Poincarés legacies: pages from year two of a mathematical blog

Terence Tao

Department of Mathematics, UCLA, Los Angeles, CA 90095

E-mail address: tao@math.ucla.edu

To Garth Gaudry, who set me on the road;
To my family, for their constant support;
And to the readers of my blog, for their feedback and contributions.

Contents

Preface xi

A remark on notation xiv

Acknowledgments xiv

Chapter 1. Expository articles 1

§1.1. The blue-eyed islanders puzzle 2

§1.2. Kleiner's proof of Gromov's theorem 3

§1.3. Dvir's proof of the finite field Kakeya conjecture 11

§1.4. The van der Corput lemma, and equidistribution on nilmanifolds 16

§1.5. The strong law of large numbers 23

§1.6. The Black-Scholes equation 32

§1.7. Hassell's proof of scarring for the Bunimovich stadium 44

§1.8. Tate's proof of the functional equation 48

§1.9. The divisor bound 59

§1.10. What is a gauge? 64

§1.11. The Lucas-Lehmer test for Mersenne primes 86

§1.12. Finite subsets of groups with no finite models 93

§1.13. Small samples, and the margin of error 99

§1.14. Non-measurable sets via non-standard analysis 110

§1.15. When are eigenvalues stable? 113
§1.16. Concentration compactness and the profile decomposition 120
§1.17. A counterexample to a strong polynomial Freiman-Ruzsa conjecture 131
§1.18. Some notes on "non-classical" polynomials in finite characteristic 135
§1.19. The Kakeya conjecture and the Ham Sandwich theorem 142
§1.20. An airport-inspired puzzle 150
§1.21. Cohomology for dynamical systems 151
§1.22. A remark on the Kakeya needle problem 158

Chapter 2. Ergodic theory 161
§2.1. Overview 162
§2.2. Three categories of dynamical systems 169
§2.3. Minimal dynamical systems, recurrence, and the Stone-Cěch compactification 177
§2.4. Multiple recurrence 190
§2.5. Other topological recurrence results 198
§2.6. Isometric systems and isometric extensions 214
§2.7. Structural theory of topological dynamical systems 233
§2.8. The mean ergodic theorem 242
§2.9. Ergodicity 254
§2.10. The Furstenberg correspondence principle 267
§2.11. Compact systems 278
§2.12. Weakly mixing systems 290
§2.13. Compact extensions 306
§2.14. Weakly mixing extensions 317
§2.15. The Furstenberg-Zimmer structure theorem and the Furstenberg recurrence theorem 326
§2.16. A Ratner-type theorem for nilmanifolds 331

§2.17. A Ratner-type theorem for $SL_2(R)$ orbits 344

Chapter 3. The Poincaré conjecture 357
§3.1. Riemannian manifolds and curvature 358
§3.2. Flows on Riemannian manifolds 375
§3.3. The Ricci flow approach to the Poincaré conjecture 390
§3.4. The maximum principle, and the pinching phenomenon 402
§3.5. Finite time extinction of the second homotopy group 418
§3.6. Finite time extinction of the third homotopy group, I 428
§3.7. Finite time extinction of the third homotopy group, II 434
§3.8. Rescaling of Ricci flows and κ-noncollapsing 447
§3.9. Ricci flow as a gradient flow, log-Sobolev inequalities, and Perelman entropy 463
§3.10. Comparison geometry, the high-dimensional limit, and Perelman reduced volume 481
§3.11. κ-noncollapsing via Perelman reduced volume 510
§3.12. High curvature regions of Ricci flow and κ-solutions 524
§3.13. Li-Yau-Hamilton Harnack inequalities and κ-solutions 536
§3.14. Stationary points of Perelman entropy or reduced volume are gradient shrinking solitons 548
§3.15. Geometric limits of Ricci flows, and asymptotic gradient shrinking solitons 557
§3.16. Classification of asymptotic gradient shrinking solitons 572
§3.17. The structure of κ-solutions 584
§3.18. The structure of high-curvature regions of Ricci flow 596
§3.19. The structure of Ricci flow at the singular time, surgery, and the Poincaré conjecture 606

Chapter 4. Lectures in additive prime number theory 619
§4.1. Structure and randomness in the prime numbers 620
§4.2. Linear equations in primes 631
§4.3. Small gaps between primes 644
§4.4. Sieving for almost primes and expanders 653

Bibliography 665

Preface

In February of 2007, I converted my "What's new" web page of research updates into a blog at `terrytao.wordpress.com`. This blog has since grown and evolved to cover a wide variety of mathematical topics, ranging from my own research updates, to lectures and guest posts by other mathematicians, to open problems, to class lecture notes, to expository articles at both basic and advanced levels.

With the encouragement of my blog readers, and also of the AMS, I published many of the mathematical articles from the first year (2007) of the blog as [**Ta2008b**], which will henceforth be referred to as *Structure and Randomness* throughout this book. This gave me the opportunity to improve and update these articles to a publishable (and citeable) standard, and also to record some of the substantive feedback I had received on these articles by the readers of the blog. Given the success of the blog experiment so far, I am now doing the same for the second year (2008) of articles from the blog, which has become the book you are now reading.

As with *Structure and Randomness*, the book begins with a collection of expository articles, ranging in level from completely elementary logic puzzles to remarks on recent research, which are only loosely related to each other and to the rest of the book. However, in contrast to the previous book, the bulk of this manuscript is dominated by the lecture notes for two graduate courses I gave during the year.

The two courses stemmed from two very different but fundamental contributions to mathematics by Henri Poincaré, which explains the title of the book.

The first course (Chapter 2) was on the topics of *topological dynamics and ergodic theory*, which originated in part from Poincaré's pioneering work in chaotic dynamical systems. Many situations in mathematics, physics, or other sciences can be modeled by a discrete or continuous *dynamical system*, which at its most abstract level is simply a space X, together with a shift $T : X \to X$ (or family of shifts) acting on that space, and possibly preserving either the topological or measure-theoretic structure of that space. At this level of generality, there are a countless variety of dynamical systems available for study, and it may seem hopeless to say much of interest without specialising to much more concrete systems. Nevertheless, there is a remarkable phenomenon that dynamical systems can largely be classified into "structured" (or "periodic") components, and "random" (or "mixing") components[1], which then can be used to prove various *recurrence theorems* that apply to very large classes of dynamical systems, not the least of which is the *Furstenberg multiple recurrence theorem* (Theorem 2.10.3). By means of various *correspondence principles*, these recurrence theorems can then be used to prove some deep theorems in combinatorics and other areas of mathematics, in particular yielding one of the shortest known proofs of *Szemerédi's theorem* (Theorem 2.10.1) that all sets of integers of positive upper density contain arbitrarily long arithmetic progressions. The road to these recurrence theorems, and several related topics (e.g. ergodicity, and Ratner's theorem on the equidistribution of unipotent orbits in homogeneous spaces) will occupy the bulk of this course. I was able to cover all but the last two sections in a 10-week course at UCLA, using the exercises provided within the notes to assess the students (who were generally second or third-year graduate students, having already taken a course or two in graduate real analysis).

[1] One also has to consider *extensions* of systems of one type by another, e.g. mixing extensions of periodic systems; see Section 2.15 for a precise statement.

The second course (Chapter 3) focused on a completely different problem posed by Poincaré, namely the famous *Poincaré conjecture* that every simply connected compact three-dimensional manifold is homeomorphic to a sphere, and its recent spectacular solution [**Pe2002**], [**Pe2003**], [**Pe2003b**] by Perelman. This conjecture is purely topological in nature, and yet Perelman's proof uses remarkably little topology, instead working almost entirely in the realm of Riemannian geometry and partial differential equations, and specifically in a detailed analysis of solutions to Ricci flow on three-dimensional manifolds, and the singularities formed by such flows. As such, the course will incorporate, along the way, a review of many of the basic concepts and results from Riemannian geometry (and to a lesser extent, from parabolic PDE), while being focused primarily on the single objective of proving the Poincaré conjecture. Due to the complexity and technical intricacy of the argument, we will not be providing a fully complete proof of this conjecture here (but see [**MoTi2007**] for a careful and detailed treatment); but we will be able to cover the high-level features of the argument, as well as many of the specific components of that argument, in full detail, and the remaining components are sketched and motivated, with references to more complete arguments given. In principle, the course material is sufficiently self-contained that prior exposure to Riemannian geometry, PDE, or topology at the graduate level is not strictly necessary, but in practice, one would probably need some comfort with at least one of these three areas in order to not be totally overwhelmed by the material. (I ran this course as a topics course; in particular, I did not assign homework.)

Finally, I close the book with a third (and largely unrelated) topic (Chapter 4), namely a series of lectures on recent developments in additive prime number theory, both by myself and my coauthors, and by others. These lectures are derived from a lecture I gave at the annual meeting of the AMS at San Diego in January of 2007, as well as a lecture series I gave at Penn State University in November 2007.

A remark on notation

For reasons of space, we will not be able to define every single mathematical term that we use in this book. If a term is italicised for reasons other than emphasis or for definition, then it denotes a standard mathematical object, result, or concept, which can be easily looked up in any number of references. (In the blog version of the book, many of these terms were linked to their Wikipedia pages, or other on-line reference pages.)

I will however mention a few notational conventions that I will use throughout. The cardinality of a finite set E will be denoted $|E|$. We will use the asymptotic notation $X = O(Y)$, $X \ll Y$, or $Y \gg X$ to denote the estimate $|X| \leq CY$ for some absolute constant $C > 0$. In some cases we will need this constant C to depend on a parameter (e.g. d), in which case we shall indicate this dependence by subscripts, e.g. $X = O_d(Y)$ or $X \ll_d Y$. We also sometimes use $X \sim Y$ as a synonym for $X \ll Y \ll X$.

In many situations there will be a large parameter n that goes off to infinity. When that occurs, we also use the notation $o_{n\to\infty}(X)$ or simply $o(X)$ to denote any quantity bounded in magnitude by $c(n)X$, where $c(n)$ is a function depending only on n that goes to zero as n goes to infinity. If we need $c(n)$ to depend on another parameter, e.g. d, we indicate this by further subscripts, e.g. $o_{n\to\infty;d}(X)$.

We will occasionally use the averaging notation $\mathbf{E}_{x\in X} f(x) := \frac{1}{|X|}\sum_{x\in X} f(x)$ to denote the average value of a function $f : X \to \mathbf{C}$ on a non-empty finite set X.

Acknowledgments

The author is supported by a grant from the MacArthur Foundation, by NSF grant DMS-0649473, and by the NSF Waterman award.

Chapter 1

Expository articles

1.1. The blue-eyed islanders puzzle

This is one of my favourite logic puzzles. It has a number of formulations, but I will use this one:

Problem 1.1.1. There is an island upon which a tribe resides. The tribe consists of 1000 people, with various eye colours. Yet, their religion forbids them to know their own eye color, or even to discuss the topic; thus, each resident can (and does) see the eye colors of all other residents, but has no way of discovering his or her own (there are no reflective surfaces). If a tribesperson does discover his or her own eye color, then their religion compels them to commit ritual suicide at noon the following day in the village square for all to witness. All the tribespeople are highly logical[1] and devout, and they all know that each other is also highly logical and devout (and they all know that they all know that each other is highly logical and devout, and so forth).

Of the 1000 islanders, it turns out that 100 of them have blue eyes and 900 of them have brown eyes, although the islanders are not initially aware of these statistics (each of them can of course only see 999 of the 1000 tribespeople).

One day, a blue-eyed foreigner visits to the island and wins the complete trust of the tribe.

One evening, he addresses the entire tribe to thank them for their hospitality.

However, not knowing the customs, the foreigner makes the mistake of mentioning eye color in his address, remarking how unusual it is to see another blue-eyed person like myself in this region of the world.

What effect, if anything, does this *faux pas* have on the tribe?

The interesting thing about this puzzle is that there are two quite plausible arguments here, which give opposing conclusions:

[1]For the purposes of this logic puzzle, "highly logical" means that any conclusion that can logically deduced from the information and observations available to an islander, will automatically be known to that islander.

Argument 1. The foreigner has no effect, because his comments do not tell the tribe anything that they do not already know (everyone in the tribe can already see that there are several blue-eyed people in their tribe). □

Argument 2. 100 days after the address, all the blue eyed people commit suicide. This is proven as a special case of Proposition 1.1.2 below. □

Proposition 1.1.2. *Suppose that the tribe had n blue-eyed people for some positive integer n. Then n days after the travellers address, all n blue-eyed people commit suicide.*

Proof. We induct on n. When $n = 1$, the single blue-eyed person realizes that the traveler is referring to him or her, and thus commits suicide on the next day. Now suppose inductively that n is larger than 1. Each blue-eyed person will reason as follows: "If I am not blue-eyed, then there will only be $n-1$ blue-eyed people on this island, and so they will all commit suicide $n-1$ days after the travelers address. But when $n-1$ days pass, none of the blue-eyed people do so (because at that stage they have no evidence that they themselves are blue-eyed). After nobody commits suicide on the $(n-1)^{\text{st}}$ day, each of the blue eyed people then realizes that they themselves must have blue eyes, and will then commit suicide on the n^{th} day. □

Which argument is logically valid? Or are the hypotheses of the puzzle logically impossible to satisfy[2]?

1.1.1. Notes. I won't spoil the solution to this puzzle in this article; but one can find much discussion on this problem at the comments to the web page for this puzzle, at `terrytao.wordpress.com/2008/02/05`. See also `xkcd.com/blue_eyes.html` for some further discussion.

1.2. Kleiner's proof of Gromov's theorem

Inn this article, I would like to present the recent simplified proof by Kleiner[**Kl2007**] of the celebrated theorem of Gromov[**Gr1981**] on groups of polynomial growth.

[2]Note that this is not the same as the hypotheses being extremely implausible, which of course they are.

Let G be an at most countable group generated by a finite set S of generators, which we can take to be symmetric (i.e. $s^{-1} \in S$ whenever $s \in S$). Then we can form the *Cayley graph* Γ, whose vertices are the elements of G, and with g and gs connected by an edge for every $g \in G$ and $s \in S$. This is a connected regular graph, with a transitive left-action of G. For any vertex x and $R > 0$, one can define the ball B(x,R) in Γ to be the set of all vertices connected to x by a path of length at most R. We say that G has *polynomial growth* if we have the bound $|B(x,R)| = O(R^{O(1)})$ as $R \to \infty$; one can easily show that the left-hand side is independent of x, and that the polynomial growth property does not depend on the choice of generating set S.

Examples of finitely generated groups of polynomial growth include

(1) Finite groups;

(2) Abelian groups (e.g. $\mathbf{Z}^d$);

(3) Nilpotent groups (a generalisation of 2.);

(4) *Virtually nilpotent* groups, i.e. it has a nilpotent subgroup of finite index (a combination of 1. and 3.).

In [**Gr1981**], Gromov proved that these are the only examples:

Theorem 1.2.1 (Gromov's theorem)**.** [**Gr1981**] *Let G be a finitely generated group of polynomial growth. Then G is virtually nilpotent.*

Gromov's original argument used a number of deep tools, including the Montgomery-Zippin-Yamabe[**MoZi1955**] structure theory of locally compact groups (related to *Hilbert's fifth problem*), as well as various earlier partial results on groups of polynomial growth. Several proofs have subsequently been found. Recently, Kleiner[**Kl2007**] obtained a proof which was significantly more elementary, although it still relies on some non-trivial partial versions of Gromov's theorem. Specifically, it needs the following result proven by Wolf[**Wo1968**] and by Milnor[**Mi1968**]:

Theorem 1.2.2 (Gromov's theorem for virtually solvable groups)**.** [**Wo1968**],[**Mi1968**] *Let G be a finitely generated group of polynomial*

growth which is virtually solvable *(i.e. it has a solvable subgroup of finite index). Then it is virtually nilpotent.*

The argument also needs a related result:

Theorem 1.2.3. *Let G be a finitely generated amenable*[3] *group which is linear, thus $G \subset GL_n(\mathbf{C})$ for some n. Then G is virtually solvable.*

This theorem is an immediate consequence of the Tits alternative[**Ti1972**], but also has a short elementary proof, due to Shalom[**Sh1998**]. An easy application of the pigeonhole principle to the sequence $|B(x,R)|$ for $R = 1, 2, \ldots$ shows that every group of polynomial growth is amenable. Thus Theorem 1.2.2 and Theorem 1.2.3 already give Gromov's theorem for linear groups.

Other than Theorem 1.2.2 and Theorem 1.2.3, Kleiner's proof of Theorem 1.2.1 is essentially self contained. The argument also extends to groups of *weakly polynomial growth*, which means that $|B(x,R)| = O(R^{O(1)})$ for some sequence of radii R going to infinity. (This extension of Gromov's theorem was first established in [**vdDrWi1984**]. But for simplicity we only discuss the polynomial growth case here.

1.2.1. Reductions. The first few reductions follow the lines of Gromov's original argument. The first observation is that it suffices to exhibit an infinite abelianisation of G, or more specifically to prove:

Proposition 1.2.4 (Existence of infinite abelian representation)**.** *Let G be an infinite finitely generated group of polynomial growth. Then there exists a subgroup G' of finite index whose abelianisation $G'/[G',G']$ is infinite.*

Indeed, if G' has infinite abelianisation, then one can find a non-trivial homomorphism $\alpha : G' \to \mathbf{Z}$. The kernel K of this homomorphism is a normal subgroup of G'. Using the polynomial growth hypothesis, one can show that K is also finitely generated; furthermore, it is polynomial growth of one lower order (i.e. the exponent

[3] In this context, one definition of amenability is that G contains an *Følner sequence* $F_1, F_2, \ldots$ of finite sets, thus $\bigcup_{n=1}^{\infty} F_n = G$ and $\lim_{n\to\infty} |gF_n \Delta F_n|/|F_n| = 0$ for all $g \in G$.

in the $O(R^{O(1)})$ bound for $|B(x,R)|$ is reduced by 1). An induction hypothesis then gives that K is virtually nilpotent, which easily implies that G' (and thus G) is virtually solvable. Gromov's theorem for infinite G then follows from Theorem 1.2.2. (The theorem is of course trivial for finite G.)

Remark 1.2.5. The above argument not only shows that G is virtually solvable, but moreover that G' is the semidirect product $K \rtimes_\phi \mathbf{Z}$ of a virtually nilpotent group K and the integers, which acts on K by some automorphism ϕ. Thus one does not actually need the full strength of Theorem 1.2.2 here, but only the special case of semidirect products of the above form. In any case, most proofs of Theorem 1.2.2 proceed by reducing to this sort of case anyway.

To show Proposition 1.2.4, it suffices to show

Proposition 1.2.6 (Existence of infinite linear representation)**.** *Let G be an infinite finitely generated group of polynomial growth. Then there exists a finite-dimensional representation $\rho : G \to GL_n(\mathbf{C})$ whose image $\rho(G)$ is infinite.*

Indeed, the image $\rho(G) \subset GL_n(\mathbf{C})$ is also finitely generated with polynomial growth, and hence by Theorem 1.2.3 and Theorem 1.2.2 is virtually nilpotent (actually, for this argument we don't need Theorem 1.2.2 and would be content with virtual solvability). If the abelianisation of $\rho(G)$ is finite, one can easily pass to a subgroup G' of finite index and reduce the (virtual) step of $\rho(G')$ by 1, so one can quickly reduce to the case when the abelianisation is infinite, at which point Proposition 1.2.4 follows. So all we need to do now is to prove Proposition 1.2.6.

1.2.2. Harmonic functions on Cayley graphs. Kleiner's approach to Proposition 1.2.6 relies on the notion of a (possibly vector-valued) *harmonic function* on the Cayley graph Γ. This is a function $f : G \to H$ taking values in a Hilbert space H such that $f(g) = \frac{1}{|S|}\sum_{s\in S} f(gs)$ for all $g \in G$. Formally, harmonic functions are local minimisers of the energy functional

$$E(f) := \frac{1}{2}\sum_{g\in G} |\nabla f(g)|^2$$

where

$$|\nabla f(g)|^2 := \frac{1}{|S|}\sum_{s\in S} \|f(gs) - f(g)\|_H^2$$

though of course with the caveat that $E(f)$ is often infinite. (This property is also equivalent to a certain graph Laplacian of f vanishing.)

Of course, every constant function is harmonic. But there are other harmonic functions too: for instance, on $\mathbf{Z}^d$, any linear function is harmonic (regardless of the actual choice of generators). Kleiner's proof of Proposition 1.2.6 follows by combining the following two results:

Proposition 1.2.7. *Let G be an infinite finitely generated group of polynomial growth. Then there exists an (affine-) isometric (left-)action of G on a Hilbert space H with no fixed points, and a harmonic map $f : G \to H$ which is G-equivariant (thus $f(gh) = gf(h)$ for all $g, h \in G$). (Note that by equivariance and the absence of fixed points, this harmonic map is necessarily non-constant.)*

Proposition 1.2.8. *Let G be an finitely generated group of polynomial growth, and let $d \geq 0$. Then the linear space of harmonic functions $u : G \to \mathbf{R}$ which grow of order at most d (thus $u(g) = O(R^d)$ on $B(\mathrm{id}, R)$) is finite-dimensional.*

Indeed, if f is the vector-valued map given by Proposition 1.2.7, then from the G-equivariance it is easy to see that f is of polynomial growth (indeed it is Lipschitz). But the linear projections $\{f \cdot v : v \in H\}$ of f to scalar-valued harmonic maps lie in a finite-dimensional space, by Proposition 1.2.8. This implies that the range $f(G)$ of f lies in a finite-dimensional space V. On the other hand, the obvious action of G on V has no fixed points (being a restriction of the action of G on H), and so the image of G in $GL_n(V)$ must be infinite, and Proposition 1.2.6 follows.

It remains to prove Proposition 1.2.7 and Proposition 1.2.8. Proposition 1.2.7 follows by some more general results of Korevaar-Schoen[**KoSc1997**] and Mok[**Mo1995**], though Kleiner provided an elementary proof which we sketch below. Proposition 1.2.8 was initially proven by Colding and Minicozzi[**CoMi1997**] (for finitely presented groups, at

least) using Gromov's theorem; Kleiner's key new observation was that Proposition 1.2.8 can be proven directly by an elementary argument based on a Poincaré inequalities.

1.2.3. A non-constant equivariant harmonic function. We now sketch the proof of Proposition 1.2.6. The first step is to just get the action on a Hilbert space with no fixed points:

Lemma 1.2.9. *Let G be a countably infinite amenable group. Then there exists an action of G on a Hilbert space H with no fixed points.*

This is essentially the well-known assertion that countably infinite amenable groups do not obey *Property (T)*, but we can give an explicit proof as follows. Using amenability, one can construct a nested *Følner sequence* $F_1 \subset F_2 \subset \ldots \subset \bigcup_n F_n = G$ of finite sets with the property that $|(F_{n-1} \cdot F_n) \Delta F_n| \leq 2^{-n}|F_n|$ (say). (In the case of groups of polynomial growth, one can take $F_n = B(\mathrm{id}, R_n)$ for some rapidly growing, randomly chosen sequence of radii R_n.) We then look at $H := l^2(\mathbf{N}; l^2(G))$, the Hilbert space of sequences $f_1, f_2, \ldots \in l^2(G)$ with $\sum_n \|f_n\|^2_{l^2(G)} < \infty$. This space has the obvious unitary action of G, defined as $g : (f_n(\cdot))_{n\in\mathbf{N}} \to (f_n(g\cdot))_{n\in\mathbf{N}}$. This action has a fixed point of 0, but we can delete this fixed point by considering instead the affine-isometric action $f \mapsto gf + gh - h$, where h is the sequence $h = (\frac{1}{|F_n|^{1/2}} 1_{F_n})_{n\in\mathbf{N}}$. This sequence h does not directly lie in H, but observe that $gh - h$ lies in H for every g. One can then easily show that this action obeys the conclusions of Lemma 1.2.9.

Another way of asserting that an action of G on H has no fixed point is to say that the energy functional $E : H \to \mathbf{R}^+$ defined by $E(v) := \frac{1}{2}\sum_{s\in S} \|sv - v\|^2_H$ is always strictly positive. So Lemma 1.2.9 concludes that there exists an action of G on a Hilbert space on which E is strictly positive. It is possible to then conclude that there exists another action of G on another Hilbert space on which the energy E is not only strictly positive, but actually attains its minimum at some vector v. This observation follows from more general results of Fisher and Margulis[**FiMa2005**], but one can also argue directly as follows. For every $0 < \lambda < 1$ and $A > 0$, there must exist a vector v which almost minimises E in the sense that $E(v') \geq \lambda E(v)$ whenever $\|v - v'\| \leq AE(v)^{1/2}$, since otherwise one could iterate and

sum a Neumann-type series to obtain a fixed point of v. But then by shifting v to the origin, and taking an ultrafilter limit (!) as $\lambda \to 1$ and $A \to \infty$, we obtain the claim.

Some elementary calculus of variations then shows that if v is the energy minimiser, the map $f : g \mapsto gv$ is a harmonic G-equivariant function from G to H, and Proposition 1.2.6 follows.

1.2.4. Poincaré's inequality, and the complexity of harmonic functions. Now we turn to the proof of Proposition 1.2.8, which is the main new ingredient in Kleiner's argument. To simplify the exposition, let us cheat[4] and suppose that the polynomial growth condition $|B(x,R)| = O(R^{O(1)})$ is replaced by the slightly stronger doubling condition $|B(x,2R)| = O(|B(x,R)|)$. Similarly, to simplify the argument, let us pretend that the harmonic functions $u : G \to \mathbf{R}$ are not only of polynomial growth, but also obey a doubling condition $\sum_{x \in B(\mathrm{id},2R)} u(x)^2 \ll \sum_{x \in B(\mathrm{id},R)} u(x)^2$.

The key point is to exploit the fact that harmonic functions are fairly smooth. For instance, a simple "integration by parts" argument shows that if u is harmonic, then

$$\sum_{x \in B(\mathrm{id},R)} |\nabla u(x)|^2 \, dx \ll R^{-2} \sum_{x \in B(\mathrm{id},2R)} |u(x)|^2 \, dx; \tag{1.1}$$

this estimate can be established by the usual trick of replacing the summation over $B(\mathrm{id}, R)$ with a smoother cutoff function and then expanding out the gradient square.

To use this gradient control, Kleiner established the Poincaré inequality

$$\sum_{x \in B(x_0,R)} \sum_{y \in B(x_0,R)} |u(x) - u(y)|^2 \ll R^2 |B(x_0,R)| \sum_{x \in B(x_0,3R)} |\nabla u(x)|^2 \tag{1.2}$$

assuming the doubling condition on balls. This inequality (a variant of a similar inequality of Coulhon and Saloff-Coste[**CoSC1993**]) is actually quite easy to prove. Observe that if $x, y \in B(x_0, R)$, then

[4] In practice, one can use pigeonholing arguments to show that polynomial growth implies doubling on large ranges of scales, which turns out to suffice.

$x = yg$ for some $g \in B(0, 2R)$. Thus it suffices to show that

$$\sum_{x \in B(x_0,R)} |u(x) - u(xg)|^2 \ll R^2 \sum_{x \in B(x_0,3R)} |\nabla u(x)|^2$$

for each such g. But by expanding g as a product of at most $2R$ generators, splitting $u(x) - u(xg)$ as a telescoping series, and using Cauchy-Schwarz, the result follows easily.

Combining (1.1) and (1.2), one sees that a harmonic function which is controlled on a large ball $B(0, R)$, becomes nearly constant on small balls $B(x, \varepsilon R)$ (morally speaking, we have $|u(x) - u(y)| \ll \varepsilon |u(x)|$ "on the average" on such small balls). In particular, given any $0 < \varepsilon < 1$, one can now obtain an inequality of the form

$$\sum_{x \in B(\mathrm{id},R)} |u(x)|^2 \ll \frac{1}{|B(\mathrm{id}, \varepsilon R)|} \sum_j |\sum_{x \in B_j} u(x)|^2 + \varepsilon^2 \sum_{x \in B(\mathrm{id},16R)} |u(x)|^2$$

where B_j ranges over a cover of $B(\mathrm{id}, R)$ by balls B_j of radius εR (the number of such balls can be chosen to be polynomially bounded in $1/\varepsilon$, by the doubling condition). As a consequence, we see that if a harmonic function u obeys a doubling condition, and has zero average on each ball B_j, then it vanishes identically. Morally speaking, this shows that the space of functions that obeys the doubling condition has finite dimension (bounded by the number of such balls), yielding Proposition 1.2.8 in the doubling case. It requires a small amount of combinatorial trickery to obtain this conclusion in the case when u and the balls $B(\mathrm{id}, R)$ exhibit polynomial growth rather than doubling, but the general idea is still the same.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/02/14`, and was given as a talk in an IPAM workshop on expanders in pure and applied mathematics in February of 2008.

Yehuda Shalom pointed out that one does not need the full strength of Proposition 1.2.7 (constructing the vector-valued equivariant harmonic map) in order to deduce Proposition 1.2.6 from Proposition 1.2.8. Instead, all one needs is a single non-constant scalar harmonic map f of polynomial growth. Indeed, observe that G acts by left rotation on the space V of harmonic maps of a fixed polynomial growth, which is finite dimensional by Proposition 1.2.8. If the image of f is

infinite then Proposition 1.2.6 is immediate, so suppose the image of f is finite. Then there is a normal subgroup N of G of finite index which stabilises f (and everyone else in the image of f also). Because of this, the harmonic map f on G pushes down to a harmonic map on the quotient space $N\backslash G$ (which still has a right-action of the generator set S). But it is easy to see that a harmonic map on a finite connected graph is constant, and so f is constant, a contradiction.

David Fisher pointed out that a simplified treatment of the Montgomery-Zippin-Yamabe theory of locally compact groups using nonstandard analysis is given in [**Hi**].

1.3. Dvir's proof of the finite field Kakeya conjecture

One of my favourite unsolved problems in mathematics is the *Kakeya conjecture* in geometric measure theory. This conjecture is descended from the following question, posed by Soichi Kakeya in 1917:

Problem 1.3.1 (Kakeya needle problem)**.** What is the least area in the plane required to continuously rotate a needle of unit length and zero thickness around completely (i.e. by 360°)?

For instance, one can rotate a unit needle inside a unit disk, which has area $\pi/4$. By using a *deltoid* one requires only $\pi/8$ area.

In [**Be1919**], [**Be1928**], Besicovitch showed that that in fact one could rotate a unit needle using an *arbitrarily small* amount of positive area. This unintuitive fact was a corollary of two observations. The first, which is easy, is that one can *translate* a needle using arbitrarily small area, by sliding the needle along the direction it points in for a long distance (which costs zero area), turning it slightly (costing a small amount of area), sliding back, and then undoing the turn. The second fact, which is less obvious, can be phrased as follows. Define a *Kakeya set* in $\mathbf{R}^2$ to be any set which contains a unit line segment in each direction.

Theorem 1.3.2. [**Be1919**] *There exists Kakeya sets* $\mathbf{R}^2$ *of arbitrarily small area (or more precisely, Lebesgue measure).*

In fact, one can construct such sets with zero Lebesgue measure. On the other hand, it was shown by Davies[**Da1971**] that even though these sets had zero area, they were still necessarily two-dimensional (in the sense of either Hausdorff dimension or Minkowski dimension). This led to an analogous conjecture in higher dimensions:

Conjecture 1.3.3 (Kakeya conjecture)**.** *A Besicovitch set in* $\mathbf{R}^n$ *(i.e. a subset of* $\mathbf{R}^n$ *that contains a unit line segment in every direction) has Minkowski and Hausdorff dimension equal to* n*.*

This conjecture remains open in dimensions three and higher (and gets more difficult as the dimension increases), although many partial results are known. For instance, when $n = 3$, it is known that Besicovitch sets have *Hausdorff dimension* at least $5/2$ (see [**Wo1995**]) and *upper Minkowski dimension* at least $5/2+10^{-10}$ (see [**KaLaTa2000**]). See also the surveys [**Ta2001**], [**KaTa2002**], [**Wo1999**].

In [**Wo1999**], Wolff proposed a simpler finite field analogue[5] of the Kakeya conjecture as a model problem that avoided all the technical issues involving Minkowski and Hausdorff dimension. If F^n is a vector space over a finite field F, define a *Kakeya set* to be a subset of F^n which contains a line in every direction.

Conjecture 1.3.4 (Finite field Kakeya conjecture)**.** *Let* $E \subset F^n$ *be a Kakeya set. Then* E *has cardinality at least* $c_n|F|^n$*, where* $c_n > 0$ *depends only on* n*.*

This conjecture has had a significant influence in the subject, in particular inspiring work on the *sum-product phenomenon* in finite fields, which has since proven to have many applications in number theory and computer science. Modulo minor technicalities, the progress on the finite field Kakeya conjecture was, until very recently, essentially the same as that of the original "Euclidean" Kakeya conjecture.

Recently, the finite field Kakeya conjecture was proven using a beautifully simple argument by Dvir[**Dv2008**], using the *polynomial method* in algebraic extremal combinatorics. The proof is so short that I can present it in full here.

[5] Cf. Section 1.6 of *Structure and Randomness*.

The polynomial method is used to control the size of various sets E by looking at one or more polynomials P which vanish on that set E. This philosophy of course closely resembles that of algebraic geometry, and indeed one could classify the polynomial method as a kind of "combinatorial algebraic geometry". An important difference, though, is that in the combinatorial setting we work over fields that are definitely *not* algebraically closed; in particular, we are primarily interested in polynomials[6] and their zero sets over *finite* fields.

For instance, in high school we learn the following connection between one-dimensional sets E, and polynomials $P(x)$ of one variable:

Theorem 1.3.5 (Factor theorem)**.** *Let F be a field, and $d \geq 1$ be an integer. Let $F[x]$ denote the polynomials in one variable with coefficients in F.*

(1) *If $P \in F[x]$ is a non-zero polynomial of degree at most d, then the set $\{x \in F : P(x) = 0\}$ has cardinality at most d.*

(2) *Conversely, given any set $E \subset F$ of cardinality at most d, there exists a non-zero polynomial $P \in F[x]$ of degree at most d that vanishes on E.*

Thus, to obtain an *upper bound* on the size of a one-dimensional set E, it would suffice to exhibit a non-zero low-degree polynomial that vanishes on E; conversely, to *lower bound* the size of E, one would have to show that the only low-degree polynomial that vanishes on E is the zero polynomial. It is the latter type of observation which is of relevance to the finite field Kakeya problem.

There are analogues of both 1. and 2. above in higher dimensions. For instance, the *Schwartz-Zippel lemma*[**Sc1980**] is a higher-dimensional analogue of 1., as is the combinatorial nullstellensatz of Alon[**Al1999**] and *Bezout's theorem* from algebraic geometry, while Stepanov's method[**St1969**] exploits a higher-dimensional analogue of 2. These sorts of techniques and results are collectively referred to as the *polynomial method* in extremal algebraic combinatorics. For

[6]Also, whereas algebraic geometry is more concerned with *specific* (and often highly structured) polynomials, the polynomial method requires that one consider rather *generic* (and usually quite high degree) polynomials.

Dvir's argument, we will need a very simple higher-dimensional version of 2. that comes from basic linear algebra, namely

Lemma 1.3.6. *Let $E \subset F^n$ be a set of cardinality less than $\binom{n+d}{n}$ for some $d \geq 0$. Then there exists a non-zero polynomial $P \in F[x_1, \ldots, x_n]$ on n variables of degree at most d which vanishes on E.*

Proof. Let V be the vector space of polynomials in $F[x_1, \ldots, x_n]$ of degree at most d. Elementary combinatorics reveals that V has dimension $\binom{n+d}{n}$. On the other hand, the vector space F^E of F-valued functions on E has dimension $|E| < \binom{n+d}{d}$. Hence the evaluation map $P \mapsto (P(x))_{x \in E}$ from V to F^E is non-injective, and the claim follows. □

Dvir's argument combines this lemma with

Proposition 1.3.7. *Let $P \in F[x_1, \ldots, x_n]$ be a polynomial of degree at most $|F| - 1$ which vanishes on a Kakeya set E. Then P is identically zero.*

Proof. Suppose for contradiction that P is non-zero. We can write $P = \sum_{i=0}^{d} P_i$, where $0 \leq d \leq |F| - 1$ is the degree of P and P_i is the i^{th} homogeneous component, thus P_d is non-zero. Since P vanishes on E, d cannot be zero.

Let $v \in F^n \backslash \{0\}$ be an arbitrary direction. As E is a Kakeya set, E contains a line $\{x + tv : t \in F\}$ for some $x = x_v \in F^n$, thus $P(x + tv) = 0$ for all $t \in F$. The left-hand side is a polynomial in t of degree at most $|F| - 1$, and thus vanishes identically by the factor theorem. In particular, the t^d coefficient of this polynomial, which is $P_d(v)$, vanishes for any non-zero v. Since P_d is homogeneous of degree $d > 0$, P_d vanishes on all of F^n. Since P_d also has degree less than $|F|$, repeated application of the factor theorem for each variable in turn (or the Schwartz-Zippel lemma[**Sc1980**], which is much the same thing) shows that $P_d = 0$, a contradiction. □

Remark 1.3.8. The point here is that a low-degree polynomial which vanishes on a line must also vanish on the point at infinity where the

line touches the hyperplane at infinity. Thus a polynomial which vanishes on a Kakeya set vanishes at the entire hyperplane at infinity. One can then divide out the defining polynomial for that hyperplane and repeat the process to conclude that the polynomial vanishes identically.

Combining the lemma and the proposition we obtain

Corollary 1.3.9. *Every Kakeya set in F^n has cardinality at least $\binom{|F|+n-1}{n}$.*

Since $\binom{|F|+n-1}{n} = \frac{1}{n!}|F|^n + O_n(|F|^{n-1})$, this establishes the finite field Kakeya conjecture.

This bound seems to be quite tight. For instance, it gives the lower bound of $\frac{|F|(|F|+1)}{2}$ for Kakeya sets in F^2 (which was already implicitly observed by Wolff); this is very close to the exact bound, which was recently established in [**Ba2008**], [**BlMa2008**] to be $\frac{|F|(|F|+1)}{2} + \frac{|F|-1}{2}|F|$ in the case when $|F|$ is odd. (Thanks to Simeon Ball and Francesco Mazzocca for these references.)

It now seems sensible to revisit other problems in extremal combinatorics over finite fields to see if the polynomial method can yield results there. Certainly close relatives of the Kakeya conjecture (e.g. the Nikodym set conjecture, or the Kakeya maximal function conjecture) should now be establishable by these methods. On the other hand, there are other problems (such as the sum-product problem, Szemerédi-Trotter type theorems, and distance set problems) which are sensitive to the choice of field F (and in particular, whether that field contains a subfield of index 2); see [**BoKaTa2004**]. It would be interesting to see if there are ways to adapt the polynomial method in order to detect the existence of subfields.

Very recently, the polynomial method has also been extended to yield some progress on the Euclidean case; see Section 1.19.

Notes. This article first appeared at `terrytao.wordpress.com/2008/03/24`. Thanks to ninguem for corrections.

Seva posed the question of determining the asymptotic best density for a Kakeya set in, say, F_3^n, as $n \to \infty$.

Some pictures of Kakeya sets can be found at www.math.ucla.edu/ tao/jav or en.wikipedia.org/wiki/Kakeya_set.

Further discussion of Dvir's result can be found online at ilaba.wordpress. and quomodocumque.wordpress.com/2008/03/25.

1.4. The van der Corput lemma, and equidistribution on nilmanifolds

In this article I would like to record a version of van der Corput lemma which is particularly applicable for equidistribution of orbits on nilmanifolds, and morally underlies my paper [**GrTa2009c**] with Ben Green on this topic. As an application, I reprove an old theorem of Leon Green (Theorem 2.16.18) that gives a necessary and sufficient condition as to whether a linear sequence $(g^n x)_{n=1}^\infty$ on a nilmanifold G/Γ is equidistributed, which generalises the famous theorem of Weyl on equidistribution of polynomials, Theorem 2.6.26.

1.4.1. The classical van der Corput trick. The classical van der Corput trick (first used implicitly by Weyl) gives a means to establish the equidistribution of a sequence $(x_n)_{n=1}^\infty$ in a torus $\mathbf{T}^d$ (e.g. a sequence $(P(n) \bmod 1)_{n=1}^\infty$ in the unit circle $\mathbf{T} = \mathbf{R}/\mathbf{Z}$ for some function P, such as a polynomial). Recall that such a sequence is said to be *equidistributed* if one has

$$\frac{1}{N}\sum_{n=1}^{N} f(x_n) \to \int_{\mathbf{T}^d} f \tag{1.3}$$

as $N \to \infty$ for every continuous function $f : \mathbf{T}^d \to \mathbf{C}$; an equivalent[7] formulation of equidistribution is that

$$\frac{1}{N}|\{1 \le n \le N : x_n \in B\}| \to \mathrm{Vol}(B)$$

for every box B in the torus $\mathbf{T}^d$. Equidistribution is an important phenomenon to study in ergodic theory and number theory, but also arises in applications such as Monte Carlo integration and pseudo-random number generation.

A fundamental result in the subject is

[7] The equivalence can be deduced easily from *Urysohn's lemma*.

Theorem 1.4.1 (Weyl equidistribution theorem)**.** *A sequence $(x_n)_{n=1}^\infty$ in $\mathbf{T}^d$ is equidistributed if and only if the exponential sums*

$$\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i \chi(x_n)} \tag{1.4}$$

converge to zero for every non-trivial character $\chi : \mathbf{T}^d \to \mathbf{T}$, i.e. a non-zero continuous homomorphism to the unit cicle.

Proof. It is clear that (1.4) is a special case of (1.3). Conversely, (1.4) implies that (1.3) holds whenever f is a finite linear combination of characters $e^{2\pi i \chi}$. Applying the *Weierstrass approximation theorem*, we obtain the claim. □

The significance of the equidistribution theorem is that it reduces the study of equidistribution to the question of estimating exponential sums, which is a problem in analysis and number theory. For instance, from Theorem 1.4.1 and the geometric series formula we immediately obtain the following result:

Corollary 1.4.2 (Equidistribution of linear sequences in torii)**.** *Let $\alpha \in \mathbf{T}^d$. Then the sequence $(\alpha n)_{n=1}^\infty$ is equidistributed in $\mathbf{T}^d$ if and only if α is* totally irrational*, which means that $\chi(\alpha) \neq 0$ for all non-zero characters χ.*

For instance, the linear sequence $(\sqrt{2}n \bmod 1, \sqrt{3}n \bmod 1)$ is equidistributed in the two-torus $\mathbf{T}^2$, since $(\sqrt{2}, \sqrt{3})$ is totally irrational, but the linear sequence $(\sqrt{2}n \bmod 1, \sqrt{8}n \bmod 1)$ is not[8] (the character $\chi : (x, y) \mapsto y - 2x$ annihilates $(\sqrt{2}, \sqrt{8})$ and thus obstructs equidistribution).

One elementary but very useful tool for estimating exponential sums is *Weyl's differencing trick*, that ultimately rests on the humble Cauchy-Schwarz inequality. One formulation of this trick can be phrased as the following inequality (cf. Lemma 2.12.7):

Lemma 1.4.3 (van der Corput inequality)**.** *Let $a_1, a_2, \ldots$ be a sequence of complex numbers bounded in magnitude by 1. Then for any*

[8] Of course, in the latter case, the orbit is still equidistributed in a smaller torus, namely the kernel of the character χ mentioned above; this is an extremely simple case of *Ratner's theorem*. See Section 2.16 for further discussion.

$1 \leq H \leq N$ we have

$$|\frac{1}{N}\sum_{n=1}^{N} a_n| \ll (\frac{1}{H}\sum_{h=0}^{H-1} |\frac{1}{N}\sum_{n=1}^{N} a_{n+h}\overline{a_n}|)^{1/2} + O\left(\frac{H}{N}\right). \tag{1.5}$$

Proof. Observe that

$$\frac{1}{N}\sum_{n=1}^{N} a_n = \frac{1}{N}\sum_{n=1}^{N} a_{n+h} + O\left(\frac{H}{N}\right)$$

for every $0 \leq h \leq H-1$. Averaging this in h we obtain

$$\frac{1}{N}\sum_{n=1}^{N} a_n = \frac{1}{N}\sum_{n=1}^{N} \frac{1}{H}\sum_{h=0}^{H-1} a_{n+h} + O\left(\frac{H}{N}\right)$$

and hence by the Cauchy-Schwarz inequality

$$|\frac{1}{N}\sum_{n=1}^{N} a_n| \leq (\frac{1}{N}\sum_{n=1}^{N} |\frac{1}{H}\sum_{h=0}^{H-1} a_{n+h}|^2)^{1/2} + O\left(\frac{H}{N}\right).$$

Expanding out the square and rearranging a bit, we soon obtain the upper bound (1.5) (in fact one can sharpen the constants slightly here, though this will not be important for this discussion). □

The significance of this inequality is that it replaces the task of bounding a sum of coefficients a_n by that of bounding a sum of "differentiated" coefficients $a_{n+h}\overline{a_n}$. This trick is thus useful in "polynomial" type situations when the differentiated coefficients are often simpler than the original coefficients. One particularly clean application of this inequality is as follows:

Corollary 1.4.4 (Van der Corput's difference theorem)**.** *Let $(x_n)_{n=1}^{\infty}$ be a sequence in a torus $\mathbf{T}^d$ such that the difference sequences $(x_{n+h} - x_n)_{n=1}^{\infty}$ are equidistributed for every non-zero h. Then $(x_n)_{n=1}^{\infty}$ is itself equidistributed.*

Proof. By Theorem 1.4.1, it suffices to show that (1.4) holds for every non-trivial character χ. But by Lemma 1.4.3, we can bound the magnitude of the left-hand side of (1.4) by

$$\ll (\frac{1}{H}\sum_{h=0}^{H-1} |\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i\chi(x_{n+h})} e^{-2\pi i\chi(x_n)}|)^{1/2} + O\left(\frac{H}{N}\right) \tag{1.6}$$

for any fixed H.

