f)$ on $\mathbf{S}(C_{\mathbb{K}})$ by convolution product*

$$\vartheta_m(f)\xi(u) = \int_{C_{\mathbb{K}}} \xi(g^{-1}u) f(g)\, d^*g = (f \star \xi)(u). \tag{4.39}$$

Proof. One first shows that one can lift f to a function $\tilde{f}$ on $\mathbb{A}_{\mathbb{K}}^*$ such that

$$\sum_{k \in \mathbb{K}^*} \tilde{f}(kx) = f(x)$$

and that convolution by $\tilde{f}$, i.e.,

$$\int \tilde{f}(\gamma)\vartheta_a(\gamma) d^*\gamma,$$

leaves $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$ globally invariant. This means showing that that $\mathcal{S}(\mathbb{A}_{\mathbb{K}})_0$ is stable under convolution by the lift of $\mathbf{S}\,(C_{\mathbb{K}})$. Then (4.39) follows directly from the definition of the actions (4.33), (4.30), (4.34) and the operator (4.38). $\square$

For $f \in \mathbf{S}\,(C_{\mathbb{K}})$, $\tilde{\chi}$ the extension of a character $\chi \in \widehat{C_{\mathbb{K},1}}$ to $C_{\mathbb{K}}$, and $\hat{f}(\tilde{\chi},\rho)$ the Fourier transform (2.30), the operators $\underline{\vartheta}_m(f)$ of (4.38) satisfy the spectral side of the trace formula. Namely, we have the following result.

Theorem 4.16. *For any $f \in \mathbf{S}\,(C_{\mathbb{K}})$, the operator $\underline{\vartheta}_m(f)$ defined in* (4.38) *acting on $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$ is of trace class. The trace is given by*

$$\mathrm{Tr}(\underline{\vartheta}_m(f)|H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})) = \sum_{\rho\in\mathbb{C}/N^{\perp}\,|\,L(\tilde{\chi},\rho)=0} \hat{f}(\tilde{\chi},\rho), \tag{4.40}$$

with $\hat{f}(\tilde{\chi},\rho)$ the Fourier transform (2.30)*.*

Proof. Due to the different normalization of the summation map, the representation $\underline{\vartheta}_m(\gamma)$ considered here differs from the action $W(\gamma)$ considered in [10] by

$$\underline{\vartheta}_m(\gamma) = |\gamma|^{1/2}\, W(\gamma). \tag{4.41}$$

This means that we have

$$\underline{\vartheta}_m(f) = \int_{C_{\mathbb{K}}} f(\gamma)\,\underline{\vartheta}_m(\gamma)\, d^*\gamma = \int_{C_{\mathbb{K}}} h(\gamma)\, W(\gamma)\, d^*\gamma, \tag{4.42}$$

where

$$h(\gamma) = \,|\gamma|^{1/2}\, f(\gamma). \tag{4.43}$$

We then have, for $W(h) = \int_{C_{\mathbb{K}}} h(\gamma)\, W(\gamma)\, d^*\gamma$,

$$\mathrm{Tr}\ W(h) = \sum_{\rho\in\mathbb{C}/N^{\perp}\,|\,L(\tilde{\chi},\frac{1}{2}+\rho)=0} \hat{h}(\tilde{\chi},\rho). \tag{4.44}$$

Note that in contrast to [10], all zeros contribute, including those that might fail to be on the critical line, and they do so with their natural multiplicity. This follows from the choice of function space as in [33]. The Fourier transform $\hat{h}(\tilde{\chi},\rho)$ satisfies

$$\hat{h}(\tilde{\chi},\rho) = \int_{C_{\mathbb{K}}} h(u)\tilde{\chi}(u)\,|u|^{\rho}\, d^*u = \int_{C_{\mathbb{K}}} f(u)\tilde{\chi}(u)\,|u|^{\rho+1/2}\, d^*u = \hat{f}(\tilde{\chi},\rho+1/2), \tag{4.45}$$

where h and f are related as in (4.43). Thus, the shift by $1/2$ in (4.44) is absorbed in (4.45) and this gives the required formula (4.40).

4.5 Trace pairing and vanishing

The commutativity of the convolution product implies the following vanishing result.

Lemma 4.17. *Suppose we are given an element $f \in \mathcal{V} \subset \mathbf{S}(C_{\mathbb{K}})$, where $\mathcal{V}$ is the range of the reduction map as in Definition 4.14. Then one has*

$$\underline{\vartheta}_m(f)|_{H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})} = 0. \tag{4.46}$$

Proof. The result follows by showing that for $f \in \mathcal{V}$, the operator $\underline{\vartheta}_m(f)$ maps any element $\xi \in \mathbf{S}(C_{\mathbb{K}})$ to an element in $\mathcal{V}$; hence the induced map on the quotient of $\mathbf{S}(C_{\mathbb{K}})$ by $\mathcal{V}$ is trivial. Since $\mathcal{V}$ is a submodule of $\mathbf{S}(C_{\mathbb{K}})$, for the action of $\mathbf{S}(C_{\mathbb{K}})$ by convolution we obtain

$$\underline{\vartheta}_m(f)\xi = f \star \xi = \xi \star f \in \mathcal{V},$$

where $\star$ is the convolution product of (4.39). □

This makes it possible to define a trace pairing as follows.

Remark 4.18. The pairing

$$f_1 \otimes f_2 \mapsto \langle f_1, f_2 \rangle_{H^1} := \mathrm{Tr}(\underline{\vartheta}_m(f_1 \star f_2)|H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})) \tag{4.47}$$

descends to a well-defined pairing on $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}}) \otimes H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$.

5 Primitive cohomology

The aim of this section is to interpret the motivic construction described in the previous section as the noncommutative version of a classical construction in algebraic geometry. In motive theory, realizations of (mixed) motives appear frequently in the form of kernels/cokernels of relevant homomorphisms. The *primitive cohomology* is the example we shall review hereafter.

If Y is a compact Kähler variety, a Kähler cocycle class $[\omega] \in H^2(Y, \mathbb{R})$ determines the Lefschetz operator ($i \in \mathbb{Z}_{\geq 0}$):

$$L : H^i(Y, \mathbb{R}) \to H^{i+2}(Y, \mathbb{R}), \quad L(a) := [\omega] \cup a.$$

Let $n = \dim Y$. Then the *primitive cohomology* is defined as the kernel of iterated powers of the Lefschetz operator

$$H^i(Y, \mathbb{R})_{\mathrm{prim}} := \mathrm{Ker}(L^{n-i+1} : H^i(Y, \mathbb{R}) \to H^{2n-i+2}(Y, \mathbb{R})).$$

In particular, for $i = n$ we have

$$H^n(Y, \mathbb{R})_{\mathrm{prim}} := \mathrm{Ker}(L : H^n(Y, \mathbb{R}) \to H^{n+2}(Y, \mathbb{R})).$$

Let us assume, from now on, that $j : Y \hookrightarrow X$ is a smooth hyperplane section of a smooth projective complex algebraic variety X. Then it is a classical result of geometric topology that $L = j^* \circ j_*$, where

$$j_* : H^i(Y, \mathbb{R}) \to H^{i+2}(X, \mathbb{R})$$

is the Gysin homomorphism, that is, the Poincaré dual of the restriction homomorphism

$$j^* : H^{2n-i}(X, \mathbb{R}) \to H^{2n-i}(Y, \mathbb{R}).$$

In fact, because the class of L comes from an integral class, the equality $L = j^* \circ j_*$ holds already in integral cohomology. For $i = n$, the above description of the Lefschetz operator together with the Lefschetz theorem of hyperplane sections implies that

$$H^n(Y, \mathbb{R})_{\mathrm{prim}} \cong \mathrm{Ker}(j_* : H^n(Y, \mathbb{R}) \to H^{n+2}(X, \mathbb{R})) =: H^n(Y, \mathbb{R})_{\mathrm{van}},$$

where by $H^i(Y, \mathbb{R})_{\mathrm{van}}$ we denote the *vanishing cohomology*

$$H^i(Y, \mathbb{R})_{\mathrm{van}} := \mathrm{Ker}(j_* : H^i(Y, \mathbb{R}) \to H^{i+2}(X, \mathbb{R})).$$

Now we introduce the theory of mixed Hodge structures in this setup.

Let $U := X \smallsetminus Y$ be the open space that is the complement of Y in X and let us denote by $k : U \hookrightarrow X$ the corresponding open immersion. Then one knows that $R^i j_* \mathbb{Z} = 0$ unless $i = 0, 1$, so that the Leray spectral sequence for j,

$$E_2^{p,q} = H^q(X, \mathrm{R}^p k_* \mathbb{Z}) \Rightarrow H^{p+q}(U, \mathbb{Z}),$$

coincides with the long exact sequence (of mixed Hodge structures)

$$\cdots \xrightarrow{\partial} H^{i-2}(Y, \mathbb{Z})(-1) \xrightarrow{j_*} H^i(X, \mathbb{Z}) \to H^i(U, \mathbb{Z}) \xrightarrow{\partial} \cdots .$$

The boundary homomorphism ∂ in this sequence is known to coincide [21, §9.2] with the residue homomorphism

$$\mathrm{Res} : H^{i+1}(U, \mathbb{Z}) \to H^i(Y, \mathbb{Z})(-1),$$

whose description, with complex coefficients, is derived from a corresponding morphism of filtered complexes (Poincaré residue map). This morphism fits into the following exact sequence of filtered complexes of Hodge modules:

$$0 \to \Omega_X^\bullet \to \Omega_X^\bullet(\log Y) \xrightarrow{\mathrm{res}} j_* \Omega_Y^\bullet[-1] \to 0,$$

$$\mathrm{res}\left(\alpha \wedge \frac{dt}{t}\right) = \alpha_{|Y}.$$

One knows that Res is a homomorphism of Hodge structures; hence the Hodge filtration on $H^{n+1}(U, \mathbb{C}) \cong \mathbb{H}^{n+1}\left(X, \Omega_X^\bullet(\log Y)\right)$ determines a corresponding filtration on the (twisted) vanishing cohomology

$$H^n(Y,\mathbb{C})(n)_{\mathrm{van}} = \mathrm{Ker}(j_* : H^n(Y,\mathbb{C})(n) \to H^{n+2}(X,\mathbb{C})(n+1)) \\ \cong H^{n+1}(U)(n+1).$$

In degree $i = n$, one also knows that the excision exact sequence (of Hodge structures) becomes the short exact sequence

$$0 \to H^n(X,\mathbb{C}) \xrightarrow{j^*} H^n(Y,\mathbb{C}) \to H_c^{n+1}(U,\mathbb{C}) \to 0.$$

Therefore, it follows by the Poincaré duality isomorphism

$$H_c^{n+1}(U,\mathbb{C})^* \cong H^{n+1}(U,\mathbb{C})(n+1)$$

that

$$(\mathrm{Coker}(j^* : H^n(X,\mathbb{C}) \to H^n(Y,\mathbb{C})))^* \cong H_c^{n+1}(U,\mathbb{C})^* \cong H^n(Y,\mathbb{C})(n)_{\mathrm{van}}. \tag{5.1}$$

When $j : Y \hookrightarrow X$ is a singular hypersurface or a divisor in X with (local) normal crossings (i.e., $Y = \bigcup_i Y_i$, $\dim Y_i = n = \dim X - 1$, Y locally described by an equation $x_{i_1} \cdots x_{i_r} = 0$, $\{i_1,\ldots,i_r\} \subseteq \{1,\ldots,n+1\}$, $\{x_1,\ldots,x_{n+1}\} =$ system of local coordinates in X), the notion of the Gysin homomorphism is lost. One then replaces the vanishing cohomology by the primitive cohomology, whose definition extends to this general setup and is given, in analogy to (5.1), as

$$H^n(Y,\mathbb{C})_{\mathrm{prim}} := \mathrm{Coker}(j^* : H^n(X,\mathbb{C}) \to H^n(Y,\mathbb{C})) \subseteq H_c^{n+1}(U,\mathbb{C}).$$

One also knows that the primitive cohomology is motivic (cf. [22] and [3] for interesting examples). Following the classical construction that we have just reviewed, we would like to argue now that the definition of the cyclic module $\mathcal{H}^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$ (as in Definition 4.10), which is based on a noncommutative version of a restriction map "from adèles to idèles" defined in the category of Λ-modules, should be interpreted as the noncommutative analogue of a primitive motive (a cyclic primitive module). The cohomological realization of such a motive (i.e., its cyclic homology) is given by the group $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}}) = \mathrm{Tor}(\mathbb{C}^\natural, \mathcal{H}^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}}))$ (cf. (4.28)), which therefore can be interpreted as a noncommutative version of a primitive cohomology.

6 A cohomological Lefschetz trace formula

6.1 Weil's explicit formula as a trace formula

As in Section 2.2 above, let α be a nontrivial character of $\mathbb{A}_{\mathbb{K}}$ that is trivial on $\mathbb{K} \subset \mathbb{A}_{\mathbb{K}}$. It is well known ([44] VII-2) that for such a character α there exists a differental idèle $a = (a_v) \in \mathbb{A}_{\mathbb{K}}^*$ such that

$$\alpha_v(x) = e_{\mathbb{K}_v}(a_v\, x), \quad \forall x \in \mathbb{K}_v, \tag{6.1}$$

where, for a local field K, the additive character e_K is chosen in the following way:

- If $K = \mathbb{R}$ then $e_{\mathbb{R}}(x) = e^{-2\pi i x}$, for all $x \in \mathbb{R}$.
- If $K = \mathbb{C}$ then $e_{\mathbb{C}}(z) = e^{-2\pi i(z+\bar{z})}$, for all $z \in \mathbb{C}$.
- If K is a non-archimedean local field with maximal compact subring $\mathcal{O}$, then the character e_K satisfies $\operatorname{Ker} e_K = \mathcal{O}$.

The notion of differental idèle can be thought of as an extension of the canonical class of the algebraic curve C, from the setting of function fields $\mathbb{F}_q(C)$ to arbitrary global fields $\mathbb{K}$. For instance, one has

$$|a| = q^{2-2g} \qquad \text{or} \qquad |a| = D^{-1}, \tag{6.2}$$

respectively, for the case of a function field $\mathbb{F}_q(C)$ and of a number field. In the number field case, D denotes the discriminant.

In [11] we gave a cohomological formulation of the Lefschetz trace formula of [10], using the version of the Riemann–Weil explicit formula as a trace formula given in [33] in the context of nuclear spaces, rather than the semilocal Hilbert space version of [10].

Theorem 6.1. *For $f \in \mathbf{S}(C_{\mathbb{K}})$ let $\underline{\vartheta}_m(f)$ be the operator (4.38) acting on the space $H^1 = H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$. Then the trace is given by*

$$\operatorname{Tr}(\underline{\vartheta}_m(f)|H^1) = \hat{f}(0)+\hat{f}(1)-(\log|a|)\,f(1)-\sum_{v\in\Sigma_{\mathbb{K}}}\int'_{(\mathbb{K}_v^*,e_{\mathbb{K}_v})}\frac{f(u^{-1})}{|1-u|}\,d^*u. \tag{6.3}$$

The formula (6.3) is obtained in [11] by first showing that the Lefschetz trace formula of [10] in the version of [33] can be formulated equivalently in the form

$$\operatorname{Tr}(\underline{\vartheta}_m(f)|H^1) = \hat{f}(0) + \hat{f}(1) - \sum_{v\in\mathbb{K}_v}\int'_{\mathbb{K}_v^*}\frac{f(u^{-1})}{|1-u|}\,d^*u, \tag{6.4}$$

where one uses the global character α to fix the local normalizations of the principal values in the last term of the formula. We then compute this principal value using the differental idèle in the form

$$\int'_{(\mathbb{K}_v^*,\alpha_v)}\frac{f(u^{-1})}{|1-u|}\,d^*u = (\log|a_v|)\,f(1) + \int'_{(\mathbb{K}_v^*,e_{\mathbb{K}_v})}\frac{f(u^{-1})}{|1-u|}\,d^*u. \tag{6.5}$$

6.2 Weil positivity and the Riemann hypothesis

We introduce an involution for elements $f \in \mathbf{S}(C_{\mathbb{K}})$ by setting

$$f^*(g) = \overline{f(g^{-1})}. \tag{6.6}$$

We also consider a one parameter group $z \mapsto \Delta^z$ of automorphisms of the convolution algebra $\mathbf{S}(C_{\mathbb{K}})$ with the convolution product (4.39) by setting

$$\Delta^z(f)(g) = |g|^z f(g), \tag{6.7}$$

for $f \in \mathbf{S}(C_{\mathbb{K}})$ and $z \in \mathbb{C}$. Since (6.7) is given by multiplication by a character, it satisfies

$$\Delta^z(f \star h) = \Delta^z(f) \star \Delta^z(f), \quad \forall f, h \in \mathbf{S}(C_{\mathbb{K}}). \tag{6.8}$$

We consider also the involution

$$f \mapsto f^\sharp = \Delta^{-1} f^*, \quad \text{with} \quad f^\sharp(g) = |g|^{-1}\overline{f(g^{-1})}. \tag{6.9}$$

The reformulation, originally due to A. Weil, of the Riemann hypothesis in our setting is given by the following statement.

Proposition 6.2. *The following two conditions are equivalent:*

- *All L-functions with Grössencharakter on $\mathbb{K}$ satisfy the Riemann hypothesis.*
- *The trace pairing (4.47) satisfies the positivity condition*

$$\langle \Delta^{-1/2} f, \Delta^{-1/2} f^* \rangle \geq 0, \quad \forall f \in \mathbf{S}(C_{\mathbb{K}}). \tag{6.10}$$

Proof. Let $W(\gamma) = |\gamma|^{-1/2} \underline{\vartheta}_m(\gamma)$. Then by [42], the RH for L-functions with Grössencharakter on $\mathbb{K}$ is equivalent to the positivity

$$\mathrm{Tr}(W(f \star f^*)) \geq 0, \quad \forall f \in \mathbf{S}(C_{\mathbb{K}}). \tag{6.11}$$

Thus, in terms of the representation $\underline{\vartheta}_m$ we are considering here, we have

$$W(f) = \underline{\vartheta}_m(\Delta^{-1/2} f).$$

Using the multiplicative property (6.8) of Δ^z we rewrite (6.11) in the equivalent form (6.10). □

In terms of the involution (6.9) we can reformulate Proposition 6.2 in the following equivalent way.

Corollary 6.3. *The following conditions are equivalent:*

- *All L-functions with Grössencharakter on $\mathbb{K}$ satisfy the Riemann hypothesis.*
- *The trace pairing (4.47) satisfies $\langle f, f^\sharp \rangle \geq 0$, for all $f \in \mathbf{S}(C_{\mathbb{K}})$.*

Proof. In (6.10) we write $\Delta^{-1/2} f = h$. This gives

$$\Delta^{-1/2} f^* = \Delta^{-1/2}(\Delta^{1/2} h)^* = \Delta^{-1} h^* = h^\sharp,$$

and the result follows, since $\Delta^{-1/2}$ is an automorphism of $\mathbf{S}(C_{\mathbb{K}})$. □

The vanishing result of Lemma 4.17, for elements in the range $\mathcal{V} \subset \mathbf{S}\,(C_{\mathbb{K}})$ of the reduction map $\mathrm{Tr} \circ \rho$ from adèles, then gives the following result.

Proposition 6.4. *The elements $f \star f^\sharp$ considered in Corollary 6.3 above have the following properties:*

1. *The trace pairing $\langle f, f^\sharp \rangle$ vanishes for all $f \in \mathcal{V}$, i.e., when f is the restriction $\mathrm{Tr} \circ \rho$ of an element of $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$.*
2. *By adding elements of $\mathcal{V}$ one can make the values*

$$f \star f^\sharp(1) = \int_{C_{\mathbb{K}}} |f(g)|^2 \, |g| \, d^*g \tag{6.12}$$

less than ϵ, for arbitrarily small $\epsilon > 0$.

Proof. (1) The vanishing result of Lemma 4.17 shows $\underline{\vartheta}_m(f)|_{H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})} = 0$ for all $f \in \mathcal{V}$. Thus, the trace pairing satisfies $\langle f, h \rangle = 0$, for $f \in \mathcal{V}$ and for all $h \in \mathbf{S}\,(C_{\mathbb{K}})$. In particular, this applies to the case $h = f^\sharp$.

(2) This follows from the surjectivity of the map $\mathfrak{E}$ for the weight $\delta = 0$ (cf. [10, Appendix 1]).

Proposition 6.4 shows that the trace pairing admits a large radical given by all functions that extend to adèles. Thus, one can divide out this radical and work with the cohomology $H^1(\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*, C_{\mathbb{K}})$ described above.

7 Correspondences

To start building the dictionary between the Weil proof and the noncommutative geometry of the adèles class space, we begin by reformulating the trace formula discussed above in more intersection-theoretic language, so as to be able to compare it with the setup of Section 2.1 above. We also discuss in this section the analogue of modding out by the trivial correspondences.

7.1 The scaling correspondence as Frobenius

To the scaling action

$$\vartheta_a(\gamma)(\xi)(x) = \xi(\gamma^{-1}x), \qquad \text{for } \gamma \in C_{\mathbb{K}} \quad \text{and} \quad \xi \in \mathcal{S}(\mathbb{A}_{\mathbb{K}}),$$

one associates the graph Z_γ given by the pairs $(x, \gamma^{-1}x)$. These should be considered as points in the product $\mathbb{A}_{\mathbb{K}}/\mathbb{K}^* \times \mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$ of two copies of the adèles class space. Thus, the analogue in our context of the correspondences $Z = \sum_n a_n \mathrm{Fr}^n$ on $C \times C$ is given by elements of the form

$$Z(f) = \int_{C_{\mathbb{K}}} f(g) \, Z_g \, d^*g, \tag{7.1}$$

for some $f \in \mathbf{S}\,(C_{\mathbb{K}})$.

With this interpretation of correspondences, we can then make sense of the terms in the trace formula in the following way.

Definition 7.1. *For a correspondence of the form (7.1) we define degree and codegree by the following prescription:*

$$d(Z(f)) := \hat{f}(1) = \int_{C_{\mathbb{K}}} f(u)\,|u|\,d^*u, \tag{7.2}$$

$$d'(Z(f)) := d(Z(f^\sharp)) = \int_{C_{\mathbb{K}}} f(u)\,d^*u = \hat{f}(0). \tag{7.3}$$

Here the Fourier transform $\hat{f}$ is as in (2.30), with the trivial character $\chi = 1$. Notice that with this definition of degree and codegree we obtain

$$d(Z_g) = |g| \qquad \text{and} \qquad d'(Z_g) = 1. \tag{7.4}$$

Thus, the term $\hat{f}(1) + \hat{f}(0)$ in the trace formula of Theorem 6.1 matches the term $d(Z) + d'(Z)$ in Weil's formula for the trace of a correspondence as in (2.25). The term

$$-\int'_{(\mathbb{K}_v^*,\alpha_v)} \frac{f(u^{-1})}{|1-u|}\,d^*u \tag{7.5}$$

of (6.4) in turn can be seen as the remaining term $-Z \bullet \Delta$ in (2.25). In fact, the formula (7.5) describes, using distributions, the local contributions to the trace of the intersections between the graph $Z(f)$ and the diagonal Δ. This was proved in [10, Section VI and Appendix III]. It generalizes the analogous formula for flows on manifolds of [23], which in turn can be seen as a generalization of the usual Atiyah–Bott Lefschetz formula for a diffeomorphism of a smooth compact manifold [2].

When we separate out the contribution $\log|a|\,h(1)$, as in passing from (6.4) to (6.3), and we rewrite the trace formula as in Theorem 6.1, this corresponds to separating the intersection $Z \bullet \Delta$ into a term that is proportional to the self-intersection $\Delta \bullet \Delta$ and a remaining term in which the intersection is transverse.

To see this, we notice that the term $\log|a|$, for $a = (a_v)$ a differental idèle, is of the form (6.2). Indeed, one sees that in the function-field case the term

$$-\log|a| = -\log q^{2-2g} = (2g-2)\log q = -\Delta \bullet \Delta \,\log q$$

is proportional to the self-intersection of the diagonal, which brings us to consider the value $\log|a| = -\log D$ with the discriminant of a number field as the analogue in characteristic zero of the self-intersection of the diagonal.

In these intersection-theoretic terms we can reformulate the positivity condition (cf. [4]) equivalent to the Riemann hypothesis in the following way.

Proposition 7.2. *The following two conditions are equivalent:*

- *All L-functions with Grössencharakter on $\mathbb{K}$ satisfy the Riemann hypothesis.*

• *The estimate*

$$Z(f) \bullet_{\text{trans}} Z(f) \leq 2d(Z(f))d'(Z(f)) - \varDelta \bullet \varDelta\, f \star f^{\sharp}(1) \tag{7.6}$$

holds for all $f \in \mathbf{S}\,(C_{\mathbb{K}})$.

Proof. As in the Weil proof, one separates the terms $Z \star Z' = d'(Z)\varDelta + Y$, where Y has transverse intersection with the diagonal; here we can write an identity

$$\operatorname{Tr}(\underline{\vartheta}_m(f \star f^{\sharp})|H^1) =: Z(f) \bullet Z(f) = \varDelta \bullet \varDelta\, f \star f^{\sharp}(1) + Z(f) \bullet_{\text{trans}} Z(f), \tag{7.7}$$

where the remaning term $Z(f) \bullet_{\text{trans}} Z(f)$, which represents the transverse intersection, is given by the local contributions given by the principal values over $(\mathbb{K}_v^*, e_{\mathbb{K}_v})$ in the formula (6.3) for $\operatorname{Tr}(\underline{\vartheta}_m(f \star f^{\sharp})|H^1)$.

The formula (6.3) for $\operatorname{Tr}(\underline{\vartheta}_m(f \star f^{\sharp})|H^1)$ gives a term of the form $-\log|a|\, f \star f^{\sharp}(1)$, with

$$f \star f^{\sharp}(1) = \int_{C_{\mathbb{K}}} |f(g)|^2\, |g|\, d^*g.$$

We rewrite this term as $-\varDelta \bullet \varDelta\, f \star f^{\sharp}(1)$ according to our interpretation of $\log|a|$ as self-intersection of the diagonal. This matches the term $(2g-2)d'(Z)$ in the estimate for $\operatorname{Tr}(Z \star Z')$ in the Weil proof.

The first two terms in the formula (6.3) for $\operatorname{Tr}(\underline{\vartheta}_m(f \star f^{\sharp})|H^1)$ are of the form

$$\widehat{f \star f^{\sharp}}(0) + \widehat{f \star f^{\sharp}}(1) = 2\hat{f}(0)\hat{f}(1) = 2d'(Z(f))d(Z(f)). \tag{7.8}$$

This matches the term $2gd'(Z) = 2d(Z)d'(Z)$ in the expression for $\operatorname{Tr}(Z \star Z')$ in the Weil proof.

With this notation understood, we see that the positivity $\operatorname{Tr}(\underline{\vartheta}_m(f \star f^{\sharp})|H^1) \geq 0$ indeed corresponds to the estimate (7.6).

7.2 Fubini's theorem and the trivial correspondences

As we have seen in recalling the main steps in the Weil proof, a first step in dealing with correspondences is to use the freedom to add multiples of the trivial correspondences in order to adjust the degree. We describe an analogue, in our noncommutative geometry setting, of the trivial correspondences and of this operation of modifying the degree.

In view of the result of Proposition 6.4 above, it is natural to regard the elements $f \in \mathcal{V} \subset \mathbf{S}\,(C_{\mathbb{K}})$ as those that give rise to the trivial correspondences $Z(f)$. Here, as above, $\mathcal{V}$ is the range of the reduction map from adèles.

The fact that it is possible to arbitrarily modify the degree $d(Z(f)) = \hat{f}(1)$ of a correspondence by adding to f an element in $\mathcal{V}$ depends on the subtle fact that we are dealing with a case in which the Fubini theorem does not apply.

In fact, consider an element $\xi \in \mathcal{S}(\mathbb{A}_{\mathbb{K}})_0$. We know that it satisfies the vanishing condition

$$\int_{\mathbb{A}_{\mathbb{K}}} \xi(x)\, dx = 0.$$

Thus, at first sight it would appear that for the function on $C_{\mathbb{K}}$ defined by $f(x) = \sum_{k\in\mathbb{K}^*} \xi(kx)$,

$$\hat{f}(1) = \int_{C_{\mathbb{K}}} f(g)|g|d^*g \tag{7.9}$$

should also vanish, since we have $f(x) = \sum_{k\in\mathbb{K}^*} \xi(kx)$, and for local fields (but not in the global case) the relation between the additive and multiplicative Haar measures is of the form $dg = |g|d^*g$. This, however, is in general not the case. To see more clearly what happens, let us just restrict to the case $\mathbb{K} = \mathbb{Q}$ and assume that the function $\xi(x)$ is of the form

$$\xi = \mathbf{1}_{\hat{\mathbb{Z}}} \otimes \eta,$$

with $\mathbf{1}_{\hat{\mathbb{Z}}}$ the characteristic function of $\hat{\mathbb{Z}}$ and with $\eta \in \mathcal{S}(\mathbb{R})_0$. We then have $C_{\mathbb{Q}} = \hat{\mathbb{Z}}^* \times \mathbb{R}_+^*$, and the function f is of the form

$$f(u,\lambda) = \sum_{n\in\mathbb{Z}, n\neq 0} \eta(n\lambda), \quad \forall \lambda \in \mathbb{R}_+^*\,,\ u \in \hat{\mathbb{Z}}^*\,. \tag{7.10}$$

We can thus write (7.9) in this case as

$$\hat{f}(1) = \int_{\hat{\mathbb{Z}}^*\times\mathbb{R}_+^*} f(u,\lambda) du\, d\lambda = \int_{\mathbb{R}} \sum_{n\in\mathbb{N}} \eta(n\lambda)\, d\lambda. \tag{7.11}$$

Moreover, since $\eta \in \mathcal{S}(\mathbb{R})_0$, we have for all n,

$$\int_{\mathbb{R}} \eta(n\lambda) d\lambda = 0\,. \tag{7.12}$$

It is, however, *not* necessarily the case that we can apply Fubini's theorem and write

$$\int_{\mathbb{R}} \sum_{n\in\mathbb{N}} \eta(n\lambda)\, d\lambda = \sum_n \int_{\mathbb{R}} \eta(n\lambda) d\lambda = 0, \tag{7.13}$$

since as soon as $\eta \neq 0$ one has

$$\sum_{n=1}^{\infty} \int_{\mathbb{R}} |\eta(n\lambda)| d\lambda = \left(\int_{\mathbb{R}} |\eta(\lambda)| d\lambda \right) \sum_{n=1}^{\infty} \frac{1}{n} = \infty,$$

so that Fubini's theorem does not apply and one cannot interchange the integral and the sum in (7.13). Thus, one can in general have $\hat{f}(1) \neq 0$, even though $\sum_n \int_{\mathbb{R}} \eta(n\lambda) d\lambda = 0$. In fact, we have the following result.

Lemma 7.3. *Given $f \in \mathbf{S}(C_{\mathbb{K}})$, it is possible to change arbitrarily the value of the degree $d(Z(f)) = \hat{f}(1)$ by adding elements of $\mathcal{V}$.*

Proof. It suffices to exhibit an element $f \in \mathcal{V}$ such that $\hat{f}(1) \neq 0$, since then, by linearity one can obtain the result. We treat only the case $\mathbb{K} = \mathbb{Q}$. We take $\eta \in \mathcal{S}(\mathbb{R})_0$ given by

$$\eta(x) = \pi x^2 \left(\pi x^2 - \frac{3}{2}\right) e^{-\pi x^2}.$$

One finds that, up to normalization, the Fourier transform $\hat{f}$ is given by

$$\hat{f}(is) = \int_{\mathbb{R}_+^*} \sum_{n \in \mathbb{N}} \eta(n\lambda)\, \lambda^{is} d^*\lambda = s(s+i)\zeta^*(is),$$

where ζ^* is the complete zeta function,

$$\zeta^*(z) = \pi^{-z/2}\, \Gamma\left(\frac{z}{2}\right) \zeta(z). \tag{7.14}$$

This function has a simple pole at $z = 1$; thus one gets that $\hat{f}(1) \neq 0$. ☐

An important question, in order to proceed and build a dictionary that parallels the main steps in the Weil proof, is to identify the correct notion of principal divisors. To this purpose, we show that we have at least a good analogue for the points of the curve, in terms of states of some thermodynamic system, that extend from the function field setting to the number field case.

8 Thermodynamics and geometry of the primes

Let $\mathbb{K}$ be a global field, with $\mathbb{A}_{\mathbb{K}}$ the ring of adèles and $C_{\mathbb{K}}$ the idèles classes, as above. We denote by $C_{\mathbb{K},1} \subset C_{\mathbb{K}}$ the kernel of the norm $|\cdot| : C_{\mathbb{K}} \to \mathbb{R}_+^*$.

The origin (cf. [10]) of the terms in the geometric side of the trace formula (Theorem 6.1) comes from the Lefschetz formula by Atiyah–Bott [2] and its adaptation by Guillemin–Sternberg (cf. [23]) to the distribution-theoretic trace for flows on manifolds, which is a variation on the theme of [2]. For the action of $C_{\mathbb{K}}$ on the adèles, class space $X_{\mathbb{K}}$ the relevant periodic points are

$$P = \{(x,u) \in X_{\mathbb{K}} \times C_{\mathbb{K}} \,|\, u\, x = x\}, \tag{8.1}$$

and one has (cf. [10]) the following proposition.

Proposition 8.1. *Let $(x,u) \in P$, with $u \neq 1$. There exists a place $v \in \Sigma_{\mathbb{K}}$ such that*

$$x \in X_{\mathbb{K},v} = \{x \in X_{\mathbb{K}} \,|\, x_v = 0\}. \tag{8.2}$$

The isotropy subgroup of any $x \in X_{\mathbb{K},v}$ contains the cocompact subgroup

$$\mathbb{K}_v^* \subset C_{\mathbb{K}}\,, \quad \mathbb{K}_v^* = \{(k_w) \,|\, k_w = 1 \;\forall w \neq v\}. \tag{8.3}$$

The spaces $X_{\mathbb{K},v}$ are noncommutative spaces, and as such they are described by the following noncommutative algebras:

Definition 8.2. *Let $\mathbb{A}_{\mathbb{K},v} \subset \mathbb{A}_{\mathbb{K}}$ denote the closed $\mathbb{K}^*$-invariant subset of adèles*

$$\mathbb{A}_{\mathbb{K},v} = \{a = (a_w)_{w\in\Sigma_{\mathbb{K}}} \,|\, a_v = 0\}. \tag{8.4}$$

Let $\mathcal{G}_{\mathbb{K},v}$ denote the closed subgroupoid of $\mathcal{G}_{\mathbb{K}}$ given by

$$\mathcal{G}_{\mathbb{K},v} = \{(k,x) \in \mathcal{G}_{\mathbb{K}} \,|\, x_v = 0\}, \tag{8.5}$$

and let $\mathcal{A}_v = \mathcal{S}(\mathcal{G}_{\mathbb{K},v})$ be the corresponding groupoid algebra.

Since the inclusion $\mathbb{A}_{\mathbb{K},v} \subset \mathbb{A}_{\mathbb{K}}$ is $\mathbb{K}^*$-equivariant and proper, it extends to an algebra homomorphism

$$\rho_v \;:\; \mathcal{S}(\mathcal{G}_{\mathbb{K}}) \to \mathcal{S}(\mathcal{G}_{\mathbb{K},v}), \tag{8.6}$$

which plays the role of the restriction map to the periodic orbit $X_{\mathbb{K},v}$. We shall now determine the classical points of each of the $X_{\mathbb{K},v}$. Taken together these will form the following locus inside the adèles class space, which we refer to as the "periodic classical points" of $X_{\mathbb{K}} = \mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$.

Definition 8.3. *Let $\mathbb{K}$ be a global field. For a place $v \in \Sigma_{\mathbb{K}}$ consider the adèle*

$$a^{(v)} = \left(a_w^{(v)}\right), \qquad \text{with} \quad a_w^{(v)} = \begin{cases} 1 & w \neq v, \\ 0 & w = v. \end{cases} \tag{8.7}$$

The set of periodic classical *points of the adèles class space $\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$ is defined as the union of orbits*

$$\Xi_{\mathbb{K}} := \bigcup_{v\in\Sigma_{\mathbb{K}}} C_{\mathbb{K}} a^{(v)}. \tag{8.8}$$

8.1 The global Morita equivalence

In order to deal with states rather than weights, we perform a global Morita equivalence, obtained by reducing the groupoid $\mathcal{G}_{\mathbb{K}}$ by a suitable open set. The set $\mathbb{A}_{\mathbb{K}}^{(1)}$ of (8.9) that we use to reduce the groupoid $\mathcal{G}_{\mathbb{K}}$ will capture only part of the classical subspace $C_{\mathbb{K}}$, but since our main focus is on the geometry of the complement of this subspace (the cokernel of the reduction map), this will not be a problem.

Lemma 8.4. *Let $\mathbb{K}$ be a global field. Let $W \subset \mathbb{A}_{\mathbb{K}}$ be a neighborhood of $0 \in \mathbb{A}_{\mathbb{K}}$. Then for $x \in \mathbb{A}_{\mathbb{K}}$ one has $\mathbb{K}^*x \cap W \neq \emptyset$, unless $x \in \mathbb{A}_{\mathbb{K}}^*$ is an idèle. For $x \in \mathbb{A}_{\mathbb{K}}^*$, the orbit $\mathbb{K}^*x$ is discrete in $\mathbb{A}_{\mathbb{K}}$.*

Proof. One can assume that W is of the form

$$W = \{a = (a_w) |\ |a_w| < \varepsilon\ \forall w \in S \text{ and } |a_w| \le 1\ \forall w \notin S\},$$

for S a finite set of places and for some $\varepsilon > 0$. Multiplying by a suitable idèle, one can in fact assume that $S = \emptyset$, so that we have

$$W = \{a = (a_w) |\ |a_w| \le 1\ \forall w \in \Sigma_{\mathbb{K}}\}.$$

One has $|x_v| \le 1$ except on a finite set $F \subset \Sigma_{\mathbb{K}}$ of places. Moreover, if x is not an idèle, one can also assume that

$$\prod_{v \in F} |x_v| < \delta$$

for any fixed δ. Thus, $-\log |x_v|$ is as large as one wants and there exists $k \in \mathbb{K}^*$ such that $k\, x \in W$. This is clear in the function field case because of the Riemann–Roch formula (2.7). In the case of $\mathbb{Q}$ one can first multiply x by an integer to get $|x_v| \le 1$ for all finite places; then, since this does not alter the product of all $|x_v|$, one gets $|x_\infty| < 1$ and $x \in W$. The case of more general number fields is analogous. In the case of idèles, one can assume that $x = 1$, and then the second statement follows from the discreteness of $\mathbb{K}$ in $\mathbb{A}_{\mathbb{K}}$. $\square$

We consider the following choice of a neighborhood of zero.

Definition 8.5. *Consider the open neighborhood of* $0 \in \mathbb{A}_{\mathbb{K}}$ *defined by*

$$\mathbb{A}_{\mathbb{K}}^{(1)} = \prod_{w \in \Sigma_{\mathbb{K}}} \mathbb{K}_w^{(1)} \subset \mathbb{A}_{\mathbb{K}}, \tag{8.9}$$

where for any place we let $\mathbb{K}_w^{(1)}$ *be the* interior *of* $\{x \in \mathbb{K}_w\, ;\, |x| \le 1\}$. *Let* $\mathcal{G}_{\mathbb{K}}^{(1)}$ *denote the reduction of the groupoid* $\mathcal{G}_{\mathbb{K}}$ *by the open subset* $\mathbb{A}_{\mathbb{K}}^{(1)} \subset \mathbb{A}_{\mathbb{K}}$ *of the units and let* $\mathcal{S}\left(\mathcal{G}_{\mathbb{K}}^{(1)}\right)$ *denote the corresponding (smooth) groupoid algebra.*

The algebra $\mathcal{S}\left(\mathcal{G}_{\mathbb{K}}^{(1)}\right)$ is a subalgebra of $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$, where one simply extends the function $f(k, x)$ by zero outside of the open subgroupoid $\mathcal{G}_{\mathbb{K}}^{(1)} \subset \mathcal{G}_{\mathbb{K}}$. With this convention, the convolution product of $\mathcal{S}\left(\mathcal{G}_{\mathbb{K}}^{(1)}\right)$ is simply given by the convolution product of $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$ of the form

$$(f_1 \star f_2)(k, x) = \sum_{h \in \mathbb{K}^*} f_1(kh^{-1}, hx) f_2(h, x).$$

We see from Lemma 8.4 above that the only effect of the reduction to $\mathcal{G}_{\mathbb{K}}^{(1)}$ is to remove from the noncommutative space $\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$ all the elements of $C_{\mathbb{K}}$ whose class modulo $\mathbb{K}^*$ does not intersect $\mathcal{G}_{\mathbb{K}}^{(1)}$ (i.e., in particular, those whose norm is greater than or equal to one). We then have the following symmetries for the algebra $\mathcal{S}\left(\mathcal{G}_{\mathbb{K}}^{(1)}\right)$.

Proposition 8.6. *Let $\mathcal{J}^+$ denote the semigroup of idèles $j \in \mathbb{A}_{\mathbb{K}}^*$ such that $j\mathbb{A}_{\mathbb{K}}^{(1)} \subset \mathbb{A}_{\mathbb{K}}^{(1)}$. The semigroup $\mathcal{J}^+$ acts on the algebra $\mathcal{S}\left(\mathcal{G}_{\mathbb{K}}^{(1)}\right)$ by endomorphisms obtained as restrictions of the automorphisms of $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$ of the form*

$$\vartheta_a(j)(f)(k,x) = f(k, j^{-1}x), \quad \forall (k,x) \in \mathcal{G}_{\mathbb{K}}\,, \quad j \in \mathcal{J}^+. \tag{8.10}$$

Let $\mathbb{K} = \mathbb{Q}$ and let $C_{\mathbb{Q}}^+ \subset C_{\mathbb{Q}}$ be the semigroup $C_{\mathbb{Q}}^+ = \{g \in C_{\mathbb{Q}} |\, |g| < 1\}$. The semigroup $C_{\mathbb{Q}}^+$ acts on $\mathcal{S}\left(\mathcal{G}_{\mathbb{Q}}^{(1)}\right)$ by the endomorphisms

$$F(g) = \vartheta_a(\bar{g}) \tag{8.11}$$

with $\bar{g}$ the natural lift of $g \in C_{\mathbb{Q}}^+$ to $\hat{\mathbb{Z}}^* \times \mathbb{R}_+^*$.

Proof. By construction, $\vartheta_a(j)$ is an automorphism of $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$. For a function f with support B in the open set $\mathcal{G}_{\mathbb{K}}^{(1)}$ the support of the function $\vartheta_a(j)(f)$ is $jB = \{(k, jx) | (k,x) \in B\} \subset \mathcal{G}_{\mathbb{K}}^{(1)}$, so that $\vartheta_a(j)(f)$ still has support in $\mathcal{G}_{\mathbb{K}}^{(1)}$.

For $\mathbb{K} = \mathbb{Q}$ let $\bar{g} \in \hat{\mathbb{Z}}^* \times \mathbb{R}_+^*$ be the natural lift of an element $g \in C_{\mathbb{Q}}^+$. Then the archimedean component $\bar{g}_\infty$ is of absolute value less than 1, so that $\bar{g} \in \mathcal{J}^+$. The action of $\vartheta_a(\bar{g})$ by endomorphisms of $\mathcal{S}\left(\mathcal{G}_{\mathbb{Q}}^{(1)}\right)$ induces a corresponding action of $C_{\mathbb{Q}}^+$. □

Remark 8.7. For m a positive integer, consider the element $g = (1, m^{-1}) \in C_{\mathbb{Q}}^+$. Both $g = (1, m^{-1})$ and $\tilde{m} = (m, 1)$ are in $\mathcal{J}^+$ and have the same class in the idèle class group $C_{\mathbb{Q}}$, since $m\,g = \tilde{m}$. Thus the automorphisms $\vartheta_a(g)$ and $\vartheta_a(\tilde{m})$ of $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$ are inner conjugate. Since the open set $\mathbb{A}_{\mathbb{K}}^{(1)} \subset \mathbb{A}_{\mathbb{K}}$ is not closed, its characteristic function is not continuous and does not define a multiplier of $\mathcal{S}(\mathcal{G}_{\mathbb{K}})$. It follows that the endomorphism $F(g)$ is inner conjugate to the endomorphism $\vartheta_a(\tilde{m})$ only in the following weaker sense. There exists a sequence of elements u_n of $\mathcal{S}\left(\mathcal{G}_{\mathbb{K}}^{(1)}\right)$ such that for any $f \in \mathcal{S}\left(\mathcal{G}_{\mathbb{K}}^{(1)}\right)$ with compact support,

$$F(g)(f) = u_n\, \vartheta_a(\tilde{m})(f)\, u_n^*$$

holds for all n large enough.

8.2 The valuation systems

We now explain why the orbits $C_{\mathbb{K}}a^{(v)}$ indeed appear as the set of classical points, in the sense of the low-temperature KMS states, of the noncommutative spaces $X_{\mathbb{K},v}$. The notion of classical points obtained from low-temperature KMS states is discussed at length in [15] (cf. also [12], [13], [14]).

The noncommutative space $X_{\mathbb{K},v}$ is described by the restricted groupoid

$$\mathcal{G}(v) = \mathbb{K}^* \ltimes \mathbb{A}_{\mathbb{K},v}^{(1)} = \{(g,a) \in \mathbb{K}^* \ltimes \mathbb{A}_{\mathbb{K},v} \,|\, a \text{ and } ga \in \mathbb{A}_{\mathbb{K},v}^{(1)}\}\,. \tag{8.12}$$

We denote by φ the positive functional on $C^*\left(\mathbb{K}^* \ltimes \mathbb{A}^{(1)}_{\mathbb{K},v}\right)$ given by

$$\varphi(f) = \int_{\mathbb{A}^{(1)}_{\mathbb{K},v}} f(1,a)\, da. \tag{8.13}$$

Proposition 8.8. *The modular automorphism group of the functional φ on the crossed product $C^*(\mathcal{G}(v))$ is given by the time evolution*

$$\sigma^v_t(f)(k,x) = |k|^{it}_v\, f(k,x), \quad \forall t \in \mathbb{R}, \;\; \forall f \in C^*\left(\mathbb{K}^* \ltimes \mathbb{A}^{(1)}_{\mathbb{K},v}\right). \tag{8.14}$$

Proof. We identify elements of $C_c\left(\mathbb{K}^* \ltimes \mathbb{A}^{(1)}_{\mathbb{K},v}\right)$ with functions $f(g,a)$ of elements $g \in \mathbb{K}^*$ and $a \in \mathbb{A}^{(1)}_{\mathbb{K},v}$. The product is simply of the form

$$f_1 * f_2(g,a) = \sum_r f_1(g\,r^{-1}, ga) f_2(r,a).$$

The additive Haar measure da on $\mathbb{A}_{\mathbb{K},v}$ satisfies the scaling property

$$d(ka) = |k|^{-1}_v da\,, \quad \forall k \in \mathbb{K}^*\,, \tag{8.15}$$

since the product measure $da \times da_v$ on $\mathbb{A}_{\mathbb{K}} = \mathbb{A}_{\mathbb{K},v} \times \mathbb{K}_v$ is invariant under the scaling by $k \in \mathbb{K}^*$, while the additive Haar measure da_v on $\mathbb{K}_v$ gets multiplied by $|k|_v$, namely $d(ka_v) = |k|_v da_v$. We then check the KMS_1 condition, for φ associated to the additive Haar measure, as follows,

$$\begin{aligned}\varphi(f_1 * f_2) &= \textstyle\sum_r \int_{\mathbb{A}^{(1)}_{\mathbb{K},v}} f_1(r^{-1}, r\,a) f_2(r,\, a)\, da \\ &= \textstyle\sum_r \int_{\mathbb{A}^{(1)}_{\mathbb{K},v}} f_2(k^{-1}, k\,b) f_1(k,\, b)\, |k|^{-1}_v\, db = \varphi(f_2 * \sigma_i(f_1))\,,\end{aligned}$$

using the change of variables $k = r^{-1}$, $a = kb$, and $da = |k|^{-1}_v\, db$. $\square$

It is worthwhile to observe that these automorphisms extend to the global algebra. Let $\mathcal{G}^{(1)}_{\mathbb{K}}$ be the groupoid $\mathbb{K}^* \ltimes \mathbb{A}^{(1)}_{\mathbb{K}}$ of Definition 8.5.

Lemma 8.9. *Let $\mathbb{K}$ be a global field and $v \in \Sigma_{\mathbb{K}}$ a place. The map*

$$d_v(k,x) = \log |k|_v \in \mathbb{R} \tag{8.16}$$

defines a homomorphism of the groupoid $\mathcal{G}^{(1)}_{\mathbb{K}}$ to the additive group $\mathbb{R}$, and the time evolution

$$\sigma^v_t(f)(k,x) = |k|^{it}_v\, f(k,x), \quad \forall t \in \mathbb{R}, \;\; \forall f \in \mathcal{S}\left(\mathcal{G}^{(1)}_{\mathbb{K}}\right) \tag{8.17}$$

generates a one-parameter group of automorphisms of the algebra $\mathcal{S}\left(\mathcal{G}^{(1)}_{\mathbb{K}}\right)$.

The following result shows that the nontrivial part of the dynamics σ^v_t concentrates on the algebra $\mathcal{S}(\mathcal{G}(v))$ with $\mathcal{G}(v)$ as in (8.12).

Proposition 8.10. *The morphism ρ_v of (8.6) restricts to a σ_t^v-equivariant morphism $\mathcal{S}\left(\mathcal{G}_{\mathbb{K}}^{(1)}\right) \to \mathcal{S}(\mathcal{G}(v))$. Moreover, the restriction of the one-parameter group σ_t^v to the kernel of ρ_v is inner.*

Proof. For the first statement note that the proper inclusion $\mathbb{A}_{\mathbb{K},v} \subset \mathbb{A}_{\mathbb{K}}$ restricts to a proper inclusion $\mathbb{A}_{\mathbb{K},v}^{(1)} \subset \mathbb{A}_{\mathbb{K}}^{(1)}$. For the second statement, notice that the formula

$$h_v(x) = \log |x|_v, \quad \forall x \in \mathbb{A}_{\mathbb{K}}^{(1)}, \tag{8.18}$$

defines the multipliers e^{ith_v} of the kernel of ρ_v. Indeed, e^{ith_v} is a bounded continuous function on $\mathbb{A}_{\mathbb{K}}^{(1)} \smallsetminus \mathbb{A}_{\mathbb{K},v}^{(1)}$.

We can then check that the 1-cocycle d_v is the coboundary of h_v. In fact, we have

$$h_v(k\,x) - h_v(x) = d_v(k,x), \quad \forall (k,x) \in \mathcal{G}_{\mathbb{K}}^{(1)} \smallsetminus \mathcal{G}(v). \tag{8.19}$$

□

We now recall that for an étale groupoid like $\mathcal{G}(v)$, every unit $y \in \mathcal{G}(v)^{(0)}$ defines, by

$$(\pi_y(f)\xi)(\gamma) = \sum_{\gamma_1\gamma_2=\gamma} f(\gamma_1)\xi(\gamma_2), \tag{8.20}$$

a representation π_y by left convolution of the algebra of $\mathcal{G}(v)$ in the Hilbert space $\mathcal{H}_y = \ell^2(\mathcal{G}(v)_y)$, where $\mathcal{G}(v)_y$ denotes the set of elements of the groupoid $\mathcal{G}(v)$ with source y. By construction, the unitary equivalence class of the representation π_y is unaffected when one replaces y by an equivalent $z \in \mathcal{G}(v)^{(0)}$, i.e., one assumes that there exists $\gamma \in \mathcal{G}(v)$ with range and source y and z. Thus we can think of the label y of π_y as living in the quotient space $X_{\mathbb{K},\,v}$ of equivalence classes of elements of $\mathcal{G}(v)^{(0)}$.

The relation between $\Xi_{\mathbb{K},\,v}$ and $X_{\mathbb{K},\,v}$ is then the following.

Theorem 8.11. *For $y \in X_{\mathbb{K},\,v}$, the representation π_y is a positive energy representation if and only if $y \in \Xi_{\mathbb{K},\,v}$.*

Proof. Let first $y \in \mathcal{G}(v)^{(0)} \cap \Xi_{\mathbb{K},\,v}$. Thus one has $y \in \mathbb{A}_{\mathbb{K},v}^{(1)}$, $y_w \neq 0$, for all w, and $|y_w| = 1$ for all $w \notin S$, where S is a finite set of places. We can identify $\mathcal{G}(v)_y$ with the set of $k \in \mathbb{K}^*$ such that $k\,y \in \mathbb{A}_{\mathbb{K},v}^{(1)}$. We extend y to the adèle $\tilde{y} = y \times 1$ whose component at the place v is equal to $1 \in \mathbb{K}_v$. Then $\tilde{y}$ is an idèle. Thus by Lemma 8.4 the number of elements of the orbit $\mathbb{K}^*\tilde{y}$ in a given compact subset of $\mathbb{A}_{\mathbb{K}}$ is finite. It follows that $\log|k|_v$ is bounded below on $\mathcal{G}(v)_y$. Indeed, otherwise there would exist a sequence $k_n \in \mathbb{K}^* \cap \mathcal{G}(v)_y$ such that $|k_n|_v \to 0$. Then $k_n\,\tilde{y} \in \mathbb{A}_{\mathbb{K}}^{(1)}$ for all n large enough, and this contradicts the discreteness of $\mathbb{K}^*\tilde{y}$. In the representation π_y, the time evolution σ_t is implemented by the Hamiltonian H_y given by

$$(H_y\,\xi)(k,y) = \log|k|_v\;\xi(k,y). \tag{8.21}$$

Namely, we have

$$\pi_y(\sigma_t(f)) = e^{itH_y}\pi_y(f)e^{-itH_y}, \quad \forall f \in C_c(\mathcal{G}(v)). \tag{8.22}$$

Thus since $\log|k|_v$ is bounded below on $\mathcal{G}(v)_y$, we get that the representation π_y is a positive energy representation.

Then let $y \in \mathcal{G}(v)^{(0)} \smallsetminus \Xi_{\mathbb{K},v}$. We shall show that $\log|k|_v$ is not bounded below on $\mathcal{G}(v)_y$, and thus that π_y is not a positive-energy representation. We consider as above the adelè $\tilde{y} = y \times 1$ whose component at the place v is equal to $1 \in \mathbb{K}_v$. Assume that $\log|k|_v$ is bounded below on $\mathcal{G}(v)_y$. Then there exists $\epsilon > 0$ such that for $k \in \mathbb{K}^*$,

$$k\,y \in \mathbb{A}_{\mathbb{K},v}^{(1)} \Rightarrow |k|_v \geq \epsilon\,.$$

This shows that the neighborhood of $0 \in \mathbb{A}_{\mathbb{K}}$ defined as

$$W = \{a \in \mathbb{A}_{\mathbb{K}}\,;\, |a_v| < \epsilon\,,\ a_w \in \mathbb{K}_w^{(1)}\,,\ \forall w \neq v\}$$

does not intersect $\mathbb{K}^*\tilde{y}$. Thus by Lemma 8.4 we get that $\tilde{y}$ is an idèle and $y \in \Xi_{\mathbb{K},v}$. ∎

The specific example of the Bost–Connes system combined with Theorem 8.11 shows that one can refine the recipe of [15] (cf. also [12], [13], [14]) for taking "classical points" of a noncommutative space. The latter recipe only provides a notion of classical points that can be thought of, by analogy with the positive characteristic case, as points defined over the mysterious "field with one element" $\mathbb{F}_1$ (see, e.g., [29]). To obtain instead a viable notion of the points defined over the maximal unramified extension $\bar{\mathbb{F}}_1$, one performs the following sequence of operations.

$$X \overset{\text{Dual System}}{\longrightarrow} \hat{X} \overset{\text{Periodic Orbits}}{\longrightarrow} \cup \hat{X}_v \overset{\text{Classical Points}}{\longrightarrow} \cup \Xi_v, \tag{8.23}$$

which make sense in the framework of endomotives of [11]. Note in particular that the dual system $\hat{X}$ is of type II and as such does not have a nontrivial time evolution. Thus it is only by restricting to the periodic orbits that one passes to noncommutative spaces of type III for which the cooling operation is nontrivial. In the analogy with geometry in nonzero characteristic, the set of points $X(\bar{\mathbb{F}}_q)$ over $\bar{\mathbb{F}}_q$ of a variety X is indeed obtained as the union of the periodic orbits of the Frobenius.

Remark 8.12. Theorem 8.11 does not give the classification of KMS_β states for the quantum statistical system $\left(C^*\left(\mathbb{K}^* \ltimes \mathbb{A}_{\mathbb{K},v}^{(1)}\right), \sigma_t\right)$. It just exhibits extremal KMS_β states but does not show that all of them are of this form.

8.3 The curve inside the adèles class space

In the case of a function field $\mathbb{K} = \mathbb{F}_q(C)$, the set of *periodic classical* points of the adèles class space $\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$ is (noncanonically) isomorphic to the algebraic points $C(\bar{\mathbb{F}}_q)$. In fact, more precisely, the set of algebraic points $C(\bar{\mathbb{F}}_q)$ is equivariantly isomorphic to the quotient $\Xi_{\mathbb{K}}/C_{\mathbb{K},1}$ where $C_{\mathbb{K},1} \subset C_{\mathbb{K}}$ is the kernel of the norm $|\cdot| : C_{\mathbb{K}} \to \mathbb{R}_+^*$, and $\Xi_{\mathbb{K}}$ is as in (8.8).

Proposition 8.13. *For $\mathbb{K} = \mathbb{F}_q(C)$ a function field, the orbits of the Frobenius on $C(\bar{\mathbb{F}}_q)$ give an equivariant identification*

$$\Xi_{\mathbb{K}}/C_{\mathbb{K},1} \simeq C(\bar{\mathbb{F}}_q), \tag{8.24}$$

between $\Xi_{\mathbb{K}}/C_{\mathbb{K},1}$ with the action of $q^{\mathbb{Z}}$ and $C(\bar{\mathbb{F}}_q)$ with the action of the group of integer powers of the Frobenius.

Proof. At each place $v \in \Sigma_{\mathbb{K}}$, the quotient group of the range N of the norm $|\cdot| : C_{\mathbb{K}} \to \mathbb{R}_+^*$ by the range N_v of $|\cdot| : \mathbb{K}_v \to \mathbb{R}_+^*$ is the finite cyclic group

$$N/N_v = q^{\mathbb{Z}}/q^{n_v\mathbb{Z}} \simeq \mathbb{Z}/n_v\mathbb{Z}, \tag{8.25}$$

where n_v is the degree of the place $v \in \Sigma_{\mathbb{K}}$. The degree n_v is the same as the cardinality of the orbit of the Frobenius acting on the fiber of the map (2.2) from algebraic points in $C(\bar{\mathbb{F}}_q)$ to places in $\Sigma_{\mathbb{K}}$. Thus, one can construct in this way an equivariant embedding

$$C(\bar{\mathbb{F}}_q) \hookrightarrow (\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*)/C_{\mathbb{K},1} \tag{8.26}$$

obtained, after choosing a point in each orbit, by mapping the orbit of the integer powers of the Frobenius in $C(\bar{\mathbb{F}}_q)$ over a place v to the orbit of $C_{\mathbb{K}}/C_{\mathbb{K},1} \sim q^{\mathbb{Z}}$ on the adèle $a^{(v)}$. □

Modulo the problem created by the fact that the identification above is noncanonical and relies upon the choice of a point in each orbit, it is then possible to think of the locus $\Xi_{\mathbb{K}}$, in the number field case, as a replacement for $C(\bar{\mathbb{F}}_q)$ inside the adèles class space $\mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$.

In the case of $\mathbb{K} = \mathbb{Q}$, the quotient $\Xi_{\mathbb{Q}}/C_{\mathbb{Q},1}$ appears as a union of periodic orbits of period $\log p$ under the action of $C_{\mathbb{Q}}/C_{\mathbb{Q},1} \sim \mathbb{R}$, as in Figure 1. What

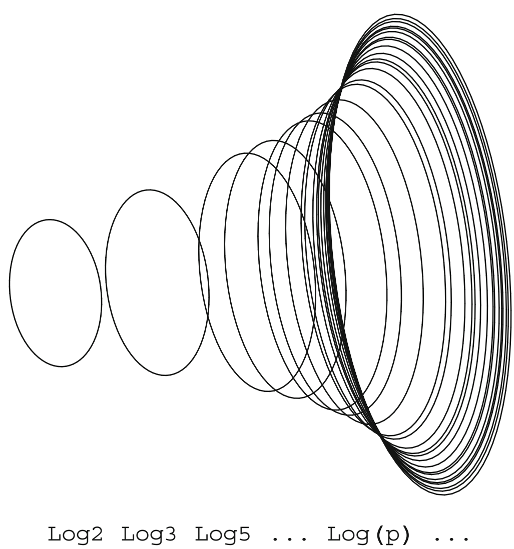


Fig. 1. The classical points $\Xi_{\mathbb{Q}}/C_{\mathbb{Q},1}$ of the adèles class space $\mathbb{A}_{\mathbb{Q}}/\mathbb{Q}^*$.

matters, however, is not the space $\Xi_{\mathbb{Q}}/C_{\mathbb{Q},1}$ in itself but the way it sits inside $\mathbb{A}_{\mathbb{Q}}/\mathbb{Q}^*$. Without taking into account the topology induced by $\mathbb{A}_{\mathbb{K}}$, the space $\Xi_{\mathbb{K}}$ would just be a disjoint union of orbits without any interesting global structure, while it is the embedding in the adèles class space that provides the geometric setting underlying the Lefschetz trace formula of [10] and its cohomological formulation of [11].

8.4 The valuation systems for $\mathbb{K} = \mathbb{Q}$

We concentrate again on the specific case of $\mathbb{K} = \mathbb{Q}$ to understand better the properties of the dynamical systems σ_t^p associated to the finite primes $p \in \Sigma_{\mathbb{Q}}$.

We know that in the case of the BC system, the KMS state at critical temperature $\beta = 1$ is given by the additive Haar measure on finite adèles [5]. Thus, one expects that for the systems associated to the finite primes, the additive Haar measure of $\mathbb{A}_{\mathbb{Q},p}$ should play an analogous role.

Definition 8.14. *Let $\mathbb{A}^*_{\mathbb{Q},p} \subset \mathbb{A}^{(1)}_{\mathbb{Q},p}$ be the subspace*

$$\mathbb{A}^*_{\mathbb{Q},p} = \{x \in \mathbb{A}_{\mathbb{Q},p} |\ |x_w| = 1\ \forall w \neq p, \infty \quad \textit{and } p^{-1} \leq |x_\infty| < 1\}. \tag{8.27}$$

As above, $\mathcal{G}(p)$ denotes the reduction of the groupoid $\mathcal{G}_{\mathbb{Q},p}$ by the open subset $\mathbb{A}^{(1)}_{\mathbb{Q},p} \subset \mathbb{A}_{\mathbb{Q},p}$, namely

$$\mathcal{G}(p) = \{(k, x) \in \mathcal{G}_{\mathbb{Q},p} \,|\, x \in \mathbb{A}^{(1)}_{\mathbb{Q},p},\ kx \in \mathbb{A}^{(1)}_{\mathbb{Q},p}\}. \tag{8.28}$$

Notice that the set $\mathbb{A}^{(1)}_{\mathbb{Q},p}$ meets all the equivalence classes in $\mathbb{A}_{\mathbb{Q},p}$ by the action of $\mathbb{Q}^*$. In fact, given $x \in \mathbb{A}_{\mathbb{Q},p}$, one can find a representative y with $y \sim x$ in $\mathbb{A}_{\mathbb{Q},p}/\mathbb{Q}^*$ such that $y \in \hat{\mathbb{Z}} \times \mathbb{R}$. Upon multiplying y by a suitable power of p, one can make y_∞ as small as required, and in particular one can obtain in this way a representative in $\mathbb{A}^{(1)}_{\mathbb{Q},p}$. Let us assume that $|y_w| = 1$ for all finite places $w \neq p$ and that $y_\infty > 0$. Then there exists a unique $n \in \mathbb{N} \cup \{0\}$ such that $p^n\, y \in \mathbb{A}^*_{\mathbb{Q},p}$.

Given a prime p we define the function $f_p(\lambda, \beta)$ for $\lambda \in (1, p]$ and $\beta > 1$ by

$$f_p(\lambda, \beta) = \sum c_k\, p^{-k\beta}, \tag{8.29}$$

where the $c_k \in \{0, \ldots, p-1\}$ are the digits of the expansion of λ in base p. There is an ambiguous case in which all digits c_k are equal to 0 for $k > m$, while $c_m > 0$, since the same number

$$\lambda = \sum c_k\, p^{-k}$$

is obtained using the same c_j for $j < m$, $c_m - 1$ instead of c_m, and $c_j = p - 1$ for $j > m$. In that case, for $\beta > 1$, (8.29) gives two different values, and we choose the value coming from the second representation of λ, i.e., the lower of the two. These coefficients c_k of the expansion of λ in base p are then given by

$$c_k = \lceil \lambda p^k - 1 \rceil - p \lceil \lambda p^{k-1} - 1 \rceil \,, \tag{8.30}$$

where $\lceil x \rceil = \inf_{n \in \mathbb{Z}} \{ n \geq x \}$ denotes the ceiling function.

Note that for $\beta > 1$, the function $f_p(\lambda, \beta)$ is discontinuous (cf. Figures 2 and 3) at any point (λ, β) where the expansion of λ in base p is ambiguous, i.e., $\lambda \in \mathbb{N} p^{-k}$. Moreover, for $\beta = 1$ one gets

$$f_p(\lambda, 1) = \lambda \,, \quad \forall \lambda \in (1, p] \,. \tag{8.31}$$

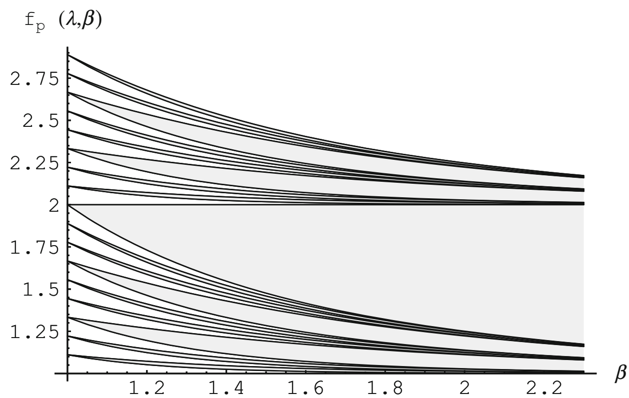


Fig. 2. Graphs of the functions $f_p(\lambda, \beta)$ as functions of β for $p = 3$, $\lambda = n/27$. The gray regions are the gaps in the range of f_p.

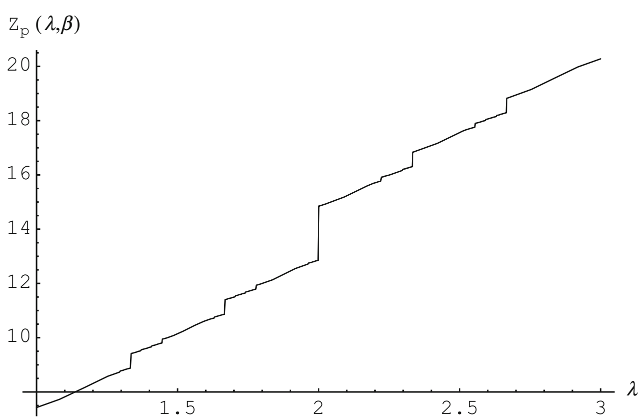


Fig. 3. Graph of the function $Z_p(\lambda, \beta)$ as a function of λ for $p = 3$, $\beta = 1.2$.

We then obtain the following result.

Theorem 8.15. *Let* $(C^*(\mathcal{G}(p)), \sigma_t^p)$ *be the* C^**-dynamical system associated to the groupoid* (8.28) *with the time evolution* (8.17)*. Then the following properties hold:*

1. *For any* $y \in \mathbb{A}^*_{\mathbb{Q},p}$ *the corresponding representation* π_y *has positive energy.*
2. *Let* H_y *denote the Hamiltonian implementing the time evolution in the representation* π_y*, for* $y \in \mathbb{A}^*_{\mathbb{Q},p}$ *with* $y_\infty = \lambda^{-1}$ *and* $\lambda \in (1, p]$*. Then the partition function is given by*

$$Z_p(\lambda, \beta) = \mathrm{Tr}(e^{-\beta H_y}) = 2\,\frac{1 - p^{-\beta}}{1 - p^{1-\beta}}\, f_p(\lambda, \beta)\,. \tag{8.32}$$

3. *The functionals*

$$\psi_{\beta,\,y}(a) = \mathrm{Tr}(e^{-\beta H_y}\,\pi_y(a))\,, \quad \forall a \in C^*(\mathcal{G}(p)), \tag{8.33}$$

satisfy the KMS$_\beta$ *condition for* σ_t^p *and depend weakly continuously on the parameter* $y \in \mathbb{A}^*_{\mathbb{Q},p}$.

Proof. (1) This follows from Theorem 8.11. For $y \in \mathbb{A}^*_{\mathbb{Q},p}$ one has

$$r \in \mathbb{Q}^*, \;\; ry \in \mathbb{A}^{(1)}_{\mathbb{Q},p} \Longrightarrow r = p^{-k} m, \tag{8.34}$$

for some $k \geq 0$ and some integer m prime to p and such that $|r\,y_\infty| < 1$. This implies

$$|m| < p^{k+1}, \tag{8.35}$$

and one obtains

$$|r|_p = p^k \geq 1 \quad \text{and} \quad \log|r|_p \geq 0. \tag{8.36}$$

In fact, the argument above shows that the spectrum of the Hamiltonian H_y implementing the time evolution σ_t^p in the representation π_y is given by

$$\mathrm{Spec}(H_y) = \{k\,\log p\}_{k \in \mathbb{N}\cup\{0\}}; \tag{8.37}$$

hence π_y is a positive energy representation.

(2) We begin with the special case with $y_\infty = p^{-1}$. Then $\lambda = p$ and $f_p(\lambda, \beta) = \frac{p-1}{1-p^{-\beta}}$, since all digits of $\lambda = p$ are equal to $p-1$. We want to show that the partition function is given by

$$\mathrm{Tr}(e^{-\beta H_y}) = 2\,\frac{p-1}{1 - p^{1-\beta}}. \tag{8.38}$$

The multiplicity of an eigenvalue $k \log p$ of H_y is the number of integers $m \neq 0 \in \mathbb{Z}$ that are prime to p and such that $p^{-k}\,|m|\,y_\infty < 1$. Since we are assuming that $y_\infty = p^{-1}$, this gives $|m| < p^{k+1}$. Thus, the multiplicity is $2\,(p^{k+1} - p^k)$. The factor 2 comes from the sign of the integer m. The factor $(p^{k+1} - p^k)$

corresponds to subtracting from the number p^{k+1} of positive integers $m \leq p^{k+1}$ the number p^k of those that are multiples of p.

We now pass to the general case. For $x > 0$, $\lceil x - 1 \rceil$ is the cardinality of $(0, x) \cap \mathbb{N}$. The same argument used above shows that the multiplicity of the eigenvalue $k \log p$ is given by the counting

$$2\left(\lceil \lambda p^k - 1 \rceil - \lceil \lambda p^{k-1} - 1 \rceil\right).$$

Thus

$$\operatorname{Tr}(e^{-\beta H_y}) = 2 \sum_{k=0}^{\infty} \left(\lceil \lambda p^k - 1 \rceil - \lceil \lambda p^{k-1} - 1 \rceil\right) p^{-k\beta}. \tag{8.39}$$

One has the following equalities of convergent series,

$$\sum_{k=0}^{\infty} \left(\lceil \lambda p^k - 1 \rceil - \lceil \lambda p^{k-1} - 1 \rceil\right) p^{-k\beta} = \sum_{k=0}^{\infty} \lceil \lambda p^k - 1 \rceil \, (p^{-k\beta} - p^{-(k+1)\beta}),$$

so that

$$\operatorname{Tr}(e^{-\beta H_y}) = 2\,(1 - p^{-\beta}) \sum_{k=0}^{\infty} \lceil \lambda p^k - 1 \rceil \, p^{-k\beta}. \tag{8.40}$$

Similarly,

$$\sum_{k=0}^{\infty} \left(\lceil \lambda p^k - 1 \rceil - p\,\lceil \lambda p^{k-1} - 1 \rceil\right) p^{-k\beta} = \sum_{k=0}^{\infty} \lceil \lambda p^k - 1 \rceil \, (p^{-k\beta} - p\, p^{-(k+1)\beta}),$$

which gives

$$f_p(\lambda, \beta) = (1 - p^{1-\beta}) \sum_{k=0}^{\infty} \lceil \lambda p^k - 1 \rceil \, p^{-k\beta}, \tag{8.41}$$

since the coefficients c_k of the expansion of λ in base p are given by (8.30). Combining (8.40) with (8.41) gives (8.32).

(3) It follows from (8.22) and the finiteness of the partition function (8.32) that the functionals (8.33) fulfill the KMS_β condition. In terms of functions on the groupoid $\mathcal{G}(p)$ one has

$$\psi_{\beta,\, y}(f) = \sum f(1, n\, p^{-k}\, y)\, p^{-k\beta}, \quad \forall f \in C_c(\mathcal{G}(p)) \tag{8.42}$$

where the sum is absolutely convergent. Each of the terms in the sum gives a weakly continuous linear form, and thus one obtains the required continuity. ☐

Remark 8.16. The partition function $Z_p(\lambda, \beta)$ is a discontinuous function of the parameter λ, and this might seem to contradict the third statement of Theorem 8.15. It would if the algebra $C^*(\mathcal{G}(p))$ were unital, since in that case, the partition function is given by evaluation on the unit, and weak continuity implies that it is continuous. In our case $C^*(\mathcal{G}(p))$ is *not* unital, and the partition function is expressed as a supremum of the form

$$Z_p(\lambda,\beta)=\sup\{\psi_{\beta,\,y}(a^*a)|a\in C^*(\mathcal{G}(p))\,,\ ||a||\leq 1\}\,.$$

In particular, it shows that $Z_p(\lambda,\beta)$ is lower semicontinuous as a function of λ.

The precise qualitative properties of the partition functions $Z_p(\lambda,\beta)$ are described by the following result.

Proposition 8.17. *As a function of $\lambda\in(1,\lambda]$ the partition function $Z_p(\lambda,\beta)$ satisfies for $\beta>1$:*

1. *Z_p is strictly increasing.*
2. *Z_p is continuous on the left, and lower semicontinuous.*
3. *Z_p is discontinuous at any point of the form $\lambda=m\,p^{-k}$ with a jump of $2\,p^{-k\beta}$ (for m prime to p).*
4. *The measure $\frac{\partial Z_p}{\partial\lambda}$ is the sum of the Dirac masses at the points $\lambda=m\,p^{-k}$, m prime to p, with coefficients $2\,p^{-k\beta}$.*
5. *The closure of the range of Z_p is a Cantor set.*

Proof. (1) This follows from (8.40), which expresses Z_p as an absolutely convergent sum of multiples of the functions $\lceil\lambda p^k-1\rceil$. The latter are nondecreasing and jump by 1 at $\lambda\in\mathbb{N}\,p^{-k}\cap(1,p]$. The density of the union of these finite sets for $k\geq 0$ shows that Z_p is strictly increasing.

(2) This follows as above from (8.40) and the semicontinuity properties of the ceiling function.

(3) Let $\lambda=m\,p^{-k}$ with m prime to p. Then for any $j\geq k$ one gets a jump of $2\,(1-p^{-\beta})\,p^{-j\beta}$ coming from (8.40), so that their sum gives

$$2\,(1-p^{-\beta})\sum_{j=k}^{\infty}p^{-j\beta}=\,2\,p^{-k\beta}.$$

(4) This follows as above from (8.40) and from (8.3), which computes the discontinuity at the jumps.

(5) Recall that when writing elements of an interval in base p one gets a map from the cantor set to the interval. This map is surjective but fails to be injective due to the identifications coming from $\sum_0^\infty(p-1)\,p^{-m}=p$. The connectedness of the interval is recovered from these identifications. In our case the coefficients c_k of the expansion in base p of elements of $(1,p]$ are such that $c_0\in\{1,\ldots,p-1\}$, while $c_k\in\{0,\ldots,p-1\}$ for $k>0$. This is a Cantor set K in the product topology of $K=\{1,\ldots,p-1\}\times\prod_{\mathbb{N}}\{0,\ldots,p-1\}$. As shown in Figure 3, the discontinuities of the function $Z_p(\lambda,\beta)$ as a function of λ replace the connected topology of $(1,p]$ by the totally disconnected topology of K (Fig. 4). $\square$

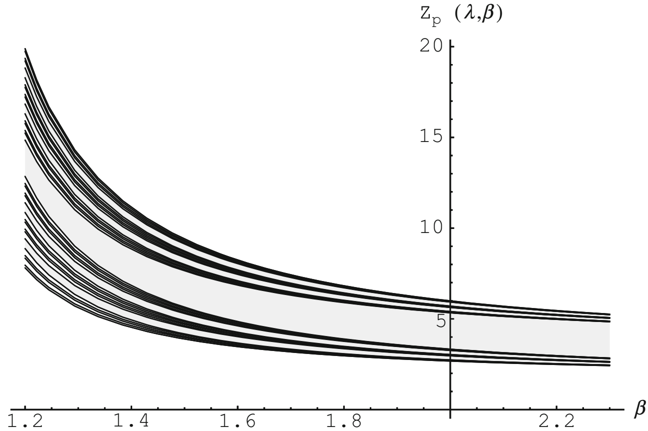


Fig. 4. Graphs of the functions $Z_p(\lambda,\beta)$ as functions of β for $p=3$, $\lambda = n/27$. The gray regions are the gaps in the range. All these functions have a pole at $\beta = 1$.

Remark 8.18. One can use (8.39) to define $Z_p(\lambda,\beta)$ for any $\lambda > 0$, as

$$Z_p(\lambda,\beta) = 2\sum_{-\infty}^{\infty}\left(\lceil \lambda p^k - 1\rceil - \lceil \lambda p^{k-1} - 1\rceil\right) p^{-k\beta}. \tag{8.43}$$

This makes sense for $\Re(\beta) > 1$, since $\lceil \lambda p^k - 1\rceil = 0$ for $k \leq -\frac{\log\lambda}{\log p}$. The extended function (8.43) satisfies

$$Z_p(p\lambda,\beta) = p^{\beta} Z_p(\lambda,\beta),$$

which suggests replacing $Z_p(\lambda,\beta)$ with

$$\zeta_p(\lambda,\beta) = \lambda^{-\beta} Z_p(\lambda,\beta), \tag{8.44}$$

so that

$$\zeta_p(p\lambda,\beta) = \zeta_p(\lambda,\beta). \tag{8.45}$$

This replacement $Z_p \mapsto \zeta_p$ corresponds to the shift in the Hamiltonian H_y by

$$H_y \mapsto H_y - \log|y_\infty|.$$

We can now refine Theorem 8.11 and consider the zero-temperature KMS state of the system $(C^*(\mathcal{G}(p)), \sigma_t^p)$ corresponding to the positive energy representation π_y for $y \in \Xi_{\mathbb{Q},\,p}$.

Proposition 8.19. *As $\beta \to \infty$ the vacuum states (zero temperature KMS states) of the system $(C^*(\mathcal{G}(p)), \sigma_t^p)$ with Hamiltonian H_y have a degeneracy of $2\lceil \lambda - 1\rceil$, where $y_\infty = \lambda^{-1}$. There is a preferred choice of a vacuum state given by the evaluation at $y \in \mathbb{A}^*_{\mathbb{Q},p}$.*

Proof. When we look at the orbit of $y \in \mathbb{A}^*_{\mathbb{Q},p}$, i.e., at the intersection $\mathbb{Q}^* y \cap \mathbb{A}^{(1)}_{\mathbb{Q},p}$, and label its elements by pairs (k,m) as above, we find that all elements with $k = 0$ give a ground state. This degeneracy of the vacuum reflects the fact that the limit of the partition function as the temperature goes to 0 is not in general equal to 1. For instance, for $y_\infty = p^{-1}$, one obtains

$$\lim_{\beta\to\infty} \mathrm{Tr}(e^{-\beta H_y}) = \lim_{\beta\to\infty} 2\frac{p-1}{1-p^{1-\beta}} = 2(p-1).$$

More generally, one similarly obtains the limit

$$\lim_{\beta\to\infty} \mathrm{Tr}(e^{-\beta H_y}) = 2\lceil \lambda - 1 \rceil.$$

Among the $2\lceil \lambda - 1\rceil$ vacuum states, the state given by evaluation at $y \in \mathbb{A}^*_{\mathbb{Q},p}$ is singled out, since $my \notin \mathbb{A}^*_{\mathbb{Q},p}$ for $m \neq 1$. It is then natural to consider, for each finite place $p \in \Sigma_{\mathbb{Q}}$, the section

$$s_p(x) = \mathbb{Q}^* x \cap \mathbb{A}^*_{\mathbb{Q},p}, \quad \forall x \in C_{\mathbb{Q}} a^{(p)} \subset \mathbb{A}_{\mathbb{Q}}/\mathbb{Q}^*, \tag{8.46}$$

of the projection from $\mathbb{A}_{\mathbb{Q}}$ to the orbit $C_{\mathbb{Q}} a^{(p)}$. $\square$

Notice that s_p is discontinuous at the boundary of the domain $\mathbb{A}^*_{\mathbb{Q},p}$. Indeed, when y_∞ crosses the value p^{-1} the class in $C_{\mathbb{Q}} a^{(p)}$ varies continuously but the representative in $\mathbb{A}^*_{\mathbb{Q},p}$ jumps discontinuously, so that its archimedean component remains in the interval $[p^{-1}, 1)$. This suggests that one consider a cyclic covering of $\Xi_{\mathbb{Q}}$, which we now discuss in Section 8.5.

8.5 The cyclic covering $\tilde{\Xi}_{\mathbb{Q}}$ of $\Xi_{\mathbb{Q}}$

By construction $\Xi_{\mathbb{K}}$ is a subspace of the adèles class space $X_{\mathbb{K}}$. We shall now show, in the case $\mathbb{K} = \mathbb{Q}$, that it admits a natural lift $\tilde{\Xi}_{\mathbb{Q}}$ to a subspace of $\mathbb{A}_{\mathbb{Q}}$ that reduces the ambiguity group $\mathbb{Q}^*$ to a cyclic group. One thus obtains a natural cyclic covering $\tilde{\Xi}_{\mathbb{Q}} \subset \mathbb{A}_{\mathbb{Q}}$ of $\Xi_{\mathbb{Q}}$. We already saw above, in Proposition 8.19, that it is natural to choose representatives for the elements of the orbit $C_{\mathbb{Q}} a^{(p)}$, for a finite prime p, in the subset of adèles given by

$$\tilde{\Xi}_{\mathbb{Q},p} := \{y \in \mathbb{A}_{\mathbb{Q}} \mid y_p = 0 \text{ and } |y_\ell| = 1 \text{ for } \ell \neq p, \infty \text{ and } y_\infty > 0\}. \tag{8.47}$$

We extend this definition at ∞ by

$$\tilde{\Xi}_{\mathbb{Q},\infty} := \{y \in \mathbb{A}_{\mathbb{Q}} \mid |y_w| = 1 \ \forall w \neq \infty \text{ and } y_\infty = 0\}. \tag{8.48}$$

Definition 8.20. *The locus $\tilde{\Xi}_{\mathbb{Q}} \subset \mathbb{A}_{\mathbb{Q}}$ is defined as*

$$\tilde{\Xi}_{\mathbb{Q}} = \bigcup_{v \in \Sigma_{\mathbb{Q}}} \tilde{\Xi}_{\mathbb{Q},v} \subset \mathbb{A}_{\mathbb{Q}}. \tag{8.49}$$

We then have the following simple fact.

Proposition 8.21. *Let π be the projection from $\tilde{\Xi}_{\mathbb{Q}}$ to $\Xi_{\mathbb{Q}}$, with $\pi(x)$ the class of x modulo the action of $\mathbb{Q}^*$.*

1. *The map $\pi : \tilde{\Xi}_{\mathbb{Q}} \to \Xi_{\mathbb{Q}}$ is surjective.*
2. *Two elements in $\tilde{\Xi}_{\mathbb{Q},v}$ have the same image in $C_{\mathbb{Q}} a^{(v)}$ iff they are on the same orbit of the following transformation T:*

$$Tx = p\,x\,, \quad \forall x \in \tilde{\Xi}_{\mathbb{Q},p}\,, \quad Tx = -x\,, \quad \forall x \in \tilde{\Xi}_{\mathbb{Q},\infty}. \tag{8.50}$$

Proof. The first statement follows by lifting $C_{\mathbb{Q}}$ inside $\mathbb{A}_{\mathbb{Q}}^*$ as the subgroup $\hat{\mathbb{Z}}^* \times \mathbb{R}_+^*$. Then any element of $C_{\mathbb{Q}} a^{(v)}$ has a representative in $\left(\hat{\mathbb{Z}}^* \times \mathbb{R}_+^*\right) a^{(v)}$.

The proof of the second statement is straightforward, since for a finite prime p the subgroup $p^{\mathbb{Z}} \subset \mathbb{Q}^*$ is the group of elements of $\mathbb{Q}^*$ that leave $\tilde{\Xi}_{\mathbb{Q},p}$ globally invariant. □

8.6 Arithmetic subalgebra, Frobenius and monodromy

We now describe a natural algebra of coordinates $\mathcal{B}$ on $\Xi_{\mathbb{Q}}$.

The BC system of [5], as well as its arithmetic generalizations of [12] and [14], have the important property that they come endowed with an arithmetic structure given by an arithmetic subalgebra. The general framework of endomotives developed in [11] shows a broad class of examples in which a similar arithmetic structure is naturally present. We consider here the issue of extending the construction of the "rational subalgebra" of the BC-system to the algebra $\mathcal{S}\left(\mathcal{G}_{\mathbb{Q}}^{(1)}\right)$ of Section 8.1.

In order to get a good geometric picture it is convenient to think in terms of $\mathbb{Q}$-lattices rather than of adèles, as in [12]. Thus, we let $\mathcal{L}$ denote the set of 1-dimensional $\mathbb{Q}$-lattices (as defined in [12]). We consider the map

$$\iota : \hat{\mathbb{Z}} \times \mathbb{R}_+^* \to \mathcal{L}\,, \quad \iota(\rho, \lambda) = (\Lambda, \phi) = (\lambda^{-1}\mathbb{Z}, \lambda^{-1}\rho), \tag{8.51}$$

which associates to an adèle $(\rho, \lambda) \in \hat{\mathbb{Z}} \times \mathbb{R}_+^* \subset \mathbb{A}_{\mathbb{Q}}$ the $\mathbb{Q}$-lattice obtained using ρ to label the torsion points of $\mathbb{R}/\lambda^{-1}\mathbb{Z}$. Replacing (ρ, λ) by $(n\rho, n\lambda)$, for a positive integer $n \in \mathbb{N}$, one obtains the pair $\left(\frac{1}{n}\Lambda, \phi\right)$, which is commensurable with (Λ, ϕ). Thus, the action of $\mathbb{Q}_+^*$ corresponds to commensurability of $\mathbb{Q}$-lattices under the map ι. Multiplying λ by a positive scalar corresponds to the scaling action of $\mathbb{R}_+^*$ on $\mathbb{Q}$-lattices.

Let us recall the definition of the "rational algebra" $\mathcal{A}_{\mathbb{Q}}$ of [12] for the BC system, given in terms of $\mathbb{Q}$-lattices. We let

$$\epsilon_a(\Lambda, \phi) = \sum_{y \in \Lambda + \phi(a)} y^{-1}, \tag{8.52}$$

for any $a \in \mathbb{Q}/\mathbb{Z}$. This is well defined, for $\phi(a) \neq 0$, using the summation $\lim_{N\to\infty} \sum_{-N}^{N}$, and is zero by definition for $\phi(a) = 0$. The function

$$\varphi_a(\rho, \lambda) = \epsilon_a(\iota(\rho, \lambda)), \quad \forall (\rho, \lambda) \in \hat{\mathbb{Z}} \times \mathbb{R}_+^*, \tag{8.53}$$

is well defined and homogeneous of degree 1 in λ. Moreover, for fixed $a \in \mathbb{Q}/\mathbb{Z}$ with denominator m, it depends only upon the projection of ρ on the finite group $\mathbb{Z}/m\mathbb{Z}$; hence it defines a continuous function on $\hat{\mathbb{Z}} \times \mathbb{R}_+^*$. Using the degree 1 homogeneity in λ, one gets that (8.53) extends by continuity to 0 on $\hat{\mathbb{Z}} \times \{0\}$.

One gets functions that are homogeneous of weight zero by taking the derivatives of the functions φ_a. The functions

$$\psi_a(\rho, \lambda) = \frac{1}{2\pi i} \frac{d}{d\lambda} \varphi_a(\rho, \lambda), \quad \forall (\rho, \lambda) \in \hat{\mathbb{Z}} \times \mathbb{R}_+^*, \tag{8.54}$$

are independent of λ; hence they define continuous functions on $\mathbb{A}_{\mathbb{Q}}^{(1)}$. They are nontrivial on $\tilde{\Xi}_{\mathbb{Q},\infty} = \hat{\mathbb{Z}}^* \times \{0\} \subset \hat{\mathbb{Z}} \times \{0\}$ and they agree there with the functions e_a of [12].

Proposition 8.22. *Let $\mathcal{B}$ be the algebra generated by the φ_a and ψ_a defined in* (8.53) *and* (8.54) *above.*

1. *The expression*
$$N(f) = \frac{1}{2\pi i} \frac{d}{d\lambda} f \tag{8.55}$$
defines a derivation N of $\mathcal{B}$.
2. *The algebra $\mathcal{B}$ is stable under the derivation Y that generates the* 1*-parameter semigroup $F(\mu)$ of endomorphisms of $\mathcal{S}\left(\mathcal{G}_{\mathbb{Q}}^{(1)}\right)$ of* (8.11)*, and one has, at the formal level, the relation*
$$F(\mu)N = \mu N F(\mu). \tag{8.56}$$
3. *For any element $f \in \mathcal{B}$ one has*
$$\alpha \circ f(x) = f(\tilde{\alpha}\, x), \quad \forall x \in \tilde{\Xi}_{\mathbb{Q},\infty}, \quad \textit{and} \ \ \forall \alpha \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{cycl}}/\mathbb{Q}), \tag{8.57}$$
where $\tilde{\alpha} \in \hat{\mathbb{Z}}^* \subset C_{\mathbb{Q}}$ is the element of the idèle class group associated to $\alpha \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{cycl}}/\mathbb{Q})$ by the class field theory isomorphism.

Proof. (1) By construction, N is a derivation of the algebra of functions. Moreover, (8.54) shows that $N(\varphi_a) = \psi_a$, while $N(\psi_a) = 0$. Thus, the derivation rule shows that $\mathcal{B}$ is stable under N.

(2) The derivation generating the one-parameter semigroup $F(\mu)$ is given, up to sign, by the grading operator

$$Y(f) = \lambda \frac{d}{d\lambda} f. \tag{8.58}$$

By construction, any of the φ_a is of degree one, i.e., $Y(\varphi_a) = \varphi_a$, and ψ_a is of degree 0. Thus, again the derivation rule shows that $\mathcal{B}$ is stable under Y.

(3) This involves only the functions ψ_a, since by construction the restriction of φ_a is zero on $\tilde{\Xi}_{\mathbb{Q},\infty}$. The result then follows from the main result of [5] in the reformulation given in [12] (see also [13, Chapter 3]). In fact, all these functions take values in the cyclotomic field $\mathbb{Q}^{\mathrm{cycl}} \subset \mathbb{C}$, and they intertwine the action of the discontinuous piece $\hat{\mathbb{Z}}^*$ of $C_{\mathbb{Q}}$ with the action of the Galois group of $\mathbb{Q}^{\mathrm{cycl}}$. $\square$

This is in agreement with viewing the algebra $\mathcal{B}$ as the algebra of coordinates on $\tilde{\Xi}_{\mathbb{Q}}$. Indeed, in the case of a global field $\mathbb{K}$ of positive characteristic, the action of the Frobenius on the points of $C(\bar{\mathbb{F}}_q)$ (which have coordinates in $\bar{\mathbb{F}}_q$) corresponds to the Frobenius map

$$\mathrm{Fr} : u \mapsto u^q, \qquad \forall u \in \mathbb{K}, \tag{8.59}$$

of the function field $\mathbb{K}$ of the curve C. The Frobenius endomorphism $u \mapsto u^q$ of $\mathbb{K}$ is the operation that replaces a function $f : C(\bar{\mathbb{F}}_q) \to \bar{\mathbb{F}}_q$ by its q-th power, i.e., the composition $\mathrm{Fr}\circ f$ with the Frobenius automorphism $\mathrm{Fr} \in \mathrm{Gal}(\bar{\mathbb{F}}_q/\mathbb{F}_q)$. For $f \in \mathbb{K}$, one has

$$\mathrm{Fr} \circ f = f^q = f \circ \mathrm{Fr}, \tag{8.60}$$

where on the right-hand side Fr is the map that raises every coordinate to the power q. This corresponds to the intertwining with the Galois action discussed above.

Notice, moreover, that as we have seen in Proposition 8.6, only the semigroup $C_{\mathbb{Q}}^{+}$ acts on the reduced system $\mathcal{S}\left(\mathcal{G}_{\mathbb{Q}}^{(1)}\right)$, and it acts by endomorphisms. It nevertheless acts in a bijective manner on the points of $\Xi_{\mathbb{Q}}$. This is similar to what happens with the Frobenius endomorphism (8.59), which is only an endomorphism of the field of functions $\bar{\mathbb{K}}$, while it acts bijectively (as a Galois automorphism of the coordinates) on the points of $C(\bar{\mathbb{F}}_q)$.

Further notice that there is a striking formal analogy between the operators F and N of Proposition 8.22 satisfying the relation (8.56) and the Frobenius and local monodromy operators introduced in the context of the “special fiber at arithmetic infinity” in Arakelov geometry (see [18], [19]). In particular, one should compare (8.56) with [19, §2.5] which discusses a notion of the Weil–Deligne group at arithmetic infinity.

9 Functoriality of the adèles class space

We investigate in this section the functoriality of the adèles class space $X_{\mathbb{K}}$ and of its classical subspace $\Xi_{\mathbb{K}} \subset X_{\mathbb{K}}$, for Galois extensions of the global field $\mathbb{K}$.

This issue is related to the question of functoriality. Namely, given a finite algebraic extension $\mathbb{L}$ of the global field $\mathbb{K}$, we want to relate the adèles class spaces of both fields. Assume that the extension is a Galois extension. In general, we do not expect the relation between the adèles class spaces to be canonical, in the sense that it will involve a symmetry-breaking choice on the Galois group $G = \mathrm{Gal}(\mathbb{L}/\mathbb{K})$ of the extension. More precisely, the norm map

$$\mathfrak{n}(a) = \prod_{\sigma \in G} \sigma(a) \in \mathbb{A}_{\mathbb{K}}, \quad \forall a \in \mathbb{A}_{\mathbb{L}}, \tag{9.1}$$

appears to be the obvious candidate that relates the two adèles class spaces. In fact, since $\mathfrak{n}(\mathbb{L}) \subset \mathbb{K}$, the map (9.1) passes to the quotient and gives a natural map from $X_{\mathbb{L}} = \mathbb{A}_{\mathbb{L}}/\mathbb{L}^*$ to $X_{\mathbb{K}} = \mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$ that looks like the covering required by functoriality. However, the problem is that the norm map fails to be surjective in general, hence it certainly does not qualify as a covering map. In fact, this problem already occurs at the level of the idèle class group $C_{\mathbb{K}}$; namely, the norm map fails to be a surjection from $C_{\mathbb{L}}$ to $C_{\mathbb{K}}$.

The correct object to consider is the Weil group $W_{\mathbb{L},\mathbb{K}}$. This is an extension of $C_{\mathbb{L}}$ by the Galois group $G = \mathrm{Gal}(\mathbb{L}/\mathbb{K})$, which is not a semidirect product. The corresponding nontrivial 2-cocycle is called the "fundamental class". One has a natural morphism t, called the *transfer*, from $W_{\mathbb{L},\mathbb{K}}$ to $C_{\mathbb{K}}$. The transfer satisfies the following two properties:

- The morphism t restricts to the norm map from $C_{\mathbb{L}}$ to $C_{\mathbb{K}}$.
- The morphism t is surjective on $C_{\mathbb{K}}$.

Thus, the correct way to understand the relation between the adèles class spaces $X_{\mathbb{L}}$ and $X_{\mathbb{K}}$ is by extending the construction of the Weil group and of the transfer map.

One obtains in this way n copies of the adèles class space $X_{\mathbb{L}}$ of $\mathbb{L}$ and a map to $X_{\mathbb{K}}$ that is now a covering from $G \times \Xi_{\mathbb{L}} \to \Xi_{\mathbb{K}}$. This space has a natural action of the Weil group. We explain this in more detail in what follows.

9.1 The norm map

We begin by recalling the well-known properties of the norm map that are relevant to our setup. Thus, we let $\mathbb{L} \supset \mathbb{K}$ be a finite Galois extension of $\mathbb{K}$ of degree n, with $G = \mathrm{Gal}(\mathbb{L}/\mathbb{K})$ the Galois group.

Since the adèles depend naturally on the field, one has a canonical action of G on $\mathbb{A}_{\mathbb{L}}$. If $v \in \Sigma_{\mathbb{K}}$ is a place of $\mathbb{K}$, there are m_v places of $\mathbb{L}$ over v and they are permuted transitively by the action of G. Let G_w be the isotropy subgroup of w. Then G_w is the Galois group $G_w = \mathrm{Gal}(\mathbb{L}_w/\mathbb{K}_v)$.

One has a canonical embedding of $\mathbb{A}_{\mathbb{K}}$ as the fixed points of the action of G on $\mathbb{A}_{\mathbb{L}}$ by

$$\mathbb{A}_{\mathbb{K}} = \mathbb{A}_{\mathbb{L}}^{G}, \qquad (a_v) \mapsto (a_{\pi(w)}), \qquad \text{with} \quad \pi : \Sigma_{\mathbb{L}} \to \Sigma_{\mathbb{K}}. \tag{9.2}$$

The norm map $\mathfrak{n} : \mathbb{A}_{\mathbb{L}} \to \mathbb{A}_{\mathbb{K}}$ is then defined as in (9.1). By [44, IV 1, Corollary 3], it is given explicitly by

$$\mathfrak{n}(x) = z\,, \quad z_v = \prod_{w|v} \mathfrak{n}_{\mathbb{L}_w/\mathbb{K}_v}(x_w)\,, \quad \forall x \in \mathbb{A}_{\mathbb{L}}\,. \tag{9.3}$$

Here the notation $w|v$ means that w is a place of $\mathbb{L}$ over the place $v \in \Sigma_{\mathbb{K}}$. Also $\mathfrak{n}_{\mathbb{L}_w/\mathbb{K}_v}$ is the norm map of the extension $\mathbb{L}_w/\mathbb{K}_v$. When restricted to principal adèles of $\mathbb{L}$ it gives the norm map from $\mathbb{L}$ to $\mathbb{K}$. When restricted to the subgroup $\mathbb{L}_w^* = (\ldots, 1, \ldots, y, \ldots, 1, \ldots) \subset \mathbb{A}_{\mathbb{L}}^*$, it gives the norm map of the extension $\mathbb{L}_w/\mathbb{K}_v$. For nontrivial extensions this map is never surjective, but its restriction $\mathfrak{n} : \mathcal{O}(\mathbb{L}_w)^* \to \mathcal{O}(\mathbb{K}_v)^*$ is surjective when the extension is unramified, which is the case for almost all places $v \in \Sigma_{\mathbb{K}}$ (cf. [44, Theorem 1 p.153]). In such cases, the module of the subgroup $\mathfrak{n}\,(\mathbb{L}_w^*) \subset \mathbb{K}_v^*$ is a subgroup of index the order of the extension $\mathbb{K}_v \subset \mathbb{L}_w$. The restriction of the norm map to the idelè group $\mathbb{A}_{\mathbb{L}}^*$ is very far from being surjective to $\mathbb{A}_{\mathbb{K}}^*$, and its range is a subgroup of infinite index. The situation is much better with the idèle class groups, since (cf. [44, Corollary p. 153]) the norm map is an open mapping $\mathfrak{n} : C_{\mathbb{L}} \to C_{\mathbb{K}}$ whose range is a subgroup of finite index.

9.2 The Weil group and the transfer map

The Weil group $W_{\mathbb{L},\mathbb{K}}$ associated to the Galois extension $\mathbb{K} \subset \mathbb{L}$ is an extension

$$1 \to C_{\mathbb{L}} \to W_{\mathbb{L},\mathbb{K}} \to G \to 1 \tag{9.4}$$

of $C_{\mathbb{L}}$ by the Galois group G. One chooses a section s from G and lets $a \in Z^2(G, C_{\mathbb{L}})$ be the corresponding 2-cocycle, so that

$$a_{\alpha,\beta} = s_{\alpha\beta}^{-1}\, s_\alpha\, s_\beta, \qquad \forall \alpha, \beta \in G. \tag{9.5}$$

The algebraic rules in $W_{\mathbb{L},\mathbb{K}}$ are then given by

$$s_\alpha\, s_\beta = s_{\alpha\beta}\, a_{\alpha,\beta}, \qquad \forall \alpha, \beta \in G, \tag{9.6}$$

and

$$s_\alpha\, x\, s_\alpha^{-1} = \alpha(x), \qquad \forall \alpha \in G, \;\; \forall x \in C_{\mathbb{L}}. \tag{9.7}$$

The transfer homomorphism

$$t : W_{\mathbb{L},\mathbb{K}} \to C_{\mathbb{K}} \tag{9.8}$$

is then given by

$$t(x) = \mathfrak{n}(x), \quad \forall x \in C_{\mathbb{L}} \quad \text{ and } \quad t(s_\alpha) = \prod_\beta a_{\alpha,\beta}, \quad \forall \alpha \in G. \tag{9.9}$$

Its main properties are the following (see [40]).

- t is a surjective group morphism $W_{\mathbb{L},\mathbb{K}} \to C_{\mathbb{K}}$.
- Let $W_{\mathbb{L},\mathbb{K}}^{ab}$ be the abelian quotient of $W_{\mathbb{L},\mathbb{K}}$ by the closure of its commutator subgroup $W_{\mathbb{L},\mathbb{K}}^c$. Then t induces an isomorphism of $W_{\mathbb{L},\mathbb{K}}^{ab}$ with $C_{\mathbb{K}}$.

9.3 The covering

We finally describe the resulting functoriality of the adèles class spaces in terms of a covering map obtained by extending the Weil group and transfer map described above. Let, as above, $\mathbb{L} \supset \mathbb{K}$ be a finite Galois extension of $\mathbb{K}$.

Lemma 9.1. *The transfer map extends to a map*

$$\tau : G \times X_{\mathbb{L}} \to X_{\mathbb{K}} \tag{9.10}$$

of the adèle class spaces.

Proof. We endow $G \times X_{\mathbb{L}}$ with a two-sided action of G compatible with τ. By construction the norm map $\mathfrak{n}$ of (9.1) is well defined on $\mathbb{A}_{\mathbb{L}}$. Since it is multiplicative and we have $\mathfrak{n}(\mathbb{L}^*) \subset \mathbb{K}^*$, it induces a map of quotient spaces $\mathfrak{n} : X_{\mathbb{L}} \to X_{\mathbb{K}}$. By construction $C_{\mathbb{L}}$ acts on $X_{\mathbb{L}}$, and the actions by left and right multiplication coincide, so we use both notations. We define the map τ as

$$\tau : G \times X_{\mathbb{L}} \to X_{\mathbb{K}}, \qquad \tau(\alpha, x) = t(s_\alpha)\mathfrak{n}(x), \quad \forall x \in X_{\mathbb{L}},\ \forall \alpha \in G. \tag{9.11}$$

This makes sense, since $t(s_\alpha) \in C_{\mathbb{K}}$ and $C_{\mathbb{K}}$ acts on $X_{\mathbb{K}}$. By construction, the restriction of τ to $G \times C_{\mathbb{L}}$ is the transfer map. ☐

One identifies $G \times C_{\mathbb{L}}$ with $W_{\mathbb{L},\mathbb{K}}$ by the map that to $(\alpha, g) \in G \times C_{\mathbb{L}}$ associates the element $s_\alpha\, g$ of $W_{\mathbb{L},\mathbb{K}}$.

In the following we use the notation

$$x^g = g^{-1}(x). \tag{9.12}$$

We have the following result.

Lemma 9.2. *Let $\mathbb{L} \supset \mathbb{K}$ be a finite Galois extension of $\mathbb{K}$.*

1. *The expressions*

$$s_\alpha g(\beta, x) = (\alpha\beta, a_{\alpha,\beta} g^\beta x) \quad \textit{and} \quad (\alpha, x) s_\beta g = (\alpha\beta, a_{\alpha,\beta} x^\beta g) \tag{9.13}$$

define a left and a right action of $W_{\mathbb{L},\mathbb{K}}$ on $G \times X_{\mathbb{L}}$.

2. *The map τ of (9.11) satisfies the equivariance property*

$$\tau(gxk) = t(g)\tau(x)t(k), \qquad \forall x \in G \times X_{\mathbb{L}}, \quad \textit{and} \quad \forall g, k \in W_{\mathbb{L},\mathbb{K}}. \tag{9.14}$$

Proof. (1) We defined the rules (9.6) as the natural extension of the multiplication in $W_{\mathbb{L},\mathbb{K}}$ using

$$s_\alpha g s_\beta h = s_\alpha s_\beta g^\beta h = s_{\alpha\beta} a_{\alpha,\beta} g^\beta h. \tag{9.15}$$

Thus, the proof of associativity in the group $W_{\mathbb{L},\mathbb{K}}$ extends, and it implies that (9.13) defines a left and a right action of $W_{\mathbb{L},\mathbb{K}}$ and that these two actions commute.

(2) The proof that the transfer map t is a group homomorphism extends to give the required equality, since the norm map is a bimodule morphism when extended to $X_{\mathbb{L}}$. ☐

At the level of the classical points, we can then describe the covering map in the following way.

Proposition 9.3. *Let* $\mathbb{L}$ *and* $\mathbb{K}$ *be as above.*

1. The restriction of τ *to* $G \times \Xi_{\mathbb{L}} \subset G \times X_{\mathbb{L}}$ *defines a surjection*

$$\tau : G \times \Xi_{\mathbb{L}} \to \Xi_{\mathbb{K}}. \tag{9.16}$$

2. The map τ *induces a surjection*

$$\tau : G \times (\Xi_{\mathbb{L}}/C_{\mathbb{L},1}) \to \Xi_{\mathbb{K}}/C_{\mathbb{K},1}. \tag{9.17}$$

Proof. (1) By construction, $\Xi_{\mathbb{L}} = \cup_{w\in\Sigma_{\mathbb{L}}} C_{\mathbb{L}} a^{(w)}$, where $a^{(w)} \in X_{\mathbb{L}}$ is the class, modulo the action of $\mathbb{L}^*$, of the adèle with all entries equal to 1 except for a zero at w as in (8.7). Let π denote the natural surjection from $\Sigma_{\mathbb{L}}$ to $\Sigma_{\mathbb{K}}$. One has

$$\tau(1, a^{(w)}) = a^{(\pi(w))}, \quad \forall w \in \Sigma_{\mathbb{L}}. \tag{9.18}$$

In fact, one has $\tau(1, a^{(w)}) = \mathfrak{n}(a^{(w)})$. Moreover, by (9.3), the adèle $a = \mathfrak{n}(a^{(w)})$ has components $a_z = 1$ for all $z \neq \pi(w)$ and $a_{\pi(w)} = 0$. Thus $a = a^{(\pi(w))}$. The equivariance of the map τ as in Lemma 9.2 together with the surjectivity of the transfer map from $W_{\mathbb{L},\mathbb{K}}$ to $C_{\mathbb{K}}$ then shows that we have

$$\tau(W_{\mathbb{L},\mathbb{K}}(1, a^{(w)})) = C_{\mathbb{K}}\, a^{(\pi(w))}, \quad \forall w \in \Sigma_{\mathbb{L}}.$$

For $s_\alpha g \in W_{\mathbb{L},\mathbb{K}}$, one has

$$s_\alpha g(1, a^{(w)}) = (\alpha, g a^{(w)}),$$

since $a_{\alpha,1} = 1$. Thus, $W_{\mathbb{L},\mathbb{K}}(1, a^{(w)}) = G \times C_{\mathbb{L}} a^{(w)}$ and one gets

$$\tau(G \times C_{\mathbb{L}} a^{(w)}) = C_{\mathbb{K}} a^{(\pi(w))}, \quad \forall w \in \Sigma_{\mathbb{L}}.$$

Since the map π is surjective we get the conclusion.

(2) The transfer map satisfies $t(C_{\mathbb{L},1}) \subset C_{\mathbb{K},1}$. When restricted to the subgroup $C_{\mathbb{L}}$ the transfer coincides with the norm map $\mathfrak{n}$ and in particular, if $|g| = 1$ one has $|\mathfrak{n}(g)| = 1$. Thus one obtains a surjection of the quotient spaces

$$\tau : (G \times \Xi_{\mathbb{L}})/C_{\mathbb{L},1} \to \Xi_{\mathbb{K}}/C_{\mathbb{K},1}.$$

Moreover, the right action of the subgroup $C_{\mathbb{L},1} \subset W_{\mathbb{L},\mathbb{K}}$ is given by

$$(\alpha, x)g = (\alpha, xg).$$

This means that we can identify

$$(G \times \Xi_{\mathbb{L}})/C_{\mathbb{L},1} \sim G \times (\Xi_{\mathbb{L}}/C_{\mathbb{L},1}).$$

∎

9.4 The function field case

Let $\mathbb{K} = \mathbb{F}_q(C)$ be a global field of positive characteristic, identified with the field of rational functions on a nonsingular curve C over $\mathbb{F}_q$. We consider the extensions

$$\mathbb{L} = \mathbb{K} \otimes_{\mathbb{F}_q} \mathbb{F}_{q^n}. \tag{9.19}$$

The Galois group G is the cyclic group of order n with generator $\sigma \in \mathrm{Gal}(\mathbb{L}/\mathbb{K})$ given by $\sigma = \mathrm{id} \otimes \mathrm{Fr}$, where $\mathrm{Fr} \in \mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)$ is the Frobenius automorphism. Given a point $x \in C(\bar{\mathbb{F}}_q)$ we let n be the order of its orbit under the Frobenius. One then has $x \in C(\mathbb{F}_{q^n})$, and evaluation at x gives a well-defined place $w(x) \in \Sigma_{\mathbb{L}}$. The projection $\pi(w(x)) \in \Sigma_{\mathbb{K}}$ is a well-defined place of $\mathbb{K}$ that is invariant under $x \mapsto \mathrm{Fr}(x)$.

In the isomorphism of $\mathbb{Z}$-spaces

$$\vartheta_{\mathbb{L}} : C(\bar{\mathbb{F}}_q) \to \Xi_{\mathbb{L}}/C_{\mathbb{L},1}$$

described in Section 8.3, we have no ambiguity for places corresponding to points $x \in C(\mathbb{F}_{q^n})$. To such a point we assign simply

$$\vartheta_{\mathbb{L}}(x) = a^{(w(x))} \in \Xi_{\mathbb{L}}/C_{\mathbb{L},1}.$$

We now describe what happens with these points of $\Xi_{\mathbb{L}}/C_{\mathbb{L},1}$ under the covering map τ. We first need to see explicitly why the surjectivity occurs only after crossing by G.

Proposition 9.4. *Let $\mathbb{K}$ and $\mathbb{L} = \mathbb{K} \otimes_{\mathbb{F}_q} \mathbb{F}_{q^n}$ be as above.*

1. *The image $\mathfrak{n}(C_{\mathbb{L}}) \subset C_{\mathbb{K}}$ is the kernel of the morphism from $C_{\mathbb{K}}$ to $G = \mathbb{Z}/n\mathbb{Z}$ given by*
$$g \mapsto \rho(g) = \log_q |g| \mod n.$$
2. *One has $\rho(t(s_\sigma)) = 1 \mod n$, where $\sigma \in \mathrm{Gal}(\mathbb{L}/\mathbb{K})$ is the Frobenius generator.*

Proof. Since $\mathbb{L}$ is an abelian extension of $\mathbb{K}$, one has the inclusions

$$\mathbb{K} \subset \mathbb{L} \subset \mathbb{K}^{\mathrm{ab}} \subset \mathbb{L}^{\mathrm{ab}}, \tag{9.20}$$

where $\mathbb{K}^{ab}$ is the maximal abelian extension of $\mathbb{K}$. Using the class field theory isomorphisms

$$C_{\mathbb{K}} \sim W(\mathbb{K}^{ab}/\mathbb{K}) \quad \text{and} \quad C_{\mathbb{L}} \sim W(\mathbb{L}^{ab}/\mathbb{L}),$$

one can translate the proposition in terms of Galois groups. The result then follows using [40, p. 502]. ☐

10 Vanishing cycles: an analogy

We begin by considering some simple examples that illustrate some aspects of the geometry of the adèles class space, by restricting to the semilocal case of a finite number of places. This will also illustrate more explicitly the idea of considering the adèles class space as a noncommutative compactification of the idelè class group.

We draw an analogy between the complement of the idèle classes in the adèle classes and the singular fiber of a degeneration. This analogy should be taken with a big grain of salt, since this complement is a highly singular space and it really makes sense only as a noncommutative space in the motivic sense described in Sections 4 and 5 above.

10.1 Two real places

We first consider the example of the real quadratic field $\mathbb{K} = \mathbb{Q}(\sqrt{2})$ and we restrict to its two real places v_1 and v_2. Thus, we replace the adèles $\mathbb{A}_{\mathbb{K}}$ simply by the product $\mathbb{K}_{v_1} \times \mathbb{K}_{v_2}$ over the real places, which is just the product of two copies of $\mathbb{R}$. The idèles $\mathbb{A}^*_{\mathbb{K}}$ are correspondingly replaced by $\mathbb{K}^*_{v_1} \times \mathbb{K}^*_{v_2}$, and the inclusion of idèles in adèles is simply given by the inclusion

$$(\mathbb{R}^*)^2 \subset \mathbb{R}^2. \tag{10.1}$$

The role of the action of the group $\mathbb{K}^*$ by multiplication is now replaced by the action by multiplication of the group U of units of $\mathbb{K} = \mathbb{Q}(\sqrt{2})$. This group is

$$U = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z},$$

where the $\mathbb{Z}/2\mathbb{Z}$ comes from ± 1 and the $\mathbb{Z}$ is generated by the unit $u = 3-2\sqrt{2}$. Its action on $\mathbb{R}^2$ is given by the transformation

$$S(x, y) = (ux, u^{-1}y). \tag{10.2}$$

Thus, in this case of two real places, the semilocal version of the adèles class space is the quotient

$$X_{v_1,v_2} := \mathbb{R}^2/U \tag{10.3}$$

of $\mathbb{R}^2$ by the symmetry $(x, y) \mapsto (-x, -y)$ and the transformation S.

Both of these transformations preserve the function

$$\tilde{f} : \mathbb{R}^2 \to \mathbb{R}, \quad \tilde{f}(x, y) = xy, \tag{10.4}$$

which descends to a function

$$f : X_{v_1,v_2} \to \mathbb{R}. \tag{10.5}$$

Moreover, one has

$$(x, y) \in (\mathbb{R}^*)^2 \subset \mathbb{R}^2 \Leftrightarrow f(x, y) \neq 0,$$

and the fiber of f over any nonzero $\varepsilon \in \mathbb{R}$ is easily identified with a one-dimensional torus

$$f^{-1}(\varepsilon) \sim \mathbb{R}_+^*/u^{\mathbb{Z}}, \quad \forall \varepsilon \neq 0, \tag{10.6}$$

where one can use the map $(x, y) \mapsto |x|$ to obtain the required isomorphism.

The fiber $f^{-1}(0)$ of f over the point $\varepsilon = 0$, on the other hand, is no longer a one-dimensional torus and it is singular. It is the union of three pieces

$$f^{-1}(0) = T_1 \cup T_2 \cup \{0\} \tag{10.7}$$

corresponding respectively to

- T_1 is the locus $x = 0$, $y \neq 0$, which is a torus $T_1 \sim \mathbb{R}_+^*/u^{\mathbb{Z}}$ under the identification given by the map $(x, y) \mapsto |y|$.
- T_2 is the locus $x \neq 0$, $y = 0$, which is also identified with a torus $T_2 \sim \mathbb{R}_+^*/u^{\mathbb{Z}}$ under the analogous map $(x, y) \mapsto |x|$.
- The last piece is the single point $x = 0$, $y = 0$.

One can see that at the naive level, the quotient topology on the singular fiber (10.7) looks as follows. For any point $x \in T_j$ its closure is $\bar{x} = \{x\} \cup \{0\}$. Moreover, the point 0 is closed, and the induced topology on its complement is the same as the disjoint union of two one-dimensional tori T_j. In fact, one can be more precise and see what happens by analyzing the C^*-algebras involved. The C^*-algebra A associated to the singular fiber is by construction the crossed product

$$A = C_0(\tilde{f}^{-1}(0)) \rtimes U \tag{10.8}$$

with $\tilde{f}$ as in (10.4). One lets

$$A_j = C_0(V_j) \rtimes U, \tag{10.9}$$

where we use the restriction of the action of U to the subsets

$$V_j = \{(x_1, x_2) \,|\, x_j = 0\} \sim \mathbb{R}.$$

Evaluation at $0 \in \mathbb{R}$ gives a homomorphism

$$\epsilon_j : A_j \to C^*(U)\,.$$

Lemma 10.1. *One has an exact sequence of the form*

$$0 \to C(T_j) \otimes \mathcal{K} \to A_j \xrightarrow{\epsilon_j} C^*(U) \to 0,$$

where $\mathcal{K}$ is the algebra of compact operators.

The C^*-algebra A is the fibered product of the A_j over $C^*(U)$ using the morphisms ϵ_j.

Proof. The first statement follows using the fact that the action of U on $\mathbb{R}^*$ is free. Notice that A_j is not unital, so that it is not the unital algebra obtained from $C(T_j) \otimes \mathcal{K}$ by adjoining a unit.

Since the decomposition of $\tilde{f}^{-1}(0)$ as the union of the V_j over their common point 0 is U-equivariant, one gets the second statement. □

After collapsing the spectrum of $C^*(U)$ to a point, the topology of the spectrum of A_j is the topology of $T_j \cup \{0\}$ described above. The topology of the spectrum of A is the topology of $f^{-1}(0)$ of (10.7) described above.

10.2 A real and a non-Archimedean place

We now consider another example, namely the case of $\mathbb{K} = \mathbb{Q}$ with two places v_1, v_2, where $v_1 = p$ is a non-Archimedean place associated to a prime p and $v_2 = \infty$ is the real place. Again, we replace adèles by the product $\mathbb{K}_{v_1} \times \mathbb{K}_{v_2}$ over the two places, which in this case is just the product

$$\mathbb{K}_{v_1} \times \mathbb{K}_{v_2} = \mathbb{Q}_p \times \mathbb{R}. \tag{10.10}$$

The idèles are correspondingly replaced by $\mathbb{K}^*_{v_1} \times \mathbb{K}^*_{v_2} = \mathbb{Q}^*_p \times \mathbb{R}^*$ and the inclusion is given by

$$\mathbb{Q}^*_p \times \mathbb{R}^* \subset \mathbb{Q}_p \times \mathbb{R}. \tag{10.11}$$

The role of the action of the group $\mathbb{K}^*$ by multiplication is now replaced by the action by multiplication by the group U of elements of $\mathbb{K}^* = \mathbb{Q}^*$ that are units outside the above two places. This group is

$$U = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}, \tag{10.12}$$

where the $\mathbb{Z}/2\mathbb{Z}$ comes from ± 1 and the cyclic group is $p^{\mathbb{Z}}$ generated by $p \in \mathbb{K}^* = \mathbb{Q}^*$.

The action of U of (10.12) on $\mathbb{R} \times \mathbb{Q}_p$ is given by the transformation

$$S(x, y) = (px, py). \tag{10.13}$$

By comparison with the previous case of $\mathbb{K} = \mathbb{Q}(\sqrt{2})$, notice how in that case (cf. (10.2)) the pair (u, u^{-1}) was just the image of the element $3 - 2\sqrt{2}$ under the diagonal embedding of $\mathbb{K}$ in $\mathbb{K}_{v_1} \times \mathbb{K}_{v_2}$.

In the present case, the role of the adèles class space $X_{\mathbb{K}} = \mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$ is then played by its semilocal version

$$X_{p,\infty} = (\mathbb{Q}_p \times \mathbb{R})/U, \tag{10.14}$$

the quotient of $\mathbb{Q}_p \times \mathbb{R}$ by the symmetry $(x, y) \mapsto p(-x, -y)$ and the transformation S. Both of these transformations preserve the function

$$\tilde{f} : \mathbb{Q}_p \times \mathbb{R} \to \mathbb{R}_+, \qquad \tilde{f}(x, y) = |x|_p \, |y| \in \mathbb{R}_+, \tag{10.15}$$

which descends to a function

$$f : X_{p,\infty} \to \mathbb{R}_+. \tag{10.16}$$

Moreover, one has

$$(x, y) \in \mathbb{Q}_p^* \times \mathbb{R}^* \subset \mathbb{Q}_p \times \mathbb{R} \Leftrightarrow f(x, y) \neq 0,$$

and the fiber of f over any nonzero $\varepsilon \in \mathbb{R}_+$ is easily identified with $\mathbb{Z}_p^*$:

$$f^{-1}(\varepsilon) \sim \mathbb{Z}_p^*, \quad \forall \varepsilon \neq 0.$$

In fact, one can use the fundamental domain

$$\mathbb{Z}_p^* \times \mathbb{R}_+^*$$

for the action of U on $\mathbb{Q}_p^* \times \mathbb{R}^*$ to obtain the required isomorphism.

The fiber $f^{-1}(0)$ of f over the point $\varepsilon = 0$ is no longer $\mathbb{Z}_p^*$, and once again it is singular. It is again described as the union of three pieces,

$$f^{-1}(0) = T_p \cup T_\infty \cup \{0\}, \tag{10.17}$$

which have, respectively, the following description:

- T_p is the locus $x = 0$, $y \neq 0$, which is identified with a torus $T_p \sim \mathbb{R}_+^*/p^{\mathbb{Z}}$, using the map $(x, y) \mapsto |y|$.
- T_∞ is the locus $x \neq 0$, $y = 0$, which gives the compact space $T_\infty \sim \mathbb{Z}_p^*/\pm 1$ obtained as the quotient of $\mathbb{Q}_p^*$ by the action of U.
- The remaining piece is the point $x = 0$, $y = 0$.

The description of the topology of $f^{-1}(0)$ is similar to what happens in the case of $\mathbb{Q}(\sqrt{2})$ analyzed above.

What is not obvious in this case is how the totally disconnected fiber $f^{-1}(\varepsilon) \sim \mathbb{Z}_p^*$ can tie in with the torus $T_p \sim \mathbb{R}_+^*/p^{\mathbb{Z}}$ when $\varepsilon \to 0$.

To see what happens, we use the map

$$X_{p,\infty} \ni (x, y) \mapsto g(x, y) = \text{class of } |y| \in \mathbb{R}_+^*/p^{\mathbb{Z}}. \tag{10.18}$$

This is well defined on the open set $y \neq 0$. It is continuous and passes to the quotient. Thus, when a sequence $(x_n, y_n) \in X_{p,\infty}$ converges to a point $(0, y) \in T_p$, $y \neq 0$, one has $g(0, y) = \lim_n g(x_n, y_n)$.

The point then is simply that we have the relation

$$g(x, y) = f(x, y) \in \mathbb{R}_+^*/p^{\mathbb{Z}}. \tag{10.19}$$

In other words, $g(x_n, y_n) = \varepsilon_n$ with (x_n, y_n) in the fiber $f^{-1}(\varepsilon_n)$, and the point of the singular fiber T_p toward which $(x_n, y_n) \in X_{p,\infty}$ converges depends only on the value of ε_n in $\mathbb{R}_+^*/p^{\mathbb{Z}}$.

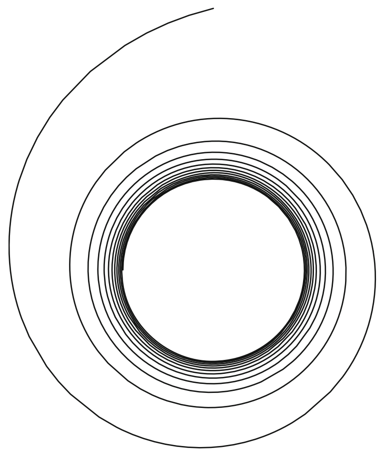


Fig. 5. The limit cycle of a foliation.

This phenomenon is reminiscent of the behavior of holonomy in the context of foliations, using a logarithmic scale $\mathbb{R}_+^*/p^{\mathbb{Z}} \sim \mathbb{R}/(\mathbb{Z}\log p)$. It corresponds to what happens in the limit cycle of the foliation associated to a flow as depicted in Figure 5.

As we argued in [11] (see also Sections 3.2 and 7.1 here above), the role of the Frobenius in characteristic zero is played by the one-parameter group $\mathrm{Fr}(t)$ with $t \in \mathbb{R}$, which corresponds to the action of $\mathbb{R}$ on the adèle class space $X_{\mathbb{Q}} = \mathbb{A}_{\mathbb{Q}}/\mathbb{Q}^*$ given in the above logarithmic scale, namely

$$\mathrm{Fr}(t)(a) = e^t\, a, \qquad \forall a \in X_{\mathbb{Q}}. \tag{10.20}$$

Its orbit over $p \in \Sigma_{\mathbb{Q}}$ is of length $\log p$, and it corresponds, in the simplified picture of $X_{p,\infty}$, to the component T_p of the singular fiber $f^{-1}(0)$.

10.3 Singularities of maps

The simple examples described above illustrate how one can use the function $f(x) = |x|$ in general, and see the place where it vanishes as the complement of $C_{\mathbb{K}}$ in the adèles class space $X_{\mathbb{K}}$. This provides a way of thinking of the inclusion of $C_{\mathbb{K}}$ in $X_{\mathbb{K}}$ in terms of the notions of "singular fiber" and "generic fiber" as seen in the examples above. The generic fiber appears to be typically identified with $C_{\mathbb{K},1}$, with the union of the generic fibers giving $C_{\mathbb{K}}$ as it should. This suggests the possibility of adapting to our noncommutative geometry context some aspects of the well-developed theory of nearby and vanishing cycles. A brief dictionary summarizing this analogy is given here below.

Total space	Adelè class space $X_{\mathbb{K}} = \mathbb{A}_{\mathbb{K}}/\mathbb{K}^*$
The map f	$f(x) = \|x\|$
Singular fiber	$X_{\mathbb{K}} \setminus C_{\mathbb{K}} = f^{-1}(0)$
Union of generic fibers	$C_{\mathbb{K}} = f^{-1}(\{0\}^c)$

References

1. Y. André, *Une introduction aux motifs*, Panoramas et Synthèses, Vol. 17, Société mathématique de France, 2005.
2. M.F. Atiyah, R. Bott, *A Lefschetz fixed point formula for elliptic complexes: I*, Annals of Math. Vol. **86** (1967), 374–407.
3. S. Bloch, H. Esnault, D. Kreimer, *On motives associated to graph polynomials*, Commun. Math. Phys. **267** (2006), 181–225.
4. E. Bombieri, *Problems of the Millennium: The Riemann Hypothesis*, Clay Mathematical Institute (2000).
5. J.B. Bost, A. Connes, *Hecke algebras, type III factors and phase transitions with spontaneous symmetry breaking in number theory*, Selecta Math. (N.S.) **1** (1995), no. 3, 411–457.
6. A. Connes, *Une classification des facteurs de type* III, Ann. Sci. École Norm. Sup. (4) **6** (1973), 133–252.
7. A. Connes, *Cohomologie cyclique et foncteurs* Ext^n, C. R. Acad. Sci. Paris Sér. I Math. **296** (1983), no. 23, 953–958.
8. A. Connes, *Noncommutative differential geometry.* Inst. Hautes Études Sci. Publ. Math. No. 62 (1985), 257–360.
9. A. Connes, *Noncommutative Geometry*, Academic Press, 1994.
10. A. Connes, *Trace formula in noncommutative geometry and the zeros of the Riemann zeta function.* Selecta Math. (N.S.) **5** (1999), no. 1, 29–106.
11. A. Connes, C. Consani, M. Marcolli, *Noncommutative geometry and motives: the thermodynamics of endomotives*, Advances in Mathematics, **214** (2007) N.2, 761–831.
12. A. Connes, M. Marcolli, *From physics to number theory via noncommutative geometry. Part I. Quantum statistical mechanics of* $\mathbb{Q}$*-lattices*, in "Frontiers in Number Theory, Physics, and Geometry, I", 269–350, 2006.
13. A. Connes, M. Marcolli, *Noncommutative geometry, quantum fields and motives*, Colloquium Publications, Vol. 55, American Mathematical Society, 2008.
14. A. Connes, M. Marcolli, N. Ramachandran, *KMS states and complex multiplication*, Selecta Math. (New Ser.) **11** (2005), no. 3-4, 325–347.
15. A. Connes, M. Marcolli, N. Ramachandran, *KMS states and complex multiplication, II*, in "Operator Algebras – The Abel Symposium 2004", 15–60, Springer, 2006.

16. A. Connes, G. Skandalis, *The longitudinal index theorem for foliations*, Publ. RIMS Kyoto Univ. **20** (1984) 1139–1183.
17. A. Connes, M. Takesaki, *The flow of weights on factors of type III.* Tôhoku Math. J. **29** (1977), 473–575.
18. C. Consani, *Double complexes and Euler L-factors*, Compositio Math. **111** (1998), no. 3, 323–358.
19. C. Consani, M. Marcolli, *Archimedean cohomology revisited*, in "Noncommutative Geometry and Number Theory" 109–140. Vieweg Verlag, 2006.
20. C. Consani, M. Marcolli, *Quantum statistical mechanics over function fields*, Journal of Number Theory **123** (2007), 487–528.
21. P. Deligne, *Théorie de Hodge III*, Publ. Math. IHES **44** (1974), 5–78.
22. A. Goncharov, Yu. Manin, *Multiple zeta motives and moduli spaces* $\bar{M}_{0,n}$, Compos. Math. **140** no. 1 (2004), 1–14.
23. V. Guillemin, S. Sternberg, *Geometric asymptotics*, Math. Surveys **14**, American Mathematical Society, 1977.
24. V. Guillemin, *Lectures on spectral theory of elliptic operators*, Duke Math. J., **44** 3 (1977), 485–517.
25. E. Ha, F. Paugam, *Bost-Connes-Marcolli systems for Shimura varieties. I. Definitions and formal analytic properties*, IMRP Int. Math. Res. Pap. 2005, no. 5, 237–286.
26. B. Jacob, *Bost-Connes type systems for function fields*, J. Noncommutative Geometry, Vol. 1 (2007), no. 2, 141–211.
27. M. Laca, *From endomorphisms to automorphisms and back: dilations and full corners*, J. London Math. Soc. (2) **61** (2000), no. 3, 893–904.
28. Yu.I. Manin, *Correspondences, motifs and monoidal transformations*, Mat. Sb. (N.S.) **77** (119) 1968, 475–507.
29. Yu.I. Manin, *Lectures on zeta functions and motives (according to Deninger and Kurokawa)*, Columbia University Number Theory Seminar (New York, 1992). Astérisque **228** (1995), 4, 121–163.
30. Yu.I. Manin, *Real Multiplication and noncommutative geometry (ein Alterstraum)*, in "The legacy of Niels Henrik Abel", 685–727, Springer, Berlin, 2004.
31. Yu.I. Manin, *Von Zahlen und Figuren, in "Géométric an XXe siècle. Histoire et horizons." (Ed. J. Kouneiher, D. Flament, Ph. Nabonnand, J.-J. Szczeciniarz.)* Hermann, Paris, 2005, 24–44.
32. M. Marcolli, *Arithmetic noncommutative geometry*, University Lectures Series **36**, American Mathematical Society, 2005.
33. R. Meyer, *On a representation of the idelè class group related to primes and zeros of L-functions*, Duke Math. J. **127** (2005), no. 3, 519–595.
34. B. Riemann, *Mathematical Works*, Dover, New York, 1953.
35. H.P.F. Swinnerton-Dyer, *Applications of Algebraic Geometry to Number Theory*, Proceedings Symposia in Pure Math. Vol. XX (1969), 21–26.
36. M. Takesaki, *Tomita's theory of modular Hilbert algebras and its applications.* Lecture Notes in Math. **28**, Springer, 1970.
37. M. Takesaki, *Duality for crossed products and the structure of von Neumann algebras of type III*, Acta Math. **131** (1973), 249–310.
38. A. Weil, *Letter to Artin*, Collected Papers, Vol. I (1980), 280–298.
39. A. Weil, *On the Riemann hypothesis in function-fields*, Proc. Nat. Acad. Sci. U.S.A. **27** (1941), 345–347.
40. A. Weil, *Sur la théorie du corps de classes*, J. Math. Soc. Japan, **3** (1951), 1–35.

41. A. Weil, *Fonction zeta et distributions*, Séminaire Bourbaki, Vol. 9 (1966), Exp. No. 312, 523–531.
42. A. Weil, *Sur les formules explicites de la théorie des nombres premiers*, Oeuvres complètes, Vol. 2, 48–62.
43. A. Weil, *Sur les formules explicites de la théorie des nombres*, Izv. Mat. Nauk., (Ser. Mat.) **36** (1972), 3–18.
44. A. Weil, *Basic Number Theory*, Reprint of the second (1973) edition. Classics in Mathematics. Springer, 1995.

Elliptic Curves with Large Analytic Order of Ш($\boldsymbol{E}$)

Andrzej Dąbrowski[1] and Mariusz Wodzicki[2,3]

[1] Instytut Matematyki, Uniwersytet Szczeciński, ul. Wielkopolska 15, 70-451 Szczecin, Poland, dabrowsk@wmf.univ.szczecin.pl

[2] Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA, Wodzicki@math.berkeley.edu

To Yuri Ivanovich Manin on His Seventieth Birthday

Summary. We present the results of our search for elliptic curves over $\mathbb{Q}$ with exceptionally large analytic orders of the Tate-Shafarevich group. We exibit 134 examples of rank zero curves with $|Ш(E)| > 1832^2$ which was the largest known value for any explicit curve. Our record is a curve with $|Ш(E)| = 63,408^2$.

We also present examples of curves of rank zero with the value of $L(E,1)$ much smaller, or much bigger, than in any previously known example. Finally, we present an example of a pair of non-isogeneous curves whose values of $L(E,1)$ coincide in the first 11 digits after the point!