Now we use the fact that χ is a character to simplify $e^{2\pi i\chi(x_{n+h})}e^{-2\pi i\chi(x_n)}$ as $e^{2\pi i\chi(x_{n+h}-x_n)}$. By hypothesis and the equidistribution theorem, the inner sum $\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i\chi(x_{n+h})}e^{-2\pi i\chi(x_n)}$ goes to zero as $N \to \infty$ for any fixed non-zero h; when instead h is zero, this sum is of course just 1. We conclude that for fixed H, the expression (1.6) is bounded by $O(1/H)$ in the limit $N \to \infty$. Thus the limit (or limit superior) of the magnitude of (1.4) is bounded in magnitude by $O(1/H)$ for every H, and is thus zero. The claim follows. □

By iterating this theorem, and using the observation that the difference sequence $(P(n+h) - P(n))_{n=1}^{\infty}$ of a polynomial sequence $(P(n))_{n=1}^{\infty}$ of degree d becomes a polynomial sequence of degree $d-1$ for any non-zero h, we can conclude by induction the following famous result of Weyl, generalising Corollary 1.4.2 (see also Theorem 2.1.12, Corollary 2.4.4, Theorem 2.6.26):

Theorem 1.4.5 (Equidistribution of polynomial sequences in torii)**.** *Let $P : \mathbf{Z} \to \mathbf{T}^d$ be a polynomial sequence taking values in a torus. Then the sequence $(P(n))_{n=1}^{\infty}$ is equidistributed in $\mathbf{T}^d$ if and only if $\chi(P(\cdot))$ is non-constant for all non-zero characters χ.*

In the one-dimensional case $d = 1$, this theorem asserts that a polynomial $P : \mathbf{Z} \to \mathbf{R}$ with real coefficients is equidistributed modulo one if and only if it has at least one irrational non-constant coefficient; thus for instance the sequence $(\pi n^3 + \sqrt{2}n^2 + \frac{1}{4}n \bmod 1)_{n=1}^{\infty}$ is equidistributed.

1.4.2. A variant of the trick. It turns out that van der Corput's difference theorem (Corollary 1.4.4) can be generalised to deal not just on torii, but on more general measure spaces with a torus action. Given a topological probability space (X, μ) (which we will take to be a Polish space to avoid various technicalities) and a sequence $(x_n)_{n=1}^{\infty}$ in X, we say that such a sequence is *equidistributed with respect to* μ if we have

$$\frac{1}{N}\sum_{n=1}^{N} f(x_n) \to \int_X f \, d\mu \tag{1.7}$$

for all continuous compactly supported functions $f : X \to \mathbf{C}$. This clearly generalises the previous notion of equidistribution, in which X was a torus and μ was uniform probability measure.

To motivate our generalised version of Corollary 1.4.4, we observe that the hypothesis "the sequence $(x_{n+h}-x_n)_{n=1}^\infty$ is equidistributed in $\mathbf{T}^d$" can be phrased in a more dynamical fashion (eliminating the subtraction operation, which is algebraic) as the equivalent assertion that the sequence of pairs $((x_{n+h}, x_n))_{n=1}^\infty$ in $\mathbf{T}^d \times \mathbf{T}^d$, after quotienting out by the action of the diagonal subgroup $(\mathbf{T})^d)^\Delta := \{(y,y) : y \in \mathbf{T}^d\}$, becomes equidistributed on the quotient space $\mathbf{T}^d \times \mathbf{T}^d/(\mathbf{T})^d)^\Delta$. This convoluted reformulation is necessary for generalisations, in which we do not have a good notion of subtraction, but we still have a good notion of group action and quotient spaces.

We can now prove

Proposition 1.4.6 (Generalised van der Corput difference theorem)**.** *Let (X, μ) be a (Polish) probability space with a continuous (right-)action of a torus $\mathbf{T}^d$, and let $\pi : X \to X/\mathbf{T}^d$ be the projection map onto the quotient space (which then has the pushforward measure $\pi_*\mu$. Let $(x_n)_{n=1}^\infty$ be a sequence in X obeying the following properties:*

(1) *(Horizontal equidistribution) The projected sequence $(\pi(x_n))_{n=1}^\infty$ in $X/\mathbf{T}^d$ is equidistributed with respect to $\pi_*\mu$.*

(2) *(Vertical differenced equidistribution) For every non-zero h, the sequence $((x_{n+h}, x_n)(\mathbf{T}^d)^\Delta)_{n=1}^\infty$ in the quotiented product space $(X \times X)/(\mathbf{T}^d)^\Delta$ is equidistributed with respect to some measure ν_h which is invariant under the action of the torus $\mathbf{T}^d \times \mathbf{T}^d/(\mathbf{T}^d)^\Delta$.*

Then $(x_n)_{n=1}^\infty$ is equidistributed with respect to μ.

Note that Corollary 1.4.4 is the special case of Proposition 1.4.6 in which X is itself the torus $\mathbf{T}^d$ with the usual translation action and uniform measure (so that the quotient space is a point).

Proof. We need to verify the property (1.3). If the function f was invariant under the action of the torus $\mathbf{T}^d$, then we could push it down to the quotient space $X/\mathbf{T}^d$ and the claim would follow from hypothesis 1. We may therefore subtract off the invariant component

$\int_{\mathbf{T}^d} f(\cdot y)\ dy$ from our function and assume instead that f has zero vertical mean in the sense that $\int_{\mathbf{T}^d} f(xy)\ dy = 0$ for all x. A Fourier expansion in the vertical variable (or the Weierstrass approximation theorem) then allows us to reduce to the case when f has a *vertical frequency* given by some non-zero character $\chi : \mathbf{T}^d \to \mathbf{T}$ of the torus, in the sense that $f(xy) = f(x)e^{2\pi i\chi(y)}$ for all $x \in X$ and $y \in \mathbf{T}^d$.

Now we apply van der Corput's inequality as in the proof of Corollary 1.4.4. Using these arguments, we find that it suffices to show that

$$\frac{1}{N}\sum_{n=1}^{N} f(x_{n+h})\overline{f(x_n)} \to 0$$

for each non-zero h. But the summand here is just the tensor product function $f \otimes \overline{f} : X \times X \to \mathbf{C}$ applied to the pair (x_{n+h}, x_n). The fact that f has a vertical frequency implies that $f \otimes \overline{f}$ is invariant with respect to the diagonal action $(\mathbf{T}^d)^\Delta$, and thus this function descends to the quotient space $(X \times X)/(\mathbf{T}^d)^\Delta$. On the other hand, as the vertical frequency is non-trivial, the latter function also has zero mean on every orbit of $\mathbf{T}^d \times \mathbf{T}^d/(\mathbf{T}^d)^\Delta$ and thus vanishes when integrated against ν_h. The claim then follows from hypothesis 2. □

As an application, let us prove the following result, first established in [**Gr1961**]:

Theorem 1.4.7 (Equidistribution of linear sequences in nilmanifolds)**.** *Let G/Γ be a* nilmanifold *(where we take the nilpotent group G to be connected for simplicity, although this is not strictly necessary), and let $g \in G$ and $x \in G/\Gamma$. Then $(g^n x)_{n=1}^\infty$ is equidistributed with respect to Haar measure on G/Γ if and only if $\chi(g^n x)$ is non-constant in n for every non-trivial* horizontal character $\chi : G/\Gamma \to \mathbf{T}$*, where a horizontal character is any continuous homomorphism $\chi : G \to \mathbf{T}$ that vanishes on Γ (and thus descends to G/Γ).*

This statement happens to contain[9] Weyl's result (Theorem 1.4.5) as a special case, because polynomial sequences can be encoded as linear sequences in nilmanifolds; but it is actually stronger, allowing

[9] It is also equivalent to Theorem 3.16.1.

extensions to generalised polynomials that involve the floor function $\lfloor \cdot \rfloor$ or the fractional part function $\{\}$. For instance, if we take

$$G := \begin{pmatrix} 1 & \mathbf{R} & \mathbf{R} \\ 0 & 1 & \mathbf{R} \\ 0 & 0 & 1 \end{pmatrix}; \Gamma := \begin{pmatrix} 1 & \mathbf{Z} & \mathbf{Z} \\ 0 & 1 & \mathbf{Z} \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$g := \begin{pmatrix} 1 & \alpha & \beta \\ 0 & 1 & \gamma \\ 0 & 0 & 1 \end{pmatrix}; x = \Gamma$$

for some real numbers α, β, γ then a computation shows that

$$g^n x = \begin{pmatrix} 1 & \{\alpha n\} & \{\beta n + \alpha \frac{n(n-1)}{2} - \{\alpha n\}\lfloor \gamma n \rfloor\} \\ 0 & 1 & \{\gamma n\} \\ 0 & 0 & 1 \end{pmatrix} \Gamma$$

and then Theorem 1.4.7 asserts that the triple

$$(\{\alpha n\}, \{\beta n + \alpha \frac{n(n-1)}{2} - \{\alpha n\}\lfloor \gamma n \rfloor\}, \{\gamma n\})$$

is equidistributed in the unit cube $[0,1]^3$ if and only if the pair (α, γ) is totally irrational (the rationality of β turns out to be irrelevant). Even for concrete values such as $\alpha = \sqrt{2}, \beta = 0, \gamma = \sqrt{3}$, it is not obvious how to establish this fact directly; for instance a direct application of Corollary 1.4.4 does not obviously simplify the situation.

Proof of Theorem 1.4.7. (Sketch) It is clear that if $\chi(g^n x)$ is constant for some non-trivial character, then the orbit $g^n x$ is trapped on a level set of χ and thus cannot equidistribute. Conversely, suppose that $\chi(g^n x)$ is never constant. We induct on the step s of the nilmanifold. The case $s = 0$ is trivial, and the case $s = 1$ follows from Corollary 1.4.2, so suppose inductively that $s \geq 2$ and that the claim has already been proven for smaller s. We then look at the vertical torus $G_s/(\Gamma \cap G_s) \equiv \mathbf{T}^d$, where G_s is the last non-trivial group in the lower central series (and thus central). The quotient of the nilmanifold G/Γ by this torus action turns out to be a nilmanifold of one lower step (in which G is replaced by G/G_s) and so the projection of the orbit $(g^n x)_{n=1}^\infty$ is then equidistributed by induction hypothesis. Applying Proposition 1.4.6, it thus suffices to check that for each non-zero h, the sequence of pairs $(g^{n+h} x, g^n x)$ in $G/\Gamma \times G/\Gamma$, after

quotienting out by the diagonal action of the torus, is equidistributed with respect to some measure which is invariant under the residual torus $\mathbf{T}^d \times \mathbf{T}^d/(\mathbf{T}^d)^\Delta$.

We first pass to the abelianisation (or *horizontal torus*) $G/G_2\Gamma$ of the nilmanifold, and observe that the projections $\pi(g^{n+h}x), \pi(g^n x)$ of the coefficients of the pair $(g^{n+h}x, g^n x)$ to this torus only differ by a constant $\pi(g^h)$. Thus the pair $(g^{n+h}x, g^n x)$ does not range freely in $G/\Gamma \times G/\Gamma$, but is instead constrained to a translate of a smaller nilmanifold $G/\Gamma \times_\pi G/\Gamma$, defined as the space of pairs (x, y) with $\pi(x) = \pi(y)$. After quotienting out also by the diagonal vertical torus, we obtain a nilmanifold coming from the group $(G \times_{G_2} G)/G_s^\Delta$, where $G \times_{G_2} G$ is the space of pairs (g, h) of group elements $g, h \in G$ whose projections to the abelianisation G/G_2 agree, and $G_s^\Delta := \{(g_s, g_s) : g_s \in G_s\}$ is the vertical diagonal group. But a short computation shows that this new group is at most $s-1$ step nilpotent. One can then apply the induction hypothesis to show the required equidistribution properties of (x_{n+h}, x_n), thus closing the induction by Proposition 1.4.6. □

There are many further generalisations of these results, including a polynomial version of Theorem 1.4.7 in [**Le2005**], [**Le2005b**] that also permits G to be disconnected, and quantitative versions of all of these results in [**GrTa2009c**].

Notes. This article first appeared at `terrytao.wordpress.com/2008/06/14`. Thanks to an anonymous commenter for corrections.

1.5. The strong law of large numbers

Let X be a real-valued random variable, and let $X_1, X_2, X_3, \ldots$ be an infinite sequence of independent and identically distributed copies of X. Let $\overline{X}_n := \frac{1}{n}(X_1 + \ldots + X_n)$ be the empirical averages of this sequence. A fundamental theorem in probability theory is the *law of large numbers*, which comes in both a weak and a strong form:

Theorem 1.5.1 (Weak law of large numbers)**.** *Suppose that the first moment $\mathbf{E}|X|$ of X is finite. Then $\overline{X}_n$ converges in probability to $\mathbf{E}X$, thus $\lim_{n\to\infty} \mathbf{P}(|\overline{X}_n - \mathbf{E}X| \geq \varepsilon) = 0$ for every $\varepsilon > 0$.*

Theorem 1.5.2 (Strong law of large numbers)**.** *Suppose that the first moment* $\mathbf{E}|X|$ *of* X *is finite. Then* $\overline{X}_n$ *converges almost surely to* $\mathbf{E}X$*, thus* $\mathbf{P}(\lim_{n\to\infty} \overline{X}_n = \mathbf{E}X) = 1$.

Remark 1.5.3. The concepts of convergence in probability and almost sure convergence in probability theory are specialisations of the concepts of *convergence in measure* and *pointwise convergence almost everywhere* in measure theory.

Remark 1.5.4. If one strengthens the first moment assumption to that of finiteness of the second moment $\mathbf{E}|X|^2$, then we of course have a more precise statement than the (weak) law of large numbers, namely the *central limit theorem*, but I will not discuss that theorem here. With even more hypotheses on X, one similarly has more precise versions of the strong law of large numbers, such as the *Chernoff inequality*, which I will again not discuss here.

The weak law is easy to prove, but the strong law (which of course implies the weak law, by the dominated convergence theorem) is more subtle, and in fact the proof of this law (assuming just finiteness of the first moment) usually only appears in advanced graduate texts. So I thought I would present a proof here of both laws, which proceeds by the standard techniques of the moment method and truncation. The emphasis in this exposition will be on motivation and methods rather than brevity and strength of results; there do exist proofs of the strong law in the literature that have been compressed down to the size of one page or less, but this is not my goal here.

1.5.1. The moment method. The moment method seeks to control the tail probabilities of a random variable (i.e. the probability that it fluctuates far from its mean) by means of moments, and in particular the zeroth, first or second moment. The reason that this method is so effective is because the first few moments can often be computed rather precisely. The first moment method usually employs *Markov's inequality*

$$\mathbf{P}(|X| \geq \lambda) \leq \frac{1}{\lambda}\mathbf{E}|X| \tag{1.8}$$

(which follows by taking expectations of the pointwise inequality $\lambda I(|X| \geq \lambda) \leq |X|$), whereas the second moment method employs

some version of *Chebyshev's inequality*, such as

$$\mathbf{P}(|X| \geq \lambda) \leq \frac{1}{\lambda^2}\mathbf{E}|X|^2 \tag{1.9}$$

(note that (1.9) is just (1.8) applied to the random variable $|X|^2$ and to the threshold λ^2).

Generally speaking, to compute the first moment one usually employs *linearity of expectation*

$$\mathbf{E}X_1 + \ldots + X_n = \mathbf{E}X_1 + \ldots + \mathbf{E}X_n,$$

whereas to compute the second moment one also needs to understand *covariances* $\mathbf{Cov}(X_i, X_j) := \mathbf{E}(X_i X_j) - \mathbf{E}(X_i)\mathbf{E}(X_j)$ (which are particularly simple if one assumes pairwise independence), thanks to identities such as

$$\mathbf{E}(X_1 + \ldots + X_n)^2 = \mathbf{E}X_1^2 + \ldots + \mathbf{E}X_n^2 + 2\sum_{1 \leq i < j \leq n} X_i X_j$$

or the normalised variant

$$\mathbf{Var}(X_1+\ldots+X_n) = \mathbf{Var}(X_1)+\ldots+\mathbf{Var}(X_n)+2\sum_{1 \leq i < j \leq n} \mathbf{Cov}(X_i, X_j). \tag{1.10}$$

Higher moments can in principle give more precise information, but often require stronger assumptions on the objects being studied, such as joint independence.

Here is a basic application of the first moment method:

Lemma 1.5.5 (Borel-Cantelli lemma)**.** *Let $E_1, E_2, E_3, \ldots$ be a sequence of events such that $\sum_{n=1}^\infty \mathbf{P}(E_n)$ is finite. Then almost surely, only finitely many of the events E_n are true.*

Proof. Let $I(E_n)$ denote the indicator function of the event E_n. Our task is to show that $\sum_{n=1}^\infty I(E_n)$ is almost surely finite. But by linearity of expectation, the expectation of this random variable is $\sum_{n=1}^\infty \mathbf{P}(E_n)$, which is finite by hypothesis. By Markov's inequality (1.8) we conclude that

$$\mathbf{P}(\sum_{n=1}^\infty I(E_n) \geq \lambda) \leq \frac{1}{\lambda}\sum_{n=1}^\infty \mathbf{P}(E_n).$$

Letting $\lambda \to \infty$ we obtain the claim. □

Returning to the law of large numbers, the first moment method gives the following tail bound:

Lemma 1.5.6 (First moment tail bound)**.** *If* $\mathbf{E}|X|$ *is finite, then*

$$\mathbf{P}(|\overline{X}_n| \geq \lambda) \leq \frac{\mathbf{E}|X|}{\lambda}.$$

Proof. By the triangle inequality, $|\overline{X}_n| \leq \overline{|X|}_n$. By linearity of expectation, the expectation of $\overline{|X|}_n$ is $\mathbf{E}|X|$. The claim now follows from Markov's inequality. □

Lemma 1.5.6 is not strong enough by itself to prove the law of large numbers in either weak or strong form - in particular, it does not show any improvement as n gets large - but it will be useful to handle one of the error terms in those proofs.

We can get stronger bounds than Lemma 1.5.6 - in particular, bounds which improve with n - at the expense of stronger assumptions on X.

Lemma 1.5.7 (Second moment tail bound)**.** *If* $\mathbf{E}|X|^2$ *is finite, then*

$$\mathbf{P}(|\overline{X}_n - \mathbf{E}(X)| \geq \lambda) \leq \frac{\mathbf{E}|X - \mathbf{E}(X)|^2}{n\lambda^2}.$$

Proof. A standard computation, exploiting (1.10) and the pairwise independence of the X_i, shows that the variance $\mathbf{E}|\overline{X}_n - \mathbf{E}(X)|^2$ of the empirical averages $\overline{X}_n$ is equal to $\frac{1}{n}$ times the variance $\mathbf{E}|X - \mathbf{E}(X)|^2$ of the original variable X. The claim now follows from Chebyshev's inequality (1.9). □

In the opposite direction, there is the *zeroth moment method*, more commonly known as the *union bound*

$$\mathbf{P}(E_1 \vee \ldots \vee E_n) \leq \sum_{j=1}^{n} \mathbf{P}(E_j)$$

or equivalently (to explain the terminology "zeroth moment")

$$\mathbf{E}(X_1 + \ldots + X_n)^0 \leq \mathbf{E}X_1^0 + \ldots + X_n^0$$

for any non-negative random variables $X_1, \ldots, X_n \geq 0$. Applying this to the empirical means, we obtain the *zeroth moment tail estimate*

$$\mathbf{P}(\overline{X}_n \neq 0) \leq n\mathbf{P}(X \neq 0). \tag{1.11}$$

Just as the second moment bound (Lemma 1.5.7) is only useful when one has good control on the second moment (or variance) of X, the zeroth moment tail estimate (1.10) is only useful when we have good control on the zeroth moment $\mathbf{E}|X|^0 = \mathbf{P}(X \neq 0)$, i.e. when X is mostly zero.

1.5.2. Truncation. The second moment tail bound (Lemma 1.5.7) already gives the weak law of large numbers in the case when X has finite second moment (or equivalently, finite variance). In general, if all one knows about X is that it has finite first moment, then we cannot conclude that X has finite second moment. However, we can perform a truncation

$$X = X_{\leq N} + X_{>N} \tag{1.12}$$

of X at any desired threshold N, where $X_{\leq N} := XI(|X| \leq N)$ and $X_{>N} := XI(|X| > N)$. The first term $X_{\leq N}$ has finite second moment; indeed we clearly have

$$\mathbf{E}|X_{\leq N}|^2 \leq N\mathbf{E}|X|$$

and hence also we have finite variance

$$\mathbf{E}|X_{\leq N} - \mathbf{E}X_{\leq N}|^2 \leq N\mathbf{E}|X|. \tag{1.13}$$

The second term $X_{>N}$ may have infinite second moment, but its first moment is well controlled. Indeed, by the monotone convergence theorem, we have

$$\mathbf{E}|X_{>N}| \to 0 \text{ as } N \to \infty. \tag{1.14}$$

By the triangle inequality, we conclude that the first term $X_{\leq N}$ has expectation close to $\mathbf{E}X$:

$$\mathbf{E}X_{\leq N} \to \mathbf{E}(X) \text{ as } N \to \infty. \tag{1.15}$$

These are all the tools we need to prove the weak law of large numbers:

Proof of Theorem 1.5.1. Let $\varepsilon > 0$. It suffices to show that whenever n is sufficiently large depending on ε, that $\overline{X}_n = \mathbf{E}X + O(\varepsilon)$ with probability $1 - O(\varepsilon)$.

From (1.14), (1.15), we can find a threshold N (depending on ε) such that $\mathbf{E}|X_{\geq N}| = O(\varepsilon^2)$ and $\mathbf{E}X_{<N} = \mathbf{E}X + O(\varepsilon)$. Now we use (1.12) to split

$$\overline{X}_n = (\overline{X_{\geq N}})_n + (\overline{X_{<N}})_n.$$

From the first moment tail bound (Lemma 1.5.6), we know that $(\overline{X_{\geq N}})_n = O(\varepsilon)$ with probability $1 - O(\varepsilon)$. From the second moment tail bound (Lemma 1.5.7) and (1.13), we know that $(\overline{X_{<N}})_n = \mathbf{E}X_{<N} + O(\varepsilon) = \mathbf{E}X + O(\varepsilon)$ with probability $1 - O(\varepsilon)$ if n is sufficiently large depending on N and ε. The claim follows. □

1.5.3. The strong law. The strong law can be proven by pushing the above methods a bit further, and using a few more tricks.

The first trick is to observe that to prove the strong law, it suffices to do so for non-negative random variables $X \geq 0$. Indeed, this follows immediately from the simple fact that any random variable X with finite first moment can be expressed as the difference of two non-negative random variables $\max(X, 0), \max(-X, 0)$ of finite first moment.

Once X is non-negative, we see that the empirical averages $\overline{X}_n$ cannot decrease too quickly in n. In particular we observe that

$$\overline{X}_m \leq (1 + O(\varepsilon))\overline{X}_n \text{ whenever } (1 - \varepsilon)n \leq m \leq n. \tag{1.16}$$

Because of this quasimonotonicity, we can *sparsify* the set of n for which we need to prove the strong law. More precisely, it suffices to show

Theorem 1.5.8 (Strong law of large numbers, reduced version)**.** *Let X be a non-negative random variable with $\mathbf{E}X < \infty$, and let $1 \leq n_1 \leq n_2 \leq n_3 \leq \ldots$ be a sequence of integers which is* lacunary *in the sense that $n_{j+1}/n_j > c$ for some $c > 1$ and all sufficiently large j. Then $\overline{X}_{n_j}$ converges almost surely to $\mathbf{E}X$.*

Indeed, if we could prove the reduced version, then on applying that version to the lacunary sequence $n_j := \lfloor (1 + \varepsilon)^j \rfloor$ and using

(1.16) we would see that almost surely the empirical means $\overline{X}_n$ cannot deviate by more than a multiplicative error of $1+O(\varepsilon)$ from the mean $\mathbf{E}X$. Setting $\varepsilon := 1/m$ for $m = 1, 2, 3, \ldots$ (and using the fact that a countable intersection of almost sure events remains almost sure) we obtain the full strong law.

Remark 1.5.9. This sparsification trick is philosophically related to the *dyadic pigeonhole principle* philosophy; see [**Ta3**]. One could easily sparsify further, so that the lacunarity constant c is large instead of small, but this turns out not to help us too much in what follows.

Now that we have sparsified the sequence, it becomes economical to apply the Borel-Cantelli lemma (Lemma 1.5.5). Indeed, by many applications of that lemma we see that it suffices to show that[10]

$$\sum_{j=1}^{\infty} \mathbf{P}(\overline{X}_{n_j} \neq \mathbf{E}(X) + O(\varepsilon)) < \infty \tag{1.17}$$

for non-negative X of finite first moment, any lacunary sequence $1 \leq n_1 \leq n_2 \leq \ldots$ and any $\varepsilon > 0$.

Remark 1.5.10. If we did not first sparsify the sequence, the Borel-Cantelli lemma would have been too expensive to apply; see Remark 1.5.12 below. Generally speaking, Borel-Cantelli is only worth applying when one expects the events E_n to be fairly "disjoint" or "independent" of each other; in the non-lacunary case, the events E_n change very slowly in n, which makes the lemma very inefficient. We will not see how lacunarity is exploited until the punchline at the very end of the proof, but certainly there is no harm in taking advantage of this "free" reduction to the lacunary case now, even if it is not immediately clear how it will be exploited.

At this point we go back and apply the methods that already worked to give the weak law. Namely, to estimate each of the tail probabilities $\mathbf{P}(\overline{X}_{n_j} \neq \mathbf{E}(X) + O(\varepsilon))$, we perform a truncation (1.12) at some threshold N_j. It is not immediately obvious what truncation to perform, so we adopt the usual strategy of leaving N_j unspecified for now and optimising in this parameter later.

[10] This is a slight abuse of the O() notation, but it should be clear what is meant by this.

We should at least pick N_j large enough so that $\mathbf{E}X_{<N_j} = \mathbf{E}X + O(\varepsilon)$. From the second moment tail estimate (Lemma 1.5.7) we conclude that $(\overline{X_{<N_j}})_{n_j}$ is also equal to $\mathbf{E}X + O(\varepsilon)$ with probability $1 - O\left(\frac{1}{\varepsilon n_j}\mathbf{E}|X_{\leq N_j}|^2\right)$. One could attempt to simplify this expression using (1.13), but this turns out to be a little wasteful, so let us hold off on that for now. However, (1.13) does strongly suggest that we want to take N_j to be something like n_j, which is worth keeping in mind in what follows.

Now we look at the contribution of $X_{\geq N_j}$. One could use the first moment tail estimate (Lemma 1.5.6), but it turns out that the first moment $\mathbf{E}X_{>N_j}$ decays too slowly in j to be of much use (recall that we are expecting N_j to be like the lacunary sequence n_j); the root problem here is that the decay (1.14) coming from the monotone convergence theorem is *ineffective*[11].

But there is one last card to play, which is the zeroth moment method tail estimate (1.11). As mentioned earlier, this bound is lousy in general - but is very good when X is mostly zero, which is precisely the situation with $X_{>N_j}$. and in particular we see that $(\overline{X_{>N_j}})_{n_j}$ is zero with probability $1 - O(n_j\mathbf{P}(X > N_j))$.

Putting this all together, we see that

$$\mathbf{P}(\overline{X}_{n_j} \neq \mathbf{E}(X) + O(\varepsilon)) \ll \frac{1}{\varepsilon n_j}\mathbf{E}|X_{\leq N_j}|^2 + n_j\mathbf{P}(X > N_j).$$

Summing this in j, we see that we will be done as soon as we figure out how to choose N_j so that

$$\sum_{j=1}^{\infty} \frac{1}{n_j}\mathbf{E}|X_{\leq N_j}|^2 \tag{1.18}$$

and

$$\sum_{j=1}^{\infty} n_j\mathbf{P}(X > N_j) \tag{1.19}$$

are both finite. As usual, we have a tradeoff: making the N_j larger makes (1.19) easier to establish at the expense of (1.18), and vice versa when making N_j smaller.

[11] One could effectivise this using the finite convergence principle, see Section 1.3 of *Structure and Randomness*, but this turns out to give very poor results here.

Based on the discussion earlier, it is natural to try setting $N_j := n_j$. Happily, this choice works cleanly; the lacunary nature of n_j ensures (basically from the geometric series formula) that we have the pointwise estimates

$$\sum_{j=1}^{\infty} \frac{1}{n_j} X_{\leq n_j} = O(X)$$

and

$$\sum_{j=1}^{\infty} n_j I(X \geq n_j) = O(X)$$

(where the implied constant here depends on the sequence $n_1, n_2, \ldots$, and in particular on the lacunarity constant c). The claims (1.17), (1.18) then follow from one last application of linearity of expectation, giving the strong law of large numbers.

Remark 1.5.11. The above proof in fact shows that the strong law of large numbers holds even if one only assumes pairwise independence of the X_n, rather than joint independence.

Remark 1.5.12. It is essential that the random variables $X_1, X_2, \ldots$ are "recycled" from one empirical average $\overline{X}_n$ to the next, in order to get the crucial quasimonotonicity property (1.16). If instead we took completely independent averages $\overline{X}_n = \frac{1}{n}(X_{n,1} + \ldots + X_{n,n})$, where the $X_{i,j}$ are all iid, then the strong law of large numbers in fact breaks down[12] with just a first moment assumption. Of course, if one restricts attention to a lacunary sequence of n then the above proof goes through in the independent case (since the Borel-Cantelli lemma is insensitive to this independence). By exploiting the joint independence further (e.g. by using *Chernoff's inequality*) one can also get the strong law for independent empirical means for the full sequence n under second moment bounds.

[12]For a counterexample, consider a random variable X which equals $2^m/m^2$ with probability 2^{-m} for $m = 1, 2, 3, \ldots$; this random variable (barely) has finite first moment, but for $n \sim 2^m/m^2$, we see that $\overline{X}_n$ deviates by at least absolute constant from its mean with probability $\gg 1/m^2$. As the empirical means $\overline{X}_n$ for $n \sim 2^m/m^2$ are now jointly independent, the probability that one of them deviates significantly is now extremely close to 1 (super-exponentially close in m, in fact), leading to the total failure of the strong law in this setting.

Remark 1.5.13. From the perspective of interpolation theory, one can view the above argument as an interpolation argument, establishing an L^1 estimate (1.17) by interpolating between an L^2 estimate (Lemma 1.5.7) and the L^0 estimate (1.11).

Remark 1.5.14. By viewing the sequence $X_1, X_2, \ldots$ as a stationary process, and thus as a special case of a measure-preserving system one can view the weak and strong law of large numbers as special cases of the mean and pointwise ergodic theorems respectively (see Exercise 2.8.9 and Theorem 2.9.4).

Notes. This article first appeared at `terrytao.wordpress.com/2008/06/18`. Thanks to toomuchcoffeeman and Joshua Batson for corrections.

Siva pointed out that for bounded random variables, a short proof of the strong law of large numbers (interpreting this law as an ergodic theorem for stationary processes) appears in [**Ke1995**].

Giovanni Pecatti noted that almost sure analogues of the central limit theorem exist, see e.g. [**Be1995**].

1.6. The Black-Scholes equation

In this article I would like to describe the mathematical derivation of the famous *Black-Scholes equation* in financial mathematics, at least in the simplified case in which time is discrete. This simplified model avoids many of the technicalities involving stochastic calculus, Itô's formula, etc., and brings the beautifully simple basic idea behind the derivation of this formula into focus.

The basic type of problem that the Black-Scholes equation solves (in particular models) is the following. One has an *underlying financial instrument* S, which represents some asset[13] which can be bought and sold at various times t, with the per-unit price S_t of the instrument varying with t. Given such an underlying instrument S, one can create *options* based on S and on some future time t_1, which give the buyer and seller of the options certain rights and obligations regarding S at an *expiration time* t_1. For instance,

[13]For the mathematical model, it is not relevant what type of asset S actually is, but one could imagine for instance that S is a stock, a commodity, a currency, or a bond.

(1) A *call option* for S at time t_1 and at a *strike price* P gives the buyer of the option the right (but not the obligation) to buy a unit of S from the seller of the option at price P at time t_1 (conversely, the seller of the option has the obligation but not the right to sell a unit of S to the buyer of the option at time t_1, if the buyer so requests).

(2) A *put option* for S at time t_1 and at a strike price P gives the buyer of the option the right (but not the obligation) to sell a unit of S to the seller of the option at price P at time t_1 (and conversely, the seller of the option has the obligation but not the right to buy a unit of S from the buyer of the option at time t_1, if the buyer so requests).

(3) More complicated options, such as *straddles* and *collars*, can be formed by taking linear combinations of call and put options, e.g. simultaneously buying or selling a call and a put option. One can also consider "American options" which offer rights and obligations for an interval of time, rather than the "European options" described above which only apply at a fixed time t_1. The Black-Scholes formula applies only to European options, though extensions of this theory have been applied to American options.

The problem is this: what is the "correct" price, at time t_0, to assign to an European option (such as a put or call option) at a future expiration time t_1? Of course, due to the volatility of the underlying instrument S, the future price S_{t_1} of this instrument is not known at time t_0. Nevertheless - and this is really quite a remarkable fact - it is still possible to compute deterministically, at time t_0, the price of an option that depends on that unknown price S_{t_1}, under certain assumptions (one of which is that one knows exactly *how* volatile the underlying instrument is).

1.6.1. How to compute price. Before we do any mathematics, we must first settle a fundamental financial question - how can one compute the price of some asset A? In most economic situations, such a price would depend on many factors, such as the supply and demand of A, transaction costs in buying or selling A, legal regulations

concerning A, or more intangible factors such as the current market sentiment regarding A. Any model that attempted to accurately describe all of these features would be hideously complicated and involve a large number of parameters that would be nearly impossible to measure directly. So, in general, one cannot hope to compute such prices mathematically.

But the situation is much simpler for purely financial products, such as options, at least when one has a highly deep and liquid market for the underlying instrument S. More precisely, we will make the following (unrealistic) assumptions:

- **Infinite liquidity** Market participants can buy or sell a unit of the underlying instrument S at any time[14].
- **Infinite depth** Each sale of a unit of S of does not affect the price of futher sales of units of S.
- **No transaction costs** The purchase price and sale price of an asset is the same: in other words, the money spent by a buyer in a sale is exactly equal to the money earned by the seller.
- **No arbitrage** There do not exist risk-free opportunities for market participants to instantaneously make money.

With these assumptions, the supply situation is simplified enormously, because any participant in this market can, in principle, use cash to create an option to sell to others (for instance one can sell a call option for S and cover it by buying a unit of S at any time before the expiration time), in contrast to physical assets (e.g. barrels of oil) which cannot be created purely from market transactions. This freedom of supply leads to upper bounds on the price of a financial asset A; if any market participant can instantaneously create a unit of A at time t_0 from market transactions using an amount X (or less) of cash, then clearly one should not assign such a unit of A a price greater than X at time t_0, otherwise there would exist an arbitrage opportunity.

[14]In principle, the participant would need a certain amount of cash, or a certain amount of S, in order to buy or sell S, but see the infinite credit and short selling assumptions below.

As a simple example of such an upper bound, if a deep and liquid market allows one to repeatedly buy individual units of A at a price of X per unit, then for any integer $k \geq 1$, the price of k units of A has an upper bound[15] of kX.

As another example, the price at time t_0 of a put option for a unit of S at time t_1 at strike price P cannot exceed[16] P, because any market participant can create (and then sell) such an option simply by setting aside P units of cash to cover the future expense of buying a unit of S. For similar reasons, the price at time t_0 of a call option for a unit of S at time t_1 cannot exceed S_{t_0}.

Dually to the above freedom of supply, there is also a freedom of demand: any participant can, in principle, purchase a financial asset and convert it into cash by combining the rights offered by that asset with other purchases. For instance, one could attempt to profit from a put option by buying a unit of the underlying instrument S and then (if the price is favourable) exercising the right to sell that unit to the option seller. This freedom of demand leads to *lower* bounds on the price of an asset: if any market participant can instantaneously convert a unit of A using market transactions into an amount X of cash, then clearly one should not assign a unit of A any price lower than X, otherwise there would be an arbitrage opportunity.

To give a trivial example: any option has a lower bound of zero[17] for its price, since one can convert an option into zero units of cash simply by refusing to exercise it.

To summarise so far: freedoms of supply give upper bounds on the price of an asset A, and freedoms of demand give lower bounds on the price of an asset A. The lower bounds cannot exceed the upper bounds, as this would provide an arbitrage opportunity. But

[15]The true price may be lower, due for instance to volume discounts, but in general the price of k units of A will be a subadditive function of A. Note though that if the market is not infinitely deep, then each purchase of a unit may increase the price of the next unit, leading to superadditive behaviour instead.

[16]This is an extremely crude upper bound, of course, as the option buyer might not exercise the option, in which case the P units of cash are recovered, or the option buyer does exercise in the option, in which case the seller is compensated for the P units of cash by a unit of S. Also, we are assuming here that there are no costs (e.g. security costs) associated with holding on an asset over time.

[17]Note that some financial assets can have a negative cash value - mortgages are a good example.

if the lower bounds and upper bounds happen to be equal, then one can compute the price of A exactly. This is a rare occurrence - one almost never expects the upper and lower bounds to be so tight. But, amazingly, this will turn out to be the case for options in the Black-Scholes model.

To give a simple example of a situation in which upper and lower bounds match, let us make another assumption:

- **Infinite credit** Market participants can borrow or lend arbitrary amounts of money at a risk-free interest rate of r. Thus, for instance, participants can deposit (or lend) X amount of cash at time t_0 and be guaranteed to receive $\exp(r(t_1-t_0))X$ cash at time t_1, and conversely can borrow X amount of cash at time t_0 but pay back $\exp(r(t_1-t_0))X$ cash at time t_1.

Remark 1.6.1. One can renormalise r to be zero, basically by using *real* units of currency instead of nominal units, but we will not do so here.

With this assumption one can now compute the *time value of money*. Suppose one has a risk-free government bond A which is guaranteed to pay out X amount of cash at the maturity time t_1 of the bond. Then, at any time t_0 prior to the maturity time, one can convert A to an amount $\exp(-r(t_1-t_0))X$ of cash, by borrowing this amount of cash at time t_0, and using the proceeds of the bond A to pay off the debt from this borrowing at time t_1. Thus there is a lower bound of $\exp(-r(t_1-t_0))X$ to the price of the bond A. Conversely, given an amount $\exp(-r(t_1-t_0))X$ of cash at time t_0, one can create the equivalent of the bond A simply by depositing or lending out this cash to obtain X amount of cash at time t_1. Thus, in this case the lower and upper bounds match exactly, and the price of the bond can be computed at time t_0 to be $\exp(-r(t_1-t_0))X$. (Because of this fact, the quantity r in the Black-Scholes model is usually set equal to the interest rate of an essentially risk-free asset, such as short-term Treasury bonds.)

One can use the time value of money to produce further upper and lower bounds on options. For instance, the price at time t_0 of a

put option for a unit of S at time t_1 at strike price P cannot be lower than $\exp(-r(t_1 - t_0))P - S_{t_0}$, since one can always convert the put option into this amount of cash by buying a unit of S at price S_{t_0} at time t_0, holding on to this unit until time t_1, and selling at price P at time t_1, which has the equivalent cash value of $\exp(-r(t_1 - t_0))P$ at time t_0. However, in order to make the lower and upper bounds match, we will need some additional assumptions on how the price S_t of the underlying stock evolves with time.

1.6.2. The Black-Scholes model. To simplify the computations, we shall assume

- **Discrete time** The time variable t increases in discrete steps of some time unit dt. (At each time t, one can make an arbitrary number of purchases and sale of assets, but the price S_t of the underlying instrument stays constant for each fixed t, as guaranteed by the infinite depth hypothesis.)

For instance, one could imagine a market in which the price S_t only changes once a day, so in this case dt would be a day in length. Similarly if S_t only changes once a minute or once a second.

The Black-Scholes model then describes how the next price S_{t+dt} of the underlying instrument depends on the current price S_t. The whole point, of course, is that there is to be some randomness (or risk) involved in this process. The simplest such model would be that of a simple random walk[18]

$$S_{t+dt} = S_t + \epsilon_t \sigma (dt)^{1/2}$$

where $\sigma > 0$ is a constant (representing volatility) and $\epsilon_t = \pm 1$ is a random variable, equal to $+1$ or -1 with equal probability; thus in this model the price either jumps up or jumps down by $\sigma(dt)^{1/2}$ for each time step dt. One can assume that the random variables ϵ_t are jointly independent as t varies, but remarkably we will not need to use such an independence hypothesis in our analysis. Similarly, we will not use the fact that the probabilities of going up or down are

[18]The factor of $(dt)^{1/2}$ is a natural normalisation, required for this model to converge to Brownian motion in the continuous time limit $dt \to 0$. with this normalisation, σ^2 basically becomes the amount of variance produced in S_t per unit time.

both equal to 1/2; it will turn out, unintuitively enough, that these probabilities are irrelevant to the final option price.

This simple model has a number of deficiencies. Firstly, it does not reflect the fact that many assets, while risky, will tend to grow in value over time. Secondly, the model allows for the possiblity that the price S_t becomes negative, which is clearly unrealistic. (A third deficiency, that it only allows two outcomes at each time step, is more serious, and will be discussed later.)

To address the first deficiency, one can add a drift term, thus leading to the model

$$S_{t+dt} = S_t + \mu dt + \sigma \epsilon_t (dt)^{1/2}$$

for some fixed $\mu \in \mathbf{R}$ (which could be positive, zero, or negative), representing the expected rate of appreciation of a unit of S per unit time. A remarkable (and highly unintuitive) consequence of Black-Scholes theory is that the exact value of μ will in fact have no impact on the final formula for the value of an option: an underlying instrument which is rising in value on average will have the same option pricing as one which is steady or even falling on the average!

To address the second deficiency, we work with the logarithm $\log S_t$ of the price of S, rather than the price itself, since this will make the price positive no matter how we move the logarithm up and down (as long as we only move the logarithm a finite amount, of course). More precisely, we adopt the model

$$\log S_{t+dt} = \log S_t + \mu dt + \sigma \epsilon_t (dt)^{1/2} \tag{1.20}$$

and so μ now measures the expected *relative* increase in value per unit time (as opposed to the expected absolute increase), and similarly σ^2 measures the relative increase in variance per unit time. This model may seem complicated, but the key point is that, given S_t, there are only two possible values of S_{t+dt}.

1.6.3. Pricing options. Now we begin the task of pricing an option with expiry date t_1 at time t_0. The interesting case is of course when t_0 is less than t_1, but to begin with let us first check what happens when $t_0 = t_1$, so that we are pricing an option that is expiring immediately.

Consider first a call option. If one has the option to buy a unit of S at price P at time t_1, and S_{t_1} was greater or equal to P, then it is clear that this option could be converted into $S_{t_1} - P$ units of cash, simply by exercising the option and then immediately selling the stock that was bought. Conversely, given $S_{t_1} - P$ units of cash, one could create such an option (and might even recover this money if the bearer of the option forgets to exercise it). So we see that when $S_{t_1} \geq P$, the price of this option is $S_{t_1} - P$.

On the other hand, if S_{t_1} is less than P (in the jargon, the option is "underwater" or "out of the money"), then it is intuitively clear that the call option is worthless (i.e. has a price of zero). To see this more rigorously, recall that any option has a lower bound of zero for its price. To get the upper bound, one can issue an underwater call option at no cost, since if someone is foolish enough to exercise that option, one can simply buy the stock from the open market at S_{t_1} and sell it for P, and pocket or discard the difference. Putting all this together, we see that the price V_{t_1} of the call option at time t_1 is a function of the price S_{t_1} of the underlying instrument at that time, and is given by the formula

$$V_{t_1}(S_{t_1}) := \max(S_{t_1} - P, 0). \tag{1.21}$$

For similar reasons, the price V_{t_1} at time t_1 of a put option for a unit of S at expiry time t_1 and strike price P is given by the formula

$$V_{t_1}(S_{t_1}) := \max(P - S_{t_1}, 0). \tag{1.22}$$

Thus we have worked out the price of both put and call options at the time of expiry. To handle the general case, we have to move backwards in time. For reasons that will become clearer shortly, we shall also need three final assumptions:

- **Infinite fungibility** Stock can be sold in arbitrary non-integer amounts.
- **Short selling** Market participants can borrow arbitrary amounts of stock, at no interest, for arbitrary amounts of time.
- **No storage costs** Market participants can hold arbitrary amounts of stock at no cost for arbitrary amounts of time.

The fundamental lemma here is the following:

Lemma 1.6.2. *If a financial asset A has a price at time t that is a function $V_t(S_t)$ that depends only on the price S_t of S at time t, then the same asset has a price at time $t-dt$ that is a function $V_{t-dt}(S_{t-dt})$ of the price S_{t-dt} of S at time $t-dt$, where V_{t-dt} is given from V_t by an explicit formula (see* (1.24) *below).*

Iterating this lemma, starting from (1.21) and (1.22), and taking the limit as $dt \to 0$, will ultimately lead to the Black-Scholes formula for the price of such options.

Let's see how this lemma is proven. Suppose we are at time $t-dt$, and the price of S is currently $s := S_{t-dt}$. We do not know what the price S_t of S at the next time step will be exactly, but thanks to (1.20), we know that it is one of two values, say s_- and s_+ with $s_+ > s_-$. From (1.20) we have the explicit formula

$$s_\pm = s\exp(1 + \mu dt \pm \sigma(dt)^{1/2}). \tag{1.23}$$

By hypothesis, we know that the instrument A has a price of $V_t(s_+)$ or $V_t(s_-)$ at time t, depending on whether S has a price of s_+ or s_- at this time t. Our task is now to show that A has a price at time t-dt that depends only on s.

Let us first consider the easy case when $V_t(s_+)$ and $V_t(s_-)$ are both equal to the same value, say X. In this case, the instrument A is (for the purposes of pricing) identical to a bond which matures at time t with a value of X. By the previous discussion, we thus see that the price of A at time t-dt is equal to $\exp(-rdt)X$.