Key words: elliptic curves, Birch–Swinnerton-Dyer Conjecture, Tate–Shafarevich group

2000 Mathematics Subject Classifications: 11G05, 11G40

Introduction

The L-series $L(E,s) = \sum_{n=1}^{\infty} a_n n^{-s}$ of an elliptic curve E over $\mathbb{Q}$ converges for $\operatorname{Re} s > 3/2$. The Modularity Conjecture, settled by Wiles–Taylor–Diamond–Breuil–Conrad [BCDT], implies that $L(E,s)$ analytically continues to an entire function and its leading term at $s = 1$ is described by the following long-standing conjecture.

Conjecture 1 (Birch and Swinnerton-Dyer). *The L-function $L(E,s)$ has a zero of order $r = \operatorname{rank} E(\mathbb{Q})$ at $s = 1$, and*

$$\lim_{s\to 1} \frac{L(E,s)}{(s-1)^r} = \frac{c_\infty(E) c_{\text{fin}}(E) R(E) |Ш(E)|}{|E(\mathbb{Q})_{\text{tors}}|^2}.$$

[3]Partially supported by NSF Grants DMS-9707965 and DMS-0503401

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*,
Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_9,

Here $E(\mathbb{Q})_{\mathrm{tors}}$ denotes the torsion subgroup of the group $E(\mathbb{Q})$ of rational points of E, the fudge factor c_{fin} is the Tamagawa number of E, and $R(E)$ is the regulator calculated with respect to the Néron–Tate height pairing. If ω is the *real* period of E, then $c_\infty = \omega$ or 2ω, according to whether the group *of real points* $E(\mathbb{R})$ is connected or not.

Finally, $Ш(E)$ denotes the Tate–Shafarevich group of E. The latter is formed by isomorphism classes of pairs (T, ϕ), where T is a smooth projective curve over $\mathbb{Q}$ of genus one that possesses a $\mathbb{Q}_p$-rational point for every prime p (including $p = \infty$), and $\phi : E \to \mathrm{Jac}(T)$ is an isomorphism defined over $\mathbb{Q}$. The Tate–Shafarevich group is very difficult to determine. It is known that the subgroup

$$Ш(E)[n] := \{a \in Ш(E) \mid na = 0\}$$

is finite for any $n > 1$, and it is conjectured that $Ш(E)$ is always finite. In theory, the standard 2-descent method calculates the dimension of the $\mathbb{F}_2$-vector space $Ш(E)[2]$ (see [Cr$_1$], [S]). It is not clear in general how to exhibit the curves of genus 1 that represent elements of $Ш(E)$ of order > 2 (see, however, [CFNS^2]).

It has been known for a long time that the order of $Ш(E)$, provided the latter is always finite, can take arbitrarily large values. Cassels [C] was the first to show this by proving that $|Ш(E)[3]|$ can be arbitrarily large for a special family of elliptic curves with j-invariant zero. Only in 1987 was it finally established that there are any elliptic curves over $\mathbb{Q}$ for which the Tate–Shafarevich group is finite (Rubin [Ru], Kolyvagin [K], Kato). Ten years later Rohrlich [Ro], by combining results of [HL] and [K], proved that given a modular elliptic curve E over $\mathbb{Q}$ (hence any curve—according to [BCDT]), and a positive integer n, there exists a quadratic twist E_d of E such that $Ш(E_d)$ *is* finite and $|Ш(E_d)[2]| \geq n$. This finally proved that $Ш(E)$ can indeed be a group of arbitrarily large finite order.

Assuming the Birch and Swinnerton-Dyer Conjecture, Mai and Murty [MM$_2$] showed that for the family of quadratic twists of any elliptic curve E, one has

$$\varliminf_{d} \frac{N(E_d)^{\frac{1}{4}-\epsilon}}{|Ш(E_d)|} = 0.$$

Goldfeld and Szpiro [GS], and Mai and Murty [M$_2$] (as reported by Rajan [R]), in the early 1990s proposed the following general conjecture:

Conjecture 2 (Goldfeld–Szpiro–Mai–Murty). *For any $\epsilon > 0$ we have*[4]

$$|Ш(E)| \ll N(E)^{1/2+\epsilon}. \tag{1}$$

[4] In this article we adhere to the following notational convention. Let $A(E)$ and $B(E)$ be some quantities dependent on a curve E belonging to a specified class $\mathcal{C}$ of elliptic curves defined over $\mathbb{Q}$. We say that $A(E) \ll B(E)$ if for any $K > 0$, there exists N_0 such that $A(E) < KB(E)$ for all curves in $\mathcal{C}$ with conductor $N(E) > N_0$. This is meaningful only if $\mathcal{C}$ contains infinitely many nonisomorphic curves. If either $A(E)$ or $B(E)$ depends on some parameter ϵ, then the choice of N_0 is allowed to depend on ϵ.

Estimate (1) holds for the family of rank zero quadratic twists of any particular elliptic curve, provided the Birch and Swinnerton-Dyer Conjecture holds for every member of that family.

The Birch and Swinnerton-Dyer Conjecture combined with the following consequence of the Generalized Lindelöf Hypothesis (see [GHP], p. 154),

$$\lim_{d\to\infty} \frac{L^{(r_d)}(E_d,1)}{N(E_d)^{\epsilon}} = 0 \qquad (d \text{ square-zero}),$$

where r_d denotes the rank of the group $E_d(\mathbb{Q})$, and the following conjecture of Lang (see [L]),

$$R(E) \gg N(E)^{-\epsilon},$$

easily imply that

$$|Ш(E_d)| \ll N(E_d)^{1/4+\epsilon}.$$

The following unconditional bounds,

$$|Ш(E)| \ll \begin{cases} N(E)^{79/120+\epsilon} & \text{if } \ j(E)=0, \\ N(E)^{37/60+\epsilon} & \text{if } \ j(E)=1728, \\ N(E)^{59/120+\epsilon} & \text{otherwise,} \end{cases}$$

where $j(E)$ denotes the j-invariant of E, are known for curves of rank zero with complex multiplication [GL].

In general, for elliptic curves satisfying the Birch and Swinnerton-Dyer Conjecture, Goldfeld and Szpiro [GS] have shown that the Goldfeld–Szpiro–Mai–Murty Conjecture is equivalent to the Szpiro Conjecture:

$$|\Delta(E)| \ll N(E)^{6+\epsilon}, \tag{2}$$

where $\Delta(E)$ denotes the discriminant of the minimal model of E. Masser proves in [Ma] that 6 in the exponent of (2) cannot be improved; in [We] de Weger conjectures that the exponent in (1) is also, in a certain sense, the best possible.

Conjecture 3 (de Weger). *For any $\epsilon > 0$ and any $C > 0$, there exists an elliptic curve over $\mathbb{Q}$ with*

$$|Ш(E)| > CN(E)^{1/2-\epsilon}.$$

He shows [We] that Conjecture 3 is a consequence of the following three conjectures: the Birch and Swinnerton-Dyer Conjecture for curves of rank zero, the Szpiro Conjecture, and the Riemann Hypothesis for Rankin–Selberg zeta functions associated to certain modular forms of weight $\frac{3}{2}$.

On the other hand, de Weger demonstrates that the following variant of Conjecture 3 which involves the minimal discriminant instead of the conductor, is a consequence of just the Birch and Swinnerton-Dyer Conjecture for elliptic curves with $L(E,1) \neq 0$.

Conjecture 4 (de Weger). *For any $\epsilon > 0$ and any $C > 0$, there exists an elliptic curve over $\mathbb{Q}$ with*

$$|\text{Ш}(E)| > C|\Delta(E)|^{1/12-\epsilon}.$$

For the purpose of the present article, the quantity

$$GS(E) := \frac{|\text{Ш}(E)|}{\sqrt{N(E)}} \tag{3}$$

will be referred to as the *Goldfeld–Szpiro ratio* of E. Eleven examples of elliptic curves with $GS(E) \geq 1$ are given in [We], the largest value being $6.893\dots$. Another forty-seven examples with $GS(E) \geq 1$ are produced by Nitaj [Ni], his largest value of $GS(E)$ being 42.265. Note that curves of small conductor with $GS(E) > 1$ were already known from Cremona's tables [Cr_2]. In all these examples, $GS(E)$ is calculated using the formula for $|\text{Ш}(E)|$ that is predicted by the Birch and Swinnerton-Dyer Conjecture; see (4) below.

Let us say a few words about the order of the Tate–Shafarevich group for those curves when it is known. The results by Stein and his collaborators [GJPST, Thm. 4.4] imply that $|\text{Ш}(E)| = 7^2$ for the curves denoted by `546f2` and `858k2`, respectively, in Cremona's tables [Cr_2]. No other curve of rank zero and conductor less than 1000 has larger $|\text{Ш}(E)|$ if the Birch and Swinnerton-Dyer Conjecture holds for such curves. Gonzalez-Avilés demonstrated [GA, Thm. B], that formula (4) for the order of the Tate–Shafarevich group holds for all the quadratic twists

$$E_d: \quad y^2 = x^3 + 21dx^2 + 112d^2x$$

with $L(E_d, 1) \neq 0$. The largest value of $|\text{Ш}(E_d)|$ for such curves, when $d \leq 2000$, is $|\text{Ш}(E_{1783})| = 8^2$ (cf. [Le, Table I]).

Assuming the validity of the Birch and Swinnerton-Dyer Conjecture, one can compute $|\text{Ш}(E)|$ for an elliptic curve of rank zero E by evaluating $L(E, 1)$ with sufficient accuracy. (In practice, this is possible only for curves with not too big conductors.) We shall be referring to this number as the *analytic order* of the Tate–Shafarevich group of E. In what follows, $|\text{Ш}(E)|$ will denote exclusively the analytic order of $\text{Ш}(E)$.

It is rather surprising how small the analytic order is in all known examples: de Weger [We] produced one with $|\text{Ш}(E)| = 224^2$, Rose [Rs] produced another one with $|\text{Ш}(E)| = 635^2$; and finally, Nitaj [Ni] found a curve with

$$|\text{Ш}(E)| = 1832^2,$$

and that seems to be the largest known value prior to the year 2002.

For the family of cubic twists considered by Zagier and Kramarz [ZK],

$$E'_d: \quad x^3 + y^3 = d \qquad (d \text{ cubic-free}),$$

the value of $|Ш(E'_d)|$ does not exceed 21^2 for $d \le 70000$. In this case, the Birch and Swinnerton-Dyer, the Lang, and the Generalized Lindelöf Conjectures imply that

$$|Ш(E'_d)| \ll N(E'_d)^{1/3+\epsilon}.$$

For quadratic twists of a given curve one can calculate the analytic order of the Tate–Shafarevich group using a well-known theorem of Waldspurger [W] in conjunction with purely combinatorial methods. The details for some curves with complex multiplication can be found in [Fr$_1$],[Fr$_2$],[Le],[N],[T]. Here we shall consider only one example, the family

$$E_d: \quad y^2 = x^3 - d^2x \qquad (d \ge 1 \text{ an odd square-free integer})$$

of so-called congruent-number elliptic curves. Define the sequence $a(d)$ by

$$\sum_{n=1}^{\infty} a(n)q^n := \eta(8z)\eta(16z)\Theta(2z),$$

where

$$\eta(z) = q^{1/24}\prod_{n=1}^{\infty}(1-q^n), \qquad \Theta(z) = \sum_{n=-\infty}^{\infty} q^{n^2} \qquad (q = e^{2\pi iz}).$$

When the curve E_d is of rank zero, then, assuming as usual the Birch and Swinnerton-Dyer conjecture, we have (see [T])

$$|Ш(E_d)| = \left(\frac{a(d)}{\tau(d)}\right)^2, \tag{4}$$

where $\tau(d)$ denotes the number of divisors of d. (Coefficients $a(d)$ can also be calculated using a formula of Ono [O].) Conjecturally, one expects that

$$|Ш(E_d)| \ll N(E_d)^{1/4+\epsilon};$$

hence the sequence of curves E_d (and, more generally, the family of quadratic twists of *any* curve) is not a likely candidate to produce curves with large Goldfeld–Szpiro ratio.

The primary aim of this article is to present the results of our search for curves with exceptionally large analytic orders of the Tate–Shafarevich group. We exhibit 134 examples of curves of rank zero with $|Ш(E)| > 1832^2$, which was the largest previously known value for any explicit curve. For our record curve we have

$$|Ш(E)| = 63,408^2.$$

For the reasons explained in the last section, we focused on the family

$$E(n,p): \quad y^2 = x(x+p)(x+p-4\cdot 3^{2n+1})$$

and three families of isogeneous curves, for n and p integers within the bounds $3 \leq n \leq 19$ and $0 < |p| < 1000$. Compared to the previously published results, in our work we faced dealing with curves of very large conductor. A large conductor translates into a very slow convergence rate of the approximation to $L(E,1)$. The main difficulty was to design a successful search strategy for curves with an exceptionally large Goldfeld–Szpiro ratio, (3), which is usually accompanied by a large value of the analytic order of the Tate–Shafarevich group.

Our explorations brought out also a number of unexpected discoveries: curves of rank zero with the value of $L(E,1)$ much smaller, or much bigger, than in any previously known example (see Tables 6 and 5 below). A particularly notable case involves a pair of nonisogeneous curves whose values of $L(E,1)$ coincide in their first 11 digits after the decimal!

Details of the computations, tables, and related comments are contained in Sections 1–3. Further remarks on Conjecture 3 are the subject of Section 4.

The actual calculations were carried out by the second author in summer and early fall 2002 on a variety of computers, almost all of them located in the Department of Mathematics at Berkeley. Supplemental computations were conducted also in 2003 and the Summer 2004.

The results were reported by M.W. at the conference *Geometric Methods in Algebra and Number Theory*, which took place in December 2003 in Miami, and by A.D. at the Number Theory Seminar at the Max-Planck-Institut in October 2006; A.D. would like to thank the Department of Mathematics at Berkeley and the Max-Planck-Institut in Bonn for their support and hospitality during his visits in 2006 when the revised version of this article was prepared; M.W. would like to thank the Institute of Mathematics at the University of Szczecin for its hospitality during his visits there in summer 2002, when the project started, and in summer 2003. The second author was partially supported by NSF grants DMS–9707965 and DMS–0503401.

1 Examples of elliptic curves with large $|\text{Ш}(E)|$

Consider the family

$$E(n,p): \quad y^2 = x(x+p)(x+p-4\cdot 3^{2n+1}),$$

with $(n,p) \in \mathbb{N} \times \mathbb{Z}$ and $p \neq 0,\ 4\cdot 3^{2n+1}$. Any member of this family is isogeneous over $\mathbb{Q}$ to three other curves $E_i(n,p)$ $(i = 2,3,4)$:

$$E_2(n,p): \quad y^2 = x^3 + 4(2\cdot 3^{2n+1} - p)x^2 + 16\cdot 3^{4n+2}x,$$
$$E_3(n,p): \quad y^2 = x^3 + 2(4\cdot 3^{2n+1} + p)x^2 + (4\cdot 3^{2n+1} - p)^2 x,$$

and

$$E_4(n,p): \quad y^2 = x^3 + 2(p - 8\cdot 3^{2n+1})x^2 + p^2 x.$$

The L-series and ranks of isogeneous curves coincide, while the orders of $E(\mathbb{Q})_{\text{tors}}$ and Ш(E), the real period, ω, and the Tamagawa number c_{fin} may differ. The curves being 2-isogeneous, the analytic orders of Ш(E_i) may differ from $|Ш(E(n,p)|$ only by a power of 2.

All the examples we found where *at least one* of the four analytic orders of Ш$(E(n,p)$ and Ш$(E_i(n,p))$ $(i = 2,3,4)$ is greater than or equal to 1000^2 are listed in Table 1. Notation used: $|Ш| = |Ш(E)|$ and $|Ш_i| = |Ш(E_i)|$.

For a curve E of rank zero, we compute the analytic order of Ш(E), i.e., the quantity

$$|Ш(E)| = \frac{L(E,1) \cdot |E(\mathbb{Q})_{\text{tors}}|^2}{c_\infty(E) c_{\text{fin}}(E)},$$

using the following approximation to $L(E,1)$, cf. [Co]:

$$S_m = 2 \sum_{l=1}^{m} \frac{a_l}{l} e^{-\frac{2\pi l}{\sqrt{N}}},$$

which, for

$$m \geq \frac{\sqrt{N}}{2\pi} \left(2 \log 2 + k \log 10 - \log(1 - e^{-2\pi/\sqrt{N}})\right),$$

differs from $L(E,1)$ by less than 10^{-k}.

It seems that the currently available techniques of n-descent for $n = 3$, 4, and 5 (cf. [CFNS^2],[MSS],[Be],[F]) can be utilized to see that 60^2 divides the actual order of Ш(E) for $E = E_3(15,12)$. On the other hand, the results of Kolyvagin and Kato could be used to prove that the actual order of Ш(E) divides $|Ш(E)|$. This would establish the validity of the exact form of the Birch and Swinnerton-Dyer Conjecture in this case. The Birch and Swinnerton-Dyer Conjecture is invariant under isogeny, hence this would establish the validity of this conjecture for each of its three isogeneous relatives. In particular, this would show that Ш$(E_4(15,12))$ is indeed a group of order 3840^2.

Table 1. Examples of elliptic curves $E(n,p)$ $(n \leq 19;\ 0 < |p| \leq 1000)$ with $\max(|Ш|, |Ш_2|, |Ш_3|, |Ш_4|) \geq 1000^2$.

(n,p)	$N(n,p)$	$\|Ш\|$	$\|Ш_2\|$	$\|Ш_3\|$	$\|Ш_4\|$
$(11,-489)$	1473152464197864	680^2	680^2	1360^2	680^2
$(11,163)$	1473152461647240	346^2	1384^2	173^2	1384^2
$(11,301)$	5440722586421136	576^2	1152^2	576^2	288^2
$(11,336)$	15816054028824	529^2	1058^2	529^2	1058^2
$(11,865)$	15635299103673360	617^2	1234^2	617^2	617^2
$(12,-605)$	4473683858657640	1031^2	1031^2	1031^2	2062^2
$(12,-257)$	20904304573762872	1545^2	1545^2	3090^2	3090^2
$(12,-56)$	569377945555104	1049^2	1049^2	2098^2	1049^2
$(12,22)$	143157883450560	416^2	1664^2	416^2	1664^2

(n,p)	$N(n,p)$	$\lvert Ш \rvert$	$\lvert Ш_2 \rvert$	$\lvert Ш_3 \rvert$	$\lvert Ш_4 \rvert$
$(12, 24)$	81339706505952	603^2	1206^2	603^2	1206^2
$(12, 63)$	63264216170568	554^2	1108^2	554^2	1108^2
$(12, 262)$	42622006206125760	468^2	1872^2	234^2	1872^2
$(12, 382)$	62143535763983040	648^2	2592^2	324^2	2592^2
$(12, 466)$	75808606453660608	1435^2	5740^2	1435^2	5740^2
$(12, 694)$	112899512607942336	576^2	2304^2	288^2	2304^2
$(12, 934)$	151942571712321216	512^2	2048^2	256^2	2048^2
$(13, -672)$	1281100377506040	389^2	1556^2	389^2	778^2
$(13, -160)$	915071698203240	1079^2	1079^2	2158^2	1079^2
$(13, -125)$	3660286792808760	639^2	1278^2	639^2	2556^2
$(13, -69)$	16837319246889384	516^2	516^2	258^2	1032^2
$(13, -42)$	20497606039673280	502^2	2008^2	251^2	2008^2
$(13, -17)$	12444975095505720	348^2	1392^2	348^2	2784^2
$(13, -5)$	3660286792794360	1583^2	1583^2	1583^2	3166^2
$(13, -3)$	1464114717117648	2364^2	2364^2	1182^2	2364^2
$(13, 60)$	457535849098320	552^2	1104^2	276^2	552^2
$(13, 66)$	32210523776515392	618^2	2472^2	309^2	2472^2
$(13, 73)$	610744996281840	494^2	1964^2	247^2	988^2
$(13, 96)$	10765549390536	588^2	1176^2	294^2	588^2
$(13, 136)$	264786704158368	258^2	1032^2	258^2	1032^2
$(13, 544)$	3111243773819208	929^2	1858^2	929^2	929^2
$(13, 708)$	21595692076981920	812^2	3248^2	406^2	1624^2
$(13, 876)$	835002924582096	340^2	1360^2	85^2	1360^2
$(13, 928)$	5307415849389480	470^2	1880^2	470^2	940^2
$(14, -948)$	2033174929441680	312^2	1248^2	156^2	624^2
$(14, -800)$	8235645283809960	390^2	1560^2	195^2	1560^2
$(14, -672)$	11529903397328568	2310^2	4620^2	2310^2	2310^2
$(14, -596)$	61355557364338608	598^2	1196^2	598^2	2392^2
$(14, -281)$	15300603799975032	253^2	1012^2	253^2	2024^2
$(14, -212)$	21824460002049648	560^2	560^2	560^2	1120^2
$(14, -33)$	72473678497325160	1002^2	2004^2	1002^2	4008^2
$(14, -12)$	3294258113514528	1077^2	2154^2	1077^2	2154^2
$(14, -11)$	144947356994638704	1806^2	3612^2	903^2	3612^2
$(14, -3)$	775119556121040	588^2	1176^2	294^2	1176^2
$(14, 12)$	205891132094640	564^2	2256^2	282^2	4512^2
$(14, 96)$	1647129056756616	306^2	1224^2	153^2	612^2
$(14, 100)$	16471290567565920	1186^2	2372^2	593^2	2372^2
$(14, 240)$	8235645283778760	1184^2	2368^2	592^2	1184^2

(n,p)	$N(n,p)$	$\lvert Ш\rvert$	$\lvert Ш_2\rvert$	$\lvert Ш_3\rvert$	$\lvert Ш_4\rvert$
$(14,268)$	726037150017264	858^2	1716^2	429^2	1716^2
$(14,528)$	18118419624294264	356^2	1424^2	356^2	712^2
$(14,652)$	33560254531348080	268^2	2144^2	67^2	2144^2
$(15,-852)$	8222777088032880	562^2	1124^2	281^2	1124^2
$(15,-248)$	141399694410862368	1185^2	4740^2	1185^2	4740^2
$(15,-240)$	74120807554080840	965^2	3860^2	965^2	3860^2
$(15,-212)$	280600200026160	498^2	1992^2	249^2	3984^2
$(15,-116)$	107475170953411824	2368^2	4736^2	2368^2	9472^2
$(15,-96)$	14824161510815304	1434^2	2838^2	717^2	2838^2
$(15,-84)$	3242785330490832	775^2	1650^2	775^2	1650^2
$(15,-80)$	74120807554076040	679^2	1358^2	679^2	1358^2
$(15,-48)$	14824161510815016	3057^2	3057^2	3057^2	3057^2
$(15,-12)$	5929664604325920	576^2	1152^2	288^2	1152^2
$(15,-6)$	237186584173036224	3705^2	3705^2	3705^2	7410^2
$(15,-1)$	59296646043258936	162^2	648^2	81^2	1296^2
$(15,1)$	118593292086517776	4032^2	8064^2	2016^2	8064^2
$(15,12)$	336912761609424	240^2	1920^2	60^2	3840^2
$(15,60)$	37060403777035920	2299^2	4598^2	2299^2	2299^2
$(15,88)$	130452621295164960	1232^2	2464^2	1232^2	2464^2
$(15,172)$	2489995878769488	1258^2	2516^2	629^2	1258^2
$(15,375)$	26953020928749960	1143^2	4572^2	1143^2	4572^2
$(16,-408)$	72579094756950240	1863^2	3726^2	3726^2	3726^2
$(16,-96)$	133417453597333128	3804^2	7608^2	1902^2	7608^2
$(16,-33)$	234814718331305640	3717^2	7437^2	3717^2	14868^2
$(16,-32)$	133417453597332744	5463^2	10926^2	5463^2	10926^2
$(16,-8)$	106733962877866080	891^2	891^2	891^2	1782^2
$(16,12)$	2084647712458320	792^2	3168^2	396^2	6336^2
$(16,48)$	7021971241964856	4608^2	9216^2	2304^2	9216^2
$(16,92)$	61372028654772720	1064^2	2128^2	532^2	2128^2
$(16,268)$	279342793469411664	2916^2	11664^2	1458^2	11664^2
$(16,300)$	166771816996663440	1018^2	4072^2	509^2	4072^2
$(16,472)$	186310763603371680	3119^2	12476^2	3119^2	12476^2
$(16,588)$	116740271897662896	549^2	2196^2	549^2	1098^2
$(16,592)$	17950711938549720	2221^2	8884^2	2221^2	4442^2
$(16,624)$	102025111574427912	1100^2	2200^2	550^2	1100^2
$(17,-404)$	118434048164038608	3246^2	6492^2	1623^2	12948^2
$(17,-68)$	10206435200195943696	8284^2	33136^2	4142^2	33136^2
$(19,-32)$	19452264734491086120	31704^2	63408^2	31704^2	63408^2

2 Values of the Goldfeld–Szpiro ratio $GS(E)$

The Goldfeld–Szpiro ratio was defined in (3). The articles of de Weger [We] and Nitaj [Ni] produce altogether 58 examples of elliptic curves with $GS(E)$ greater than 1 (the record value being 42.265...). For all of these examples the conductor does not exceed 10^{10}. The largest values of $GS(E)$ that we observed for our curves are tabulated in Table 2.

Table 2. Elliptic curves $E_i(n, p)$ ($9 \le n \le 19$; $0 < |p| \le 1000$; $1 \le i \le 4$) with the largest $GS(E)$. The notation $E_{i,j}(n, p)$ means that the given values of $|Ш(E)|$ and $GS(E)$ are shared by the isogeneous curves $E_i(n, p)$ and $E_j(n, p)$.

E	$\lvert Ш(E)\rvert$	$GS(E)$
$E_2(9, 544)$	344^2	1.20290...
$E_{2,4}(16, 48)$	9216^2	1.01357...
$E_2(10, 204)$	504^2	0.98366...
$E_4(15, -212)$	3984^2	0.94753...
$E_{2,4}(19, -32)$	63408^2	0.91159...
$E_4(16, 12)$	6336^2	0.87925...
$E_4(15, 12)$	3840^2	0.80334...
$E_2(16, 592)$	8882^2	0.58908...
$E_2(11, 160)$	322^2	0.57131...
$E_4(17, -404)$	12984^2	0.48986...
$E_4(16, -33)$	14868^2	0.45618...
$E_2(13, 96)$	1176^2	0.42149...
$E_{2,4}(16, 472)$	12476^2	0.36060...
$E_{2,4}(17, -68)$	33136^2	0.34368...
$E_{2,4}(16, -32)$	10926^2	0.32682...
$E_2(11, 336)$	1058^2	0.28146...
$E_4(15, -116)$	9472^2	0.27367...
$E_{2,4}(16, 268)$	11664^2	0.25741...

3 Large and small (nonzero) values of $L(E, 1)$

In this section we produce elliptic curves of rank zero with $L(E, 1)$ either much smaller or much larger than in all previously known examples (Tables 3 and 4).

Table 3. Elliptic curves $E(n,p)$ ($n \leq 19$; $0 < |p| \leq 1000$) with the largest values of $L(E,1)$ known to us.

E	$L(E,1)$
$E(11,-733)$	88.203561907255071...
$E(13,-160)$	71.523635814751843...
$E(12,466)$	56.224807584564927...
$E(7,-433)$	36.275918867296195...
$E(10,687)$	30.274774697662334...
$E(9,767)$	29.638568367562609...
$E(9,-93)$	28.032198538875886...
$E(11,336)$	22.922225180212583...

Table 4. Elliptic curves $E(n,p)$ ($n \leq 19$; $0 < |p| \leq 1000$) with the smallest positive values of $L(E,1)$ known to us.

E	$L(E,1)$
E(12,800)	0.0001706491750110...
E(10,142)	0.0002457348122099...
E(11,168)	0.0003276464160384...
E(14,672)	0.0006067526222261...
E(9,160)	0.0007372044423472...
E(10,-534)	0.0009829392448696...
E(10,408)	0.0009829392504019...

Note that

$$L(E(10,408),1) - L(E(10,-534),1) = 0.00000000000553237117\ldots .$$

This is the *smallest known* difference between the values of $L(E,1)$ of two elliptic curves of rank zero. The analytic orders of the Tate–Shafarevich group are 2^2, 4^2, 1^2, 4^2 for the isogeneous curves $E(10,408)$ and $E_i(10,408)$, respectively, and 2^2, 8^2, 8^2, 8^2 for the curves $E(10,-534)$ and $E_i(10,-534)$, repectively, where $i = 2$, 3, or 4.

We observed that for a large percentage of rank-zero curves $E_i(n,p)$ with $7 \leq n \leq 19$ and $0 < |p| \leq 1000$, one has

$$L(E,1) \geq \frac{1}{(\log N(E))^2}. \tag{5}$$

We verified, in particular, that (5) holds for every single curve $E(7,p)$ of rank zero when $0 < |p| \leq 1000$. This is consistent with results of Iwaniec and Sarnak, who proved [IS] that

$$L(f,1) \geq \frac{1}{(\log N)^2},$$

for a large percentage of newforms of weight 2, with the *level* N of a newform f playing the role of the conductor of an elliptic curve.

On the other hand, we have

$$L(E(8,-131),1) = 0.0002764516... < 0.0012048710... = (\log N(8,-131))^{-2},$$
$$L(E(9,160),1) = 0.0007372044... < 0.0015186182... = (\log N(9,160))^{-2},$$
$$L(E(10,142),1) = 0.0002457384... < 0.0009026601... = (\log N(10,142))^{-2},$$
$$L(E(11,168),1) = 0.0003276464... < 0.0009902333... = (\log N(11,168))^{-2},$$
$$L(E(12,800),1) = 0.0001706491... < 0.0009613138... = (\log N(12,800))^{-2}.$$

An estimate much weaker than (5) was proposed by Hindry [H], see Conjecture 6 below.

4 Remarks on Conjecture 3

Below we sketch how to utilize the curves $E(n,p)$ in order to establish the first of the two conjectures of de Weger (Conjecture 3 above).

According to Chen [Ch], every sufficiently large even integer can be represented as the sum $p+q$ where p is an odd prime and q is the product of at most two primes. Apply this, for sufficiently large n, to the number

$$4 \cdot 3^{2n+1} = p + q. \tag{6}$$

The factors $c_\infty(E(n,p))$ and $c_{\text{fin}}(E(n,p))$ on the right-hand side of the formula for the analytic order of the Tate–Shafarevich group, (4), are given by the following lemma.

Lemma. *Assume $p < q$, with q having at most two prime factors. Then we have*

(i) $$c_\infty(E(n,p)) = \frac{\pi}{3^{n+1/2} \cdot AGM(1, \sqrt{q/(p+q)})},$$

(ii) $$c_{\text{fin}}(E(n,p)) = 2c_2c_3c_q,$$

where $AGM(a,b)$ denotes the arithmetic-geometric mean of a and b,

$$c_2 = \begin{cases} 2 \textit{ if } p \equiv 1 \mod 4 \\ 4 \textit{ if } p \equiv 3 \mod 4 \end{cases}, \qquad c_3 = \begin{cases} 2(2n+1) \textit{ if } p \equiv 2 \mod 3 \\ 4 \quad \textit{ if } p \equiv 1 \mod 3 \end{cases},$$

and

$$c_q = \begin{cases} 2 & \textit{if } q \textit{ is a prime} \\ 4 & \textit{if } q \textit{ is a product of two primes} \end{cases}.$$

The conductor is given by the formula

(iii) $N(E(n,p)) = 2^{f_2} \cdot 3 \cdot p \cdot rad(q)$,

where $rad(q)$ *denotes the product of prime factors, and*

$$f_2 = \begin{cases} 3 & \textit{if } p \equiv 1 (\textit{mod}\, 4) \\ 4 & \textit{if } p \equiv 3 (\textit{mod}\, 4) \,. \end{cases}$$

This is easily proven using calculations of Nitaj [Ni, Propositions 2.1, 3.1, and 3.2]. The following then seems to be a plausible conjecture.

Conjecture 5. *For any* $\epsilon > 0$ *there exists* $c(\epsilon) > 0$ *and infinitely many* n *admitting a decomposition* (6) *with*

$$p \le c(\epsilon) q^{\epsilon}$$

such that the curve $E(n,p)$ *has rank zero.*

If we accept Conjecture 5, then

$$\frac{1}{c_\infty(E(n,p))} \gg N(E(n,p))^{1/2-\epsilon} \quad \text{and} \quad \frac{1}{c_{\text{fin}}(E(n,p))} \gg N(E(n,p))^{-\epsilon} \tag{7}$$

on an infinite set of curves $E(n,p)$.

Since $|E(\mathbb{Q})_{\text{tors}}| \ge 1$ (in fact, $|E(\mathbb{Q})_{\text{tors}}|$ can take only twelve values between 1 and 16, cf. [Mz]) it remains to estimate $L(E,1)$. The result of Iwaniec and Sarnak mentioned in Section 3 provides support for the following conjecture recently proposed by Hindry [H, Conjecture 5.4].

Conjecture 6 (Hindry). *One has*

$$L^{(r)}(E,1) \gg N(E)^{-\epsilon} \qquad (r \textit{ being the rank of } E). \tag{8}$$

Hindry observed that (8) implies that the distance from 1 to the nearest zero of $L(E,s)$ is $\gg N(E)^{-\epsilon}$.

The combination of inequalities (7) and (8), for curves of rank zero, yields the assertion of Conjecture 3 for the analytic order of the Tate–Shafarevich group. In order to pass to the actual order, one needs, of course, the equality of the two, as predicted by the Birch and Swinnerton-Dyer Conjecture.

References

[BCDT] C. Breuil, B. Conrad, F. Diamond and R. Taylor, *On the modularity of elliptic curves over* $\mathbb{Q}$: *wild 3-adic exercises,* J. Amer. Math. Soc. **14** (2001), 843–939.

[Be] C. D. Beaver, *5-torsion in the Shafarevich-Tate group of a family of elliptic curves,* J. Number Theory **82** (2000), 25–46.

[C] J. W. S. Cassels, *Arithmetic on curves of genus 1. VI. The Tate-Šafarevič group can be arbitrarily large,* J. reine angew. Math. **214/215** (1964), 65–70.

[Ch] J. R. Chen, *On the representation of a large even number as the sum of a prime and the product of at most two primes,* Sc. Sinica **16** (1973), 157–176.

[Co] H. Cohen, *A course in Computational Algebraic Number Theory.* Graduate Texts in Math. **138** Springer-Verlag (1993).

[Cr_1] J. E. Cremona, *Algorithms for modular elliptic curves.* Cambridge University Press 1997.

[Cr_2] J. E. Cremona, Elliptic Curve Data (*an online resource*).

[$CFNS^2$] J. E. Cremona, T. A. Fischer, C. O'Neil, D. Simon, M. Stoll, *Explicit n-descent on elliptic curves. I. Algebra, II. Geometry,* preprints 2006.

[F] T. Fisher, *Some examples of 5 and 7 descent for elliptic curves over* $\mathbb{Q}$, J. Eur. Math. Soc. **3** (2001), 169–201.

[Fr_1] G. Frey, *Der Rang der Lösungen von* $y^2 = x^3 \pm p^3$ *über* $\mathbb{Q}$, Manuscr. math. **48** (1984), 71–101.

[Fr_2] G. Frey, *A relation between the value of the L-series of the curve* $y^2 = x^3 - k^3$ *in* $s = 1$ *and its Selmer group,* Arch. Math. **45** (1985), 232–238.

[GHP] D. Goldfeld, J. Hoffstein, S. J. Patterson, *On automorphic functions of half-integral weight with applications to elliptic curves.* In: Number Theory Related to Fermat's Last Theorem (Birkhäuser, Boston, MA, 1982), 153–193.

[GL] D. Goldfeld and D. Lieman, *Effective bounds on the size of the Tate-Shafarevich group,* Math. Res. Letters **3** (1996), 309–318.

[GS] D. Goldfeld and L. Szpiro, *Bounds for the order of the Tate-Shafarevich group,* Compos. math. **97** (1995), 71–87.

[GA] C. Gonzalez-Avilés, *On the conjecture of Birch and Swinnerton-Dyer*, Trans. Amer. Math. Soc. **349** (1997), 4181–4200.

[GJPST] G. Grigorov, A. Jorza, S. Patrikis, W.A. Stein, C. Tarniţă, *Verification of the Birch and Swinnerton-Dyer conjecture for specific elliptic curves*, Math. Comput. (to appear).

[H] M. Hindry, *Why it is difficult to compute the Mordell-Weil group?* In: Diophantine Geometry. Proceedings (ed. Umberto Zannier, Edizioni della Normale, Pisa, 2007).

[HL] J. Hoffstein, W. Luo, *Nonvanishing of L-series and the combinatorial sieve,* Math. Res. Letters **4** (1997), 435–442.

[IS] H. Iwaniec and P. Sarnak, *The non-vanishing of central values of automorphic L-functions and Landau-Siegel zeros,* Israel J. Math. **120** (2000), 155–177.

[K] V. A. Kolyvagin, *Finiteness of* $E(\mathbb{Q})$ *and* $Ш(E/\mathbb{Q})$ *for a subclass of Weil curves*, Izv. Akad. Nauk SSSR, **52** (1988), 522-540; *translation in* Math. USSR-Izv. **32** (1989), 523–541.

[L] S. Lang, *Conjectured diophantine estimates on elliptic curves.* In: Arithmetic and Geometry–Papers dedicated to I. R. Shafarevich. Vol. 1 (Birkhäuser, Boston, MA, 1983), 155–171.

[Le] J. L. Lehman, *Rational points on elliptic curves with complex multiplication by the ring of integers in* $\mathbb{Q}(\sqrt{-7})$, J. Number Theory **27** (1987), 253–272.

[M$_2$] L. Mai and M. R. Murty, *On the Shafarevich-Tate group.* (unpublished, reported in [R]).

[MM$_2$] L. Mai and M. R. Murty, *A note on quadratic twists of an elliptic curve.* In: Elliptic Curves and Related Topics. CRM Proceed. and Lecture Notes **4** (1994), 121–124.

[Ma] D. W. Masser, *Note on a conjecture of Szpiro,* Astérisque **183** (1990), 19–23.

[Mz] B. Mazur, *Modular curves and the Eisenstein ideal,* Publ. Math. IHES **47** (1977), 33–186.

[MSS] J. R. Merriman, S. Siksek, N. P. Smart, *Explicit* 4*-descents on an elliptic curve,* Acta arithmetica **77** (1996), 385–404.

[N] J. Nekovář, *Class numbers of quadratic fields and Shimura's correspondence,* Math. Ann. **287** (1990), 577–594.

[Ni] A. Nitaj, *Invariants des courbes de Frey-Hellegouarch et grands groupes de Tate-Shafarevich,* Acta arithmetica **93** (2000), 303–327.

[O] K. Ono, *Tate-Shafarevich groups of the congruent number elliptic curves,* Acta arithmetica **81** (1997), 247–252.

[R] C. S. Rajan, *On the size of the Shafarevich-Tate group of elliptic curves over function fields,* Compos. math. **105** (1997), 29–41.

[Ro] D. E. Rohrlich, *Unboundedness of the Tate-Shafarevich group in families of quadratic twists* (an appendix to [HL]), Math. Res. Letters **4** (1997), 443–444.

[Rs] H. E. Rose, *On some elliptic curves with large Sha,* Experimental Math. **9** (2000), 85–89.

[Ru] K. Rubin, *Tate-Shafarevich groups and L-functions of elliptic curves with complex multiplication.* Invent. math **89** (1987), 527–559.

[S] J. H. Silverman, *The arithmetic of elliptic curves.* Graduate Texts in Math. **106**, Springer, New York 1986.

[T] J. Tunnell, *A classical diophantine problem and modular forms of weight* 3/2*,* Invent. math. **72** (1983), 323–334.

[W] J. L. Waldspurger, *Sur les coefficients de Fourier des formes modulaires de poids demi-entiers,* J. Math. Pure et Appl. **60** (1981), 375–484.

[We] B. M. M. de Weger, $A + B = C$ *and big* Ш*'s,* Quart. J. Math. **49** (1998), 105–128.

[ZK] D. Zagier and G. Kramarz, *Numerical investigations related to the L-series of certain elliptic curves,* Journal of the Indian Math. Soc. **52** (1987), 51–69.

$\boldsymbol{p}$-adic Entropy and a $\boldsymbol{p}$-adic Fuglede–Kadison Determinant

Christopher Deninger

Mathematisches Institut, Einsteinstr. 62, 48149 Münster, Germany
c.deninger@math.uni-muenster.de

Dedicated to Yuri Ivanovich Manin

Summary. Using periodic points, we study a notion of entropy with values in the p-adic numbers. This is done for actions of countable discrete residually finite groups Γ. For suitable $\Gamma = \mathbb{Z}^d$-actions we obtain p-adic analogues of multivariable Mahler measures. For certain actions of more general groups, the p-adic entropy can be expressed in terms of a p-adic analogue of the Fuglede–Kadison determinant from the theory of von Neumann algebras. Many basic questions remain open.

Key words: Mahler measure, p-adic numbers, Kadison–Fuglede determinants.

2000 Mathematics Subject Classifications: 11D88, 16S34, 19B28, 37A35, 37C35, 37C85

1 Introduction

In several instances, the entropy $h(\varphi)$ of an automorphism φ on a space X can be calculated in terms of periodic points:

$$h(\varphi) = \lim_{n\to\infty} \frac{1}{n} \log |\mathrm{Fix}\,(\varphi^n)|. \tag{1}$$

Here $\mathrm{Fix}\,(\varphi^n)$ is the set of fixed points of φ^n on X. Let $\log_p : \mathbb{Q}_p^* \to \mathbb{Z}_p$ be the branch of the p-adic logarithm normalized by $\log_p(p) = 0$. The p-adic analogue of the limit (1), if it exists, may be viewed as a kind of entropy with values in the p-adic number field $\mathbb{Q}_p$,

$$h_p(\varphi) = \lim_{n\to\infty} \frac{1}{n} \log_p |\mathrm{Fix}\,(\varphi^n)| \;. \tag{2}$$

It depends only on the action of φ on X viewed as a set.

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*,
Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_10,

An earlier, different, approach to a p-adic entropy theory was mentioned to me by Amnon Besser. The usual definitions of measure-theoretic or topological entropy have no obvious p-adic analogue, since $\overline{\lim}$ and sup do not make sense p-adically and since the cardinalities of partitions, coverings, and separating or spanning sets do not behave reasonably in the p-adic metric.

Instead of actions of a single automorphism φ we look more generally at actions of a countable discrete residually finite but not necessarily amenable group Γ on a set X. Let us write $\Gamma_n \to e$ if (Γ_n) is a sequence of cofinite normal subgroups of Γ such that only the neutral element e of Γ is contained in infinitely many Γ_n's. Let $\mathrm{Fix}_{\Gamma_n}(X)$ be the set of points in X that are fixed by Γ_n. If the limit

$$h_p := \lim_{n\to\infty} \frac{1}{(\Gamma : \Gamma_n)} \log_p |\mathrm{Fix}_{\Gamma_n}(X)| \tag{3}$$

exists with respect to a choice of $\Gamma_n \to e$, we call it the p-adic entropy of the Γ-action on the set X (with respect to the sequence (Γ_n)).

In this note we show that for an interesting class of Γ-actions the p-adic entropy exists independently of the choice of $\Gamma_n \to e$. In these examples X is an abelian group and Γ acts by automorphisms of groups. Namely, let $(\mathbb{R}/\mathbb{Z})^\Gamma$ be the full shift on Γ with values in the circle $\mathbb{R}/\mathbb{Z}$ and left Γ-action by $\gamma(x_{\gamma'}) = (x_{\gamma^{-1}\gamma'})$. For an element $f = \sum_\gamma a_\gamma \gamma$ in the integral group ring $\mathbb{Z}\Gamma$ consider the closed subshift $X_f \subset (\mathbb{R}/\mathbb{Z})^\Gamma$ consisting of all sequences $(x_{\gamma'})$ that satisfy the equation

$$\sum_{\gamma'} x_{\gamma'} a_{\gamma^{-1}\gamma'} = 0 \quad \text{in } (\mathbb{R}/\mathbb{Z})^\Gamma \text{ for all } \gamma \in \Gamma \,.$$

In fact, as in [ER01], we study more general systems defined by an $r\times r$-matrix over $\mathbb{Z}\Gamma$. However, in this introduction, for simplicity, we describe only the case $r = 1$. If Γ is amenable, we denote by $h(f)$ the topological entropy of the Γ-action on X_f.

The case $\Gamma = \mathbb{Z}^d$ is classical. Here we may view f as a Laurent polynomial, and according to [LSW90], the entropy is given by the (logarithmic) Mahler measure of f:

$$h(f) = m(f) := \int_{T^d} \log |f(z)| \, d\mu(z) \,. \tag{4}$$

Here μ is the normalized Haar measure on the d-torus T^d. According to [LSW90], the $\mathbb{Z}^d$-action on X_f is expansive if and only if f does not vanish at any point of T^d. By a theorem of Wiener this is also equivalent to f being a unit in $L^1(\mathbb{Z}^n)$. In this case $h(f)$ can be calculated in terms of periodic points, cf. [LSW90, Theorem 7.1]. See also [Sch95] for this theory.

What about a p-adic analogue? In [Den97] it was observed that in the expansive case, $m(f)$ has an interpretation via the Deligne–Beilinson regulator map from algebraic K-theory to Deligne cohomology. Looking at the analogous regulator map from algebraic K-theory to syntomic cohomology one gets a suggestion of what a (purely) p-adic Mahler measure $m_p(f)$ of f should be; cf. [BD99]. It can be defined only if f does not vanish at any point of the p-adic d-torus $T_p^d = \{z \in \mathbb{C}_p^d \,|\, |z_i|_p = 1\}$, where $\mathbb{C}_p$ is the completion of a fixed algebraic closure $\overline{\mathbb{Q}}_p$ of $\mathbb{Q}_p$. In this case $m_p(f)$ is given by the convergent Snirelman integral

$$m_p(f) = \int_{T_p^d} \log_p f(z)\,. \tag{5}$$

Recall that the Snirelman integral of a continuous function $F : T_p^d \to \mathbb{C}_p$ is defined by the following limit if it exists:

$$\int_{T_p^d} F(z) := \lim_{\substack{N\to\infty \\ (N,p)=1}} \frac{1}{N^d} \sum_{\zeta \in \mu_N^d} F(\zeta)\,.$$

Here μ_N is the group of N-th roots of unity in $\overline{\mathbb{Q}}_p^*$.

For example, let $P(t) = a_m t^m + \cdots + a_r t^r$ be a polynomial in $\mathbb{C}_p[t]$ with $a_m, a_r \neq 0$ whose zeros α satisfy $|\alpha|_p \neq 1$. Then, according to [BD99, Proposition 1.5] we have the following expression for the p-adic Mahler measure:

$$\begin{aligned} m_p(f) &= \log_p a_r - \sum_{0<|\alpha|_p<1} \log_p \alpha \\ &= \log_p a_m + \sum_{|\alpha|_p>1} \log_p \alpha\,. \end{aligned} \tag{6}$$

For $d \geq 2$ there does not seem to be a simple formula for $m_p(f)$.

In [BD99] we mentioned the obvious problem to give an interpretation of $m_p(f)$ as a p-adically valued entropy. This is now provided by the following result:

Theorem 1. *Assume that $f \in \mathbb{Z}[\mathbb{Z}^d] = \mathbb{Z}[t_1^{\pm 1}, \ldots, t_d^{\pm 1}]$ does not vanish at any point of the p-adic d-torus T_p^d. Then the p-adic entropy $h_p(f)$ of the $\Gamma = \mathbb{Z}^d$-action on X_f in the sense of (3) exists for all $\Gamma_n \to 0$ and we have $h_p(f) = m_p(f)$.*

Now we turn to more general groups Γ. In [DS] extending [Den06] it was shown that for countable residually finite amenable groups Γ and elements f in $\mathbb{Z}\Gamma$ that are invertible in $L^1(\Gamma)$ we have

$$h(f) = \log \det\nolimits_{\mathcal{N}\Gamma} f\,. \tag{7}$$

Here $\det_{\mathcal{N}\Gamma}$ is the Fuglede–Kadison determinant [FK52] on the units of the von Neumann algebra $\mathcal{N}\Gamma \supset L^1\Gamma \supset \mathbb{Z}\Gamma$ of Γ. In fact, equation (7) holds without the condition of amenability if $h(f)$ is replaced by the quantity

$$h_{\mathrm{per}}(f) := \lim_{n\to\infty} \frac{1}{(\Gamma : \Gamma_n)} \log |\mathrm{Fix}_{\Gamma_n}(X)|\,.$$

For the Γ-action on X_f this limit exists and is independent of the choice of sequence $\Gamma_n \to e$.

In the p-adic case, instead of working with a p-adic L^1-convolution algebra it is more natural to work with the bigger convolution algebra $c_0(\Gamma)$. It consists of all formal series $x = \sum_\gamma x_\gamma \gamma$ with $x_\gamma \in \mathbb{Q}_p$ and $|x_\gamma|_p \to 0$ as $\gamma \to \infty$ in Γ.

For $\Gamma = \mathbb{Z}^d$ it is known that $f \in \mathbb{Z}[\mathbb{Z}^d]$ does not vanish in any point of the p-adic d-torus T_p^d if and only if f is a unit in the algebra $c_0(\mathbb{Z}^d)$. Hence in general, it is natural to look for a p-adic analogue of formula (7) for all $f \in \mathbb{Z}\Gamma$ that are units in $c_0(\Gamma)$. In the p-adic case there is no analogue for the theory of von Neumann algebras and for the functional calculus used to define $\det_{\mathcal{N}\Gamma}$. However, using some algebraic K-theory and the results of [FL03], [BLR], and [KLM88] we can define a p-adic analogue $\log_p \det_\Gamma$ of $\log \det_{\mathcal{N}\Gamma}$ for suitable classes of groups Γ. For example, we get the following result generalizing Theorem 1:

Theorem 2. *Assume that the residually finite group Γ is elementary amenable and torsion-free. Let f be an element of $\mathbb{Z}\Gamma$ that is a unit in $c_0(\Gamma)$. Then the p-adic entropy $h_p(f)$ of the Γ-action on X_f in the sense of (3) exists for all $\Gamma_n \to e$ and we have*

$$h_p(f) = \log_p \det_\Gamma f .$$

Acknowledgement: I would like to thank my colleagues Wolfgang Lück and Peter Schneider for helpful conversations.

2 Preliminaries

Fix an integer $r \geq 1$ and set $T = T^r = (\mathbb{R}/\mathbb{Z})^r$. For a discrete group Γ let T^Γ be the full shift with left Γ-action by $\gamma(x_{\gamma'}) = (x_{\gamma^{-1}\gamma'})$. Write $M_r(R)$ for the ring of $r \times r$ matrices over a ring R. For an element $f = \sum a_\gamma \gamma$ in $M_r(\mathbb{Z})[\Gamma] = M_r(\mathbb{Z}\Gamma)$ the closed subshift $X_f \subset T^\Gamma$ is defined as the closed subgroup consisting of all sequences with

$$\sum_{\gamma'} x_{\gamma'} a^*_{\gamma^{-1}\gamma'} = 0 \quad \text{in } T^\Gamma \text{ for all } \gamma \in \Gamma \, .$$

Here a^* denotes the transpose of a matrix a in $M_r(\mathbb{Z})$. The group ring $M_r(\mathbb{Z})[\Gamma]$ is equipped with an anti-involution $*$ defined by $f^* = \sum_\gamma a^*_{\gamma^{-1}} \gamma$ for $f = \sum_\gamma a_\gamma \gamma$.

Let ρ_f be right multiplication by f^* on the group $T[[\Gamma]]$ of formal T-valued series on Γ. For $x = \sum_\gamma x_\gamma \gamma$ in $T[[\Gamma]]$ we have

$$\rho_f(x) = \sum_\gamma x_\gamma \gamma \sum_\gamma a^*_{\gamma^{-1}} \gamma = \sum_\gamma \Big(\sum_{\gamma'} x_{\gamma'} a^*_{\gamma^{-1}\gamma'} \Big) \gamma \, .$$

Hence we see that

$$X_f = \operatorname{Ker}(\rho_f : T[[\Gamma]] \longrightarrow T[[\Gamma]]),$$

where on the right-hand side the group Γ acts by left multiplication. Let N be a normal subgroup of Γ with quotient map $\sim: \Gamma \to \tilde{\Gamma} = \Gamma/N$. Set

$$\tilde{f} = \sum_{\gamma} a_\gamma \tilde{\gamma} = \sum_{\delta \in \tilde{\Gamma}} \Big(\sum_{\gamma \in \delta} a_\gamma \Big) \delta \quad \text{in } M_r(\mathbb{Z})[\tilde{\Gamma}] \, .$$

This is the image of f under the reduction map $M_r(\mathbb{Z})[\Gamma] \to M_r(\mathbb{Z})[\tilde{\Gamma}]$. Under the natural isomorphism

$$T[[\tilde{\Gamma}]] \xrightarrow{\sim} \operatorname{Fix}_N(T[[\Gamma]])$$

mapping $\sum_\delta x_\delta \delta$ to $\sum_\gamma x_{\tilde{\gamma}} \gamma$, the action $\rho_{\tilde{f}}$ corresponds to the restriction of ρ_f. Hence we have

$$\operatorname{Fix}_N(X_f) = \operatorname{Ker}(\rho_{\tilde{f}} : T[[\tilde{\Gamma}]] \longrightarrow T[[\tilde{\Gamma}]]) = X_{\tilde{f}} \, .$$

If we assume that $\tilde{\Gamma}$ is finite we get that

$$\operatorname{Fix}_N(X_f) = \rho_{\tilde{f},\mathbb{R}}^{-1}(\mathbb{Z}\tilde{\Gamma})^r / (\mathbb{Z}\tilde{\Gamma})^r$$

for the endomorphism $\rho_{\tilde{f},\mathbb{R}}$ of right multiplication by $\tilde{f}^*$ on $(\mathbb{R}\tilde{\Gamma})^r$. This implies the following fact, cf. [DS, Corollary 4.3]:

Proposition 3. *Let $\tilde{\Gamma}$ be finite. Then $\rho_{\tilde{f}}$ is an isomorphism of $(\mathbb{Q}\tilde{\Gamma})^r$ if and only if $\operatorname{Fix}_N(X_f)$ is finite. In this case the order is given by*

$$|\operatorname{Fix}_N(X_f)| = \pm \det \rho_{\tilde{f}} \, .$$

This follows from the fact that for an isomorphism φ of a finite dimensional real vector space V and a lattice Λ in V with $\varphi(\Lambda) \subset \Lambda$, we have

$$|\varphi^{-1}\Lambda/\Lambda| = |\Lambda/\varphi(\Lambda)| = |\det(\varphi \,|\, V)|.$$

For any countable discrete group Γ let $c_0(\Gamma)$ be the set of formal series $\sum_\gamma x_\gamma \gamma$ with $x_\gamma \in \mathbb{Q}_p$ and $|x_\gamma|_p \to 0$ for $\gamma \to \infty$. This means that for any $\varepsilon > 0$ there is a finite subset $S \subset \Gamma$ such that $|x_\gamma|_p < \varepsilon$ for all $\gamma \in \Gamma \setminus S$. The set $c_0(\Gamma)$ is a $\mathbb{Q}_p$-vector space, and it becomes a $\mathbb{Q}_p$-algebra with the product

$$\sum_\gamma x_\gamma \gamma \cdot \sum_\gamma y_\gamma \gamma = \sum_\gamma \Big(\sum_{\gamma'\gamma''=\gamma} x_{\gamma'} y_{\gamma''} \Big) \gamma \, . \tag{8}$$

Note that the sums

$$\sum_{\gamma'\gamma''=\gamma} x_{\gamma'} y_{\gamma''} = \sum_{\gamma'} x_{\gamma'} y_{\gamma'^{-1}\gamma}$$

converge p-adically for every γ, since $\lim_{\gamma'\to\infty} |x_{\gamma'} y_{\gamma'^{-1}\gamma}|_p = 0$. The value is independent of the order of summation. Moreover, because of the inequality

$$\Big|\sum_{\gamma'} x_{\gamma'} y_{\gamma'^{-1}\gamma}\Big|_p \le \sup_{\gamma'} |x_{\gamma'} y_{\gamma'^{-1}\gamma}|_p \,, \tag{9}$$

we have

$$\lim_{\gamma\to\infty} \sum_{\gamma'} x_{\gamma'} y_{\gamma'^{-1}\gamma} = 0 \,,$$

so that the product (8) is well defined. We may also view $c_0(\Gamma)$ as an algebra of $\mathbb{Q}_p$-valued functions on Γ under convolution.

The $\mathbb{Q}_p$-algebra $c_0(\Gamma)$ is complete in the norm

$$\Big\|\sum_{\gamma} x_\gamma \gamma\Big\| = \sup_\gamma |x_\gamma|_p = \max_\gamma |x_\gamma|_p \,.$$

The norm satisfies the following properties:

$$\|x\| = 0 \quad \text{if and only if } x = 0; \tag{10}$$

$$\|x+y\| \le \max(\|x\|, \|y\|); \tag{11}$$

$$\|\lambda x\| = |\lambda|_p \, \|x\| \quad \text{for all } \lambda \in \mathbb{Q}_p; \tag{12}$$

$$\|xy\| \le \|x\| \, \|y\| \quad \text{and } \|1\| = 1. \tag{13}$$

Hence $c_0(\Gamma)$ is a p-adic Banach algebra over $\mathbb{Q}_p$, i.e., a unital $\mathbb{Q}_p$-algebra B that is complete with respect to a norm $\| \; \| : B \to \mathbb{R}^{\ge 0}$ satisfying conditions (10)–(13).

We will consider only Banach algebras for which $\| \; \|$ takes values in $p^{\mathbb{Z}} \cup \{0\}$. The subring $A = B^0$ of elements x in B of norm $\|x\| \le 1$ is a p-adic Banach algebra over $\mathbb{Z}_p$, defined similarly as before. An example is given by

$$c_0(\Gamma, \mathbb{Z}_p) = c_0(\Gamma)^0 = \Big\{ \sum x_\gamma \gamma \,|\, x_\gamma \in \mathbb{Z}_p \text{ with } \lim_{\gamma\to\infty} |x_\gamma|_p = 0 \Big\} .$$

In this case, the residue algebra A/pA over $\mathbb{F}_p$ is isomorphic to the group ring of Γ over $\mathbb{F}_p$:

$$c_0(\Gamma, \mathbb{Z}_p)/p c_0(\Gamma, \mathbb{Z}_p) = \mathbb{F}_p[\Gamma] \,. \tag{14}$$

The 1-units $U^1 = 1 + pA$ form a subgroup of A^*, since

$$(1+pa)^{-1} := \sum_{\nu=0}^{\infty} (-pa)^\nu$$

provides an inverse of $1 + pa \in U^1$ in U^1. It is easy to see that one has an exact sequence of groups

$$1 \longrightarrow U^1 \longrightarrow A^* \longrightarrow (A/pA)^* \longrightarrow 1 \,. \tag{15}$$

For $A = c_0(\Gamma, \mathbb{Z}_p)$ this is the exact seqence

$$1 \longrightarrow 1 + pc_0(\Gamma, \mathbb{Z}_p) \longrightarrow c_0(\Gamma, \mathbb{Z}_p)^* \longrightarrow \mathbb{F}_p[\Gamma]^* \longrightarrow 1 \,. \tag{16}$$

Concerning the units of a p-adic Banach algebra over $\mathbb{Q}_p$, we have the following known fact:

Proposition 4. *Let B be a p-adic Banach algebra over $\mathbb{Q}_p$ whose norm takes values in $p^{\mathbb{Z}} \cup \{0\}$ and set $A = B^0$. If the residue algebra A/pA has no zero divisors, then we have*

$$B^* = p^{\mathbb{Z}} A^* \quad \textit{and} \quad p^{\mathbb{Z}} \cap A^* = 1 \,.$$

Proof. For f in B^* set $g = 1/f$. Let ν, μ be such that $f_1 = p^\nu f$ and $g_1 = p^\mu g$ have norm one. The reductions $\overline{f}_1, \overline{g}_1$ of f_1, g_1 are nonzero. In the equation $f_1 g_1 = p^{\nu+\mu}$ we have $\nu + \mu \ge 0$. Reducing $\bmod\, p$ we find that $0 \neq \overline{f}_1 \overline{g}_1 = p^{\nu+\mu} \bmod p$ in A/pA. Hence we have $\nu+\mu = 0$ and therefore $f_1 g_1 = 1$. The first assertion follows. Because of (13) we have $\|a\| = 1$ for $a \in A^*$ and $\|p^\nu\| = p^{-\nu}$. This implies the second assertion. □

Example 5. For the group $\Gamma = \mathbb{Z}^d$ the algebra $c_0(\mathbb{Z}^d)$ can be identified with the affinoid commutative algebra $\mathbb{Q}_p\langle t_1^{\pm 1}, \dots, t_d^{\pm 1}\rangle$ of power series $\sum_{\nu \in \mathbb{Z}^d} x_\nu t^\nu$ with $x_\nu \in \mathbb{Q}_p$ and $\lim_{|\nu| \to \infty} |x_\nu|_p = 0$. Note that these power series can be viewed as functions on T_p^d. The residue algebra is $\mathbb{F}_p[\mathbb{Z}^d] = \mathbb{F}_p[t_1^{\pm 1}, \dots, t_d^{\pm 1}]$. It has no zero divisors, and its groups of units is

$$\mathbb{F}_p[\mathbb{Z}^d]^* = \mathbb{F}_p^* t_1^{\mathbb{Z}} \cdots t_d^{\mathbb{Z}} \,.$$

The preceeding proposition and the exact sequence (16) now give a decomposition into a direct product of groups

$$c_0(\mathbb{Z}^d)^* = p^{\mathbb{Z}} \mu_{p-1} t_1^{\mathbb{Z}} \cdots t_d^{\mathbb{Z}} (1 + p\, c_0(\mathbb{Z}^d, \mathbb{Z}_p)) \,.$$

Proposition 6. *For f in $\mathbb{Q}_p[\mathbb{Z}^d] = \mathbb{Q}_p[t_1^{\pm 1}, \dots, t_d^{\pm 1}]$ the following properties are equivalent:*

(a) We have $f(z) \neq 0$ for every z in T_p^d;

(b) f is a unit in $c_0(\mathbb{Z}^d)^*$;

(c) f has the form $f(t) = ct^\nu(1 + pg(t))$ for some $c \in \mathbb{Q}_p^*$, $\nu \in \mathbb{Z}^d$ and $g(t)$ in $c_0(\mathbb{Z}^d, \mathbb{Z}_p)$.

Proof. We have seen that (b) and (c) are equivalent, and it is clear that both (b) and (c) imply (a). For proving that (a) implies (b) note that the maximal ideals of $c_0(\mathbb{Z}^d) = \mathbb{Q}_p\langle t_1^{\pm 1}, \dots, t_d^{\pm 1}\rangle$ correspond to the orbits of the $\mathrm{Gal}\,(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$-operation on $T_p^d \cap (\overline{\mathbb{Q}}_p^*)^d$. Hence f is not contained in any maximal ideal of $c_0(\mathbb{Z}^d)$ by assumption (a) and therefore f is a unit. □

3 The Frobenius group determinant and a proof of Theorem 1

The map $\mathbb{Z}\Gamma \to \mathbb{Z}\tilde{\Gamma}$ for $\tilde{\Gamma} = \Gamma/N$ from the beginning of the last section can be extended to a homomorphism of $\mathbb{Q}_p$-algebras $c_0(\Gamma) \to c_0(\tilde{\Gamma})$ by sending $f = \sum a_\gamma \gamma$ to $\tilde{f} = \sum a_\gamma \tilde{\gamma}$. Note that this is well defined by the ultrametric inequality and that we have $\|\tilde{f}\| \le \|f\|$. If $\tilde{\Gamma}$ is finite, we have $c_0(\tilde{\Gamma}) = \mathbb{Q}_p\tilde{\Gamma}$ and hence we obtain a homomorphism of groups $\mathrm{GL}_r(c_0(\Gamma)) \to \mathrm{GL}_r(\mathbb{Q}_p\tilde{\Gamma})$. It follows that for f in $M_r(\mathbb{Z}\Gamma) \cap \mathrm{GL}_r(c_0(\Gamma))$ the endomorphism $\rho_{\tilde{f}}$ of $(\mathbb{Q}\tilde{\Gamma})^r$ is an isomorphism. Together with Proposition 3 we have proved the first equation in the following proposition:

Proposition 7. *Let Γ be a discrete group and N a normal subgroup with finite quotient group $\tilde{\Gamma}$. For f in $M_r(\mathbb{Z}\Gamma) \cap \mathrm{GL}_r(c_0(\Gamma))$ the set $\mathrm{Fix}_N(X_f)$ is finite and we have*

$$\begin{aligned} |\mathrm{Fix}_N(X_f)| &= \pm \det \rho_{\tilde{f}} \\ &= \pm \prod_\pi \det\nolimits_{\overline{\mathbb{Q}}_p} \Big(\sum_\gamma a_\gamma^* \otimes \rho_\pi(\tilde{\gamma}) \Big)^{d_\pi} . \end{aligned}$$

Here π runs over the equivalence classes of irreducible representations ρ_π of $\tilde{\Gamma}$ on $\overline{\mathbb{Q}}_p$-vector spaces V_π, and d_π is the degree $\dim V_\pi$ *of π.*

Proof. It remains to prove the second equation, which is essentially due to Frobenius. Consider $\overline{\mathbb{Q}}_p\tilde{\Gamma}$ as a representation of $\tilde{\Gamma}$ via the map $\tilde{\gamma} \mapsto \rho(\tilde{\gamma}) :=$ right multiplication by $\tilde{\gamma}^{-1}$. This ("right regular") representation decomposes as follows into irreducible representations; cf. [Sch95, I, 2.2.4]

$$\overline{\mathbb{Q}}_p\tilde{\Gamma} \cong \bigoplus_\pi V_\pi^{d_\pi} .$$

The endomorphism

$$\rho_{\tilde{f}} = \sum_\gamma a_\gamma^* \otimes \rho(\tilde{\gamma})$$

on

$$(\overline{\mathbb{Q}}_p\tilde{\Gamma})^r = \overline{\mathbb{Q}}_p^r \otimes \overline{\mathbb{Q}}_p\tilde{\Gamma}$$

therefore corresponds to the endomorphism

$$\bigoplus_\pi \Big(\sum_\gamma a_\gamma^* \otimes \rho_\pi(\tilde{\gamma}) \Big)^{d_\pi} \quad \text{on} \quad \bigoplus_\pi \overline{\mathbb{Q}}_p^r \otimes V_\pi^{d_\pi} = \bigoplus_\pi (\overline{\mathbb{Q}}_p^r \otimes V_\pi)^{d_\pi} .$$

Hence the formula follows. □

Remark 8. In the real case and for the Heisenberg group, Klaus Schmidt previously used the group determinant to calculate $|\mathrm{Fix}_{\Gamma_n}(X_f)|$ for f in $L^1(\Gamma)^*$ and certain Γ_n.

The following result generalizes Theorem 1 from the introduction, at least for a particular sequence $\Gamma_n \to 0$:

Theorem 9. *Let $f = \sum_{\nu \in \mathbb{Z}^n} a_\nu t^\nu$ in $M_r(\mathbb{Z}[t_1^{\pm 1}, \dots, t_d^{\pm 1}])$ be invertible at every point of the p-adic d-torus T_p^d. Then the p-adic entropy $h_p(f)$ of the $\Gamma = \mathbb{Z}^d$-action on X_f exists in the sense of (3) for the sequence $\Gamma_n = (n\mathbb{Z})^d \to 0$ with n prime to p, and we have*

$$h_p(f) = m_p(\det f) .$$

Proof. By assumption, the Laurent polynomial $\det f$ does not vanish at any point of T_p^d. Hence $\det f$ is a unit in $c_0(\Gamma)$ by Proposition 6. It follows from Proposition 7 that we have

$$|\mathrm{Fix}_{\Gamma_n}(X_f)| = \pm \prod_{\chi} \det_{\overline{\mathbb{Q}}_p}\Big(\sum_{\nu \in \mathbb{Z}^d} a_\nu^* \otimes \chi(\nu) \Big),$$

where χ runs over the characters of $\Gamma/\Gamma_n = (\mathbb{Z}/n\mathbb{Z})^d$. These correspond via $\chi(\nu) = \zeta^\nu$ to the elements ζ of μ_n^d. Viewing f as a matrix of functions on T_p^d, we therefore get the formulas

$$\begin{aligned} |\mathrm{Fix}_{\Gamma_n}(X_f)| &= \pm \prod_{\zeta \in \mu_n^d} \det_{\overline{\mathbb{Q}}_p}\Big(\sum_{\nu \in \mathbb{Z}^d} a_\nu^* \zeta^\nu \Big) \\ &= \pm \prod_{\zeta \in \mu_n^d} \det_{\overline{\mathbb{Q}}_p}(f(\zeta)) \\ &= \pm \prod_{\zeta \in \mu_n^d} (\det f)(\zeta) . \end{aligned}$$

Thus the p-adic entropy $h_p(f)$ of the Γ-action on X_f with respect to the above sequence is given by

$$\begin{aligned} h_p(f) &= \lim_{\substack{n \to \infty \\ (n,p)=1}} \frac{1}{(\Gamma : \Gamma_n)} \log_p |\mathrm{Fix}_{\Gamma_n}(X_f)| \\ &= \lim_{\substack{n \to \infty \\ (n,p)=1}} \frac{1}{n^d} \sum_{\zeta \in \mu_n^d} \log_p (\det f)(\zeta) \\ &= \int_{T_p^d} \log_p \det f = m_p(\det f) . \end{aligned}$$

Note here that for the Laurent polynomials $\det f$ under consideration the Snirelman integral exists by [BD99, Proposition 1.3]. □

Remark 10. A suitable generalization of that proposition would give Theorems 9 and 1 for general sequences $\Gamma_n \to 0$ in $\Gamma = \mathbb{Z}^d$. We leave this to the interested reader, since the general case of Theorem 1 is also a corollary of Theorem 2, which will be proved by a different method in Section 5.

Example 11. The polynomial in one variable $f(T) = 2T^2 - T + 2$ does not vanish at any point of the 2-adic circle T^1_2. In this sense, X_f is "2-adically expansive." Consider the square root of -15 in $\mathbb{Z}_2$ given by the 2-adically convergent series

$$\sqrt{-15} = (1 + (-16))^{1/2} = \sum_{\nu=0}^{\infty} \binom{1/2}{\nu} (-1)^{\nu} 2^{4\nu} \; .$$

The zeros of $f(T)$ in $\overline{\mathbb{Q}}_2$ are given by $\alpha_{\pm} = \frac{1}{4}(1 \pm \sqrt{-15}) \in \mathbb{Q}_2$, where $|\alpha_+|_2 = 2$ and $|\alpha_-|_2 = 1/2$. Successive approximations for α_+ coming from the series for $\sqrt{-15}$ are $1/2, -3/2, -19/2, -83/2$. By Theorem 9 and formula (6) the 2-adic entropy of X_f is given by

$$h_2(f) = \log_2 \alpha_+ \in \mathbb{Z}_2 \; .$$

Note that f viewed as a complex-valued function has both its zeros on S^1, so that X_f is not expansive in the usual sense. The topological entropy is $h(f) = \log 2$.

4 The logarithm on the 1-units of a *p*-adic Banach algebra

For a discrete group Γ we would like to define a homomorphism

$$\log_p \det{}_\Gamma : c_0(\Gamma)^* \longrightarrow \mathbb{Q}_p$$

that should be a p-adic replacement for the map

$$\log \det{}_{\mathcal{N}\Gamma} : L^1(\Gamma)^* \subset (\mathcal{N}\Gamma)^* \longrightarrow \mathbb{R} \; .$$

More generally, we would like to define such a map on $\mathrm{GL}_r(c_0(\Gamma))$. In this section we give the definition on the subgroup of 1-units and relate $\log_p \det_\Gamma$ to p-adic entropy. The extension to a map on all of $c_0(\Gamma)^*$ will be done in the next section for suitable classes of groups Γ using rather deep facts about group rings.

Let B be a p-adic Banach algebra over $\mathbb{Q}_p$ whose norm $\| \; \|$ takes values in $p^{\mathbb{Z}} \cup \{0\}$. A trace functional on B is a continuous linear map $\mathrm{tr}_B : B \to \mathbb{Q}_p$ that vanishes on commutators $[a, b] = ab - ba$ of elements in B. For $b \in B$ and $c \in B^*$ we have

$$\mathrm{tr}_B(cbc^{-1}) = \mathrm{tr}_B(b) \; . \tag{1}$$

Set $A = B^0 = \{b \in B \mid \|b\| \le 1\}$ and let U^1 be the normal subgroup of 1-units in A^*. The logarithmic series

$$\log : U^1 \longrightarrow A \; , \; \log u = -\sum_{\nu=1}^{\infty} \frac{(1-u)^{\nu}}{\nu},$$

converges and defines a continuous map. An argument with formal power series shows that we have

$$\log uv = \log u + \log v \tag{2}$$

if the elements u and v in U^1 *commute with each other.*

The next result is a consequence of the Campbell–Baker–Hausdorff formula.

Theorem 13. *The map*

$$\mathrm{tr}_B \log : U^1 \longrightarrow \mathbb{Z}_p$$

is a homomorphism. For u in U^1 and a in A^* we have

$$\mathrm{tr}_B \log(aua^{-1}) = \mathrm{tr}_B \log(u)\ . \tag{3}$$

Proof. Formula (3) follows from (1). From [Bou72, Ch. II, § 8] we get the following information about log. Set $G = \{b \in B \mid \|b\| < p^{-\frac{1}{p-1}}\}$. Then the exponential series defines a bijection $\exp : G \xrightarrow{\sim} 1 + G$ with inverse $\exp^{-1} = \log|_{1+G}$. For x, y in G we have

$$\exp x \cdot \exp y = \exp h(x, y), \tag{4}$$

where $h(x, y) \in G$ is given by a convergent series in B. It has the form

$$h(x, y) = x + y + \text{series of (iterated) commutators}\ .$$

Elements u, v of $1+G$ have the form $u = \exp x$ and $v = \exp y$. Taking the log of relation (4) and applying tr_B, we get

$$\begin{aligned} \mathrm{tr}_B \log(uv) &= \mathrm{tr}_B h(x, y) = \mathrm{tr}_B(x + y) \\ &= \mathrm{tr}_B \log u + \mathrm{tr}_B \log v\ . \end{aligned} \tag{5}$$

Hence $\mathrm{tr}_B \log$ is a homomorphism on the subgroup $1+G$ of U^1. By assumption the norm of B takes values in $p^{\mathbb{Z}} \cup \{0\}$. For $p \neq 2$ we therefore have $1+G = U^1$ and we are done.

For $p = 2$ the restriction of the map

$$\varphi = \mathrm{tr}_B \log : U^1 \to \mathbb{Q}_2$$

to $1 + G = 1 + 4A$ is a homomorphism by (5). We have to show that it is a homomorphism on $U^1 = 1 + 2A$ as well. For u in U^1 we have $\varphi(u) = \frac{1}{2}\varphi(u^2)$ by (2), and u^2 lies in $1 + 4A$. Now consider elements u, v in U^1. Then we have

$$\begin{aligned} \varphi(uv) &= \frac{1}{2}\varphi((uv)^2) = \frac{1}{2}\varphi(uvuv) \overset{(3)}{=} \frac{1}{2}\varphi(u^2vuvu^{-1}) \\ &= \frac{1}{2}\varphi(u^2) + \frac{1}{2}\varphi(vuvu^{-1}) \end{aligned}$$

since u^2 and $vuvu^{-1}$ lie in $1+4A$ where φ is a homomorphism. By similar arguments we get

$$\begin{aligned}\varphi(uv) &= \varphi(u) + \frac{1}{2}\varphi(v^2uvu^{-1}v^{-1}) \\ &= \varphi(u) + \varphi(v) + \frac{1}{2}\varphi(uvu^{-1}v^{-1})\,.\end{aligned}$$

Thus we must show that $\varphi(uvu^{-1}v^{-1}) = 0$. By (3) we have $\varphi(uvu^{-1}v^{-1}) = \varphi(vu^{-1}v^{-1}u)$, and hence using that both $uvu^{-1}v^{-1}$ and $vu^{-1}v^{-1}u$ lie in $1+4A$ we obtain

$$\begin{aligned}2\varphi(uvu^{-1}v^{-1}) &= \varphi(uvu^{-1}v^{-1}) + \varphi(vu^{-1}v^{-1}u) \\ &= \varphi(uvu^{-2}v^{-1}u) \overset{(3)}{=} \varphi(u^{-2}v^{-1}u^2v) \\ &= \varphi(u^{-2}) + \varphi(v^{-1}u^2v) \\ &\overset{(3)}{=} \varphi(u^{-2}) + \varphi(u^2) = \varphi(e) \\ &= 0\,.\end{aligned}$$

□

For a discrete group Γ, the map

$$\mathrm{tr}_\Gamma : c_0(\Gamma) \longrightarrow \mathbb{Q}_p\,,\ \mathrm{tr}_\Gamma(\sum a_\gamma\gamma) = a_e$$

defines a trace functional on $c_0(\Gamma)$. Let $B = M_r(c_0(\Gamma))$ be the p-adic Banach algebra over $\mathbb{Q}_p$ of $r \times r$ matrices (a_{ij}) with entries in $c_0(\Gamma)$ and equipped with the norm $\|(a_{ij})\| = \max_{ij} \|a_{ij}\|$. The composition

$$\mathrm{tr}_\Gamma : M_r(c_0(\Gamma)) \xrightarrow{\mathrm{tr}} c_0(\Gamma) \xrightarrow{\mathrm{tr}_\Gamma} \mathbb{Q}_p$$

defines a trace functional on $M_r(c_0(\Gamma))$.

The algebra $A = B^0$ is given by $M_r(c_0(\Gamma, \mathbb{Z}_p))$ and we have $U^1 = 1 + pM_r(c_0(\Gamma, \mathbb{Z}_p))$. The exact sequence (15) becomes the exact sequence of groups

$$1 \longrightarrow 1 + pM_r(c_0(\Gamma, \mathbb{Z}_p)) \longrightarrow \mathrm{GL}_r(c_0(\Gamma, \mathbb{Z}_p)) \longrightarrow \mathrm{GL}_r(\mathbb{F}_p\Gamma) \longrightarrow 1\,. \tag{6}$$

According to Theorem 13, the map

$$\log_p \det{}_\Gamma := \mathrm{tr}_\Gamma \log : 1 + pM_r(c_0(\Gamma, \mathbb{Z}_p)) \longrightarrow \mathbb{Z}_p \tag{7}$$

is a homomorphism of groups.

Example 14. For $\Gamma = \mathbb{Z}^d$, in the notation of Example 5 we have a commutative diagram, cf. [BD99, Lemma 1.1]

$$\begin{tikzcd} c_0(\Gamma) \arrow[r, "\sim"] \arrow[d, "\mathrm{tr}_\Gamma"'] & \mathbb{Q}_p\langle t_1^{\pm 1}, \ldots, t_d^{\pm 1}\rangle \arrow[d, "\int_{T_p^d}"] \\ \mathbb{Q}_p \arrow[r, equal] & \mathbb{Q}_p\,. \end{tikzcd}$$

It follows that for a 1-unit f in $M_r(c_0(\Gamma))$ we have

$$\log_p \det{}_\Gamma f = \int_{T_p^d} \log\det f = m_p(\det f)\;. \tag{8}$$

Here we have used the relation

$$\operatorname{tr}\log f = \log\det f \quad \text{in } c_0(\Gamma)\;, \tag{9}$$

where $\det : \mathrm{GL}_r(c_0(\Gamma)) \to c_0(\Gamma)^*$ is the determinant and tr the trace for matrices over the *commutative* ring $c_0(\Gamma)$. Note that det maps 1-units to 1-units. Relation (9) can be proved by embedding the integral domain $c_0(\Gamma) = \mathbb{Q}_p\langle t_1^{\pm 1}, \dots, t_d^{\pm 1}\rangle$ into its quotient field and applying [Har77, Appendix C, Lemma 4.1].

For finite groups Γ the map $\log_p \det_\Gamma$ can be calculated as follows. For f in $M_r(c_0(\Gamma)) = M_r(\mathbb{Q}_p\Gamma)$ let ρ_f be the endomorphism of $(\mathbb{Q}_p\Gamma)^r$ by right multiplication by f^* and $\det_{\mathbb{Q}_p}(\rho_f)$ its determinant over $\mathbb{Q}_p$.

Proposition 15. *Let Γ be finite. Then we have*

$$\log_p \det{}_\Gamma f = \frac{1}{|\Gamma|} \log_p \det{}_{\mathbb{Q}_p}(\rho_f) \tag{10}$$

for f in $1 + pM_r(\mathbb{Z}_p\Gamma)$.

Remark 16. Since $\rho_{fg} = \rho_f\rho_g$, the equation in the proposition shows that $\log_p \det_\Gamma$ is a homomorphism, something we know in general by Theorem 13. For finite Γ the group $\mathrm{GL}_r(\mathbb{F}_p\Gamma)$ is finite. Hence, by (6) there is at most one way to extend $\log_p \det_\Gamma$ from $1 + pM_r(\mathbb{Z}_p\Gamma)$ to a homomorphism from $\mathrm{GL}_r(\mathbb{Z}_p\Gamma)$ to $\mathbb{Q}_p$. Namely, we have to set

$$\log_p \det{}_\Gamma f := \frac{1}{N} \log_p \det{}_\Gamma f^N,$$

where $N \geq 1$ is any integer with $\overline{f}^N = 1$ in $\mathrm{GL}_r(\mathbb{F}_p\Gamma)$. Because of (2) this is well defined, but it is not clear from the definition that we get a homomorphism. However, for the same f, N we have

$$\log_p \det{}_{\mathbb{Q}_p}(\rho_f) = \frac{1}{N} \log_p \det{}_{\mathbb{Q}_p}(\rho_{f^N}).$$

Hence equation (10) holds for all f in $\mathrm{GL}_r(\mathbb{Z}_p\Gamma)$, and it follows that $\log_p \det_\Gamma$ extends to a *homomorphism* on $\mathrm{GL}_r(\mathbb{Z}_p\Gamma)$. In the next section such arguments will be generalized to infinite groups with the help of K-theory.

Proof of Proposition 15. Under the continuous homomorphism

$$\rho : M_r(\mathbb{Z}_p\Gamma) \longrightarrow \mathrm{End}_{\mathbb{Z}_p}(\mathbb{Z}_p\Gamma)^r$$

of p-adic Banach algebras over $\mathbb{Z}_p$, the groups of 1-units are mapped to each other. Hence we have

$$\log \rho_f = \rho_{\log f} \tag{11}$$

for f in $1 + pM_r(\mathbb{Z}_p\Gamma)$.

On the other hand, we have

$$\mathrm{tr}_\Gamma(g) = \frac{1}{|\Gamma|}\mathrm{tr}(\rho_g) \tag{12}$$

for any element g of $M_r(\mathbb{Q}_p\Gamma)$. This is proved first for $r = 1$ by checking the cases in which $g = \gamma$ is an element of Γ. Then one extends to arbitrary r by thinking of ρ_g as a block matrix with blocks of size $|\Gamma| \times |\Gamma|$.

Combining (11) and (12) we obtain

$$\begin{aligned} \log_p \det_\Gamma f &= \mathrm{tr}_\Gamma \log f \\ &= \frac{1}{|\Gamma|}\mathrm{tr}(\rho_{\log f}) \\ &= \frac{1}{|\Gamma|}\mathrm{tr}(\log \rho_f) \\ &= \frac{1}{|\Gamma|} \log_p \det_{\mathbb{Q}_p}(\rho_f) \ . \end{aligned}$$

The last equation is proved by writing ρ_f in triangular form in a suitable basis over $\overline{\mathbb{Q}}_p$ and observing that the eigenvalues of ρ_f are 1-units in $\overline{\mathbb{Q}}_p$. □

The next result is necessary to prove the relation of $\log_p \det_\Gamma f$ with p-adic entropies.

Proposition 17. *Let Γ be a residually finite countable discrete group and $\Gamma_n \to e$ a sequence as in the introduction. For f in $1+pM_r(c_0(\Gamma, \mathbb{Z}_p))$ consider its image $f^{(n)}$ in $1+pM_r(\mathbb{Z}_p\Gamma^{(n)})$, where $\Gamma^{(n)}$ is the finite group $\Gamma^{(n)} = \Gamma/\Gamma_n$. Then we have*

$$\log_p \det_\Gamma f = \lim_{n\to\infty} \log_p \det_{\Gamma^{(n)}} f^{(n)} \quad \textit{in} \quad \mathbb{Z}_p \ .$$

Proof. The algebra map $M_r(c_0(\Gamma)) \to M_r(c_0(\Gamma^{(n)}))$ sending f to $f^{(n)}$ is continuous, since we have $\|f^{(n)}\| \le \|f\|$. For f in $1 + pM_r(c_0(\Gamma, \mathbb{Z}_p))$ we therefore get:

$$(\log f)^{(n)} = \log f^{(n)} \quad \text{in } M_r(c_0(\Gamma^{(n)})) \ .$$

The next claim for $g = \log f$ thus implies the assertion. □

Claim. For g in $M_r(c_0(\Gamma))$ we have

$$\mathrm{tr}_\Gamma(g) = \lim_{n\to\infty} \mathrm{tr}_{\Gamma^{(n)}}(g^{(n)}) \ .$$

Proof. We may assume that $r=1$. Writing $g=\sum a_\gamma \gamma$ with $a_\gamma \in \mathbb{Q}_p$, $|a_\gamma|_p \to 0$ for $\gamma \to \infty$, we have

$$\begin{aligned} |\mathrm{tr}_\Gamma(g) - \mathrm{tr}_{\Gamma^{(n)}}(g^{(n)})|_p &= \Big| a_e - \sum_{\overline{\gamma}=e} a_\gamma \Big|_p \\ &= \Big| \sum_{\gamma \in \Gamma_n \setminus e} a_\gamma \Big|_p \\ &\le \max_{\gamma \in \Gamma_n \setminus e} |a_\gamma|_p \, . \end{aligned}$$

For $\varepsilon > 0$ there is a finite subset $S = S_\varepsilon$ of Γ such that $|a_\gamma|_p < \varepsilon$ for $\gamma \in \Gamma \setminus S$. Only e is contained in infinitely many Γ_n's. Hence there is some n_0 such that $(\Gamma_n \setminus e) \cap S = \emptyset$, i.e., $\Gamma_n \setminus e \subset \Gamma \setminus S$ for all $n \ge n_0$. It follows that for $n \ge n_0$ we have $|\mathrm{tr}_\Gamma(g) - \mathrm{tr}_{\Gamma^{(n)}}(g^{(n)})|_p \le \varepsilon$. □

Corollary 18. *Let Γ be a residually finite countable discrete group and f an element of $M_r(\mathbb{Z}\Gamma)$ that is a 1-unit in $M_r(c_0(\Gamma))$. Then the p-adic entropy $h_p(f)$ of the Γ-action on X_f exists for all $\Gamma_n \to e$, and we have*

$$h_p(f) = \log_p \det{}_\Gamma f \quad \textit{in } \mathbb{Z}_p \, .$$

Proof. By Propositions 7 and 15 we have

$$\begin{aligned} \frac{1}{(\Gamma : \Gamma_n)} \log_p |\mathrm{Fix}_{\Gamma_n}(X_f)| &= \frac{1}{(\Gamma : \Gamma_n)} \log_p \det{}_{\mathbb{Q}_p}(\rho_{f^{(n)}}) \\ &= \log_p \det{}_{\Gamma^{(n)}}(f^{(n)}) \, . \end{aligned}$$

Hence the assertion follows from Proposition 17. □

5 A p-adic logarithmic Fuglede–Kadison determinant and its relation to p-adic entropy

Having defined a homomorphism $\log_p \det_\Gamma$ on $1 + pc_0(\Gamma, \mathbb{Z}_p)$ in (7) one would like to use the exact sequence (16) to extend it to $c_0(\Gamma, \mathbb{Z}_p)^*$. However, for infinite groups Γ the abelianization of the group $\mathbb{F}_p[\Gamma]^*$ divided by the image of Γ is not known to be torsion in any generality, as far as I know. However, corresponding results are known for K_1 of $\mathbb{F}_p[\Gamma]$, and this determines our approach, which even for $r = 1$ requires the preceding considerations for matrix algebras.

For a unital ring R recall the embedding $\mathrm{GL}_r(R) \hookrightarrow \mathrm{GL}_{r+1}(R)$ mapping a to $\left(\begin{smallmatrix} a & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Let $\mathrm{GL}_\infty(R)$ be the union of the $\mathrm{GL}_r(R)$'s. We will view elements of $\mathrm{GL}_\infty(R)$ as infinite matrices with 1's on the diagonal and only finitely many further nonzero entries. The subgroup $E_r(R) \subset \mathrm{GL}_r(R)$ of elementary matrices is the subgroup generated by matrices that have 1's on the diagonal and at most one further nonzero entry. Let $E_\infty(R)$ be their union and set $K_1(R) = \mathrm{GL}_\infty(R)/E_\infty(R)$. It is known that we have

$E_\infty(R) = (\mathrm{GL}_\infty(R), \mathrm{GL}_\infty(R))$ and hence that $K_1(R) = \mathrm{GL}_\infty(R)^{\mathrm{ab}}$; cf. [Mil71, § 3]. The Whitehead group over $\mathbb{F}_p$ of a discrete group Γ is defined to be

$$Wh^{\mathbb{F}_p}(\Gamma) := K_1(\mathbb{F}_p[\Gamma])/\langle\Gamma\rangle \; .$$

Here $\langle\Gamma\rangle$ is the image of Γ under the canonical map $\mathbb{F}_p[\Gamma]^* \to K_1(\mathbb{F}_p[\Gamma])$.

We can treat groups for which $Wh^{\mathbb{F}_p}(\Gamma)$ is torsion. According to [FL03, Theorem 1.1], this is the case for torsion-free elementary amenable groups Γ. Recently, in [BLR] it has been shown for a larger class of groups that $Wh^{\mathbb{F}_p}(\Gamma)$ is torsion. Apart from the elementary amenable groups, this class comprises all word hyperbolic groups. It is closed under subgroups, finite products, colimits, and suitable extensions.

Theorem 19. *Let Γ be a countable discrete residually finite group such that $Wh^{\mathbb{F}_p}(\Gamma)$ is torsion. Then there is a unique homomorphism*

$$\log_p \det_\Gamma : K_1(c_0(\Gamma, \mathbb{Z}_p)) \longrightarrow \mathbb{Q}_p$$

with the following properties:
(a) *For every $r \geq 1$ the composition*

$$1 + pM_r(c_0(\Gamma, \mathbb{Z}_p)) \hookrightarrow \mathrm{GL}_r(c_0(\Gamma, \mathbb{Z}_p)) \to K_1(c_0(\Gamma, \mathbb{Z}_p)) \xrightarrow{\log_p \det_\Gamma} \mathbb{Q}_p$$

coincides with the map $\log_p \det_\Gamma$ introduced in (7).
(b) *On the image of Γ in $K_1(c_0(\Gamma, \mathbb{Z}_p))$ the map $\log_p \det_\Gamma$ vanishes.*

Proof. Set $A = c_0(\Gamma, \mathbb{Z}_p)$ and $\overline{A} = A/pA = \mathbb{F}_p[\Gamma]$. The reduction map $A \to \overline{A}$ induces an exact sequence

$$0 \to \Gamma E_\infty(A)(1 + pM_\infty(A))/\Gamma E_\infty(A) \to K_1(A)/\langle\Gamma\rangle \to K_1(\overline{A})/\langle\Gamma\rangle \; . \quad (30)$$

Here $M_\infty(A)$ is the (nonunital) algebra of infinite matrices $(a_{ij})_{i,j\geq 1}$ with only finitely many nonzero entries. Note that $1 + pM_\infty(A)$ is a subgroup of $\mathrm{GL}_\infty(A)$, since $1 + pM_r(A)$ is a subgroup of $\mathrm{GL}_r(A)$. Moreover, $\Gamma E_\infty(A)$ is a normal subgroup of $\mathrm{GL}_\infty(A)$. Hence the sequence (30) becomes an exact sequence:

$$0 \to (1 + pM_\infty(A))/\Gamma E_\infty(A) \cap (1 + pM_\infty(A)) \to K_1(A)/\langle\Gamma\rangle \to K_1(\overline{A})/\langle\Gamma\rangle \; . \quad (31)$$

Since $\mathbb{Q}_p$ is uniquely divisible, this implies the uniqueness assertion in the theorem for any group Γ such that $Wh^{\mathbb{F}_p}(\Gamma) = K_1(\overline{A})/\langle\Gamma\rangle$ is torsion. For the existence, we first note that the homomorphisms defined in (7) induce a homomorphism

$$\log_p \det_\Gamma : 1 + pM_\infty(A) \longrightarrow \mathbb{Z}_p \; .$$

We have to show that $\log_p \det_\Gamma f = 0$ for every f in $1 + pM_\infty(A)$ that also lies in $\Gamma E_\infty(A)$. Under our identification of $\mathrm{GL}_r(A)$ with a subgroup of $\mathrm{GL}_\infty(A)$ we find some $r \geq 1$ such that we have

$$f = i(\gamma) e_1 \cdots e_N \quad \text{in } 1 + pM_r(A) \; .$$

Here the e_i are elementary $r \times r$ matrices and $i(\gamma) = \begin{pmatrix} \gamma & 0 \\ 0 & 1_{r-1} \end{pmatrix}$ for some γ in Γ.

According to Propositions 15 and 17 we have, for any choice of sequence $\Gamma_n \to e$,

$$\log_p \det_\Gamma f = \lim_{n\to\infty} \frac{1}{(\Gamma : \Gamma_n)} \log_p \det_{\mathbb{Q}_p}(\rho_{f^{(n)}}) \; .$$

On the other hand,

$$\det_{\mathbb{Q}_p}(\rho_{f^{(n)}}) = \det_{\mathbb{Q}_p}(\rho_{i(\gamma)^{(n)} e_1^{(n)} \cdots e_N^{(n)}}) = \det_{\mathbb{Q}_p}(\rho_{i(\gamma)^{(n)}}) \prod_i \det_{\mathbb{Q}_p}(\rho_{e_i^{(n)}}) \; .$$

Let $\mathfrak{b}$ be a basis of $\mathbb{Q}_p[\Gamma^{(n)}]$. In the basis $(\mathfrak{b}, \ldots, \mathfrak{b})$ of $\mathbb{Q}_p[\Gamma^{(n)}]^r$ the endomorphism $\rho_{e_i^{(n)}}$ is given by a matrix of $|\Gamma^{(n)}| \times |\Gamma^{(n)}|$-blocks. The diagonal blocks are identity matrices. At most one of the other blocks is nonzero. In particular, the matrix is triangular and we have $\det_{\mathbb{Q}_p}(\rho_{e_i^{(n)}}) = 1$. In the same basis $\rho_{i(\gamma)^{(n)}}$ is a permutation matrix, and hence $\det_{\mathbb{Q}_p}(\rho_{i(\gamma)^{(n)}}) = \pm 1$. It follows that we have $\log_p \det_\Gamma f = 0$, as we wanted to show. □

I think that Theorem 19 should also hold without the condition that Γ is residually finite.

Remark 20. For $\Gamma = \mathbb{Z}^d$ and f in $\mathrm{GL}_r(c_0(\Gamma, \mathbb{Z}_p))$, writing $[f]$ for the class of f in K_1, we have

$$\log_p \det_\Gamma [f] = m_p(\det f),$$

extending equation (8).

This follows from the uniqueness assertion in Theorem 19. Namely, the map $[f] \mapsto m_p(\det f)$ defines a homomorphism on K_1 that according to equation (8) satisfies condition **(a)**. It satisfies condition **(b)** as well, since $\log_p$ vanishes on roots of unity, and hence we have $m_p(t^\nu) = 0$ for all ν in $\mathbb{Z}^d$, cf. Example 5.

Definition 21. *For any group Γ as in the theorem we define the homomorphism $\log_p \det_\Gamma$ on $\mathrm{GL}_r(c_0(\Gamma, \mathbb{Z}_p))$ to be the composition*

$$\log_p \det_\Gamma : \mathrm{GL}_r(c_0(\Gamma, \mathbb{Z}_p)) \longrightarrow K_1(c_0(\Gamma, \mathbb{Z}_p)) \xrightarrow{\log_p \det_\Gamma} \mathbb{Q}_p \; .$$

If we unravel the definitions we get the following description of this map. Given a matrix f in $\mathrm{GL}_r(c_0(\Gamma, \mathbb{Z}_p))$ there are integers $N \geq 1$ and $s \geq r$ such that in $\mathrm{GL}_s(c_0(\Gamma, \mathbb{Z}_p))$ we have $f^N = i(\gamma)\varepsilon g$ with ε in $E_s(c_0(\Gamma, \mathbb{Z}_p))$, g in $1 + pM_s(c_0(\Gamma, \mathbb{Z}_p))$, and $i(\gamma)$ the $s \times s$ matrix $\begin{pmatrix} \gamma & 0 \\ 0 & 1_{s-1} \end{pmatrix}$. Then we have

$$\log_p \det_\Gamma f = \frac{1}{N} \log_p \det_\Gamma g = \frac{1}{N} \mathrm{tr}_\Gamma \log g \; . \tag{32}$$

We can now prove the following extension of Corollary 18.

Theorem 22. *Let Γ be a residually finite countable discrete group such that $Wh^{\mathbb{F}_p}(\Gamma)$ is torsion. Let f be an element of $M_r(\mathbb{Z}\Gamma) \cap \mathrm{GL}_r(c_0(\Gamma, \mathbb{Z}_p))$. Then $h_p(f)$ exists for all $\Gamma_n \to e$ and we have*

$$h_p(f) = \log_p \det{}_\Gamma f \quad \textit{in } \mathbb{Q}_p \, .$$

Proof. Let us write $f^N = i(\gamma)\varepsilon g$ as above. Then by Proposition 7 we have

$$\begin{aligned} \log_p |\mathrm{Fix}_{\,\Gamma_n}(X_f)| &= \log_p \det{}_{\mathbb{Q}_p}(\rho_{f^{(n)}}) \\ &= \frac{1}{N} \log_p \det{}_{\mathbb{Q}_p}(\rho_{i(\gamma)^{(n)}}) + \frac{1}{N} \log_p \det{}_{\mathbb{Q}_p}(\rho_{\varepsilon^{(n)}}) \\ &\quad + \frac{1}{N} \log_p \det{}_{\mathbb{Q}_p}(\rho_{g^{(n)}}) \, . \end{aligned}$$

Note here that the composition

$$M_s(c_0(\Gamma)) \longrightarrow M_s(c_0(\Gamma^{(n)})) \xrightarrow{\rho} \mathrm{End}_{\,\mathbb{Q}_p}(\mathbb{Q}_p \Gamma^{(n)})^s$$

is a homomorphism of algebras.

As in the proof of Theorem 19 we see that the terms $\log_p \det_{\mathbb{Q}_p}(\rho_{i(\gamma)^{(n)}})$ and $\log_p \det_{\mathbb{Q}_p}(\rho_{\varepsilon^{(n)}})$ vanish. This gives

$$\begin{aligned} \frac{1}{(\Gamma : \Gamma_n)} \log_p |\mathrm{Fix}_{\,\Gamma_n}(X_f)| &= \frac{1}{(\Gamma : \Gamma_n)} \frac{1}{N} \log_p \det{}_{\mathbb{Q}_p}(\rho_{g^{(n)}}) \\ &= \frac{1}{N} \log_p \det{}_{\Gamma^{(n)}}(g^{(n)}) \quad \text{by Proposition 15.} \end{aligned}$$

Using Proposition 17 we get in the limit $n \to \infty$ that

$$h_p(f) = \frac{1}{N} \log_p \det{}_\Gamma(g) \overset{(32)}{=} \log_p \det f \, . \qquad \square$$

For groups Γ as in Theorem 22 whose group ring $\mathbb{F}_p\Gamma$ has no zero divisors it is possible to extend the definition of $\log_p \det_\Gamma$ from $c_0(\Gamma, \mathbb{Z}_p)^*$ to $c_0(\Gamma)^*$. Namely, by Proposition 4 we know that

$$c_0(\Gamma)^* = p^{\mathbb{Z}} c_0(\Gamma, \mathbb{Z}_p)^* \quad \text{and} \quad p^{\mathbb{Z}} \cap c_0(\Gamma, \mathbb{Z}_p)^* = 1 \, .$$

Hence there is a unique homomorphism

$$\log_p \det{}_\Gamma : c_0(\Gamma)^* \longrightarrow \mathbb{Q}_p$$

that agrees with $\log_p \det_\Gamma$ previously defined on $c_0(\Gamma, \mathbb{Z}_p)^*$ in Definition 21 and satisfies

$$\log_p \det{}_\Gamma(p) = 0 \, .$$

Let Γ be a torsion-free elementary amenable group. Then according to [KLM88, Theorem 1.4] the group ring $\mathbb{F}_p\Gamma$ has no zero divisors, and according to [FL03, Theorem 1.1] the group $Wh^{\mathbb{F}_p}(\Gamma)$ is torsion. Hence $\log_p \det_\Gamma$ is defined on $c_0(\Gamma)^*$, and this is the map used in Theorem 2.

Proof of Theorem 2. Writing f in $\mathbb{Z}\Gamma \cap c_0(\Gamma)^*$ as a product $f = p^\nu g$ with g in $c_0(\Gamma, \mathbb{Z}_p)^*$ it follows that $g \in \mathbb{Z}\Gamma$, and Proposition 7 shows that we have

$$\begin{aligned}\log_p |\mathrm{Fix}_{\Gamma_n}(X_f)| &= \log_p \det_{\mathbb{Q}_p}(\rho_{f^{(n)}}) \\ &= \log_p \det_{\mathbb{Q}_p}(\rho_{g^{(n)}}) \\ &= \log_p |\mathrm{Fix}_{\Gamma_n}(X_g)| \ .\end{aligned}$$

Note here that we have $\log_p(p) = 0$. It follows from Theorem 22 applied to g that for all $\Gamma_n \to e$ we get

$$h_p(f) = h_p(g) = \log_p \det_\Gamma g = \log_p \det_\Gamma f \ .$$

□

For $\Gamma = \mathbb{Z}^d$ it follows from Remark 20 that for any f in $c_0(\mathbb{Z}^d)^* = \mathbb{Q}_p\langle t_1^{\pm 1}, \ldots, t_d^{\pm 1}\rangle^*$ we have

$$\log_p \det_\Gamma f = m_p(f) \ .$$

Hence Theorem 1 is a special case of Theorem 2.

Concerning approximations of $\log_p \det_\Gamma f$ we note that Proposition 17 extends to more general cases.

Proposition 23. *Let Γ be a residually finite countable discrete group and let $\Gamma_n \to e$ be as in the introduction. For f in $M_r(c_0(\Gamma))$ let $f^{(n)}$ be its image in $M_r(\mathbb{Q}_p\Gamma^{(n)})$. Then the formula*

$$\log_p \det_\Gamma f = \lim_{n\to\infty} \frac{1}{(\Gamma : \Gamma_n)} \log_p \det_{\mathbb{Q}_p}(\rho_{f^{(n)}}) \tag{33}$$

holds whenever $\log_p \det_\Gamma f$ is defined. These are the cases
(a) *in which f is in $1 + pM_r(c_0(\Gamma, \mathbb{Z}_p))$;*
(b) *in which $Wh^{\mathbb{F}_p}(\Gamma)$ is torsion and f is in $\mathrm{GL}_r(c_0(\Gamma, \mathbb{Z}_p))$;*
(c) *in which $Wh^{\mathbb{F}_p}(\Gamma)$ is torsion, $\mathbb{F}_p\Gamma$ has no zero divisors and f is in $c_0(\Gamma)^*$.*

Proof. The assertions follow from Propositions 15 and 17 together with calculations as in the proofs of Theorems 19 and 22. □

We end the paper with some open questions: Is there a dynamical criterion for the existence of the limit defining p-adic entropy? Is there a notion of "p-adic expansiveness" for Γ-actions on compact spaces X that for the systems X_f with f in $M_r(\mathbb{Z}\Gamma)$ translates into the condition that f be invertible in $M_r(c_0(\Gamma))$ (or in $M_r(c_0(\Gamma, \mathbb{Z}_p))$)? In fact, I assume that p-adic entropy can be defined only for "p-adically expansive" systems; cf. [BD99, Remark after Proposition 1.3]. What is the dynamical meaning of Proposition 12? Is there a direct proof that the limit in formula (33) exists?

Finally, in [BD99] a second version of a p-adic Mahler measure was defined that involves both the p-adic and the archimedean valuations of $\mathbb{Q}$. Can this be obtained for the systems X_f by doing something more involved with the fixed points than taking their cardinalities and forming the limit (3)?

References

[BD99] A. Besser, C. Deninger, *p-adic Mahler measures*, J. reine angew. Math., 517:19–50, 1999.

[BLR] A. Bartels, W. Lück, H. Reich, *On the Farrell–Jones conjecture and its applications*, in preparation.

[Bou72] N. Bourbaki, *Groupes et algèbres de Lie, Chapitres 2 et 3*, Hermann, 1972.

[Den97] C. Deninger, *Deligne periods of mixed motives, k-theory and the entropy of certain $\mathbb{Z}^n$-actions*, J. Amer. Math. Soc., 10:259–281, 1997.

[Den06] C. Deninger, *Fuglede–Kadison determinants and entropy for actions of discrete amenable groups*, J. Amer. Math. Soc., 19:717–736, 2006.

[DS] C. Deninger, K. Schmidt, *Expansive algebraic actions of discrete residually finite amenable groups and their entropy*, Ergodic Theory Dynam. Systems, 27, no. 3:769–786, 2007.

[ER01] M. Einsiedler, H. Rindler, *Algebraic actions of the discrete Heisenberg group and other non-abelian groups*, Aequationes Math., 62:117–135, 2001.

[FK52] B. Fuglede, R.V. Kadison, *Determinant theory in finite factors*, Ann. Math., 55:520–530, 1952.

[FL03] F.T. Farrell, P. Linnell, *Whitehead groups and the Bass conjecture*, Math. Ann., 326:723–757, 2003.

[Har77] R. Hartshorne, *Algebraic Geometry*, volume 52 of GTM. Springer, 1977.

[KLM88] P.H. Kropholler, P.A. Linnell, J.A. Moody, *Applications of a new k-theoretic theorem to soluble group rings* Proc. Amer. Math. Soc., 104:675–684, 1988.

[LSW90] D. Lind, K. Schmidt, T. Ward, *Mahler measure and entropy for commuting automorphisms of compact groups*, Invent. Math., 101:593–629, 1990.

[Mil71] J. Milnor, *Introduction to Algebraic K-theory*, volume No. 72 of Annals of Mathematical Studies, Princeton University Press, 1971.

[Sch95] K. Schmidt, *Dynamical Systems of Algebraic Origin*, volume 128 of Progress in Mathematics, Birkhäuser, 1995.

Finite Subgroups of the Plane Cremona Group

Igor V. Dolgachev[1] and Vasily A. Iskovskikh[2]

[1]Department of Mathematics, University of Michigan, 525 E. University Ave., Ann Arbor, MI, 49109 idolga@umich.edu

To Yuri I. Manin

Summary. This paper completes the classic and modern results on classification of conjugacy classes of finite subgroups of the group of birational automorphisms of the complex projective plane.

Key words: Cremona group, Del Pezzo surfaces, conic bundles.

2000 Mathematics Subject Classifications: 14E07 (Primary); 14J26, 14J50, 20B25 (Secondary)

1 Introduction

The Cremona group $\mathrm{Cr}_k(n)$ over a field k is the group of birational automorphisms of the projective space $\mathbb{P}^n_k$, or equivalently, the group of k-automorphisms of the field $k(x_1, x_2, \ldots, x_n)$ of rational functions in n independent variables. The group $\mathrm{Cr}_k(1)$ is the group of automorphisms of the projective line, and hence it is isomorphic to the projective linear group $\mathrm{PGL}_k(2)$. Already in the case $n = 2$ the group $\mathrm{Cr}_k(2)$ is not well understood in spite of extensive classical literature (e.g., [21], [35]) on the subject as well as some modern research and expositions of classical results (e.g., [2]). Very little is known about the Cremona groups in higher-dimensional spaces.

In this paper we consider the plane Cremona group over the field of complex numbers, denoted by $\mathrm{Cr}(2)$. We return to the classical problem of classification of finite subgroups of $\mathrm{Cr}(2)$. The classification of finite subgroups of $\mathrm{PGL}(2)$ is well known and goes back to F. Klein. It consists of cyclic, dihedral,

[1]The author was supported in part by NSF grant 0245203.
[2]The author was supported in part by RFBR 05-01-00353-a RFBR 08-01-00395-a, grant CRDF RUMI 2692-MO-05 and grant of NSh 1987-2008.1.

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*, Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_11,

tetrahedral, octahedral, and icosahedral groups. Groups of the same type and order constitute a unique conjugacy class in PGL(2). Our goal is to find a similar classification in the two-dimensional case.

The history of this problem begins with the work of E. Bertini [10], who classified conjugacy classes of subgroups of order 2 in Cr(2). Already in this case the answer is drastically different. The set of conjugacy classes is parametrized by a disconnected algebraic variety whose connected components are respectively isomorphic to either the moduli spaces of hyperelliptic curves of genus g (de Jonquières involutions), or the moduli space of canonical curves of genus 3 (Geiser involutions), or the moduli space of canonical curves of genus 4 with vanishing theta characteristic (Bertini involutions). Bertini's proof was considered to be incomplete even according to the standards of rigor of nineteenth-century algebraic geometry. A complete and short proof was published only a few years ago by L. Bayle and A. Beauville [5].

In 1894, G. Castelnuovo [16], as an application of his theory of adjoint linear systems, proved that any element of finite order in Cr(2) leaves invariant either a net of lines, or a pencil of lines, or a linear system of cubic curves with $n \leq 8$ base points. A similar result was claimed earlier by S. Kantor in his memoir which was awarded a prize by the Accademia delle Scienze di Napoli in 1883. However Kantor's arguments, as was pointed out by Castelnuovo, required justifications. Kantor went much further and announced a similar theorem for arbitrary finite subgroups of Cr(2). He proceeded to classify possible groups in each case (projective linear groups, groups of de Jonquières type, and groups of type M_n). A much clearer exposition of his results can be found in a paper of A. Wiman [50]. Unfortunately, Kantor's classification, even with some correction made by Wiman, is incomplete for the following two reasons. First, only maximal groups were considered, and even some of them were missed. The most notorious example is a cyclic group of order 8 of automorphisms of a cubic surface, also missed by B. Segre [48] (see [34]). Second, although Kantor was aware of the problem of conjugacy of subgroups, he did not attempt to fully investigate this problem.

The goal of our work is to complete Kantor's classification. We use a modern approach to the problem initiated in the works of Yuri Manin and the second author (see a survey of their results in [39]). Their work gives a clear understanding of the conjugacy problem via the concept of a rational G-surface. It is a pair (S, G) consisting of a nonsingular rational projective surface and a subgroup G of its automorphism group. A birational map $S - \to \mathbb{P}^2_k$ realizes G as a finite subgroup of Cr(2). Two birational isomorphic G-surfaces define conjugate subgroups of Cr(2), and conversely, a conjugacy class of a finite subgroup G of Cr(2) can be realized as a birational isomorphism class of G-surfaces. In this way, classification of conjugacy classes of subgroups of Cr(2) becomes equivalent to the birational classification of G-surfaces. A G-equivariant analogue of a minimal surface allows one to concentrate on the study of minimal G-surfaces, i.e., surfaces that cannot be G-equivariantly birationally and regularly mapped to another G-surface.

Minimal G-surfaces turn out to be G-isomorphic either to the projective plane, or a conic bundle, or a Del Pezzo surface of degree $d = 9 - n \leq 6$ and $d = 8$. This leads to groups of projective transformations, or groups of de Jonquières type, or groups of type M_n, respectively. To complete the classification one requires

- to classify all finite groups G that may occur in a minimal G-pair (S, G);
- to determine when two minimal G-surfaces are birationally isomorphic.

To solve the first part of the problem one has to compute the full automorphism group of a conic bundle surface or a Del Pezzo surface (in the latter case this was essentially accomplished by Kantor and Wiman), then make a list of all finite subgroups which act minimally on the surface (this did not come up in the works of Kantor and Wiman). The second part is less straightforward. For this we use the ideas from Mori theory to decompose a birational map of rational G-surfaces into elementary links. This theory was successfully applied in the arithmetic case, where the analogue of the group G is the Galois group of the base field (see [39]). We borrow these results with obvious modifications adjusted to the geometric case. Here we use the analogy between k-rational points in the arithmetic case (fixed points of the Galois action) and fixed points of the G-action. As an important implication of the classification of elementary G-links is the rigidity property of groups of type M_n with $n \geq 6$: any minimal Del Pezzo surface (S, G) of degree $d \leq 3$ is not isomorphic to a minimal G-surface of different type. This allows us to avoid much of the painful analysis of possible conjugacy for a lot of groups.

The large amount of group-theoretical computations needed for the classification of finite subgroups of groups of automorphisms of conic bundles and Del Pezzo surfaces makes us expect some possible gaps in our classification. This seems to be the destiny of enormous classification problems. We hope that our hard work will be useful for the future faultless classification of conjugacy classes of finite subgroups of $\mathrm{Cr}(2)$.

It is appropriate to mention some recent work on the classification of conjugacy classes of subgroups of $Cr(2)$. We have already mentioned the work of L. Bayle and A. Beauville on groups of order 2. The papers [8], [23], [52] study groups of prime orders, Beauville's paper [9] classifies p-elementary groups, and a thesis of J. Blanc [6] contains a classification of all finite abelian groups. The second author studies two nonconjugate classes of subgroups isomorphic to $S_3 \times \mathbb{Z}/2\mathbb{Z}$. In the work of S. Bannai and H. Tokunaga [4], examples are given of nonconjugate subgroups isomorphic to S_4 and A_5.

This paper is partly based on the lectures by the first author in workshops on Cremona transformations in Torino in September 2005 and Lisbon in May 2006. He takes the opportunity to thank the organizers for the invitation and for providing a good audience. We would both like to thank A. Beauville, Chenyang Xu, and especially J. Blanc for pointing out some errors in the previous versions of our paper.

This paper is dedicated to Yuri Ivanovich Manin to whom both authors are grateful for initiating them into algebraic geometry more than 40 years ago. Through his seminars, inspiring lectures, and as the second author's thesis adviser, he was an immeasurable influence on our mathematical lives.

2 First examples

2.1 Homaloidal linear systems

We will be working over the field of complex numbers. Recall that a dominant rational map $\chi : \mathbb{P}^2 - \to \mathbb{P}^2$ is given by a 2-dimensional linear system $\mathcal{H}$ equal to the proper transform of the linear system of lines $\mathcal{H}' = |\ell|$ in the target plane. A choice of a basis in $\mathcal{H}$ gives an explicit formula for the map in terms of homogeneous coordinates

$$(x_0', x_1', x_2') = (P_0(x_0, x_1, x_2), P_1(x_0, x_1, x_2), P_2(x_0, x_1, x_2)),$$

where P_0, P_1, P_2 are linearly independent homogeneous polynomials of degree d, called the (algebraic) *degree* of the map. This is the smallest number d such that $\mathcal{H}$ is contained in the complete linear system $|\mathcal{O}_{\mathbb{P}^2}(d)|$ of curves of degree d in the plane. By definition of the proper transform, the linear system $\mathcal{H}$ has no fixed components, or equivalently, the polynomials P_i are mutually coprime. The birational map χ is not a projective transformation if and only if the degree is larger than 1, or equivalently, when χ has *base points*, the common zeros of the members of the linear system. A linear system defining a birational map is called a *homaloidal* linear system. Being a proper transform of a general line under a birational map, its general member is an irreducible rational curve. Also, two general curves from the linear system intersect outside the base points at one point. These two conditions characterize homaloidal linear systems (more about this later).

2.2 Quadratic transformations

A quadratic Cremona transformation is a birational map $\chi : \mathbb{P}^2 - \to \mathbb{P}^2$ of degree 2. The simplest example is the *standard quadratic transformation* defined by the formula

$$\tau_1 : (x_0, x_1, x_2) \mapsto (x_1x_2, x_0x_2, x_0x_1). \tag{2.1}$$

In affine coordinates this is given by $\tau_1 : (x, y) \mapsto (\frac{1}{x}, \frac{1}{y})$. It follows from the definition that $\tau_1^{-1} = \tau_1$, i.e., τ_1 is a birational involution of $\mathbb{P}^2$. The base points of τ_1 are the points $p_1 = (1, 0, 0), p_2 = (0, 1, 0), p_3 = (0, 0, 1)$. The transformation maps an open subset of the coordinate line $x_i = 0$ to the point p_i. The homaloidal linear system defining τ_1 is the linear system of conics passing through the points p_1, p_2, p_3.

The Möbius transformation $x \mapsto x^{-1}$ of $\mathbb{P}^1$ is conjugate to the transformation $x \mapsto -x$ (by means of the map $x \mapsto \frac{x-1}{x+1}$). This shows that the standard Cremona transformation τ_1 is conjugate in $\mathrm{Cr}(2)$ to a projective transformation given by

$$(x_0, x_1, x_2) \mapsto (x_0, -x_1, -x_2).$$

When we change the homaloidal linear system defining τ_1 to the homaloidal linear system of conics passing through the point p_2, p_3 and tangent at p_3 to the line $x_0 = 0$, we obtain the transformation

$$\tau_2 : (x_0, x_1, x_2) \mapsto (x_1^2, x_0x_1, x_0x_2). \tag{2.2}$$

In affine coordinates it is given by $(x, y) \mapsto (\frac{1}{x}, \frac{y}{x^2})$. The transformation τ_2 is also a birational involution conjugate to a projective involution. To see this we define a rational map $\chi : \mathbb{P}^2 - \to \mathbb{P}^3$ by the formula $(x_0, x_1, x_2) \mapsto (x_1^2, x_0x_1, x_0x_2, x_1x_2)$. The Cremona transformation τ_2 acts on $\mathbb{P}^3$ via this transformation by $(u_0, u_1, u_2, u_3) \mapsto (u_1, u_0, u_3, u_2)$. Composing with the projection of the image from the fixed point $(1, 1, 1, 1)$ we get a birational map $(x_0, x_1, x_2) \mapsto (y_0, y_1, y_2) = (x_1(x_0 - x_1), x_0x_2 - x_1^2, x_1(x_2 - x_1))$. It defines the conjugation of τ_2 with the projective transformation $(y_0, y_1, y_2) \mapsto (-y_0, y_2 - y_0, y_1 - y_0)$.

Finally, we could further "degenerate" τ_1 by considering the linear system of conics passing through the point p_3 and intersecting at this point with multiplicity 3. This linear system defines a birational involution

$$\tau_3 : (x_0, x_1, x_2) \mapsto (x_0^2, x_0x_1, x_1^2 - x_0x_2). \tag{2.3}$$

Again it can be shown that τ_3 is conjugate to a projective involution.

Recall that a birational transformation is not determined by the choice of a homaloidal linear system; one has to choose additionally a basis of the linear system. In the above examples, the basis is chosen to make the transformation an involution.

2.3 De Jonquières involutions

Here we exhibit a series of birational involutions that are not conjugate to each other and not conjugate to a projective involution. In affine coordinates they are given by the formula

$$\mathrm{dj}_P : (x, y) \mapsto (x, \frac{P(x)}{y}), \tag{2.4}$$

where $P(x)$ is a polynomial of degree $2g+1$ or $2g+2$ without multiple roots. The conjugation by the transformation $(x, y) \mapsto (\frac{ax+b}{cx+d}, y)$ shows that the conjugacy class of dj_P depends only on the orbit of the set of roots of P with respect to the group $\mathrm{PGL}(2)$, or in other words, on the birational class of the hyperelliptic curve

$$y^2 + P(x) = 0. \tag{2.5}$$

The transformation dj_P has the following beautiful geometric interpretation. Consider the projective model H_{g+2} of the hyperelliptic curve (2.5) given by the homogeneous equation of degree $g+2$

$$x_2^2 F_g(T_0, T_1) + 2x_2 F_{g+1}(T_0, T_1) + F_{g+2}(x_0, x_1) = 0, \tag{2.6}$$

where

$$D = F_{g+1}^2 - F_g F_{g+2} = x_0^{2g+2} P(x_1/x_0)$$

is the homogenization of the polynomial $P(x)$. The curve H_{g+2} has an ordinary singular point of multiplicity g at $q = (0, 0, 1)$, and the projection from this point to $\mathbb{P}^1$ exhibits the curve as a double cover of $\mathbb{P}^1$ branched over the $2g+2$ zeroes of D.

Consider the affine set $x_2 = 1$ with affine coordinates $(x, y) = (x_0/x_2, x_1/x_2)$. A general line $y = kx$ intersects the curve H_{g+2} at the point $q = (0, 0)$ with multiplicity g and at two other points $(\alpha, k\alpha)$ and $(\alpha', k\alpha')$, where α, α' are the roots of the quadratic equation

$$t^2 F_{g+2}(1, k) + 2t F_{g+1}(1, k) + F_g(1, k) = 0.$$

Take a general point $p = (x, kx)$ on the line and define the point $p' = (x', kx')$ such that the pairs $(\alpha, k\alpha), (\alpha', k\alpha')$ and $(x, kx), (x', kx')$ are harmonic conjugate. This means that x, x' are the roots of the equation $at^2 + 2bt + c = 0$, where $aF_g(1, k) + cF_{g+2}(1, k) - 2bF_{g+1}(1, k) = 0$. Since $x + x' = -2b/a, xx' = c/a$ we get $F_g(1, k) + xx'F_{g+2}(1, k) + (x + x')F_{g+1}(1, k) = 0$. We express x' as $(ax + b)/(cx + d)$ and solve for (a, b, c, d) to obtain

$$x' = \frac{-F_{g+1}(1, k)x - F_g(1, k)}{xF_{g+2}(1, k) + F_{g+1}(1, k)}.$$

Since $k = y/x$, after changing the affine coordinates $(x, y) = (x_0/x_2, x_1/x_2)$ to $(X, Y) = (x_1/x_0, x_2/x_0) = (y/x, 1/x)$, we get

$$IH_{g+2} : (X, Y) \mapsto (X', Y') := \big(X, \frac{-YP_{g+1}(X) - P_{g+2}(X)}{P_g(X)Y + P_{g+1}(X)}\big), \tag{2.7}$$

where $P_i(X) = F_i(1, X)$. Let $T : (x, y) \mapsto (x, yP_g + P_{g+1})$. Taking $P(x) = P_{g+1}^2 - P_g P_{g+2}$, we check that $T^{-1} \circ \mathrm{dj}_P \circ T = IH_{g+2}$. This shows that our geometric de Jonquières involution IH_{g+2} given by (2.7) is conjugate to the de Jonquières involution dj_P defined by (2.4).

Let us rewrite (2.7) in homogeneous coordinates:

$$\begin{aligned} x_0' &= x_0(x_2 F_g(x_0, x_1) + F_{g+1}(x_0, x_1)), \\ x_1' &= x_1(x_2 F_g(x_0, x_1) + F_{g+1}(x_0, x_1)), \\ x_2' &= -x_2 F_{g+1}(x_0, x_1) - F_{g+2}(x_0, x_1). \end{aligned} \tag{2.8}$$

Now it is clear that the homaloidal linear system defining IH_{g+2} consists of curves of degree $g+2$ which pass through the singular point q of the hyperelliptic curve (2.6) with multiplicity g. Other base points satisfy

$$x_2F_g(x_0,x_1)+F_{g+1}(x_0,x_1)=-x_2F_{g+1}(x_0,x_1)-F_{g+2}(x_0,x_1)=0.$$

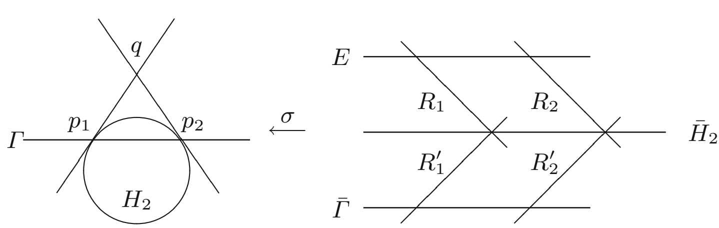
Eliminating x_2, we get the equation $F_{g+1}^2-F_gF_{g+2}=0$, which defines the set of the $2g+2$ ramification points $p_1,\dots,p_{2g+2}$ of the projection $H_{g+2}\setminus\{q\}\to\mathbb{P}^1$.

Let

$$\Gamma: x_2F_g(x_0,x_1)+F_{g+1}(x_0,x_1)=0$$

be the first polar Γ of H_{g+2} with respect to the point q. The transformation IH_{g+2} blows down Γ and the lines $\langle q,p_i\rangle$ to points. It follows immediately from (1) that the set of fixed points of the involution IH_{g+2} outside the base locus is equal to the hyperelliptic curve (2.6). Also we see that the pencil of lines through q is invariant with respect to IH_{g+2}.

Let $\sigma: S\to\mathbb{P}^2$ be the blowup of the point q and the points $p_1,\dots,p_{2g+2}$. The full preimage of the line $\ell_i=\langle q,p_i\rangle$ consists of two irreducible components, each isomorphic to $\mathbb{P}^1$. They intersect transversally at one point. We will call such a reducible curve a *bouquet* of two $\mathbb{P}^1$'s. One component is the exceptional curve $R_i=\sigma^{-1}(p_i)$, and another one is the proper transform R_i' of the line ℓ_i. The proper transform of H_{g+2} intersects $\sigma^{-1}(\ell_i)$ at its singular point. Thus the proper transform $\bar{H}_{g+2}$ of the hyperelliptic curve H_{g+2} intersects the exceptional curve $E=\sigma^{-1}(q)$ at the same points where the proper transform of lines ℓ_i intersect E. The proper transform $\bar{\Gamma}$ of Γ intersects R_i at one nonsingular point, and intersects E at g points, the same points where the proper inverse transform $\bar{H}_{g+2}$ of H_{g+2} intersects E. The involution IH_{g+2} lifts to a biregular automorphism τ of S. It switches the components R_i and R_i' of $\sigma^{-1}(\ell_i)$, switches E with $\bar{\Gamma}$, and fixes the curve $\bar{H}_{g+2}$ pointwise. The pencil of lines through q defines a morphism $\phi: S\to\mathbb{P}^1$ whose fibres over the points corresponding to the lines ℓ_i are isomorphic to a bouquet of two $\mathbb{P}^1$'s. All other fibres are isomorphic to $\mathbb{P}^1$. This is an example of a *conic bundle* or a Mori fibration (or in the archaic terminology of [38], a minimal rational surface with a pencil of rational curves).



To show that the birational involutions $IH_{g+2}, g>0$, are not conjugate to each other or to a projective involution we use the following.

Lemma 2.1. *Let G be a finite subgroup of* $\mathrm{Cr}(2)$ *and let $C_1, \ldots, C_k$ be non-rational irreducible curves on $\mathbb{P}^2$ such that each of them contains an open subset C_i^0 whose points are fixed under all $g \in G$. Then the set of birational isomorphism classes of the curves C_i is an invariant of the conjugacy class of G in* $\mathrm{Cr}(2)$.

Proof. Suppose $G = T \circ G' \circ T^{-1}$ for some subgroup G' of $\mathrm{Cr}(2)$ and some $T \in \mathrm{Cr}(2)$. Then, replacing C_i^0 by a smaller open subset we may assume that $T^{-1}(C_i^0)$ is defined and consists of fixed points of G'. Since C_i is not rational, $T^{-1}(C_i^0)$ is not a point, and hence its Zariski closure is a rational irreducible curve C_i' birationally isomorphic to C_i that contains an open subset of fixed points of G'. □

Since a connected component of the fixed locus of a finite group of projective transformations is a line or a point, we see that IH_{g+2} is not conjugate to a subgroup of projective transformations for any $g > 0$. Since IH_{g+2} is conjugate to some involution (2.4), where $P(x)$ is determined by the birational isomorphism class of H_{g+2}, we see from the previous lemma that IH_{g+2} is conjugate to $IH'_{g'+2}$ if and only if $g = g'$ and the curves H_{g+2} and H'_{g+2} are birationally isomorphic. Finally, let us look at the involution IH_2. It is a quadratic transformation that is conjugate to the quadratic transformation $\tau_2 : (x, y) \mapsto (x, x/y)$.

A de Jonquières involution (2.4) is a special case of a Cremona transformation of the form

$$(x, y) \mapsto \left(\frac{ax + b}{cx + d}, \frac{r_1(x)y + r_2(x)}{r_3(x)y + r_4(x)} \right),$$

where $a, b, c, d \in \mathbb{C}, ad - bc \neq 0$, and $r_i(x) \in \mathbb{C}(x)$ with $r_1(x)r_4(x) - r_2(x)r_3(x) \neq 0$. These transformations form a subgroup of $\mathrm{Cr}(2)$ called a *de Jonquières subgroup* and denoted by $\mathrm{dJ}(2)$. Of course, its definition requires a choice of a transcendence basis of the field $\mathbb{C}(\mathbb{P}^2)$. If we identify $\mathrm{Cr}(2)$ with the group $\mathrm{Aut}_{\mathbb{C}}(\mathbb{C}(x, y))$, and consider the field $\mathbb{C}(x, y)$ as a field $K(y)$, where $K = \mathbb{C}(x)$, then

$$\mathrm{dJ}(2) \cong \mathrm{PGL}_{\mathbb{C}(x)}(2) \rtimes \mathrm{PGL}(2),$$

where $\mathrm{PGL}(2)$ acts on $\mathrm{PGL}_{\mathbb{C}(x)}(2)$ via Möebius transformations of the variable x.

It is clear that all elements from $\mathrm{dJ}(2)$ leave the pencil of lines parallel to the y-axis invariant. One can show that a subgroup of $\mathrm{Cr}(2)$ that leaves a pencil of rational curves invariant is conjugate to a subgroup of $\mathrm{dJ}(2)$.

2.4 Geiser and Bertini involutions

The classical definition of a *Geiser involution* is as follows [30]. Fix seven points $p_1, \ldots, p_7$ in $\mathbb{P}^2$ in general position (we will make this more precise later). The linear system L of cubic curves through the seven points is two-dimensional. Take a general point p and consider the pencil of curves from L

passing through p. Since a general pencil of cubic curves has nine base points, we can define $\gamma(p)$ as the ninth base point of the pencil. One can show that the algebraic degree of a Geiser involution is equal to 8. Another way to see a Geiser involution is as follows. The linear system L defines a rational map of degree 2,

$$f : \mathbb{P}^2 - \to |L|^* \cong \mathbb{P}^2.$$

The points p and $\gamma(p)$ lie in the same fibre. Thus γ is a birational deck transformation of this cover. Blowing up the seven points, we obtain a Del Pezzo surface S of degree 2 (more about this later), and a regular map of degree 2 from S to $\mathbb{P}^2$. The Geiser involution γ becomes an automorphism of the surface S.

It is easy to see that the fixed points of a Geiser involution lie on the ramification curve of f. This curve is a curve of degree 6 with double points at the points $p_1, \ldots, p_7$. It is birationally isomorphic to a canonical curve of genus 3. Applying Lemma 2.1, we obtain that a Geiser involution is not conjugate to any de Jonquières involution IH_{g+2}. Also, as we will see later, the conjugacy classes of Geiser involutions are in a bijective correspondence with the moduli space of canonical curves of genus 3.

To define a *Bertini involution* we fix eight points in $\mathbb{P}^2$ in general position and consider the pencil of cubic curves through these points. It has the ninth base point p_9. For any general point p there will be a unique cubic curve $C(p)$ from the pencil which passes through p. Take p_9 for the zero in the group law of the cubic $C(p)$ and define $\beta(p)$ as the negative $-p$ with respect to the group law. This defines a birational involution on $\mathbb{P}^2$, a Bertini involution [10]. One can show that the algebraic degree of a Bertini involution is equal to 17. We will see later that the fixed points of a Bertini involution lie on a canonical curve of genus 4 with vanishing theta characteristic (isomorphic to a nonsingular intersection of a cubic surface and a quadric cone in $\mathbb{P}^3$). So, a Bertini involution is not conjugate to a Geiser involution or a de Jonquières involution. It can be realized as an automorphism of the blowup of the eight points (a Del Pezzo surface of degree 1), and the quotient by this involution is isomorphic to a quadratic cone in $\mathbb{P}^3$.

3 Rational *G*-surfaces

3.1 Resolution of indeterminacy points

Let $\chi : S - \to S'$ be a birational map of nonsingular projective surfaces. It is well known (see [33]) that there exist birational morphisms of nonsingular surfaces $\sigma : X \to S$ and $\phi : X \to S'$ such that the following diagram is commutative:

$$\begin{tikzcd} & X \arrow[dl, "\sigma"'] \arrow[dr, "\phi"] & \\ S \arrow[rr, dashed] & & S'. \end{tikzcd} \tag{3.1}$$

It is called a *resolution of indeterminacy points* of χ. Recall also that any birational morphism can be factored into a finite sequence of blowups of points. Let

$$\sigma : X = X_N \xrightarrow{\sigma_N} X_{N-1} \xrightarrow{\sigma_{N-1}} \cdots \xrightarrow{\sigma_2} X_1 \xrightarrow{\sigma_1} X_0 = S \tag{3.2}$$

be such a factorization. Here $\sigma_i : X_i \to X_{i-1}$ is the blowup of a point $x_i \in X_{i-1}$. Let

$$E_i = \sigma_i^{-1}(x_i), \quad \mathcal{E}_i = (\sigma_{i+1} \circ \ldots \circ \sigma_N)^{-1}(E_i). \tag{3.3}$$

Let H' be a very ample divisor class on S' and $\mathcal{H}'$ be the corresponding complete linear system $|H'|$. Let $\mathcal{H}_N = \phi^*(\mathcal{H}')$. Define $m(x_N)$ as the smallest positive number such that $\mathcal{H}_N + m(x_N)E_N = \sigma_N^*(\mathcal{H}_{N-1})$ for some linear system $\mathcal{H}_{N-1}$ on X_{N-1}. Then proceed inductively to define linear systems $\mathcal{H}_k$ on each X_k such that $\mathcal{H}_{k+1} + m(x_{k+1})E_{k+1} = \sigma_{k+1}^*(\mathcal{H}_k)$, and finally a linear system $\mathcal{H} = \mathcal{H}_0$ on S such that $\mathcal{H}_1 + m(x_1)E_1 = \sigma_1^*(\mathcal{H})$. It follows from the definition that

$$\phi^*(\mathcal{H}') = \sigma^*(\mathcal{H}) - \sum_{i=1}^{N} m(x_i)\mathcal{E}_i. \tag{3.4}$$

The proper transform of $\mathcal{H}'$ on S under χ is contained in the linear system $\mathcal{H}$. It consists of curves which pass through the points x_i with multiplicity $\geq m_i$. We denote it by

$$\chi^{-1}(\mathcal{H}') = |H - m(x_1)x_1 - \cdots - m(x_N)x_N|,$$

where $\mathcal{H} \subset |H|$. Here for a curve on S to pass through a point $x_i \in X_{i-1}$ with multiplicity $\geq m(x_i)$ means that the proper transform of the curve on X_{i-1} has x_i as a point of multiplicity $\geq m(x_i)$. The divisors $\mathcal{E}_i$ are called the *exceptional curves* of the resolution $\sigma : X \to S$ of the birational map χ. Note that $\mathcal{E}_i$ is an irreducible curve if and only if $\sigma_{i+1} \circ \ldots \circ \sigma_N : X \to X_i$ is an isomorphism over $E_i = \sigma^{-1}(x_i)$.

The set of points $x_i \in X_i, i = 1, \ldots, N$, is called the set of *indeterminacy points*, or *base points*, or *fundamental points* of χ. Note that, strictly speaking, only one of them, x_1, lies in S. However, if $\sigma_1 \circ \ldots \circ \sigma_i : X_i \to S$ is an isomorphism in a neighborhood of x_{i+1} we can identify this point with a point in S. Let $\{x_i, i \in I\}$ be the set of such points. Points $x_j, j \notin I$, are *infinitely near points*. A precise meaning of this classical notion is as follows.

Let S be a nonsingular projective surface and $\mathcal{B}(S)$ be the category of birational morphisms $\pi : S' \to S$ of nonsingular projective surfaces. Recall that a morphism from $(S' \xrightarrow{\pi'} S)$ to $(S'' \xrightarrow{\pi''} S)$ in this category is a regular map $\phi : S' \to S''$ such that $\pi'' \circ \phi = \pi'$.

Definition 3.1. *The* bubble space S^{bb} *of a nonsingular surface* S *is the factor set*

$$S^{\mathrm{bb}} = \Big(\bigcup_{(S' \xrightarrow{\pi'} S) \in \mathcal{B}(S)} S' \Big)/R,$$

where R *is the following equivalence relation:* $x' \in S'$ *is equivalent to* $x'' \in S''$ *if the rational map* $\pi''^{-1} \circ \pi' : S' - \to S''$ *maps isomorphically an open neighborhood of* x' *to an open neighborhood of* x''*.*

It is clear that for any $\pi : S' \to S$ from $\mathcal{B}(S)$ we have an injective map $i_{S'} : S' \to S^{\mathrm{bb}}$. We will identify points of S' with their images. If $\phi : S'' \to S'$ is a morphism in $\mathcal{B}(S)$ which is isomorphic in $\mathcal{B}(S')$ to the blowup of a point $x' \in S'$, any point $x'' \in \phi^{-1}(x')$ is called *infinitely near point* to x' of the first order. This is denoted by $x'' \succ x'$. By induction, one defines an infinitely near point of order k, denoted by $x'' \succ_k x'$. This defines a partial order on S^{bb}.

We say that a point $x \in S^{\mathrm{bb}}$ is of height k if $x \succ_k x_0$ for some $x_0 \in S$. This defines the *height function* on the bubble space

$$\mathrm{ht}_S : S^{\mathrm{bb}} \to \mathbb{N}.$$

Clearly, $S = \mathrm{ht}^{-1}(0)$.

It follows from the known behavior of the canonical class under a blowup that

$$K_X = \sigma^*(K_S) + \sum_{i=1}^{N} \mathcal{E}_i. \tag{3.5}$$

The intersection theory on a nonsingular surface gives

$$\mathcal{H}'^2 = (\phi^*(\mathcal{H}'))^2 = \Big(\sigma^*(\mathcal{H}) - \sum_{i=1}^{N} m(x_i)\mathcal{E}_i\Big)^2 = \mathcal{H}^2 - \sum_{i=1}^{N} m(x_i)^2, \tag{3.6}$$

$$K_{S'} \cdot \mathcal{H}' = K_S \cdot \mathcal{H} + \sum_{i=1}^{N} m(x_i).$$

Example 3.2. Let $\chi : \mathbb{P}^2 - \to \mathbb{P}^2$ be a Cremona transformation, $\mathcal{H}' = |\ell|$ be the linear system of lines in $\mathbb{P}^2$, and $\mathcal{H} \subset |n\ell|$. The formulas (3.6) give

$$n^2 - \sum_{i=1}^{N} m(x_i)^2 = 1, \tag{3.7}$$

$$3n - \sum_{i=1}^{N} m(x_i) = 3.$$

The linear system $\mathcal{H}$ is written in this situation as $\mathcal{H} = |n\ell - \sum_{i=1}^{N} m_i x_i|$. For example, a quadratic transformation with three base points p_1, p_2, p_3 is given by the linear system $|2\ell - p_1 - p_2 - p_3|$. In the case of the standard quadratic transformation τ_1, the curves $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ are irreducible, the map $\sigma_1 : X_1 \to X_0 = \mathbb{P}^2$ is an isomorphism in a neighborhood of p_2, p_3, and the map $\sigma_2 : X_2 \to X_1$ is an isomorphism in a neighborhood of $\sigma^{-1}(p_3)$. This shows that we can identify p_1, p_2, p_3 with points on $\mathbb{P}^2$. In the case of the transformation (2.2), we have $\sigma_1(p_2) = p_1$ and p_3 can be identified with a point on $\mathbb{P}^2$. So in this case $p_2 \succ p_1$. For the transformation (2.3) we have $p_3 \succ p_2 \succ p_1$.

For a Geiser involution (resp. a Bertini involution) we have $\mathcal{H} = |8\ell - 3p_1 - \cdots - 3p_7|$ (respectively $\mathcal{H} = |17\ell - 6p_1 - \cdots - 6p_8|$).

3.2 *G*-surfaces

Let G be a finite group. A G-surface is a pair (S, ρ), where S is a nonsingular projective surface and ρ is an isomorphism from G to a group of automorphisms of S. A morphism of the pairs $(S, \rho) \to (S', \rho')$ is defined to be a morphism of surfaces $f : S \to S'$ such that $\rho'(G') = f \circ \rho(G) \circ f^{-1}$. In particular, two subgroups of $\mathrm{Aut}(S)$ define isomorphic G-surfaces if and only if they are conjugate inside of $\mathrm{Aut}(S)$. Often, if no confusion arises, we will denote a G-surface by (S, G).

Let $\chi : S \to S'$ be a birational map of G-surfaces. Then one can G-equivariantly resolve χ, in the sense that one can find diagram (3.1) where all maps are morphisms of G-surfaces. The group G acts on the surface X permuting the exceptional configurations $\mathcal{E}_i$ in such a way that $\mathcal{E}_i \subset \mathcal{E}_j$ implies $g(\mathcal{E}_i) \subset g(\mathcal{E}_j)$. This defines an action of G on the set of indeterminacy points of χ ($g(x_i) = x_j$ if $g(\mathcal{E}_i) = g(\mathcal{E}_j)$). The action preserves the order, i.e. $x_i \succ x_j$ implies $g(x_i) \succ g(x_j)$, so the function $\mathrm{ht} : \{x_1, \ldots, x_N\} \to \mathbb{N}$ is constant on each orbit Gx_i.

Let $\mathcal{H}' = |H'|$ be an ample linear system on S' and

$$\phi^*(\mathcal{H}') = \sigma^*(\mathcal{H}) - \sum_{i=1}^{N} m(x_i)\mathcal{E}_i$$

be its inverse transform on X as above. Everything here is G-invariant, so $\mathcal{H}$ is a G-invariant linear system on S and the multiplicities $m(x_i)$ are constant on the G-orbits. So we can rewrite the system in the form

$$\phi^*(\mathcal{H}') = \sigma^*(\mathcal{H}) - \sum_{\kappa \in \mathcal{I}} m(\kappa)\mathcal{E}_\kappa,$$

where $\mathcal{I}$ is the set of G-orbits of indeterminacy points. For any $\kappa \in \mathcal{I}$ we set $m(\kappa) = m(x_i)$, where $x_i \in \kappa$ and $\mathcal{E}_\kappa, = \sum_{x_i \in \kappa} \mathcal{E}_i$. Similarly one can rewrite the proper transform of $\mathcal{H}'$ on S:

$$|H - \sum_{\kappa \in \mathcal{I}} m(\kappa)\kappa|. \tag{3.8}$$

Now we can rewrite the intersection formula (3.6) in the form

$$H'^2 = H^2 - \sum_{\kappa \in \mathcal{I}} m(\kappa)^2 d(\kappa), \tag{3.9}$$
$$K_{S'} \circ H' = K_S \cdot H + \sum_{\kappa \in \mathcal{I}} m(\kappa) d(\kappa),$$

where $d(\kappa) = \#\{i : x_i \in \kappa\}$.

Remark 3.3. In the arithmetical analog of the previous theory all the notation become more natural. Our maps are maps over a perfect ground field k. A blowup is the blowup of a closed point in the scheme-theoretic sense, not necessary k-rational. An exceptional curve is defined over k but when we replace k with an algebraic closure $\bar{k}$, it splits into the union of conjugate exceptional curves over $\bar{k}$. So, in the above notation, κ means a closed point on S or an infinitely near point. The analog of $d(\kappa)$ is of course the degree of a point, i.e., the extension degree $[k(x) : k]$, where $k(x)$ is the residue field of x.

3.3 The *G*-equivariant bubble space

Here we recall Manin's formalism of the theory of linear systems with base conditions in its G-equivariant form (see [46]).

First we define the *G-equivariant bubble space* of a G-surface S as a G-equivariant version $(S,G)^{\mathrm{bb}}$ of Definition 3.1. One replaces the category $\mathcal{B}(S)$ of birational morphisms $S' \to S$ with the category $\mathcal{B}(S,G)$ of birational morphisms of G-surfaces. In this way the group G acts on the bubble space $(S,G)^{\mathrm{bb}}$. Let

$$Z^*(S,G) = \varinjlim \mathrm{Pic}(S'), \tag{3.10}$$

where the inductive limit is taken with respect to the functor Pic from the category $\mathcal{B}(S,G)$ with values in the category of abelian groups defined by $S' \to \mathrm{Pic}(S')$. The group $Z^*(S,G)$ is equipped with a natural structure of G-module. Also it is equipped with the following natural structures.

(a) A symmetric G-invariant pairing

$$Z^*(S,G) \times Z^*(S,G) \to \mathbb{Z}$$

induced by the intersection pairing on each $\mathrm{Pic}(S')$.

(b) A distinguished cone of effective divisors classes in $Z^*(S,G)$

$$Z^*_+(S,G) = \varinjlim \mathrm{Pic}_+(S'),$$

where $\mathrm{Pic}_+(S')$ is the cone of effective divisor classes on each S' from $\mathcal{B}(S,G)$.

(c) A distinguished G-equivariant homomorphism

$$K : Z^*(S,G) \to \mathbb{Z}, \quad K(z) = K_{S'} \cdot z, \text{ for any } S' \to S \text{ from } \mathcal{B}(S,G).$$

Let $f : S' \to S$ be a morphism from $\mathcal{B}(S,G)$ and let $\mathcal{E}_1, \ldots, \mathcal{E}_n$ be its exceptional curves. We have a natural splitting

$$\mathrm{Pic}(S') = f^*(\mathrm{Pic}(S)) \oplus \mathbb{Z}[\mathcal{E}_1] \oplus \cdots \oplus \mathbb{Z}[\mathcal{E}_n].$$

Now let $Z_0(S,G) = \mathbb{Z}^{(S,G)^{\mathrm{bb}}}$ be the free abelian group generated by the set $(S,G)^{\mathrm{bb}}$. Identifying exceptional curves with points in the bubble space, and passing to the limit we obtain a natural splitting

$$Z^*(S,G) = Z_0(S,G) \oplus \mathrm{Pic}(S). \tag{3.11}$$

Passing to invariants we get the splitting

$$Z^*(S,G)^G = Z_0(S,G)^G \oplus \mathrm{Pic}(S)^G. \tag{3.12}$$

Write an element of $Z^*(S,G)^G$ in the form

$$z = D - \sum_{\kappa \in O} m(\kappa)\kappa,$$

where O is the set of G-orbits in $Z_0(S,G)^G$ and D is a G-invariant divisor class on S. Then

(a) $z \cdot z' = D \cdot D' - \sum_{\kappa \in O} m(\kappa)m'(\kappa)d(\kappa)$;
(b) $z \in Z_+^*(S,G)$ if and only if $D \in \mathrm{Pic}_+(S)^G, m(\kappa) \geq 0$, and $m(\kappa') \leq m(\kappa)$ if $\kappa' \succ \kappa$;
(c) $K(z) = D \cdot K_S + \sum_{\kappa \in O} m(\kappa)d(\kappa)$.

Let $\phi : S' \to S$ be an object of $\mathcal{B}(S,G)$. Then we have a natural map $\phi_{\mathrm{bb}} : (S',G)^{\mathrm{bb}} \to (S,G)^{\mathrm{bb}}$ that induces an isomorphism $\phi_{\mathrm{bb}}^* : Z(S,G) \to Z(S',G)$. We also have a natural isomorphism $\phi_*^{\mathrm{bb}} : Z(S',G) \to Z(S,G)$. None of these maps preserves the splitting (3.11). Resolving indeterminacy points of any birational map $\chi : (S,G) - \to (S',G')$ we can define

- *proper direct transform map* $\chi_* : Z^*(S,G) \xrightarrow{\sim} Z^*(S',G)$;
- *proper inverse transform map* $\chi^* : Z^*(S',G) \xrightarrow{\sim} Z^*(S,G)$.

The group $Z^*(S,G)$ equipped with all above structures is one of the main G-birational invariants of S. It can be viewed as the Picard group of the bubble space $(S,G)^{\mathrm{bb}}$.

The previous machinery gives a convenient way to consider the linear systems defining rational maps of surfaces. Thus we can rewrite (3.4) in the form $|z|$, where $z = H - \sum m_i x_i$ is considered as an element of $Z_+^*(S,G)$. The condition that $|z|$ be homaloidal becomes equivalent to the conditions

$$\begin{aligned} z^2 = H^2 - \sum m_i^2 &= H'^2, \\ K(z) = H \cdot K_S + \sum m_i &= H' \cdot K_{S'}. \end{aligned} \tag{3.13}$$

When $S = S' = \mathbb{P}^2$ we get equalities (3.7).

3.4 Minimal rational G-surfaces

Let (S,ρ) be a rational G-surface. Choose a birational map $\phi : S - \rightarrow \mathbb{P}^2$. For any $g \in G$, the map $\phi \circ g \circ \phi^{-1}$ belongs to $\mathrm{Cr}(2)$. This defines an injective homomorphism

$$\iota_\phi : G \rightarrow \mathrm{Cr}(2). \tag{3.14}$$

Suppose (S',ρ') is another rational G-surface and $\phi' : S' - \rightarrow \mathbb{P}^2$ is a birational map.

The following lemma is obvious.

Lemma 3.4. *The subgroups $\iota_\phi(G)$ and $\iota_{\phi'}(G)$ of* $\mathrm{Cr}(2)$ *are conjugate if and only if there exists a birational map of G-surfaces $\chi : S' - \rightarrow S$.*

The lemma shows that a birational isomorphism class of G-surfaces defines a conjugacy class of subgroups of $\mathrm{Cr}(2)$ isomorphic to G. The next lemma shows that any conjugacy class is obtained in this way.

Lemma 3.5. *Suppose G is a finite subgroup of* $\mathrm{Cr}(2)$*, then there exists a rational G-surface (S,ρ) and a birational map $\phi : S \rightarrow \mathbb{P}^2$ such that*

$$G = \phi \circ \rho(G) \circ \phi^{-1}.$$

Proof. We give two proofs. The first one is after A. Verra. Let $D = \cap_{g\in G}\mathrm{dom}(g)$, where $\mathrm{dom}(g)$ is an open subset on which g is defined. Then $U = \cap_{g\in G} g(D)$ is an open invariant subset of $\mathbb{P}^2$ on which $g \in G$ acts biregularly. Order G in some way and consider a copy of $\mathbb{P}^2_g$ indexed by $g \in G$. For any $u \in U$ let $g(u) \in \mathbb{P}^2_g$. We define a morphism

$$\phi : U \rightarrow \prod_{g\in G} \mathbb{P}^2_g,\ u \mapsto (g(u))_{g\in G}.$$

Define an action of G on $\phi(U)$ by $g'((x_g)_{g\in G}) = (x_{gg'})_{g\in G}$. Then ϕ is obviously G-equivariant. Now define V as the Zariski closure of $\phi(U)$ in the product. It is obviously a G-invariant surface which contains an open G-invariant subset G-isomorphic to U. It remains to replace V by its G-equivariant resolution of singularities (which always exists).

The second proof is standard. Let U be as above and $U' = U/G$ be the orbit space. It is a normal algebraic surface. Choose any normal projective completion X' of U'. Let S' be the normalization of X' in the field of rational functions of U. This is a normal projective surface on which G acts by biregular transformations. It remains to define S to be a G-invariant resolution of singularities (see also [24]).

Summing up, we obtain the following result.

Theorem 3.6. *There is a natural bijective correspondence between birational isomorphism classes of rational G-surfaces and conjugate classes of subgroups of* $\mathrm{Cr}(2)$ *isomorphic to G.*

So our goal is to classify G-surfaces (S,ρ) up to birational isomorphism of G-surfaces.

Definition 3.7. *A* minimal *G-surface is a G-surface* (S,ρ) *such that any birational morphism of G-surfaces* $(S,\rho)\to(S',\rho')$ *is an isomorphism. A group G of automorphisms of a rational surface S is called a minimal group of automorphisms if the pair* (S,ρ) *is minimal.*

Obviously, it is enough to classify minimal rational G-surfaces up to birational isomorphism of G-surfaces.

Before we state the next fundamental result, let us recall some terminology.

A *conic bundle structure* on a rational G-surface (S,G) is a G-equivariant morphism $\phi: S\to\mathbb{P}^1$ such that the fibres are isomorphic to a reduced conic in $\mathbb{P}^2$. A *Del Pezzo surface* is a surface with ample anticanonical divisor $-K_S$.

Theorem 3.8. *Let S be a minimal rational G-surface. Then either S admits a structure of a conic bundle with* $\mathrm{Pic}(S)^G\cong\mathbb{Z}^2$*, or S is isomorphic to a Del Pezzo surface with* $\mathrm{Pic}(S)^G\cong\mathbb{Z}$.

An analogous result from the classical literature is proven by using the method of the termination of adjoints, first introduced for linear system of plane curves in the work of G. Castelnuovo. It consists in replacing a linear system $|D|$ with the linear system $|D+K_S|$ and repeating this, stopping only if the next step leads to the empty linear system. The application of this method to finding a G-invariant linear system of curves in the plane was initiated in the works of S. Kantor [42], who essentially stated the theorem above but without the concept of minimality. In arithmetical situation this method was first applied by F. Enriques [28]. A first modern proof of the theorem was given by Yu. Manin [45] and by the second author [38] (an earlier proof of Manin used the assumption that G is an abelian group). Nowadays the theorem follows easily from a G-equivariant version of Mori theory (see [43, Example 2.18]), and the proof can be found in literature ([8], [23]). For this reason we omit the proof.

Recall the classification of Del Pezzo surfaces (see [25], [46]). The number $d=K_S^2$ is called the *degree*. By Noether's formula, $1\le d\le 9$. For $d\ge 3$, the anticanonical linear system $|-K_S|$ maps S onto a nonsingular surface of degree d in $\mathbb{P}^d$. If $d=9$, $S\cong\mathbb{P}^2$. If $d=8$, then $S\cong\mathbb{P}^1\times\mathbb{P}^1$, or $S\cong\mathbf{F}_1$, where as always we denote by $\mathbf{F}_n$ the minimal ruled surface $\mathbb{P}(\mathcal{O}_{\mathbb{P}^1}\oplus\mathcal{O}_{\mathbb{P}^1}(n))$. For $d\le 7$, a Del Pezzo surface S is isomorphic to the blowup of $n=9-d$ points in $\mathbb{P}^2$ in general position that means that

- no three are on a line;
- no six are on a conic;
- if $n=8$, then then the points are not on a plane cubic which has one of them as its singular point.

For $d=2$, the linear system $|-K_S|$ defines a finite morphism of degree 2 from S to $\mathbb{P}^2$ with a nonsingular quartic as the branch curve. Finally, for $d=1$, the linear system $|-2K_S|$ defines a finite morphism of degree 2 onto a quadric cone $Q\subset\mathbb{P}^3$ with the branch curve cut out by a cubic.

For a minimal Del Pezzo G-surface the group $\mathrm{Pic}(S)^G$ is generated by K_S if S is not isomorphic to $\mathbb{P}^2$ or $\mathbb{P}^1 \times \mathbb{P}^1$. In the latter cases it is generated by $\frac{1}{3}K_S$ or $\frac{1}{2}K_S$, respectively.

A conic bundle surface is either isomorphic to $\mathbf{F}_n$ or to a surface obtained from $\mathbf{F}_n$ by blowing up a finite set of points, no two lying in a fibre of a ruling. The number of blowups is equal to the number of singular fibres of the conic bundle fibration. We will exclude the surfaces $\mathbf{F}_0$ and $\mathbf{F}_1$, considering them as Del Pezzo surfaces.

There are minimal conic bundles with ample $-K_S$ (see Proposition 5.2).

4 Automorphisms of minimal ruled surfaces

4.1 Some of group theory

We employ the standard notations for groups used by group-theorists (see [19]):

- C_n, a cyclic group of order n;
- $n = C_n$ if no confusion arises;
- $n^r = C_n^r$, the direct sum of r copies of C_n (not to be confused with cyclic group of order n^r);
- S_n, the *permutation group* of degree n;
- A_n, the *alternating group* of degree n;
- D_{2n}, the *dihedral group* of order $2n$;
- $Q_{4n} = \langle a, b \mid a^{2n} = 1, b^2 = a^n, b^{-1}ab = a^{-1}\rangle$, *dicyclic group* of order $4n$, a *generalized quaternion group* if $n = 2^k$;
- $H_n(p)$, the *Heisenberg group* of unipotent $n \times n$-matrices with entries in $\mathbb{F}_p$;
- $\mathrm{GL}(n) = \mathrm{GL}(n, \mathbb{C})$, general linear group over $\mathbb{C}$;
- $\mathrm{SL}(n) = \mathrm{SL}(n, \mathbb{C})$, special linear group over $\mathbb{C}$;
- $\mathrm{PGL}(n) = \mathrm{GL}(n, \mathbb{C})/\mathbb{C}^*$, general projective linear group over $\mathbb{C}$;
- $\mathrm{O}(n)$, the orthogonal linear group over $\mathbb{C}$;
- $\mathrm{PO}(n)$, the projective orthogonal linear group over $\mathbb{C}$,
- $L_n(q) = \mathrm{PSL}(n, \mathbb{F}_q)$, where $q = p^r$ is a power of a prime number p;
- $T \cong A_4$, $O \cong S_4 \cong \mathrm{PGL}(2, \mathbb{F}_3)$, $I \cong A_5 \cong L_2(5) \cong L_2(2^2)$, tetrahedral, octahedral, icosahedral subgroups of $\mathrm{PGL}(2)$;
- $\overline{T} \cong \mathrm{SL}(2, \mathbb{F}_3)$, $\bar{O} \cong \mathrm{GL}(2, \mathbb{F}_3)$, $\bar{I} \cong \mathrm{SL}(2, \mathbb{F}_5), \overline{D}_{2n} \cong Q_{4n}$, binary tetrahedral, binary octahedral, binary icosahedral, binary dihedral subgroups of $\mathrm{SL}(2)$;
- $A \bullet B$ is an upward extension of B with help of a normal subgroup A;
- $A : B$ is a split extension $A \bullet B$, i.e., a semidirect product $A \rtimes B$ (it is defined by a homomorphism $\phi : B \to \mathrm{Aut}(A)$);
- $A^\bullet B$ is a nonsplit extension $A \bullet B$;
- $nA = n^\bullet A$, where the normal group n is equal to the center;
- $p^{a+b} = C_p^a \bullet C_p^b$, where p is prime;

- $A \triangle B$ (or $A \triangle_D B$), the diagonal product of A and B over their common homomorphic image D (i.e., the subgroup of $A \times B$ of pairs (a, b) such that $\alpha(a) = \beta(b)$ for some surjections $\alpha : A \to D, \beta : B \to D$); When D is omitted it means that D is the largest possible;
- $\frac{1}{m}[A \times B] = A \triangle_D B$, where $\#D = m$;
- $A \wr S_n$, the wreath product, i.e., $A^n : S_n$, where S_n is the symmetric group acting on A^n by permuting the factors;
- μ_n, the group of nth roots of unity with generator $\epsilon_n = e^{2\pi i/n}$.

We will often use the following simple result from group theory which is known as Goursat's Lemma.

Lemma 4.1. *Let G be a finite subgroup of the product $A \times B$ of two groups A and B. Let $p_1 : A \times B \to A, p_2 : A \times B \to B$ be the projection homomorphisms. Let $G_i = p_i(G), H_i = \mathrm{Ker}(p_j|G), i \neq j = 1, 2$. Then H_i is a normal subgroup in G_i. The map $\phi : G_1/H_1 \to G_2/H_2$ defined by $\phi(aH_1) = p_2(a)H_2$ is an isomorphism, and*

$$G = G_1 \triangle_D G_2,$$

where $D = G_1/H_1, \alpha : G_1 \to D$, is the projection map to the quotient, and β is the composition of the projection $G_2 \to G_2/H_2$ and ϕ^{-1}.

Note some special cases:

$$G \triangle_1 G' \cong G \times G', \quad G \triangle_{G'} G' = \{(g, \alpha(g)) \in G \times G', g \in G_1\},$$

where $\alpha : G \to G'$ is a surjection and $G' \to G'$ is the identity.

We will be dealing with various group extensions. The following lemma is known in group theory as the Schur-Zassenhaus Theorem. Its proof can be found in [31, 6.2].

Lemma 4.2. *Let $A \bullet B$ be an extension of groups. Suppose that the orders of A and B are coprime. Then the extension splits. If, moreover, A or B is solvable, then all subgroups of $A : B$ defining splittings are conjugate.*

We will often use the following simple facts, their proofs are left to the reader (or can be found in www.planetmath.org).

Lemma 4.3. *A subgroup of $D_{2n} = \langle a, b \mid a^n = b^2 = b^{-1}aba = 1\rangle$ is either cyclic or dihedral. A normal subgroup H is either cyclic subgroup $\langle a\rangle$, or $n = 2k$ and H is one of the following two subgroups $\langle a^2, b\rangle, \langle a^2, ab\rangle$ of index 2. These two subgroups are interchanged under the outer automorphism $a \mapsto a, b \mapsto ab$. If H is cyclic of order n/k, the quotient group is isomorphic to $D_{2n/k}$.*

The group of $\mathrm{Aut}(D_{2n})$ is isomorphic to $(\mathbb{Z}/n)^* : n$ and it is generated by the transformations $a \mapsto a^s, b \mapsto a^t b$. The subgroup of inner automorphisms is generated by transformations $a \mapsto a^{-1}, b \mapsto b$ and $a \mapsto a, b \mapsto a^2 b$.

It will be convenient to list all isomorphism classes of nonabelian groups of order 16; see Table 1.

Table 1. Nonabelian groups of order 16.

Notation	Center	LCS	Extensions	Presentation
$2 \times D_8$	2^2	16, 2, 1	$2^{1+3}, 2^{2+2}, 2^{3+1}$, $(2 \times 4) : 2$	$a^4 = b^2 = c^2 = 1$, $[a, b]a^2 = [a, c] = [b, c] = 1$
$2 \times Q_8$	2^2	16, 2, 1	2^{2+2}, $(2 \times 4)^{\bullet}2$	$a^4 = a^2 b^{-2} = a^2[a, b] = 1$
D_{16}	2	16, 4, 2, 1	$8 : 2, 2D_8$, $(2^2)^{\bullet}4, D_8^{\bullet}2$	$a^8 = b^2 = a^2[a, b] = 1$
SD_{16}	2	16, 4, 2, 1	$8 : 2, D_8^{\bullet}2$, $2D_8, (2^2)^{\bullet}4$	$a^8 = b^2 = [a, b]a^{-2} = 1$
Q_{16}	2	16, 4, 2, 1	$8^{\bullet}2, 2D_8$, $4^{\bullet}(2^2)$	$a^8 = a^4 b^{-2} = [a, b]a^2 = 1$
AS_{16}	4	16, 2, 1	$2^{1+3}, D_8 : 2$ $4(2^2), (2 \times 4) : 2$	$a^4 = b^2 = c^2 = [a, b] = 1$ $[c, b]a^{-2} = [c, a] = 1$
K_{16}	2^2	16, 2, 1	$2^{2+2}, (2 \times 4)^{\bullet}2$ $2^{\bullet}(2 \times 4), 4 : 4$	$a^4 = b^4 = [a, b]a^2 = 1$
L_{16}	2^2	16, 2, 1	$2^2 : 4, 2^{\bullet}(2 \times 4)$ $(2 \times 4) : 2$	$a^4 = b^2 = c^2 = 1$, $[c, a]b = [a, b] = [c, b] = 1$
M_{16}	4	16, 2, 1	$8 : 2,\ 4(2^2)$ $2^{\bullet}(2 \times 4)$	$a^8 = b^2 = 1, [a, b]a^4 = 1$

Recall that there are two nonisomorphic nonabelian groups of order 8: D_8 and Q_8.

Finally, we describe the central extension of polyhedral and binary polyhedral groups. Recall that the isomorphism classes of central extensions $A{\bullet}G$, where A is an abelian group, are parametrized by the 2-cohomology group $H^2(G, A)$. We will assume that $A \cong p$, where p is prime. We will use the following facts about the cohomology groups of polyhedral and binary polyhedral groups, which can be found in textbooks on group cohomology (see, for example, [1]).

Lemma 4.4. *Let G be a polyhedral group or a binary polyhedral group. If $G \cong n$ is cyclic, then $H^2(G, p) \cong p$ if $p \mid n$ and zero otherwise. If G is not cyclic, then $H^2(G, p) = 0$ if $p \neq 2, 3$. Moreover,*

(i) If G is a polyhedral group, then

$$H^2(G, 2) \cong \begin{cases} 2 & \text{if } G \cong D_{2n}, n \text{ odd}, \\ 2^3 & \text{if } G \cong D_{2n}, n \text{ even}, \\ 2 & \text{if } G \cong T, \\ 2^2 & \text{if } G \cong O, \\ 2 & \text{if } G \cong I. \end{cases}$$

$$H^2(G,3) \cong \begin{cases} 3 & \textit{if } G \cong T, \\ 1 & \textit{otherwise.} \end{cases}$$

(ii) If G is a binary polyhedral group, then

$$H^2(G,2) \cong \begin{cases} 2 & \textit{if } G \cong \overline{D}_{2n}, n \textit{ odd}, \\ 2^2 & \textit{if } G \cong \overline{D}_{2n}, n \textit{ even}, \\ 2 & \textit{if } G \cong \overline{O}, \\ 1 & \textit{otherwise.} \end{cases}$$

$$H^2(G,3) \cong \begin{cases} 3 & \textit{if } G \cong \overline{T}, \\ 1 & \textit{otherwise.} \end{cases}$$

4.2 Finite groups of projective automorphisms

We start with the case $S = \mathbb{P}^2$, where $\mathrm{Aut}(S) \cong \mathrm{PGL}(3)$. To save space we will often denote a projective transformation

$$(x_0, x_1, x_2) \mapsto (L_0(x_0, x_1, x_2), L_1(x_0, x_1, x_2), L_2(x_0, x_1, x_2))$$

by $[L_0(x_0, x_1, x_2), L_1(x_0, x_1, x_2), L_2(x_0, x_1, x_2)]$.

Recall some standard terminology from the theory of linear groups. Let G be a subgroup of the general linear group $\mathrm{GL}(V)$ of a complex vector space V. The group G is called *intransitive* if the representation of G in V is reducible. Otherwise it is called *transitive.* A transitive group G is called *imprimitive* if it contains an intransitive normal subgroup G'. In this case V decomposes into a direct sum of G'-invariant proper subspaces, and elements from G permute them. A group is *primitive* if it is neither intransitive, nor imprimitive. We reserve this terminology for subgroups of $\mathrm{PGL}(V)$, keeping in mind that each such group can be represented by a subgroup of $\mathrm{GL}(V)$.

Let G' be a finite intransitive subgroup of $\mathrm{GL}(3)$ and G be its image in $\mathrm{PGL}(3)$. Then G' is conjugate to a subgroup $\mathbb{C}^* \times \mathrm{GL}(2)$ of block matrices.

To classify such subgroups we have to classify subgroups of $\mathrm{GL}(2)$. We will use the well-known classification of finite subgroups of $\mathrm{PGL}(2)$. They are isomorphic to one of the following *polyhedral groups*:

- a cyclic group C_n;
- a dihedral group D_{2n} of order $2n \geq 2$;
- the tetrahedral group $T \cong A_4$ of order 12;
- the octahedral group $O \cong S_4$ of order 24;
- the icosahedral group $I \cong A_5$ of order 60.

Two isomorphic subgroups are conjugate subgroups of $\mathrm{PGL}(2)$.

The pre-image of such group in $\mathrm{SL}(2, \mathbb{C})$ under the natural map

$$\mathrm{SL}(2) \to \mathrm{PSL}(2) = \mathrm{SL}(2)/(\pm 1) \cong \mathrm{PGL}(2)$$

is a double extension $\overline{G} = 2\bullet G$. The group $\overline{G} = 2\bullet G$ is called a *binary polyhedral group*. A cyclic group of odd order is isomorphic to a subgroup SL(2) intersecting trivially the center.

Consider the natural surjective homomorphism of groups

$$\beta : \mathbb{C}^* \times \mathrm{SL}(2) \to \mathrm{GL}(2), \quad (c, A) \mapsto cA.$$

Its kernel is the subgroup $\{(1, I_2), (-1, -I_2)\}$.

Let G be a finite subgroup of GL(2) with center $Z(G)$. Since $cA = (-c)(-A)$ and $\det(cA) = c^2 \det A$, we see that $\widetilde{G} = \beta^{-1}(G)$ is a subgroup of $\mu_{2m} \times \bar{G}'$, where $\overline{G'}$ is a binary polyhedral group with nontrivial center whose image G' in PGL(2) is isomorphic to $G/Z(G)$. The homomorphism β defines an isomorphism from the kernel H_2 of the first projection $\widetilde{G} \to \mu_{2m}$ onto the subgroup $G_0 = \mathrm{Ker}(\det : G \to \mathbb{C}^*)$. Also it defines an isomorphism from the kernel H_1 of the second projection $\widetilde{G} \to \bar{G}'$ onto $Z(G)$. Applying Lemma 4.1, we obtain

$$\widetilde{G} \cong \mu_{2m} \triangle_D \overline{G}', \ D = \bar{G}'/G_0.$$

Lemma 4.5. *Let G be a finite non-abelian subgroup of* GL(2)*. Then $G = \beta(\widetilde{G})$, where $\widetilde{G} \subset \mathbb{C}^* \times \mathrm{SL}(2, \mathbb{C})$ is conjugate to one of the following groups.*

(i) $\widetilde{G} = \mu_{2m} \times \bar{I}, \ G \cong m \times \bar{I}$;
(ii) $\widetilde{G} = \mu_{2m} \times \overline{O}, \ G \cong m \times \overline{O}$;
(iii) $\widetilde{G} = \mu_{2m} \times \overline{T}, \ G \cong m \times \overline{T}$;
(iv) $\widetilde{G} = \mu_{2m} \times Q_{4n}, \ G \cong m \times Q_{4n}$;
(v) $\widetilde{G} = \frac{1}{2}[\mu_{4m} \times \overline{O}], \ G \cong 2m\bullet O \cong (m \times \overline{T})\bullet 2$ *(split if* $m = 1, 2$*)*;
(vi) $\widetilde{G} = \frac{1}{3}[\mu_{6m} \times \overline{T}], \ G \cong 2m\bullet T \cong (m \times 2^2)\bullet 3$ *(split if* $m = 1, 3$*)*;
(vii) $\widetilde{G} \cong \frac{1}{2}[\mu_{4m} \times Q_{8n}], G \cong 2m\bullet D_{4n} \cong (m \times Q_{4n})\bullet 2$ *(split if* $m = 1, 2$*)*;
(viii) $\widetilde{G} = \frac{1}{2}[\mu_{4m} \times Q_{4n}], \ G \cong 2m\bullet D_{2n} \cong (m \times 2n)\bullet 2$ *(split if* $m = 1, 2$*)*;
(ix) $\widetilde{G} = \frac{1}{4}[\mu_{4m} \times Q_{4n}], n$ *is odd,* $G \cong m\bullet D_{2n} \cong (m \times n)\bullet 2$ *(split if* $m = 1, 2$*)*.

Note that although Q_{8n} has two different non-cyclic subgroups of index 2, they are conjugate under an element of SL(2), so they lead to conjugate subgroups in GL(2).

Lemma 4.4 gives us some information when some of these extensions split.

An abelian subgroup $G \subset \mathrm{GL}(2)$ is conjugate to a subgroup of diagonal matrices of the form $(\epsilon_m^a, \epsilon_n^b)$, where ϵ_m, ϵ_n are primitive roots of unity and $a, b \in \mathbb{Z}$. Let $d = (m, n), m = du, n = dv, d = kq$ for some fixed positive integer k. Let $H_1 = \langle \epsilon_m^k \rangle \subset \langle \epsilon_m \rangle, H_2 = \langle \epsilon_n^k \rangle \subset \langle \epsilon_n \rangle$ be cyclic subgroups of index k. Applying Lemma 4.1 we obtain

$$G \cong \langle \epsilon_m \rangle \triangle_k \langle \epsilon_n \rangle,$$

where the homomorphisms $\langle\epsilon_m\rangle \to k, \langle\epsilon_n\rangle \to k$ differ by an automorphism of the cyclic group $\langle\epsilon_k\rangle \cong k$ defined by a choice of a new generator $\epsilon_m^s, (s,k) = 1$. In this case

$$G = (\langle\epsilon_m^k\rangle \times \langle\epsilon_n^k\rangle_\bullet \langle\epsilon_k\rangle \tag{4.1}$$

is of order $mn/k = uvkq^2$. In other words, G consists of diagonal matrices of the form $(\epsilon_m^a, \epsilon_n^b)$, where $a \equiv sb \mod k$.

Corollary 4.6. *Let G be an intransitive finite subgroup of* GL(3)*. Then its image in* PGL(3) *consists of transformations $[ax_0 + bx_1, cx_0 + dx_1, x_2]$, where the matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ form a non-abelian finite subgroup H of* GL(2) *from Lemma 4.5 or an abelian group of the form* (4.1)*.*

Now suppose G is transitive but imprimitive subgroup of PGL(3). Let G' be its largest intransitive normal subgroup. Then G/G' permutes transitively the invariant subspaces of G'; hence we may assume that all of them are one-dimensional. Replacing G by a conjugate group we may assume that G' is a subgroup of diagonal matrices. We can represent its elements by diagonal matrices $g = (\epsilon_m^a, \epsilon_n^b, 1)$, where $a \equiv sb \mod k$ as in (4.1). The group G contains a cyclic permutation τ of coordinates. Since G' is a normal subgroup of G, we get $\tau^{-1}g\tau = (\epsilon_n^{-b}, \epsilon_n^{-b}\epsilon_m^a, 1) \in G'$. This implies that $n|bm, m|an$, hence $u|b, v|a$. Since $(\epsilon_m, \epsilon_n^s, 1)$ or $(\epsilon_m^{s'}, \epsilon_n, 1), ss' \equiv 1 \mod k$, belongs to G we must have $u = v = 1$, i.e. $m = n = d$. Therefore G' consists of diagonal matrices $g = (\epsilon_d^a, \epsilon_d^{sa}, 1)$. Since $\tau^{-1}g\tau = (\epsilon_d^{-sa}, \epsilon_d^{a-sa}, 1) \in G'$, we get $a - sa \equiv -s^2a \mod k$ for all $a \in \mathbb{Z}/m\mathbb{Z}$. Hence the integers s satisfy the congruence $s^2 - s + 1 \equiv 0 \mod k$. If, moreover, $G/G' \cong S_3$, then we have an additional condition $s^2 \equiv 1 \mod k$, and hence either $k = 1$ and $G' = \mu_n \times \mu_n$ or $k = 3$, $s = 2$, and $G' = n \times n/k$.

This gives the following.

Theorem 4.7. *Let G be a transitive imprimitive finite subgroup of* PGL(3)*. Then G is conjugate to one of the following groups:*

- $G \cong n^2 : 3$ *generated by transformations*

$$[\epsilon_n x_0, x_1, x_2],\ [x_0, \epsilon_n x_1, x_2],\ [x_2, x_0, x_1];$$

- $G \cong n^2 : S_3$ *generated by transformations*

$$[\epsilon_n x_0, x_1, x_2],\ [x_0, \epsilon_n x_1, x_2],\ [x_0, x_2, x_1],\ [x_2, x_0, x_1];$$

- $G = G_{n,k,s} \cong (n \times \frac{n}{k}) : 3$, *where $k > 1, k|n$ and $s^2 - s + 1 = 0 \mod k$. It is generated by transformations*

$$[\epsilon_{n/k} x_0, x_1, x_2],\ [\epsilon_n^s x_0, \epsilon_n x_1, x_2],\ [x_2, x_0, x_1].$$

- $G \simeq (n \times \frac{n}{3}) : S_3$ *generated by transformations*

$$[\epsilon_{n/3} x_0, x_1, x_2],\ [\epsilon_n^2 x_0, \epsilon_n x_1, x_2],\ [x_0, x_2, x_1],\ [x_1, x_0, x_2].$$

The next theorem is a well-known result of Blichfeldt [11].

Theorem 4.8. *Any primitive finite subgroup G of* $\mathrm{PGL}(3)$ *is conjugate to one of the following groups:*

1. *The icosahedral group A_5 isomorphic to $L_2(5)$. It leaves invariant a nonsingular conic.*
2. *The* Hessian group *of order 216 isomorphic to $3^2 : \overline{T}$. It is realized as the group of automorphisms of the Hesse pencil of cubics*
$$x^3 + y^3 + z^3 + txyz = 0.$$
3. *The* Klein group *of order 168 isomorphic to $L_2(7)$ (realized as the full group of automorphisms of the Klein quartic $x^3y + y^3z + z^3x = 0$).*
4. *The* Valentiner group *of order 360 isomorphic to A_6. It can be realized as the full group of automorphisms of the nonsingular plane sextic*
$$10x^3y^3 + 9zx^5 + y^5 - 45x^2y^2z^2 - 135xyz^4 + 27z^6 = 0.$$
5. *Subgroups of the Hessian group:*
   - $3^2 : 4$*;*
   - $3^2 : Q_8$*.*

4.3 Finite groups of automorphisms of $\mathbf{F_0}$

Since $\mathbf{F}_0$ is isomorphic to a nonsingular quadric in $\mathbb{P}^3$, the group $\mathrm{Aut}(\mathbf{F}_0)$ is isomorphic to the projective orthogonal group $\mathrm{PO}(4)$. The classification of finite subgroups of $O(4)$ is due to É. Goursat [32] (in the real case see a modern account in [20]). Goursat' Lemma 4.1 plays an important role in this classification.

Obviously,
$$\mathrm{Aut}(\mathbf{F}_0) \cong \mathrm{PGL}(2) \wr S_2.$$

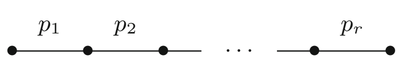
First we classify subgroups of $\mathrm{PGL}(2) \times \mathrm{PGL}(2)$ by applying Goursat's Lemma.

Observe the following special subgroups of $\mathrm{PGL}(2) \times \mathrm{PGL}(2)$:

1. $G = G_1 \times G_2$ is the *product subgroup.*
2. $G \triangle_1 G = \{(g_1, g_2) \in G \times G : \alpha(g_1) = g_2\} \cong G$ is a *α-twisted diagonal subgroup*. If $\alpha = \mathrm{id}_G$, we get the *diagonal subgroup.*

Note that α-twisted diagonal groups are conjugate in $\mathrm{Aut}(\mathbf{F}_0)$ if $\alpha(g) = xgx^{-1}$for some x in the normalizer of G inside $\mathrm{Aut}(\mathbb{P}^1)$. In particular, we may always assume that α is an exterior automorphism of G.

We will use the notation $[p_1, \ldots, p_r]$ for the Coxeter group defined by the Coxeter diagram



Following [20] we write $[p_1,\ldots,p_r]^+$ to denote the index 2 subgroup of even-length words in standard generators of the Coxeter group. If exactly one of the numbers $p_1,\ldots,p_r$ is even, say p_k, there are two other subgroups of index 2, denoted by $[p_1,\ldots,p_r^+]$ (respectively $[^+p_1,\ldots,p_r]$). They consist of words which contain each generator $R_1,\ldots,R_{k-1}$ (respectively $R_{k+1},\ldots,R_r$)an even number of times. The intersection of these two subgroups is denoted by $[^+p_1,\ldots,p_r^+]$. For example,

$$D_{2n}=[n],\ T=[3,3]^+,\ O=[3,4]^+,\ I=[3,5]^+.$$

Recall that each group $[p_1,\ldots,p_r]$ has a natural linear representation in $\mathbb{R}^r$ as a reflection group. If $r=3$, the corresponding representation defines a subgroup of $\mathrm{PO}(4)$. If $r=2$, it defines a subgroup of $\mathrm{PO}(3)$ that acts diagonally on $\mathbb{P}^2\times\mathbb{P}^2$ and on $\mathbb{P}^1\times\mathbb{P}^1$ embedded in $\mathbb{P}^2\times\mathbb{P}^2$ by the product of the Veronese maps. We denote by $\overline{[p_1,\ldots,p_r]}$ the quotient of $[p_1,\ldots,p_r]$ by its center. Similar notation is used for the even subgroups of $[p_1,\ldots,p_r]$.

Theorem 4.9. *Let G be a finite subgroup of $\mathrm{PGL}(2)\times\mathrm{PGL}(2)$ not conjugate to the product $A\times B$ of subgroups of $\mathrm{PGL}(2)$. Then G is conjugate to one of the following groups or its image under the switching of the factors:*

- $\frac{1}{60}[I\times I]\cong I\cong[3,5]^+$;
- $\frac{1}{60}[I\times I]\cong I\cong[3,3,3]^+$;
- $\frac{1}{24}[O\times O]\cong O\cong[3,4]^+$;
- $\frac{1}{24}[O\times O]\cong O\cong[2,3,3]^+$;
- $\frac{1}{12}[T\times T]\cong T\cong[3,3]^+$;
- $\frac{1}{2}[O\times O]\cong(T\times T):2\cong\overline{[3,4,3]^+}$;
- $\frac{1}{6}[O\times O]\cong 2^4:S_3\cong\overline{[3,3,4]^+}$;
- $\frac{1}{3}[T\times T]\cong 2^4:3\cong\overline{[^+3,3,4^+]}$;
- $\frac{1}{2}[D_{2m}\times D_{4n}]\cong(m\times D_{2n})^\bullet 2$ *($m,n\ge 2$)*;
- $\frac{1}{4}[D_{4m}\times D_{4n}]\cong(m\times n):4$ *(m,n odd)*;
- $\frac{1}{2k}[D_{2mk}\times D_{2nk}]_s\cong(m\times n):D_{2k},\ (s,k)=1$*)*;
- $\frac{1}{2k}[D_{2mk}\times D_{2nk}]_s\cong(m\times n):D_{2k},\ (s,2k)=1,m,n$ *odd)*;
- $\frac{1}{k}[C_{mk}\times C_{nk}]_s\cong(m\times n)\bullet k\ ((s,k)=1)$;
- $\frac{1}{k}[C_{mk}\times C_{nk}]_s\cong(m\times n)\bullet k\quad((s,2k)=1,m,n$ *odd;)*$)$;
- $\frac{1}{2}[D_{2m}\times O]\cong(m\times T):2$;
- $\frac{1}{2}[D_{4m}\times O]\cong(D_{2m}\times T):2(m\ge 2)$;

- $\frac{1}{6}[D_{6n} \times O] \cong (m \times 2^2) : S_3 (m \geq 2)$;
- $\frac{1}{2}[C_{2m} \times O] \cong (m \times T)\bullet 2$ *(split if* $m = 1$*)*;
- $\frac{1}{3}[C_{3m} \times T] \cong (m \times 2^2)\bullet 3$ *(split if* $m = 1$*)*;
- $\frac{1}{2}[D_{4m} \times D_{4n}] \cong (D_{2m} \times D_{2n})^{\bullet}2$ $(m, n \geq 2)$;
- $\frac{1}{2}[C_{2m} \times D_{4n}] \cong (m \times D_{2n})^{\bullet}2$ $(n \geq 2)$;
- $\frac{1}{2}[C_{2m} \times D_{2n}] \cong (m \times n) : 2 \cong m\bullet D_{2n}$.

All other finite subgroups of $\mathrm{Aut}(\mathbb{P}^1 \times \mathbb{P}^1)$ are conjugate to a group $G^0\bullet 2$, where the quotient 2 is represented by an automorphism that interchanges the two rulings of $\mathbf{F}_0$. It is equal to $\tau \circ g$, where τ is the switch $(x, y) \mapsto (y, x)$ and $g \in \mathrm{PGL}(2) \times \mathrm{PGL}(2)$.

4.4 Finite groups of automorphisms of $\mathbf{F}_n, n \neq 0$

Let S be a minimal ruled surface $\mathbf{F}_n, n \neq 0$. If $n = 1$, the group $\mathrm{Aut}(\mathbf{F}_1)$ is isomorphic to a subgroup of $\mathrm{Aut}(\mathbb{P}^2)$ leaving one point fixed. We will not be interested in such subgroups so we assume that $n \geq 2$.

Theorem 4.10. *Let* $S = \mathbf{F}_n, n \neq 0$. *We have*

$$\mathrm{Aut}(\mathbf{F}_n) \cong \mathbb{C}^{n+1} : (\mathrm{GL}(2)/\mu_n),$$

where $\mathrm{GL}(2)/\mu_n$ *acts on* $\mathbb{C}^{n+1}$ *by means of its natural linear representation in the space of binary forms of degree* n. *Moreover,*

$$\mathrm{GL}(2)/\mu_n \cong \begin{cases} \mathbb{C}^* : \mathrm{PSL}(2), & \text{if } n \text{ is even,} \\ \mathbb{C}^* : \mathrm{SL}(2), & \text{if } n \text{ is odd.} \end{cases}$$

Proof. This is of course well known. We identify $\mathbb{F}_n$ with the weighted projective plane $\mathbb{P}(1, 1, n)$. An automorphism is given by the formula

$$(t_0, t_1, t_2) \mapsto (at_0 + bt_1, ct_0 + dt_1, et_2 + f_n(t_0, t_1)),$$

where f_n is a homogeneous polynomial of degree n. The vector space $\mathbb{C}^{n+1}$ is identified with the normal subgroup of transformations $[t_0, t_1, t_2 + f_n(t_0, t_1)]$. The quotient by this subgroup is isomorphic to the subgroup of transformations $[at_0 + bt_1, ct_0 + dt_1, et_2]$ modulo transformations of the form $[\lambda t_0, \lambda t_1, \lambda^n t_2]$. This group is obviously isomorphic to $\mathrm{GL}(2)/\mu_n$. Consider the natural projection $\mathrm{GL}(2)/\mu_n \to \mathrm{PGL}(2) \cong \mathrm{PSL}(2)$. Define a homomorphism $\mathrm{SL}(2) \to \mathrm{GL}(2)/\mu_n$ by assigning to a matrix A the coset of A modulo μ_n. If n is even, the kernel of this homomorphism is $\langle -I_2 \rangle$, so we have a splitting $\mathrm{GL}(2)/\mu_n \cong \mathbb{C}^* : \mathrm{PGL}(2)$. If n is odd, the homomorphism is injective and defines a splitting $\mathrm{GL}(2)/\mu_n \cong \mathbb{C}^* : \mathrm{SL}(2)$. $\square$

Suppose G is a finite subgroup of $\mathrm{Aut}(\mathbb{F}_n)$. Obviously, G is contained in the subgroup $\mathrm{GL}(2)/\mu_n$.

Suppose $G \cap \mathbb{C}^* = \{1\}$. Then G is isomorphic to a subgroup of $\mathrm{PGL}(2)$ (resp. $\mathrm{SL}(2)$) over which the extension splits. Note that the kernel $\mathbb{C}^*$ of the projections $\mathrm{GL}(2)/\mu_n \to \mathrm{PGL}(2)$ or $\mathrm{GL}(2)/\mu_n \to \mathrm{SL}(2)$ is the center. Thus each finite subgroup H of $\mathrm{PGL}(2)$ (respectively $\mathrm{SL}(2)$) defines k conjugacy classes of subgroups isomorphic to H, where $k = \#\mathrm{Hom}(H, \mathbb{C}^*) = \#G/[G, G]$.

If $G \cap \mathbb{C}^* \cong \mu_m$ is non-trivial, the group is a central extension $m \bullet H$, where H is a polyhedral group, if n is even, and a binary polyhedral group otherwise. We can apply Lemma 4.4 to find some cases when the extension must split. In other cases the structure of the group is determined by using Theorem 4.9. We leave this to the reader.

5 Automorphisms of conic bundles

5.1 Geometry of conic bundles

Let $\phi : S \to \mathbb{P}^1$ be a conic bundle with singular fibres over points in a finite set $\Sigma \subset \mathbb{P}^1$. We assume that $k = \#\Sigma > 0$. Recall that each singular fibre $F_x, x \in \Sigma$, is the bouquet of two $\mathbb{P}^1$'s.

Let E be a section of the conic bundle fibration ϕ. The Picard group of S is freely generated by the divisor classes of E, the class F of a fibre, and the classes of k components of singular fibres, no two in the same fibre. The next lemma follows easily from the intersection theory on S.

Lemma 5.1. *Let E and E' be two sections with negative self-intersection $-n$. Let r be the number of components of singular fibres which intersect both E and E'. Then $k - r$ is even and*

$$2E \cdot E' = k - 2n - r.$$

In particular,

$$k \geq 2n + r.$$

Since a conic bundle S is isomorphic to a blowup of a minimal ruled surface, it always contains a section E with negative self-intersection $-n$. If $n \geq 2$, we obviously get $k \geq 4$. If $n = 1$, since (S, G) is minimal, there exists $g \in G$ such that $g(E) \neq E$ and $E \cap g(E) \neq \emptyset$. Applying the previous lemma we get

$$k \geq 4.$$

5.2 Exceptional conic bundles

We give three different constructions of the same conic bundle, which we will call an *exceptional conic bundle*.

First construction. Choose a ruling $p : \mathbf{F}_0 \to \mathbb{P}^1$ on $\mathbf{F}_0$ and fix two points on the base, say 0 and ∞. Let F_0 and F_∞ be the corresponding fibres. Take $g+1$ points $a_1, \ldots, a_{g+1}$ on F_0 and $g+1$ points $a_{g+2}, \ldots, a_{2g+2}$ on F_∞ such that no two lie in the same fibre of the second ruling $q : \mathbf{F}_0 \to \mathbb{P}^1$. Let $\sigma : S \to \mathbf{F}_0$ be the blowup of the points $a_1, \ldots, a_{2g+2}$. The composition $\pi = q \circ \sigma : S \to \mathbb{P}^1$ is a conic bundle with $2g+2$ singular fibres $R_i + R_i'$ over the points $x_i = q(a_i), i = 1, \ldots, 2g+2$. For $i = 1, \ldots, g+1$, $R_i = \sigma^{-1}(a_i)$, and R_{n+i} is the proper transform of the fibre $q^{-1}(a_i)$. Similarly, for $i = 1, \ldots, n$, R_i' is the proper transform of the fibre $q^{-1}(a_i)$, and $R'_{g+1+i} = \sigma^{-1}(a_{g+1+i})$.

Let E_0, E_∞ be the proper transforms of F_0, F_∞ on S. Each is a section of the conic bundle π. The section E_0 intersects $R_1, \ldots, R_{2g+2}$, and the section E_∞ intersects $R_1', \ldots, R_{2g+2}'$.

Let

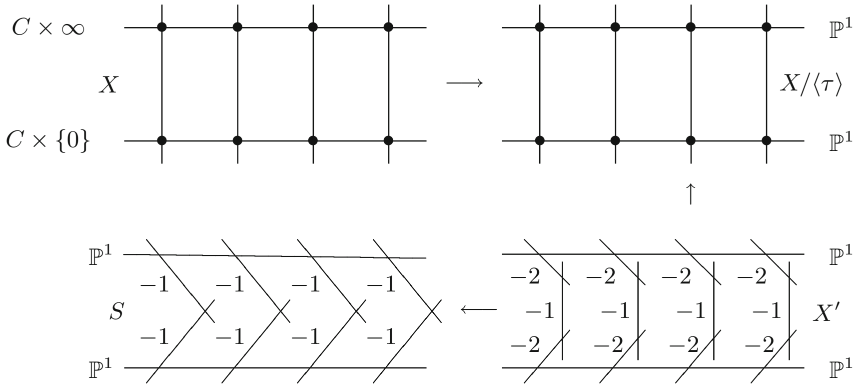
$$D_0 = 2E_0 + \sum_{i=1}^{2g+2} R_i, \quad D_\infty = 2E_\infty + \sum_{i=1}^{2g+2} R_i'.$$

It is easy to check that $D_0 \sim D_\infty$. Consider the pencil $\mathcal{P}$ spanned by the curves D_0 and D_∞. It has $2g+2$ simple base points $p_i = R_i \cap R_i'$. Its general member is a nonsingular curve C. In fact, a standard formula for computing the Euler characteristic of a fibred surface in terms of the Euler characteristics of fibres shows that all members except D_0 and D_∞ are nonsingular curves. Let F be a fibre of the conic bundle. Since $C \cdot F = 2$, the linear system $|F|$ cuts out a g_2^1 on C, so it is a hyperelliptic curve or the genus g of C is 0 or 1. The points p_i are obviously the ramification points of the g_2^1. Computing the genus of C we find that it is equal to g, thus $p_1, \ldots, p_{2g+2}$ is the set of ramification points. Obviously all nonsingular members are isomorphic curves. Let $\sigma : S' \to S$ be the blowup the base points $p_1, \ldots, p_{2g+2}$ and let $\overline{D}$ denote the proper transform of a curve on S. We have

$$2\bar{E}_0 + 2\bar{E}_\infty + \sum_{i=1}^{2g+2} (\bar{R}_i + \bar{R}_i' + 2\sigma^{-1}(p_i)) \sim 2\sigma^*(C).$$

This shows that there exists a double cover $X' \to S'$ branched along the divisor $\sum_{i=1}^{2g+2} (\bar{R}_i + \bar{R}_i')$. Since $\bar{R}_i^{\,2} = \bar{R}_i'^2 = -2$, the ramification divisor on X' consists of $4g+4$ (-1)-curves. Blowing them down we obtain a surface X isomorphic to the product $C \times \mathbb{P}^1$. This gives us the following.

Second construction. A pair (C, h) consisting of a nonsingular curve and an involution $h \in \mathrm{Aut}(C)$ with quotient $\mathbb{P}^1$ will be called a *hyperelliptic curve.* If C is of genus $g \geq 2$, then C is a hyperelliptic curve and h is its involution defined by the unique g_2^1 on C. Let δ be an involution of $\mathbb{P}^1$ defined by $(t_0, t_1) \mapsto (t_0, -t_1)$. Consider the involution $\tau = h \times \delta$ of the product $X = C \times \mathbb{P}^1$. Its fixed points are $4g+4$ points $c_i \times \{0\}$ and $c_i \times \{\infty\}$, where $X^{\langle h \rangle} = \{c_1, \ldots, c_{2g+2}\}$. Let X' be a minimal resolution of $X/(\tau)$. It is easy to see that the images of the curves $\{c_i\} \times \mathbb{P}^1$ are (-1)-curves on X'. Blowing them down we obtain our exceptional conic bundle.



Third construction. Let us consider a quasi-smooth hypersurface Y of degree $2g+2$ in weighted projective space $\mathbb{P} = \mathbb{P}(1,1,g+1,g+1)$ given by an equation

$$F_{2g+2}(T_0,T_1) + T_2T_3 = 0, \tag{5.1}$$

where $F_{2g+2}(T_0,T_1)$ is a homogeneous polynomial of degree $2g+2$ without multiple roots. The surface is a double cover of $\mathbb{P}(1,1,g+1)$ (the cone over a Veronese curve of degree $g+1$) branched over the curve $F_{2g+2}(T_0,T_1)+T_2^2 = 0$. The preimages of the singular point of $\mathbb{P}(1,1,g+1)$ with coordinates $(0,0,1)$ is a pair of singular points of Y with coordinates $(0,0,1,0)$ and $(0,0,0,1)$. The singularities are locally isomorphic to the singular points of a cone of the Veronese surface of degree $g+1$. Let S be a minimal resolution of Y. The preimages of the singular points are disjoint smooth rational curves E and E' with self-intersection $-(g+1)$. The projection $\mathbb{P}(1,1,g+1,g+1) \to \mathbb{P}^1, (t_0,t_1,t_2,t_3) \mapsto (t_0,t_1)$ lifts to a conic bundle on S with sections E, E'. The pencil $\lambda T_2 + \mu T_3 = 0$ cuts out a pencil of curves on Y which lifts to a pencil of bisections of the conic bundle S with $2g+2$ base points $(t_0,t_1,0,0)$, where $F_{2g+2}(t_0,t_1) = 0$.

It is easy to see that this is a general example of an exceptional conic bundle. In Construction 2, we blow down the sections E_0, E_∞ to singular points. Then consider an involution g_0 of the surface which is a descent of the automorphism of the product $C \times \mathbb{P}^1$ given by $\mathrm{id}_C \times \psi$, where $\psi : (t_0,t_1) \mapsto (t_1,t_0)$. The quotient by (g_0) gives $\mathbb{P}(1,1,g+1)$ and the ramification divisor is the image on S of the curve $C \times (1,1)$ or $C \times (1,-1)$. On one of these curves g_0 acts identically, on the other one it acts as the involution defined by the g_2^1.

Proposition 5.2. *Let $\phi : S \to \mathbb{P}^1$ be a minimal conic G-bundle with $k \leq 5$ singular fibres. Then S is a Del Pezzo surface, unless $k = 4$ and S is an exceptional conic bundle.*

Proof. Since $k \leq 5$, we have $K_S^2 = 8 - k \geq 3$. By Riemann-Roch, $|-K_S| \neq \emptyset$. Suppose S is not a Del Pezzo surface. Then there exists an irreducible curve C such that $-K_S \cdot C \leq 0$. Suppose, that equality holds. By Hodge's Index

Theorem, $C^2 < 0$, and by the adjunction formula, $C^2 = -2$ and $C \cong \mathbb{P}^1$. If strict inequality holds, then C is a component of a divisor $D \in |-K_S|$; hence $|-K_S - C| \neq \emptyset$ and $|K_S + C| = \emptyset$. Moreover, since $K_S^2 > 0$, we have $C \notin |-K_S|$. Applying Riemann-Roch to the divisor $K_S + C$, we easily obtain that C is of arithmetic genus 0, and hence $C \cong \mathbb{P}^1$. By adjunction, $C^2 \leq -2$. In both cases we have a smooth rational curve with $C^2 \leq -2$.

If $k = 4$ and S is an exceptional conic bundle, then S is not a Del Pezzo surface since it has sections with self-intersection -2. Assume this is not the case. Let C be the union of smooth rational curves with self-intersection < -2. It is obviously a G-invariant curve, so we can write $C \sim -aK_S - bf$, where f is the divisor class of a fibre of ϕ. Intersecting with f, we get $a > 0$. Intersecting with K_S, we get $2b > ad$, where $d = 8 - k \geq 3$. It follows from Lemma 5.1 that S contains a section E with self-intersection -2 or -1. Intersecting C with E, we get $0 \leq C \cdot E = a(-K_S \cdot E) - b \leq a - b$. This contradicts the previous inequality. Now let us take C to be the union of (-2)-curves. Similarly, we get $2b = ad$ and $C^2 = -aK_S \cdot C - bC \cdot f = -bC \cdot f = -2ab$. Let r be the number of irreducible components of C. We have $2a = C \cdot f \geq r$ and $-2r \leq C^2 = -2ab \leq -br$. If $b = 2$, we have the equality everywhere, hence C consists of $r = 2a$ disjoint sections, and $8 = rd$. Since $d \geq 3$, the only solution is $d = 4, r = 2$, and this leads to the exceptional conic bundle. Assume $b = 1$. Since $C^2 = -2a$ is even, a is a positive integer, and we get $2 = ad$. Since $d \geq 3$, this is impossible.

5.3 Automorphisms of an exceptional conic bundle

Let us describe the automorphism group of an exceptional conic bundle. The easiest way to do it to use Construction 3. We denote by Y_g an exceptional conic bundle given by equation (5.1). Since we are interested only in minimal groups we assume that $g \geq 1$.

Since $K_{Y_g} = \mathcal{O}_{\mathbb{P}}(-2)$, any automorphism σ of Y_g is a restriction of an automorphism of $\mathbb{P}$. Let G_1 be the subgroup of $\mathrm{SL}(2)$ of transformations preserving the zero divisor of $F_{2g+2}(T_0, T_1)$ and $\chi_1 : G_1 \to \mathbb{C}^*$ be the multiplicative character of G_1 defined by $\sigma_1^*(F_{2g+2}) = \chi_1(\sigma_1)F_{2g+2}$. Similarly, let G_2 be the subgroup of $\mathrm{GL}(2)$ of matrices preserving the zeroes of T_2T_3 and let $\chi_2 : G_2 \to \mathbb{C}^*$ be the character defined by $\sigma^*(T_2T_3) = \chi_2(\sigma_2)T_2T_3$. Let

$$(G_1 \times G_2)^0 = \{(\sigma_1, \sigma_2) \in G_1 \times G_2 : \chi_1(\sigma_1) = \chi_2(\sigma_2)\}.$$

In the notation of the diagonal products,

$$(G_1 \times G_2)^0 = \frac{1}{m}[G_1 \times G_2],$$

where $\chi_1(G_1) = \mu_m \subset \mathbb{C}^*$. The subgroup

$$K = \langle(-I_2, (-1)^{g+1}I_2)\rangle$$

acts identically on Y_g and the quotient group is isomorphic to $\mathrm{Aut}(Y_g)$.

Let $\mathrm{Aut}(Y_g) \cong (G_1 \times G_2)^0/K \to \mathrm{PGL}(2)$ be the homomorphism induced by the projection of G_1 to $\mathrm{PGL}(2)$. Its image is a finite subgroup P of $\mathrm{PGL}(2)$. Its kernel H consists of cosets modulo K of pairs $(\pm I_2, \sigma_2)$, where $\chi_2(\sigma_2) = 1$. Clearly, $H \cong \mathrm{Ker}(\chi_2)$.

It is easy to see that $\mathrm{Ker}(\chi_2) \cong \mathbb{C}^* : 2$ is generated by diagonal matrices with determinant 1 and the matrix that switches the coordinates. Inside of $\mathrm{GL}(2)$ it is conjugate to the normalizer N of the maximal torus in $\mathrm{SL}(2)$. So we obtain an isomorphism

$$\mathrm{Aut}(Y_g) \cong N \bullet P. \tag{5.2}$$

Suppose there exists a homomorphism $\eta : G_1 \to \mathbb{C}^*$ such that $\eta(-1) = (-1)^{g+1}$. Then the homomorphism

$$G_1 \to (G_1 \times G_2)^0/K, \ \sigma_1 \mapsto (\sigma_1, \eta(\sigma_1) I_2) \mod K$$

factors through a homomorphism $P \to (G_1 \times G_2)^0/K$ and defines a splitting of the extension (5.2). Since elements of the form $(\sigma_1, \eta(\sigma_1) I_2)$ commute with elements of N, we see that the extension is trivial when it splits. It is easy to see that the converse is also true. Since the trivial η works when g is odd, we obtain that the extension always splits in this case. Assume g is even and G_1 admits a 1-dimensional representation η with $\eta(-I_2) = -1$. If its kernel is trivial, G_1 is isomorphic to a subgroup of $\mathbb{C}^*$, hence cyclic. Otherwise, the kernel is a subgroup of $\mathrm{SL}(2)$ not containing the center. It must be a cyclic subgroup of odd order. The image is a cyclic group. Thus G_1 is either cyclic, or a binary dihedral group D_{2n} with n odd.

To summarize we have proved the following.

Proposition 5.3. *The group of automorphisms of an exceptional conic bundle* (5.1) *is isomorphic to an extension $N \bullet P$, where P is the subgroup of* $\mathrm{PGL}(2)$ *leaving the set of zeroes of $F_{2g+2}(T_0, T_1)$ invariant and $N \cong \mathbb{C}^* : 2$ is a group of matrices with determinant ± 1 leaving $T_2 T_3$ invariant. Moreover, the extension splits and defines an isomorphism*

$$\mathrm{Aut}(Y_g) \cong N \times P$$

if and only if g is odd, or g is even and P is either a cyclic group or a dihedral group D_{4k+2}.

Now let G be a finite minimal subgroup of $\mathrm{Aut}(Y_g)$. Assume first that $\mathrm{Aut}(Y_g) \cong N \times P$. Let N' be the projection of G to N and P' be the projection to P. Since G is minimal, N' contains an element which switches $V(T_2)$ and $V(T_3)$. Thus N' is isomorphic to a dihedral group D_{2n}. Applying Goursat's Lemma we obtain that

$$G \cong N' \triangle_D Q,$$

where D is a common quotient of N' and Q. If N' is a dihedral group, then D is either dihedral group or a cyclic of order 2. Using Goursat's Lemma it is easy to list all possible subgroups. We leave it as an exercise to the reader.

If $\mathrm{Aut}(Y_g)$ is not isomorphic to the direct product $N \times P$, we can only say that

$$G \cong H \bullet Q,$$

where H is a subgroup of D_{2n} or Q_{4n}, and Q is a polyhedral group. Note that we can write these extensions in the form $n \bullet (2 \bullet Q)$ or $n \bullet (2^2 \bullet Q)$.

Example 5.4. Let $\phi : S \to \mathbb{P}^1$ be an exceptional conic bundle with $g = 1$. It has 4 singular fibres. According to the first construction, the blow up S' of S at the four singular points of the singular fibres admits an elliptic fibration $f : S' \to \mathbb{P}^1$ with two singular fibres of type I_0^* in Kodaira's notation. The j-invariant of the fibration is zero, and after the degree 2 base change $\mathbb{P}^1 \to \mathbb{P}^1$ ramified at two points, the surface becomes isomorphic to the product $E \times \mathbb{P}^1$, where E is an elliptic curve. Its j-invariant corresponds to the cross ratio of the four points, where a section of π with self-intersection -2 intersects the singular fibres. Conversely, starting from the product, we can divide it by an elliptic involution, to get the conic bundle. This is our second construction.

According to the third construction, the surface can be given by an equation

$$F_4(t_0, t_1) + t_2 t_3 = 0$$

in the weighted projective space $\mathbb{P}(1, 1, 2, 2)$. The projection to (t_0, t_1) is a rational map undefined at the four points $P_i = (a, b, 0, 0)$, where $F_4(a, b) = 0$. After we blow them up we get the conic bundle. The projection to (t_2, t_3) is a rational map undefined at the two singular points $(0, 0, 1, 0)$ and $(0, 0, 0, 1)$. After we blow them up, we get the elliptic fibration. We have two obvious commuting involutions $\sigma_1 = [t_0, t_1, -t_2, -t_3]$ and $\sigma_2 = [t_0, t_1, t_3, t_2]$. The locus of fixed points of each of them is an elliptic curve with equation $t_2 = t_3$ or $t_3 = -t_2$. The group $\langle \sigma_1, \sigma_2 \rangle \cong 2^2$ permutes these two curves.

The groups of automorphisms of S is easy to describe. It follows from Proposition 5.3 that G is a finite subgroup of $P \times K$, where P is a subgroup of $\mathrm{PGL}(2)$ leaving the zeros of F_4 invariant and K is a subgroup of $\mathrm{GL}(2)$ leaving the zeroes of $t_2 t_3$ invariant. First we choose coordinates t_0, t_1 to write F_4 in the form $t_0^4 + t_1^4 + a t_0^2 t_1^2, a^2 \neq 4$. This is always possible if F_4 has four distinct roots (true in our case). Let P be the subgroup of $\mathrm{PGL}(2)$ leaving the set of zeroes of F_4 invariant. It is one of the following groups $1, 2, 4, 2^2, D_8, A_4$. If $a^2 \neq 0, -12, 36$, then P is a subgroup of 2^2. If $a^2 = 0, 36$, then P is a subgroup of D_8. If $a^2 = -12$, then P is a subgroup of A_4.

Suppose a is not exceptional. Let $\bar{P}$ be the corresponding binary group. Then it leaves F_4 invariant, so G is a subgroup of $K' \times P$, where K' consists of matrices leaving $t_2 t_3$ invariant. It is generated by diagonal matrices with determinant 1 and the transformations $[t_1, t_0]$. Thus

$$G \subset D_{2n} \times 2^2,$$

where we use that a finite subgroup of K' is either cyclic or binary dihedral.

Suppose $a^2 = 0, 36$ and G contains an element of order 4. The form can be transformed to the form $T_0^4 + T_1^4$. The value of the character at an element τ of order 4 is equal to -1. We obtain

$$G \subset \tfrac{1}{2}[Q \times P] \cong (Q' \times P')^{\bullet}2 \cong Q'{}_{\bullet}P,$$

where Q is a finite subgroup of K and the diagonal product is taken with respect to the subgroup $Q' = Q \cap K'$ of Q and the subgroup $P \cap 2^2$ of index 2 of P. The group Q' is cyclic or dihedral.

Suppose $a = -2\sqrt{3}i$ and G contains an element g of order 3 given by $[\frac{1-i}{2}t_0 + \frac{1-i}{2}t_1, -\frac{1+i}{2}t_0 - \frac{1+i}{2}t_1]$. Its character is defined by $\chi(g) = \epsilon_3$. We obtain

$$G \subset \frac{1}{3}[Q \times P],$$

where the diagonal product is taken with respect to the subgroup $Q = \chi_2^{-1}(\mu_3)$ of K and the subgroup $P' = P \cap 2^2$ of index 3 of P. Again G is a subgroup of $D_{2n}{\bullet}P$.

5.4 Minimal conic bundles G-surfaces

Now assume (S, G) is a minimal G-surface such that S admits a conic bundle map $\phi : S \to \mathbb{P}^1$. As we had noticed before, the number of singular fibres k is greater or equal to 4. Thus

$$K_S^2 = 8 - k \leq 4. \tag{5.3}$$

Let (S, G) be a rational G-surface and let

$$a : G \to \mathrm{O}(\mathrm{Pic}(X)), \quad g \mapsto (g^*)^{-1}, \tag{5.4}$$

be the natural representation of G in the orthogonal group of the Picard group. We denote by G_0 the kernel of this representation. Since $k > 2$ and G_0 fixes any component of a singular fibre, it acts identically on the base of the conic bundle fibration. Since G_0 fixes the divisor class of a section, and sections with negative self-intersection do not move in a linear system, we see that G_0 fixes pointwise any section with negative self-intersection. If we consider a section as a point of degree 1 on the generic fibre, we see that G_0 must be is a cyclic group.

Proposition 5.5. *Assume $G_0 \neq \{1\}$. Then S is an exceptional conic bundle.*

Proof. Let g_0 be a non-trivial element from G_0. Let E be a section with $E^2 = -n < 0$. Take an element $g \in G$ such that $E' = g(E) \neq E$. Since g_0 has two fixed points on each component of a singular fibre we obtain that E and E' do not intersect the same component. By Lemma 5.1, we obtain that $k = 2n$. Now we blow down n components in n fibres intersecting E and n components in the remaining n fibres intersecting E' to get a minimal ruled surface with two disjoint sections with self-intersection 0. It must be isomorphic to $\mathbf{F}_0$. So, we see that S is an exceptional conic bundle (Construction 1) with $n = g + 1$.

From now on in this section, we assume that $G_0 = \{1\}$.

Let S_η be the general fibre of ϕ. By Tsen's theorem it is isomorphic to $\mathbb{P}^1_K$, where $K = \mathbb{C}(t)$ is the field of rational functions of the base. Consider S_η as a scheme over $\mathbb{C}$. Then

$$\mathrm{Aut}_{\mathbb{C}}(S_\eta) \cong \mathrm{Aut}_K(S_\eta) : \mathrm{PGL}(2) \cong \mathrm{dJ}(2),$$

where $\mathrm{dJ}(2)$ is a de Jonquières subgroup of $\mathrm{Cr}(2)$ and $\mathrm{Aut}_K(S_\eta) \cong \mathrm{PGL}(2, K)$. A finite minimal group G of automorphisms of a conic bundle is isomorphic to a subgroup of $\mathrm{Aut}_{\mathbb{C}}(S_\eta)$. Let $G_K = G \cap \mathrm{Aut}_K(S_\eta)$ and $G_B \cong G/G_K$ be the image of G in $\mathrm{PGL}(2)$. We have an extension of groups

$$1 \to G_K \to G \to G_B \to 1. \tag{5.5}$$

Let $\mathcal{R}$ be the subgroup of $\mathrm{Pic}(S)$ spanned by the divisor classes of $R_i - R_i', i = 1, \ldots, k$. It is obviously G-invariant and $\mathcal{R}_{\mathbb{Q}}$ is equal to the orthogonal complement of $\mathrm{Pic}(S)^G_{\mathbb{Q}}$ in $\mathrm{Pic}(S)_{\mathbb{Q}}$. The orthogonal group of the quadratic lattice $\mathcal{R}$ is isomorphic to the wreath product $2 \wr S_k$. The normal subgroup 2^k consists of transformations which switch some of the R_i's with R_i'. A subgroup isomorphic to S_k permutes the classes $R_i - R_i'$.

Lemma 5.6. *Let G be a minimal group of automorphisms of S. There exists an element $g \in G_K$ of order 2 which switches the components of some singular fibre.*

Proof. Since G is minimal, the G-orbit of any R_i cannot consist of disjoint components of fibres. Thus it contains a pair R_j, R_j', and hence there exists an element $g \in G$ such that $g(R_j) = R_j'$. If g is of odd order $2k+1$, then g^{2k} and g^{2k+1} fix R_j; hence g fixes R_j. This contradiction shows that g is of even order $2m$. Replacing g by an odd power, we may assume that g is of order $m = 2^a$.

Assume $a = 1$. Obviously the singular point $p = R_j \cap R_j'$ of the fibre belongs to the fixed locus S^g of g. Suppose p is an isolated fixed point. Then we can choose local coordinates at p such that g acts by $(z_1, z_2) \mapsto (-z_1, -z_2)$, and hence acts identically on the tangent directions. So it cannot switch the components. Thus S^g contains a curve not contained in fibres which passes through p. This implies that $g \in G_K$.

Suppose $a > 1$. Replacing g by $g' = g^{m/2}$ we get an automorphism of order 2 that fixes the point x_j and the components R_j, R_j'. Suppose $S^{g'}$ contains one of the components, say R_j. Take a general point $y \in R_j$. We have $g'(g(y)) = g(g'(y)) = g(y)$. This shows that g' fixes R_j' pointwise. Since $S^{g'}$ is smooth, this is impossible. Thus g' has three fixed points y, y', p on F_j, two on each component. Suppose y is an isolated fixed point lying on R_j. Let $\pi : S \to S'$ be the blowing down of R_j. The element g' descends to an automorphism of order 2 of S' that has an isolated fixed point at $q = \pi(R_j)$. Then it acts identically on the tangent directions at q, hence on R_j. This contradiction shows that $S^{g'}$

contains a curve intersecting F_j at y or at p, and hence $g' \in G_K$. If g' does not switch components of any fibre then it acts identically on the Picard group. By our assumption this implies that g' is the identity, a contradiction. □

The restriction of the homomorphism $G \to \mathrm{O}(\mathcal{R}) \cong 2^k : S_k$ to G_K defines a surjective homomorphism

$$\rho : G_K \to 2^s, \ s \leq k.$$

An element from $\mathrm{Ker}(\rho)$ acts identically on $\mathcal{R}$ and hence on $\mathrm{Pic}(S)$. By Lemma 5.6, G_K is not trivial and $s > 0$. A finite subgroup of $\mathrm{PGL}(2, K)$ does not admit a surjective homomorphism to 2^s for $s > 2$. Thus $s = 1$ or 2.

Case 1: $s = 1$. Let $\sum'$ be the nonempty subset of Σ such that G_K switches the components of fibers over Σ'. Since G_K is a normal subgroup of G, the set Σ' is a G-invariant set. If $\Sigma \neq \Sigma'$, we repeat the proof of Lemma 5.6 starting from some component R_i of a fiber over a point $x \notin \Sigma'$, and find an element in G_K of even order that switches the components of perhaps another fiber F_x, where $x \notin \Sigma'$. Since $G_K = 2$, we get a contradiction.

Let $G_K = \langle h \rangle$. The element h fixes two points on each nonsingular fibre. The closure of these points is a one-dimensional component C of S^h. It is a smooth bisection of the fibration. Since we know that h switches all components, its trace on the subgroup $\mathcal{R}$ generated by the divisor classes $R_i - R_i'$ is equal to -k. Thus its trace on $H^2(S, \mathbb{Q})$ is equal to $2 - k$. Applying the Lefschetz fixed-point formula, we get $e(S^h) = 4 - k$. If C is the disjoint union of two components, then S^h consists of k isolated fixed points (the singular points of fibers) and C. We get $e(S^h) = 4 + k$. This contradiction shows that C is irreducible and $e(C) = 4 - k$. Since h fixes C pointwise and switches the components R_i and R_i', the intersection point $R_i \cap R_i'$ must be on C. Thus the projection $C \to \mathbb{P}^1$ has $\geq k$ ramification points. Hence $4 - k = e(C) = 4 - (2 + 2g(C)) \leq 4 - k$. This shows that $k = 2g(C) + 2$, i.e., the singular points of fibers are the ramification points of the g_2^1.

Case 2: $s = 2$. Let g_1, g_2 be two elements from G_K which are mapped to generators of the image of G_K in 2^k. Let C_1 and C_2 be the one-dimensional components of the sets S^{g_1} and S^{g_2}. As in the previous case we show that C_1 and C_2 are smooth hyperelliptic curves of genera $g(C_1)$ and $g(C_2)$. Let Σ_1 and Σ_2 be the sets of branch points of the corresponding double covers. Since the group 2^2 does not have isolated fixed points, it is easy to see that $\Sigma = \Sigma_1 \cup \Sigma_2$. For any point $x \in \Sigma_1 \cap \Sigma_2$ the transformation $g_3 = g_1 g_2$ fixes the components of the fibre F_x. For any point $x \in \Sigma_1 \setminus \Sigma_2$, g_3 switches the components of F_x. Let C_3 be the one-dimensional component of S^{g_3} and let Σ_3 be the set of branch points of C_3. We see that $\Sigma_i = \Sigma_j + \Sigma_k$ for distinct i, j, k, where $\Sigma_j + \Sigma_k = (\Sigma_j \cup \Sigma_k) \setminus (\Sigma_j \cap \Sigma_k)$. This implies that there exist three binary

---

We thank V. Tsygankov for this observation.

forms $p_1(t_0,t_1), p_2(t_0,t_1), p_3(t_0,t_1)$, no two of which have a common root, such that $\Sigma_1 = V(p_2p_3), \Sigma_2 = V(p_1p_3), \Sigma_3 = V(p_1p_2)$. Setting $d_i = \deg p_i$, we get

$$2g(C_i) + 2 = d_j + d_k.$$

Let us summarize what we have learnt[2].

Theorem 5.7. *Let G be a minimal finite group of automorphisms of a conic bundle $\phi : S \to \mathbb{P}^1$ with a set Σ of singular fibres. Assume $G_0 = \{1\}$. Then $k = \#\Sigma > 3$ and one of the following cases occurs.*

(1) $G = 2P$, where the central involution h fixes pointwise an irreducible smooth bisection C of π and switches the components in all fibres. The curve C is a curve of genus $g = (k-2)/2$. The conic bundle projection defines a g_2^1 on C with ramification points equal to singular points of fibers. The group P is isomorphic to a group of automorphisms of C modulo the involution defined by the g_2^1.

(2) $G \cong 2^2 \bullet P$, each nontrivial element g_i of the subgroup 2^2 fixes pointwise an irreducible smooth bisection C_i. The set Σ is partitioned into three subsets $\Sigma_1, \Sigma_2, \Sigma_3$ such that the projection $\varphi : C_i \to \mathbb{P}^1$ ramifies over $\Sigma_j + \Sigma_k, i \neq j \neq k$. The group P is a subgroup of $\mathrm{Aut}(\mathbb{P}^1)$ leaving the set Σ and its partition into 3 subsets Σ_i invariant.

It follows from Lemma 4.4 that in Case 1, the non-split extension is isomorphic to a binary polyhedral group, unless $G = O$ or D_{2n}, where n is even.

Remark 5.8. It follows from the previous description that any abelian group G of automorphisms of a conic bundle must be a subgroup of some extension $Q \bullet P$, where Q is a dihedral, binary dihedral, or cyclic group, and P is a polyhedral group. This implies that G is either a cyclic group, or a group $2 \times m$, or $2^2 \times m$, or 2^4. All these groups occur (see Example 5.12 and [6]).

5.5 Automorphisms of hyperelliptic curves

We consider a curve of genus g equipped with a linear series g_2^1 as a curve C of degree $2g+2$ in $\mathbb{P}(1,1,g+1)$ given by an equation

$$t_2^2 + F_{2g+2}(t_0,t_1) = 0.$$

An automorphism σ of C is defined by a transformation

$$(t_0,t_1,t_2) \mapsto (at_1 + bt_0, ct_1 + dt_0, \alpha t_2),$$

where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2)$ and $F(at_0 + bt_1, ct_0 + dt_1) = \alpha^2 F(t_0,t_1)$. So to find the group of automorphisms of C we need to know relative invariants $\Phi(t_0,t_1)$ for

[2] We thank J. Blanc for pointing out a mistake in the statement of this theorem in an earlier version of this paper. The correct statement had appeared first in his paper [7].

finite subgroups $\bar{P}$ of $\mathrm{SL}(2,\mathbb{C})$ (see [49]). The set of relative invariants is a finitely generated $\mathbb{C}$-algebra. Its generators are called Gründformen. We will list the Gründformen (see [49]). We will use them later for the description of automorphism groups of Del Pezzo surfaces of degree 1.

- *$\bar{P}$ is a cyclic group of order n.*

A generator is given by the matrix

$$g = \begin{pmatrix} \epsilon_n & 0 \\ 0 & \epsilon_n^{-1} \end{pmatrix}.$$

The Gründformen are t_0 and t_1 with characters determined by

$$\chi_1(g) = \epsilon_n, \quad \chi_2(g) = \epsilon_n^{-1}.$$

- *$\bar{P} \cong Q_{4n}$ is a binary dihedral group of order $4n$.*

  Its generators are given by the matrices

$$g_1 = \begin{pmatrix} \epsilon_{2n} & 0 \\ 0 & \epsilon_{2n}^{-1} \end{pmatrix}, \quad g_2 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

The Gründformen are

$$\Phi_1 = t_0 t_1 (t_0^4 - t_1^4),\ \Phi_2,\ \Phi_3 = t_0^4 \pm 2\sqrt{-3} t_0^2 t_1^2 + t_1^4. \tag{5.6}$$

The generators g_1, g_2, g_3 act on the Gründformen with characters

$$\chi_1(g_1) = \chi_1(g_2) = \chi_1(g_3) = 1,$$

$$\chi_2(g_1) = \chi_2(g_2) = 1,\ \chi_2(g_3) = \epsilon_3,$$
$$\chi_3(g_1) = \chi_3(g_2) = 1,\ \chi_3(g_3) = \epsilon_3^2.$$

- *$\bar{P}$ is a binary octahedral group of order 48.*

  Its generators are

$$g_1 = \begin{pmatrix} \epsilon_8 & 0 \\ 0 & \epsilon_8^{-1} \end{pmatrix}, \quad g_2 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad g_3 = \frac{1}{i-1}\begin{pmatrix} i & i \\ 1 & -1 \end{pmatrix}.$$

The Gründformen are

$$\Phi_1 = t_0 t_1 (t_0^4 - t_1^4),\ \Phi_2 = t_0^8 + 14 t_0^4 t_1^4 + t_1^8,$$
$$\Phi_3 = (t_0^4 + t_1^4)((t_0^4 + t_1^4)^2 - 36 t_0^4 t_1^4).$$

The generators g_1, g_2, g_3 act on the Gründformen with characters

$$\chi_1(g_1) = -1, \chi_1(g_2) = \chi_1(g_3) = 1,$$
$$\chi_2(g_1) = \chi_2(g_2) = \chi_2(g_3) = 1,$$
$$\chi_3(g_1) = -1, \quad \chi_3(g_2) = \chi(g_3) = 1.$$

- *$\bar{P}$ is a binary icosahedral group of order 120.*

  Its generators are

$$g_1 = \begin{pmatrix} \epsilon_{10} & 0 \\ 0 & \epsilon_{10}^{-1} \end{pmatrix}, \quad g_2 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad g_3 = \frac{1}{\sqrt{5}} \begin{pmatrix} \epsilon_5 - \epsilon_5^4 & \epsilon_5^2 - \epsilon_5^3 \\ \epsilon_5^2 - \epsilon_5^3 & -\epsilon_5 + \epsilon_5^4 \end{pmatrix}.$$

The Gründformen are

$$\begin{aligned}
\Phi_1 &= t_0^{30} + t_1^{30} + 522(t_0^{25}t_1^5 - t_0^5t_1^{25}) - 10005(t_0^{20}t_1^{10} + t_0^{10}t_1^{20}), \\
\Phi_2 &= -(t_0^{20} + t_1^{20}) + 228(t_0^{15}t_1^5 - t_0^5t_1^{15}) - 494t_0^{10}t_1^{10}, \\
\Phi_3 &= t_0t_1(t_0^{10} + 11t_0^5t_1^5 - t_1^{10}).
\end{aligned}$$

Since $P/(\pm 1) \cong A_5$ is a simple group and all Gründformen are of even degree, we easily see that the characters are trivial.

5.6 Commuting de Jonquières involutions

Recall that a de Jonquières involution IH_{g+2} is regularized by an automorphism of the surface S that is obtained from $\mathbf{F}_1$ by blowing up $2g+2$ points. Their images on $\mathbb{P}^2$ are the $2g+2$ fixed points of the involution of H_{g+2}. Let $\pi : S \to X = S/IH_{g+2}$. Since the fixed locus of the involution is a smooth hyperelliptic curve of genus g, the quotient surface X is a nonsingular surface. Since the components of singular fibres of the conic bundle on S are switched by IH_{g+2}, their images on X are isomorphic to $\mathbb{P}^1$. Thus X is a minimal ruled surface $\mathbf{F}_e$. What is e?

Let $\bar{C} = \pi(C)$ and $\bar{E} = \pi(E)$, where E is the exceptional section on S. The curve $\bar{E}$ is a section on X whose preimage in the cover splits. It is either tangent to $\bar{C}$ at any of its of g intersection points (since $IH_{g+2}(E) \cdot E = g$) or disjoint from $\bar{C}$ if $g = 0$. Let s be the divisor class of a section on $\mathbf{F}_e$ with self-intersection $-e$ and f be the class of a fibre. It is easy to see that

$$\bar{C} = (g+1+e)f + 2s, \quad \bar{E} = \frac{g+e-1}{2}f + s.$$

Let $\bar{R}$ be a section with the divisor class s. Suppose $\bar{R} = \bar{E}$; then $\bar{R} \cdot \bar{C} = g+1-e = 2g$ implies $g = 1-e$, so $(g,e) = (1,0)$ or $(0,1)$. In the first case, we get an elliptic curve on $\mathbf{F}_0$ with divisor class $2f + 2s$ and S is a non-exceptional conic bundle with $k = 4$. In the second case S is the conic bundle (nonminimal) with $k = 2$.

Assume that $(g,e) \neq (1,0)$. Let $R = \pi^{-1}(\bar{R})$ be the preimage of $\bar{R}$. We have $R^2 = -2e$. If it splits into two sections $R_1 + R_2$, then $R_1 \cdot R_2 = \bar{C} \cdot \bar{R} = g+1-e$; hence $-2e = 2(g+1-e) + 2R_1^2$ gives $R_1^2 = -g-1$. Applying Lemma 5.1, we get $R_1 \cdot R_2 = g+1-e = g-1+(2g+2-a)/2 = -a/2$, where $a \geq 0$. This gives $e = g+1$, but intersecting $\bar{E}$ with $\bar{R}$ we get $e \leq g-1$. This contradiction shows that $\bar{R}$ does not split, and hence R is an irreducible bisection of the conic bundle with $R^2 = -2e$. We have $R \cdot E = (g-e-1)/2, R \cdot R_i = R \cdot R_i' = 1$, where $R_i + R_i'$ are reducible fibres of the conic fibration.

This shows that the image of R in the plane is a hyperelliptic curve $H'_{g'+2}$ of degree $d = (g-e+3)/2$ and genus $g' = d-2 = (g-e-1)/2$ with the point q of multiplicity g'. It also passes through the points $p_1, \ldots, p_{2g+2}$. Its Weierstrass points $p_1', \ldots, p'_{2g'+2}$ lie on H_{g+2}. Here we use the notation from Section 2.3. Also, the curve $H'_{g'+2}$ is invariant with respect to the de Jonquières involution.

Write the equation of $H'_{g'+2}$ in the form

$$A_{g'}(t_0,t_1)t_2^2 + 2A_{g'+1}(t_0,t_1)t_2 + A_{g'+2}(t_0,t_1) = 0. \tag{5.7}$$

It follows from the geometric definition of the de Jonquières involution that we have the following relation between the equations of $H'_{g'+2}$ and H_{g+2} (cf. [18], p.126):

$$F_g A_{g'+2} - 2F_{g+1}A_{g'+1} + F_{g+2}A_{g'} = 0. \tag{5.8}$$

Consider this as a system of linear equations with coefficients of $A_{g'+2}, A_{g'+1}, A_{g'}$ considered as the unknowns. The number of the unknowns is equal to $(3g-3e+9)/2$. The number of the equations is $(3g-e+5)/2$. So, for a general H_{g+2} we can solve these equations only if $g = 2k+1, e = 0, d = k+2$ or $g = 2k, e = 1, d = k+1$. Moreover, in the first case we get a pencil of curves R satisfying these properties, and in the second case we have a unique such curve (as expected). Also, the first case covers our exceptional case $(g,e) = (1,0)$.

For example, if we take $g = 2$ we obtain that the six Weierstrass points $p_1, \ldots, p_6$ of H_{g+2} must be on a conic. Or, if $g = 3$, the eight Weierstrass points together with the point q must be the base points of a pencil of cubics. All these properties are of course not expected for a general set of six or eight points in the plane.

To sum up, we have proved the following.

Theorem 5.9. *Let H_{g+2} be a hyperelliptic curve of degree $g+2$ and genus g defining a de Jonquières involution IH_{g+2}. View this involution as an automorphism τ of order 2 of the surface S obtained by blowing up the singular point q of H_{g+2} and its $2g+2$ Weierstrass points $p_1, \ldots, p_{2g+2}$. Then*

(i) the quotient surface $X = S/(\tau)$ is isomorphic to $\mathbf{F}_e$ and the ramification curve is $C = S^\tau$;
(ii) if H_{g+2} is a general hyperelliptic curve, then $e = 0$ if g is odd and $e = 1$ if g is even;
(iii) the branch curve $\bar{C}$ of the double cover $S \to \mathbf{F}_e$ is a curve from the divisor class $(g+1+e)f + 2s$;
(iv) there exists a section from the divisor class $\frac{g+e-1}{2}f + s$ which is tangent to $\bar{C}$ at each g intersection points unless $g = 0, e = 1$ in which case it is disjoint from $\bar{C}$;

(v) *the reducible fibres of the conic bundle on S are the preimages of the $2g+2$ fibres from the pencil $|f|$ which are tangent to $\bar{C}$;*
(vi) *the preimage of a section from the divisor class s either splits if $(g,e) = (1,0)$ or a curve of genus $g = 0$, or a hyperelliptic curve C' of genus $g' = (g-e-1)/2 \geq 1$ that is invariant with respect to τ. It intersects the hyperelliptic curve C at its $2g'+2$ Weierstrass points;*
(vii) *the curve C' is uniquely defined if $e > 0$ and varies in a pencil if $e = 0$.*

Let $IH'_{g'+2}$ be the de Jonquières involution defined by the curve $H'_{g'+2}$ from equation (5.7). Then it can be given in affine coordinates by formulas (2.7), where F_i is replaced with A_i. Thus we have two involutions defined by the formulas

$$
\begin{aligned}
IH_{g+2}:\ (x',y') &= \big(x, \frac{-yP_{g+1}(x) - P_{g+2}(x)}{P_g(x)y + P_{g+1}(x)}\big), \\
IH'_{g'+2}: (x',y') &= \big(x, \frac{-yQ_{g'+1}(x) - Q_{g'+2}(x)}{Q_g(x)y + Q_{g'+1}(x)}\big),
\end{aligned}
\tag{5.9}
$$

where the P_i are the dehomogenizations of the F_i and the Q_i are the dehomogenizations of the A_i. Composing them in both ways, we see that the relation (5.8) is satisfied if and only if the two involutions commute. Thus a de Jonquières involution can always be included in a group of de Jonquières transformations isomorphic to 2^2. In fact, for a general IH_{g+2} there exists a unique such group if g is even and there is an ∞^1 of such groups when g is odd. It is easy to check that the involution $IH_{g+2} \circ H'_{g'+2}$ is the de Jonquières involution defined by the third hyperelliptic curve with equation

$$
\det \begin{pmatrix} F_g & F_{g+1} & F_{g+2} \\ A_{g'} & A_{g'+1} & A_{g''+2} \\ 1 & -t_2 & t_2^2 \end{pmatrix} = B_{g''}t_2^2 + 2B_{g''-1}t_2 + B_{g''+2} = 0 \tag{5.10}
$$

(cf. [18], p.126).

If we blow up the Weierstrass point of the curve C' (the proper transform of $H'_{g'+2}$ in S), then we get a conic bundle surface S' from case (2) of Theorem 5.7.

5.7 A question on extensions

It still remains to decide which extensions

$$
1 \to G_K \to G \to G_B \to 1 \tag{5.11}
$$

describe minimal groups of automorphisms of conic bundles. We do not have the full answer and only make a few remarks and examples. Lemma 4.4 helps to decide on splitting in the case that $\mathrm{G_K}$ is abelian and central.

Example 5.10. Consider a de Jonquières transformation

$$\mathrm{dj}_P : (x, y) \mapsto (x, P(x)/y),$$

where $P(t_1/t_0) = t_0^{-2g} F_{2g+2}(t_0, t_1)$ is a dehomogenization of a homogenous polynomial $F_{2g+2}(t_0, t_1)$ of degree $2g+2$ defining a hyperelliptic curve of genus g. Choose F_{2g+2} to be a relative invariant of a binary polyhedral group $\bar{P}$ with character $\chi : \bar{P} \to \mathbb{C}^*$. We assume that $\chi = \alpha^2$ for some character $\alpha : \bar{P} \to \mathbb{C}^*$. For any $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \bar{P}$ define the transformation

$$g : (x, y) \mapsto \left(\frac{ax+b}{cx+d}, \alpha(g)(cx+d)^{-g-1} y\right).$$

We have

$$P\left(\frac{ax+b}{cx+d}\right) = \alpha^2(g)(cx+d)^{-2g-2} P(x).$$

It is immediate to check that g and dj_p commute. The matrix $-I_2$ defines the transformation $g_0 : (x, y) \mapsto (x, \alpha(-I_2)(-1)^{g+1} y)$. So, if

$$\alpha(-I_2) = (-1)^{g+1},$$

the action of $\bar{P}$ factors through P and together with dj_P generate the group $2 \times P$. On the other hand, if $\alpha(-I_2) = (-1)^g$, we get the group $G = 2 \times \bar{P}$. In this case the group G is regularized on an exceptional conic bundle with $G_0 \cong 2$. The generator corresponds to the transformation g_0.

Our first general observation is that the extension $G = 2P$ always splits if g is even, and, of course, if P is a cyclic group of odd order. In fact, suppose G does not split. We can always find an element $g \in G$ which is mapped to an element $\bar{g}$ in P of even order $2d$ such that $g^{2d} = g_0 \in G_K$. Now $g_1 = g^d$ defines an automorphism of order 2 of the hyperelliptic curve $C = S^{g_0}$ with fixed points lying over two fixed points of $\bar{g}$ in $\mathbb{P}^1$. None of these points belong to Σ, since otherwise g_0, being a square of g_1, cannot switch the components of the corresponding fibre. Since g_1 has two fixed points on the invariant fibre and both of them must lie on C, we see that g_1 has 4 fixed points. However, this contradicts the Hurwitz formula.

Recall that a double cover $f : X \to Y$ of nonsingular varieties with branch divisor $W \subset Y$ is given by an invertible sheaf $\mathcal{L}$ together with a section $s_W \in \Gamma(Y, \mathcal{L}^2)$ with zero divisor W. Suppose a group G acts on Y leaving invariant W. A *lift* of G is a group $\widetilde{G}$ of automorphisms of X such that it commutes with the covering involution τ of X and the corresponding homomorphism $\widetilde{G} \to \mathrm{Aut}(Y)$ is an isomorphism onto the group G.

The following lemma is well-known and is left to the reader.

Lemma 5.11. *A subgroup $G \subset \mathrm{Aut}(Y)$ admits a lift if and only if $\mathcal{L}$ admits a G-linearization and in the corresponding representation of G in $\Gamma(Y, \mathcal{L}^2)$ the section s_W is G-invariant.*

Example 5.12. Let $p_i(t_0, t_1), i = 0, 1, 2$, be binary forms of degree d. Consider a curve C in $\mathbf{F}_0 \cong \mathbb{P}^1 \times \mathbb{P}^1$ given by an equation

$$F = p_0(t_0, t_1)x_0^2 + 2p_1(t_0, t_1)x_0x_1 + p_2(t_0, t_1)x_1^2 = 0.$$

Assume that the binary form $D = p_1^2 - p_0p_2$ does not have multiple roots. Then C is a nonsingular hyperelliptic curve of genus $d - 1$. Suppose $d = 2a$ is even, so that the genus of the curve is odd. Let P be a polyhedral group not isomorphic to a cyclic group of odd order. Let $V = \Gamma(\mathbb{P}^1, \mathcal{O}_{\mathbb{P}^1})$ and $\rho : P \to \mathrm{GL}(S^{2a}V \otimes S^2V)$ be its natural representation, the tensor product of the two natural representations of P in the space of binary forms of even degree. Suppose that $F \in S^{2a}V \otimes S^2V$ is an invariant. Consider the double cover $S \to \mathbf{F}_0$ defined by the section F and the invertible sheaf $\mathcal{L} = \mathcal{O}_{\mathbf{F}_0}(a, 1)$. Now assume additionally that P does not have a linear representation in $S^aV \otimes V$ whose tensor square is equal to ρ. Thus $\mathcal{L}$ does not admit a P-linearization and we cannot lift P to a group of automorphisms of the double cover. However, the binary polyhedral group $\bar{P}$ lifts. Its central involution acts identically on $\mathbf{F}_0$, hence lifts to the covering involution of S. It follows from the discussion in the previous subsection that S is a non-exceptional conic bundle, and the group $\bar{P}$ is a minimal group of automorphisms of S with $G_K \cong 2$ and $G_B \cong P$.

Here is a concrete example. Take

$$p_0 = t_0t_1(t_0^2 + t_1^2), \quad p_1 = t_0^4 + t_1^4, \quad p_2 = t_0t_1(t_0^2 - t_1^2).$$

Let $\bar{P} \subset \mathrm{SL}(2)$ be a cyclic group of order 4 that acts on the variables t_0, t_1 via the transformation $[it_0, -it_1]$ and on the variables x_0, x_1 via the transformation $[ix_0, -ix_1]$. Then $\bar{P}$ acts on $S^2V \otimes V$ via $[-1, 1, -1] \otimes [i, -i]$. The matrix $-I_2$ acts as $1 \otimes -1$ and hence $P = \bar{P}/(\pm I_2)$ does not act on $S^2V \otimes V$. This realizes the cyclic group C_4 as a minimal group of automorphisms of a conic bundle with $k = 2g + 2 = 8$.

The previous example shows that for any $g \equiv 1 \mod 4$ one can realize a binary polyhedral group $\bar{P} = 2.P$ as a minimal group of automorphisms of a conic bundle with $2g + 2$ singular fibres. We do not know whether the same is true for $g \equiv 3 \mod 4$.

Example 5.13. Let $p_i(t_0, t_1), i = 1, 2, 3$, be three binary forms of even degree d with no multiple roots. Assume no two have common zeroes. Consider a surface S in $\mathbb{P}^1 \times \mathbb{P}^2$ given by a bihomogeneous form of degree $(d, 2)$,

$$p_1(t_0, t_1)z_0^2 + p_2(t_0, t_1)z_1^2 + p_3(t_0, t_1)z_2^2 = 0. \tag{5.12}$$

The surface is nonsingular. The projection to $\mathbb{P}^1$ defines a conic bundle structure on S with singular fibres over the zeroes of the polynomials p_i. The curves C_i equal to the preimages of the lines $z_i = 0$ under the second projection are hyperelliptic curves of genus $g = d - 1$. The automorphisms σ_1, σ_2 defined by the negation of one of the first two coordinates z_0, z_1, z_2 form a subgroup of

$\mathrm{Aut}(S)$ isomorphic to 2^2. Let P be a finite subgroup of $\mathrm{SL}(2,\mathbb{C})$ and $g \mapsto g^*$ be its natural action on the space of binary forms. Assume that p_1, p_2, p_3 are relative invariants of P with characters χ_1, χ_2, χ_3 such that we can write them in the form η_i^2 for some characters η_1, η_2, η_3 of P. Then P acts on S by the formula

$$g((t_0,t_1),(z_0,z_1,z_2)) = ((g^*(t_0), g^*(t_1)), (\eta_1(g)^{-1}z_0, \eta_2(g)^{-1}z_1, \eta_3(g)^{-1}z_2)).$$

For example, let $P = \langle g\rangle$ be a cyclic group of order 4. We take $p_1 = t_0^2 + t_1^2, p_2 = t_0^2 - t_1^2, p_3 = t_0t_1$. It acts on S by the formula

$$g : ((t_0,t_1),(z_0,z_1,z_2)) \mapsto ((it_1, it_0),(iz_0, z_1, iz_2)).$$

Thus g^2 acts identically on t_0, t_1, z_1 and multiplies z_0, z_2 by -1. We see that $G_K = \langle g^2\rangle$ and the extension $1 \to G_K \to G \to G_B \to 1$ does not split. If we add to the group the transformation $(t_0,t_1,z_0,z_1,z_2) \mapsto (t_0,t_1,z_0,-z_1,z_2)$, we get a non-split extension 2^{2+1}.

On the other hand, let us replace p_2 with $t_0^2 + t_1^2 + t_0t_1$. Define g_1 as acting only on t_0, t_1 by $[it_1, it_0]$, g_2 acts only on z_0 by $z_0 \mapsto -z_0$, and g_3 acts only on z_1 by $z_1 \mapsto -z_1$. We get the groups $\langle g_1, g_2\rangle = 2^2$ and $\langle g_1, g_2, g_3\rangle = 2^3$.

In another example we take P to be the dihedral group D_8. We take $p_1 = t_0^2 + t_1^2, p_2 = t_0^2 - t_1^2, p_3 = t_0t_1$. It acts on S by the formula

$$\left(\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right) : ((t_0,t_1),(z_0,z_1,z_2)) \mapsto ((it_0, -it_1),(iz_0, iz_1, z_2)),$$

$$\left(\begin{smallmatrix} 0 & i \\ i & 0 \end{smallmatrix}\right) : ((t_0,t_1),(z_0,z_1,z_2)) \mapsto ((it_1, it_0),(z_0, iz_1, z_2)).$$

The scalar matrix $c = -I_2$ belongs to $G_K \cong 2^2$ and the quotient $P/(c) \cong 2^2$ acts faithfully on the base. This gives a non-split extension 2^{2+2}.

Finally, let us take

$$p_1 = t_0^4 + t_1^4,\ p_2 = t_0^4 + t_1^4 + t_0^2t_1^2,\ p_3 = t_0^4 + t_1^4 - t_0^2t_1^2.$$

These are invariants for the group D_4 acting via $g_1 : (t_0,t_1) \mapsto (t_0,-t_1),\ g_2 : (t_0,t_1) \mapsto (t_1,t_0)$. Together with transformations σ_1, σ_2 this generates the group 2^4 (see another realization of this group in [9]).

6 Automorphisms of Del Pezzo surfaces

6.1 The Weyl group

Let S be a Del Pezzo surface of degree d not isomorphic to $\mathbb{P}^2$ or $\mathbf{F}_0$. It is isomorphic to the blowup of $N = 9 - d \leq 8$ points in $\mathbb{P}^2$ satisfying the conditions of generality from section 3.4. The blowup of one or 2 points is obviously nonminimal (since the exceptional curve in the first case and the proper transform of the line through the two points is G-invariant). So we may assume that S is a Del Pezzo surface of degree $d \leq 6$.

Let $\pi : S \to \mathbb{P}^2$ be the blowing-up map. Consider the factorization (3.2) of π into a composition of blowups of $N = 9-d$ points. Because of the generality condition, we may assume that none of the points $p_1, \ldots, p_N$ is infinitely near, or, equivalently, all exceptional curves $\mathcal{E}_i$ are irreducible curves. We identify them with curves $E_i = \pi^{-1}(p_i)$. The divisor classes $e_0 = [\pi^*(\text{line}], e_i = [E_i], i = 1, \ldots, N$, form a basis of $\text{Pic}(S)$. It is called a *geometric basis.*

Let

$$\alpha_1 = e_0 - e_1 - e_2 - e_3, \alpha_2 = e_1 - e_2, \ldots, \alpha_N = e_{N-1} - e_N.$$

For any $i = 1, \ldots, N$ define a *reflection* isometry s_i of the abelian group $\text{Pic}(S)$:

$$s_i : x \mapsto x + (x \cdot \alpha_i)\alpha_i.$$

Obviously, $s_i^2 = 1$ and s_i acts identically on the orthogonal complement of α_i. Let W_S be the group of automorphisms of $\text{Pic}(S)$ generated by the transformations $s_1, \ldots, s_N$. It is called the *Weyl group of* S. Using the basis $(e_0, \ldots, e_N)$ we identify W_S with a group of isometries of the odd unimodular quadratic form $q : \mathbb{Z}^{N+1} \to \mathbb{Z}$ of signature $(1, N)$ defined by

$$q_N(m_0, \ldots, m_N) = m_0^2 - m_1^2 - \cdots - m_N^2.$$

Since $K_S = -3e_0 + e_1 + \cdots + e_N$ is orthogonal to all α_i's, the image of W_S in $\text{O}(q_N)$ fixes the vector $k_N = (-3, 1, \ldots, 1)$. The subgroup of $\text{O}(q_N)$ fixing k_N is denoted by W_N and is called the *Weyl group of type* E_N. The orthogonal complement $\mathcal{R}_N$ of k_N equipped with the restricted inner-product is called the *root lattice* of W_N.

We denote by $\mathcal{R}_S$ the sublattice of $\text{Pic}(S)$ equal to the orthogonal complement of the vector K_S. The vectors $\alpha_1, \ldots, \alpha_N$ form a $\mathbb{Z}$-basis of $\mathcal{R}_S$. By restriction the Weyl group W_S is isomorphic to a subgroup of $\text{O}(\mathcal{R}_S)$. A choice of a geometric basis $\alpha_1, \ldots, \alpha_N$ defines an isomorphism from $\mathcal{R}_S$ to the root lattice Q of a finite root system of type $E_N(N = 6, 7, 8), D_5(N = 5), A_4(N = 4)$ and $A_2 + A_1(N = 3)$. The group W_S becomes isomorphic to the corresponding Weyl group $W(E_N)$.

The next lemma is well-known and its proof goes back to Kantor [42] and Du Val [28]. We refer for modern proofs to [2] or [26].

Lemma 6.1. *Let* $(e_0', e_1', \ldots, e_N')$ *be another geometric basis in* $\text{Pic}(S)$ *defined by a birational morphism* $\pi' : S \to \mathbb{P}^2$ *and a choice of a factorization of* π' *into a composition of blowups of points. Then the transition matrix is an element of* W_N. *Conversely, any element of* W_N *is a transition matrix of two geometric bases in* $\text{Pic}(S)$.

The next lemma is also well-known and is left to the reader.

Lemma 6.2. *If* $d \leq 5$, *then the natural homomorphism*

$$\rho : \text{Aut}(S) \to W_S$$

is injective.

We will use the known classification of conjugacy classes in the Weyl groups. According to [15] they are indexed by certain graphs. We call them *Carter graphs*. One writes each element $w \in W$ as the product of two involutions $w_1 w_2$, where each involution is the product of reflections with respect to orthogonal roots. Let $\mathcal{R}_1, \mathcal{R}_2$ be the corresponding sets of such roots. Then the graph has vertices identified with elements of the set $\mathcal{R}_1 \cup \mathcal{R}_2$, and two vertices α, β are joined by an edge if and only if $(\alpha, \beta) \neq 0$. A Carter graph with no cycles is a Dynkin diagram. The subscript in the notation of a Carter graph indicates the number of vertices. It is also equal to the difference between the rank of the root lattice Q and the rank of its fixed sublattice $Q^{(w)}$.

Note that the same conjugacy classes may correspond to different graphs (e.g., D_3 and A_3, or $2A_3 + A_1$ and $D_4(a_1) + 3A_1$).

The Carter graph determines the characteristic polynomial of w. In particular, it gives the trace $\mathrm{Tr}_2(g)$ of g^* on the cohomology space $H^2(S, \mathbb{C}) \cong \mathrm{Pic}(S) \otimes \mathbb{C}$. The latter should be compared with the Euler-Poincarè characteristic of the fixed locus S^g of g by applying the Lefschetz fixed-point formula.

Table 2. Carter graphs and characteristic polynomials.

Graph	Order	Characteristic polynomial
A_k	$k+1$	$t^k + t^{k-1} + \cdots + 1$
D_k	$2k-2$	$(t^{k-1}+1)(t+1)$
$D_k(a_1)$	l.c.m$(2k-4, 4)$	$(t^{k-2}+1)(t^2+1)$
$D_k(a_2)$	l.c.m$(2k-6, 6)$	$(t^{k-3}+1)(t^3+1)$
$\vdots$	$\vdots$	$\vdots$
$D_k(a_{\frac{k}{2}-1})$	even k	$(t^{\frac{k}{2}}+1)^2$
E_6	12	$(t^4-t^2+1)(t^2+t+1)$
$E_6(a_1)$	9	t^6+t^3+1
$E_6(a_2)$	6	$(t^2-t+1)^2(t^2+t+1)$
E_7	18	$(t^6-t^3+1)(t+1)$
$E_7(a_1)$	14	t^7+1
$E_7(a_2)$	12	$(t^4-t^2+1)(t^3+1)$
$E_7(a_3)$	30	$(t^5+1)(t^2-t+1)$
$E_7(a_4)$	6	$(t^2-t+1)^2(t^3+1)$
E_8	30	$t^8+t^7-t^5-t^4-t^3+t+1$
$E_8(a_1)$	24	t^8-t^4+1
$E_8(a_2)$	20	$t^8-t^6+t^4-t^2+1$
$E_8(a_3)$	12	$(t^4-t^2+1)^2$
$E_8(a_4)$	18	$(t^6-t^3+1)(t^2-t+1)$
$E_8(a_5)$	15	$t^8-t^7+t^5-t^4+t^3-t+1$
$E_8(a_6)$	10	$(t^4-t^3+t^2-t+1)^2$
$E_8(a_7)$	12	$(t^4-t^2+1)(t^2-t+1)^2$
$E_8(a_8)$	6	$(t^2-t+1)^4$

$$\mathrm{Tr}_2(g) = s - 2 + \sum_{i \in I}(2 - 2g_i), \tag{6.1}$$

where S^g the disjoint union of smooth curves $R_i, i \in I$, of genus g_i and s isolated fixed points.

To determine whether a finite subgroup G of $\mathrm{Aut}(S)$ is minimal, we use the well-known formula from the character theory of finite groups

$$\text{rank } \mathrm{Pic}(S)^G = \frac{1}{\#G} \sum_{g \in G} \mathrm{Tr}_2(g).$$

The tables for conjugacy classes of elements from the Weyl group W_S give the values of the trace on the lattice $\mathcal{R}_S = K_S^{\perp}$. Thus the group is minimal if and only if the sum of the traces add up to 0.

6.2 Del Pezzo surfaces of degree 6

Let S be a Del Pezzo surface of degree 6. We fix a geometric basis e_0, e_1, e_2, e_3 which is defined by a birational morphism $\pi : S \to \mathbb{P}^2$ with indeterminacy points $p_1 = (1, 0, 0), p_2 = (0, 1, 0)$, and $p_3 = (0, 0, 1)$. The vectors

$$\alpha_1 = e_0 - e_1 - e_2 - e_3, \alpha_2 = e_1 - e_2, \alpha_3 = e_2 - e_3$$

form a basis of the lattice $\mathcal{R}_S$ with Dynkin diagram of type $A_2 + A_1$. The Weyl group

$$W_S = \langle s_1 \rangle \times \langle s_2, s_3 \rangle \cong 2 \times S_3.$$

The representation $\rho : \mathrm{Aut}(S) \to W_S$ is surjective. The reflection s_1 is realized by the lift of the standard quadratic transformation τ_1. The reflection s_2(respectively s_3) is realized by the projective transformations $[x_1, x_0, x_2]$ (respectively $[x_0, x_2, x_1]$). The kernel of ρ is the maximal torus T of $\mathrm{PGL}(3)$, the quotient of $(\mathbb{C}^*)^3$ by the diagonal subgroup $\mathbb{C}^*$. Thus

$$\mathrm{Aut}(S) \cong T : (S_3 \times 2) \cong N(T) : 2,$$

where $N(T)$ is the normalizer of T in $\mathrm{PGL}(2)$. It is easy to check that s_1 acts on T as the inversion automorphism.

Let G be a minimal finite subgroup of $\mathrm{Aut}(S)$. Obviously, $\rho(G)$ contains s_1 and $s_2 s_3$ since otherwise, G leaves invariant α_1 or one of the vectors $2\alpha_1 + \alpha_2$, or $\alpha_1 + 2\alpha_2$. This shows that $G \cap N(T)$ is an imprimitive subgroup of $\mathrm{PGL}(3)$. This gives the following result.

Theorem 6.3. *Let G be a minimal subgroup of a Del Pezzo surface of degree 6. Then*

$$G = H_{\bullet}\langle s_1 \rangle,$$

where H is an imprimitive finite subgroup of $\mathrm{PGL}(3)$.

Note that one of the groups from the theorem is the group $2^2 : S_3 \cong S_4$. Its action on S given by the equation

$$x_0y_0z_0 - x_1y_1z_1 = 0$$

in $(\mathbb{P}^1)^3$ is given in [4].

6.3 Del Pezzo surfaces of degree $d = 5$.

In this case S is isomorphic to the blowup of the reference points $p_1 = (1,0,0)$, $p_2 = (0,1,0)$, $p_3 = (0,0,1), p_4 = (1,1,1)$. The lattice $\mathcal{R}_S$ is of type A_4 and $W_S \cong S_5$. It is known that the homomorphism $\rho : \mathrm{Aut}(S) \to W_S$ is an isomorphism. We already know that it is injective. To see the surjectivity one can argue, for example, as follows.

Let τ be the standard quadratic transformation with base points p_1, p_2, p_3. It follows from its formula that the point p_4 is a fixed point. We know that τ can be regularized on the Del Pezzo surface S' of degree 6 obtained by the blowup of the first three points. Since the preimage of p_4 in S' is a fixed point, τ lifts to an automorphism of S. Now let ϕ be a projective transformation such that $\phi(p_1) = p_1, \phi(p_2) = p_2, \phi(p_4) = p_3$. For example, we take $A = [t_0 - t_2, t_1, -t_2, -t_2]$. Then the quadratic transformation $\phi^{-1}\tau\phi$ is not defined at the points p_1, p_2, p_4 and fixes the point p_3. As above, it can be lifted to an involution of S. Proceeding in this way we define 4 involutions $\tau = \tau_1, \ldots, \tau_4$ of S; each fixes one of the exceptional curves. One checks that their images in the Weyl group W_S generate the group.

Another way to see the isomorphism $\mathrm{Aut}(S) \cong S_5$ is to use a well-known isomorphism between S and the moduli space $\overline{\mathcal{M}_{0,5}} \cong (\mathbb{P}^1)^5//\mathrm{SL}(2)$. The group S_5 acts by permuting the factors.

Theorem 6.4. *Let (S, G) be a minimal Del Pezzo surface of degree $d = 5$. Then $G = S_5, A_5, 5 : 4, 5 : 2$, or C_5.*

Proof. As we have just shown $\mathrm{Aut}(S) \cong W_4 \cong S_5$. The group S_5 acts on $\mathcal{R}_S \cong \mathbb{Z}^4$ by means of its standard irreducible 4-dimensional representation (view $\mathbb{Z}^4$ as a subgroup of $\mathbb{Z}^5$ of vectors with coordinates added up to zero and consider the representation of S_5 by switching the coordinates). It is known that a maximal proper subgroup of S_5 is equal (up to a conjugation) to one of three subgroups $S_4, S_3 \times 2, A_5, 5 : 4$. A maximal subgroup of A_5 is either 5×2 or S_3 or $D_{10} = 5 : 2$. It is easy to see that the groups S_4 and $S_3 \times 2$ have invariant elements in the lattice Q_4. It is known that an element of order 5 in S_5 is a cyclic permutation, and hence has no invariant vectors. Thus any subgroup G of S_5 containing an element of order 5 defines a minimal surface (S, G). So, if (S, G) is minimal, G must be equal to one of the groups from the assertion of the theorem.

6.4 Automorphisms of a Del Pezzo surface of degree $d = 4$

If the degree is 4, then $\mathcal{R}$ is of type D_5 and $W_S \cong 2^4 : S_5$. We use the following well-known classical result.

Lemma 6.5. *Let S be a Del Pezzo surface of degree 4. Then S is isomorphic to a nonsingular surface of degree 4 in $\mathbb{P}^4$ given by equations*

$$F_1 = \sum_{i=0}^{4} T_i^2 = 0, \quad F_2 = \sum_{i=0}^{4} a_i T_i^2 = 0, \tag{6.2}$$

where all a_i's are distinct.

Proof. It is known that a Del Pezzo surface in its anti-canonical embedding is projectively normal. Using Riemann-Roch, one obtains that S is a complete intersection $Q_1 \cap Q_2$ of two quadrics. Let $\mathcal{P} = \lambda Q_1 + \mu Q_2$ be the pencil spanned by these quadrics. The locus of singular quadrics in the pencil is a homogeneous equation of degree 5 in the coordinates λ, μ. Since S is nonsingular, it is not hard to see that the equation has no multiple roots (otherwise; $\mathcal{P}$ contains a reducible quadric or there exists a quadric in the pencil with singular point at S; in both cases S is singular). Let $p_1, \ldots, p_5$ be the singular points of singular quadrics from the pencil. Suppose they are contained in a hyperplane H. Since no quadrics in the pencil contains H, the restriction $\mathcal{P}|H$ of the pencil of quadrics to H contains ≥ 5 singular members. This implies that all quadrics in $\mathcal{P}|H$ are singular. By Bertini's theorem, all quadrics are singular at some point $p \in H$. This implies that all quadrics in $\mathcal{P}$ are tangent to H at p. One of the quadrics must be singular at p, and hence S is singular at p. This contradiction shows that $p_1, \ldots, p_5$ span $\mathbb{P}^4$. Choose coordinates in $\mathbb{P}^4$ such that the singular points of singular quadrics from $\mathcal{P}$ are the points $(1,0,0,0,0), (0,1,0,0,0)$, and so on. Then each hyperplane $V(T_i) = (T_i = 0)$ is a common tangent hyperplane of quadrics from $\mathcal{P}$ at the point p_i. This easily implies that the equations of quadrics are given by (6.2).

Let $Q_i = V(a_i F_1 - F_2), i = 0, \ldots, 4$, be one of the singular quadrics in the pencil $\mathcal{P}$; It is a cone over a nonsingular quadric in $\mathbb{P}^3$; hence it contains two families of planes. The intersection of a plane with any other quadric in the pencil is a conic contained in S. Thus each Q_i defines a pair of pencils of conics $|C_i|$ and $|C_i'|$, and it is easy to see that $|C_i + C_i'| = |-K_S|$.

Proposition 6.6. *Let S be a Del Pezzo surface given by equations* (6.2)*. The divisor classes $c_i = [C_i]$ together with K_S form a basis of* $\mathrm{Pic}(S) \otimes \mathbb{Q}$*. The Weyl group W_S acts on this basis by permuting the c_i's and sending c_i to $c_i' = [C_i'] = -K_S - c_i$.*

Proof. If we choose a geometric basis $(e_0, e_1, \ldots, e_5)$ in $\mathrm{Pic}(S)$, then the 5 pairs of pencils of conics are defined by the classes $e_0 - e_i, 2e_0 - e_1 - \cdots - e_5 + e_i$.

It is easy to check that the classes $[C_i]$'s and K_S form a basis in $\mathrm{Pic}(S)\otimes\mathbb{Q}$. The group W_S contains a subgroup isomorphic to S_5 generated by the reflections in vectors $e_1-e_2,\ldots,e_4-e_5$., It acts by permuting $e_1,\ldots,e_5$, hence permuting the pencils $|C_i|$. It is equal to the semi-direct product of S_5 and the subgroup isomorphic to 2^4 which is generated by the conjugates of the product s of two commuting reflections with respect to the vectors $e_0-e_1-e_2-e_3$ and e_4-e_5. It is easy to see that $s([C_4])=[C_4'], s([C_5])=[C_5']$ and $s([C_i])=[C_i]$ for $i\neq 4,5$. This easily implies that W_S acts by permuting the classes $[C_i]$ and switching even number of them to $[C_i']$.

Corollary 6.7. *Let $W(D_5)$ act in $\mathbb{C}^5$ by permuting the coordinates and switching the signs of even number of coordinates. This linear representation of $W(D_5)$ is isomorphic to the representation of $W(D_5)$ on $\mathcal{R}_S\otimes\mathbb{C}$.*

The group of projective automorphisms generated by the transformations which switch x_i to $-x_i$ generates a subgroup H of $\mathrm{Aut}(S)$ isomorphic to 2^4. We identify the group H with the linear space of subsets of even cardinality of the set $J=\{0,1,2,3,4\}$, or, equivalently, with the subspace of $\mathbb{F}_2^J$ of functions with support at a subset of even cardinality. We equip H with the symmetric bilinear form defined by the dot-product in $\mathbb{F}_2^J$, or, equivalently, by $(A,B)=\#A\cap B \mod 2$. We denote elements of H by i_A, where i_A is the characteristic function of $A\subset J$.

There are two kinds of involutions i_A. An involution of the first kind corresponds to a subset A of 4 elements. The set of fixed points of such an involution is a hyperplane section of S, an elliptic curve. The trace formula (6.1) gives that the the trace of i_A in $\mathrm{Pic}(S)$ is equal to -2. The corresponding conjugacy class in W_5 is of type $4A_1$. There are 5 involutions of the first kind. The quotient surface $S/\langle i_A\rangle=Q$ is isomorphic to a nonsingular quadric. The map $S\to Q$ coincides with the map $S\to\mathbb{P}^1\times\mathbb{P}^1$ that is given by the pencils $|C_i|$ and $|C_i'|$.

Involutions of the second type correspond to subsets A of cardinality 2. The fixed-point set of such involution consists of 4 isolated points. This gives that the trace is equal to 2, and the conjugacy class is of type $2A_1$. The quotient $S/(i_A)$ is isomorphic to the double cover of $\mathbb{P}^2$ branched along the union of two conics.

The subgroup of the Weyl group $W(D_5)$ generated by involutions from the conjugacy class of type $2A_1$ is the normal subgroup 2^4 in the decomposition $W(D_5)\cong 2^4:S_5$. The product of two commuting involutions from this conjugacy class is an involution of type $4A_1$. Thus the image of H in W_S is a normal subgroup isomorphic to 2^4.

Let $G\cong 2^a$ be a subgroup of 2^4. All cyclic groups G are not minimal.

There are three kinds of subgroups H of order 4 in 2^4. A subgroup of the first kind does not contain an involution of the first kind. An example is the group generated by i_{01},i_{12}. The trace of its nonzero elements equal to 1. So this group is not minimal.

A subgroup of the second type contains only one involution of the first kind. An example is the group generated by i_{01}, i_{23}. The trace formula gives rank $\mathrm{Pic}(S)^H = 2$. So it is also nonminimal.

A subgroup of the third kind contains two involutions of the first kind, for example, a group generated by i_{1234}, i_{0234}. It contains 2 elements with trace -3 and one element with trace 1. Adding up the traces, we see the group is a minimal group. It is easy to see that S^H consists of 4 isolated points.

Now let us consider subgroups of 2^4 of order 8. They are parametrized by the same sets which parametrize involutions. A subgroup H_A corresponding to a subset A consists of involutions i_B such that $\#A \cap B$ is even. The subsets A correspond to linear functions on 2^4. If $\#A = 2$, say $A = \{0, 1\}$, we see that H_A contains the involutions $i_{01}, i_{01ab}, i_{cd}, c, d \neq 0, 1$. Adding up the traces, we obtain that these subgroups are minimal.

If $\#A = 4$, say $A = \{1, 2, 3, 4\}$, the subgroup H_A consists of i_{1234} and i_{ab}, where $a, b \neq 0$. Adding up the traces, we obtain that H_A is not minimal.

Since 2^4 contains a minimal subgroup, it is minimal itself.

Now suppose that the image G' of G in S_5 is non-trivial. The subgroup S_5 of $\mathrm{Aut}(S)$ can be realized as the stabilizer of a set of 5 skew lines on S (corresponding to the basis vectors $e_1, \ldots, e_5$). Thus any subgroup H of S_5 realized as a group of automorphisms of S is isomorphic to a group of projective transformations of $\mathbb{P}^2$ leaving invariant a set of 5 points. Since there is a unique conic through these points, the group is isomorphic to a finite group of $\mathrm{PGL}(2)$ leaving invariant a set of 5 distinct points. In Section 4, we listed all possible subgroups of $\mathrm{GL}(2)$, and in section 5 we described their relative invariants. It follows that a subgroup leaves invariant a set of 5 distinct points if and only if it is one of the following groups $C_2, C_3, C_4, C_5, S_3, D_{10}$. The corresponding binary forms of degree 5 are projectively equivalent to the following binary forms:

- $C_2 : t_0(t_0^2 - t_1^2)(t_0^2 + at_1^2),\ a \neq -1, 0, 1;$
- $C_4 : t_0(t_0^2 - t_1^2)(t_0^2 + t_1^2);$
- $C_3, S_3 : t_0t_1(t_0 - t_1)(t_0 - \epsilon_3 t_1)(t_0 - \epsilon_3^2 t_1);$
- $C_5, D_{10} : (t_0 - t_1)(t_0 - \epsilon_5 t_1)(t_0 - \epsilon_5^2 t_1)(t_0 - \epsilon_5^3 t_1)(t_0 - \epsilon_5^4 t_1).$

The corresponding surfaces are projectively equivalent to the following surfaces

$$C_2 : x_0^2 + x_1^2 + x_2^2 + x_3^2 + x_4^2 = x_0^2 + ax_1^2 - x_2^2 - ax_3^2 = 0,\ a \neq -1, 0, 1, \tag{6.3}$$

$$C_4 : x_0^2 + x_1^2 + x_2^2 + x_3^2 + x_4^2 = x_0^2 + ix_1^2 - x_2^2 - ix_3^2 = 0, \tag{6.4}$$

$$S_3 : x_0^2 + \epsilon_3 x_1^2 + \epsilon_3^2 x_2^2 + x_3^2 = x_0^2 + \epsilon_3^2 x_1^2 + \epsilon_3 x_2^2 + x_4^2 = 0, \tag{6.5}$$

$$D_{10} : \sum_{i=0}^{4} \epsilon_5^i x_i^2 = \sum_{i=0}^{4} \epsilon_5^{4-i} x_i^2 = 0. \tag{6.6}$$

Remark 6.8. Note that equations (6.4), (6.5) and (6.6) are specializations of equation (6.3) . It is obvious for equation (6.4) where we have to take $a = i$. Equation (6.3) specializes to equation (6.5) when we take $a = \pm\frac{1}{\sqrt{-3}}$ (use that the Möbius transformation of order 3 $x \mapsto \frac{ax+1}{x+a}$ permutes cyclically $\infty, a, -a$ and fixes $1, -1$). Equation (6.3) specializes to equation (6.6) if we take $a = -2 \pm \sqrt{5}$ (use that the Möbius transformation $x \mapsto \frac{x+2a-1}{x+1}$ permutes cyclically $(\infty, 1, a, -a, -1)$). We thank J. Blanc for this observation.

Since the subgroup S_5 leaves the class e_0 invariant, it remains to consider subgroups G of $2^4 : S_5$ which are not contained in 2^4 and not conjugate to a subgroup of S_5. We use the following facts.

1) Suppose G contains a minimal subgroup of 2^4. Then G is minimal.

2) Let $\overline{G}$ be the image of G in S_5. Then it is a subgroup of one of the groups listed above.

3) The group $W(D_5)$ is isomorphic to the group of transformations of $\mathbb{R}^5$ which consists of permutations of coordinates and changing even number of signs of the coordinates. Each element $w \in W(D_5)$ defines a permutation of the coordinate lines which can be written as a composition of cycles $(i_1 \ldots i_k)$. If w changes signs of even number of the coordinates $x_{i_1}, \ldots, x_{i_k}$, the cycle is called positive. Otherwise it is called a negative cycle. The conjugacy class of w is determined by the lengths of positive and negative cycles, except when all cycles of even length and positive in which case there are two conjugacy classes. The latter case does not occur in the case when n is odd. Assign to a positive cycle of length k the Carter graph A_{k-1}. Assign to a pair of negative cycles of lengths $i \geq j$ the Carter graph of type D_{i+1} if $j = 1$ and $D_{i+j}(a_{j-1})$ if $j > 1$. Each conjugacy class is defined by the sum of the graphs. We identify D_2 with $2A_1$, and D_3 with A_3. In Table 2 below we give the conjugacy classes of elements in $W(D_5)$, their characteristic polynomials and the traces in the root lattice of type D_5.

In the following, $\overline{G}$ denotes the image of G in $K = W(D_5)/2^4 \cong S_5$.

Case 1. $\overline{G} \cong C_2$.

It follows from the description of the image of $\mathrm{Aut}(S)$ in $W(D_5)$ given in Corollary 6.7 that $\overline{G}$ is generated by the permutation $s = (02)(13)$. Let $g \notin G \cap 2^4$. Then $g = s$ or $g = si_A$ for some A. It follows from Table 3 that g is either of type $2A_1$, or of type $A_1 + A_3$, or of type $D_4(a_1)$. Let $K = G \cap 2^4 \cong 2^a$. If $a = 0$ or 1, the group is not minimal.