Now consider the case when $V_t(s_+)$ and $V_t(s_-)$ are unequal. Then there is some risk in the value of A at time t. But - and this is the key point - one can *hedge* this risk by buying or selling some units of S. Suppose for instance one owns one unit of A at time $t-dt$, and then buys k units of S at this time at the price s. At time t, one sells the k units of S, earning ks_+ units of cash at time t if the price is s_+, and ks_- units if the price is s_-. In effect, this hedging strategy adjusts $V_t(s_+)$ and $V_t(s_-)$ to $V_t(s_+) + ks_+$ and $V_t(s_-) + ks_-$ respectively, at the cost of paying ks at time $t - dt$. If $V_t(s_+) < V_t(s_-)$, then one can find a positive k so that the adjusted values $V_t(s_+) + ks_+$

and $V_t(s_-) + ks_-$ of the instrument are equal (indeed, k is simply $k = (V_t(s_-) - V_t(s_+))/(s_+ - s_-)$). We have thus effectively converted A, at the cost of ks units of cash at time $t - dt$, into a bond that matures at time t with a value of

$$V_t(s_+) + ks_+ = V_t(s_-) + ks_- = \frac{s_+ V_t(s_-) - s_- V_t(s_+)}{s_+ - s_-}.$$

Conversely, we can convert such a bond into one unit of A and ks units of cash at time $t - dt$ by reversing the above procedure. Namely, instead of buying k units of S at time $t - dt$ to sell at time t, one instead *short sells* k units of S at time $t - dt$ to buy back at time t. More precisely, one borrows k units of stock at time $t - dt$ to sell immediately, and then at time t buys them back again to repay the stock loan. (Mathematically, this is equivalent to buying $-k$ units of stock at time $t - dt$ to sell at time t; thus short selling effectively allows one to buy negative units of stock, in much the same way that fungibility allows one to buy fractional units of stock.) We thus conclude that in this case, A has a value of

$$\exp(-rdt)\frac{s_+ V_t(s_-) - s_- V_t(s_+)}{s_+ - s_-} - ks = \frac{(\exp(-rdt)s_+ - s)V_t(s_-) - (\exp(-rdt)s$$

This analysis was conducted in the case $V_t(s_+) < V_t(s_-)$, but one can get the same formula at the end in the opposite case $V_t(s_+) > V_t(s_-)$; k is now negative in this case, but since buying a negative amount of stock is equivalent to short-selling a positive amount of stock (and vice versa), the arguments go through as before. Substituting the formula for k, we have thus proven the lemma, with

(1.24)
$$V_{t-dt}(s) := \frac{(\exp(-rdt)s_+ - s)V_t(s_-) - (\exp(-rdt)s_- - s)V_t(s_+)}{s_+ - s_-}.$$

This is a somewhat complicated formula, but it can be simplified by means of Taylor expansion (assuming for the moment that V_t is smooth). To illustrate the idea, let us make the simplifying assumption that $r = 0$. If we then Taylor expand

$$V_t(s_\pm) = V_t(s) + (s_\pm - s)\partial_s V_t(s) + \frac{1}{2}(s_\pm - s)^2 \partial_{ss} V_t(s) + O((dt)^{3/2})$$

(cautioning here that the implied constants in the $O()$ notation depend on all sorts of things, such as the third derivative of V_t) and

note that $s_+ - s_-$ is comparable to $(dt)^{1/2}$ in magnitude, then the right-hand side of (1.24) simplifies to

$$V_t(s) - \frac{1}{2}\partial_{ss}V_t(s)(s_+ - s)(s_- - s) + O((dt)^{3/2}).$$

Since

$$(s_+ - s)(s_- - s) = -s^2\sigma^2 dt + O((dt)^{3/2})$$

we thus obtain

$$V_{t-dt}(s) = V_t(s) + \frac{1}{2}s^2\sigma^2\partial_{ss}V_t(s)dt + O((dt)^{3/2}).$$

Performing Taylor expansion in t, we thus conclude

$$\partial_t V_t(s) = -\frac{1}{2}s^2\sigma^2\partial_{ss}V_t(s) + O((dt)^{1/2})$$

and so in the continuum limit $dt \to 0$ one (formally, at least) obtains the backwards heat equation

$$\partial_t V = -\frac{1}{2}s^2\sigma^2\partial_{ss}V.$$

A similar (but more complicated) computation can be made in the $r \neq 0$ case (or one can renormalise using real currency units, as remarked earlier), obtaining the *Black-Scholes PDE*

$$\partial_t V = -\frac{1}{2}s^2\sigma^2\partial_{ss}V - rs\partial_s V + rV.$$

Using (1.21) or (1.22) as an initial condition, one can then solve for V at time t_0; the quantity $V_{t_0}(S_{t_0})$ is then the price of the option[19] at time t_0.

The above analysis was not rigorous because the error terms were not properly estimated when taking the continuum limit $dt \to 0$, and also because the initial conditions (1.21), (1.22) were not smooth. The latter turns out to be a very minor difficulty, due to the smoothing nature of the Black-Scholes PDE (which is a parabolic equation) and also because one can use the comparison principle (which formalises the intuitively obvious fact that if a financial asset A is always worth more than an asset B at time t, then this is also the case at time $t-dt$) to approximate the non-smooth options (1.21), (1.22) by smooth ones. The former difficulty does require a certain amount of non-trivial

[19] V can be computed explicitly in terms of the error function, leading to the *Black-Scholes formula*, which we will not give here.

analysis (e.g. Fourier analysis or Itô's formula) but I will not discuss this here.

There is an enormous amount of literature aimed at relaxing the idealised hypotheses in the above analysis, for instance adding transaction costs, fluctuations in volatility, or more complicated financial features such as dividends. In some of these more general models, the upper and lower bounds for the prices of options cease to match perfectly, due to transaction costs or the inability to perfectly hedge away the risk; this for instance starts occurring when the underlying price S_t can fluctuate to three or more values from a fixed value of S_{t-dt}, as it then becomes impossible in general to make V constant for all of these values at once purely by buying and selling S. In particular, the reliability of the Black-Scholes model becomes suspect when the price movements of S differ significantly from the model (1.20), for instance if there are occasional very large price swings[20].

The other major issue with the Black-Scholes formula is that it requires one to compute the volatility σ, which is difficult to do in practice. In fact, the formula is often used in *reverse*, using the actual prices in option markets to deduce an implied volatility for an underlying instrument.

Notes. This article first appeared at `terrytao.wordpress.com/2008/07/01`.

Kenny Easwaran noted the unintuitive fact that the probabilities of the two possible values for the right-hand side of (1.20) turn out to be completely irrelevant for the purposes of pricing an option.

An anonymous commenter noted that in practice, the Black-Scholes formula (and more robust variants of this formula) are never applied directly, because the day-to-day volatility is almost impossible to compute. Instead, the formula is more often applied in reverse, to obtain an implied volatility from actual option prices which can then be used either as a convenient shorthand for options trading, or for pricing more exotic options based on existing prices of simpler options.

[20] The world stock and bond markets, in the months following the initial posting of this article, exhibited this phenomenon quite strongly.

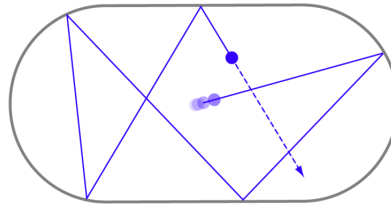

Figure 1. The Bunimovich stadium. (Figure from wikipedia.)

1.7. Hassell's proof of scarring for the Bunimovich stadium

In Section 3.5 of *Structure and Randomness*, I featured one of my favorite problems, namely that of establishing scarring for the Bunimovich stadium. I'm now happy to say that this problem has been solved (for generic stadiums, at least, and for phase space scarring rather than physical space scarring) by Andrew Hassell[**Ha2008**]. Actually, the argument is beautifully simple and short, though it of course uses the basic theory of eigenfunctions on domains, such as Weyl's law, and I would like the gist of it here

Let's first recall the problem. We consider a stadium domain formed by adjoining two semicircles on the ends of a rectangle, as in Figure 1.

We can normalise the rectangle to have height 1 and width t, and will call this stadium S_t. For reasons that will be clearer later, it is convenient to view t as a time parameter, so that the stadium is steadily getting elongated in time. The Laplacian on this domain (with Dirichlet boundary conditions) has a countable sequence of eigenfunctions $u_1, u_2, \ldots$ associated to an increasing sequence of eigenvalues $0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \ldots$, which we can normalise so that $\int_{S_t} |u_k|^2 = 1$ for all k. The conjecture is that the u_k do *not* equidistribute in physical space (or in phase space) in the limit $k \to \infty$, or in other words that quantum unique ergodicity fails. In physical space, the conjecture is as follows:

Conjecture 1.7.1 (Scarring conjecture)**.** *There exists a subset $A \subset \Omega$ and a sequence u_{k_j} of eigenfunctions with $\lambda_{k_j} \to \infty$, such that*

$\int_A |u_{k_j}|^2$ *does* not *converge to* $|A|/|\Omega|$. *Informally, the eigenfunctions either concentrate (or "scar") in* A*, or on the complement of* A.

There is some numerical evidence for this conjecture, as discussed in Section 3.5 of *Structure and Randomness*; more relevantly for Hassell's argument, there is also a heuristic argument, which we recall shortly.

Conjecture 1.7.1 only considered scarring in physical space. There is a (slightly weaker) form of this conjecture which considers scarring in *phase space* instead (thus the indicator function 1_A is replaced by a more general *pseudodifferential operator*); alternatively, one can phrase things using the *Wigner transform*. The precise statement is slightly technical and will not be given here.

Hassell's result is as follows:

Theorem 1.7.2. [**Ha2008**] *The phase space version of the scarring conjecture is true for* S_t *for almost every* $t > 0$.

Thus, for most stadiums[21], there is an infinite sequence of eigenfunctions which exhibit significant non-uniformity in phase space.

Hassell's argument relies on three ingredients:

1.7.1. The Heller-Zelditch argument. As discussed in Section 3.5 of *Structure and Randomness*, there is already a heuristic argument due to Heller[**He1991**] and refined by Zelditch[**Ze2004**], which almost gives the scarring already for any given stadium S_t - but it requires one to exclude eigenvalue concentration in an interval $[\pi^2 n^2 - O(1), \pi^2 n^2 + O(1)]$ for some integer n. The point is that the stadium already exhibits some explicit quasimodes (i.e. approximate eigenfunctions), namely the tensor products $v_n = \sin(\pi n y)\psi(x)$ for some suitable cutoff function $\psi(x)$. Note that $\Delta v_n = \pi^2 n^2 v_n + O(1)$, so morally this means that the spectrum of v_n with respect to the Laplacian is concentrated in the interval $[\pi^2 n^2 - O(1), \pi^2 n^2 + O(1)]$. On the other hand, this quasimode is highly scarred in phase space (it is extremely concentrated in momentum space). So if one knew

[21] In an appendix to [**Ha2008**], Hassell and Hillairet also extend this result to other partially rectangular domains.

that there were only $O(1)$ eigenfunctions in this interval, then[22] one of these eigenfunctions must itself be scarred (basically by the pigeonhole principle, or triangle inequality).

The difficulty, as discussed in *Structure and Randomness*, was that nobody knew how to prevent a lot of eigenvalues concentrating in the intervals $[\pi^2 n^2 - O(1), \pi^2 n^2 + O(1)]$ - the standard tool for understanding eigenvalue distribution, namely Weyl's law, had far too large an error term for this task. So we need some new ingredients...

1.7.2. The Hadamard eigenvalue variation formula. Now one starts exploiting the parameter t. As t varies, the eigenvalues and eigenfunctions of the Laplacian on S_t will of course change. How do they change? One can already get some understanding of what is going on by looking at the variation of eigenvalues and eigenvectors for self-adjoint *matrices* rather than operators. Suppose we have a family $A(t)$ of self-adjoint $n \times n$ matrices depending smoothly on a time parameter t, with some eigenvalue $\lambda_k(t)$ and eigenvector $u_k(t)$, also varying smoothly, thus

$$A(t)u_k(t) = \lambda_k(t)u_k(t).$$

We normalise the eigenvectors to have unit magnitude. We can differentiate both sides with respect to t using the product rule to obtain

$$\dot{A}(t)u_k(t) + A(t)\dot{u}_k(t) = \dot{\lambda}_k(t)u_k(t) + \lambda_k(t)\dot{u}_k(t).$$

Now we take the dot product with $u_k(t)$. Since we have normalised $u_k(t)$ to be a unit vector, we have $u_k(t) \cdot u_k(t) = 1$ and $u_k(t) \cdot \dot{u}_k(t) = 0$, and we conclude the variation formula[23]

$$\langle u_k(t), \dot{A}_k(t)u_k(t)\rangle = \dot{\lambda}_k(t).$$

Thus, the rate of change of the k^{th} eigenvalue λ_k can be computed by testing the rate of change of the matrix A against the normalised eigenvalue u_k.

---

[22] The above argument can be made rigorous with a dash of microlocal analysis; see [**Ha2008**].

[23] See Section 1.15 for further discussion of these sorts of variation formulae.

It turns out that one can do a similar thing for the Laplacian $\Delta = \Delta_t$ on the domain S_t. Since the domain S_t is growing with t, one could imagine that the Laplacian Δ is also "growing", and its "time derivative" should be given by something on the boundary ∂S_t. It requires some care to make this intuition precise, but in [**Ha2008**], Hassell was able to show a Hadamard-type variation formula

$$\dot{\lambda}_k(t) = -\int_{\partial S_t} (X \cdot n) |\partial_n u_k(t,x)|^2 \, ds \tag{1.25}$$

where ds is the length element on ∂S_t, n is the outward unit normal, and X is the vector field which equals $+\frac{1}{2}\partial_x$ on the right semicircle (this is the vector field that grows the width t of the stadium S_t at a unit rate).

Note that $X \cdot n$ is always non-negative; so the formula (1.25) implies that the eigenvalues are decreasing as the width t increases. This is consistent with *Weyl's law* $\lambda_k = \frac{4\pi}{|S_t|}(1 + o(1))k$ for these eigenvalues. Actually, one can be a bit more precise; heat kernel methods reveal that $|\partial_n u_k(t,x)| \sim \lambda_k^{1/2}$ on average, and so from (1.25) we expect to have

$$-\dot{\lambda}_k \sim \lambda_k \tag{1.26}$$

on the average, which is broadly consistent with Weyl's law.

1.7.3. Quantum unique ergodicity. The last trick in [**Ha2008**] is to prove Theorem 1.7.2 by contradiction. To illustrate the idea, let us suppose that the extreme opposite to Theorem 1.7.2 holds, namely that no scarring occurs for *any* stadium S_t. Informally, this means that any eigenfunction (with large eigenvalue) for any stadium will be approximately uniformly distributed in phase space.

According to Egorov's theorem, eigenfunctions should propagate their position and momentum in phase space by geodesic flow. Since all geodesics in the stadium hit the boundary, this in principle allows us to understand the distribution of an eigenfunction on the boundary in terms of the eigenfunction in the interior. Indeed, one can show that an eigenfunction which is uniformly distributed in phase space in the interior, will have a normal derivative which is uniformly distributed on the boundary (rigorous formulations of this fact date

back to [**GeLe1993**]. Thus, by assumption, *every* eigenvector is uniformly distributed on the boundary. Because of this, the eigenvalue decay (1.26) does not just hold on the average - it holds for *all* k. Thus all eigenvalues decay exponentially in t at a steady rate.

But once one has this, it is not hard to show that the eigenvalues cannot concentrate close to any given interval $[\pi^2 n^2 - O(1), \pi^2 n^2 + O(1)]$ for extended periods of time t. We then apply the Heller-Zelditch argument and get a contradiction. That's it! (Modulo details, of course.)

As always, there are several further directions of research to pursue, for instance to improve the scarring so that one obtains non-equidistribution in physical space. This seems to be related to the question of improving the quality of the quasimode used in the Heller-Zelditch argument; see [**BuZw**] for further discussion.

Notes. This article first appeared at `terrytao.wordpress.com/2008/07/07`.

1.8. Tate's proof of the functional equation

The *Riemann zeta function* $\zeta(s)$, defined for $\mathrm{Re}(s) > 1$ by the formula

$$\zeta(s) := \sum_{n \in \mathbf{N}} \frac{1}{n^s} \tag{1.27}$$

where $\mathbf{N} = \{1, 2, \ldots\}$ are the natural numbers, and extended meromorphically to other values of s by analytic continuation, obeys the remarkable *functional equation*

$$\Xi(s) = \Xi(1-s) \tag{1.28}$$

where

$$\Xi(s) := \Gamma_\infty(s)\zeta(s) \tag{1.29}$$

is the Riemann Xi function,

$$\Gamma_\infty(s) := \pi^{-s/2}\Gamma(s/2) \tag{1.30}$$

is the *Gamma factor at infinity*, and the *Gamma function* $\Gamma(s)$ is defined for $\mathrm{Re}(s) > 1$ by

$$\Gamma(s) := \int_0^\infty e^{-t} t^s \, \frac{dt}{t} \tag{1.31}$$

and extended meromorphically to other values of s by analytic continuation.

There are many proofs known of the functional equation (1.28). One of them (dating back to [**Ri1859**][24]) relies on the *Poisson summation formula*

$$\sum_{a\in\mathbf{Z}} f_\infty(at_\infty) = \frac{1}{|t|_\infty} \sum_{a\in\mathbf{Z}} \hat{f}_\infty(a/t_\infty) \tag{1.32}$$

for the reals[25] $k_\infty := \mathbf{R}$ and $t \in k_\infty^*$, where f is a *Schwartz function*, $|t|_\infty := |t|$ is the usual *Archimedean absolute value* on k_∞, and

$$\hat{f}_\infty(\xi_\infty) := \int_{k_\infty} e_\infty(-x_\infty\xi_\infty) f_\infty(x_\infty)\, dx_\infty \tag{1.33}$$

is the Fourier transform on k_∞, with $e_\infty(x_\infty) := e^{2\pi i x_\infty}$ being the standard *character* $e_\infty : k_\infty \to S^1$ on k_∞. Applying this formula to the (Archimedean) Gaussian function

$$g_\infty(x_\infty) := e^{-\pi|x_\infty|^2}, \tag{1.34}$$

which is its own (additive) Fourier transform, and then applying the *multiplicative* Fourier transform (i.e. the *Mellin transform*), one soon obtains (1.28). One can "clean up" this proof a bit by replacing the Gaussian by a Dirac delta function, although one now has to work formally and "renormalise" by throwing away some infinite terms[26]. Note how this proof combines the additive Fourier transform with the multiplicative Fourier transform[27].

In the famous 1950 thesis of Tate (see e.g. [**CaFr1967**, Chapter XV]), the above argument was reinterpreted using the language of the *adele ring* $\mathbf{A}$, with the Poisson summation formula (1.30) on k_∞

[24] Riemann also had another proof of the functional equation relying primarily on contour integration, which I will not discuss here.

[25] The reason for this rather strange notation for the real line and its associated structures will be made clearer shortly.

[26] One can use the theory of distributions to make this approach rigorous, but I will not discuss this here.

[27] Continuing with this theme, the Gamma function (1.31) is an inner product between an additive character e^{-t} and a multiplicative character t^s, and the zeta function (1.27) can be viewed both additively, as a sum over n, or multiplicatively, as an Euler product.

replaced by the Poisson summation formula

$$\sum_{a\in k} f(at) = \sum_{a\in k} \hat{f}(t/a) \tag{1.35}$$

on $\mathbf{A}$, where $k = \mathbf{Q}$ is the rationals, $t \in \mathbf{A}$, and f is now a *Schwartz-Bruhat function* on $\mathbf{A}$. Applying this formula to the adelic (or global) Gaussian function $g(x) := g_\infty(x_\infty) \prod_p 1_{\mathbb{Z}_p}(x_p)$, which is its own Fourier transform, and then using the adelic Mellin transform, one again obtains (1.28). Again, the proof can be cleaned up by replacing the Gaussian with a Dirac mass, at the cost of making the computations formal (or requiring the theory of distributions).

In this post I will write down both Riemann's proof and Tate's proof together (but omitting some technical details), to emphasise the fact that they are, in some sense, the same proof. However, Tate's proof gives a high-level clarity to the situation (in particular, explaining more adequately why the Gamma factor at infinity (1.30) fits seamlessly with the Riemann zeta function (1.27) to form the Xi function (1.28)), and allows one to generalise the functional equation relatively painlessly to other zeta-functions and L-functions, such as Dedekind zeta functions and Hecke L-functions.

1.8.1. Riemann's proof. Applying the Poisson summation formula (1.28) for k_∞ to the Schwartz function (1.34), we see that the *theta function*

$$\Theta_\infty(x_\infty) := \sum_{n\in\mathbf{Z}} g_\infty(nx_\infty) = 1 + 2\sum_{n=1}^{\infty} e^{-\pi n^2 |x_\infty|_\infty^2} \tag{1.36}$$

obeys the functional equation

$$\Theta_\infty(x_\infty) = \frac{1}{|x_\infty|_\infty} \Theta_\infty(\frac{1}{x_\infty}) \tag{1.37}$$

for $x_\infty \in k_\infty^\times := k_\infty \backslash \{0\}$. In particular, since $\Theta_\infty(x_\infty) - 1$ is rapidly decreasing as $x_\infty \to \infty$, we see that $\Theta_\infty(x_\infty) - 1/x_\infty$ is rapidly decreasing as $x_\infty \to 0$.

Formally, we can take Mellin transforms of (1.37) and conclude that

$$\int_{k_\infty^\times} \Theta_\infty(x_\infty) |x_\infty|_\infty^s d^\times x_\infty = \int_{k_\infty^\times} \Theta_\infty(x_\infty) |x_\infty|_\infty^{1-s} d^\times x_\infty \tag{1.38}$$

for any s, where $d^\times x_\infty := \frac{dx_\infty}{|x_\infty|_\infty}$ is the standard multiplicative Haar measure on $k_\infty^\times$. This does not make rigorous sense, because the integrands here diverge at 0 and at infinity (which is ultimately due to the poles of the Riemann Xi function at $s = 0$ and $s = 1$), but let us forge ahead regardless. By making the change of variables $y := \pi n^2 t^2$ and using (1.29), (1.30), we see that

$$\int_{k_\infty} e^{-\pi n^2 x_\infty^2} |x_\infty|_\infty^s d^\times x_\infty = \Gamma_\infty(s) n^{-s} \tag{1.39}$$

and so from (1.36) and (1.27) we formally have

$$\int_{k_\infty} \Theta_\infty(x_\infty) |x_\infty|_\infty^s d^\times x_\infty = \int_{k_\infty} |x_\infty|^s d^\times x_\infty + 2\Gamma_\infty(s)\zeta(s). \tag{1.40}$$

If we casually discard the divergent integral $\int_{k_\infty} |x_\infty|^s d^\times x_\infty$ and apply (1.38), we formally obtain the functional equation (1.28).

Of course, the above computations were totally formal in nature. Nevertheless it is possible to make the argument rigorous. For instance, when $\mathrm{Re}(s) > 1$, we have a rigorous version of (1.40), namely

$$\int_{k^\infty} (\Theta_\infty(x_\infty) - 1) |x_\infty|_\infty^s d^\times x_\infty = 2\Gamma_\infty(s)\zeta(s), \tag{1.41}$$

which can be deduced from (1.39) and Fubini's theorem (or by dominated convergence). Using (1.37) and a little undergraduate calculus, we can rewrite the left-hand side of (1.41) as

$$\int_1^\infty (\Theta_\infty(t) - 1)(t^s + t^{1-s}) \frac{dt}{t} - \frac{1}{s} - \frac{1}{1-s}. \tag{1.42}$$

Observe that this expression extends meromorphically to all of s and can thus be taken as a definition of $\Xi(s)$ for all $s \neq 0, 1$, and the functional equation (1.28) is then manifestly obvious.

Here is a slightly different way to view the above computations. Since the Gaussian (1.34) is its own Fourier transform, we see for every $t > 0$ that the Fourier transform of $e^{-\pi t^2 |x|_\infty^2}$ is $\frac{1}{t} e^{-\pi |x|_\infty^2 / t^2}$. Integrating this fact against $|t|^s d^\times t$ on $k_\infty^\times$ using (1.39), we obtain (formally) at least that the Fourier transform[28] of $\Gamma_\infty(s)|x|_\infty^{-s} d^\times x$

[28]Note from scaling considerations it is formally clear that the Fourier transform of $|x|_\infty^{-s} d^\times x$ must be some sort of constant multiple of $|x|_\infty^{1-s} d^\times x$; the Gamma factors can thus be viewed as the normalisation of these multiplicative characters that is compatible with the Fourier transform.

is $\Gamma_\infty(1-s)|x|_\infty^{1-s} d^\times x$. Formally applying the Poisson summation formula (1.30) to this, and casually discarding the singular terms at the origin, we obtain (1.28). One can make the above computations rigorous using the theory of distributions, and by using Gaussians to regularise the various integrals and summations appearing here, in which case the computations become essentially equivalent to the previous ones.

1.8.2. p-adic analogues. The above "Archimedean" Fourier analysis on $k_\infty = \mathbf{R}$ has analogues in the *p-adic completions* $k_p = \mathbf{Q}_p$ of the rationals $k = \mathbf{Q}$. Recall that the reals $k_\infty = \mathbf{R}$ are the metric completion of the rationals $k = \mathbf{Q}$ with respect to the metric arising from the usual Archimedean absolute value $x \mapsto |x|_\infty$. This absolute value obeys the following basic properties:

(1) Positivity: we have $|x| \geq 0$ for all x, with equality if and only if $x = 0$.

(2) Multiplicativity: we have $|xy| = |x||y|$ for all x, y.

(3) Triangle inequality: We have $|x+y| \leq |x| + |y|$ for all x, y.

A function from k to $[0, +\infty)$ with the above three properties is known as an *absolute value* (or *valuation*) on k. In addition to the Archimedean absolute value, each prime p defines a p-adic absolute value $x \mapsto |x|_p$ on k, defined by the formula $|x|_p := p^{-n}$, where n is the number[29] of times p divides x. Equivalently, $|x|_p$ is the unique valuation such that $|p|_p = 1/p$ and $|n|_p = 1$ whenever n is an integer coprime to p. One easily verifies that $|x|_p$ is an absolute value; in fact it not only obeys the triangle inequality, but also the *ultra-triangle* inequality $|x+y| \leq \max(|x|, |y|)$, making the p-adic absolute value a non-Archimedean absolute value.

A classical theorem of Ostrowski asserts that the Archimedean absolute value $x \mapsto |x|_\infty$ and the p-adic absolute values $x \mapsto |x|_p$ are in fact the *only* absolute values on the rationals k, up to the renormalisation of replacing an absolute value $|x|$ with a power $|x|^\alpha$. If we define a *place* to be an absolute value up to renormalisation, we thus see that the rationals k have one Archimedean (or infinite) place

[29] This number could be negative if the denominator of the rational number x contains factors of p.

∞, together with one non-Archimedean (or finite) place p for every prime. One could have set $|p|_p$ to some other value between 0 and 1 than $1/p$ (thus replacing $|x|_p$ with some power $|x|_p^\alpha$) and still get an absolute value; but this normalisation is natural because it allows one to write the fundamental theorem of arithmetic in the appealing form

$$\prod_\nu |x|_\nu = 1 \text{ for all } x \in k^\times \tag{1.43}$$

where ν ranges over all places, and $k^\times := k\backslash\{0\}$ is the multiplicative group of k. If one takes the metric completion of the rationals $k = \mathbf{Q}$ using a p-adic absolute value $|x|_p$ rather than the Archimedean one, one obtains the p-adic field $k_p = \mathbf{Q}_p$. One can view this field as a kind of inverted version of the real field $\mathbf{R}$, in which p has been inverted to be small rather than large. Some illustrations of this inversion:

(1) In k_∞, the sequence p^n goes to infinity as $n \to +\infty$ and goes to zero as $n \to -\infty$; in k_p, it is the other way around.

(2) Elements of k_∞ can be expressed base p as strings of digits that need not terminate to the right of the decimal point, but must terminate to the left. In k_p, it is the other way around[30].

(3) In k_∞, the integers $\mathbf{Z}$ is closed and forms a discrete cocompact additive subgroup. In k_p, the integers are not closed, but their closure $\mathcal{O}_p = \mathbf{Z}_p$ (the ring of p-adic integers) forms a compact codiscrete additive subgroup.

Despite this inversion, we can obtain analogues of most of the additive and multiplicative Fourier analytic computations of the previous section for the p-adics.

Let's first begin with the additive Fourier structure. By the theory of Haar measures, there is a unique translation-invariant measure dx_p on $k_p = \mathbf{Q}_p$ which assigns a unit mass to the compact codiscrete subgroup $\mathcal{O}_p = \mathbf{Z}_p$. One can check that this measure interacts with

[30] The famous ambiguity $0.999\ldots = 1.000\ldots$ in k_∞ does not occur in the p-adic field k_p, because the latter has the topology of a Cantor space rather than a continuum. See also Section 1.6 of *Structure and Randomness*.

dilations in the expected manner, thus

$$\int_{k_p} f(tx_p)\ dx_p = \frac{1}{|t|} \int_{k_p} f(x_p)\ dx_p \tag{1.44}$$

for all absolutely integrable f and all invertible $t \in k_p^\times := k_p \backslash \{0\}$.

Just as k_∞ has a standard character $e_\infty : x \mapsto e^{2\pi i x}$, we can define a standard character $e_p : k_p \to S^1$ as the unique character (i.e. continuous homomorphism from k_p to S^1) such that $e_p(p^n) = e^{2\pi i p^n}$ for all integers n (in particular, e_p is trivial on the integers, just as e_∞ is). One easily verifies that this is indeed a character. From this and the additive Haar measure dx_p, we can now define the p-adic Fourier transform

$$\hat{f}(\xi_p) := \int_{k_p} e_p(-x_p \xi_p) f(x_p)\ dx_p \tag{1.45}$$

for reasonable (e.g. absolutely integrable) f, and it is a routine matter to verify all the usual Fourier-analytic identities for this transform (or one can appeal to the general theory of Fourier analysis on locally compact abelian groups).

In k_∞, we have the Gaussian function (1.34), which is its own Fourier transform. In k_p, the analogous Gaussian function $g_p : k_p \to \mathbf{C}$ is given by the formula

$$g_p := 1_{\mathcal{O}_p}, \tag{1.46}$$

i.e. the p-adic Gaussian is just the indicator function of the p-adic integers. One easily verifies that this function is also its own Fourier transform.

Now we turn to the multiplicative Fourier theory for k_p. The natural multiplicative Haar measure $d^\times x_p$ on $k_p^\times$ is given by the formula $d^\times x_p := \frac{p}{p-1} \frac{dx_p}{|x_p|_p}$; the normalisation factor $\frac{p}{p-1}$ is natural in order for the group of units $\mathcal{O}_p^\times = \mathbf{Z}_p \backslash p\mathbf{Z}_p$ to have unit mass.

In k_∞, we see from (1.39) that the Gamma factor at infinity can be expressed (for $\mathrm{Re}(s) > 1$) as the Mellin transform of the Gaussian:

$$\Gamma_\infty(s) = \int_{k_\infty^\times} g_\infty(x_\infty) |x_\infty|_\infty^s d^\times x_\infty. \tag{1.47}$$

In analogy with this, we can define the Gamma factor at p by the formula

$$\Gamma_p(s) = \int_{k_p^\times} g_p(x_p)|x_p|_p^s d^\times x_p. \tag{1.48}$$

Due to the simple and explicit nature of all the expressions on the right-hand side, it is a straightforward matter to compute this factor explicitly; it becomes

$$\Gamma_p(s) = (1-p^{-s})^{-1} \tag{1.49}$$

for $\mathrm{Re}(s) > 1$, at least; of course, one can then extend Γ_p meromorphically in the obvious manner.

In k_∞, we showed (formally, at least) that $\Gamma_\infty(s)|x|_\infty^s$ and $\Gamma_\infty(1-s)|x|_\infty^{1-s}$ were Fourier transforms of each other. One can similarly show that $\Gamma_p(s)|x|_p^s$ and $\Gamma_p(1-s)|x|_p^{1-s}$ are Fourier transforms of each other in k_p. On the other hand, there is no obvious analogue of the Poisson summation formula manipulations for k_p, because (unlike k_∞), k_p lacks a discrete cocompact subgroup.

1.8.3. Tate's proof. We have just performed some *local*[31] additive and multiplicative Fourier analysis at a single place. In his famous thesis, Tate observed that all these local Fourier-analytic computations could be unified into a single global Fourier-analytic computation, using the languge of the adele ring $\mathbf{A}$. This ring is the set[32] of all tuples $x = (x_\nu)_\nu$, where ν ranges over places and $x_\nu \in k_\nu$, and furthermore all but finitely many of the x_ν lie in their associated ring of integers $\mathcal{O}_\nu$. This restriction that the x_ν consists mostly of integers in k_ν is important for a large variety of analytic and algebraic reasons; for instance, it keeps the adele ring σ-compact.

[31] This use of "local" may seem unrelated to the topological or analytical notion of "local", as in "in the vicinity of a single point", but it is actually much the same concept; compare for instance the formal power series in p for a p-adic number with the Taylor series expansion in $t - t_0$ of a function $f(t)$ around a point t_0. Indeed one can view local analysis at a place p as being the analysis of the integers or rationals when p is "close to zero"; one can make this precise using the language of schemes, but we will not do so here.

[32] Equivalently, the adele ring is the tensor product of the rationals $k = \mathbf{Q}$ with the ring of integral adeles $\mathbf{R} \times \prod_p \mathbf{Z}_p$.

Many of the structures and objects on the local fields k_ν can be multiplied together to form corresponding global structures on the adele ring. For instance:

(1) The commutative ring structures on the k_ν multiply together to give a commutative ring structure on $\mathbf{A}$.

(2) The locally compact Hausdorff structures on the k_ν multiply together to give a locally compact Hausdorff structure on $\mathbf{A}$.

(3) The local additive Haar measures dx_ν on the k_ν multiply together to give a global additive Haar measure dx on $\mathbf{A}$.

(4) The local characters $e_\nu : k_\nu \to S^1$ on the k_ν multiply together to give a global character $e : \mathbf{A} \to S^1$ (here it is essential that most components of an adele are integers, and so are trivial with respect to their local character).

(5) The local additive Fourier transforms on the k_ν then multiply to form a global additive Fourier transform on $\mathbf{A}$, defined as $\hat{f}(\xi) := \int_{\mathbf{A}} e(-x\xi) f(x)\, dx$ for reasonable f.

(6) The local absolute values $x_\nu \mapsto |x_\nu|_\nu$ on k_ν multiply to form a global absolute value $x \mapsto |x|$ on $\mathbf{A}$, though with the important caveat that $|x|$ can vanish for non-zero x (indeed, a simple calculation using Euler's observation $\prod_p (1 - \frac{1}{p}) = 0$ shows that almost every x does this, with respect to additive Haar measure). The x for which $|x|$ is non-zero are invertible and known as ideles, and form a multiplicative group $\mathbf{A}^\times$; the ideles have measure zero inside the adeles.

(7) The local gaussians $g_\nu : k_\nu \to \mathbf{C}$ multiply together to form a global gaussian $g : \mathbf{A} \to \mathbf{C}$, which is its own Fourier transform.

(8) The embeddings $k \subset k_\nu$ at each place ν multiply together to form a diagonal embedding $k \subset \mathbf{A}$. This embedding is both discrete (by the fundamental theorem of arithmetic) and cocompact (this is basically because the integers are cocompact in the adelic integers).

(9) The local multiplicative Haar measures $d^\times x_\nu$ multiply together to form a global multiplicative Haar measure $d^\times x$,

though one should caution that this measure is supported on the ideles $\mathbf{A}^\times$ rather than the adeles $\mathbf{A}$.

(10) The local Gamma factors $\Gamma_\nu(s)$ for each place ν multiply together to form the Riemann Xi function (1.29) (for $\mathrm{Re}(s) > 1$ at least), thanks to the Euler product formula $\zeta(s) = \prod_p (1-p^{-s})^{-1}$.

Recall that the local Gamma factors were the local Mellin transforms of the local Gaussians. Multiplying this together, we see that the Riemann Xi function is the global Mellin transform of the global Gaussian:

$$\Xi(s) = \int_{\mathbf{A}^\times} g(x)|x|^s d^\times x. \tag{1.50}$$

Our derivation of (1.50) used the Euler product formula. Another way to establish (1.50) using the original form (1.27) of the zeta function is to observe (thanks to the fundamental theorem of arithmetic) that the set $J := \mathbf{R}^+ \times \prod_p \mathcal{O}_p^\times$ is a fundamental domain for the action of $k^\times$ on $\mathbf{A}^\times$, thus

$$\mathbf{A}^\times = \biguplus_{a \in k^\times} a \cdot J. \tag{1.51}$$

Partitioning (1.50) using (1.51) and then using (1.39) and (1.27) one can give an alternate derivation[33] of (1.50).

In his thesis, Tate established the Poisson summation formula (1.35) for the adeles for all sufficiently nice f (e.g. any f in the *Schwartz-Bruhat class* would do). Applying this to the global gaussian g, we conclude that the global Theta function $\Theta(x) := \sum_{a \in k} g(ax)$ obeys the functional equation

$$\Theta(x) = \frac{1}{|x|}\Theta(\frac{1}{x}) \tag{1.52}$$

for all ideles t. This formally implies that

$$\int_J \Theta(x)|x|^s d^\times x = \int_J \Theta(x)|x|^{1-s} d^\times x, \tag{1.53}$$

which on applying (1.51) and (1.50) (and casually discarding the singular contributions of $a = 0$) yields the functional equation (1.28).

[33] The two derivations are ultimately the same, of course, since the Euler product formula is itself essentially a restatement of the fundamental theorem of arithmetic.

One can make this formal computation rigorous in exactly the same way that Riemann's proof was made rigorous in previous sections.

Recall that Riemann's proof could also be established by inspecting the Fourier transforms of $\Gamma_\infty(s)|x_\infty|_\infty^s d^\times x_\infty$. A similar approach can work here. If we (very formally!) apply the Poisson summation formula (1.35) to the measure $1_J(x)|x|^s\ d^\times x$, one obtains

$$1 = \sum_{a \in k} \int_J e(-ax)|x|^s\ d^\times x. \tag{1.54}$$

Unpacking this summation using (1.51) (and (1.43)), and casually discarding the $a = 0$ term, we formally conclude that

$$1 = \int_{\mathbb{A}^\times} e(-x)|x|^s\ d^\times x. \tag{1.55}$$

Rescaling this, we formally conclude that the Fourier transform of $|x|^s\ d^\times x$ is $|x|^{1-s}\ d^\times x$. Inserting this into (1.50), the functional equation (1.28) formally follows from *Parseval's theorem*; alternatively, one can derive it by multiplying together all the local facts that the Fourier transform of $\Gamma_\nu(s)|x_\nu|_\nu^s\ d^\times x_\nu$ in k_ν is $\Gamma_\nu(1-s)|x_\nu|_\nu^{1-s}\ d^\times x_\nu$. These arguments can be made rigorous using the theory of distributions (and a lot of care), but we will not do so here.

Notes. This article first appeared at `terrytao.wordpress.com/2008/07/27`. Thanks to Chandan Singh Dalawat for corrections.

Richard Borcherds recalled Andre Weil's characterisation of the gamma factors as the unique constant (up to some standard normalisations) that one could place in front of the obvious homogeneous distributions $|x|^s$ to make the resultant distribution holomorphic for *all* complex s; this uniqueness can then be used to easily derive the functional equation. In higher rank groups, the analogue of Weils description of the gamma factor gives the local L-factor of an automorphic form.

There was some discussion as to when special values (or residues) of zeta-type functions could be recovered from the same sort of adelic analysis discussed here; for instance, the class number formula could be obtained by these methods, but many other special values seem to require deeper tools to establish.

1.9. The divisor bound

Given a positive integer n, let $d(n)$ denote the *number of divisors* of n (including 1 and n), thus for instance $d(6) = 4$, and more generally, if n has a prime factorisation

$$n = p_1^{a_1} \dots p_k^{a_k} \tag{1.56}$$

then (by the fundamental theorem of arithmetic)

$$d(n) = (a_1 + 1) \dots (a_k + 1). \tag{1.57}$$

Clearly, $d(n) \leq n$. The *divisor bound* asserts that, as n gets large, one can improve this trivial bound to

$$d(n) \leq C_\varepsilon n^\varepsilon \tag{1.58}$$

for any $\varepsilon > 0$, where C_ε depends only on ε; equivalently, in asymptotic notation, one has $d(n) = n^{o(1)}$. In fact one has a more precise bound

$$d(n) \leq n^{O(1/\log\log n)} = \exp\left(O\left(\frac{\log n}{\log\log n}\right)\right). \tag{1.59}$$

The divisor bound is useful in many applications in number theory, harmonic analysis, and even PDE (on periodic domains); it asserts that for any large number n, only a "logarithmically small" set of numbers less than n will actually divide n exactly, even in the worst-case[34] scenario when n is *smooth* (i.e. has many small prime factors).

The divisor bound is elementary to prove (and not particularly difficult), and I was asked about it recently, so I thought I would provide the proof here, as it serves as a case study in how to establish worst-case estimates in elementary multiplicative number theory.

---

[34] The average value of $d(n)$ is much smaller, being about $\log n$ on the average, as can be seen easily from the *double counting identity*

$$\sum_{n \leq N} d(n) = \#\{(m,l) \in \mathbf{N} \times \mathbf{N} : ml \leq N\} = \sum_{m=1}^{N} \lfloor \frac{N}{m} \rfloor \sim N \log N,$$

or from the heuristic that a randomly chosen number m less than n has a probability about $1/m$ of dividing n, and $\sum_{m<n} \frac{1}{m} \sim \log n$. However, (1.59) is the correct "worst case" bound, as I discuss below.

1.9.1. Proof of (1.58)**.** Let's first prove the weak form of the divisor bound (1.58), which is already good enough for many applications (because a loss of $n^{o(1)}$ is often negligible if, say, the final goal is to extract some polynomial factor in n in one's eventual estimates). By rearranging a bit, our task is to show that for any $\varepsilon > 0$, the expression

$$\frac{d(n)}{n^\varepsilon} \tag{1.60}$$

is bounded uniformly in n by some constant depending on ε. Using (1.56) and (1.57), we can express (1.60) as a product

$$\prod_{j=1}^{k} \frac{a_j+1}{p_j^{\varepsilon a_j}} \tag{1.61}$$

where each term involves a different prime p_j, and the a_j are at least 1. We have thus "localised" the problem to studying the effect of each individual prime independently. (We are able to do this because $d(n)$ is a multiplicative function.)

Let's fix a prime p_j and look at a single term $\frac{a_j+1}{p_j^{\varepsilon a_j}}$. The numerator is linear in a_j, while the denominator is exponential. Thus, as per Malthus, we expect the denominator to dominate, at least when a_j is large. But, because of the ε, the numerator might be able to exceed the denominator when a_j is small - but only if p_j is also small.

Following these heuristics, we now divide into cases. Suppose that p_j is large, say $p_j \geq \exp(1/\varepsilon)$. Then $p_j^{\varepsilon a_j} \geq \exp(a_j) \geq 1 + a_j$ (by Taylor expansion), and so the contribution of p_j to the product (1.61) is at most 1. So all the large primes give a net contribution of at most 1 here.

What about the small primes, in which $p_j < \exp(1/\varepsilon)$? Well, by Malthus, we know that the sequence $\frac{a+1}{p_j^{\varepsilon a}}$ goes to zero as $a \to \infty$. Since convergent sequences are bounded, we therefore have some bound of the form $\frac{a_j+1}{p_j^{\varepsilon a_j}} \leq C_{p_j,\varepsilon}$ for some $C_{p_j,\varepsilon}$ depending only on p_j and ε, but not on a_j. So, each small prime gives a bounded contribution to (1.61) (uniformly in n). But the number of small primes is itself bounded (uniformly in n). Thus the total product in (1.61) is also bounded uniformly in n, and the claim follows.

1.9.2. Proof of (1.59)**.** One can refine the above analysis to get a more explicit value of C_ε, which will let us get (1.59), as follows.

Again consider the product (1.61) for some $\varepsilon > 0$. As discussed previously, each prime larger than $\exp(1/\varepsilon)$ gives a contribution of at most 1. What about the small primes? Here we can estimate the denominator from below by Taylor expansion:

$$p_j^{\varepsilon a_j} = \exp(\varepsilon a_j \log p_j) \geq 1 + \varepsilon a_j \log p_j$$

and hence

$$\frac{a_j + 1}{p_j^{\varepsilon a_j}} \leq \frac{a_j + 1}{1 + \varepsilon a_j \log p_j} \ll \frac{1}{\varepsilon \log p_j}$$

(the point here being that our bound is uniform in a_j). One can of course use undergraduate calculus to try to sharpen the bound here, but it turns out not to improve by too much, basically because the Taylor approximation $\exp(x) \approx 1 + x$ is quite accurate when x is small, which is the important case here.

Anyway, inserting this bound into (1.61), we see that (1.61) is in fact bounded by

$$\prod_{p < \exp(1/\varepsilon)} O\left(\frac{1}{\varepsilon \log p}\right).$$

Now let's be very crude[35] and bound $\log p$ from below by $\log 2$, and bound the number of primes less than $\exp(1/\varepsilon)$ by $\exp(1/\varepsilon)$. We thus conclude that

$$\leq O\left(\frac{1}{\varepsilon}\right)^{\exp(1/\varepsilon)} = \exp(\exp(O(1/\varepsilon))); \tag{1.62}$$

unwinding what this means for (1.58), we obtain

$$d(n) \leq \exp(\exp(O(1/\varepsilon)))n^\varepsilon$$

for all $n \geq 1$ and $\varepsilon > 0$. If we now set $\varepsilon = C/\log\log n$ for a sufficiently large C, then the second term of the RHS dominates the first (as can be seen by taking logarithms), and the claim (1.59) follows.

[35] One can of course be more efficient about this, but again it turns out not to improve the final bounds too much. A general principle is that when one estimating an expression such as A^B, or more generally the product of B terms, each of size about A, then it is far more important to get a good bound for B than to get a good bound for A, except in those cases when A is very close to 1.

The above argument also suggests the counterexample that will demonstrate that (1.59) is basically sharp. Pick $\varepsilon > 0$, and let n be the product of all the primes up to $\exp(1/\varepsilon)$. The prime number theorem[36] tells us that $\log n \sim \exp(1/\varepsilon)$. On the other hand, the prime number theorem also tells us that the number of primes dividing n is $\sim \varepsilon \exp(1/\varepsilon)$, so by (1.57), $\log d(n) \sim \varepsilon \exp(1/\varepsilon)$. Using these numbers we see that (1.59) is tight up to constants.

1.9.3. Why is the divisor bound useful? One principal application of the divisor bound (and some generalisations of that bound) is to control the number of solutions to a Diophantine equation. For instance, (1.58) immediately implies that for any fixed positive n, the number of solutions to the equation

$$xy = n$$

with x, y integer[37], is only $n^{o(1)}$ at most. This can be leveraged to some other Diophantine equations by high-school algebra. For instance, thanks to the identity $x^2 - y^2 = (x+y)(x-y)$, we conclude that the number of integer solutions to

$$x^2 - y^2 = n$$

is also at most $n^{o(1)}$; similarly, the identity $x^3 + y^3 = (x+y)(x^2 - xy + y^2)$ implies[38] that the number of integer solutions to

$$x^3 + y^3 = n$$

is at most $n^{o(1)}$.