$a = 2$.

Suppose first that s acts identically on K. Then the group is commutative isomorphic to 2^3 if it does not contain elements of order 4, and 2×4 otherwise. In the first case,

$$K = F := \langle i_{02}, i_{13} \rangle. \tag{6.7}$$

is the subspace of fixed points of s in 2^4. Since $(si_A)^2 = i_{A+s(A)}$, we see that $G \setminus K = \{si_A, A \in K\}$. Consulting Table 3, we compute the traces of all

Table 3. Conjugacy classes in $W(D_5)$.

Order	Notation	Characteristic polynomial	Trace	Representatives
2	A_1	$t+1$	3	(ab)
2	$2A_1$	$(t+1)^2$	1	$(ab)(cd), (ab)(cd)i_{abcd}$
2	$2A_1^*$	$(t+1)^2$	1	$(ab)(cd), (ab)(cd)i_{ab}$
2	$3A_1$	$(t+1)^3$	-1	$(ab)i_{cd}$
2	$4A_1$	$(t+1)^4$	-3	i_{abcd}
3	A_2	t^2+t+1	2	$(abc), (abc)i_{ab}$
4	A_3	t^3+t^2+t+1	1	$(abcd), (abcd)i_{ab}, (abcd)i_{abcd}$
4	A_1+A_3	$(t^3+t^2+t+1)(t+1)$	-1	$(ab)(cd)i_{ae}$
4	$D_4(a_1)$	$(t^2+1)^2$	1	$(ab)(cd)i_{ac}$
5	A_4	$(t^4+t^3+t^2+t+1)$	0	$(abcde), (abcde)i_A$
6	A_2+A_1	$(t^2+t+1)(t+1)$	0	$(ab)(cde)$
6	A_2+2A_1	$(t^2+t+1)(t+1)^2$	-2	$(abc)i_{abde}, (abc)i_{de}$
6	D_4	$(t^3+1)(t+1)$	0	$(abc)i_{abce}$
8	D_5	$(t^4+1)(t+1)$	-1	$(abcd)i_{abce}, (abcd)i_{de}$
12	$D_5(a_1)$	$(t^3+1)(t^2+1)$	0	$(abc)(de)i_{ac}$

elements from G to conclude that the total sum is equal to 8. Thus the group is not minimal.

In the second case G contains 4 elements of order 4 of the form si_A, where $A \notin K$. Suppose $A + s(A) = \{0, 1, 2, 3\}$. Since K is a subspace of the second type, it contributes 4 to the total sum of the traces. Thus the sum of the traces of the elements of order 4 must be equal to -4. In other words they have to be elements with trace -1 of the form si_A, where $\#A = 2, 4 \in A$. This gives the unique conjugacy class of a minimal group isomorphic to 2×4. It is represented by the group $G = \langle K, si_{04}\rangle$.

Assume now that G is non-abelian, obviously isomorphic to D_8. The subspace K contains one element from the set F. The nontrivial coset contains 2 elements of order 4 and two elements of order 2. Suppose K is of the second type with the sum of the traces of its elements equal to 4. Two elements of order 2 in $G \setminus K$ have the trace equal to 1. Elements of order 4 have the trace equal to 1 or -1. So the group cannot be minimal. Thus K must be of the third type, the minimal one. This gives us the minimal group conjugate to the subgroup $G = \langle i_{1234}, i_{02}, si_{04}\rangle$ isomorphic to D_8.

$a = 3$.

There are three s invariant subspaces of 2^4 of dimension 3. Their orthogonal complements are spanned by the one of the vectors in the set (6.7). As we saw earlier, if $K^\perp = \langle i_A\rangle$, where $\#A = 2$, the subspace K is a minimal group. Otherwise, the total sum of the traces of elements from K is equal to 8. In the first case we may assume that $K = \langle i_{14}, i_{34}, i_{02}\rangle$. All elements of order 2 in the nontrivial coset have the trace equal to 1. Thus we must have elements of order 4 in the coset with trace -1. Let si_A be such an element, where we may assume that $A = \{0, 4\}$. Thus $G = \langle K, si_{04}\rangle$. Its nontrivial coset has 4 elements of order 4 with trace 1, and four elements of order 4 with trace -1. The group is minimal. It is a non-split extension $(2^3)^\bullet 2$. Its center is isomorphic to 2^2. The classification of groups of order 16 from Table 1 shows that this is group is isomorphic to L_{16}.

$a = 4$.

In this case $G = 2^4 : 2$, where the extension is defined by the action of s in 2^4. The group has 2-dimensional center with the quotient isomorphic to 2^3.

Case 2. $\overline{G} \cong C_3$.

We may assume that $\overline{G} = \langle s \rangle$, where $s = (012)$. Applying Lemma 4.2, we obtain that G is a split extension $K : 3$, where $K = G \cap 2^4 \cong 2^a$. Since there are no minimal elements of order 3, we must have $a > 0$. If $a = 1$, the group is $2 : 3 \cong 6$. There are no minimal elements of order 6, so we may assume that $a > 1$.

Assume $a = 2$. The group is abelian $2^2 \times 3$ or non-abelian $2^2 : 3 \cong A_4$. In the first case, $K = \langle i_{0123}, i_{0124} \rangle$ is the subspace of the third type, the minimal one. Thus the total number of elements in the nontrivial cosets is equal to 0. An element of order 3 has the trace equal to 1. An element of order 6 has trace equal to -2 of 0. So we must have an element of order 6 with trace -2. It must be equal to si_{34}. Its cube is $i_{34} \in K$. So we get one conjugacy class of a minimal group isomorphic to $2^2 \times 3$. It is equal to $\langle K, si_{34} \rangle$.

If $G \cong A_4$, the subspace K is not minimal. The group does not contain elements of order 6. So the traces of all elements not from K are positive. This shows that the group cannot be minimal.

Assume $a = 3$. The subspace K is minimal if and only if its orthogonal complement is generated by i_{34}. Again the group must contain si_{34} with trace -2 and hence equal to $\langle K, si_{34} \rangle = \langle K, s \rangle$. The group G is minimal and is isomorphic to $2 \times (2^2 : 3) \cong 2 \times A_4$.

Finally, assume that $K = \langle i_{0123} \rangle^\perp$ is not minimal. We have computed earlier the sum of the traces of its elements. It is equal to 8. Again it must contain an element of order 6 equals si_{0123}. Since K contains i_{0123}, the group contains s. Now we can add all the traces and conclude that the group is not minimal.

Of course we should not forget the minimal group $2^4 : 3$.

Case 3. $\overline{G} \cong S_3$.

The group $\overline{G}$ is generated by the permutations of coordinates (012) and (12)(34). It is immediately checked that $H = G \cap 2^4$ is not trivial for minimal G. The only subgroup of H invariant with respect to the conjugation action of $\overline{G}$ on H is H itself. This gives a minimal group isomorphic to $2^4 : S_3$. The extension is defined by the restriction to S_3 of the natural action of $S_4 = W(A_4)$ on its root lattice modulo 2.

Case 4. $\overline{G} \cong C_4$.

The group $2^4 : 4$ contains $2^4 : 2$, so all minimal groups of the latter group are minimal subgroups of $2^4 : 4$. Without loss of generality, we may assume that the group $\overline{G}$ is generated by the permutations of coordinates $s = (0123)$.

The only proper subgroup of 2^4 invariant with respect to the conjugation action of $\overline{G}$ on $K = G \cap 2^4$ is either $\langle i_{0123} \rangle$ or its orthogonal complement. In the first case $G \cong 2\bullet 4 \cong 2 \times 4$ or 8. In the first case the group is not minimal. In the second case $G = \langle (0123) i_{0123} \rangle$ is minimal.

Assume $K = \langle i_{0123} \rangle^{\perp}$. If $s \notin G$, then $si_A \in G$, where $A \notin K$. The Table 3 shows that all such elements are minimal of order 8. This gives a minimal group $G \cong 2^3 : 4 = 2^2 : 8$.

Next we have to consider the case when $s \in G$ so that $G = \langle K, s \rangle$. The total number of traces of elements from K is equal to 8. Consulting the Table 3 we obtain that the elements in the cosets sK, s^2K, s^3K have the trace equal to 1. So the group is not minimal.

Our last minimal group in this case is $2^4 : 4$.

Case 5. $\overline{G} = C_5$ or D_{10}.

Again, we check using the table of conjugacy classes that no group isomorphic to D_{10} is minimal. Also, no proper subgroup of H is invariant with respect to conjugation by a permutation of order 5, or by a subgroup of S_5 generated by (012) and (12)(34). Thus we get two minimal groups isomorphic to $2^4 : 5$ or $2^4 : D_{10}$.

The following theorem summarizes what we have found.

Theorem 6.9. *Let (S, G) be a minimal Del Pezzo surface of degree $d = 4$. Then G is isomorphic to one of the following groups:*

1. $\mathrm{Aut}(S) \cong 2^4$*:*
$$2^4,\ 2^3,\ 2^2.$$
2. $\mathrm{Aut}(S) \cong 2^4 : 2$*:*
$$2 \times 4,\ D_8,\ L_{16},\ 2^4 : 2,$$
and from the previous case.

3. $\mathrm{Aut}(S) \cong 2^4 : 4$*:*
$$8,\ 2^2 : 8,\ 2^4 : 4,$$
and from the previous two cases.

4. $\mathrm{Aut}(S) \cong 2^4 : S_3$*:*
$$2^2 \times 3,\ 2 \times A_4,\quad 2^4 : 3,\ 2^4 : S_3,$$
and from Cases (1) and (2).

5. $\mathrm{Aut}(S) \cong 2^4 : D_{10}$
$$2^4 : D_{10},\ 2^4 : 5,$$
and from Cases (1) and (2).

6.5 Cubic surfaces

The following theorem gives the classification of cyclic subgroups of $\mathrm{Aut}(S)$ and identifies the conjugacy classes of their generators.

Theorem 6.10. *Let S be a nonsingular cubic surface admitting a non-trivial automorphism g of order n. Then S is equivariantly isomorphic to one of the following surfaces $V(F)$ with*

$$g = [t_0, \epsilon_n^a t_1, \epsilon_n^b t_2, \epsilon_n^c t_3]. \tag{6.8}$$

- $4A_1$ $(n = 2)$, $(a, b, c) = (0, 0, 1)$,

$$F = T_3^2 L_1(T_0, T_1, T_2) + T_0^3 + T_1^3 + T_2^3 + \alpha T_0 T_1 T_2.$$

- $2A_1$ $(n = 2)$, $(a, b, c) = (0, 1, 1)$,

$$F = T_0 T_2(T_2 + \alpha T_3) + T_1 T_3(T_2 + \beta T_3) + T_0^3 + T_1^3.$$

- $3A_2$ $(n = 3)$, $(a, b, c) = (0, 0, 1)$,

$$F = T_0^3 + T_1^3 + T_2^3 + T_3^3 + \alpha T_0 T_1 T_2.$$

- A_2 $(n = 3)$, $(a, b, c) = (0, 1, 1)$,

$$F = T_0^3 + T_1^3 + T_2^3 + T_3^3.$$

- $2A_2$ $(n = 3)$, $(a, b, c) = (0, 1, 2)$,

$$F = T_0^3 + T_1^3 + T_2 T_3(T_0 + a T_1) + T_2^3 + T_3^3.$$

- $D_4(a_1)$ $(n = 4)$, $(a, b, c) = (0, 2, 1)$,

$$F = T_3^2 T_2 + L_3(T_0, T_1) + T_2^2(T_0 + \alpha T_1).$$

- $A_3 + A_1$ $(n = 4)$, $(a, b, c) = (2, 1, 3)$,

$$F = T_0^3 + T_0 T_1^2 + T_1 T_3^2 + T_1 T_2^2.$$

- A_4 $(n = 5)$, $(a, b, c) = (4, 1, 2)$,

$$F = T_0^2 T_1 + T_1^2 T_2 + T_2^2 T_3 + T_3^2 T_0.$$

- $E_6(a_2)$ $(n = 6)$, $(a, b, c) = (0, 3, 2)$,

$$F = T_0^3 + T_1^3 + T_3^3 + T_2^2(\alpha T_0 + T_1).$$

- D_4 $(n = 6)$, $(a, b, c) = (0, 2, 5)$,

$$F = L_3(T_0, T_1) + T_3^2 T_2 + T_2^3.$$

- $A_5 + A_1$ $(n = 6)$, $(a, b, c) = (4, 2, 1)$,

$$F = T_3^2 T_1 + T_0^3 + T_1^3 + T_2^3 + \lambda T_0 T_1 T_2.$$

- $2A_1 + A_2$ $(n = 6)$, $(a, b, c) = (4, 1, 3)$,

$$F = T_0^3 + \beta T_0 T_3^2 + T_2^2 T_1 + T_1^3.$$

- D_5 $(n = 8)$, $(a, b, c) = (4, 3, 2)$,

$$F = T_3^2 T_1 + T_2^2 T_3 + T_0 T_1^2 + T_0^3.$$

- $E_6(a_1)$ $(n = 9)$, $(a, b, c) = (4, 1, 7)$,

$$F = T_3^2 T_1 + T_1^2 T_2 + T_2^2 T_3 + T_0^3.$$

- E_6 $(n = 12)$, $(a, b, c) = (4, 1, 10)$,

$$F = T_3^2 T_1 + T_2^2 T_3 + T_0^3 + T_1^3.$$

We only sketch a proof, referring for the details to [27]. Let g be a nontrivial projective automorphism of $S = V(F)$ of order n. All possible values of n can be obtained from the classification of conjugacy classes of $W(E_6)$. Choose coordinates to assume that g acts as in (6.8). Then F is a sum of monomials which belong to the same eigensubspace of g in its action in the space of cubic polynomials. We list all possible eigensubspaces. Since $V(F)$ is nonsingular, the square or the cube of each variable divides some monomial entering in F. This allows one to list all possible nonsingular $V(F)$ admitting an automorphism g. Some obvious linear change of variables allows one to find normal forms. Finally, we determine the conjugacy class by using the trace formula (6.1) applied to the locus of fixed points of g and its powers.

The conjugacy class labeled by the Carter graph with 6 vertices defines a minimal cyclic group.

Corollary 6.11. *The following conjugacy classes define minimal cyclic groups of automorphisms of a cubic surface S.*

- $3A_2$ *of order* 3,
- $E_6(a_2)$ *of order* 6,
- $A_5 + A_1$ *of order* 6,
- $E_6(a_1)$ *of order* 9,
- E_6 *of order* 12.

Next we find all possible automorphism groups of nonsingular cubic surfaces. Using a normal form of a cubic admitting a cyclic group of automorphisms from given conjugacy class, we determine all other possible symmetries of the equation. We refer for the details to [27]. The list of possible automorphism groups of cubic surfaces is given in Table 4.

Table 4. Groups of automorphisms of cubic surfaces.

Type	Order	Structure	$F(T_0,T_1,T_2,T_3)$	Parameters
I	648	$3^3 : S_4$	$T_0^3+T_1^3+T_2^3+T_3^3$	
II	120	S_5	$T_0^2T_1+T_0T_2^2+T_2T_3^2+T_3T_1^2$	
III	108	$H_3(3):4$	$T_0^3+T_1^3+T_2^3+T_3^3+6aT_1T_2T_3$	$20a^3+8a^6=1$
IV	54	$H_3(3):2$	$T_0^3+T_1^3+T_2^3+T_3^3+6aT_1T_2T_3$	$a-a^4\neq 0,$ $8a^3\neq -1,$ $20a^3+8a^6\neq 1$
V	24	S_4	$T_0^3+T_0(T_1^2+T_2^2+T_3^2)$ $+aT_1T_2T_3$	$9a^3\neq 8a$ $8a^3\neq -1,$
VI	12	$S_3\times 2$	$T_2^3+T_3^3+aT_2T_3(T_0+T_1)+T_0^3+T_1^3$	$a\neq 0$
VII	8	8	$T_3^2T_2+T_2^2T_1+T_0^3+T_0T_1^2$	
VIII	6	S_3	$T_2^3+T_3^3+aT_2T_3(T_0+bT_1)+T_0^3+T_1^3$	$a\neq 0, b\neq 0,1$
IX	4	4	$T_3^2T_2+T_2^2T_1+T_0^3+T_0T_1^2+aT_1^3$	$a\neq 0$
X	4	2^2	$T_0^2(T_1+T_2+aT_3)+T_1^3+T_2^3$ $+T_3^3+6bT_1T_2T_3$	$8b^3\neq -1$
XI	2	2	$T_1^3+T_2^3+T_3^3+6aT_1T_2T_3$ $+T_0^2(T_1+bT_2+cT_3)$	$b^3,c^3\neq 1$ $b^3\neq c^3$ $8a^3\neq -1,$

Remark 6.12. Note that there are various ways to write the equation of cubic surfaces from the table. For example, using the identity

$$(x+y+z)^3+\epsilon_3(x+\epsilon_3 y+\epsilon_3^2 z)^3+\epsilon_3^2(x+\epsilon_3^2 y+\epsilon_3 z)^3=9(x^2z+y^2x+z^2y)$$

we see that the Fermat cubic can be given by the equation

$$T_0^3+T_1^2T_3+T_3^2T_2+T_2^2T_1=0.$$

Using Theorem 6.10, this exhibits a symmetry of order 9 of the surface, whose existence is not obvious in the original equation.

Using the Hesse form of an equation of a nonsingular plane cubic curve we see that a surface with equation

$$T_0^3+F_3(T_1,T_2,T_3)=0$$

is projectively equivalent to a surface with the equation

$$T_0^3+T_1^3+T_2^3+T_3^3+6aT_0T_1T_2=0.$$

The special values of the parameters $a=0,1,\epsilon_3,\epsilon_3^2$ give the Fermat cubic. The values a satisfying $20a^3+8a^6=1$ give a plane cubic with an automorphism of order 4 (a harmonic cubic). Since a harmonic cubic is isomorphic to the cubic with equation $T_1^3+T_1T_2^2+T_3^3=0$, using Theorem 6.10 we see symmetries of order 6 from the conjugacy class $E_6(a_2)$ for surfaces of type III,IV and of order 12 for the surface

$$T_3^2T_1+T_2^2T_3+T_0^3+T_1^3=0$$

of type III.

It remains to classify minimal groups G. Note that if G is realized as a group of projective (or weighted projective) automorphisms of a family of surfaces (S_t), then G is a subgroup of the group of projective automorphisms of any surface S_{t_0} corresponding to a special value t_0 of the parameters. We indicate this by writing $S' \to S$. The types of S' when it happens are

$$\mathrm{IV} \to \mathrm{III},\ \mathrm{IV} \to \mathrm{I}, \quad \mathrm{VI}, \mathrm{VIII}, \mathrm{IX} \to \mathrm{I}, \quad \mathrm{XI} \to \mathrm{X}.$$

So it suffices to consider the surfaces of types I, II, III, V, VII, X.

We will be using the following lemma, kindly communicated to us by R. Griess. For completeness sake, we provide its proof.

Lemma 6.13. *Let S_{n+1} act naturally on its root lattice $\mathcal{R}_n = \{(a_1, \ldots, a_{n+1}) \in \mathbb{Z}^{n+1} : a_1 + \cdots + a_{n+1} = 0\}$. Let $\mathcal{R}_n(p) \cong \mathbb{F}_p^n$ be the reduction of $\mathcal{R}_n$ modulo a prime number $p > 2$ not dividing $n+1$. Then the set of conjugacy classes of subgroups dividing a splitting of $\mathbb{F}_p^n : S_{n+1}$ is bijective to the set $\mathbb{F}_p$ if $p \mid n+1$ and consists of one element otherwise.*

Proof. It is easy to see that, fixing a splitting, there is a natural bijection between conjugacy classes of splitting subgroups in $A : B$ and the cohomology set $H^1(B, A)$, where B acts on A via the homomorphism $B \to \mathrm{Aut}(A)$ defining the semi-direct product. So, it suffices to prove that $H^1(S_{n+1}, \mathcal{R}_n(p)) \cong \mathbb{F}_p$ if $p \mid n+1$ and zero otherwise. Consider the permutation representation of S_{n+1} on $M = \mathbb{F}_p^{n+1}$. We have an exact sequence of S_{n+1}-modules

$$0 \to \mathcal{R}_n(p) \to M \to \mathbb{F}_p \to 0$$

defined by the map $(a_1, \ldots, a_{n+1}) \to a_1 + \cdots + a_{n+1}$. The module M is the induced module of the trivial representation of the subgroup S_n of S_{n+1}. By Eckmann-Shapiro's Lemma,

$$H^1(S_{n+1}, M) = H^1(S_n, \mathbb{F}_p) = \mathrm{Hom}(S_n, \mathbb{F}_p).$$

Suppose $p \nmid n+1$, then the exact sequence splits, and we get

$$0 = H^1(S_{n+1}, M) = H^1(S_{n+1}, \mathcal{R}_n(p)) \oplus H^1(S_{n+1}, \mathbb{F}_p).$$

Since $H^1(S_{n+1}, \mathbb{F}_p) = \mathrm{Hom}(S_{n+1}, \mathbb{F}_p) = 0$, we get the result. If $p \mid n+1$, then $H^0(S_{n+1}, M) = 0, H^0(S_{n+1}, \mathbb{F}_p) = \mathbb{F}_p$ and the exact sequence of cohomology gives the desired result.

Type I.

Let us first classify $\mathbb{F}_3$-subspaces of the group $K = 3^3$. We view 3^3 as the S_4-module $\mathcal{R}_3(3) \cong \mathbb{F}_3^3$ from the previous lemma. We denote the image of a vector (a, b, c, d) in K by $[a, b, c, d]$. In our old notations

$$[a, b, c, d] = [\epsilon^a t_0, \epsilon^b t_1, \epsilon^c t_2, \epsilon^d t_3].$$

There are 13 ($= \#\mathbb{P}^2(\mathbb{F}_3)$) one-dimensional subspaces in 3^3. The group S_4 acts on this set with 3 orbits. They are represented by vectors $[1,2,0,0], [1,1,1,0], [1,1,2,2]$ with respective stabilizer subgroups $2^2, S_3$ and D_8. We call them lines of type I, II, III, respectively. As subgroups they are cyclic groups of order 3 of the following types.

$$\begin{cases} 2A_2 & \text{Type I,} \\ 3A_2 & \text{Type II,} \\ A_2 & \text{Type III.} \end{cases}$$

The conjugacy class of a 2-dimensional subspace K is determined by its orthogonal complement in 3^3 with respect to the dot-product pairing on $\mathbb{F}_3^4$. Thus we have 3 types of 2-dimensional subspaces of types determined by the type of its orthogonal complement.

An easy computation gives the following table.

Type	$3A_2$	$2A_2$	A_2	Trace
I	4	2	2	0
II	2	6	0	0
III	0	4	4	2

(6.9)

Here we list the types of the nontrivial elements in the subspace and the last column is the sum $\frac{1}{9}\sum_{g\in L}\mathrm{Tr}(g|\mathcal{R}_S)$. This gives us two conjugacy classes of minimal subgroups isomorphic to 3^2.

Let G be a subgroup of $\mathrm{Aut}(S)$, $\overline{G}$ be its image in S_4 and $K = G\cap 3^3$. Let $k = \dim_{\mathbb{F}_3} K$.

Case 1: $k = 0$.

In this case G defines a splitting of the projection $3^3 : S_4 \to S_4$. Assume $G \cong S_4$. It follows from Lemma 6.13 that there are 3 conjugacy classes of subgroups isomorphic to S_4 which define a splitting. Let us show each of them is minimal.

We start with the standard S_4 generated by permutations of the coordinates. It contains 6 elements of type $4A_1$, 8 elements of type $2A_2$, three elements of type $2A_1$, and 6 elements of type $A_3 + A_1$. Adding up the traces, we obtain that the group is minimal.

Suppose G is another subgroup isomorphic but not conjugate to the previous S_4. It corresponds to a 1-cocylce $\phi : S_4 \to \mathbb{F}_3^3$ defined by a vector $v = (a_1, a_2, a_3, a_4)$ with $a_1 + a_2 + a_3 + a_4$

$$\phi_v(\sigma) = (a_1, a_2, a_3, a_4) - (a_{\sigma(1)}, a_{\sigma(2)}, a_{\sigma(3)}, a_{\sigma(4)}) = (a, b, c, d).$$

The cohomology class of this cocycle depends only on the sum of the coordinates of the vector v. Without loss of generality we may choose $v = (1,0,0,0)$ and drop the subscript in ϕ. Thus a new S_4 consists of transformation $\sigma\circ\phi(\sigma)$. We check that the type of an element $\sigma\phi(\sigma)$ is equal to the type of σ for each

$\sigma \in S_4$. This shows that the representation of a new S_4 in $\mathcal{R}_S$ has the same character as that of the old S_4. This shows that all three S_4's are minimal.

Let G be mapped isomorphically to a subgroup G' of S_4. If G' is a 2-group, then it is contained in a 2-Sylow subgroup of one of the S_4's. By the above its representation in $\mathcal{R}_S$ is the same as the restriction of the representation of S_4. A 2-Sylow subgroup of S_4 contains three elements of type $2A_1$, two element of type $4A_1$, and 2 elements of type $A_3 + A_1$. Adding up the traces we get 8. Thus the group is not minimal.

If G' is not a 2-group, then G' is either S_3 or 3. A lift of a permutation $(i23)$ is given by a matrix

$$\begin{pmatrix} 0 & 0 & \epsilon_3^a \\ \epsilon_3^b & 0 & 0 \\ 0 & \epsilon_3^c & 0 \end{pmatrix},$$

where $a + b + c \equiv 0 \mod 3$. It is immediately checked that all such matrices define an element of type $2A_2$. Adding up the traces, we see that a lift of S_3 is minimal.

Case 2: $k = 1$.

Let G' be the image of G in S_4 and $K = G \cap 3^3$. Clearly, the subspace K must be invariant with respect to the restriction of the homomorphism $S_4 \to \mathrm{Aut}(3^3)$ to G'.

Assume K is of type I, say generated by $\mathbf{n}_K = [1, 2, 0, 0]$. Then its stabilizer in S_4 is generated by $(12), (34), (123)$ isomorphic to 2^2. The conjugation by (12) sends $\mathbf{n}_K$ to $-\mathbf{n}_K$, and the conjugation by (34) fixes $\mathbf{n}_K$. Thus the group $K : 2^2$ is isomorphic $S_3 \times 2$. It is easy to check that the product $g = (34)\mathbf{n}_K$ is of order 6, of minimal type $A_5 + A_1$. Thus the groups $K : 2^2$ and its cyclic subgroup of order 6 are minimal. Also its subgroup of order 6 is minimal. Its subgroup $\langle \mathbf{n}_K, (12) \rangle \cong S_3$ contains three elements of type $4A_1$ and 2 elements of type $2A_2$. Adding up the traces we see that S_3 is also minimal. It is obviously not conjugate to S_3 from the previous case.

Assume K is of type II, say generated by $\mathbf{n}_K = [1, 1, 1, 0]$. Since $\mathbf{n}_K$ is minimal, any subgroup in this case is minimal. The stabilizer of K in S_4 is generated by $(123), (12)$ and is isomorphic to S_3. Our group G is a subgroup of $3 \bullet S_3$ with K contained in the center. There are three non-abelian groups of order 18: $D_{18}, 3 \times S_3, 3^2 : 2$. The extension in the last group is defined by the automorphism of 3^2 equal to the minus identity. In our case the image of G in S_3 acts identically on K. Since the center of D_{18} or $3^2 : 2$ is trivial, this implies that either G is a cyclic subgroup of D_{18} of order 9 or 3, or a subgroup of $S_3 \times 3$, in which case $G \cong 3, 3^2, 6, 3 \times S_3$. Note that the group 3^2 is not conjugate to a subgroup of 3^3. To realize a cyclic subgroup of order 9, it is enough to take $g = \mathbf{n}_K(234)$. Note that the Sylow subgroup of $3 \bullet S_3$ is of order 9, so all 3-subgroups of order 9 are conjugate.

Assume K is of type III, say generated by $\mathbf{n}_K = [1, 1, 2, 2]$. The stabilizer group is generated by $(12), (34), (13)(24)$ and is isomorphic to D_8. Our group G is a subgroup of $3 : D_8$. The split extension is defined by the homomorphism $D_8 \to 2$ with kernel $\langle(12),(34)\rangle \cong 2^2$. The subgroup isomorphic to D_8 is contained in a nonminimal S_4, hence is not minimal. Let $H = \langle \mathbf{n}_K, (12), (34)\rangle \cong 6 \times 2$ so that $3 : D_8 \cong H : 2$. The subgroup H contains 4 elements of type D_4, 4 elements of type A_2, 2 elements of type A_1 and one element of type $2A_1$. Adding up the traces of elements from H, we get the sum equal to 24. The nontrivial coset contains eight elements of type $A_3 + A_1$, 4 elements of type $4A_1$ and one element of type $2A_1$. Adding up the traces we get 0. Thus the group is not minimal. So this case does not reveal any minimal groups.

Case 3: $k = 2$.

The image of G in S_4 is contained in the stabilizer of the orthogonal vector $\mathbf{n}_K$. Thus G is a subgroup of one of the following groups

$$G = \begin{cases} K : 2^2 & \text{if } K \text{ is of type I,} \\ K\bullet S_3 & \text{if } K \text{ is of type II,} \\ K : D_8 & \text{if } K \text{ is of type III.} \end{cases}$$

Since K of type I or II contains a minimal element of order 3, the subgroups of $K : 2^2$ and $K\bullet S_3$ are minimal. Recall that they all contain K.

Assume K is of type I. Recall that the S_4-module 3^3 is isomorphic to the root lattice $\mathcal{R}_3(3)$ of A_3 modulo 3. A permutation σ of order 2 represented by the transposition (12) decomposes the module into the sum of one-dimensional subspaces with eigenvalues $-1, 1, 1$. So, if σ fixes $\mathbf{n}_K$ it acts on K with eigenvalues $-1, 1$. Otherwise it acts identically on K. The product (12)(34) acts with eigenvalues $(-1, -1, 1)$, so if it fixes $\mathbf{n}_K$ then it acts as the minus identity on K. In our case $\mathbf{n}_K = [1, 2, 0, 0]$ and $(12), (12)(34) \in 2^2$ fix $\mathbf{n}_K$. Accordingly, (12) acts as $(-1, 1)$ giving a subgroup $K : 2 \cong 3 \times S_3$, (12)(34) gives the subgroup $K : 2 \cong 3^2 : 2 \ncong 3^2 : 2$. Finally, (34) acts identically on K, giving the subgroup $3^2 \times 2$. So we obtain 3 subgroups of index 2 of $K : 2^2$ isomorphic to $3^2 : 2, 3^2 \times 2, S_3 \times 3$. The remaining subgroups are $K \cong 3^2$ and $K : 2^2 \cong 3^2 : 2^2$.

Assume K is of type II. Again we have to find all subgroups H of $G = K\bullet S_3$ containing K. Elements of order 2 are transpositions in S_3. They fix $\mathbf{n}_K$. Arguing as above we see that $K : 2 \cong 3 \times S_3$. An element of order 3 in S_3 fixes $\mathbf{n}_K$. Hence it acts in the orthogonal complement as an element of type A_2 in the root lattice of type A_2 modulo 3. This defines a unique non-abelian group of order 27 isomorphic to the Heisenberg group $H_3(3)$. The third group is $K\bullet S_3 \cong H_3(2) : 2 \cong 3(3^2 : 2)$.

Assume K is of type III. This time K is not minimal. Each subgroup of order 2 of type $4A_1$ of D_8 defines a subgroup $K : 2 \cong 3 \times S_3$ of $K : D_8$. It contains three elements of order 2, of type $4A_1$, six elements of order 6, of type D_4, and the elements from K. Adding up the traces we get the sum equal to 18. So the subgroup is not minimal. An element of order 2 of type $2A_1$ defines a subgroup $3^2 : 2$ not isomorphic to $3 \times S_3$. It contains nine elements of type $2A_1$, and the elements from K. Adding up the traces we get the sum equal to 36. So the group is not minimal.

Assume $G \cong K : 4$. It contains the previous group $3^2 : 2$ as a subgroup of index 2. It has nine 2-Sylow subgroups of order 4. Thus the nontrivial coset consists of 18 elements of order 4. Each element of order 4 has the trace equal tom 0. This shows that the group is not minimal.

Finally it remains to investigate the group $K : D_8$. It contains the previous group $K : 4$ as a subgroup of index 2. The sum of the traces of its elements is equal to 36. The nontrivial coset consists of the union of 4 subsets; each is equal to the set of nontrivial elements in the group of type $K : 2 \cong 3 \times S_3$. The sum of traces of elements in each subset is equal to 12. So the total sum is 72 and the group is not minimal.

Case 3: $k = 3$.

This gives the groups $3^3 : H$, where H is a subgroup of S_4 which acts on 3^3 via the restriction of the homomorphism $S_4 \to \mathrm{Aut}(3^3)$ describing the action of $W(A_3)$ on its root lattice modulo 3. We have the groups

$$3^3 : S_4,\ 3^3 : D_8,\ 3^3 : S_3,\ 3^3 : 2^2(2),\ 3^3 : 3,\ 3^3 : 4,\ 3^3 : 2(2).$$

Type II.

The surface is isomorphic to the Clebsch diagonal cubic surface in $\mathbb{P}^4$ given by the equations

$$\sum_{i=0}^{4} T_i^3 = \sum_{i=0}^{4} T_i = 0.$$

The group S_5 acts by permuting the coordinates. The orbit of the line $x_0 = x_1 + x_2 = x_3 + x_4 = 0$ consists of 15 lines. It is easy to see that the remaining 12 lines form a double-six. The lines in the double-six are described as follows.

Let ω be a primitive 5th root of unity. Let $\sigma = (a_1, \ldots, a_5)$ be a permutation of $\{0, 1, 2, 3, 4\}$. Each line ℓ_σ is the span by a pair of points $(\omega^{a_1}, \ldots, \omega^{a_5})$ and $(\omega^{-a_1}, \ldots, \omega^{-a_5})$. This gives 12 different lines. One immediately checks that $\ell_\sigma \cap \ell_{\sigma'} \neq \emptyset$ if and only if $\sigma' = \sigma \circ \tau$ for some odd permutation τ. Thus the orbit of the alternating subgroup A_5 of any line defines a set of 6 skew lines (a sixer) and therefore A_5 is not minimal. Let $\ell_1, \ldots, \ell_6$ be a sixer. It is known that the divisor classes ℓ_i, K_S span $\mathrm{Pic}(S) \otimes \mathbb{Q}$. This immediately implies that $\mathrm{Pic}(S)^{A_5}$ is spanned (over $\mathbb{Q}$) by K_S and the sum $\sum \ell_i$. Since S_5 does not leave this sum invariant, we see that S_5 is a minimal group.

A maximal subgroup of S_5 not contained in A_5 is isomorphic to S_4, or $5 : 4$, or $2 \times S_3$. The subgroups isomorphic to S_4 are conjugate so we may assume that it consists of permutations of $1, 2, 3, 4$. The group has six elements of type $4A_1$ conjugate to (12), three elements of type $2A_1$ conjugate to $(12)(34)$, eight elements of type $2A_2$ conjugate to (123), and 6 elements of type $A_3 + A_1$ conjugate to (1234). The total sum of the traces is equal to 0. So the group is minimal. This gives another, non-geometric proof of the minimality of S_5.

Consider a 2-Sylow subgroup G of S_4 isomorphic to D_8. It consists of five elements of order 2, two of type $4A_1$ (from the conjugacy class of (12)), and 3 of type $2A_1$ (from the conjugacy class of $(12)(34)$). Its cyclic subgroup of order 4 is generated by an element of type $A_3 + A_1$. Adding up the traces, we see that the subgroup is not minimal. Thus S_4 has no minimal proper subgroups.

A subgroup isomorphic to $5 : 4$ is conjugate to a subgroup generated by two cycles (01234) and (0123). Computing the traces, we find that the group is not minimal. The subgroup isomorphic to $2 \times S_3$ is conjugate to a subgroup generated by $(012), (01), (34)$. Its element of order 6 belongs to the conjugacy class D_4. So this group is different from the isomorphic group in the previous case. Computing the traces, we find that it is not minimal.

Type III.

The surface is a specialization of a surface of type IV. Recall that each nonsingular plane cubic curve is isomorphic to a member of the Hesse pencil

$$T_1^3 + T_2^3 + T_3^3 + 6aT_1T_2T_3 = 0.$$

The group of projective automorphisms leaving the pencil invariant is the Hesse group G_{216} of order 216. It is isomorphic to $3^2 : \overline{T}$. The stabilizer of a general member of the pencil is isomorphic to a non-abelian extension $3^2 : 2$. It is generated by

$$g_1 = [t_1, \epsilon_3 t_2, \epsilon_3^2 t_3],\ g_2 = [t_2, t_3, t_1],\ g_3 = [t_2, t_1, t_3].$$

The pencil contains 6 members isomorphic to a harmonic cubic. They correspond to the values of the parameters satisfying the equation $8a^6 + 20a^3 - 1{=}0$. The stabilizer of a harmonic member is the group $3^2 : 4$. The additional generator is given by the matrix

$$g_4 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & \epsilon_3 & \epsilon_3^2 \\ 1 & \epsilon_3^2 & \epsilon_3 \end{pmatrix}.$$

The pencil also contains four anharmonic cubics isomorphic to the Fermat cubic. They correspond to the parameters a satisfying the equation $a^4 - a = 0$. The stabilizer of an anharmonic member is isomorphic to $3^2 : 6$. The additional generator is given by $g_5 = [x_1, \epsilon_3 x_2, \epsilon_3 x_3]$.

All curves from the Hesse pencil have nine common inflection points. If we fix one of them, say $(1, -1, 0)$, all nonsingular members acquire a group law. The group of automorphisms generated by g_1, g_2 correspond to translations by 3-torsion points. The automorphism g_3 is the negation automorphism. The automorphism g_4 is the complex multiplication by $\sqrt{-1}$. The automorphism g_5 is the complex multiplication by $e^{2\pi i/3}$.

The Hesse group admits a central extension $3G_{216} \cong H_3(3) : \overline{T}$ realized as the complex reflection group in $\mathbb{C}^3$. It acts linearly on the variables T_1, T_2, T_3 leaving the polynomial $T_1^3 + T_2^3 + T_3^3 + 6aT_1T_2T_3$ unchanged. We denote by $\tilde{g}_i$ the automorphism of the cubic surface obtained from the automorphism g_i of the Hesse pencil by acting identically on the variable T_0. The center of the group $3G_{216}$ is generated by $c = [\tilde{g}_1, \tilde{g}_2] = [1, \epsilon_3, \epsilon_3, \epsilon_3]$. This is an element of order 3, of minimal type $3A_2$.

Now we have a complete description of the automorphism group of a surface of type IV. Any minimal subgroup of $H_3(3) : 2$ can be found among minimal subgroups of surfaces of type I. However, we have 2 non conjugate subgroups of type S_3 equal to $\langle \tilde{g}_1, \tilde{g}_3 \rangle$ and $\langle \tilde{g}_2, \tilde{g}_3 \rangle$, and two non-conjugate subgroups in $S_3 \times 3$ obtained from the previous groups by adding the central element c.

Surfaces of type III acquire additional minimal subgroups of the form $A : 4$, where A is a subgroup of $H_3(3)$. The element $\tilde{g}_4$ acts by conjugation on the subgroup $H_3(3)$ via $(\tilde{g}_1, \tilde{g}_2) \mapsto (\tilde{g}_2^2, \tilde{g}_1)$. Using $\tilde{g}_4$, we can conjugate the subgroups isomorphic to $S_3, 3 \times S_3, 3^2$. Also we get two new minimal groups $H_3(3) : 4$ and 12.

Type V.

The group $S_4 \cong 2^2 : S_3$ acts by permuting the coordinates T_1, T_2, T_3 and multiplying them by -1, leaving the monomial $T_1T_2T_3$ unchanged. To make the action explicit, we identify 2^2 with the subspace of $\mathbb{F}_2^3$ of vectors whose coordinates add up to 0. The semi-direct product corresponds to the natural action of S_3 by permuting the coordinates. Thus $g = ((a,b,c),\sigma) \in 2^2 : S_3$ acts as the transformation $[t_0, (-1)^a t_{\sigma(1)}, (-1)^b t_{\sigma(2)}, (-1)^c t_{\sigma(3)}]$. It is easy to compute the types of elements of S_4 in their action on S. The group contains three elements of type $2A_1$, six elements of type $4A_1$, eight elements of type $2A_2$ and six elements of type $A_3 + A_1$. Adding up the traces we see that the group is minimal. The subgroup S_3 is minimal. No other subgroup is minimal.

Type VII.

The automorphism group of the surface of type VII is a nonminimal cyclic group of order 8.

Type X.

The automorphism group of the surface of type X consists of the identity, two involutions of type $4A_1$, and one involution of type $2A_1$. Adding up the traces, we get that the group is not minimal.

Let us summarize our result in the following.

Theorem 6.14. *Let G be a minimal subgroup of automorphisms of a nonsingular cubic surface. Then G is isomorphic to one of the following groups.*

1. *G is a subgroup of automorphisms of a surface of type I:*

$$S_4(3),\ S_3(2),\ S_3 \times 2,\ S_3 \times 3\ (2),\ 3^2 : 2\ (2),\ 3^2 : 2^2,$$

$$H_3(3){:}2,\ H_3(3),\ 3^3{:}2\ (2),\ 3^3{:}2^2\ (2),\ 3^3{:}3,\ 3^3{:}S_3,\ 3^3{:}D_8,\ 3^3{:}S_4,\ 3^3{:}4,$$
$$3^3,\ 3^2\ (3),\ 3^2 \times 2,\ 9,\ 6\ (2),\ 3.$$

2. *G is a subgroup of automorphisms of a surface of type II:*

$$S_5, \quad S_4.$$

3. G is a subgroup of automorphisms of a surface of type III:

$$H_3(3):4,\ H_3(3):2,\ H_3(3),\ S_3\times 3,\ S_3,\ 3^2,\ 12,\ 6,\ 3.$$

4. G is a subgroup of automorphisms of a surface of type IV:

$$H_3(3):2,\ H_3(3),\ S_3\ (2),\ 3\times S_3\ (2),\ 3^2\ (2),\ 6,\ 3.$$

5. G is a subgroup of automorphisms of a surface of type V:

$$S_4,\ S_3.$$

6. G is a subgroup of automorphisms of a surface of type VI:

$$6,\quad S_3\times 2,\quad S_3.$$

7. G is a subgroup of automorphisms of a surface of type VIII:

$$S_3.$$

6.6 Automorphisms of Del Pezzo surfaces of degree 2

Recall that the center of the Weyl group $W(E_7)$ is generated by an element w_0 which acts on the root lattice as the negative of the identity. Its conjugacy class is of type A_1^7. The quotient group $W(E_7)' = W(E_7)/\langle w_0\rangle$ is isomorphic to the simple group $\mathrm{Sp}(6,\mathbb{F}_2)$. The extension $2.\mathrm{Sp}(6,\mathbb{F}_2)$ splits by the subgroup $W(E_7)^+$ equal to the kernel of the determinant homomorphism $\det : W(E_7)\to\{\pm 1\}$. Thus we have

$$W(E_7) = W(E_7)^+\times\langle w_0\rangle.$$

Let H be a subgroup of $W(E_7)'$. Denote by H^+ its lift to an isomorphic subgroup of W^+. Any other isomorphic lift of H is defined by a nontrivial homomorphism $\alpha : H\to\langle w_0\rangle\cong 2$. Its elements are the products $h\alpha(h), h\in H^+$. We denote such a lift by H_α. Thus all lifts are parametrized by the group $\mathrm{Hom}(H,\langle w_0\rangle)$ and H^+ corresponds to the trivial homomorphism. Note that $wH_\alpha w^{-1} = (w'Hw'^{-1})_\alpha$, where w' is the image of w in $W(E_7)'$. In particular, two lifts of the same group are never conjugate.

Now we apply this to our geometric situation. Let S be a Del Pezzo surface of degree 2. Recall that the map $S\to\mathbb{P}^2$ defined by $|-K_S|$ is a degree 2 cover. Its branch curve is a nonsingular curve of degree 4. It is convenient to view a Del Pezzo surface of degree 2 as a hypersurface in the weighted projective space $\mathbb{P}(1,1,1,2)$ given by an equation of degree 4

$$T_3^2 + F_4(T_0,T_1,T_2) = 0. \tag{6.10}$$

The automorphism of the cover $\gamma = [t_0,t_1,t_2,-t_3]$ defines the conjugacy class of a Geiser involution of $\mathbb{P}^2$. For any divisor class D on S we have $D+\gamma_0^*(D)\in$

$|-mK_S|$ for some integer m. This easily implies that γ^* acts as the minus identity in $\mathcal{R}_S$. Its image in the Weyl group $W(E_7)$ is the generator w_0 of its center. Thus the Geiser involution is the geometric realization of w_0.

Let $\rho : \mathrm{Aut}(S) \to W(E_7)$ be the natural injective homomorphism corresponding to a choice of a geometric basis in $\mathrm{Pic}(S)$. Denote by $\mathrm{Aut}(S)^+$ the full preimage of $W(E_7)^+$. Since $W(E_7)^+$ is a normal subgroup, this definition is independent of a choice of a geometric basis. Under the restriction homomorphism $\mathrm{Aut}(S) \to \mathrm{Aut}(B)$, the group $\mathrm{Aut}(S)^+$ is mapped isomorphically to $\mathrm{Aut}(B)$, and we obtain

$$\mathrm{Aut}(S)^+ \cong \mathrm{Aut}(S)/\langle\gamma\rangle \cong \mathrm{Aut}(B).$$

From now on we will identify any subgroup G of $\mathrm{Aut}(B)$ with a subgroup of $\mathrm{Aut}(S)$ which we call the *even lift* of G. Under the homomorphism $\rho :$ $\mathrm{Aut}(S) \to W(E_7)$, all elements of G define *even conjugacy classes*, i.e. the conjugacy classes of elements from $W(E_7)^+$. It is immediate to see that a conjugacy class is even if and only if the sum of the subscripts in its Carter graph is even. An isomorphic lift of a subgroup G to a subgroup of $\mathrm{Aut}(S)$ corresponding to some nontrivial homomorphism $G \to \langle\gamma\rangle$ (or, equivalently; to a subgroup of index 2 of G) will be called an *odd lift* of G.

The odd and even lifts of the same group are never conjugate, two minimal lifts are conjugate in $\mathrm{Aut}(S)$ if and only if the groups are conjugate in $\mathrm{Aut}(B)$. Two odd lifts of G are conjugate if and only if they correspond to conjugate subgroups of index 2 (inside of the normalizer of G in $\mathrm{Aut}(B)$).

The following simple lemma will be heavily used.

Lemma 6.15. *Let G be a subgroup of* $\mathrm{Aut}(B)$ *and H be its subgroup of index 2. Assume H is a minimal subgroup of* $\mathrm{Aut}(S)$ *(i.e., its even lift is such a subgroup). Then G is minimal in its even lift and its odd lift corresponding to H. Conversely, if G is minimal in both lifts, then H is a minimal subgroup.*

Proof. Let $\mathrm{Tr}(g)_\pm$ be the trace of $g \in G$ in the representation of G in $\mathcal{R}_S$ corresponding to the minimal (respectively odd) lift. Suppose G is minimal in both lifts. Then

$$\begin{aligned}\sum_{g\in G}\mathrm{Tr}_+(g) &= \sum_{g\in H}\mathrm{Tr}_+(g) + \sum_{g\notin H}\mathrm{Tr}_+(g) = 0,\\ \sum_{g\in G}\mathrm{Tr}_-(g) &= \sum_{g\in H}\mathrm{Tr}_-(g) + \sum_{g\notin H}\mathrm{Tr}_-(g)\\ &= \sum_{g\in H}\mathrm{Tr}_+(g) - \sum_{g\notin H}\mathrm{Tr}_+(g) = 0.\end{aligned}$$

This implies that $\sum_{g\in H}\mathrm{Tr}_+(g) = 0$, i.e. H is minimal. The converse is obviously true. $\square$

Since γ generates a minimal subgroup of automorphisms of S, any group containing γ is minimal. So we classify first subgroups of $\mathrm{Aut}(B)$ which admit minimal lifts. These will be all minimal subgroups of $\mathrm{Aut}(S)$ which do not contain the Geiser involution γ. The remaining minimal groups will be of the form $\langle\gamma\rangle \times \widetilde{G}$, where $\widetilde{G}$ is any lift of a subgroup G of $\mathrm{Aut}(B)$. Obviously, the product does not depend on the parity of the lift.

As in the case of cubic surfaces we first classify cyclic subgroups.

Lemma 6.16. *Let g be an automorphism of order $n > 1$ of a nonsingular plane quartic $C = V(F)$. Then one can choose coordinates in such a way that $g = [t_0, \epsilon_n^a t_1, \epsilon_n^b t_2]$ and F is given in the following list.*

(i) $(n = 2), (a, b) = (0, 1)$,

$$F = T_2^4 + T_2^2 L_2(T_0, T_1) + L_4(T_0, T_1).$$

(ii) $(n = 3), (a, b) = (0, 1)$,

$$F = T_2^3 L_1(T_0, T_1) + L_4(T_0, T_1).$$

(iii) $(n = 3), (a, b) = (1, 2)$,

$$F = T_0^4 + \alpha T_0^2 T_1 T_2 + T_0 T_1^3 + T_0 T_2^3 + \beta T_1^2 T_2^2.$$

(iv) $(n = 4), (a, b) = (0, 1)$,

$$F = T_2^4 + L_4(T_0, T_1).$$

(v) $(n = 4), (a, b) = (1, 2)$,

$$F = T_0^4 + T_1^4 + T_2^4 + \alpha T_0^2 T_2^2 + \beta T_0 T_1^2 T_2.$$

(vi) $(n = 6), (a, b) = (3, 2)$,

$$F = T_0^4 + T_1^4 + \alpha T_0^2 T_1^2 + T_0 T_2^3.$$

(vii) $(n = 7), (a, b) = (3, 1)$,

$$F = T_0^3 T_1 + T_1^3 T_2 + T_2^3 T_0.$$

(viii) $(n = 8), (a, b) = (3, 7)$,

$$F = T_0^4 + T_1^3 T_2 + T_1 T_2^3.$$

(ix) $(n = 9), (a, b) = (3, 2)$,

$$F = T_0^4 + T_0 T_1^3 + T_2^3 T_1.$$

(x) $(n = 12), (a, b) = (3, 4)$,

$$F = T_0^4 + T_1^4 + T_0 T_2^3.$$

Here the subscript in polynomial L_i indicates its degree.

Also observe that the diagonal matrix (t, t, t, t^2) acts identically on S.

Let $g \in \mathrm{Aut}(B)$ be an element of order n of type $(*)$ from the previous Lemma. The following Table identifies the conjugacy class of two lifts $\tilde{g}$ of g in the Weyl group $W(E_7)$. If n is even, then g admits two lifts in $\mathrm{Aut}(S)$ of order n. If n is odd, then one of the lifts is of order n and another is of order $2n$. We denote by $(*)_+$ the conjugacy class of the lift which is represented by an element from $W(E_7)^+$ (of order n if n is odd). The conjugacy class of another lift is denoted by $(*)_-$. The last column of the Table gives the trace of g on $\mathcal{R}_S$.

Table 5. Conjugacy classes of automorphisms of a Del Pezzo surface of degree 2.

Type	Order	Notation	Trace
$(0)_-$	2	$7A_1$	-7
$(i)_+$	2	$4A_1$	-1
$(i)_-$	2	$3A_1$	1
$(ii)_+$	3	$3A_2$	-2
$(ii)_-$	6	$E_7(a_4)$	2
$(iii)_+$	3	$2A_2$	1
$(iii)_-$	6	$D_6(a_2)+A_1$	-1
$(iv)_+$	4	$D_4(a_1)$	3
$(iv)_-$	4	$2A_3+A_1$	-3
$(v)_+$	4	$2A_3$	-1
$(v)_-$	4	$D_4(a_1)+A_1$	1
$(vi)_+$	6	$E_6(a_2)$	2
$(vi)_-$	6	A_2+A_5	-2
$(vii)_+$	7	A_6	0
$(vii)_-$	14	$E_7(a_1)$	0
$(viii)_+$	8	D_5+A_1	-1
$(viii)_-$	8	D_5	1
$(ix)_+$	9	$E_6(a_1)$	1
$(ix)_-$	18	E_7	-1
$(x)_+$	12	E_6	0
$(x)_-$	12	$E_7(a_2)$	0

The following is the list of elements of finite order which generate a minimal cyclic group of automorphisms. To identify the conjugacy class of a minimal lift we use the trace formula (6.1). If both lifts have the same trace, we distinguish them by computing the traces of their powers.

1. Order 2 (A_1^7) (The Geiser involution) $g = [t_0, t_1, t_2, -t_3]$,

$$F = T_3^2 + F_4(T_0, T_1, T_2).$$

2. Order 4 ($2A_3 + A_1$) $g = [t_0, t_1, it_2, t_3]$,

$$F = T_3^2 + T_2^4 + L_4(T_0, T_1).$$

3. Order 6 ($E_7(a_4)$) $g = [t_0, t_1, \epsilon_3 t_2, -t_3]$,

$$F = T_3^2 + T_2^3 L_1(T_0, T_1) + L_4(T_0, T_1).$$

4. Order 6 ($A_5 + A_2$) $g = [t_0, -t_1, \epsilon_3 t_2, -t_3]$,

$$F = T_3^2 + T_0^4 + T_1^4 + T_0 T_2^3 + a T_0^2 T_1^2.$$

5. Order 6 ($D_6(a_2) + A_1$) $g = [t_0, \epsilon_3 x_1, \epsilon_3^2 x_2, -x_3]$,

$$F = T_3^2 + T_0(T_0^3 + T_1^3 + T_2^3) + T_1 T_2(\alpha T_0^2 + \beta T_1 T_2).$$

6. Order 12 ($E_7(a_2)$) $g = [t_0, \epsilon_4 t_1, \epsilon_3 t_2, t_3]$,

$$F = T_3^2 + T_0^4 + T_1^4 + T_0 T_2^3, \ (t_0, t_1, t_2, t_3).$$

7. Order 14 ($E_7(a_1)$) $g = [t_0, \epsilon_4 t_1, \epsilon_3 t_2, t_3]$,

$$F = T_3^2 + T_0^3 T_1 + T_1^3 T_2 + T_2^3 T_0.$$

8. Order 18 (E_7) $g = [t_0, \epsilon_3 t_1, \epsilon_9^2 t_2, -t_3]$,

$$F = T_3^2 + T_0^4 + T_0 T_1^3 + T_2^3 T_1.$$

Using the information about cyclic groups of automorphisms of plane quartics, it is not hard to get the classification of possible automorphism groups (see [27]). It is given in Table 5.

Table 6. Groups of automorphisms of Del Pezzo surfaces of degree 2.

Type	Order	Structure	Equation	Parameters
I	336	$2 \times L_2(7)$	$T_3^2 + T_0^3 T_1 + T_1^3 T_2 + T_2^3 T_0$	
II	192	$2 \times (4^2 : S_3)$	$T_3^2 + T_0^4 + T_1^4 + T_2^4$	
III	96	$2 \times 4A_4$	$T_3^2 + T_2^4 + T_0^4 + aT_0^2T_1^2 + T_1^4$	$a^2 = -12$
IV	48	$2 \times S_4$	$T_3^2 + T_2^4 + T_1^4 + T_0^4 +$ $+a(T_0^2T_1^2 + T_0^2T_2^2 + T_1^2T_2^2)$	$a \neq \frac{-1 \pm \sqrt{-7}}{2}$
V	32	$2 \times AS_{16}$	$T_3^2 + T_2^4 + T_0^4 + aT_0^2T_1^2 + T_1^4$	$a^2 \neq 0, -12, 4, 36$
VI	18	18	$T_3^2 + T_0^4 + T_0 T_1^3 + T_1 T_2^3$	
VII	16	$2 \times D_8$	$T_3^2 + T_2^4 + T_0^4 + T_1^4 + aT_0^2T_1^2 + bT_2^2T_0T_1$	$a, b \neq 0$
VIII	12	2×6	$T_3^2 + T_2^3 T_0 + T_0^4 + T_1^4 + aT_0^2T_1^2$	
IX	12	$2 \times S_3$	$T_3^2 + T_2^4 + aT_2^2T_0T_1 + T_2(T_0^3 + T_1^3) + bT_0^2T_1^2$	
X	8	2^3	$T_3^2 + T_2^4 + T_1^4 + T_0^4$ $+aT_2^2T_0^2 + bT_1^2T_2^2 + cT_0^2T_1^2$	distinct $a, b, c \neq 0$
XI	6	6	$T_3^2 + T_2^3 T_0 + L_4(T_0, T_1)$	
XII	4	2^2	$T_3^2 + T_2^4 + T_2^2 L_2(T_0, T_1) + L_4(T_0, T_1)$	
XIII	2	2	$T_3^2 + F_4(T_0, T_1, T_2)$	

Next we find minimal subgroups of automorphisms of a Del Pezzo surface of degree 2.

As in the previous case it is enough to consider surfaces S' which are not specialized to surfaces S of other types. When this happens we write $S' \to S$. We have

$$\mathrm{IX} \to \mathrm{IV} \to \mathrm{I}, \mathrm{II},$$

$$\mathrm{XII} \to \mathrm{X} \to \mathrm{VII} \to \mathrm{V} \to \mathrm{II}, \mathrm{III},$$

$$\mathrm{XI} \to \mathrm{VIII} \to \mathrm{III}.$$

All of this is immediate to see, except the degeneration $VIII \to III$. This is achieved by some linear change of variables transforming the form $x^3y + y^4$ to the form $u^4 + 2\sqrt{3}iu^2v^2 + v^4$. So it suffices to consider surfaces of types I, II, III, VI.

Before we start the classification we advice the reader to go back to the beginning of the section and recall the concepts of odd and even lifts of subgroups of $\mathrm{Aut}(B)$.

Type I.

Since $L_2(7)$ has no subgroups of index 2 (in fact, it is a simple group), it admits a unique lift to a subgroup of $\mathrm{Aut}(S)$. It is known that the group $L_2(7)$ is generated by elements of order 2, 3 and 7. Consulting Table 4, we find that an element of order 2 must be of type $4A_1$, an element of order 3 must be of types $3A_2$ or $2A_2$, and element of order 7 is of type A_6. To decide the type of a generator g of order 3, we use that it acts as a cyclic permutation of the coordinates in the plane, hence has 3 fixed points $(1,1,1), (1,\eta_3,\eta_3^2), (1,\eta_3^2,\eta_3)$. The last two of them lie on the quartic. This easily implies that g has four fixed points on S, hence its trace in $\mathrm{Pic}(S)$ is equal to 2. This implies that g is of type $2A_2$. Comparing the traces with the character table of the group $L_2(7)$, we find that the representation of $L_2(7)$ in $(\mathcal{R}_S) \otimes \mathbb{C}$ is an irreducible 7-dimensional representation of $L_2(7)$. Thus the group is minimal.

Assume G is a proper subgroup of $L_2(7)$. It is known that maximal subgroups of $L_2(7)$ are isomorphic to S_4 or $7:3$. There are two conjugacy classes of subgroups isomorphic to S_4 (in the realization $L_2(7) \cong L_3(2)$ they occur as the stabilizer subgroups of a point or a line in $\mathbb{P}^2(\mathbb{F}_2)$). Since S_4 contains a unique subgroup of index 2, each subgroup can be lifted in two ways. Consider the even lift of S_4 lying in $L_2(7)$. To find the restriction of the 7-dimensional representation $V_7 = (\mathcal{R}_S)_{\mathbb{C}}$ to G we apply the Frobenius Reciprocity formula. Let χ_k denote a k-dimensional irreducible representation of $L_2(7)$ and $\bar{\chi}_k$ be its restriction to S_4. It is known that the induced character of the trivial representation of S_4 is equal to $\chi_1 + \chi_6$ (see [19]). Applying the Frobenius Reciprocity formula we get $\langle \bar{\chi}_1, \bar{\chi}_7 \rangle = \langle \chi_1 + \chi_6, \chi_7 \rangle = 0$. This computation shows that the even lifts of the two conjugacy classes of S_4 in $L_2(7)$ are minimal subgroups. It follows from Lemma 6.15 the the odd lifts are minimal only if the lift of the subgroup A_4 of S_4 is minimal. One checks that the induced character of the trivial representation of A_4 is equal to $\chi_1 + \chi_6 + \chi_7$. By the Frobenius Reciprocity formula, the restriction of V_7 to A_4 contains the trivial summand. Thus A_4 is not minimal, and we conclude that there are only two non conjugate lifts of S_4 to a minimal subgroup of $\mathrm{Aut}(S)$.

Next consider the subgroup $7:3$. It admits a unique lift. The induced representation of its trivial representation has the character equal to $\chi_1+\chi_7$. Applying the Frobenius Reciprocity formula, we see that this group is not minimal.

Let H be any subgroup of $L_2(7)$ which admits a minimal lift. Since $\mathrm{Aut}(S)$ does not contain minimal elements of order 3 or 7, H must be a subgroup of S_4. Since A_4 does not admit a minimal lift, H is either a cyclic group or isomorphic to either 2^2 or D_8. The only cyclic group which may admit a minimal lift is a cyclic group of order 4. However, the character table for $L_2(7)$ shows that the value of the character χ_7 at an element of order 4 is equal to -1; hence it is of type $2A_3$. It follows from the Table that this element does not admit minimal lifts.

Suppose $G \cong 2^2$. In the even lift, it contains 3 nontrivial elements of type $4A_1$. Adding up the traces we see that this group is not minimal. In the odd lift, it contains one element of type $4A_1$ and two of type $3A_1$. Again, we see that the group is not minimal.

Assume $G \cong D_8$. The group S_4 is the normalizer of D_8. This shows that there are two conjugacy classes of subgroups isomorphic to D_8. The group G admits 2 lifts. In the even lift it contains two elements of type $2A_3$ and five elements of type $4A_1$. Adding up the traces, we obtain that the lift is minimal. Since the lift of 4 is not minimal, the odd lift of D_8 is not minimal.

Type II.

The group $\mathrm{Aut}(B)$ is generated by the transformations

$$g_1 = [t_0, it_1, t_2, -t_3],\ \tau = [t_1, t_0, t_2, t_3],\ \sigma = [t_0, t_2, t_1, t_3]$$

of types $D_4(a_1), 4A_1, 4A_1$, respectively. Let $g_2 = \sigma g_1 \sigma^{-1} = [t_0, t_1, it_2, -t_3]$.

We have

$$\tau g_1 \tau^{-1} = g_1^{-1} g_2^{-1},\ \tau g_2 \tau^{-1} = g_2.$$

The elements g_1, g_2, γ generate a normal subgroup isomorphic to 4^2. The quotient group is isomorphic to S_3. Its generators of order 2 can be represented by τ and σ. The elements $g_1^2, g_2^2, \tau, \sigma$ generate a subgroup (not normal) isomorphic to S_4. Thus

$$\mathrm{Aut}(B) \cong 4^2 : S_3 \tag{6.11}$$

and

$$\mathrm{Aut}(S) \cong 2 \times (4^2 : S_3). \tag{6.12}$$

Consider the natural homomorphism $f : \mathrm{Aut}(B) \to S_3$ with kernel 4^2. We will consider different cases corresponding to a possible image of a subgroup $G \subset \mathrm{Aut}(B)$ in S_3. For the future use we observe that $\mathrm{Aut}(B)$ does not contain elements of order 6 because its square is an element of type $3A_2$ but all our elements of order 3 are of type $2A_2$. Also it does not contain 2^3 (this follow from the presentation of the group). We will also use that $\mathrm{Aut}(B)$ contains two conjugacy classes of elements of order 4 of types $D_4(a_1)$ (represented by g_1) and $2A_3$ (represented by g_1g_2).

Case 3: $f(G) = \{1\}$.

In this case, G is a subgroup of 4^2. The group itself contains three elements of type $4A_1$, six elements of type $D_4(a_1)$ and six elements of type $2A_3$. The sum of the traces is equal to 16. Thus the group is not minimal. So no subgroup is minimal in the even lift. An odd lift corresponding to the homomorphism $4^2 \to \langle\gamma\rangle$ sending an element of type $D_4(a_1)$ to γ defines an odd lift. There is only one conjugacy class of subgroups of index 2 in 4^2. It defines an odd lift of 4^2. We may assume that the subgroup of index 2 is generated by g_1, g_2^2. It admits two odd lifts corresponding to the subgroups $\langle g_1^2, g_2^2\rangle$ and $\langle g_1g_2^2\rangle$. Finally a cyclic subgroup 4 of type $D_4(a_1)$ admits an odd lift. No other subgroup admits a minimal lift.

Case 2: $|f(G)| = 2$.

Replacing the group by a conjugate group, we may assume that $f(G) = \langle\tau\rangle$. We have

$$G_1 = f^{-1}(\langle\tau\rangle) = \langle\tau, g_1, g_2\rangle \cong 4^2 : 2 \cong 4D_8,$$

where the center is generated by g_2.

Let $H = \langle\tau, g_1^2g_2\rangle$. One immediately checks that H is normal in G_1 and isomorphic to D_8. We have $G_1 \cong D_8 : 4$. The subgroup H consists of five elements of type $4A_1$ and two elements of type $2A_3$. Adding up the traces we obtain that H is minimal in its even lift. Thus G_1 is minimal in its even lift. The subgroup $G_2 = \langle\tau, g_1^2, g_2\rangle$ is of order 16. It contains H defining a split extension $D_8 : 2$ with center generated by g_2. It is isomorphic to the group AS_{16} (see Table 1) and is of index 2 in G_1. Since it is minimal, the odd lift of G_1 corresponding to this subgroup is minimal.

We check that τg_1 is of order 8 and the normalizer of the cyclic group $\langle\tau g_1\rangle$ is generated by this subgroup and g_1^2. This gives us another subgroup G_3 of index 2 of G_1. It is a group of order 16 isomorphic to M_{16}. An element of order 8 is of type $D_5(a_1) + A_1$. Thus the sum of the traces is equal to 8. Adding up the traces of elements in the nontrivial coset of $\langle\tau g_1\rangle$ we get that the sum is equal to -8 (all elements have the trace equal to -1). This shows that G_3 is minimal. Thus the corresponding odd lift of G_1 is minimal.

Let G be a subgroup of index 2 of G_1 and $g = \tau g_1^a g_2^b \in G$ be the element of largest possible order in H. We verify that $g^2 = g_2^{2b-a}$. If g is of order 8, we check that it generates either $\langle\tau g_1\rangle$ considered earlier or its conjugate subgroup. Its normalizer is conjugate to the subgroup G_3 considered earlier. If g is of order 4, then $2b-a \equiv 2 \mod 4$. We list all possible cases and find that all elements of order 4 are conjugate. Thus we may assume that G contains $g = \tau g_1^2$. Now we check that the normalizer of this group is our group G_2.

So, all subgroups of index 2 are accounted for. They are two of them isomorphic to AS_{16} and M_{16}. They are all minimal in their even lift, and hence define odd lifts of G_1.

Let G be a subgroup of index 4 of $f^{-1}(\tau)$. It follows from above argument that G is conjugate to a subgroup of index 2 of G_2 or G_3. It could be $D_8, 8$, or 2×4. The first group is minimal; hence $D_8 : 2$ admits an odd minimal lift. other two groups are not minimal. The last group admits an odd minimal lift. Note that it is not conjugate to odd 2×4 from Case 1. Finally a cyclic group of type $D_4(a_1)$ admits an odd minimal lift. It is not conjugate to a group from Case 1.

Case 3: $|f(G)| = 3$.

Without loss of generality we may assume that $f(G) = \langle \sigma\tau \rangle$. By Lemma 4.2, G is a split extension $H : 3$, where H is a subgroup of $\langle g_1, g_2 \rangle$. Let $G_1 = f^{-1}(\langle \sigma\tau \rangle)$. It is a split extension $4^2 : 3$. By Sylow's Theorem, all subgroups of order 3 are conjugate. Thus we may assume that H contains $\sigma\tau$. The possibilities are G_1 or $G_2 = \langle g_1^2, g_2^2, \sigma\tau \rangle \cong 2^2 : 3 \cong A_4$. The group A_4 has three elements of type $4A_1$, four elements of type $2A_2$, and four elements of type $3A_2$. Adding up the traces, we see that the group is minimal. Thus G_1 is minimal too. The group G_1 does not have subgroups of index 2, so it does not admit odd lifts. Other groups in this case are conjugate to the nonminimal group $\langle \sigma\tau \rangle$.

Case 4:$f(G) = S_3$.

In this case $G \cap f^{-1}(\langle \sigma\tau \rangle)$ is a subgroup of index 2 equal to one of the two groups considered in the previous case. We get $G = \mathrm{Aut}(B)$, or $G \cong 2^2 : S_3 \cong S_4$, or S_3. Considering the preimage of $\langle \tau \rangle$, we find that all groups isomorphic to S_4 are conjugate and their Sylow 2-subgroup is D_8 from the previous case. Thus both $\mathrm{Aut}(B)$ and S_4 admit two minimal lifts. A group isomorphic to S_3 contains two elements of type $2A_2$, and it is not minimal in any lift.

Type III.

We assume that $a = 2\sqrt{3}i$ in the equation of the surface. The group $\mathrm{Aut}(B)$ is isomorphic to $4A_4$. It is generated (as always in its even lift) by

$$g_1 = [t_1, t_0, t_2, -t_3], \quad g_2 = [it_1, -it_0, t_2, -t_3],$$

$$g_3 = [\epsilon_8^7 t_0 + \epsilon_8^7 t_1, \epsilon_8^5 t_0 + \epsilon_8 t_1, \sqrt{2}\epsilon_{12} t_2, 2\epsilon_6 t_3], \quad c = [t_0, t_1, it_2, -t_3].$$

The "complicated" transformation g_3 is of order 3 (see our list of Gründformen for binary polyhedral groups). The generators g_1, g_2 are of type $4A_1$, the generator g_3 is of type $2A_2$, and the generator c is of type $D_4(a_1)$.

The element c generates the center. We have $g_1 g_2 = g_2 g_1 c^2$. This shows that the quotient by $\langle c \rangle$ is isomorphic to A_4 and the subgroup $\langle c, g_1, g_2 \rangle \cong 4D_4$ is a group of order 16 isomorphic to the group AS_{16} (see Table 1).

Let $f : \mathrm{Aut}(B) \to A_4$ be the natural surjection with kernel $\langle c \rangle$. Let G be a subgroup of $\mathrm{Aut}(B)$.

Case 1: $G \subset \mathrm{Ker}(f) \cong 4$.

There are no even minimal subgroups. The whole kernel admits a minimal odd lift.

Case 2: $\#f(G) = 2$.

Without loss of generality we may assume that $f(G) = \langle g_1 \rangle$. The subgroup $f^{-1}(\langle g_1 \rangle)$ is generated by c, g_1 and is isomorphic to 4×2. It is not minimal in the even lift and minimal in the unique odd lift. Its subgroup $\langle cg_1 \rangle$ of order 4 is of type $2A_3$ and does not admit minimal lifts.

Case 3:$f(G) = \langle g_3 \rangle$.

We have $G = f^{-1}(\langle g_3 \rangle) = \langle c, g_3 \rangle = \langle cg_3 \rangle \cong 12$. The element cg_3 is of type E_6, hence not minimal. Its square is an element of type $E_6(a_2)$, also not minimal. The subgroup $\langle (cg_3)^2 \rangle \cong 6$ defines an odd minimal lift of G. The subgroup $\langle (cg_3)^4 \rangle$ is of order 3. It defines an odd minimal lift of $\langle (cg_3)^2 \rangle$. The group $\langle g_3 \rangle$ admits an odd minimal lift.

Case 4: $f(G) = \langle g_1, g_2 \rangle \cong 2^2$.

The subgroup $H = f^{-1}(\langle g_1, g_2 \rangle)$ is generated by c, g_1, g_2. As we observed earlier, it is isomorphic to the group AS_{16} from Table 1. A proper subgroup is conjugate in $\mathrm{Aut}(B)$ to either $\langle g_1, g_2 \rangle \cong D_8$ or $\langle cg_1, g_2 \rangle$. All of the subgroups are isomorphic to D_8 with center generated by c^2. The cyclic subgroup of order 4 is of type $2A_3$; thus the subgroups are minimal in the even lift (we have done this computation for surfaces of type II). Thus the group H is minimal in the even lift and also minimal in two odd lifts corresponding to its two subgroups of index 2.

Case 5: $f(G) = A_4$.

It is easy to see that G has non-trivial center (the center of its Sylow 2-subgroup). It is equal to $\langle c \rangle$ or $\langle c^2 \rangle$. In the first case, $G = \mathrm{Aut}(B)$. Since it contains minimal subgroups it is minimal.

A subgroup G of index 2 is isomorphic to $2A_4 \cong D_8 : 3$. Its Sylow 2-subgroup is equal to one of the two subgroups isomorphic to D_8 from Case 4. Thus $\mathrm{Aut}(B)$ admits two odd lifts. Since G has no subgroups of index 2, the odd lifts of G do not exist.

Type VI.

In this case $\mathrm{Aut}(B) \cong 9$ is not minimal so does not admit minimal lifts.

To summarize our investigation we give two lists. In the first we list all groups that do not contain the Geiser involution γ. We indicate by $+$ or $-$ the types of their lifts. Also we indicate the number of conjugacy classes.

All other minimal groups are of the form $\langle \gamma \rangle \times G$, where G is one of the lifts of a subgroup of $\mathrm{Aut}(B)$. In the second list we give only groups $2 \times G$, where G does not admit a minimal lift. All other groups are of the form $2 \times G$, where G is given in the previous table.

Theorem 6.17. *Let G be a minimal group of automorphisms of a Del Pezzo surface of degree 2. Then G is either equal to a minimal lift of a subgroup from Table 7 or equal to $\gamma \times G'$, where G' is either from the Table or is one of the following groups of automorphisms of the branch quartic curve B:*

1. *Type I:* $7 : 3,\ A_4,\ S_3,\ 7,\ 4,\ 3,\ 2.$
2. *Type II:* $2^2,\ S_3,\ 8,\ 4,\ 3,\ 2.$

Table 7. Minimal groups of automorphisms not containing γ.

Type of S	Group	Lift	Conjugacy classes
I	$L_2(7)$	$+$	1
	S_4	$+$	2
	D_8	$+$	2
II	$4^2 : S_3$	$+,-$	2
	S_4	$+,-$	2
	$4^2 : 3$	$+$	1
	A_4	$+$	1
	$4^2 : 2 \cong D_8 : 4$	$+,-,-$	3
	M_{16}	$+$	1
	AS_{16}	$+,-$	2
	D_8	$+$	1
	4^2	$-$	1
	2×4	$-$	2
	4	$-$	2
III	$4A_4$	$+,-$	2
	$2A_4 \cong D_8 : 3$	$+$	1
	AS_{16}	$+,-$	2
	D_8	$+$	1
	12	$-$	1
	6	$-$	1
	2×4	$-$	1
	4	$-$	1
IV	S_4	$+$	1
	D_8	$+$	1
V	AS_{16}	$+,-$	2
	D_8	$+$	1
	2×4	$-$	2
	4	$-$	1
VII	D_8	$+$	1
VIII	6	$-$	1

3. *Type III:* $2^2, \ 4, \ 2.$
4. *Type IV:* $S_3, \ 2^2, \ 3, \ 2.$
5. *Type V:* $2^2, \ 2.$
6. *Type VI:* $9, \ 3.$
7. *Type VII:* $2^2, \ 4, \ 2$
8. *Type VIII:* $3.$
9. *Type IX:* $S_3, \ 3, \ 2.$
10. *Type X:* $2^2, \ 2.$
11. *Type XI:* $3.$
12. *Type XII:* $\{1\}.$

6.7 Automorphisms of Del Pezzo surfaces of degree 1

Let S be a Del Pezzo surface of degree 1. The linear system $|-2K_S|$ defines a finite map of degree 2 onto a quadric cone Q in $\mathbb{P}^3$. Its branch locus is a nonsingular curve B of genus 4 cut out by a cubic surface. Recall that a singular quadric is isomorphic to the weighted projective space $\mathbb{P}(1,1,2)$. A curve of genus 4 of degree 6 cut out in Q by a cubic surface is given by equation $F(T_0,T_1,T_2)$ of degree 6. After change of coordinates it can be given by an equation $T_2^3+F_4(T_0,T_1)T_2+F_6(T_0,T_1)=0$, where F_4 and F_6 are binary forms of degree 4 and 6. The double cover of Q branched along such curve is isomorphic to a hypersurface of degree 6 in $\mathbb{P}(1,1,2,3)$

$$T_3^2+T_2^3+F_4(T_0,T_1)T_2+F_6(T_0,T_1)=0. \tag{6.13}$$

The vertex of Q has coordinates $(0,0,1)$ and its preimage in the cover consists of one point $(0,0,1,a)$, where $a^2+1=0$ (note that $(0,0,1,a)$ and $(0,0,1,-a)$ represent the same point on $\mathbb{P}(1,1,2,3)$). This is the base point of $|-K_S|$. The members of $|-K_S|$ are isomorphic to genus 1 curves with equations $y^2+x^3+F_4(t_0,t_1)x+F_6(t_0,t_1)=0$. The locus of zeros of $\Delta=4F_4^3+27F_6^2$ is the set of points in $\mathbb{P}^1$ such that the corresponding genus-1 curve is singular. It consists of a simple roots and b double roots. The zeros of F_4 either are common zeros with F_6 and Δ, or represent nonsingular elliptic curves isomorphic to an anharmonic plane cubic curve. The zeros of F_6, are either common zeros with F_4 and Δ, or represent nonsingular elliptic curves isomorphic to a harmonic plane cubic curve.

Observe that no common root of F_4 and F_6 is a multiple root of F_6, since otherwise the surface is singular.

Since the ramification curve of the cover $S\to Q$ (identified with the branch curve B) is obviously invariant with respect to $\mathrm{Aut}(S)$ we have a natural surjective homomorphism

$$\mathrm{Aut}(S)\to\mathrm{Aut}(B). \tag{6.14}$$

Its kernel is generated by the deck involution β, which we call the *Bertini involution*. It defines the Bertini involution in $\mathrm{Cr}(2)$. The Bertini involution is the analogue of the Geiser involution for Del Pezzo surfaces of degree 2. The same argument as above shows that β acts in $\mathcal{R}_S$ as the negative of the identity map. Under the homomorphism $\mathrm{Aut}(S)\to W(E_8)$ defined by a choice of a geometric basis, the image of β is the elements w_0 generating the center of $W(E_8)$. This time w_0 is an even element, i.e., belongs to $W(E_8)^+$. The quotient group $W(E_8)^+/\langle w_0\rangle$ is isomorphic to the simple group $\mathrm{O}(8,\mathbb{F}_2)^+$.

Since Q is a unique quadric cone containing B, the group $\mathrm{Aut}(B)$ is a subgroup of $\mathrm{Aut}(Q)$. Consider the natural homomorphism

$$r:\mathrm{Aut}(B)\to\mathrm{Aut}(\mathbb{P}^1).$$

Let G be a subgroup of $\mathrm{Aut}(B)$ and P be its image in $\mathrm{Aut}(\mathbb{P}^1)$. We assume that elements from G act on the variables T_0,T_1 by linear transformations with determinant 1. The polynomials F_4 and F_6 are the relative invariants of the binary group $\bar{P}$. They are polynomials in Gründformen that were listed in Section 5.5. Let χ_4, χ_6 be the corresponding characters of $\bar{P}$ defined by the binary forms F_4, F_6. Let χ_2, χ_3 be the characters of G defined by the action on the variables T_2, T_3. Assume that $F_4 \neq 0$. Then

$$\chi_4\chi_2 = \chi_6 = \chi_3^3 = \chi_3^2.$$

If $g \in G \cap \mathrm{Ker}(r) \setminus \{1\}$, then g acts on the variables T_0, T_1 by either the identity or the minus identity. Thus $\chi_4(g) = \chi_6(g) = 1$ and we must have $\chi_2(g) = \chi_3(g)^2 = 1$. This shows that $g = [t_0, t_1, t_3, -t_3] = [-t_0, -t_1, t_2, -t_3] = \beta$.

If $F_4 = 0$, then we must have only $\chi_2(g)^3 = \chi_3(g)^2 = 1$. Since $[-t_0, -t_1, t_2, -t_3]$ is the identity transformation, we may assume that $\chi_3(g) = 1$ and represent g by $g = [t_0, t_1, \epsilon_3 t_3, \pm t_3]$. Thus $G \cap \mathrm{Ker}(r) = \langle \beta, \alpha \rangle \cong 6$.

Conversely, start with a polyhedral group P such that its lift to a binary polyhedral group $\bar{P}$ acts on the variables T_0, T_1 leaving $V(F_4)$ and $V(F_6)$ invariant. Let χ_4, χ_6 be the corresponding characters. Assume that there exist character $\chi_2, \chi_3 : \bar{P} \to \mathbb{C}^*$ such that

$$\chi_0^2 = \chi_4\chi_1 = \chi_6 = \chi_1^3. \tag{6.15}$$

Then $g = [at_0 + bt_1, ct_0 + t_1] \in \bar{P}$ acts on S by $[at_0 + bt_1, ct_0 + t_1, \chi_2(g)t_2, \chi_3(t_3)]$. This transformation is the identity in $\mathrm{Aut}(S)$ if and only if $g = [-t_0, -t_1]$ and $\chi_2(-1) = 1, \chi_3(-1) = -1$. This shows that $\bar{P}$ can be identified with a subgroup of $\mathrm{Aut}(S)$ with $-I_2 = \beta$ if and only if $\chi_3(-1) = -1$. If $\chi_3(-1) = 1$, then P can be identified with a subgroup of $\mathrm{Aut}(S)$ not containing β. In the latter case,

$$r^{-1}(P) = \begin{cases} P \times \langle \beta \rangle & \text{if } F_4 \neq 0 \\ P \times \langle \beta, \alpha \rangle & \text{otherwise.} \end{cases}$$

In particular, if $F_4 = 0$, there are three subgroups of $\mathrm{Aut}(S)$ which are mapped surjectively to P.

In the former case

$$r^{-1}(P) = \begin{cases} \bar{P} & \text{if } F_4 \neq 0, \\ \bar{P} \times \langle \alpha \rangle & \text{otherwise.} \end{cases}$$

Of course it could happen that neither P nor $\bar{P}$ lifts to a subgroup of $\mathrm{Aut}(S)$. In this case $r^{-1}(P) \cong 2P \not\cong \bar{P}$ or $r^{-1}(P) \cong 3 \times 2P$ (if $F_4 = 0$).

In the following list we give a nontrivial subgroup P of $\mathrm{Aut}(\mathbb{P}^1)$ as a group of automorphisms of B and a smallest lift $\widetilde{P}$ of P to a subgroup of $r^{-1}(P)$. If $F_4 \neq 0$ then we will see that $\tilde{P} = r^{-1}(P)$ or $\tilde{P} \cong P$. In the latter case, $r^{-1}(P) \cong 2 \times P$. If $F_4 = 0$, and $\tilde{P} \cong P$, then $r^{-1}(P) \cong 6 \times \tilde{P}$. Otherwise, $r^{-1}(P) \cong 3 \times \widetilde{P}$.

Also we give generators of $\tilde{P}$ to $\mathrm{Aut}(S)$ as a group acting on t_0, t_1 with determinant 1 and the Bertini involution as an element of the lift.

1. *Cyclic groups P*

(i) $P = \{1\}, F_4 = 0, r^{-1}(P) = \langle \beta, \alpha \rangle \cong 6;$

(ii) $P \cong 2, g = [it_0, -it_1, -t_2, it_3],$

$$F_4 = F_2(T_0^2, T_1^2) \neq 0, \quad F_6 = F_3(T_0^2, T_1^2);$$

(iii) $P \cong 2, \widetilde{P} \cong 4, g = [it_0, -it_1, t_2, t_3], \beta = g^2,$

$$F_4 = a(T_0^4 + T_1^4) + bT_0^2T_1^2, \quad F_6 = T_0T_1F_2(T_0^2, T_1^2);$$

(iv) $P \cong 2, \widetilde{P} = 4, g = [-t_1, t_0, t_2, t_3], \beta = g^2,$

$$F_4 \text{ as in (iii)}, \ F_6 = (T_0^2 + T_1^2)(a(T_0^4 + T_1^4) + T_0T_1(bT_0T_1 + c(T_0^2 - T_1^2));$$

(v) $P \cong 2, g = [-t_1, t_0, -t_2, it_3],$

$$F_4 \text{ as in (iii)}, \ F_6 = a(T_0^6 - T_1^6) + bT_0T_1(T_0^4 + T_1^4);$$

(vi) $P \cong 3, g = [\epsilon_3 t_0, \epsilon_3^2 t_1, \epsilon_3^2 t_2, t_3],$

$$F_4 = T_0(aT_0^3 + bT_1^3), \quad F_6 = F_2(T_0^3, T_1^3);$$

(vii) $P \cong 3, g = [\epsilon_3 t_0, \epsilon_3^2 t_1, t_2, t_3],$

$$F_4 = aT_0^2T_1^2, \quad F_6 = F_2(T_0^3, T_1^3);$$

(viii) $P \cong 4, g = [\epsilon_8 t_0, \epsilon_8^{-1} t_1, it_2, \epsilon_8^3 t_3],$

$$F_4 = aT_0^4 + bT_1^4, \quad F_6 = T_0^2(cT_0^4 + dT_1^4);$$

(ix) $P \cong 4, \widetilde{P} \cong 8, g = [\epsilon_8 t_0, \epsilon_8^{-1} t_1, -t_2, t_3], \beta = g^4,$

$$F_4 = aT_0^2T_1^2, \quad F_6 = T_0T_1(T_0^4 + T_1^4);$$

(x) $P \cong 5, g = [\epsilon_{10} t_0, \epsilon_{10}^{-1} t_1, \epsilon_5 t_2, \epsilon_{10}^3 t_3],$

$$F_4 = aT_0^4, \quad F_6 = T_0(T_0^5 + T_1^5);$$

(xi) $P \cong 6, g = [\epsilon_{12} t_0, \epsilon_{12}^{-1} t_1, \epsilon_6 t_2, it_3],$

$$F_4 = aT_0^4, \quad F_6 = bT_0^6 + T_1^6, b \neq 0;$$

(xii) $P \cong 6, g = [\epsilon_{12} t_0, \epsilon_{12}^{-1} t_1, -t_2, it_3],$

$$F_4 = aT_0^2T_1^2, \quad F_6 = T_0^6 + T_1^6;$$

(xiii) $P \cong 10, \widetilde{P} \cong 20, g = [\epsilon_{20} t_0, \epsilon_{20}^{-1} t_1, \epsilon_{10}^8 t_2, \epsilon_{10}^{-1} t_3], g^{10} = \beta,$

$$F_4 = T_0^4, \quad F_6 = T_0T_1^5;$$

(xiv) $P \cong 12, g = [\epsilon_{24}t_0, \epsilon_{24}^{-1}t_1, \epsilon_{12}t_2, \epsilon_{24}t_3]$,

$$F_4 = T_0^4, \quad F_6 = T_1^6.$$

2. *Dihedral groups*
   (i) $P \cong 2^2, \widetilde{P} \cong D_8, g_1 = [it_1, it_0, -t_2, it_3],\ g_2 = [-t_1, t_0, -t_2, it_3], \beta = (g_1g_2)^2, g_1^2 = g_2^2 = 1$,

$$F_4 = a(T_0^4 + T_1^4) + bT_0^2T_1^2, \quad F_6 = T_0T_1[c(T_0^4 + T_1^4) + dT_0^2T_1^2];$$

   (ii) $P \cong 2^2, \widetilde{P} \cong Q_8, g_1 = [it_1, it_0, t_2, t_3],\ g_2 = [-t_1, t_0, t_2, t_3], \beta = g_1^2 = g_2^2$,

$$F_4 = a(T_0^4 + T_1^4) + bT_0^2T_1^2, \quad F_6 = T_0T_1(T_0^4 - T_1^4);$$

   (iii) $P \cong D_6, g_1 = [\epsilon_6t_0, \epsilon_6^{-1}t_1, t_2, -t_3], g_2 = [it_1, it_0, -t_2, it_3]$,

$$F_4 = aT_0^2T_1^2, \quad F_6 = T_0^6 + T_1^6 + bT_0^3T_1^3;$$

   (iv) $P \cong D_8, \widetilde{P} \cong D_{16},\ g_1 = [\epsilon_8t_0, \epsilon_8^{-1}t_1, -t_2, it_3], g_2 = [-t_1, t_0, -t_2, it_3]$, $g_1^4 = \beta, g_2^2 = 1$,

$$F_4 = aT_0^2T_1^2, \quad F_6 = T_0T_1(T_0^4 + T_1^4);$$

   (v) $P \cong D_{12}, \widetilde{P} \cong 2D_{12} \cong (2 \times 6)^\bullet 2, g_1 = [\epsilon_{12}t_0, \epsilon_{12}^{-1}t_1, -t_2, it_3], g_2 = [-t_1, t_0, t_2, t_3], g_1^6 = 1, g_2^2 = \beta$,

$$F_4 = aT_0^2T_1^2, \quad F_6 = T_0^6 + T_1^6.$$

3. *Other groups*
   (i) $P \cong A_4, \widetilde{P} \cong \overline{T}, g_1 = [\epsilon_8^7t_0 + \epsilon_8^7t_1, \epsilon_8^5t_0 + \epsilon_8t_1, 2\epsilon_3t_2, 2\sqrt{2}t_3],\ g_2 = [it_0, -it_1, t_2, t_3],\ g_1^3 = g_2^2 = \beta$,

$$F_4 = T_0^4 + 2\sqrt{-3}T_0^2T_1^2 + T_2^4, \quad F_6 = T_0T_1(T_0^4 - T_1^4),$$

   (ii) $P \cong O, \widetilde{P} \cong \overline{T} : 2, g_1 = [\epsilon_8^7t_0 + \epsilon_8^7t_1, \epsilon_8^5t_0 + \epsilon_8t_1, 2\epsilon_3t_2, 2\sqrt{2}t_3],\ g_2 = [\epsilon_8t_0, \epsilon_8^{-1}t_1, -t_2, it_3],\ g_3 = [-\epsilon_8t_1, \epsilon_8^7t_0, -t_2, it_3], g_1^3 = g_2^4 = \beta, g_3^2 = 1, r^{-1}(P) = 3 \times \bar{O}$,

$$F_4 = 0, \quad F_6 = T_0T_1(T_0^4 - T_1^4),$$

Table 8 gives the list of the full automorphism groups of Del Pezzo surfaces of degree 1.

The following is the list of cyclic minimal groups $\langle g \rangle$ of automorphisms of Del Pezzo surfaces $V(F)$ of degree 1.

1. Order 2:
   - A_1^8 (the Bertini involution) $g = [t_0, t_1, t_2, -t_3]$,

$$F = T_3^2 + T_2^3 + F_4(T_0, T_1)T_2 + F_6(T_0, T_1);$$

Table 8. Groups of automorphisms of Del Pezzo surfaces of degree 1.

Type	Order	Structure	F_4	F_6	Parameters
I	144	$3 \times (\bar{T} : 2)$	0	$T_0T_1(T_0^4 - T_1^4)$	
II	72	$3 \times 2D_{12}$	0	$T_0^6 + T_1^6$	
III	36	$6 \times D_6$	0	$T_0^6 + aT_0^3T_1^3 + T_1^6$	$a \neq 0$
IV	30	30	0	$T_0(T_0^5 + T_1^5)$	
V	24	$\bar{T}$	$a(T_0^4 + \alpha T_0^2T_1^2 + T_1^4)$	$T_0T_1(T_0^4 - T_1^4)$	$\alpha = 2\sqrt{-3}$
VI	24	$2D_{12}$	$aT_0^2T_1^2$	$T_0^6 + T_1^6$	$a \neq 0$
VII	24	2×12	T_0^4	T_1^6	
VIII	20	20	T_0^4	$T_0T_1^5$	
IX	16	D_{16}	$aT_0^2T_1^2$	$T_0T_1(T_0^4 + T_1^4)$	$a \neq 0$
X	12	D_{12}	$T_0^2T_1^2$	$T_0^6 + aT_0^3T_1^3 + T_1^6$	$a \neq 0$
XI	12	2×6	0	$G_3(T_0^2, T_1^2)$	
XII	12	2×6	T_0^4	$aT_0^6 + T_1^6$	$a \neq 0$
XIII	10	10	T_0^4	$T_0(aT_0^5 + T_1^5)$	$a \neq 0$
XIV	8	Q_8	$T_0^4 + T_1^4 + aT_0^2T_1^2$	$bT_0T_1(T_0^4 - T_1^4)$	$a \neq 2\sqrt{-3}$
XV	8	2×4	$aT_0^4 + T_1^4$	$T_0^2(bT_0^4 + cT_1^4)$	
XVI	8	D_8	$T_0^4 + T_1^4 + aT_0^2T_1^2$	$T_0T_1(b(T_0^4 + T_1^4) + cT_0^2T_1^2)$	$b \neq 0$
XVII	6	6	0	$F_6(T_0, T_1)$	
XVIII	6	6	$T_0(aT_0^3 + bT_1^3)$	$cT_0^6 + dT_0^3T_1^3 + T_1^6$	
XIX	4	4	$G_2(T_0^2, T_1^2)$	$T_0T_1F_2(T_0^2, T_1^2)$	
XX	4	2^2	$G_2(T_0^2, T_1^2)$	$G_3(T_0^2, T_1^2)$	
XXI	2	2	$F_4(T_0, T_1)$	$F_6(T_0, T_1)$	

2. Order 3:
   - $4A_2$ $g = [t_0, t_1, \epsilon_3 t_2, t_3]$,
$$F = T_3^2 + T_2^3 + F_6(T_0, T_1);$$
3. Order 4:
   - $2D_4(a_1)$ $g = [t_0, -t_1, -t_2, \pm it_3]$,
$$F = T_3^2 + T_2^3 + (aT_0^4 + bT_0^2T_1^2 + cT_1^4)T_2 + T_0T_1(dT_0^4 + eT_1^4);$$
4. Order 5:
   - $2A_4$ $g = [t_0, \epsilon_5 t_1, t_2, t_3]$,
$$F = T_3^2 + T_2^3 + aT_0^4T_2 + T_0(bT_0^5 + T_1^5);$$
5. Order 6:
   - $E_6(a_2) + A_2$ $g = [t_0, -t_1, \epsilon_3 t_2, t_3]$,
$$F = T_3^2 + T_2^3 + G_3(T_0^2, T_1^2);$$
   - $E_7(a_4) + A_1$ $g = [t_0, \epsilon_3 t_1, t_2, -t_3]$,
$$F = T_3^2 + T_2^3 + (T_0^4 + aT_0T_1^3)T_2 + bT_0^6 + cT_0^3T_1^3 + dT_1^6;$$

- $2D_4$ $g = [\epsilon_6 t_0, \epsilon_6^{-1} t_1, t_2, t_3]$,

$$F = T_3^2 + T_2^3 + aT_0^2T_1^2T_2 + bT_0^6 + cT_0^3T_1^3 + eT_1^6;$$

- $E_8(a_8)$ $g = [t_0, t_1, \epsilon_3 t_2, -t_3]$,

$$F = T_3^2 + T_2^3 + F_6(T_0, T_1);$$

- $A_5 + A_2 + A_1$ $g = [t_0, \epsilon_6 t_1, t_2, t_3]$,

$$F = T_3^2 + T_2^3 + aT_0^4T_2 + T_0^6 + bT_1^6;$$

6. Order 8:
- $D_8(a_3)$ $g = [it_0, t_1, -it_2, \pm\epsilon_8 t_3]$,

$$F = T_3^2 + T_2^3 + aT_0^2T_1^2T_2 + T_0T_1(T_0^4 + T_1^4);$$

7. Order 10:
- $E_8(a_6)$ $g = [t_0, \epsilon_5 t_1, t_2, -t_3]$,

$$F = T_3^2 + T_2^3 + aT_0^4T_2 + T_0(bT_0^5 + T_1^5);$$

8. Order 12:
- $E_8(a_3)$ $g = [-t_0, t_1, \epsilon_6 t_2, it_3]$,

$$F = T_3^2 + T_2^3 + T_0T_1(T_0^4 + aT_0^2T_1^2 + T_1^4);$$

9. Order 15:
- $E_8(a_5)$ $g = [t_0, \epsilon_5 t_1, \epsilon_3 t_2, t_3]$,

$$F = T_3^2 + T_2^3 + T_0(T_0^5 + T_1^5);$$

10. Order 20:
- $E_8(a_2)$ $g = [t_0, \epsilon_{10} t_1, -t_2, it_3]$,

$$F = T_3^2 + T_2^3 + aT_0^4T_2 + T_0T_1^5;$$

11. Order 24:
- $E_8(a_1)$ $g = [it_0, t_1, \epsilon_{12} t_2, \epsilon_8 t_3]$,

$$F = T_3^2 + T_2^3 + T_0T_1(T_0^4 + T_1^4);$$

12. Order 30:
- E_8 $g = [t_0, \epsilon_5 t_1, \epsilon_3 t_2, -t_3]$,

$$F = T_3^2 + T_2^3 + T_0(T_0^5 + T_1^5).$$

To list all minimal subgroups of $\mathrm{Aut}(S)$ is very easy. We know that any subgroup in $\mathrm{Ker}(r)$ contains one of the elements $\alpha, \beta, \alpha\beta$, which are all minimal of types $8A_1, 4A_2, E_8(a_8)$. So, a subgroup is not minimal only if its image P in $\mathrm{Aut}(B)$ can be lifted isomorphically to $\mathrm{Aut}(S)$.

We will use the following lemma.

Lemma 6.18. *Let $P \subset \mathrm{Aut}(\mathbb{P}^1)$ and $G \subset \mathrm{Aut}(S)$ be contained in $r^{-1}(P)$. Then G is a minimal group unless $G = \tilde{P} \cong P$ and G is a nonminimal cyclic group or nonminimal dihedral group D_6.*

Proof. It follows from above classification of possible subgroups of $\mathrm{Aut}(B)$ and its lifts to $\mathrm{Aut}(S)$ that any non isomorphic lift contains β, α, or $\beta\alpha$ which generate minimal cyclic groups. If the lift is isomorphic to P, then P is either a cyclic group or $P \cong D_6$. The group D_6 contains three elements of type $4A_1$ and two elements of type $2A_2$. Adding up the traces, we see that the group is not minimal. □

Let us classify minimal groups of automorphisms of a Del Pezzo surface of degree 1. As in the previous cases, to find a structure of such groups it is enough to consider the types of surfaces that are not specialized to surfaces of other types. The notation Type A→Type B indicates that a surface of type A specializes to a surface of type B.

$$\mathrm{V}, \mathrm{IX}, \mathrm{XIV}, \mathrm{XVI}, \mathrm{XVII}, \mathrm{XIX}, \mathrm{XXI} \to \mathrm{I},$$

$$\mathrm{III}, \mathrm{VI}, \mathrm{X}, \mathrm{XI}, \mathrm{XII}, \mathrm{XVI}, \mathrm{XVII}, \mathrm{XVIII}, \mathrm{XX}, \mathrm{XXI} \to \mathrm{II}$$

$$\mathrm{XIII}, \mathrm{XXI} \to \mathrm{IV}, \quad \mathrm{XIII}, \mathrm{XXI} \to \mathrm{VIII},$$

$$\mathrm{XII}, \mathrm{XX}, \mathrm{XXI} \to \mathrm{VII}, \;\; \mathrm{XX}, \mathrm{XI} \to \mathrm{XV}.$$

It remains to consider surfaces of types I, II, IV, VII, VIII, XV.

Type I. $P \cong S_4$.

Possible conjugacy classes of subgroups H are $\{1\}, 2, 2, 3, 2^2, 4, D_8, D_6, A_4, S_4$. Groups of order 2 have two conjugacy classes in P represented by $[it_0, -it_1]$ and $[-t_1, t_0]$. The groups are realized in cases (iii) and (iv). None of them lifts isomorphically. A cyclic group of order 3 is generated by a nonminimal element realized in case (vii). Its isomorphic minimal lift is not minimal. A cyclic group of order 4 does not admit an isomorphic lift. The dihedral subgroup 2^2 is of type (ii). This information, together with Lemma 6.18, allows us to classify all minimal subgroups:

- $P = \{1\}$: $\langle\beta\alpha\rangle \cong 6, \langle\alpha\rangle \cong 3, \langle\beta\rangle \cong 2$;
- $P = 2$: 4, 12;
- $P = 2$: 4, 12;
- $P = 3$: 3^2, 3×6;
- $P = 2^2$: Q_8, $Q_8 \times 3$;
- $P = 2^2$: D_8, $D_8 \times 3$;
- $P = 4$: 8, 8×3;
- $P = D_8$: D_{16}, $D_8 \times 3$;
- $P = D_6$: $D_6 \times 2$, $D_6 \times 3, D_6 \times 6$;

- $P = A_4$: $\overline{T}$, $\overline{T} \times 3$;
- $P = S_4$: $\overline{T} : 2$, $3 \times (\overline{T} : 2)$.

Surfaces specializing to a surface of type I have the following minimal subgroups.

V: 4, 6, Q_8, $\overline{T}$.
IX: 4(2), 8, D_{16}.
XIV: 4, Q_8.
XVI: D_8.
XVII: 2, 3, 6.
XIX: 2, 4.
XXI: 2.

Type II: $P = D_{12}$.

Possible subgroups are $\{1\}, 2, 2, 3, 2^2, 6, D_6, D_{12}$. Cyclic subgroups of order 2, 3 and 6 admit isomorphic nonminimal lifts. All these groups are not minimal. There are two conjugacy classes of subgroups of order 2 in P, represented by $[it_0, -it_1]$ and $[-t_1, t_0]$. One subgroups lifts isomorphically, while the other one does not. The cyclic group of order 6 admits an isomorphic lift and is not minimal. The dihedral group D_6 admits a nonminimal isomorphic lift.

- $P = \{1\}$: $\langle \beta\alpha \rangle \cong 6, \langle \alpha \rangle \cong 3, \langle \beta \rangle \cong 2$;
- $P = 2$: 4, 12;
- $P = 2$: 2^2, $2^2 \times 3$, 6;
- $P = 3$: 3^2, $3^2 \times 2$;
- $P = 2^2$: Q_8, $Q_8 \times 3$;
- $P = 6$: 2×6,
- $P = D_6$: $2 \times D_6$, $D_6 \times 3$, $D_6 \times 6$;
- $P = D_{12}$: $2D_{12}$, $3 \times 2D_{12}$.

Surfaces specializing to a surface of type II have the following subgroups:

III: 4, 12, 2^2, $2^2 \times 3$, 6, 3^2, $3^2 \times 2$, Q_8, $Q_8 \times 3$, $2 \times D_6$, $D_6 \times 3, D_6 \times 6$.
VI: 4, 2^2, 3^2, Q_8, 2×6, $2 \times D_6$, $2D_{12} \cong (2 \times 6)^\bullet 2$.
X: 2, $2 \times D_6$.
XI: 2, 3, 6, 2^2, 2×6.
XII: $6 \times 2, 6, 2^2, 2 \times 6$.
XVI: 2, 4, D_8.
XVII: 2, 3, 6.
XVIII: 2, 6.
XX: 2, 2^2.
XXI: 2.

Type IV: $P = 5$ This is easy. We have $P \cong 5$. It admits an isomorphic lift to a nonminimal subgroup.

- $P = \{1\}$: $\langle \beta\alpha \rangle \cong 6, \langle \alpha \rangle \cong 3, \langle \beta \rangle \cong 2$;
- $P = 5$: 5, $10, 15, 30$;

Surfaces specializing to a surface of type IV have the following subgroups:
XIII: $5,\ 10$.
XXI: 2.

Type VII: $P \cong 12$.

- $P = 2$: 2^2;
- $P = 3$: 6;
- $P = 4$: 2×4;
- $P = 6$: 2×6;
- $P = 12$: 2×12.

Surfaces specializing to a surface of type VII have the following subgroups:
XII: $2,\ 6,\ 2 \times 6$. XX: $2,\ 2^2$.
XXI: 2.

Type VIII: $P \cong 10$.

- $P = \{1\}$: $\langle \beta\alpha \rangle \cong 6, \langle \alpha \rangle \cong 3, \langle \beta \rangle \cong 2$;
- $P = 2$: 2^2;
- $P = 5$: 10;
- $P = 10$: 20.

Surfaces specializing to a surface of type VIII have the following subgroups:
XIII: $5,\ 10$.
XXI: 2.

Type XV: $P \cong 4$.

- $P = \{1\}$: $\langle \beta\alpha \rangle \cong 6, \langle \alpha \rangle \cong 3, \langle \beta \rangle \cong 2$;
- $P = 2$: 2^2;
- $P = 4$: 2×4.

Surfaces specializing to a surface of type VIII have the following subgroups:
XX: $2,\ 2^2$.
XXI: 2.

7 Elementary links and factorization theorem

7.1 Noether-Fano inequality

Let $|d\ell - m_1x_1 - \cdots - m_Nx_N|$ be a homaloidal net in $\mathbb{P}^2$. The following is a well-known classical result.

Lemma 7.1. *(Noether's inequality) Assume $d > 1, m_1 \geq \cdots \geq m_N \geq 0$. Then*

$$m_1 + m_2 + m_3 \geq d + 1,$$

and the equality holds if and only if either $m_1 = \cdots = m_N$ or $m_1 = n-1, m_2 = \cdots = m_N$.

Proof. We have

$$m_1^2 + \cdots + m_N^2 = d^2 - 1, \quad m_1 + \cdots + m_N = 3d - 3.$$

Multiplying the second equality by m_3 and subtracting from the first one, we get

$$m_1(m_1 - m_3) + m_2(m_2 - m_3) - \sum_{i \geq 4} m_i(m_3 - m_i) = d^2 - 1 - 3m_3(d - 1).$$

From this we obtain

$$(d-1)(m_1 + m_2 + m_3 - d - 1) = (m_1 - m_3)(d - 1 - m_1)$$
$$+(m_2 - m_3)(d - 1 - m_2) + \sum_{i \geq 4} m_i(m_3 - m_i).$$

Since $d - 1 - m_i \geq 0$, this obviously proves the assertion. □

Corollary 7.2.

$$m_1 > d/3.$$

Let us generalize Corollary 7.2 to birational maps of any rational surfaces. The same idea works even for higher-dimensional varieties. Let $\chi : S \to S'$ be a birational map of surfaces. Let $\sigma : X \to S, \phi : X \to S'$ be its resolution. Let $|H'|$ be a linear system on S' without base points. Let

$$\phi^*(H') \sim \sigma^*(H) - \sum_i m_i \mathcal{E}_i$$

for some divisor H on S and exceptional curves $\mathcal{E}_i$ of the map σ.

Theorem 7.3. *(Noether-Fano inequality) Assume that there exists some integer $m_0 \geq 0$ such that $|H' + mK_{S'}| = \emptyset$ for $m \geq m_0$. For any $m \geq m_0$ such that $|H + mK_S| \neq \emptyset$ there exists i such that*

$$m_i > m.$$

Proof. We know that $K_X = \sigma^*(K_S) + \sum_i \mathcal{E}_i$. Thus we have in $\mathrm{Pic}(X)$ the equality

$$\phi^*(H') + mK_X = (\sigma^*(H + mK_S)) + \sum (m - m_i)\mathcal{E}_i.$$

Applying $*$ to the left-hand side, we get the divisor class $H' + mK_{S'}$, which, by assumption, cannot be effective. Since $|\sigma^*(H + mK_S)| \neq \emptyset$, applying ϕ_* to the right-hand side, we get the sum of an effective divisor and the image of the divisor $\sum_i (m - m_i)\mathcal{E}_i$. If all $m - m_i$ are nonnegative, it is also an effective divisor, and we get a contradiction. Thus there exists i such that $m - m_i < 0$.

Example 7.4. Assume $S = S' = \mathbb{P}^2$, $H = d\ell$, and $H' = \ell$. We have $|H' + K_{S'}| = |-2\ell| = \emptyset$. Thus we can take $m_0 = 1$. If $d \geq 3$, we have for any $1 \leq a \leq d/3$, $|H' + aK_S| = |(d-3a)\ell| \neq \emptyset$. This gives $m_i > d/3$ for some i. This is Corollary 7.2.

Example 7.5. Let $\chi : S - \to S'$ be a birational map of Del Pezzo surfaces. Assume that S' is not a quadric or the plane. Consider the complete linear system $H' = |-K_{S'}|$. Then $|H' + mK_{S'}| = \emptyset$ for $m \geq 2$. Let $\chi^{-1}(H') = |D - \eta|$ be its proper transform on S. Choose a standard basis $(e_0, \ldots, e_k)$ in $\mathrm{Pic}(S)$ corresponding to the blowup $S \to \mathbb{P}^2$. Since $K_S = -3e_0 + e_1 + \cdots + e_k$, we can write $\chi^{-1}(H') = |-aK_S - \sum m_i x_i|$, where $a \in \frac{1}{3}\mathbb{Z}$. Assume that $\chi^{-1}(H') = -aK_S$. Then there exists a point with multiplicity $> a$ if $a > 1$, that we assume.

Remark 7.6. The Noether inequality is of course well-known (see, for example, [2], [35]). We give it here to express our respect of this important and beautiful result of classical algebraic geometry. Its generalization from Theorem 7.3 is also well known (see [39, 1.3]). Note that the result can be also applied to G-equivariant maps χ provided that the linear system $|H'|$ is G-invariant. In this case the linear system $|H - \eta|$ is also G-invariant and the bubble cycle $\eta = \sum m_i x_i$ consists of the sum of G-orbits.

The existence of base points of high multiplicity in the linear system $|H - \eta| = \chi^{-1}(H')$ follows from the classical theory of termination of the adjoint system for rational surfaces, which goes back to G. Castelnuovo. Nowadays, this theory has an elegant interpretation in the Mori theory, which we give in the next section.

7.2 Elementary links

We will be dealing with minimal Del Pezzo G-surfaces or minimal-conic-bundles G-surfaces. In the G-equivariant version of the Mori theory they are interpreted as extremal contractions $\phi : S \to C$, where $C = \mathrm{pt}$ is a point in the first case and $C \cong \mathbb{P}^1$ in the second case. They are also two-dimensional analogs of rational Mori G-fibrations.

A birational G-map between Mori fibrations are diagrams

$$\begin{tikzcd} S \arrow[r, dashed, "\chi"] \arrow[d, "\phi"'] & S' \arrow[d, "\phi'"] \\ C & C' \end{tikzcd} \tag{7.1}$$

which in general do not commute with the fibrations. These maps are decomposed into *elementary links*. These links are divided into the four following four types.

- Links of type I:

They are commutative diagrams of the form

$$\begin{tikzcd} S \arrow[d, "\phi"'] & Z = S' \arrow[l, "\sigma"'] \arrow[d, "\phi'"] \\ C = \mathrm{pt} & C' = \mathbb{P}^1 \arrow[l, "\alpha"'] \end{tikzcd} \tag{7.2}$$

Here $\sigma : Z \to S$ is the blowup of a G-orbit, S is a minimal Del Pezzo surface, $\phi' : S' \to \mathbb{P}^1$ is a minimal conic bundle G-fibration, α is the constant map. For example, the blowup of a G-fixed point on $\mathbb{P}^2$ defines a minimal conic G-bundle $\phi' : \mathbf{F}_1 \to \mathbb{P}^1$ with a G-invariant exceptional section.

- Links of type II:

They are commutative diagrams of the form

$$\begin{tikzcd} S \arrow[d, "\phi"'] & Z \arrow[l, "\sigma"'] \arrow[r, "\tau"] & S' \arrow[d, "\phi'"] \\ C & = & C' \end{tikzcd} \tag{7.3}$$

Here $\sigma : Z \to S, \tau : Z \to S'$ are the blowups of G-orbits such that rank $\mathrm{Pic}(Z)^G = \mathrm{rank}\ \mathrm{Pic}(S)^G + 1 = \mathrm{rank}\ \mathrm{Pic}(S')^G + 1$, $C = C'$ is either a point or $\mathbb{P}^1$. An example of a link of type II is the Geiser (or Bertini) involution of $\mathbb{P}^2$, where one blows up seven (or eight) points in general position that form one G-orbit. Another frequently used link of type II is an elementary transformation of minimal ruled surfaces and conic bundles.

- Links of type III:

These are the birational maps which are the inverses of links of type I.

- Links of type IV:

They exist when S has two different structures of G-equivariant conic bundles. The link is the exchange of the two conic bundle structures

$$\begin{tikzcd} S \arrow[d, "\phi"'] & = & S' \arrow[d, "\phi'"] \\ C & & C' \end{tikzcd} \tag{7.4}$$

One uses these links to relate elementary links with respect to one conic fibration to elementary links with respect to another conic fibration. Often the change of the conic-bundle structures is realized via an involution in $\mathrm{Aut}(S)$, for example, the switch of the factors of $S = \mathbb{P}^1 \times \mathbb{P}^1$ (see the following classification of elementary links).

7.3 The factorization theorem

Let $\chi : S - \to S'$ be a G-equivariant birational map of minimal G-surfaces. We would like to decompose it into a composition of elementary links. This is achieved with help of G-equivariant theory of log-pairs (S, D), where D is a G-invariant $\mathbb{Q}$-divisor on S. It is chosen as follows. Let us fix a G-invariant very ample linear system H' on S'. If S' is a minimal Del Pezzo surface, we take $\mathcal{H}' = |-a'K_{S'}|, a' \in \mathbb{Z}_+$. If S' is a conic bundle we take $\mathcal{H}' = |-a'K_{S'} + b'f'|$, where f' is the class of a fibre of the conic bundle, a', b' are some appropriate positive integers.

Let $\mathcal{H} = \mathcal{H}_S = \chi^{-1}(\mathcal{H}')$ be the proper transform of $\mathcal{H}'$ on S. Then

$$\mathcal{H} = |-aK_S - \sum m_x x|$$

if S is a Del Pezzo surface, $a \in \frac{1}{2}\mathbb{Z}_+ \cup \frac{1}{3}\mathbb{Z}_+$, and

$$\mathcal{H} = |-aK_S + bf - \sum m_x x|$$

if S is a conic bundle, $a \in \frac{1}{2}\mathbb{Z}_+, b \in \frac{1}{2}\mathbb{Z}$. The linear system $\mathcal{H}$ is G-invariant, and the 0-cycle $\sum m_x x$ is a sum of G-orbits with integer multiplicities. One uses the theory of log-pairs (S, D), where D is a general divisor from the linear system $\mathcal{H}$, by applying some "untwisting links" to χ in order to decrease the number a, the algebraic degree of $\mathcal{H}$. Since a is a rational positive number with bounded denominator, this process terminates after finitely many steps (see [22], [39]).

Theorem 7.7. *Let $f : S - \to S'$ be a birational map of minimal G-surfaces. Then χ is equal to a composition of elementary links.*

The proof of this theorem is the same as in the arithmetic case ([39], Theorem 2.5). Each time one chooses a link to apply and the criterion used for termination of the process is based on the following version of Noether's inequality in Mori theory.

Lemma 7.8. *In the notation from above, if $m_x \leq a$ for all base points x of $\mathcal{H}$ and $b \geq 0$ in the case of conic bundles, then χ is an isomorphism.*

The proof of this lemma is the same as in the arithmetic case ([39], Lemma 2.4).

We will call a base point x of $\mathcal{H}$ with $m_x > a$ a *maximal singularity* of $\mathcal{H}$. It follows from 3.2 that if $\mathcal{H}$ has a maximal singularity of height > 0, then it also has a maximal singularity of height 0. We will be applying the "untwisting links" of types I-III to these points. If $\phi : S \to \mathbb{P}^1$ is a conic bundle with all its maximal singularities with the help of links of type II, then either the algorithm terminates, or $b < 0$. In the latter case the linear system $|K_S + \frac{1}{a}\mathcal{H}| = |\frac{b}{a}f|$ is not nef and has *canonical singularities* (i.e. no maximal singularities). Applying the theory of log-pairs to the pair $(S, |\frac{b}{a}f|)$, we obtain an extremal contraction $\phi' : S \to \mathbb{P}^1$, i.e., another conic-bundle structure on S. Rewriting $\mathcal{H}$ in a new basis $-K_S, f'$, we obtain the new coefficient $a' < a$. Applying the link of type IV relating ϕ and ϕ', we begin the algorithm again with decreased a.

It follows from the proofs of Theorem 7.7 and Lemma 7.8 that all maximal singularities of H are in general position in the following sense.

(i) If S is a minimal Del Pezzo G-surface, then the blowup of all maximal singularities of $\mathcal{H}$ is a Del Pezzo surface (of course, this agrees with the description of points in general position at the end of Section 3.8).
(ii) If $\phi : S \to \mathbb{P}^1$ is a conic bundle, then none of the maximal singularities lie on a singular fiber of ϕ and no two lie on one fiber.

The meaning of these assertions is that the linear system $|H|$ has no fixed components. In the case of Del Pezzo surfaces with an orbit of maximal singular points we can find a link by blowing up this orbit to obtain a surface Z with $\mathrm{Pic}(Z)^G \cong \mathbb{Z}\oplus\mathbb{Z}$ and two extremal rays. By applying Kleiman's criterion this implies that $-K_Z$ is ample. The similar situation occurs in the case of conic bundles (see [39, Comment 2]).

Let S be a minimal Del Pezzo G-surface of degree d. Let us write $\mathcal{H}_S = |-aK_S - \sum m_\kappa \kappa|$ as in (3.8).

Lemma 7.9. *Let $\kappa_1, \ldots, \kappa_n$ be the G-orbits of maximal multiplicity. Then*

$$\sum_{i=1}^n d(\kappa_i) < d.$$

Proof. Let $D_1, D_2 \in \mathcal{H}_S$ be two general divisors from $\mathcal{H}_S$. Since $\mathcal{H}_S$ has no fixed components, we have

$$0 \le D_1 \cdot D_2 = a^2 d - \sum m_\kappa^2 d(\kappa) \le a^2 d - \sum_{i=1}^n m_{\kappa_i}^2 d(\kappa_i)$$

$$= a^2(d - \sum_{i=1}^n d(\kappa_i)) - \sum_{i=1}^n (m_{\kappa_i}^2 - a^2) d(\kappa_i).$$

It follows from Example 7.5 that $m_{\kappa_i} > a$ for all $i = 1, \ldots, n$. This implies that $d > \sum_{i=1}^n d(\kappa_i)$.

Definition 7.10. *A minimal Del Pezzo G-surface is called* superrigid *(respectively* rigid*) if any birational G-map $\chi : S - \to S'$ is a G-isomorphism (respectively there exists a birational G-automorphism $\alpha : S - \to S$ such that $\chi \circ \alpha$ is a G-isomorphism).*

A minimal conic bundle $\phi : S \to \mathbb{P}^1$ is called superrigid *(respectively* rigid*) if for any birational G-map $\chi : S - \to S'$, where $\phi' : S' \to \mathbb{P}^1$ is a minimal* *conic bundle, there exists an isomorphism* $\delta : \mathbb{P}^1 \to \mathbb{P}^1$ *such that the following diagram is commutative*

$$\begin{tikzcd} S \arrow[r, dashed, "\chi"] \arrow[d, "\phi"'] & S' \arrow[d, "\phi'"] \\ \mathbb{P}^1 \arrow[r, "\delta"] & \mathbb{P}^1 \end{tikzcd} \tag{7.5}$$

(respectively there exists a birational G*-automorphism* $\alpha : S - \to S'$ *such that the diagram is commutative after we replace* χ *with* $\chi \circ \alpha$*).*

Applying Lemma 7.8 and Lemma 7.9, we get the following.

Corollary 7.11. *Let* S *be a minimal Del Pezzo* G*-surface of degree* $d = K_S^2$*. If* S *has no* G*-orbits* κ *with* $d(\kappa) < d$*, then* S *is superrigid. In particular, a Del Pezzo surface of degree* 1 *is always superrigid and a Del Pezzo surface of degree* 2 *is superrigid unless* G *has a fixed point.*

A minimal conic G*-bundle with* $K_S^2 \le 0$ *is superrigid.*

The first assertion is clear. To prove the second one, we untwist all maximal base points of $\mathcal{H}_S$ with help of links of type II to get a conic bundle $\phi' : S' \to \mathbb{P}^1$ with $b' < 0$. Since $H_{S'}^2 = a^2 K_{S'}^2 + 4ab' - \sum {m_x'}^2 \ge 0$ and $K_{S'}^2 = K_S^2 \le 0, 4ab' < 0$, we get a contradiction with Lemma 7.8. Thus χ after untwisting maximal base points terminates at an isomorphism (see [36], [37], [39, Theorem 1.6]).

7.4 Classification of elementary links

Here we consider an elementary link $f : S - \to S'$ defined by a resolution $(S \stackrel{\sigma}{\leftarrow} Z \stackrel{\tau}{\to} S')$. We take $H_{S'}$ to be the linear system $|-aK_{S'}|$ if S' is a Del Pezzo surface and $|f|$ if S' is a conic bundle, where f is the divisor class of a fiber. It is assumed that the points that we blow up are in general position in sense of the previous subsection.

We denote by $\mathcal{DP}_k$ (respectively $\mathcal{C}_k$) the set of isomorphism classes of minimal Del Pezzo surfaces (respectively conic bundles) with $k = K_S^2$ (respectively $k = 8 - K_S^2$).

Proposition 7.12. *Let* S, S' *be as in Link I of type I. The map* $\sigma : Z = S' \to S$ *is the blowing up of a* G*-invariant bubble cycle* η *with* $\mathrm{ht}(\eta) = 0$ *of some degree* d*. The proper transform of the linear system* $|f|$ *on* S' *is equal to the linear system* $\mathcal{H}_S = |-aK_S - m\eta|$*. Here* f *is the class of a fiber of the conic bundle structure on* S'*. The following cases are possible:*

1. $K_S^2 = 9$
 - $S = \mathbb{P}^2, S' = \mathbf{F}_1, d = 1, m = 1, a = \frac{1}{3}$.
 - $S = \mathbb{P}^2, S' \in \mathcal{C}_3, d = 4, m = 1, a = \frac{2}{3}$.

2. $K_S^2 = 8$
 - $S = \mathbf{F}_0, \pi : S' \to \mathbb{P}^1$ *a conic bundle with* $k = 2$*,* $d = 2$*,* $m = 1$*,* $a = \frac{1}{2}$.

3. $K_S^2 = 4$
   - *$S \in \mathcal{D}P_4, p : S' \to \mathbb{P}^1$ a conic bundle with $f = -K_{S'} - l$, where l is a $(-1)-$curve, $d = a = 1, m = 2$.*

Proof. Let $\mathcal{H}_S = |-aK_S - b\eta|$, where η is a G-invariant bubble cycle of degree d. We have

$$(-aK_S - b\eta)^2 = a^2K_S^2 - b^2d = 0, \quad (-aK_S - b\eta, -K_S) = aK_S^2 - bd = 2.$$

Let $t = b/a$. We have

$$(td)^2 = dK_S^2, \quad K_S^2 - td = 2/a > 0.$$

The second inequality, gives $td < K_S^2$, hence $d < K_S^2$. Giving the possible values for K_S^2 and using that $a \in \frac{1}{3}\mathbb{Z}$, we check that the only possibilities are

$$(K_S^2, d, t) = (9, 1, 3),\ (8, 2, 2),\ (4, 1, 2),\ (4, 2, 1).$$

This gives our cases and one extra case $(4, 2, 1)$. In this case $a = 2$ and $\mathcal{H}_S = |-2K_S - 2x_1|$ contradicting the primitiveness of f. Note that this case is realized in the case that the ground field is not algebraically closed (see [39]). □

Proposition 7.13. *Let S, S' be as in Link of type II. Assume that S, S' are both minimal Del Pezzo surfaces. Then $(S \xleftarrow{\sigma} Z \xrightarrow{\tau} S')$, where σ is the blowup of a G-invariant bubble cycle η with $\mathrm{ht}(\eta) = 0$ and some degree d. The proper transform of the linear system $|-K_{S'}|$ on S is equal to $|-aK_S - m\eta|$. And d', m', a' for τ are similarly defined. The following cases are possible:*

1. $K_S^2 = 9$
   - *$S' \cong S = \mathbb{P}^2, d = d' = 8, m = m' = 18, a = a' = 17$ ($S \leftarrow Z \to S'$) is a minimal resolution of a Bertini involution).*
   - *$S' \cong S = \mathbb{P}^2, d = d' = 7, m = m' = 9, a = a' = 8$ ($S \leftarrow Z \to S'$) is a minimal resolution of a Geiser transformation).*
   - *$S' \cong S = \mathbb{P}^2, d = d' = 6, m = m' = 6, a = a' = 5$ ($S \leftarrow Z \to S'$) is a minimal resolution of a Cremona transformation given by the linear system $|5\ell - 2p_1 - 2p_2 - 2p_3 - 2p_4 - 2p_5|$.*
   - *$S \cong \mathbb{P}^2, S' \in \mathcal{D}P_5, d = 5, m' = 6, a = \frac{5}{3}, d' = 1, m = 2, a' = 3$.*
   - *$S \cong S' = \mathbb{P}^2, d = d' = 3, m = m' = 1, a = a' = \frac{2}{3}$, ($S \leftarrow Z \to S'$) is a minimal resolution of a standard quadratic transformation.*
   - *$S = \mathbb{P}^2, S' = \mathbf{F}_0, d = 2, m = 3, a' = \frac{3}{2}, d' = 1, a = \frac{4}{3}$.*
2. $K_S^2 = 8$
   - *$S \cong S' \cong \mathbf{F}_0, d = d' = 7, a = a' = 15, m = m' = 16$.*
   - *$S \cong S' \cong \mathbf{F}_0, d = d' = 6, a = a' = 7, m = m' = 8$.*
   - *$S \cong \mathbf{F}_0, S' \in \mathcal{D}P_5, d = 5, d' = 2, a = \frac{5}{2}, m = 4, a' = 4, m' = 6$.*
   - *$S \cong \mathbf{F}_0, S' \cong \mathbf{F}_0, d = d' = 4, a = a' = 3, m = m' = 4$.*

- $S \cong \mathbf{F}_0, S' \in \mathcal{D}P_6, d = 3, d' = 1, a = \frac{3}{2}, m = 2, m' = 4, a' = 2.$
- $S \cong \mathbf{F}_0, S' \cong \mathbb{P}^2, d = 1, d' = 2, a = \frac{3}{2}, m = 3, a' = \frac{4}{3}, m' = 2.$ *This link is the inverse of the last case from the preceding list.*

3. $K_S^2 = 6$

- $S \cong S' \in \mathcal{D}P_6, d = d' = 5, a = 11, m = 12.$
- $S \cong S' \in \mathcal{D}P_6, d = d' = 4, a = 5, m = 6.$
- $S \cong S' \in \mathcal{D}P_6, d = d' = 3, a = 3, m = 4.$
- $S \cong S' \in \mathcal{D}P_6, d = d' = 2, a = 2, m = 3.$
- $S \in \mathcal{D}P_6, S' = \mathbf{F}_0, d = 1, d' = 3, a = \frac{3}{2}, m = 2.$ *This link is the inverse of the link from the preceding list with* $S' \in \mathcal{D}P_6, d = 3.$

4. $K_S^2 = 5$

- $S \cong S' \in \mathcal{D}P_5, d = d' = 4, m = m' = 10, a = a' = 5..$
- $S = S' \in \mathcal{D}P_5, d = d' = 3, m = m' = 5, a = a' = 4.$
- $S \in \mathcal{D}P_5, S' = \mathbf{F}_0, d = 2, d' = 5.$ *This link is inverse of the link with* $S = \mathbf{F}_0, S' \in \mathcal{D}P_5, d = 5.$
- $S \in \mathcal{D}P_5, S' = \mathbb{P}^2, d = 1, d' = 5.$ *This link is inverse of the link with* $S = \mathbb{P}^2, S' \in \mathcal{D}P_5, d = 5.$

5. $K_S^2 = 4$

- $S \cong S' \in \mathcal{D}P_4, d = d' = 3.$ *This is an analogue of the Bertini involution.*
- $S \cong S' \in \mathcal{D}P_4, d = d' = 2.$ *This is an analogue of the Geiser involution.*

6. $K_S^2 = 3$

- $S \cong S' \in \mathcal{D}P_3, d = d' = 2.$ *This is an analogue of the Bertini involution.*
- $S \cong S' \in \mathcal{D}P_3, d = d' = 1.$ *This is an analogue of the Geiser involution.*

7. $K_S^2 = 2$

- $S = S' \in \mathcal{D}P_2, d = d' = 1.$ *This is an analogue of the Bertini involution.*

Proof. Similar to the proof of the previous proposition, we use that

$$\mathcal{H}_S^2 = a^2 K_S^2 - b^2 d = K_{S'}^2, \quad aK_S^2 - bd = K_{S'}^2,$$

$$H_{S'}^2 = a'^2 K_{S'}^2 - b'^2 d = K_S^2, \quad a'K_S^2 - b'd = K_S^2.$$

Since the link is not a biregular map, by Noether's inequality we have $a > 1, a' > 1, b > a, b' > a'$. This implies

$$d < K_S^2 - \frac{1}{a}K_{S'}^2, \ d' < K_{S'}^2 - \frac{1}{a'}K_S^2.$$

It is not difficult to list all solutions. For example, assume $K_S^2 = 1$. Since d is a positive integer, we see that there are no solutions. If $K_S^2 = 2$, we must have $d = d' = 1$. □

Proposition 7.14. *Let* S, S' *be as in Link of type II. Assume that* S, S' *are both minimal conic bundles. Then* $(S \leftarrow Z \to S')$ *is a composition of elementary transformations* $\mathrm{elm}_{x_1} \circ \ldots \circ \mathrm{elm}_{x_s}$*, where* $(x_1, \ldots, x_s)$ *is a* G*-orbit of points not lying on a singular fiber with no two points lying on the same fiber.*

We skip the classification of links of type III. They are the inverses of links of type I.

Proposition 7.15. *Let S, S' be as in Link of type IV. Recall that they consist of changing the conic bundle structure. The following cases are possible:*

- $K_S^2 = 8, S' = S, f' = -K_{S'} - f$; *it is represented by a switch automorphism;*
- $K_S^2 = 4, S' = S, f' = -K_{S'} - f$;
- $K_S^2 = 2, S' = S, f' = -2K_{S'} - f$; *it is represented by a Geiser involution;*
- $K_S^2 = 1, S' = S, f' = -4K_{S'} - f$; *it is represented by a Bertini involution;*

Proof. In this case, S admits two extremal rays and rank $\mathrm{Pic}(S)^G = 2$, so that $-K_S$ is ample. Let $|f'|$ be the second conic bundle. Write $f' \sim -aK_S + bf$. Using that $f'^2 = 0, f \cdot K_S = f' \cdot K_S = -2$, we easily get $b = -1$ and $aK_S^2 = 4$. This gives all possible cases from the assertion. □

8 Birational classes of minimal G-surfaces

8.1

Let S be a minimal G-surface S and $d = K_S^2$. We will classify all isomorphism classes of (S, G) according to the increasing parameter d. Since the number of singular fibers of a minimal conic bundle is at least 4, we have $d \leq 4$ for conic bundles.

- $d \leq 0$.

By Corollary 7.11, S is a superrigid conic bundle with $k = 8 - d$ singular fibers. The number k is a birational invariant. The group G is of de Jonquières type, and its conjugacy class in $\mathrm{Cr}(2)$ is determined uniquely by Theorem 5.7 or Theorem 5.3.

Also observe that if $\phi : S \to \mathbb{P}^1$ is an exceptional conic bundle and $G_0 = \mathrm{Ker}(G \to \mathrm{O}(\mathrm{Pic}(S))$ is non trivial, then no links of type II is possible. Thus the conjugacy class of G is uniquely determined by the isomorphism class of S.

- $d = 1$, S is a Del Pezzo surface.

By Corollary 7.11, the surface S is superrigid. Hence the conjugacy class of G is determined uniquely by its conjugacy class in $\mathrm{Aut}(S)$. All such conjugacy classes are listed in Subsection 6.7.

- $d = 1$, S is a conic bundle.

Let $\phi : S \to \mathbb{P}^1$ be a minimal conic bundle on S. It has $k = 7$ singular fibers. If $-K_S$ is ample, i.e., S is a (nonminimal) Del Pezzo surface, then the center of $\mathrm{Aut}(S)$ contains the Bertini involution β. We know that β acts as -1 on $\mathcal{R}_S$, thus β does not act identically on $\mathrm{Pic}(S)^G$, hence $\beta \notin G$. Since

k is odd, the conic bundle is not exceptional, so we cam apply Theorem 5.7, Case (2). It follows that G must contain a subgroup isomorphic to 2^2, adding β, we get that S is a minimal Del Pezzo 2^3-surface of degree 1. However, the classification shows that there are no such surfaces.

Thus $-K_S$ is not ample. It follows from Proposition 7.13 that the structure of a conic bundle on S is unique. Any other conic bundle birationally G-isomorphic to S is obtained from S by elementary transformations with G-invariant set of centers.

- $d = 2$, S is a Del Pezzo surface.

By Corollary 7.11, S is superrigid unless G has a fixed point on S. If $\chi : S- \to S'$ is a birational G-map, then H_S has only one maximal base point and it does not lie on a (-1)-curve. We can apply an elementary link $Z \to S$, $Z \to S$ of type II, which together with the projections $S \to \mathbb{P}^2$ resolves the Bertini involution. These links together with the G-automorphisms (including the Geiser involution) generate the group of birational G-automorphisms of S (see [39], Theorem 4.6). Thus the surface is rigid. The conjugacy class of G in $\mathrm{Cr}(2)$ is determined uniquely by the conjugacy class of G in $\mathrm{Aut}(S)$. All such conjugacy classes are listed in Table 7 and Theorem 6.17.

- $d = 2$, $\phi : S \to \mathbb{P}^1$ is a conic bundle.

If $-K_S$ is ample, then ϕ is not exceptional. The center of $\mathrm{Aut}(S)$ contains the Geiser involution γ. Since γ acts non-trivially on $\mathrm{Pic}(S)^G = \mathbb{Z}^2$, we see that $\gamma \notin G$. Applying γ, we obtain another conic-bundle structure. In other words, γ defines an elementary link of type IV. Using the factorization theorem; we show that the group of birational G-automorphisms of S is generated by links of type II, the Geiser involution, and G-automorphisms (see [37], [40], Theorem 4.9). Thus $\phi : S \to \mathbb{P}^1$ is a rigid conic bundle.

If S is not a Del Pezzo surface, ϕ could be an exceptional conic bundle with $g = 2$. In any case, the group G is determined in Theorem 5.3. We do not know whether S can be mapped to a conic bundle with $-K_S$ ample (see [37]).

Applying Proposition 5.2, we obtain that any conic bundle with $d \geq 3$ is a nonminimal Del Pezzo surface, unless $d = 4$ and S is an exceptional conic bundle. In the latter case, the group G can be found in Theorem 5.3. It is not known whether it is birationally G-isomorphic to a Del Pezzo surface. It is true in the arithmetic case.

- $d = 3$, S is a minimal Del Pezzo surface.

The classification of elementary links shows that S is rigid. Birational G-automorphisms are generated by links of type (6) from Proposition 7.12. The conjugacy class of G in $\mathrm{Cr}(2)$ is determined by the conjugacy class of G in $\mathrm{Aut}(S)$.

- $d = 3$, S is a minimal conic bundle.

Since $k = 5$ is odd, G has three commuting involutions; the fixed-point locus of one of them must be a rational 2-section of the conic bundle. It is easy to see that it is a (-1)-curve C from the divisor class $-K_S - f$. The other two fixed-point curves are of genus 2. The group G leaves the curve C invariant. Thus blowing it down, we obtain a minimal Del Pezzo G-surface S' of degree 4. The group G contains a subgroup isomorphic to 2^2. Thus G can be found in the list of minimal groups of degree-4 Del Pezzo surfaces with a fixed point. For example, the group 2^2 has four fixed points.

- $d = 4$, S is a minimal Del Pezzo surface.

If $S^G = \emptyset$, then S admits only self-links of type II, so it is rigid or superrigid. The conjugacy class of G in $\mathrm{Cr}(2)$ is determined by the conjugacy class of G in $\mathrm{Aut}(S)$, and we can apply Theorem 6.9. If x is a fixed point of G, then we can apply a link of type I to get a minimal conic bundle with $d = 3$. So, all groups with $S^G \neq \emptyset$ are conjugate to groups of de Jonquières type realized on a conic bundle $S \in \mathcal{C}_5$. Among minimal ones there are only two such groups isomorphic to a cyclic group of order 8 of type D_5 or 2^2.

- $d = 4$, S is a minimal conic bundle.

Since $k = 4$, it follows from Lemma 5.1 that either S is an exceptional conic bundle with $g = 1$, or S is a Del Pezzo surface with two sections with self-intersection -1 intersecting at one point. In the latter case, S is obtained by regularizing a de Jonquéres involution IH_3 (see Section 2.3). In the case that S is an exceptional conic bundle, the groups of automorphisms are described in Example 5.4. They are minimal if and only if the kernel of the map $G \to \mathrm{PGL}(2)$ contains an involution not contained in $G_0 = \mathrm{Ker}(G \to \mathrm{O}(\mathrm{Pic}(S))$. If G_0 is not trivial, then no elementary transformation is possible. So, S is not birationally isomorphic to a Del Pezzo surface.

- $d = 5$, S is a Del Pezzo surface, $G \cong 5$.

Let us show that (S, G) is birationally isomorphic to $(\mathbb{P}^2, G)$. Since rational surfaces are simply connected, G has a fixed point x on S. The anti canonical model of S is a surface of degree 5 in $\mathbb{P}^5$. Let P be the tangent plane of S at x. The projection from P defines a birational G-equivariant map $S - \to \mathbb{P}^2$ given by the linear system of anti canonical curves with double point at x. It is an elementary link of type II.

- $d = 5$, S is a Del Pezzo surface, $G \cong 5:2, 5:4$.

The cyclic subgroup of order 5 of G has two fixed points on S. This immediately follows from the Lefschetz fixed-point formula. Since it is normal in G, the groups G has an orbit κ with $d(\kappa) = 2$. Using an elementary link of type II with $S' = \mathbf{F}_0$, we obtain that G is conjugate to a group acting on $\mathbf{F}_0$.

- $d = 5$, S is a Del Pezzo surface, $G \cong A_5, S_5$.

It is clear that $S^G = \emptyset$, since otherwise, G admits a faithful 2-dimensional linear representation. It is known that it does not exist. Since A_5 has no index 2 subgroups G does not admit orbits κ with $d(\kappa) = 2$. The same is obviously true for $G = S_5$. It follows from the classification of links that (S, G) is superrigid.

- $d = 6$.

One of the groups from this case, namely $G \cong 2 \times S_3$, was considered in [40], [41] (the papers also discuss the relation of this problem to some questions in the theory of algebraic groups raised in [44]). It is proved there that (S, G) is not birationally isomorphic to $(\mathbb{P}^2, G)$ but birationally isomorphic to minimal $(\mathbf{F}_0, G)$. The birational isomorphism is easy to describe. We know that G contains the lift of the standard Cremona involution. It has 4 fixed points in S, the lifts of the points given in affine coordinates by $(\pm 1, \pm 1)$. The group S_3 fixes $(1, 1)$ and permutes the remaining points p_1, p_2, p_3. The proper transforms of the lines $\langle p_i, p_j \rangle$ in S are disjoint (-1)-curves E_i. The anti canonical model of S is a surface of degree 6 in $\mathbb{P}^6$. The projection from the tangent plane to S at the fixed point is a link of type II with $S' = \mathbf{F}_0$. It blows up the fixed point and then blows down the preimages of the curves E_i. Now the group G acts on $\mathbf{F}_0$ with $\mathbf{F}_0^G = \emptyset$.

If minimal G contains some non trivial imprimitive projective transformations, then G has no fixed points. It follows from the classification of links that S is rigid. If $G \cong 6$ or S_3, then it acts on $\mathbf{F}_0$ with a fixed point. The projection from this point defines a birational isomorphism (S, G) and $(\mathbb{P}^2, G)$. Thus the only groups that are not conjugate to a group of projective transformations are the groups which are mapped surjectively onto $W_S = S_3 \times 2$. Those of them that are mapped isomorphically are conjugate to subgroups of $\mathbf{F}_0$.

- $d = 8$.

In this case, $S = \mathbf{F}_0$ or $\mathbf{F}_n, n > 1$. In the first case (S, G) is birationally isomorphic to $(\mathbb{P}^2, G)$ if $S^G \neq \emptyset$ (via the projection from the fixed point). This implies that the subgroup G' of G belonging to the connected component of the identity of $\mathrm{Aut}(\mathbf{F}_0)$ is an extension of cyclic groups. As we saw in Theorem 4.9 this implies that G' is an abelian group of transformations $(x, y) \mapsto (\epsilon_{nk}^a x, \epsilon_{mk}^b y)$, where $a = sb \mod k$ for some s coprime to k. If $G \neq G'$, then we must have $m = n = 1$ and $s = \pm 1 \mod k$.

If $\mathbf{F}_0^G = \emptyset$ and $\mathrm{Pic}(\mathbf{F}_0)^G \cong \mathbb{Z}$, then the classification of links shows that links of type II with $d = d' = 7, 6, 5, d = 3, d' = 1$ map $\mathbf{F}_0$ to $\mathbf{F}_0$ or to minimal Del Pezzo surfaces of degrees 5 or 6. These cases have already been considered. If rank $\mathrm{Pic}(S)^G = 2$, then any birational G-map $S - \to S'$ is composed of elementary transformations with respect to one of the conic-bundle fibrations. They do not change K_S^2 and do not give rise a fixed points. So, G is not conjugate to any subgroup of $\mathrm{Aut}(\mathbb{P}^2)$.

Assume $n > 1$. Then $G = A.B$, where $A \cong n$ acts identically on the base of the fibration and $B \subset \mathrm{PGL}(2)$. The subgroup B fixes pointwise two disjoint sections, one of them is the exceptional one. Let us consider different cases corresponding to possible groups B.

$B \cong C_n$. In this case B has two fixed points on the base; hence G has two fixed points, on the non exceptional section. Performing an elementary transformation with center at one of these points, we descend G to a subgroup of $\mathbf{F}_{n-1}$. Proceeding in this way, we arrive to the case $n = 1$, and then obtain that G is a group of automorphisms of $\mathbb{P}^2$.

$B \cong D_n$. In this case B, has an orbit of cardinality 2 in $\mathbb{P}^1$. A similar argument shows that G has an orbit of cardinality 2 on the non exceptional section. Applying the product of the elementary transformations at these points, we descend G to a subgroup of automorphisms of $\mathbf{F}_{n-2}$. Proceeding in this way we obtain that G is conjugate to a subgroup of $\mathrm{Aut}(\mathbb{P}^2)$ or of $\mathrm{Aut}(\mathbf{F}_0)$. In the latter case it does not have fixed points, and hence is not conjugate to a linear subgroup of $\mathrm{Cr}(2)$.

$B \cong T$. The group B has an orbit of cardinality 4 on the non exceptional section. A similar argument shows that G is conjugate to a group of automorphisms of $\mathbf{F}_2, \mathbf{F}_0$, or $\mathbb{P}^2$.

$B \cong O$. The group B has an orbit of cardinality 6. As in the previous case we show that G is conjugate to a group of automorphisms of $\mathbb{P}^2$, $\mathbf{F}_0$, $\mathbf{F}_2$, or $\mathbf{F}_3$.

$B \cong I$. The group B has an orbit of cardinality 12. We obtain that G is conjugate to a group of automorphisms of $\mathbb{P}^2$ or of $\mathbf{F}_n$, where $n = 0, 2, 3, 4, 5, 6$.[3]

- $d = 9$.

In this case, $S = \mathbb{P}^2$ and G is a finite subgroup of $\mathrm{PGL}(3)$. The methods of representation theory allows us to classify them up to conjugacy in the group $\mathrm{PGL}(3)$. However, some of non conjugate groups can still be conjugate inside the Cremona group.

For example, all cyclic subgroups of $\mathrm{PGL}(3)$ of the same order n are conjugate in $\mathrm{Cr}(2)$. Any element g of order n in $\mathrm{PGL}(3)$ is conjugate to a transformation g given in affine coordinates by the formula $(x, y) \mapsto (\epsilon_n x, \epsilon_n^a y)$. Let $T \in \mathrm{dJ}(2)$ be given by the formula $(x, y) \mapsto (x, x^a/y)$. Let $g' : (x, y) \mapsto (\epsilon_n^{-1} x, y)$. We have

$$g' \circ T \circ g : (x, y) \mapsto (\epsilon_n x, \epsilon_n^a y) \mapsto (\epsilon_n x, x^a/y) \mapsto (x, x^a/y) = T.$$

This shows that g' and g are conjugate.

We do not know whether any two isomorphic non conjugate subgroups of $\mathrm{PGL}(3)$ are conjugate in $\mathrm{Cr}(2)$.

[3] A better argument due to I. Cheltsov shows that in this and the previous cases B is conjugate to a group of automorphisms of $\mathbb{P}^2$, or $\mathbf{F}_0$, or $\mathbf{F}_2$.

9 What is left?

Here we list some problems which have not been yet resolved.

- Find the conjugacy classes in Cr(2) of subgroups of PGL(3). For example, there are two non conjugate subgroups of PGL(3) isomorphic to A_5 or A_6 that differ by an outer automorphism of the groups. Are they conjugate in Cr(2)?
- Find the finer classification of the conjugacy classes of de Jonquières groups.

We already know that the number of singular fibers in a minimal conic-bundle G-surface is an invariant. Even more, the projective equivalence class of the corresponding k points on the base of the conic fibration is an invariant. Are there other invariants? In the case $G_K \cong 2$, we know that the quotient of the conic bundle by the involution generating G_K is a minimal ruled surface $\mathbf{F}_e$. Is the number e a new invariant?

- Give a finer geometric description of the algebraic variety parametrizing conjugacy classes.

Even in the case of Del Pezzo surfaces we give only normal forms. What is precisely the moduli space of Del Pezzo surfaces with a fixed isomorphism class of a minimal automorphism group?

We know that conic bundles (S, G) with $k \geq 8$ singular fibers are super-rigid, so any finite subgroup G' of Cr(2) conjugate to G is realized as an automorphism group of a conic bundle obtained from S by a composition of elementary transformations with G-invariant centers. If S is not exceptional and $G \cong 2.P$, then the invariant of the conjugacy class is the hyperelliptic curve of fixed points of the central involution. If $G \cong 2^2.P$, then we have three commuting involutions, and their curves of fixed points are the invariants of the conjugacy class. Do they determine the conjugacy class?

When $k = 6, 7$ we do not know whether (S, G) is birationally isomorphic to (S', G), where S' is a Del Pezzo surface. This is true when $k \in \{0, 1, 2, 3, 5\}$ or $k = 4$ and S is not exceptional.

- Find more explicit description of groups G as subgroups of Cr(2).

This has been done in the case of abelian groups in [6]. For example, one may ask to reprove and revise Autonne's classification of groups whose elements are quadratic transformations [3]. An example of such non cyclic group is the group of automorphisms S_5 of a Del Pezzo surface of degree 5.

- Finish the classical work on the birational classification of rational cyclic planes $z^n = f(x, y)$.

More precisely, the quotient S/G of a rational surface S by a cyclic group of automorphisms defines a cyclic extension of the fields of rational functions. Thus there exists a rational function $R(x, y)$ such that there exists an isomorphism of fields $\mathbb{C}(x, y)(\sqrt[n]{R(x, y)}) \cong \mathbb{C}(x, y)$, where n is the order of G.

Obviously we may assume that $R(x,y)$ is a polynomial $f(x,y)$, hence we obtain an affine model of S in the form $z^n = f(x,y)$. A birational isomorphism of G-surfaces replaces the branch curve $f(x,y) = 0$ by a Cremona equivalent curve $g(x,y)$. The problem is to describe the Cremona equivalence classes of the branch curves that define rational cyclic planes.

For example, when (S,G) is birationally equivalent to $(\mathbb{P}^2, G)$, we may take $f(x,y) = x$, since all cyclic groups of given order are conjugate in $\mathrm{Cr}(2)$. When $n = 2$, the problem was solved by M. Noether [47], and later, G. Castelnuovo and F. Enriques [17] realized that the classification follows from Bertini's classification of involutions in $\mathrm{Cr}(2)$. When n is prime, the problem was studied by A. Bottari [12]. We refer for modern work on this problem to [13], [14].

- Extend the classification to the case of non-algebraically closed fields, e.g., $\mathbb{Q}$, and algebraically closed fields of positive characteristic.

Note that there could be more automorphism groups in the latter case. For example, the Fermat cubic surface $T_0^3 + T_1^3 + T_2^3 + T_3^3 = 0$ over a field of characteristic 2 has the automorphism group isomorphic to $U(4, \mathbb{F}_4)$, which is a subgroup of index 2 of the Weyl group $W(E_6)$.

10 Tables

In the following tables we give the tables of conjugacy classes of subgroups in $\mathrm{Cr}(2)$ that are realized as minimal automorphism groups of Del Pezzo surfaces of degree $d \le 5$ and not conjugate to subgroups of automorphisms of minimal rational surfaces or conic bundles. The information about other groups can be found in the corresponding sections of the paper. The tables contain the order of a group G, its structure, the type of a surface on which the group is realized as a minimal group of automorphisms, the equation of the surface, and the number of conjugacy classes, if finite, or the dimension of the variety parametrizing the conjugacy classes.

Table 9. Cyclic subgroups.

Order	Type	Surface	Equation	Conjugacy
2	A_1^7	$\mathcal{D}P_2$	XIII	∞^6
2	A_1^8	$\mathcal{D}P_1$	XXI	
3	$3A_2$	$\mathcal{D}P_3$	I,III,IV	∞^1
3	$4A_2$	$\mathcal{D}P_1$	XVII	∞^3
4	$2A_3 + A_1$	$\mathcal{D}P_2$	II,III,V	∞^1
4	$2D_4(a_1)$	$\mathcal{D}P_1$	I,V, IX,XVI,XIX	∞^5
5	$2A_4$	$\mathcal{D}P_1$	XIII	∞^2
6	$E_6(a_2)$	$\mathcal{D}P_3$	I,VI	∞^1
6	$A_5 + A_1$	$\mathcal{D}P_3$	I, III, IV	∞^1
6	$E_7(a_4)$	$\mathcal{D}P_2$	XI	∞^1
6	$A_5 + A_2$	$\mathcal{D}P_2$	VIII	∞^1
6	$D_6(a_2) + A_1$	$\mathcal{D}P_2$	II,III,IV,IX	∞^1
6	$A_5 + A_2 + A_1$	$\mathcal{D}P_1$	II,VII,XII	∞^2
6	$E_6(a_2) + A_2$	$\mathcal{D}P_1$	II,XI	∞^2
6	$E_8(a_8)$	$\mathcal{D}P_1$	XVII	∞^3
6	$2D_4$	$\mathcal{D}P_1$	VI,X	∞^1
6	$E_7(a_4) + A_1$	$\mathcal{D}P_1$	II,VII,XVIII	∞^4
8	$D_8(a_3)$	$\mathcal{D}P_1$	IX	1
9	$E_6(a_1)$	$\mathcal{D}P_3$	I	1
10	$E_8(a_6)$	$\mathcal{D}P_1$	IV,VIII,XIII	∞^2
12	E_6	$\mathcal{D}P_3$	III	1
12	$E_7(a_2)$	$\mathcal{D}P_2$	III	1
12	$E_8(a_3)$	$\mathcal{D}P_1$	I,V	∞^2
14	$E_7(a_1)$	$\mathcal{D}P_2$	I	1
15	$E_8(a_5)$	$\mathcal{D}P_1$	IV	1
18	E_7	$\mathcal{D}P_2$	VI	1
20	$E_8(a_2)$	$\mathcal{D}P_1$	VIII	1
24	$E_8(a_1)$	$\mathcal{D}P_1$	I	1
30	E_8	$\mathcal{D}P_1$	IV	1

Table 10. Abelian noncyclic subgroups.

Order	Structure	Surface	Equation	Conjugacy classes
4	2^2	$\mathcal{D}P_2$	XII	∞^5
4	2^2	$\mathcal{D}P_1$	XX	∞^5
4	2^2	$\mathcal{D}P_1$	V, VI,X,XV,XVII	∞^3
8	2×4	$\mathcal{D}P_4$	(38)	∞^1
8	2×4	$\mathcal{D}P_2$	V	$2 \times \infty^1$
8	2×4	$\mathcal{D}P_2$	I-V	∞^1
8	2×4	$\mathcal{D}P_2$	VII	∞^2
8	2×4	$\mathcal{D}P_1$	VII,XV	∞^2
8	2^3	$\mathcal{D}P_4$		∞^2
8	2^3	$\mathcal{D}P_2$	I-V,X	∞^3
9	3^2	$\mathcal{D}P_3$	I	1
9	3^2	$\mathcal{D}P_3$	I,IV	$2 \times \infty^1$
9	3^2	$\mathcal{D}P_3$	III	1
9	3^2	$\mathcal{D}P_1$	I,II,III	∞^1
12	2×6	$\mathcal{D}P_4$	(6.5)	1
12	2×6	$\mathcal{D}P_2$	III, VII	∞^1
12	2×6	$\mathcal{D}P_1$	II,VIII,XII	∞^2
12	2×6	$\mathcal{D}P_1$	III,XI	∞^2
12	2×6	$\mathcal{D}P_1$	II,VII	∞^1
16	2^4	$\mathcal{D}P_4$		∞^2
16	$2^2 \times 4$	$\mathcal{D}P_2$	II,III,V	∞^1
16	4^2	$\mathcal{D}P_2$	II	1
16	2×8	$\mathcal{D}P_2$	II	1
18	3×6	$\mathcal{D}P_3$	I	1
18	3×6	$\mathcal{D}P_1$	III	∞^1
18	3×6	$\mathcal{D}P_1$	II	1
24	2×12	$\mathcal{D}P_1$	VIII	∞^1
24	2×12	$\mathcal{D}P_2$	III	1
27	3^3	$\mathcal{D}P_3$	I	1
32	2×4^2	$\mathcal{D}P_2$	II	1
36	6^2	$\mathcal{D}P_1$	II	1

Table 11. Products of cyclic groups and polyhedral or binary polyhedral noncyclic group.

Order	Structure	Surface	Equation	Conjugacy classes
6	D_6	$\mathcal{D}P_3$	III,IV,VIII,XI	∞^2
6	D_6	$\mathcal{D}P_3$	I	1
8	D_8	$\mathcal{D}P_4$	(6.4)	∞^1
8	D_8	$\mathcal{D}P_2$	II,III,V,VII	∞^2
8	D_8	$\mathcal{D}P_1$	I, XVI	∞^3
8	Q_8	$\mathcal{D}P_1$	II,V,VI,XIV	∞^2
8	$3 \times D_8$	$\mathcal{D}P_1$	I, XVI	∞^3
8	$3 \times Q_8$	$\mathcal{D}P_1$	II,III	∞^1
12	$2 \times D_6$	$\mathcal{D}P_3$	I,VI	∞^1
12	T	$\mathcal{D}P_2$	II	1
12	$2 \times D_6$	$\mathcal{D}P_2$	I,II,IV,IX	∞^2
12	$2 \times D_6$	$\mathcal{D}P_2$	II	1
12	$2 \times D_6$	$\mathcal{D}P_1$	I,II,III,VI	∞^1
16	$2 \times D_8$	$\mathcal{D}P_2$	II,III,V,VII	∞^2
16	D_{16}	$\mathcal{D}P_1$	I,IX	∞^1
18	$3 \times D_6$	$\mathcal{D}P_3$	III	1
18	$3 \times D_6$	$\mathcal{D}P_3$	I,IV	$2 \times \infty^1$
18	$3 \times D_6$	$\mathcal{D}P_1$	I,II,III	∞^1
24	$\overline{T}$	$\mathcal{D}P_1$	I,V	∞^1
24	$2 \times T$	$\mathcal{D}P_4$	(6.5)	1
24	S_4	$\mathcal{D}P_3$	I	3
24	S_4	$\mathcal{D}P_3$	II	1
24	S_4	$\mathcal{D}P_2$	II	1
24	S_4	$\mathcal{D}P_2$	I, II,IV	∞^1
24	$2 \times T$	$\mathcal{D}P_2$	II	1
24	$3 \times Q_8$	$\mathcal{D}P_1$	I	1
24	$3 \times Q_8$	$\mathcal{D}P_1$	II	1
24	$3 \times D_8$	$\mathcal{D}P_1$	I	∞^2
36	$6 \times D_6$	$\mathcal{D}P_1$	I,II,III	∞^1
48	$2 \times O$	$\mathcal{D}P_2$	I,II,III,V	∞^1
48	$2 \times \overline{T}$	$\mathcal{D}P_1$	I	1
72	$3 \times \overline{T}$	$\mathcal{D}P_1$	I	1

Table 12. Nonabelian p-groups.

Order	Structure	Surface	Equation	Conjugacy classes
8	D_8	$\mathcal{D}P_4$	(6.3)	∞^1
8	D_8	$\mathcal{D}P_2$	II,III,IV,V,VII	∞^2
8	Q_8	$\mathcal{D}P_1$	I,V,XIV	∞^2
8	D_8	$\mathcal{D}P_1$	I,V,XVI	∞^2
16	L_{16}	$\mathcal{D}P_4$	(6.3)	$2 \times \infty^1$
16	$2 \times D_8$	$\mathcal{D}P_2$	II,III,IV,V,VII	1
16	$2 \times D_8$	$\mathcal{D}P_2$	I	1
16	AS_{16}	$\mathcal{D}P_2$	II,III,V	$2 \times \infty^1$
16	M_{16}	$\mathcal{D}P_2$	II	1
16	D_{16}	$\mathcal{D}P_1$	I,IX	∞^1
32	$2^2 : 8$	$\mathcal{D}P_4$	(6.4)	1
32	$2^4 : 2$	$\mathcal{D}P_4$	(6.3)	∞^1
32	$D_8 : 4$	$\mathcal{D}P_2$	II	1
32	$2 \times AS_{16}$	$\mathcal{D}P_2$	III,V	$3 \times \infty^1$
32	$2 \times M_{16}$	$\mathcal{D}P_2$	II	1
64	$2^4 : 4$	$\mathcal{D}P_4$	(6.4)	1
64	$2 \times (D_8 : 4)$	$\mathcal{D}P_2$	II	1
27	$H_3(3)$	$\mathcal{D}P_3$	I,III,IV	∞^1
81	$3^3 : 3$	$\mathcal{D}P_3$	I	1

Table 13. Other groups.

Order	Structure	Surface	Equation	Conjugacy classes
18	$3^2 : 2$	$\mathcal{D}P_3$	I	2
24	$2D_{12}$	$\mathcal{D}P_1$	II,VI	∞^1
24	$D_8 : 3$	$\mathcal{D}P_2$	III	1
36	$3^2 : 2^2$	$\mathcal{D}P_3$	I	1
42	$2 \times (7 : 3)$	$\mathcal{D}P_1$	II	1
48	$2^4 : 3$	$\mathcal{D}P_4$	(6.5)	1
48	$2 \times D_8 : 3$	$\mathcal{D}P_2$	III	1
48	$\overline{T} : 2$	$\mathcal{D}P_1$	I	1
48	$4^2 : 3$	$\mathcal{D}P_2$	II	1
54	$H_3(3) : 2$	$\mathcal{D}P_3$	IV	∞^1
54	$3^3 : 2$	$\mathcal{D}P_3$	I	2
60	A_5	$\mathcal{D}P_5$		1
72	$3 \times 2D_{12}$	$\mathcal{D}P_1$	II	1
80	$2^4 : 5$	$\mathcal{D}P_4$	(6.6)	1
96	$2^4 : S_3$	$\mathcal{D}P_4$	(6.5)	1
96	$4^2 : S_3$	$\mathcal{D}P_2$	I	2
96	$2 \times (4^2 : 3)$	$\mathcal{D}P_2$	II	1
108	$3^3 : 4$	$\mathcal{D}P_3$	I	2
108	$3^3 : 2^2$	$\mathcal{D}P_3$	I	2
120	S_5	$\mathcal{D}P_5$		1
120	S_5	$\mathcal{D}P_3$	II	1
144	$3 \times (\overline{T} : 2)$	$\mathcal{D}P_1$	I	1
160	$2^4 : D_{10}$	$\mathcal{D}P_4$	(6.6)	1
162	$3^3 : S_3$	$\mathcal{D}P_3$	I	2
168	$L_2(7)$	$\mathcal{D}P_2$	I	1
192	$2 \times (4^2 : S_3)$	$\mathcal{D}P_3$	I	1
216	$3^3 : D_8$	$\mathcal{D}P_3$	I	2
336	$2 \times L_2(7)$	$\mathcal{D}P_2$	I	1
648	$3^3 : S_4$	$\mathcal{D}P_3$	I	2

References

1. A. Adem, J. Milgram, *Cohomology of finite groups*, Grundlehren der Mathematischen Wissenschaften, **309**, Springer-Verlag, Berlin, 1994.
2. M. Alberich-Carramiana, *Geometry of the plane Cremona maps*, Lecture Notes in Mathematics, 1769. Springer-Verlag, Berlin, 2002.
3. L. Autonne, *Recherches sur les groupes d'ordre fini contenus dans le groupe Cremona, Premier Mémoire Généralités et groupes quadratiques*, J. Math. Pures et Appl., (4) **1** (1885), 431–454.
4. S. Bannai, H. Tokunaga, *A note on embeddings of S_4 and A_5 into the Cremona group and versal Galois covers*, Publ. Res. Inst. Math. Sci. **43** (2007), 1111–1123.
5. L. Bayle, A. Beauville, *Birational involutions of P^2*, Asian J. Math. **4** (2000), 11–17.
6. J. Blanc, *Finite abelian subgroups of the Cremona group of the plane*, thesis, Univ. of Geneva, 2006.
7. J. Blanc, *Elements and cyclic subgroups of finite order of the Cremona group*, to appear in Com. Math. Helv.
8. A. Beauville, J. Blanc, *On Cremona transformations of prime order*, C. R. Math. Acad. Sci. Paris **339** (2004), 257–259.
9. A. Beauville, *p-elementary subgroups of the Cremona group*, J. Algebra, **314** (2007), 553–564.
10. E. Bertini, *Ricerche sulle trasformazioni univoche involutorie nel piano*, Annali di Mat. Pura Appl. (2) **8** (1877), 254–287.
11. H. Blichfeldt, *Finite collineation groups, with an introduction to the theory of operators and substitution groups*, Univ. of Chicago Press, Chicago, 1917.
12. A. Bottari, *Sulla razionalità dei piani multipli* $\{x, y, \sqrt[n]{F(x,y)}\}$, Annali di Mat. Pura Appl. (3) **2** (1899), 277–296.
13. A. Calabri, *Sulle razionalità dei piani doppi e tripli cyclici*, Ph.D. thesis, Univ. di Roma "La Sapienza", 1999.
14. A. Calabri, *On rational and ruled double planes*, Annali. di Mat. Pura Appl. (4) **181** (2002), 365–387.
15. R. Carter, *Conjugacy classes in the Weyl group.* in "Seminar on Algebraic Groups and Related Finite Groups", The Institute for Advanced Study, Princeton, N.J., 1968/69, pp. 297–318, Springer, Berlin.
16. G. Castelnuovo, *Sulle razionalità delle involutioni piani*, Math. Ann, **44** (1894), 125–155.
17. G. Castelnuovo, F. Enriques, *Sulle condizioni di razionalità dei piani doppia*, Rend. Circ. Mat. di Palermo, **14** (1900), 290–302.
18. A. Coble, *Algebraic geometry and theta functions* (reprint of the 1929 edition), A.M.S. Coll. Publ., v. 10. A.M.S., Providence, RI, 1982. MR0733252 (84m.14001)
19. J. Conway, R. Curtis, S. Norton, R. Parker, R. Wilson, *Atlas of finite groups*, Oxford Univ. Press, Eynsham, 1985.
20. J. Conway, D. Smith, *On quaternions and octonions: their geometry, arithmetic, and symmetry*, A K Peters, Ltd., Natick, MA, 2003.
21. J. Coolidge, *A treatise on algebraic plane curves*, Dover Publ. New York. 1959.
22. A. Corti, *Factoring birational maps of threefolds after Sarkisov*, J. Algebraic Geom. **4** (1995), 223–254.
23. T. de Fernex, *On planar Cremona maps of prime order*, Nagoya Math. J. **174** (2004), 1–28.

24. T. de Fernex, L. Ein, *Resolution of indeterminacy of pairs*, in “Algebraic geometry”, pp. 165–177, de Gruyter, Berlin, 2002.
25. M. Demazure, *Surfaces de Del Pezzo, I-V*, in “Séminaire sur les Singularités des Surfaces”, ed. by M. Demazure, H. Pinkham and B. Teissier. Lecture Notes in Mathematics, 777. Springer, Berlin, 1980, pp. 21–69.
26. I. Dolgachev, *Weyl groups and Cremona transformations*. in “Singularities, Part 1 (Arcata, Calif., 1981)”, 283–294, Proc. Sympos. Pure Math., 40, Amer. Math. Soc., Providence, RI, 1983.
27. I. Dolgachev, *Topics in classical algebraic geometry, Part I*, manuscript in preparations, see www.math.lsa.umich.edu/ idolga/lecturenotes.html.
28. P. Du Val, *On the Kantor group of a set of points in a plane*, Proc. London Math. Soc. **42** (1936), 18–51.
29. F. Enriques, *Sulle irrazionalita da cui puo farsi dipendere la resoluzione d'un' equazione algebrica* $f(x, y, z) = 0$ *con funzioni razionali di due parametri*, Math. Ann. **49** (1897), 1–23.
30. C. Geiser, *Über zwei geometrische Probleme*, J. Reine Angew. Math. **67** (1867), 78–89.
31. D. Gorenstein, *Finite groups*, Chelsea Publ. Co., New York, 1980.
32. É. Goursat, *Sur les substitutions orthogonales et les divisions rgulires de l'espace*, Ann. de l'Ècole Norm. Sup. (3) **6** (1889), 9–102.
33. R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics, No. 52. Springer-Verlag, New York-Heidelberg, 1977.
34. T. Hosoh, *Automorphism groups of cubic surfaces*, J. Algebra **192** (1997), 651–677.
35. H. Hudson, *Cremona transformations in plane and space*, Cambridge Univ. Press. 1927.
36. V. A. Iskovskih, *Rational surfaces with a pencil of rational curves* (Russian) Mat. Sb. (N.S.) **74** (1967), 608–638.
37. V. A. Iskovskikh, *Rational surfaces with a pencil of rational curves with positive square of the canonical class*, Math. USSR Sbornik, **12** (1970), 93–117.
38. V. A. Iskovskih, *Minimal models of rational surfaces over arbitrary fields* (Russian) Izv. Akad. Nauk SSSR Ser. Mat. **43** (1979), 19–43,
39. V. A. Iskovskikh, *Factorization of birational mappings of rational surfaces from the point of view of Mori theory* (Russian) Uspekhi Mat. Nauk **51** (1996), 3–72; translation in Russian Math. Surveys **51** (1996), 585–652.
40. V. A. Iskovskikh, *Two nonconjugate embeddings of the group* $S_3 \times Z_2$ *into the Cremona group* (Russian) Tr. Mat. Inst. Steklova **241** (2003), Teor. Chisel, Algebra i Algebr. Geom., 105–109; translation in Proc. Steklov Inst. Math. **241** (2003), 93–97.
41. V. A. Iskovskikh, *Two non-conjugate embeddings of* $S_3 \times \mathbb{Z}_2$ *into the Cremona group II*, Adv. Study in Pure Math. **50** (2008), 251–267.
42. S. Kantor, *Theorie der endlichen Gruppen von eindeutigen Transformationen in der Ebene*, Berlin. Mayer & Müller. 111 S. gr. 8°. 1895.
43. J. Kollár, S. Mori, *Birational geometry of algebraic varieties*, Cambridge Tracts in Mathematics, 134. Cambridge University Press, Cambridge, 1998.
44. N. Lemire, V. Popov, Z. Reichstein, *Cayley groups*, J. Amer. Math. Soc. **19** (2006), 921–967.
45. Yu. I. Manin, *Rational surfaces over perfect fields. II* (Russian) Mat. Sb. (N.S.) **72** (1967), 161–192.

46. Yu. I. Manin, *Cubic forms: algebra, geometry, arithmetic* Translated from Russian by M. Hazewinkel. North-Holland Mathematical Library, Vol. 4. North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., New York, 1974.
47. M. Noether, *Über die ein-zweideutigen Ebenentransformationen*, Sitzungberichte der physic-medizin, Soc. zu Erlangen, 1878.
48. B. Segre, *The non-singular cubic surface*, Oxford Univ. Press. Oxford. 1942.
49. T. Springer, *Invariant theory.* Lecture Notes in Mathematics, Vol. 585. Springer-Verlag, Berlin-New York, 1977.
50. A. Wiman, *Zur Theorie endlichen Gruppen von birationalen Transformationen in der Ebene*, Math. Ann. **48** (1896), 195–240.
51. H. Zassenhaus, *The theory of groups*, Chelsea Publ. New York, 1949.
52. D.-Q. Zhang, *Automorphisms of finite order on rational surfaces. With an appendix by I. Dolgachev*, J. Algebra **238** (2001), 560–589.

Lie Algebra Theory without Algebra

Simon K. Donaldson

Department of Mathematics, Imperial College, London SW7 2AZ
s.donaldson@imperial.ac.uk

Dedicated to Professor Yu I. Manin, on his 70th birthday

Summary. We present a self-contained and geometric proof of the standard results about maximal compact subgroups of simple Lie groups. The approach follows an idea suggested by Cartan, and is a variant of a proof given by Richardson in the complex case. We apply techniques associated to geometric invariant theory. The main step in the proof involves an argument with Riemannian metrics of nonpositive curvature.

Key words: simple Lie algebras, symmetric spaces, maximal compact subgroups, minimal vectors.

2000 Mathematics Subject Classifications: 22 E15, 53C35

1 Introduction

This is an entirely expository piece: the main results discussed are very well known and the approach we take is not really new, although the presentation may be somewhat different from what is in the literature. The author's main motivation for writing this piece comes from a feeling that the ideas deserve to be more widely known.

Let $\mathbf{g}$ be a Lie algebra over $\mathbf{R}$ or $\mathbf{C}$. A vector subspace $I \subset \mathbf{g}$ is an *ideal* if $[I, \mathbf{g}] \subset I$. The Lie algebra is called *simple* if it is not abelian and contains no proper ideals. A famous result of Cartan asserts that any simple complex Lie algebra has a compact real form (that is to say, the complex Lie algebra is the complexification of the Lie algebra of a compact group). This result underpins the theory of real Lie algebras, their maximal compact subgroups, and the classification of symmetric spaces. In the standard approach, Cartan's result emerges after a good deal of theory: the theorems of Engel and Lie, Cartan's

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*,
Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_12,

criterion involving the nondegeneracy of the Killing form, root systems, etc. On the other hand, if one assumes this result known—by some means—then one can immediately read off much of the standard structure theory of complex Lie groups and their representations. Everything is reduced to the compact case (Weyl's "unitarian trick"), and one can proceed directly to develop the detailed theory of root systems etc.

In [4], Cartan wrote:

> *J'ai trouvé effectivement une telle forme pour chacun des types de groupes simples. M. H. Weyl a démontré ensuite l'existence de cette forme par une raisonnement général s'appliquant à tous les cas à fois. On peut se demander si les calculs qui l'ont conduit à ce résultat ne pourraient pas encore se simplifier, ou plutôt si l'on ne pourrait pas, par une raissonnement a priori, démontrer ce théorème; une telle démonstration permettrait de simplifier notablement l'exposition de la theorie des groupes simples. Je ne suis à cet égard arrivé à aucun résultat; j'indique simplement l'idée qui m'a guidé dans mes recherches infructueuses.*

The direct approach that Cartan outlined (in which he assumed known the nondegeneracy of the Killing form) was developed by Helgason [5, p. 196], and a complete proof was accomplished by Richardson in [15]. In this article we revisit these ideas and present an almost entirely geometric proof of the result. This is essentially along the same lines as Richardson's, so it might be asked what we can add to the story. One point is that, guided by modern developments in geometric invariant theory and its relations with differential geometry, we can nowadays fit this into a much more general context and hence present the proofs in a (perhaps) simpler way. Another is that we are able to remove more of the algebraic theory, in particular, the nondegeneracy of the Killing form. We show that the results can be deduced from a general principle in Riemannian geometry (Theorem 4). The arguments apply directly to real Lie groups, and in our exposition we will work mainly in that setting. In the real case the crucial concept is the following. Suppose V is a Euclidean vector space. Then there is a transposition map $A \mapsto A^T$ on the Lie algebra $\mathrm{End}\, V$. We say that a subalgebra $\mathbf{g} \subset \mathrm{End}\, V$ is *symmetric* with respect to the Euclidean structure if it is preserved by the transposition map.

Theorem 1. *Let* $\mathbf{g}$ *be a simple real Lie algebra. Then there are a Euclidean vector space* V*, a Lie algebra embedding* $\mathbf{g} \subset \mathrm{End}(V)$*, and a Lie group* $G \subset \mathrm{SL}(V)$ *with Lie algebra* $\mathbf{g}$ *such that* $\mathbf{g}$ *is symmetric with respect to the Euclidean structure. Moreover, any compact subgroup of* G *is conjugate in* G *to a subgroup of* $G \cap \mathrm{SO}(V)$*.*

We explain in Section 5.1 below how to deduce the existence of the compact real form, in the complex case. Theorem 1 also leads immediately to the standard results about real Lie algebras and symmetric spaces, as we will discuss further in Section 5.1.

The author thanks Professors Martin Bridson, Frances Kirwan, Zhou Zhang, and Xuhua He for comments on the earlier version of this article.

2 More general setting

Consider any representation

$$\rho : \mathrm{SL}(V) \to \mathrm{SL}(W),$$

where V, W are finite-dimensional real vector spaces. Let w be a nonzero vector in W and let G_w be the identity component of the stabilizer of w in $\mathrm{SL}(V)$.

Theorem 2. *If V is an ireducible representation of G_w then there is a Euclidean metric on V such that the Lie algebra of G_w is symmetric with respect to the Euclidean structure, and any compact subgroup of G_w is conjugate in G_w to a subgroup of $G_w \cap \mathrm{SO}(V)$.*

Now we will show that Theorem 2 implies Theorem 1. Given a simple real Lie algebra $\mathbf{g}$, consider the action of $\mathrm{SL}(\mathbf{g})$ on the vector space W of skew-symmetric bilinear maps from $\mathbf{g} \times \mathbf{g}$ to $\mathbf{g}$. The Lie bracket of $\mathbf{g}$ is a point w in W. The group G_w is the identity component of the group of Lie algebra automorphisms of $\mathbf{g}$, and the Lie algebra of G_w is the algebra $\mathrm{Der}(\mathbf{g})$ of derivations of $\mathbf{g}$, that is, linear maps $\delta : \mathbf{g} \to \mathbf{g}$ with

$$\delta[x, y] = [\delta x, y] + [x, \delta y].$$

The adjoint action gives a Lie algebra homomorphism

$$\mathrm{ad} : \mathbf{g} \to \mathrm{Der}(\mathbf{g}).$$

The kernel of ad is an ideal in $\mathbf{g}$. This is not the whole of $\mathbf{g}$ (since $\mathbf{g}$ is not abelian), so it must be the zero ideal (since $\mathbf{g}$ is simple). Hence ad is injective. If U is a vector subspace of $\mathbf{g}$ preserved by G_w then any derivation δ must map U to U. In particular, ad_ξ maps U to U for any ξ in $\mathbf{g}$, so $[\mathbf{g}, U] \subset U$ and U is an ideal. Since $\mathbf{g}$ is simple we see that there can be no proper subspace preserved by G_w and the restriction of the representation is irreducible. By Theorem 2 there is a Euclidean metric on $\mathbf{g}$ such that $\mathrm{Der}(\mathbf{g})$ is preserved by transposition. Now we want to see that in fact $\mathrm{Der}(\mathbf{g}) = \mathbf{g}$. For $\alpha \in \mathrm{Der}(\mathbf{g})$ and $\xi \in \mathbf{g}$ we have

$$[\mathrm{ad}_\xi, \alpha] = \mathrm{ad}_{\alpha(\xi)},$$

so $\mathbf{g}$ is an ideal in $\mathrm{Der}(\mathbf{g})$. Consider the bilinear form

$$B(\alpha_1, \alpha_2) = \mathrm{Tr}(\alpha_1 \alpha_2)$$

on $\mathrm{Der}(\mathbf{g})$. This is nondegenerate, since $\mathrm{Der}(\mathbf{g})$ is preserved by transposition and $B(\alpha, \alpha^T) = |\alpha|^2$. We have

$$B([\alpha, \beta], \gamma) + B(\beta, [\alpha, \gamma]) = 0$$

for all $\alpha, \beta, \gamma \in \mathrm{Der}(\mathbf{g})$. Thus the subspace

$$\mathbf{g}^{\mathrm{perp}} = \{\alpha \in \mathrm{Der}\mathbf{g} : B(\alpha, \mathrm{ad}_\xi) = 0 \text{ for all } \xi \in \mathbf{g}\}$$

is another ideal in $\mathrm{Der}(\mathbf{g})$. On the other hand, the map $\alpha \mapsto -\alpha^T$ is an automorphism of $\mathrm{Der}(\mathbf{g})$, so $\mathbf{g}^T$ is also an ideal in $\mathrm{Der}(\mathbf{g})$. Suppose that $\mathbf{g} \cap \mathbf{g}^T \neq 0$. Then we can find a nonzero element α of $\mathbf{g} \cap \mathbf{g}^T$ with $\alpha^T = \pm\alpha$ and then $B(\alpha, \alpha) = \pm|\alpha|^2 \neq 0$, so the restriction of B to $\mathbf{g}$ is not identically zero. This means that $I = \mathbf{g} \cap \mathbf{g}^{\mathrm{perp}}$ is not the whole of $\mathbf{g}$, but I is an ideal in $\mathbf{g}$, so since $\mathbf{g}$ is simple, we must have $I = 0$.

We conclude from the above that if $\mathbf{g}$ were a proper ideal in $\mathrm{Der}(\mathbf{g})$ there would be another proper ideal J in $\mathrm{Der}(\mathbf{g})$ such that $J \cap \mathbf{g} = 0$. (We take J to be either $\mathbf{g}^T$ or $\mathbf{g}^{\mathrm{perp}}$.) But then for $\alpha \in J$ we have $[\alpha, \mathbf{g}] = 0$, but this means that α acts trivially on $\mathbf{g}$, which gives a contradiction.

Finally, the statement about compact subgroups in Theorem 1 follows immediately from that in Theorem 2.

(The argument corresponding to the above in the complex case (see Section 5.1) is more transparent.)

3 Lengths of vectors

We will now begin the proof of Theorem 2. The idea is to find a metric by minimizing the associated norm of the vector w. In the Lie algebra situation, which we are primarily concerned with here, this is in essence the approach suggested by Cartan and carried through by Richardson. In the general situation considered in Theorem 2 the ideas have been studied and applied extensively over the last quarter century or so, following the work of Kempf–Ness [9], Ness [13], and Kirwan [7]. Most of the literature is cast in the setting of complex representations. The real case has been studied by Richardson and Slodowy [14] and Marian [12] and works in just the same way.

Recall that we have a representation ρ of $\mathrm{SL}(V)$ in $\mathrm{SL}(W)$, where V and W are real vector spaces, a fixed vector w in W, and we define G_w to be the identity component of the stabilizer of w in $\mathrm{SL}(V)$. Suppose we also have some compact subgroup (which could be trivial) $K_0 \subset G_w$. We fix any Euclidean metric $| \, |_1$ on V that is preserved by K_0. Now it is standard that we can choose a Euclidean metric $| \, |_W$ on W that is invariant under the restriction of ρ to $\mathrm{SO}(V)$. We want to choose this metric $| \, |_W$ with the further property that the derivative $d\rho$ intertwines transposition in $\mathrm{End}V$ (defined by $| \, |_1$) and transposition in $\mathrm{End}W$ (defined by $| \, |_W$); that is to say,

$$d\rho(\xi^T) = (d\rho(\xi))^T .$$

To see that this is possible we can argue as follows. We complexify the representation to get $\rho_{\mathbf{C}} : \mathrm{SL}(V \otimes \mathbf{C}) \to \mathrm{SL}(W \otimes C)$. Then the compact group generated by the action of $\mathrm{SU}(V \otimes \mathbf{C})$ and complex conjugation acts on $W \otimes \mathbf{C}$

and we can choose a Hermitian metric on $W \otimes \mathbf{C}$ whose norm function is invariant under this group. Invariance under complex conjugation means that this Hermitian metric is induced from a Euclidean metric on W. Then the fact that $\rho_{\mathbf{C}}$ maps $\mathrm{SU}(V \otimes \mathbf{C})$ to $\mathrm{SU}(W \otimes \mathbf{C})$ implies that $d\rho$ has the property desired. (The author is grateful to Professors He and Zhang for pointing out the need for this argument. In our main application, to Theorem 1, the standard metric on W already has the desired property.)

Now define a function $\tilde{F}$ on $\mathrm{SL}(V)$ by

$$\tilde{F}(g) = |g(w)|_W^2.$$

For $u \in \mathrm{SO}(V)$ and $\gamma \in G_w$ we have

$$\tilde{F}(ug\gamma) = |ug\gamma(w)|_W^2 = |ug(w)|_W^2 = \tilde{F}(g).$$

So $\tilde{F}$ induces a function F on the quotient space $\mathcal{H} = \mathrm{SL}(V)/\mathrm{SO}(V)$, invariant under the natural action of $G_w \subset \mathrm{SL}(V)$. We can think about this in another, equivalent, way. We identify $\mathcal{H}$ with the Euclidean metrics on V of a fixed determinant. Since $\rho : \mathrm{SL}(V) \to \mathrm{SL}(W)$ maps $\mathrm{SO}(V)$ to $\mathrm{SO}(W)$ it induces a map from $\mathrm{SL}(V)/\mathrm{SO}(V)$ to $\mathrm{SL}(W)/\mathrm{SO}(W)$, and so a metric on V with the same determinant as $|\ |_1$ induces a metric on W. Then F is given by the square of the induced norm of the fixed vector w. Explicitly, the identification of $\mathrm{SL}(V)/\mathrm{SO}(V)$ with metrics is given by $[g] \mapsto |\ |_g$, where

$$|v|_g^2 = |gv|_1^2 = \langle v, g^T gv \rangle_1.$$

This function F has two crucial, and well-known, properties, which we state in the following lemmas.

Lemma 3. *Suppose F has a critical point at $H \in \mathcal{H}$. Then the Lie algebra of the stabilizer G_w is symmetric with respect to the Euclidean structure H on V.*

To prove this, there is no loss in supposing that H is the original metric $|\ |_1$. (For we can replace w by gw for any $g \in \mathrm{SL}(V)$.) The fact that ρ maps $\mathrm{SO}(V)$ to $\mathrm{SO}(W)$ implies that its derivative takes transposition in $\mathrm{End}V$ (defined by $|\ |_1$) to transposition in $\mathrm{End}W$ defined by $|\ |_W$. The condition for $\tilde{V}$ to be stationary is that

$$\langle d\rho(\xi)w, w\rangle_W = 0$$

for all ξ in the Lie algebra of $SL(V)$. In particular, consider elements of the form $\xi = [\eta, \eta^T]$ and write $A = d\rho(\eta)$. Then we have

$$0 = \langle d\rho[\eta, \eta^T]w, w\rangle_W = \langle [A, A^T]w, w\rangle_W = |A^T w|_W^2 - |Aw|_W^2.$$

By definition η lies in the Lie algebra of G_w if and only if $Aw = 0$. By the identity above, this occurs if and only if $A^T w = 0$, which is just when η^T lies in the Lie algebra of G_w.

For the second property of the function we need to recall the standard notion of geodesics in $\mathcal{H}$. We can identify $\mathcal{H}$ with the positive definite symmetric elements of $\mathrm{SL}(V)$, with the quotient map $\mathrm{SL}(V) \to \mathcal{H}$ given by $g \mapsto g^T g$. Then the geodesics in $\mathcal{H}$ are paths of the form

$$\gamma(t) = g^T \exp(St) g, \tag{1}$$

where g and S are fixed, with $g \in \mathrm{SL}(V)$ and S a trace-free endomorphism that is symmetric with respect to $| \, |_1$. Another way of expressing this is that a geodesic through any point $H \in \mathcal{H}$ is the orbit of H under a 1-parameter subgroup $e(t)$ in $\mathrm{SL}(V)$, where $e(t) = \exp(\sigma t)$ with σ a symmetric endomorphism with respect to the metric H.

Lemma 4. *1. For any geodesic γ the function $F \circ \gamma$ is convex, i.e.,*

$$\frac{d^2}{dt^2} F(\gamma(t)) \geq 0.$$

2. If F achieves its minimum in $\mathcal{H}$ then G_w acts transitively on the set of minima.

To prove the first part, note that, replacing w by gw, we can reduce to considering a geodesic through the base point $[1] \in \mathcal{H}$, so of the form $\exp(St)$, where S is symmetric with respect to $| \, |_1$. Now the derivative $d\rho$ maps the symmetric endomorphism S to a symmetric endomorphism $A \in \mathrm{End}(W)$. We can choose an orthonormal basis in W so that A is diagonal, with eigenvalues λ_i say. Then if w has coordinates w_i in this basis we have

$$F(\exp(St)) = \tilde{F}(\exp(St/2)) = \sum | \exp(\lambda_i t/2) w_i|_W^2 = \sum |w_i|^2 \exp(\lambda_i t),$$

and this is obviously a convex function of t.

To prove the second part note that in the above, the function $F(\exp(St)$ is either strictly convex or constant, and the latter occurs only when $\lambda_i = 0$ for each index i such that $w_i \neq 0$, which is the same as saying that $\exp(St)w = w$ for all t, or that the 1-parameter subgroup $\exp(St)$ lies in G_w. More generally, if we write a geodesic through a point H as the orbit of H under a 1-parameter subgroup $e(t)$ in $\mathrm{SL}(V)$, then the function is constant if and only if the 1-parameter subgroup lies in G_w. Suppose that H_1, H_2 are two points in $\mathcal{H}$ where F is minimal. Then F must be constant on the geodesic between H_1, H_2. Thus H_2 lies in the orbit of H_1 under a 1-parameter subgroup in G_w. So G_w acts transitively on the set of minima.

We now turn back to the proof of Theorem 2. Suppose that the convex function F on $\mathcal{H}$ achieves a minimum at $H_1 \in \mathcal{H}$. Then by Lemma 3 the Lie algebra of G_w is symmetric with respect to the Euclidean structure H_1 on V. It remains only to see that the compact subgroup K_0 of G_w is conjugate to a subgroup of the orthogonal group for this Euclidean structure. For each $H \in \mathcal{H}$ we have a corresponding special orthogonal group $\mathrm{SO}(H, V) \subset \mathrm{SL}(V)$. For $g \in \mathrm{SL}(V)$ the groups $\mathrm{SO}(H, V), \mathrm{SO}(g(H), V)$ are conjugate by g in $\mathrm{SL}(V)$. Recall that we chose the metric $|\ |_1$ to be K_0-invariant. This means that K_0 fixes the base point [1] in $\mathcal{H}$. Suppose we can find a point H_0 in $\mathcal{H}$ that minimizes F and that is also K_0-invariant. Then K_0 is contained in $\mathrm{SO}(H_0, V)$. But by the second part of Lemma 4 there is a $\gamma \in G_w$ such that $\gamma(H_0) = H_1$. Thus conjugation by γ takes $\mathrm{SO}(H_0, V)$ to $\mathrm{SO}(H_1, V)$ and takes K_0 to a subgroup on $\mathrm{SO}(H_1, V)$, as required.

To sum up, Theorem 2 will be proved if we can establish the following result.

Theorem 5. *Let F be a convex function on $\mathcal{H}$, invariant under a group $G_w \subset \mathrm{SL}(V)$. Let K_0 be a compact subgroup of G_w and let $[1] \in \mathcal{H}$ be fixed by K_0. Then if V is an ireducible representation of G_w there is a point $H_0 \in \mathcal{H}$ where F achieves its minimum and that is fixed by K_0.*

(Notice that the hypothesis here that there is a point $[1] \in \mathcal{H}$ fixed by K_0 is actually redundant, since any compact subgroup of $\mathrm{SL}(V)$ fixes some metric.)

4 Riemannian geometry argument

In this section we will see that Theorem 5 is a particular case of a more general result in Riemannian geometry. Let M be a complete Riemannian manifold, so for each point $p \in M$ we have a surjective exponential map

$$\exp_p : TM_p \to M.$$

We suppose M has the following property:

Property (∗)
For each point p in M the exponential map $\exp_p$ is distance-increasing:

$$d(\exp_p(\xi), \exp_p(\eta)) \geq |\xi - \eta|.$$

Readers with some background in Riemannian geometry will know that it is equivalent to say that M is simply connected with nonpositive sectional curvature, but we do not need to assume knowledge of these matters. The crucial background we need to know is the following.

Fact
There is a metric on $\mathcal{H} = \mathrm{SL}(V)/\mathrm{SO}(V)$ for which the action of $\mathrm{SL}(V)$ is isometric, with the geodesics described in (1) above and having Property (∗).

This Riemannian metric on $\mathcal{H}$ can be given by the formula

$$\|\delta H\|_H^2 = \mathrm{Tr}\left(\delta H H^{-1}\right)^2.$$

The distance-increasing property can be deduced from the fact that $\mathcal{H}$ has nonpositive curvature and standard comparison results for Jacobi fields. For completeness, we give a self-contained proof of the Fact in the appendix.

The piece of theory we need to recall in order to state our theorem is the notion of the "sphere at infinity" associated to a manifold M with Property $(*)$. This will be familiar in the prototype cases of Euclidean space and hyperbolic space. In general, for $x \in M$ write S_x for the unit sphere in the tangent space TM_x and define

$$\Theta_x : M \setminus \{x\} \to S_x$$

by

$$\Theta_x(z) = \frac{\exp_x^{-1}(z)}{|\exp_x^{-1}(z)|}.$$

If y is another point in M and R is greater than the distance $d = d(x,y)$ we define

$$F_{R,x,y} : S_y \to S_x$$

by

$$F_{R,x,y}(\nu) = \Theta_x \exp_y(R\nu).$$

Lemma 6. *For fixed x, y, ν the norm of the derivative of $F_{R,x,y,\nu}$ with respect to R is bounded by*

$$|\frac{\partial}{\partial R} F_{R,x,y}(\nu)| \leq \frac{d}{R(R-d)}.$$

Let γ be the geodesic $\gamma(t) = \exp_y(t\nu)$, let w be the point $\gamma(R)$, and let σ be the geodesic from x to w. The distance-increasing property of $\exp_x$ implies that the norm of the derivative appearing in the statement is bounded by $d(x,w)^{-1}$ times the component of $\gamma'(R)$ orthogonal to the tangent vector of σ at w. Thus

$$|\frac{\partial}{\partial R} F_{R,x,y}(\nu)| \leq \frac{\sin\phi}{d(x,w)},$$

where ϕ is the angle between the geodesics γ, σ at w. By the triangle inequality, $d(x,w) \geq R - d$. In a Euclidean triangle with side lengths d, R the angle opposite the side of length d is at most $\sin^{-1}(d/R)$. It follows from the distance-increasing property of $\exp_z$ that $\sin\phi \leq d/R$. Thus

$$\frac{\sin\phi}{d(x,w)} \leq \frac{d}{R(R-d)},$$

as required.

Since the integral of the function $1/R(R-d)$, with respect to R, from $R = 2d$ (say) to $R = \infty$, is finite, it follows from the lemma that $F_{R,x,y}$ converges uniformly as $R \to \infty$ to a continuous map $F_{x,y} : S_y \to S_x$, and obviously $F_{x,x}$ is the identity. Let z be another point in M, and ν a unit tangent vector at z. Then we have an identity that follows immediately from the definitions:

$$F_{R,x,z}(\nu) = F_{R',x,y} \circ F_{R,y,z}(\nu),$$

where $R' = d(y, \exp_z(R\nu))$. Since, by the triangle inequality again,

$$R' \geq R - d(y, z),$$

we can take the limit as $R \to \infty$ to obtain

$$F_{x,z} = F_{x,y} \circ F_{y,z} : S_z \to S_x.$$

In particular, $F_{y,x}$ is inverse to $F_{x,y}$, so the maps $F_{x,y}$ give a compatible family of homeomorphisms between spheres in the tangent spaces. We define the *sphere at infinity* $S_\infty(M)$ to be the quotient of the unit sphere bundle of M by these homeomorphisms, with the topology induced by the identification with S_{x_0} for any fixed base point x_0.

Now suppose that a topological group Γ acts by isometries on M. Then Γ acts on $S_\infty(M)$, as a set. Explicitly, if we fix a base point x_0 and identify the sphere at infinity with S_{x_0}, the action of a group element $g \in \Gamma$ is given by

$$g(\nu) = \lim_{R\to\infty} \Theta_{x_0} g(\exp_{x_0} R\nu).$$

Write the action as

$$A : \Gamma \times S_{x_0} \to S_{x_0}.$$

Given a compact set $P \subset \Gamma$ we can define

$$A_R : P \times S_{x_0} \to S_{x_0},$$

for sufficiently large R, by

$$A_R(g, \nu) = \Theta_{x_0} g(\exp_{x_0} R\nu).$$

Since $g(\exp_{x_0} R\nu) = \exp_{g(x_0)}(Rg_*\nu)$, the maps A_R converge uniformly as $R \to \infty$ to the restriction of A to $P \times S_{x_0}$. It follows that the action A is continuous. With these preliminaries in place we can state our main technical result.

Theorem 7. *Suppose that the Riemannian manifold M has Property (∗). Suppose that Γ acts by isometries on M and that F is a convex Γ-invariant function on M. Then either there is a fixed point for the action of Γ on $S_\infty(M)$ or the function F attains its minimum in M. Moreover, in the second case, if K_0 is a subgroup of Γ that fixes a point $x \in M$, then there is a point $x' \in M$ where F attains its minimum in M and with x' fixed by K_0.*

Return now to our example $\mathcal{H}$. The tangent space at the identity matrix [1] is the set of trace-free symmetric matrices. We define a *weighted flag* $(\mathcal{F}, \underline{\mu})$ to be a strictly increasing sequence of vector subspaces

$$0 = F_0 \subset F_1 \subset F_2 \subset \cdots \subset F_r = V$$

with associated weights $\mu_1 > \mu_2 > \cdots > \mu_r$, subject to the conditions

$$\sum n_i \mu_i = 0, \quad \sum n_i \mu_i^2 = 1,$$

where $n_i = \dim F_i / F_{i-1}$. If S is a trace-free symmetric endomorphism with $\mathrm{Tr}\, S^2 = 1$ then we associate a weighted flag to S as follows. We take μ_i to be the eigenvalues of S, with eigenspaces E_i, and form a flag with

$$F_1 = E_1, \quad F_2 = E_1 \oplus E_2, \quad \ldots .$$

It is clear then that the unit sphere $S_{[1]}$ in the tangent space of $\mathcal{H}$ at $[1]$ can be identified with the set of all weighted flags. Now there is an obvious action of $\mathrm{SL}(V)$ on the set of weighted flags and we have the following:

Lemma 8. *The action of* $\mathrm{SL}(V)$ *on the sphere at infinity in* $\mathcal{H}$ *coincides with the obvious action under the identifications above.*

This is clearly true for the subgroup $\mathrm{SO}(V)$. We use the fact that given any weighted flag $(\mathcal{F}, \underline{\mu})$ and $g \in \mathrm{SL}(V)$ we can write $g = uh$, where $u \in \mathrm{SO}(V)$ and h preserves $\underline{\mathcal{F}}$. (This is a consequence of the obvious fact that $\mathrm{SO}(V)$ acts transitively on the set of flags of a given type.) Thus it suffices to show that such h fix the point S in the unit sphere corresponding to $(\mathcal{F}, \underline{\mu})$ in the differential-geometric action. By the $\mathrm{SO}(V)$ invariance of the setup we can choose a basis so that $\mathcal{F}$ is the standard flag

$$0 \subset \mathbf{R}^{n_1} \subset \mathbf{R}^{n_1} \oplus \mathbf{R}^{n_2} \subset \cdots \subset \mathbf{R}^{n}.$$

Then S is the diagonal matrix with diagonal entries $\mu_1, \ldots, \mu_r$, repeated according to the multiplicities $n_1, \ldots, n_r$. The matrix h is upper triangular in blocks with respect to the flag. Now consider, for a large real parameter R, the matrix

$$M_R = \exp\left(-\frac{RS}{2}\right) h \exp\left(\frac{RS}{2}\right).$$

Consider a block h_{ij} of h. The corresponding block of M_R is

$$(M_R)_{ij} = e^{R(\mu_i - \mu_j)/2} h_{ij}.$$

Since h is upper-triangular in blocks and the μ_i are increasing, we see that M_R has a limit as R tends to infinity, given by the diagonal blocks in h. Since these diagonal blocks are invertible, the limit of $M(R)$ is invertible; hence

$$\delta_R = \mathrm{Tr}\left(\log(M_R M_R^*)\right)^2$$

is a bounded function of R. But $\delta_R^{1/2}$ is the distance in $\mathcal{H}_n$ between $\exp(RS)$ and $h \exp(RS) h^T$. It follows from the comparison argument, as before, that the angle between $\Theta_{[1]}(h \exp(RS) h^T)$ and S tends to zero as $R \to \infty$; hence h fixes S in the differential-geometric action.

Now Theorem 3 is an immediate consequence of Theorem 7 and Lemma 8, since if G_w fixes a point on the sphere at infinity in $\mathcal{H}$, it fixes a flag, hence some nontrivial subspace of V, and V is reducible as a representation of G_w.

Remarks 9.
- The advantage of this approach is that Theorem 4 seems quite accessible to geometric intuition. For example, it is obviously true in the case that M is hyperbolic space, taking the ball model, and we suppose that F extends continuously to the boundary of the ball. For then F attains its minimum on the closed ball, and if there are no minimizing points in the interior, the minimizer on the boundary must be unique (since there is a geodesic asymptotic to any two given points in the boundary).
- The author has not found Theorem 7 in the literature, but it does not seem likely that it is new. There are very similar results in [1], for example. The author has been told by Martin Bridson that a more general result of this nature holds, in the context of proper CAT(0) spaces. The proof of this more general result follows in an obvious way from Lemma 8.26 of [3] (see also Corollary 8.20 in that reference).
- The hypothesis on the existence of a fixed point x for K_0 in the statement of Theorem 7 is redundant, since any compact group acting on a manifold with Property $(*)$ has a fixed point, by a theorem of Cartan (see the remarks at the end of Section 3 above, and at the end of Section 5.2 below). However, we do not need to use this.

We now prove Theorem 7. We begin by disposing of the statement involving the compact group K_0. Suppose that F attains its minimum somewhere in M. Then, by convexity, the minimum set is a totally geodesic submanifold $\Sigma \subset M$. The action of K_0 preserves Σ, since F is Γ-invariant and K_0 is contained in Γ. Let x' be a point in Σ that minimizes the distance to the K_0 fixed-point x. Then if x'' is any other point in Σ the geodesic segment from x' to x'' lies in Σ and is orthogonal to the geodesic from x to x' at x'. By the distance-increasing property of the exponential map at x' it follows that the distance from x to x'' is strictly greater than the distance from x to x'. Thus the distance-minimizing point x' is unique, hence fixed by K_0.

To prove the main statement in Theorem 7 we use the following lemma.

Lemma 10. *Suppose that M has Property $(*)$ and N is any set of isometries of M. If there is a sequence x_i in M with $d(x_0, x_i) \to \infty$ and for each $g \in N$ there is a C_g with $d(x_i, gx_i) \leq C_g$ for all i, then there is a point in $S_\infty(M)$ fixed by N.*

Set $R_i = d(x_0, x_i)$ and $\nu_i = \Theta_{x_0}(x_i) \in S_{x_0}$. By the compactness of this sphere we may suppose, after perhaps taking a subsequence, that the ν_i converge as i tends to infinity to some $\nu \in S_{x_0}$. Then for each $g \in N$ we have, from the definitions,

$$A_{R_i}(g, \nu_i) = \Theta_{x_0}(gx_i).$$

Fixing g, let ϕ_i be the angle between the unit tangent vectors $\nu_i = \Theta_{x_0}(x_i)$ and $\Theta_{x_0}(gx_i)$. The distance-increasing property implies, as in Lemma 6, that

$$\sin \phi_i \leq C_g / R_i.$$

The angle ϕ_i can be regarded as the distance dist(,) between the points ν_i and $A_{R_i}(g, \nu_i)$ in the sphere S_{x_0}. In other words, we have

$$\text{dist }((\nu_i, A_{R_i}(g, \nu_i)) \leq \sin^{-1}(\frac{C_g}{R_i}).$$

Now take the limit as $i \to \infty$: we see that dist $(\nu, A(g, \nu)) = 0$, which is to say that ν is fixed by g.

To prove Theorem 7, consider the gradient vector field grad F of the function F, and the associated flow

$$\frac{dx}{dt} = -\text{grad } F_x$$

on M. By the standard theory, given any initial point there is a solution $x(t)$ defined for some time interval $(-T, T)$.

Lemma 11. *If $x(t)$ and $y(t)$ are two solutions of the gradient flow equation, for $t \in (-T, T)$, then $d(x(t), y(t))$ is a nonincreasing function of t.*

If $x(t)$ and $y(t)$ coincide for some t then they must do so for all t, by uniqueness of the solution of the flow equation, and in that case the result is certainly true. If $x(t)$ and $y(t)$ are always different, then the function $D(t) = d(x(t), y(t))$ is smooth: we compute the derivative at some fixed t_0. Let $\gamma(s)$ be the geodesic from $x(t_0) = \gamma(0)$ to $y(t_0) = \gamma(D)$. Clearly

$$D'(t_0) = \langle \text{grad} F_{x(t)}, \gamma'(0) \rangle - \langle \text{grad} F_{y(t)}, \gamma'(D) \rangle.$$

But $\langle \text{grad} F_{\gamma(s)}, \gamma'(s) \rangle$ is the derivative of the function $F \circ \gamma(s)$, which is non-decreasing in s by the convexity hypothesis, so $D'(t_0) \geq 0$, as required.

A first consequence of this lemma—applied to $y(t) = x(t + \delta)$ and taking the limit as $\delta \to 0$—is that the velocity $|\frac{dx}{dt}|$of a gradient path is decreasing. Thus for finite positive time, $x(t)$ stays in an a priori determined compact subset of M (since this manifold is complete). It follows that the flow is actually defined for all positive time, for any initial condition. Consider an arbitrary initial point x_0 and let $x(t)$ be this gradient path, for $t \geq 0$. If there is a sequence $t_i \to \infty$ such that $x(t_i)$ is bounded, then, taking a subsequence, we can suppose that $x(t_i)$ converges and it follows in a standard way that the limit is a minimum of F. If there is no such minimum then we can take a sequence such that $x_i = x(t_i)$ tends to infinity. Suppose that g is in Γ, so the action of g on M preserves F and the metric. Then $y(t) = g(x(t))$ is a gradient path with initial value $g(x_0)$ and $d(x_i, gx_i) \leq C_g = d(x_0, gx_0)$. Then, by Lemma 10, there is a fixed point for the action of Γ on $S_\infty(M)$.

There is an alternative argument that is perhaps more elementary, although it takes more space to write down in detail. With a fixed base point x_0 choose c with $\inf_M F < c < F(x_0)$ and let Σ_c be the hypersurface $F^{-1}(c)$. Let $z_c \in \Sigma_c$ be a point that minimizes the distance to x_0, so that x_0 lies on a geodesic γ from z_c normal to Σ_c. The convexity of F implies that the second fundamental form of Σ_c at z_c is positive with respect to the normal given by the geodesic from z_c to x_0. A standard comparison argument in "Fermi coordinates" shows that the exponential map on the normal bundle of Σ_c is distance-increasing on the side toward x_0. In particular, let w be another point in Σ_c and $y = \exp(R\xi)$, where ξ is the unit normal to Σ_c at w pointing in the direction of increasing F and $R = d(x_0, z_c)$. Then we have $d(z_c, w) \leq d(x_0, y)$. Now suppose g is in Γ. Then g preserves Σ_c and if we take $w = g(z_c)$ above we have $y = g(x_0)$. So we conclude from this comparison argument that $d(z_c, g(z_c)) \leq d(x_0, gx_0)$. Now take a sequence c_i decreasing to $\inf F$ (which could be finite or infinite). We get a sequence $x_i = z_{c_i}$ of points in M. If (x_i) contains a bounded subsequence then we readily deduce that there is a minimum of F. If x_i tends to infinity we get a sequence to which we can apply Lemma 5, since $d(x_i, gx_i) \leq C_g = d(x_0, gx_0)$.

5 Discussion

5.1 Consequences of Theorem 1

- We start with a simple Lie algebra $\mathbf{g}$ and use Theorem 1 to obtain an embedding $\mathbf{g} \subset \mathrm{End}(V)$, for a Euclidean space V, with $\mathbf{g}$ preserved by the transposition map. We also have a corresponding Lie group $G \subset \mathrm{SL}(V)$. We write K for the identity component of $G \cap \mathrm{SO}(V)$. It follows immediately from Theorem 1 that K is a maximal compact connected subgroup of G, and any maximal compact connected subgroup is conjugate to K.
- The involution $\alpha \mapsto -\alpha^T$ on $\mathrm{End}(V)$ induces a Cartan involution of $\mathbf{g}$, so we have an eigenspace decomposition

$$\mathbf{g} = \mathbf{k} \oplus \mathbf{p}$$

  with $\mathbf{k} = \mathrm{Lie}(K)$ and

$$[\mathbf{k}, \mathbf{k}] \subset \mathbf{k}\ ,\ \ [\mathbf{k}, \mathbf{p}] \subset \mathbf{p}\ ,\ \ [\mathbf{p}, \mathbf{p}] \subset \mathbf{k}. \tag{1}$$

  Notice that $\mathbf{k}$ is nontrivial, for otherwise $\mathbf{g}$ would be abelian.
- Consider the bilinear form $B(\alpha, \beta) = \mathrm{Tr}(\alpha\beta)$ on $\mathbf{g}$. Clearly this is positive definite on $\mathbf{p}$, negative definite on $\mathbf{k}$, and the two spaces are B-orthogonal. Thus B is nondegenerate. The Killing form $\hat{B}$ of $\mathbf{g}$ is negative definite on $\mathbf{k}$ (since the restriction of the adjoint action to K preserves some metric and $\mathbf{k}$ is not an ideal). So the Killing form is not identically zero and must be a positive multiple of B (otherwise the relative eigenspaces would be proper ideals). In fact, we do not really need this step, since in our proof of Theorem 1 the vector space W is $\mathbf{g}$ itself, and B is trivially equal to the Killing form.
- Either $\mathbf{p}$ is trivial, in which case G is itself compact, or there is a non-trivial Riemannian symmetric space of negative type $M_{\mathbf{g}}^{-} = G/K$ associated to $\mathbf{g}$. This can be described rather explicitly. Let us now fix on

the specific representation $\mathbf{g} \subset \mathrm{End}(\mathbf{g})$ used in the proof of Theorem 1. Say a Euclidean metric on $\mathbf{g}$ is "optimal" if the adjoint embedding is symmetric with respect to the metric, as in Theorem 1. (It is easy to see that the optimal metrics are exactly those that minimize the norm of the bracket, among all metrics of a given determinant.) Then $M^-_{\mathbf{g}}$ can be identified with the set of optimal metrics, a totally geodesic submanifold of $\mathcal{H} = \mathrm{SL}(\mathbf{g})/\mathrm{SO}(\mathbf{g})$.

- So far we have worked exclusively in the real setting. We will now see how to derive the existence of compact real forms of a simple complex Lie algebra.

Lemma 12. *If $\mathbf{g}$ is a simple complex Lie algebra then it is also simple when regarded as a real Lie algebra.*

To see this, suppose that $A \subset \mathbf{g}$ is a proper real ideal: a real vector subspace with $[A, \mathbf{g}] \subset A$. By complex linearity of the bracket, $A \cap iA$ is a complex ideal, so we must have $A \cap iA = 0$. But then since $i\mathbf{g} = \mathbf{g}$ we have $[A, \mathbf{g}] = [A, i\mathbf{g}] = i[A, \mathbf{g}] \subset iA$, so $[A, \mathbf{g}] = 0$. But $A + iA$ is another complex ideal, so we must have $\mathbf{g} = iA \oplus A$, and $\mathbf{g}$ is abelian.

Lemma 13. *Let $\mathbf{g}$ be a simple complex Lie algebra and let $\mathbf{g} = \mathrm{Lie}(G) \subset \mathrm{End}V$ be an embedding provided by Theorem 1, regarding $\mathbf{g}$ as a real Lie algebra. Then $\mathbf{g}$ is the complexification of the Lie algebra of the compact group $K = G \cap \mathrm{SO}(V)$.*

The inclusions (2) imply that

$$I = (\mathbf{p} \cap i\mathbf{k}) + (\mathbf{k} \cap i\mathbf{p})$$

is a complex ideal in $\mathbf{g}$, so either $I = \mathbf{g}$ or $I = 0$. In the first case we have $i\mathbf{k} = \mathbf{p}$, and $\mathbf{g}$ is the complexification of $\mathbf{k}$, as required. So we have to rule out the second case. If this were to hold we would have $\mathbf{k} \cap i\mathbf{p} = 0$, so $\mathbf{g} = \mathbf{k} \oplus (i\mathbf{p})$. Then $[i\mathbf{p}, i\mathbf{p}] \subset \mathbf{k}$, so the map σ on $\mathbf{g}$ given by multiplication by 1 on $\mathbf{k}$ and by -1 on $i\mathbf{p}$ is another involution of $\mathbf{g}$, regarded as a real Lie algebra. Now let $\hat{B}_{\mathbf{C}}$ be the Killing form regarded as a complex Lie algebra. So $\hat{B} = 2\mathrm{Re}\hat{B}_{\mathbf{C}}$. The fact that σ is an involution of $\mathbf{g}$ means that $\hat{B}(\mathbf{k}, i\mathbf{p}) = 0$. But we know that $\mathbf{p}$ is the orthogonal complement of $\mathbf{k}$ with respect to B and $\hat{B}$, so we must have $i\mathbf{p} = \mathbf{p}$. But $\hat{B}$ is positive definite on $\mathbf{p}$, while $\hat{B}(i\alpha, i\alpha) = 2\mathrm{Re}\hat{B}_{\mathbf{C}}(i\alpha, i\alpha) = -\hat{B}(\alpha, \alpha)$ so $\mathbf{p} \cap i\mathbf{p} = 0$. This means that $\mathbf{p} = 0$ and $\mathbf{g} = \mathbf{k}$ which is clearly impossible (by the same argument with the Killing form).

- The argument above probably obscures the picture. If one is interested in the complex situation it is much clearer to redo the whole proof in this setting, working with Hermitian metrics on complex representation spaces. The proof goes through essentially word for word, using the fact that the standard metric on $\mathrm{SL}(n, \mathbf{C})/\mathrm{SU}(n)$ has Property $(*)$.Then one can deduce the real case from the complex case rather than the other way around, as we have done above.

- Returning to the case of a simple real Lie algebra $\mathbf{g}$ that is not the Lie algebra of a compact group, we can also give an explicit description of the symmetric space $M_{\mathbf{g}}^{+}$ of positive type dual to $M_{\mathbf{g}}^{-}$. Fix an optimal metric on $\mathbf{g}$ and extend it to a Hermitian metric H on $\mathbf{g} \otimes \mathbf{C}$. Then $M_{\mathbf{g}}^{+}$ is the set of real forms $\mathbf{g}' \subset \mathbf{g} \otimes \mathbf{C}$ that are conjugate by G^c to $\mathbf{g}$ and such that the restriction of Re H to $\mathbf{g}'$ is an optimal metric on $\mathbf{g}'$. This is a totally geodesic submanifold of $\mathrm{SU}(\mathbf{g} \otimes \mathbf{C})/SO(\mathbf{g})$.

5.2 Comparison with other approaches

The approach we have used, minimizing the norm of the Lie bracket, is essentially the same as that suggested by Cartan, and carried through by Richardson, with the difference that we do not assume known that the Killing form is nondegenerate, so we operate with a special linear group rather than an orthogonal group. The crucial problem is to show that the minimum is attained when the Lie algebra is simple. This can be attacked by considering points in the closure of the relevant orbit for the action on the projectivized space. Richardson gives two different arguments. One uses the fact that a semisimple Lie algebra is rigid with respect to small deformations; the other uses the fact that a semisimple Lie algebra is its own algebra of derivations, so the orbits in the variety of semisimple Lie algebras all have the same dimension.

There is a general procedure for testing when an orbit contains a minimal vector, using Hilbert's 1-parameter subgroup criterion for stability in the sense of geometric invariant theory [10]. In the Lie algebra situation this gives a criterion involving the nonexistence of filtrations of a certain kind, but the author does not know an easy argument to show that simple Lie algebras do not have such filtrations. However, it is also a general fact that in the unstable case, there is a preferred maximally destabilizing 1-parameter subgroup. This theory was developed by Kempf [8], Hesselink [6], Bogomolov [2], and Rousseau [16] in the algebraic setting, and—in connection with the moment map and the length function—by Kirwan [7] and Ness [13]. The argument we give in Section 4 is essentially a translation of this theory into a differential-geometric setting. Lauret [11] has applied this general circle of ideas (geometric invariant theory/moment maps/minimal vectors) to more sophisticated questions in Lie algebra theory—going beyond the case of simple algebras.

One advantage of this method, in the real case, is that the uniqueness of maximal compact subgroups up to conjugacy emerges as part of the package. In the usual approach [5, Theorem 13.5], this is deduced from a separate argument: Cartan's fixed-point theorem for spaces of negative curvature. We avoid this, although the techniques we apply in Section 4 are very similar in spirit.

6 Appendix

We give a simple proof of the well-known fact stated in Section 4 that the manifold $\mathcal{H}$ has Property $(*)$. We identify $\mathcal{H}$ with $n \times n$ positive definite symmetric matrices of determinant 1. It suffices to prove the statement for the exponential map at the identity matrix. Recall that the metric on $\mathcal{H}$ is given by $|\delta H|_H^2 = \mathrm{Tr}\left((\delta H)H^{-1}\right)^2$. For fixed symmetric matrices S, α and a small real parameter h define

$$H(h) = (\exp(S + h\alpha) - \exp(S))\exp(-S)$$

and

$$v = \frac{dH}{dh}|_{h=0}.$$

we need to show that for any S and α, we have

$$\mathrm{Tr}\ v^2 \geq \mathrm{Tr}\ \alpha^2.$$

To see this we introduce another real parameter t and set

$$H(t,h) = (\exp(t(S + h\alpha)) - \exp(tS))\exp(-tS).$$

Then one readily computes

$$\frac{\partial H}{\partial t} = [S, H] + h\alpha \exp(t(S + h\alpha))\exp(-tS).$$

Now differentiate with respect to h and evaluate at $h = 0$ to get a matrix-valued function $V(t)$. Then we have

$$\frac{dV}{dt} = \frac{\partial^2 H}{\partial h \partial t}|_{h=0} = [S, V] + \alpha.$$

Clearly $v = V(1)$ and $V(0) = 0$, so our result follows from the following lemma:

Lemma 14. *Let S, α be real symmetric $n \times n$ matrices and let $V(t)$ be the matrix-valued function that is the solution of the ODE*

$$\frac{dV}{dt} = [S, V] + \alpha$$

with $V(0) = 0$. Then

$$\mathrm{Tr}\ V(t)^2 \geq t^2 \mathrm{Tr}\ \alpha^2$$

for all t.

To see this, consider first a scalar equation

$$\frac{df^+}{dt} = \lambda f^+ + a,$$

with λ, a constants and with the initial condition $f^+(0) = 0$. The solution is

$$f^+(t) = \left(\frac{e^{\lambda t} - 1}{\lambda} \right) a,$$

where we understand that the expression in parentheses is to be interpreted as t in the case $\lambda = 0$. Let $f^-(t)$ satisfy the similar equation

$$\frac{df^-}{dt} = -\lambda f^- + a,$$

with $f^-(0) = 0$. Then

$$f^+(t) f^-(t) = t^2 a^2 Q(t),$$

where

$$Q(t) = \frac{(e^{\lambda t} - 1)(1 - e^{-\lambda t})}{\lambda^2 t^2} = \frac{2(\cosh(\lambda t) - 1)}{\lambda^2 t^2}.$$

It is elementary that $Q(t) \geq 1$, so $f^+(t) f^-(t) \geq t^2 a^2$.

Now consider the operator ad_S acting on $n \times n$ matrices. We can suppose that S is diagonal with eigenvalues λ_i. Then a basis of eigenvectors for ad_S is given by the standard elementary matrices E_{ij} and

$$\mathrm{ad}_S(E_{ij}) = \lambda_{ij} E_{ij},$$

where $\lambda_{ij} = \lambda_i - \lambda_j$. Thus the matrix equation reduces to a collection of scalar equations for the components $V_{ij}(t)$. Since $\lambda_{ji} = -\lambda_{ij}$ and $\alpha_{ij} = \alpha_{ji}$, each pair V_{ij}, V_{ji} satisfies the conditions considered for f^+, f^- above and we have

$$V_{ij}(t) V_{ji}(t) \geq \alpha_{ij}^2 t^2.$$

(This is also true, with equality, when $i = j$.) Now summing over i, j gives the result.

This proof is not very different from the usual discussion of the Jacobi equation in a symmetric space. It is also much the same as the proof of Helgason's formula for the derivative of the exponential map [5, Theorem 1.7].

References

1. Bishop, R., O'Neill, B., *Manifolds of negative curvature*, Trans. Amer. Math. Soc. **145** (1969), 1–49.
2. Bogomolov, F.A., *Holomorphic tensors and vector bundles on projective varities*, Izv. Akad. nauk SSSR, Ser. Mat. **42** (1978), no. 6, 1227–1287 (English transl.: math. USSR, Izv, **13** (1979), 499–555).
3. Bridson, M., Haefliger, A., *Metric spaces of non-positive curvature*, Springer (1999).
4. Cartan, E., *Groupes simples clos et ouverts et géométrie Riemannienne*, J. Math. Pures Appl. **8** (1929), 1–33.

5. Helgason, S., *Differential Geometry, Lie groups and symmetric spaces*, Academic Press (1978).
6. Hesselink, W., *Uniform instability in reductive groups*, Jour. Reine Angew. Math. **304** (1978), 74–96.
7. Kirwan, F., *Cohomology of quotients in symplectic and algebraic geometry*, Princeton U.P. (1984).
8. Kempf, G., *Instability in invariant theory*, Annals of Math. **108** (1978), 299–316.
9. Kempf, G., Ness, L., *The lengths of vectors in representation spaces*, Springer Lecture Notes **732** (1978), 233–242.
10. Mumford, D. Fogarty, J. Kirwan, F.C., *Geometric Invariant Theory*, 3rd. ed. Springer (1994).
11. Lauret, J., *On the moment map for the variety of Lie algebras*, Jour. Fuctional Analysis **202** (2003), 392–423.
12. Marian, A., *On the real moment map*, Math. Res. Letters **8** (2001), 779–788.
13. Ness, L., *A stratification of the null cone by the moment map*, Amer. Jour. Math. **106** (1984), 128–132.
14. Richardson, R., Slodowy, P., *Minimal vectors for real reductive group actions*, Jour. Lond. Math. Soc. **42** (1990), 409–429.
15. Richardson, R., *Compact real forms of a complex semisimple Lie algebra,* Jour. Differential Geometry **2** (1968), 411–420.
16. Rousseau, G., *Immeubles, sphériques et théorie des invariants*, C. R. Acad. Sci. Paris Sér. A **286** (1978), 247–250.

Cycle Classes of the E-O Stratification on the Moduli of Abelian Varieties

Torsten Ekedahl[1] and Gerard van der Geer[2]

[1] Matematiska institutionen, Stockholms universitet, SE-106 91 Stockholm, Sweden, teke@math.su.se

[2] Korteweg-de Vries Instituut, Universiteit van Amsterdam, Postbus 94248, 1090 GE, Amsterdam, The Netherlands, geer@science.uva.nl

To Yuri Ivanovich Manin on the occasion of his 70th birthday

Summary. We introduce a stratification on the space of symplectic flags on the de Rham bundle of the universal principally polarized abelian variety in positive characteristic. We study its geometric properties, such as irreducibility of the strata, and we calculate the cycle classes. When the characteristic p is treated as a formal variable these classes can be seen as a deformation of the classes of the Schubert varieties for the corresponding classical flag variety (the classical case is recovered by putting p equal to 0). We relate our stratification with the E-O stratification on the moduli space of principally polarized abelian varieties of a fixed dimension and derive properties of the latter. Our results are strongly linked with the combinatorics of the Weyl group of the symplectic group.

Key words: moduli space, abelian variety, E-O-stratification, cycle classes, Weyl group.

2000 Mathematics Subject Classifications: 14K, 14G35, 14L

1 Introduction

The moduli space $\mathcal{A}_g$ of principally polarized abelian varieties of dimension g is defined over the integers. For the characteristic-zero fiber $\mathcal{A}_g \otimes \mathbb{C}$ we have an explicit description as an orbifold $\mathrm{Sp}_{2g}(\mathbb{Z})\backslash\mathcal{H}_g$ with $\mathcal{H}_g$ the Siegel upper half-space of degree g. It is a recent insight, though, that perhaps the positive characteristic fibres $\mathcal{A}_g \otimes \mathbb{F}_p$ are more accessible than the characteristic-zero one. A good illustration of this is provided by the E-O stratification of $\mathcal{A}_g \otimes \mathbb{F}_p$, a stratification consisting of 2^g quasi-affine strata. It was originally defined by Ekedahl and Oort (see [Oo01]) by analyzing the structure of the kernel of

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*, Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_13,

multiplication by p of an abelian variety. It turns out that this group scheme can assume 2^g forms only, and this led to the strata. For $g = 1$ the two strata are the loci of ordinary and of supersingular elliptic curves. Some strata possess intriguing properties. For example, the stratum of abelian varieties of p-rank 0 is a complete subvariety of $\mathcal{A}_g \otimes \mathbb{F}_p$ of codimension g, the smallest codimension possible. No analogue in characteristic 0 of either this stratum or the stratification is known. In fact, Keel and Sadun [KS03] proved that complete subvarieties of $\mathcal{A}_g \otimes \mathbb{C}$ of codimension g do not exist for $g \geq 3$.

While trying to find cycle classes for the E-O strata we realized that the strata could be described as degeneration loci for maps between vector bundles, and since such loci are indexed by Young diagrams, our attention was turned towards the combinatorics of the Weyl group. When considered in this light it is clear that much of the combinatorics of [Oo01] is closely related to the Weyl group W_g of Sp_{2g}, which is the semisimple group relevant for the analytic description of $\mathcal{A}_g \otimes \mathbb{C}$. The main idea of this paper is to try to make this connection more explicit. More precisely, the combinatorics of the E-O strata is most closely related to the combinatorics associated to W_g and the Weyl subgroup corresponding to the maximal parabolic subgroup P of elements of Sp_{2g} stabilizing a maximal isotropic subspace in the $2g$-dimensional symplectic vector space. Indeed, this sub-Weyl group is S_g, the group of permutations on g letters (embedded as a sub-Weyl group in W_g), and the E-O strata are in bijection with the cosets in W_g/S_g; we shall use the notation $\mathcal{V}_\nu$ for the (open) stratum of $\mathcal{A}_g \otimes \mathbb{F}_p$ corresponding to $\nu \in W_g/S_g$ (and $\overline{\mathcal{V}}_\nu$ for its closure). The coset space W_g/S_g is also in bijection with the set of Bruhat cells in the space of maximal totally isotropic flags Sp_{2g}/P and we believe this to be no accident. (The formal relation between $\mathcal{A}_g$ and Sp_{2g}/P is that Sp_{2g}/P is the *compact dual* of $\mathcal{H}_g$.)

In order to push the analogy further we introduce a "flag space" $\mathcal{F}_g \to \mathcal{A}_g$ whose fibers are isomorphic to the fibers of the quotient morphism $\mathrm{Sp}_{2g}/B \to \mathrm{Sp}_{2g}/P$, where B is a Borel subgroup of P. In positive characteristic we define (and this definition makes sense only in positive characteristic) a stratification of $\mathcal{F}_g$, whose open strata $\mathcal{U}_w$ and their corresponding closures $\overline{\mathcal{U}}_w$ are parametrized by the elements of W_g. This stratification is very similar to the stratification by Bruhat cells of Sp_{2g}/B and their closures, the Schubert strata, which are also parametrized by the elements of W_g. Our first main result is that this is more than a similarity when one works locally; we show (cf., Therorem 8.2) that for each point of $\mathcal{F}_g$ there is a stratum-preserving local isomorphism (in the étale topology) taking the point to some point of Sp_{2g}/B. Since much is known about the local structure of the Schubert varieties we immediately get a great deal of information about the local structure of our strata. The first consequence is that $\overline{\mathcal{U}}_w$ is equidimensional of dimension equal to the length of w. A very important consequence is that the $\overline{\mathcal{U}}_w$ are all normal; this situation differs markedly from the case of the closed E-O strata, which in general are not normal. Another consequence is that the inclusion relation between the strata is given exactly by the Bruhat–Chevalley order on W_g.

(A much more sophisticated consequence is that the local structure of the ℓ-adic intersection complex for a closed stratum is the same as for the Schubert varieties and in particular that the dimensions of its fibers over the open strata of the closed stratum are given by the Kazhdan–Lusztig polynomials. We shall not, however, pursue that in this article.)

We give several applications of our results to the structure of the strata $\overline{\mathcal{U}}_w$. The first, and most important, is that by construction the strata $\mathcal{U}_w$ are defined as the loci where two symplectic flags on the same vector bundle are in relative position given by w. After having shown that they have the expected codimension and are reduced, we can use formulas of Fulton, as well as those of Pragacz and Ratajski, as crystallized in the formulas of Kresch and Tamvakis, to get formulas for the cycle classes of the strata. A result of Fulton gives such formulas (cf., Theorem 12.1) for all strata but in terms of a recursion formula that we have not been able to turn into a closed formula; however, these formulas should have independent interest and we use them to get formulas for the E-O strata as follows. If $w \in W_g$ is minimal for the Bruhat–Chevalley order in its coset wS_g, then $\mathcal{U}_w$ maps by a finite étale map to the open E-O stratum $\mathcal{V}_\nu$ corresponding to the coset $\nu := wS_g$. We can compute the degree of this map in terms of the combinatorics of the element w and we then can push down our formulas for $\overline{\mathcal{U}}_w$ to obtain formulas for the cycle classes of the E-O strata. Also the formulas of Kresch and Tamvakis can be used to give the classes of E-O strata. One interesting general consequence (cf., Theorem 13.1) is that each class is a polynomial in the Chern classes λ_i of the Hodge bundle, the cotangent bundle of the zero-section of the universal abelian variety, and the coefficients are polynomials in p. This is a phenomenon already visible in the special cases of our formula that were known previously; the oldest such example being Deuring's mass formula for the number of supersingular elliptic curves (weighted by one over the cardinalities of their groups of automorphisms) that says that this mass is $(p-1)/12$. This appears in our context as the combination of the formula $(p-1)\lambda_1$ for the class of the supersingular locus and the formula $\deg \lambda_1 = 1/12$. We interpret these results as giving rise to elements in the *p-tautological ring*; this is the ring obtained from the usual *tautological ring*, the ring generated by the Chern classes of the Hodge bundle, by extending the scalars to $\mathbb{Z}\{p\}$, the localization of the polynomial ring $\mathbb{Z}[p]$ at the polynomials with constant coefficient 1. Hence we get elements parametrized by W_g/S_g in the p-tautological ring and we show that they form a $\mathbb{Z}\{p\}$-basis for the p-tautological ring. Putting p equal to 0 maps these elements to elements of the ordinary tautological ring that can be identified with the Chow ring of Sp_{2g}/P, and these elements are the usual classes of the Schubert varieties. It seems that these results call for a p-Schubert calculus in the sense of a better understanding of these elements of the p-tautological ring and their behavior under multiplication.

However, there seems to be a more intriguing problem. We have for each $w \in W_g$ a stratum in our flag space and these strata push down to elements of the p-tautological ring under the projection map to $\tilde{\mathcal{A}}_g$ (a suitable toroidal compactification of $\mathcal{A}_g$). When setting p to 0 these elements specialize to the classes of the images of the Schubert varieties of Sp_{2g}/B_g in Sp_{2g}/P, and for them the situation is very simple: either w is minimal in its S_g coset and then the Schubert variety maps birationally to the corresponding Schubert variety of Sp_{2g}/P or it is not and then it maps to 0. When it comes to the elements of the p-tautological ring this allows us to conclude only – in the nonminimal case – that the coefficients are divisible by p, and indeed in general, they are not zero. We show that unless they map to 0 they will always map to a multiple of a class of an E-O stratum. When the element w is minimal in its S_g-coset, this stratum is indexed by the coset spanned by w. Our considerations give an extension of this map from elements minimal in their cosets to a larger class of elements. We give some examples of this extension, but in general it seems a very mysterious construction.

Another application is to the irreducibility of our strata (and hence also to the strata of the E-O stratification, since they are images of some of our strata). Since the strata are normal, this is equivalent to the connectedness of a stratum, and this connectedness can sometimes (cf. Theorem 11.5) be proved via an arithmetic argument. It is natural to ask whether this method produces all the irreducible strata, and for the characteristic large enough (the size depending on g) we can show that indeed it does. This is done using a Pieri-type formula for our strata obtained by applying a result of Pittie and Ram. A Pieri-type formula for multiplying the class of a connected cycle by an ample line bundle has as a consequence that a part of the boundary is supported by an ample line bundle and hence that this part of the boundary is connected. Applying this to λ_1, which is an ample line bundle on $\mathcal{A}_g$, allows us to show that our results are optimal; cf. [Ha07]. We are forced to assume that the characteristic is large (and are unable to specify how large), since we do not know by which power of λ_1 one needs to twist the exterior powers of the dual of the Hodge bundle to make these generated by global sections.

There is a particular element of $w_\emptyset \in W_g$ that is the largest of the elements that are minimal in their right S_g-cosets and that has the property that $\overline{\mathcal{U}}_{w_\emptyset}$ maps generically of finite degree onto $\mathcal{A}_g$. It is really the strata that are contained in this stratum that seem geometrically related to $\mathcal{A}_g$, and indeed the elements $w \in W_g$ lying below $w_\emptyset$ are the ones of most interest. (The rest of $\mathcal{F}_g$ appears mostly as a technical device for relating our strata to the Schubert varieties.) It should be of particular interest to understand the composite map $\overline{\mathcal{U}}_{w_\emptyset} \subset \mathcal{F}_g \to \mathcal{A}_g$. It follows from a result of Oort on Dieudonné modules that the inverse image of an open E-O stratum under this map is a locally constant fibration. This focuses interest on its fibers, a fiber depending only on the element of $\nu \in W_g/S_g$ that specifies the E-O stratum. We call these fibers *punctual flag spaces* (see Section 9 for details). We determine their connected components, showing in particular that two points in the same connected component can be connected by a sequence of quite simple rational curves. We also show that knowing which strata $\mathcal{U}_w$ have nonempty intersections with a given punctual flag space would determine the inclusion relations between the E-O strata.

A geometric point x of the stratum $\overline{\mathcal{U}}_{w_\emptyset}$ corresponds to a symplectic flag of subgroup schemes of the kernel of multiplication by p on the principally polarized abelian variety that is the image of x in $\mathcal{A}_g$ under the map $\mathcal{F}_g \to \mathcal{A}_g$. This is reminiscent of de Jong's moduli stack $\mathcal{S}(g,p)$ of $\Gamma_0(p)$-structures; cf. [Jo93]. The major difference (apart from the fact that $\overline{\mathcal{U}}_{w_\emptyset}$ makes sense only in positive characteristic) is that the g-dimensional element of the flag is determined by the abelian variety in our case. We shall indeed identify $\overline{\mathcal{U}}_{w_\emptyset}$ with the component of the fiber at p of $\mathcal{S}(g,p)$ that is the closure of the ordinary abelian varieties provided with a flag on the local part of the kernel of multiplication by p. As a consequence we get that that component of $\mathcal{S}(g,p)$ is normal and Cohen–Macaulay.

This paper is clearly heavily inspired by [Oo01]. The attentive reader will notice that we re-prove some of the results of that paper, sometimes with proofs that are very close to the proofs used by Oort. We justify such duplications by our desire to emphasize the relations with the combinatorics of W_g and the flag spaces. Hence, we start with (a rather long) combinatorial section in which the combinatorial aspects have been separated from the geometric ones. We hope that this way of presenting the material will be as clarifying to the reader as it has been to us. We intend to continue to exploit the approach using the flag spaces in a future paper that will deal with K3 surfaces. Since its announcement in [Ge99], our idea of connecting the E-O stratification on $\mathcal{A}_g$ with the Weyl group and filtrations on the de Rham cohomology has been taken up in other work. In this connection we want to draw attention to papers by Moonen and Wedhorn; cf. [Mo01, MW04].

We would like to thank Piotr Pragacz for some useful comments.

Conventions. We shall exclusively work in positive characteristic $p > 0$ (note, however, that in Section 13 the symbol p will also be a polynomial variable). After having identified final types and final elements in Section 2 we shall often use the same notation for the final type (which is a function on $\{1,\dots,2g\}$) and the corresponding final element (which is an element of the Weyl group W_g). In Sections 10 and 11 our strata will be considered in flag spaces over not just $\mathcal{A}_g$ and $\tilde{\mathcal{A}}_g$ but also over the corresponding moduli stacks with a level structure. We shall define several natural objects, such as the Hodge bundle, over several different spaces (such as the moduli space of abelian varieties as well as toroidal compactifications of it). In order not to make the notation overly heavy, the same notation (such as $\mathbb{E}$) will normally be used for the objects on different spaces. Since the objects in question will be compatible with pullback, this should not cause confusion. Sometimes, when there might still be such a risk we shall use subscripts (such as $\mathbb{E}_{\mathcal{A}_g}$ and $\mathbb{E}_{\tilde{\mathcal{A}}_g}$) to distinguish between them. Also, when there is no risk of confusion we shall use the same name (like $\mathcal{U}_w$ and $\overline{\mathcal{U}}_w$) for a stratum and its extension to the compactified moduli space.

Our moduli objects such as $\mathcal{A}_g$ are really Deligne–Mumford stacks. However, in order to avoid what we have found to be a sometimes awkward terminology (such as “flag stacks”), we shall usually speak of them as spaces rather than stacks. In a similar vein, by for instance a “locally closed subset” of an algebraic stack we shall mean a reduced locally closed substack.

2 Combinatorics

This section is of a preparatory nature and deals with the combinatorial aspects of the E-O stratification. The combinatorics is determined by the Weyl group of the symplectic group of degree g. A general reference for the combinatorics of Weyl groups is [BL00]. We start by recalling some general notation and facts about W_g and its Bruhat–Chevalley order. We then go on to give various descriptions of the minimal elements in the S_g cosets (which we presume are well known). The short subsection on shuffles will be used to understand the rôle that the multiplicative and étale parts of the Barsotti–Tate group play in our stratification in the case of positive p-rank.

2.1 Final Elements in the Weyl Group

The Weyl group W_g of type C_g in Cartan's terminology is isomorphic to the semidirect product $S_g \ltimes (\mathbb{Z}/2\mathbb{Z})^g$, where the symmetric group S_g on g letters acts on $(\mathbb{Z}/2\mathbb{Z})^g$ by permuting the g factors. Another description of this group, and the one we shall use here, is as the subgroup of the symmetric group S_{2g} formed by elements that map any symmetric 2-element subset of $\{1,\dots,2g\}$ of the form $\{i, 2g+1-i\}$ to a subset of the same type:

$$W_g = \{\sigma \in S_{2g} \colon \sigma(i) + \sigma(2g+1-i) = 2g+1 \text{ for } i = 1,\dots,g\}.$$

The function $i \mapsto 2g+1-i$ on the set $\{1,\dots,2g\}$ will occur frequently. We shall sometimes use the notation $\bar{\imath}$ for $2g+1-i$. Using it we can say that $\sigma \in S_{2g}$ is an element of W_g precisely when $\sigma(\bar{\imath}) = \overline{\sigma(i)}$ for all i. This makes the connection with another standard description of W_g, namely as a group of *signed permutations*. An element w in this Weyl group has a *length* and a *codimension* defined by

$$\ell(w) = \#\{i < j \leq g : w(i) > w(j)\} + \#\{i \leq j \leq g : w(i) + w(j) > 2g+1\},$$

and

$$\mathrm{codim}(w) = \#\{i < j \leq g \colon w(i) < w(j)\} + \#\{i \leq j \leq g \colon w(i) + w(j) < 2g+1\}$$

and these satisfy the equality

$$\ell(w) + \mathrm{codim}(w) = g^2.$$

We shall use the following notation for elements in W_g. By $[a_1, a_2, \dots, a_{2g}]$ we mean the permutation of $\{1,2,\dots,2g\}$ with $\sigma(i) = a_i$. Since $\sigma(i)$ determines

$\sigma(2g+1-i)$ for $1 \leq i \leq g$, sometimes we use the notation $[a_1, \ldots, a_g]$ instead (when the a_i are single digits we shall often dispense with the commas and write $[a_1 \ldots a_g]$, which should cause no confusion). We shall also use cycle notation for permutations. In particular, for $1 \leq i < g$ we let $s_i \in S_{2g}$ be the permutation $(i, i+1)(2g-i, 2g+1-i)$ in W_g, which interchanges i and $i+1$ (and then also $2g-i$ and $2g+1-i$) and we let $s_g = (g, g+1) \in S_{2g}$. Then the pair $(W = W_g, S = \{s_1, \ldots, s_g\})$ is a Coxeter system.

Let (W, S) be a Coxeter system and $a \in W$. If X is a subset of S we denote by W_X the subgroup of W generated by X. It is well known that for any subset X of S there exists precisely one element w of minimal length in aW_X and it has the property that every element $w' \in aW_X$ can be written in the form $w' = wx$ with $x \in W_X$ and $\ell(w') = \ell(w) + \ell(x)$. Such an element w is called an *X-reduced element*; cf. [GrLie4-6, Chapter IV, Exercises §1].

Let $W = W_g$ be the Weyl group and S the set of simple reflections. If we take $X = S \backslash \{s_g\}$, then we obtain

$$W_X = \{\sigma \in W_g \colon \sigma\{1, 2, \ldots, g\} = \{1, 2, \ldots, g\}\} \cong S_g.$$

There is a natural partial order on W_g with respect W_X, the Bruhat–Chevalley order. It is defined in terms of Schubert cells $X(w_i)$ by

$$w_1 \leq w_2 \iff X(w_1) \subseteq X(w_2).$$

Equivalently, if for $w \in W_g$ we define

$$r_w(i,j) := \#\{a \leq i \colon\ w(a) \leq j\}, \tag{1}$$

then we have the combinatorial characterization

$$w_1 \leq w_2 \iff r_{w_1}(i,j) \geq r_{w_2}(i,j) \quad \text{for all } 1 \leq i, j \leq 2g.$$

(Indeed, it is easy to see that it is enough to check this for all $1 \leq i \leq g$ and $1 \leq j \leq 2g$.) Chevalley has shown that $w_1 \geq w_2$ if and only if any (hence every) X-reduced expression for w_1 contains a subexpression (obtained by just deleting elements) that is a reduced expression for w_2; here reduced means that w_2 is written as a product of $\ell(w_2)$ elements of S. Again, a reference for these facts is [BL00].

We now restrict to the following case. Let V be a symplectic vector space over $\mathbb{Q}$ and consider the associated algebraic group $G = \mathrm{Sp}(V)$. If $E \subset V$ is a maximal isotropic subspace, then the stabilizer of the flag $(0) \subset E \subset V$ is a parabolic subgroup conjugate to the standard parabolic subgroup corresponding to X. Since X is a heavily used letter we shall use $H := S \backslash \{s_g\} \subset S$ instead. Hence W_H will denote the subgroup of W_g generated by the elements of H and we will also use the notation P_H for the parabolic subgroup corresponding to W_H, i.e., the subgroup of the symplectic group stabilizing a maximal totally isotropic subspace. Note that W_H consists of the permutations of W_g that stabilize the subsets $\{1, \ldots, g\}$ and $\{g+1, \ldots, 2g\}$ and that the restriction of the action of an element of W_H to $\{1,\dots,g\}$ determines the full permutation. Therefore, we may identify W_H with S_g, the group of permutations of $\{1,\dots,g\}$, and we shall do so without further notice. (This is of course compatible with the fact that H spans an A_{g-1}-subdiagram of the Dynkin diagram of G.) There are $2^g = |W_g|/|W_H|$ elements in W_g that are H-reduced elements. These 2^g elements will be called *final* elements of W_g. The Bruhat–Chevalley order between elements in W_g as well as the condition for being H-reduced can be conveniently expressed in terms of the concrete representation of elements of W_g as permutations in the following way.

Let A, B be to two finite subsets of $\{1,2,\dots,g\}$ of the same cardinality. We shall write $A \prec B$ if for all $1 \le i \le |A|$ the ith-largest element of A is at most equal to the ith-largest element of B.

Lemma 2.1. *(i) If $w = [a_1 a_2 \dots a_g]$ and $w' = [b_1 b_2 \dots b_g]$ are two elements of W_g, then $w \le w'$ in the Bruhat–Chevalley order precisely when for all $1 \le d \le g$ we have $\{a_1, a_2, \dots, a_d\} \prec \{b_1, b_2, \dots, b_d\}$.*

(ii) Let $w = [a_1 a_2 \dots a_g] \in W_g$. Denote the final element of wW_H by w^f. For $w' = [b_1 b_2 \dots b_g] \in W_g$ we have $w^f \le w'$ in the Bruhat–Chevalley order precisely when $\{a_1, a_2, \dots, a_g\} \prec \{b_1, b_2, \dots, b_g\}$.

(iii) An element $\sigma \in W_g$ is H-reduced *(or* final*) if and only if $\sigma(i) < \sigma(j)$ for all $1 \le i < j \le g$. Also, σ is H-reduced if and only if σ sends the first $g-1$ simple roots into positive roots.*

Proof. See for instance [BL00, p. 30]. □

2.2 Final Types and Young Diagrams

There are other descriptions of final elements that are sometimes equally useful. They involve maps of $\{1,2,\dots,2g\}$ to $\{1,2,\dots,g\}$ and certain Young diagrams. We begin with the maps.

Definition 2.2. *A* final type *(of degree g) is a nondecreasing surjective map*

$$\nu\colon \{0,1,2,\dots,2g\} \to \{0,1,2,\dots,g\}$$

satisfying

$$\nu(2g-i) = \nu(i) - i + g \qquad \text{for} \quad 0 \le i \le g.$$

We always have $\nu(0) = 0$ and $\nu(2g) = g$. Note that we have either $\nu(i+1) = \nu(i)$ and then $\nu(2g-i) = \nu(2g-i-1)+1$ or $\nu(i+1) = \nu(i)+1$ and then $\nu(2g-i) = \nu(2g-i-1)$. A final type is determined by its values on $\{0,1,\dots,g\}$. There are 2^g final types of degree g corresponding to the vectors $(\nu(i+1)-\nu(i))_{i=0}^{g-1} \in \{0,1\}^g$. The notion of a final type was introduced by Oort [Oo01].

To an element $w \in W_g$ we can associate the final type ν_w defined by

$$\nu_w(i) := i - r_w(g,i).$$

This is a final type because of the rule $r_w(g, 2g-i) - r_w(g,i) = g-i$, which follows by induction on $i \in \{1, \ldots, g\}$ from the fact that $w(2g+1-a) = 2g+1-w(a)$. It depends only on the coset wW_H of w, since a permutation of $\{a : 1 \leq a \leq g\}$ does not change the definition of

$$r_w(g,i) = \#\{a \leq g \colon w(a) \leq i\}.$$

Conversely, to a final type ν we now associate an element w_ν of the Weyl group, a permutation of $\{1, 2, \ldots, 2g\}$, as follows. Let

$$\beta = \{i_1, i_2, \ldots, i_k\} = \{1 \leq i \leq g \colon \nu(i) = \nu(i-1)\}$$

with $i_1 < i_2 < \cdots$ given in *increasing* order and let

$$\beta^c = \{j_1, j_2, \ldots, j_{g-k}\}$$

be the elements of $\{1, 2, \ldots, g\}$ not in ξ, in *decreasing* order. We then define a permutation w_ν by mapping $1 \leq s \leq k$ to i_s and $k+1 \leq s \leq g$ to $2g+1-j_{s-k}$. The requirement that w_ν belong to W_g now completely specifies w_ν and by construction $w_\nu(i) < w_\nu(j)$ if $1 \leq i < j \leq g$. Thus w_ν is a final element of W_g. It is clear from Lemma 2.1 that we get in this way all final elements of W_g. The Bruhat–Chevalley order for final elements can also be read off from the final type ν. We have $w \geq w'$ if and only if $\nu_w \geq \nu_{w'}$. This follows from Lemma 2.1, (ii).

We summarize:

Lemma 2.3. *By associating to a final type ν the element w_ν and to a final element $w \in W_g$ the final type ν_w we get an order-preserving bijection between the set of 2^g final types and the set of final elements of W_g.*

The final types are in bijection with certain Young diagrams. Our Young diagrams will be put in a position that is opposite to the usual positioning, i.e., larger rows will be below smaller ones and the rows will be lined up to the right (see next example). Furthermore, we shall make Young diagrams correspond to partitions by associating to a diagram the parts that are the lengths of its rows. Given g we shall say that a Young diagram is *final of degree g* if its parts are $\leq g$ and no two parts are equal. They therefore correspond to subsets ξ of $\{1, 2, \ldots, g\}$. We write such a ξ as $\{\xi_1, \ldots, \xi_r\}$ with $g \geq \xi_1 > \cdots > \xi_r$.

To a final type ν we now associate the Young diagram Y_ν whose associated subset ξ is defined by

$$\xi_j = \#\{i \colon 1 \leq i \leq g,\, \nu(i) \leq i - j\}.$$

A pictorial way of describing the Young diagram is by putting a stack of $i - \nu(i)$ squares in vertical position i for $1 \leq i \leq g$.

Example 2.4. This example corresponds to

$$\{\nu(i) \colon i = 1, \ldots, g\} = \{1, 2, \ldots, g-5, g-5, g-4, g-4, g-3, g-3\}$$

and hence $\xi = \{5, 3, 1\}$ (see Figure 1).

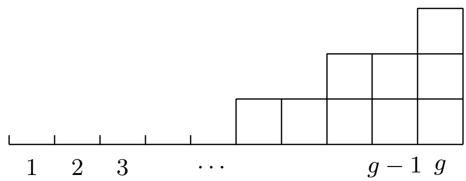


Fig. 1. Young diagram with $\xi = \{5, 3, 1\}$.

The final elements w in W_g are in one-to-one correspondence with the elements of W_g/W_H. The group W_g acts on W_g/W_H by multiplication on the left, i.e., by the permutation representation. Therefore W_g also acts on the set of final types and the set of final Young diagrams. To describe these actions we need the notion of a break point.

By a *break point* of a final type ν we mean an integer i with $1 \leq i \leq g$ such that either

1. $\nu(i-1) = \nu(i) \neq \nu(i+1)$, or
2. $\nu(i-1) \neq \nu(i) = \nu(i+1)$.

If i is not a break point of $\nu = \nu_w$ then either $\nu(i+1) = \nu(i-1)$ or $\nu(i+1) = \nu(i-1) + 2$ and therefore $\nu_{s_i w} = \nu_w$. In particular, g is always a break point. The set of break points of ν is

$$\{1 \leq i \leq g \colon \nu_{s_i w} \neq \nu_w\}.$$

Since $\nu = \nu_w$ determines a coset wW_H, we have that i is not a break point of ν if and only if $w^{-1} s_i w \in W_H$, i.e., if and only if wW_H is a fixed point of s_i acting on W_g/W_H. The action of s_i on a final type ν is as follows: if i is not a break point then ν is fixed; otherwise, replace the value of ν at i by $\nu(i) + 1$ if $\nu(i-1) = \nu(i)$ and $\nu(i) - 1$ otherwise.

If w is a final element given by the permutation $[a_1, a_2, \ldots, a_g]$, then it defines a second final element, called the *complementary permutation*, defined by the permutation $[b_1, b_2, \ldots, b_g]$, where $b_1 < b_2 < \cdots < b_g$ are the elements of the complement $\{1, 2, \ldots, 2g\} \backslash \{a_1, \ldots, a_g\}$. If ξ is the partition defining the Young diagram of w then ξ^c defines the Young diagram of the complementary permutation. The set of break points of w (that is, of the corresponding ν) and its complementary element are the same.

Lemma 2.5. *Let $w \in W_g$ be a final element with associated final type ν and complementary element v.*

(i) We have that $v = \sigma_1 w \sigma_0 = w \sigma_1 \sigma_0$, where σ_0 (respectively σ_1) is the element of S_g (respectively W_g) that maps i with $1 \leq i \leq g$ to $g + 1 - i$ (resp. to $2g + 1 - i$).

(ii) Let $i \in \{1, \ldots, g\}$. If $\nu(i-1) \neq \nu(i)$ then $v^{-1}(i) = \nu(i)$ and if $\nu(i-1) = \nu(i)$ then $v^{-1}(i) = 2g + 1 - \nu(2g + 1 - i)$.

Proof. Since w maps i to a_i we have that $2g+1-a_i$ is not among the a_j and hence $b_i = 2g+1-a_{g+1-i}$ (using that both the a_i and b_i are increasing sequences). This gives $\sigma_1 w \sigma_0(i) = \sigma_1(a_{g+1-i}) = 2g+1-a_{g+1-i} = b_i$, but we note that since $w \in W_g$, it commutes with σ_1. Thus (i) holds.

If $\nu(i-1) \neq \nu(i)$ and, say, $\nu(i) = i-k$ then we have k natural numbers $1 \leq i_1 < i_2 < \cdots < i_k < i$ such that $\nu(i_j - 1) = \nu(i_j)$. By the definition of v we then have $v(i-k) = (i-k)+k$, since the k values i_j $(j = 1, \ldots, k)$ are values for v, hence not of w. The second part is checked similarly. □

Remark 2.6. The elements σ_1 and σ_0 of course have clear root-theoretic relevance: they are respectively the longest elements of W_g and S_g. Multiplication by $\sigma_1\sigma_0$ reverses the Bruhat–Chevalley order. Similarly it is clear that going from a final element to its complementary element also reverses the Bruhat–Chevalley order among the final elements and the first part of our statement says that that operation is obtained by multiplying by σ_1 and σ_0. Somewhat curiously, our use of the complementary permutation seems unrelated to these facts.

In terms of Young diagrams the description is analogous and gives us a way to write the element w_ν as a reduced product of simple reflections. To each s_i we can associate an operator on final Young diagrams. If Y is a final diagram, s_i is defined on Y by adding or deleting a box in the ith column if this gives a final diagram (only one of the two can give a final diagram) and then s_iY will be that new diagram; if neither adding nor deleting such a box gives a final Young diagram we do nothing. In terms of the description as subsets ξ, adding a box corresponds to $g+1-i \in \xi$ and $g+2-i \notin \xi$ and then $s_i\xi = (\xi \setminus \{g+1-i\}) \cup \{g+2-i\}$. It is then clear that for any final Young diagram Y there is a word $s_{i_1}s_{i_2}s_{i_3}\cdots s_{i_k}$ such that $Y = s_{i_1}s_{i_2}s_{i_3}\cdots s_{i_k}\emptyset$, where $\emptyset$ denotes the empty Young diagram. Comparison with the action of s_i on final types and the correspondence between final types and Young diagrams shows that the action of s_i on diagrams is indeed obtained from that on final types. If we now have a word $t = s_{i_1}\cdots s_{i_k}$ in the s_i we can make it act on Young diagrams by letting each individual s_i act as specified. Note that this action depends only on the image of t in W_g, but for the moment we want to consider the action by words. We define the *area* of a Young diagram Y to be the number of boxes it contains. We shall say that the word t is *building* if the area of $t\emptyset$ is equal to k, the length of the *word* (not the resulting element). This is equivalent to the action of s_{i_r} adding a box to $s_{i_{r+1}}\cdots s_{i_k}\emptyset$ for all r.

Lemma 2.7. *(i) If ν is a final type and t is a word in the s_i such that $Y_\nu^c = t\emptyset$ then $w_\nu = w$, where w is the image of t in W_g and $\ell(w_\nu) = g(g+1)/2 - area(Y_\nu)$.*

(ii) t is H-reduced if and only if t is building.

Proof. To prove (i) we begin by noting that $t\emptyset$ depends only on the image of t in W_g, so that (i) is independent of the choice of t. Hence we may prove it by choosing a particular t using induction on the area of Y_ν^c. Note that $g(g+1)/2-\text{area}(Y_\nu) = \text{area}(Y_\nu^c)$, so that the last part of (i) says that $\ell(w_\nu) = \text{area}(Y_\nu^c)$. The final type ν with $\nu(i) = 0$ for $i \leq g$ corresponds to a final diagram Y_ν with an empty complementary diagram. We have $w_\nu = 1 \in W_g$, the empty product, and it has length 0. This proves the base case of the induction. Suppose we have proved the statement for diagram Y_ν with $\text{area}(Y_\nu^c) \leq a$. Adding one block to Y_ν^c to obtain $Y_{\nu'}^c$ means that for some i we have $g+1-i \in \xi^c$ and $g+2-i \notin \xi^c$, where ξ^c is the subset corresponding to Y_ν^c, and the new subset is $(\xi')^c = (\xi \setminus \{g-i\}) \cup \{g-i+1\}$. This means that if $i < g$ there are $b < a \leq g$ such that $w_\nu(b) = i$, $w_\nu(a) = 2g-i$, $w_{\nu'}(b) = i+1$, and $w_{\nu'}(a) = 2g+1-i$, and the rest of the integers between 1 and g remain unchanged. (The case $i = g$ is similar and left to the reader.) This makes it clear that we have $w_{\nu'} = s_i w_\nu$, so by the induction t maps to w_ν. It remains to establish the formula for $\ell(w_\nu)$. In the definition of $\ell(w_\nu)$ only the second term contributes, since $w_\nu(i) < w_\nu(j)$ if $i < j \leq g$. Now, the only difference in the collections of sums $w(i) + w(j)$ for $i \leq j$ and w equal to w_ν and $w_{\nu'}$ appears for $(i,j) = (b,a)$, and we have $w_\nu(b) + w_\nu(a) = 2g$ and $w_{\nu'}(b) + w_{\nu'}(a) = 2g+2$, so that the length of $w_{\nu'}$ is indeed one larger than that of w_ν.

As for (ii), we have that $t\emptyset = Y_\nu^c$, where ν is the final type of w and then (ii) is equivalent to t being H-reduced if and only if $\text{area}(Y_\nu^c)$ is equal to the length of t. However, by (i) we know that $\text{area}(Y_\nu^c)$ is equal to $\ell(w_\nu)$, and t is indeed H-reduced precisely when its length is equal to $\ell(w_\nu)$. □

Example 2.8. Consider again the Young diagram of the previous example but now for $g = 5$ (see Figure 2). We have $\xi = \{5,3,1\}$ and thus $\xi^c = \{2,4\}$, so $w_\nu = [13579]$ and w_ν can be written as $s_4 s_5 s_2 s_3 s_4 s_5$ (we emphasize that permutations act from the left on diagrams).

We now characterize final types. Besides the function $\nu = \nu_w$ defined by

$$\nu(i) = i - \#\{a \leq g \colon w(a) \leq i\} = i - r_w(g,i)$$

and extended by $\nu(2g-i) = \nu(i) - i + g$ for $i = 0, \ldots, g$, we define a function $\mu = \mu_w$ on the integers $1 \leq i \leq 2g$ by

$$\mu(i) := \left(\max\{w^{-1}(a) \colon 1 \leq a \leq i\} - g\right)^+,$$

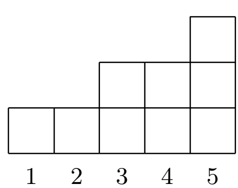


Fig. 2. Young diagram with $g = 3$ and $\xi = \{5, 3, 1\}$.

where $(x)^+ := \max(x, 0)$. Alternatively, we have

$$\mu(i) = \min \{0 \le j \le g : r_w(g+j, i) = i\}.$$

Note that both μ and ν are nondecreasing functions taking values between 0 and g. Also for $i = 1, \ldots, 2g-1$ we have $\nu(i+1) = \nu(i)$ or $\nu(i+1) = \nu(i)+1$ and $\nu(2g) = \mu(2g) = g$. If w is a final element then ν_w is the final type associated to w. For an arbitrary w the function ν is the final type of the final element in the coset wS_g.

Lemma 2.9. *We have $\mu_w(i) \ge \nu_w(i)$ for $1 \le i \le 2g$ with equality precisely when w is a final element and then ν_w is the final type of w.*

Proof. We suppress the index w and first prove the inequality $\mu \ge \nu$. Let $1 \le i \le g$. Suppose that $\mu(i) = m$, i.e., the maximal j with $w(j)$ in $[1, i]$ is $g+m$. Then there are at most m elements from $[g+1, 2g]$ that map under w into $[1, i]$ and there are at least $i - m$ elements from $[1, g]$ with their image under w in $[1, i]$, so $i - \nu(i) \ge i - m$; in other words, $\nu(i) \le \mu(i)$. For i in the interval $[g+1, 2g]$ we consider $\nu(2g-i) = \#\{a \le g \colon w(a) > i\}$. If $\mu(2g-i) = m$ then there are at least $g - m$ elements from $[1, g]$ mapping into $[1, 2g-i]$ and thus $\nu(2g-i)$ is at most equal to m.

If w is final then w respects the order on $[1, g]$, and this implies that if $\#\{a \le g \colon w(a) \le t\} = n$ then $t - n$ elements from $[g+1, 2g]$ map under w to $[1, t]$, so the maximum element from $[g+1, 2g]$ mapping into $[1, t]$ is $g+t-n$. Hence $\mu(t) = t - n = \nu(t)$.

Conversely, if $\mu(i) = \nu(i)$ then this guarantees that $w(i) < w(j)$ for all pairs $1 \le i < j \le g$. Thus w is a final element. □

Corollary 2.10. *Let $w \in W_g$. Then w is a final element if and only if $r_w(g + \nu_w(i), i) = i$ for all $1 \le i \le g$.*

Proof. Lemma 2.9 says that if w is final then we have

$$\nu(i) = \mu(i) = \min \{0 \le j \le g : r_w(g+j, i) = i\}$$

and in particular that $\nu(i) \in \{0 \le j \le g : r_w(g+j, i) = i\}$, which gives one direction.

Conversely, if we have $r_w(g + \nu_w(i), i) = i$, then $\nu_w(i) \ge \mu_w(i)$, and then Lemma 2.9 gives that w is final. □

2.3 Canonical Types

We now deal with an iterative way of constructing the function ν starting from its values on the endpoints and applying it repeatedly.

A final type ν is given by specifying $\nu(j)$ for $j = 1, \ldots, 2g$. But it suffices to specify the values of ν for the break points of ν. Under ν an interval $[i_1, i_2]$ between two consecutive break points of ν is mapped either to an interval of length $i_2 - i_1$ or to one point. However, the image points $\nu(i_1)$ and $\nu(i_2)$ need not be break points of ν. Therefore we enlarge the set of break points to a larger set C_ν, called the *canonical domain*. We define C_ν to be the smallest subset of $\{0, 1, \dots, 2g\}$ containing 0 and $2g$ such that if $j \in C_\nu$ then also $2g - j \in C_\nu$ and if $j \in C_\nu$ then $\nu(j) \in C_\nu$. It is obtained by starting from $R = \{0, 2g\}$ and adding the values $\nu(k)$ and $\nu(2g-k)$ for $k \in R$ and continuing till this stabilizes. The restriction of ν to C_ν is called a *canonical type*. We wish to see that the canonical domain C_ν contains the break points of ν and hence that we can retrieve ν from the canonical type of ν. To see this we need a technical lemma (its formulation is somewhat obscured by the fact that we also want to use it in another slightly different context).

Definition-Lemma 2.11. *We shall say that a subset $S \subseteq \{0, 1, \dots, 2g\}$ is* stable *if it has the property that it contains* 0 *and is stable under $i \mapsto i^\perp := 2g - i$. For a stable subset S a map $f\colon S \to S \cap \{0, 1, \dots, g\}$ is* adapted to S *if $f(0) = 0$ and $f(2g) = g$, if it is* contracting, *i.e., it is increasing and we have $f(j) - f(i) \leq j - i$ for $i < j$ and if it fulfills the following* complementarity condition*: For any two consecutive $i, j \in S$ (i.e., $i < j$ and there are no $k \in S$ with $i < k < j$) we have $f(j) - f(i) = j - i \Rightarrow f(j^\perp) = f(i^\perp)$.*

(i) If S is stable and f is a nonsurjective function adapted to S then there is a proper subset $T \subset S$ such that $f_{|T}$ is adapted to T.

(ii) If S is stable and f is a surjective function adapted to S then for any two consecutive $i, j \in S$ we have either $f(i) = f(j)$ or $f(j) - f(i) = j - i$.

(iii) We say that (S, f) is minimally stable *if S is stable and f is adapted to S and furthermore there is no proper stable subset $T \subset S$ for which $f_{|T}$ is adapted to it, then the function $\nu\colon \{1, 2, \dots, 2g\} \to \{1, 2, \dots, g\}$ obtained from f by extending it linearly between any two consecutive $i, j \in S$ is a final type, $S = C_\nu$, and ν is the unique final extension of f. Conversely, if f is the canonical type of a final type ν, then (C_ν, f) is minimally stable and in particular ν is the linear extension of its canonical type.*

Proof. For (i) consider $T = f(S) \cup (f(S))^\perp$. It is clearly stable under f and $\perp$ and contains 0. If f is not surjective, then T is a proper subset of S.

Assume now that we are in the situation of (ii). We show that if $i < j \in S$ are consecutive then either $f(j) - f(i) = j - i$ or $f(i) = f(j)$ by descending induction on $j - i$.

By induction we are going to construct a sequence $i_k < j_k \in S$, $k = 1, 2, \dots$, of consecutive elements such that either $(i_{k-1}, j_{k-1}) = (j_k^\perp, i_k^\perp)$ or $(f(i_k), f(j_k)) = (i_{k-1}, j_{k-1})$ but not both $(i_{k-1}, j_{k-1}) = (j_k^\perp, i_k^\perp)$ and $(i_{k-2}, j_{k-2}) = (j_{k-1}^\perp, i_{k-1}^\perp)$ and in any case $j_k - i_k = j - i$. We start by putting $i_1 := i, j_1 := j$. Assume now that $i_k < j_k$ have been constructed. If we do not have $i_k, j_k \leq g$, then since $g \in S$, we must have $j_k^\perp, i_k^\perp \leq g$ and then we put $(i_{k+1}, j_{k+1}) = (j_k^\perp, i_k^\perp)$. If we do have $i_k, j_k \leq g$ then by the surjectivity of f there are $i_{k+1}, j_{k+1} \in S$ such that $f(i_{k+1}) = i_k$ and $f(j_{k+1}) = j_k$. Since f is increasing, $i_{k+1} < j_{k+1}$, and by choosing i_{k+1} to be maximal and j_{k+1} to be minimal we may assume that they are neighbors. We must have that $j_{k+1} - i_{k+1} = j_1 - i_1$. Indeed, we have $f(j_{k+1}) - f(i_{k+1}) \leq j_k - i_k$, since f is contracting. If we have strict inequality we have $j - i = j_k - i_k < j_{k+1} - i_{k+1}$, and hence by the induction assumption we have either that $j_k - i_k = f(j_{k+1}) - f(i_{k+1}) = j_{k+1} - i_{k+1}$, which is a contradiction, or that $j_k = f(j_{k+1}) = f(i_{k+1}) = i_k$, which is also a contradiction. Hence we have $j_{k+1} - i_{k+1} = j_k - i_k = j - i$ and we have verified the required properties of (i_{k+1}, j_{k+1}).

There must now exist $1 \leq k < \ell$ such that $(i_k, j_k) = (i_\ell, j_\ell)$, and we pick k minimal for this property. If $k = 1$ we have either $j - i = j_{\ell-1} - i_{\ell-1} = f(j_\ell) - f(i_\ell) = f(j) - f(i)$ or $j - i = j_{\ell-2} - i_{\ell-2} = f(j_{\ell-1}) - f(i_{\ell-1}) = f(i^\perp) - f(j^\perp)$, which implies that $f(i) = f(j)$ by assumptions on f. We may hence assume that $k > 1$. We cannot have both $(i_{k-1}, j_{k-1}) = (j_k^\perp, i_k^\perp)$ and $(i_{\ell-1}, j_{\ell-1}) = (j_\ell^\perp, i_\ell^\perp)$, since that would contradict the minimality of k. If $(i_{k-1}, j_{k-1}) = (j_k^\perp, i_k^\perp)$ and $(i_{\ell-1}, j_{\ell-1}) = (f(i_\ell), f(j_\ell))$ then we get $j_{k-1} - i_{k-1} = j - i = j_{\ell-1} - i_{\ell-1} = f(i_{k-1}^\perp) - f(j_{k-1}^\perp)$, which implies $f(j_{k-1}) = f(i_{k-1})$, which is either what we want in case $k = 2$ or a contradiction. Similarly, the case $(i_{\ell-1}, j_{\ell-1}) = (j_\ell^\perp, i_\ell^\perp)$ and $(i_{k-1}, j_{k-1}) = (f(i_k), f(j_k))$ leads to a contradiction, as does the case $(i_{\ell-1}, j_{\ell-1}) = (f(i_\ell), f(j_\ell))$ and $(i_{k-1}, j_{k-1}) = (f(i_k), f(j_k))$.

Finally, to prove the first part of (iii) we note that by (ii), for $i < j \in S$ consecutive we have either $f(i) = f(j)$ or $f(j) - f(i) = j - i$. This means that the linear extension ν of f has the property that for $1 \leq i \leq 2g$ we have either $\nu(i) = \nu(i-1)$ or $\nu(i) = \nu(i-1) + 1$, and if $\nu(i) = \nu(i-1) + 1$ we get by the conditions on f that $\nu(2g - i + 1) = \nu(2g - i)$. If for some i we have $\nu(i) = \nu(i-1)$ and $\nu(2g-i+1) = \nu(2g-i)$, we get that $g = f(2g) = \nu(2g) < g$, which is impossible by assumption, and hence ν is indeed a final type. It is clear that S fulfills the defining property of C_ν, so that $S = C_\nu$. The conditions on a final element imply that $\nu(j) - \nu(i) \leq j - i$ for $j < i$ and thus that f has a unique final extension.

Conversely, if ν is a final type then C_ν clearly fulfills the required conditions and we have just noted that $\nu(j) - \nu(i) \leq j - i$ for $j < i$. The complementarity condition for $f|C_\nu$ follows from the equivalence $\nu(i) = \nu(i-1) \Longleftrightarrow \nu(2g - i + 1) = \nu(2g - i) + 1$. □

We now give an interpretation of the canonical domain in terms of the Weyl group. Let $v \in W_g$ be a final element. A *canonical fragment* of v is an interval $]i, j] := [i + 1, \ldots, j] \subseteq \{1, 2, \ldots, 2g\}$ that is maximal with respect to the requirement that for all $k \geq 1$ the set $v^k(]i, j])$ be an interval.

Proposition 2.12. *Let $v \in W_g$ be a final element, w its complementary element, and ν the final type of w.*

(i) The set $\{1, 2, \ldots, 2g\}$ is the disjoint union of the canonical fragments of v. Moreover, the canonical fragments of v are permuted by v.
(ii) If $]i, j]$ is a canonical fragment of v, and if $\nu(j) \neq \nu(j-1)$, then ν maps $]i, j]$ bijectively to $]\nu(i), \nu(j)]$.

(iii) If $]i,j]$ *is a canonical fragment of* v*, then so is* $]\bar{\jmath},\bar{\imath}]$*.*
(iv) The upper endpoints of the canonical fragments of v *together with* 0 *form exactly the canonical domain for* w*.*

Proof. If two canonical fragments I and J meet, their union K will be an interval, and since $v^k(K) = v^k(I) \cup v^k(J)$, we see that $v^k(K)$ will be an interval for all k. By the maximality we get that $I = J$. On the other hand, $]i-1,i]$ fulfills the stability condition, so that i lies in a fragment. Hence $\{1,2,\dots,2g\}$ is the disjoint union of fragments.

Now let R be the set of upper endpoints of fragments together with 0. Since $\{1,2,\dots,2g\}$ is the disjoint union of the fragments of v, it follows that if $]i,j]$ is a fragment, then i is also the upper endpoint of a fragment. Thus it follows from (iii) that R is stable under $i \mapsto \bar{\imath}$. Let now i be an upper endpoint of a fragment. We want to show that $\nu(i) \in R$, and we may certainly assume that $\nu(i) \neq 0$, and we may also, by way of contradiction, assume that i is a minimal upper endpoint for which $\nu(i)$ is not an upper endpoint. If $\nu(i) \neq \nu(i-1)$, then $v^{-1}(i) = \nu(i)$ and hence $\nu(i)$ is an upper endpoint of a fragment. Hence we may pick $j < i$ such that $\nu(i) = \nu(i-1) = \cdots = \nu(j) \neq \nu(j-1)$. Then j cannot belong to the same fragment as i, and thus there must be an upper endpoint $j \leq k < i$. Then $\nu(k) = \nu(i)$ and by minimality of i we see that $\nu(k)$ is an upper endpoint, which is a contradiction.