Now consider the number of solutions to the equation

$$x^2 + y^2 = n.$$

In this case, $x^2 + y^2$ does not split over the rationals $\mathbf{Q}$, and so one cannot directly exploit the divisor bound for the rational integers $\mathbf{Z}$. However, we can factor $x^2 + y^2 = (x + iy)(x - iy)$ over the Gaussian rationals $\mathbf{Q}[\sqrt{-1}]$. Happily, the Gaussian integers $\mathbf{Z}[\sqrt{-1}]$ enjoy

[36] If one does not care about the constants, then one does not need the full strength of the prime number theorem to show that (1.59) is sharp; the more elementary bounds of Chebyshev, that say that the number of primes up to N is comparable to $N/\log N$ up to constants, would suffice.

[37] For x and y real, the number of solutions is of course infinite.

[38] Note from Bezout's theorem (or direct calculation) that $x+y$ and $x^2 - xy + y^2$ determine x, y up to at most a finite ambiguity.

essentially the same divisor bound as the rational integers $\mathbf{Z}$; the Gaussian integers have unique factorisation, but perhaps more importantly they only have a finite set of units ($\{-1, +1, -i, +i\}$ to be precise). Because of this, one can easily check that $x^2 + y^2 = n$ also has at most $n^{o(1)}$ solutions.

One can similarly exploit the divisor bound on other number fields; for instance the divisor bound for $\mathbf{Z}[\sqrt{-3}]$ lets one count solutions to $x^2 + xy + y^2 = n$ or $x^2 - xy + y^2 = n$. On the other hand, not all number fields have the divisor bound. For instance, $\mathbf{Z}[\sqrt{2}]$ has an infinite number of units, which means that the number of solutions to *Pell's equation*

$$x^2 - 2y^2 = 1$$

is infinite.

Another application of the divisor bound comes up in *sieve theory*. Here, one is often dealing with functions of the form $\nu(n) := \sum_{d|n} a_d$, where the sieve weights a_d typically have size $O(n^{o(1)})$, and the sum is over all d that divide n. The divisor bound (1.58) then implies that the sieve function $\nu(n)$ also has size $O(n^{o(1)})$. This bound is too crude to deal with the most delicate components of a sieve theory estimate, but is often very useful for dealing with error terms (especially those which have gained a factor of n^{-c} relative to the main terms for some $c > 0$).

Notes. This article first appeared at `terrytao.wordpress.com/2008/09/23`.

Marius Overholt and Emmanuel Kowalski noted that the implied constant in (1.59) is known to be $\log 2 + o(1)$, a classical result of Wigert from 1906 (with a simpler proof given by Ramanujan in 1914). Explicit constants are also known for the moments of the divisor function, for instance

$$\frac{1}{x} \sum_{n \leq x} d(n)^2 = (\frac{1}{\pi^2} + o(1)) \log^3 x$$

and more generally

$$\frac{1}{x} \sum_{n \leq x} d(n)^k = (c_k + o(1)) \log^{2^k - 1} x$$

for any fixed $k \geq 1$ and some explicit constant c_k.

Ben Green noted that there appeared to be no bound approaching the strength of the divisor bound if one perturbs the integers in a manner that destroys the number-theoretic structure. For instance, for a fixed shift θ, it does not seem possible to obtain a bound of $O(n^{o(1)})$ to the number of solutions to the equation $(a+\theta)(b+\theta) = n + O(1)$ for integer a, b.

1.10. What is a gauge?

Gauge theory is a term which has connotations of being a fearsomely complicated part of mathematics - for instance, playing an important role in quantum field theory, general relativity, geometric PDE, and so forth. But the underlying concept is really quite simple: a *gauge* is nothing more than a "coordinate system" that varies depending on one's "location" with respect to some "base space" or "parameter space", a *gauge transform* is a change of coordinates applied to each such location, and a *gauge theory* is a model for some physical or mathematical system to which gauge transforms can be applied (and is typically *gauge invariant*, in that all physically meaningful quantities are left unchanged (or transform naturally) under gauge transformations). By *fixing* a gauge (thus *breaking*[39] or *spending* the gauge symmetry), the model becomes something easier to analyse mathematically, such as a system of partial differential equations (in classical gauge theories) or a perturbative quantum field theory (in quantum gauge theories), though the tractability of the resulting problem can be heavily dependent on the choice of gauge that one fixed. Deciding exactly how to fix a gauge (or whether one should spend the gauge symmetry at all) is a key question in the analysis of gauge theories, and one that often requires the input of geometric ideas and intuition into that analysis.

I was asked recently to explain what a gauge theory was, and so I will try to do so in this post. For simplicity, I will focus exclusively on classical gauge theories; quantum gauge theories are the quantization of classical gauge theories and have their own set of conceptual

[39]This mathematical notion of breaking a symmetry is a little different from the notion of symmetry breaking in physics, which usually refers to a situation in which a symmetric state is perturbed into a non-symmetric one.

difficulties (coming from quantum field theory) that I will not discuss here. While gauge theories originated from physics, I will not discuss the physical significance of these theories much here, instead focusing just on their mathematical aspects. My discussion will be informal, as I want to try to convey the geometric intuition rather than the rigorous formalism (which can, of course, be found in any graduate text on differential geometry).

1.10.1. Coordinate systems. Before I discuss gauges, I first review the more familiar concept of a *coordinate system*, which is basically the special case of a gauge when the base space (or parameter space) is trivial.

Classical mathematics, such as practised by the ancient Greeks, could be loosely divided into two disciplines, *geometry* and *number theory*, where I use the latter term very broadly, to encompass all sorts of mathematics dealing with any sort of number. The two disciplines are unified by the concept of a *coordinate system*, which allows one to convert geometric objects to numeric ones or vice versa. The most well known example of a coordinate system is the *Cartesian coordinate system* for the plane (or more generally for a Euclidean space), but this is just one example of many such systems. For instance:

(1) One can convert a length (of, say, an interval) into an (unsigned) real number, or vice versa, once one fixes a unit of length (e.g. the metre or the foot). In this case, the coordinate system is specified by the choice of length unit.

(2) One can convert a displacement along a line into a (signed) real number, or vice versa, once one fixes a unit of length *and* an orientation along that line. In this case, the coordinate system is specified by the length unit together with the choice of orientation. Alternatively, one can replace the unit of length and the orientation by a unit displacement vector e along the line.

(3) One can convert a position (i.e. a point) on a line into a real number, or vice versa, once one fixes a unit of length, an orientation along the line, *and* an origin on that line.

Equivalently, one can pick an origin O and a unit displacement vector e. This coordinate system essentially identifies the original line with the standard real line $\mathbf{R}$.

(4) One can generalise these systems to higher dimensions. For instance, one can convert a displacement along a plane into a vector in $\mathbf{R}^2$, or vice versa, once one fixes two linearly independent displacement vectors e_1, e_2 (i.e. a basis) to span that plane; the Cartesian coordinate system is just one special case of this general scheme. Similarly, one can convert a position on a plane to a vector in $\mathbf{R}^2$ once one picks a basis e_1, e_2 for that plane as well as an origin O, thus identifying that plane with the standard Euclidean plane $\mathbf{R}^2$. (To put it another way, units of measurement are nothing more than one-dimensional (i.e. scalar) coordinate systems.)

(5) To convert an angle in a plane to a signed number (modulo multiples of 2π), or vice versa, one needs to pick an orientation on the plane (e.g. to decide that anti-clockwise angles are positive).

(6) To convert a *direction* in a plane to a signed number (again modulo multiples of 2π), or vice versa, one needs to pick an orientation on the plane, as well as a reference direction (e.g. *true* or *magnetic north* is often used in the case of ocean navigation).

(7) Similarly, to convert a position on a circle to a number (modulo multiples of 2π), or vice versa, one needs to pick an orientation on that circle, together with an origin on that circle. Such a coordinate system then equates the original circle to the standard unit circle $S^1 := \{z \in \mathbf{C} : |z| = 1\}$ (with the standard origin $+1$ and the standard anticlockwise orientation $\circlearrowleft$).

(8) To convert a position on a two-dimensional sphere (e.g. the surface of the Earth, as a first approximation) to a point on the standard unit sphere $S^2 := \{(x, y, z) \in \mathbf{R}^3 : x^2 + y^2 + z^2\}$, one can pick an orientation on that sphere, an "origin" (or "north pole") for that sphere, and a "*prime meridian*" connecting the north pole to its antipode. Alternatively,

one can view this coordinate system as determining a pair of *Euler angles* ϕ, λ (or a latitude and longitude) to be assigned to every point on one's original sphere.

(9) The above examples were all geometric in nature, but one can also consider "combinatorial" coordinate systems, which allow one to identify combinatorial objects with numerical ones. An extremely familiar example of this is *enumeration*: one can identify a set A of (say) five elements with the numbers 1,2,3,4,5 simply by choosing an enumeration $a_1, a_2, \ldots, a_5$ of the set A. One can similarly enumerate other combinatorial objects (e.g. graphs, relations, trees, partial orders, etc.), and indeed this is done all the time in combinatorics. Similarly for algebraic objects, such as cosets of a subgroup H (or more generally, *torsors* of a group G); one can identify such a coset with H itself by designating an element of that coset to be the "identity" or "origin".

More generally, a coordinate system[40] Φ can be viewed as an isomorphism $\Phi : A \to G$ between a given geometric (or combinatorial) object A in some class (e.g. a circle), and a standard object G in that class (e.g. the standard unit circle).

Coordinate systems identify geometric or combinatorial objects with numerical (or standard) ones, but in many cases, there is no natural (or *canonical*) choice of this identification; instead, one may be faced with a variety of coordinate systems, all equally valid. One can of course just fix one such system once and for all, in which case there is no real harm in thinking of the geometric and numeric objects as being equivalent. If however one plans to change from one system to the next (or to avoid using such systems altogether), then it becomes important to carefully distinguish these two types of objects, to avoid confusion. For instance, if an interval AB is measured to have a length of 3 yards, then it is OK to write $|AB| = 3$ (identifying the geometric concept of length with the numeric concept

[40]To be pedantic, this is what a *global* coordinate system is; a *local* coordinate system, such as the coordinate charts on a manifold, is an isomorphism between a local piece of a geometric or combinatorial object in a class, and a local piece of a standard object in that class. I will restrict attention to global coordinate systems for this discussion.

of a positive real number) so long as you plan to stick to having the yard as the unit of length for the rest of one's analysis. But if one was also planning to use, say, feet, as a unit of length also, then to avoid confusing statements such as "$|AB| = 3$ and $|AB| = 9$", one should specify the coordinate systems explicitly, e.g. "$|AB| = 3$ yards and $|AB| = 9$ feet". Similarly, identifying a point P in a plane with its coordinates (e.g. $P = (4, 3)$) is safe as long as one intends to only use a single coordinate system throughout; but if one intends to change coordinates at some point (or to switch to a coordinate-free perspective) then one should be more careful, e.g. writing $P = 4e_1 + 3e_2$, or even $P = O + 4e_1 + 3e_2$, if the origin O and basis vectors e_1, e_2 of one's coordinate systems might be subject to future change.

As mentioned above, it is possible to in many cases to dispense with coordinates altogether. For instance, one can view the length $|AB|$ of a line segment AB not as a number (which requires one to select a unit of length), but more abstractly as the equivalence class of all line segments CD that are congruent to AB. With this perspective, $|AB|$ no longer lies in the standard semigroup $\mathbf{R}^+$, but in a more abstract semigroup $\mathcal{L}$ (the space of line segments quotiented by congruence), with addition now defined geometrically (by concatenation of intervals) rather than numerically. A unit of length can now be viewed as just one of many different isomorphisms $\Phi : \mathcal{L} \to \mathbf{R}^+$ between $\mathcal{L}$ and $\mathbf{R}^+$, but one can abandon the use of such units and just work with $\mathcal{L}$ directly. Many statements in Euclidean geometry involving length can be phrased in this manner. For instance, if B lies in AC, then the statement $|AC| = |AB| + |BC|$ can be stated in $\mathcal{L}$, and does not require any units to convert $\mathcal{L}$ to $\mathcal{R}^+$; with a bit more work, one can also make sense of such statements as $|AC|^2 = |AB|^2 + |BC|^2$ for a right-angled triangle ABC (i.e Pythagoras' theorem) while avoiding units[41], by defining a symmetric bilinear product operation $\times : \mathcal{L} \times \mathcal{L} \to \mathcal{A}$ from the abstract semigroup $\mathcal{L}$ of lengths to the abstract semigroup $\mathcal{A}$ of areas.

The above abstract *coordinate-free perspective* is equivalent to a more concrete *coordinate-invariant perspective*, in which we do allow

[41]Indeed, this is basically how the ancient Greeks, who did not quite possess the modern real number system $\mathbf{R}$, viewed geometry, though of course without the assistance of such modern terminology as "semigroup" or "bilinear".

the use of coordinates to convert all geometric quantities to numeric ones, but insist that every statement that we write down is invariant under changes of coordinates. For instance, if we shrink our chosen unit of length by a factor $\lambda > 0$, then the numerical length of every interval increases by a factor of λ, e.g. $|AB| \mapsto \lambda|AB|$. The coordinate-invariant approach to length measurement then treats lengths such as $|AB|$ as numbers, but requires[42] all statements involving such lengths to be invariant under the above scaling symmetry. For instance, a statement such as $|AC|^2 = |AB|^2 + |BC|^2$ is legitimate under this perspective, but a statement such as $|AB| = |BC|^2$ or $|AB| = 3$ is not. One can retain this coordinate-invariance symmetry throughout one's arguments; or one can, at some point, choose to *spend* (or *break*) this coordinate invariance by selecting (or *fixing*) the coordinate system (which, in this case, means selecting a unit length). The advantage in spending such a symmetry is that one can often normalise one or more quantities to equal a particularly nice value; for instance, if a length $|AB|$ is appearing everywhere in one's arguments, and one has carefully retained coordinate-invariance up until some key point, then it can be convenient to spend this invariance[43] to normalise $|AB|$ to equal 1. Conversely, if one has already spent the coordinate invariance, one can often buy it back by converting all the facts, hypotheses, and desired conclusions one currently possesses in the situation back to a coordinate-invariant formulation. Thus one could imagine performing one normalisation to do one set of calculations, then undoing that normalisation to return to a coordinate-free perspective, doing some coordinate-free manipulations, and then performing a different normalisation to work on another part of the problem, and so forth[44].

[42] In other words, co-ordinate invariance here is the same thing as being dimensionally consistent. Indeed, dimensional analysis is nothing more than the analysis of the scaling symmetries in one's coordinate systems.

[43] In this case, one only has a one-dimensional family of symmetries, and so can only normalise one quantity at a time; but when one's symmetry group is larger, one can often normalise many more quantities at once; as a rule of thumb, one can normalise one quantity for each degree of freedom in the symmetry group.

[44] For instance, in Euclidean geometry problems, it is often convenient to temporarily assign one key point to be the origin (thus spending translation invariance symmetry), then another, then switch back to a translation-invariant perspective, and so forth. As long as one is correctly accounting for what symmetries are being spent and bought at any given time, this can be a very powerful way of simplifying one's calculations.

Given a coordinate system $\Phi : A \to G$ that identifies some geometric object A with a standard object G, and some isomorphism $\Psi : G \to G$ of that standard object, we can obtain a new coordinate system $\Psi \circ \Phi : A \to G$ of A by composing the two isomorphisms[45]. Conversely, every other coordinate system $\Phi' : A \to G$ of A arises in this manner. Thus, the space of coordinate systems on A is (non-canonically) identifiable with the isomorphism group $\mathrm{Isom}(G)$ of G. This isomorphism group is called the *structure group* (or *gauge group*) of the class of geometric objects. For example, the structure group for lengths is $\mathbf{R}^+$; the structure group for angles is $\mathbf{Z}/2\mathbf{Z}$; the structure group for lines is the *affine group* $\mathrm{Aff}(\mathbf{R}) = \mathbf{R} \rtimes \mathbf{R}$; the structure group for n-dimensional Euclidean geometry is the *Euclidean group* $E(n) = O(n) \rtimes \mathbf{R}^n$; the structure group for (oriented) 2-spheres is the (special) *orthogonal group* $SO(3)$; and so forth[46].

1.10.2. Gauges. In our discussion of coordinate systems, we focused on a single geometric (or combinatorial) object A: a single line, a single circle, a single set, etc. We then used a single coordinate system to identify that object with a standard representative of such an object.

Now let us consider the more general situation in which one has a *family* (or *fibre bundle*) $(A_x)_{x \in X}$ of geometric (or combinatorial) objects (or *fibres*) A_x: a family of lines (i.e. a line bundle), a family of circles (i.e. a circle bundle), a family of sets, etc. This family is parameterised by some *parameter set* or *base point* x, which ranges in some *parameter space* or *base space* X. In many cases one also requires some topological or differentiable compatibility between the various fibres; for instance, continuous (or smooth) variations of the base point should lead to continuous (or smooth) variations in the fibre. For sake of discussion, however, let us gloss over these compatibility conditions.

In many cases, each individual fibre A_x in a bundle $(A_x)_{x \in X}$, being a geometric object of a certain class, can be identified with a

[45] I will be vague on what "isomorphism" means; one can formalise the concept using the language of *category theory*.

[46] Indeed, one can basically describe each of the classical geometries (Euclidean, affine, projective, spherical, hyperbolic, Minkowski, etc.) as a homogeneous space for its structure group, as per the *Erlangen program*.

standard object G in that class, by means of a separate coordinate system $\Phi_x : A_x \to G$ for each base point x. The entire collection $\Phi = (\Phi_x)_{x \in X}$ is then referred to as a (global) *gauge* or *trivialisation* for this bundle (provided that it is compatible with whatever topological or differentiable structures one has placed on the bundle, but never mind that for now). Equivalently, a gauge[47] is a *bundle isomorphism* Φ from the original bundle $(A_x)_{x \in X}$ to the *trivial bundle* $(G)_{x \in X}$, in which every fibre is the standard geometric object G.

Let's give three concrete examples of bundles and gauges; one from differential geometry, one from dynamical systems, and one from combinatorics.

Example 1.10.1 (The circle bundle of the sphere)**.** Recall from the previous section that the space of directions in a plane (which can be viewed as the circle of unit vectors) can be identified with the standard circle S^1 after picking an orientation and a reference direction. Now let us work not on the plane, but on a sphere, and specifically, on the surface X of the earth. At each point x on this surface, there is a circle S_x of directions that one can travel along the sphere from x; the collection $SX := (S_x)_{x \in X}$ of all such circles is then a circle bundle with base space X (known as *the* circle bundle; it could also be viewed as the sphere bundle, cosphere bundle, or orthonormal frame bundle of X). The structure group of this bundle is the circle group $U(1) \equiv S^1$ if one preserves orientation, or the *semi-direct product* $S^1 \rtimes \mathbf{Z}/2\mathbf{Z}$ otherwise.

Now suppose, at every point x on the earth X, the wind is blowing[48] in some direction $w_x \in S_x$. Thus wind direction can be thought of as a collection $w = (w_x)_{x \in X}$ of representatives from the fibres of the fibre bundle $(S_x)_{x \in X}$; such a collection is known as a *section* of the fibre bundle (it is to bundles as the concept of a graph $\{(x, f(x)) : x \in X\} \subset X \times G$ of a function $f : X \to G$ is to the trivial bundle $(G)_{x \in X}$).

At present, this section has not been represented in terms of numbers; instead, the wind direction $w(w_x)_{x \in X}$ is a collection of points on

[47] There are also *local* gauges, which only trivialise a portion of the bundle, but let's ignore this distinction for now.

[48] This is not actually possible globally, thanks to the *hairy ball theorem*, but let's ignore this technicality for now.

various different circles in the circle bundle SX. But one can convert this section w into a collection of numbers (and more specifically, a function $u : X \to S^1$ from X to S^1) by choosing a gauge for this circle bundle - in other words, by selecting an orientation ϵ_x and a reference direction N_x for each point x on the surface of the Earth X. For instance, one can pick the anticlockwise orientation $\circlearrowleft$ and true north for every point x (ignore for now the problem that this is not defined at the north and south poles, and so is merely a local gauge rather than a global one), and then each wind direction w_x can now be identified with a unit complex number $u(x) \in S^1$ (e.g. $e^{i\pi/4}$ if the wind is blowing in the northwest direction at x). Now that one has a numerical function u to play with, rather than a geometric object w, one can now use analytical tools (e.g. differentiation, integration, Fourier transforms, etc.) to analyse the wind direction if one desires. But one should be aware that this function reflects the choice of gauge as well as the original object of study. If one changes the gauge (e.g. by using magnetic north instead of true north), then the function u changes, even though the wind direction w is still the same. If one does not want to spend the $U(1)$ gauge symmetry, one would have to take care that all operations one performs on these functions are gauge-invariant; unfortunately, this restrictive requirement eliminates wide swathes of analytic tools (in particular, integration and the Fourier transform) and so one is often forced to break the gauge symmetry in order to use analysis. The challenge is then to select the gauge that maximises the effectiveness of analytic methods.

Example 1.10.2 (Circle extensions of a dynamical system)**.** Recall (see Section 2.1) that a dynamical system is a pair $X = (X, T)$, where X is a space and $T : X \to X$ is an invertible map. Given such a system, and given a *cocycle* $\rho : X \to S^1$ (which, in this context, is simply a function from X to the unit circle), we can define the *skew product* $X \times_\rho S^1$ of X and the unit circle S^1, twisted by the cocycle ρ, to be the Cartesian product $X \times S^1 := \{(x, u) : x \in X, u \in S^1\}$ with the shift $\tilde{T} : (x, u) \mapsto (Tx, \rho(x)u)$; this is easily seen to be another dynamical system[49]. Observe that there is a free

[49]If one wishes to have a topological or measure-theoretic dynamical system, then ρ will have to be continuous or measurable here, see Section 2.1; but let us ignore such issues for this discussion.

action $(S_v : (x,u) \mapsto (x,vu))_{v \in S^1}$ of the circle group S^1 on the skew product $X \times_\rho S^1$ that commutes with the shift $\tilde{T}$; the quotient space $(X \times_\rho S^1)/S^1$ of this action is isomorphic to X, thus leading to a *factor map* $\pi : X \times_\rho S^1 \to X$, which is of course just the projection map $\pi : (x,u) \mapsto x$. (An example is provided by the *skew shift system*, described in Section 2.2.)

Conversely, suppose that one had a dynamical system $\tilde{X} = (\tilde{X}, \tilde{T})$ which had a free S^1 action $(S_v : \tilde{X} \to \tilde{X})_{v \in S^1}$ commuting with the shift $\tilde{T}$. If we set $X := \tilde{X}/S^1$ to be the quotient space, we thus have a factor map $\pi : \tilde{X} \to X$, whose level sets $\pi^{-1}(\{x\})$ are all isomorphic to the circle S^1; we call $\tilde{X}$ a *circle extension* of the dynamical system X. We can thus view $\tilde{X}$ as a *circle bundle* $(\pi^{-1}(\{x\}))_{x \in X}$ with base space X, thus the level sets $\pi^{-1}(\{x\})$ are now the fibres of the bundle, and the structure group is S^1. If one picks a *gauge* for this bundle, by choosing a reference point $p_x \in \pi^{-1}(\{x\})$ in the fibre for each base point x (thus in this context a gauge is the same thing as a *section* $p = (p_x)_{x \in X}$; this is basically because this bundle is a *principal bundle*), then one can identify $\tilde{X}$ with a skew product $X \times_\rho S^1$ by identifying the point $S_v p_x \in \tilde{X}$ with the point $(x,v) \in X \times_\rho S^1$ for all $x \in X, v \in S^1$, and letting ρ be the cocycle defined by the formula

$$S_{\rho(x)} p_{Tx} = \tilde{T} p_x.$$

One can check that this is indeed an isomorphism of dynamical systems; if all the various objects here are continuous (resp. measurable), then one also has an isomorphism of topological dynamical systems (resp. measure-preserving systems). Thus we see that gauges allow us to write circle extensions as skew products. However, more than one gauge is available for any given circle extension; two gauges $(p_x)_{x \in X}$, $(p'_x)_{x \in X}$ will give rise to two skew products $X \times_\rho S^1$, $X \times_{\rho'} S^1$ which are isomorphic but not identical. Indeed, if we let $v : X \to S^1$ be a rotation map that sends p_x to p'_x, thus $p'_x = S_{v(x)} p_x$, then we see that the two cocycles ρ' and ρ are related by the formula

$$\rho'(x) = v(Tx)^{-1} \rho(x) v(x). \tag{1.63}$$

Two cocycles that obey the above relation are called *cohomologous*; their skew products are isomorphic to each other. An important general question in dynamical systems is to understand when two

given cocycles are in fact cohomologous, for instance by introducing non-trivial cohomological invariants for such cocycles.

As an example of a circle extension, consider the sphere $X = S^2$ from Example 1.10.1, with a rotation shift T given by, say, rotating anti-clockwise by some given angle α around the axis connecting the north and south poles. This rotation also induces a rotation on the circle bundle $\tilde{X} := SX$, thus giving a circle extension of the original system (X, T). One can then use a gauge to write this system as a skew product. For instance, if one selects the gauge that chooses p_x to be the true north direction at each point x (ignoring for now the fact that this is not defined at the two poles), then this system becomes the ordinary product $X \times_0 S^1$ of the original system X with the circle S^1, with the cocycle being the trivial cocycle 0. If we were however to use a different gauge, e.g. magnetic north instead of true north, one would obtain a different skew-product $X \times_{\rho'} S^1$, where ρ' is some cocycle which is cohomologous[50] to the trivial cocycle (except at the poles).

There was nothing terribly special about circles in this example; one can also define group extensions, or more generally homogeneous space extensions, of dynamical systems, and have a similar theory, although one has to take a little care with the order of operations when the structure group is non-abelian; see e.g. Section 2.6.

Example 1.10.3 (Orienting an undirected graph)**.** The language of gauge theory is not often used in combinatorics, but nevertheless combinatorics does provide some simple discrete examples of bundles and gauges which can be useful in getting an intuitive grasp of the concept. Consider for instance an *undirected graph* $G = (V, E)$ of vertices and edges. I will let $X = E$ denote the space of edges (not the space of vertices)!. Every edge $e \in X$ can be oriented (or directed) in two different ways; let A_e be the pair of directed edges of e arising in this manner. Then $(A_e)_{e \in X}$ is a fibre bundle with base space X and with each fibre isomorphic (in the category of sets) to the standard two-element set $\{-1, +1\}$, with structure group $\mathbf{Z}/2\mathbf{Z}$.

[50] A cocycle which is globally cohomologous to the trivial cocycle is known as a *coboundary*. Not every cocycle is a coboundary, especially once one imposes topological or measure-theoretic structure, thanks to the presence of various topological or measure-theoretic invariants, such as degree; see Section 1.21 for further discussion.

A priori, there is no reason to prefer one orientation of an edge e over another, and so there is no canonical way to identify each fibre A_e with the standard set $\{-1, +1\}$. Nevertheless, we can go ahead and arbitrary select a gauge for X by *orienting* the graph G. This orientation assigns an oriented edge $\vec{e} \in A_e$ to each edge $e \in X$, thus creating a gauge (or section) $(\vec{e})_{e \in X}$ of the bundle $(A_e)_{e \in X}$. Once one selects such a gauge, we can now identify the fibre bundle $(A_e)_{e \in X}$ with the trivial bundle $X \times \{-1, +1\}$ by identifying the preferred oriented edge $\vec{e}$ of each unoriented edge $e \in X$ with $(e, +1)$, and the other oriented edge with $(e, -1)$. In particular, any other orientation of the graph G can be expressed relative to this reference orientation as a function $f : X \to \{-1, +1\}$, which measures when the two orientations agree or disagree with each other.

Recall that every isomorphism $\Psi \in \mathrm{Isom}(G)$ of a standard geometric object G allowed one to transform a coordinate system $\Phi : A \to G$ on a geometric object A to another coordinate system $\Psi \circ \Phi : A \to G$. We can generalise this observation to gauges: every family $\Psi = (\Psi_x)_{x \in X}$ of isomorphisms on G allows one to transform a gauge $(\Phi_x)_{x \in X}$ to another gauge $(\Psi_x \circ \Phi_x)_{x \in X}$ (again assuming that Ψ respects whatever topological or differentiable structure is present). Such a collection Ψ is known as a *gauge transformation*. For instance, in Example 1.10.1, one could rotate the reference direction N_x at each point $x \in X$ anti-clockwise by some angle $\theta(x)$; this would cause the function $u(x)$ to rotate to $u(x)e^{-i\theta(x)}$. In Example 1.10.2, a gauge transformation is just a map $v : X \to S^1$ (which may need to be continuous or measurable, depending on the structures one places on X); it rotates a point $(x, u) \in X \times_\rho S^1$ to $(x, v^{-1}u)$, and it also transforms the cocycle ρ by the formula (1.63). In Example 1.10.3, a gauge transformation would be a map $v : X \to \{-1, +1\}$; it rotates a point $(x, \epsilon) \in X \times \{-1, +1\}$ to $(x, v(x)\epsilon)$.

Gauge transformations transform functions on the base X in many ways, but some things remain gauge-invariant. For instance, in Example 1.10.1, the winding number of a function $u : X \to S^1$ along a closed loop $\gamma \subset X$ would not change under a gauge transformation (as long as no singularities in the gauge are created, moved, or destroyed, and the orientation is not reversed). But such topological

gauge-invariants are not the only gauge invariants of interest; there are important *differential* gauge-invariants which make gauge theory a crucial component of modern differential geometry and geometric PDE. But to describe these, one needs an additional gauge-theoretic concept, namely that of a *connection* on a fibre bundle.

1.10.3. Connections. There are many essentially equivalent ways to introduce the concept of a connection; I will use the formulation based primarily on parallel transport, and on differentiation of sections. To avoid some technical details I will work (somewhat non-rigorously) with infinitesimals[51] such as dx.

In single variable calculus, we learn that if we want to differentiate a function $f : [a, b] \to \mathbf{R}$ at some point x, then we need to compare the value f(x) of f at x with its value $f(x + dx)$ at some infinitesimally close point $x + dx$, take the difference $f(x + dx) - f(x)$, and then divide by dx, taking limits as $dx \to 0$, if one does not like to use infinitesimals:

$$\nabla f(x) := \lim_{dx \to 0} \frac{f(x + dx) - f(x)}{dx}.$$

In several variable calculus, we learn several generalisations of this concept in which the domain and range of f to be multi-dimensional. For instance, if $f : X \to \mathbf{R}^d$ is now a vector-valued function on some multi-dimensional domain (e.g. a manifold) X, and v is a *tangent vector* to X at some point x, we can define the *directional derivative* $\nabla_v f(x)$ of f at x by comparing $f(x + vdt)$ with $f(x)$ for some infinitesimal dt, take the difference[52] $f(x + vdt) - f(x)$, divide by dt, and then take limits as $dt \to 0$:

$$\nabla_v f(x) := \lim_{dt \to 0} \frac{f(x + vdt) - f(x)}{dt}.$$

If f is sufficiently smooth (being continuously differentiable will do), the directional derivative is linear in v, thus for instance $\nabla_{v+v'} f(x) = \nabla_v f(x) + \nabla_{v'} f(x)$. One can also generalise the range of f to other

[51] There are ways to make the use of infinitesimals rigorous, such as non-standard analysis (see Section 1.5 of *Structure and Randomness*), but this will not be the focus of this article.

[52] Strictly speaking, if X is not flat, then $x+vdt$ is only defined up to an ambiguity of o(dt), but let us ignore this minor issue here, as it is not important in the limit.

multi-dimensional domains than $\mathbf{R}^d$; the directional derivative then lives in a tangent space of that domain.

In all of the above examples, though, we were differentiating functions $f : X \to Y$, thus each element $x \in X$ in the base (or domain) gets mapped to an element $f(x)$ in the same range Y. However, in many geometrical situations we would like to differentiate *sections* $f = (f_x)_{x \in X}$ instead of functions, thus f now maps each point $x \in X$ in the base to an element $f_x \in A_x$ of some fibre in a fibre bundle $(A_x)_{x \in X}$. For instance, one might want to know how the wind direction $w = (w_x)_{x \in X}$ changes as one moves x in some direction v; thus computing a directional derivative $\nabla_v w(x)$ of w at x in direction v. One can try to mimic the previous definitions in order to define this directional derivative. For instance, one can move x along v by some infinitesimal amount dt, creating a nearby point $x + vdt$, and then evaluate w at this point to obtain $w(x+vdt)$. But here we hit a snag: we cannot directly compare $w(x+vdt)$ with $w(x)$, because the former lives in the fibre A_{x+vdt} while the latter lives in the fibre A_x.

With a gauge, of course, we can identify all the fibres (and in particular, A_{x+vdt} and A_x) with a common object G, in which case there is no difficulty comparing $w(x+vdt)$ with $w(x)$. But this would lead to a notion of derivative which is not gauge-invariant, known as the *non-covariant* or *ordinary* derivative in physics.

But there is another way to take a derivative, which does not require the full strength of a gauge (which identifies *all* fibres simultaneously together). Indeed, in order to compute a derivative $\nabla_v w(x)$, one only needs to identify (or *connect*) two infinitesimally close fibres together: A_x and A_{x+vdt}. In practice, these two fibres are already "within $O(dt)$ of each other" in some sense, but suppose in fact that we have some means $\Gamma(x \to x + vdt) : A_x \to A_{x+vdt}$ of identifying these two fibres together (possibly up to errors of $o(dt)$). Then, we can pull back $w(x + vdt)$ from A_{x+vdt} to A_x through $\Gamma(x \to x + vdt)$ to define the covariant derivative:

$$\nabla_v w(x) := \lim_{dt \to 0} \frac{\Gamma(x \to x + vdt)^{-1}(w(x + vdt)) - w(x)}{dt}.$$

In order to retain the basic property that $\nabla_v w$ is linear in v, and to allow one to extend the infinitesimal identifications $\Gamma(x \to x + dx)$

to non-infinitesimal identifications, we impose the property that the $\Gamma(x \to x+dx)$ to be approximately transitive in that

$$\Gamma(x+dx \to x+dx+dx') \circ \Gamma(x \to x+dx) \approx \Gamma(x \to x+dx+dx') \tag{1.64}$$

for all x, dx, dx', where the $\approx$ symbol indicates that the error between the two sides is[53] $o(|dx| + |dx'|)$. To oversimplify a little bit, any collection Γ of infinitesimal maps $\Gamma(x \to x+dx)$ obeying this property (and some technical regularity properties) is a *connection*.

Remark 1.10.4. There are many other important ways to view connections, for instance the *Christoffel symbol perspective* that we will discuss a bit later. Another approach is to focus on the differentiation operation ∇_v rather than the identifications $\Gamma(x \to x+dx)$ or $\Gamma(\gamma)$, and in particular on the algebraic properties of this operation, such as linearity in v or *derivation*-type properties (in particular, obeying various variants of the *Leibnitz rule*). This approach is particularly important in algebraic geometry, in which the notion of an infinitesimal or of a path may not always be obviously available, but we will not discuss it here. For the Riemannian geometry perspective, see Section 3.1.

The way we have defined it, a connection is a means of identifying two infinitesimally close fibres A_x, A_{x+dx} of a fibre bundle $(A_x)_{x \in X}$. But, thanks to (1.64), we can also identify two distant fibres A_x, A_y, provided that we have a path $\gamma : [a,b] \to X$ from $x = \gamma(a)$ to $y = \gamma(b)$, by concatenating the infinitesimal identifications by a non-commutative variant of a *Riemann sum*:

$$\Gamma(\gamma) := \lim_{\sup |t_{i+1}-t_i| \to 0} \Gamma(\gamma(t_{n-1}) \to \gamma(t_n)) \circ \ldots \circ \Gamma(\gamma(t_0) \to \gamma(t_1)), \tag{1.65}$$

where $a = t_0 < t_1 < \ldots < t_n = b$ ranges over partitions. This gives us a *parallel transport* map $\Gamma(\gamma) : A_x \to A_y$ identifying A_x with A_y, which in view of its Riemann sum definition, can be viewed as the "integral" of the connection Γ along the curve γ. This map does not depend on how one parametrises the path γ, but it can depend on the choice of path used to travel from x to y.

[53] The precise nature of this error is actually rather important, being essentially the *curvature* of the connection Γ at x in the directions dx, dx'; see Section 3.1 for further discussion.

We illustrate these concepts using several examples, including the three examples introduced earlier.

Example 1.10.5 (Example 1.10.1 continued)**.** The geometry of the sphere X in Example 1.10.1 provides a natural connection on the circle bundle SX, the *Levi-Civita connection* Γ, that lets one transport directions around the sphere in as "parallel" a manner as possible; the precise definition is a little technical (see e.g. Section 3.1 for a brief description). Suppose for instance one starts at some location x on the equator of the earth, and moves to the antipodal point y by a *great semi-circle* γ going through the north pole. The parallel transport $\Gamma(\gamma) : S_x \to S_y$ along this path will map the north direction at x to the *south* direction at y. On the other hand, if we went from x to y by a great semi-circle γ' going along the equator, then the north direction at x would be transported to the *north* direction at y. Given a section u of this circle bundle, the quantity $\nabla_v u(x)$ can be interpreted as the rate at which u rotates as one travels from x with velocity v.

Example 1.10.6 (Example 1.10.2 continued)**.** In Example 1.10.2, we change the notion of "infinitesimally close" by declaring x and Tx to be infinitesimally close for any x in the base space X (and more generally, x and $T^n x$ are non-infinitesimally close for any positive integer n, being connected by the path $x \to Tx \to \ldots \to T^n x$, and similarly for negative n). A cocycle $\rho : X \to S^1$ can then be viewed as defining a connection on the skew product $X \times_\rho S^1$, by setting $\Gamma(x \mapsto Tx) = \rho(x)$ (and also $\Gamma(x \to x) = 1$ and $\Gamma(Tx \to x) = \rho(x)^{-1}$ to ensure compatibility with (1.64); to avoid notational ambiguities let us assume for sake of discussion that $x, Tx, T^{-1}x$ are always distinct from each other). The non-infinitesimal connections $\rho_n(x) := \Gamma(x \to Tx \to \ldots \to T^n x)$ are then given by the formula $\rho_n(x) = \rho(x)\rho(Tx)\ldots\rho(T^{n-1}x)$ for positive n (with a similar formula for negative n). Note that these iterated cocycles ρ_n also describe the iterations of the shift $\tilde{T} : (x, u) \mapsto (Tx, \rho(x)u)$, indeed $\tilde{T}^n(x, u) = (T^n x, \rho_n(x)u)$.

Example 1.10.7 (Example 1.10.3 continued)**.** In Example 1.10.3, we declare two edges e, e' in X to be "infinitesimally close" if they are adjacent. Then there is a natural notion of parallel transport

on the bundle $(A_e)_{e\in X}$; given two adjacent edges $e = \{u, v\}$, $e' = \{v, w\}$, we let $\Gamma(e \to e')$ be the isomorphism from $A_e = \{\vec{uv}, \vec{vu}\}$ to $A_{e'} = \{\vec{vw}, \vec{wv}\}$ that maps $\vec{uv}$ to $\vec{vw}$ and $\vec{vu}$ to $\vec{wv}$. Any path $\gamma = (\{v_1, v_2\}, \{v_2, v_3\}, \ldots, \{v_{n-1}, v_n\})$ of edges then gives rise to a connection $\Gamma(\gamma)$ identifying $A_{\{v_1,v_2\}}$ with $A_{\{v_{n-1},v_n\}}$. For instance, the triangular path $(\{u, v\}, \{v, w\}, \{w, u\}, \{u, v\})$ induces the identity map on $A_{\{u,v\}}$, whereas the U-turn path $(\{u, v\}, \{v, w\}, \{w, x\}, \{x, v\}, \{v, u\})$ induces the anti-identity map on $A_{\{u,v\}}$.

Given an orientation $\vec{G} = (\vec{e})_{e\in X}$ of the graph G, one can "differentiate" $\vec{G}$ at an edge $\{u, v\}$ in the direction $\{u, v\} \to \{v, w\}$ to obtain a number $\nabla_{\{u,v\}\to\{v,w\}}\vec{G}(\{u, v\}) \in \{-1, +1\}$, defined as $+1$ if the parallel transport from $\{u, v\}$ and $\{v, w\}$ preserves the orientations given by $\vec{G}$, and -1 otherwise. This number of course depends on the choice of orientation. But certain combinations of these numbers are independent of such a choice; for instance, given any closed path $\gamma = \{e_1, e_2, \ldots, e_n, e_{n+1} = e_1\}$ of edges in X, the "integral" $\prod_{i=1}^n \nabla_{e_i\to e_{i+1}}\vec{G}(e_i) \in \{-1, +1\}$ is independent of the choice of orientation $\vec{G}$ (indeed, it is equal to $+1$ if $\Gamma(\gamma)$ is the identity, and -1 if $\Gamma(\gamma)$ is the anti-identity.

Example 1.10.8 (Monodromy)**.** One can interpret the *monodromy maps* of a *covering space* in the language of connections. Suppose for instance that we have a covering space $\pi : \tilde{X} \to X$ of a topological space X whose fibres $\pi^{-1}(\{x\})$ are discrete; thus $\tilde{X}$ is a discrete fibre bundle over X. The discreteness induces a natural connection Γ on this space, which is given by the lifting map; in particular, if one integrates this connection on a closed loop based at some point x, one obtains the monodromy map of that loop at x.

Example 1.10.9 (Definite integrals)**.** In view of the definition (1.65), it should not be surprising that the *definite integral* $\int_a^b f(x)\,dx$ of a scalar function $f : [a, b] \to \mathbf{R}$ can be interpreted as an integral of a connection. Indeed, set $X := [a, b]$, and let $(\mathbf{R})_{x\in X}$ be the trivial line bundle over X. The function f induces a connection Γ_f on this bundle by setting

$$\Gamma_f(x \mapsto x + dx) : y \mapsto y + f(x)dx.$$

The integral $\Gamma_f([a,b])$ of this connection along $[a,b]$ is then just the operation of translation by $\int_a^b f(x)\,dx$ in the real line.

Example 1.10.10 (Line integrals)**.** One can generalise Example 1.10.9 to encompass *line integrals* in several variable calculus. Indeed, if X is an n-dimensional domain, then a vector field $f = (f_1, \ldots, f_n) : X \to \mathbf{R}^n$ induces a connection Γ_f on the trivial line bundle $(\mathbf{R})_{x\in X}$ by setting

$$\Gamma_f(x \mapsto x + dx) : y \mapsto y + f_1(x)dx_1 + \ldots + f_n(x)dx_n.$$

The integral $\Gamma_f(\gamma)$ of this connection along a curve γ is then just the operation of translation by the line integral $\int_\gamma f \cdot dx$ in the real line.

Note that a gauge transformation in this context is just a vertical translation $(x,y) \mapsto (x, y + V(x))$ of the bundle $(\mathbf{R})_{x\in X} \equiv X \times \mathbf{R}$ by some potential function $V : X \to \mathbf{R}$, which we will assume to be smooth for sake of discussion. This transformation conjugates the connection Γ_f to the connection $\Gamma_{f-\nabla V}$. Note that this is a *conservative* transformation: the integral of a connection along a closed loop is unchanged by gauge transformation.

Example 1.10.11 (ODE)**.** A different way to generalise Example 1.10.9 can be obtained by using the fundamental theorem of calculus to interpret $\int_{[a,b]} f(x)\,dx$ as the final value $u(b)$ of the solution to the initial value problem

$$u'(t) = f(t); \quad u(a) = 0$$

for the ordinary differential equation $u' = f$. More generally, the solution $u(b)$ to the initial value problem

$$u'(t) = F(t, u(t)); \quad u(a) = u_0$$

for some $u : [a,b] \to \mathbf{R}^n$ taking values in some Euclidean space[54] $\mathbf{R}^n$, where $F : [a,b] \times \mathbf{R}^n \to \mathbf{R}^n$ is a function (let us take it to be Lipschitz, to avoid technical issues), can also be interpreted as the integral of a connection Γ on the trivial vector space bundle $(\mathbf{R}^n)_{t\in[a,b]}$, defined by the formula

$$\Gamma(t \mapsto t + dt) : y \mapsto y + F(t,y)dt.$$

[54] One can also interpret ODE for functions u taking values in more general manifolds Y as integration along a connection; we leave the details to the reader.

Then $\Gamma[a,b]$ will map u_0 to $u(b)$, this is nothing more than the *Euler method* for solving ODE. Note that the method of *integrating factors* in solving ODE can be interpreted as an attempt to simplify the connection Γ via a gauge transformation. Indeed, it can be profitable to view the entire theory of connections as a multidimensional "variable-coefficient" generalisation of the theory of ODE.

Once one selects a gauge, one can express a connection in terms of that gauge. In the case of *vector bundles* (in which every fibre is a d-dimensional vector space for some fixed d), the covariant derivative $\nabla_v w(x)$ of a section w of that bundle along some vector v emanating from x can be expressed in any given gauge by the formula

$$\nabla_v w(x)^i = v^\alpha \partial_\alpha w(x)^i + v^\alpha \Gamma^i_{\alpha j} w(x)^j$$

where we use the gauge to express w(x) as a vector $(w(x)^1, \ldots, w(x)^d)$, the indices $i, j = 1, \ldots, d$ are summed over the fibre dimensions (and α summed over the base dimensions) as per the usual Einstein conventions (see Section 3.1), and the $\Gamma^i_{\alpha j} := (\nabla_{e_\alpha} e_j)^i$ are the *Christoffel symbols* of this connection relative to this gauge.

One example of this, which models electromagnetism, is a connection on a *complex line bundle* $V = (V_{t,x})_{(t,x)\in\mathbf{R}^{1+3}}$ in spacetime $\mathbf{R}^{1+3} = \{(t,x) : t \in \mathbf{R}, x \in \mathbf{R}^3\}$. Such a bundle assigns a complex line $V_{t,x}$ (i.e. a one-dimensional complex vector space, and thus isomorphic to $\mathbf{C}$) to every point (t,x) in spacetime. The structure group here is $U(1)$ (strictly speaking, this means that we view the fibres as *normed* one-dimensional complex vector spaces, otherwise the structure group would be $\mathbf{C}^\times$). A gauge identifies V with the trivial complex line bundle $(\mathbf{C})_{(t,x)\in\mathbf{R}^{1+3}}$, thus converting sections $(w_{t,x})_{(t,x)\in\mathbf{R}^{1+3}}$ of this bundle into complex-valued functions $\phi : \mathbf{R}^{1+3} \to \mathbf{C}$. A connection on V, when described in this gauge, can be given in terms of fields $A_\alpha : \mathbf{R}^{1+3} \to \mathbf{R}$ for $\alpha = 0, 1, 2, 3$; the covariant derivative of a section in this gauge is then given by the formula

$$\nabla_\alpha \phi := \partial_\alpha \phi + i A_\alpha \phi.$$

In the theory of electromagnetism, A_0 and (A_1, A_2, A_3) are known (up to some normalising constants) as the *electric potential* and *magnetic potential* respectively. Sections of V do not show up directly in

Maxwell's equations of electromagnetism, but appear in more complicated variants of these equations, such as the *Maxwell-Klein-Gordon equation*.

A gauge transformation of V is given by a map $U : \mathbf{R}^{1+3} \to S^1$; it transforms sections by the formula $\phi \mapsto U^{-1}\phi$, and connections by the formula $\nabla_\alpha \mapsto U^{-1}\nabla_\alpha U$, or equivalently

$$A_\alpha \mapsto A_\alpha + \frac{1}{i}U^{-1}\partial_\alpha U = A_\alpha + \partial_\alpha \frac{1}{i}\log U. \tag{1.66}$$

In particular, the electromagnetic potential A_α is not gauge invariant (which broadly corresponds to the concept of being *nonphysical* or *nonmeasurable* in physics), as gauge symmetry allows one to add an arbitrary gradient function to this potential. However, the *curvature tensor*[55]

$$F_{\alpha\beta} := [\nabla_\alpha, \nabla_\beta] = \partial_\alpha A_\beta - \partial_\beta A_\alpha$$

of the connection is gauge-invariant, and physically measurable in electromagnetism; the components $F_{0i} = -F_{i0}$ for $i = 1, 2, 3$ of this field have a physical interpretation as the *electric field*, and the components $F_{ij} = -F_{ji}$ for $1 \leq i < j \leq 3$ have a physical interpretation as the *magnetic field*.

Gauge theories can often be expressed succinctly in terms of a connection and its curvatures. For instance, *Maxwell's equations in free space*, which describes how electromagnetic radiation propagates in the presence of charges and currents (but no media other than vacuum), can be written (after normalising away some physical constants) as[56]

$$\partial^\alpha F_{\alpha\beta} = J_\beta$$

[55]The curvature tensor F can be interpreted as describing the parallel transport of infinitesimal rectangles; it measures how far off the connection is from being *flat*, which means that it can be (locally) "straightened" via some choice of gauge to be the trivial connection. In nonabelian gauge theories, in which the structure group is more complicated than just the abelian group $U(1)$, the curvature tensor is non-scalar, but remains gauge-invariant in a tensor sense (gauge transformations will transform the curvature as they would transform a tensor of the same rank.

[56]Actually, this is only half of Maxwell's equations, but the other half are a consequence of the interpretation (*) of the electromagnetic field as a curvature of a $U(1)$ connection, and indeed collapse to the *Bianchi identities* for that connection. Thus this purely geometric interpretation of electromagnetism has some non-trivial physical implications, for instance ruling out the possibility of (classical) magnetic monopoles.

where J_β is the *4-current*. If one generalises from complex line bundles to higher-dimensional vector bundles (with a larger structure group), one can then write down the (classical) Yang-Mills equation

$$\nabla^\alpha F_{\alpha\beta} = 0$$

which is the classical model[57] for three of the four fundamental forces in physics: the electromagnetic, weak, and strong nuclear forces (with structure groups $U(1)$, $SU(2)$, and $SU(3)$ respectively).

The gauge invariance (or gauge freedom) inherent in these equations complicates their analysis. For instance, due to the gauge freedom (1.66), Maxwell's equations, when viewed in terms of the electromagnetic potential A_α, are ill-posed: specifying the initial value of this potential at time zero does not uniquely specify the future value of this potential (even if one also specifies any number of additional time derivatives of this potential at time zero), since one can use (1.66) with a gauge function U that is trivial at time zero but non-trivial at some future time to demonstrate the non-uniqueness. Thus, in order to use standard PDE methods to solve these equations, it is necessary to first fix the gauge to a sufficient extent that it eliminates this sort of ambiguity. If one were in a one-dimensional situation (as opposed to the four-dimensional situation of spacetime), with a trivial topology (i.e. the domain is a line rather than a circle), then it is possible to gauge transform the connection to be completely trivial, for reasons[58] generalising both the fundamental theorem of calculus and the fundamental theorem of ODEs. However, in higher dimensions, one cannot hope to completely trivialise a connection by gauge transforms (mainly because of the possibility of a non-zero curvature form); in general, one cannot hope to do much better than setting a single component of the connection to equal zero. For instance, for Maxwell's equations (or the Yang-Mills equations), one can trivialise the connection A_α in the time direction, leading to the *temporal gauge condition*

$$A_0 = 0.$$

[57] The classical model for the fourth force, gravitation, is given by a somewhat different geometric equation, namely the *Einstein equations* $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$, though this equation is also "gauge-invariant" in some sense.

[58] Indeed, to trivialise a connection Γ on a line $\mathbf{R}$, one can pick an arbitrary origin $t_0 \in \mathbf{R}$ and gauge transform each point $t \in \mathbf{R}$ by $\Gamma([t_0, t])$.

This gauge is indeed useful for providing an easy proof of local existence for these equations, at least for smooth initial data. But there are many other useful gauges also that one can fix; for instance one has the *Lorenz gauge*

$$\partial^\alpha A_\alpha = 0$$

which has the nice property of being *Lorentz-invariant*, and transforms the Maxwell or Yang-Mills equations into linear or nonlinear wave equations respectively. Another important gauge is the *Coulomb gauge*

$$\partial_i A_i = 0$$

where i only ranges over spatial indices $1, 2, 3$ rather than over spacetime indices $0, 1, 2, 3$. This gauge has an elliptic[59] variational formulation. In some cases, the correct selection of a gauge is crucial in order to establish basic properties of the underlying equation, such as local existence. For instance, the simplest proof of local existence of the Einstein equations uses the *harmonic gauge*, which is analogous to the Lorenz gauge mentioned earlier; the simplest proof of local existence of Ricci flow uses a gauge of de Turck[**DeT1983**] that is also related to harmonic maps (see e.g. Section 3.2); and in my own work on wave maps[**Ta2008b**], [**Ta2008c**], a certain "caloric gauge" based on harmonic map heat flow is crucial. But in many situations, it is not yet fully understood whether the use of the correct choice of gauge is a mere technical convenience, or is more innate to the equation. It is definitely conceivable, for instance, that a given gauge field equation is well-posed with one choice of gauge but ill-posed with another. It would also be desirable to have a more gauge-invariant theory of PDEs that did not rely so heavily on gauge theory at all, but this seems to be rather difficult; many of our most powerful tools in PDE (for instance, the Fourier transform) are highly non-gauge-invariant, which makes it very inconvenient to try to analyse these equations in a purely gauge-invariant setting.

[59] More precisely, Coulomb gauges are critical points of the functional $\int_{\mathbf{R}^3} \sum_{i=1}^3 |A_i|^2$) and thus are expected to be "smaller" and "smoother" than many other gauges; this intuition can be borne out by standard elliptic theory (or *Hodge theory*, in the case of Maxwell's equations).

Notes. This article first appeared at `terrytao.wordpress.com/2008/09/27`. Thanks to Roland Bacher for corrections.

Pedro Lauridsen Ribeiro pointed out some recent work in developing gauge-invariant techniques for controlling solutions to gauge equations, e.g. [**KlRo2007**], and also noted that a good characterisation for hyperbolicity in gauge-invariant equations was still lacking.

Allen Knutson pointed out two further interesting examples of bundles and connections. The first was that of a *quiver*, where the base space is a graph, the fibres over each vertex is a vector space (of varying dimension), and the parallel transport maps are linear transformations from one such space to another. Another was that of currency trading, where the base space is the set of currencies, the fibers are one-dimensional vector spaces, and the parallel transport maps are currency exchange rates. In this example, curvature of the connection can be interpreted as an arbitrage opportunity; see [**Il1997**].

1.11. The Lucas-Lehmer test for Mersenne primes

Recently, the *Great Internet Mersenne Prime Search* (GIMPS) announced the discovery of two new Mersenne primes, both over ten million digits in length, including one discovered by the computing team right here at UCLA.

The GIMPS approach to finding Mersenne primes relies of course on modern computing power, parallelisation, and efficient programming, but the number-theoretic heart of it - aside from some basic optimisation tricks such as *fast multiplication* and preliminary sieving to eliminate some obviously non-prime Mersenne number candidates - is the *Lucas-Lehmer primality test for Mersenne numbers*, which is much faster for this special type of number than any known general-purpose (deterministic) primality test (such as, say, the *AKS primality test*[**AgKaSa2004**]). This test is easy enough to describe, and I will do so later in this post, and also has some short elementary proofs of correctness; but the proofs are sometimes presented in a way that involves pulling a lot of rabbits out of hats, giving the argument a

magical feel rather than a natural one. In this article, I will try to explain the basic ideas that make the primality test work, seeking a proof which is perhaps less elementary and a little longer than some of the proofs in the literature, but is perhaps a bit better motivated.

1.11.1. Order. We begin with a general discussion of how to tell when a given number n (which is not necessarily of the Mersenne form $n = 2^m - 1$) is prime or not. One should think of n as being moderately large, e.g. $n \sim 10^{10^7}$ (which is broadly the size of the Mersenne primes discovered recently).

Our starting point will be *Lagrange's theorem*, which asserts that

$$a^{|G|} = 1 \tag{1.67}$$

for any finite group G and any $a \in G$, thus the *order* $\mathrm{ord}_G(a)$ of a in G divides $|G|$. Specialising this to the multiplicative group $\mathbf{F}_p^\times$ of a finite field of prime order p, we obtain *Fermat's little theorem*

$$a^{p-1} = 1 \bmod p \tag{1.68}$$

for p prime and a coprime to p; applying it instead to the multiplicative group $(\mathbf{Z}/n\mathbf{Z})^\times$ of a cyclic group of order n, we obtain *Euler's theorem*

$$a^{\phi(n)} = 1 \bmod n \tag{1.69}$$

whenever a is coprime to n, where $\phi(n) := |(\mathbf{Z}/n\mathbf{Z})^\times|$ is the *Euler totient function* of n.

Fermat's little theorem (1.68) already gives a necessary condition for the primality of a candidate prime n: take any a coprime to n (typically one picks a small number such as $a = 2$ or $a = 3$), and compute a^{n-1} modulo n. If it is not equal to 1, then n cannot be prime. This is a (barely) feasible test to execute for n as large as 10^{10^7}, because one can compute exponents such as a^{n-1} relatively quickly, by the trick of repeatedly squaring a modulo n to obtain $a, a^2, a^{2^2}, a^{2^3}, \ldots \bmod n$, and then decomposing $n-1$ into binary[60]

[60] If n is a Mersenne number, then there are some pretty obvious shortcuts one can take for this last step, using division instead of multiplication.

to compute a^{n-1}. Unfortunately, while this test is necessary for primality, it is not sufficient, due to the existence of *pseudoprimes*[61]. So Fermat's little theorem alone does not provide the answer, at least if one wants a *deterministic* certificate of primality rather than a *probabilistic* one.

Nevertheless, the above facts do provide some important information about the order $\mathrm{ord}_n(a) := \mathrm{ord}_{(\mathbf{Z}/n\mathbf{Z})^*}(a)$ of a modulo n. If n is prime, Fermat's little theorem (1.68) tells us that $\mathrm{ord}_n(a)$ is at most $n-1$ (in fact, it divides $n-1$). On the other hand, if n is not prime, then Euler's theorem (1.69) tells us that $\mathrm{ord}_n(a)$ cannot be as large as $n-1$, but is instead at most $\phi(n)$, which is now strictly less than $n-1$. Thus: if we can find a number a coprime to n such that $\mathrm{ord}_n(a)$ is exactly $n-1$, we have certified that n is prime.

Testing whether a given number a is coprime to n is very easy and fast (thanks to the *Euclidean algorithm*). Unfortunately, it is difficult in general to compute[62] the order $\mathrm{ord}_n(a)$ of a number a if the base n is large; a brute-force approach would require one to compute up to $n-1$ powers of a, which is prohibitively expensive if n has size comparable to 10^{10^7}. However, there are a few cases in which the order can be found very quickly. Suppose that we somehow find positive integers a, k such that

$$a^{2^k} = -1 \bmod n \tag{1.70}$$

(which in particular implies that a is coprime to n). Squaring this, we obtain

$$a^{2^{k+1}} = 1 \bmod n$$

and so we see that $\mathrm{ord}_n(a)$ divides 2^{k+1} but not 2^k, and thus must be exactly equal to 2^{k+1}. Conversely, if $\mathrm{ord}_n(a) = 2^{k+1}$, then we must have (1.70). So in the special case when the order of a is a power of 2, we can compute the order using only one exponentiation, which is computationally feasible for the orders of magnitude we are considering.

---

[61] For instance, the number $n = 561 = 3*11*17$ is not prime, but $a^{n-1} = 1 \bmod n$ is true for all a coprime to n (in other words, 561 is a *Carmichael number*).

[62] More generally, the problem of computing order exactly in general is closely related to the *discrete logarithm problem*, which is notoriously difficult.

Unfortunately, this is not quite what we need for the Mersenne prime problem, because if n is a Mersenne number, then it is n *plus* 1 which is[63] a power of 2, rather than n *minus* 1.

So this is a frustrating near miss: if n is a Mersenne number, we can easily check if a number has order $n+1$ modulo n, but we needed a test for when a number has order $n-1$ instead. And indeed, even when n is prime, Fermat's little theorem (1.68) shows that it is impossible for a number to have order $n+1$ modulo n, since the order needs to divide $n-1$. So we seem to be a bit stuck.

But while $n+1$ clearly does not divide $n-1$, it *does* divide n^2-1. Looking at Lagrange's theorem (1.67), we then see that it could be possible to find elements of order $n+1$ in a multiplicative group of order n^2-1 rather than $n-1$. Recall that if n was prime, then the multiplicative group $\mathbf{F}_n^\times$ of the finite field $\mathbf{F}_n$ had order $n-1$. But $\mathbf{F}_{n^2}$ is also a finite field, and its multiplicative group $\mathbf{F}_{n^2}^\times$ has order n^2-1. Aha!

So the plan (assuming for sake of argument that n is prime) is to somehow work in the finite field $\mathbf{F}_{n^2}$ instead of $\mathbf{F}_n$, in order to find elements of order $n+1$. We can get our hands on this larger finite field more concretely by viewing it as a *quadratic extension* $\mathbf{F}_n[\sqrt{a}]$, where a is a *quadratic non-residue* of n.

Let's now take n to be a Mersenne prime. What numbers are quadratic non-residues? A quick appeal to *quadratic reciprocity* and some elementary number theory soon reveals that 2 is a quadratic residue of n, but that 3 is not. Thus we can[64] take $\mathbf{F}_{n^2} \equiv \mathbf{F}_n[\sqrt{3}]$. Henceforth all calculations will be in this field $\mathbf{F}_n[\sqrt{3}]$, which of course contains $\mathbf{F}_n$ as a subfield.

Now we need to look for a field element a of order $n+1$, which is a power of 2. Thus (by adapting (1.70) to $\mathbf{F}_n[\sqrt{3}]$) we need to find

[63] The above method would be ideal for finding *Fermat primes* rather than Mersenne primes, but it is likely that in fact there are no more such primes to be found.

[64] One could work with other numbers than 3 here, but being the smallest quadratic non-residue available, it is the simplest one to use, and the one which is most likely to be able to take advantage of the *strong law of small numbers* [**Gu1988**], which is the informal assertion that numerical coincidences are most likely to occur amongst small numbers than amongst large ones.

a solution to the equation

$$a^{(n+1)/2} = -1 \tag{1.71}$$

in this field.

Let's compute some expressions of the form $a^{(n+1)/2}$. From Fermat's little theorem (1.68) we have

$$3^{n-1} = 1;$$

because 3 is not a quadratic residue, we see (from taking discrete logarithms) that

$$3^{(n-1)/2} = -1 \tag{1.72}$$

and thus

$$3^{(n+1)/2} = -3.$$

Similarly we have

$$2^{(n+1)/2} = +2 \tag{1.73}$$

These are pretty close to (1.71), but not quite right. To go further, it is convenient to work with n^{th} powers rather than $((n+1)/2)^{\text{th}}$ powers - i.e. we work with the *Frobenius endomorphism* $x \mapsto x^n$. Indeed, since $\mathbf{F}_n[\sqrt{3}]$ has characteristic n, we have the endomorphism properties

$$(x+y)^n = x^n + y^n; \quad (xy)^n = x^n y^n. \tag{1.74}$$

From (1.72) we have $(\sqrt{3})^n = -\sqrt{3}$, while from (1.68) we have $a^n = a$ for $a \in \mathbf{F}_n$. From (1.74) we thus see that

$$(a + b\sqrt{3})^n = a - b\sqrt{3}$$

for $a, b \in \mathbf{F}_n$; thus the Frobenius automorphism is nothing more than Galois conjugation[65].

Now we go back from n^{th} powers to $((n+1)/2)^{th}$ powers. Multiplying both sides of the preceding equation by $a + b\sqrt{3}$, we obtain

$$(a + b\sqrt{3})^{n+1} = a^2 - 3b^2.$$

Squaring $a + b\sqrt{3}$, we conclude

$$(a^2 + 3b^2 + 2ab\sqrt{3})^{(n+1)/2} = a^2 - 3b^2.$$

[65]Actually, this can be deduced quite readily from standard Galois theory.

Now we in a good position to solve the equation (1.71). We cannot make a^2-3b^2 equal to -1 - since -1 is not a quadratic residue modulo 3 - but we can make it equal to, say, -2, by setting $a=1$ and $b=-1$ (say):

$$(4+2\sqrt{3})^{(n+1)/2} = -2.$$

Dividing this by (1.73) we obtain the desired solution

$$\omega^{(n+1)/2} = -1 \tag{1.75}$$

to (1.71), where[66] $\omega := 2+\sqrt{3}$.

To summarise, we have shown that

Proposition 1.11.1. *If n is a Mersenne prime, then* (1.75) *holds.*

Based on our previous discussion, we expect to be able to reverse this implication. Indeed, we have the following converse:

Lemma 1.11.2. *Let n be a Mersenne number. If* (1.75) *holds (in the ring $(\mathbf{Z}/n\mathbf{Z})[\sqrt{3}]$), then n is prime.*

Proof. We use an argument of Bruce[**Br1993**]. Let q be a prime divisor of n. Then $\omega^{(n+1)/2} = -1$ in the field $\mathbf{F}_q[\sqrt{3}]$ (which we define as $\mathbf{F}_q$ if 3 is a quadratic residue there), thus ω has order exactly n+1 (cf. (1.70)). By Lagrange's theorem (1.67), this means that $n+1$ divides the multiplicative order of $\mathbf{F}_q[\sqrt{3}]^\times$, which is q^2-1 (if 3 is a non-residue modulo q) or $q-1$ (if 3 is a residue modulo q). In particular, q has to exceed $\sqrt{n}$. Thus the only prime divisors of n exceed $\sqrt{n}$, and so by the *sieve of Eratosthenes*, n is prime. □

We have thus shown

Corollary 1.11.3 (Lucas-Lehmer test, preliminary version)**.** *Let $n = 2^m-1$ with m odd. Then n is prime if and only if* (1.75) *holds in $\mathbf{Z}/n\mathbf{Z}[\sqrt{3}]$.*

This is already a reasonable criterion, but it is a little non-elementary (and also a little unpleasant numerically) due to the presence of the quadratic extension by $\sqrt{3}$. One can get rid of this extension by the Galois theory trick of taking *traces*. Indeed, observe

[66] Note that one could also use $\omega^{-1} = 2-\sqrt{3}$ here; indeed, Galois theory tells us that $+\sqrt{3}$ and $-\sqrt{3}$ are interchangeable in these computations.

that $\omega^{-1} = 2 - \sqrt{3}$ is the *Galois conjugate* of ω. Basic Galois theory tells us that $\omega^{(n+1)/4} + \omega^{-(n+1)/4}$ lies in $\mathbf{Z}/n\mathbf{Z}$, and vanishes precisely when $\omega^{(n+1)/2}$ is equal to -1. So it suffices to show that

$$\omega^{(n+1)/4} + \omega^{-(n+1)/4} = \omega^{2^{m-2}} + \omega^{-2^{m-2}} =: S_{m-2}$$

vanishes in $\mathbf{Z}/n\mathbf{Z}$. The quantity $\omega^{(n+1)/4} = \omega^{2^{m-2}}$ could be computed by repeated squaring in $\mathbf{Z}/n\mathbf{Z}[\sqrt{3}]$. The quantity S_{m-2} can be computed by a similar device in $\mathbf{Z}/n\mathbf{Z}$. Indeed, the sequence $S_j := \omega^{2^j} + \omega^{-2^j}$ is easily seen to obey the recursion

$$S_j = S_{j-1}^2 - 2; \quad S_0 = 4 \tag{1.76}$$

and so we have

Theorem 1.11.4 (Lucas-Lehmer test, final version)**.** *Let $n = 2^m - 1$ with m odd. Then n is prime if and only if S_{m-2} vanishes modulo n, where S_{m-2} is given by the recursion* (1.76)*.*

To apply this test, one needs to perform about m squaring operations modulo n. Doing everything as efficiently as possible (in particular, using *fast multiplication*), the total cost of testing a single Mersenne number $n = 2^m - 1$ for primality is about $O(m^2)$ (modulo some $\log m$ terms). This turns out to barely be within reach[67] of modern computers for $m \sim 10^7$, especially since the algorithm is somewhat parallelisable. There are general-purpose probabilistic tests (such as the *Miller-Rabin test*) which have run-time comparable to the Lucas-Lehmer test, but as mentioned at the beginning, we are only interested here in deterministic (and unconditional, in particular not relying on the generalised Riemann hypothesis) certificates of primality.

Notes. This article first appeared at `terrytao.wordpress.com/2008/10/02`. Thanks to René Schoof, Jernej, and several anonymous commenters for corrections.

More information on the recently found Mersenne primes can be found at `www.math.ucla.edu/ edson/prime`. (I was not involved in

[67] In contrast, the best known general-purpose deterministic primality testing algorithm, the AKS algorithm[**AgKaSa2004**], has a run time of about $O(m^6)$ (with a sizable implicit constant), which is not feasible for $m \sim 10^7$.

this computing effort.) As for the question "Why do we want to find such big primes anyway?", see `primes.utm.edu/notes/faq/why.html`.

An anonymous commenter pointed out that an application of large Mersenne primes to coding theory has appeared recently in [**Ye2007**].

1.12. Finite subsets of groups with no finite models

Additive combinatorics is largely focused on the additive properties of finite subsets A of an additive group $G = (G, +)$. This group can be finite or infinite, but there is a very convenient trick, the *Ruzsa projection trick*, which allows one to reduce the latter case to the former. For instance, consider the set $A = \{1, \ldots, N\}$ inside the integers $\mathbf{Z}$. The integers of course form an infinite group, but if we are only interested in sums of at most two elements of A at a time, we can embed A inside the finite cyclic group $\mathbf{Z}/2N\mathbf{Z}$ without losing any combinatorial information. More precisely, there is a *Freiman isomorphism of order* 2 between the set $\{1, \ldots, N\}$ in $\mathbf{Z}$ and the set $\{1, \ldots, N\}$ in $\mathbf{Z}/2N\mathbf{Z}$. One can view the latter version of $\{1, \ldots, N\}$ as a *model* for the former version of $\{1, \ldots, N\}$. More generally, it turns out that any finite set A in an additive group can be modeled in the above set by an equivalent set in a finite group, and in fact one can ensure that this ambient modeling group is not much larger than A itself if A has some additive structure; see [**Ru1994**] or [**TaVu2006**, Lemma 5.26] for a precise statement. This projection trick has a number of important uses in additive combinatorics, most notably in Ruzsa's simplified proof[**Ru1994**] of Freiman's theorem[**Fr1973**].

Given the interest in non-commutative analogues of Freiman's theorem (see Section 3.2 of *Structure and Randomness*), it is natural to ask whether one can similarly model finite sets A in multiplicative (and non-commutative) groups $G = (G, \times)$ using finite models. Unfortunately (as I learned recently from Akshay Venkatesh, via Ben Green), this turns out to be impossible in general, due to an old example of Higman[**Hi1951**]. More precisely, Higman shows:

Theorem 1.12.1. *There exists an infinite group G generated by four distinct elements a, b, c, d that obey the relations*

$$ab = ba^2; bc = cb^2; cd = dc^2; da = ad^2; \tag{1.77}$$

in fact, a and c generate the free nonabelian group in G. On the other hand, if G' is a finite group containing four elements a, b, c, d obeying (1.77)*, then a, b, c, d are all trivial.*

As a consequence, the finite set $A := \{1, a, b, c, d, ab, bc, cd, da\}$ in G has no model (in the sense of *Freiman isomorphisms*) in a finite group. Theorem 1.12.1 is proven by a small amount of elementary group theory and number theory, and it was neat enough that I thought I would reproduce it here.

1.12.1. No non-trivial finite models. Let's first show the second part of Theorem 1.12.1. The key point is that in a finite group G', all elements have finite *order*, thanks to *Lagrange's theorem*. From (1.77) we have

$$b^{-1}ab = a^2$$

and hence by induction

$$b^{-n}ab^n = a^{2^n} \tag{1.78}$$

for any positive n. One consequence of (1.78) is that if $b^n = 1$, then $a = a^{2^n}$, and thus $a^{2^n-1} = 1$. Applying this with n equal to the order $\mathrm{ord}(b)$ of b, we conclude that

$$\mathrm{ord}(a) | 2^{\mathrm{ord}(b)} - 1.$$

As a consequence, if $\mathrm{ord}(a)$ is divisible by some prime p, then $2^{\mathrm{ord}(b)}-1$ is divisible by p, which forces p to be odd and $\mathrm{ord}(b)$ to be divisible by the multiplicative order of 2 modulo p. This is at most $p-1$ (by *Fermat's little theorem*), and so $\mathrm{ord}(b)$ is divisible by a prime strictly smaller than the prime dividing $\mathrm{ord}(a)$. But we can cyclically permute this argument and conclude that $\mathrm{ord}(c)$ is divisible by an even smaller prime than the prime dividing $\mathrm{ord}(b)$, and so forth, creating an *infinite descent*, which is absurd. Thus none of $\mathrm{ord}(a), \mathrm{ord}(b), \mathrm{ord}(c), \mathrm{ord}(d)$ can be divisible by any prime, and so a, b, c, d are trivial as claimed.

Remark 1.12.2. There is nothing special here about using four generators; the above arguments work with any number of generators (adapting (1.77) appropriately). But we will need four generators in order to establish the infinite model below.

Remark 1.12.3. The above argument also shows that the group G has no non-trivial finite-dimensional linear representation. Indeed, let a, b, c, d be matrices obeying (1.77), then b is conjugate to $b^2 = c^{-1}bc$, which by the spectral theorem forces the eigenvalues of b to be roots of unity, which implies in particular that b^n grows at most polynomially in n; similarly for a^n. Applying (1.78) we see that a^{2^n} grows at most polynomially in n, which by the *Jordan normal form* (see Section 1.13 of *Structure and Randomness*) for a implies that a is diagonalisable; since its eigenvalues are roots of unity, it thus has finite order. Similarly for b, c, d. Now apply the previous argument to conclude that a, b, c, d are trivial.

1.12.2. Existence of an infinite model. To build the infinite group G that obeys the relations (1.77), we need the notion of an *amalgamated free product* of groups. Recall that the *free product* $G_1 * G_2$ of two groups G_1 and G_2 can be defined (up to group isomorphism) in one of three equivalent ways:

(1) (Relations-based definition) $G_1 * G_2$ is the group *generated* by the disjoint union $G_1 \uplus G_2$ of G_1 and G_2, with no further relations between these elements beyond those already present in G_1 and G_2 separately.

(2) (Category-theoretic definition) $G_1 * G_2$ is a group with *homomorphisms* from G_1 and G_2 into $G_1 * G_2$, which is *universal* in the sense that any other group G' with homomorphisms from G_1, G_2 will have these homomorphisms factor uniquely through $G_1 * G_2$.

(3) (Word-based definition) $G_1 * G_2$ is the collection of all *words* $g_1 g_2 \dots g_n$, where each g_i lies in either G_1 or G_2, with no two adjacent g_i, g_{i+1} lying in the same G_j (let's label G_1, G_2 here to be disjoint to avoid notational confusion), with the obvious group operations.

It is not hard to see that all three definitions are equivalent, and that the free product exists and is unique up to group isomorphism.

Example 1.12.4. The free product of the free cyclic group $\langle a \rangle$ with one generator a, and the free cyclic group $\langle b \rangle$ with one generator b, is the free (non-abelian) group $\langle a, b \rangle$ on two generators.

We will need a "relative" generalisation of the free product concept, in which the groups G_1, G_2 are not totally disjoint, but instead share a common subgroup H (or if one wants to proceed more category-theoretically, with a group H that embeds into both G_1 and G_2). In this situation, we define the amalgamated free product $G_1 *_H G_2$ by one of the following two equivalent definitions:

(1) (Relations-based definition) $G_1 *_H G_2$ is the group *generated* by the relative disjoint union $G_1 \uplus_H G_2$ of G_1 and G_2 (which is the same as the disjoint union but with the common subgroup H identified), with no further relations between these elements beyond those already present in G_1 and G_2 separately.

(2) (Category-theoretic definition) $G_1 *_H G_2$ is a group with *homomorphisms* from G_1 and G_2 into $G_1 *_H G_2$ that agree on H, which is *universal* in the sense that any other group G' with homomorphisms from G_1, G_2 that agree on H will have these homomorphisms factor uniquely through $G_1 *_H G_2$.

Example 1.12.5. Let $G_1 := \langle a, b | ab = ba^2 \rangle$ be the group generated by two elements a, b with one relation $ab = ba^2$. It is not hard to see that all elements of G_1 can be expressed uniquely as $b^n a^m$ for some integers n, m, and in particular that $H := \langle b \rangle$ is a free cyclic group. Let $G_2 := \langle b, c | bc = cb^2 \rangle$ be the group generated by two elements b, c with one relation $bc = cb^2$, then again $H := \langle b \rangle$ is a free cyclic group, and isomorphic to the previous copy of H. The amalgamated free product $G_1 *_H G_2 = \langle a, b, c | ab = ba^2, bc = cb^2 \rangle$ is then generated by three elements a, b, c with two relations $ab = ba^2, bc = cb^2$.

It is not hard to see that the above two definitions are equivalent, and that $G_1 *_H G_2$ exists and is unique up to group isomorphism. But note that I did not give the word-based definition of the amalgamated

free product yet. We will need to do so now; I will use the arguments from [**Ne1954**], though the basic result I need here (namely, Corollary 1.12.8) dates all the way back to the work of Schreier in 1927.

In order to analyse these groups, we will need to study how they act on various spaces. If G is a group, we define a *G-space* to be a set X together with an *action* $(g,x) \mapsto gx$ of G on X (or equivalently, a homomorphism from G to the permutation group $\mathrm{Sym}(X)$ of X). Thus for instance G is itself a G-space. A G-space X is *transitive* if for every $x, y \in X$, there exists $g \in G$ such that $gx = y$. A *morphism* from one G-space X to another G-space Y is a map $\phi : X \to Y$ such that $\phi(gx) = g\phi(x)$ for all $g \in G$ and $x \in X$. If a morphism has an inverse that is also a morphism, we say that it is an *isomorphism*.

The first observation is that a G-space with certain properties will necessarily be isomorphic to G itself.

Lemma 1.12.6 (Criterion for isomorphism with G)**.** *Let G be a group, let X be a non-empty transitive G-space, and suppose there is a morphism $\pi : X \to G$ from the G-space X to the G-space G. Then π is in fact an isomorphism of G-spaces.*

Proof. It suffices to show that π is both injective and surjective. To show surjectivity, observe that the image $\pi(X)$ is G-invariant and non-empty. But the action of G on G is transitive, and so $\pi(X) = G$ as desired. To show injectivity, observe from transitivity that if x, x' are distinct elements of X then $x' = gx$ for some non-identity $g \in G$, thus $\pi(x') = g\pi(x)$, thus $\pi(x') \neq \pi(x)$, establishing injectivity. □

Now we can give the word formulation of the amalgamated free product.

Lemma 1.12.7 (Word-based description of amalgamated free product)**.** *Let G_1, G_2 be two groups with common subgroup H, and let $G := G_1 *_H G_2$ be the amalgamated free product. Let $G_1 = \bigcup_{s_1 \in S_1} H \cdot s_1$, $G_2 = \bigcup_{s_2 \in S_2} H \cdot s_2$ be some partitions of G_1, G_2 into right-cosets of H. Let X be the space of all formal words of the form $hs_1 s_2 \dots s_n$, where $h \in H$, each s_i lies in either S_1 or S_2, and no two adjacent s_i, s_{i+1} lie in the same S_j. let $\pi : X \to G$ be the obvious evaluation*

map. Then there is an action of G on X for which π becomes an isomorphism of G-spaces.

Proof. It is easy to verify that G_1 and G_2 act separately on X in a manner consistent (via π) with their action on G, and these actions agree on H. Hence the amalgamated free product G also acts on this space and turns π into a morphism of G-spaces. From construction of X we see that the G-action is transitive, and the claim now follows from Lemma 1.12.6. □

Corollary 1.12.8. *Let G_1, G_2 be two groups with common subgroup H, and let $G := G_1 *_H G_2$ be the amalgamated free product. Let $g_1 \in G_1$ and $g_2 \in G_2$ be such that the cyclic groups $\langle g_1 \rangle, \langle g_2 \rangle$ are infinite and have no intersection with H. Then g_1, g_2 generate a free subgroup in G.*

Proof. By hypothesis (and the axiom of choice), we can find a partition $G_1 = \bigcup_{s_1 \in S_1} H \cdot s_1$ where S_1 contains the infinite cyclic group $\langle g_1 \rangle$, and similarly we can find a partition $G_2 = \bigcup_{s_2 \in S_2} H \cdot s_2$. Let X be the space in Lemma 2. Each reduced word formed by g_1, g_2 then generates a distinct element of X, and thus (by Lemma 1.12.7) a distinct element of G. The claim follows. □

Remark 1.12.9. The above corollary can also be established by the *ping-pong lemma* (which is not surprising, since the proof of that lemma uses many of the same ideas, and in particular exploiting an action of G on a space X in order to distinguish various words in G from each other). Indeed, observe that g_1, g_1^{-1} map those words $hs_1 s_2 \dots s_n$ in X with $s_1 \notin S_1$ into words $hs_0 s_1 \dots s_n$ with $s_0 \in S_1$, and similarly for g_2, g_2^{-1}, which is the type of hypothesis needed to apply the ping-pong lemma. [Thanks to Ben Green for this observation.]

Now we can finish the proof of Theorem 1.12.1. As discussed in Example 1.12.5, the group $G_1 := \langle a, b, c | ab = ba^2, bc = cb^2 \rangle$ is the amalgamated free product of $\langle a, b | ab = ba^2 \rangle$ and $\langle b, c | bc = cb^2 \rangle$ relative to $\langle b \rangle$. By Corollary 1.12.8, a and c generate the free group here, thus G_1 contains $H = \langle a, c \rangle$ as a subgroup. Similarly, the group $G_2 := \langle c, d, a | cd = dc^2; da = ad^2 \rangle$ also contains $H = \langle a, c \rangle$ as a

subgroup. We may then form the amalgamated free product

$$G := G_1 *_H G_2 = \langle a, b, c, d | ab = ba^2, bc = cb^2, cd = dc^2, da = ad^2 \rangle$$

and another application of Corollary 1.12.8 shows that b, d generate the free group (and are in particular distinct); similarly a, c are distinct. Finally, the group $\langle a, b | ab = ba^2 \rangle$ embeds into G_1, which embeds into G, and so a, b, are also distinct; cyclically permuting this we conclude that all of the a, b, c, d are distinct as claimed.

Notes. This article first appeared at `terrytao.wordpress.com/2008/10/06`. Thanks to an anonymous commenter for corrections.

Ben Green pointed out that for the specific "non-commutative Freiman theorem" application of trying to characterise finite sets A of small doubling (thus $|A \cdot A| \leq K|A|$), it may still be possible that some large *subset* A' of A has a finite model, even if A itself need not be. Currently, all known examples of finite sets of small doubling have this property.

David Fisher pointed out that Remark 1.12.3 can be deduced from Theorem 1.12.1 using the general fact that all linear groups are residually finite (i.e. have finite quotients that separate any finite set). The proof of this latter fact is non-trivial, however.

1.13. Small samples, and the margin of error

In view of this year's U.S. presidential election, I would like to talk about some of the basic mathematics underlying electoral polling, and specifically to explain the fact, which can be highly unintuitive to those not well versed in statistics, that polls can be accurate even when sampling only a tiny fraction of the entire population.

Take for instance a nationwide poll of U.S. voters on which presidential candidate they intend to vote for. A typical poll will ask a number n of randomly selected voters for their opinion; a typical value here is $n = 1000$. In contrast, the total voting-eligible population of the U.S. - let's call this set X - is about 200 million[68]. Thus, such a poll would sample about 0.0005% of the total population X

[68]The actual *turnout* for the 2008 election ended up being approximately 130 million, but let's ignore this fact for sake of discussion.

- an incredibly tiny fraction. Nevertheless, the margin of error (at the 95% *confidence level*) for such a poll, if conducted under idealised conditions (see below), is about 3%. In other words, if we let p denote the proportion of the entire population X that will vote for a given candidate A, and let $\overline{p}$ denote the proportion of the polled voters that will vote for A, then the event $\overline{p} - 0.03 \leq p \leq \overline{p} + 0.03$ will occur with probability at least 0.95. Thus, for instance (and oversimplifying somewhat by ignoring the probability-altering effects of conditional expectation - see below), if the poll reports that 55% of respondents would vote for A, then the true percentage of the electorate that would vote for A has at least a 95% chance of lying between 52% and 58%. Larger polls will of course give a smaller margin of error; for instance the margin of error for an (idealised) poll of $2,000$ voters is about 2%.

I'll give a rigorous proof of a weaker version of the above statement (giving a margin of error of about 7%, rather than 3%) in an appendix at the end of this post. But the main point of my post here is a little different, namely to address the common misconception that the accuracy of a poll is a function of the *relative* sample size rather than the *absolute* sample size, which would suggest that a poll involving only 0.0005% of the population could not possibly have a margin of error as low as 3%. I also want to point out some limitations of the mathematical analysis; depending on the methodology and the context, some polls involving 1000 respondents may have a much higher margin of error than the idealised rate of 3%.

1.13.1. Assumptions and conclusion. Not all polls are created equal; there are a certain number of hypotheses on the methodology and effectiveness of the poll that we have to assume in order to make our mathematical conclusions valid. We will make the following idealised assumptions:

(1) **Simple question.** Voters polled can only offer one of two responses, which I will call A and not-A; thus we ignore the effect of third-party candidates, undecided voters, or refusals to respond. In particular, we do not try to combine this data with other questions about the polled voters, such as demographic data. We also assume that the question is

unambiguous and cannot be misinterpreted by respondents (see Hypothesis 3 below).

(2) **Perfect response rate.** All voters polled offer a response; there are no refusals to respond to the poll, or failures to make contact with the voter being polled. (This is a special case of Hypothesis 1, but deserves to be emphasised.) In particular, this excludes polls that are self-selected, such as internet polls (since in most cases, a large fraction of viewers of a web page with a poll will refuse to respond to that poll).

(3) **Honest responses.** The response given by a voter to the poll is an accurate representation whether that voter intends to vote for A or not; thus we ignore response-distorting effects such as the *Bradley effect*, *push-polling*, *tactical voting*, frivolous responses, misunderstanding of the question, or attempts to "game" a poll by the respondents.

(4) **Fixed poll size.** The number n of polled voters is fixed in advance; in particular, one cannot keep polling until one has achieved some desired outcome, and then stop.

(5) **Simple random sampling (without replacement).** Each one of the n voters polled is selected uniformly at random among the entire population X, thus each voter is equally likely to be selected by the poll, and no non-voter can be selected by the poll. (In particular, we make the important assumption that there is no *selection bias*.) Furthermore, each polled voter is chosen *independently* of all the others, except for the one condition that we do not poll any given voter more than once. (Thus, once a voter is polled, that voter is "crossed off the list" of the pool X of voters that one randomly selects from to determine the next voter polled.) In particular, we assume that the poll is not *clustered*.

(6) **Honest reporting.** The results of the poll are always reported, with no inaccuracies; one cannot cancel, modify, or ignore a poll once it has begun. In particular, one cannot conduct multiple polls and only report the "best" results (thus running the risk of *confirmation bias*).

Polls which deviate significantly from these hypotheses (e.g. due to complex questions, self-selection or other selection bias, confirmation bias, inaccurate responses, a high refusal rate, variable poll size, or clustering) will generally be less accurate than an idealised poll with the same sample size. Of course, there is a substantial literature in statistics (and polling methodology) devoted to measuring, mitigating, avoiding, or compensating for these less ideal situations, but we will not discuss those (important) issues here. We will remark though that in practice it is difficult to make the poll selection truly uniform. For instance, if one is conducting a telephone poll, then the sample will of course be heavily biased towards those voters who actually own phones; a little more subtly, it will also be biased toward those voters who are near their phones at the time the poll was conducted, and have the time and inclination to answer phone calls. As long as these factors are not strongly correlated with the poll question (i.e. whether the voter will vote for A), this is not a major concern, but in some cases, the poll methodology will need to be adjusted (e.g. by reweighting the sample) to compensate for the non-uniformity.

As stated in the introduction, we let p be the proportion of the entire population X that will vote for A, and $\overline{p}$ be the proportion of the polled voters that will vote for A (which, by Hypotheses 2 and 3, is exactly equal to the proportion of polled voters that say that they will vote for A). Under the above idealised conditions, if the number n of polled voters is $1,000$, and the size of the population X is 200 million, then the margin of error is about 3%, thus $\mathbf{P}(\overline{p} - 0.03 \leq p \leq \overline{p} + 0.03) \geq 0.95$.

There is an important subtlety here: it is only the *unconditional* probability of the event $\overline{p} - 0.03 \leq p \leq \overline{p} + 0.03$ that is guaranteed to be greater than 0.95. If one has additional *prior information* about p and $\overline{p}$, then the *conditional* probability of this event, relative to this information, may be very different. For instance, if one had, prior to the poll, a very good reason to believe that p is almost certainly between 0.4 and 0.6, and then the poll reports $\overline{p}$ to be 0.1, then the conditional probability that $\overline{p} - 0.03 \leq p \leq \overline{p} + 0.03$ occurs should

be lower[69] than the unconditional probability. The question of how to account for prior information is a very delicate one in *Bayesian probability*, and will not be discussed here.

One special case of the above point is worth emphasising: the statement that $\overline{p} - 0.03 \leq p \leq \overline{p} + 0.03$ is true with at least 95% probability is only valid *before* one actually conducts the poll and finds out the value of $\overline{p}$. Once $\overline{p}$ is computed, the statement $\overline{p} - 0.03 \leq p \leq \overline{p} + 0.03$ is either true or false, i.e. occurs with probability[70] 1 or 0 (unless one takes a Bayesian approach, as mentioned above).

1.13.2. Nobody asked for my opinion! One intuitive argument against a poll of small relative size being accurate goes something like this: a poll of just $1,000$ people among a population of $200,000,000$ is almost certainly not going to poll myself, or any of my friends or acquaintances. If the opinions of myself, and everyone that I know, is not being considered at all in this poll, how could this poll possibly be accurate?

It is true that if you know, say, $5,000$ voting-eligible people, then chances are that none of them (or maybe one of them, at best) will be contacted by the above poll. However, even though the opinions of all these people are not being directly polled, there will be many other people with *equivalent* opinions that *will* be contacted by the poll. Through those people, the views of yourself and your friends are being represented. (This may seem like a very weak form of representation, but recall that you and your 5,000 friends and acquaintances still only represent 0.0025% of the total electorate.)

Now one may argue that no two voters are identical, and that each voter arrives at a decision of who to vote for their own unique reasons. True enough - but recall that this poll is asking only a *simple* question: whether one is going to vote for A or not. Once one narrowly focuses on this question alone, any two voters who both

[69] Note though that having priori information just about p, and not $\overline{p}$, will not cause the probability to drop below 95%, as this bound on the confidence level is uniform in p.

[70] This phenomenon of course occurs all the time in probability. For instance, if x denotes the outcome of rolling a fair six-sided die, then before one performs this roll, the probability that x equals 1 will be 1/6, but after one has seen what the value of this die is, the probability that x equals 1 will be either 1 or 0.

decide to vote for A, or to not vote for A, are considered equivalent, even if they arrive at this decision for totally different reasons. So, for the purposes of this poll, there are only two types of voters in the world - A-voters, and not-A-voters - with all voters in one of these two types considered equivalent. In particular, any given voter is going to have millions of other equivalent voters distributed throughout the population X, and a representative fraction of those equivalent voters is likely to be picked up by the poll.

As mentioned before, polls which offer complex questions (for instance, trying to discern the motivation behind one's voting choices) will inherently be less accurate; there are now fewer equivalent voters for each individual, and it is harder for a poll to pick up each equivalence class in a representative manner[71].

1.13.3. Is there enough information? Another common objection to the accuracy of polls argues that there is not enough information (or "degrees of freedom") present in the poll sample to accurately describe the much larger amount of data present in the full population; $1,000$ bits of data cannot possibly contain $200,000,000$ bits of information. However, we are not asking to find out so much information; the purpose of the poll is to estimate just a *single* piece of information, namely the number p. If one is willing to accept an error of up to 3%, then one can represent this piece of information in about five bits rather than $200,000,000$. So, in principle at least, there is more than enough information present in the poll to recover this information; one does not need to sample the entire population to get a good reading[72].