We therefore have shown that R contains 0 and is stable under $i \mapsto \bar{\imath}$ and ν. Hence it contains the canonical domain. Let now $j \in C_\nu \setminus \{0\}$ and let i be the largest $j \in C_\nu$ such that $i < j$. We now want to show by induction on k that $v^{-k}(I)$, $I :=]i,j]$, remains an interval for all k and that also $v^{-k}(j)$ is one of its endpoints. Now, it follows from Lemma 2.5 that $C_\nu \setminus \{0\}$ is stable under v and hence $v^{-k}(j)$ will be the only element of C_ν in $v^{-k}(I)$. Under the induction assumption, $v^{-k}(I)$ is an interval with $v^{-k}(j)$ as one of its endpoints, and hence ν is constant on $v^{-k}(I)$ by Lemma 2.11. By Lemma 2.5, an interval v^{-1} maps $v^{-k}(I)$ to the interval with $v^{-k-1}(j)$ as one of its endpoints. This means that I is contained in a fragment and $R \cap I = \{j\}$. This means that there are no elements of R between consecutive elements of C_ν and hence $R \subseteq C_\nu$. $\square$

Corollary 2.13. *By associating to a final type* ν *its canonical type, its Young diagram, and the element* w_ν *we obtain a bijection between the following sets of cardinality* 2^g*: the set of final types, the set of canonical types, the set of final Young diagrams, and the set of final elements of* W_g*.*

2.4 Admissible Elements

The longest final element of W_g is the element

$$w_\emptyset := s_g s_{g-1} s_g s_{g-2} s_{g-1} s_g \dots s_g s_1 s_2 s_3 \dots s_g,$$

which as a permutation equals $[g+1, g+2, \dots, 2g]$. Elements of W_g that satisfy $w \leq w_\emptyset$ are called *admissible*. We now characterize these.

Lemma 2.14. *(i) An element $w \in W_g$ fulfills $w \leq w_\emptyset$ if and only if we have $w(i) \leq g + i$ for all $1 \leq i \leq g$.*

(ii) The condition that $w \leq w_\emptyset$ is equivalent to $r_w(i, g+i) = i$ for all $1 \leq i \leq g$.

Proof. The first part follows immediately from the description of the Bruhat–Chevalley order (Lemma 2.1) and the presentation of $w_\emptyset$.

For the second part one easily shows that $w(i) \leq w_\emptyset(i) = g + i$ for all $1 \leq i \leq g$ is equivalent to $r_w(i, g+i) = i$ for all $1 \leq i \leq g$, which gives the first equivalence. □

Remark 2.15. The number of elements $w \in W_g$ with $w \leq w_\emptyset$ and of given length has recently been determined by J. Sjöstrand [Sj07]. This implies in particular that the number of elements $w \in W_g$ with $w \leq w_\emptyset$ equals

$$\left(x\frac{d}{dx}\right)^g \left(\frac{1}{1-x}\right)_{\big|x=1/2},$$

a fact that we originally guessed from a computation for small g and a search in [S] leading to the sequence A000629.

We give an illustration of the various notions for the case $g = 2$.

Example 2.16. $g = 2$. The Weyl group W_2 consists of eight elements. We list (see Figure 3) the element, a reduced expression as a word (i.e., a decomposition $w = s_{i_1} \cdots s_{i_k}$ with $k = \ell(w)$), its length, the functions ν and μ, and for final elements we also give the partition defining the Young diagram.

The orbits of the complementary element will play an important role in our discussion of the canonical flag. Here we introduce some definitions pertaining to them.

w	s	ℓ	ν	μ	Y
[4321]	$s_1s_2s_1s_2$	4	$\{1,2\}$	$\{2,2\}$	
[4231]	$s_1s_2s_1$	3	$\{1,1\}$	$\{2,2\}$	
[3412]	$s_2s_1s_2$	3	$\{1,2\}$	$\{1,2\}$	$\emptyset$
[2413]	s_1s_2	2	$\{1,1\}$	$\{1,1\}$	$\{1\}$
[3142]	s_2s_1	2	$\{0,1\}$	$\{0,2\}$	
[2143]	s_1	1	$\{0,0\}$	$\{0,0\}$	
[1324]	s_2	1	$\{0,1\}$	$\{0,1\}$	$\{2\}$
[1234]	1	0	$\{0,0\}$	$\{0,0\}$	$\{1,2\}$

Fig. 3. The $g = 2$ case.

Definition 2.17. *Let $w \in W_g$ be a final element and let v be its complementary element. Assume that S is an orbit of the action of v on its fragments. Since v commutes with $i \mapsto \bar{\imath}$, we have either that S is invariant under $i \mapsto \bar{\imath}$, in which case we say that it is an* odd orbit*, or that $\overline{S}$ is another orbit, in which case we say $\{S, \overline{S}\}$ is an* even orbit pair*.*

2.5 Shuffles

Recall that a (p, q)-*shuffle* is a permutation σ of $\{1, 2, \ldots, p+q\}$ for which we have $\sigma(i) < \sigma(j)$ whenever $i < j \leq p$ or $p < i < j$. It is clear that for each subset I of $\{1, 2, \ldots, g\}$ there is a unique $(|I|, g-|I|)$-shuffle σ^I such that $I = \{\sigma^I(1), \sigma^I(2), \ldots, \sigma^I(|I|)\}$, and we will call it the *shuffle associated to I*. We will use the same notation for the corresponding element in W_g (i.e., fulfilling $\sigma^I(2g+1-i) = 2g+1-\sigma^I(i)$ for $1 \leq i \leq g$). By doing the shuffling from above instead of from below we get another shuffle σ_I given by $\sigma_I(i) = g+1-\sigma^I(g+1-i)$. We will use the same notation for its extension to W_g. Note that σ_I will shuffle the elements $\{g+1, g+2, \ldots, 2g\}$ in the same way that σ^I shuffles $\{1, 2, \ldots, g\}$, i.e., $\sigma^I(g+i) = g+\sigma_I(i)$, which is the relation with σ^I that motivates the definition. Note that if $I = \{i_1 < \cdots < i_r\}$ and if we assume that $i_r > r$ (if this does not hold then σ^I and σ_I are the identity elements) and we let k be the smallest index such that $i_k > k$, then $\sigma^I = s_{i_k-1}\sigma^{I'}$ and $\sigma_I = s_{g+1-(i_k-1)}\sigma_{I'}$, where $I' = \{i_1, \ldots, i_k-1, \ldots, i_r\}$. We call the element $s_{i-1}ws_{g+1-(i-1)}$ for $w \in W_g$ the *ith elementary shuffle* of w, and say that I' is the *elementary reduction* of I whose *reduction index* is i_k.

We define the *height* of a shuffle associated to a subset $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, g\}$ to be $\sum_s (i_s - s)$. Using $w' = s_iws_{g+1-t} \iff s_iw's_{g+1-t} = w$, we see that starting with an element w obtained by applying a shuffle to a final element we arrive at a final element after $\mathrm{ht}(w)$ elementary shuffles.

Definition 2.18. *Let Y be a final Young diagram of degree g. The* shuffles *of Y are the elements of W_g of the form $\sigma^I w_Y \sigma_I^{-1}$ for $I \subseteq \{1, 2, \ldots, g\}$.*

If $w \leq w_\emptyset$ we say that i with $1 \leq i \leq g$ is a *semi-simple index* for w if $w(i) = g+i$ (note that since $w \leq w_\emptyset$, we always have $w(i) \leq g+i$). The set of semi-simple indices will be called the *semi-simple index set* and its cardinality the *semi-simple rank*. We say that w is *semi-simply final* if the semi-simple index set has the form $[g-f+1, g]$ (where then f is the semi-simple rank). This is equivalent to w having the form $[\ldots, 2g-f+1, 2g-f+2, \ldots, 2g]$. If $w = w_Y$, Y a final Young diagram, then w is semi-simply final and the semi-simple rank is equal to g minus the length of the largest row of Y (defined to be zero if Y is empty).

Proposition 2.19. *Let $w \leq w_\emptyset$ be a semi-simply final element of semi-simple rank f and let $I \subseteq \{1, 2, \ldots, 2g\}$ be a subset with $\#I = f$. Put $\tilde{I} := \{g+1-i : i \in I\}$. Then $w' := \sigma^I w \sigma_I^{-1}$ is an element with $w' \leq w_\emptyset$ of semi-simple rank f and semi-simple index set $\tilde{I}$. Conversely, all $w' \leq w_\emptyset$ whose semi-simple index set is equal to I are of this form.*

Proof. Put $j := \sigma_I^{-1}(i)$. Note that $j > g - f \iff i \in \tilde{I}$. If $j > g - f$ we have $w(j) = g + j$ and hence $\sigma^I w_Y \sigma_I^{-1}(i) = \sigma^I(g + j) = g + \sigma_I(j) = g + i$. If on the other hand, $j \leq g - f$, then if $w_Y(j) \leq g$ holds there is nothing to prove, and if it does not hold we may write $w_Y(j) = g + k$, and since the semi-simple rank of Y is f, we have $k < j$. Then $\sigma^I w_Y \sigma_I^{-1}(i) = \sigma^I(g + j) = g + \sigma_I(k)$ and since $k < j \leq g - f$, we have $\sigma_I(k) < \sigma_I(j) = i$, which gives $\sigma^I w_Y \sigma_I^{-1}(i) < g + i$.

The converse is easy and left to the reader. □

Finally, we define the *a-number* of an element $w \in W_g$ by the rule

$$a(w) := r_w(g, g).$$

If w is final with associated Young diagram Y then its a-number, also denoted by a_Y, is the largest integer a with $0 \leq a \leq g$ such that Y contains the diagram that corresponds to the set $\xi = \{a, a-1, a-2, \ldots, 1\}$.

3 The Flag Space

3.1 The Flag space of the Hodge bundle

In this section we introduce the flag space of a principally polarized abelian scheme over a base scheme of characteristic p. We use the Frobenius morphism to produce from a chosen flag on the de Rham cohomology a second flag, whose position with respect to the first flag will be the object of study.

We let S be a scheme (or Deligne–Mumford stack) in characteristic p and let $\mathcal{X} \to S$ be an abelian scheme over S with principal polarization (everything would go through using a polarization of degree prime to p but we shall stick to the principally polarized case). We consider the de Rham cohomology sheaf $\mathcal{H}^1_{dR}(\mathcal{X}/S)$. It is defined as the hyper-direct image $\mathcal{R}^1\pi_*(O_{\mathcal{X}} \to \Omega^1_{\mathcal{X}/S})$ and is a locally free sheaf of rank $2g$ on S. The polarization (locally in the étale topology given by a relatively ample line bundle on $\mathcal{X}/S$) provides us with a symmetric homomorphism $\rho \colon \mathcal{X} \to \hat{\mathcal{X}}$, and the Poincaré bundle defines a perfect pairing between $\mathcal{H}^1_{dR}(\mathcal{X}/S)$ and $\mathcal{H}^1_{dR}(\hat{\mathcal{X}}/S)$, and thus $\mathcal{H}_{dR}(\mathcal{X}/S)$ comes equipped with a nondegenerate alternating form (cf. [Oo95])

$$\langle , \rangle \colon \mathcal{H}^1_{dR}(\mathcal{X}/S) \times \mathcal{H}^1_{dR}(\mathcal{X}/S) \to O_S.$$

Moreover, we have an exact sequence of locally free sheaves on S:

$$0 \to \pi_*(\Omega^1_{\mathcal{X}/S}) \to \mathcal{H}^1_{dR}(\mathcal{X}/S) \to R^1\pi_* O_{\mathcal{X}} \to 0.$$

We shall write $\mathbb{H}$ for the sheaf $\mathcal{H}^1_{dR}(\mathcal{X}/S)$ and $\mathbb{E}$ for the *Hodge bundle* $\pi_*(\Omega^1_{\mathcal{X}/S})$. We thus have an exact sequence

$$0 \to \mathbb{E} \to \mathbb{H} \to \mathbb{E}^\vee \to 0$$

of locally free sheaves on S. The relative Frobenius $F\colon \mathcal{X} \to \mathcal{X}^{(p)}$ and the Verschiebung $V\colon \mathcal{X}^{(p)} \to \mathcal{X}$ satisfy $F \cdot V = p \cdot \mathrm{id}_{\mathcal{X}^{(p)}}$ and $V \cdot F = p \cdot \mathrm{id}_{\mathcal{X}}$ and they induce maps, also denoted by F, respectively V, in cohomology:

$$F\colon \mathbb{H}^{(p)} \to \mathbb{H} \qquad \text{and} \qquad V\colon \mathbb{H} \to \mathbb{H}^{(p)}.$$

Of course, we have $FV = 0$ and $VF = 0$ and F and V are adjoints (with respect to the alternating form). This implies that $\mathrm{Im}(F) = \ker(V)$ and $\mathrm{Im}(V) = \ker(F)$ are maximally isotropic subbundles of $\mathbb{H}$ and $\mathbb{H}^{(p)}$ respectively. Moreover, since $dF = 0$ on $\mathrm{Lie}(\mathcal{X})$, it follows that $F = 0$ on $\mathbb{E}^{(p)}$ and thus $\mathrm{Im}(V) = \ker(F) = \mathbb{E}^{(p)}$. Verschiebung thus provides us with a bundle map (again denoted by V) $V\colon \mathbb{H} \to \mathbb{E}^{(p)}$.

Consider the space $\mathcal{F} = \mathrm{Flag}(\mathbb{H})$ of symplectic flags on the bundle $\mathbb{H}$ consisting of flags of subbundles $\{\mathbb{E}_i\}_{i=1}^{2g}$ satisfying $\mathrm{rk}(\mathbb{E}_i) = i$, $\mathbb{E}_{g+i} = \mathbb{E}_{g-i}^{\perp}$, and $\mathbb{E}_g = \mathbb{E}$. This space is a scheme over S and it is fibered by the spaces $\mathcal{F}^{(i)}$ of partial flags

$$\mathbb{E}_i \subsetneq \mathbb{E}_{i+1} \subsetneq \cdots \subsetneq \mathbb{E}_g.$$

So $\mathcal{F} = \mathcal{F}^{(1)} = \mathrm{Flag}(\mathbb{H})$ and $\mathcal{F}^{(g)} = S$ and there are natural maps

$$\pi_{i,i+1}\colon \mathcal{F}^{(i)} \to \mathcal{F}^{(i+1)},$$

the fibers of which are Grassmann varieties of dimension i. So the relative dimension of $\mathcal{F}$ is $g(g-1)/2$. The space $\mathcal{F}^{(i)}$ is equipped with a universal partial flag. On $\mathcal{F}$ the Chern classes of the bundle $\mathbb{E}$ decompose into their roots:

$$\lambda_i = \sigma_i(\ell_1, \ldots, \ell_g) \qquad \text{with} \qquad \ell_i = c_1(\mathbb{E}_i/\mathbb{E}_{i-1}),$$

where σ_i is the ith elementary symmetric function.

On $\mathcal{F}^{(i)}$ we have the Chern classes $\ell_{i+1}, \ldots, \ell_g$ and

$$\lambda_j(i) := c_j(\mathbb{E}_i), \qquad j = 0, 1, \ldots, i.$$

Its Chow ring is generated over that of $\mathcal{A}_g$ by the monomials $\ell_1^{m_1} \cdots \ell_g^{m_g}$ with $0 \leq m_j \leq j-1$. For later use we record the following Gysin formula.

Formula 3.1. *We have $(\pi_{i,i+1})_*\ell_{i+1}^k = s_{k-i}(i+1)$, where $s_j(i+1)$ denotes the jth Segre class of $\mathbb{E}_{i+1}$ (jth complete symmetric function in the Chern roots $\ell_1, \ldots, \ell_{i+1}$).*

Given an arbitrary flag of subbundles

$$0 = \mathbb{E}_0 \subsetneq \mathbb{E}_1 \subsetneq \cdots \subsetneq \mathbb{E}_g = \mathbb{E}$$

with $\mathrm{rank}(\mathbb{E}_i) = i$ we can extend this uniquely to a symplectic filtration on $\mathbb{H}$ by putting

$$\mathbb{E}_{g+i} = (\mathbb{E}_{g-i})^{\perp}.$$

By base change we can transport this filtration to $\mathbb{H}^{(p)}$.

We introduce a second filtration by starting with the isotropic subbundle

$$\mathbb{D}_g := \ker(V) = V^{-1}(0) \subset \mathbb{H}$$

and continuing with

$$\mathbb{D}_{g+i} = V^{-1}(\mathbb{E}_i^{(p)}).$$

We extend it to a symplectic filtration by setting $\mathbb{D}_{g-i} = (\mathbb{D}_{g+i})^{\perp}$. We thus have two filtrations $\mathbb{E}_\bullet$ and $\mathbb{D}_\bullet$ on the pullback of $\mathbb{H}$ to $\mathcal{F}$.

We shall use the following notation:

$$\mathcal{L}_i = \mathbb{E}_i/\mathbb{E}_{i-1} \quad \text{and} \quad \mathcal{M}_i = \mathbb{D}_i/\mathbb{D}_{i-1} \qquad \text{for } 1 \le i \le 2g.$$

For ease of reference we formulate a lemma that follows immediately from definitions.

Lemma 3.2. *We have* $\mathcal{M}_{g+i} \cong \mathcal{L}_i^p$, $\mathcal{L}_{2g+1-i} \cong \mathcal{L}_i^\vee$, *and* $\mathcal{M}_{2g+1-i} \cong \mathcal{M}_i^\vee$.

More generally, for a family $X \to S$ of principally polarized abelian varieties we shall say that a *Hodge flag* for the family is a complete symplectic flag $\{\mathbb{E}_i\}$ of $\mathbb{H}$ for which $\mathbb{E}_g$ is equal to the Hodge bundle. By construction this is the same thing as a section of $\mathcal{F}_g \to S$. We shall also call the associated flag $\{\mathbb{D}_i\}$ the *conjugate flag* of the Hodge flag.

3.2 The canonical flag of an abelian variety

In this section we shall confirm that the *canonical filtration* of $X[p]$, (kernel of multiplication by p) by subgroup schemes of a principally polarized abelian variety X as defined by Ekedahl and Oort [Oo01] has an analogue for de Rham cohomology. Just as in [Oo01] we do this in a family $\mathcal{X} \to S$. It is the coarsest flag that is isotropic (i.e., if $\mathbb{D}$ is a member of the flag then so is $\mathbb{D}^\perp$) and stable under F (i.e., if $\mathbb{D}$ is a member of the flag then so is $F(\mathbb{D}^{(p)})$). The existence of such a minimal flag is proven by adding elements $\mathcal{F}^\perp$ and $F(\mathbb{D}^{(p)})$ for $\mathbb{D}$ already in the flag in a controlled fashion. We start by adding 0 to the flag. We then insist on three rules:

1. If we added $\mathbb{D} \subseteq \mathbb{D}_g$, then we immediately add $\mathbb{D}^\perp$ (unless it is already in the flag constructed so far).
2. If we added $\mathbb{D}_g \subseteq \mathbb{D}$, then we immediately add $F(\mathbb{D}^{(p)})$ (unless it is already in the flag constructed so far).
3. If neither rule (1) nor rule (2) applies, then we add $F(\mathbb{D}^{(p)})$ for the largest element $\mathbb{D}$ of the flag for which $F(\mathbb{D}^{(p)})$ is not already in the flag.

We should not, however, do this construction on S; we want to ensure that we get a filtration by vector bundles: at each stage when we want to add the image $F(\mathbb{D}^{(p)})$, we have maps $F\colon \mathbb{D}^{(p)} \to \mathbb{H}$ of vector bundles, and we then have a unique minimal decomposition of the base as a disjoint union of locally closed subschemes such that on each subscheme this map has constant rank; these are subschemes because they are given by degeneracy conditions. At the same time as we add $F(\mathbb{D}^{(p)})$ to the flag we replace the base by this disjoint union. Over this disjoint union, $F(\mathbb{D}^{(p)})$ then becomes a subbundle of $\mathbb{H}$ and whether it is equal to one of the previously defined subbundles is a locally constant condition. A simple induction then shows that we get a flag, i.e., for any two elements constructed, one is included in the other, on a disjoint union of subschemes of S. Since each element added is either the image under F of an element previously constructed or the orthogonal of such an element, it is clear that this flag is the coarsening of any isotropic flag stable under F, and it is equally clear that the decomposition of S is the coarsest possible decomposition. We shall call the (partial) flag obtained in this way the *canonical flag* of $\mathcal{X}/S$ and the decomposition of S the *canonical decomposition* of the base.

To each stratum S' of the canonical decomposition of S we associate a canonical type as follows: let $T \subseteq \{1, 2, \ldots, 2g\}$ be the set of ranks of the elements of the canonical flag and let $f\colon T \to T \cap \{1, \ldots, g\}$ be the function that to t associates $\mathrm{rk}(F(\mathbb{D}^{(p)}))$, where $\mathbb{D}$ is the element of the canonical flag of rank t. We now claim that T and f fulfill the conditions of Lemma 2.11. Clearly T contains 0, and by construction it is invariant under $i \mapsto 2g - i$. Again by construction f is increasing and has $f(0) = 0$ and $f(2g) = g$. Furthermore, if $i, j \in T$ with $i < j$ then F induces a surjective map $(\mathbb{D}/\mathbb{D}')^{(p)} \to F(\mathbb{D})/F(\mathbb{D}')$, where $\mathbb{D}$ respectively $\mathbb{D}'$ are the elements of the canonical flag for which the rank is j respectively i and hence $f(j) - f(i) = \mathrm{rk}(F(\mathbb{D})/F(\mathbb{D}')) \leq \mathrm{rk}(\mathbb{D}/\mathbb{D}') = j - i$. Finally, assume that $f(j) - f(i) = j - i$ and let $\mathbb{D}$ and $\mathbb{D}'$ be as before. Putting $\mathbb{D}_1 := F(\mathbb{D}^{(p)})$ and $\mathbb{D}_1' := F(\mathbb{D}'^{(p)})$ these are also elements of the canonical filtration, and by assumption F induces an isomorphism $F\colon (\mathbb{D}/\mathbb{D}')^{(p)} \to \mathbb{D}_1/\mathbb{D}_1'$. The fact that it is injective means that $\mathbb{D} \cap \ker F = \mathbb{D}' \cap \ker F$, which by taking annihilators and using that $\ker F$ is its own annihilator, gives $\mathbb{D}^\perp + \ker F = \mathbb{D}'^\perp + \ker F$. In turn, this implies $F(\mathbb{D}^\perp) = (\mathbb{D}'^\perp)$ and hence that $f(2g-i) = f(2g-j)$. Now, if f is not surjective then by Lemma 2.11 there is a proper subset of T fulfilling the conditions of Lemma 2.11. This is not possible since T by construction is a minimal subset with these conditions. Hence T and f fulfill the conditions of Lemma 2.11, and hence by them we get that (f, T) is a canonical type. Let ν be its associated final type. If $0 = \mathbb{D}_0 \subset \mathbb{D}_1 \subset \cdots \subset \mathbb{D}_{2g} = \mathbb{H}$ is the canonical flag with $\mathrm{rk}\,\mathbb{D}_i = i$, we have also proved that F induces an isomorphism $(\mathbb{D}_j/\mathbb{D}_i)^{(p)} \to \mathbb{D}_{\nu(j)}/\mathbb{D}_{\nu(i)}$, which can be rephrased as an isomorphism

$$F\colon \mathbb{D}_{v(I)}^{(p)} \xrightarrow{\sim} \mathbb{D}_I,$$

where we have used the notation $\mathbb{D}_J := \mathbb{D}_i/\mathbb{D}_j$ for an interval $J =]j, i]$ and $v \in W_g$ is the complementary element of (the final element of) ν. We shall say that ν (or more properly f) is the *canonical type* of the principally polarized abelian variety $\mathcal{X}_{S'} \to S'$. (We could consider the canonical type as a locally constant function on the canonical decomposition to the set of canonical (final) types.)

Remark 3.3. Note that the canonical flag is a flag containing $\mathbb{D}_g$ and is not defined in terms of $\mathbb{E}_g$. This will later mean that the canonical flag will be a coarsening of a conjugate flag that is derived from a Hodge flag. On the one hand, this is to be expected. Since the canonical flag is just that, it will be constructed in a canonical fashion from the family of principally polarized abelian varieties. Hence it is to be expected, and it is clearly true, that the canonical flag is horizontal with respect to the Gauss–Manin connection. On the other hand, we do not want to consider just conjugate flags (or make constructions starting only with conjugate flags). The reason is essentially the same; since $\mathbb{D}_g$ (or more generally the elements of the canonical flag) is horizontal, it will not reflect first-order deformations, whereas $\mathbb{E}_g$ isn't and does. This will turn out to be of crucial importance to us and is the reason why the Hodge flags will be the primary objects, while the conjugate flags are secondary. On the other hand, when working pointwise, over an algebraically closed field, say, we may recover the Hodge flag from the conjugate flag and then it is usually more convenient to work with the conjugate flag.

Example 3.4. Let X be an abelian variety with p-rank f and $a(X) = 1$ (equivalently, on $\mathbb{E}_g$ the operator V has rank $g-1$ and semi-simple rank $g-f$). Then the canonical type is given by the numbers $\{\mathrm{rk}(\mathbb{D}_i)\}$, i.e.,

$$\{0, f, f+1, \ldots, 2g-f-1, 2g-f, 2g\},$$

and ν is given by $\nu(f) = f$, $\nu(f+1) = f$, $\nu(f+2) = f+1$, $\ldots$, $\nu(g) = g-1, \ldots$, $\nu(2g-f-1) = g-1$, $\nu(2g-f) = g$, and $\nu(2g) = g$. The corresponding element $w \in W_g$ is $[f+1, g+1, \ldots, 2g-f-1, 2g-f+1, \ldots, 2g]$.

4 Strata on the Flag Space

4.1 The Stratification

The respective positions of two symplectic flags are encoded by a combinatorial datum, an element of a Weyl group. We shall now define strata on the flag space $\mathcal{F}$ over the base S of a principally polarized abelian scheme $X \to S$ that mark the respective positions of the two filtrations $\mathbb{E}_\bullet$ and $\mathbb{D}_\bullet$ that we have on the de Rham bundle over $\mathcal{F}$.

Intuitively, the stratum $\overline{\mathcal{U}}_w$ is defined as the locus of points x such that at x we have

$$\dim(\mathbb{E}_i \cap \mathbb{D}_j) \geq r_w(i,j) = \#\{a \leq i \colon w(a) \leq j\} \quad \text{for all} \quad 1 \leq i, j \leq 2g.$$

A more precise definition would be as degeneracy loci for some appropriate bundle maps. While this definition would work fine in our situation, where we are dealing with flag spaces for the symplectic group, it would not quite work when the symplectic group is replaced by the orthogonal group on an even-dimensional space (cf. [FP98]). With a view toward future extensions of the ideas of this paper to other situations we therefore adopt the definition that would work in general. Hence assume that we have a semi-simple group G, a Borel group B of it, a G/B-bundle $T \to Y$ (with G as structure group) over some scheme Y, and two sections $s, t\colon Y \to T$ of it. Then for any element w of the Weyl group of G we define a (locally) closed subscheme $\mathcal{U}_w$ respectively $\overline{\mathcal{U}}_w$ of Y in the following way. We choose locally (possibly in the étale topology) a trivialization of T for which t is a constant section. Then s corresponds to a map $Y \to G/B$ and we let $\mathcal{U}_w$ (respectively $\overline{\mathcal{U}}_w$) be the inverse image of the B-orbit BwB (respectively of its closure in G/B). Another trivialization will differ by a map $Y \to B$, and since BwB and its closure are B-invariant, these definitions are independent of the chosen trivializations and hence give global subschemes on Y. If s and t have the property that $Y = \mathcal{U}_w$, then we shall say that s and t are in *relative position* w and if $Y = \overline{\mathcal{U}}_w$, we shall say that s and t are in *relative position* $\leq w$.

Remark 4.1. The notation is somewhat misleading, since it suggests that $\overline{\mathcal{U}}_w$ is the closure of $\mathcal{U}_w$, which may not be the case in general. In the situation that we shall meet it will, however, be the case (cf. Corollary 8.4).

The situation to which we will apply this construction is that in which the base scheme is the space $\mathcal{F}$ of symplectic flags $\mathbb{E}_\bullet$ as above, s is the tautological section of the flag space of $\mathbb{H}$ over $\mathcal{F}$, and t is the section given by the conjugate flag $\mathbb{D}_\bullet$. From now on we shall, unless otherwise mentioned, let $\mathcal{U}_w$ and $\overline{\mathcal{U}}_w$ denote the subschemes of $\mathcal{F}$ coming from the given s and t and $w \in W_g$. In this case it is actually often more convenient to use the language of flags rather than sections of G/B-bundles and we shall do so without further mention. We shall also say that a Hodge flag $\mathbb{E}_\bullet$ is of *stamp* w respectively *stamp* less than or equal to w if $\mathbb{E}_\bullet$ and its conjugate flag $\mathbb{D}_\bullet$ are in relative position w respectively $\leq w$.

Lemma 4.2. *Over $\mathcal{U}_w$ we have an isomorphism $\mathcal{L}_i \cong \mathcal{M}_{w(i)}$ for all $1 \leq i \leq 2g$.*

Proof. By the definition of the strata we have that the image of $\mathbb{E}_i \cap \mathbb{D}_{w(i)}$ has rank one greater than the ranks of $\mathbb{E}_{i-1} \cap \mathbb{D}_{w(i)}$, $\mathbb{E}_i \cap \mathbb{D}_{w(i)-1}$, and $\mathbb{E}_{i-1} \cap \mathbb{D}_{w(i)-1}$. So the maps $\mathbb{E}_i/\mathbb{E}_{i-1} \leftarrow (\mathbb{E}_i \cap \mathbb{D}_{w(i)})/(\mathbb{E}_{i-1} \cap \mathbb{D}_{w(i)-1}) \to \mathbb{D}_{w(i)}/\mathbb{D}_{w(i)-1}$ give the isomorphism. □

When the base of the principally polarized abelian scheme is $\mathcal{A}_g$, we shall use the notation $\mathcal{F}_g$ for the space of Hodge flags. Note that a Hodge flag with respect to $X \to S$ is the same thing as a lifting over $\mathcal{F}_g \to \mathcal{A}_g$ of the classifying map $S \to \mathcal{A}_g$. The conjugate flag as well as the strata $\mathcal{U}_g$ and $\overline{\mathcal{U}}_g$ on S are then the pullbacks of the conjugate flag, respectively the strata on $\mathcal{F}_g$.

4.2 Some Important Strata

We now give an interpretation for some of the most important strata. To begin with, if one thinks instead in terms of filtrations of $X[p]$ by subgroup schemes it becomes clear that the condition $F(\mathbb{D}_i^{(p)}) \subseteq \mathbb{D}_i$ should be of interest. It can almost be characterized in terms of the strata $\overline{\mathcal{U}}_w$.

Proposition 4.3. *Let $X \to S$ be a family of principally polarized abelian varieties and $\mathbb{E}_\bullet$ a Hodge flag such that the flag is of stamp $\leq w$ and w is the smallest element with that property.*

(i) For $j \leq g$ and for all $i \in [1, \ldots, g]$ we have $r_w(i, g+j) = i$ if and only if $V(\mathbb{E}_i) \subseteq \mathbb{E}_j^{(p)}$.

(ii) For $j \leq g$ we have that $r_w(g+j, i) = i$ implies that $F(\mathbb{D}_i^{(p)}) \subseteq \mathbb{D}_j$ and the converse is true if S is reduced.

(iii) We have that $V(\mathbb{E}_i) \subseteq \mathbb{E}_i^{(p)}$ for all i precisely when $w \leq w_\emptyset$. If S is reduced, $F(\mathbb{D}_i^{(p)}) \subseteq \mathbb{D}_i$ for all i precisely when $w \leq w_\emptyset$.

Proof. We have that $V(\mathbb{E}_i) \subseteq \mathbb{E}_j^{(p)}$ if and only if $\mathbb{E}_i \subseteq V^{-1}(\mathbb{E}_j^{(p)}) = \mathbb{D}_{g+j}$. On the other hand, by definition $\operatorname{rk} \mathbb{E}_i \cap \mathbb{D}_{g+j} \leq r_w(i, g+j)$ with equality for at least one point of S. Since $\operatorname{rk} \mathbb{E}_i \cap \mathbb{D}_{g+j} = i \iff \mathbb{E}_i \subseteq \mathbb{D}_{g+j}$, we get the first part.

For the second part we start by claiming that $\mathbb{E}_i^{(p)} \subseteq \mathbb{D}_j$ is implied by $F(\mathbb{D}_i^{(p)}) \subseteq \mathbb{D}_j$. Indeed, $F(\mathbb{D}_i^{(p)}) \subseteq \mathbb{D}_j$ is equivalent to $F(\mathbb{D}_i^{(p)})$ being orthogonal to $\mathbb{D}_{2g-j}$, i.e., to the condition that for $u \in \mathbb{D}_i^{(p)})$ and $v \in \mathbb{D}_{2g-j}$ we have $\langle Fu, v\rangle = 0$. This implies that $0 = \langle Fu, v\rangle = \langle u, Vv\rangle^p$ and hence $\langle u, Vv\rangle = 0$, since S is reduced, which means that $\mathbb{D}_i^{(p)} \subseteq (V(\mathbb{D}_{2g-j}))^\perp = (\mathbb{E}_{g-j}^{(p)})^\perp = \mathbb{E}_{g+j}^{(p)}$. Since S is reduced, this implies that $\mathbb{D}_i \subseteq \mathbb{E}_{g+j}$, and this in turn is equivalent to $r_w(g+j, i) = i$. The argument can be reversed (and then it does not require S to be reduced).

Finally, we have from the first part that $V(\mathbb{E}_i) \subseteq \mathbb{E}_i^{(p)}$ for all $i \leq g$ precisely when $r_w(i, g+i) = i$ for all $i \leq g$, but by induction on i that is easily seen to be equivalent to $w(i) \leq g+i$ for all $i \leq g$, which by definition means that $w \leq w_\emptyset$. Since $\mathbb{E}_g^{(p)} = V(\mathbb{H})$, the condition $V(\mathbb{E}_i) \subseteq \mathbb{E}_i^{(p)}$ for $i > g$ is trivially fulfilled.

The proof of the second equivalence is analogous in that using (ii), the condition that $F(\mathbb{D}_i^{(p)}) \subseteq \mathbb{D}_i$ is equivalent to $r_w(g+i, i) = i$. In general, $r_u(i,j) = r_{u^{-1}}(j,i)$, so that this condition is equivalent to $r_{w^{-1}}(i, g+i) = i$, and hence by the same argument as before, this condition for all i is equivalent to $w^{-1} \leq w_\emptyset$. Chevalley's characterization of the Bruhat–Chevalley order makes it clear that $u \leq v \iff u^{-1} \leq v^{-1}$, and hence we get $w^{-1} \leq w_\emptyset \iff w \leq w_\emptyset^{-1}$. However, $w_\emptyset$ is an involution. □

Remark 4.4. (i) As we shall see (cf. Corollary 8.4) the strata $\overline{\mathcal{U}}_w$ in the universal case of $\mathcal{F}_g$ are reduced.

(ii) Flags of stamp $w \leq w_\emptyset$ are called *admissible*.

We can also show that the relations between final and canonical types are reflected in relations for flags. We say that a Hodge flag is a *final flag* if it is of stamp w for a final element w. Also if $I =]i,j] \subseteq \{1,2,\dots,2g\}$ is an interval and $\mathbb{F}_\bullet$ is a complete flag of a vector bundle of rank $2g$ then we define $\mathbb{F}_I$ to be $\mathbb{F}_j/\mathbb{F}_i$.

Proposition 4.5. *Let $X \to S$ be a principally polarized abelian scheme over S and let $\mathbb{E}_\bullet$ be a final flag for it of stamp w.*

(i) The conjugate flag $\mathbb{D}_\bullet$ is a refinement of the canonical flag. In particular, w is determined by $X \to S$. More directly, we have that the final type ν associated to w is given by

$$\mathrm{rk}(\mathbb{E}_g \cap \mathbb{D}_i) = i - \nu(i)$$

for all i. In particular, the canonical decomposition of S with respect to $X \to S$ consists of a single stratum, and its canonical type is the canonical type associated to w.

(ii) Conversely, assume that S is reduced and that the canonical decomposition of S consists of a single stratum, and let ν be the final type associated to the canonical type of the canonical flag. Then any Hodge flag $\mathbb{E}_\bullet$ whose conjugate flag $\mathbb{D}_\bullet$ is a refinement of the canonical flag and for which we have $F(\mathbb{D}_i^{(p)}) \subseteq \mathbb{D}_{\nu(i)}$ for all i, is a final flag.

(iii) If I is a canonical fragment for v, the complementary element to w, then F induces a bijection $(\mathbb{D}_{v(I)})^{(p)} \xrightarrow{\sim} \mathbb{D}_I$.

Proof. We start by showing that $F(\mathbb{D}_i) = \mathbb{D}_{\nu(i)}$ for all i. Indeed, this is equivalent to $F(\mathbb{D}_i) \subseteq \mathbb{D}_{\nu(i)}$ and $\mathrm{rk}\big((\ker(F) = \mathbb{E}_g^{(p)}) \cap \mathbb{D}_i^{(p)}\big) = i - \nu(i)$ since the second condition says that $F(\mathbb{D}_i)$ has rank $\nu(i)$. Now, the condition $F(\mathbb{D}_i) \subseteq \mathbb{D}_{\nu(i)}$ is by Proposition 4.3 implied by $r_w(g+\nu(i), i) = \nu(i)$, which is true for a final element by Corollary 2.10. On the other hand, the condition $\mathrm{rk}(\mathbb{E}_g^{(p)} \cap \mathbb{D}_i^{(p)}) = i - \nu(i)$ is implied by $r_w(g,i) = \mathrm{rk}(\mathbb{E}_g \cap \mathbb{D}_i) = i - \nu(i)$, which is true by the very definition of ν. Now, the fact that $F(\mathbb{D}_i) = \mathbb{D}_j$ for all i and some j (depending on i) implies by induction on the steps of the construction of the canonical flag that $\mathbb{D}$ is a refinement of the canonical flag. The rest of the first part then follows from what we have proved.

As for (ii), assume first that $\mathbb{E}_\bullet$ has a fixed stamp w' and let ν' be its final type. Since $\mathbb{D}_\bullet$ is an extension of the Hodge flag, we get that when i is in the canonical domain of ν then $i - \nu'(i) = \mathrm{rk}(\mathbb{E}_g \cap \mathbb{D}_i) = i - \nu(i)$, so that ν and ν' coincide on the canonical domain of ν and hence they coincide by Lemma 2.11. The assumption that $F(\mathbb{D}_i^{(p)}) \subseteq \mathbb{D}_{\nu(i)} = \mathbb{D}_{\nu'(i)}$ for all i is by Proposition 4.3 equivalent to $r_w(g+\nu'(i), i) = i$ for all i, and hence Corollary 2.10 gives that w is final of type $\nu' = \nu$.

Finally, assume that $I =]i,j]$. The induced map $\mathbb{D}_j^{(p)}/\mathbb{D}_i^{(p)} \to \mathbb{D}_{\nu(j)}/\mathbb{D}_{\nu(i)}$ is always surjective, but it follows from Lemma 2.11 that either the right-hand side of this map has rank 0 or it has the same rank as the right-hand side.

If they have the same rank it induces an isomorphism $\mathbb{D}_I^{(p)} \xrightarrow{\sim} \mathbb{D}_{v^{-1}(I)}$. If the right-hand side has rank zero, then again from Lemma 2.11 the two sides of $\mathbb{D}_{\bar{\jmath}}^{(p)}/\mathbb{D}_{\bar{\imath}}^{(p)} \to \mathbb{D}_{\nu(\bar{\jmath})}/\mathbb{D}_{\nu(\bar{\imath})}$ have the same rank, and hence this map is an isomorphism and again gives an isomorphism $\mathbb{D}_I^{(p)} \xrightarrow{\sim} \mathbb{D}_{v^{-1}(I)}$. Since I is an arbitrary fragment, we conclude the proof of (ii). $\square$

The number of final extensions of a canonical flag will now be expressed in the familiar terms of the number of flags (respectively self-dual flags) in a vector space over a finite field (respectively a vector space with a unitary form). Hence we let $\gamma_n^e(m)$ be the number of complete $\mathbb{F}_{p^m}$-flags in $\mathbb{F}_{p^m}^n$ and let $\gamma_n^o(m)$ be the number of complete $\mathbb{F}_{p^{2m}}$-flags self-dual under the unitary form $\langle (u_1,\ldots,u_n),(v_1,\ldots,v_n)\rangle := u_1 v_1^{p^m} + \cdots + u_n v_n^{p^m}$. Recall now the definition of even and odd orbits of fragments given in Definition 2.17.

Lemma 4.6. *Let X be a principally polarized abelian variety over an algebraically closed field and $w \in W_g$ the element whose canonical type is the canonical type of X. Put*

$$\gamma(w) = \gamma_g(w) := \prod_{S=\overline{S}} \gamma^o_{\#I}(\#S/2) \prod_{\{S,\overline{S}\}} \gamma^e_{\#I}(\#S),$$

where the first product runs over the odd orbits and the second over the even orbit pairs and in both cases I indicates a member of S.

The number of final flags for X is then equal to $\gamma(w)$.

Proof. Since we are over a perfect field, any symplectic flag extending $\mathbb{D}_g$ is the conjugate flag of a unique Hodge flag. Hence we get from Proposition 4.5 that a final flag is the same thing as a flag $\mathbb{D}_\bullet$ extending the canonical flag and for which we have $F(\mathbb{D}_i^{(p)}) \subseteq \mathbb{D}_{\nu(i)}$ for all i. The condition that $\dim(\mathbb{E}_g \cap \mathbb{D}_i) = i - v(i)$ then gives that we actually have $F(\mathbb{D}_i^{(p)}) = \mathbb{D}_{\nu(i)}$. However, since $\mathbb{D}_\bullet$ refines the canonical flag, it is determined by the induced flags of the $\mathbb{D}_I$ for fragments I of v, and the stability condition $F(\mathbb{D}_i^{(p)}) = \mathbb{D}_{\nu(i)}$ transfers into a stability condition under the isomorphisms $F\colon \mathbb{D}_{v(I)}^{(p)} \xrightarrow{\sim} \mathbb{D}_I$ of Section 3.2.

Hence the problem splits up into separate problems for each orbit under v and $i \mapsto \bar{\imath}$ (the map $i \mapsto \bar{\imath}$ transfers into the isomorphism $\mathbb{D}_{\overline{I}} \cong \mathbb{D}_I^\vee$ induced by the symplectic form). Given any fragment $I \in S$, S a v-orbit of fragments of v, the flag of $\mathbb{D}_{v^k(I)}$ for any k is then determined by the flag corresponding to I by the condition that F^k-stability takes the kth Frobenius pullback of the flag on $\mathbb{D}_{v^k(I)}$ to the one of $\mathbb{D}_I$. Furthermore, the flag on $\mathbb{D}_I$ then has to satisfy the consistency condition of being stable under $F_S := F^{\#S}$.

For an even orbit pair $\{S, \overline{S}\}$ the self-duality requirement for the flag means that the flags for the elements of $\overline{S}$ are determined by those for the elements of S. Moreover, given $I \in S$ there is no other constraint on the flag on $\mathbb{D}_I$ than that it be stable under F_S. Hence we have the situation of a vector bundle $\mathcal{D}$

over our base field $\mathbf{k}$ and an isomorphism $F_S\colon \mathcal{D}^{(p^m)} \xrightarrow{\sim} \mathcal{D}$, where $m = \#S$, and we want to count the number of flags stable under F_S. Now, since $\mathbf{k}$ is algebraically closed, $\mathcal{D}_p := \{v \in \mathcal{D} : F_S(v) = v\}$ is an $\mathbb{F}_{p^m}$-vector space for which the inclusion map induces an isomorphism $\mathbf{k} \bigotimes_{\mathbb{F}_{p^m}} \mathcal{D}_p \xrightarrow{\sim} \mathcal{D}$. It then follows that F_S-stable flags correspond to $\mathbb{F}_{p^m}$-flags of $\mathcal{D}_p$.

If instead S is an odd orbit, we will together with I also have $\overline{I}$ in S. If we denote the cardinality of S by $2m$, the flag on $\mathbb{D}_I$ must be mapped to the dual flag on $\mathbb{D}_{\overline{I}}$ by $F_S := F^m$. The situation is similar to the even orbit pair situation but with a "unitary twist" as in Proposition 7.2, and we get instead a correspondence with self-dual flags. We leave the details to the reader. $\square$

Example 4.7. For the canonical type associated to the final type of an element w of Example 3.4 we have

$$\gamma_g(w) = (p+1)(p^2+p+1)\cdots(p^{f-1}+p^{f-2}+\cdots+1).$$

Definition 4.8. *For $1 \leq f < g$ we let $w = u_f$ be the final element*

$$u_f = s_g s_{g-1} s_g \cdots s_{g-f+1} \cdots s_g s_{g-f-1} \cdots s_g \cdots s_1 \cdots s_g,$$

i.e., if we introduce $\tau_j = s_j s_{j+1} \cdots s_g$ then we have $u_f = \tau_g \tau_{g-1} \cdots \hat{\tau}_f \cdots \tau_1$. It corresponds to the Young diagram consisting of one row with $g-f$ blocks and equals the element w given in Example 3.4.

Recall the notion of the a-number $a(X)$ for a g-dimensional abelian variety X of an (algebraically closed) field k of characteristic p. It equals the dimension over k of the vector space $\mathrm{Hom}(\alpha_p, X)$ of maps of the group scheme α_p to X. Equivalently, $a(X)$ equals the dimension of the kernel of V on $H^0(X, \Omega^1_X)$. In our terms $H^0(X, \Omega^1_X) = \mathbb{E}_g$ and $\ker V = \mathbb{D}_g$, so that $a(X) = \dim \mathbb{E}_g \cap \mathbb{D}_g$. The p-rank or *semi-simple rank* f, on the other hand, can be characterized by the condition that $\dim_{\mathbf{k}} \cap_{i \leq g} (V^i \mathbb{H})^{(p^{g-i})}$ is equal to f.

Lemma 4.9. *(i) Let Y be a final Young diagram and $w \in W_g$ its final element and assume that $x = (X, \mathbb{E}_\bullet, \mathbb{D}_\bullet) \in \mathcal{U}_w(\mathbf{k})$, $\mathbf{k}$ a field. Then the p-rank of X equals the p-rank of Y.*

(ii) Let $w = u_f$ (so that $f < g$) and $x = (X, \mathbb{E}_\bullet, \mathbb{D}_\bullet) \in \mathcal{F}_g(\mathbf{k})$. Then we have $x \in \mathcal{U}_w$ (respectively $\overline{\mathcal{U}}_w$) if and only if the filtration $\mathbb{E}_\bullet$ is V-stable and the p-rank of X is f and the a-number of X is 1 (respectively the p-rank of X is $\leq f$). The image of $\mathcal{U}_w$ in $\mathcal{A}_g$ is the locus of abelian varieties of p-rank f and a-number 1.

Proof. By definition we have $r_w(i, g+i) = i$ for all $1 \leq i \leq g$. Moreover, $r_w(i, g+i-1) = i$ precisely when $i \leq g-f$. Hence by Proposition 4.3 we see that $V(\mathbb{E}_i) \subseteq \mathbb{E}_i^{(p)}$ for all $1 \leq i \leq g$, $V(\mathbb{E}_i) \subseteq \mathbb{E}_{i-1}^{(p)}$ for $i \leq g-f$, and $V(\mathbb{E}_i) \subsetneq \mathbb{E}_{i-1}^{(p)}$ for $i > g-f$. The first and last conditions mean that V induces an isomorphism $V\colon \mathbb{E}_{\{i\}} \xrightarrow{\sim} \mathbb{E}_{\{i\}}$. On the other hand, the second condition gives $V^{g-f}(\mathbb{E}_{g-f}) = 0$. Together this gives that the semi-simple rank of X is f.

For the second part, since we have $w \leq w_\emptyset$ we must check the condition on the p-rank and the a-number. By the definition on $\mathcal{U}_w$ we have $x \in \mathcal{U}_w$ if and only if $\mathrm{rk}(\mathbb{E}_g \cap \mathbb{D}_j) = 1$ for $f+1 \leq j \leq g$ and $\mathrm{rk}(\mathbb{E}_g \cap \mathbb{D}_j) = 0$ for $1 \leq j \leq f$. This implies that the kernel of V (which equals $\mathbb{D}_g \cap \mathbb{E}_g$) has rank 1 and the semi-simple rank of V on $\mathbb{E}_g$ is f. For $x \in \overline{\mathcal{U}}_w$ we get instead $\mathrm{rk}(\mathbb{E}_g \cap \mathbb{D}_j) \leq 1$ for $f+1 \leq j \leq g$ and $\mathrm{rk}(\mathbb{E}_g \cap \mathbb{D}_j) = 0$ for $1 \leq j \leq f$. □

Also, the strata $\mathcal{U}_w$ with $w \in S_g$ admit a relatively simple interpretation. Recall that an abelian variety is called *superspecial* if its a-number is equal to its dimension. This happens if and only if the abelian variety (without polarization) is geometrically isomorphic to a product of supersingular elliptic curves.

Lemma 4.10. *Let x be a geometric point of $\mathcal{F}_g$ lying over $[X] \in \mathcal{A}_g$. The following four statements are equivalent:*

1. $x \in \cup_{w \in S_g} \mathcal{U}_w$.
2. $\dim(\mathbb{E}_g \cap \mathbb{D}_g) \geq g$.
3. $\ker(V) = \mathbb{E}_g$.
4. *The underlying abelian variety X is superspecial.*

Proof. An abelian variety X is superspecial if and only if X is a product of supersingular elliptic curves, and this condition is equivalent to $\dim(\mathbb{E}_g \cap \mathbb{D}_g) \geq g$. This explains the equivalences of (2), (3), and (4). If $x \in \mathcal{U}_w$ with $w \in S_g$ then $r_w(g,g) = g$; hence (2) holds. Conversely, if X is superspecial then any filtration $\mathbb{E}_\bullet$ on $\mathbb{E}_g$ is V-stable and can be extended to a symplectic filtration. The lemma now follows from the observation that the degeneracy strata for $w \in S_g$, the Weyl group of GL_g, cover the flag space of flags on $\mathbb{E}_g$. □

Lemma 4.11. *Let x be a point of $\mathcal{U}_w$ with underlying abelian variety X. Then the a-number of X equals $a(w)$. Moreover, if $Y = \{1, 2, \ldots, a\}$ with corresponding final element $w_Y \in W_g$ then the image of $\mathcal{U}_{w_Y}$ in $\mathcal{A}_g$ is the locus T_a of abelian varieties with a-number a.*

Proof. The a-number of an abelian variety is by definition the dimension of the kernel of V on $H^0(X, \Omega^1_X)$. But this is equal to $r_w(g,g) = a(w)$. The condition that $a(X) = a$ implies that $r_w(g,g) = a$; hence $\nu(g) = g - a$. This implies that $\nu(g - a + i) \geq i$ for $i = 1, \ldots, a$. Therefore the "smallest" ν satisfying these conditions is ν_{w_Y}. □

4.3 Shuffling flags

Our first result on the stratification will concern the case in which the p-rank is positive. All in all, the étale and multiplicative parts of the kernel of multiplication by p on the abelian variety have very little effect on the space of flags on its de Rham cohomology. There is, however, one exception to this.

The most natural thing to do is to put the multiplicative part at the bottom (i.e., the first steps of the flag, and thus, by self-duality, the étale part at the top), which is what automatically happens for a final filtration (on the conjugate filtration, that is). We may, however, start with a final filtration and then "move" the μ_p-factors upward. Note that over a perfect field the kernel of multiplication by p is the direct sum of its multiplicative, local-local, and étale parts, so that this is always possible. In general, however, it is possible only after a purely inseparable extension. This means that we get an inseparable map from a stratum where not all the μ_p-factors are at the bottom to a stratum where they all are. We intend first to give a combinatorial description of the strata that can be obtained in this way from a final stratum and then to compute the degrees of the inseparable maps involved. However, since we have to compute an inseparable degree, we should work with Hodge filtrations instead of conjugate filtrations, since conjugate filtrations kill some infinitesimal information. This causes a slight conceptual problem, since the V-simple parts in a final filtration are to be found "in the middle" rather than at the top and bottom (recall that V maps the top part of the conjugate filtration to the bottom of the Hodge filtration). This will not be a technical problem, but the reader will probably be helped by keeping it in mind.

It turns out that the arguments used do not change if instead of considering shuffles of final elements we consider shuffles of semi-simply final elements. We shall treat the more general case, since we shall need it later.

Hence we pick a subset $\tilde{I} \subseteq \{1, 2, \dots, g\}$ and let $\overline{\mathcal{U}}_{\tilde{I}}^{ss}$ be the closed subscheme of $\mathcal{F}_g$ defined by the conditions that V map $\mathbb{E}_i^{(p)}$ to $\mathbb{E}_i^{(p)}$ for all $1 \le i \le g$ and to $\mathbb{E}_{i-1}^{(p)}$ for $i \notin \tilde{I}$. Hence $\overline{\mathcal{U}}_w \subseteq \overline{\mathcal{U}}_{\tilde{I}}^{ss}$ precisely when $w \le w_\emptyset$ and the semi-simple index set of w is a subset of $\tilde{I}$. We also put

$$\mathcal{U}_{\tilde{I}}^{ss} := \overline{\mathcal{U}}_{\tilde{I}}^{ss} \setminus \cup_{\tilde{I}' \subset \tilde{I}} \overline{\mathcal{U}}_{\tilde{I}'}^{ss},$$

so that $\mathcal{U}_w \subseteq \mathcal{U}_{\tilde{I}}^{ss}$ precisely when $w \le w_\emptyset$ and its semi-simple index set is equal to $\tilde{I}$. If $I := \{g+1-i : i \in \tilde{I}\}$ we get from Proposition 2.19 that these w are precisely those of the form $\sigma^I w' \sigma_I^{-1}$ for the semi-simply final w'.

We are now going to construct, for every $I \subseteq \{1, 2, \dots, g\}$, a morphism $S_I \colon \mathcal{U}_{\tilde{I}}^{ss} \to \mathcal{U}_{\{g-f+1,\dots,g\}}^{ss}$, where $\tilde{I} := \{g+1-i : i \in I\}$ and $\#\tilde{I} = f$.

Let $\tilde{\imath}$ be the reduction index of the elementary reduction I' of I and put $i := g + 1 - \tilde{\imath}$. By Proposition 2.19 we have that $r_w(i+1, g+i) = i+1$ and $w(i) = g+i$. This means that if $\mathbb{E}_\bullet$ is the (tautological) Hodge flag on $\mathcal{U}_{\tilde{I}}^{ss}$, then $V(\mathbb{E}_{i+1}) \subseteq \mathbb{E}_i^{(p)}$ and $V(\mathbb{E}_i) \subsetneq \mathbb{E}_{i-1}^{(p)}$ everywhere on $\mathcal{U}_{\tilde{I}}^{ss}$. This means that V induces a bijection on $\mathbb{E}_{\{i\}}$ and is zero on $\mathbb{E}_{\{i+1\}}$. Let Id denote the map $s \mapsto 1 \otimes s$ from $\mathcal{T}$ to $\mathcal{T}^{(p)}$ for the sheaves involved. Then the map induced by the quotient map

$$\mathrm{Ker}(\mathrm{Id} - V)_{\mathbb{E}_{\{i+1,i\}}} \to \mathrm{Ker}(\mathrm{Id} - V)_{\mathbb{E}_{\{i\}}},$$

where the kernel is computed in the étale topology on $\mathcal{U}_w$, is an isomorphism. Also for a sheaf $\mathcal{T}$ equipped with a linear isomorphism $V \colon \mathcal{T} \to \mathcal{T}^{(p)}$, we have an isomorphism $\mathrm{Ker}(V - I) \bigotimes \mathcal{O} \to \mathcal{T}$ with $\mathcal{O}$ the structure sheaf. It follows that the short exact sequence

$$0 \to \mathbb{E}_{\{i\}} \longrightarrow \mathbb{E}_{\{i+1,i\}} \longrightarrow \mathbb{E}_{\{i+1\}} \to 0$$

splits uniquely in a way compatible with V. This means that we may define a new flag where $\mathbb{E}'_j = \mathbb{E}_j$ for $j \neq i$ and $\mathbb{E}'_i/\mathbb{E}'_{i-1} = \mathbb{E}_{i+1}/\mathbb{E}_i$. Then the classifying map $\mathcal{U}^{ss}_{\tilde{I}} \to \mathcal{F}_g$ for this new flag will have its image in $\mathcal{U}^{ss}_{\tilde{I}'}$. Repeating this process, we end up with a flag whose classifying map will have its image in $\mathcal{U}_{\{g-f+1,\dots,g\}}$, which by definition is our map S_I.

Proposition 4.12. *If $I \subseteq \{1,\dots,g\}$ then the map $S_I \colon \mathcal{U}^{ss}_{\tilde{I}} \to \mathcal{U}^{ss}_{\{g-f+1,\dots,g\}}$ is finite, radicial, and surjective.*

Proof. To get from a point of $\mathcal{U}^{ss}_{\{g-f+1,\dots,g\}}$ to one of $\mathcal{U}^{ss}_{\tilde{I}}$ one has to find a V-invariant complement to some $\mathbb{E}_i/\mathbb{E}_{i-1}$ in $\mathbb{E}_{i+1}/\mathbb{E}_{i-1}$. Since V will be zero on $\mathbb{E}_{i+1}/\mathbb{E}_i$ and bijective on $\mathbb{E}_i/\mathbb{E}_{i-1}$, a complement over the fraction field of a discrete valuation ring will extend to a complement over the discrete valuation ring (since the complement cannot meet $\mathbb{E}_i/\mathbb{E}_{i-1}$ over the special fiber), so that the map is proper. It then remains to show that the map is a bijection over an algebraically closed field. In that case $\mathbb{E}_g$ splits canonically as a sum of a V-nilpotent part and a V-semisimple part, and the bijectivity is clear. □

In order to determine the degree (necessarily of inseparability) we shall do the same factorization as in the definition of S_I, so that we may consider the situation of $\tilde{I}$ with $\tilde{i}$ and I' being the reduction index respectively elementary reduction of I. For the tautological flag $\mathbb{E}_\bullet$ on $\mathcal{U}^{ss}_{\tilde{I}}$ we have that V is an isomorphism on $\mathbb{E}_{\{i\}}$ and zero on $\mathbb{E}_{\{i+1\}}$, while the opposite is true on $\mathcal{U}^{ss}_{\tilde{I}'}$.

Lemma 4.13. *The map $\mathcal{U}^{ss}_{\tilde{I}} \to \mathcal{U}^{ss}_{\tilde{I}'}$ is flat of degree p.*

Proof. We consider the partial symplectic flag space $\mathcal{F}_g(i)$ consisting of the flags of $\mathcal{F}_g$ by removing the ith member $\mathbb{D}_i$ and its annihilator. This means that we have a $\mathbb{P}^1$-bundle $\mathcal{F}_g \to \mathcal{F}_g(i)$. Now, under this map $\mathcal{U}^{ss}_{\tilde{I}}$ and $\mathcal{U}^{ss}_{\tilde{I}'}$ map to the same subscheme $\mathcal{U} \subseteq \mathcal{F}_g(i)$, and the map $\mathcal{U}^{ss}_{\tilde{I}} \to \mathcal{U}^{ss}_{\tilde{I}'}$ is compatible with these projections. Over $\mathcal{U}$ put $\mathcal{E} := \mathbb{E}_{\{i,i+1\}}$, $\mathcal{M} := \ker(V \colon \mathcal{E} \to \mathcal{E}^{(p)})$, and $\mathcal{L} := \mathrm{Im}(V \colon \mathcal{E} \to \mathcal{E}^{(p)})$. Then on the $\mathbb{P}^1$-bundle $\pi \colon \mathcal{F}_g \to \mathcal{F}_g(i)$, the subscheme $\mathcal{U}^{ss}_{\tilde{I}'}$ is defined by the vanishing of the composite map $\mathcal{O}(-1) \to \pi^*\mathcal{E} \to \pi^*(\mathcal{E}/\mathcal{M})$ and in fact gives a section of $\mathcal{F}_g$ over $\mathcal{U}$ given by the line subbundle $\mathcal{M} \subset \mathcal{E}$. Hence it is enough to show that the projection map $\mathcal{U}^{ss}_{\tilde{I}} \to \mathcal{F}_g(i)$ is flat of degree p. We have that $\mathcal{U}^{ss}_{\tilde{I}'} \subseteq \mathcal{F}_g$ is defined by the vanishing of the composite $\mathcal{O}(-1)^{(p)} \to \pi^*\mathcal{E}^{(p)} \to \pi^*(\mathcal{E}^{(p)}/\mathcal{M})$. It is then enough to show that $\mathcal{U}^{ss}_{\tilde{I}'} \subseteq \mathcal{F}_g$ is a relative Cartier divisor, and for that it is enough to show that it is a proper subset in each fiber of $\mathcal{F}_g \to \mathcal{F}_g(i)$. This, however, is clear, since for a geometric point of $\mathcal{F}_g(i)$ there are just two points that lie in $\overline{\mathcal{U}}_\emptyset$, given by $\mathcal{M}$ and $\mathcal{L}^{p^{-1}}$. □

Composing these maps, we get the following proposition.

Proposition 4.14. *Let $I \subseteq \{1, \ldots, g\}$ and $\tilde{I} := \{g+1-i : i \in I\}$. Then the map $S_I \colon \mathcal{U}_I^{ss} \to \mathcal{U}_{\{1,\ldots,g\}}^{ss}$ is a finite purely inseparable map of degree $p^{\mathrm{ht}(I)}$.*

Proof. The flatness and the degree of S_I follow by factoring it by maps as in Lemma 4.13 and noting that the number of maps is $\mathrm{ht}(I)$. The rest then follows from Proposition 4.12. □

Remark 4.15. The result implies in particular that if w' is a shuffle of w by I, then $S_I \colon \mathcal{U}_{w'} \to \mathcal{U}_w$ is flat and purely inseparable of degree $p^{\mathrm{ht}(I)}$. We shall later (see Corollary 8.4) show that $\mathcal{U}_{w'}$ and $\mathcal{U}_w$ are reduced. This shows that over a generic point of $\mathcal{U}_w$ each simple shuffle toward w' really requires a finite inseparable extension of degree p. This is a kind of nondegeneracy statement that is the inseparable analogue of maximal monodromy (of which we shall see some examples later on). It can also be seen as saying that a certain Kodaira–Spencer map is injective.

4.4 The E-O strata on $\mathcal{A}_g \otimes \mathbb{F}_p$

Definition 4.16. *Let $w \in W_g$ be a final type. Then the E-O stratum $\mathcal{V}_w$ associated to w is the locally closed subset of $\mathcal{A}_g$ of points x for which the canonical type of the underlying abelian variety is equal to the canonical type of w. We let $\overline{\mathcal{V}}_w$ be the closure of $\mathcal{V}_w$.*

It is known that the dimension of $\mathcal{V}_w$ is equal to $\dim(w)$ [Oo01]. This and the fact that the E-O strata form a stratification will also follow from our results in Sections 8 and 9.3.

5 Extension to the boundary

The moduli space $\mathcal{A}_g$ admits several compactifications. The Satake or Baily–Borel compactification $\mathcal{A}_g^*$ is in some sense minimal, cf. [FC90, Chapter V]. It is a stratified space

$$\mathcal{A}_g^* = \bigsqcup_{i=0}^{g} \mathcal{A}_i.$$

Chai and Faltings define in [FC90] a class of smooth toroidal compactifications of $\mathcal{A}_g$. If $\tilde{\mathcal{A}}_g$ is such a toroidal compactification then there is a natural map $q \colon \tilde{\mathcal{A}}_g \to \mathcal{A}_g^*$ extending the identity on $\mathcal{A}_g$. This induces a stratification of $\tilde{\mathcal{A}}_g$:

$$\tilde{\mathcal{A}}_g = \bigsqcup_{i=0}^{g} q^{-1}(\mathcal{A}_{g-i}) = \bigsqcup_{i=0}^{g} \mathcal{A}_g^{\langle i \rangle}.$$

The stratum $\mathcal{A}_g^{\langle i \rangle}$ parametrizes the semi-abelian varieties of torus rank i.

The Hodge bundle $\mathbb{E}$ on $\mathcal{A}_g$ can be extended to a rank g vector bundle, again denoted by $\mathbb{E}$, on $\tilde{\mathcal{A}}_g$. On $\mathcal{A}_g^{\langle i\rangle}$ the Hodge bundle fits into a short exact sequence

$$0 \to \mathbb{E}' \to \mathbb{E} \to \mathbb{E}'' \to 0,$$

where $\mathbb{E}'$ is a rank $g-i$ bundle and $\mathbb{E}''$ can be identified with the cotangent bundle to the toric part of the universal semi-abelian variety along the identity section over $\mathcal{A}_g^{\langle i\rangle}$. The bundle $\mathbb{E}'$ is the pullback under $q\colon \mathcal{A}_g^{\langle i\rangle} \to \mathcal{A}_g^*$ of the Hodge bundle on $\mathcal{A}_{g-i}$.

The Verschiebung V acts in a natural way on the above short exact sequence and it preserves $\mathbb{E}'$. It induces an action on $\mathbb{E}''$ with trivial kernel because $\mathbb{E}''$ comes from the toric part and is generated by logarithmic forms.

The de Rham bundle $\mathbb{H}$ on $\mathcal{A}_g$ also admits an extension (denoted again by $\mathbb{H}$). This is the logarithmic de Rham sheaf $R^1\pi_*(\Omega^\bullet_{\tilde{\mathcal{X}}_g/\tilde{\mathcal{A}}_g}(\log))$, where the log refers to allowing logarithmic singularities along the divisor at infinity; cf. [FC90, Theorem VI, 1.1]. We have a short exact sequence

$$0 \to \mathbb{E} \to \mathbb{H} \to \mathbb{E}^\vee \to 0$$

extending the earlier mentioned sequence on $\mathcal{A}_g$. The symplectic form on $\mathbb{H}$ extends as well.

We now want to compare strata on $\mathcal{A}_g$ and $\tilde{\mathcal{A}}_g$, and for this we introduce some notation. For a given integer $1 \le i \le g$ we can consider the Weyl group W_{g-i} as a subgroup of W_g by letting it act on the set $\{i+1, i+2, \dots, g, \dots, 2g-i\}$ via the bijection $j \longleftrightarrow i+j$ for $1 \le j \le g-i$. More precisely, define $\rho_i\colon W_{g-i} \to W_g$ via

$$\rho_i(w)(l) = \begin{cases} i + w(l) & \text{for } 1 \le l \le g-i, \\ g + l & \text{for } g-i+1 \le l \le g. \end{cases}$$

This map respects the Bruhat–Chevalley order, and final elements are mapped to final elements.

Since symplectic flags on $\mathbb{H}$ are determined by their restriction to $\mathbb{E}$ and since we can extend $\mathbb{E}$ to $\tilde{\mathcal{A}}_g$, we can extend $\mathcal{F}_g$ to a flag space bundle $\tilde{\mathcal{F}}_g$ over $\tilde{\mathcal{A}}_g$. Then we can also consider the degeneracy loci $\mathcal{U}_w$ and $\overline{\mathcal{U}}_w$ for $\tilde{\mathcal{F}}_g$. We shall use the same notation for these extensions.

Similarly, we can define the notion of a canonical filtration for a semi-abelian variety. If $1 \to T \to A \to A' \to 0$ is a semi-abelian variety with abelian part A' and toric part T of rank t and if the function ν' on $\{0, c_1, \dots, c_r, c_{r+1}, \dots, c_{2r} = 2\dim(A')\}$ is the canonical type of A', then we define the canonical type of A to be the function ν on

$$\{0, t, t+c_1, \dots, t+c_r, t+c_{r+1}, \dots, t+c_{2r}, 2g-t, 2g\}$$

defined by $\nu(t+c_i) = t + \nu'(c_i)$. Using this definition we can extend the E-O stratification to $\tilde{\mathcal{A}}_g$.

The stratification of $\tilde{\mathcal{A}}_g$ by the strata $\mathcal{A}_g^{\langle i\rangle}$ induces a stratification of $\tilde{\mathcal{F}}_g$ by flag spaces $\tilde{\mathcal{F}}_g^{\langle i\rangle}$. Recall that the stratum $\mathcal{A}_g^{\langle i\rangle}$ admits a map $q\colon \mathcal{A}_g^{\langle i\rangle} \to \mathcal{A}_{g-i}$ induced by the natural map $\tilde{\mathcal{A}}_g \to \mathcal{A}_g^*$. Similarly, we have a natural map $\pi_i = \pi\colon \mathcal{F}_g^{\langle i\rangle} \to \mathcal{F}_{g-i}$ given by restricting the filtration on $\mathbb{E}$ to $\mathbb{E}'$.

We now describe the interplay between the two stratifications $(\mathcal{F}_g^{\langle i\rangle})_{i=1}^g$ and $(\mathcal{U}_w)_{w\in W_g}$.

Lemma 5.1. *Let $w \in W_g$ be an element with $w \leq w_\emptyset$.*

(i) We have $\mathcal{U}_w \cap \mathcal{F}_g^{\langle i\rangle} \neq \emptyset$ if and only if w is a shuffle of an element in $\rho_i(W_{g-i})$.

(ii) If $w = \rho_i(w')$ with associated degeneracy loci $\mathcal{U}_w \subset \mathcal{F}_g$ and $\mathcal{U}_{w'} \subset \mathcal{F}_{g-i}$ then we have $\mathcal{U}_w \cap \mathcal{F}_g^{\langle i\rangle} = \pi_i^{-1}(\mathcal{U}_{w'})$.

(iii) At a point x of $\tilde{\mathcal{A}}_g$ for which the torus part of the "universal" semi-abelian scheme has rank r, there is a formally smooth map from the formal completion of $\tilde{\mathcal{A}}_g$ at x to the formal multiplicative group $\hat{\mathbb{G}}_m^r$ with the following properties. The locus where the torus rank of the universal semi-abelian scheme is $s \leq r$ is the inverse image of the locus of points of $\hat{\mathbb{G}}_m^r$ where $r - s$ coordinates are 1. The restriction of this map to the formal completion of any $\overline{\mathcal{U}}_w$ containing x is formally smooth.

(iv) In particular, $\mathcal{U}_w$ is the closure of its intersection with $\mathcal{F}_g$.

Proof. A V-stable filtration on $\mathbb{E}$ restricts to a V-stable filtration on $\mathbb{E}'$. If $\mathcal{U}_w \cap \mathcal{F}_g^{\langle i\rangle}$ is not empty, then it determines a $w' \in W_{g-i}$ such that $\mathcal{U}_w \cap \mathcal{F}_g^{\langle i\rangle} \subseteq \pi_i^{-1}(\mathcal{U}_{w'})$. Since V is invertible on $\mathbb{E}''$, one sees that w is a shuffle of $\rho_i(w')$ and that $\mathcal{U}_w \cap \mathcal{F}_g^{\langle i\rangle} = \pi_i^{-1}(\mathcal{U}_{w'})$.

The third part is a direct consequence of the local construction of $\tilde{\mathcal{A}}_g$ using toroidal compactifications and of the universal semi-abelian variety using Mumford's construction, where it is defined by taking the quotient of a semi-abelian variety by a subgroup of the torus part, the subgroup being generated by the coordinate functions of $\hat{\mathbb{G}}_m^r$ (see [FC90, Section III.4] for details). Since $\mathbb{H}$ and $\mathbb{E}$ depend only on that universal semi-abelian scheme, it is clear that the restriction of the map to a $\overline{\mathcal{U}}_w$ is smooth.

The last part follows directly from the third. □

Note also that Lemma 5.1 is compatible with shuffling. It also results from this lemma that we can define the E-O stratification on the Satake compactification by considering either the closure of the stratum $\mathcal{V}_w$ on $\mathcal{A}_g$ or the images of the final strata $\mathcal{V}_w$ on $\tilde{\mathcal{A}}_g$.

6 Existence of boundary components

Our intent in this section is to show the existence in irreducible components of our strata of points in the smallest possible stratum $\overline{\mathcal{U}}_1$, the stratum associated to the identity element of W_g.

Proposition 6.1. *Let X be an irreducible component of any $\overline{\mathcal{U}}_w$ in $\mathcal{F}_g$. Then X contains a point of $\overline{\mathcal{U}}_1$.*

Proof. We prove this by induction on g and on the Bruhat–Chevalley order of w. The statement is clear for $g = 1$. We start off by choosing a Chai–Faltings compactification $\tilde{\mathcal{A}}_g$ of $\mathcal{A}_g$ with a semi-abelian family over it (and a "principal" cubical structure so that we get a principal polarization on the semi-abelian variety modulo its toroidal part).

What we now actually want to prove is the same statement as in the proposition but for $\tilde{\mathcal{F}}_g$ instead. Since $\overline{\mathcal{U}}_1$ is contained in $\mathcal{F}_g$, the result will follow. We start off by considering $Y := X \cap (\tilde{\mathcal{F}}_g \backslash \mathcal{F}_g)$. Assume that Y is nonempty and irreducible. If it is not, then we replace it by an irreducible component of Y. Then Y is contained in $\pi_1^{-1}(\overline{\mathcal{U}}_{w'})$ with $\rho_1(w') = w$ for some $w' \in W_{g-1}$. We claim that Y is an irreducible component of $\pi_1^{-1}(\overline{\mathcal{U}}_{w'})$. This follows from the fact that "we can freely move the toroidal part into an abelian variety," which is Lemma 5.1, part (iv).

By induction on g we can assume that $\overline{\mathcal{U}}_{w'}$ in $\mathcal{F}_{g-1}$ contains $\overline{\mathcal{U}}_{1'}$, where $1'$ is the identity element of W_{g-1}. Any component Z of $\overline{\mathcal{U}}_{\rho_1(1')}$ that lies in X and meets Y does not lie completely in the boundary $\tilde{\mathcal{F}}_g - \mathcal{F}_g$. By induction on the Bruhat–Chevalley order we can assume that $w = \rho_1(1')$ and $X = Z$. Note also that for any $w'' < w$ we have that $\overline{\mathcal{U}}_{w''}$ does not meet the boundary, and by induction on the Bruhat–Chevalley order we get that $X = \mathcal{U}_w \cap X$. On the other hand, if X does not meet the boundary we immediately get the same conclusion.

Hence we may and shall assume that Y has the property that it lies completely inside $\mathcal{U}_w$ and that it is proper. Lemma 6.2 now shows that it has an ample line bundle of finite order, which together with properness forces Y to be zero-dimensional. Now observe that $\dim Y \geq \ell(w)$ (the proof is analogous to [Ful, Theorem 14.3]). This gives $\ell(w) = 0$, and so $w = 1$ and we are reduced to a trivial case. □

The following is a version of the so-called Raynaud trick.

Lemma 6.2. *Suppose that X is a proper irreducible component of $\overline{\mathcal{U}}_w$ inside $\mathcal{F}_g$ such that $X \cap \mathcal{U}_w = X$. Then X is 0-dimensional.*

Proof. We have the variety X and two symplectic flags $\mathbb{E}_\bullet$ and $\mathbb{D}_\bullet$ that at all points of X are in the same relative position w. It follows from Lemma 4.2 that we have an isomorphism between $\mathcal{L}_i := \mathbb{E}_i/\mathbb{E}_{i-1}$ and $\mathcal{M}_{w(i)} := \mathbb{D}_{w(i)}/\mathbb{D}_{w(i)-1}$ over X, and then since we also have isomorphisms between $\mathcal{L}_i^p$ and $\mathcal{M}_{g+i}$ and $\mathcal{L}_i$ and $\mathcal{L}_{2g+1-i}^{-1}$, we conclude that all the $\mathcal{L}_i$ have finite order. On the other hand, we know from the theory of flag spaces that $\mathcal{L}_{2g} \bigotimes \mathcal{L}_{2g-1} \bigotimes \cdots \bigotimes \mathcal{L}_{g+1}$ is relatively ample, and since $\mathcal{L}_g \bigotimes \mathcal{L}_{g-1} \bigotimes \cdots \bigotimes \mathcal{L}_1$ is the pullback of an ample line bundle over the base $\mathcal{A}_g$, we conclude. □

7 Superspecial fibers

We shall now discuss the fiber of $\mathcal{F}_g \to \mathcal{A}_g$ over superspecial points. The superspecial abelian varieties are characterized by the condition that $\mathbb{E}_g = \mathbb{D}_g$, i.e., the strata $\mathcal{U}_w$ for which $w \in S_g$. Furthermore, V induces an isomorphism $\mathbb{E}/\mathbb{E}_g \xrightarrow{\sim} \mathbb{E}_g^{(p)}$. On the other hand, the polarization gives an isomorphism $(\mathbb{E}_g)^* \xrightarrow{\sim} \mathbb{E}/\mathbb{E}_g$. This leads to the following definition.

Definition 7.1. *(i) Let S be an $\mathbb{F}_p$-scheme. A p-*unitary *vector bundle is a vector bundle $\mathcal{E}$ over S together with an isomorphism $F^*\mathcal{E} \xrightarrow{\sim} \mathcal{E}^*$, where $F\colon S \to S$ is the (absolute) Frobenius map.*

*(ii) Let $\mathcal{E}$ be a p-unitary vector bundle over S and let $P \to S$ be the bundle of complete flags on $\mathcal{E}$. The p-*unitary Schubert stratification *of P is the stratification given by letting $\mathcal{U}_w$, $w \in S_g$, consist of the points for which the universal flag $\mathcal{F}$ and the dual of the Frobenius pullback $(F^*\mathcal{F})^*$ are in position corresponding to w.*

A map $F^*\mathcal{E} \to \mathcal{E}^*$ of vector bundles is the same thing as a map $F^*\mathcal{E} \bigotimes_{\mathcal{O}_S} \mathcal{E} \to \mathcal{O}_S$, which in turn corresponds to a biadditive map $\langle -,- \rangle\colon \mathcal{E} \times \mathcal{E} \to \mathcal{O}_S$ fulfilling $\langle fa, b \rangle = f^p \langle a, b \rangle$ and $\langle a, fb \rangle = f \langle a, b \rangle$. We shall normally use this latter description.

All p-unitary vector bundles are trivial in the étale topology, as the following proposition shows.

Proposition 7.2. *If $\langle -,- \rangle$ is a p-unitary structure on the vector bundle $\mathcal{E}$ then*

$$E := \{ a \in \mathcal{E} : \forall b \in \mathcal{E}\colon \langle b, a \rangle = \langle a, b \rangle^p \}$$

is a local (in the étale topology) system of $\mathbb{F}_{p^2}$-vector spaces. Furthermore, $\langle -,- \rangle$ induces a unitary (with respect to the involution $(-)^p$ on $\mathbb{F}_{p^2}$) structure on E. Conversely, if E is a local system of $\mathbb{F}_{p^2}$-vector spaces with a unitary structure, then $\mathcal{E} := E \bigotimes_{\mathbb{F}_{p^2}} \mathcal{O}_S$ is a p-unitary vector bundle.

These two constructions establish an equivalence between the categories of p-unitary vector bundles and that of local systems of unitary $\mathbb{F}_{p^2}$-vector spaces. In particular, all p-unitary vector bundles (of the same rank) are locally isomorphic in the étale topology.

Proof. The pairing $\langle -,- \rangle$ gives rise to an isomorphism $\psi'\colon \mathcal{E} \xrightarrow{\sim} (F^*\mathcal{E})^*$ by $a \mapsto b \mapsto \langle b, a \rangle$ and an isomorphism $\psi\colon (F^2)^*\colon \mathcal{E} \xrightarrow{\sim} (F^*\mathcal{E})^*$ by $a \mapsto b \mapsto \langle a, b \rangle^p$. The composite $\rho := \psi^{-1} \circ \psi'$ thus gives an isomorphism $\mathcal{E} \xrightarrow{\sim} (F^2)^*\mathcal{E}$. Then E is simply the kernel of $\rho - 1 \otimes id$, and the fact that $\mathcal{E} = E \bigotimes_{\mathbb{F}_{p^2}} \mathcal{O}_S$ follows from [HW61]. The pairing $\langle -,- \rangle$ then induces a unitary pairing on E, which is perfect, since $\langle -,- \rangle$ is. Conversely, it is clear that a unitary pairing on E translates to one on $\mathcal{E}$.

Finally, since all (perfect) unitary pairings on $\mathbb{F}_{p^2}$-vector spaces of fixed dimension are isomorphic, we get the local isomorphism. □

Proposition 7.2 has the following immediate corollary.

Corollary 7.3. *The flag variety fibrations of two p-unitary vector bundles of the same rank on the same base are locally isomorphic by an isomorphism preserving the p-unitary Schubert strata.*

Let us now consider the situation in which the base scheme is $\mathrm{Spec}(\mathbb{F}_{p^2})$ and $\mathcal{E}$ is an $\mathbb{F}_{p^2}$-vector space given a unitary perfect pairing. The smallest p-unitary Schubert stratum consists of flags that coincide with their unitary dual. Taking duals once more, we see that they are taken to themselves after pullback by the square of the Frobenius, hence are defined over $\mathbb{F}_{p^2}$. Furthermore, they are self-dual with respect to the unitary pairing. This should come as no surprise, since that stratum corresponds to final filtrations on superspecial abelian varieties. The next-to-lowest strata are somewhat more interesting.

Lemma 7.4. *Let V be a g-dimensional $\mathbb{F}_{p^2}$ unitary vector space and let $\mathbb{P}$ be projective space based on V. If $s = (i, i+1) \in S_g$ for some $1 \le i < g$ then the closed Schubert stratum $\mathcal{U}_s \subseteq \mathbb{P}$ consists of the flags $0 = \mathbb{E}_0 \subset \mathbb{E}_1 \subset \cdots \subset \mathbb{E}_g$, where the $\mathbb{E}_j$ for $j \ne i, g-i$ are $\mathbb{F}_{p^2}$-rational with $\mathbb{E}_j^{\perp} = \mathbb{E}_{g-j}$, $\mathbb{E}_{g-i} = (\mathbb{E}_i^{(p)})^{\perp}$ unless $i = 2g$, and $\mathbb{E}_i \ne (\mathbb{E}_{g-i}^{(p)})^{\perp}$.*

Proof. If $\mathbb{E}_\bullet$ and $\mathbb{D}_\bullet$ are two flags in $V \bigotimes R$ (R some $\mathbb{F}_{p^2}$-algebra) and they are in position s, then $\dim(\mathbb{E}_j \cap \mathbb{D}_j) = r_s(j,j) = j$ for $i \ne j$, i.e., $\mathbb{E}_j = \mathbb{D}_j$ and for i the conditions give us $\mathbb{E}_i \cap \mathbb{D}_i = \mathbb{E}_{i-1}$. In our case, where $\mathbb{D}_j = F^*\mathbb{E}_{g-j}^{\perp}$, this means $\mathbb{E}_j = F^*\mathbb{E}_{g-j}^{\perp}$ for $j \ne i$ and $\mathbb{E}_i \ne (\mathbb{E}_{g-i}^{(p)})^{\perp}$. If also $j \ne g-i$ we can use this twice and get that $\mathbb{E}_j = \mathbb{E}_j^{(p^2)}$, i.e., $\mathbb{E}_j$ is $\mathbb{F}_{p^2}$-rational. □

As a result we get the following connectedness result, analogous to [Oo01, Proposition 7.3].

Theorem 7.5. *Let V be a g-dimensional $\mathbb{F}_{p^2}$ unitary vector space and let $\mathbb{P}$ be projective space based on V. Let $S \subseteq \{1, \ldots, g-1\}$. Let $\overline{\mathcal{U}}$ be the union of the $\overline{\mathcal{U}}_{s_i}$ for $i \in S$. Then two flags $0 = A_0 \subset A_1 \subset \cdots \subset A_{g-1} \subset A_g$ and $0 = B_0 \subset B_1 \subset \cdots \subset B_{g-1} \subset B_g$ in $\overline{\mathcal{U}}_1$ lie in the same component of $\overline{\mathcal{U}}$ precisely when $B_i = A_i$ for all $i \notin S$. Furthermore, every connected component of $\overline{\mathcal{U}}$ contains an element of $\overline{\mathcal{U}}_1$.*

Proof. The last statement is clear, since every irreducible component of any $\overline{\mathcal{U}}_{s_i}$ contains a point of $\overline{\mathcal{U}}_1$. This follows from Proposition 6.1 but can also easily be seen directly.

We start by looking at the locus $\overline{\mathcal{U}}_F^i$ of a $\overline{\mathcal{U}}_{s_i}$ with $1 \le i < g$ of flags for which all the components of the flag except the dimension i and dimension $g-i$ parts are equal to a fixed (partial) $\mathbb{F}_{p^2}$-rational self-dual flag $F_\bullet$. The following claims are easily proved using Lemma 7.4.

(i) For any $1 \le i \le (g-2)/2$ or $(g+2)/2 \le i < g$ we get an element in $\overline{\mathcal{U}}_F^i$ by picking any $\mathbb{E}_{i-1} \subset \mathbb{E}_i \subset \mathbb{E}_{i+1}$ and then letting $\mathbb{E}_{g-i}$ be determined by Lemma 7.4. Hence the locus is isomorphic to $\mathbb{P}^1$ and the intersection with $\overline{\mathcal{U}}_1$ consists of the points for which $\mathbb{E}_i$ and $\mathbb{E}_{g-i}$ are $\mathbb{F}_{p^2}$-rational.

(ii) When g is even we get an element in $\overline{\mathcal{U}}^i_F$ by picking $\mathbb{E}_{g/2-1} \subset \mathbb{E}_{g/2} \subset \mathbb{E}_{g/2+1}$. Hence the locus is isomorphic to $\mathbb{P}^1$ and the intersection with $\overline{\mathcal{U}}_1$ consists of the points for which $\mathbb{E}_{g/2}$ is $\mathbb{F}_{p^2}$-rational.

(iii) When g is odd we get an element in $\overline{\mathcal{U}}^i_F$ by picking $\mathbb{E}_{(g-3)/2} \subset \mathbb{E}_{(g-1)/2} \subset \mathbb{E}_{(g+3)/2}$ for which $\overline{\mathbb{E}}_{(g-1)/2} \subset F^*\overline{\mathbb{E}}^{\perp}_{(g-1)/2}$, where $\overline{\mathbb{E}}_{(g-1)/2} = \mathbb{E}_{(g-1)/2}/\mathbb{E}_{(g-3)/2}$ and F^* comes from the $\mathbb{F}_{p^2}$-rational structure on $\mathbb{E}_{(g+3)/2}/\mathbb{E}_{(g-3)/2}$ and the scalar product is induced from that on $\mathbb{E}_g$, and then define $\mathbb{E}_{(g+1)/2}$ by the condition that $\mathbb{E}_{(g+1)/2}/\mathbb{E}_{(g-3)/2} = F^*\overline{\mathbb{E}}^{\perp}_{(g-1)/2}$. Since all nondegenerate unitary forms are equivalent, choosing a basis of $\mathbb{E}_{(g+3)/2}/\mathbb{E}_{(g-3)/2}$ for which the form has the standard form $\langle (x,y,z),(x,y,z)\rangle = x^{p+1}+y^{p+1}+z^{p+1}$ yields that $\overline{\mathcal{U}}^i_F$ is isomorphic to the Fermat curve of degree $p+1$ and hence is irreducible. The intersection with $\overline{\mathcal{U}}_1$ consists of the points for which $\mathbb{E}_{(g-1)/2}$ is $\mathbb{F}_{p^2}$-rational and then $\mathbb{E}_{(g+1)/2} = \mathbb{E}^{\perp}_{(g-1)/2}$.

(iv) When g is odd we get an element in $\overline{\mathcal{U}}^i_F$ by picking $\mathbb{E}_{(g+3)/2}$ fulfilling conditions dual to those of (iii). Hence again $\overline{\mathcal{U}}^i_F$ is irreducible and the intersection with $\overline{\mathcal{U}}_1$ consists of the points for which $\mathbb{E}_{(g+1)/2}$ is $\mathbb{F}_{p^2}$-rational and then $\mathbb{E}_{(g-1)/2} = \mathbb{E}^{\perp}_{(g+1)/2}$.

It follows from this description that two flags in $\overline{\mathcal{U}}_1$ lie in the same component of $\overline{\mathcal{U}}$ if and only if they are equivalent under the equivalence relation generated by the relations that for any unitary $\mathbb{F}_{p^2}$-flag $0 = A_0 \subset A_1 \subset \cdots \subset A_{g-1} \subset A_g$ we may replace it by any flag that is the same except for A_i and A_{g-i} for $i \in S$. The theorem then follows from the following lemma. □

Lemma 7.6. *(i) Let* $\mathbf{k}$ *be a field and let* $\mathcal{F}\ell_n$ *be the set of complete flags of vector spaces in a finite-dimensional vector space. The equivalence relation generated by the operations of modifying a flag* $E_\bullet$ *by, for any* i*, replacing* E_i *by any* i*-dimensional subspace of* E_{i+1} *containing* E_{i-1} *contains just one equivalence class.*

(ii) Let $\mathcal{F}\ell_n$ *the set of complete flags of vector spaces in an* n*-dimensional* $\mathbb{F}_{p^2}$*-vector space, self-dual with respect to a perfect unitary pairing. An* elementary modification *of such a flag* $E_\bullet$ *is obtained by either, for any* $1 \le i \le (n-1)/2$*, replacing* E_i *by any isotropic* i*-dimensional subspace of* E_{i+1} *containing* E_{i-1} *and* E_{n-i} *by its annihilator or, when* n *is even, replacing* $E_{n/2}$ *by any maximal totally isotropic subspace contained in* $E_{n/2+1}$ *and containing* $E_{n/2-1}$*. Then the equivalence relation generated by all elementary modifications contains just one equivalence class.*

Proof. Starting with (i), we prove it by induction on n, the dimension of the vector space V. Given two flags $E_\bullet$ and $F_\bullet$, if E_1 and F_1 are equal we may use induction applied to $E_\bullet/E_1$ and $F_\bullet/E_1$. We now use induction on the smallest j such that $E_1 \subseteq F_j$. The case $j = 1$ has already been taken care of. We now get a new flag $F'_\bullet$ by replacing F_{j-1} by $F_{j-2} \bigoplus E_1$, which works because $E_1 \subsetneq F_{j-1}$, and we then have $E_1 \subseteq F'_{j-1}$.

Continuing with (ii) we again use induction on n and start with two self-dual flags $E_\bullet$ and $F_\bullet$. Let us first assume that n is even, $n = 2k$. Then E_k and F_k are isotropic subspaces. If they have nontrivial intersection, then we may pick a 1-dimensional subspace contained in it and then use (i) to replace $E_\bullet$ and $F_\bullet$ by flags for which E_k and F_k are the same and $E_1 = F_1$. This implies that also $E_{n-1} = F_{n-1}$ and we may consider E_{n-1}/E_1 with its two flags induced from $E_\bullet$ and $F_\bullet$ and use induction to conclude. Assuming $F_k \cap E_k = \{0\}$ we may again use (i) to modify $F_\bullet$, keeping F_k fixed, so that $E_1 \subseteq F_{k+1}$. This means that $E_1 \bigoplus F_{k-1}$ is totally isotropic and we may replace F_k by it to obtain a new flag $F'_\bullet$ for which F'_k and E_k intersect nontrivially.

When n is odd, $n = 2k + 1$, we may again use induction on n to finish if E_k and F_k intersect nontrivially. If not, we may again use (i) to reduce to the case $E_1 \subseteq F_{k+2}$ and then we may replace F_k by $E_1 \bigoplus F_{k-1}$ and F_{k+1} by its annihilator. □

8 Local structure of strata

8.1 Stratified Spaces

We now want to show that $\mathcal{F}_g$ looks locally like the space of complete symplectic flags (in $2g$-dimensional space). More precisely, we shall get an isomorphism between étale neighborhoods of points that preserves the degeneration strata. This is proved by establishing a result on suitable infinitesimal neighborhoods that involves not just the complete flag spaces but also partial ones. In order to have a convenient way of formulating such a result we introduce the following two notions:

By a *stratified space* we shall mean a scheme together with a collection of closed subschemes, called *strata*. A map between stratified spaces is said to be stratified if it maps strata into strata.

If P is a partially ordered set then a *diagram $X_\bullet$ of spaces over P* associates to each element q of P a scheme X_q and to each relation $q > q'$ a map $X_q \to X_{q'}$ fulfilling the condition that the composite $X_q \to X_{q'} \to X_{q''}$ equal the map $X_q \to X_{q''}$ for any $q > q' > q''$. We shall also similarly speak about a diagram of stratified spaces where both the schemes and the maps are assumed to be stratified. Given a field $\mathbf{k}$ and a $\mathbf{k}$-point x of a diagram $X_\bullet$ we may speak of its (strict) *Henselization* at x, which at each $q \in P$ is the Henselization at x of X_q.

For a positive integer g we now consider the partially ordered set P_g whose elements are the subsets of $\{1,2,\dots,g-1\}$ and with ordering that of inclusion. We have two diagrams of stratified spaces over this set: The first, $\mathcal{F}\ell_g^\bullet$, associates to the subset S the flag space of a maximal totally isotropic subspace E of a symplectic $2g$-dimensional vector space and partial flags of subspaces of E whose dimensions form the set S. The map associated to an inclusion $S \subset S'$ is simply the map forgetting some of the elements of the flag. Similarly, we let $\mathcal{F}_g^\bullet$ be the diagram that to a subset S associates the space of flags over the moduli space $\mathcal{A}_g$ of principally polarized abelian varieties that associates to a principally polarized abelian variety the space of flags on its Hodge bundle whose dimensions form the subset S.

The diagram $\mathcal{F}\ell_g^\bullet$ becomes a stratified diagram by considering the stratifications given by the (closed) Schubert cells with respect to some fixed complete flag. In positive characteristic p the diagram $\mathcal{F}_g^\bullet$ becomes a stratified diagram by considering the degeneracy loci given by the relative positions of the Hodge flag $\mathbb{E}_\bullet$ and the conjugate flag $\mathbb{D}$.

8.2 Height 1-Maps

For schemes in a fixed positive characteristic p we shall say that a closed immersion $S \hookrightarrow S'$ defined by the ideal sheaf $\mathcal{I}$ on S' is a *height* 1*-map* if $\mathcal{I}_S^{(p)} = 0$, where for an ideal I, we let $I^{(p)}$ be the ideal generated by the pth powers of elements of I. If R is a local ring in characteristic p with maximal ideal m_R, the *height* 1*-hull* of R is the quotient $R/m_R^{(p)}$. It has the property that its spectrum is the largest closed subscheme of $\operatorname{Spec} R$ for which the map from $\operatorname{Spec} R/m_R$ to $\operatorname{Spec} R/m_R^{(p)}$ is a height 1-map. If k is a field of characteristic p and $x\colon \operatorname{Spec} k \to S$ a k-map to a k-scheme S of characteristic p, then by the *height* 1*-neighborhood* of x we will mean the spectrum of the height 1-hull of the local ring of S at x. It is clear that taking height 1-neighborhoods of k-points is functorial under maps between pointed k-schemes. Finally, we shall say that two local rings are *height* 1*-isomorphic* if their respective height 1-hulls are isomorphic and that the height 1-hull of a k-point is *height* 1*-smooth* if it is isomorphic to the height 1-hull of a smooth k-point (i.e., is of the form $k[[t_1,\dots,t_n]]/m^{(p)}$ with $m=(t_1,\dots,t_n)$).

Theorem 8.1. *For each perfect field k of positive characteristic p and each k-point x of $\mathcal{F}_g^\bullet$ there is a k-point y of $\mathcal{F}\ell_g^\bullet$ such that the height* 1*-neighborhood of x is isomorphic to the height* 1*-neighborhood of y by a stratified isomorphism of diagrams.*

Proof. Denote also by x the point of $\mathcal{F}_g$, the space of complete flags of the Hodge bundle, associated to x as a point of the diagram $\mathcal{F}_g^\bullet$. Let $X^\bullet$ be the height 1-neighborhood of x in $\mathcal{F}_g^\bullet$ and X the height 1-neighborhood of x in $\mathcal{F}_g$. Now the ideal of the closed point of x in X has a divided power structure for which all the divided powers of order $\geq p$ are zero. This allows us to get a trivialization of the restriction of the de Rham cohomology $\mathbb{H}_X \xrightarrow{\sim} X \times W$ that is horizontal (i.e., compatible with the Gauss–Manin connection on the left and the trivial connection on the right). Now, since the absolute Frobenius map on X factors through the closed point, we get that the pullback $\mathbb{E}_\bullet^{(p)}$ is a horizontal flag, and then so is $\mathbb{D}_\bullet$, its elements being either inverse images of horizontal subbundles by the horizontal map V or duals of horizontal subbundles. We now get a map from X to the space $\mathcal{F}\ell_g$ of complete symplectic flags on W such that the pullback of the universal flag equals $\mathbb{E}_\bullet$. We may, furthermore, choose a symplectic isomorphism of W and the standard symplectic space such that $\mathbb{D}_\bullet$ is taken to the fixed complete flag. We can extend this map in a compatible fashion for all partial flag spaces giving a map from the diagram $X^\bullet$ to $\mathcal{F}\ell_g^\bullet$, and we will denote by y the k-point that is the composite of x and this map. This map is clearly a stratified map, and by the infinitesimal Torelli theorem (cf. [FC90, pp. 14–15]) it induces an isomorphism from $X^\bullet$ to $Y^\bullet$, the first height 1-neighborhood of y in $\mathcal{F}\ell_g^\bullet$. □

Theorem 8.2. *For each perfect field k of positive characteristic p and each k-point x of $\mathcal{F}_g$ there is a k-point y of $\mathcal{F}\ell_g$ such that the Henselization of x is isomorphic to the Henselization of y by a stratified isomorphism.*

Proof. The theorem provides such an isomorphism over the height 1-hull X of x. Now, over $\mathcal{O}_{\mathcal{F}_g,x}$ we may extend the trivialization of $\mathbb{H}_X$ to a trivialization of $\mathbb{H}_{\mathcal{F}_g,x}$ that also extends the trivialization of $\mathbb{D}$ (making, of course, no requirements of horizontality). This gives a map from the localization, $\tilde{X}$, of $\mathcal{F}_g$ at x to $\mathcal{F}\ell_g$ that extends the map from X to $\mathcal{F}\ell_g$. It thus induces a map from $\tilde{X}$ to $\tilde{Y}$, the localization $\mathcal{F}\ell_g^\bullet$ at y. Now, this map induces an isomorphism on tangent spaces and $\mathcal{F}_g$ is smooth. This implies that we get an induced isomorphism on Henselization and proves the theorem. □

Lemma 8.3. *Let A be a principally polarized abelian variety over an algebraically closed field. If a flag $\mathbb{D}_\bullet$ for it has type w' that is less than or equal to its canonical type, then w' is the final element corresponding to the canonical type of A.*

Proof. The flag $\mathbb{D}_\bullet$ has the property, since it is of a type $\leq$ to the canonical type, that F maps $\mathbb{D}_i^{(p)}$ into $\mathbb{D}_{\nu_w(i)}$. Consider now the set I of i's for which $\mathbb{D}_i$ is a member of the canonical flag. It clearly contains 0 and is closed under $i \mapsto \bar{\imath}$. Furthermore, if $i \in I$, then $F(\mathbb{D}_i^{(p)})$ has dimension $\nu_w(i)$ but is then equal to $\mathbb{D}_{\nu_w(i)}$, since it is contained in it. Hence I fulfills the conditions of Corollary 2.10 and hence contains the canonical domain, which means that $\mathbb{D}_\bullet$ is a refinement of the canonical flag and thus ν, the final type of A, and $\nu_{w'}$ coincide on the canonical domain of ν and are equal by Corollary 2.10, which means that w' is the final element of the canonical type. □

Corollary 8.4. *(i) Each stratum $\mathcal{U}_w$ of $\mathcal{F}_g$ is smooth of dimension $\ell(w)$ (over $\mathbb{F}_p$).*

(ii) The closed stratum $\overline{\mathcal{U}}_w$ *(again of* $\mathcal{F}_g$*) is Cohen–Macaulay, reduced, and normal of dimension* $\ell(w)$*, and* $\overline{\mathcal{U}}_w$ *is the closure of* $\mathcal{U}_w$ *in* $\mathcal{F}_g$ *for all* $w \in W_g$*.*

(iii) If w *is final then the restriction of the projection* $\mathcal{F}_g \to \mathcal{A}_g$ *to* $\mathcal{U}_w$ *is a finite surjective étale covering from* $\mathcal{U}_w$ *to* $\mathcal{V}_w$ *of degree* $\gamma_g(w)$*.*

Proof. We know that each open Schubert cell of $\mathcal{F}\ell_g$ is smooth, and each closed one is Cohen–Macaulay by a proof that runs completely along the lines of [Ful, Theorem 14.3.]. By a theorem of Chevalley (cf. [Ch94, Corollary of Proposition 3]) they are smooth in codimension 1, so by Serre's criterion (cf. [Gr65, Theorem 5.8.6]) they are normal and reduced. The same statement for the stratification of $\mathcal{F}_g$ then follows from the theorem. To finish (ii), the fact that $\overline{\mathcal{U}}_w = \overline{\mathcal{U}_w}$ follows more or less formally from the rest: if $x \in \overline{\mathcal{U}}_w$, then we know that the dimensions of all $\mathcal{U}_{w'}$ with $w' < w$ that pass through in x are $\ell(w') < \ell(w)$, but the dimension of $\overline{\mathcal{U}}_w$ at that point is $\ell(w)$, and hence x must lie in the closure of $\mathcal{U}_w$.

As for (iii), that $\mathcal{U}_w$ maps into $\mathcal{V}_w$ follows from the fact that the restriction of a final filtration to its canonical domain is a canonical filtration (Proposition 4.5). That the map $\mathcal{U}_w \to \mathcal{A}_g$ is unramified follows from the same statement for Schubert cells, which is [BGG73, Proposition 5.1]. We next prove that $\mathcal{U}_w \to \mathcal{V}_w$ is proper. Note that by Proposition 4.5 and by the fact that by definition, $\mathcal{V}_w$ is the image of $\mathcal{U}_w$, we get that the geometric points of $\mathcal{V}_w$ consist of the principally polarized abelian varieties with a canonical filtration whose canonical type corresponds to the final type of w. Hence for properness we may assume that we have a principally polarized abelian variety over a discrete valuation ring R such that both its generic and special points are of type w and we suppose that we are given a final flag over the generic point. Hence the canonical decomposition of $\operatorname{Spec} R$ for the abelian variety is equal to $\operatorname{Spec} R$, and we have a canonical flag over $\operatorname{Spec} R$. Since $\overline{\mathcal{U}}_w$ is proper, the map to it from the generic point of $\operatorname{Spec} R$ extends to a map from $\operatorname{Spec} R$ to $\overline{\mathcal{U}}_w$, hence giving a flag over its special point. This flag is then of a type $\leq w$, and hence by Lemma 8.3, its type is equal to w and the image of $\operatorname{Spec} R$ lies in $\mathcal{U}_w$, which proves properness.

Now, $\mathcal{V}_w$ being by definition the schematic image of $\mathcal{U}_w$, it is reduced because $\mathcal{U}_w$ is. Since $\mathcal{U}_w \to \mathcal{V}_w$ is unramified, it has reduced geometric fibers, and since it is finite and $\mathcal{V}_w$ is reduced, to show that it is flat it is enough to show that the cardinalities of the geometric fibers are the same for all geometric points of $\mathcal{V}_w$. This, however, is Lemma 4.6. Being finite, flat, and unramified, it is étale. That its degree is $\gamma(w)$ follows from Lemma 4.6. □

Remark 8.5. (i) The corollary is true also for the strata of $\tilde{\mathcal{F}}_g$.

(ii) Note that the degree of the map $\mathcal{U}_w \to \mathcal{V}_w$ is $\gamma_g(w)$. By looking at the proof of Lemma 4.6 it is not difficult to show that it is a covering with structure group a product of linear and unitary groups over finite fields of characteristic p.

9 Punctual flag spaces

Let M be the (contravariant) Dieudonné module of a truncated Barsotti–Tate group of level 1 over an algebraically closed field of characteristic p provided with an alternating perfect pairing (of Dieudonné modules). We let $\mathcal{F}_M$, the *punctual flag space* for M, be the scheme of self-dual admissible complete flags in M for which the middle element equals $\mathrm{Im}(V)$. It is well known that every such M occurs as the Dieudonné module of the kernel of multiplication by p on a principally polarized abelian variety. Then $\mathcal{F}_M$ is the intersection of $\overline{\mathcal{U}}_\emptyset$ and the fiber over a point of $\mathcal{A}_g$ giving rise to M. Also, by a result of Oort [Oo01], the canonical type of M determines it (over an algebraically closed field), and hence we shall also use the notation $\mathcal{F}_\nu$ where ν is a final type. For $\Gamma = (I, \mathcal{S})$ where $I \subseteq \{1, \dots, g\}$ with $\#I$ equal to the semi-simple rank of M and $\mathcal{S}$ a complete V-stable flag of the V-semi-simple part of M, we define $\mathcal{F}_M^\Gamma$ as follows: we let $\mathcal{F}_M^\Gamma$ be the part of $\mathcal{F}_M \cap \mathcal{U}_I^{ss}$ ($\mathcal{U}_I^{ss}$ may clearly be defined directly in terms of M) for which the flag induces $\mathcal{S}$ on the V-semi-simple part. The p-rank of M is denoted by f, and we easily see that $\mathcal{F}_M$ is the disjoint union of the $\mathcal{F}_M^\Gamma$, and putting $\mathcal{F}_M^{\mathcal{S}} := \mathcal{F}_M^{(\{1,\dots,g-f+1\},\Gamma)}$, we have maps $S_I \colon \mathcal{F}_M^I \to \mathcal{F}_M^{\mathcal{S}}$. These maps are homeomorphisms by Proposition 4.14. This can be seen directly by decomposing M as $M^{mul} \bigoplus M^{\ell\ell} \bigoplus M^{et}$, where V is bijective on M^{mul}, F on M^{et}, and F and V nilpotent on $M^{\ell\ell}$. Any element of an admissible flag over a perfect field will decompose in the same way (since that element is stable under F by definition and under V by duality) and is hence determined by its intersection with M^{mul}, $M^{\ell\ell}$, and M^{et}. By self-duality the intersection of all the elements of the flag with M^{et} is determined by that with M^{mul}, and that part is given by an arbitrary full flag of submodules of M^{mul}, which is our $\mathcal{S}$. That means that we may indeed reconstitute the whole flag from Γ and the induced flag on $M^{\ell\ell}$ and that any choice of flag on $M^{\ell\ell}$ gives rise to a flag in $\mathcal{F}_M^\Gamma$. This means that the map $\mathcal{F}_M^\Gamma \to \mathcal{F}_{M^{\ell\ell}}$ is a homeomorphism, and we may for all practical purposes focus our attention on the case that F and V are nilpotent on M (i.e., M is *local–local*). Hence in this section, *unless otherwise mentioned, the Dieudonné modules considered will be local–local.* Note that the principal interest in this section will be focused on the question of which $\mathcal{U}_w$ have nonempty intersection with $\mathcal{F}_M$ and that this problem is indeed by the above considerations immediately reduced to the local–local case.

We shall make extensive use of one way to move in each $\mathcal{F}_M$:

Consider $w_\emptyset \geq w \in W_g$. Assume that we have an index $1 \leq i \leq g-1$ for which $r_w(g+i-1, i+1) \geq i+1$. This means that for a flag $\mathbb{D}_\bullet$ in $\overline{\mathcal{U}}_w$ we have that $F(\mathbb{D}_{i+1}) \subseteq \mathbb{D}_{i-1}$ or equivalently that F is zero on $\mathbb{D}_{i+1}/\mathbb{D}_{i-1}$. Hence if we replace $\mathbb{D}_i$ by any $\mathbb{D}_{i-1} \subset \mathbb{D} \subset \mathbb{D}_{i+1}$ (replacing also $\mathbb{D}_{2g-i}$ to make the flag self-dual), we shall still have an admissible flag, since $V(\mathbb{D}) \subseteq \mathbb{D}_{i-1}$. In order to construct the $\mathbb{E}$-flag, we apply V to the $\mathbb{D}$-flag, which gives us half of the $\mathbb{E}$-flag, and we complement by taking orthogonal spaces. In the $\mathbb{E}$-flag now $\mathbb{E}_{g-i}$ and $\mathbb{E}_{g+i}$ move. This construction gives a mapping from the projective line $\mathbb{P}(\mathbb{E}_{g-i+1}/\mathbb{E}_{g-i-1})$ to $\mathcal{F}_M$, and we shall therefore call this family the *simple family* of index i (with, of course, respect to M), and we shall write $P_{w,i}$ for this simple family. The condition $r_w(g+i-1,i+1) \geq i+1$ is equivalent to $r_w(g-i+1,i-1) = g-i-1$, and when it is fulfilled we shall say that $g-i$ is *movable for* w.

Proposition 9.1. *Any two points of the local flag space $\mathcal{F}_M^\Gamma$ can be connected by a sequence of simple families.*

Proof. We immediately reduce to the case that M is local (in which case the statement is about $\mathcal{F}_M$). We are going to identify $\mathcal{F}_M$ with the scheme of V-stable flags in $\mathrm{Im}(V)$, and we prove the statement for any Dieudonné module N with $F = 0$ and V nilpotent. Let $E_\bullet$ and $F_\bullet$ be two V-stable flags in N. If $E_1 = F_1$ then we may consider N/E_1 and use induction on the length of N to conclude. If not, we use induction on the smallest i such that $F_1 \subseteq E_i$ which we thus may assume to be > 1. We now have $F_1 \subsetneq E_{i-1}$ and hence that F_1 is a complement to E_{i-1} in E_i, so that in particular, $E_i/E_{i-2} = (E_{i-1}/E_{i-2}) \bigoplus (F_1 + E_{i-2})/E_{i-2}$. This has a consequence that V is zero on E_i/E_{i-2}, which means that every subspace of it is stable under V, so that we get a $\mathbb{P}^1$-family of flags in E_i/E_{i-2} in which both E_{i-1}/E_{i-2} and $(F_1 + E_{i-2})/E_{i-2}$ are members, so that we may move E_{i-1} so that it contains F_1. $\square$

Recall (cf. [Oo01, 14.3]) that one defines the partial order relation on final types $\nu_1 \subseteq \nu_2$ (respectively $\nu_1 \subset \nu_2$) by the condition that $\mathcal{V}_{\nu_1} \subseteq \overline{\mathcal{V}}_{\nu_2}$ (respectively $\mathcal{V}_{\nu_1} \subsetneq \overline{\mathcal{V}}_{\nu_2}$). We shall now see that this relation can be expressed in terms of local flag spaces. For this we let M_ν be a Dieudonné module of a principally polarized truncated Barsotti–Tate group of level 1 with final type ν (there is up to isomorphism only one such M_ν, [Oo01, Theorem 9.4]).

Theorem 9.2. *(i) We have that $\nu' \subset \nu$ precisely when there is a $w \in W_g$ such that $w \leq \nu$ and there is a flag of type w in $\mathcal{F}_{M_{\nu'}}$.*

(ii) If there is a flag of type w in $\mathcal{F}_{M_{\nu'}}$, then there is a $w' \leq w$ such that the intersection $\mathcal{U}_{w'} \cap \mathcal{F}_{M_{\nu'}}$ is finite.

Proof. Consider the image in $\mathcal{A}_g$ of $\overline{\mathcal{U}}_\nu$. It is a closed subset containing $\mathcal{V}_\nu$ and hence contains $\overline{\mathcal{V}}_\nu$, and in particular it meets each fiber over a point of $\overline{\mathcal{V}}_\nu$. Consequently there is a point s in the intersection of $\overline{\mathcal{U}}_\nu$ and the fiber over a point t of $\mathcal{V}_{\nu'}$. Now, s lies in some $\mathcal{U}_w \subseteq \overline{\mathcal{U}}_\nu$ and hence fulfills $w \leq \nu$, and since $\nu \leq w_\emptyset$, s also lies in the local flag space of t, and as has been noted, this is the "same" as $\mathcal{F}_{M_{\nu'}}$. The converse is clear.

As for the second part, the proof of Lemma 6.2 shows that a $w' \leq w$ that is minimal for the condition that $\mathcal{U}_{w'} \cap \mathcal{F}_{M_{\nu'}}$ is nonempty has $\mathcal{U}_{w'} \cap \mathcal{F}_{M_{\nu'}}$ finite. $\square$

The theorem allows us to re-prove a result of Oort [Oo01]; the E-O strata are defined in Section 4.4.

Corollary 9.3. *The E-O stratification on $\mathcal{A}_g$ is a stratification.*

Proof. The condition in 9.2, (i) says that $\nu' \subset \nu$ if and only if the closure $\overline{\mathcal{U}}_\nu$ of $\mathcal{U}_\nu$ has a nonempty intersection with the punctual flag space $\mathcal{F}_{\nu'}$. The proof there gives more precisely that a given point s of $\mathcal{V}_{\nu'}$ lies in $\overline{\mathcal{V}}_\nu$ precisely when $\overline{\mathcal{U}}_\nu$ intersects the fiber over s of the map $\mathcal{F}_g \to \mathcal{A}_g$. This condition does not depend on the point s by a result of Oort on Dieudonné modules [Oo01]. □

From Theorem 9.2 it is clear that the condition that $\mathcal{U}_w \cap \mathcal{F}_M \neq \emptyset$ is important. We shall say that an admissible $w \in W_g$ *occurs* in ν, with ν a final type, if $\mathcal{U}_w \cap \mathcal{F}_M \neq \emptyset$, and we shall write it symbolically as $w \to \nu$.

Remark 9.4. It is important to realize that a priori this relation $w \to \nu$ depends on the characteristic, which is implicit in all of this article. Hence the notation $w \xrightarrow{p} \nu$ would be more appropriate. It is our hope that the relation will a posteriori turn out to be independent of p. If not and if one is working with several primes p, the more precise notation will have to be used.

Hence we can formulate the theorem as saying that $\nu' \subset \nu$ precisely when there exists an admissible w with $w \to \nu'$ and $w \leq \nu$. Suppose final types ν and ν' are given. For an element w of minimal length in the set of minimal elements of $\{w \in W_g : \nu > w,\ w \to \nu'\}$ in the Bruhat–Chevalley order, we then have the following property. The space $\mathcal{U}_w \cap \mathcal{F}_{M_{\nu'}}$ has dimension 0 for the generic point of $\mathcal{V}_{\nu'}$. Clearly, then $\ell(w) \geq \ell(\nu')$ for every w as in Theorem 9.2.

Example 9.5. Since E-O strata on $\mathcal{A}_g$ are defined using the projection from the flag space, the closure of an E-O stratum on $\mathcal{A}_g$ need not be given by the Bruhat–Chevalley order on the set of final elements, and indeed it isn't. Oort gave the first counterexample for $g = 7$ based on products of abelian varieties. We reproduce his example and give two others, one for $g = 5$ and one for $g = 6$ that do not come from products.

(i) Let $g = 7$ and let $w_1 = [1, 2, 4, 6, 7, 10, 12]$ and $w_2 = [1, 2, 3, 7, 9, 10, 11]$. Then w_1 and w_2 are final elements of W_7 and have lengths $\ell(w_1) = 8$ and $\ell(w_2) = 9$. In the Bruhat–Chevalley order neither $w_1 \leq w_2$ nor $w_2 \leq w_1$ holds. Despite this, we have $\overline{\mathcal{V}}_{w_1} \subset \overline{\mathcal{V}}_{w_2}$. The explanation for this lies in the fact that the simple family $P_{w_1,4}$ hits the stratum U_{w_3} with w_3 the element $[1, 2, 3, 7, 6, 10, 11] = s_3 w_1 s_4$, with $w_2 > w_3$ and $w_3 \to w_1$, so by Theorem 9.2 it follows that $\overline{\mathcal{V}}_{w_1} \subset \overline{\mathcal{V}}_{w_2}$. (That there is such a simple family can be proved directly, but for now we leave it as an unsupported claim, since a proof "by hand" would be somewhat messy. A more systematic study of these phenomena will appear in a subsequent paper.) This explains the phenomenon observed in [Oo01, p. 406], (but note the misprints there). Also the element $w_2 > w_4 = [1, 2, 3, 7, 9, 5, 11] \to w_1$ will work for w_1. The element w_1 is the final element corresponding to taking the product of a Dieudonné module with final element [135] and a Dieudonné module with final element [1246], whereas similarly, w_2 appears as the "product" of the final elements [135] and [1256]. Since $[1246] < [1256]$, there is a degeneration of a Dieudonné module of type

[1256] to one of type [1246]. This shows that this example simply expresses the fact that $\subset$ must be stable under products, whereas the Bruhat–Chevalley order isn't. (We'd like to thank Ben Moonen for pointing this out to us.)

(ii) For $g = 5$ we consider the final elements $w_1 = [1, 3, 4, 6, 9]$ and $w_2 = [1, 2, 6, 7, 8]$ of lengths 5 and 6 and the nonfinal element $w_3 = [1, 2, 6, 4, 8]$ in W_5. Then $w_3 < w_2$ and $w_3 \to w_1$, so that $\mathcal{V}_{w_1}$ lies in the closure of $\mathcal{V}_{w_2}$. But in the Bruhat–Chevalley order neither $w_1 < w_2$ nor $w_2 < w_1$ holds.

(iii) Let $g = 6$ and consider the final elements $w_1 = [1, 3, 5, 6, 9, 11]$ and $w_2 := [1, 2, 6, 8, 9, 10]$ of lengths $\ell(w_1) = 8$ and $\ell(w_2) = 9$. In the Bruhat–Chevalley order we do not have $w_1 \leq w_2$. Nevertheless, $\mathcal{V}_{w_1}$ occurs in the closure of the E-O stratum $\mathcal{V}_{w_2}$. Indeed, the admissible element $w_3 = [1, 2, 6, 8, 4, 10]$ satisfies $w_2 \geq w_3 \to w_1$: $\mathcal{U}_{w_3}$ has a nonempty intersection with the punctual flag space $\mathcal{F}_{w_1}$. This time, neither of the elements w_1 and w_2 is a product in the sense of (i). Furthermore, since $\mathcal{V}_{w_1}$ is of codimension 1 in $\mathcal{V}_{w_2}$, this example cannot be derived by taking the transitive closure of the closure under products of the Bruhat–Chevalley relation. The claim that we have $w_3 \to w_1$ and the two preceding ones will be substantiated in a subsequent paper.

There is an approach to the study of the relation of the E-O strata and the strata on $\mathcal{F}_g$ that is in some sense "dual" to the study of punctual flag spaces: that of considering the image in $\mathcal{A}_g$ of the $\mathcal{U}_w$. The following result gives a compatibility result on these images and the E-O stratification.

Proposition 9.6. *(i) The image of any $\mathcal{U}_w$, $w \in W_g$, is a union of strata $\mathcal{V}_\nu$. In particular, the image of a $\overline{\mathcal{U}}_w$ is equal to some $\overline{\mathcal{V}}_\nu$.*

(ii) For any final ν and $w \in W_g$, the maps $\mathcal{U}_w \cap \pi^{-1}\mathcal{V}_\nu \to \mathcal{V}_\nu$ and $\overline{\mathcal{U}}_w \cap \pi^{-1}\mathcal{V}_\nu \to \mathcal{V}_\nu$, where π is the projection $\mathcal{F}_g \to \mathcal{A}_g$, have the property that there is a surjective flat map $X \to \mathcal{V}_\nu$ such that the pullback of them to X is isomorphic to the product $X \times (\mathcal{F}_\nu \cap \mathcal{U}_w)$ respectively $X \times (\mathcal{F}_\nu \cap \overline{\mathcal{U}}_w)$.

(iii) A generic point of a component of $\mathcal{U}_w$ maps to the generic point of some $\mathcal{V}_\nu$, and that ν is independent of the chosen component of $\mathcal{U}_w$.

Proof. The first part follows directly from Oort's theorem (in [Oo01]) on the uniqueness of the Dieudonné module in a stratum $\mathcal{V}_\nu$, since it implies that if one fiber of $\pi^{-1}(\mathcal{V}_\nu) \to \mathcal{V}_\nu$ meets $\mathcal{U}_w$, then they all do. As for the second part, it would follow if we could prove that there is a surjective flat map $X \to \mathcal{V}_\nu$ such that the pullback of $(\mathbb{H}, \mathbb{E}, F, V, \langle -, - \rangle)$ is isomorphic to the constant data (provided by the Dieudonné module of type ν). For this we first pass to the space X_ν of bases of $\mathbb{H}$ for which the first g elements form a basis of $\mathbb{E}$, which is flat surjective over $\mathcal{V}_\nu$. Over X_ν, the data is the pullback from a universal situation, where F, V, and $\langle -, - \rangle$ are given by matrices. In this universal situation we have an action of the group G of base changes, and two points over an algebraically closed field give rise to isomorphic $(\mathbb{H}, \mathbb{E}, F, V, \langle -, - \rangle)$ precisely when they are in the same orbit. By assumption (and Oort's theorem) the image of X_ν lies in an orbit, so it is enough to show that the data over an orbit can be made constant by a flat surjective map. However, the map from G to the orbit obtained by letting g act on a fixed point of the orbit has this property.

The third part follows directly from the second. $\square$

The proposition gives us a map $\tau_p\colon W_g \to W_g/S_g$ that to w associates the final type of the open stratum into which each generic point of $\mathcal{U}_w$ maps. We shall return to this map in Section 13.

Example 9.7. Note that the punctual flag space is in general rather easy to understand, since it depends only on the image of V and we are almost talking about the space of flags stable under a nilpotent endomorphism (remember that we have reduced to the local–local case). Almost, but not quite, since the endomorphism is semilinear rather than linear. What is complicated is the induced stratification. Already the case of $\nu = s_3 \in W_3$ is an illustrative example. We have then that $\ker V \cap \operatorname{Im} V$ is of dimension 2; in fact, we have one Jordan block for V on $\operatorname{Im} V$ of size 2 and one of size 1. The first element, $\mathbb{E}_1$, of the flag must lie in $\ker V \cap \operatorname{Im} V$, so we get a $\mathbb{P}^1$ of possibilities for it. If $\mathbb{E}_1 = \operatorname{Im} V^2$, then V is zero on $\mathbb{E}_3/\mathbb{E}_1$ and we can choose $\mathbb{E}_2/\mathbb{E}_1$ as an arbitrary subspace of $\mathbb{E}_3$ giving us a $\mathbb{P}^1$ of choices for $\mathbb{E}_2$. On the other hand, if $\mathbb{E}_1 \neq \operatorname{Im} V^2$, then $\mathbb{E}_3/\mathbb{E}_1$ has a Jordan block of size 2, and hence there is only one V-stable 1-dimensional subspace, and thus the flag is determined by $\mathbb{E}_1$. The conclusion is that the punctual flag space is the union of two $\mathbb{P}^1$'s meeting at a single point. The intersection point is the canonical filtration (which is a full flag), and one can show that the rest of the points on one component are flags of type [241] and the rest of the points on the other component are flags of type [315].

10 Pieri formulas

In this section we are going to apply a theorem of Pittie and Ram [PR99] to obtain a Pieri-type formula for our strata. (It seems to be historically more correct to speak of Pieri–Chevalley-type formulas, cf. [Ch94].) The main application of it will not be to obtain cycle class formulas, since Pieri formulas usually do not give formulas for individual strata but only for certain linear combinations. For us the principal use of these formulas will be that they show that a certain strictly positive linear combination of the boundary components will be a section of an ample line bundle (or close to ample, since one of the contributors to ampleness will be λ_1, which is ample only on the Satake compactification). This will have as consequence affineness for the open strata as well as a connectivity result for the boundary of the closed strata. We shall see in Section 13 that there is also a Pieri formula for the classes of the E-O strata, though we know very little about it.

In this section we are going to work with level structures. There are two reasons for this. The first one is that we are going to exploit the ampleness λ_1, and even formulating the notion of ampleness for a Deligne–Mumford stack is somewhat awkward. The second is that one of the consequences of our considerations will be an irreducibility criterion for strata. Irreducibility for a stratum on $\mathcal{A}_g$ does not imply irreducibility for the same stratum on the space $\mathcal{A}_{g,n}$ of principally polarized abelian varieties with level n-structure, where always $p \nmid n$. In fact, irreducibility for the level-n case means irreducibility on $\mathcal{A}_g$ together with the fact that the monodromy group of the level-n cover is the maximum possible. Hence in this section we shall use $\mathcal{A}_{g,n}$ but also some toroidal compactification $\tilde{\mathcal{A}}_{g,n}$ (cf. [FC90]). Everything we have said so far applies to this situation giving us in particular $\mathcal{F}_{g,n}$ and $\tilde{\mathcal{F}}_{g,n}$, but we have the extra property that for $n \geq 3$, then $\tilde{\mathcal{A}}_{g,n}$ and hence $\tilde{\mathcal{F}}_{g,n}$ are smooth projective varieties.

We now introduce the classes $\ell_i := c_1(\mathbb{E}_{\{i\}})$ for $1 \leq i \leq 2g$ in the Chow ring $\mathrm{CH}^*(\mathcal{F}_g)$. By self-duality of the flag $\mathbb{E}_\bullet$ we have that $\ell_{2g+1-i} = -\ell_i$, and by construction $c_1(\mathbb{D}_{\{i\}}) = p\ell_{i-g} = -p\ell_{3g+1-i}$ for $g+1 \leq i \leq 2g$. (For the notation D_J see Section 3.2.) Furthermore, $\ell_1 + \cdots + \ell_g$ is the pullback from $\tilde{\mathcal{A}}_g$ of λ_1, the first Chern class of the Hodge bundle.

Now we let $M_i := c_1(\mathbb{D}_{]2g-i,2g]})$, $1 \leq i \leq g$, and start by noting that if $n = (n_1, \ldots, n_g)$ then $n \cdot M := \sum_i n_i M_i$ is relatively ample for $\tilde{\mathcal{F}}_g \to \tilde{\mathcal{A}}_g$ if $n_i > 0$ for $1 \leq i < g$. Indeed, by construction $L_i := \ell_{2g} + \cdots + \ell_{2g-i+1}$, $1 \leq i \leq g$, is the pullback from the partial flag space $\tilde{\mathcal{F}}_g[i]$ of flags with elements of rank i, $2g-i$, and g (and with the rank g-component equal to $\mathbb{E}_g$) and on $\tilde{\mathcal{F}}_g[i]$ we have that $\ell_{2g} + \cdots + \ell_{2g-i+1}$ is ample. It is well known that any strictly positive linear combination of these elements is relatively ample. From the formulas above we get that $M_i = p(L_{g-i} + \lambda_1)$ (where we put $L_0 = 0$). On the other hand, λ_1 is almost ample; it is the pullback from $\mathcal{A}_g^*$ of an ample line bundle.

Now we identify the L_i with the fundamental weights of the root system of C_g. Note that W_g acts on the ℓ_i considered as parts of the weight lattice by $\sigma(\ell_i) = \ell_{\sigma(i)}$ (keeping in mind that $\ell_{2g+1-i} = -\ell_i$) and then acts accordingly on the L_i. Let us also note that (by Chevalley's characterization of the Bruhat–Chevalley order) if $w' < w$ with $\ell(w') = \ell(w) - 1$ and if $w = s_{i_1} \cdots s_{i_k}$, then w' is of the form $s_{i_1} \cdots \widehat{s_{i_r}} \cdots s_{i_k}$, which can be rewritten as ws_α, where $s_\alpha = (s_{i_{r+1}} \cdots s_{i_k})^{-1} s_{i_r} (s_{i_{r+1}} \cdots s_{i_k})$, which thus is the reflection with respect to a unique positive root.

Theorem 10.1. *For each $1 \leq i \leq g$ and $w \in W_g$ we have that*

$$(p\lambda_1 + pL_{g-i} - wL_i)[\overline{\mathcal{U}}_w] = \sum_{w' \prec w} c^i_{w,w'} [\overline{\mathcal{U}}_{w'}] \in \mathrm{CH}^1_{\mathbb{Q}}(\overline{\mathcal{U}}_w \bigotimes \overline{\mathbb{F}}_p),$$

where $c^i_{w,w'} \geq 0$ and $w' \prec w$ means $w' \leq w$ and $\ell(w) = \ell(w')+1$. Furthermore, $c^i_{w,w'} > 0$ precisely when $w' = ws_\alpha$ for α a positive root for which the simple root α_i appears with a strictly positive coefficient when α is written as a linear combination of the simple roots.

Proof. We shall use [PR99], which has the following setup: We fix a semi-simple algebraic group G (which in our case is the symplectic group Sp_{2g}, but using this in the notation will only confuse) with Borel group B and fix a principal B-bundle $E \to X$ over an algebraic variety X. Letting $E(G/B) \to X$ be the associated G/B-bundle, we have, because its structure group is B and not just G, Schubert varieties $\Omega_w \to X$ (which fiber by fiber are the usual Schubert varieties). For every weight $\lambda \in P$, P being the group of weights for G, we have two line bundles on $E(G/B)$; on the one hand, y^λ, obtained by regarding λ as a character of B, which gives a G-equivariant line bundle on G/B and hence a line bundle on $E(G/B)$, on the other hand, the character λ can also be used to construct, with the aid of the principal B-bundle E, a line bundle x^λ on X and then by pullback to $E(G/B)$ a line bundle also denoted by x^λ. A result of [PR99, Corollary] then says that if λ is a dominant weight then

$$y^\lambda[\mathcal{O}_{\Omega_w}] = \sum_{\eta \in \mathcal{T}_w^\lambda} x^{\eta(1)}[\mathcal{O}_{\Omega_{v(\eta,w)}}] \in K_0(E(G/B)). \tag{1}$$

Here $\mathcal{T}_w^\lambda$ is a certain set of piecewise linear paths $\eta\colon [0,1] \to P \bigotimes \mathbb{R}$ in the real vector space spanned by P; moreover, $v(\eta, w)$ is a certain element in the Weyl group of G that is always $\leq w$, and $\mathcal{T}_w^\lambda$ has the property that $\eta(1) \in P$ for all its elements η. An important property of $\mathcal{T}_w^\lambda$ is that it depends only on w and λ and not on E. It follows immediately from the description of [PR99] that $v(\eta, w) = w$ in only one case, namely when η is the straight line $\eta(t) = tw\lambda$. Hence we can rewrite the formula as

$$(y^\lambda - x^{w\lambda})[\mathcal{O}_{\Omega_w}] = {\sum_{\eta \in \mathcal{T}_w^\lambda}}' x^{\eta(1)}[\mathcal{O}_{\Omega_{v(\eta,w)}}],$$

where the sum now runs over all elements of $\mathcal{T}_w^\lambda$ for which $v(\eta, w) < w$. Taking Chern characters and looking at the top term that appears in codimension $\mathrm{codim}(w) + 1$, we get

$$(c_1(y^\lambda) - c_1(x^{w\lambda}))[\Omega_w] = {\sum_{\eta \in \mathcal{T}_w^\lambda}}'' [\Omega_{v(\eta,w)}], \tag{2}$$

where the sum is now over the elements of $\mathcal{T}_w^\lambda$ for which $\ell(v(\eta, w)) = \ell(w) - 1$. To determine the multiplicity with which a given $[\Omega_{w'}]$ appears in the right-hand side we could no doubt use the definition of $\mathcal{T}_w^\lambda$. However, it seems easier to note that the multiplicity is independent of E, and hence we may assume that X is a point and by additivity in λ that λ is a fundamental weight λ_i. In that case one can use a result of Chevalley [Ch94, Proposition 10] to get the description of the theorem. However, we want this formula to be true not in the Chow group of $E(G/B)$ but instead in the Chow group of the relative Schubert subvariety of index w of $E(G/B)$. This, however, is no problem, since the (relative) cell decomposition shows that this Chow group injects into the Chow group of $E(G/B)$.

We now would like to claim Formula 2 in the case $X = \mathcal{F}_g$ and the principal B-bundle is the tautological bundle $\mathbb{E}_\bullet$. The deduction of (2) from (1) is purely formal; however, [PR99] claims (1) only for X a smooth variety over $\mathbb{C}$. This at least allows us to conclude (2) for $X = \mathcal{F}_{g,n} \bigotimes \mathbb{C}$, the flag space associated to a smooth toroidal compactification of $\mathcal{A}_{g,n}$, the moduli space of principally polarized abelian varieties with a principal level n-structure for $n \geq 3$ (see [FC90, Theorem 6.7, Corollary 6.9]). Using the specialization map for the Chow group, we conclude that (2) is valid for $X = \mathcal{F}_{g,n} \bigotimes \overline{\mathbb{F}}_p$. Finally, pushing down under the map $\mathcal{F}_{g,n} \to \mathcal{F}_g$ induced from the map $\overline{\mathcal{A}}_{g,n} \to \overline{\mathcal{A}}_g$ we get (2) for $X = \mathcal{F}_g$.

The final step is to pull back (2) along the section of $E(G/B)$ given by $\mathbb{D}_\bullet$. To make the pullback possible (note that the relative Schubert variety will in general not be smooth over the base), we remove the relative Schubert varieties of codimension 2 in the relative Schubert variety in question. This forces us to remove the part of $\overline{\mathcal{U}}_w$ where the section encounters the removed locus. This is, however, a codimension-2 subset by Corollary 8.4, so its removal will not affect $\mathrm{CH}^1_{\mathbb{Q}}(\overline{\mathcal{U}}_w)$. Unraveling the pullback of $(c_1(y^\lambda) - c_1(x^{w\lambda}))$ gives the theorem. □

To apply the theorem we start with some preliminary results that will be used to exploit the positivity of the involved line bundles.

Lemma 10.2. *Let X be a proper (irreducible) variety of dimension > 1 and $\mathcal{L}$ a line bundle on X that is ample on some open subset $U \subseteq X$. Let $D := X \setminus U$ and let $H \subset U$ be the zero set of a section of $\mathcal{L}_{|U}$. If D is connected then so is $D \cup H$.*

Proof. By replacing the section by a power of it, we may assume that $\mathcal{L}$ is very ample, giving an embedding $U \hookrightarrow \mathbb{P}^n$. Let Z be the closure of the graph of this map in $X \times \mathbb{P}^n$; moreover, let Y be the image of Z under the projection on the second factor giving us two surjective maps $X \leftarrow Z \to Y$ and let D' be the inverse image in Z of D. Assume that $D \cup H$ is the disjoint union of the nonempty closed subsets A and B and let A' and B' be their inverse images in Z. Now, Y is irreducible of dimension > 1 and hence H'' is connected, where H'' is the hyperplane section of Y corresponding to H, so that the images of A' and B' in Y must meet. However, outside of D' the map $Z \to Y$ is a bijection, and hence the meeting point must lie below a point of D' and hence A' and B' both meet D'. This implies that A and B both meet D, which is a contradiction, since D is assumed to be connected. □

Proposition 10.3. *Let $\mathcal{L}$ be the determinant $\det \mathbb{E}$ of the Hodge bundle over $\mathcal{A}_{g,n}$, $n \geq 3$ (and prime to p).*

(i) There is, for each $1 \leq i < g$, an integer m_i such that the global sections of $\Lambda^{g-i}(\mathbb{H}/\mathbb{E}) \bigotimes \mathcal{L}^{\otimes m}$ generate this bundle over $\mathcal{A}_{g,n}$ whenever $m \geq m_i$. These m_i can be chosen independently of p (but depending on g and n).

(ii) Putting $N_i := L_i + n_i\lambda_1$ for $1 \leq i < g$ and $N_g := \lambda_1$, then $\sum_i m_i N_i$ is ample on $\mathcal{F}_g$ if $m_i > 0$ for all $1 \leq i \leq g$.

(iii) Fix $w \in W_g$ and put $L := \sum_{i<g} L_i$, $N := \sum_{i<g} N_i$, and $m = \sum_i m_i$. Choose r, s, t, and u such that $rN + t\lambda_1 - wL$ respectively $sN + u\lambda - wL_g$ can be written as a positive linear combination of the N_i (using that $L_g = -\lambda_1$). Then if $p > r + sm$ and $(g-1)p > t + um$ we have that $p(L + (g-1)\lambda_1) + pm\lambda_1 - wL - mwL_g$ is ample on $\mathcal{F}_{g,n}$. The constants r, s, t, and u can be chosen independently of p.

Proof. Statement (i) follows directly from the fact that λ_1 is ample on $\mathcal{A}_{g,n}$. The independence of p follows from the existence of a model of $\mathcal{A}_{g,n}$ that exists over $\operatorname{Spec}\mathbb{Z}[1/n, \zeta_n]$.

As for (ii), we have that $\pi_*\mathcal{O}(L_i) = \Lambda^{g-i}(\mathbb{H}/\mathbb{E})$, $1 \leq i < g$, since $\pi\colon \mathcal{F}_{g,n} \to \mathcal{A}_{g,n}$ can be identified with the space of flags on $\mathbb{H}/\mathbb{E}$, and $\mathcal{O}(L_i)$ is $\det(\mathbb{H}/\mathbb{E}_{2g-i})$. By definition we then have that $\pi_*\mathcal{O}(N_i)$ is generated by global sections on $\mathcal{A}_{g,n}$. We know that on the flag space SL_g/B we have that the canonical ring $\bigoplus_\lambda H^0(\operatorname{SL}_g/B, \mathcal{L}_\lambda)$, where λ runs over the dominant weights and $\mathcal{L}_\lambda$ is the corresponding line bundle, is generated by the $H^0(\operatorname{SL}_g/B, \mathcal{L}_{\lambda_i})$, $1 \leq i < g$, where λ_i is the ith fundamental weight (see for instance [RR85]). Also $\mathcal{O}(\sum_{i<g} m_i N_i)$ is relatively very ample, and we have just shown that $\pi_*\mathcal{O}(\sum_{i<g} m_i N_i)$ is generated by global sections. Since λ_1 is ample on $\mathcal{A}_{g,n}$, we get that $\pi_*\mathcal{O}(\sum_{i\leq g} m_i N_i)$ is ample.

Continuing with (iii), we have that $N = L + n\lambda_1$, which gives $p(L + (g-1)\lambda_1) + pm\lambda_1 - wL - mwL_g = (p - (r+sm))N + (p(g-1) - (t+um)) + (rN + t\lambda_1 - wL) + m(sN + u\lambda_1 - wL_g)$. We then conclude by the definitions of r, s, t, and u and (ii). ☐

Remark 10.4. (i) The constants r, s, t, and u are quite small and easy to compute. We know nothing about the m_i but imagine that they would not be too large.

(ii) It would seem that the last part would not be applicable for $g = 1$, but it can be easily modified to do so. On the other hand, for $g = 1$ everything is trivial anyway.

We are now ready for the first application of the Pieri formula.

Proposition 10.5. *(i) There is a bound depending only on g and n such that if p is larger than that bound, then for an irreducible component Z of some $\mathcal{U}_w \subseteq \tilde{\mathcal{F}}_{g,n}$, $w \in W_g$, the union of the complement of Z in $\overline{Z}$, the closure of Z in $\tilde{\mathcal{F}}_{g,n}$, and the intersection of $\overline{Z}$ and $\tilde{\mathcal{F}}_{g,n} \setminus \mathcal{F}_{g,n}$ is connected if the intersection of $\overline{Z}$ with the boundary $\tilde{\mathcal{F}}_{g,n} \setminus \mathcal{F}_{g,n}$ is connected or empty.*

(ii) There is a bound depending only on g and n such that if p is larger than that bound, then for $w \in W_g$ of semi-simple rank 0 we have that $\mathcal{U}_w$ is affine.

Proof. By Proposition 10.3 there is a bound depending only on g and n such that if p is larger than it, then $M := p(L + (g-1)\lambda_1) + pm\lambda_1 - wL - mwL_g$ is ample on $\mathcal{F}_{g,n}$. Summing up Pieri's formula (Theorem 10.1) for $1 \leq i < g$ and

m times the formula for $i = g$ we get that $M[\overline{Z}]$ is supported on $\overline{Z}$ intersected with smaller strata. (Note that the Pieri formula is a priori—and quite likely in reality—true only modulo torsion. We may, however, simply multiply it by a highly divisible integer, and that doesn't change the support.) We then conclude by Lemma 10.2.

As for the second part, we argue as in the first part and conclude that $\overline{\mathcal{U}}_w \setminus \mathcal{U}_w$ is the support of an ample divisor in $\overline{\mathcal{U}}_w$ (as in the theorem, each component must appear, since we are summing up for $1 \leq i \leq g$, and some α_i must appear in the expansion of α) and hence $\mathcal{U}_w$ is affine. □

Remark 10.6. (i) The first part of the proposition is somewhat difficult to use because of the condition on the intersection with the boundary. In the applications of the next section it turns out that we need to apply it only when the intersection is empty.

(ii) For the second part we would like to say more generally that the image of $\mathcal{U}_w$ is affine in $\mathcal{F}^*_{g,n}$ for some appropriate definition of $\mathcal{F}^*_{g,n}$ analogous to the Satake compactification. The problem is that it doesn't seem as if some power of M would be generated by its global sections, so that we cannot define $\mathcal{F}^*_{g,n}$ as the image of $\tilde{\mathcal{F}}_{g,n}$.

11 Irreducibility properties

In this section we shall prove irreducibility of a large class of strata and also that if the characteristic is large enough and our irreducibility criterion is not fulfilled, then (with some extra conditions on the stratum) the stratum is reducible. Our proofs show two advantages of working on the flag spaces. The major one is that our strata are normal, so that irreducibility follows from connectedness. The connectedness of the closed E-O strata except $\overline{\mathcal{V}}_1$ is proven in [Oo01], but the $\overline{\mathcal{V}}_w$ are most definitely not locally connected and hence that does not say very much about the irreducibility. In the converse direction we also make use of the Pieri formula.

Definition-Lemma 11.1. *Let $\{Z_\alpha\}$ be a* stratification *of a Deligne–Mumford stack X of finite type over a field, by which we mean that the strata Z_α are locally closed reduced substacks of X such that the closure $\overline{Z_\alpha}$ of a stratum is the union of strata. By the k*-skeleton *of the stratification we mean the union of the strata of dimension $\leq k$ (which is a closed substack). The* boundary *of a stratum Z_α is the complement of Z_α in its closure. Assume furthermore that each Z_α is irreducible and that for Z_α of dimension strictly greater than some fixed N we have that its boundary is connected (and in particular nonempty). Then the intersection of a connected union Z of closed strata $\overline{Z_\alpha}$ with the N-skeleton is connected.*

Proof. It is enough by induction to prove that if we remove a stratum Z' from Z, whose dimension is maximal and $> N$, then the result remains connected.

Assume that $Z \setminus Z'$ is the disjoint union of two closed subschemes Z_1 and Z_2. By assumption, the boundary of Z' is connected and hence lies in Z_1, say. This means that $Z' \cup Z_1$ is closed and disjoint from Z_2, which by the connectedness of Z implies that Z_2 is empty. □

Proposition 11.2. *There is a bound depending only on g and n such that the following is true if p is larger than that bound:*

Let $X \subseteq \tilde{\mathcal{F}}_{g,n}$ be a connected union of irreducible components of closed strata $\overline{\mathcal{U}}_w$ (for possibly different w) that lie inside of $\mathcal{F}_{g,n}$. Then the intersection of X with the 1-skeleton of the stratification is connected.

Proof. This follows directly from Proposition 10.5 and Lemma 11.1. (Note that for level 1 or 2 we may pass to a higher level in order to apply the proposition.) □

We now want to interpret this proposition (and its converse, which will be true for any p) in arithmetical terms. Hence we define the 1-*skeleton graph* of level n as the following edge-colored graph: Its vertices are the points of $\overline{\mathcal{U}}_1 \subset \mathcal{F}_{g,n}$, i.e., isomorphism classes of principally polarized superspecial g-dimensional abelian varieties A together with a level-n structure and a complete flag $0 = \mathbb{D}_0 \subset \mathbb{D}_1 \subset \cdots \subset \mathbb{D}_g = H^0(A, \Omega^1_A)$ on $H^0(A, \Omega^1_A)$ for which $\mathbb{D}^{\perp}_{g-i} = V^{-1}\mathbb{D}^{(p)}_i$. For each $1 \leq i \leq g$ we connect two vertices by an edge of color i if there is an irreducible component of $\overline{\mathcal{U}}_{s_i}$ that contains them.

Lemma 11.3. *If $S \subseteq \{1, \dots, g\}$ has the property that it contains g and for every $1 \leq i < g$ we have that either i or $g - i$ belongs to S, then the subgraph of the 1-skeleton graph consisting of all vertices and all edges of colors $i \in S$ is connected.*

Proof. This follows from [Oo01, Proposition 7.3] and Theorem 7.5. □

For a subset $S \subseteq \{1, \dots, g\}$ the S-*subgraph* of the 1-skeleton graph is the subgraph with the same vertices and with only the edges whose color is in S. This definition allows us to formulate our irreducibility conditions.

Theorem 11.4. *(i) Let $w \in W_g$ and let $S := \{1 \leq i \leq g : s_i \leq w\}$. If the S-subgraph of the 1-skeleton graph is connected, then $\overline{\mathcal{U}}_w \subseteq \mathcal{A}_g$ is irreducible.*

(ii) There is a bound depending only on g such that if p is larger than that bound the following is true: if $w \in W_g$ is admissible and either final or of semi-simple rank 0 and if $S := \{1 \leq i \leq g : s_i \leq w\}$, then there is a bijection between the irreducible components of $\overline{\mathcal{U}}_w \subseteq \mathcal{A}_g$ and the connected components of the 1-skeleton graph.

Proof. The first part is clear, since Proposition 6.1 says that each connected component meets $\overline{\mathcal{U}}_1$, and then by the assumption on connectedness of the S-subgraph $\overline{\mathcal{U}}_w$ is connected; but by Corollary 8.4 it is normal and hence is irreducible.

As for the second part, assume first that w is final but of positive semi-simple rank. This means that its Young diagram does not contain a row of length g, and hence by Lemma 2.7 and the Chevalley characterization of the Bruhat–Chevalley order we have that $s_i \leq w$ for all $1 \leq i \leq g$ and hence the S-subgraph is connected by Lemma 11.3, which makes the statement trivially true. We may therefore assume that the semi-simple rank is 0 and hence that $\overline{\mathcal{U}}_w$ lies entirely in $\mathcal{F}_{g,n}$. In that case the result follows from Proposition 11.2 and the fact that two irreducible components of two $\overline{\mathcal{U}}_{s_i}$ meet only at $\overline{\mathcal{U}}_1$. □

Projecting down to $\mathcal{A}_g$ we get the following corollary, which shows irreducibility for many E-O strata.

Theorem 11.5. *Let $w \in W_g$ be a final element whose Young diagram Y has the property that there is a $\lceil (g+1)/2 \rceil \leq i \leq g$ such that Y does not contain a row of length i. Then $\overline{\mathcal{V}}_w$ is irreducible and the total space of the étale cover $\mathcal{U}_w \to \mathcal{V}_w$ is connected.*

Proof. This follows from Theorem 11.4 and Lemmas 2.7 and 11.3. □

Example 11.6. For $g = 2$ the locus of abelian surfaces of p-rank ≤ 1 is irreducible. For $g = 3$ all E-O strata except the superspecial locus ($Y = \{1, 2, 3\}$) and the Moret–Bailly locus ($Y = \{2, 3\}$) are irreducible.

In [Ha07], S. Harashita has proved that the number of irreducible components of an E-O stratum that is contained in the supersingular locus is given as a class number and as a consequence that, except possibly for small p, these strata are reducible. As has been proved by Oort (cf. [Ha07, Proposition 5.2]) these strata are exactly the ones to which Theorem 11.5 does not apply.

We shall finish this section by showing that the 1-skeleton graph can be described in purely arithmetic terms very strongly reminiscent of the results of Harashita. Note that even for final elements our results are not formally equivalent to Harashita's, since we are dealing with the set of components of the final strata in $\mathcal{F}_g$, whereas Harashita is dealing with their images in $\mathcal{A}_g$. In any case, our counting of the number of components uses Theorem 11.4 and hence is valid only for sufficiently large p, whereas Harashita's are true unconditionally.

We start by giving a well known description of the vertices of the 1-skeleton graph (see for instance [Ek87]) valid when $g > 1$. We fix a supersingular elliptic curve E and its endomorphism ring $\mathbf{D}$ that is provided with the Rosati involution $*$. To simplify life we assume, as we may, that E is defined over $\mathbb{F}_p$ and hence $\mathbf{D}$ contains the Frobenius map F. It has the property that $\mathbf{D}F = F\mathbf{D} = \mathbf{D}F\mathbf{D}$, the unique maximal ideal containing p. Furthermore, we have that $\mathbf{D}/\mathbf{D}F \cong \mathbb{F}_{p^2}$. There is then a bijection between isomorphism classes of $\mathbf{D}$-*lattices* M (i.e., right modules torsion-free and finitely generated as abelian groups) of rank g (i.e., of rank $4g$ as abelian groups) and isomorphism classes of g-dimensional abelian varieties A. The correspondence associates to the abelian variety A the $\mathbf{D}$-module $\mathrm{Hom}(E, A)$. Polarizations on A then correspond to positive definite *unitary forms*, i.e., a biadditive map $\langle -,- \rangle\colon M \times M \to \mathbf{D}$ such that $\langle md, n\rangle = \langle m,n\rangle dn$, $\langle n,m\rangle = (\langle m,n\rangle)^*$, and $m \neq 0 \Rightarrow \langle m,m\rangle > 0$. The polarization is principal precisely when the form is *perfect*, i.e., the induced map of right $\mathbf{D}$-modules $M \to \mathrm{Hom}_{\mathbf{D}}(M,\mathbf{D})$ given $n \mapsto (m \mapsto \langle m,n\rangle)$ is an isomorphism. In general we put $M^* := \mathrm{Hom}_{\mathbf{D}}(M,\mathbf{D})$, and then the form induces an embedding $M \to M^*$, which makes the image of finite index. More precisely, on $M \bigotimes \mathbb{Q}$ we get a nondegenerate pairing with values in $\mathbf{D} \bigotimes \mathbb{Q}$ and then we may identify M^* with the set $\{n \in M \bigotimes \mathbb{Q} : \forall m \in M\colon \langle m,n\rangle \in \mathbf{D}\}$. Using this we get a $\mathbf{D} \bigotimes \mathbb{Q}/\mathbb{Z}$-valued unitary perfect form on M^*/M given by $\langle \overline{m}, \overline{n}\rangle := \langle m,n\rangle \bmod \mathbf{D}$. As usual, superlattices $M \subseteq N$ over $\mathbf{D}$ correspond to totally isotropic submodules of M^*/M.

If now $S \subseteq \{0,\dots,g\}$ is stable under $i \mapsto g-i$ then an *arithmetic S-flag* consists of the choice of unitary forms on $\mathbf{D}$-modules M_i of rank g for $i \in S$ and compatible isometric embeddings $M_i \hookrightarrow M_j$ whenever $i < j$ fulfills the following conditions:

- For all $i \in S$ with $i \geq g/2$ we have that M_i^*/M_i is killed by F of $\mathbf{D}$ and can hence be considered as a $\mathbf{D}/m = \mathbb{F}_{p^2}$-vector space with a perfect unitary form.
- We have that $FM_i^* = M_{g-i}$ for all $i \in S$.
- The length of M_j/M_i for $i < j$ is equal to $j-i$.

Remark 11.7. (i) Note that we allow S to be empty, in which case there is exactly one S-flag.

(ii) As follows (implicitly) from the proof of the next proposition, the isomorphism class of an element of an arithmetic S-flag tensored with $\mathbb{Q}$ is independent of the flag. Hence we may consider only lattices in a fixed unitary form over $\mathbf{D} \bigotimes \mathbb{Q}$ and then think of M_i as a sublattice of M_j.

Proposition 11.8. *Let $S \subseteq \{1,\dots,g\}$ and let $\overline{S} \subseteq \{0,\dots,g\}$ be the set of integers of the form i or $g-i$ for $i \in S$. Then the set of isomorphism classes of $\overline{S}$-flags is in bijection with the set of connected components of the S-subgraph of the 1-skeleton graph.*

Proof. This follows from the discussion above and Theorem 7.5 once we have proven that an S-flag can be extended to a $\{0,\dots,g\}$-flag. Assume first that $g \notin S$ and let $i \in S$ be the largest element in S. By assumption we have $FM_i^* = M_{g-i} \subseteq M_i \subset M_i^*$ and we have that the length of M_i^*/FM_i^* is g, whereas again by assumption, that of M_i/M_{g-i} is $2i-g$. Together this gives that the length of M_i^*/M_i is $g-(2i-g) = 2(g-i)$. Since the form on M_i^*/M_i is a nondegenerate unitary $\mathbb{F}_{p^2}$-form and since all such forms are equivalent, we get that there is a $(g-i)$-dimensional totally isotropic (and hence its own orthogonal) subspace of M_i^*/M_i; this then gives an $M_i \subset M_g \subset M_i^*$ and since M_g/M_i is its own orthogonal, we get that the pairing on M_g is perfect. We then put $M_0 := FM_g$, and the rest of the flag extension is immediate. $\square$

12 The Cycle Classes

If one wishes to exploit our stratification on $\mathcal{F}_g$ and the E-O stratification on $\mathcal{A}_g$ fully, one needs to know the cohomology classes (or Chow classes) of the (closed) strata. In this section we show how to calculate these classes.

The original idea for the determination of the cycle classes can be illustrated well by the p-rank strata. If X is a principally polarized abelian variety of dimension g that is general in the sense that its p-rank is g, then its kernel of multiplication by p contains a direct sum of g copies of μ_p, the multiplicative group scheme of order p. The unit tangent vector of μ_p gives a tangent vector to X at the origin. By doing this in the universal family we thus see that on a suitable level cover of the moduli we have g sections of the Hodge bundle over the open part of ordinary abelian varieties. If the abelian variety loses p-rank under specialization, the g sections thus obtained become dependent and the loci where this happens have classes represented by a multiple of the Chern classes of the Hodge bundle.

To calculate the cycle classes of the E-O strata on $\mathcal{A}_g \otimes \mathbb{F}_p$ we shall use the theory of degeneration cycles of maps between vector bundles. To this end we shall apply formulas of Fulton for degeneracy loci of symplectic bundle maps to calculate the classes of the $\overline{\mathcal{U}}_w$ and formulas of Pragacz and Ratajski and of Kresch and Tamvakis for calculating those of $\overline{\mathcal{V}}_\nu$.

12.1 Fulton's formulas

Over the flag space $\mathcal{F}_g$ we have the pullback of the de Rham bundle and the two flags $\mathbb{E}_\bullet$ and $\mathbb{D}_\bullet$ on it. We denote by ℓ_i the roots of the Chern classes of $\mathbb{E}$ so that $c_1(\mathbb{E}_i) = \ell_1 + \ell_2 + \cdots + \ell_i$. We then have $c_1(\mathbb{D}_{g+i}) - c_1(\mathbb{D}_{g+i-1}) = p\ell_i$.

Recall (cf. Section 4.1 and Section 5) that for each element $w \in W_g$ we have degeneracy loci $\overline{\mathcal{U}}_w$ in $\mathcal{F}_g$ respectively $\tilde{\mathcal{F}}_g$. Their codimensions equal the length $\ell(w)$, and it thus makes sense to consider the cycle class $u_w = [\overline{\mathcal{U}}_w]$ in $\mathrm{CH}_{\mathbb{Q}}^{\mathrm{codim(w)}}(\tilde{\mathcal{F}}_g)$, where we write $\mathcal{F}_g$ instead of $\mathcal{F}_g \otimes \mathbb{F}_p$.

Fulton's setup in [Fu96] is the following (or more precisely the part that interests us): We have a symplectic vector bundle H over some scheme X and two full symplectic flags $0 \subset \cdots \subset E_2 \subset E_1 = H$ and $0 \subset \cdots \subset D_2 \subset D_1 = H$. For each $w \in W_g$ one defines the degeneracy locus $\overline{\mathcal{U}}_w$ by $\{x \in X : \forall i,j\colon \dim(\mathbb{E}_{i,x} \cap \mathbb{D}_{j,x}) \leq r_w(i,j)\}$ (of course this closed subset is given a scheme structure by considering these conditions as rank conditions for maps of vector bundles). Fulton then defines a polynomial in two sets of variables x_i and y_j, $i,j = 1,\ldots,g$, such that if this polynomial is evaluated as $x_i = c_1(E_i/E_{i+1})$ and $y_j = c_1(D_j/D_{j+1})$, then it gives the class of $\overline{\mathcal{U}}_w$ provided that $\overline{\mathcal{U}}_w$ has the expected codimension $\mathrm{codim}(w)$ (and X is Cohen–Macaulay). The precise definition of these polynomials is as follows: For a partition $\mu = \{\mu_1 > \mu_2 > \cdots > \mu_r > 0\}$ with $r \leq g$ and $\mu_1 \leq g$ one defines a Schur function

$$\Delta_\mu(x) := \det(x_{\mu_i+j-i})_{1 \leq i,j \leq r}$$

in the variables x_i and puts

$$\Delta(x,y) := \Delta_{(g,g-1,\ldots,1)}(\sigma_i(x_1,\ldots,x_g)+\sigma_i(y_1,\ldots,y_g)),$$

where σ_i is the ith elementary symmetric function. One then considers the "divided difference operators" ∂_i on the polynomial ring $\mathbb{Z}[x_1,\ldots,x_g]$ by

$$\partial_i(F(x)) = \begin{cases} \frac{F(x)-F(s_i x)}{x_i - x_{i+1}} & \text{if } i<g, \\ \frac{F(x)-F(s_g' x)}{2x_g} & \text{if } i=g, \end{cases}$$

where s_i interchanges x_i and x_{i+1} for $i=1,\ldots,g-1$, but s_g' sends x_g to $-x_g$ and leaves the other x_i unchanged. We write an element $w \in W_g$ as a product $w = s_{i_\ell} s_{i_{\ell-1}} \cdots s_{i_1}$ with $\ell = \ell(w)$ and set

$$P_w := \partial_{i_\ell} \cdots \partial_{i_1}(\prod_{i+j\leq g}(x_i - y_j)\cdot \Delta). \tag{4}$$

An application of Fulton's formulas gives the following.

Theorem 12.1. *Let $w = s_{i_\ell} s_{i_{\ell-1}} \cdots s_{i_1}$ with $\ell = \ell(w)$ be an element of the Weyl group W_g. Then the cycle class $u_w := [\overline{\mathcal{U}}_w]$ in $\mathrm{CH}_{\mathbb{Q}}^{\mathrm{codim}(w)}(\tilde{\mathcal{F}}_g)$ is given by*

$$u_w = \partial_{i_1}\cdots\partial_{i_\ell}\left(\prod_{i+j\leq g}(x_i - y_j)\cdot\Delta(x,y)\right)_{|x_i=-\ell_i, y_j = p\ell_j}.$$

Proof. By construction $\overline{\mathcal{U}}_w$ is the degeneracy locus of the flags $\mathbb{E}_\bullet$ and $\mathbb{D}_\bullet$. By Corollary 8.4 they are Cohen–Macaulay and have the expected dimension, and hence the degeneracy cycle class is equal to the class of $\overline{\mathcal{U}}_w$. □

For a final element $w \in W_g$ the map $\overline{\mathcal{U}}_w \to \overline{\mathcal{V}}_w$ is generically finite of degree $\gamma_g(w)$. By applying the Gysin map to the formula of Theorem 12.1 using Formula 3.1 we can in principle calculate the cohomology classes of all the pushdowns of final strata, hence of the E-O strata.

Example 12.2. $g=2$.

The Weyl group W_2 consists of eight elements; we give the cycle classes in $\tilde{\mathcal{F}}_2$ and the pushdowns on $\tilde{\mathcal{A}}_2$:

w	s	ℓ	$[\overline{\mathcal{U}}_w]$	$\pi_*([\overline{\mathcal{U}}_w])$
$[4,3]$	$s_1s_2s_1s_2$	4	1	0
$[4,2]$	$s_1s_2s_1$	3	$(p-1)\lambda_1$	0
$[3,4]$	$s_2s_1s_2$	3	$-\ell_1+p\ell_2$	$1+p$
$[2,4]$	s_1s_2	2	$(1-p)\ell_1^2+(p^2-p)\lambda_2$	$(p-1)\lambda_1$
$[3,1]$	s_2s_1	2	$(1-p^2)\ell_1^2+(1-p)\ell_1\ell_2+(1-p)\ell_2^2$	$p(p-1)\lambda_1$
$[2,1]$	s_1	1	$(p-1)(p^2+1)\lambda_1\lambda_2$	0
$[1,3]$	s_2	1	$(p^2-1)\ell_1^2(\ell_1-p\ell_2)$	$(p-1)(p^2-1)\lambda_2$
$[1,2]$	1	0	$-(p^4-1)\lambda_1\lambda_2\ell_1$	$(p^4-1)\lambda_1\lambda_2$

In the flag space $\mathcal{F}_2$ the stratum corresponding to the empty diagram is $U_{s_2s_1s_2}$ and the strata contained in its closure are the four final ones $U_{s_2s_1s_2}$, $U_{s_1s_2}$, U_{s_2}, and U_1 and the two nonfinal ones U_{s_1} and $U_{s_2s_1}$. The Bruhat–Chevalley order on these is given by the following diagram:

$$\begin{tikzcd}
& s_2s_1s_2 \arrow[dl] \arrow[dr] & \\
s_2s_1 \arrow[d] \arrow[drr] & & s_1s_2 \arrow[d] \arrow[dll] \\
s_1 \arrow[dr] & & s_2 \arrow[dl] \\
& 1 &
\end{tikzcd}$$

The four final strata $U_{s_2s_1s_2}$, $U_{s_1s_2}$, U_{s_2}, and U_1 lie étale of degree 1 over the p-rank 2 locus, the p-rank 1 locus, the locus of abelian surfaces with p-rank 0 and a-number 1, and the locus of superspecial abelian surfaces ($a = 2$). The locus U_{s_1} is an open part of the fibers over the superspecial points. The locus $U_{s_2s_1}$ is of dimension 2 and lies finite but inseparably of degree p over the p-rank 1 locus. Then E_1 corresponds to an α_p and E_2/E_1 to a μ_p. In the final type locus $U_{s_1s_2}$ the filtration is $\mu_p \subset \mu_p \oplus \alpha_p$. Note that this description is compatible with the calculated classes of the loci.

We have implemented the calculation of the Gysin map in Macaulay2 (cf. [M2]) and calculated all cycle classes for $g \leq 5$. For $g = 3, 4$ the reader will find the classes in the appendix. (The Macaulay2 code for performing the calculations can be found at `http://www.math.su.se/ teke/strata.m2`.) We shall return to the qualitative consequences one can draw from Theorem 12.1 in the next section.

12.2 The p-rank strata

It is very useful to have closed formulas for the cycle classes of important strata. We give the formulas for the strata defined by the p-rank and by the a-number. The formulas for the p-rank strata can be derived immediately from the definition of the strata.

Let V_f be the closed E-O stratum of $\tilde{\mathcal{A}}_g$ of semi-abelian varieties of p-rank $\leq f$. It has codimension $g - f$. To calculate its class we consider the element $w_\emptyset$, the longest final element. The corresponding locus $\overline{\mathcal{U}}_\emptyset$ is a generically finite cover of $\mathcal{A}_g$ of degree $\gamma_g(w_\emptyset) = \prod_{i=1}^{g-1}(p^i + p^{i-1} + \cdots + 1)$. The map of $\mathcal{U}_\emptyset$ to the p-rank g locus is finite. The space $\overline{\mathcal{U}}_\emptyset$ contains the degeneracy loci $\mathcal{U}_w$ for all final elements $w \in W_g$. The condition that a point x of $\mathcal{F}_g$ lie in $\overline{\mathcal{U}}_\emptyset$ is that the filtration $\mathbb{E}_i$ for $i = 1, \ldots, g$ be stable under V. By forgetting part of the flag and considering flags $\mathbb{E}_j$ with $j = i, \ldots, g$ we find that $\overline{\mathcal{U}}_\emptyset \to \mathcal{A}_g$ is fibered by generically finite morphisms

$$\overline{\mathcal{U}}_\emptyset = \overline{\mathcal{U}}^{(1)} \xrightarrow{\pi_1} \overline{\mathcal{U}}^{(2)} \xrightarrow{\pi_2} \cdots \xrightarrow{\pi_{g-1}} \overline{\mathcal{U}}^{(g)} = \mathcal{A}_g.$$

We shall write $\pi_{i,j}$ for the composition $\pi_j \pi_{j-1} \cdots \pi_i \colon \overline{\mathcal{U}}^{(i)} \to \overline{\mathcal{U}}^{(j)}$ and $\pi_\emptyset = \pi_{1,g}$.

Since V_{g-1} is given by the vanishing of the map $\det(V)\colon \det(\mathbb{E}_g) \to \det(\mathbb{E}^{(p)})$, the class of V_{g-1} is $(p-1)\lambda_1$. The pullback of V_{g-1} to $\overline{\mathcal{U}}_\emptyset$ decomposes in g irreducible components

$$\pi_\emptyset^{-1}(V_{g-1}) = \cup_{i=1}^g Z_i,$$

where Z_i is the degeneracy locus of the induced map $\phi_i = V_{|\mathcal{L}_i} \colon \mathcal{L}_i \to \mathcal{L}_i^{(p)}$. Note that the Z_i are the $\overline{\mathcal{U}}_w$ for the w that are shuffles of the final element u_{g-1} (see Section 4.2) defining the (open) E-O stratum of p-rank f, and Z_g is the stratum corresponding to the element $u_{g-1} \in W_g$. An abelian variety of p-rank $g-1$ (and thus with a-number 1) has a unique subgroup scheme α_p. The index i of Z_i indicates where this subgroup scheme can be found (i.e., its Dieudonné module lies in $\mathbb{E}_i$ but not in $\mathbb{E}_{i-1}$).

It follows from the definition of Z_i as degeneracy set that the class of Z_i on $\overline{\mathcal{U}}_\emptyset$ equals $(p-1)\ell_i$, since ϕ_i can be interpreted as a section of $\mathcal{L}_i^{(p)} \otimes \mathcal{L}_i^{-1}$. We also know by Section 4.3 that the map $Z_i \to Z_{i+1}$ is inseparable. Therefore $\pi_\emptyset([Z_i]) = p^{n(i)} \pi_\emptyset([Z_g])$ for some integer $n(i) \geq g-i$. Using the fact that $(\pi_\emptyset)_*([Z_g]) = \gamma_g(u_{g-1})[V_{g-1}] = \deg(\pi_{1,g-1})[V_{g-1}]$, we see that

$$(\pi_\emptyset)_*(\pi_\emptyset^*([V_{g-1}]) = \sum_{i=1}^g (\pi_\emptyset)_*([Z_i]) = \sum_{i=1}^g p^{n(i)} \deg(\pi_{1,g-1})[V_{g-1}],$$

while on the other hand,

$$(\pi_\emptyset)_*(\pi_\emptyset^*([V_{g-1}]) = \deg(\pi_\emptyset)[V_{g-1}] = (1+p+\cdots+p^{g-1}) \deg(\pi_{1,g-1})[V_{g-1}].$$

Comparison yields that $n(i) = g-i$ and so we obtain

$$(\pi_\emptyset)_*(\ell_i) = p^{g-i} \deg(\pi_{1,g-1})\lambda_1$$

and

$$(\pi_\emptyset)_*([Z_i]) = (p-1)p^{g-i} \deg(\pi_{1,g-1})\lambda_1.$$

Lemma 12.3. *In the Chow groups with rational coefficients of $\overline{\mathcal{U}}^{(i)}$ and $\overline{\mathcal{U}}^{(i+1)}$ we have for the pushdown of the jth Chern class $\lambda_j(i)$ of $\mathbb{E}_i$ the relations*

$$\pi_*^i \lambda_j(i) = p^j(p^{i-j} + p^{i-j-1} + \cdots + p + 1)\lambda_j(i+1)$$

and

$$p^{f(g-f)}\,(\pi_{1,g})_*(\ell_g \ell_{g-1} \cdots \ell_{f+1}) = (\pi_{1,g})_*(\ell_1 \ell_2 \cdots \ell_{g-f}).$$

Proof. The relation $(\pi_1)_*([Z_1]) = p\,[Z_2]$ translates into the case $j=1$ and $i=1$. Using the push–pull formula and the relations $\pi_i^*(\lambda_j(i+1) = \ell_{i+1}\lambda_{j-1}(i) + \lambda_j(i)$, the formulas for the pushdowns of the $\lambda_j(i)$ follow by induction on j and i. □

We now calculate the class of all p-rank strata V_f.

Theorem 12.4. *The class of the locus V_f of semi-abelian varieties of p-rank $\leq f$ in the Chow ring $\mathrm{CH}_{\mathbb{Q}}(\tilde{\mathcal{A}}_g)$ equals*

$$[V_f] = (p-1)(p^2-1)\cdots(p^{g-f}-1)\,\lambda_{g-f},$$

where λ_i denotes the ith Chern class of the Hodge bundle.

Proof. The class of the final stratum $\overline{\mathcal{U}}_{u_f}$ on $\overline{\mathcal{U}}_{\emptyset}$ is given by the formula

$$(p-1)^{g-f}\ell_g\ell_{g-1}\cdots\ell_{f+1},$$

since it is the simultaneous degeneracy class of the maps ϕ_j for $j=f+1,\ldots,g$. By pushing down this class under $\pi_{\emptyset}=\pi_{1,g}$ we find using Lemma 12.3 and the notation $\lambda_j(i)=c_j(\mathbb{E}_i)$ that

$$\begin{aligned}(\pi_{1,g})_*(\ell_g\ell_{g-1}\cdots\ell_{f+1}) &= p^{-f(g-f)}(\pi_{1,g})_*(\ell_1\ell_2\cdots\ell_{g-f})\\ &= p^{-f(g-f)}\,(\pi_{1,g})_*(\pi^*_{1,g-f}(\lambda_{g-f}(g-f)))\\ &= p^{-f(g-f)}\,\deg(\pi_{1,g-f})\,(\pi_{g-f,g})_*(\lambda_{g-f}(g-f)).\end{aligned}$$

Applying Lemma 12.3 repeatedly we obtain

$$\begin{aligned}(\pi_{g-f,g})_*(\lambda_{g-f}(g-f)) &= p^{f(g-f)}(1+p)(1+p+p^2)\cdots(1+\cdots+p^{f-1})\lambda_g\\ &= p^{f(g-f)}\,\gamma_g(u_f)\,\lambda_g,\end{aligned}$$

with $\gamma_g(u_f)$ the number of final filtrations refining the canonical filtration of u_f. Hence we get $(\pi_{\emptyset})_*(\overline{\mathcal{U}}_{u_f})=(p-1)^{g-f}\deg(\pi_{1,g-f})\gamma_g(u_f)\lambda_g$. On the other hand, we have that $(\pi_{\emptyset})_*(\overline{\mathcal{U}}_{u_f})=\gamma_g(u_f)\,[V_f]$. All together these formulas prove the result. $\square$

12.3 The a-number Strata

Another case in which we can find attractive explicit formulas is that of the E-O strata $\mathcal{V}_w$ with w the element of W_g associated to $Y=\{1,2,\ldots,a\}$. We denote these by T_a. Here we can work directly on $\mathcal{A}_g$. The locus T_a on $\mathcal{A}_g$ may be defined as the locus $\{x\in\mathcal{A}_g\colon \mathrm{rank}(V)|_{\mathbb{E}_g}\leq g-a\}$. We have $T_{a+1}\subset T_a$ and $\dim(T_g)=0$. We apply now formulas of Pragacz and Ratajski [PR97] for the degeneracy locus for the rank of a self-adjoint bundle map of symplectic bundles globalizing the results in isotropic Schubert calculus from [Pr91]. Before we apply their result to our case we have to introduce some notation.

Define for a vector bundle A with Chern classes a_i the expression

$$Q_{ij}(A) := a_ia_j + 2\sum_{k=1}^{j}(-1)^k a_{i+k}a_{j-k} \quad \text{for} \quad i>j.$$

A subset $\beta = \{g \geq \beta_1 > \cdots > \beta_r \geq 0\}$ of $\{1, 2, \ldots, g\}$ (with r even, note that β_r may be zero) is called admissible, and for such subsets we set

$$Q_\beta = \mathrm{Pfaffian}(x_{ij}),$$

where the matrix (x_{ij}) is antisymmetric with entries $x_{ij} = Q_{\beta_i,\beta_j}$. Applying the formula of Pragacz–Ratajski to our situation gives the following result:

Theorem 12.5. *The cycle class $[T_a]$ of the reduced locus T_a of abelian varieties with a-number $\geq a$ is given by*

$$[T_a] = \sum_\beta Q_\beta(\mathbb{E}^{(p)}) \cdot Q_{\rho(a)-\beta}(\mathbb{E}^*),$$

where the sum is over the admissible subsets β contained in the subset $\rho(a) = \{a, a-1, a-2, \ldots, 1\}$.

Example 12.6.

$$\begin{aligned}
[T_1] &= (p-1)\lambda_1 \\
[T_2] &= (p-1)(p^2+1)\lambda_1\lambda_2 - (p^3-1)2\lambda_3 \\
&\cdots \\
[T_g] &= (p-1)(p^2+1)\cdots(p^g+(-1)^g)\lambda_1\lambda_2\cdots\lambda_g.
\end{aligned}$$

As a corollary we obtain a result of one of us [Ek87] on the number of principally polarized abelian varieties with $a = g$.

Corollary 12.7. *We have*

$$\sum_X \frac{1}{\#\mathrm{Aut}(X)} = (-1)^{g(g+1)/2} 2^{-g} \left[\prod_{j=1}^{g} (p^j + (-1)^j)\right] \cdot \zeta(-1)\zeta(-3)\cdots\zeta(1-2g),$$

where the sum is over the isomorphism classes (over $\bar{\mathbb{F}}_p$) of principally polarized abelian varieties of dimension g with $a = g$, and $\zeta(s)$ is the Riemann zeta function.

Proof. Combine the formula for T_g with the Hirzebruch–Mumford proportionality theorem (see [Ge99]), which says that

$$\deg(\lambda_1\lambda_2\cdots\lambda_g) = (-1)^{\frac{g(g+1)}{2}} \prod_{j=1}^{g} \frac{\zeta(1-2j)}{2},$$

when interpreted for the stack $\mathcal{A}_g$. □

The formulas for the cycles classes of the p-rank strata and the a-number strata can be seen as generalizations of the classical formula of Deuring (known as Deuring's mass formula), which states that

$$\sum_E \frac{1}{\#\mathrm{Aut}(E)} = \frac{p-1}{24},$$

where the sum is over the isomorphism classes over $\bar{\mathbb{F}}_p$ of supersingular elliptic curves. It is obtained from the formula for V_{g-1} or T_1 for $g=1$, i.e., $[V_1] = (p-1)\lambda_1$, by observing that the degree of λ_1 is $1/12$ the degree of a generic point of the stack $\tilde{\mathcal{A}}_1$.

One can obtain formulas for all the E-O strata by applying the formulas of Pragacz–Ratajski or those of Kresch–Tamvakis [KT02, Corollary 4]. If Y is a Young diagram given by a subset $\{\xi_1,\ldots,\xi_r\}$ we call $|\xi| = \sum_{i=1}^r \xi_i$ the weight and r the length of ξ. Moreover, we need the *excess* $e(\xi) = |\xi| - r(r+1)/2$ and the *intertwining number* $e(\xi,\eta)$ of two strict partitions with $\xi \cap \eta = \emptyset$ by

$$e(\xi,\eta) = \sum_{i\geq 1} i\,\#\{j : \xi_i > \eta_j > \xi_{i+1}\}$$

(where we use $\xi_k = 0$ if $k > r$). We put $\rho_g = \{g, g-1, \ldots, 1\}$ and $\xi' = \rho_g - \xi$ and have then $e(\xi,\xi') = e(\xi)$. The formula obtained by applying the result of Kresch and Tamvakis interpolates between the formulas for the two special cases, the p-rank strata and a-number strata, as follows:

Theorem 12.8. *For a Young diagram given by a partition ξ we have*

$$[\overline{\mathcal{V}}_Y] = (-1)^{e(\xi)+|\xi'|} \sum_\alpha Q_\alpha(\mathbb{E}^{(p)}) \sum_\beta (-1)^{e(\alpha,\beta)} Q_{(\alpha\cup\beta)'}(\mathbb{E}^*)\det(c_{\beta_i-\xi'_j}(\mathbb{E}^*_{g-\xi'_j})),$$

where the sum is over all admissible α and all admissible β that contain ξ' with length $\ell(\beta) = \ell(\xi')$ and $\alpha \cap \beta = \emptyset$.

12.4 Positivity of tautological classes

The Hodge bundle possesses certain positivity properties. It is well known that the determinant of the Hodge bundle (represented by the class λ_1) is ample on $\mathcal{A}_g$. Over $\mathbb{C}$ this is a classical result, while in positive characteristic this was proven by Moret–Bailly [MB85]. On the other hand, the Hodge bundle itself is not positive in positive characteristic. For example, for $g=2$ the restriction of $\mathbb{E}$ to a line from the p-rank 0 locus is $O(-1) \oplus O(p)$, [MB81]. But our Theorem 12.4 implies the following nonnegativity result.

Theorem 12.9. *The Chern classes $\lambda_i \in \mathrm{CH}_{\mathbb{Q}}(\mathcal{A}_g \otimes \mathbb{F}_p)$ $(i = 1,\ldots,g)$ of the Hodge bundle $\mathbb{E}$ are represented by effective classes with $\mathbb{Q}$–coefficients.*

13 Tautological rings

We shall now interpret the results of previous sections in terms of tautological rings. Recall that the tautological ring of $\tilde{\mathcal{A}}_g$ is the subring of $\mathrm{CH}_{\mathbb{Q}}(\tilde{\mathcal{A}}_g)$ generated by the Chern classes λ_i. To obtain maximal precision we shall use the subring and *not* the $\mathbb{Q}$-subalgebra (but note that this is still a subring of $\mathrm{CH}_{\mathbb{Q}}(\tilde{\mathcal{A}}_g)$ not of the integral Chow ring $\mathrm{CH}^*(\tilde{\mathcal{A}}_g)$). As a graded ring it is isomorphic to the Chow ring $\mathrm{CH}^*(\mathrm{Sp}_{2g}/P_H)$ and as an abstract graded ring it is generated by the λ_i with relations coming from the identity $1=(1+\lambda_1+\cdots+\lambda_g)(1-\lambda_1+\cdots+(-1)^g\lambda_g)$. This implies that it has a $\mathbb{Z}$-basis consisting of the square-free monomials in the λ_i. (Note, however, that the degree maps from the degree $g(g+1)/2$ part are not the same; on Sp_{2g}/P_H the degree of $\lambda_1\cdots\lambda_g$ is ± 1, whereas for $\tilde{\mathcal{A}}_g$ it is given by the Hirzebruch–Mumford proportionality theorem as in the previous section.) Since $\tilde{\mathcal{F}}_g\to\tilde{\mathcal{A}}_g$ is an SL_g/B-bundle, we can express $\mathrm{CH}_{\mathbb{Q}}(\tilde{\mathcal{F}}_g)$ as an algebra over $\mathrm{CH}_{\mathbb{Q}}(\tilde{\mathcal{A}}_g)$; it is the algebra generated by the ℓ_i, and the relations are that the elementary symmetric functions in them are equal to the λ_i. This makes it natural to define the *tautological ring* of $\tilde{\mathcal{F}}_g$ to be the subring of $\mathrm{CH}_{\mathbb{Q}}(\tilde{\mathcal{F}}_g)$ generated by the ℓ_i. It will then be the algebra over the tautological ring of $\tilde{\mathcal{A}}_g$ generated by the ℓ_i and with the relations that say that the elementary symmetric functions in the ℓ_i are equal to the λ_i. Again this means that the tautological ring for $\tilde{\mathcal{F}}_g$ is isomorphic to the integral Chow ring of Sp_{2g}/B_g, the space of full symplectic flags in a $2g$-dimensional symplectic vector space. Note furthermore that the Gysin maps for $\mathrm{Sp}_{2g}/B_g\to\mathrm{Sp}_{2g}/P_H$ and $\tilde{\mathcal{F}}_g\to\tilde{\mathcal{A}}$ are both given by Formula 3.1.

Theorem 12.1 shows in particular that the classes of the $\overline{\mathcal{U}}_w$ and $\overline{\mathcal{V}}_\nu$ lie in the respective tautological rings. However, we want both to compare the formulas for these classes with the classical formulas for the Schubert varieties and to take into account the variation of the coefficients of the classes when expressed in a fixed basis for the tautological ring. Hence since the rest of this section is purely algebraic, we shall allow ourselves the luxury of letting p temporarily be also a polynomial variable. We then introduce the ring $\mathbb{Z}\{p\}$, which is the localization of the polynomial ring $\mathbb{Z}[p]$ at the multiplicative subset of polynomials with constant coefficient equal to 1. Hence evaluation at 0 extends to a ring homomorphism $\mathbb{Z}\{p\}\to\mathbb{Z}$, which we shall call the *classical specialization*. An element of $\mathbb{Z}\{p\}$ is thus invertible precisely when its classical specialization is invertible. By a modulo n consideration we see that an integer polynomial with 1 as constant coefficient can have no integer zero $n\neq\pm1$. This means in particular that evaluation at a prime p induces a ring homomorphism $\mathbb{Z}\{p\}\to\mathbb{Q}$ taking the variable p to the integer p, which we shall call the *characteristic p specialization* (since p will be an integer only when this phrase is used, there should be no confusion because of our dual use of p). We now extend scalars of the two tautological rings from $\mathbb{Z}$ to $\mathbb{Z}\{p\}$ and we shall call them the *p-tautological rings*. We shall also need to express the condition that an element is in the subring obtained by extension to a subring of $\mathbb{Z}\{p\}$ and we shall then say that the element *has coefficients* in the subring. We may consider the Fulton polynomial P_w of (4) as a polynomial with coefficients in $\mathbb{Z}\{p\}$, and when we evaluate them on elements of the tautological ring as in Theorem 12.1 we get elements $[\overline{\mathcal{U}}_w]$ of the p-tautological ring of $\tilde{\mathcal{F}}_g$. If ν is a final element we can push down the formula for $[\overline{\mathcal{U}}_\nu]$ using Formula 3.1 and then we get an element in the p-tautological ring of $\tilde{\mathcal{A}}_g$. We then note that $\gamma(w)$ is a polynomial in p with constant coefficient equal to 1, and hence we can define $[\overline{\mathcal{V}}_\nu] := \gamma(w)^{-1}\pi_*[\overline{\mathcal{U}}_\nu]$, where $\pi\colon \tilde{\mathcal{F}}_g \to \tilde{\mathcal{A}}_g$ is the projection map. By construction these elements map to the classes of $\overline{\mathcal{U}}_w$, respectively $\overline{\mathcal{V}}_\nu$, under specialization to characteristic p. We shall need to compare them with the classes of the Schubert varieties. To be specific we shall define the Schubert varieties of Sp_{2g}/B_g by the condition $\dim E_i \cap D_j \geq r_w(i,j)$, where $D_\bullet$ is a fixed reference flag (and then the Schubert varieties of Sp_{2g}/P_H are the images of the Schubert varieties of Sp_{2g}/B_g for final elements of W_g).

Theorem 13.1. *(i) The classes $[\overline{\mathcal{U}}_w]$ and $[\overline{\mathcal{V}}_\nu]$ in the p-tautological ring of $\tilde{\mathcal{F}}_g$ map to the classes of the corresponding Schubert varieties under classical specialization.*

(ii) The classes $[\overline{\mathcal{U}}_w]$ and $[\overline{\mathcal{V}}_\nu]$ form a $\mathbb{Z}\{p\}$-basis for the respective p-tautological rings.

(iii) The coefficients of $[\overline{\mathcal{U}}_w]$ and $[\overline{\mathcal{V}}_\nu]$ when expressed in terms of the polynomials in the λ_i are in $\mathbb{Z}[p]$.

(iv) For $w \in W_g$ we have that $\ell(w) = \ell(\tau_p(w))$ (see Section 9 for the definition of τ_p) precisely when the specialization to characteristic p of $\pi_*[\overline{\mathcal{U}}_w]$ is nonzero. In particular, there is a unique map $\tau\colon W_g \to (W_g/S_g)\bigsqcup\{0\}$ such that $\tau(w) = 0$ precisely when $\pi_*[\overline{\mathcal{U}}_w] = 0$, which implies that $\ell(w) \neq \ell(\tau_p(w))$ and is implied by $\ell(w) \neq \ell(\tau_p(w))$ for all sufficiently large p. Furthermore, if $\tau(w) \neq 0$ then $\ell(w) = \ell(\tau(w))$ and $\pi_*[\overline{\mathcal{U}}_w]$ is a nonzero multiple of $[\overline{\mathcal{V}}_{\tau(w)}]$.

Proof. The first part is clear, since putting $p = 0$ in our formulas gives the Fulton formulas for $x_i = -\ell_i$ and $y_i = 0$, which are the Fulton formulas for the Schubert varieties in Sp_{2g}/B_g. One then obtains the formulas for the Schubert varieties of Sp_{2g}/P_H by pushing down by Gysin formulas. The remaining compatibility needed is that the classical specialization of $\gamma(w)$ is the degree of the map from the Schubert variety of Sp_{2g}/B_g for a final element to the corresponding Schubert variety of Sp_{2g}/P_H. However, the classical specialization of $\gamma(w)$ is 1, and the map between Schubert varieties is an isomorphism between Bruhat cells.

As for the second part, we need to prove that the determinant of the matrix expressing the classes of the strata in terms of a basis of the tautological ring (say given by monomials in the ℓ_i respectively the λ_i) is invertible. Given that an element of $\mathbb{Z}\{p\}$ is invertible precisely when its classical specialization is, we are reduced to proving the corresponding statement in the classical case. However, there it follows from the cell decomposition given by the Bruhat cells, which give that the classes of the Schubert cells form a basis for the integral Chow groups.

To prove (iii) it is enough to verify the conditions of Proposition 13.2. Hence consider $w \in W_g$ (respectively a final element ν) and consider an element m in the tautological ring of degree complementary to that of $[\overline{\mathcal{U}}_w]$ (respectively $[\overline{\mathcal{V}}_\nu]$). By the projection formula the degree of $m[\overline{\mathcal{U}}_w]$ (respectively $m[\overline{\mathcal{V}}_\nu]$) is the degree of the restriction of m to $\overline{\mathcal{U}}_w$ (respectively $\overline{\mathcal{V}}_\nu$), and it is enough to show that the denominators of these degrees are divisible only by a finite number of primes (independently of the characteristic p). However, if the characteristic is different from 3 we may pull back to the moduli space with a level 3 structure, and there the degree is an integer, since the corresponding strata are schemes. Hence the denominator divides the degree of the level 3 structure covering $\tilde{\mathcal{A}}_{g,3} \to \tilde{\mathcal{A}}_g$, which is independent of p.

Finally for (iv), it is clear that in the Chow ring of $\tilde{\mathcal{A}}_g$ the class $\pi_*[\overline{\mathcal{U}}_w]$ is nonzero precisely when $\pi\colon [\overline{\mathcal{U}}_w] \to \overline{\mathcal{V}}_{\tau_p(w)}$ is generically finite, since all fibers have the same dimension by Proposition 9.6. This latter fact also gives that it is generically finite precisely when $\overline{\mathcal{U}}_w$ and $\overline{\mathcal{V}}_{\tau_p(w)}$ have the same dimension, which is equivalent to $\ell(w) = \ell(\tau_p(w))$. When this is the case, we get that $\pi_*[\overline{\mathcal{U}}_w]$ is a nonzero multiple of $[\overline{\mathcal{V}}_{\tau_p(w)}]$, again since the degree over each component of $\overline{\mathcal{V}}_{\tau_p(w)}$ is the same by Proposition 9.6. Consider now instead $\pi_*[\overline{\mathcal{U}}_w]$ in the p-tautological ring and expand $\pi_*[\overline{\mathcal{U}}_w]$ as a linear combination of the $[\overline{\mathcal{V}}_\nu]$. Then what we have just shown is that for every specialization to characteristic p, at most one of the coefficients is nonzero. This implies that in the p-tautological ring at most one of the coefficients is nonzero. If it is zero then $\pi_*[\overline{\mathcal{U}}_w]$ is always zero in all characteristic p specializations, and we get $\ell(w) \neq \ell(\tau_p(w))$ for all p. If it is nonzero, then the coefficient is nonzero for all sufficiently large p. This proves (iv). □

To complete the proof of the theorem we need to prove the following proposition.

Proposition 13.2. *Let a be an element of the p-tautological ring for $\tilde{\mathcal{F}}_g$ or $\tilde{\mathcal{A}}_g$. Assume that there exists an $n \neq 0$ such that for all elements b of the tautological ring of complementary degree and all sufficiently large primes p we have that $\deg(\overline{a}\overline{b}) \in \mathbb{Z}[n^{-1}]$, where $\overline{a}$ and $\overline{b}$ are the specializations to characteristic p of a, respectively b. Then the coefficients of a are in $\mathbb{Z}[p]$.*

Proof. If $r(x)$ is one of the coefficients of a, then the assumptions say that $r(p) \in \mathbb{Z}[n^{-1}]$ for all sufficiently large primes p. Write r as $g(x)/f(x)$ where f and g are integer polynomials with no common factor. Thus there are integer polynomials $s(x)$ and $t(x)$ such that $s(x)f(x) + t(x)g(x) = m$, where m is a nonzero integer. If g is nonconstant there are arbitrarily large primes ℓ such that there is an integer k with $\ell | f(k)$ (by, for instance, the fact that there is a prime that splits completely in a splitting field of f). By Dirichlet's theorem on primes in arithmetic progressions there are arbitrarily large primes p such that $f(p) \neq 0$ and $f(p) \equiv f(k) \equiv 0 \bmod \ell$. By making ℓ so large that $\ell \nmid m$, we get that $\ell \nmid g(q)$ (since $s(q)f(q) + t(q)g(q) = m$) and hence ℓ appears in the denominator of $r(q)$. By making ℓ so large that $\ell \nmid n$, we conclude. □

Example 13.3. If w is a shuffle of a final element ν we have $\tau(w)=\nu$.

We can combine Theorem 13.1 with our results on the punctual flag spaces to give an algebraic criterion for inclusion between E-O strata.

Corollary 13.4. *Let ν' and ν be final elements. Then for sufficiently large p we have that $\nu'\subseteq\nu$ if there are $w,w'\in W_g$ for which $w'\leq w$, $\tau(w)=\nu$, and $\tau(w')=\nu'$.*

Proof. Assume that there are $w,w'\in W_g$ for which $w'\leq w$, $\tau(w)=\nu$, and $\tau(w')=\nu'$. By Proposition 9.6 we have for $\pi\colon\tilde{\mathcal{F}}_g\to\tilde{\mathcal{A}}_g$, the image relations $\pi(\overline{\mathcal{U}}_w)=\overline{\mathcal{V}}_{\nu_1}$ and $\pi(\overline{\mathcal{U}}_{w'})=\overline{\mathcal{V}}_{\nu_1'}$ for some ν_1 and ν_1', and by the theorem $\nu_1=\tau(w)$ and $\nu_1'=\tau(w')$ for sufficiently large p. Since $w'\leq w$ we have that $\overline{\mathcal{U}}_{w'}\subseteq\overline{\mathcal{U}}_w$, which implies that $\pi\overline{\mathcal{U}}_{w'}\subseteq\pi(\overline{\mathcal{U}}_w)$. □

14 Comparison with $\mathcal{S}(g,p)$

We shall now make a comparison with de Jong's moduli stack $\mathcal{S}(g,p)$ of $\Gamma_0(p)$-structures (cf. [Jo93]). Recall that for a family $\mathcal{A}\to S$ of principally polarized g-dimensional abelian varieties a *$\Gamma_0(p)$-structure* consists of the choice of a flag $0\subset H_1\subset\cdots\subset H_g\subset\mathcal{A}[p]$ of flat subgroup schemes with H_i of order p^i and H_g totally isotropic with respect to the Weil pairing. We shall work exclusively in characteristic p and denote by $\overline{\mathcal{S}(g,p)}$ the mod p fiber of $\mathcal{S}(g,p)$. We now let $\mathcal{S}(g,p)^0$ be the closed subscheme of $\overline{\mathcal{S}(g,p)}$ defined by the condition that the group scheme H_g be of height 1. This means that the (relative) Frobenius map $F_{\mathcal{A}/\mathcal{S}(g,p)}$, where $\pi\colon\mathcal{A}\to\mathcal{S}(g,p)$ is the universal abelian variety, is zero on it. For degree reasons we then get that H_g equals the kernel of $F_{\mathcal{A}/\mathcal{S}(g,p)}$. Using the principal polarization we may identify the Lie algebra of π with $R^1\pi_*\mathcal{O}_{\mathcal{A}}$, and hence we get a flag $0\subset\mathrm{Lie}(H_1)\subset\mathrm{Lie}(H_2)\subset\cdots\subset\mathrm{Lie}(H_g)=R^1\pi_*\mathcal{O}_{\mathcal{A}}$. By functoriality this is stable under V. Completing this flag by taking its annihilator in $\mathcal{E}$ gives a flag in $\overline{\mathcal{U}}_{w_\emptyset}$, thus giving a map $\mathcal{S}(g,p)^0\to\overline{\mathcal{U}}_{w_\emptyset}$.

Theorem 14.1. *The canonical map $\mathcal{S}(g,p)^0\to\overline{\mathcal{U}}_{w_\emptyset}$ is an isomorphism. In particular, $\mathcal{S}(g,p)^0$ is the closure of its intersection with the locus of ordinary abelian varieties and is normal and Cohen–Macaulay.*

Proof. Starting with the tautological flag $\{\mathbb{E}_i\}$ on $\overline{\mathcal{U}}_{w_\emptyset}$ we consider the induced flag $\{\mathbb{E}_{g+i}/\mathbb{E}_g\}$ in $R^1\pi_*\mathcal{O}_{\mathcal{A}}$. This is a V-stable flag of the Lie algebra of a height 1 group scheme, so by, for instance, [Mu70, Theorem §14], any V-stable subbundle comes from a subgroup scheme of the kernel of $F_{\mathcal{A}/\mathcal{S}(g,p)}$ and thus the flag $\{\mathbb{E}_{g+i}/\mathbb{E}_g\}$ gives rise to a complete flag of subgroup schemes with H_g equal to the kernel of the Frobenius map and hence a map from $\overline{\mathcal{U}}_{w_\emptyset}$ to $\mathcal{S}(g,p)^0$ that clearly is the inverse of the canonical map.

The rest of the theorem now follows from Corollary 8.4. □

15 Appendix $g = 3, 4$

15.1 Admissible Strata for $g = 3$

In the following matrix one finds the loci lying in $\overline{\mathcal{U}}_{w_\emptyset}$. In the sixth column we give for each w an example of a final ν such that $w \to \nu$.

ℓ	Y	w	ν	word	$w \to \nu$	
0	$\{1,2,3\}$	[123]	$\{0,0,0\}$	Id	[123]	Superspecial
1		[132]	$\{0,0,0\}$	s_2	[123]	Fiber over s.s.
1		[213]	$\{0,0,0\}$	s_1	[123]	Fiber over s.s.
1	$\{2,3\}$	[124]	$\{0,0,1\}$	s_3	[124]	Moret-Bailly
2		[142]	$\{0,0,1\}$	s_3s_2	[135]	
2		[214]	$\{0,0,1\}$	s_3s_1	[135]	
2		[231]	$\{0,0,0\}$	s_1s_2	[123]	Fiber over s.s.
2		[312]	$\{0,0,0\}$	s_2s_1	[123]	Fiber over s.s.
2	$\{1,3\}$	[135]	$\{0,1,1\}$	s_2s_3	[135]	$f=0$, $a=2$
3		[153]	$\{0,1,1\}$	$s_2s_3s_2$	[236]	Shuffle of $\{1,2\}$
3		[241]	$\{0,0,1\}$	$s_3s_1s_2$	[124]	
3		[315]	$\{0,1,1\}$	$s_2s_3s_1$	[124]	
3		[321]	$\{0,0,0\}$	$s_1s_2s_1$	[123]	Fiber over s.s.
3		[412]	$\{0,0,1\}$	$s_3s_2s_1$	[236]	Shuffle of $\{1,2\}$
3	$\{3\}$	[145]	$\{0,1,2\}$	$s_3s_2s_3$	[145]	$f=0$
3	$\{1,2\}$	[236]	$\{1,1,1\}$	$s_1s_2s_3$	[236]	$a=2$
4		[154]	$\{0,1,2\}$	$s_3s_2s_3s_2$	[246]	Shuffle of $\{2\}$
4		[326]	$\{1,1,1\}$	$s_1s_2s_3s_1$	[236]	
4		[351]	$\{0,1,1\}$	$s_2s_3s_1s_2$	[236]	
4		[415]	$\{0,1,2\}$	$s_3s_2s_3s_1$	[246]	Shuffle of $\{2\}$
4		[421]	$\{0,0,1\}$	$s_3s_1s_2s_1$	[236]	
4	$\{2\}$	[246]	$\{1,1,2\}$	$s_3s_1s_2s_3$	[246]	$f=1$
5		[426]	$\{1,1,2\}$	$s_3s_1s_2s_3s_1$	[356]	Shuffle of $\{1\}$
5		[451]	$\{0,1,2\}$	$s_3s_2s_3s_1s_2$	[356]	Shuffle of $\{1\}$
5	$\{1\}$	[356]	$\{1,2,2\}$	$s_2s_3s_1s_2s_3$	[356]	$f=2$
6	$\{\}$	[456]	$\{1,2,3\}$	$s_3s_2s_3s_1s_2s_3$	[456]	$f=3$

15.2 E-O Cycle Classes for $g = 3$

We give the cycle classes of the (reduced) E-O strata for $g = 3$.

Y	class
$\emptyset$	1
$\{1\}$	$(p-1)\lambda_1$
$\{2\}$	$(p-1)(p^2-1)\lambda_2$
$\{1,2\}$	$(p-1)(p^2+1)\lambda_1\lambda_2-2(p^3-1)\lambda_3$
$\{3\}$	$(p-1)(p^2-1)(p^3-1)\lambda_3$
$\{1,3\}$	$(p-1)^2(p^3+1)\lambda_1\lambda_3$
$\{2,3\}$	$(p-1)^2(p^6-1)\lambda_2\lambda_3$
$\{1,2,3\}$	$(p-1)(p^2+1)(p^3-1)\lambda_1\lambda_2\lambda_3$

15.3 E-O Cycle Classes for $g=4$

We give the cycle classes of the (reduced) E-O strata for $g=4$.

Y	class
$\emptyset$	1
$\{1\}$	$(p-1)\lambda_1$
$\{2\}$	$(p-1)(p^2-1)\lambda_2$
$\{1,2\}$	$(p-1)(p^2+1)\lambda_1\lambda_2-2(p^3-1)\lambda_3$
$\{3\}$	$(p-1)(p^2-1)(p^3-1)\lambda_3$
$\{1,3\}$	$(p-1)^2(p+1)((p^2-p+1)\lambda_1\lambda_3-2(p^2+1)\lambda_4)$
$\{2,3\}$	$(p-1)^2((p^6-1)\lambda_2\lambda_3-(2p^6+p^5-p-2)\lambda_1\lambda_4$
$\{1,2,3\}$	$(p-1)(p^2+1)((p^3+1)((p^3-1)\lambda_1\lambda_2\lambda_3-2(3p^3+p^2-p+3)\lambda_2\lambda_4)$
$\{4\}$	$(p-1)(p^2-1)(p^3-1)(p^4-1)\lambda_4$
$\{1,4\}$	$(p-1)^3(p+1)(p^4+1)\lambda_1\lambda_4$
$\{2,4\}$	$(p-1)^3(p^8-1)\lambda_2\lambda_4$
$\{1,2,4\}$	$(p-1)^2(p^4-1)((p^2+1)\lambda_1\lambda_2-2(p^2+p+1)\lambda_3)\lambda_4$
$\{3,4\}$	$(p-1)^2(p^2+1)(p^3-1)(p^2-p+1)((p+1)^2\lambda_3-p\lambda_1\lambda_2)\lambda_4$
$\{1,3,4\}$	$(p-1)^2(p^4-1)(p^6-1)\lambda_1\lambda_3\lambda_4$
$\{2,3,4\}$	$(p-1)(p^6-1)(p^8-1)\lambda_2\lambda_3\lambda_4$
$\{1,2,3,4\}$	$(p-1)(p^2+1)(p^3-1)(p^4+1)\lambda_1\lambda_2\lambda_3\lambda_4$

References

[BGG73] I. N. Bernstein, I. M. Gelfand, S. I. Gelfand, *Schubert cells, and the cohomology of the spaces G/P*, Uspehi Mat. Nauk **28** (1973), no. 3(171), 3–26.

[BL00] S. Billey, V. Lakshmibai, *Singular Loci of Schubert Varieties*, Progs. in Math., vol. 182, Birkhäuser Boston Inc., Boston, MA, 2000.

[Ch94] C. Chevalley, *Sur les décompositions cellulaires des espaces G/B*, Algebraic groups and their generalizations: classical methods (University Park, PA, 1991), Proc. Sympos. Pure Math., vol. 56, Amer. Math. Soc., Providence, RI, 1994, With a foreword by Armand Borel, pp. 1–23.

[Ek87] T. Ekedahl, *On supersingular curves and abelian varieties*, Math. Scand. **60** (1987), 151–178.

[FC90] G. Faltings, C.-L. Chai, *Degeneration of abelian varieties.*, Ergebnisse der Math., no. 22, Springer-Verlag, 1990.

[FP98] W. Fulton, P. Pragacz, *Schubert varieties and degeneracy loci*, SLN, vol. 1689, Springer-Verlag, Berlin, 1998, Appendix J by the authors in collaboration with I. Ciocan-Fontanine.

[Fu96] W. Fulton, *Determinantal formulas for orthogonal and symplectic degeneracy loci*, J. Differential Geom. **43** (1996), no. 2, 276–290.

[Ful] W. Fulton, *Intersection Theory*, vol. 3, Ergebnisse der Mathematik und ihrer Grenzgebiete, no. 2, Springer-Verlag, 1984.

[Ge99] G. van der Geer, *Cycles on the moduli space of abelian varieties*, Moduli of curves and abelian varieties (C. Faber and E. Looijenga, eds.), Aspects Math., E33, Vieweg, Braunschweig, 1999, pp. 65–89.

[Gr65] A. Grothendieck, *Éléments de géométrie algébrique. IV. Étude locale des schémas et des morphismes de schémas. II*, Inst. Hautes Études Sci. Publ. Math. (1965), no. 24, 231.

[GrLie4-6] N. Bourbaki, *Groupes et algèbres de Lie. IV–VI*, Hermann, Paris, 1968.

[HW61] H. Hasse, E. Witt, *Zyklische unverzweigte Erweiterungskörper von primzahlgrade p über einem algebraischen Funktionenkörper der Charakteristik p*, Monatsh. Math. Phys. **43** (1963), 477–492.

[Ha07] S. Harashita, *Ekedahl-Oort strata and the first Newton slope strata*, J. of Algebraic Geom. **16** (2007), no. 1, 171–199.

[Jo93] A. J. de Jong, *The moduli spaces of principally polarized abelian varieties with $\Gamma_0(p)$-level structure*, J. of Algebraic Geom. **2** (1993), no. 4, 667–688.

[KS03] S. Keel, L. Sadun, *Oort's conjecture for $A_g \otimes \mathbb{C}$*, J. Amer. Math. Soc. **16** (2003), no. 4, 887–900.

[KT02] A. Kresch, H. Tamvakis, *Double Schubert polynomials and degeneracy loci for the classical groups*, Ann. Inst. Fourier (Grenoble) **52** (2002), no. 6, 1681–1727.

[M2] D. R. Grayson, M. E. Stillman, *Macaulay 2, a software system for research in algebraic geometry*, Available at `http://www.math.uiuc.edu/Macaulay2`.

[MB81] L. Moret-Bailly, *Familles de courbes et variétés abéliennes sur $\mathbb{P}^1$*, Séminaire sur les pinceaux de courbes de genre au moins deux. (L. Szpiro, ed.), Astérisque, vol. 86, Soc. math. de France, 1981, pp. 109–140.

[MB85] L. Moret-Bailly, *Pinceaux de variétés abéliennes*, Astérisque (1985), no. 129, 266.

[MW04] B. Moonen, T. Wedhorn, *Discrete invariants of varieties in positive characteristic*, IMRN (2004), no. 72, 3855–3903, Eprint: `math.AG/0306339`.

[Mo01] B. Moonen, *Group schemes with additional structures and Weyl group cosets*, Moduli of abelian varieties (Texel Island, 1999), Progr. Math., vol. 195, Birkhäuser, Basel, 2001, pp. 255–298.

[Mu70] D. Mumford, *Abelian varieties*, Tata Institute of Fundamental Research Studies in Mathematics, vol. 5, Oxford University Press., London, 1970.

[Oo01] F. Oort, *A stratification of a moduli space of abelian varieties*, Moduli of abelian varieties (Texel Island, 1999), Progr. Math., vol. 195, Birkhäuser, Basel, 2001, pp. 345–416.

[Oo95] F. Oort, *Complete subvarieties of moduli spaces*, Abelian varieties (Egloffstein, 1993), de Gruyter, Berlin, 1995, pp. 225–235.

[PR97] P. Pragacz, J. Ratajski, *Formulas for Lagrangian and orthogonal degeneracy loci; $\tilde{Q}$-polynomial approach*, Compositio Math. **107** (1997), no. 1, 11–87.

[PR99] H. Pittie, A. Ram, *A Pieri-Chevalley formula in the K-theory of a G/B-bundle*, Electron. Res. Announc. Amer. Math. Soc. **5** (1999), 102–107.

[Pr91] P. Pragacz, *Algebro-geometric applications of Schur S- and Q-polynomials*, Topics in invariant theory (Paris, 1989/1990), Lecture Notes in Math., vol. 1478, Springer-Verlag, Berlin, 1991, pp. 130–191.

[RR85] S. Ramanan, A. Ramanathan, *Projective normality of flag varieties and Schubert varieties*, Invent. Math. **79** (1985), no. 2, 217–224.

[S] N. Sloane, *The on-line encyclopedia of integer sequences*, `http://www.research.att.com/njas/sequences`.

[Sj07] J. Sjöstrand, *Bruhat intervals as rooks on skew Ferrers boards*, J. Combin. Theory Ser. A, **114**(2007), 1182–1198.

Experiments with General Cubic Surfaces

Andreas-Stephan Elsenhans and Jörg Jahnel

Universität Göttingen, Mathematisches Institut*
Bunsenstraße 3–5, D-37073 Göttingen, Germany
elsenhan@uni-math.gwdg.de, jahnel@uni-math.gwdg.de

Dedicated to Yuri Ivanovich Manin on the occasion of his 70th birthday

Summary. For general cubic surfaces, we test numerically the conjecture of Manin (in the refined form due to E. Peyre) about the asymptotics of points of bounded height on Fano varieties. We also study the behavior of the height of the smallest rational point versus the Tamagawa type number introduced by Peyre.

Key words: general cubic surface, smallest point, points of bounded height, asymptotics, Peyre's constant, Manin's conjecture, Diophantine equation, algorithmic resolution, computation of Galois groups.

2000 Mathematics Subject Classifications: 14G05, 11656 (Primary); 11Y50, 14J26 (Secondary)

1 Introduction

The arithmetic of cubic surfaces is a fascinating subject. To a large extent, it was initiated by the work of Yu. I. Manin, particularly by his fundamental and influential book on *Cubic forms* [Ma].

In this article, we study the distribution of rational points on general cubic surfaces over $\mathbb{Q}$. The main problems are

- Existence of $\mathbb{Q}$-rational points,
- Asymptotics of $\mathbb{Q}$-rational points of bounded height,
- The height of the smallest point.

*The computer part of this work was executed on the Sun Fire V20z Servers of the Gauß Laboratory for Scientific Computing at the Göttingen Mathematical Institute. Both authors are grateful to Prof. Y. Tschinkel for permission to use these machines as well as to the system administrators for their support.

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*, Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_14,

Existence of rational points. Let V be an algebraic variety defined over $\mathbb{Q}$. One says that the *Hasse principle* holds for V if

$$V(\mathbb{Q}) = \emptyset \iff \exists\, \nu \in \mathrm{Val}(\mathbb{Q})\colon V(\mathbb{Q}_\nu) = \emptyset\,.$$

For quadrics in $\mathbf{P}^n_{\mathbb{Q}}$, the Hasse principle holds by the famous theorem of Hasse–Minkowski. It is well known that for smooth cubic surfaces over $\mathbb{Q}$ the Hasse principle does not hold in general. In all known examples, this is explained by the Brauer–Manin obstruction. (See Section 2 for details.)

Asymptotics of rational points. The following famous conjecture is due to Yu. I. Manin [FMT].

Conjecture 1 (Manin)**.** *Let V be an arbitrary Fano variety over $\mathbb{Q}$ and* H *an anticanonical height on V. Then there exist a dense, Zariski-open subset $V^\circ \subseteq V$ and a constant C such that*

$$(*) \qquad \#\{x \in V^\circ(\mathbb{Q}) \mid \mathrm{H}(x) < B\} \sim CB \log^{\operatorname{rk}\mathrm{Pic}(V)-1} B$$

for $B \to \infty$.

Peyre's constant. Motivated by results obtained by the classical circle method, E. Peyre refined Manin's conjecture by giving a conjectural value for the leading coefficient C.

Let us explain this more precisely in the particular case that V is a smooth hypersurface in $\mathbf{P}^{d+1}_{\mathbb{Q}}$ defined by a polynomial $f \in \mathbb{Z}[X_0, \ldots, X_{d+1}]$. Assume that $\operatorname{rk}\mathrm{Pic}(V) = 1$ and suppose there is no Brauer–Manin obstruction on V. Then, Peyre's constant is equal to the Tamagawa type number τ given by $\tau := \prod_{p \in \mathbb{P} \cup \{\infty\}} \tau_p$, where

$$\tau_p = \Big(1 - \frac{1}{p}\Big) \cdot \lim_{n \to \infty} \frac{\#\mathscr{V}(\mathbb{Z}/p^n\mathbb{Z})}{p^{dn}}$$

for p finite and

$$\tau_\infty = \frac{1}{2} \int\limits_{\substack{x \in [-1,1]^{d+2} \\ f(x)=0}} \frac{1}{\|(\operatorname{grad} f)(x)\|_2}\, dS\,.$$

Here, $\mathscr{V} \subset \mathbf{P}^{d+1}_{\mathbb{Z}}$ is the integral model of V defined by the polynomial f, and dS denotes the usual hypersurface measure on the cone $C_V(\mathbb{R})$, considered as a hypersurface in $\mathbb{R}^{d+2}$.

Note that the constant τ is invariant under scaling. When we multiply f by a prime number p, then τ_p gets multiplied by a factor of p. On the other hand, τ_∞ gets multiplied by a factor of $1/p$, and all the other factors remain unchanged.

Known cases. Conjecture 1 is established for smooth complete intersections of multidegree $d_1, \dots, d_n$ in the case that the dimension of V is very large compared to $d_1, \dots, d_n$ [Bi]. Further, it has been proven for projective spaces and quadrics. Finally, there are a number of further special cases in which Manin's conjecture is known to be true. See, e.g., [Pe, Section 4].

Recently, numerical evidence for Conjecture 1 has been presented in the case of the threefolds $V^e_{a,b}$ given by $ax^e = by^e + z^e + v^e + w^e$ in $\mathbf{P}^4_{\mathbb{Q}}$ for $e = 3$ and 4 [EJ1].

The smallest point. It would be desirable to have an a priori upper bound for the height of the smallest $\mathbb{Q}$-rational point on V, since this would allow us to effectively decide whether $V(\mathbb{Q}) \neq \emptyset$.

When V is a conic, Legendre's theorem on zeros of ternary quadratic forms yields an effective bound for the smallest point. For quadrics of arbitrary dimension, the same is true by an observation due to J. W. S. Cassels [Ca]. Further, there is a theorem of C. L. Siegel [Sg, Satz 1] that provides a generalization to hypersurfaces defined by norm equations. This certainly includes some special cubic surfaces, but in general, no theoretical upper bound is known for the height of the smallest $\mathbb{Q}$-rational point on a cubic surface.

Remark 2. If one had an error term [S-D] for $(*)$ uniform over all cubic surfaces V of Picard rank 1, then this would imply that the height $\mathrm{m}(V)$ of the smallest $\mathbb{Q}$-rational point is always less than $\frac{C}{\tau(V)^\alpha}$ for certain constants $\alpha > 1$ and $C > 0$.

The investigations on quartic threefolds made in [EJ2] indicate that one might have even $\mathrm{m}(V) < \frac{C(\varepsilon)}{\tau(V)^{1+\varepsilon}}$ for any $\varepsilon > 0$. Assuming equidistribution, one would expect that the height of the smallest $\mathbb{Q}$-rational point on V should even be $\sim\frac{1}{\tau(V)}$. An inequality of the form $\mathrm{m}(V) < \frac{C}{\tau(V)}$ is, however, known to be wrong in a similar situation (cf. [EJ2, Theorem 2.2]).

The results. We consider two families of cubic surfaces that are produced by a random-number generator. For each of these surfaces, we do the following:

(i) We verify that the Galois group acting on the 27 lines is equal to $W(E_6)$.

(ii) We compute E. Peyre's constant $\tau(V)$.

(iii) Up to a certain bound for the anticanonical height, we count all $\mathbb{Q}$-rational points on the surface V.

Thereby, we establish the Hasse principle for each of the surfaces considered. Further, we test numerically the conjecture of Manin, in the refined form due to E. Peyre, on the asymptotics of points of bounded height. Finally, we study the behavior of the height of the smallest $\mathbb{Q}$-rational point versus E. Peyre's constant. This means that we test the estimates formulated in Remark 2.

2 Background

27 lines. Recall that a nonsingular cubic surface defined over $\overline{\mathbb{Q}}$ contains exactly 27 lines. The symmetries of the configuration of the 27 lines respecting the intersection pairing are given by the Weyl group $W(E_6)$ [Ma, Theorem 23.9.ii].

Fact 3. *Let V be a smooth cubic surface defined over $\mathbb{Q}$ and let K be the field of definition of the 27 lines on V. Then K is a Galois extension of $\mathbb{Q}$. The Galois group $\mathrm{Gal}(K/\mathbb{Q})$ is a subgroup of $W(E_6)$.*

Remark 4. $W(E_6)$ contains a subgroup U of index two that is isomorphic to the simple group of order 25 920. It is of Lie type $B_2(\mathbb{F}_3)$, i.e., $U \cong \Omega_5(\mathbb{F}_3) \subset \mathrm{SO}_5(\mathbb{F}_3)$.

Remark 5. The operation of $W(E_6)$ on the 27 lines gives rise to a transitive permutation representation $\iota\colon W(E_6) \to S_{27}$. It turns out that the image of ι is contained in the alternating group A_{27}. We will call an element $\sigma \in W(E_6)$ *even* if $\sigma \in U$ and *odd* otherwise. This should not be confused with the sign of $\iota(\sigma) \in S_{27}$, which is always even.

The Brauer–Manin obstruction. For Fano varieties, all known obstructions against the Hasse principle are explained by the following observation.

Observation 6 (Manin)**.** *Let V be a nonsingular variety over $\mathbb{Q}$. Choose an element $\alpha \in \mathrm{Br}(V)$ [Ma, Definition 41.3]. Then, any $\mathbb{Q}$-rational point $x \in V(\mathbb{Q})$ gives rise to an adelic point $(x_\nu)_\nu \in V(\mathbf{A}_\mathbb{Q})$ satisfying the condition*

$$\sum_{\nu \in \mathrm{Val}(\mathbb{Q})} \mathrm{inv}(\alpha|_{x_\nu}) = 0\,.$$

Here, $\mathrm{inv}\colon \mathrm{Br}(\mathbb{Q}_\nu) \to \mathbb{Q}/\mathbb{Z}$ *(respectively* $\mathrm{inv}\colon \mathrm{Br}(\mathbb{R}) \to \frac{1}{2}\mathbb{Z}/\mathbb{Z}$*) denotes the canonical isomorphism.*

The local invariant $\mathrm{inv}(\alpha|_{x_\nu})$ depends continuously on $x_\nu \in V(\mathbb{Q}_\nu)$. Further, Yu. I. Manin proved [Ma, Corollary 44.2.5] that for each nonsingular variety V over $\mathbb{Q}$, there exists a finite set $S \subset \mathrm{Val}(\mathbb{Q})$ such that $\mathrm{inv}(\alpha|_{x_\nu}) = 0$ for every $\alpha \in \mathrm{Br}(V)$, $\nu \notin S$, and $x_\nu \in V(\mathbb{Q}_\nu)$. This implies that the Brauer–Manin obstruction, if present, is an obstruction to the principle of weak approximation.

Denote by $\pi\colon V \to \mathrm{Spec}(\mathbb{Q})$ the structural map. It is obvious that altering $\alpha \in \mathrm{Br}(V)$ by some Brauer class $\pi^*\rho$ for $\rho \in \mathrm{Br}(\mathbb{Q})$ does not change the obstruction defined by α. In consequence, it is only the factor group $\mathrm{Br}(V)/\pi^*\mathrm{Br}(\mathbb{Q})$ that is relevant for the Brauer–Manin obstruction. The latter is canonically isomorphic to $H^1(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), \mathrm{Pic}(V_{\overline{\mathbb{Q}}}))$ [Ma, Lemma 43.1.1]. In particular, if $H^1(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), \mathrm{Pic}(V_{\overline{\mathbb{Q}}})) = 0$ then there is no Brauer–Manin obstruction on V.

For a smooth cubic surface V, the geometric Picard group $\mathrm{Pic}(V_{\overline{\mathbb{Q}}})$ is generated by the classes of the 27 lines on $V_{\overline{\mathbb{Q}}}$. Its first cohomology group can be described in terms of the Galois action on these lines. Indeed, there is a canonical isomorphism [Ma, Proposition 31.3]

$$(+) \qquad H^1(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), \mathrm{Pic}(V_{\overline{\mathbb{Q}}})) \cong \mathrm{Hom}((NF \cap F_0)/NF_0, \mathbb{Q}/\mathbb{Z}).$$

Here, $F \subset \mathrm{Div}(V_{\overline{\mathbb{Q}}})$ is the group generated by the 27 lines, $F_0 \subset F$ denotes the subgroup of principal divisors, and N is the norm map under the operation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})/H$, H being the stabilizer of F.

Remark 7. Consider the particular case in which the Galois group acts transitively on the 27 lines. Then (+) shows that $H^1(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), \mathrm{Pic}(V_{\overline{\mathbb{Q}}})) = 0$. In particular, there is no Brauer–Manin obstruction in this case.

It is expected that the Hasse principle holds for all cubic surfaces such that $H^1(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), \mathrm{Pic}(V_{\overline{\mathbb{Q}}})) = 0$. (See [CS, Conjecture C].)

3 Computation of the Galois group

Let V be a smooth cubic surface defined over $\mathbb{Q}$ and let K be the field of definition of the 27 lines on V. By Fact 3, $K/\mathbb{Q}$ is a Galois extension and the Galois group $G := \mathrm{Gal}(K/\mathbb{Q})$ is a subgroup of $W(E_6)$. For general cubic surfaces, G is actually equal to $W(E_6)$. To verify this for particular examples, the following lemma is useful.

Lemma 8. *Let $H \subseteq W(E_6)$ be a subgroup that acts transitively on the 27 lines and contains an element of order five. Then, either H is the subgroup $U \subset W(E_6)$ of index two or $H = W(E_6)$.*

Proof. $H \cap U$ still acts transitively on the 27 lines and still contains an element of order five. Thus, we may suppose $H \subseteq U$.

Assume that $H \subsetneq U$. Denote by k the index of H in U. The natural action of U on the set of cosets U/H yields a permutation representation $i \colon U \to S_k$. Since U is simple, i is necessarily injective. In particular, since $\#U \nmid 8!$, we see that $k > 8$. Let us consider the stabilizer $H' \subset H$ of one of the lines. Since H acts transitively, it follows that $\#H' = \frac{\#H}{27} = \frac{\#U}{27 \cdot k} = \frac{960}{k}$. We distinguish two cases.

First case: $k > 16$. Then, $k \geq 20$ and $\#H' \leq 48$. This implies that the 5-Sylow subgroup is normal in H'. Its conjugate by some $\sigma \in H$ therefore depends only on $\overline{\sigma} \in H/H'$. In consequence, the number n of 5-Sylow subgroups in H is a divisor of $\#H/\#H' = 27$. Sylow's congruence $n \equiv 1 \pmod 5$ yields that $n = 1$.

Let $H_5 \subset H$ be the 5-Sylow subgroup. Then, $\iota(H_5) \subset S_{27}$ is generated by a product of disjoint 5-cycles leaving at least two lines fixed. It is, therefore, not normal in the transitive group $\iota(H)$. This is a contradiction.

Second case: $9 \leq k \leq 16$. We have $k \,|\, 960$. On the other hand, the assumption $5 \,|\, \#H$ implies $5 \nmid k$. This shows that there are only two possibilities: $k = 12$ and $k = 16$. Since in U there is no subgroup of index eight or less, $H \subset U$ must be a maximal subgroup. In particular, the permutation representation $i \colon U \to S_k$ is primitive.

Primitive permutation representations of degree up to 20 were classified already in the late nineteenth century. It is well known that no group of order 25 920 allows a faithful primitive permutation representation of degree 12 or 16 [Sm, Table 1]. □

Remark 9. The subgroups of the simple group U were completely classified by L. E. Dickson [Di] in 1904. It would not be complicated to deduce the lemma from Dickson's list.

Let the smooth cubic surface V be given by a homogeneous equation $f = 0$ with integral coefficients. We want to compute the Galois group G.

An affine part of a general line ℓ can be described by four coefficients a, b, c, d via the parametrization

$$\ell \colon t \mapsto (1 : t : (a + bt) : (c + dt)),$$

where ℓ is contained in S if and only if it intersects S in at least four points. This implies that

$$f(\ell(0)) = f(\ell(\infty)) = f(\ell(1)) = f(\ell(-1)) = 0$$

is a system of equations for a, b, c, d that encodes that ℓ is contained in S.

By a Gröbner base calculation in `SINGULAR`, we compute a univariate polynomial g of minimal degree belonging to the ideal generated by the equations. If g is of degree 27 then the splitting field of g is equal to the field K of definition of the 27 lines on V. We then use van der Waerden's criterion [PZ, Proposition 2.9.35]. More precisely, our algorithm works as follows.

Algorithm 10 (Verifying $G = W(E_6)$**).** Given the equation $f = 0$ of a smooth cubic surface, this algorithm verifies $G = W(E_6)$.

(i) Compute a univariate polynomial $0 \neq g \in \mathbb{Z}[d]$ of minimal degree such that

$$g \in (f(\ell(0)), f(\ell(\infty)), f(\ell(1)), f(\ell(-1))) \subset \mathbb{Q}[a, b, c, d],$$

where $\ell \colon t \mapsto (1 : t : (a + bt) : (c + dt))$.

If g is not of degree 27 then terminate with an error message. In this case, the coordinate system for the lines is not sufficiently general. If we are erroneously given a singular cubic surface then the algorithm will fail at this point.

(ii) Factor g modulo all primes below a given limit. Ignore the primes dividing the leading coefficient of g.

(iii) If one of the factors is composite, then go to the next prime immediately. Otherwise, check whether the decomposition type corresponds to one of the cases listed below:

$$A := \{(9,9,9)\}, \qquad B := \{(1,1,5,5,5,5,5),(2,5,5,5,10)\},$$
$$C := \{(1,4,4,6,12),(2,5,5,5,10),(1,2,8,8,8)\}.$$

(iv) If each of the cases occurred for at least one of the primes then output the message "The Galois group is equal to $W(E_6)$" and terminate.

Otherwise, output "Cannot prove that the Galois group is equal to $W(E_6)$."

Remark 11. The cases above function as follows:

(a) Case B shows that the order of the Galois group is divisible by five.

(b) Cases A and B together guarantee that g is irreducible. Therefore, by Lemma 8, cases A and B prove that G contains the index-two subgroup $U \subset W(E_6)$.

(c) Case C is a selection of the most frequent odd conjugacy classes in $W(E_6)$.

Remark 12. One could replace cases B and C by their common element $(2,5,5,5,10)$. This would lead to a simpler but less efficient algorithm.

Remark 13. Actually, a decomposition type as considered in step (iii) does not always represent a single conjugacy class in $W(E_6)$. Two elements $\iota(\sigma)$, $\iota(\sigma') \in S_{27}$ might be conjugate in S_{27} via a permutation $\tau \notin \iota(W(E_6))$.

For example, as is easily seen using `GAP`, the decomposition type $(3,6,6,6,6)$ falls into three conjugacy classes, two of which are even and one odd (cf. Remark 4). However, all the decomposition types sought in Algorithm 10 do represent single conjugacy classes.

Remark 14. Since we expect $G = W(E_6)$, we can estimate the probability of each case by the Čebotarev density theorem. Case A has a probability of $\frac{1}{9}$. This is the lowest value among the three cases.

Remark 15. Since we do not use the factors of $\overline{g}$ explicitly, it is enough to compute their degrees and to check that each of them occurs with multiplicity one. This means that we have only to compute $\gcd(\overline{g}(X), \overline{g}'(X))$ and $\gcd(\overline{g}(X), X^{p^d} - X)$ in $\mathbb{F}_p[X]$ for $d = 1, 2, \ldots, 13$ [Co, Algorithms 3.4.2 and 3.4.3].

4 Computation of Peyre's constant

The Euler product. We want to compute the product over all τ_p. For a finite place p, we have

$$\tau_p = \left(1 - \frac{1}{p}\right) \cdot \lim_{n \to \infty} \frac{V(\mathbb{Z}/p^n\mathbb{Z})}{p^{2n}}\,.$$

If the reduction $V_{\mathbb{F}_p}$ is smooth, then the sequence under the limit is constant by virtue of Hensel's lemma. Otherwise, it becomes stationary after finitely many steps.

We approximate the infinite product over all τ_p by the finite product taken over the primes less than 300. Numerical experiments show that the contributions of larger primes do not lead to a significant change. (Compare the values calculated for the concrete example in Section 6.)

The factor at the infinite place. We want to compute

$$\tau_\infty = \frac{1}{2} \int\limits_R \frac{1}{\|\operatorname{grad} f\|_2}\, dS,$$

where the domain of integration is given by

$$R = \{(x, y, z, w) \in [-1, 1]^4 \mid f(x, y, z, w) = 0\}\,.$$

Here, dS denotes the usual hypersurface measure on R, considered as a hypersurface in $\mathbb{R}^4$. Thus, τ_∞ is given by a three-dimensional integral.

Since f is a homogeneous polynomial, we may reduce to an integral over the boundary of R, which is a two-dimensional domain. In our particular case, we have $\deg f = 3$. Then, a direct computation leads to

$$\begin{aligned}\tau_\infty = &\int\limits_{R_0} \frac{1}{\|(\frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}, \frac{\partial f}{\partial w})\|_2}\, dA + \int\limits_{R_1} \frac{1}{\|(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial z}, \frac{\partial f}{\partial w})\|_2}\, dA \\ &+ \int\limits_{R_2} \frac{1}{\|(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial w})\|_2}\, dA + \int\limits_{R_3} \frac{1}{\|(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z})\|_2}\, dA,\end{aligned}$$

where the domains of integration are

$$R_i = \{(x_0, x_1, x_2, x_3) \in [-1, 1]^4 \mid x_i = 1 \text{ and } f(x_0, x_1, x_2, x_3) = 0\}\,,$$

and dA denotes the two-dimensional hypersurface measure on R_i, considered as a hypersurface in $\mathbb{R}^3$.

We therefore have to integrate a smooth function over a compact part of a smooth two-dimensional submanifold in $\mathbb{R}^3$. To do this, we approximate the domain of integration by a triangular mesh.

Algorithm 16 (Generating a triangular mesh**).** Given the equation $f = 0$ of a smooth surface in $\mathbb{R}^3$, this algorithm constructs a triangular mesh approximating the part of the surface that is contained in a given cube.

(i) We split the cube into eight smaller cubes and iterate this procedure a predefined number of times, recursively. During recursion, we exclude those cubes that obviously do not intersect the manifold. To do this, we estimate $\|\operatorname{grad} f\|_2$ on each cube.

(ii) Then each resulting cube is split into six simplices.

(iii) For each edge of each simplex that intersects the manifold, we compute an approximation of the point of intersection. We use them as the vertices of the triangles to be constructed. This leads to a mesh consisting of one or two triangles per simplex.

The next step is to compute the contribution of each triangle Δ_i to the integral. For this, we use an adaptation of the midpoint rule. We approximate the integrand g by its value $g(C_i)$ at the barycenter C_i of the triangle. Note that this point usually lies outside the surface given by $f = 0$. Algorithm 16 guarantees only that the three vertices of each facet are contained in that surface.

The product $g(C_i)A(\Delta_i)$ seems to be a reasonable approximation of the contribution of Δ_i to the integral. We correct by an additional factor, the cosine of the angle between the normal vector of the triangle and the gradient vector $\operatorname{grad} f$ at the barycenter C.

Remark 17. In our application, these correctional factors are close to 1 and seem to converge to 1 when the number of recursions is growing. This is, however, not a priori clear. H. A. Schwarz's cylindrical surface [Sch] constitutes a famous example of a sequence of triangulations in which the triangles become arbitrarily small and the factors are nevertheless necessary for correct integration.

We use the method described above to approximate the value of τ_∞. In Algorithm 16, we work with six recursions.

Remark 18. Our method of numerical integration is a combination of standard algorithms for 2.5-dimensional mesh generation and two-dimensional integration that are described in the literature [Hb].

On a triangle, we integrate linear functions correctly. This indicates that the method should converge to second order. The facts that we work with the area of a linearized triangle and that the barycenters C_i are located at a certain distance from the manifold generate errors of the same order.

5 Numerical Data

The computations carried out. A general cubic surface is described by twenty coefficients. With current technology, it is impossible to study all cubic surfaces with coefficients below a given bound. For that reason, we decided to work with coefficient vectors provided by a random-number generator. Our first sample consists of 20 000 surfaces with coefficients randomly chosen in the interval $[0, 50]$. The second sample consists of 20 000 surfaces with randomly chosen coefficients from the interval $[-100, 100]$.

These limits were, of course, chosen somewhat arbitrarily. There is, at least, some reason not to work with too large limits, since this would lead to low values of τ. (The reader might want to compare [EJ2, Theorem 3.3.4], where this is rigorously proven in a different situation.) Low values of τ are undesirable, since they require high search bounds in order to satisfactorily test Manin's conjecture.

We verified explicitly that each of the surfaces studied is smooth. For this, we inspected a Gröbner basis of the ideal corresponding to the singular locus. The computations were done in `SINGULAR`.

Then, using Algorithm 10, we proved that for each surface, the full Galois group $W(E_6)$ acts on the 27 lines. The largest prime used was 457. This means that all our examples are general from the Galois point of view. In consequence, their Picard ranks are equal to 1. Further, according to Remark 7, the Brauer–Manin obstruction is not present on any of the surfaces considered.

Almost as a byproduct, we verified that no two of the 40 000 surfaces are isomorphic. Actually, when running part (ii) of Algorithm 10, we wrote the decomposition types found into a file. Primes at which the algorithm failed were labeled with a special marker. A program, written in `C`, ran in an iterated loop over all pairs of surfaces and looked for a prime at which the decomposition types differ. The largest prime needed to distinguish two surfaces was 73.

We counted all $\mathbb{Q}$-rational points of height less than 250 on the surfaces of the first sample. It turns out that on two of these surfaces, there are no $\mathbb{Q}$-rational points occurring, since the equation is unsolvable in $\mathbb{Q}_p$ for some small p. In this situation, Manin's conjecture is true, trivially. On each of the remaining surfaces, we found at least one $\mathbb{Q}$-rational point; 228 examples contained fewer than ten points. On the other hand, 1213 examples contained at least 100 $\mathbb{Q}$-rational points. The largest number of points found was 335.

For the second sample, the search bound was 500. Again, on two of these surfaces, there are no $\mathbb{Q}$-rational points occurring, since the equation is unsolvable in a certain $\mathbb{Q}_p$. There were 202 examples containing between one and nine points, while 1857 examples contained at least 100 $\mathbb{Q}$-rational points. The largest number of points found was 349.

To find the $\mathbb{Q}$-rational points, we used a 2-adic search method that works as follows. Let a cubic surface V be given. Then, in a first step, we determined on V all points defined over $\mathbb{Z}/512\mathbb{Z}$ (respectively $\mathbb{Z}/1024\mathbb{Z}$). Then, for each of the points found we checked which of its lifts to $\mathbf{P}^3(\mathbb{Z})$ actually lie on V. This leads to an $\mathrm{O}(B^3)$-algorithm that may be efficiently implemented in `C`.

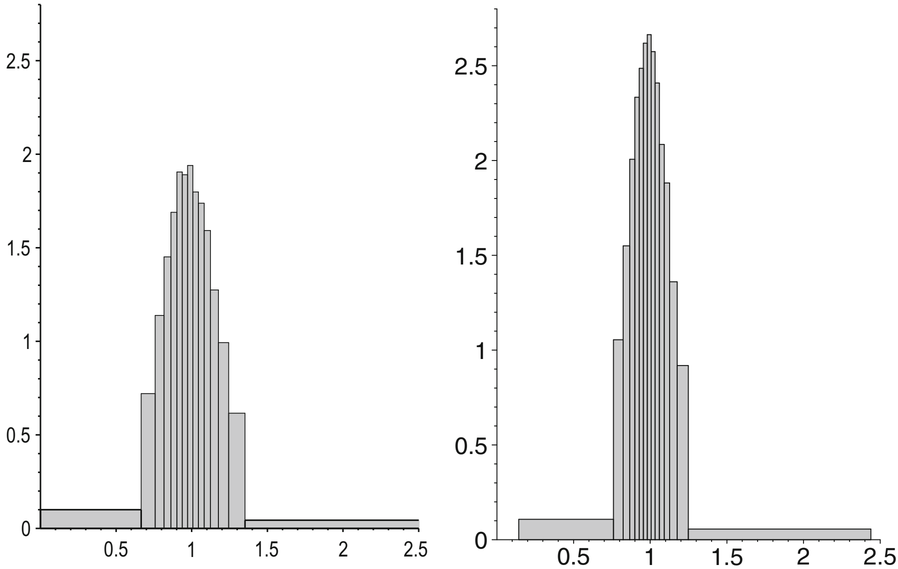
There are algorithms which are asymptotically faster, for example Elkies' method, which is $\mathrm{O}(B^2)$ and implemented in `Magma`. A practical comparison shows, however, that Elkies' method is not yet faster for our relatively low search bounds.

Furthermore, using the method described in Section 4, we computed an approximation of Peyre's constant for each surface.

The density results. For each of the surfaces considered we calculated the quotient

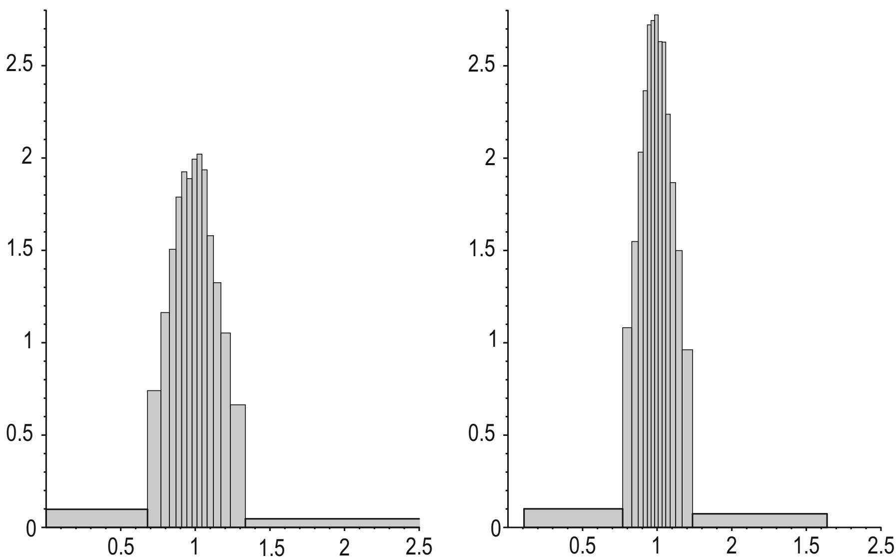
$$\#\{\text{ points of height} < B \text{ found}\,\} \,/\, \#\{\text{ points of height} < B \text{ expected}\,\}.$$

Let us visualize the distribution of the quotients by some histograms, shown in Figures 1 and 2.



First sample, $B = 125$ First sample, B= 250

Fig. 1. Distribution of the quotients for the first sample.



Second sample, $B = 250$ Second sample, $B = 500$

Fig. 2. Distribution of the quotients for the second sample.

Some statistical parameters are shown in Tables 1 and 2.

Table 1. Parameters of the distribution for the first sample.

search bound	125	250
mean	0.999 93	0.998 87
standard deviation	0.235 58	0.169 25

Table 2. Parameters of the distribution for the second sample.

search bound	250	500
mean	1.000 93	0.999 43
standard deviation	0.225 27	0.161 58

The results for the smallest point. For each of the surfaces in our samples, we determined the height $\mathrm{m}(V)$ of its smallest point. We visualize the behavior of $\mathrm{m}(V)$ in the diagrams of Figure 3.

At first glance, it looks very natural to consider the distribution of the values of $\mathrm{m}(V)$ versus the Tamagawa type number $\tau(V)$. In view of the inequalities asked for in the introduction, it seems, however, to be better to make a slight modification and plot the product $\mathrm{m}(V)\tau(V)$ instead of $\mathrm{m}(V)$ itself.

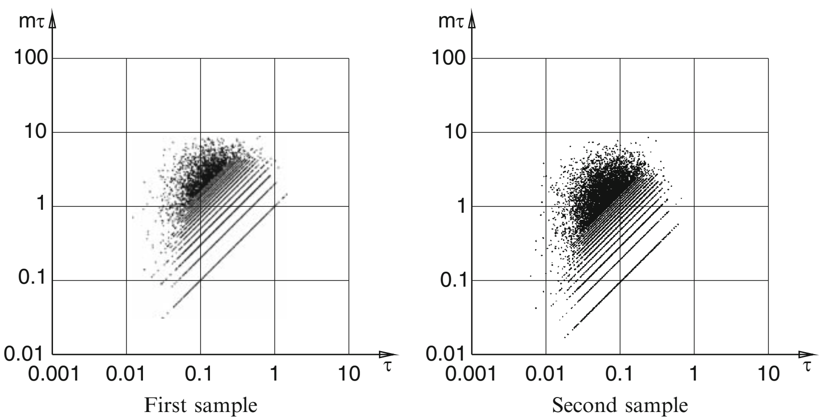


Fig. 3. The smallest height of a rational point versus the Tamagawa number.

Conclusion. Our experiments suggest that for general cubic surfaces V over $\mathbb{Q}$, the following assertions hold:

(i) There are no obstructions to the Hasse principle.

(ii) Manin's conjecture is true in the form refined by E. Peyre.

Further, it is apparent from the diagrams in Figure 3 that the experiment agrees with the expectation for the heights of the smallest points formulated in Remark 2 above. Indeed, for both samples, a line tangent to the top of the scatter plot is nearly horizontal. This indicates that even the strong form of the estimate should be true, i.e., $\mathrm{m}(V) < \frac{C(\varepsilon)}{\tau(V)^{1+\varepsilon}}$ for any $\varepsilon > 0$.

Running times. The largest portion of the running time was spent on the calculation of the Euler products. It took 20 days of CPU time to calculate all 40 000 Euler products for $p < 300$. For comparison, we estimated all the integrals, using six recursions, within 36 hours. Further, it took eight days to systematically search for all points of height less than 500 on the surfaces of the second sample. Search for points of height less than 250 on the surfaces of the first sample took only one day.

In running Algorithm 10, the lion's share of the time was used for the computation of the univariate degree-27 polynomials. This took approximately seven days of CPU time. In comparison with that, all other parts were negligible. It took only twelve minutes to ensure that all 40 000 surfaces are smooth. The `C` program verifying that no two of the surfaces are isomorphic to each other ran approximately 80 seconds.

6 A concrete example

An example. Let us conclude this article with some results on the particular cubic surface V given by

$$x^3 + 2xy^2 + 11y^3 + 3xz^2 + 5y^2w + 7zw^2 = 0. \tag{$-$}$$

Example $(-)$ was not among the surfaces produced by the random-number generator. Our intention is just to present the output of our algorithms in a specific (and not too artificial) example and, most notably, to show the intermediate results of Algorithm 10.

A Gröbner basis calculation in `Magma` shows that V has bad reduction at $p = 2$, 3, 7, 23, and 22 359 013 270 232 677. The idea behind that calculation is the same as described above for the verification of smoothness. The only difference is that we consider Gröbner bases over $\mathbb{Z}$ instead of $\mathbb{Q}$.

The Galois group. The first step of Algorithm 10 works well on V, i.e., the polynomial g is indeed of degree 27. Its coefficients become rather large. The one of greatest magnitude is that of d^{13}. It is equal to 38 300 982 629 255 010. The leading coefficient of g is $5^3 \cdot 7^{12}$. We find case A at $p = 373$. The common decomposition type $(2, 5, 5, 5, 10)$ of cases B and C occurs at $p = 19$, 31, 59, 61, 191, 199, and 223.

Consequently, V is an explicit example of a smooth cubic surface over $\mathbb{Q}$ admitting the property that the Galois group that acts on the 27 lines is equal to $W(E_6)$.

Remark 19. The first such examples were constructed by T. Ekedahl [Ek, Theorem 2.1].

Remark 20. Our example $(-)$ is different from Ekedahl's. Indeed, in Ekedahl's examples, the Frobenius Frob_{11} acts on the 27 lines as an element of the conjugacy class $C_{15} \subset W(E_6)$ (in Sir P. Swinnerton-Dyer's numbering). In our case, however, the first two steps of Algorithm 10 show that Frob_{11} yields the decomposition type $(1,1,1,1,1,2,4,4,4,4,4)$. This corresponds to the class C_{18} [Ma, §31, Table 1]. Note that Ekedahl's examples, as well as ours, have good reduction at $p = 11$.

Computation of Peyre's constant. As an approximation of the Euler product, we get

$$\prod_{p<300} \tau_p \approx 0.729\,750.$$

Using the Lefschetz trace formula, we calculated all partial products of this particular Euler product up to $p < 40\,000$. The oscillations, we observed, remain within a distance of less than two percent. For example, we obtained

$$\prod_{p<40\,000} \tau_p \approx 0.731\,732.$$

For the factor at the infinite place, we get, using six recursions,

$$\tau_\infty \approx 1.786\,726.$$

We list several approximate values in Table 3

Table 3. Approximate values of τ_∞.

recursions	3	4	5	6	7	8
approx. of τ_∞	1.780 729	1.785 147	1.786 453	**1.786 726**	1.786 800	1.786 820
Δ		0.004 418	0.001 306	0.000 273	0.000 074	0.000 020

The successive differences decline by a factor of close to four from one step to the next. This conforms to second order convergence expected for our method of numerical integration.

Altogether, E. Peyre's constant is approximately $\tau \approx 1.3074$.

Rational points. There are 345 $\mathbb{Q}$-rational points on V of height less than 250 and 693 $\mathbb{Q}$-rational points of height less than 500. The smallest points are $(0:0:1:0)$ and $(0:0:0:1)$. The smallest nonobvious point is $(1:2:(-3):(-2))$. A complete list of all $\mathbb{Q}$-rational points on V of height up to 20 is presented in Table 4.

Table 4. Points on V of height ≤ 20.

Point	Height	Point	Height
$(0:0:0:1)$	1	$(4:-6:-13:2)$	13
$(0:0:1:0)$	1	$(5:5:0:-14)$	14
$(1:2:-3:-2)$	3	$(10:-14:12:11)$	14
$(3:-2:-3:2)$	3	$(0:7:-16:7)$	16
$(0:7:-6:-7)$	7	$(16:-8:3:-4)$	16
$(0:4:-3:8)$	8	$(6:-9:16:3)$	16
$(5:-5:0:8)$	8	$(12:7:-6:17)$	17
$(2:-8:8:7)$	8	$(14:-9:-2:17)$	17
$(10:-5:0:-1)$	10	$(6:-3:-18:7)$	18
$(0:5:0:-11)$	11	$(9:-6:-1:18)$	18
$(8:6:-11:-8)$	12	$(3:6:-19:-6)$	19
$(12:-6:-4:-3)$	12	$(8:7:-4:19)$	19
$(9:-12:9:10)$	12		

The field of definition of the 27 lines. Having done the Gröbner basis calculation in Algorithm 10(i), the 27 lines may be computed at high precision. This allows us to find the 45 triangles on V, explicitly. We calculated a degree-45 resolvent G of the degree-27 polynomial g, the zeros of which are all the sums $a_{i_1}+a_{i_2}+a_{i_3}$ for $\ell_{i_1}, \ell_{i_2}, \ell_{i_3}$ representing three lines that form a triangle. Here, $\ell_i\colon t \mapsto (1:t:(a_i+b_it):(c_i+d_it))$ denote parametrizations of the 27 lines. Since $G \in \mathbb{Z}[X]$, our floating-point calculation is in fact exact.

Proposition 21. *The unique quadratic subfield in the field K of definition of the 27 lines on V is $\mathbb{Q}(\sqrt{-23\cdot 22\,359\,013\,270\,232\,677})$.*

Proof. K is unramified at all places of good reduction of V. This leaves us with only $2^6-1=63$ possibilities for the quadratic subfield $\mathbb{Q}(\sqrt{d})$. To exclude 62 of them is algorithmically easy.

Indeed, for a good prime p, there is a way to compute $\left(\frac{d}{p}\right)$ without knowledge of d. We factor the degree-45 resolvent G modulo p. If p divides the leading coefficient or there are multiple factors, then we get no answer. Otherwise, $\left(\frac{d}{p}\right)=\pm1$ depending on whether the decomposition type found is even or odd in S_{45}.

It turns out that it is sufficient to do this for $p = 13$, 17, 19, 29, 31, and 53. □

Proposition 22. *The field extension $K/\mathbb{Q}$ is ramified exactly at $p=2$, 3, 7, 23, and* 22 359 013 270 232 677.

Proof. It remains to verify ramification at $p = 2$, 3, and 7. For that, we computed in `magma` the p-adic factorization of g. The decomposition types are $(3, 24)$ for $p=2$ and 3 and $(1,1,1,4,4,4,4,8)$ for $p=7$.

Let Z_p be the decomposition field of p. If p were unramified then $\mathrm{Gal}(K/Z_p)$ would be a cyclic group, i.e., $\mathrm{Gal}(K/Z_p) = \langle\sigma\rangle$ for some $\sigma \in W(E_6)$. On the other hand, on the 27 lines, the orbit structure under the operation of $\mathrm{Gal}(K/Z_p)$ is the same as under the operation of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$. There is, however, no element in $W(E_6)$ that yields the decomposition type $(3, 24)$ or $(1, 1, 1, 4, 4, 4, 4, 8)$ [Ma, §31, Table 1]. $\square$

Remark 21. This shows that there is no integral model of V that is smooth over $p = 2$, 3, 7, 23, or 22 359 013 270 232 677.

References

[Bi] Birch, B. J.: *Forms in many variables,* Proc. Roy. Soc. Ser. A **265** (1961/1962), 245–263.

[Ca] Cassels, J. W. S.: *Bounds for the least solutions of homogeneous quadratic equations,* Proc. Cambridge Philos. Soc. **51** (1955), 262–264.

[Co] Cohen, H.: *A course in computational algebraic number theory,* Springer, Berlin, Heidelberg 1993.

[CS] Colliot-Thélène, J.-L. and Sansuc, J.-J.: *On the Chow groups of certain rational surfaces: A sequel to a paper of S. Bloch,* Duke Math. J. **48** (1981), 421–447.

[Di] Dickson, L. E.: *Determination of all the subgroups of the known simple group of order* 25 920, Trans. Amer. Math. Soc. **5** (1904), 126–166.

[Ek] Ekedahl, T.: *An effective version of Hilbert's irreducibility theorem,* in: Séminaire de Théorie des Nombres, Paris 1988–1989, Prog. Math. 91, Birkhäuser, Boston 1990, 241–249.

[EJ1] Elsenhans, A.-S. and Jahnel, J.: *The Asymptotics of Points of Bounded Height on Diagonal Cubic and Quartic Threefolds,* in: Algorithmic number theory, Lecture Notes in Computer Science 4076, Springer, Berlin 2006, 317–332.

[EJ2] Elsenhans, A.-S. and Jahnel, J.: *On the smallest point on a diagonal quartic threefold,* J. Ramanujan Math. Soc. **22** (2007), 189–204.

[FMT] Franke, J., Manin, Y. I., and Tschinkel, Y.: *Rational points of bounded height on Fano varieties,* Invent. Math. **95** (1989), 421–435.

[Hb] *Handbook of numerical analysis,* edited by P. G. Ciarlet and J. L. Lions, Vols. II–IV, North-Holland Publishing Co., Amsterdam 1991–1996.

[He] Hermann, G.: *Die Frage der endlich vielen Schritte in der Theorie der Polynomideale,* Math. Ann. **95** (1926), 736–788.

[Ma] Manin, Yu. I.: *Cubic forms, algebra, geometry, arithmetic,* North-Holland Publishing Co. and American Elsevier Publishing Co., Amsterdam, London, and New York 1974.

[Pe] Peyre, E.: *Points de hauteur bornée et géométrie des variétés (d'après Y. Manin et al.),* Séminaire Bourbaki 2000/2001, Astérisque **282** (2002), 323–344.

[PZ] Pohst, M. and Zassenhaus, H.: *Algorithmic Algebraic Number Theory,* Cambridge University Press, Cambridge 1989.

[Sch] Schwarz, H. A.: *Sur une définition erronée de l'aire d'une surface courbe,* Communication faite à M. Charles Hermite, 1881–82, in: Gesammelte Mathematische Abhandlungen, Zweiter Band, Springer, Berlin 1890, 309–311.

[Sg] Siegel, C. L.: *Normen algebraischer Zahlen,* Nachr. Akad. Wiss. Göttingen, Math.-Phys. Kl. II **1973** (1973), 197–215.

[Sm] Sims, C. C.: *Computational methods in the study of permutation groups,* in: Computational Problems in Abstract Algebra (Proc. Conf., Oxford 1967), Pergamon, Oxford 1970, 169–183.

[S-D] Swinnerton-Dyer, Sir P.: *Counting points on cubic surfaces II,* in: Geometric methods in algebra and number theory, Progr. Math. 235, Birkhäuser, Boston 2005, 303–309.

Cluster Ensembles, Quantization and the Dilogarithm II: The Intertwiner

V.V. Fock[1] and A.B. Goncharov[2]

[1] Université Louis Pasteur, I.R.M.A., 7, rue René Descartes, 67084 Strasbourg Cedex, France, fock@math.u-strasbg.fr
[2] Department of Math, Brown University, Providence RI 02906, USA, sasha@math.brown.edu

Dedicated to Yu. I. Manin for his 70th Birthday

Summary. This paper is the second part of our paper "Cluster ensembles, quantization, and the dilogarithm" [FG2].[1] Its main result is a construction, by means of the quantum dilogarithm, of certain intertwiner operators, which play a crucial role in the quantization of the cluster $\mathcal{X}$-varieties and construction of the corresponding canonical representation.

Key words: quantum dilogarithm, cluster varieties, quantization

2000 Mathematics Subject Classifications: 17B37, 11G55, 16W35 81R10, 33E30

A cluster ensemble as defined in [FG2] is a pair $(\mathcal{A}, \mathcal{X})$ of schemes over $\mathbb{Z}$, called cluster $\mathcal{A}$- and $\mathcal{X}$-varieties, related by a map $p : \mathcal{A} \longrightarrow \mathcal{X}$. The ring of regular functions on the cluster $\mathcal{A}$-variety is the upper cluster algebra [BFZ].

Cluster $\mathcal{A}$- and $\mathcal{X}$-varieties are glued from families of coordinatized split algebraic tori by means of certain subtraction-free rational transformations. In particular, it makes sense to consider the spaces of their positive real points, denoted by $\mathcal{A}^+$ and $\mathcal{X}^+$.

The cluster $\mathcal{X}$-variety has a Poisson structure, given in any cluster coordinate system $\{X_i\}$ by

$$\{X_i, X_j\} = \widehat{\varepsilon}_{ij} X_i X_j, \quad \widehat{\varepsilon}_{ij} \in \mathbb{Z}.$$

The Poisson tensor $\widehat{\varepsilon}_{ij}$ depends on the choice of coordinate system. There is a canonical noncommutative deformation $\mathcal{X}_q$ of the cluster $\mathcal{X}$-variety in the direction of this Poisson structure [FG2].

[1]Both papers were originally combined into a single ArXiv preprint math/0311245, version 1.

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*, Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_15,

The gluing procedure underlying the definition of a cluster variety can be understood as a functor from a certain groupoid, called the *cluster modular groupoid* $\mathcal{G}$, to a category of commutative algebras.

In Section 3 we suggest a $*$-quantization of the cluster $\mathcal{X}$-variety, understood as a functor from the groupoid $\mathcal{G}$ to the category of noncommutative topological $*$-algebras. More precisely, the coordinate systems on cluster varieties are parametrized by the objects $\mathbf{i}$ of the groupoid $\mathcal{G}$, called *seeds*. To each seed $\mathbf{i}$ we assign two coordinatized tori, $\mathcal{A}_{\mathbf{i}}$ and $\mathcal{X}_{\mathbf{i}}$. The algebra of smooth functions on the latter admits a canonical $\hbar$-deformation, given by a topological Heisenberg $*$-algebra $\mathcal{H}^{\hbar}_{\mathbf{i}}$. So to define a functor we need to relate these algebras for different seeds. We write the formulas relating the generators of the algebras $\mathcal{H}^{\hbar}_{\mathbf{i}}$, but do not specify the category of topological $*$-algebras. As a result, the $*$-quantization of the cluster $\mathcal{X}$-variety serves only as a motivation, and in Sections 3 we state claims instead of theorems when those unspecified topological algebras enter the formulations. We hope to have a precise version of Section 3. However, the rest of the paper does not depend on that, while motivations given in Section 3 clarify what we do next.

In Section 4 we proceed to a construction of the canonical unitary projective representation of the modular groupoid. It is realized in the Hilbert space $L^2(\mathcal{A}^+)$ assigned to the set of positive real points of the cluster $\mathcal{A}$-variety.

For each seed $\mathbf{i}$ there is a Hilbert space $L^2(\mathcal{A}^+_{\mathbf{i}})$ (which is canonically identified with $L^2(\mathcal{A}^+)$). In Section 4.1, the Heisenberg $*$-algebra $\mathcal{H}^{\hbar}_{\mathbf{i}}$ is represented by unbounded operators in $L^2(\mathcal{A}^+_{\mathbf{i}})$. In fact, a bigger algebra—the chiral double $\mathcal{H}^{\hbar}_{\mathbf{i}} \otimes \mathcal{H}^{\hbar}_{\mathbf{i}^{\circ}}$ of the Heisenberg $*$-algebra—acts by unbounded operators on the same Hilbert space.

The morphisms in the groupoid $\mathcal{G}$ are defined as compositions of certain elementary ones, called *mutations* and *symmetries*. Given a mutation $\mathbf{i} \to \mathbf{i}'$ we construct a unitary operator

$$\mathbf{K}_{\mathbf{i}\to\mathbf{i}'} : L^2(\mathcal{A}^+_{\mathbf{i}}) \longrightarrow L^2(\mathcal{A}^+_{\mathbf{i}'}).$$

It intertwines the actions of the Heisenberg $*$-algebras related to the seeds $\mathbf{i}$ and $\mathbf{i}'$[2]. (Similar intertwiners for symmetries are rather tautological). This construction is the main result of the paper. The operator $\mathbf{K}_{\mathbf{i}\to\mathbf{i}'}$ is characterized by its intertwining property uniquely up to a constant.

Certain compositions of mutations and symmetries are identity morphisms in the groupoid $\mathcal{G}$. So to get a representation of the modular groupoid we have to show that the corresponding compositions of the intertwiners $\mathbf{K}_{\mathbf{i}\to\mathbf{i}'}$ are multiples of the identity operators. This is a rather difficult problem. It is solved in [FG3], where we give another construction of the intertwiner and

[2] The precise meaning we put into "intertwining" is clear from the computation carried out in the proof: we deal with the generators of the Heisenberg algebra only, and thus are not concerned with the nature of the topological completion, which was left unspecified in Section 3. Furthermore, we understand the intertwining property formally without paying attention to the domains of the definition of the generators of the Heisenberg algebra. For a different approach see [FG3].

introduce the geometric object reflecting its properties, the cluster double. Altogether, the intertwiners give rise to a unitary projective representation of the cluster modular groupoid.

The paper is organized as follows: the properties of the quantum logarithm and dilogarithm essential for us are collected, without proofs, in Section 5. The proofs and more of the properties of these functions can be found in Section 4 of [FG3]. In Section 2.1 we recall, for the convenience of the reader, basic definitions/facts about cluster ensembles. Claim 2.2 delivers a quantization of the space of real positive points of the cluster $\mathcal{X}$-variety. The main result of this paper is Theorem 10, providing an explicit formula for the intertwiner $\mathbf{K}_{\mathbf{i}\to\mathbf{i}'}$.

Acknowledgments. V.F. was supported by grants CRDF 2622; 2660. A.G. was supported by NSF grants DMS-0099390, DMS-0400449, and DMS-065372.

We would like to thank the referee for useful remarks regarding the content of the paper.

1 Cluster ensembles

1.1 Basic definitions

A *seed* $\mathbf{i}$ is a triple (I, ε, d), where I is a finite set, ε is a matrix ε_{ij}, $i, j \in I$, with $\varepsilon_{ij} \in \mathbb{Z}$, and $d = \{d_i\}$, $i \in I$, are positive integers such that the matrix $\widehat{\varepsilon}_{ij} := \varepsilon_{ij} d_j^{-1}$ is skew-symmetric.

For a seed $\mathbf{i}$ we assign a torus $\mathcal{X}_{\mathbf{i}} = (\mathbb{G}_m)^I$ with the coordinates $\{X_i | i \in I\}$ on the factors and a Poisson structure given by

$$\{X_i, X_j\} = \widehat{\varepsilon}_{ij} X_i X_j. \tag{1}$$

Let $\mathbf{i} = (I, \varepsilon, d)$ and $\mathbf{i}' = (I', \varepsilon', d')$ be two seeds, and $k \in I$. A *mutation in the direction* $k \in I$ is an isomorphism $\mu_k : I \to I'$ such that $d'_{\mu_k(i)} = d_i$, and

$$\varepsilon'_{\mu_k(i)\mu_k(j)} = \begin{cases} -\varepsilon_{ij} & \text{if } i = k \text{ or } j = k, \\ \varepsilon_{ij} & \text{if } \varepsilon_{ik}\varepsilon_{kj} \leq 0, \\ \varepsilon_{ij} + |\varepsilon_{ik}|\varepsilon_{kj} & \text{if } \varepsilon_{ik}\varepsilon_{kj} > 0. \end{cases} \tag{2}$$

A *symmetry* of a seed $\mathbf{i} = (I, \varepsilon, d)$ is an automorphism σ of the set I preserving the matrix ε and the numbers d_i. Symmetries and mutations induce rational maps between the corresponding seed $\mathcal{X}$-tori, denoted by the same symbols μ_k and σ and given by the formulas $\sigma^* X_{\sigma(i)} = X_i$ and

$$\mu_{k(i)}^* X_{\mu_k(i)} = \begin{cases} X_k^{-1} & \text{if } \ i = k, \\ X_i(1 + X_k^{-\mathrm{sgn}(\varepsilon_{\mathrm{ik}})})^{-\varepsilon_{ik}} & \text{if } \ i \neq k. \end{cases} \tag{3}$$

A *seed cluster transformation* is a composition of symmetries and mutations. Two seeds are *equivalent* if they are related by a cluster transformation. The equivalence class of a seed $\mathbf{i}$ is denoted by $|\mathbf{i}|$. A seed cluster transformation induces a rational map between the two seed $\mathcal{X}$-tori, called a cluster transformation map.

A *cluster $\mathcal{X}$-variety* $\mathcal{X}_{|\mathbf{i}|}$ is a scheme over $\mathbb{Z}$ obtained by gluing the seed $\mathcal{X}$-tori for the seeds equivalent to a given seed $\mathbf{i}$ via the cluster transformation maps, and then taking the affine closure. Every seed provides a cluster $\mathcal{X}$-variety with a rational coordinate system. Its coordinates are called *cluster coordinates.* Cluster transformation maps preserve the Poisson structure. Thus a cluster $\mathcal{X}$-variety has a canonical Poisson structure.

The cluster $\mathcal{A}$-varieties. Given a seed $\mathbf{i}$, define a *seed $\mathcal{A}$-torus* $\mathcal{A}_{\mathbf{i}} := (\mathbb{G}_m)^I$ with the standard coordinates $\{A_i | i \in I\}$ on the factors. Symmetries and mutations give rise to birational maps between the seed $\mathcal{A}$-tori, given by $\sigma^* A_{\sigma(i)} = A_i$ and

$$\mu_{k(i)}^* A_{\mu_k(i)} = \begin{cases} A_i & \text{if } \ i \neq k, \\ A_k^{-1} \left(\prod_{i|\varepsilon_{ki}>0} A_i^{\varepsilon_{ki}} + \prod_{i|\varepsilon_{ki}<0} A_i^{-\varepsilon_{ki}} \right) & \text{if } \ i = k. \end{cases} \tag{4}$$

The cluster $\mathcal{A}$-variety $\mathcal{A}_{|\mathbf{i}|}$ is a scheme over $\mathbb{Z}$ obtained by gluing all seed $\mathcal{A}$-tori for the seeds equivalent to a given seed $\mathbf{i}$ using the above birational isomorphisms, and taking the affine closure.

There is a map $p : \mathcal{A} \longrightarrow \mathcal{X}$, given in every cluster coordinate system by $p^* X_k = \prod_{I \in I} A_i^{\varepsilon_{ki}}$.

Cluster $\mathcal{A}$- and $\mathcal{X}$-varieties have canonical positive atlases, so it makes sense to consider the sets of their real positive points, denoted by $\mathcal{A}^+$ and $\mathcal{X}^+$.

The cluster modular groupoid. Seed cluster transformations inducing the same map of the seed $\mathcal{A}$-tori are called *trivial seed cluster transformations.* The *cluster modular groupoid* $\mathcal{G}_{|\mathbf{i}|}$ is a groupoid whose objects are seeds equivalent to a given seed $\mathbf{i}$, and $\mathrm{Hom}(\mathbf{i}, \mathbf{i}')$ is the set of all seed cluster transformations from $\mathbf{i}$ to $\mathbf{i}'$ modulo the trivial ones. Given a seed $\mathbf{i}$, the *cluster mapping class group* $\Gamma_{\mathbf{i}}$ is the automorphism group of the object $\mathbf{i}$ of $\mathcal{G}_{|\mathbf{i}|}$. The group $\Gamma_{\mathbf{i}}$ acts by automorphisms of the cluster $\mathcal{A}$-variety.

The quantum space $\mathcal{X}_q$. This is a canonical noncommutative q-deformation of the cluster $\mathcal{X}$-variety defined in Section 3 of [FG2].

We start from the *seed quantum torus algebra* $\mathrm{T}_{\mathbf{i}}^q$, defined as an associative $*$-algebra with generators $X_i^{\pm 1}$, $i \in I$, and $q^{\pm 1}$ and relations

$$q^{-\widehat{\varepsilon}_{ij}} X_i X_j = q^{-\widehat{\varepsilon}_{ji}} X_j X_i, \qquad *X_i = X_i, \quad *q = q^{-1}.$$

Let QTor^* be the category whose objects are quantum torus algebras and whose morphisms are $*$-homomorphisms of their fraction fields. The quantum space $\mathcal{X}_q$ is understood as a contravariant functor

$$\eta^q : \text{the modular groupoid } \mathcal{G}_{|\mathbf{i}|} \longrightarrow \mathrm{QTor}^*.$$

It assigns to a seed $\mathbf{i}$ the quantum torus $*$-algebra $\mathrm{T}_{\mathbf{i}}^q$, and to a mutation $\mathbf{i} \longrightarrow \mathbf{i}'$ a map of the fraction fields $\mathrm{Frac}(\mathrm{T}_{\mathbf{i}'}^q) \longrightarrow \mathrm{Frac}(\mathrm{T}_{\mathbf{i}}^q)$, given by a q-deformation of formulas[3] (3): Set $q_k := q^{1/h_k}$,

[3]See also a more transparent and less computational definition given in Section 3 of [FG3].

$$(\mu_k^q)^*(X_i) := \begin{cases} X_k^{-1} & \text{if } \ i = k, \\ X_i \left(\prod_{a=1}^{|\varepsilon_{ik}|}(1+q_k^{2a-1}X_k^{-\mathrm{sgn}(\varepsilon_{ik})})\right)^{-\varepsilon_{ik}} & \text{if } \ i \neq k. \end{cases}$$

One uses Theorem 7.2 in [FG3] to prove that η^q sends trivial seed cluster transformations to the identity maps.

The *chiral dual* to a seed $\mathbf{i} = (I, \varepsilon, d)$ is a seed $\mathbf{i}^0 := (I, -\varepsilon, d)$. Mutations commute with the chiral duality on seeds. Therefore a cluster $\mathcal{X}$-variety $\mathcal{X}$ (respectively $\mathcal{A}$-variety $\mathcal{A}$), gives rise to the chiral dual cluster $\mathcal{X}$-variety (respectively $\mathcal{A}$-variety) denoted by $\mathcal{X}^0$ (respectively $\mathcal{A}^0$). They are related, see Lemmas 1 and 2 .

The *Langlands dual* to a seed $\mathbf{i} = (I, \varepsilon, d)$ is the seed $\mathbf{i}^\vee = (I, \varepsilon^\vee, d^\vee)$, where $d_i^\vee := d_i^{-1}$ and

$$\varepsilon_{ij}^\vee = -\varepsilon_{ji}^\vee := \widehat{d}_i^{-1}\varepsilon_{ij}\widehat{d}_j, \qquad \widehat{d}_i := d_i^{-1}. \tag{5}$$

The Langlands duality on seeds commutes with mutations. Therefore it gives rise to the Langlands dual cluster $\mathcal{A}$- and $\mathcal{X}$-varieties, denoted by $\mathcal{A}^\vee$ and $\mathcal{X}^\vee$.

Below we omit the subscript $|\mathbf{i}|$, encoding the corresponding cluster ensemble whenever possible.

1.2 Connections between quantum $\mathcal{X}$-varieties

There are three ways to alter the space $\mathcal{X}_{|\mathbf{i}|,q}$:

(i) change q to q^{-1},
(ii) change $\mathbf{i}$ to $\mathbf{i}^0$,
(iii) change the quantum space $\mathcal{X}_{|\mathbf{i}|,q}$ to the opposite quantum space $\mathcal{X}_{|\mathbf{i}|,q}^{\mathrm{opp}}$.

(In (iii) we change every quantum torus from which we glue the space to the opposite one).

The following lemma tells that the resulting three quantum spaces are canonically isomorphic:

Lemma 1. *(a) There is a canonical isomorphism of quantum spaces*

$$\alpha_{\mathcal{X}}^q : \mathcal{X}_{|\mathbf{i}|,q} \longrightarrow \mathcal{X}_{|\mathbf{i}|,q^{-1}}^{\mathrm{opp}}, \qquad (\alpha_{\mathcal{X}}^q)^* : X_i \longmapsto X_i$$

(given on the generators of any cluster coordinate system by $X_i \longmapsto X_i$).
(b) There is a canonical isomorphism of quantum spaces

$$i_{\mathcal{X}}^q : \mathcal{X}_{|\mathbf{i}|,q} \longrightarrow \mathcal{X}_{|\mathbf{i}^0|,q^{-1}}, \qquad (i_{\mathcal{X}}^q)^* : X_i^0 \longmapsto X_i^{-1}$$

(given in any cluster coordinate system by $X_i^0 \longmapsto X_i^{-1}$, where X_i^0 are the generators of $\mathcal{X}_{\mathbf{i}^0,q^{-1}}$).
(c) There is a canonical isomorphism of quantum spaces

$$\beta_{\mathcal{X}}^q := \alpha_{\mathcal{X}}^q \circ i_{\mathcal{X}}^q : \mathcal{X}_{|\mathbf{i}^0|,q} \longrightarrow \mathcal{X}_{|\mathbf{i}|,q}^{\mathrm{opp}}, \qquad X_i \longmapsto X_i^{0^{-1}}.$$

Proof. Clearly each of the three maps is an isomorphism of the corresponding seed quantum tori algebras. For example, in case (b) we have

$$i^q_{\mathcal{X}}\left(\left(q^{-1}\right)^{-\widehat{\varepsilon}^o_{ij}} X^0_i X^0_j\right) = q^{-\widehat{\varepsilon}_{ij}} X_i^{-1} X_j^{-1}.$$

So we need to check that they commute with the mutations.

(a) Let us assume first that $\varepsilon_{ik} = a < 0$. The claim results from the fact that the following two compositions are equal (observe that α^* is an antiautomorphism):

$$X_i \stackrel{\mu_k^*}{\longmapsto} X_i \prod_{b=1}^{a}(1+q_k^{2b-1}X_k) \stackrel{\alpha^*}{\longmapsto} \prod_{b=1}^{a}(1+q_k^{2b-1}X_k)X_i;$$
$$X_i \stackrel{\alpha^*}{\longmapsto} X_i \stackrel{(\mu_k^0)^*}{\longmapsto} X_i \prod_{b=1}^{a}(1+q_k^{-(2b-1)}X_k).$$

The computation in the case $\varepsilon_{ik} > 0$ is similar.