As before, the accuracy degrades as one asks more and more complicated questions. For instance, if one were to poll $1,000$ voters for their opinions on two unrelated questions A and B, each of the answers to A and B would be accurate to within 3% with probability

[71] In particular, the more questions that are asked, the more likely it becomes that the responses to at least one of these questions will be inaccurate by an amount exceeding its margin of error. This provides a limit as to how much information one can confidently extract from data mining any given data set.

[72] The same general philosophy underlies *compressed sensing*, see Section 1.2 of *Structure and Randomness*, but that's another story.

95%, but the probability that the answers to A and B were simultaneously accurate to within 3% would be lower (around 90% or so), and so any data analysis that relies on the responses to both A and B may not have as high a confidence level as data analysis that relies on A and B separately. This is consistent with the information-theoretic perspective: we are demanding more and more bits of information on our population, and it is harder for our fixed data set to supply so much information accurately and confidently.

1.13.4. Swings. One intuitive way to gauge the margin of error of a poll is to see how likely such a poll is to accurately detect a swing in the electorate. Suppose for instance that over the course of a given time period (e.g. a week), 7% of the voters switch their vote from not-A to A, while another 2% of the voters switch their vote from A to not-A, leading to a net increase of 5% in the proportion p of voters voting for A. How does would this swing in the vote affect the proportion $\overline{p}$ of the voters being polled, if one imagines the same voters being polled at both the start of the week and at the end of the week? (Recall that we are assuming that voters will honestly report their change of mind from one poll to the next.)

If the poll was conducted by simple random sampling, then each of the $1,000$ voters polled would have a 7% probability of switching from not-A to A, and and a 2% probability of switching from A to not-A. Thus, one would expect about 70 of the $1,000$ voters polled to switch to A, and about 20 to switch to not-A, leading to a net swing of 50 voters, that would increase $\overline{p}$ by 5%, thus matching the increase in p. Now, in practice, there will be some variability here; due to the luck of the draw, the poll may pick up more or less than 70 of the voters switching to A, and more or less than 20 of the voters switching to not-A. But having $1,000$ voters to sample is just about large enough for the law of large numbers[73] (Section 1.5) to kick in and ensure that the number of voters switching to A picked up by the poll will be significantly larger than the number of voters switching to not-A. Thus, this poll will have a good chance of detecting a swing

[73] In appealing to the law of large numbers, we are implicitly exploiting the uniformity and independence assumptions in Hypothesis 5.

of size 5% or more, which is consistent with the assertion of a margin of error of about 3%.

It is worth noting that this swing of 5% in an electorate of $200,000,000$ voters represents quite a large shift in absolute terms: fourteen million voters switching to A and four million switching away from A. Quite a few of these shifting voters will be picked up by the poll (in contrast to one's sphere of friends and acquaintances, which is likely to be missed completely).

1.13.5. Irregularity. Another intuitive objection to polling accuracy is that the voting population is far from homogeneous. For instance, it is clear that voting preferences for the U.S. presidential election vary widely among the 50 states - shouldn't one need to multiply the poll size by 50 just to accomodate this fact? Similarly for distinctions in voting patterns based on gender, race, party affiliation, etc.

Again, these irregularities in voter distribution do not affect the final accuracy of the poll, for two reasons. Firstly, we are asking only the simple question of whether a voter votes for A or not-A, and are not breaking down the answers to this question by state, gender, race, or any other factor; as stated before, two voters are considered equivalent as long as they have the same preference for A, even if they are in different states, have different genders, etc. Secondly, while it is conceivable that the poll will cluster its sample in one particular state (or one particular gender, etc.), thus potentially skewing the poll, the fact that the voters are selected uniformly *and* independently of each other prevents this from happening very often[74].

The independence hypothesis is rather important. If for instance one were to poll by picking one particular location[75] in the U.S. at random, and polling $1,000$ people from that location, then the responses would be highly correlated (as one could have picked a location which

[74]And in any event, clustering in a demographic or geographic category is not what is of direct importance to the accuracy of the poll; the only thing that really matters in the end is whether there is clustering in the category of A-voters or not-A-voters.

[75]Incidentally, in the specific case of the U.S. presidential election, statewide polls are in fact more relevant to the outcome of the election than nationwide polls, due to the mechanics of the *U.S. Electoral College*, but this does not detract from the above points.

Figure 2. A low-resolution image of a U.S. president, from [**Ha1973**].

happens to highly favour A, or highly favour not-A) and would have a much larger margin of error than if one polled $1,000$ people at random across the U.S..

1.13.6. Analogies. Some analogies may help explain why the relative size of a sample is largely irrelevant to the accuracy of a poll.

Suppose one is in front of a large body of water (e.g. a sea or ocean), and wants to determine whether it is a freshwater or saltwater body. This can be done very easily: dip one's finger into the body of water and taste a single drop. This gives an extremely accurate result, even though the relative proportion of the sample size to the population size is, literally, a drop in the ocean; the quintillions of water molecues and salt molecues present in that drop are more than sufficient to give a good reading of the salinity[76] of the water body.

Another analogy comes from digital imaging. As we all know, a digital camera takes a picture of a real-world object (e.g. a human face) and converts it into an array of pixels; an image with a larger number of pixels will generally lead to a more accurate image than one with fewer. But even with just a handful of pixels, say 1,000 pixels, one is already able to make crude distinctions between different images, for instance to distinguish a light-skinned face from a dark-skinned face (despite the fact that skin colour is determined by millions of cells and quintillions of pigment molecues). See for instance the well-known image in Figure 2.

[76] To be fair, in order for this reading to be accurate, one needs to assume that the salinity is uniformly distributed across the body of water; if for instance the body happened to be nearly fresh on one side and much saltier on the other, then dipping one's finger in just one of these two sides would lead to an inaccurate measurement of average salinity. But if one were to stir the body of water vigorously, this irregularity of distribution disappears. The procedure of taking a random sample, with each sample point being independent of all the others, is analogous to this stirring procedure.

1.13.7. Appendix: Mathematical justification. One can compute the margin of error for this simple sampling problem very precisely using the *binomial distribution*; however I would like to present here a cruder but more robust estimate, based on the *second moment method*, that works in much greater generality than the setting discussed here. (It is closely related to the arguments in Section 1.5.) The main mathematical result we need is

Theorem 1.13.1. *Let X be a finite set, let A be a subset of X, and let $p := |A|/|X|$ be the proportion of elements of X that lie in A. Let $x_1, \ldots, x_n$ be sampled independently and uniformly at random from X (in particular, we allow repetitions). Let $\overline{p} := |\{1 \leq i \leq n : x_i \in A\}|/n$ be the proportion of the $x_1, \ldots, x_n$ (counting repetition) that lie in A. Then for any $r > 0$, one has*

$$\mathbf{P}(|\overline{p} - p| \leq r) \geq 1 - \frac{1}{4nr^2}. \tag{1.79}$$

Proof. We use the second moment method. For each $1 \leq i \leq n$, let I_i be the *indicator* of the event $x_i \in A$, thus $I_i := 1$ when $x_i \in A$ and $I_i = 0$ otherwise. Observe that each I_i has a probability of p of equaling 1, thus

$$p = \mathbf{E} I_i.$$

On the other hand, we have

$$\overline{p} = \frac{1}{n} \sum_{i=1}^{n} I_i.$$

Thus

$$\overline{p} - p = \frac{1}{n} \sum_{i=1}^{n} I_i - \mathbf{E}(I_i);$$

squaring this and taking expectations, we obtain

$$\mathbf{E}|\overline{p} - p|^2 = \frac{1}{n^2} \sum_{i=1}^{n} \mathbf{Var}(I_i) + \frac{2}{n} \sum_{1 \leq i < j \leq n} \mathbf{Cov}(I_i, I_j)$$

where $\mathbf{Var}(I_i) := \mathbf{E}(I_i - \mathbf{E}I_i)^2$ is *variance* of I_i, and $\mathbf{Cov}(I_i, I_j) := \mathbf{E}((I_i - p)(I_j - p))$ is the *covariance* of I_i, I_j.

By assumption, the random variable I_i, I_j for $i \neq j$ are independent, and so the covariances $\mathbf{Cov}(I_i, I_j)$ vanish. On the other hand,

a direct computation shows that

$$\mathbf{Var}(I_i) = p - p^2 = \frac{1}{4} - (p - \frac{1}{2})^2 \leq \frac{1}{4}$$

for each i. Putting all this together we conclude that

$$\mathbf{E}|\overline{p} - p|^2 \leq \frac{1}{4n}$$

and the claim (1.79) follows from *Markov's inequality*. □

Applying this theorem with $n = 1000$ and $r = 1/\sqrt{200} \approx 0.07$, we conclude that p and $\overline{p}$ lie within about 7% of each other with probability at least 95%, regardless of how large the population X is. In the context of an election poll, this means that if one samples 1000 voters independently at random (with replacement) whether they would vote for A, the margin of error for the answer would be at most 7% at the 95% confidence level.

Remark 1.13.2. Observe that the proof of the above theorem did not really need the x_i to be fully independent of each other; the key thing was that each x_i was close to uniformly distributed, and that the covariances between the indicators I_i, I_j were small. (In particular, one only needs *pairwise independence* rather than joint independence for the theorem to hold.) Because of this, one can also obtain variants of the above theorem when one selects $x_1, \ldots, x_n$ for random sampling without replacement (known as *simple random sampling*); now there is a slight correlation between I_i, I_j, but it turns out to be negligible when X is large, for instance[77] when $n = 1000$ and $|X| \sim 10^8$.

Remark 1.13.3. If one assumes joint independence instead of pairwise independence, one can obtain slightly sharper inequalities than (1.79) (e.g. by using the *Chernoff inequality*), but at the 95% confidence level, this gives a relatively modest improvement only in the margin of error (in our specific example, the optimal margin of error is about 3% rather than 7%).

Remark 1.13.4. An inspection of the argument shows that if p is known to be very small or very large, then the margin of error is

[77] For this range of parameters, there is a non-trivial probability of a *birthday paradox* occurring, so the two sampling methods are genuinely different from each other; but they turn out to have almost the same margin of error anyway.

better than what (1.79) predicts. (In the most extreme case, if $p = 0$ or $p = 1$, then it is easy to see that the margin of error is zero.) But in the case of election polls, p is generally expected to be close to $1/2$, and so one does not expect to be able to improve the margin of error much from this effect. And in any case, we don't know the value of p exactly in practice (otherwise why would we be doing the poll in the first place?).

Remark 1.13.5. In real world situations, it can be difficult or impractical to get the x_i to be close to uniformly distributed (because of *sampling bias*), and to keep the correlations low (because of effects such as *clustering*). Because of this, one often needs to perform a more complicated sampling procedure than simple random sampling, which requires more sophisticated statistical analysis than given by the above theorem. This is beyond the scope of this post, though.

Notes. This article first appeared at `terrytao.wordpress.com/2008/10/10`. Thanks to jonm and Kieran for corrections.

A calculator to compute margins of error for various sample sizes and population sizes can be found at `www.americanresearchgroup.com/moe.ht`

1.14. Non-measurable sets via non-standard analysis

In Section 1.4 of *Structure and Randomness*, I sketched out a non-rigorous probabilistic argument justifying the following well-known theorem:

Theorem 1.14.1 (Non-measurable sets exist)**.** *There exists a subset E of the unit interval $[0,1]$ which is not Lebesgue-measurable.*

The idea was to let E be a "random" subset of $[0,1]$. If one (non-rigorously) applies the law of large numbers (Section 1.5), one expects E to have "density" $1/2$ with respect to every subinterval of $[0,1]$, which would contradict the *Lebesgue differentiation theorem*.

I was recently asked whether I could in fact make the above argument rigorous. This turned out to be more difficult than I had anticipated, due to some technicalities in trying to make the concept of a random subset of $[0,1]$ (which requires an uncountable number of

"coin flips" to generate) both rigorous and useful. However, there is a simpler variant of the above argument which can be made rigorous. Instead of letting E be a "random" subset of $[0,1]$, one takes E to be an "alternating" set that contains "every other" real number in $[0,1]$; this again should have density $1/2$ in every subinterval and thus again contradict the Lebesgue differentiation theorem.

Of course, in the standard model of the real numbers, it makes no sense to talk about "every other" or "every second" real number, as the real numbers are not discrete. If however one employs the language of *non-standard analysis*, then it is possible to make the above argument rigorous, and this is the purpose of my post today. I will assume some basic familiarity with non-standard analysis, for instance as discussed in Section 1.5 of *Structure and Randomness*.

We begin by selecting a *non-principal ultrafilter* $p \in \beta\mathbf{N} \backslash \mathbf{N}$ and use it to construct non-standard models $*\mathbf{N}, *[0,1]$ of the natural numbers $\mathbf{N}$ and the unit interval $[0,1]$ by the usual *ultrapower* construction. We then let $N \in *\mathbf{N}$ be an unlimited non-standard number, i.e. a non-standard natural number larger than any standard natural number[78].

We can partition the non-standard unit interval $*[0,1]$ into 2^N (non-standard) intervals $J_j := [j/2^N, (j+1)/2^N]$ for $j = 0, \ldots, 2^N - 1$ (the overlap of these intervals will have a negligible impact in our analysis). We then define the non-standard set $*E \subset *[0,1]$ to be the union of those J_j with j odd; this is the formalisation of the idea of "every other real number" in the introduction. The key property about $*E$ that we need here is the following symmetry property: if $I = [a/2^n, (a+1)/2^n]$ is any standard *dyadic interval*, then the (non-standard counterpart to the) interval I can be partitioned (modulo (non-standard) rationals) as the set $*E \cap I$ and its reflection $(2a+1)/2^n - (*E \cap I)$. This demonstrates in particular that $*E$ has (non-standard) density $1/2$ inside I, but we will need to use the symmetry property directly, rather than the non-standard density property, because the former is much easier to transfer to the standard setting than the latter.

[78] For instance, one could take N to be the equivalence class in $*\mathbf{N}$ of the sequence $1, 2, 3, \ldots$.

We now return to the standard world, and introduce the *standard* set $E \subset [0,1]$, defined as the collection of all standard $x \in [0,1]$ whose non-standard representative $*x$ lies in $*E$. This is certainly a standard subset of $[0,1]$; informally, it is the set of all standard numbers whose N^{th} digit is 1. We claim that it is not Lebesgue measurable, thus establishing Theorem 1.14.1. To see this, recall for any standard dyadic interval $I = [a/2^n, (a+1)/2^n]$, that every non-standard irrational element of I lies in exactly one of $*E$ or $(2a+1)/2^n - *E$. Applying the transfer principle, we conclude that every standard irrational element of I lies in exactly one of E or $(2a+1)/2^n - E$. Thus, if E were measurable, the density of E in the dyadic interval I must be exactly $1/2$. But this contradicts the *Lebesgue differentiation theorem* (see e.g. Section 1.4 of *Structure and Randomness*), and we are done.

Remark 1.14.2. One can eliminate the non-standard analysis from this argument and rely directly on the non-principal ultrafilter p. Indeed, if one inspects the ultrapower construction carefully, one sees that (outside of the terminating binary rationals, which do not have a unique binary expansion), E consists of those numbers in $[0,1]$ whose 1s in the binary expansion lie in p. The symmetry property of E then reflects the non-principal ultrafilter nature of p, in particular the fact that membership of a set $A \subset \mathbf{N}$ in p is insensitive to any finite modification of A, and reversed by replacing A with its complement. On the other hand, one cannot eliminate the non-principal ultrafilter entirely; once one drops the axiom of choice, there exist models of the real line in which every set is Lebesgue measurable, and so it is necessary to have *some* step in the proof of Theorem 1.14.1 that involves this axiom. In the above argument, choice is used to find the non-principal ultrafilter p.

Remark 1.14.3. It is tempting to also make the original argument, based on randomness, work, but I was unable to push it through completely. Certainly, if one sets $*E$ to be a (non-standardly) random collection of the J_j, then with probability infinitesimally close to 1, the (non-standard) density of $*E$ in any standard dyadic interval (or indeed, any standard interval) is infinitesimally close to $1/2$, thanks to the law of large numbers. However, I was not able to transfer

this fact to tell me anything about E; indeed, I could not even show that E was non-empty. (Question: does there exist a non-standard subset of $*[0,1]$ of positive (non-infinitesimal) measure which avoids all standard real numbers? I don't know the answer to this.)

Notes. This article first appeared at `terrytao.wordpress.com/2008/10/14`. Thanks to Timothy Gowers, Wonghang, and an anonymous commenter for corrections.

K.P. Hart and Kevin O'Bryant noted that this construction is quite classical, going back to [**Ul1929**], [**Ta**]. K.P. Hart also noted that in the product space $2^{\mathbf{R}}$, the set of measurable and non-measurable sets both have outer measure 1 with respect to product measure, so a naive attempt to formalise the statement that "a randomly chosen set is non-measurable" does not work.

1.15. When are eigenvalues stable?

I was asked recently (in relation to my recent work [**TaVu2008**] with Van Vu) to explain some standard heuristics regarding how the eigenvalues $\lambda_1, \ldots, \lambda_n$ of an $n \times n$ matrix A behave under small perturbations. These heuristics can be summarised as follows:

(1) For *normal matrices* (and in particular, *unitary* or *self-adjoint* matrices), eigenvalues are very stable under small perturbations. For more general matrices, eigenvalues can become unstable if there is *pseudospectrum* present.

(2) For self-adjoint (Hermitian) matrices, eigenvalues that are too close together tend to repel quickly from each other under such perturbations. For more general matrices, eigenvalues can either be repelled or be attracted to each other, depending on the type of perturbation.

In this article, I would like to briefly explain why these heuristics are plausible.

1.15.1. Pseudospectrum. As any student of linear algebra knows, the *spectrum* $\sigma(A)$ of a $n \times n$ matrix A consists of all the complex numbers λ such that $A - \lambda I$ fails to be invertible, thus there exists a non-zero vector v such that $(A - \lambda I)v$ is zero. This is a finite set,

consisting of at most n points. But there is also an important set containing the spectrum (and which, at times, can be much larger than that spectrum), which is the *pseudospectrum* of the matrix A. Unlike the spectrum, which is canonically defined, there is more than one definition of the pseudospectrum; also, this concept depends on an additional parameter, a small number $\varepsilon > 0$. We will define[79] the pseudospectrum (or more precisely, the ε-pseudospectrum) $\sigma_\varepsilon(A)$ to be the set of all the complex numbers λ such that $A - \lambda I$ has least singular value at most ε, or equivalently that there exists a unit vector v such that $|(A - \lambda I)v| \leq \varepsilon$.

The significance of the pseudospectrum $\sigma_\varepsilon(A)$ is that *it describes where the spectrum $\sigma(A)$ can go to under small perturbations.* Indeed, if λ lies in the pseudospectrum $\sigma_\varepsilon(A)$, so that there exists a unit vector v whose image $w := (A - \lambda I)v$ has magnitude at most ε, then we see that

$$(A - wv^* - \lambda I)v = (A - \lambda I)v - wv^*v = w - w = 0$$

and so λ lies in the spectrum $\sigma(A - wv^*)$ of the perturbation $A - wv^*$ of A. Note that the operator norm of wv^* is at most ε.

Conversely, if λ does *not* lie in the pseudospectrum $\sigma_\varepsilon(A)$, and $A + E$ is a small perturbation of A (with E having operator norm at most ε), then for any unit vector v, one has

$$|(A + E - \lambda I)v| \geq |(A - \lambda I)v| - |Ev| > \varepsilon - \varepsilon = 0$$

by the triangle inequality, and so λ cannot lie in the spectrum of $A + E$.

Thus, if the pseudospectrum is tightly clustered around the spectrum, the spectrum is stable under small perturbations; but if the pseudospectrum is widely dispersed, then the spectrum becomes unstable.

No matter what A is, the pseudospectrum $\sigma_\varepsilon(A)$ always contains the ε-neighbourhood of the spectrum $\sigma(A)$. Indeed, if v is a unit

[79] Another equivalent definition is that the ε-pseudospectrum consists of the spectrum, together with those complex numbers λ for which the *resolvent* $(A - \lambda)^{-1}$ has *operator* norm at least $1/\varepsilon$.

eigenvector with eigenvalue $\lambda \in \sigma(A)$, then $(A - \lambda I)v = 0$, which implies that $|(A-\lambda' I)v| = |\lambda-\lambda'| \leq \varepsilon$ for any λ' in the ε-neighbourhood of λ, and the claim follows.

Conversely, when A is normal, $\sigma_\varepsilon(A)$ consists *only* of the ε-neighbourhood of $\sigma(A)$. This is easiest to see by using the *spectral theorem* to diagonalise A and then computing everything explicitly. In particular, we conclude that if we perturb a normal matrix by a (possibly non-normal) perturbation of operator norm at most ε, then the spectrum moves by at most ε.

In the non-normal case, things can be quite different. A good example is provided by the shift matrix

$$U := \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

This matrix is *nilpotent*: $U^n = 0$. As such, the only eigenvalue is zero. But observe that for any complex number λ,

$$(U - \lambda I) \begin{pmatrix} 1 \\ \lambda \\ \dots \\ \lambda^{n-1} \\ \lambda^n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ -\lambda^{n+1} \end{pmatrix}.$$

From this and a little computation, we see that if $|\lambda| < 1$, then λ will lie in the $O\left(\frac{|\lambda|^{n+1}}{1-|\lambda|}\right)$-pseudospectrum of U. For fixed ε, we thus see that $\sigma_\varepsilon(U)$ fills up the unit disk in the high dimensional limit $n \to \infty$. (The pseudospectrum will not venture far beyond the unit disk, as the operator norm of U is 1.) And indeed, it is easy to perturb U so that its spectrum moves far away from the origin. For instance,

observe that the perturbation

$$U + E := \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & \\ 0 & 0 & 1 & \dots & 0 & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ & 0 & 0 & 0 & \dots & 1 \\ \varepsilon & 0 & 0 & \dots & 0 & \end{pmatrix}$$

of U has a *characteristic polynomial* $\det(U + E - \lambda I)$ equal[80] to $(-\lambda)^n - (-1)^n\varepsilon$ and so has eigenvalues equal to the n^{th} roots of ε; for fixed ε and n tending to infinity, this spectrum becomes asymptotically uniformly distributed on the unit circle, rather than at the origin.

Remark 1.15.1. Much more on the theory of pseudospectra can be found at `http://www.comlab.ox.ac.uk/pseudospectra/`; thanks to Nick Trefethen for the reference.

1.15.2. Spectral dynamics. The pseudospectrum tells us, roughly speaking, *how far* the spectrum $\sigma(A)$ of a matrix A can move with respect to a small perturbation, but does not tell us the *direction* in which the spectrum moves. For this, it is convenient to use the language of calculus: we suppose that $A = A(t)$ varies smoothly with respect to some time parameter t, and would like to "differentiate" the spectrum $\sigma(A)$ with respect to t. Since it is a little unclear what it means to differentiate a set, let us work instead with the eigenvalues $\lambda_j = \lambda_j(t)$ of $A = A(t)$. Note that generically (e.g. for almost all A), the eigenvalues will be distinct[81]. So it is not unreasonable to assume that for all t in some open interval, the $\lambda_j(t)$ are distinct; an application of the implicit function theorem then allows one to make the $\lambda_j(t)$ smooth in t. Similarly, we can make the eigenvectors[82] $v_j = v_j(t)$ vary smoothly in t.

[80] Alternatively, one can simply observe that $U^n = \varepsilon I$; this fact is of course related to the formula for the characteristic polynomial via the *Cayley-Hamilton theorem*.

[81] Proof: the eigenvalues are distinct when the characteristic polynomial has no repeated roots, or equivalently when the resultant of the characteristic polynomial with its derivative is non-zero. This is clearly a *Zariski-open* condition; since the condition is obeyed at least once, it is thus Zariski-dense.

[82] There is some freedom to multiply each eigenvector by a scalar, but this freedom will cancel itself out in the end, as we are ultimately interested only in the eigenvalues rather than the eigenvectors.

The eigenvectors $v_1, \ldots, v_n$ form a basis of $\mathbf{C}^n$. Let $w_1, \ldots, w_n$ be the *dual basis*, thus $w_j^* v_k = \delta_{jk}$ for all $1 \leq j, k \leq n$, and so we have the reproducing formula

$$u = \sum_{j=1}^{n} (w_j^* u) v_j \tag{1.80}$$

for all vectors u. Combining this with the eigenvector equations

$$A v_k = \lambda_k v_k \tag{1.81}$$

we obtain the adjoint eigenvector equations

$$w_k^* A = \lambda_k w_k. \tag{1.82}$$

Next, we differentiate (1.81) using the product rule to obtain

$$\dot{A} v_k + A \dot{v}_k = \dot{\lambda}_k v_k + \lambda_k \dot{v}_k. \tag{1.83}$$

Taking the inner product of this with the dual vector w_k, and using (1.82) to cancel some terms, we obtain the *first variation formula* for eigenvalues:

$$\dot{\lambda}_k = w_k^* \dot{A} v_k. \tag{1.84}$$

Note that if A is normal, then we can take the eigenbasis v_k to be orthonormal, in which case the dual basis w_k is identical to v_k. In particular we see that $|\dot{\lambda}_k| \leq \|\dot{A}\|_{op}$; the infinitesimal change of each eigenvalue does not exceed the infinitesimal size of the perturbation. This is consistent with the stability of the spectrum for normal operators mentioned in the previous section.

Remark 1.15.2. If A evolves by the *Lax pair equation* $\dot{A} = [A, P]$ for some matrix $P = P(t)$, then $w_k^* \dot{A} v_k = w_k^* \lambda \mathrm{P} v_k - w_k^* P \lambda v_k = 0$, and so from (1.84) we see that the spectrum of A is invariant in time. This fact underpins the *inverse scattering method* for solving *integrable systems*, which we will not discuss here.

Now we look at how the eigen*vectors* vary. Taking the inner product instead with a dual vector w_j for $j \neq k$, we obtain

$$w_j^* \dot{A} v_k + (\lambda_j - \lambda_k) w_j^* \dot{v}_k = 0;$$

applying (1.80) we conclude a first variation formula for the eigenvectors v_k, namely that

$$\dot{v}_k = \sum_{j \neq k} \frac{w_j^* \dot{A} v_k}{\lambda_k - \lambda_j} v_j + c_k v_k \tag{1.85}$$

for some scalar c_k (the presence of this term reflects the freedom to multiply c_k by a scalar). Similar considerations for the adjoint give

$$\dot{w}_k = \sum_{j \neq k} \frac{w_k^* \dot{A} v_j}{\lambda_k - \lambda_j} w_j - c_k w_k \tag{1.86}$$

(here we use the derivative of the identity $w_k^* v_k = 1$ to get the correct multiple of w_k on the right-hand side). We can use (1.84), (1.85), (1.86) to obtain a second variation formula for the eigenvalues. Indeed, by differentiating (1.84) we obtain

$$\ddot{\lambda}_k = \dot{w}_k^* \dot{A} v_k + w_k^* \ddot{A} v_k + w_k^* \dot{A} \dot{v}_k;$$

applying (1.85), (1.86) we conclude the second variation formula

$$\ddot{\lambda}_k = w_k^* \ddot{A} v_k + 2 \sum_{j \neq k} \frac{(w_k^* \dot{A} v_j)(w_j^* \dot{A} v_k)}{\lambda_k - \lambda_j}. \tag{1.87}$$

Now suppose that A is self-adjoint, so as before we can take $v_k = w_k$ to be orthonormal. The above formula then becomes

$$\ddot{\lambda}_k = v_k^* \ddot{A} v_k + 2 \sum_{j \neq k} \frac{|v_k^* \dot{A} v_j|^2}{\lambda_k - \lambda_j}.$$

One can view the terms on the right-hand side here as various "forces" acting on the eigenvalue λ_k; the acceleration of the original matrix A provides one such force, while all the other eigenvalues λ_j provide a repulsive force[83]. The closer two eigenvalues are to each other, the stronger the repulsive force becomes.

When the matrix A is not self-adjoint, then the interaction between λ_j and λ_k can be either repulsive or attractive. Consider for instance the matrices

$$\begin{pmatrix} 1 & t \\ -t & -1 \end{pmatrix}.$$

[83] As with *Newton's third law*, the force that λ_j exerts on λ_k is equal and opposite to the force that λ_k exerts on λ_j; note that this is consistent with the trace formula $\sum_{k=1}^n \lambda_k = \mathrm{tr}(A)$.

The eigenvalues of this matrix are $\pm\sqrt{1-t^2}$ for $-1 < t < 1$ - so we see that they are attracted to each other as t evolves, until the matrix becomes degenerate at $t = \pm 1$. In contrast, the self-adjoint matrices

$$\begin{pmatrix} 1 & t \\ t & -1 \end{pmatrix}.$$

have eigenvalues $\pm\sqrt{1+t^2}$, which repel each other as t evolves.

The repulsion effect of eigenvalues is also consistent with the smallness of the set of matrices with repeated eigenvalues. Consider for instance the space of Hermitian $n \times n$ matrices, which has real dimension n^2. The subset of Hermitian matrices with distinct eigenvalues can be described by a collection of n orthogonal (complex) one-dimensional eigenspaces (which can be computed to have $2(n-1) + 2(n-2) + \ldots + 2 = n(n-1)$ degrees of freedom) plus n real eigenvalues (for an additional n degrees of freedom), thus the set of matrices with distinct eigenvalues has full dimension n^2.

Now consider the space of matrices with one repeated eigenvalue. This can be described by $n-2$ orthogonal complex one-dimensional eigenspaces, plus a complex two-dimensional orthogonal complement (which has $2(n-1) + 2(n-2) + \ldots + 4 = n(n-1) - 2$ degrees of freedom) plus $n-1$ real eigenvalues, thus the set of matrices with repeated eigenvalues only has dimension $n^2 - 3$. Thus it is in fact very rare for eigenvalues to actually collide, which helps explain why there must be a repulsion effect in the first place.

An example can help illustrate this phenomenon. Consider the one-parameter family of Hermitian matrices

$$\begin{pmatrix} t & 0 \\ 0 & -t \end{pmatrix}.$$

The eigenvalues of this matrix at time t are of course t and $-t$, which cross over each other when t changes sign. Now consider instead the Hermitian perturbation

$$\begin{pmatrix} t & \varepsilon \\ \varepsilon & -t \end{pmatrix}$$

for some small $\varepsilon > 0$. The eigenvalues are now $\sqrt{t^2+\varepsilon^2}$ and $-\sqrt{t^2+\varepsilon^2}$; they come close to each other as t approaches 0, but then "bounce" off of each other due to the repulsion effect.

Notes. This article first appeared at `terrytao.wordpress.com/2008/10/28`. Thanks to orr for corrections.

1.16. Concentration compactness and the profile decomposition

One of the most important topological concepts in analysis is that of *compactness*. There are various flavours of this concept, but let us focus on *sequential compactness*: a subset E of a topological space X is sequentially compact if every sequence in E has a convergent subsequence whose limit is also in E. This property allows one to do many things with the set E. For instance, it allows one to maximise a functional on E:

Proposition 1.16.1 (Existence of extremisers)**.** *Let E be a non-empty sequentially compact subset of a topological space X, and let $F : E \to \mathbf{R}$ be a continuous function. Then the supremum $\sup_{x\in E} f(x)$ is attained at at least one point $x_* \in E$, thus $F(x) \leq F(x_*)$ for all $x \in E$. (In particular, this supremum is finite.) Similarly for the infimum.*

Proof. Let $-\infty < L \leq +\infty$ be the supremum $L := \sup_{x\in E} F(x)$. By the definition of supremum (and the *axiom of (countable) choice*), one can find a sequence $x^{(n)}$ in E such that $F(x^{(n)}) \to L$. By compactness, we can refine this sequence to a subsequence (which, by abuse of notation, we shall continue to call $x^{(n)}$) such that $x^{(n)}$ converges to a limit x in E. Since we still have $f(x^{(n)}) \to L$, and f is continuous at x, we conclude that $f(x) = L$, and the claim for the supremum follows. The claim for the infimum is similar. □

Remark 1.16.2. An inspection of the argument shows that one can relax the continuity hypothesis on F somewhat: to attain the supremum, it suffices that F be *upper semicontinuous*, and to attain the infimum, it suffices that F be *lower semicontinuous*.

We thus see that sequential compactness is useful, among other things, for ensuring the existence of extremisers. In finite-dimensional spaces (such as $\mathbf{R}^n$), compact sets are plentiful; indeed, the *Heine-Borel theorem* asserts that every closed and bounded set is compact.

However, once one moves to infinite-dimensional spaces, such as *function spaces*, then the Heine-Borel theorem fails quite dramatically; most of the closed and bounded sets one encounters in a topological vector space are non-compact, if one insists on using a reasonably "strong" topology. This causes a difficulty in (among other things) *calculus of variations*, which is often concerned to finding extremisers to a functional $F : E \to \mathbf{R}$ on a subset E of an infinite-dimensional function space X.

In recent decades, mathematicians have found a number of ways to get around this difficulty. One of them is to weaken the topology to recover compactness, taking advantage of such results as the *Banach-Alaoglu theorem* (or its sequential counterpart). Of course, there is a tradeoff: weakening the topology makes compactness easier to attain, but makes the continuity of F harder to establish. Nevertheless, if F enjoys enough "smoothing" or "cancellation" properties, one can hope to obtain continuity in the weak topology, allowing one to do things such as locate extremisers[84].

Another option is to abandon trying to make *all* sequences have convergent subsequences, and settle just for extremising sequences to have convergent subsequences, as this would still be enough to retain Proposition 1.16.1. Pursuing this line of thought leads to the *Palais-Smale condition*, which is a substitute for compactness in some calculus of variations situations.

But in many situations, one cannot weaken the topology to the point where the domain E becomes compact, without destroying the continuity (or semi-continuity) of F, though one can often at least find an intermediate topology (or metric) in which F is continuous, but for which E is still not quite compact. Thus one can find[85] sequences $x^{(n)}$ in E which do not have any subsequences that converge to a constant element $x \in E$, even in this intermediate metric. Because of this, it is *a priori* conceivable that a continuous function F need not attain its supremum or infimum.

[84] The phenomenon that cancellation can lead to continuity in the weak topology is sometimes referred to as *compensated compactness*.

[85] As we shall see shortly, one major cause of this failure of compactness is the existence of a non-trivial action of a non-compact group G on E; such a group action can cause compensated compactness or the Palais-Smale condition to fail also.

Nevertheless, even though a sequence $x^{(n)}$ does not have any subsequences that converge to a constant x, it may have a subsequence (which we also call $x^{(n)}$) which converges to some non-constant sequence $y^{(n)}$ (in the sense that the distance $d(x^{(n)}, y^{(n)})$ between the subsequence and the new sequence in a this intermediate metric), where the approximating sequence $y^{(n)}$ is of a very structured form (e.g. "concentrating" to a point, or "travelling" off to infinity, or a superposition $y^{(n)} = \sum_j y_j^{(n)}$ of several concentrating or travelling *profiles* of this form). This weaker form of compactness, in which superpositions of a certain type of profile completely describe all the failures (or *defects*) of compactness, is known as *concentration compactness*, and the decomposition $x^{(n)} \approx \sum_j y_j^{(n)}$ of the subsequence is known as the *profile decomposition*. In many applications, it is a sufficiently good substitute for compactness that one can still do things like locate extremisers for functionals F - though one often has to make some additional assumptions of F to compensate for the more complicated nature of the compactness. This phenomenon was systematically studied by P.L. Lions in the 80s, and found great application in calculus of variations and nonlinear elliptic PDE. More recently, concentration compactness has been a crucial and powerful tool in the non-perturbative analysis of nonlinear *dispersive* PDE, in particular being used to locate "minimal energy blowup solutions" or "minimal mass blowup solutions" for such a PDE (analogously to how one can use calculus of variations to find minimal energy solutions to a nonlinear elliptic equation); see for instance [**KiVi2008**].

In typical applications, the concentration compactness phenomenon is exploited in moderately sophisticated function spaces (such as *Sobolev spaces* or *Strichartz spaces*), with the failure of traditional compactness being connected to a moderately complicated group G of symmetries (e.g. the group generated by translations and dilations). Because of this, concentration compactness can appear to be a rather complicated and technical concept when it is first encountered. In this note, I would like to illustrate concentration compactness in a simple toy setting, namely in the space $X = l^1(\mathbf{Z})$ of absolutely summable sequences, with the uniform (l^∞) metric playing the role of the intermediate metric, and the translation group $\mathbf{Z}$ playing the role

of the symmetry group G. This toy setting is significantly simpler than any model that one would actually use in practice [for instance, in most applications X is a *Hilbert space*], but hopefully it serves to illuminate this useful concept in a less technical fashion.

1.16.1. Defects of compactness in $l^1(\mathbf{Z})$. Consider the space

$$X := l^1(\mathbf{Z}) := \{(x_m)_{m\in\mathbf{Z}} : \sum_{m\in\mathbf{Z}} |x_m| < \infty\}$$

of absolutely summable doubly infinite sequences $x = (x_m)_{m\in\mathbf{Z}}$; this is a normed vector space generated by the basis vectors $e_n := (\delta_{n,m})_{m\in\mathbf{Z}}$ for $n \in \mathbf{Z}$ (here δ is the Kronecker delta). We can place several topologies on this space X:

Definition 1.16.3. Let $x^{(n)} = (x_m^{(n)})_{m\in\mathbf{Z}}$ be a sequence in X (i.e. a sequence of sequences!), and let $x = (x_m)_{m\in\mathbf{Z}}$be another element in X.

(1) (Strong topology) We say that $x^{(n)}$ converges to x in the *strong topology* (or l^1 *topology*) if the l^1 distance $\|x^{(n)} - x\|_{l^1(\mathbf{Z})} := \sum_{m\in\mathbf{Z}} |x_m^{(n)} - x_m|$ converges to zero.

(2) (Intermediate topology) We say that $x^{(n)}$ converges in x in the *intermediate topology* (or *uniform topology*) if the l^∞ distance $\|x^{(n)} - x\|_{l^\infty(\mathbf{Z})} := \sup_{m\in\mathbf{Z}} |x_m^{(n)} - x_m|$ converges to zero.

(3) (Weak topology) We say that $x^{(n)}$ converges[86] in x in the *weak topology* (or *pointwise topology*) if $x_m^{(n)} \to x_m$ as $n \to \infty$ for each m.

Example 1.16.4. The sequence e_n for $n = 1, 2, \ldots$ converges weakly to zero, but is not convergent in the strong or intermediate topologies. The sequence $\frac{1}{n}\sum_{n'=1}^n e_{n'}$ converges in the intermediate and weak topologies to zero, but is not convergent in the strong topology.

It is easy to see that strong convergence implies intermediate convergence, which in turn implies weak convergence, thus justifying the names "strong", "intermediate", and "weak". For bounded

[86] Strictly speaking, this only describes the weak topology for *bounded* sequences, but these are the only sequences we will be considering here.

sequences, the intermediate topology can also be described by a number of other norms, e.g. the $l^p(\mathbf{Z})$ norm for any $p > 1$ (this is an easy application of *Hölder's inequality*).

The space X also has the *translation action* of the group of integers $G := \mathbf{Z}$, defined using the shift operators $T^h : X \to X$ for $h \in G$, defined by the formula

$$T^h(x_m)_{m\in\mathbf{Z}} := (x_{m-h})_{m\in\mathbf{Z}}$$

(in particular, T^h is linear with $T^h e_n = e_{n+h}$). This action is continuous with respect to all three of the above topologies. (We give G the discrete topology.)

Inside the infinite-dimensional space X, we let E be the "unit sphere" (though it looks more like an octahedron, actually)

$$E := \{(x_m)_{m\in\mathbf{Z}} \in X : \sum_{m\in\mathbf{Z}} |x_m| = 1\}.$$

E is clearly invariant under the translation action of G. It is easy to see that E is closed and bounded in the strong topology (or metric). However, it is not closed in the weak topology: the sequence $e_n \in E$ of basis vectors for $n = 1, 2, \ldots$ converges weakly to the origin 0, which lies outside of E. It is also not closed in the intermediate topology; the sequence $\frac{1}{n}\sum_{n'=1}^{n} e_{n'}$ lies in E but converges in the intermediate topology to 0, which lies outside of E.

The failure of closure in the weak topology causes failure of compactness in the strong or intermediate topologies. Indeed, the sequence $e_n \in E$ cannot have any convergent subsequence in those topologies, since the limit of such a subsequence would have to equal its weak limit, which is zero; but e_n clearly does not converge in either the strong or intermediate topologies to 0. (To put it another way, the embedding of $l^1(\mathbf{Z})$ into $l^\infty(\mathbf{Z})$ is not *compact*.)

More generally, for any fixed *profile* $x \in E$, the "travelling wave" (or "travelling profile") $T^n x \in E$ for $n = 1, 2, \ldots$ converges weakly to zero, and so by the above argument has no convergent subsequence in the strong or intermediate topologies. A little more generally still, given any sequence $h^{(n)}$ of integers going off to infinity, $T^{h^{(n)}} x \in E$ is a sequence in E which has no convergent subsequence in the strong

or intermediate topologies. Thus we see that the action of the (non-compact) group G is causing a failure of compactness of E in the strong and intermediate topologies.

Because of the linear nature of the vector space X, one can also create examples of sequences in E with no convergent subsequences by taking superpositions of travelling profiles. For instance, if $x_1, x_2 \in X$ are two non-negative sequences with $\|x_1\|_{l^1(\mathbf{Z})} + \|x_2\|_{l^1(\mathbf{Z})} = 1$, and $h_1^{(n)}, h_2^{(n)}$ are two sequences of integers which both go off to infinity,

$$|h_1^{(n)}|, |h_2^{(n)}| \to \infty$$

then the superposition

$$x^{(n)} := T^{h_1^{(n)}} x_1 + T^{h_2^{(n)}} x_2$$

of the two travelling profiles $T^{h_1^{(n)}} x_1$ and $T^{h_2^{(n)}} x_2$ will be a sequence in E that continues to converge weakly to zero, and so again has no convergent subsequence in the strong or intermediate topologies.

If x_1 and x_2 are not non-negative, then there can be cancellations between $T^{h_1^{(n)}} x_1$ and $T^{h_2^{(n)}} x_2$, which could cause $x^{(n)}$ to have norm significantly less than 1 (thus straying away from E). However, if one also imposes the *asymptotic orthogonality* condition

$$|h_2^{(n)} - h_1^{(n)}| \to \infty$$

we see that these cancellations vanish in the limit $n \to \infty$, and so in this case[87] we can build a modified superposition

$$x^{(n)} := T^{h_1^{(n)}} x_1 + T^{h_2^{(n)}} x_2 + w^{(n)}$$

that lies in E, with $w^{(n)}$ converging to zero in the strong and uniform topology, and will once again be a sequence with no convergent subsequence. More generally, given any collection x_j of non-zero elements of X with

$$\sum_j \|x_j\|_{l^1(\mathbf{Z})} \leq 1 \tag{1.88}$$

[87] If the asymptotic orthogonality condition fails, then one can collapse the superposition of two travelling profiles into a single travelling profile, after passing to a subsequence if necessary. Indeed, if $|h_2^{(n)} - h_1^{(n)}|$ does not go to infinity, then we can find a subsequence for which $h_2^{(n)} - h_1^{(n)}$ is equal to a constant c, in which case $T^{h_1^{(n)}} f_1 + T^{h_2^{(n)}} f_2$ is equal to a single travelling profile $T^{h_1^{(n)}}(f_1 + T^c f_2)$.

and any sequences $h_j^{(n)}$ of integers obeying the asymptotic orthogonality condition

$$|h_{j'}^{(n)} - h_j^{(n)}| \to \infty \text{ as} n \to \infty \tag{1.89}$$

for all $j' > j$, we can find a sequence in $x^{(n)}$ that takes the form

$$x^{(n)} = \sum_j T^{h_j^{(n)}} x_j + w^{(n)} \tag{1.90}$$

where $w^{(n)}$ converges to zero in the intermediate topology[88]. If $h_j^{(n)}$ goes off to infinity for at least one j with x_j non-zero, then this sequence will have no convergent subsequence.

We have thus demonstrated a large number of ways that compactness of E fails in the strong and intermediate topologies. The concentration compactness phenomenon, in this setting, tells us that these are essentially the *only* ways in which compactness fails in the intermediate topology. More precisely, one has

Theorem 1.16.5 (Profile decomposition)**.** *Let $x^{(n)}$ be a sequence in E. Then, after passing to a subsequence (which we still call $x^{(n)}$), there exist $x_j \in X$ obeying* (1.88)*, and sequences $h_j^{(n)}$ of integers obeying* (1.89)*, such that we have the decomposition* (1.90) *where the error $w^{(n)}$ converges to zero in the intermediate topology. Furthermore, we can improve* (1.88) *to*

$$\sum_j \|x_j\|_{l^1(\mathbf{Z})} + \lim_{n\to\infty} \|w^{(n)}\|_{l^1(\mathbf{Z})} \leq 1 \tag{1.91}$$

Remark 1.16.6. The situation is vastly different in the strong topology; in this case, virtually every sequence in E fails to have a convergent subsequence (consider for instance the sequence $\frac{1}{n}\sum_{n'=1}^n e_{n'}$ from Example 1.16.4), and there are so many different ways a sequence can behave that there is no meaningful profile decomposition. A more quantitative way to see this is via a computation of metric entropy constants (i.e. covering numbers). Pick a small number $\varepsilon > 0$ (e.g. $\varepsilon = 0.1$) and a large number N, and consider how many balls of radius ε in the $l^1(\{1,\ldots,N\})$ norm are needed to cover the

[88] If one has equality in (1.88), one can make $w^{(n)}$ converge in the strong topology also.

unit sphere E_N in $l^1(\{1,\ldots,N\})$. A simple volume packing argument shows that this number must grow exponentially in N. On the other hand, if one wants to cover E_N with the (much larger) balls of radius ε in the $l^\infty(\{1,\ldots,N\})$ topology instead, the number of balls needed grows only polynomially with N. Indeed, after rounding down each coefficient of an element of $l^1(\{1,\ldots,N\})$ to a multiple of ε, there are only at most $1/\varepsilon$ non-zero coefficients, and so the total number of possibilities for this rounded down approximant is about $(n/\varepsilon)^{1/\varepsilon}$. Thus, the metric entropy constants for both the strong and intermediate topologies go to infinity in the infinite dimensional limit $N \to \infty$ (thus demonstrating the lack of compactness for both), but much more rapidly for the former than for the latter.

1.16.2. Proof sketch of Theorem 1.16.5. We now sketch how one would prove Theorem 1.16.5. The idea is to hunt down and "domesticate"[89] the large values of $x^{(n)}$, as these are the only obstructions to convergence in the intermediate topology. Each large piece of the $x^{(n)}$ that we capture in this manner will decrease the total "mass" in play, which guarantees that eventually one runs out of such large pieces, at which point[90] one obtains the decomposition (1.90). In this process we rely heavily on the freedom to pass to a subsequence at will, which is useful to eliminate any fluctuations so long as they range over a compact space of possibilities.

Let's see how this procedure works. We begin with our bounded sequence $x^{(n)}$, whose l^1 norms are all equal to 1. If this sequence already converging to zero in the intermediate topology, we are done (we let j range over the empty set, and set $w^{(n)}$ equal to all of $x^{(n)}$. So suppose that $x^{(n)}$ are not converging to zero in this topology. Passing to a subsequence if necessary, this implies the existence of an $\varepsilon_1 > 0$ such that $\|x^{(n)}\|_{l^\infty(\mathbf{Z})} > \varepsilon_1$ for all n. Thus we can find integers $h_1^{(n)}$ such that $|x^{(n)}_{h_1^{(n)}}| > \varepsilon_1$ for all n, or equivalently that the shifts $T^{-h_1^{(n)}} x^{(n)}$ have their zero coefficient uniformly bounded below in magnitude by ε_1.

[89] I thank Kyril Tintarev for this terminology.

[90] Curiously, the strategy here is very similar to that underlying the structural theorems that arise in additive combinatorics and ergodic theory; see Section 2.1 of *Structure and Randomness*.

We have used the symmetry group G to move a large component of each of the $x^{(n)}$ the origin. Now we take advantage of sequential compactness of the unit ball in the *weak* topology. This allows one (after passing to another subsequence) to assume that the shifted elements $T^{-h_1^{(n)}} x^{(n)}$ converge weakly to some limit x_1; since the $T^{-h_1^{(n)}} x^{(n)}$ are uniformly non-trivial at the origin, the weak limit x_1 is also; in particular, we have $\|x_1\|_{l^1(\mathbf{Z})} \geq \varepsilon_1 > 0$. Undoing the shift, we have obtained a decomposition

$$x^{(n)} = T^{h_1^{(n)}} x_1 + w_1^{(n)}$$

where the *residual* $w_1^{(n)}$ is such that $T^{-h_1^{(n)}} w_1^{(n)}$ converges weakly to zero (thus, in some sense $w_1^{(n)}$ vanishes asymptotically near $h_1^{(n)}$). It is then not difficult to show the "asymptotic orthogonality" relationship

$$\|x^{(n)}\|_{l^1(\mathbf{Z})} = \|T^{h_1^{(n)}} x_1\|_{l^1(\mathbf{Z})} + \|w_1^{(n)}\|_{l^1(\mathbf{Z})} + o(1)$$

where $o(1)$ is a quantity that goes to zero as $n \to \infty$; this implies, in particular, that the residual $w_1^{(n)}$ eventually has mass strictly less than that of the original sequence $x_1^{(n)}$:

$$\|w_1^{(n)}\|_{l^1(\mathbf{Z})} \leq 1 - \varepsilon_1 + o(1);$$

in fact we have the more precise relationship

$$\|x_1\|_{l^1(\mathbf{Z})} + \lim_{n\to\infty} \|w_1^{(n)}\|_{l^1(\mathbf{Z})} = 1.$$

Now we take this residual $w_1^{(n)}$ and repeat the whole process. Namely, if $w_1^{(n)}$ converges in the intermediate topology to zero, then we are done; otherwise, as before, we can find (after passing to a subsequence) $\varepsilon_2 > 0$, $h_2^{(n)}$ for which $T^{-h_2^{(n)}} w_1^{(n)}$ is bounded from below by ε_2 at the origin. Because $T^{-h_1^{(n)}} w_1^{(n)}$ already converged weakly to zero, one can conclude that $h_2^{(n)}$ and $h_1^{(n)}$ must be asymptotically orthogonal in the sense of (1.89).

Passing to a subsequence again, we can assume that $T^{-h_2^{(n)}} w_2^{(n)}$ converges weakly to a limit x_2 with mass at least ε_2, leading to a decomposition

$$x^{(n)} = T^{h_1^{(n)}} x_1 + T^{h_2^{(n)}} x_2 + w_2^{(n)}$$

where the residual $w_2^{(n)}$ is such that $T^{-h_1^{(n)}} w_2^{(n)}$ and $T^{-h_2^{(n)}} w_2^{(n)}$ both converge weakly to zero, and has norm

$$\|w_2^{(n)}\|_{l^1(\mathbf{Z})} \leq 1 - \varepsilon_1 - \varepsilon_2 + o(1);$$

in fact we have the more precise relationship

$$\|x_1\|_{l^1(\mathbf{Z})} + \|x_2\|_{l^1(\mathbf{Z})} + \lim_{n\to\infty} \|w_2^{(n)}\|_{l^1(\mathbf{Z})} = 1.$$

One can continue in this vein, extracting more and more travelling profiles $T^{h_j^{(n)}} x_j$ on finer and finer subsequences, with residuals $w_j^{(n)}$ that are getting smaller and smaller. The subsequences involved depend on j, but by the usual Cantor (or Arzelá-Ascoli) diagonalisation argument, one can work with a single sequence throughout. Note that the amounts of mass ε_j that are extracted in this process cannot exceed 1 in total: $\sum_j \varepsilon_j \leq 1$ (in fact we have the slightly stronger statement (1.89)). In particular, the ε_j must go to zero as $j \to \infty$. If the ε_j were selected in a "greedy" manner, this shows that the asymptotic $l^\infty(\mathbf{Z})$ norm of the residuals $w_j^{(n)}$ as $n \to \infty$ must decay to zero as $j \to \infty$. Carefully rearranging the epsilons, this gives the decomposition (1.90) with residual $w^{(n)}$ converging to zero in the intermediate topology, and the verification of the rest of the theorem is routine.

Remark 1.16.7. It is tempting to view Theorem 1.16.5 as asserting that the space E with the $l^\infty(\mathbf{Z})$ can be "compactified" by throwing in some idealised superposition of profiles that are "infinitely far apart" from each other. However, I do not know of a clean way to formalise this compactification.

1.16.3. An application of concentration compactness. As mentioned in the introduction, one can use the profile decomposition of Theorem 1.16.5 as a substitute for compactness in establishing results analogous to Proposition 1.16.1. The catch is that one needs more hypotheses on the functional F in order to be able to handle the complicated profiles that come up. It is difficult to formalise the "best" set of hypotheses that would cover all conceivable situations; it seems better to just adapt the general arguments to each individual situation separately. Here is a typical (but certainly not optimal) result of this type:

Theorem 1.16.8. . *Let X, E be as above. Let $F : X \to \mathbf{R}^+$ be a non-negative function with the following properties:*

(1) *(Continuity) F is continuous in the intermediate topology on E.*

(2) *(Homogeneity) F is homogeneous of some degree $1 < p < \infty$, thus $F(\lambda x) = \lambda^p F(x)$ for all $\lambda > 0$ and $x \in X$. (In particular, $F(0) = 0$.)*

(3) *(Invariance) F is G-invariant: $F(T^h x) = F(x)$ for all $h \in \mathbf{Z}$ and $x \in X$.*

(4) *(Asymptotic additivity) If $h_j^{(n)}$ are a collection of sequences obeying the asymptotic orthogonality condition* (1.88)*, and $x_j \in X$ are such that $\sum_j \|x_j\|_{l^1(\mathbf{Z})} < \infty$, then $\sum_j F(x_j) < \infty$ and $F(\sum_j T^{h_j^{(n)}} x_j) = \sum_j F(x_j) + o(1)$. More generally, if $w^{(n)}$ is bounded in l^1 and converges to zero in the intermediate topology, then $F(\sum_j T^{h_j^{(n)}} x_j + w^{(n)}) = \sum_j F(x_j) + o(1)$. (Note that this generalises both 1. and 3.)*

Then F is bounded on E, and attains its supremum.

A typical example of a functional F obeying the above properties is

$$F((x_m)_{m\in\mathbf{Z}}) := \sum_{m\in\mathbf{Z}} |x_m - x_{m+1}|^p$$

for some $1 < p < \infty$.

Proof. We repeat the proof of Proposition 1.16.1. Let $L := \sup_{x\in E} F(x)$. Clearly $L \geq 0$; we can assume that $L > 0$, since the claim is trivial when $L = 0$. As before, we have an extremising sequence $x^{(n)} \in E$ with $F(x^{(n)}) \to L$. Applying Theorem 1.16.5, and passing to a subsequence, we obtain a decomposition (1.90) with the stated properties. Applying the asymptotic additivity hypothesis 4., we have

$$F(x^{(n)}) = \sum_j F(x_j) + o(1)$$

and in particular

$$L = \sum_j F(x_j). \tag{1.92}$$

This implies in particular that L is finite.

Now, we use the homogeneity assumption. Since $F(x) \leq L$ when $\|x\|_{l^1(\mathbf{Z})} = 1$, we obtain the bound $F(x) \leq L\|x\|^p_{l^1(\mathbf{Z})}$. We conclude that

$$L \leq L \sum_j \|x_j\|^p_{l^1(\mathbf{Z})}.$$

Combining this with (1.88) we obtain

$$L \leq \sum_j L\|x_j\|^p_{l^1(\mathbf{Z})} \leq L \sum_j \|x_j\|_{l^1(\mathbf{Z})} \leq L.$$

Thus all these inequalities must be equality. Analysing this, we see that all but one of the x_j must vanish, with the remaining x_j (say x_0) having norm 1. From (1.92) we thus have $F(x_0) = L$, and we have obtained the desired extremiser. □

Notes. This article first appeared at `terrytao.wordpress.com/2008/11/05`. Thanks to Jerry Gagelman, JC, A.P., Dylan Thurston, and David Speyer for corrections.

1.17. A counterexample to a strong polynomial Freiman-Ruzsa conjecture

One of my favourite open problems in additive combinatorics is the *polynomial Freiman-Ruzsa conjecture*[**Gr2005**]. It has many equivalent formulations (which is always a healthy sign when considering a conjecture), but here is one involving "approximate homomorphisms":

Conjecture 1.17.1 (Polynomial Freiman-Ruzsa conjecture)**.** *Let $f : F_2^n \to F_2^m$ be a function which is an approximate homomorphism in the sense that $f(x+y)-f(x)-f(y) \in S$ for all $x, y \in F_2^n$ and some set $S \subset F_2^m$. Then there exists a genuine homomorphism $g : F_2^n \to F_2^m$ such that $f - g$ takes at most $O(|S|^{O(1)})$ values.*

Remark 1.17.2. The key point here is that the bound on the range of $f-g$ is at most polynomial in $|S|$. An exponential bound of $2^{|S|}$ can be trivially established by splitting F_2^m into the subspace spanned by S (which has size at most $2^{|S|}$) and some complementary subspace,

and then letting g be the projection of f to that complementary subspace.

In a forthcoming paper with Ben Green, we showed that this conjecture is equivalent to a certain polynomially quantitative strengthening of the inverse conjecture for the Gowers norm $U^3(F_2^n)$. For this (somewhat technical) post, I want to comment on a possible further strengthening of this conjecture, namely

Conjecture 1.17.3 (Strong Polynomial Freiman-Ruzsa conjecture)**.** *Let $f : F_2^n \to F_2^m$ be a function which is an approximate homomorphism in the sense that $f(x+y) - f(x) - f(y) \in S$ for all $x, y \in F_2^n$ and some set $S \subset F_2^m$. Then there exists a genuine homomorphism $g : F_2^n \to F_2^m$ such that $f - g$ takes values in the sumset $CS := S + \ldots + S$ for some fixed $C = O(1)$.*

This conjecture is known to be true for certain types of set S (e.g. for Hamming balls, this was shown in [**Fa2000**]). Unfortunately, it is false in general; the purpose of this post is to describe one counterexample (related to the failure of the inverse conjecture for the Gowers norm for $U^4(F_2^n)$ for classical polynomials; in particular, the arguments here have several features in common with those in [**LoMeSa2008**], [**GrTa2007**]; a somewhat different counterexample also appears in [**Fa2000**]. The verification of the counterexample is surprisingly involved, ultimately relying on the multidimensional Szemerédi theorem[**FuKa1979**].

1.17.1. Description of counterexample. We let n be a large number, and replace F_2^m by the $\frac{n(n+1)}{2}$-dimensional vector space V of quadratic forms $Q : F_2^n \to F_2$ (with a basis given by the monomials $x_i x_j$ with $1 \leq i \leq j \leq n$). We let $f : F_2^n \to V$ be defined by the formula

$$f(h_1, \ldots, h_n)(x_1, \ldots, x_n) := \sum_{1 \leq i < j \leq n} h_i x_i h_j x_j.$$

A brief computation shows that for any $h, k \in F_2^n$, the quadratic form $f(h+k) - f(h) - f(k)$ is of rank at most three, by which we mean that it is a function of at most three linear forms. More specifically,

we have

$$f(h+k) - f(h) - f(k) = a_{h,k}b_{h,k} + b_{h,k}c_{h,k} + c_{h,k}a_{h,k} \tag{1.93}$$

where

$$a_{h,k}(x) := \sum_{i=1}^{n} h_i(1-k_i)x_i$$

$$b_{h,k}(x) := \sum_{i=1}^{n} (1-h_i)x_i$$

$$c_{h,k}(x) := \sum_{i=1}^{n} h_i k_i x_i$$

Thus, if we let S be the space of quadratic forms of rank at most 3, the hypotheses of Conjecture 1.17.3 hold.

1.17.2. Verification of counterexample. To establish the counterexample, we assume for contradiction that there exists a linear function $g : F_2^n \to V$ such that $f(h) - g(h)$ has bounded rank for all h, and deduce a contradiction (for n sufficiently large).

By hypothesis, we have linear forms $L_{h,1}, \ldots, L_{h,d}$ for all $h \in F_2^n$ and some $d = O(1)$ and coefficients $c_{h,i,j} \in F_2$ for all $1 \leq i \leq j \leq d$ such that

$$f(h) - g(h) = \sum_{1 \leq i \leq j \leq n} c_{h,i,j} L_{h,i} L_{h,d}$$

and in particular (by (1.93) and linearity of g)

$$a_{h,k}b_{h,k} + b_{h,k}c_{h,k} + c_{h,k}a_{h,k} = \sum_{1 \leq i \leq j \leq n} c_{h+k,i,j} L_{h+k,i} L_{h+k,d} - c_{h,i,j} L_{h,i} L_{h,d} - c_{k,i,j} \tag{1.94}$$

The key point is that the linear forms $a_{h,k}, b_{h,k}, c_{h,k}$ are usually "independent" of the linear forms $L_{h,i}, L_{k,i}, L_{h+k,i}$. The crucial lemma in this regard is

Lemma 1.17.4. *If h, k are selected uniformly and independently at random, then with probability $1-o(1)$, $a_{h,k}$ is not a linear combination of the $L_{h,i}, L_{k,i}, L_{h+k,i}$. Similarly for $b_{h,k}, c_{h,k}$.*

Proof. By cyclically permuting h, k, $h+k$ it suffices to show this for $c_{h,k}$. Since there are at most $O(1)$ possible linear combinations

amongst the $L_{h,i}, L_{k,i}, L_{h+k,i}$, it suffices to show that for any given assignments $h \mapsto L'_h, k \mapsto L''_k, h+k \mapsto L'''_{h+k}$ of linear forms, that the probability of the event

$$c_{h,k} = L'_h + L''_k + L'''_{h+k} \tag{1.95}$$

is $o(1)$. Suppose for contradiction that the event (1.95) holds for a set E of pairs (h,k) in $F_2^n \times F_2^n$ of positive density. Applying the multidimensional Szemerédi theorem[**FuKa1979**] we can find (for n large enough) a square $(h,k), (h+r,k), (h,k+r), (h+r,k+r)$ in E with r non-zero. Applying (1.95) for all four pairs and summing, we obtain

$$c_{h,k} + c_{h+r,k} + c_{h,k+r} + c_{h+r,k+r} = 0$$

(recall we are in characteristic 2). But the left-hand side is equal to the linear form $\sum_i r_i x_i$, which is non-zero, a contradiction. □

Now we can obtain the desired contradiction. For a generic choice of h, k, we now know that none of the $a_{h,k}, b_{h,k}, c_{h,k}$ are linear combinations of the $L_{h,i}, L_{k,i}, L_{h+k,i}$. Thus, on a given level set of the $L_{h,i}, L_{k,i}, L_{h+k,i}$ (which form a subspace of F_2^n), the linear functions $a_{h,k}, b_{h,k}, c_{h,k}$ are non-constant, and so the range of the triplet $(a_{h,k}, b_{h,k}, c_{h,k})$ must be an affine subspace of F_2^3 which is not contained in any coordinate plane. This forces this subspace to have dimension at least two. But then the function $(a,b,c) \mapsto ab+bc+ca$ cannot be constant on this space, contradicting (1.94), and so Conjecture 1.17.3 fails.

Remark 1.17.5. The function f appearing in the above example is closely related to the symmetric polynomial

$$S_4(x) := \sum_{1 \le i < j < k < l \le n} x_i x_j x_k x_l.$$

Indeed, one can show that the derivative $S_4(x+h) - S_4(x)$ of S_4 is equal to $f(h)$, plus some additional terms which involve only a finite number of linear forms, and the quadratic polynomial $S_2(x) := \sum_{1 \le i < j \le n} x_i x_j$. If it was the case that f could be approximated by a linear map g modulo low rank errors, then it one could use this to eventually show that S_4 correlated with a cubic polynomial; but it is known [**LoMeSa2008**], [**GrTa2007**] that this is not the case. Thus

there is an alternate way to verify that the above example is indeed a counterexample to the strong polynomial Freiman-Ruzsa conjecture.

Notes. This article first appeared at `terrytao.wordpress.com/2008/11/09`, and is derived from forthcoming joint work with Ben Green.

1.18. Some notes on "non-classical" polynomials in finite characteristic

Let $k \geq 0$ be an integer. The concept of a polynomial $P : \mathbf{R} \to \mathbf{R}$ of one variable of degree $< k$ (or $\leq k-1$) can be defined in one of two equivalent ways:

- (Global definition) $P : \mathbf{R} \to \mathbf{R}$ is a polynomial of degree $< k$ iff it can be written in the form $P(x) = \sum_{0 \leq j < k} c_j x^j$ for some coefficients $c_j \in \mathbf{R}$.
- (Local definition) $P : \mathbf{R} \to \mathbf{R}$ is a polynomial of degree $< k$ if it is k-times continuously differentiable and $\frac{d^k}{dx^k} P \equiv 0$.

From single variable calculus we know that if P is a polynomial in the global sense, then it is a polynomial in the local sense; conversely, if P is a polynomial in the local sense, then from the Taylor series expansion

$$P(x) = \sum_{0 \leq j < k} \frac{P^{(j)}(0)}{j!} x^j$$

we see that P is a polynomial in the global sense. We make the trivial remark that we have no difficulty dividing by $j!$ here, because the field $\mathbf{R}$ is of characteristic zero.

The above equivalence carries over to higher dimensions:

- (Global definition) $P : \mathbf{R}^n \to \mathbf{R}$ is a polynomial of degree $< k$ iff it can be written in the form $P(x_1, \ldots, x_n) = \sum_{0 \leq j_1, \ldots, j_n; j_1 + \ldots + j_n < k} c_{j_1, \ldots, j_n} x_1^{j_1} \ldots x_n^{j_n}$ for some coefficients $c_{j_1, \ldots, j_n} \in \mathbf{R}$.
- (Local definition) $P : \mathbf{R}^n \to \mathbf{R}$ is a polynomial of degree $< k$ if it is k-times continuously differentiable and $(h_1 \cdot \nabla) \ldots (h_k \cdot \nabla) P \equiv 0$ for all $h_1, \ldots, h_k \in \mathbf{R}^n$.

Again, it is not difficult to use several variable calculus to show that these two definitions of a polynomial are equivalent.

The purpose of this (somewhat technical) post here is to record some basic analogues of the above facts in finite characteristic, in which the underlying domain of the polynomial P is F or F^n for some finite field F. In the "classical" case when the range of P is also the field F, it is a well-known fact (which we reproduce here) that the local and global definitions of polynomial are equivalent. But in the "non-classical" case, when P ranges in a more general group (and in particular in the unit circle $\mathbf{R}/\mathbf{Z}$), the global definition needs to be corrected somewhat by adding some new monomials to the classical ones $x_1^{j_1} \dots x_n^{j_n}$. Once one does this, one can recover the equivalence between the local and global definitions.

1.18.1. General theory. One can extend the local definition of a polynomial to cover maps $P : G \to H$ for any additive[91] groups G, H. Given any such map, and any $h \in G$, define the shift $T^h P : G \to H$ and the (discrete) derivative $\Delta_h P : G \to H$ by the formulae

$$T^h P(x) := P(x+h); \Delta_h P = T^h P - P,$$

thus schematically we have

$$\Delta_h = T^h - 1. \tag{1.96}$$

We say that P is an *(additive) polynomial of degree* $< k$ (or degree $\leq k-1$) if $\Delta_{h_1} \dots \Delta_{h_k} P = 0$ for all $h_1, \dots, h_k \in G$. Note that this corresponds to the definition of a classical polynomial from $\mathbf{R}^n$ to $\mathbf{R}$ once one adds some regularity conditions, such as k-times differentiability (actually, measurability will already suffice).

Examples 1.18.1. The zero function has degree < 0. A constant function has degree ≤ 0. A homomorphism has degree ≤ 1. Composing a polynomial of degree $\leq k$ with a homomorphism (either on the left or right) will give another polynomial of degree $\leq k$. The sum of two polynomials of degree $\leq k$ is again of degree $\leq k$. The derivative $\Delta_h P$ of a polynomial of degree $< k$ is of degree $< k-1$.

[91] There is also an important generalisation of this concept to the case of nilpotent groups; we will not concern ourselves with this generalisation here, but see [**Le1998**], [**Le2002**].

Since $T^h P = P + \Delta_h P$, we conclude that the shift of any polynomial of degree $< k$ is also of degree $< k$.

Now we show that the product and composition of polynomials is again a polynomial.

Lemma 1.18.2 (Product of polynomials is again polynomial)**.** *Let $P : G \to H, Q : G \to K$ be polynomials of degree $\leq h, \leq k$ respectively for some $h, k \geq 0$, and let $B : H \times K \to L$ be a bilinear map. Then $B(P, Q)$ is a polynomial of degree $\leq h + k$.*

Proof. We induct on $h + k$. The claim is easy when h or k is zero, so suppose that $h, k > 0$ and the claim has already been proven for smaller values of $h + k$. From the discrete product rule

$$\Delta_g B(P, Q) = B(\Delta_g P, Q) + B(T^g P, \Delta_g Q)$$

and induction we see that $\Delta_g B(P, Q)$ is of degree $\leq h + k - 1$, and thus $B(P, Q)$ has degree $\leq h + k$ as desired. □

Corollary 1.18.3. *If H is a ring, then the product of two polynomials from G to H of degree $\leq h, \leq k$ respectively is of degree $\leq h + k$.*

Lemma 1.18.4 (Composition of polynomials is again polynomial)**.** *Let $P : G \to H, Q : H \to K$ be polynomials of degree $\leq h, \leq k$ respectively for some $h, k \geq 0$. Then $Q \circ P : G \to K$ is a polynomial of degree $\leq hk$.*

Proof. For inductive reasons it is convenient to prove the following more general statement: if $P : G \to H, Q : H \to K$ are polynomials of degree $\leq h + m, \leq k$ respectively for some $m, h, k \geq 0$, and $R_1, \ldots, R_m : G \to H$ are polynomials of degree $\leq r_1, \ldots, \leq r_m$ respectively, where $0 \leq r_j \leq k$ for all j, then the function $S : G \to K$ defined by

$$S(x) := [\Delta_{R_1(x)} \ldots \Delta_{R_m(x)} P](Q(x))$$

is a polynomial of degree $hk + r_1 + \ldots + r_m$. Clearly Lemma 1.18.4 follows from the $m \geq 0$ case of this claim.

We prove this claim by induction on h, then for fixed h by induction on m, then for fixed h and m by induction on $r_1 + \ldots + r_m$. Thus, assume that the claim has already been shown for all smaller

values of h, or for the same value of h and all smaller values of m, or for the same values of h, m and all smaller values of $r_1 + \ldots + r_m$.

If $r_m = 0$ then R_m is constant, and by replacing P with $\Delta_{R_m} P$ and decrementing m, we see that the claim follows from the induction hypothesis. Similarly if any other of the r_j vanish (since the derivative operators commute with each other). So we may assume that $r_j > 0$ for all j.

Let g in G. By considering the successive differences between the quantities

$$\begin{aligned} S(x) &= [\Delta_{R_1(x)} \Delta_{R_2(x)} \ldots \Delta_{R_m(x)}](Q(x)), \\ &\quad [\Delta_{R_1(x)} \Delta_{R_2(x)} \ldots \Delta_{R_m(x)}](T_g Q(x)), \\ &\quad [\Delta_{T_g R_1(x)} \Delta_{R_2(x)} \ldots \Delta_{R_m(x)}](T_g Q(x)), \\ &\quad [\Delta_{T_g R_1(x)} \Delta_{T_g R_2(x)} \ldots \Delta_{R_m(x)}](T_g Q(x)), \\ &\quad \vdots \\ T_g S(x) &= [\Delta_{T_g R_1(x)} \Delta_{T_g R_2(x)} \ldots \Delta_{T_g R_m(x)}](T_g Q(x)), \end{aligned}$$

we see that $\Delta_g S(x)$ is the sum of

$$\begin{aligned} &[\Delta_{R_1(x)} \Delta_{R_2(x)} \ldots \Delta_{R_m(x)} \Delta_{\Delta_g Q(x)} P](Q(x)), \\ &[\Delta_{\Delta_g R_1(x)} \Delta_{R_2(x)} \ldots \Delta_{R_m(x)}](R_1(x) + T_g Q(x)), \\ &[\Delta_{T_g R_1(x)} \Delta_{\Delta_g R_2(x)} \ldots \Delta_{R_m(x)} \Delta_g](R_2(x) + T_g Q(x)) \\ &\vdots \\ &[\Delta_{T_g R_1(x)} \Delta_{T_g R_2(x)} \ldots \Delta_{\Delta_g R_m(x)} \Delta_g](R_m(x) + T_g Q(x)). \end{aligned}$$

By the induction hypothesis, each of these terms are polynomials of degree $\leq hk + r_1 + \ldots + r_m - 1$. The claim follows. □

1.18.2. The classical case. Now we consider polynomials taking values in a finite field F.

Lemma 1.18.5 (Global description of classical one-dimensional polynomials)**.** *Let F be a field of prime order p. For any $k \geq 0$, a function $P : F \to F$ is of degree $< k$ if and only if we can expand $P(x) = \sum_{0 \leq j < k} c_j x^j$ for some coefficients $c_j \in F$; this expansion is unique for $k \leq p$. Also, every function $P : F \to F$ is a polynomial of degree $< p$.*

Proof. The "if" portion of the lemma follows from Corollary 1.18.3 (since the identity function $x \mapsto x$ is clearly of degree ≤ 1). For the "only if" part, observe from the binomial identity

$$T^h = (1 + \Delta_1)^h = \sum_{j=0}^{h} \binom{h}{j} \Delta_1^j$$

for any non-negative integer h, that

$$f(h) = \sum_{j=0}^{h} \binom{h}{j} \Delta_1^j f(0).$$

Since $h \mapsto \binom{h}{j} = \frac{h(h-1)\ldots(h-j+1)}{j!}$ can be meaningfully defined on F for $0 \leq j < p$, we conclude in particular that

$$f(h) = \sum_{j=0}^{p-1} \binom{h}{j} \Delta_1^j f(0).$$

Since $\binom{h}{j}$ can be expanded as a linear combination over F of $1, h, \ldots, h^j$, we obtain the remaining claims in Lemma 1.18.5. (Note that as the space of functions from F to F is p-dimensional, and generated by $1, x, \ldots, x^{p-1}$, these functions must be linearly independent.) □

Corollary 1.18.6 (Integration lemma)**.** *Let $f : F \to F$ be a polynomial of degree $\leq k$ for some $0 \leq k \leq p-2$, and let $h \in F \backslash 0$. Then there exists a polynomial $P : F \to F$ of degree $\leq k+1$ such that $f = \Delta_h P$. (In particular, this implies the mean zero condition $\sum_{x \in F} f(x) = 0$. Conversely, any function $f : F \to F$ with $\sum_{x \in F} f(x) = 0$ is a polynomial of degree $\leq p-2$.*

Proof. From Lemma 1.18.5, the space of polynomials of degree $\leq k$ and $\leq k+1$ is a vector space over F of dimension $k+1$ and $k+2$ respectively. The derivative operator Δ_h is a linear transformation from the latter to the former with a one-dimensional kernel (the space of constants), and must therefore be surjective. The first claim follows. The second claim follows by a similar dimension counting argument. □

We can iterate Lemma 1.18.5 to describe polynomials in higher dimensions:

Lemma 1.18.7 (Global description of classical multi-dimensional polynomials)**.** *Let F be a field of prime order p, and let $n \geq 1$. For any $k \geq 0$, a function $P : F^n \to F$ is of degree $< k$ if and only if we can expand $P(x_1, \ldots, x_n) = \sum_{0 \leq j_1, \ldots, j_n : j_1 + \ldots + j_n < k} c_{j_1,\ldots,j_n} x_1^{j_1} \ldots x_n^{j_n}$ for some coefficients $c_{j_1,\ldots,j_n} \in F$.*

Proof. As before, the "if" portion follows from Corollary 1.18.3, so it suffices to show the "only if" portion. But this follows by a multi-dimensional version of the analogous argument used to show Lemma 1.18.5, starting with the identity

$$T^{(h_1,\ldots,h_n)} = (1 + \Delta_{e_1})^{h_1} \ldots (1 + \Delta_{e_n})^{h_n}$$

for non-negative integers $h_1, \ldots, h_n$, where $e_1, \ldots, e_n$ is the standard basis of F^n; we leave the details to the reader. □

Remark 1.18.8. The above discussion was for fields $F = F_p$ of prime order, but we can use these results to describe classical polynomials for fields $F = F_{p^m}$ of prime power order, by viewing any vector space over F_{p^m} as a vector space over F_p. Of course, the resulting polynomials one obtains are merely polynomials over F_p, rather than over F_p^m.

1.18.3. The non-classical case. Now we consider polynomials from F or F^n into other additive groups, where $F = F_p$ is as before a field of prime order p. Thanks to Pontryagin duality, it suffices (in principle, at least) to consider polynomials taking values in the unit circle $\mathbf{R}/\mathbf{Z}$. The first basic lemma is the following:

Lemma 1.18.9 (Multiplication by p reduces degree)**.** *Let $f : F^n \to \mathbf{R}/\mathbf{Z}$ be of degree $\leq k + p - 1$ for some $k \geq 0$. Then pf is of degree $\leq k$.*

Proof. Since $\Delta_h(pf) = p\Delta_h f$ for any h, we see by induction that it suffices to show this lemma when $k = 0$. Let $h \in F^n$. Raising (1.96) to the p^{th} power we have

$$T^{ph} = 1 + p\Delta_h + \frac{p(p-1)}{2}\Delta_h^2 + \ldots + p\Delta_h^{p-1} + \Delta_h^p.$$

Of course, $T^{ph} = 1$. Applying this identity to f and noting that $\Delta_h^p f = 0$ by hypothesis, we conclude that

$$(1 + \frac{p-1}{2}\Delta_h + \ldots + \Delta_h^{p-2})\Delta_h(pf) = 0.$$

Inverting $(1 + \frac{p-1}{2}\Delta_h + \ldots + \Delta_h^{p-2})$ using Neumann series (and the finite degree of f) we conclude that $\Delta_h(pf) = 0$ for all h, thus pf has degree ≤ 0 as required. □

Corollary 1.18.10 (Polynomials are discretely valued)**.** *If $f : F^n \to \mathbf{R}/\mathbf{Z}$ is of degree $\leq k$, then after subtracting a constant from f, f takes values in the $(p^{\lfloor (k-1)/(p-1) \rfloor + 1})^{\text{th}}$ roots of unity.*

In one dimension, there is a converse to Lemma 1.18.9:

Lemma 1.18.11. *Let $f : F^n \to \mathbf{R}/\mathbf{Z}$ be such that pf has degree $\leq k$. Then f has degree $\leq k + p - 1$.*

Proof. As in Lemma 1.18.9, it suffices to establish the case $k = 0$. But this then follows from the last part of Lemma 1.18.5. □

As a corollary we can classify all non-classical polynomials:

Theorem 1.18.12 (Global description of non-classical multi-dimensional polynomials)**.** *A function $f : F^n \to \mathbf{R}/\mathbf{Z}$ is a polynomial of degree $< k$ if and only if it has the form*

$$f(x) = c_0 + \sum_{0 \leq j_1, \ldots, j_n \leq p-1; m \geq 1: j_1 + \ldots + j_n + (p-1)(m-1) < k} c_{j_1,\ldots,j_m,m} |x_1|^{j_1} \ldots |x_n|^{j_n} / p^m$$

for some $c_0 \in \mathbf{R}/\mathbf{Z}$ and $c_{j_1,\ldots,j_m,m} \in \{0, \ldots, p-1\}$, where $x \mapsto |x|$ is the obvious map from F to $\{0, \ldots, p-1\}$.

Proof. The "if" part follows easily from Lemma 1.18.7 in the case $k \leq p$, and then from Lemma 1.18.11 and induction in the general case. The "only if" part follows from Corollary 1.18.10 and Lemma 1.18.7 in the case $k \leq p$. Now suppose inductively that $k > p$ and the claim has already been proven for smaller values of k. By Lemma 1.18.9 and the induction hypothesis, pf takes the form

$$pf(x) = c'_0 + \sum_{0 \leq j_1, \ldots, j_n \leq p-1; m \geq 1: j_1 + \ldots + j_n + (p-1)(m-1) < k-p+1} c'_{j_1,\ldots,j_n,m} |x_1|^{j_1} \ldots |x_n$$

and thus

$$f(x) = c_0 + \sum_{0 \leq j_1,\ldots,j_n \leq p-1; m \geq 1: j_1+\ldots+j_n+(p-1)(m-1) < k-p+1} c'_{j_1,\ldots,j_n,m} |x_1|^{j_1} \ldots |x_n|^{j_n} g(x_1^m \ldots x_n^m)$$

where c_0 is a p^{th} root of c'_0, and g takes values in p^{th} roots of unity. Applying Lemma 1.18.7 to expand g in monomials, we obtain the claim. □

As a corollary to this theorem we obtain a converse to Lemma 1.18.9:

Corollary 1.18.13 (p^{th} roots of minimal degree)**.** *Let $f : F^n \to \mathbf{R}/\mathbf{Z}$ be of degree $\leq k$ for some $k \geq 0$. Then there exists $g : F^n \to \mathbf{R}/\mathbf{Z}$ of degree $\leq k + p - 1$ such that $pg = f$.*

Interestingly, there does not seem to be a way to establish this theorem without going through a global classification theorem such as Theorem 1.18.12.

Another corollary to Theorem 1.18.12 is that any function from a finite dimensional vector space F^n to a p^m-torsion group for some m will be a polynomial of finite degree.

Notes. This article first appeared at `terrytao.wordpress.com/2008/11/13`, and is derived from [**BeTaZi2009**]. Thanks to James Cranch for corrections.

1.19. The Kakeya conjecture and the Ham Sandwich theorem

One of my favourite family of conjectures (and one that has preoccupied a significant fraction of my own research) is the family of *Kakeya conjectures* in geometric measure theory and harmonic analysis. There are many (not quite equivalent) conjectures in this family. The cleanest one to state is the set conjecture, Conjecture 1.3.3.

One reason why I find these conjectures fascinating is the sheer variety of mathematical fields that arise both in the partial results towards this conjecture, and in the applications of those results to other problems. See for instance [**Wo1999**], [**Ta2001**], [**La2008**] on the connections between this problem and other problems in Fourier

analysis, PDE, and additive combinatorics; there have even been some connections to number theory[**Bo2001**] and to cryptography[**Bo2005**]. At the other end of the pipeline, the mathematical tools that have gone *into* the proofs of various partial results have included:

- Maximal functions, covering lemmas, L^2 methods [**Co1977**], [**CoFe1977**];
- Fourier analysis [**NaStWa1978**];
- Multilinear integration [**Dr1983**], [**Ch1984**];
- Paraproducts [**Ka1999**];
- Combinatorial incidence geometry [**Bo1991**], [**Wo1995**];
- Multi-scale analysis [**Ba1996**], [**KaLaTa2000**], [**LaTa2001**], [**AlSoVa2003**];
- Probabilistic constructions [**BaKa2008**], [**Ba2008**];
- Additive combinatorics and graph theory [**Bo1999**], [**KaLaTa2000**], [**KaTa1999**], [**KaTa200b**];
- Sum-product theorems [**BoKaTa2004**];
- Bilinear estimates [**TaVaVe1998**];
- Perron trees [**Sc1962**], [**Ke1999**];
- Group theory [**Ka2005**];
- Low-degree algebraic geometry [**Sc1998**], [**Ta2005**], [**MoTa2004**];
- High-degree algebraic geometry [**Dv2008**], [**SaSu2008**];
- Heat flow monotonicity formulae [**BeCaTa2006**].

[This list is not exhaustive.]

Very recently, I was pleasantly surprised to see yet another mathematical tool used to obtain new progress on the Kakeya conjecture, namely (a generalisation of) the famous *Ham Sandwich theorem* from algebraic topology. This was recently used by Guth[**Gu2008**] to establish a certain endpoint multilinear Kakeya estimate left open in [**BeCaTa2006**]. With regards to the Kakeya set conjecture, Guth's arguments assert, roughly speaking, that the only Kakeya sets that can fail to have full dimension are those which obey a certain "planiness" property, which informally means that the line segments that pass through a typical point in the set must be essentially coplanar.

(This property first surfaced in [**KaLaTa2000**].) Guth's arguments can be viewed as a partial analogue of Dvir's arguments[**Dv2008**] in the finite field setting (which I discussed in Section 1.3) to the Euclidean setting; in particular, both arguments rely crucially on the ability to create a polynomial of controlled degree that vanishes at or near a large number of points. Unfortunately, while these arguments fully settle the Kakeya conjecture in the finite field setting, it appears that some new ideas are still needed to finish off the problem in the Euclidean setting. Nevertheless this is an interesting new development in the long history of this conjecture, in particular demonstrating that the polynomial method can be successfully applied to continuous Euclidean problems (i.e. it is not confined to the finite field setting).

In this article I would like to sketch some of the key ideas in Guth's paper, in particular the role of the Ham Sandwich theorem (or more precisely, a polynomial generalisation of this theorem first observed [**Gr2003**]).

1.19.1. The polynomial Ham Sandwich theorem. Let us first recall the classical Ham Sandwich theorem:

Theorem 1.19.1 (Ham Sandwich theorem)**.** *Let $U_1, \ldots, U_n$ be n bounded open sets in $\mathbf{R}^n$. Then there exists a hyperplane in $\mathbf{R}^n$ that divides each of the open sets $U_1, \ldots, U_n$ into two sets of equal volume.*

Remark 1.19.2. The name of the theorem derives from the special case when $n = 3$ and U_1, U_2, U_3 are two slices of bread and a slice of ham. One can view this theorem as a "thickened" version of the Euclidean geometry axiom that every n points in $\mathbf{R}^n$ determine at least one hyperplane.

There are many proofs of this theorem, but I will focus on the proof that is based on the Borsuk-Ulam theorem:

Theorem 1.19.3 (Borsuk-Ulam theorem)**.** *Let $f : S^n \to \mathbf{R}^n$ be a continuous map from the n-dimensional sphere $S^n \subset \mathbf{R}^{n+1}$ to the Euclidean space $\mathbf{R}^n$ which is* antipodal *(which means that $f(-x) = -f(x)$ for all $x \in S^n$. Then $f(x) = 0$ for at least one $x \in S^n$.*

Proof. (Sketch) The set of zeroes of an antipodal map automatically come in antipodal pairs x, $-x$. To prove the theorem, we shall establish the stronger fact that $f(x) = 0$ for an odd number of disjoint antipodal pairs, counting multiplicity (avoiding the degenerate antipodal maps which vanish at an infinite set of points). To see this, first observe that this is true for at least one antipodal map (e.g. one can use the horizontal projection map $(x_1, \ldots, x_{n+1}) \mapsto (x_1, \ldots, x_n)$). Also, the space of all antipodal maps is a vector space, and thus connected (though it takes some effort to show that the space of *non-degenerate* antipodal maps is still connected). So one just needs to show that the parity of the number of pairs of antipodal points where f vanishes (counting multiplicity) is unchanged with respect to continuous deformations of f. But some elementary degree theory (or *Morse theory*) shows that any (non-degenerate) perturbation of f can annihilate two such antipodal pairs by collision, or (by the reverse procedure) spontaneously create two such antipodal pairs from nothing, but cannot otherwise affect the number of pairs; thus the parity of the number of such pairs remains invariant[92]. □

Remark 1.19.4. The Borsuk-Ulam theorem is tied to the more general theory of *Lyusternik-Schnirelmann category*, which is the viewpoint taken in [**Gu2008**], but we will not explicitly use this theory here.

Proof of the Ham-Sandwich theorem using the Borsuk-Ulam theorem

We can identify $\mathbf{R}^{n+1}$ with the space of affine-linear forms $(x_1, \ldots, x_n) \mapsto a_1 x_1 + \ldots + a_n x_n + a_0$ on $\mathbf{R}^n$. Each non-trivial affine-linear form $P \in \mathbf{R}^{n+1} \backslash 0$ determines a hyperplane $\{P = 0\}$ that divides $\mathbf{R}^n$ into two half-spaces $\{P > 0\}$ and $\{P < 0\}$. We can then define $f : \mathbf{R}^{n+1} \backslash \{0\} \to \mathbf{R}^n$ to be the function whose j^{th} coordinate $f_j(P)$ at P is the volume of $U_j \cap \{P > 0\}$ minus the volume of $U_j \cap \{P < 0\}$; thus f measures the extent to which the hyperplane $\{P = 0\}$ fails to bisect all of the $U_1, \ldots, U_n$. It is easy to see that f is continuous, homogeneous of degree zero, and odd, and so its restriction to S^n is

[92] It takes some non-trivial effort to make this informal argument rigorous; see for instance [**Ma2003**]. [Thanks to Benny Sudakov for this great reference.] One can also formalise this argument using the language of $\mathbf{Z}_2$ *singular cohomology*.

an antipodal map. By the Borsuk-Ulam theorem, there exists P such that $f(P) = 0$, and the claim follows. □

We have the following polynomial generalisation of the Ham Sandwich theorem:

Theorem 1.19.5 (Polynomial Ham Sandwich theorem)**.** [**Gr2003**] *Let $d \geq 1$, and let $U_1, \ldots, U_{\binom{n+d}{d}-1}$ be bounded open sets in $\mathbf{R}^n$. Then there exists a non-trivial polynomial $P : \mathbf{R}^n \to \mathbf{R}$ of degree at most d such that the sets $\{P > 0\}$, $\{P < 0\}$ partition each of the $U_1, \ldots, U_{\binom{n+d}{d}-1}$ into two sets of equal measure.*

Note that the ordinary Ham-Sandwich theorem corresponds to the $d = 1$ case of this theorem. This theorem can be deduced from the Borsuk-Ulam theorem in exactly the same way that the ordinary one is (note that the space of polynomials of degree at most d has dimension $\binom{n+d}{d}$; the continuity of the appropriate antipodal function $f : S^{\binom{n+d}{d}-1} \to \mathbf{R}^{\binom{n+d}{d}-1}$ follows from the dominated convergence theorem and the basic observation that a non-trivial polynomial is non-zero almost everywhere).

Remark 1.19.6. One can also deduce the polynomial Ham Sandwich theorem directly from the ordinary Ham Sandwich theorem (in $\binom{n+d}{d} - 1$ dimensions) by embedding $\mathbf{R}^n$ into $\mathbf{R}^{\binom{n+d}{d}-1}$ via the *Veronese embedding*, and then thickening the images of $U_1, \ldots, U_{\binom{n+d}{d}-1}$ slightly in an appropriate fashion; we leave the details as an exercise to the reader.

The polynomial Ham Sandwich theorem should be compared with Lemma 1.3.6.

1.19.2. Connection with the Kakeya problem. Now we connect the polynomial Ham Sandwich theorem to the Kakeya problem. We begin by replacing the continuous Kakeya set conjecture with a more quantitative "δ-discretised" problem:

Conjecture 1.19.7 (Kakeya maximal conjecture)**.** *Let $0 < \delta < 1$, and let $T_1, \ldots, T_M$ be a collection of $\delta \times 1$ cylindrical tubes pointing in a δ-separated set of directions (thus the directions of any two of the*

tubes make an angle of at least δ). For each $\mu \geq 1$, let E_μ be the set of points x which are contained in at least μ of the tubes $T_1, \ldots, T_M$. Then the volume $|E_\mu|$ of E_μ obeys the bound $|E_\mu| \lesssim_\varepsilon \delta^{-\varepsilon} \mu^{-n/(n-1)}$ for any $\varepsilon > 0$.

Here we are using the asymptotic notation that $X \gtrsim Y$ if $X \geq cY$ for some positive constant c (if the $\gtrsim$ is subscripted by parameters, this indicates that c is allowed to depend on those parameters); we always allow constants to depend on the dimension n. This conjecture (which is limiting the extent to which tubes in different directions can overlap) implies the Kakeya set conjecture (for both Minkowski and Hausdorff dimension) by fairly standard arguments from geometric measure theory, see e.g. [**Bo1991**]. The factor of $\mu^{-n/(n-1)}$ is natural (and best possible), as can be seen by considering the example in which $M \sim \delta^{1-n}$ and all the tubes pass through a common point.

Remark 1.19.8. The name "maximal conjecture" has to do with the formulation of the above conjecture involving the *Kakeya maximal function*, which I will not discuss here.