(b) To check that $i^q_{\mathcal{X}} \circ \mu_k^* = \mu_k^* \circ i^q_{\mathcal{X}}$ we calculate each of the maps on the generator X_i. Let us assume $\varepsilon_{ik} = -a < 0$. Then $\varepsilon^0_{ik} = a$, and one has

$$i^q_{\mathcal{X}} \circ \mu_k^*(X_i^{0\prime}) = X_i^{-1}\prod_{b=1}^{a}(1+q^{2b-1}X_k^{-1}),$$
$$\mu_k^* \circ i^q_{\mathcal{X}}(X_i^{0\prime}) = \left(X_i \prod_{b=1}^{a}(1+q^{-(2b-1)}X_k^{-1})^{-1}\right)^{-1}$$
$$= \prod_{b=1}^{a}(1+q^{-(2b-1)}X_k^{-1})X_i^{-1} = X_i^{-1}\prod_{b=1}^{a}(1+q^{2b-1}X_k^{-1}).$$

The case $\varepsilon_{ik} > 0$ is similar. One can reduce it to the case $\varepsilon_{ik} < 0$, since $\mu_k \circ \mu_k = \mathrm{Id}$, and $\varepsilon'_{ik} = -\varepsilon_{ik}$. Part (b) is proved.

(c) Follows from (a) and (b). The lemma is proved. □

Lemma 2. *The cluster ensembles related to the seeds* $\mathbf{i}$ *and* $\mathbf{i}^{\mathbf{0}}$ *are canonically isomorphic as pairs of varieties. The isomorphism is provided by the following maps:*

$$\mathrm{Id} : \mathcal{A}_{|\mathbf{i}|} \longrightarrow \mathcal{A}_{\mathbf{i}^{\mathbf{0}}}; \qquad i_{\mathcal{X}} : \mathcal{X}_{|\mathbf{i}|} \longrightarrow \mathcal{X}_{|\mathbf{i}^{\mathbf{0}}|}.$$

Proof. In a given cluster coordinate system our maps are obviously isomorphisms. The compatibility with $\mathcal{X}$-cluster transformations is part (b) of Lemma 1; for the $\mathcal{A}$-cluster transformations we have

$$A^0_k {A^0}'_k = \prod_{\varepsilon^o_{ki}>0}(A^0_i)^{\varepsilon^0_{ki}} + \prod_{\varepsilon^o_{ki}<0}(A^0_i)^{-\varepsilon^0_{ki}} = \prod_{\varepsilon_{ki}<0} A_i^{-\varepsilon_{ki}} + \prod_{\varepsilon_{ki}>0} A_i^{\varepsilon_{ki}} = A_k A'_k.$$

Compatibility with the projection p is clear. The lemma is proved. □

2 Motivation: $*$-quantization of cluster $\mathcal{X}$-varieties

2.1 $*$-quantization of the space $\mathcal{X}^+$ via the quantum logarithm

Let $\{X_i\}$ be coordinates on the cluster $\mathcal{X}$-variety corresponding to a seed $\mathbf{i}$. Since by definition the functions X_i are strictly positive at the points of $\mathcal{X}^+$, we can introduce the *logarithmic coordinates* $x_i := \log X_i$ on $\mathcal{X}^+$. For every seed $\mathbf{i}$ they provide an isomorphism

$$\beta_{\mathbf{i}} : \mathcal{X}^+_{\mathbf{i}} \stackrel{\sim}{\longrightarrow} \mathbb{R}^I; \qquad t \longmapsto \{x_i(t)\}.$$

For a mutation $\mu_k : \mathbf{i} \to \mathbf{i}'$ there is a gluing map

$$\beta_{\mathbf{i}\to\mathbf{i}'} : \mathcal{X}^+_{\mathbf{i}} \longrightarrow \mathcal{X}^+_{\mathbf{i}'}, \qquad \beta_{\mathbf{i}\to\mathbf{i}'}(x'_i) = \begin{cases} x_i - \varepsilon_{ik}\log(1+e^{-\mathrm{sgn}(\varepsilon_{\mathrm{ik}})\mathrm{x}_{\mathrm{k}}}), & i \neq k, \\ -x_k, & i = k. \end{cases} \tag{6}$$

To prepare the soil for quantization, let us look at this from a different point of view. Let Com* be the category of commutative topological $*$-algebras over $\mathbb{C}$. Recall the cluster modular groupoid $\mathcal{G}_{|\mathbf{i}|}$. There is a contravariant functor

$$\beta : \mathcal{G}_{|\mathbf{i}|} \longrightarrow \mathrm{Com}^*.$$

Namely, we assign to a seed $\mathbf{i}$ a commutative topological $*$-algebra $S(\mathcal{X}^+_{\mathbf{i}})$ of smooth complex-valued functions in $\mathcal{X}^+_{\mathbf{i}}$ with $*f := \overline{f}$, and to a mutation $\mathbf{i} \to \mathbf{i}'$ a homomorphism $\beta^*_{\mathbf{i}\to\mathbf{i}'} : S(\mathcal{X}^+_{\mathbf{i}'}) \longrightarrow S(\mathcal{X}^+_{\mathbf{i}})$.

Let $\mathcal{C}$ be a category whose morphisms are $\mathbb{C}$-vector spaces. *Projectivization* $P\mathcal{C}$ of the category $\mathcal{C}$ is a new category with the same objects as $\mathcal{C}$, and morphisms given by $\mathrm{Hom}_{P\mathcal{C}}(C_1, C_2) := \mathrm{Hom}_{\mathcal{C}}(C_1, C_2)/U(1)$, where $U(1)$ is the multiplicative group of complex numbers with absolute value 1. A *projective functor* $F : \mathcal{G} \to \mathcal{C}$ is a functor from $\mathcal{G}$ to $P\mathcal{C}$.

Let $\mathcal{C}^*$ be the category of topological $*$-algebras. Two functors $F_1, F_2 : \mathcal{C} \longrightarrow \mathcal{C}^*$ *essentially coincide* if there exists a third functor F and natural transformations $F_1 \to F, F_2 \to F$ providing for every object C dense inclusions $F_1(C) \hookrightarrow F(C)$, $F_1(C) \hookrightarrow F(C)$.

Definition 3. *A quantization of the space $\mathcal{X}^+_{|\mathbf{i}|}$ is a family of contravariant projective functors*

$$\kappa^{\hbar}_{|\mathbf{i}|} : \mathcal{G}_{|\mathbf{i}|} \longrightarrow \mathcal{C}^*$$

depending smoothly on a real parameter $\hbar$, related to the original Poisson manifold $\mathcal{X}^+_{|\mathbf{i}|}$ as follows:

(i) The limit $\kappa_{|\mathbf{i}|} := \lim_{\hbar\to 0} \kappa^{\hbar}_{|\mathbf{i}|}$ exists and essentially coincides with the functor β defining $\mathcal{X}^+_{|\mathbf{i}|}$.

(ii) The Poisson bracket given by $\lim_{\hbar\to 0}[f_1, f_2]/\hbar$ is defined and coincides with the one on $\mathcal{X}^+_{|\mathbf{i}|}$.

Let us define a quantization functor $\kappa^\hbar = \kappa^\hbar_{|\mathbf{i}|}$. We assign to every seed $\mathbf{i}$ the Heisenberg $*$-algebra $\mathcal{H}^\hbar_{\mathbf{i}}$. It is a topological $*$-algebra over $\mathbb{C}$ generated by elements x_i such that

$$[x_j, x_k] = 2\pi i\hbar\widehat{\varepsilon}_{jk}; \quad x_j^* = x_j; \qquad q = e^{\pi i\hbar}.$$

Further, let us assign to the mutation $\mu_k : \mathbf{i} \to \mathbf{i}'$ a homomorphism of topological $*$-algebras

$$\kappa^\hbar(\mu_k) : \mathcal{H}^\hbar_{\mathbf{i}'} \longrightarrow \mathcal{H}^\hbar_{\mathbf{i}}. \tag{7}$$

We employ the quantum logarithm $\phi^\hbar(z)$, see (22). Denote by x'_i the generators of $\mathcal{H}^\hbar_{\mathbf{i}'}$. Set

$$\hbar_k := \widehat{d}_k\hbar; \qquad \kappa^\hbar(\mu_k) : x'_i \longmapsto \begin{cases} x_i - \varepsilon_{ik}\phi^{\hbar_k}(-\mathrm{sgn}(\varepsilon_{\mathrm{ik}})\mathrm{x_k}) & \text{if } k \neq i, \\ -x_i & \text{if } k = i. \end{cases} \tag{8}$$

Claim. (a) Formulas (8) provide a morphism of $*$–algebras

$$\kappa^\hbar_{|\mathbf{i}|}(\mu_k) : \mathcal{H}^\hbar_{\mathbf{i}'} \longrightarrow \mathcal{H}^\hbar_{\mathbf{i}}.$$

(b) The collection of $*$-algebras $\{\mathcal{H}_{\mathbf{i}}\}$ and morphisms $\{\kappa^\hbar_{|\mathbf{i}|}(\mu_k)\}$ provide a quantization functor

$$\kappa^\hbar_{|\mathbf{i}|} : \mathcal{G}_{|\mathbf{i}|} \longrightarrow \mathcal{PC}^*.$$

(c) Let $\hbar^\vee := 1/\hbar$. Then there are isomorphisms

$$\mathcal{H}^\hbar_{\mathbf{i}} \longrightarrow \mathcal{H}^{\hbar^\vee}_{\mathbf{i}^\vee}, \qquad x_i \longmapsto x_i^\vee := \frac{x_i}{\widehat{d}_i\hbar}. \tag{9}$$

They give rise to a natural transformation of functors $\kappa^\hbar_{|\mathbf{i}|} \longrightarrow \kappa^{\hbar^\vee}_{|\mathbf{i}^\vee|}$.

Justification. (a) Property **A3** of the function $\phi^\hbar(x)$, see Section 4, guarantees that the morphism $\kappa^\hbar(\mu_k)$ preserves the real structure. It follows from Property **A1** that when $\hbar \to 0$, the limit of the quantum formula (8) exists and coincides with the mutation formula (6) for the logarithmic coordinates x_i.

(b) To check that we have a functor, one needs to check first that mutation formulas are compatible with the transformations $\kappa^\hbar(\mu_k)$. This is a straightforward calculation using Property **A5**. Then one has to check that the defining relations for the groupoid $\mathcal{G}_{|\mathbf{i}|}$ are mapped to zero. Here we need the results of Sections 3.2–3.3 of [FG2] and the following well-known lemma:

Lemma 4. *Suppose that A, B are self-adjoint operators, $[A, B] = -\lambda$ is a scalar, and $f(z)$ is a continuous function with primitive $F(z)$. Then*

$$e^{A+f(B)} = e^A \exp\left\{\frac{1}{\lambda}\int_B^{B+\lambda} f(z)dz\right\} := e^A \exp\left(\frac{F(B+\lambda) - F(B)}{\lambda}\right).$$

(c) Thanks to formula (5), the map $x_i \longmapsto x_i^\vee$ is a $*$-algebra homomorphism:

$$[x_i^\vee, x_j^\vee] = \frac{[x_i, x_j]}{\widehat{d}_i \widehat{d}_j \hbar^2} = 2\pi i \hbar^\vee \widehat{\varepsilon}_{ij}/\widehat{d}_i \widehat{d}_j \stackrel{(5)}{=} 2\pi i \hbar^\vee \widehat{\varepsilon}_{ij}^\vee.$$

To verify that it commutes with mutation homomorphisms we use Properties **A2** and **A4** of the function $\phi^\hbar(x)$, observing that

$$\frac{x_i + |\varepsilon_{ik}|\phi^{\hbar_k}(x_k)}{\widehat{d}_i \hbar} = \frac{x_i}{\widehat{d}_i \hbar} + \frac{|\widehat{d}_i^{-1}\varepsilon_{ik}\widehat{d}_k|\phi^{\widehat{d}_k \hbar}(x_k)}{\widehat{d}_k \hbar} \stackrel{(5)+A4}{=} x_i^\vee + |\varepsilon_{ik}^\vee|\phi^{\hbar_k^\vee}(x_k^\vee).$$

2.2 Modular double of a cluster $\mathcal{X}$-variety and $*$-quantization of the space $\mathcal{X}^+$

Set

$$q := e^{\pi i \hbar}, \quad q^\vee := e^{\pi i/\hbar}, \quad \hbar \in \mathbb{R}.$$

Definition 5. *The modular double $\mathcal{X}_{|\mathbf{i}|,q} \times \mathcal{X}_{|\mathbf{i}^\vee|,q^\vee}$ of a quantum cluster $\mathcal{X}$-variety $\mathcal{X}_{|\mathbf{i}|,q}$ is a contravariant functor*

$$\eta^q_{|\mathbf{i}|} \otimes \eta^{q^\vee}_{|\mathbf{i}^\vee|} : \mathcal{G}_{|\mathbf{i}|} \longrightarrow \mathrm{QTor}^*.$$

So we assign to a seed $\mathbf{i}$ a quantum torus algebra $\mathrm{T}^q_{\mathbf{i}} \otimes \mathrm{T}^{q^\vee}_{\mathbf{i}^\vee}$, and to a mutation $\mu_k : \mathbf{i} \to \mathbf{i}'$ a positive $*$-homomorphism of the fraction fields of the quantum torus algebras

$$\eta^q_{|\mathbf{i}|}(\mu_k) \otimes \eta^{q^\vee}_{|\mathbf{i}^\vee|}(\mu_k) : \mathbb{T}^q_{\mathbf{i}'} \otimes \mathbb{T}^{q^\vee}_{\mathbf{i}'^\vee} \longrightarrow \mathbb{T}^q_{\mathbf{i}} \otimes \mathbb{T}^{q^\vee}_{\mathbf{i}^\vee}, \qquad \mathbb{T} := \mathrm{Frac}(\mathrm{T}).$$

We want to relate the modular double $\mathcal{X}_{|\mathbf{i}|,q} \times \mathcal{X}_{|\mathbf{i}^\vee|,q^\vee}$ to the quantization of the space $\mathcal{X}^+_{|\mathbf{i}|}$. We are going to define a natural transformation of functors $\eta^q_{|\mathbf{i}|} \otimes \eta^{q^\vee}_{\mathcal{E}^\vee} \longrightarrow \kappa^\hbar_{|\mathbf{i}|}$.

We use the following easy fact. Assume that $[y_i, y_j]$ is a scalar. Then we have

$$e^{y_i} e^{y_j} = e^{[y_i, y_j]} e^{y_j} e^{y_i}. \tag{10}$$

Let $\mathbf{i}$ be a seed. Denote by X_i the generators of $\mathbb{T}^q_{\mathbf{i}}$, and by $X_i^\vee$ the generators of $\mathbb{T}^{q^\vee}_{\mathbf{i}^\vee}$. It is easy to check using (10) that there are the following homomorphisms:

$$l_{\mathbf{i}} : \mathbb{T}^q_{\mathbf{i}} \longrightarrow \mathcal{H}^h_{\mathbf{i}}, \quad X_i \longmapsto e^{x_i}, \quad \text{and} \quad l^\vee_{\mathbf{i}} : \mathbb{T}^{q^\vee}_{\mathbf{i}^\vee} \longrightarrow \mathcal{H}^\hbar_{\mathbf{i}}, \quad X_i^\vee \longmapsto e^{x_i^\vee}.$$

They evidently commute with the $*$-structures. Their images commute. Indeed, since $\widehat{\varepsilon}_{ij} \in \mathbb{Z}$ one has $e^{[x_i, x_j/\hbar]} = e^{2\pi i \widehat{\varepsilon}_{ij}} = 1$. So e^{x_i} commutes with $e^{x_j/\hbar}$. Therefore they give rise to a homomorphism of the tensor product:

$$L_{\mathbf{i}} := l_{\mathbf{i}} \otimes l^\vee_{\mathbf{i}} : \mathbb{T}^q_{\mathbf{i}} \otimes \mathbb{T}^{q^\vee}_{\mathbf{i}^\vee} \longrightarrow \mathcal{H}^\hbar_{\mathbf{i}}.$$

Proposition 6. *For any mutation $\mu_k : \mathbf{i} \to \mathbf{i}'$ the following diagram, where the left vertical arrow is the map $\eta^q(\mu_k) \otimes \eta^{q^\vee}(\mu_k)$, and the right one is $\kappa^\hbar(\mu_k)$, is commutative:*

$$\begin{array}{ccc} \mathbb{T}^q_{\mathbf{i}} \otimes \mathbb{T}^{q^\vee}_{\mathbf{i}} & \xrightarrow{L_{\mathbf{i}}} & \mathcal{H}^\hbar_{\mathbf{i}} \\ \eta^{q,q^\vee} \uparrow & & \uparrow \kappa^\hbar \\ \mathbb{T}^q_{\mathbf{i}'} \otimes \mathbb{T}^{q^\vee}_{\mathbf{i}'} & \xrightarrow{L_{\mathbf{i}'}} & \mathcal{H}^\hbar_{\mathbf{i}'} \end{array}$$

Proof. We need Lemma 4. The case $i = k$ is trivial, so we assume that $i \neq k$. Let $\varepsilon_{ik} = -a \leq 0$. Then applying the lemma we get

$$\begin{aligned} \kappa^\hbar_{|\mathbf{i}|}(\mu_k) L_{\mathbf{i}'}(X'_i \otimes 1) &= \kappa^\hbar(\mu_k) e^{x_i^\vee} = e^{x_i + a\phi^{\hbar_k}(x_k)} \\ &= \frac{e^{x_i}}{2\pi i a \hbar_k} \exp\left(\int_{x_k}^{x_k + 2\pi i \hbar_k a} a\phi^{\hbar_k}(z) dz \right) \\ &= \frac{e^{x_i}}{2\pi i \hbar_k} \exp\left(\int_{-\infty}^{x_k} \left(\phi^{\hbar_k}(z + 2\pi i \hbar_k a) - \phi^{\hbar_k}(z) \right) dz \right) \\ &\overset{A5}{=} \frac{e^{x_i}}{2\pi i \hbar_k} \exp\left(\int_{-\infty}^{x_k} \sum_{b=1}^{a} \frac{2\pi i \hbar_k}{e^{-z - i\pi(2b-1)\hbar_k} + 1} dz \right) \\ e^{x_i} \prod_{b=1}^{a} (1 + q_k^{2a-1} e^{x_k}) &= L_{\mathbf{i}} \left(X_i \prod_{b=1}^{a} (1 + q_k^{2a-1} X_k) \right) \\ &= L_{\mathbf{i}}(\eta^q_{|\mathbf{i}|}(\mu_k) \otimes \eta^{q^\vee}_{|\mathbf{i}^\vee|}(\mu_k))(X'_i \otimes 1). \end{aligned}$$

The calculation in the case $\varepsilon_{ik} = a \geq 0$ is similar. The proposition is proved.

Claim. The collection of homomorphisms $\{L_{\mathbf{i}}\}$ provides a morphism of functors

$$\mathbb{L}^\hbar : \eta^q_{|\mathbf{i}|} \otimes \eta^{q^\vee}_{|\mathbf{i}^\vee|} \longrightarrow \kappa^\hbar_{|\mathbf{i}|}. \tag{11}$$

Justification. Is given by Proposition 6.

Representations of the quantized $\mathcal{X}^+_{|\mathbf{i}|}$-space. The following definition serves as a motivation of the construction of intertwiners presented below.

Definition 7. *A projective $*$-representation of the quantized $\mathcal{X}^+_{|\mathbf{i}|}$-space is the following data:*

(i) A projective functor

$$\mathcal{L}_{|\mathbf{i}|} : \mathcal{G}_{|\mathbf{i}|} \longrightarrow \textit{the category of Hilbert spaces.}$$

It includes for each object $\mathbf{i}$ of $\mathcal{G}_{|\mathbf{i}|}$ a Hilbert space $\mathbb{L}_{\mathbf{i}}$, and for every mutation $\mu_k : \mathbf{i} \to \mathbf{i}'$ a unitary operator, defined up to a scalar of absolute value 1:

$$\mathbf{K}_{\mathbf{i},\mathbf{i}'} : \mathbb{L}_{\mathbf{i}} \longrightarrow \mathbb{L}_{\mathbf{i}'}.$$

(ii) A $*$-representation $\rho_{\mathbf{i}}$ of the Heisenberg algebra $\mathcal{H}^{\hbar}_{\mathbf{i}}$ in the Hilbert space $\mathbb{L}_{\mathbf{i}}$.

(iii) The operators $\mathbf{K}_{\mathbf{i},\mathbf{i}'}$ intertwine the representations $\rho_{\mathbf{i}}$ and $\rho_{\mathbf{i}'}$:

$$\rho_{\mathbf{i}}(s) = \mathbf{K}^{-1}_{\mathbf{i},\mathbf{i}'}\rho_{\mathbf{i}'}\Big(\kappa^{\hbar}(\mu_k)(s)\Big)\mathbf{K}_{\mathbf{i},\mathbf{i}'}, \qquad s \in \mathcal{H}^{\hbar}_{\mathbf{i}}.$$

The morphisms of the representations of the quantum $\mathcal{X}^{+}_{|\mathbf{i}|}$-space are defined in an obvious way.

Representations of the mapping class group $\Gamma_{|\mathbf{i}|}$. Restricting the functor $\rho_{|\mathbf{i}|}$ to the group of automorphisms of an object of the groupoid $\mathcal{G}_{|\mathbf{i}|}$, we get a projective unitary representation of $\Gamma_{|\mathbf{i}|}$.

The Heisenberg algebra $\mathcal{H}^{\hbar}_{\mathbf{i}}$ has a family of irreducible $*$-representations by operators in a Hilbert space. These representations are characterized by the central character χ.

The collection of Hilbert spaces $\{\mathbb{L}_{\mathbf{i}}\}$ and representations $\{\rho_{\mathbf{i}}\}$ is by no means canonical: it depends, for example, on the choice of polarization of the Heisenberg algebra. Once chosen, it determines the intertwiners $\mathbf{K}_{\mathbf{i},\mathbf{i}'}$. Below we introduce a *canonical representation* of the chiral double of the quantized space $\mathcal{X}^{+}_{|\mathbf{i}|}$, defined using the Hilbert spaces $L^2(\mathcal{A}^{+}_{\mathbf{i}})$.

3 The intertwiner

3.1 A bimodule structure on functions on the $\mathcal{A}$–space

Let X be an algebra. Recall that M is a bimodule over X if X acts on M from the left as well as from the right, and these two actions commute. So M is an $X \otimes X^{\mathrm{opp}}$-module, where X^{opp} is the algebra with the product $x * y := yx$.

Let us choose a seed $\mathbf{i}$. Recall the algebra $\mathbb{Q}[\mathcal{A}_{\mathbf{i}}]$ of regular functions on the seed torus $\mathcal{A}_{\mathbf{i}}$. We assume that $q \in \mathbb{C}^*$. For each $i \in I$ let us define commuting algebra homomorphisms

$$t_i^{\pm} : \mathbb{Q}[\mathcal{A}_{\mathbf{i}}] \longrightarrow \mathbb{Q}[\mathcal{A}_{\mathbf{i}}]; \quad t_i^{\pm} : \begin{cases} A_i \longmapsto q^{\pm \widehat{d}_i} A_i, & \\ A_j \longmapsto A_j, & j \neq i. \end{cases}$$

Since $\varepsilon_{ii} = 0$, A_i does not appear in the monomial $p^* X_i$, and so the operator of multiplication by $p^* X_i$ commutes with $t_i^{\pm}$. Let us define a $T^q_{\mathbf{i}}$-bimodule structure on $\mathbb{Q}[\mathcal{A}_{\mathbf{i}}]$. The left and right actions of the generator X_i on $f \in \mathbb{Q}[\mathcal{A}_{\mathbf{i}}]$ are given by

$$X_i \circ f := p^* X_i \cdot t_i^{-}(f), \qquad f \circ X_i := p^* X_i \cdot t_i^{+}(f), \tag{12}$$

Lemma 8. *The operators (12) provide $\mathbb{Q}[\mathcal{A}_{\mathbf{i}}]$ with a structure of a bimodule over the algebra $T^q_{\mathbf{i}}$.*

Proof. Observe that one has

$$t_j^+(p^*X_i)\cdot p^*X_j = t_i^-(p^*X_j)\cdot p^*X_i = q^{\widehat{\varepsilon}_{ij}}p^*X_ip^*X_j.$$

Indeed, the first term equals $q^{\varepsilon_{ij}d_j}p^*X_i\cdot p^*X_j$, and the second is $q^{-\varepsilon_{ji}d_i}p^*X_i\cdot p^*X_j$. One has

$$q^{-\widehat{\varepsilon}_{ij}}X_iX_j\circ f = q^{-\widehat{\varepsilon}_{ij}}p^*X_i\cdot t_i^-(p^*X_j)\cdot t_i^-t_j^-(f) = p^*X_i\cdot p^*X_j\cdot t_i^-t_j^-(f),$$

$$f\circ q^{-\widehat{\varepsilon}_{ij}}X_iX_j = q^{-\widehat{\varepsilon}_{ij}}p^*X_j\cdot t_j^+(p^*X_i)\cdot t_j^+t_i^+(f) = p^*X_i\cdot p^*X_j\cdot t_i^+t_j^+(f).$$

Since the right-hand sides are evidently symmetric in i,j, we have the desired relations, and hence the left and right actions of the quantum algebra torus. Further, the two actions commute:

$$X_i\circ(f\circ X_j) = X_i\circ\left(t_j^+(f)p^*X_j\right) = p^*X_it_i^-(p^*X_j)t_i^-t_j^+(f),$$
$$(X_i\circ f)\circ X_j = (p^*X_it_i^-(f))\circ X_j = p^*X_jt_j^+(p^*X_i)t_j^+t_i^-(f).$$

The lemma is proved. □

The logarithmic version of the bimodule structure. Since the coordinate functions A_i are positive on the space $\mathcal{A}^+$, one can introduce new coordinates $a_j := \log A_j$. They provide an isomorphism $\alpha_{\mathbf{i}} : \mathcal{A}^+_{\mathbf{i}} \stackrel{\sim}{\longrightarrow} \mathbb{R}^I$. Set $da := da_1\wedge\dots\wedge da_{|I|}$. There is a Hilbert space $L^2(\mathcal{A}^+_{\mathbf{i}})$ with a scalar product

$$(f,g) := \int_{\mathcal{A}^+_{\mathbf{i}}} f(a)\overline{g(a)}da.$$

Clearly the form da changes sign under a mutation $\mathbf{i}\to\mathbf{i}'$. So the Hilbert spaces $L^2(\mathcal{A}^+_{\mathbf{i}})$ for different seeds $\mathbf{i}$ are naturally identified. Consider the following operators in $L^2(\mathcal{A}^+_{\mathbf{i}})$:

$$\widehat{x}_j^- := -\pi i\hbar\widehat{d}_j\frac{\partial}{\partial a_j} + \sum_k \varepsilon_{jk}a_k, \qquad \widehat{x}_j^+ := \pi i\hbar\widehat{d}_j\frac{\partial}{\partial a_j} + \sum_k \varepsilon_{jk}a_k.$$

Lemma 9. *The operators $\{\widehat{x}_j^\pm\}$ provide the Hilbert space $L^2(\mathcal{A}^+_{\mathbf{i}})$ with a structure of a bimodule over the $*$-algebra $\mathcal{H}^\hbar_{\mathbf{i}}$.*

Proof. These operators are self-adjoint, and one has

$$[\widehat{x}_j^-,\widehat{x}_k^-] = 2\pi i\hbar\widehat{\varepsilon}_{jk};\qquad [\widehat{x}_j^+,\widehat{x}_k^+] = -2\pi i\hbar\widehat{\varepsilon}_{jk},\qquad [\widehat{x}_j^-,\widehat{x}_k^+] = 0,$$

for any $j,k\in I$.

The lemma is proved. □

Remark. There is an automorphism $x_j\longmapsto -x_j$ of the Heisenberg algebra $\mathcal{H}_{\mathbf{i}}$. Similarly there is an automorphism $X_j\longmapsto X_j^{-1}$ of the quantum torus algebra $T^q_{\mathbf{i}}$.

3.2 The intertwiner via the quantum dilogarithm

Let $\mu_k : \mathbf{i} \to \mathbf{i}'$ be a mutation. By Lemma 9, for each seed $\mathbf{i}$ the Hilbert space $L^2(\mathcal{A}^+_{\mathbf{i}})$ has a natural $\mathcal{H}^{\hbar}_{\mathbf{i}}$-bimodule structure. According to Lemma 1, this is the same as the $\mathcal{H}^{\hbar}_{\mathbf{i}} \otimes \mathcal{H}^{\hbar}_{\mathbf{i}^\mathrm{o}}$-module structure. Our goal is to define an operator

$$\mathbf{K}_{\mathbf{i}\to\mathbf{i}'} : L^2(\mathcal{A}^+_{\mathbf{i}}) \longrightarrow L^2(\mathcal{A}^+_{\mathbf{i}'}) \tag{13}$$

intertwining the $\mathcal{H}^{\hbar}_{\mathbf{i}} \otimes \mathcal{H}^{\hbar}_{\mathbf{i}^\mathrm{o}}$- and $\mathcal{H}^{\hbar}_{\mathbf{i}'} \otimes \mathcal{H}^{\hbar}_{\mathbf{i}'^\mathrm{o}}$-module structures. By this we mean only that the operator $\mathbf{K}_{\mathbf{i}\to\mathbf{i}'}$ intertwines the action of the generators of $\mathcal{H}^{\hbar}_{\mathbf{i}} \otimes \mathcal{H}^{\hbar}_{\mathbf{i}^\mathrm{o}}$ with the action of their images under the gluing map $\kappa^{\hbar}_{|\mathbf{i}|}(\mathbf{i} \to \mathbf{i}')$.

The function G. Let us introduce our key function

$$G(a_1, \dots, a_n) = \int \Phi^{\hbar_k}\left(\widehat{d}_k c - \sum_j \varepsilon_{kj} a_j\right)^{-1} \Phi^{\hbar_k}\left(-\widehat{d}_k c - \sum_j \varepsilon_{kj} a_j\right) \exp\left(c \frac{\sum\limits_{j|\varepsilon_{kj}<0} \varepsilon_{kj} a_j + a_k}{\pi i \hbar}\right) dc. \tag{14}$$

Substituting the explicit integral expression for the function $\Phi^{\hbar_k}(z)$, one gets

$$G(a_1, \dots, a_n) = \int \exp\left(\int_\Omega \frac{\exp(it\sum_j \varepsilon_{kj} a_j) \sin(t\widehat{d}_k c)}{2it\ \mathrm{sh}(\pi t)\mathrm{sh}(\pi \hbar_k t)} dt + c\frac{\sum\limits_{j|\varepsilon_{kj}<0} \varepsilon_{kj} a_j + a_k}{\pi i \hbar}\right) dc.$$

We denote by $(a_1, \dots, a_n)$ the logarithmic coordinates corresponding to v, and by $(a_1, \dots, a'_k, \dots, a_n)$ those corresponding to $\mathbf{i}'$. Recall that only the coordinate a_k changes under the mutation μ_k. Let us define the operator $\mathbf{K}_{\mathbf{i}\to\mathbf{i}'}$ by

$$(\mathbf{K}_{\mathbf{i}\to\mathbf{i}'} f)(a_1, \dots, a'_k, \dots, a_n) := \int G(a_1, \dots, a'_k + a_k, \dots, a_n) f(a_1, \dots, a_k, \dots, a_n) da_k, \tag{15}$$

where $a'_k + a_k$ and a_k are at the kth places.

Theorem 10. *The operators $\mathbf{K}_{\mathbf{i}\to\mathbf{i}'}$ intertwine the $\mathcal{H}^{\hbar}_{\mathbf{i}} \otimes \mathcal{H}^{\hbar}_{\mathbf{i}^\mathrm{o}}$-module structures on $L^2(\mathcal{A}^+_{\mathbf{i}})$ provided by Lemma 9.*

Remark. We prove in [FG3] that the collection of Hilbert space $L^2(\mathcal{A}^+_{\mathbf{i}})$ and operators $\mathbf{K}_{\mathbf{i}\to\mathbf{i}'}$ provide a unitary projective representation of the groupoid $\mathcal{G}_{|\mathbf{i}|}$. This implies that the operators $\mathbf{K}_{\mathbf{i}\to\mathbf{i}'}$ give rise to a unitary projective representation of the cluster modular group $\Gamma_{|\mathbf{i}|}$ in $L^2(\mathcal{A}^+_{\mathbf{i}})$.

Proof. We present a computation that allows us to find the function G as a unique function up to a scalar such that the corresponding integral transformation intertwines the $\mathcal{H}^{\hbar}_{\mathbf{i}}$- and $\mathcal{H}^{\hbar}_{\mathbf{i}'}$-bimodule structures on $L^2(\mathcal{A}^+_{\mathbf{i}})$ and $L^2(\mathcal{A}^+_{\mathbf{i}'})$. Recall that $\varepsilon^o_{ij} = -\varepsilon_{ij}$, so we may write $\varepsilon^{\pm}_{ij} := \pm\varepsilon_{ij}$ and denote by $x^{\pm}_i$ the x_i-coordinates for the seeds $\mathbf{i}$ and $\mathbf{i^0}$.

So we have to find G such that the integral transformation (15) induces a map of operators:

$$\widehat{x}'^{\pm}_i \longmapsto \begin{cases} \widehat{x}^{\pm}_i - \varepsilon^{\pm}_{ik}\phi^{\hbar_k}(-\mathrm{sgn}(\varepsilon^{\pm}_{\mathrm{ik}})\widehat{\mathrm{x}}^{\pm}_{\mathrm{k}}) & \text{if } i \neq k, \\ -\widehat{x}^{\pm}_k & \text{if } i = k. \end{cases} \tag{16}$$

This means that we should have (changing $\widehat{x}^-_i$ to $-\widehat{x}^-_i$ for convenience), for $i \neq k$,

$$\begin{aligned} &\pi i\hbar\widehat{d}_i\frac{\partial}{\partial a_i} \pm \sum_{j\neq k}\varepsilon'_{ij}a_j \pm \varepsilon'_{ik}a'_k \mapsto \pi i\hbar\widehat{d}_i\frac{\partial}{\partial a_i} \\ &\pm \sum_{j\neq k}\varepsilon_{ij}a_j \pm \varepsilon_{ik}\left(a_k - \phi^{\hbar_k}\left(-\mathrm{sgn}(\pm\varepsilon_{\mathrm{ik}})(\pi\mathrm{i}\hbar\widehat{\mathrm{d}}_{\mathrm{k}}\frac{\partial}{\partial \mathrm{a_k}} \pm \sum_{\mathrm{j}}\varepsilon_{\mathrm{kj}}\mathrm{a_j})\right)\right), \end{aligned} \tag{17}$$

and

$$\pi i\hbar\widehat{d}_k\frac{\partial}{\partial a'_k} \pm \sum_j \varepsilon'_{kj}a_j \mapsto -\Big(\pi i\hbar\widehat{d}_k\frac{\partial}{\partial a_k} \pm \sum_j\varepsilon_{kj}a_j\Big). \tag{18}$$

Here we use the following conventions. The signs $\pm$ in our formulas always use either $+$ everywhere, or $-$ everywhere, so $\mp := -\pm$. Thus we have one set of equations corresponding to the upper signs and another one to the lower signs.

Observe that $\varepsilon_{kk} = 0$. The relation (18) is satisfied by (15) if and only if

$$-\varepsilon'_{kj} = \varepsilon_{kj} \quad \text{and} \quad \frac{\partial}{\partial a'_k} \longmapsto -\frac{\partial}{\partial a_k}.$$

Since these two conditions are evidently valid, we have the relation (18).

Substituting (17) into (15), one gets the identities

$$\begin{aligned} &\int\left(\pi i\hbar\widehat{d}_i\frac{\partial G}{\partial a_i}f + \pi i\hbar\widehat{d}_iG\frac{\partial f}{\partial a_i} \pm \sum_{j\neq k}\varepsilon'_{ij}a_jGf \pm \varepsilon'_{ik}a'_kGf\right)da_k \\ &\qquad = \int \pi i\hbar\widehat{d}_iG\frac{\partial f}{\partial a_i} \pm \sum_{j\neq k}\varepsilon_{ij}a_jGf \\ &\qquad\quad \pm G\varepsilon_{ik}\left(a_k - \phi^{\hbar_k}\left(\mathrm{sgn}(\varepsilon_{\mathrm{ik}})(\mp\pi\mathrm{i}\hbar\widehat{\mathrm{d}}_{\mathrm{k}}\frac{\partial}{\partial\mathrm{a_k}} - \sum_{\mathrm{j}}\varepsilon_{\mathrm{kj}}\mathrm{a_j})\right)f\right)da_k. \end{aligned}$$

Since these identities should be valid for any f, one gets the following equations for the function G:

$$\left(\pi i\hbar\widehat{d}_i\frac{\partial}{\partial a_i} \pm \sum_{j\neq k}(\varepsilon'_{ij}-\varepsilon_{ij})a_j \mp \varepsilon_{ik}a_k \right.$$
$$\left. \pm\varepsilon_{ik}\phi^{\hbar_k}\left(\mathrm{sgn}(\varepsilon_{ik})(\pm\pi i\hbar\widehat{d}_k\frac{\partial}{\partial a_k} - \sum_j \varepsilon_{kj}a_j)\right)\right)G = 0.$$

Let us introduce the function $\widehat{G}$ related to G by the Fourier transform:

$$\widehat{G}(c) = \int e^{-\frac{a_k c}{\pi i\hbar}}G(a_k)da_k; \quad G(a_k) = \frac{1}{2\pi^2\hbar}\int e^{\frac{a_k c}{\pi i\hbar}}\widehat{G}(c)dc$$

(we omit the variables $a_1,\dots,a_{k-1},a_{k+1},\dots,a_n$ on which both G and $\widehat{G}$ depend). Taking into account the relations

$$\pi i\hbar\widehat{\frac{\partial G}{\partial a_k}} = c\widehat{G}, \quad \widehat{a_k G} = -\pi i\hbar\frac{\partial\widehat{G}}{\partial c},$$

one can get the equation for the function $\widehat{G}$:

$$\left(\pi i\hbar\widehat{d}_i\frac{\partial}{\partial a_i} \pm \sum_{j\neq k}(\varepsilon'_{ij}-\varepsilon_{ij})a_j \pm \varepsilon_{ik}\pi i\hbar\frac{\partial}{\partial c} \right.$$
$$\left. \pm\varepsilon_{ik}\phi^{\hbar_k}\left(\mathrm{sgn}(\varepsilon_{\mathrm{ik}})(\pm\widehat{\mathrm{d}}_{\mathrm{k}}\mathrm{c} - \sum_{\mathrm{j}} \varepsilon_{\mathrm{kj}}\mathrm{a}_{\mathrm{j}})\right)\right)\widehat{G} = 0. \quad (19)$$

Taking the sum and difference of the equations corresponding to the upper and lower signs one obtains

$$\left(2\pi i\hbar\widehat{d}_i\frac{\partial}{\partial a_i} + \varepsilon_{ik}\phi^{\hbar_k}\left(\mathrm{sgn}(\varepsilon_{\mathrm{ik}})\left(\widehat{\mathrm{d}}_{\mathrm{k}}\mathrm{c} - \sum_{\mathrm{j}}\varepsilon_{\mathrm{kj}}\mathrm{a}_{\mathrm{j}}\right)\right)\right.$$
$$\left. -\varepsilon_{ik}\phi^{\hbar_k}\left(\mathrm{sgn}(\varepsilon_{\mathrm{ik}})\left(-\widehat{\mathrm{d}}_{\mathrm{k}}\mathrm{c} - \sum_{\mathrm{j}}\varepsilon_{\mathrm{kj}}\mathrm{a}_{\mathrm{j}}\right)\right)\right)\widehat{G} = 0$$

and

$$\Bigg(2\sum_{j\neq k}(\varepsilon'_{ij}-\varepsilon_{ij})a_j + 2\varepsilon_{ik}\pi i h\frac{\partial}{\partial c} \\ + \varepsilon_{ik}\phi^{\hbar_k}\left(\mathrm{sgn}(\varepsilon_{\mathrm{ik}})\left(\widehat{\mathrm{d}}_{\mathrm{k}}\mathrm{c}-\sum_{\mathrm{j}}\varepsilon_{\mathrm{kj}}\mathrm{a}_{\mathrm{j}}\right)\right)+ \\ + \varepsilon_{ik}\phi^{\hbar_k}\left(\mathrm{sgn}(\varepsilon_{\mathrm{ik}})\left(-\widehat{\mathrm{d}}_{\mathrm{k}}\mathrm{c}-\sum_{\mathrm{j}}\varepsilon_{\mathrm{kj}}\mathrm{a}_{\mathrm{j}}\right)\right)\Bigg)\widehat{G}=0.$$

Observe that this is a system of $2n-2$ equations for a function of n variables. So it is an overdetermined system if $n>2$. Using the identities

$$\phi^{\hbar_{\mathrm{k}}}(\mathrm{sgn(a)b}) = \phi^{\hbar_{\mathrm{k}}}(\mathrm{b}) + (\mathrm{sgn(a)}-1)\mathrm{b}/2;\quad \widehat{\mathrm{d}}_{\mathrm{i}}\varepsilon_{\mathrm{ji}} = -\widehat{\mathrm{d}}_{\mathrm{j}}\varepsilon_{\mathrm{ij}},$$

they can be transformed into the form

$$2\pi i\hbar\frac{\partial\log\widehat{G}}{\partial a_i} = \widehat{d}_k^{-1}\varepsilon_{ki} \\ \times\left(\phi^{\hbar_k}\left(\widehat{d}_k c-\sum_j\varepsilon_{kj}a_j\right)-\phi^{\hbar_k}\left(-\widehat{d}_k c-\sum_j\varepsilon_{kj}a_j\right)+\widehat{d}_k c(\mathrm{sgn}(\varepsilon_{\mathrm{ik}})-1)\right) \tag{20}$$

and

$$2\pi i\hbar\frac{\partial\log\widehat{G}}{\partial c} = 2(\varepsilon_{ik})^{-1}\sum_{j\neq k}(\varepsilon_{ij}-\varepsilon'_{ij})a_j \tag{21}$$

$$-\phi^{\hbar_k}\left(\widehat{d}_k c-\sum_j\varepsilon_{kj}a_j\right)-\phi^{\hbar_k}\left(-\widehat{d}_k c-\sum_j\varepsilon_{kj}a_j\right)+(\mathrm{sgn}(\varepsilon_{\mathrm{ik}})-1)\sum_{\mathrm{j}}\varepsilon_{\mathrm{kj}}\mathrm{a}_{\mathrm{j}}.$$

Taking into account that

$$\varepsilon'_{ij}-\varepsilon_{ij} = \frac{|\varepsilon_{ik}|\varepsilon_{kj}+\varepsilon_{ik}|\varepsilon_{kj}|}{2},\qquad i,j\neq k,$$

we have the following identity:

$$2(\varepsilon_{ik})^{-1}(\varepsilon_{ij}-\varepsilon'_{ij})+(\mathrm{sgn}(\varepsilon_{\mathrm{ik}})-1)\varepsilon_{\mathrm{kj}} = (\mathrm{sgn}(\varepsilon_{\mathrm{jk}})-1)\varepsilon_{\mathrm{kj}}.$$

Multiplying it by a_j and taking the sum over $j\neq k$, we get

$$2(\varepsilon_{ik})^{-1}\sum_{j\neq k}(\varepsilon_{ij}-\varepsilon'_{ij})a_j+(\mathrm{sgn}(\varepsilon_{\mathrm{ik}})-1)\sum_{\mathrm{j}}\varepsilon_{\mathrm{kj}}\mathrm{a}_{\mathrm{j}} = \sum_{\mathrm{j}}(\mathrm{sgn}(\varepsilon_{\mathrm{jk}})-1)\varepsilon_{\mathrm{kj}}\mathrm{a}_{\mathrm{j}}.$$

Thus (21) is equivalent to

$$2\pi i\hbar\frac{\partial \log \widehat{G}}{\partial c} = -\phi^{\hbar_k}\left(\widehat{d}_k c - \sum_j \varepsilon_{kj}a_j\right)$$
$$- \phi^{\hbar_k}\left(-\widehat{d}_k c - \sum_j \varepsilon_{kj}a_j\right) + (\mathrm{sgn}(\varepsilon_{\mathrm{jk}}) - 1)\sum_{\mathrm{j}} \varepsilon_{\mathrm{kj}}\mathrm{a}_{\mathrm{j}}.$$

Therefore the solution of equations (20) and (21) is given by the formula

$$\widehat{G} = C\Phi^{\hbar_k}\left(\widehat{d}_k c - \sum_j \varepsilon_{kj}a_j\right)^{-1}\Phi^{\hbar_k}\left(-\widehat{d}_k c - \sum_j \varepsilon_{kj}a_j\right)$$
$$\times e^{c\sum_j(\mathrm{sgn}(\varepsilon_{\mathrm{jk}})-1))\varepsilon_{\mathrm{kj}}\mathrm{a}_{\mathrm{j}}/2\pi \mathrm{i}\hbar},$$

where C is an arbitrary constant. Taking $C = 2\pi^2\hbar$, one obtains the desired formula (14). The statement is proved. □

Lemma 11. *An integral operator given by the formula (15) for a certain function G intertwines the operators (16) if and only if the standard formula for the mutation of the function ε_{ij} holds.*

Proof. The proof of the theorem shows that this formula, as well as the formula for mutations of the A-coordinates, follows from the anzatz (15) and the mutation formulas for the quantized X-coordinates. □

Representation of the modular double of the chiral double of $\mathcal{X}_{|\mathbf{i}|,q}$. Combining Claim 2.2 and Theorem 10, we see that the collection of Hilbert spaces $\{L_2(\mathcal{A}^+_{\mathbf{i}})\}$ should provide a projective unitary $*$-representation of the *modular double* of the chiral double of $\mathcal{X}_{|\mathbf{i}|,q}$, defined as

$$\mathcal{X}_{|\mathbf{i}|,q} \times \mathcal{X}_{|\mathbf{i}^0|,q} \times \mathcal{X}_{|\mathbf{i}^\vee|,q^\vee} \times \mathcal{X}_{|\mathbf{i}^{0\vee}|,q^\vee}.$$

According to [FG3], the collection of Hilbert spaces $\{L_2(\mathcal{A}^+_{\mathbf{i}})\}$ should provide a representation of the modular double $\mathcal{D}_{|\mathbf{i}|,q} \times \mathcal{D}_{|\mathbf{i}^\vee|,q^\vee}$ of the *cluster double* of the quantum cluster $\mathcal{X}$-variety $\mathcal{X}_{|\mathbf{i}|,q}$. Since there is a canonical map of quantum spaces $\mathcal{D}_{|\mathbf{i}|,q} \longrightarrow \mathcal{X}_{|\mathbf{i}|,q} \times \mathcal{X}_{|\mathbf{i}^o|,q}$, this implies the above claim.

4 The quantum logarithm and dilogarithm functions

The proofs of all results listed above can be found in [FG3].

Recall the dilogarithm function

$$\mathrm{Li}_2(x) := -\int_0^x \log(1-t)dt.$$

The quantum logarithm function. It is the following function:

$$\phi^{\hbar}(z) := -2\pi\hbar \int_{\Omega} \frac{e^{-ipz}}{(e^{\pi p} - e^{-\pi p})(e^{\pi \hbar p} - e^{-\pi \hbar p})} dp, \tag{22}$$

where the contour Ω goes along the real axis from $-\infty$ to ∞, bypassing the origin from above.

Proposition 12. *The function $\phi^{\hbar}(x)$ enjoys the following properties:*

(**A1**)
$$\lim_{\hbar \to 0} \phi^{\hbar}(z) = \log(e^z + 1).$$

(**A2**)
$$\phi^{\hbar}(z) - \phi^{\hbar}(-z) = z.$$

(**A3**)
$$\overline{\phi^{\hbar}(z)} = \phi^{\hbar}(\overline{z}).$$

(**A4**)
$$\phi^{\hbar}(z)/\hbar = \phi^{1/\hbar}(z/\hbar).$$

(**A5**)
$$\phi^{\hbar}(z+i\pi\hbar) - \phi^{\hbar}(z-i\pi\hbar) = \frac{2\pi i\hbar}{e^{-z}+1}, \qquad \phi^{\hbar}(z+i\pi) - \phi^{\hbar}(z-i\pi) = \frac{2\pi i}{e^{-z/\hbar}+1}.$$

(A6) *The form $\phi^{\hbar}(z)dz$ is meromorphic with poles at the points $\{\pi i((2m-1) + (2n-1)\hbar) | m,n \in \mathbb{N}\}$ with residues $2\pi i\hbar$ and at the points $\{-\pi i((2m-1) + (2n-1)\hbar) | m,n \in \mathbb{N}\}$ with residues $-2\pi i\hbar$.*

The quantum dilogarithm. Recall the quantum dilogarithm function:

$$\Phi^{\hbar}(z) := \exp\Big(-\frac{1}{4}\int_{\Omega} \frac{e^{-ipz}}{\mathrm{sh}(\pi p)\mathrm{sh}(\pi\hbar p)} \frac{dp}{p}\Big).$$

It goes back to Barnes [Ba], and was used by Baxter [Bax], Faddeeev [Fad], and others.

Proposition 13. *The function $\Phi^{\hbar}(x)$ enjoys the following properties:*

(**B**)
$$2\pi i\hbar \, d\log \Phi^{\hbar}(z) = \phi^{\hbar}(z)dz$$

(**B0**)
$$\lim_{\Re z \to -\infty} \Phi^{\hbar}(z) = 1.$$

Here the limit is taken along a line parallel to the real axis.

(**B1**)
$$\lim_{\hbar \to 0} \Phi^{\hbar}(z) / \exp \frac{-\mathrm{Li}_2(-e^z)}{2\pi i\hbar} = 1.$$

(**B2**)
$$\Phi^\hbar(z)\Phi^\hbar(-z) = \exp\left(\frac{z^2}{4\pi i\hbar}\right) e^{-\frac{\pi i}{12}(\hbar+\hbar^{-1})}.$$

(**B3**) $\overline{\Phi^\hbar(z)} = (\Phi^\hbar(\overline{z}))^{-1}$. *In particular,* $|\Phi^\hbar(z)| = 1$ *for* $z \in \mathbb{R}$.

(**B4**)
$$\Phi^\hbar(z) = \Phi^{1/\hbar}(z/\hbar).$$

(**B5**) $\Phi^\hbar(z + 2\pi i\hbar) = \Phi^\hbar(z)(1 + qe^z), \qquad \Phi^\hbar(z + 2\pi i) = \Phi^\hbar(z)(1 + q^\vee e^{z/\hbar}).$

*(***B6***) The function* $\Phi^\hbar(z)dz$ *is meromorphic with poles at the points*

$$\{\pi i((2m-1) + (2n-1)\hbar) | m, n \in \mathbb{N}\}$$

and zeros at the points

$$\{-\pi i((2m-1) + (2n-1)\hbar) | m, n \in \mathbb{N}\}.$$

References

[Ba] Barnes E.W., *The genesis of the double gamma function*, Proc. London Math. Soc. **31** (1899) 358–381.

[Bax] Baxter R., *Exactly solved models in statistical mechanics*, Academic Press (1982).

[BFZ] Berenstein A., Fomin S., Zelevinsky A., *Cluster algebras III, Upper bounds and double Bruhat cells*, `math.RT/0305434`. J. Amer. Math. Soc. **15** (2002), no. 2, 497–529.

[Fad] Faddeev L.D., *Discrete Heisenberg-Weyl group and modular group*, Lett. Math. Phys. **34** (1995), no. 3, 249–254.

[FCh] Fock, V. V., Chekhov, L. O., *Quantum Teichmüller spaces*, Teoret. Mat. Fiz. **120** (1999), no. 3, 511–528.

[FG1] Fock V.V., Goncharov A.B., *Moduli spaces of local systems and higher Teichmüller theory*, Publ. Math. IHES, n. **103** (2006) 1–212. `math.AG/0311149`.

[FG2] Fock V.V., Goncharov A.B., *Cluster ensembles, quantization and the dilogarithm I*, Ann Scient de Ec. Norm. Sup., t. **42**, N6, (2009), 865–929. `math.AG/0311245`.

[FG3] Fock V.V., Goncharov A.B., *The quantum dilogarithm and representations of quantum cluster varieties*, `math.QA/0702397`.

[FZI] Fomin S., Zelevinsky A., *Cluster algebras. I, Foundations*, J. Amer. Math. Soc. **15** (2002), no. 2, 497–529.

[K] Kashaev R.M., *Quantization of Teichmüller spaces and the quantum dilogarithm*, Lett. Math. Phys. 43 (1998), no. 2, 105–115.

Operads Revisited

Ezra Getzler

Northwestern University, Evanston, Illinois, USA
getzler@northwestern.edu

To Yuri Manin, many happy returns

Summary. Operads may be represented as symmetric monoidal functors on a small symmetric monoidal category. We discuss the axioms which must be imposed on a symmetric monoidal functor in order that it give rise to a theory similar to the theory of operads: we call such functors patterns. We also develop the enriched version of the theory, and show how it may be applied to axiomatize topological field theory.

Key words: operads, symmetric monoidal categories, enriched categories, topological field theory, locally presentable categories

2000 Mathematics Subject Classifications: 18D50, 57R56; 18D10, 18D20, 18C35

This paper presents an approach to operads related to recent work in topological field theory (Costello [4]). The idea is to represent operads as symmetric monoidal functors on a symmetric monoidal category T; we recall how this works for cyclic and modular operads and dioperads in Section 1. This point of view permits the construction of all sorts of variants on the notion of an operad. As our main example, we present a simplicial variant of modular operads, related to Segal's definition of quantum field theory, as modified for topological field theory (Getzler [10]). (This definition is only correct in a cocomplete symmetric monoidal category whose tensor product preserves colimits, but this covers most cases of interest.).

An operad $\mathcal{P}$ in the category Set of sets may be presented as a symmetric monoidal category $\mathtt{t}\mathcal{P}$, called the theory associated to $\mathcal{P}$ (Boardman and Vogt [2]). The category $\mathtt{t}\mathcal{P}$ has the natural numbers as its objects; tensor product is given by addition. The morphisms of $\mathtt{t}\mathcal{P}$ are built using the operad $\mathcal{P}$:

$$\mathtt{t}\mathcal{P}(m,n) = \coprod_{f:\{1,\dots,m\}\to\{1,\dots,n\}} \prod_{i=1}^{n} \mathcal{P}(|f^{-1}(i)|).$$

Y. Tschinkel and Y. Zarhin (eds.), *Algebra, Arithmetic, and Geometry*,
Progress in Mathematics 269, DOI 10.1007/978-0-8176-4745-2_16,

The category of $\mathcal{P}$-algebras is equivalent to the category of symmetric monoidal functors from $\mathsf{t}\mathcal{P}$ to Set; this reduces the study of algebras over operads to the study of symmetric monoidal functors.

The category of contravariant functors from the opposite category T° of a small category T to the category Set of sets

$$\mathsf{T}\hat{} = [\mathsf{T}^\circ, \mathsf{Set}]$$

is called the category of presheaves of T. If T is a symmetric monoidal category, then $\mathsf{T}\hat{}$ is too (Day [5]); its tensor product is the coend

$$V * W = \int^{A,B\in\mathsf{T}} \mathsf{T}(-, A\otimes B) \times V(A) \times W(B).$$

A symmetric monoidal functor $F : \mathsf{S} \to \mathsf{T}$ between symmetric monoidal categories S and T is a functor F together with a natural equivalence

$$\Phi : \otimes \circ F \times F \Longrightarrow F \circ \otimes.$$

The functor F is lax symmetric monoidal if Φ is only a natural transformation.

If $\tau : \mathsf{S} \to \mathsf{T}$ is a symmetric monoidal functor, it is not always the case that the induced functor

$$\tau\hat{} : \mathsf{T}\hat{} \longrightarrow \mathsf{S}\hat{}$$

is a symmetric monoidal functor; in general, it is only a lax symmetric monoidal functor, with respect to the natural transformation

$$\begin{tikzcd}
\tau\hat{}(V) * \tau\hat{}(W) \arrow[r, equal] \arrow[dddd, Rightarrow, "\Phi_{V,W}"'] & \int^{A,B\in\mathsf{S}} \mathsf{S}(-, A\otimes B) \times V(\tau A) \times W(\tau B) \arrow[d, Rightarrow, "\tau \text{ is a functor}"] \\
& \int^{A,B\in\mathsf{S}} \mathsf{T}(\tau(-), \tau(A\otimes B)) \times V(\tau A) \times W(\tau B) \arrow[d, Rightarrow, "\tau \text{ is symmetric monoidal}"] \\
& \int^{A,B\in\mathsf{S}} \mathsf{T}(\tau(-), \tau(A)\otimes \tau(B)) \times V(\tau A) \times W(\tau B) \arrow[d, Rightarrow, "\text{universality of coends}"] \\
& \\
\tau\hat{}(V * W) \arrow[r, equal] & \int^{A,B\in\mathsf{T}} \mathsf{T}(\tau(-), A\otimes B) \times V(A) \times W(B)
\end{tikzcd}$$

The following definition introduces the main object of study of this paper.

Definition. A **pattern** is a symmetric monoidal functor $\tau : \mathsf{S} \to \mathsf{T}$ between small symmetric monoidal categories S and T such that $\tau\hat{} : \mathsf{T}\hat{} \to \mathsf{S}\hat{}$ is a symmetric monoidal functor (in other words, the natural transformation Φ defined above is an equivalence).

Let τ be a pattern and let $\mathcal{C}$ be a symmetric monoidal category. A **τ-preoperad** in $\mathcal{C}$ is a symmetric monoidal functor from S to $\mathcal{C}$, and a **τ-operad** in $\mathcal{C}$ is a symmetric monoidal functor from T to $\mathcal{C}$. Denote by $\mathrm{PreOp}_\tau(\mathcal{C})$ and $\mathrm{Op}_\tau(\mathcal{C})$ the categories of τ-preoperads and τ-operads respectively. If $\mathcal{C}$ is a cocomplete symmetric monoidal category, there is an adjunction

$$\tau_* : \mathrm{PreOp}_\tau(\mathcal{C}) \rightleftharpoons \mathrm{Op}_\tau(\mathcal{C}) : \tau^*.$$

In Section 2, we prove that if τ is essentially surjective, $\mathcal{C}$ is cocomplete, and the functor $A \otimes B$ on $\mathcal{C}$ preserves colimits in each variable, then the functor τ^* is monadic. We also prove that if S is a free symmetric monoidal category and $\mathcal{C}$ is locally finitely presentable, then $\mathrm{Op}_\tau(\mathcal{C})$ is locally finitely presentable.

Patterns generalize the coloured operads of Boardman and Vogt [2]: colored operads may be identified with those patterns such that S is the free symmetric category generated by a discrete category (called the set of colors). Operads are themselves algebras for a coloured operad, whose colours are the natural numbers (cf. Berger and Moerdijk [1]), but it is more natural to think of the color n as having nontrivial automorphisms, namely the symmetric group $\mathbb{S}_n$.

The definition of a pattern may be applied in the setting of simplicial categories, or even more generally, enriched categories. We define and study patterns enriched over a symmetric monoidal category $\mathcal{V}$ in Section 2; we recall those parts of the formalism of enriched categories which we will need in the appendix.

In Section 3, we present examples of a simplicial pattern which arises in topological field theory. The most interesting of these is related to modular operads. Let cob be the (simplicial) groupoid of diffeomorphisms between compact connected oriented surfaces with oriented boundary. We define a pattern whose underlying functor is an inclusion $\mathbb{S}\wr\mathsf{cob} \hookrightarrow \mathsf{Cob}$. Here, Cob is a (simplicial) category whose objects are (not necessarily connected) compact oriented surfaces with oriented boundary, and whose morphisms incorporate gluing of surfaces along boundaries of opposite orientation and diffeomoporhisms.

Let $S_{g,n}$ be a compact oriented surface of genus g with n boundary circles, and let the mapping class group be the group of components of the oriented diffeomorphism group:

$$\Gamma_{g,n} = \pi_0(\mathrm{Diff}_+(S_{g,n})).$$

The simplicial groupoid cob has a skeleton

$$\coprod_{g,n} [\{+,-\}^n / \mathrm{Diff}_+(S_{g,n})].$$

Here, $[\{+,-\}^n / \mathrm{Diff}_+(S_{g,n})]$ is the groupoid associated to the action of the group $\mathrm{Diff}_+(S_{g,n})$ on the set $\{+,-\}^n$: its objects are the elements of the set $\{+,-\}^n$, which describe whether the ith marked point is incoming $(+)$ or outgoing $(-)$, while its morphisms are elements of $\{+,-\}^n \times \mathrm{Diff}_+(S_{g,n})$.

If $2g-2+n>0$, there is a homotopy equivalence $\mathrm{Diff}_+(S_{g,n}) = \Gamma_{g,n}$. Denote the generators of $B_n \subset B_n \wr \mathbb{Z}$ by $\{b_1,\dots,b_{n-1}\}$ and the generators of $\mathbb{Z}^n \subset B_n \wr \mathbb{Z}$ by $\{t_1,\dots,t_n\}$; Moore and Seiberg show [21, Appendix B.1] that $\Gamma_{0,n}$ is isomorphic to the quotient of $B_n \wr \mathbb{Z}$ by its subgroup $\langle (b_1b_2\cdots b_{n-1}t_n)^n, b_1b_2\cdots b_{n-1}t_n^2b_{n-1}\cdots b_1\rangle$. It follows that this pattern bears an analogous relationship to modular dioperads that braided operads (Fiedorowicz [8]) bear to symmetric operads.

1 Modular operads as symmetric monoidal functors

In this section, we define modular operads in terms of the symmetric monoidal category of dual graphs. Although we do not assume familiarity with the original definition (Getzler and Kapranov [11]; see also Markl et al. [18]), this section will certainly be easier to understand if this is not the reader's first brush with the subject.

We also show how modifications of this construction, in which dual graphs are replaced by forests, or by directed graphs, and yield cyclic operads and modular dioperads. Finally, we review the definition of algebras for modular operads and dioperads.

Graphs

A graph Γ consists of the following data:

(i) finite sets $V(\Gamma)$ and $F(\Gamma)$, the sets of vertices and flags of the graph;
(ii) a function $p : F(\Gamma) \to V(\Gamma)$, whose fiber $p^{-1}(v)$ is the set of flags of the graph meeting at the vertex v;
(iii) an involution $\sigma : F(\Gamma) \to F(\Gamma)$, whose fixed points are called the legs of Γ, and whose remaining orbits are called the edges of Γ.

We denote by $L(\Gamma)$ and $E(\Gamma)$ the sets of legs and edges of Γ, and by $n(v) = |p^{-1}(v)|$ the number of flags meeting a vertex v.

To a graph is associated a one-dimensional cell complex, with 0-cells $V(\Gamma) \sqcup L(\Gamma)$, and 1-cells $E(\Gamma) \sqcup L(\Gamma)$. The 1-cell associated to an edge $e = \{f, \sigma(f)\}$, $f \in F(G)$, is attached to the 0-cells corresponding to the vertices $p(f)$ and $p(\sigma(f))$ (which may be equal), and the 1-cell associated to a leg $f \in L(G) \subset F(G)$ is attached to the 0-cells corresponding to the vertex $p(f)$ and the leg f itself.

The edges of a graph Γ define an equivalence relation on its vertices $V(\Gamma)$; the components of the graph are the equivalence classes with respect to this relation. Denote the set of components by $\pi_0(\Gamma)$. The Euler characteristic of a component C of a graph is $e(C) = |V(C)| - |E(C)|$. Denote by $n(C)$ the number of legs of a component C.

Dual graphs

A dual graph is a graph $\varGamma$ together with a function $g : V(\varGamma) \to \mathbb{N}$. The natural number $g(v)$ is called the genus of the vertex v.

The genus $g(C)$ of a component C of a dual graph $\varGamma$ is defined by the formula

$$g(C) = \sum_{v \in C} g(v) + 1 - e(C).$$

The genus of a component is a nonnegative integer, which equals 0 if and only if C is a tree and $g(v) = 0$ for all vertices v of C.

A **stable** graph is a dual graph $\varGamma$ such that for all vertices $v \in V(\varGamma)$, the integer $2g(v) - 2 + n(v)$ is positive. In other words, if $g(v) = 0$, then $n(v)$ is at least 3, while if $g(v) = 1$, then $n(v)$ is nonzero. If $\varGamma$ is a stable graph, then $2g(C) - 2 + |L(C)| > 0$ for all components C of $\varGamma$.

The symmetric monoidal category $\mathcal{G}$

The objects of the category $\mathcal{G}$ are the dual graphs $\varGamma$ whose set of edges $E(\varGamma)$ is empty. Equivalently, an object of $\mathcal{G}$ is a pair of finite sets L and V and functions $p : L \to V$, $g : V \to \mathbb{N}$.

A morphism of $\mathcal{G}$ with source (L_1, V_1, p_1, g_1) and target (L_2, V_2, p_2, g_2) is a pair consisting of a dual graph $\varGamma$ with $F(\varGamma) = L_1$, $V(\varGamma) = V_1$, and $g(\varGamma) = g_1$, together with isomorphisms $\alpha : L_2 \to L(\varGamma)$ and $\beta : V_2 \to \pi_0(\varGamma)$ such that $p \circ \alpha = \beta \circ p_2$ and $\alpha^* g = g_2$.

The definition of the composition of two morphisms $\varGamma = \varGamma_2 \circ \varGamma_1$ in $\mathcal{G}$ is straightforward. We have dual graphs $\varGamma_1$ and $\varGamma_2$, with

$$\begin{aligned} F(\varGamma_1) &= L_1, & V(\varGamma_1) &= V_1, & g(\varGamma_1) &= g_1, \\ F(\varGamma_2) &= L_2, & V(\varGamma_2) &= V_2, & g(\varGamma_2) &= g_2, \end{aligned}$$

together with isomorphisms

$$\begin{aligned} \alpha_1 &: L_2 \longrightarrow L(\varGamma_1), & \beta_1 &: V_2 \longrightarrow \pi_0(\varGamma_1), \\ \alpha_2 &: L_3 \longrightarrow L(\varGamma_2), & \beta_2 &: V_3 \longrightarrow \pi_0(\varGamma_2). \end{aligned}$$

Since the source of $\varGamma = \varGamma_2 \circ \varGamma_1$ equals the source of $\varGamma_1$, we see that $p : F(\varGamma) \to V(\varGamma)$ is identified with $p : L_1 \to V_1$. The involution σ of L_1 is defined as follows: if $f \in F(\varGamma) = L_1$ lies in an edge of $\varGamma_1$, then $\sigma(f) = \sigma_1(f)$, while if f is a leg of $\varGamma_1$, then $\sigma(f) = \alpha_1(\sigma_2(\alpha_1^{-1}(f)))$. The isomorphisms α and β of $\varGamma$ are simply the isomorphisms α_2 and β_2 of $\varGamma_2$.

In other words, $\varGamma_2 \circ \varGamma_1$ is obtained from $\varGamma_1$ by gluing those pairs of legs of $\varGamma_1$ together which correspond to edges of $\varGamma_2$. It is clear that composition in $\mathcal{G}$ is associative.

The tensor product on objects of $\mathcal{G}$ extends to morphisms, making $\mathcal{G}$ into a symmetric monoidal category.

A morphism in $\mathcal{G}$ is invertible if the underlying stable graph has no edges, in other words, if it is simply an isomorphism between two objects of $\mathcal{G}$. Denote by $\mathcal{H}$ the groupoid consisting of all invertible morphisms of $\mathcal{G}$, and by $\tau : \mathcal{H} \hookrightarrow \mathcal{G}$ the inclusion. The groupoid $\mathcal{H}$ is the free symmetric monoidal functor generated by the groupoid h consisting of all morphisms of $\mathcal{G}$ with connected domain.

The category $\mathcal{G}$ has a small skeleton; for example, the full subcategory of $\mathcal{G}$ in which the sets of flags and vertices of the objects are subsets of the set of natural numbers. The tensor product of $\mathcal{G}$ takes us outside this category, but it is not hard to define an equivalent tensor product for this small skeleton. We will tacitly replace $\mathcal{G}$ by this skeleton, since some of our constructions will require that $\mathcal{G}$ be small.

Modular operads as symmetric monoidal functors

The following definition of modular operads may be found in Costello [4]. Let $\mathcal{C}$ be a symmetric monoidal category which is cocomplete, and such that the functor $A \otimes B$ preserves colimits in each variable. (This last condition is automatic if $\mathcal{C}$ is a closed symmetric monoidal category.) A **modular preoperad** in $\mathcal{C}$ is a symmetric monoidal functor from $\mathcal{H}$ to $\mathcal{C}$, and a **modular operad** in $\mathcal{C}$ is a symmetric monoidal functor from $\mathcal{G}$ to $\mathcal{C}$.

Let $\mathcal{A}$ be a small symmetric monoidal category, and let $\mathcal{B}$ be symmetric monoidal category. Denote by $[\mathcal{A}, \mathcal{B}]$ the category of functors and natural equivalences from $\mathcal{A}$ to $\mathcal{B}$, and by $[\![\mathcal{A}, \mathcal{B}]\!]$ the category of symmetric monoidal functors and monoidal natural equivalences.

With this notation, the categories of modular operads, respectively preoperads, in a symmetric monoidal category $\mathcal{C}$ are $\mathrm{Mod}(\mathcal{C}) = [\![\mathcal{G}, \mathcal{C}]\!]$ and $\mathrm{PreMod}(\mathcal{C}) = [\![\mathcal{H}, \mathcal{C}]\!]$.

Theorem. *i) There is an adjunction*

$$\tau_* : \mathrm{PreMod}(\mathcal{C}) \rightleftharpoons \mathrm{Mod}(\mathcal{C}) : \tau^*,$$

where τ^* is restriction along $\tau : \mathcal{H} \hookrightarrow \mathcal{G}$, and τ_* is the coend

$$\tau_*\mathcal{P} = \int^{A \in \mathcal{H}} \mathcal{G}(A, -) \times \mathcal{P}(A).$$

(ii) The functor

$$\tau^* : \mathrm{Mod}(\mathcal{C}) \longrightarrow \mathrm{PreMod}(\mathcal{C})$$

is monadic. That is, there is an equivalence of categories

$$\mathrm{Mod}(\mathcal{C}) \simeq \mathrm{PreMod}(\mathcal{C})^{\mathbb{T}},$$

where $\mathbb{T}$ is the monad $\tau^*\tau_*$.

(iii) The category $\mathrm{Mod}(\mathcal{C})$ *is locally finitely presentable if $\mathcal{C}$ is locally finitely presentable.*

In the original definition of modular operads (Getzler and Kapranov [11]), there was an additional stability condition, which may be phrased in the following terms. Denote by $\mathcal{G}_+$ the subcategory of $\mathcal{G}$ consisting of stable graphs, and let $\mathcal{H}_+ = \mathcal{H} \cap \mathcal{G}_+$. Then a stable modular preoperad in $\mathcal{C}$ is a symmetric monoidal functor from $\mathcal{H}_+$ to $\mathcal{C}$, and a stable modular operad in $\mathcal{C}$ is a symmetric monoidal functor from $\mathcal{G}_+$ to $\mathcal{C}$. These categories of stable modular preoperads and operads are equivalent to the categories of stable $\mathbb{S}$-modules and modular operads of [11].

Algebras for modular operads

To any object M of a closed symmetric monoidal category $\mathcal{C}$ is associated the monoid $\mathrm{End}(M) = [M, M]$: an A-module is a morphism of monoids $\rho : A \to \mathrm{End}(M)$. The analogue of this construction for modular operads is called a $\mathcal{P}$-algebra.

A bilinear form with domain M in a symmetric monoidal category $\mathcal{C}$ is a morphism $t : M \otimes M \to \mathbb{1}$ such that $t \circ \sigma = t$. Associated to a bilinear form (M, t) is a modular operad $\mathsf{End}(M, t)$, defined on a connected object $\varGamma$ of $\mathcal{G}$ to be

$$\mathsf{End}(M, t)(\varGamma) = M^{\otimes L(\varGamma)}.$$

If $\mathcal{P}$ is a modular operad, a $\mathcal{P}$-algebra M is an object M of $\mathcal{C}$, a bilinear form $t : M \otimes M \to \mathbb{1}$, and a morphism of modular operads $\rho : \mathcal{P} \to \mathsf{End}(M, t)$.

This modular operad has an underlying stable modular operad $\mathsf{End}_+(M, t)$, defined by restriction to the stable graphs in $\mathcal{G}$.

Cyclic operads

A variant of the above definition of modular operads is obtained by taking the subcategory of forests $\mathcal{G}_0$ in the category $\mathcal{G}$: a forest is a graph each component of which is simply connected. Symmetric monoidal functors on $\mathcal{G}_0$ are cyclic operads.

Denote the cyclic operad underlying $\mathsf{End}(M, t)$ by $\mathsf{End}_0(M, t)$. If $\mathcal{P}$ is a cyclic operad, a $\mathcal{P}$-algebra M is a bilinear form $t : M \otimes M \to \mathbb{1}$ and a morphism of cyclic operads $\rho : \mathcal{P} \to \mathsf{End}_0(M, t)$.

There is a functor $\mathcal{P} \mapsto \mathcal{P}_0$, which associates to a modular operad its underlying cyclic operad: this is the restriction functor from $[\![\mathcal{G}, \mathcal{C}]\!]$ to $[\![\mathcal{G}_0, \mathcal{C}]\!]$.

There is also a stable variant of cyclic operads, in which $\mathcal{G}_0$ is replaced by its stable subcategory $\mathcal{G}_{0+}$, defined by restricting to forests in which each vertex meets at least three flags.

Dioperads and modular dioperads

Another variant of the definition of modular operads is obtained by replacing the graphs $\varGamma$ in the definition of modular operads by digraphs (directed graphs):

A **digraph** Γ is a graph together with a partition

$$F(\Gamma) = F_+(\Gamma) \sqcup F_-(\Gamma)$$

of the flags into outgoing and incoming flags, such that each edge has one outgoing and one incoming flag. Each edge of a digraph has an orientation, running towards the outgoing flag. The set of legs of a digraph is partitioned into the outgoing and incoming legs: $L_\pm(\Gamma) = L(\Gamma) \cap F_\pm(\Gamma)$. Denote the number of outgoing and incoming legs by $n_\pm(\Gamma)$.

A dual digraph is a digraph together with a function $g : V(\Gamma) \to \mathbb{N}$. Imitating the construction of the symmetric monoidal category of dual graphs $\mathcal{G}$, we may construct a symmetric monoidal category of dual digraphs $\mathcal{D}$. A **modular dioperad** is a symmetric monoidal functor on $\mathcal{D}$. (These are the wheeled props studied in a recent preprint of Merkulov [20]).

If $M_\pm$ are objects of $\mathcal{C}$ and $t : M_+ \otimes M_- \to \mathbb{1}$ is a pairing, we may construct a modular dioperad $\mathsf{End}^{\rightharpoonup}(M_\pm, t)$, defined on a connected object Γ of $\mathcal{D}$ to be

$$\mathsf{End}^{\rightharpoonup}(M)(\Gamma) = M_+^{\otimes L_+(\Gamma)} \otimes M_-^{\otimes L_-(\Gamma)}. \tag{1}$$

If $\mathcal{P}$ is a modular dioperad, a $\mathcal{P}$-algebra is a pairing $t : M_+ \otimes M_- \to \mathbb{1}$ and a morphism of modular dioperads $\rho : \mathcal{P} \to \mathsf{End}^{\rightharpoonup}(M_\pm, t)$.

A directed forest is a directed graph each component of which is simply connected; let $\mathcal{D}_0$ be the subcategory of $\mathcal{D}$ of consisting of directed forests. A **dioperad** is a symmetric monoidal functor on $\mathcal{D}_0$ (Gan [9]).

If M is an object of $\mathcal{C}$, we may construct a dioperad $\mathsf{End}_0^{\rightharpoonup}(M)$, defined on a connected object Γ of $\mathcal{D}_0$ to be

$$\mathsf{End}_0^{\rightharpoonup}(M)(\Gamma) = \mathrm{Hom}\big(M^{\otimes L_-(\Gamma)}, M^{\otimes L_+(\Gamma)}\big).$$

When $M = M_-$ is a rigid object with dual $M^\vee = M_+$, this is a special case of (1). If $\mathcal{P}$ is a dioperad, a $\mathcal{P}$-algebra M is an object M of $\mathcal{C}$ and a morphism of dioperads $\rho : \mathcal{P} \to \mathsf{End}_0^{\rightharpoonup}(M)$.

props

MacLane's notion [17] of a prop also fits into the above framework. There is a subcategory $\mathcal{D}_{\mathsf{P}}$ of $\mathcal{D}$, consisting of all dual digraphs Γ such that each vertex has genus 0, and Γ has no directed circuits. A prop is a symmetric monoidal functor on $\mathcal{D}_{\mathsf{P}}$. (This follows from the description of free props in Etingof [7].) Note that $\mathcal{D}_0$ is a subcategory of $\mathcal{D}_{\mathsf{P}}$: thus, every prop has an underlying dioperad. Note also that the dioperad $\mathsf{End}_0^{\rightharpoonup}(M)$ is in fact a prop; if $\mathcal{P}$ is a prop, we may define a $\mathcal{P}$-algebra M to be an object M of $\mathcal{C}$ and a morphism of props $\rho : \mathcal{P} \to \mathsf{End}_0^{\rightharpoonup}(M)$.

2 Patterns

Patterns abstract the approach to modular operads sketched in Section 1. This section develops the theory of patterns enriched over a complete, cocomplete, closed symmetric monoidal category $\mathcal{V}$. In fact, we are mainly interested in the cases in which $\mathcal{V}$ is the category Set of sets or the category sSet of simplicial sets. We refer to the appendices for a review of the needed enriched category theory.

Symmetric monoidal $\mathcal{V}$-categories

Let $\mathcal{A}$ be a $\mathcal{V}$-category. Denote by $\mathbb{S}_n$ the symmetric group on n letters. The wreath product $\mathbb{S}_n \wr \mathcal{A}$ is the $\mathcal{V}$-category

$$\mathbb{S}_n \wr \mathcal{A} = \mathbb{S}_n \times \mathcal{A}^n.$$

If $\alpha, \beta \in \mathbb{S}_n$, the composition of morphisms $(\alpha, \varphi_1, \dots, \varphi_n)$ and $(\beta, \psi_1, \dots, \psi_n)$ is

$$(\beta \circ \alpha, \psi_{\alpha_1} \circ \varphi_1, \dots, \psi_{\alpha_n} \circ \varphi_n).$$

Define the wreath product $\mathbb{S} \wr \mathcal{A}$ to be

$$\mathbb{S} \wr \mathcal{A} = \coprod_{n=0}^{\infty} \mathbb{S}_n \wr \mathcal{A}.$$

In fact, $\mathbb{S} \wr (-)$ is a 2-functor from the 2-category $\mathcal{V}\text{-}\mathsf{Cat}$ to itself. This 2-functor underlies a 2-monad $\mathbb{S} \wr (-)$ on $\mathcal{V}\text{-}\mathsf{Cat}$: the composition

$$m : \mathbb{S} \wr \mathbb{S} \wr (-) \longrightarrow \mathbb{S} \wr (-)$$

is induced by the natural inclusions

$$(\mathbb{S}_{n_1} \times \mathcal{A}_1^n) \times \cdots \times (\mathbb{S}_{n_k} \times \mathcal{A}_k^n) \hookrightarrow \mathbb{S}_{n_1 + \cdots + n_k} \times \mathcal{A}^{n_1 + \cdots + n_k},$$

and the unit $\eta : 1_{\mathcal{V}\text{-}\mathsf{Cat}} \to \mathbb{S} \wr (-)$ is induced by the natural inclusion

$$\mathcal{A} \cong \mathbb{S}_1 \times \mathcal{A} \hookrightarrow \mathbb{S} \wr \mathcal{A}.$$

Definition 2.1. A **symmetric monoidal** $\mathcal{V}$-category $\mathcal{C}$ is a pseudo $\mathbb{S} \wr (-)$-algebra in $\mathcal{V}\text{-}\mathsf{Cat}$.

If $\mathcal{C}$ is a symmetric monoidal category, we denote the object obtained by acting on the object $(A_1, \dots, A_n)$ of $\mathbb{S}_n \wr \mathcal{C}$ by $A_1 \otimes \cdots \otimes A_n$. When $n = 0$, we obtain an object 1 of $\mathcal{C}$, called the identity. When $n = 1$, we obtain a $\mathcal{V}$-endofunctor of $\mathcal{C}$, which is equivalent by the natural $\mathcal{V}$-equivalence ι in the definition of a pseudo $\mathbb{S}_n \wr \mathcal{C}$-algebra to the identity $\mathcal{V}$-functor. When $n = 2$, we obtain a $\mathcal{V}$-functor $(A, B) \mapsto A \otimes B$ from $\mathcal{C} \times \mathcal{C}$ to $\mathcal{C}$, called the tensor product. Up to $\mathcal{V}$-equivalence, all of the higher tensor products are obtained by iterating the tensor product $A \otimes B$: if $n > 2$, there is a natural $\mathcal{V}$-equivalence between the functors $A_1 \otimes \cdots \otimes A_n$ and $(A_1 \otimes \cdots \otimes A_{n-1}) \otimes A_n$.

$\mathcal{V}$-patterns

If $\mathcal{A}$ is a small symmetric monoidal $\mathcal{V}$-category, the $\mathcal{V}$-category of presheaves $\mathcal{A}\hat{}$ on $\mathcal{A}$ is a symmetric monoidal $\mathcal{V}$-category: the convolution of presheaves $V_1, \ldots, V_n \in \mathcal{A}\hat{}$ is the $\mathcal{V}$-coend

$$V_1 * \cdots * V_n = \int^{A_1,\ldots,A_n \in \mathcal{A}} V_1(A_1) \otimes \cdots \otimes V_n(A_n) \otimes y(A_1 \otimes \cdots \otimes A_n).$$

The Yoneda functor $y : \mathcal{A} \to \mathcal{A}\hat{}$ is a symmetric monoidal $\mathcal{V}$-functor. (See Day [5] and Im and Kelly [13].)

If $\tau : \mathsf{S} \to \mathsf{T}$ is a symmetric monoidal $\mathcal{V}$-functor between small symmetric monoidal $\mathcal{V}$-categories, the pullback functor $F\hat{}$ is a lax symmetric monoidal $\mathcal{V}$-functor; we saw this in the unenriched case in the introduction, and the proof in the enriched case is similar.

Definition 2.2. A **$\mathcal{V}$-pattern** is a symmetric monoidal $\mathcal{V}$-functor $\tau : \mathsf{S} \to \mathsf{T}$ between small symmetric monoidal $\mathcal{V}$-categories S and T such that

$$\tau\hat{} : \mathsf{T}\hat{} \longrightarrow \mathsf{S}\hat{}$$

is a symmetric monoidal $\mathcal{V}$-functor.

Let $\mathcal{C}$ be a symmetric monoidal $\mathcal{V}$-category that is cocomplete, and such that the $\mathcal{V}$-functors $A \otimes B$ preserves colimits in each variable. The $\mathcal{V}$-categories of τ-preoperads and τ-operads in $\mathcal{C}$ are respectively the $\mathcal{V}$-categories $\mathrm{PreOp}_\tau(\mathcal{C}) = [\![\mathsf{S}, \mathcal{C}]\!]$ and $\mathrm{Op}_\tau(\mathcal{C}) = [\![\mathsf{T}, \mathcal{C}]\!]$ of symmetric monoidal $\mathcal{V}$-functors from S and T to $\mathcal{C}$.

The monadicity theorem

We now construct a $\mathcal{V}$-functor τ_*, which generalizes the functor taking a preoperad to the free operad that it generates.

Proposition 2.3. *Let τ be a $\mathcal{V}$-pattern. The $\mathcal{V}$-adjunction*

$$\tau_* : [\mathsf{S}, \mathcal{C}] \rightleftharpoons [\mathsf{T}, \mathcal{C}] : \tau^*$$

induces a $\mathcal{V}$-adjunction between the categories of τ-preoperads and τ-operads

$$\tau_* : \mathrm{PreOp}_\tau(\mathcal{C}) \rightleftharpoons \mathrm{Op}_\tau(\mathcal{C}) : \tau^*.$$

Proof. Let G be a symmetric monoidal $\mathcal{V}$-functor from S to $\mathcal{C}$. The left Kan $\mathcal{V}$-extension $\tau_* G$ is the $\mathcal{V}$-coend

$$\tau_* G(B) = \int^{A \in \mathsf{S}} \mathsf{T}(\tau A, B) \otimes G(A).$$

For each n, there is a natural $\mathcal{V}$-equivalence

$$
\begin{aligned}
\tau_* G(B_1) \otimes \cdots \otimes \tau_* G(B_n) &= \textstyle\bigotimes_{k=1}^n \int^{A_k \in \mathsf{S}} \mathsf{T}(\tau A_k, B_k) \otimes G(A_k) \\
&\cong \textstyle\int^{A_1,\ldots,A_n \in \mathsf{S}} \bigotimes_k \mathsf{T}(\tau A_k, B_k) \otimes \bigotimes_{k=1}^n G(A_k) \\
&\qquad \text{since } \otimes \text{ preserves } \mathcal{V}\text{-coends} \\
&\cong \textstyle\int^{A_1,\ldots,A_n \in \mathsf{S}} \bigotimes_k \mathsf{T}(\tau A_k, B_k) \otimes G\big(\bigotimes_{k=1}^n A_k\big) \\
&\qquad \text{since } G \text{ is symmetric monoidal} \\
&\cong \textstyle\int^{A_1,\ldots,A_n \in \mathsf{S}} \bigotimes_k \mathsf{T}(\tau A_k, B_k) \otimes \int^{A \in \mathsf{S}} \mathsf{S}\big(A, \bigotimes_{k=1}^n A_k\big) \otimes G(A) \\
&\qquad \text{by the Yoneda lemma} \\
&\cong \textstyle\int^{A \in \mathsf{S}} \int^{A_1,\ldots,A_n \in \mathsf{S}} \bigotimes_{k=1}^n \mathsf{T}(\tau A_k, B_k) \otimes \mathsf{S}\big(A, \bigotimes_{k=1}^n A_k\big) \otimes G(A) \\
&\qquad \text{by Fubini's theorem for } \mathcal{V}\text{-coends} \\
&\cong \textstyle\int^{A \in \mathsf{S}} \mathsf{T}\big(\tau A, \bigotimes_{k=1}^n B_k\big) \otimes G(A) \\
&\qquad \text{since } \tau\hat{\ } \text{ is symmetric monoidal} \\
&= \tau_* G\Big(\bigotimes_{k=1}^n B_k\Big).
\end{aligned}
$$

This natural $\mathcal{V}$-equivalence makes $\tau_* G$ into a symmetric monoidal $\mathcal{V}$-functor.

The unit and counit of the $\mathcal{V}$-adjunction between τ_* and τ^* on $\mathrm{PreOp}_\tau(\mathcal{C})$ and $\mathrm{Op}_\tau(\mathcal{C})$ are now induced by the unit and counit of the $\mathcal{V}$-adjunction between τ_* and τ^* on $[\mathsf{S}, \mathcal{C}]$ and $[\mathsf{T}, \mathcal{C}]$. □

For the unenriched version of the following result on reflexive $\mathcal{V}$-coequalizers, see, for example, Johnstone [14], Corollary 1.2.12; the proof in the enriched case is identical.

Proposition 2.4. *Let $\mathcal{C}$ be a symmetric monoidal $\mathcal{V}$-category with reflexive $\mathcal{V}$-coequalizers. If the tensor product $A \otimes -$ preserves reflexive $\mathcal{V}$-coequalizers, then so does the functor $\mathbb{S} \wr (-)$.*

Corollary 2.5. *The $\mathcal{V}$-categories $\mathrm{PreOp}_\tau(\mathcal{C})$ and $\mathrm{Op}_\tau(\mathcal{C})$ have reflexive coequalizers.*

Proof. The $\mathcal{V}$-coequalizer $\mathcal{R}$ in $[\mathsf{S}, \mathcal{C}]$ of a reflexive parallel pair $\mathcal{P} \underset{g}{\overset{f}{\rightrightarrows}} \mathcal{Q}$ in $[\![\mathsf{S}, \mathcal{C}]\!]$ is computed pointwise: for each $X \in \mathrm{Ob}(\mathsf{S})$,

$$\mathcal{P}(X) \underset{g(X)}{\overset{f(X)}{\rightrightarrows}} \mathcal{Q}(X) \longrightarrow \mathcal{R}(X)$$

is a reflexive $\mathcal{V}$-coequalizer in $\mathcal{C}$. By Proposition 2.4, $\mathcal{R}$ is a symmetric monoidal $\mathcal{V}$-functor; thus, $\mathcal{R}$ is the $\mathcal{V}$-coequalizer of the reflexive pair $\mathcal{P} \underset{g}{\overset{f}{\rightrightarrows}} \mathcal{Q}$ in $[\![\mathsf{S}, \mathcal{C}]\!]$. The same argument works for $[\![\mathsf{T}, \mathcal{C}]\!]$. □

Proposition 2.6. *If τ is an essentially surjective $\mathcal{V}$-pattern, the $\mathcal{V}$-functor*

$$\tau^* : \mathrm{Op}_\tau(\mathcal{C}) \longrightarrow \mathrm{PreOp}_\tau(\mathcal{C})$$

creates reflexive $\mathcal{V}$-coequalizers.

Proof. The proof of Corollary 2.5 shows that the horizontal $\mathcal{V}$-functors in the diagram

$$\begin{tikzcd} [\![\mathsf{T},\mathcal{C}]\!] \arrow[r] \arrow[d, dashed, "\tau^*"'] & [\mathsf{T},\mathcal{C}] \arrow[d, "\tau^*"] \\ [\![\mathsf{S},\mathcal{C}]\!] \arrow[r] & [\mathsf{S},\mathcal{C}] \end{tikzcd}$$

create, and hence preserve, reflexive $\mathcal{V}$-coequalizers. The $\mathcal{V}$-functor

$$\tau^* : [\mathsf{T},\mathcal{C}] \longrightarrow [\mathsf{S},\mathcal{C}]$$

creates all $\mathcal{V}$-colimits, since $\mathcal{V}$-colimits are computed pointwise and τ is essentially surjective. It follows that the $\mathcal{V}$-functor $\tau^* : [\![\mathsf{T},\mathcal{C}]\!] \to [\![\mathsf{S},\mathcal{C}]\!]$ creates reflexive $\mathcal{V}$-coequalizers. $\square$

Recall that a $\mathcal{V}$-functor $R : \mathcal{A} \to \mathcal{B}$, with left adjoint $L : \mathcal{B} \to \mathcal{A}$, is $\mathcal{V}$-monadic if there is an equivalence of $\mathcal{V}$-categories $\mathcal{A} \simeq \mathcal{B}^{\mathbb{T}}$, where $\mathbb{T}$ is the $\mathcal{V}$-monad associated to the $\mathcal{V}$-adjunction

$$L : \mathcal{A} \rightleftharpoons \mathcal{B} : R.$$

The following is a variant of Theorem II.2.1 of Dubuc [6]; reflexive $\mathcal{V}$-coequalizers are substituted for contractible $\mathcal{V}$-coequalizers, but otherwise, the proof is the same.

Proposition 2.7. *A $\mathcal{V}$-functor $R : \mathcal{A} \to \mathcal{B}$, with left adjoint $L : \mathcal{B} \to \mathcal{A}$, is $\mathcal{V}$-monadic if $\mathcal{B}$ has, and R creates, reflexive $\mathcal{V}$-coequalizers.*

Corollary 2.8. *If τ is an essentially surjective $\mathcal{V}$-pattern, then the $\mathcal{V}$-functor*

$$\tau^* : \mathrm{Op}_\tau(\mathcal{C}) \longrightarrow \mathrm{PreOp}_\tau(\mathcal{C})$$

is $\mathcal{V}$-monadic.

In practice, the $\mathcal{V}$-patterns of interest all have the following property.

Definition 2.9. A $\mathcal{V}$-pattern $\tau : \mathsf{S} \to \mathsf{T}$ is **regular** if it is essentially surjective and S is equivalent to a free symmetric monoidal $\mathcal{V}$-category.

Denote by s a $\mathcal{V}$-category such that S is equivalent to the free symmetric monoidal $\mathcal{V}$-category $\mathbb{S}\wr\mathsf{s}$. The $\mathcal{V}$-category s may be thought of as a generalized set of colours; the theory associated to a colored operad is a regular pattern with discrete s.

Theorem 2.10. *If τ is a regular pattern and $\mathcal{C}$ is locally finitely presentable, then $\mathrm{Op}_\tau(\mathcal{C})$ is locally finitely presentable.*

Proof. The $\mathcal{V}$-category $\mathrm{PreOp}_\tau(\mathcal{C})$ is equivalent to the $\mathcal{V}$-category $[\mathsf{s}, \mathcal{C}]$, and hence is locally finitely presentable. By Lemma 3.7, $\mathrm{Op}_\tau(\mathcal{C})$ is cocomplete. The functor τ_* takes finitely presentable objects of $\mathcal{C}$ to finitely presentable objects of $\mathrm{Op}_\tau(\mathcal{C})$, and hence takes finitely presentable strong generators of $\mathrm{PreOp}_\tau(\mathcal{C})$ to finitely presentable strong generators of $\mathrm{Op}_\tau(\mathcal{C})$. □

This theorem has a generalization: if $\mathcal{C}$ is a locally α-presentable category (for α a regular cardinal), then $\mathrm{Op}_\tau(\mathcal{C})$ is locally α-presentable. (The proof is identical.)

3 The simplicial pattern Cob

In this section, we construct simplicial patterns $\mathsf{Cob} = \mathsf{Cob}[d]$, associated with gluing of oriented d-dimensional manifolds along components of their boundaries.

Definition 3.1. An object S of Cob consists of a compact oriented d-dimensional manifold W, a closed oriented $(d-1)$-dimensional manifold M, and a diffeomorphism $i : M \to \partial W$.

The manifold M is partitioned into outgoing and incoming parts

$$M = M_+ \sqcup M_-$$

by the orientations on W and M, according to whether the embedding i preserves or reverses orientation.

In the above definition, we permit the manifold W to be disconnected. In particular, it may be empty. (Note that an empty manifold of dimension d has a unique orientation.) Strictly speaking, the above definition should be refined so that the objects of Cob form a set: one way to do this is to decorate the data defining an object of Cob with embeddings of W and M into a Euclidean space $\mathbb{R}^N$.

We now define a simplicial set $\mathsf{Cob}(S_0, S_1)$ of morphisms between objects S_0 and S_1 of Cob.

Definition 3.2. A hypersurface γ in a d-dimensional manifold W consists of a closed oriented $(d-1)$-dimensional manifold N together with an embedding

$$\gamma : N \hookrightarrow W$$

in the interior of W.

Given a hypersurface γ in W, let $W[\gamma]$ be the manifold obtained by cutting W along the image of γ. The orientation of W induces an orientation of $W[\gamma]$, and the boundary of $W[\gamma]$ is naturally isomorphic to $\partial W \sqcup N \sqcup \overline{N}$, where $\overline{N}$ is a copy of the manifold N with reversed orientation.

A k-simplex in the simplicial set $\mathsf{Cob}(S_0, S_1)$ of morphisms from S_0 to S_1 consists of the following data:

(i) a closed oriented $(d-1)$-manifold N;
(ii) a k-simplex hypersurfaces in W_1, that is a commutative diagram
$$\begin{tikzcd} \Delta^k \times N \arrow[rr, hook, "\gamma"] \arrow[dr] & & \Delta^k \times W_1 \arrow[dl] \\ & \Delta^k & \end{tikzcd}$$
in which γ is an embedding;
(iii) a fibred diffeomorphism
$$\begin{tikzcd} \Delta^k \times W_0 \arrow[rr, "\varphi"] \arrow[dr] & & \Delta^k \times W_1[\gamma] \arrow[dl] \\ & \Delta^k & \end{tikzcd}$$
compatible with the orientations on its domain and target.

The composition of k-simplices $(N_i, \gamma_i, \varphi_i) \in \mathsf{Cob}(S_{i-1}, S_i)_k$, $i = 1, 2$, is the k-simplex consisting of the embedding
$$\begin{tikzcd} \Delta^k \times (N_1 \sqcup N_2) \arrow[rr, hook, "(\varphi_2 \circ \gamma_1) \sqcup \gamma_2"] \arrow[dr] & & \Delta^k \times W_2 \arrow[dl] \\ & \Delta^k & \end{tikzcd}$$
and the fibred diffeomorphism
$$\begin{tikzcd} \Delta^k \times W_0 \arrow[rr, "\varphi_2 \circ \varphi_1"] \arrow[dr] & & \Delta^k \times W_2[(\varphi_2 \circ \gamma_1) \sqcup \gamma_2] \arrow[dl] \\ & \Delta^k & \end{tikzcd}$$
Ignoring the diffeomorphisms φ_1 and φ_2, this composition amounts to taking the union of the disjoint hypersurfaces $\Delta^k \times N_1$ and $\Delta^k \times N_2$ in $\Delta^k \times W_2$. It is clear that composition is associative, and compatible with the face and degeneracy maps between simplices.

The identity 0-simplex 1_S in $\mathsf{Cob}(S, S)$ is associated to the empty hypersurface in S and the identity diffeomorphism of W.

The simplicial category Cob has a symmetric monoidal structure. The tensor product of Cob is simple to describe: it is disjoint union. When S_i,

$1 \leq i \leq k$, are objects of Cob decorated by embeddings in $\mathbb{R}^{N_i}$, we may, for example, embed the manifolds W and M underlying the object $S_1 \otimes \cdots \otimes S_k$ in $\mathbb{R}^{\max(N_i)+1}$ by composing the embeddings of W_i and M_i with the inclusion $\mathbb{R}^{N_i} \hookrightarrow \mathbb{R}^{\max(N_i)+1}$ defined by

$$(t_1, \dots, t_{N_i}) \mapsto (t_1, \dots, t_{N_i}, 0, \dots, 0, i).$$

Let $\mathsf{cob}[d]$ be the simplicial groupoid whose objects are those objects of $\mathsf{Cob}[d]$ such that W is connected, and whose morphisms are those morphisms of $\mathsf{Cob}[d]$ such that the hypersurface is empty. For example, the groupoid $\mathsf{cob}[2]$ is equivalent to the discrete groupoid

$$\bigsqcup_{g,n} [\{+,-\}^n / \operatorname{Diff}_+(S_{g,n})].$$

The embedding $\mathsf{cob} \hookrightarrow \mathsf{Cob}$ extends to an essentially surjective $\mathcal{V}$-functor $\mathbb{S} \wr \mathsf{cob} \to \mathsf{Cob}$.

We can now state the main theorem of this section.

Theorem 3.3. *The simplicial functor* $\mathsf{cob} \hookrightarrow \mathsf{Cob}$ *induces a regular simplicial pattern* $\mathbb{S} \wr \mathsf{cob} \to \mathsf{Cob}$.

Definition 3.4. A d-dimensional modular dioperad is an operad for the simplicial pattern $\mathbb{S} \wr \mathsf{cob} \to \mathsf{Cob}$.

One-dimensional modular dioperads

When $d = 1$, the category cob has a skeleton with five objects, the intervals I_-^-, I_-^+, I_+^- and I_+^+, representing the 1-manifold $[0,1]$ with the four different orientations of its boundary, and the circle S. The simplicial groups $\mathsf{cob}(I_a, I_a)$ are contractible, and the simplicial group $\mathsf{cob}(S,S)$ is homotopy equivalent to $\mathrm{SO}(2)$.

A 1-dimensional modular dioperad $\mathcal{P}$ in a discrete symmetric monoidal category $\mathcal{C}$ consists of associative algebras $A = \mathcal{P}(I_-^+)$ and $B = \mathcal{P}(I_+^-)$ in $\mathcal{C}$, a (A,B)-bimodule $Q = \mathcal{P}(I_-^-)$, a (B,A)-bimodule $R = \mathcal{P}(I_+^+)$, and an object $M = \mathcal{P}(S)$, together with morphisms

$$\alpha : Q \otimes_B R \longrightarrow A, \quad \beta : R \otimes_A Q \longrightarrow B, \quad \mathrm{tr}_A : A \longrightarrow M, \quad \mathrm{tr}_B : B \longrightarrow M.$$

Denote the left and right actions of A and B on Q and R by $\lambda_Q : A \otimes Q \to Q$, $\rho_Q : Q \otimes B \to Q$, $\lambda_R : B \otimes R \to R$ and $\rho_R : R \otimes A \to R$. The above data must in addition satisfy the following conditions:

- α and β are morphisms of (A,A)-bimodules and (B,B)-bimodules respectively;
- tr_A and tr_B are traces;
- $\lambda_Q \circ (\alpha \otimes Q) = \rho_Q \circ (R \otimes \beta) : Q \otimes_B R \otimes_A Q \to Q$;

- $\lambda_R \circ (\beta \otimes R) = \rho_R \circ (Q \otimes \alpha) : R \otimes_A Q \otimes_B R \to R$;
- $\operatorname{tr}_A \circ \alpha : Q \otimes_B R \to M$ and $\operatorname{tr}_B \circ \beta : R \otimes_A Q \to M$ are equal on the isomorphic objects $(Q \otimes_B R) \otimes_{A^\circ \otimes A} \mathbb{1}$ and $(R \otimes_A Q) \otimes_{B^\circ \otimes B} \mathbb{1}$.

That is, a 1-dimensional modular dioperad is the same thing as a **Morita context** (Morita [22]) $(A, B, Q, R, \alpha, \beta)$, together with compatible traces tr_A and tr_B to a module M.

Two-dimensional topological field theories

Let $\mathcal{D}$ be the discrete pattern introduced in Section 1 whose operads are modular dioperads. There is a natural morphism of patterns

$$\alpha : \mathsf{Cob}[2] \longrightarrow \mathcal{D};$$

thus, application of α^* to a modular dioperad gives rise to a 2-dimensional modular dioperad. But modular dioperads and 2-dimensional modular dioperads are quite different: underlying a modular dioperad is a sequence of $\mathbb{S}_n$-modules labelled by genus g, while underlying a 2-dimensional modular dioperad is a sequence of $\Gamma(S_{g,n})$-modules.

An example of a 2-dimensional modular dioperad is the terminal one, for which $\mathcal{P}(S)$ is the unit $\mathbb{1}$ for each surface S. Another one comes from conformal field theory: it associates to an oriented surface W the (contractible) space $\mathcal{N}(W)$ of conformal structures on W. (More accurately, $\mathcal{N}(W)$ is the simplicial set whose k-simplices are the smooth families of conformal structures on W parametrized by the k-simplex.) To define the structure of a 2-dimensional modular dioperad on $\mathcal{N}$ amounts to showing that conformal structures may be glued along circles: this is done by choosing a Riemannian metric in the conformal class which is flat in a neighbourhood of the boundary such that the boundary is geodesic and each of its components has length 1.

Another example of a 2-dimensional modular dioperad is the endomorphism operad $\alpha^* \mathsf{End}^{\rightharpoonup}(V_\pm, t)$ associated to a pairing $t : V_+ \otimes V_- \to \mathbb{1}$ in Cat. Explicitly, we have

$$\alpha^* \mathsf{End}(V_\pm, t)(S) = V_+^{\otimes \pi_0(M_+)} \otimes V_-^{\otimes \pi_0(M_-)}.$$

When Cat is the category of cochain complexes over the complex numbers, we often restruct attention to pairings which induce non-degenerate pairings $t : H^k(V_+) \otimes H^{-k}(V_-) \to \mathbb{1}$ on cohomology groups.

Definition 3.5. Let $\mathcal{P}$ be a 2-dimensional modular operad. A $\mathcal{P}$-algebra is a pairing $t : V_+ \otimes V_- \to \mathbb{1}$ together with a morphism of 2-dimensional modular operads

$$\rho : \mathcal{P} \longrightarrow \alpha_* \mathsf{End}(V_\pm, t).$$

A topological conformal field theory is a $C_*(\mathcal{N})$-algebra in the category of cochain complexes.

In the theory of infinite loop spaces, one defines an E_∞-algebra as an algebra for an operad $\mathcal{E}$ such that $\mathcal{E}(n)$ is contractible for all n. Similarly, as shown by Fiedorowicz [8], an E_2-algebra is an algebra for a braided operad $\mathcal{E}$ such that $\mathcal{E}(n)$ is contractible for all n. Motivated by this, we make the following definition.

Definition 3.6. A 2-dimensional topological field theory is an $\mathcal{E}$-algebra for a 2-dimensional modular dioperad $\mathcal{E}$ in the category of cochain complexes such that $\mathcal{E}(S)$ is quasi-isomorphic to $\mathbb{1}$ for all surfaces S.

In particular, a topological conformal field theory is a 2-dimensional topological field theory.

Appendix. Enriched categories

In this appendix, we recall some results of enriched category theory. Let $\mathcal{V}$ be a closed symmetric monoidal category; that is, $\mathcal{V}$ is a symmetric monoidal category such that the tensor product functor $-\otimes Y$ from $\mathcal{V}$ to itself has a right adjoint for all objects Y, denoted $[Y,-]$: in other words,

$$[X\otimes Y, Z] \cong [X,[Y,Z]].$$

Throughout this paper, we assume that $\mathcal{V}$ is complete and cocomplete. Denote by $A\mapsto A_0$ the continuous functor $A_0 = \mathcal{V}(\mathbb{1},A)$ from $\mathcal{V}$ to Set, where $\mathbb{1}$ is the unit of $\mathcal{V}$.

Let $\mathcal{V}$-Cat be the 2-category whose objects are $\mathcal{V}$-categories, whose 1-morphisms are $\mathcal{V}$-functors, and whose 2-morphisms are $\mathcal{V}$-natural transformations. Since $\mathcal{V}$ is closed, $\mathcal{V}$ is itself a $\mathcal{V}$-category.

Applying the functor $(-)_0 : \mathcal{V} \to \mathsf{Set}$ to a $\mathcal{V}$-category $\mathcal{C}$, we obtain its underlying category $\mathcal{C}_0$; in this way, we obtain a 2-functor

$$(-)_0 : \mathcal{V}\text{-}\mathsf{Cat} \longrightarrow \mathsf{Cat}.$$

Given a small $\mathcal{V}$-category $\mathcal{A}$, and a $\mathcal{V}$-category $\mathcal{B}$, there is a $\mathcal{V}$-category $[\mathcal{A},\mathcal{B}]$, whose objects are the $\mathcal{V}$-functors from $\mathcal{A}$ to $\mathcal{B}$ and such that

$$[\mathcal{A},\mathcal{B}](F,G) = \textstyle\int_{A\in\mathcal{A}} \mathcal{B}(FA,GA).$$

There is an equivalence of categories $\mathcal{V}\text{-}\mathsf{Cat}(\mathcal{A},\mathcal{B}) \simeq [\mathcal{A},\mathcal{B}]_0$.

The Yoneda embedding for $\mathcal{V}$-categories

The opposite of a $\mathcal{V}$-category $\mathcal{C}$ is the $\mathcal{V}$-category $\mathcal{C}^\circ$ with

$$\mathcal{C}^\circ(A,B) = \mathcal{C}(B,A).$$

If $\mathcal{A}$ is a small $\mathcal{V}$-category, denote by $\mathcal{A}\hat{}$ the $\mathcal{V}$-category of presheaves

$$\mathcal{A}\hat{} = [\mathcal{A}^\circ, \mathcal{V}].$$

By the $\mathcal{V}$-Yoneda lemma, there is a full, faithful $\mathcal{V}$-functor $y : \mathcal{A} \to \mathcal{A}\hat{}$, with

$$y(A) = \mathcal{A}(-, A), \quad A \in \mathrm{Ob}(\mathcal{A}).$$

If $F : \mathcal{A} \to \mathcal{B}$ is a $\mathcal{V}$-functor, denote by $F\hat{} : \mathcal{B}\hat{} \to \mathcal{A}\hat{}$ the $\mathcal{V}$-functor induced by F.

Cocomplete $\mathcal{V}$-categories

A $\mathcal{V}$-category $\mathcal{C}$ is **tensored** if there is a $\mathcal{V}$-functor $\otimes : \mathcal{V} \otimes \mathcal{C} \to \mathcal{C}$ together with a $\mathcal{V}$-natural equivalence of functors $\mathcal{C}(X \otimes A, B) \cong [X, \mathcal{C}(A, B)]$ from $\mathcal{V}^\circ \otimes \mathcal{C}^\circ \otimes \mathcal{C}$ to $\mathcal{V}$. For example, the $\mathcal{V}$-category $\mathcal{V}$ is itself tensored.

Let $\mathcal{A}$ be a small $\mathcal{V}$-category, let F be a $\mathcal{V}$-functor from $\mathcal{A}^\circ$ to $\mathcal{V}$, and let G be a $\mathcal{V}$-functor from $\mathcal{A}$ to a $\mathcal{V}$-category $\mathcal{C}$. If B is an object of $\mathcal{C}$, denote by $\langle G, B\rangle : \mathcal{C}^\circ \to \mathcal{V}$ the presheaf such that $\langle G, C\rangle(A) = [GA, C]$. The **weighted colimit** of G, with weight F, is an object $\int^{A\in\mathcal{A}} FA \otimes GA$ of $\mathcal{C}$ such that there is a natural isomorphism

$$\mathcal{C}\hat{}(F, \langle G, -\rangle) \cong \mathcal{C}\big(\textstyle\int^{A\in\mathcal{A}} FA \otimes GA, -\big).$$

If $\mathcal{C}$ is tensored, $\int^{A\in\mathcal{A}} F(A) \otimes G(A)$ is the coequalizer of the diagram

$$\coprod_{A_0, A_1 \in \mathrm{Ob}(\mathcal{A})} \mathcal{A}(A_0, A_1) \otimes FA_1 \otimes GA_0 \rightrightarrows \coprod_{A \in \mathrm{Ob}(\mathcal{A})} FA \otimes GA,$$

where the two morphisms are induced by the action on F and coaction on G respectively.

A $\mathcal{V}$-category $\mathcal{C}$ is **cocomplete** if it has all weighted colimits, or equivalently, if it satisfies the following conditions:

(i) the category $\mathcal{C}_0$ is cocomplete;
(ii) for each object A of $\mathcal{C}$, the functor $\mathcal{C}(-, A) : \mathcal{C}_0 \to \mathcal{V}$ transforms colimits into limits;
(iii) $\mathcal{C}$ is tensored.

In particular, $\mathcal{V}$-categories $\mathcal{A}\hat{}$ of presheaves are cocomplete.

Let $\mathcal{C}$ be a cocomplete $\mathcal{V}$-category. A $\mathcal{V}$-functor $F : \mathcal{A} \to \mathcal{B}$ between small $\mathcal{V}$-categories gives rise to a $\mathcal{V}$-adjunction

$$F_* : [\mathcal{A}, \mathcal{C}] \rightleftharpoons [\mathcal{B}, \mathcal{C}] : F^*,$$

that is, an adjunction in the 2-category $\mathcal{V}$-Cat. The functor F_* is called the (pointwise) left $\mathcal{V}$-Kan extension of along F; it is the $\mathcal{V}$-coend

$$F_*G(-) = \int^{A\in\mathcal{A}} \mathcal{A}(FA, -) \otimes GA.$$

Cocomplete categories of $\mathcal{V}$-algebras

A $\mathcal{V}$-monad $\mathbb{T}$ on a $\mathcal{V}$-category $\mathcal{C}$ is a monad in the full sub-2-category of $\mathcal{V}$-Cat with unique object $\mathcal{C}$. If $\mathcal{C}$ is small, this is the same thing as a monoid in the monoidal category $\mathcal{V}$-$\mathsf{Cat}(\mathcal{C},\mathcal{C})$.

The following is the enriched version of a result of Linton [16], and is proved in exactly the same way.

Lemma 3.7. *Let $\mathbb{T}$ be a $\mathcal{V}$-monad on a cocomplete $\mathcal{V}$-category $\mathcal{C}$ such that the $\mathcal{V}$-category of algebras $\mathcal{C}^{\mathbb{T}}$ has reflexive $\mathcal{V}$-coequalizers. Then the $\mathcal{V}$-category of $\mathbb{T}$-algebras $\mathcal{C}^{\mathbb{T}}$ is cocomplete.*

Proof. We must show that $\mathcal{C}^{\mathbb{T}}$ has all weighted colimits. Let $\mathcal{A}$ be a small $\mathcal{V}$-category, let F be a weight, and let $G : \mathcal{A} \to \mathcal{C}^{\mathbb{T}}$ be a diagram of $\mathbb{T}$-algebras. Then the weighted colimit $\int^{A\in\mathcal{A}} F(A) \otimes G(A)$ is a reflexive coequalizer

$$\mathbb{T}\big(\textstyle\int^{A\in\mathcal{A}} FA \otimes \mathbb{T}RGA\big) \rightrightarrows \mathbb{T}\big(\textstyle\int^{A\in\mathcal{A}} FA \otimes RGA\big).$$

□

Locally finitely presentable $\mathcal{V}$-categories

In studying algebraic theories using categories, locally finite presentable categories plays a basic role. When the closed symmetric monoidal categery $\mathcal{V}$ is locally finitely presentable, with finitely presentable unit $\mathbb{1}$, these have a generalization to enriched category theory over $\mathcal{V}$, due to Kelly [15]. (There is a more general theory of locally presentable categories, where the cardinal $\aleph_0$ is replaced by an arbitrary regular cardinal; this extension is straightforward, but we do not present it here in order to simplify exposition).

Examples of locally finitely presentable closed symmetric monoidal categories include the categories of sets, groupoids, categories, simplicial sets, and abelian groups — the finitely presentable objects are respectively finite sets, groupoids and categories, simplicial sets with a finite number of nondegenerate simplices, and finitely presentable abelian groups. A less obvious example is the category of symmetric spectra of Hovey et al. [12].

An object A in a $\mathcal{V}$-category $\mathcal{C}$ is **finitely presentable** if the functor

$$\mathcal{C}(A,-) : \mathcal{C} \longrightarrow \mathcal{V}$$

preserves filtered colimits. A **strong generator** in a $\mathcal{V}$-category $\mathcal{C}$ is a set $\{G_i \in \mathrm{Ob}(\mathcal{C})\}_{i\in I}$ such that the functor

$$A \mapsto \bigotimes_{i\in I} \mathcal{C}(G_i, A) : \mathcal{C} \mapsto \mathcal{V}^{\otimes I}$$

reflects isomorphisms.

Definition 3.8. A **locally finitely presentable** $\mathcal{V}$-category $\mathcal{C}$ is a cocomplete $\mathcal{V}$-category with a finitely presentable strong generator (i.e. the objects making up the strong generator are finitely presentable).

Pseudoalgebras

A 2-monad $\mathbb{T}$ on a 2-category $\mathbb{C}$ is by definition a Cat-monad on $\mathbb{C}$. Denote the composition of the 2-monad $\mathbb{T}$ by $m : \mathbb{T}\mathbb{T} \to \mathbb{T}$, and the unit by $\eta : 1 \to \mathbb{T}$.

Associated to a 2-monad $\mathbb{T}$ is the 2-category $\mathbb{C}^{\mathbb{T}}$ of pseudo $\mathbb{T}$-algebras (Bunge [3] and Street [23]; see also Marmolejo [19]). A pseudo $\mathbb{T}$-algebra is an object $\mathcal{A}$ of $\mathbb{C}$, together with a morphism $a : \mathbb{T}\mathcal{A} \to \mathcal{A}$, the composition, and invertible 2-morphisms

$$\begin{tikzcd} & \mathbb{T}\mathcal{A} \arrow[dr, "m_{\mathcal{A}}"] \arrow[dd, phantom, "\Downarrow\theta"] & \\ \mathbb{T}\mathbb{T}\mathcal{A} \arrow[ur, "\mathbb{T}a"] \arrow[dr, "m_{\mathcal{A}}"'] & & \mathcal{A} \\ & \mathbb{T}\mathcal{A} \arrow[ur, "a"'] & \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} & \mathbb{T}\mathcal{A} \arrow[dr, "a"] \arrow[d, phantom, "\Downarrow\iota"] & \\ \mathcal{A} \arrow[ur, "\eta_{\mathcal{A}}"] \arrow[rr, equal] & {} & \mathcal{A} \end{tikzcd}$$

such that

$$\begin{tikzcd} & \mathbb{T}\mathbb{T}\mathcal{A} \arrow[rr, "\mathbb{T}a"] \arrow[dr, "\mathbb{T}m_{\mathcal{A}}"] & & \mathbb{T}\mathcal{A} \arrow[dr, "a"] & \\ \mathbb{T}\mathbb{T}\mathbb{T}\mathcal{A} \arrow[ur, "\mathbb{T}\mathbb{T}a"] \arrow[dr, "m_{\mathbb{T}\mathcal{A}}"'] & & \mathbb{T}\mathcal{A} \arrow[rr, "a"] \arrow[ur, phantom, "\Downarrow\theta"] \arrow[dr, phantom, "\Downarrow\theta"] & & \mathcal{A} \\ & \mathbb{T}\mathbb{T}\mathcal{A} \arrow[rr, "m_{\mathcal{A}}"'] \arrow[ur, "\mathbb{T}a"] & & \mathbb{T}\mathcal{A} \arrow[ur, "a"'] & \end{tikzcd} \quad \text{equals} \quad \begin{tikzcd} & \mathbb{T}\mathbb{T}\mathcal{A} \arrow[rr, "\mathbb{T}a"] & & \mathbb{T}\mathcal{A} \arrow[dr, "a"] & \\ \mathbb{T}\mathbb{T}\mathbb{T}\mathcal{A} \arrow[ur, "\mathbb{T}\mathbb{T}a"] \arrow[rr, "\mathbb{T}m_{\mathcal{A}}"'] \arrow[dr, "m_{\mathbb{T}\mathcal{A}}"'] & & \mathbb{T}\mathbb{T}\mathcal{A} \arrow[ur, "m_{\mathcal{A}}"] \arrow[dr, "\mathbb{T}a"] \arrow[ul, phantom, "\Downarrow\mathbb{T}\theta"] \arrow[rr, phantom, "\Downarrow\theta"] & & \mathcal{A} \\ & \mathbb{T}\mathbb{T}\mathcal{A} \arrow[rr, "m_{\mathcal{A}}"'] & & \mathbb{T}\mathcal{A} \arrow[ur, "a"'] & \end{tikzcd}$$

and

$$\begin{tikzcd} & \mathbb{T}\mathbb{T}\mathcal{A} \arrow[dr, "\mathbb{T}a"] \arrow[d, phantom, "\Downarrow\mathbb{T}\iota"] & & \\ \mathbb{T}\mathcal{A} \arrow[ur, "\mathbb{T}\eta_{\mathcal{A}}"] \arrow[rr, equal] \arrow[dr, "\mathbb{T}\eta_{\mathcal{A}}"'] & {} & \mathbb{T}\mathcal{A} \arrow[r, "a"] & \mathcal{A} \\ & \mathbb{T}\mathbb{T}\mathcal{A} \arrow[ur, "m_{\mathcal{A}}"'] & & \end{tikzcd}$$

equals

$$\begin{tikzcd} & & \mathbb{T}\mathcal{A} \arrow[dr, "m_{\mathcal{A}}"] \arrow[dd, phantom, "\Downarrow\theta"] & \\ \mathbb{T}\mathcal{A} \arrow[r, "\mathbb{T}\eta_{\mathcal{A}}"] & \mathbb{T}\mathbb{T}\mathcal{A} \arrow[ur, "\mathbb{T}a"] \arrow[dr, "m_{\mathcal{A}}"'] & & \mathcal{A} \\ & & \mathbb{T}\mathcal{A} \arrow[ur, "a"'] & \end{tikzcd}$$

The lax morphisms of $\mathbb{C}^{\mathbb{T}}$ are pairs (f, φ) consisting of a morphism $f : \mathcal{A} \to \mathcal{B}$ and a 2-morphism.

$$\begin{tikzcd}
& \mathbb{T}\mathcal{B} \arrow[dr, "b"] \arrow[dd, phantom, "\Downarrow\varphi"] & \\
\mathbb{T}\mathcal{A} \arrow[ur, "\mathbb{T}f"] \arrow[dr, "a"'] & & \mathcal{B} \\
& \mathcal{A} \arrow[ur, "f"'] &
\end{tikzcd}$$

such that

$$\begin{tikzcd}
& \mathbb{T}\mathbb{T}\mathcal{B} \arrow[rr, "\mathbb{T}b"] \arrow[dr, "m_{\mathcal{B}}"] & & \mathbb{T}\mathcal{B} \arrow[dr, "b"] & \\
\mathbb{T}\mathbb{T}\mathcal{A} \arrow[ur, "\mathbb{T}\mathbb{T}f"] \arrow[dr, "m_{\mathcal{A}}"'] & & \mathbb{T}\mathcal{B} \arrow[rr, "b"] \arrow[ur, phantom, "\Downarrow\theta"] \arrow[dr, phantom, "\Downarrow\varphi"] & & \mathcal{B} \\
& \mathbb{T}\mathcal{A} \arrow[rr, "a"'] \arrow[ur, "\mathbb{T}f"] & & \mathcal{A} \arrow[ur, "f"'] &
\end{tikzcd}
\quad \text{equals} \quad
\begin{tikzcd}
& \mathbb{T}\mathbb{T}\mathcal{B} \arrow[rr, "\mathbb{T}b"] \arrow[d, phantom, "\Downarrow\mathbb{T}\varphi"] & & \mathbb{T}\mathcal{B} \arrow[dr, "b"] \arrow[d, phantom, "\Downarrow\theta"] & \\
\mathbb{T}\mathbb{T}\mathcal{A} \arrow[ur, "\mathbb{T}f"] \arrow[rr, "\mathbb{T}a"] \arrow[dr, "m_{\mathcal{A}}"'] & {} & \mathbb{T}\mathcal{A} \arrow[ur, "\mathbb{T}f"] \arrow[dr, "a"] \arrow[dl, phantom, "\Downarrow\theta"] & {} & \mathcal{B} \\
& \mathbb{T}\mathcal{A} \arrow[rr, "a"'] & & \mathcal{A} \arrow[ur, "f"'] &
\end{tikzcd}$$

and

$$\begin{tikzcd}
& \mathbb{T}\mathcal{A} \arrow[dr, "\mathbb{T}f"] & & \\
\mathcal{A} \arrow[ur, "\eta_{\mathcal{A}}"] \arrow[dr, "f"'] & & \mathbb{T}\mathcal{B} \arrow[dr, "b"] \arrow[d, phantom, "\Downarrow\iota"] & \\
& \mathcal{B} \arrow[ur, "\eta_{\mathcal{B}}"] \arrow[rr, equal] & {} & \mathcal{B}
\end{tikzcd}$$

equals

$$\begin{tikzcd}
& & \mathbb{T}\mathcal{B} \arrow[dr, "b"] \arrow[d, phantom, "\Downarrow\varphi"] & \\
& \mathbb{T}\mathcal{A} \arrow[ur, "\mathbb{T}f"] \arrow[dr, "a"] \arrow[d, phantom, "\Downarrow\iota"] & {} & \mathcal{B} \\
\mathcal{A} \arrow[ur, "\eta_{\mathcal{A}}"] \arrow[rr, equal] & {} & \mathcal{A} \arrow[ur, "f"'] &
\end{tikzcd}$$

A lax morphism (f, φ) is a morphism if the 2-morphism φ is invertible.

The 2-morphisms $\gamma : (f, \varphi) \to (\tilde{f}, \tilde{\varphi})$ of $\mathbb{C}^{\mathbb{T}}$ are 2-morphisms $\gamma : f \to \tilde{f}$ such that

$$\begin{tikzcd}
& \mathbb{T}\mathcal{B} \arrow[dr, "b"] \arrow[dd, phantom, "\Downarrow\varphi"] & \\
\mathbb{T}\mathcal{A} \arrow[ur, "\mathbb{T}f"] \arrow[dr, "a"'] & & \mathcal{B} \\
& \mathcal{A} \arrow[ur, bend left, "f"] \arrow[ur, bend right, "\tilde{f}"'] \arrow[ur, phantom, "\Downarrow\gamma"] &
\end{tikzcd}
\quad \text{equals} \quad
\begin{tikzcd}
& \mathbb{T}\mathcal{B} \arrow[dr, "a"] \arrow[dd, phantom, "\Downarrow\tilde{\varphi}"] & \\
\mathbb{T}\mathcal{A} \arrow[ur, bend left, "\mathbb{T}f"] \arrow[ur, bend right, "\mathbb{T}\tilde{f}"'] \arrow[ur, phantom, "\Downarrow\mathbb{T}\gamma"] \arrow[dr, "a"'] & & \mathcal{B} \\
& \mathcal{A} \arrow[ur, "\tilde{f}"'] &
\end{tikzcd}
\tag{2}$$

$\mathcal{V}$-categories of pseudo algebras

If $\mathcal{A}$ and $\mathcal{B}$ are pseudo $\mathbb{T}$-algebras, and the underlying category of $\mathcal{A}$ is small, we saw that there is a $\mathcal{V}$-category $[\mathcal{A},\mathcal{B}]$ whose objects are $\mathcal{V}$-functors $f:\mathcal{A}\to\mathcal{B}$ between the underlying $\mathcal{V}$-categories. Using $[\mathcal{A},\mathcal{B}]$, we now define a $\mathcal{V}$-category $[\![\mathcal{A},\mathcal{B}]\!]$ whose objects are morphisms $f:\mathcal{A}\to\mathcal{B}$ of pseudo $\mathbb{T}$-algebras. If f_0 and f_1 are morphisms of pseudo $\mathbb{T}$-algebras, $[\![\mathcal{A},\mathcal{B}]\!](f_0,f_1)$ is defined as the equalizer

$$[\![\mathcal{A},\mathcal{B}]\!](f_0,f_1) \xrightarrow{\psi_{f_0f_1}} [\mathcal{A},\mathcal{B}](f_0,f_1) \underset{(-\circ a)\varphi_0}{\overset{\varphi_1(b\circ\mathbb{T}-)}{\rightrightarrows}} [\mathbb{T}\mathcal{A},\mathcal{B}](b\circ\mathbb{T}f_0,f_1\circ a)$$

Of course, this is the internal version of (2): the 2-morphisms $\gamma:f_0\to f_1$ of $\mathbb{C}^{\mathbb{T}}$ are the elements of the set $|[\![\mathcal{A},\mathcal{B}]\!](f_0,f_1)|$.

The composition morphism

$$[\![\mathcal{A},\mathcal{B}]\!](f_0,f_1)\otimes[\![\mathcal{A},\mathcal{B}]\!](f_1,f_2) \xrightarrow{m^{[\![\mathcal{A},\mathcal{B}]\!]}_{f_0f_1f_2}} [\![\mathcal{A},\mathcal{B}]\!](f_0,f_2)$$

is the universal arrow for the coequalizer $[\![\mathcal{A},\mathcal{B}]\!](f_0,f_2)$, whose existence is guaranteed by the commutativity of the diagram

$$[\![\mathcal{A},\mathcal{B}]\!](f_0,f_1)\otimes[\![\mathcal{A},\mathcal{B}]\!](f_1,f_2) \xrightarrow{\psi_{f_0f_1}\otimes\psi_{f_1f_2}} [\mathcal{A},\mathcal{B}](f_0,f_1)\otimes[\mathcal{A},\mathcal{B}](f_1,f_2)$$
$$\xrightarrow{m^{[\mathcal{A},\mathcal{B}]}_{f_0f_1f_1}} [\mathcal{A},\mathcal{B}](f_0,f_2) \underset{(-\circ a)\varphi_0}{\overset{\varphi_2(b\circ\mathbb{T}-)}{\rightrightarrows}} [\mathbb{T}\mathcal{A},\mathcal{B}](b\circ\mathbb{T}f_0,f_2\circ c)$$

Indeed, we have

$$\begin{aligned}
\varphi_2(b\circ\mathbb{T}-)\cdot m^{[\mathcal{A},\mathcal{B}]}_{f_0f_1f_1}\cdot(\psi_{f_0f_1}\otimes\psi_{f_1f_2}) \\
&= m^{[\mathbb{T}\mathcal{A},\mathcal{B}]}_{b\circ\mathbb{T}f_0,b\circ\mathbb{T}f_1,f_2\circ a}\cdot((b\circ\mathbb{T}-)\cdot\psi_{f_0f_1}\otimes\varphi_2(b\circ\mathbb{T}-)\cdot\psi_{f_1f_2}) \\
&= m^{[\mathbb{T}\mathcal{A},\mathcal{B}]}_{b\circ\mathbb{T}f_0,b\circ\mathbb{T}f_1,f_2\circ a}\cdot((b\circ\mathbb{T}-)\cdot\psi_{f_0f_1}\otimes(-\circ a)\varphi_1\cdot\psi_{f_1f_2}) \\
&= m^{[\mathbb{T}\mathcal{A},\mathcal{B}]}_{b\circ\mathbb{T}f_0,f_1\circ a,f_2\circ a}\cdot(\varphi_1(b\circ\mathbb{T}-)\cdot\psi_{f_0f_1}\otimes(-\circ a)\cdot\psi_{f_1f_2}) \\
&= m^{[\mathbb{T}\mathcal{A},\mathcal{B}]}_{b\circ\mathbb{T}f_0,f_1\circ a,f_2\circ a}\cdot((-\circ a)\varphi_0\cdot\psi_{f_0f_1}\otimes(-\circ a)\cdot\psi_{f_1f_2}) \\
&= (-\circ a)\varphi_0\cdot m^{[\mathcal{A},\mathcal{B}]}_{f_0f_1f_1}\cdot(\psi_{f_0f_1}\otimes\psi_{f_1f_2}).
\end{aligned}$$

Associativity of this composition is proved by a calculation along the same lines for the iterated composition map

$$[\![\mathcal{A},\mathcal{B}]\!](f_0,f_1)\otimes[\![\mathcal{A},\mathcal{B}]\!](f_1,f_2)\otimes[\![\mathcal{A},\mathcal{B}]\!](f_2,f_3)\longrightarrow[\![\mathcal{A},\mathcal{B}]\!](f_0,f_3).$$

The unit $v_f:\mathbb{1}\to[\![\mathcal{A},\mathcal{B}]\!](f,f)$ of the $\mathcal{V}$-category $[\![\mathcal{A},\mathcal{B}]\!]$ is the universal arrow for the coequalizer $[\![\mathcal{A},\mathcal{B}]\!](f,f)$, whose existence is guaranteed by the commutativity of the diagram

$$\mathbb{1} \xrightarrow{u_f} [\mathcal{A},\mathcal{B}](f,f) \underset{(-\circ a)\varphi}{\overset{\varphi(b\circ\mathbb{T}-)}{\rightrightarrows}} [\mathbb{T}\mathcal{A},\mathcal{B}](b\circ\mathbb{T}f, f\circ a)$$

Indeed, both $\varphi(b\circ\mathbb{T}-)\cdot u_f$ and $(-\circ a)\varphi\cdot u_f$ are equal to

$$\varphi \in \mathcal{V}(\mathbb{1}, [\mathbb{T}\mathcal{A},\mathcal{B}](b\circ\mathbb{T}f, f\circ a)).$$

In a remark at the end of Section 3 of [13], Im and Kelly discount the possible existence of the $\mathcal{V}$-categories of pseudo algebras. The construction presented here appears to get around their objections.

Acknowledgements

I thank Michael Batanin, André Henriques and Steve Lack for very helpful feedback on earlier drafts of this paper.

I am grateful to a number of institutions for hosting me during the writing of this paper: RIMS, University of Nice, and KITP. I am grateful to my hosts at all of these institutions, Kyoji Saito and Kentaro Hori at RIMS, André Hirschowitz and Carlos Simpson in Nice, and David Gross at KITP, for their hospitality.

I am also grateful to Jean-Louis Loday for the opportunity to lecture on an earlier version of this work at Luminy. I received support from NSF Grants DMS-0072508 and DMS-0505669.

References

1. Clemens Berger and Ieke Moerdijk. Resolution of coloured operads and rectification of homotopy algebras. In: *Categories in algebra, geometry, and mathematical physics*, Contemp. Math. 431, AMS.
2. J. M. Boardman and R. M. Vogt. *Homotopy invariant algebraic structures on topological spaces*. Springer-Verlag, Berlin, 1973.
3. Marta C. Bunge. Coherent extensions and relational algebras. *Trans. Amer. Math. Soc.*, 197:355–390, 1974.
4. Kevin Costello. Topological conformal field theories and Calabi-Yau categories. *Adv. Math.*, 210(1):165–214, 2007.
5. Brian Day. On closed categories of functors. In *Reports of the Midwest Category Seminar, IV*, Lecture Notes in Mathematics, Vol. 137, pages 1–38. Springer, Berlin, 1970.
6. Eduardo J. Dubuc. *Kan extensions in enriched category theory*. Lecture Notes in Mathematics, Vol. 145. Springer-Verlag, Berlin, 1970.
7. Benjamin Enriquez and Pavel Etingof. On the invertibility of quantization functors. *J. Algebra*, 289(2):321–345, 2005.
8. Zbigniew Fiedorowicz. The symmetric bar construction. http://www.math.ohio-state.edu/~fiedorow/symbar.ps.gz.
9. Wee Liang Gan. Koszul duality for dioperads. *Math. Res. Lett.*, 10(1):109–124, 2003.
10. Ezra Getzler. Batalin-Vilkovisky algebras and two-dimensional topological field theories. *Comm. Math. Phys.*, 159(2):265–285, 1994.

11. Ezra Getzler and Mikhail Kapranov. Modular operads. *Compositio Math.*, 110(1):65–126, 1998.
12. Mark Hovey, Brooke Shipley, and Jeff Smith. Symmetric spectra. *J. Amer. Math. Soc.*, 13(1):149–208, 2000.
13. Geun Bin Im and G. M. Kelly. A universal property of the convolution monoidal structure. *J. Pure Appl. Algebra*, 43(1):75–88, 1986.
14. Peter T. Johnstone. *Sketches of an elephant: a topos theory compendium. Vol. 1*, volume 43 of *Oxford Logic Guides*. The Clarendon Press Oxford University Press, New York, 2002.
15. G. M. Kelly. Structures defined by finite limits in the enriched context. I. *Cahiers Topologie Géom. Différentielle*, 23(1):3–42, 1982. Third Colloquium on Categories, Part VI (Amiens, 1980).
16. F. E. J. Linton. Coequalizers in categories of algebras. In *Sem. on Triples and Categorical Homology Theory (ETH, Zürich, 1966/67)*, pages 75–90. Lecture Notes in Math., Vol. 80. Springer, Berlin, 1969.
17. Saunders Mac Lane. Categorical algebra. *Bull. Amer. Math. Soc.*, 71:40–106, 1965.
18. Martin Markl, Steve Shnider, and Jim Stasheff. *Operads in algebra, topology and physics*, volume 96 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.
19. F. Marmolejo. Distributive laws for pseudomonads. *Theory Appl. Categ.*, 5: No. 5, 91–147 (electronic), 1999.
20. Sergei Merkulov. PROP profile of deformation quantization and graph complexes with loops and wheels. Comm. Math. Phys. 262(1):117–135, 2006.
21. Gregory Moore and Nathan Seiberg. Classical and quantum conformal field theory. *Comm. Math. Phys.*, 123(2):177–254, 1989.
22. Kiiti Morita. Duality for modules and its applications to the theory of rings with minimum condition. *Sci. Rep. Tokyo Kyoiku Daigaku Sect. A*, 6:83–142, 1958.
23. Ross Street. Fibrations and Yoneda's lemma in a 2-category. In *Category Seminar (Proc. Sem., Sydney, 1972/1973)*, pages 104–133. Lecture Notes in Math., Vol. 420. Springer, Berlin, 1974.