The maximal conjecture (and the set conjecture) is verified in the two-dimensional case $n = 2$ (with the one-dimensional case $n = 1$ being trivial), but only partial results are known in higher dimensions. However, one can do better if one only considers certain types of overlap. Let us say (somewhat informally) that a point x has *non-planar multiplicity* $\gtrsim \mu$ with respect to a given collection of tubes $T_1, \ldots, T_M$ if there exist n separate families of $\gtrsim \mu$ tubes each passing through x, such that given any n tubes from each of these three families, the solid angle between the n directions is comparable to 1. (Informally, this is a stronger assertion than saying that x has $\gtrsim \mu$ tubes passing through it, because we prohibit these tubes from being essentially contained in a hyperplane.) Then, as a special case of Guth's results, one has

Theorem 1.19.9 (Multilinear Kakeya conjecture)**.** [**Gu2008**] *Let $\delta, n, T_1, \ldots, T_M, \mu$ be as in the Kakeya maximal conjecture, and let E_μ^* be the set of points with non-planar multiplicity $\gtrsim \mu$. Then $|E_\mu^*| \lesssim \mu^{-n/(n-1)}$.*

Informally, this implies that the only counterexamples to the Kakeya maximal conjecture can come from configurations of tubes such that the tubes that pass through a typical point largely lie in a hyperplane. In [**BeCaTa2006**], we established this estimate with an additional loss of δ^ε by a totally different method (based on heat flow monotonicity formulae). For a precise statement of the full multilinear Kakeya conjecture (which is now proven without any epsilon loss), see [**BeCaTa2006**] or [**Gu2008**].

Let's now sketch why the above result is true (details can be found in [**Gu2008**]). I'll drop the dependence of implied constants on n. Let $x_1, \ldots, x_A$ be a maximal δ-net of E_μ^* (i.e. a set of δ-separated points in E_μ^* that is maximal with respect to set inclusion), then it will suffice to show that

$$A \lesssim \delta^{-n} \mu^{-n/(n-1)}. \tag{1.97}$$

Let Q_j be the cube of sidelength δ centred at x_j with sides parallel to the axes. Applying the polynomial Ham Sandwich theorem, we can find a non-trivial polynomial P of degree $O(A^{1/n})$ whose zero locus $V := \{P = 0\}$ bisects each of the cubes $Q_1, \ldots, Q_A$.

For each j, we claim that the hypersurfaces $V \cap Q_j$ have surface area $\gtrsim \delta^{n-1}$. Indeed, if instead one of the $V \cap Q_j$ had surface area $o(\delta^{n-1})$, this would imply that the projection of $V \cap Q_j$ to any $(n-1)$-dimensional coordinate subspace of Q_j has area $o(\delta^{n-1})$, in contrast with the projection of Q_j itself which has area δ^{n-1}. Thus for each $1 \leq i \leq n$ the complement of V in Q_j contains a subset of Q_j of relative density $1 - o(1)$ that consists entirely of line segments of length δ in the basis direction e_i. From this it is not hard to see that $Q_j \backslash V$ contains a path-connected component of relative density $1 - o(1)$, which contradicts the claim that V bisects Q_j.

On the other hand, we know that Q_j meets $\gtrsim \mu$ tubes T_k, which are arranged in a non-planar fashion. Because of this, one can show[93] that for a "typical" tube T_k hitting Q_j, the projection of $V \cap Q_j$ to the orthogonal complement of the direction of T_k has area $\gtrsim \delta^{n-1}$.

[93] Basically, the point is that at any given point of $V \cap Q_j$, the normal vector cannot be perpendicular (or close to perpendicular) to all the directions of all the T_k simultaneously, due to non-planarity.

To simplify the exposition, let us assume that in fact *all* tubes T_k touching Q_j are typical.

Each Q_j touches $\sim \mu$ tubes T_k (they may touch more than this, but for sake of exposition let us suppose that they touch exactly this number of tubes). By *double counting*, this means that each tube T_k touches about

$$\lambda := A\mu/M \gtrsim \delta^{n-1} A\mu \tag{1.98}$$

cubes Q_j on the average, where the inequality in (1.98) comes from the δ-separated directions of the tubes. In particular, we can find a (typical) tube T_k which touches at least λ such balls. Let v_k be the direction vector of T_k.

Now look[94] at $V \cap T_k$. This set contains $\gtrsim \lambda$ disjoint sets of the form $V \cap Q_j$. Each of these sets, when projected to the orthogonal complement of T_k, has measure $\gtrsim \delta^{n-1}$. On the other hand, T_k itself, when projected to this complement, has a measure of $O(\delta^{n-1})$. By the pigeonhole principle, we may thus find a positive measure family of lines ℓ in the direction v_k passing through T_k which intersect at $\gtrsim \lambda$ of the $V \cap Q_j$. In particular, all lines ℓ in this family intersect V in $\gtrsim \lambda$ different points.

On the other hand, the restriction of P to ℓ is a polynomial of degree $O(A^{1/n})$. If this degree is much less than λ, this forces P to vanish on each line ℓ [cf. Section 1.3]; since the set of such lines has positive measure, this forces P to be identically zero, a contradiction. Hence we must have

$$A^{1/n} \gtrsim \lambda$$

which when combined with (1.98), gives (1.97).

Notes. This article first appeared at `terrytao.wordpress.com/2008/11/27`. Thanks to Jordi-Lluis Figeras Romero and an anonymous commenter for corrections.

[94] Technically, one has to replace T_k by a slight thickening of itself here, but let us ignore this issue.

1.20. An airport-inspired puzzle

I was recently at an international airport, trying to get from one end of a very long terminal to another. It inspired in me the following simple maths puzzle, which I thought I would share here:

Suppose you are trying to get from one end A of a terminal to the other end B. (For simplicity, assume the terminal is a one-dimensional line segment.) Some portions of the terminal have moving walkways (in both directions); other portions do not. Your walking speed is a constant v, but while on a walkway, it is boosted by the speed u of the walkway for a net speed of $v + u$. (Obviously, given a choice, one would only take those walkways that are going in the direction one wishes to travel in.) Your objective is to get from A to B in the shortest time possible.

(1) Suppose you need to pause for some period of time, say to tie your shoe. Is it more efficient to do so while on a walkway, or off the walkway? Assume the period of time required is the same in both cases.

(2) Suppose you have a limited amount of energy available to run and increase your speed to a higher quantity v' (or $v'+u$, if you are on a walkway). Is it more efficient to run while on a walkway, or off the walkway? Assume that the energy expenditure is the same in both cases.

(3) Do the answers to the above questions change if one takes into account the various effects of special relativity, such as time dilation and the velocity addition formula? (This is of course an academic question rather than a practical one. But presumably it should be the time in the airport frame that one wants to minimise, not time in one's personal frame.)

It is not too difficult to answer these questions on both a rigorous mathematical level and a physically intuitive level, but ideally one should be able to come up with a satisfying mathematical explanation that also corresponds well with one's intuition.

Notes. This article first appeared at `terrytao.wordpress.com/2008/12/09`. Much discussion on this puzzle (including, of course, the correct solution) can be found in the comments to this article.

1.21. Cohomology for dynamical systems

Recall from Section 2.1 that a dynamical system is[95] a space X, together with an action $(g,x) \mapsto gx$ of some group $G = (G,\cdot)$. A useful notion in the subject is that of an (abelian) *cocycle*; this is a function[96] $\rho : G \times X \to U$ taking values in an abelian group $U = (U,+)$ that obeys the *cocycle equation*

$$\rho(gh,x) = \rho(h,x) + \rho(g,hx) \tag{1.99}$$

for all $g,h \in G$ and $x \in X$. The significance of cocycles in the subject is that they allow one to construct (abelian) *extensions* or *skew products* $X \times_\rho U$ of the original dynamical system X, defined as the Cartesian product $\{(x,u) : x \in X, u \in U\}$ with the group action $g(x,u) := (gx, u + \rho(g,x))$. (The cocycle equation (1.99) is needed to ensure that one indeed has a group action, and in particular that $(gh)(x,u) = g(h(x,u))$.) This turns out to be a useful means to build complex dynamical systems out of simpler ones[97].

A special type of cocycle is a *coboundary*; this is a cocycle $\rho : G \times X \to U$ that takes the form $\rho(g,x) := F(gx) - F(x)$ for some function $F : X \to U$. (Note that the cocycle equation (1.99) is automaticaly satisfied if ρ is of this form.) An extension $X \times_\rho U$ of a dynamical system by a coboundary $\rho(g,x) := F(gx) - F(x)$ can be conjugated to the trivial extension $X \times_0 U$ by the change of variables $(x,u) \mapsto (x, u - F(x))$.

---

[95] In practice, one often places topological or measure-theoretic structure on X or G, see Section 2.2, but this will not be relevant for the current discussion. In most applications, G is an abelian (additive) group such as the integers $\mathbf{Z}$ or the reals $\mathbf{R}$, but I prefer to use multiplicative notation here.

[96] Again, if one is placing topological or measure-theoretic structure on the system, one would want ρ to be continuous or measurable, but we will ignore these issues.

[97] For instance, one can build nilsystems by starting with a point and taking a finite number of abelian extensions of that point by a certain type of cocycle; see Section 2.16.

While every coboundary is a cocycle, the converse is not always true[98]. One can measure the extent to which this converse fails by introducing the *first cohomology group* $H^1(G,X,U) := Z^1(G,X,U)/B^1(G,X,U)$, where $Z^1(G,X,U)$ is the space of cocycles $\rho : G \times X \to U$ and $B^1(G,X,U)$ is the space of coboundaries (note that both spaces are abelian groups). In [**BeTaZi2009**], we make substantial use of some basic facts about this cohomology group (in the category of measure-preserving systems) that were established in a [**HoKr2005**].

The above terminology of cocycles, coboundaries, and cohomology groups of course comes from the theory of *cohomology* in algebraic topology. Comparing the formal definitions of cohomology groups in that theory with the ones given above, there is certainly quite a bit of similarity, but in the dynamical systems literature the precise connection does not seem to be heavily emphasised. The purpose of this post is to record the precise fashion in which dynamical systems cohomology is a special case of cochain complex cohomology from algebraic topology, and more specifically is analogous to singular cohomology (and can also be viewed as the group cohomology of the space of scalar-valued functions on X, when viewed as a G-module); this is not particularly difficult, but I found it an instructive exercise (especially given that my algebraic topology is extremely rusty), though perhaps this article is more for my own benefit that for anyone else.

1.21.1. Chains. Throughout this discussion, the dynamical system X, the group G, and the group U will be fixed.

For any $n \geq 0$, we define an *n-chain* to be a formal integer linear combination of $n+1$-tuples $(g_1,\ldots,g_n,x)$, where $x \in X$ and $g_1,\ldots,g_n \in G$. One may wish to think of each such tuple as an "oriented simplex" connecting the $n+1$ points $x, g_n x, g_{n-1}g_n x, \ldots, g_1 \ldots g_n x$. Thus, a 0-chain is a formal combination $\sum_{i=1}^m c_i x_i$ of points, a 1-chain is a formal combination $\sum_{i=1}^m c_i(g_i,x_i)$ of "line segments" from x_i to $g_i x_i$, and so forth. Let $C_n(G,X)$ be the space of n-chains; this is an

[98]For instance, if X is a point, the only coboundary is the zero function, whereas a cocycle is essentially the same thing as a homomorphism from G to U, so in many cases there will be more cocycles than coboundaries. For a contrasting example, if X and G are finite (for simplicity) and G acts *freely* on X, it is not difficult to see that every cocycle is a coboundary.

abelian group. We also adopt the convention that $C_n(G,X)$ is trivial for $n < 0$.

For each $n > 0$, we define the *boundary map* $\partial : C_n(G,X) \to C_{n-1}(G,X)$ to be the unique homomorphism such that

$$\begin{aligned}\partial(g_1,\ldots,g_n,x) &= (g_1,\ldots,g_{n-1},g_n x)\\ &\quad + \sum_{i=1}^{n-1}(-1)^{n-i}(g_1,\ldots,g_{i-1},g_i g_{i+1},g_{i+2},\ldots,g_n,x)\\ &\quad + (-1)^n(g_2,\ldots,g_n,x)\end{aligned}$$

thus for instance

$$\begin{aligned}\partial(g,x) &= gx - x\\ \partial(g,h,x) &= (g,hx) - (gh,x) + (h,x)\\ \partial(g,h,k,x) &= (g,h,kx) - (g,hk,x) + (gh,k,x) - (h,k,x)\end{aligned}$$

and so forth. Note that this is analogous to the boundary map in singular homology, if one views the $n+1$-tuple $(x,g_1,\ldots,g_n)$ as a simplex as discussed earlier. We also define the boundary maps $\partial : C_n(G,X) \to C_{n-1}(G,X)$ for $n \leq 0$ to be the trivial map, thus for instance $\partial x = 0$. It is not hard to verify the fundamental relation

$$\partial^2 = 0$$

thus turning the sequence of groups $C_n(G,X)$ into a *chain complex*.

An n-chain with vanishing boundary is called an *n-cycle*, while an n-chain which is the boundary of an $(n-1)$-chain is called an *n-boundary*; the spaces of n-cycles and n-boundaries are denoted $Z_n(G,X)$ and $B_n(G,X)$ respectively. Thus for instance $(gh,x) - (h,x) - (g,hx)$ is both a 1-cycle and a 1-boundary. However, if g is a non-trivial group element that fixes x and G is abelian, one can show that (g,x) is a 1-cycle but not a 1-boundary.

We define the *homology groups* $H_n(G,X) := Z_n(G,X)/B_n(G,X)$ for all n. It is a nice exercise to compute these groups in some simple cases, e.g.

- If G acts transitively on X, then $H_0(G,X) \equiv \mathbf{Z}$.
- If G acts freely on X, then $H_n(G,X)$ is trivial for $n > 0$.

- If X is a point, then $H_1(G,X) \equiv G/[G,G]$ is the abelianisation of G. [Question: Is there a nice description of the higher homology groups $H_n(G,X)$, $n>1$ in this case?]

However, I don't know of any application of these homology groups to the theory of dynamical systems.

1.21.2. Cochains. An *n-cochain* is a homomorphism from the space $C_n(G,X)$ of n-chains to U. Since $C_n(G,X)$ is a free abelian group generated by the simplices $(x,g_1,\ldots,g_n)$, we can view an n-cochain as a function $F:(x,g_1,\ldots,g_n)\to F(x,g_1,\ldots,g_n)$ from $G\times\ldots\times G\times X$ to U. (Again, we are ignoring all measure-theoretic or topological considerations here.) The space of all n-cochains is denoted $C^n(G,X,U) := \mathrm{Hom}(C_n(G,X),U)$; this is an abelian group.

The boundary map $\partial: C_n(G,X)\to C_{n-1}(G,X)$ defines by duality a coboundary map $\delta: C^{n-1}(G,X,U)\to C^n(G,X,U)$, defined by the formula

$$\delta F(c) := F(\partial c)$$

for all $F\in C^{n-1}(G,X,U)$ and $c\in C_n(G,X)$; viewing F as a function on simplices, we thus have

$$\begin{aligned}\delta F(g_1,\ldots,g_n,x) &= F(g_1,\ldots,g_{n-1},g_n x)\\ &\quad+\sum_{i=1}^{n-1}(-1)^{n-i}F(g_1,\ldots,g_{i-1},g_ig_{i+1},\ldots,g_n,x)\\ &\quad+(-1)^n F(g_2,\ldots,g_n,x).\end{aligned}$$

Thus for instance

$$\delta F(g,x) = F(gx)-F(x)$$

for 0-cochains $F: X\to U$,

$$\delta\rho(g,h,x) = \rho(g,hx)-\rho(gh,x)+\rho(g,x)$$

for 1-cochains $\rho: G\times X\to U$, and so forth.

Because $\partial^2=0$, we have $\delta^2=0$, and so $C^n(G,X,U)$ becomes a *cochain complex*. n-cochains whose coboundary vanishes are known as *n-cocycles*, and n-cochains which are the coboundary of an $(n-1)$-cochain are known as *n-coboundaries*. The spaces of n-cocycles

and n-cochains are denoted $Z^n(G,X,U)$ and $B^n(G,X,U)$ respectively, allowing us to define the n^{th} *cohomology group* $H^n(G,X,U) := Z^n(G,X,U)/B^n(G,X,U)$.

When $n=0$, and if the action of G is *transitive* (in the discrete category), *minimal* (in the topological category, see Section 2.2), or *ergodic* (in the measure-theoretic category, see Section 2.9), the only 0-cocycles are the constants, and the only 0-coboundary is the zero function, so $H^0(G,X,U) \equiv U$. When $n=1$, it is not hard to see that the notion of 1-cocycle and 1-coboundary correspond to the notion of cocycle and coboundary discussed at the beginning of this post.

This whole theory raises the obvious question as to whether the higher cocycles, coboundaries, and cohomology groups have any relevance in dynamical systems. For instance, a 2-cocycle is (after minor notational changes) a function $\psi : G \times G \times X \to U$ that obeys the 2-cocycle equation

$$\psi(g,h,kx) - \psi(g,hk,x) + \psi(gh,k,x) - \psi(h,k,x) = 0$$

while a 2-coboundary is a function of the form

$$\psi(g,h,x) := \rho(gh,x) - \rho(h,x) - \rho(g,hx)$$

for some $\rho : G \times X \to U$. Is there some dynamical systems interpretation of these objects, much as 1-cocycles and 1-coboundaries can be interpreted as describing abelian extensions and essentially trivial abelian extensions respectively? (See Section 1.21.3 below for a partial answer.) In [**BeTaZi2009**], we do briefly encounter 2-coboundaries (we have to deal with various "quasi-cocycles" - 1-chains ρ whose 2-coboundary $\delta\rho$ does not vanish completely, as with 1-cocycles, but is still of a relatively simple form, such as a constant or a polynomial) but we do not make systematic use of this concept. (We also rely heavily in our paper on the cubic complexes $X^{[k]}$ of Host and Kra, which have some superficial resemblance to the simplex structures appearing here, but I do not know if there is a substantive connection in this regard.)

Another oddity is that homology and cohomology, as it is classically defined, requires the space of chains, cochains, etc. to all be abelian groups; but for dynamical systems one can certainly talk about cocycles and coboundaries taking values in a non-abelian group

U by modifying the definitions slightly, leading to the concept of a *group extension* of a dynamical system. (In this context, the first cohomology $H^1(G, X, U)$ becomes a quotient space rather than a group; see also Section 1.10) It seems to me that in this case, the dynamical system concept of a cocycle or coboundary cannot be interpreted in terms of classical cohomology theory (but presumably can be handled by *non-abelian group cohomology*).

1.21.3. Epilogue - an interpretation of the second cohomology group. Minhyong Kim has provided a nice answer to my question about the relevance of higher order cohomology, such as $H^2(G, X, U)$, to the problem of extending dynamical systems. Suppose one has a short exact sequence

$$0 \to V \to \tilde{U} \to U \to 0$$

of abelian groups, thus one can view $\tilde{U}$ as the space of pairs (u, v) with $u \in U, v \in V$ with some group addition law

$$(u, v) + (u', v') := (u + u', v + v' + B(u, u')) \tag{1.100}$$

for some function $B : U \times U \to V$, that needs to obey a certain set of axioms to make $\tilde{U}$ an abelian group, which we will not write down here. We then claim that we have a long exact sequence

$$\to H^1(G, X, \tilde{U}) \to H^1(G, X, U) \to H^2(G, X, V) \to, \tag{1.101}$$

thus $H^2(G, X, V)$ is capable of detecting whether a U-extension of a G-system X can be lifted to a $\tilde{U}$-extension.

The first map in (1.101) is obvious: the projection from $\tilde{U}$ to U induces a projection from 1-cocycles $\tilde{\rho} : G \times X \to \tilde{U}$ to 1-cocycles $\rho : G \times X \to U$ which maps 1-coboundaries to 1-coboundaries, and thus maps $H^1(G, X, \tilde{U})$ to $H^1(G, X, U)$. The second map requires a bit more thought. Suppose one is given a 1-cocycle $\rho : G \times X \to U$ and asks whether it can be lifted to a 1-cocycle $\tilde{\rho} : G \times X \to \tilde{U}$ by the above projection. Writing $\tilde{\rho} = (\rho, \sigma)$ for some $\sigma : G \times X \to V$ and using (1.99), (1.100), we see that the question is equivalent to finding a σ that obeys the equation

$$\sigma(gh, x) = \sigma(h, x) + \sigma(g, hx) + B(\rho(h, x), \rho(g, hx)),$$

or in other words, to show that the map $\Phi(\rho) : (g, h, x) \mapsto B(\rho(h, x), \rho(g, hx))$ is a V-valued 2-coboundary. The same observation (now setting $\sigma = 0$) shows that the map $(g, h, x) \mapsto (0, \Phi(\rho))$ is a $\tilde{U}$-valued 2-coboundary (indeed, it is the coboundary of $(\rho, 0)$), hence a $\tilde{U}$-valued 2-cocycle, and thus $\Phi(\rho)$ is a V-valued 2-cocycle, and so the map $\rho \mapsto \Phi(\rho)$ is a map from 1-cocycles $\rho : G \times X \to U$ to 2-cocycles $\Phi(\rho) : G \times G \times X \to V$. Similarly, given two 1-cocycles $\rho, \rho' : G \times X \to U$, we see that $(\rho + \rho', 0)$ differs from $(\rho, 0) + (\rho', 0)$ by some V-valued 1-cochain, so on taking derivatives we see that $\Phi(\rho + \rho')$ differs from $\Phi(\rho) + \Phi(\rho')$ by some 2-coboundary, thus Φ is linear modulo 2-coboundaries. Finally, if ρ is a U-valued 1-coboundary, then $(\rho, 0)$ is the sum of a $\tilde{U}$-valued 1-coboundary and a V-valued 1-cochain, and so on taking derivatives we see that Φ maps 1-coboundaries to 2-coboundaries[99]. Hence it induces a map from $H^1(G, X, U)$ to $H^2(G, X, V)$, and then (1.101) is exact by the preceding discussion.

Notes. This article first appeared at `terrytao.wordpress.com/2008/12/21`. Thanks to AA for corrections.

Mikael Vejdemo Johansson pointed out that the group cohomology formalism developed above also extends to bimodules over G, though it is not clear what the dynamical interpretation of such bimodules would be.

Marlowe noted more generally that as a general rule of thumb, if a certain cohomology group helps to classify extensions up to conjugation, the next cohomology group helps you find out if a certain candidate for an extension can be extended to a full extension; the discussion in Section 1.21.3 of course supports this rule.

Peter Samuelson also pointed out that this homology is a special case of Hochschild homology.

Further discussion on this topic can also be found at `http://golem.ph.utex`

[99] Presumably the above arguments are a special case of one of the standard diagram chasing lemmas in homological algebra, but I don't know which one it is. One could also verify these facts from the axioms of B induced from (1.100) and the abelian group structure on $\tilde{U}$, but this turns out to be remarkably tedious.

1.22. A remark on the Kakeya needle problem

Recall from Section 1.3 that given any $\varepsilon > 0$, there exists a planar set of area at most ε within which a unit needle can be continuously rotated. I was recently asked (by Claus Dollinger) whether one can take $\varepsilon = 0$; in other words,

Question 1.22.1. *Does there exist a set of measure* zero *within which a unit line segment can be continuously rotated by a full rotation?*

This question does not seem to be explicitly answered in the literature. In [**vA1942**], [**Cu1971**] it is shown that it is possible to continuously rotate a unit line segment inside a set of arbitrarily small measure and of uniformly bounded diameter; this result is of course implied by a positive answer to the above question (since continuous functions on compact sets are bounded), but the converse is not true.

In this note, I show that the answer to the question is negative.

Proof. Let $E \subset \mathbf{R}^2$ be a set in the plane within which a unit line segment can be continuously rotated. This means that there exists a continuous map $l : t \mapsto l(t)$ from times $t \in [0,1]$ to unit line segments $l(t) \subset E$. We can parameterise each such line segment as

$$l(t) = \{(x(t) + s\cos\omega(t), y(t) + s\sin\omega(t)) : -0.5 \leq s \leq 0.5\}$$

where $x, y, \omega : [0,1] \to \mathbf{R}$ are continuous functions.

Recall that on a compact set, all continuous functions are uniformly continuous. In particular, there exists $\varepsilon > 0$ such that

$$|x(t) - x(t')|, |y(t) - y(t')|, |\omega(t) - \omega(t')| \leq 0.001 \tag{1.102}$$

(say) whenever $t, t' \in [0,1]$ are such that $|t - t'| \leq \varepsilon$.

Fix this ε. Observe that $\omega(t)$ cannot be a constant function of t, otherwise the needle would never rotate. We conclude that there must exist $t_0, t_1 \in [0,1]$ with $|t_0 - t_1| \leq \varepsilon$ and $\omega(t_0) \neq \omega(t_1)$.

Without loss of generality, we may assume that $t_0 < t_1$ and $x(t_0) = y(t_0) = \omega(t_0) = 0$. Now let a be any real number between -0.4 and $+0.4$. From (1.102) we see that for any $t_0 \leq t \leq t_1$, the line $l(t)$ intersects the line $x = a$ in some point $(a, y_a(t))$, which must therefore lie in E. Furthermore, $y_a(t)$ varies continuously in t. By

the intermediate value theorem, we conclude that the interval between $(a, y_a(t_0))$ and $(a, y_a(t_1))$ lies in E. Taking unions over all a between -0.4 and $+0.4$, we see that E contains a non-trivial sector, and thus has non-zero area. The claim follows. □

Remark 1.22.2. A variant of this argument shows a stronger statement, namely that for any fixed $c > 0$, any set E whose measure is sufficiently small (depending on c) within which a unit line segment can be rotated by at least c, must have a diameter of at least $2 - c$. (A similar point was already made in [**Cu1971**].)

Notes. This article first appeared at `terrytao.wordpress.com/2008/12/31`.

Chapter 2

Ergodic theory

2.1. Overview

In this lecture, I define the basic notion of a *dynamical system* (as well as the more structured notions of a *topological dynamical system* and a *measure-preserving system*), and describe the main topics we will cover in this course.

We'll begin abstractly. Suppose that X is a non-empty set (whose elements will be referred to as *points*), and $T : X \to X$ is a transformation. Later on we shall put some structures on X (such as a topology, a σ-algebra, or a probability measure), and some assumptions on T, but let us work in total generality for now[1].

One can think of X as a state space for some system, and T as the evolution of some discrete deterministic (autonomous) dynamics on X: if x is a point in X, denoting the current state of a system, then Tx can be interpreted as the state of the same system after one unit of time has elapsed[2]. More geometrically, one can think of T as some sort of shift operation (e.g. a rotation) on the space X.

Given X and T, we can define the iterates $T^n : X \to X$ for every non-negative integer n; if T is also invertible, then we can also define T^n for negative integer n as well. In the language of representation theory, T induces a representation[3] of either the additive semigroup $\mathbf{Z}^+$ or the additive group $\mathbf{Z}$. More generally, one can consider representations of other groups, such as the real line $\mathbf{R}$ (corresponding the dynamics $t \mapsto T^t$ of a continuous time evolution) or a lattice $\mathbf{Z}^d$ (which corresponds to the dynamics of d commuting shift operators $T_1, \ldots, T_d : X \to X$), or of many other semigroups or groups (not necessarily commutative). However, for simplicity we shall mostly restrict our attention to $\mathbf{Z}$-actions in this course, though many of the results here can be generalised to other actions (under suitable hypotheses on the underlying semigroup or group, of course).

---

[1] Indeed, a guiding philosophy in the first half of the course will be to try to study dynamical systems in as maximal generality as possible; later on, though, when we turn to more algebraic dynamical systems such as nilsystems, we shall exploit the specific structure of such systems more thoroughly.

[2] In particular, evolution equations which are well-posed can be viewed as a continuous dynamical system.

[3] From the dynamical perspective, this representation is the mathematical manifestation of *time*.

Henceforth we assume T to be invertible, in which case we refer to the pair (X, T) as a *cyclic dynamical system*, or *dynamical system* for short. Here are some simple examples of such systems:

Example 2.1.1 (Finite systems)**.** X is a finite set, and $T : X \to X$ is a permutation on X.

Example 2.1.2 (Group actions)**.** Let G be a group, and let X be a homogeneous space for G, i.e. a non-empty space with a transitive G-action; thus X is isomorphic to G/Γ, where $\Gamma := \mathrm{Stab}(x)$ is the stabiliser of one of the points x in X. Then every group element $g \in G$ defines a dynamical system (X, T_g) defined by $T_g x := gx$.

Example 2.1.3 (Circle rotations)**.** As a special case of Example 2.1.2 (or Example 2.1.1), every real number $\alpha \in \mathbf{R}$ induces a dynamical system $(\mathbf{R}/\mathbf{Z}, T_\alpha)$ given by the rotation $T_\alpha x := x + \alpha$. This is the prototypical example of a very *structured* system, with plenty of algebraic structure (e.g. the shift map T_α is an isometry on the circle, thus two points always stay the same distance apart under shifts).

Example 2.1.4 (Cyclic groups)**.** Another special case of Example 2.1.2 is the cyclic group $\mathbf{Z}/N\mathbf{Z}$ with shift $x \mapsto x + 1$; this is the prototypical example of a finite dynamical system.

Example 2.1.5 (Bernoulli systems)**.** Every non-empty set Ω induces a dynamical system $(\Omega^{\mathbf{Z}}, T)$, where T is the left shift $T(x_n)_{n\in\mathbf{Z}} := (x_{n+1})_{n\in\mathbf{Z}}$. This is the prototypical example of a very *pseudorandom* system, with plenty of mixing (e.g. the shift map tends to move a pair of two points randomly around the space).

Example 2.1.6 (Boolean Bernoulli system)**.** This is isomorphic to a special case of Example 2.1.5, in which $X = 2^{\mathbf{Z}} := \{A : A \subset \mathbf{Z}\} \equiv \{0,1\}^{\mathbf{Z}}$ is the power set of the integers, and $TA := A - 1 := \{a - 1 : a \in A\}$ is the left shift. (Here we endow $\{0,1\}$ with the discrete topology.)

Example 2.1.7 (Baker's map)**.** Here, $X := [0,1)^2$, and $T(x,y) := (\{2x\}, \frac{y+\lfloor 2x\rfloor}{2})$, where $\lfloor x\rfloor$ is the greatest integer function, and $\{x\} := x - \lfloor x\rfloor$ is the fractional part. This is isomorphic to Example 2.1.6, as can be seen by inspecting the effect of T on the binary expansions of x and y.

The map T^n can be interpreted as an isomorphism in several different categories:

(1) as a set isomorphism (i.e. a bijection) $T^n : X \to X$ from points $x \in X$ to points $T^n x \in X$;

(2) as a Boolean algebra isomorphism $T^n : 2^X \to 2^X$ from sets $E \subset X$ to sets $T^n E := \{T^n x : x \in E\}$; or

(3) as an algebra isomorphism $T^n : \mathbf{R}^X \to \mathbf{R}^X$) from real-valued functions $f : X \to \mathbf{R}$ to real-valued functions $T^n f : X \to \mathbf{R}$, defined by

$$T^n f(x) := f(T^{-n} x); \tag{2.1}$$

(4) as an algebra isomorphism $T^n : \mathbf{C}^X \to \mathbf{C}^X$ of complex valued functions, defined again by (2.1).

We will abuse notation and use the same symbol T^n to refer to all of the above isomorphisms; the specific meaning of T^n should be clear from context in all cases. Our sign conventions here are chosen so that we have the pleasant identities

$$T^n\{x\} = \{T^n x\}; \quad T^n 1_E = 1_{T^n E} \tag{2.2}$$

for all points x and sets E, where of course 1_E is the *indicator function* of E.

One of the main topics of study in dynamical systems is the asymptotic behaviour of T^n as $n \to \infty$. We can pose this question in any of the above categories, thus

(1) For a given point $x \in X$, what is the behaviour of $T^n x$ as $n \to \infty$?

(2) For a given set $E \subset X$, what is the behaviour of $T^n E$ as $n \to \infty$?

(3) For a given real or complex-valued function $f : X \to \mathbf{R}$ or $f : X \to \mathbf{C}$, what is the behaviour of $T^n f$ as $n \to \infty$?

These are of course very general and vague questions, but we will formalise them in many different ways later in the course[4]. The answer to these questions also depends very much on the dynamical system;

[4]For instance, one can distinguish between worst-case, average-case, and best-case behaviour in x, E, f, or n.

thus a major focus of study in this subject is to seek classifications of dynamical systems which allow one to answer the above questions satisfactorily[5].

One can also ask for more *quantitative* versions of the above asymptotic questions, in which n ranges in a finite interval (e.g. $[N] := \{1, \ldots, N\}$ for some large integer N), as opposed to going off to infinity, and one wishes to estimate various numerical measurements of $T^n x$, $T^n E$, or $T^n f$ in this range.

In this very general setting, in which X is an unstructured set, and T is an arbitrary bijection, there is not much of interest one can say with regards to these questions. However, one obtains a surprisingly rich and powerful theory when one adds a little bit more structure to X and T (thus changing categories once more). In particular, we will study the following two structured versions of a dynamical system:

(I) *Topological dynamical systems* $(X, T) = (X, \mathcal{F}, T)$, in which $X = (X, \mathcal{F})$ is a compact metrisable (and thus *Hausdorff*) topological space, and T is a topological isomorphism (i.e. a homeomorphism); and

(II) *Measure-preserving systems* $(X, T) = (X, \mathcal{X}, \mu, T)$, in which $X = (X, \mathcal{X}, \mu)$ is a probability space[6], and T is a probability space isomorphism, i.e. T and T^{-1} are both measurable, and $\mu(TE) = \mu(E)$ for all measurable $E \in \mathcal{X}$. For technical reasons we also require the measurable space $(X, \mathcal{X})$ to be *separable* (i.e. $\mathcal{X}$ is countably generated).

Remark 2.1.8. By *Urysohn's metrisation theorem*, a compact space is metrisable if and only if it is Hausdorff and *second countable*, thus providing a purely topological characterisation of a topological dynamical system.

Remark 2.1.9. It is common to add a bit more structure to each of these systems, for instance endowing a topological dynamical system

[5]In particular, ergodic theory is a framework in which our understanding of the dichotomy between structure and randomness is at is most developed; see Section 2.1.2 of *Structure and Randomness*.

[6]In this course we shall tilt towards a measure-theoretic perspective rather than a probabilistic one, thus it might be better to think of μ of as a normalised finite measure rather than as a probability measure. On the other hand, we will rely crucially on the probabilistic notions of *conditional expectation* and *conditional independence* later in this course.

with a metric, or endowing a measure preserving system with the structure of a standard Borel space; we will see examples of this in later lectures.

The study of topological dynamical systems and measure-preserving systems is known as *topological dynamics* and *ergodic theory* respectively. The two subjects are closely analogous at a heuristic level, and also have some more rigorous connections between them, so we shall pursue them in a somewhat parallel fashion in this course.

Remark 2.1.10. Observe that we assume compactness in (I) and finite measure in (II); these "boundedness" assumptions ensure that the dynamics somewhat resembles the (overly simple) case of a finite dynamical system. Dynamics on non-compact topological spaces or infinite measure spaces is a more complicated topic; see for instance [**Aa1997**]. (Thanks to Tamar Ziegler for this reference.)

Note that the action of the isomorphism T^n on sets E and functions f will be compatible with the topological or measure-theoretic structure:

(1) If $(X,T) = (X,\mathcal{F},T)$ is a topological dynamical system, then $T^n : \mathcal{F} \to \mathcal{F}$ is a topological isomorphism on open sets, and $T^n : C(X) \to C(X)$ is also a C^*-algebra isomorphism on the space $C(X)$ of real-valued (or complex-valued) continuous functions on X.

(2) If $(X,T) = (X,\mathcal{X},\mu,T)$ is a measure-preserving system, then $T^n : \mathcal{X} \to \mathcal{X}$ is a σ-algebra isomorphism on measurable sets, and $T^n : L^p(\mathcal{X},\mu) \to L^p(\mathcal{X},\mu)$ is a Banach space isomorphism on p^{th}-power integrable functions for $1 \leq p \leq \infty$. (For $p = \infty$, T^n is a von Neumann algebra isomorphism, whilst for $p = 2$, T^n is a Hilbert space isomorphism (i.e. a unitary transformation).)

We can thus see that tools from the analysis of Banach spaces, von Neumann algebras, and Hilbert spaces may have some relevance to ergodic theory; for instance, the spectral theorem for unitary operators is quite useful.

In the first half of this course, we will study topological dynamical systems and measure-preserving systems in great generality (with few assumptions on the structure of such systems), and then specialise to specific systems as appropriate. This somewhat abstract approach is broadly analogous to the combinatorial (as opposed to algebraic or arithmetic) approach to additive number theory. For instance, we will shortly be able to establish the following general result in topological dynamics (see Theorem 2.3.4):

Theorem 2.1.11 (Birkhoff recurrence theorem)**.** *Let (X,T) be a topological dynamical system. Then there exists a point $x \in X$ which is* recurrent *in the sense that there exists a sequence $n_j \to \infty$ such that $T^{n_j}x \to x$ as $j \to \infty$.*

As a corollary, we will be able to obtain the more concrete result (see Section 2.4):

Theorem 2.1.12 (Weyl recurrence theorem)**.** *Let $P : \mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ be a polynomial (modulo 1). Then there exists a sequence $n_j \to \infty$ such that $P(n_j) \to P(0)$.*

This is already a somewhat non-trivial theorem; consider for instance the case $P(n) := \sqrt{2}n^2 \bmod 1$.

In a similar spirit, in Section 2.4 we will be able to prove the general topological dynamical result (see Theorem 2.4.1):

Theorem 2.1.13 (Topological van der Waerden theorem)**.** *Let $(U_\alpha)_{\alpha\in A}$ be an open cover of a topological dynamical system (X,T), and let $k \geq 1$ be an integer. Then there exists an open set U in this cover and a shift $n \geq 1$ such that $U \cap T^n U \cap \ldots \cap T^{(k-1)n}U \neq \emptyset$. (Equivalently, there exists U, n, and a point x such that $x, T^n x, \ldots, T^{(k-1)n}x \in U$.)*

and conclude an (equivalent) combinatorial result:

Theorem 2.1.14 (van der Waerden theorem)**.** *Let $\mathbf{N} = U_1 \cup \ldots \cup U_m$ be a finite colouring of the natural numbers. Then one of the colour classes U_j contains arbitrarily long arithmetic progressions.*

More generally, topological dynamics is an excellent tool for establishing colouring theorems of Ramsey type.

Analogously, in Sections 2.10-2.15 we will be able to show the following general ergodic theory result (see Theorem 2.10.3)):

Theorem 2.1.15 (Furstenberg multiple recurrence theorem)**.** *Let (X,T) be a measure-preserving system, let $E \in \mathcal{X}$ be a set of positive measure, and let $k \geq 1$. Then there exists $n \geq 1$ such that $E \cap T^n E \cap \ldots \cap T^{(k-1)n} E \neq \emptyset$ (or equivalently, there exists $x \in X$ and $n \geq 1$ such that $x, T^n x, \ldots, T^{(k-1)n} \in E$).*

Similarly, if $f : X \to \mathbf{R}^+$ is a bounded measurable non-negative function which is not almost everywhere zero, and $k \geq 1$, then

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X f T^n f \ldots T^{(k-1)n} f > 0. \tag{2.3}$$

and deduce an equivalent (and highly non-trivial) combinatorial analogue (see Theorem 2.10.1):

Theorem 2.1.16 (Szemerédi's theorem)**.** *Let $E \subset \mathbf{Z}$ be a set of positive upper density, thus $\limsup_{N\to\infty} \frac{|E\cap[-N,N]|}{2N+1} > 0$. Then E contains arbitrarily long arithmetic progressions.*

More generally, ergodic theory methods are extremely powerful in deriving *density Ramsey theorems*. Indeed, there are several theorems of this type which currently have no known non-ergodic theory proof[7].

The first half of this course will be devoted to results of the above type, which apply to general topological dynamical systems or general measure-preserving systems. One important insight that will emerge from analysis of the latter is that in many cases, a large portion of the measure-preserving system is irrelevant for the purposes of understanding long-time average behaviour; instead, there will be a smaller system, known as a *characteristic factor* for the system, which completely controls these asymptotic averages. A deep and powerful fact is that in many situations, this characteristic factor is extremely structured algebraically, even if the original system has no obvious algebraic structure whatsoever. Because of this, it becomes important

[7]From general techniques in proof theory, one could, in principle, take an ergodic theory proof and mechanically convert it into what would technically be a non-ergodic proof, for instance avoiding the use of infinitary objects, but this is not really in the spirit of what most mathematicians would call a genuinely new proof.

to study algebraic dynamical systems, such as the group actions on homogeneous spaces described earlier, as it allows one to obtain more precise results[8]. This study will be the focus of the second half of the course, particularly in the important case of *nilsystems* - group actions arising from a nilpotent Lie group with discrete stabiliser. One of the key results here is *Ratner's theorem*, which describes the distribution of orbits $\{T^n x : n \in \mathbf{Z}\}$ in nilsystems, and also in a more general class of group actions on homogeneous spaces. While we will not prove Ratner's theorem in full generality, we will cover a few special cases of this theorem in Sections 2.16, 2.17.

In closing, I should mention that the topics I intend to cover in this course are only a small fraction of the vast area of ergodic theory and dynamical systems; for instance, there are parts of this field connected with complex analysis and fractals, ODE, probability and information theory, harmonic analysis, group theory, operator algebras, or mathematical physics which I will say absolutely nothing about here.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/01/08`.

2.2. Three categories of dynamical systems

Before we begin or study of dynamical systems, topological dynamical systems, and measure-preserving systems (as defined in Section 2.1), it is convenient to give these three classes the structure of a *category*. One of the basic insights of category theory is that a mathematical objects in a given class (such as dynamical systems) are best studied not in isolation, but in relation to each other, via *morphisms*. Furthermore, many other basic concepts pertaining to these objects (e.g. subobjects, factors, direct sums, irreducibility, etc.) can be defined in terms of these morphisms. One advantage of taking this perspective here is that it provides a unified way of defining these concepts for the three different categories of dynamical systems, topological dynamical systems, and measure-preserving systems that we will study

[8]For instance, this algebraic structure was used to show that the limit in (2.3) actually converges, a result which does not seem accessible purely through the techniques used to prove the Furstenberg recurrence theorem.

in this course, thus sparing us the need to give any of our definitions (except for our first one below) in triplicate.

Informally, a *morphism* between two objects in a class is any map which respects all the structures of that class. For the three categories we are interested in, the formal definition is as follows.

Definition 2.2.1 (Morphisms)**.** (1) A *morphism* $\phi : (X, T) \to (Y, S)$ between two dynamical systems is a map $\phi : X \to Y$ which intertwines T and S in the sense that $S \circ \phi = \phi \circ T$.

(2) A *morphism* $\phi : (X, \mathcal{F}, T) \to (Y, \mathcal{G}, S)$ between two topological dynamical systems is a morphism $\phi : (X, T) \to (Y, S)$ of dynamical systems which is also continuous, thus $\phi^{-1}(U) \in \mathcal{F}$ for all $U \in \mathcal{G}$.

(3) A *morphism* $\phi : (X, \mathcal{X}, \mu, T) \to (Y, \mathcal{Y}, \nu, S)$ between two measure-preserving systems is a morphism $\phi : (X, T) \to (Y, S)$ of dynamical systems which is also measurable (thus $\phi^{-1}(E) \in \mathcal{X}$ for all $E \in \mathcal{Y}$) and measure-preserving (thus $\mu(\phi^{-1}(E)) = \nu(E)$ for all $E \in \mathcal{Y}$). Equivalently, $\nu = \phi_*(\mu)$ is the *push-forward* of μ by ϕ.

When it is clear what category we are working in, and what the shifts are, we shall often refer to a system by its underlying space, thus for instance a morphism $\phi : (X, \mathcal{X}, \mu, T) \to (Y, \mathcal{Y}, \nu, S)$ might be abbreviated as $\phi : X \to Y$.

If a morphism $\phi : X \to Y$ has an inverse $\phi^{-1} : Y \to X$ which is also a morphism, we say that ϕ is an *isomorphism*, and that X and Y are *isomorphic* or *conjugate*.

It is easy to see that morphisms obey the axioms of a (concrete) category, or in other words that the identity map $\mathrm{id}_X : X \to X$ on a system is always a morphism, and the composition $\psi \circ \phi : X \to Z$ of two morphisms $\phi : X \to Y$ and $\psi : Y \to Z$ is again a morphism.

Let's give some simple examples of morphisms.

Example 2.2.2 (Shift)**.** If (X, T) is a dynamical system, a topological dynamical system, or a measure-preserving dynamical system, then $T^n : X \to X$ is an isomorphism for any integer n. (Indeed, one can view the map $X \mapsto T^n$ as a *natural transformation* from the

identity functor on the category of dynamical systems (or topological dynamical systems, etc.) to itself, although we will not take this perspective here.)

Example 2.2.3 (Subsystems)**.** Let (X, T) be a dynamical system, and let E be a subset of X which is T-invariant in the sense that $T^n E = E$ for all n. Then the restriction of $(E, T\ \lfloor_E)$ of (X, T) to E is itself a dynamical system, and the inclusion map $\iota : E \to X$ is a morphism. In the category of topological dynamical systems $(X, \mathcal{F}, T)$, we have the same assertion so long as E is *closed* (hence compact, since X is compact). In the category of measure-preserving systems $(X, \mathcal{X}, \mu, T)$, we have the same assertion so long as E has full measure (thus $E \in \mathcal{X}$ and $\mu(E) = 1$). We thus see that subsystems are not very common in measure-preserving systems and will in fact play very little role there; however, subsystems (and specifically, *minimal* subsystems) will play a fundamental role in topological dynamics.

Example 2.2.4 (Skew shift)**.** Let $\alpha \in \mathbf{R}$ be a fixed real number. Let (X, T) be the dynamical system $X := (\mathbf{R}/\mathbf{Z})^2, T : (x_1, x_2) \mapsto (x_1 + \alpha, x_2 + x_1)$, let (Y, S) be the dynamical system $Y := \mathbf{R}/\mathbf{Z}, S : y \mapsto y + \alpha$, and let $\pi : X \to Y$ be the projection map $\pi : (x_1, x_2) \to x_1$. Then π is a morphism. If one converts X and Y into either a topological dynamical system or a measure-preserving system in the obvious manner, then π remains a morphism. Observe that π foliates the big space X "upstairs" into "vertical" fibres $\pi^{-1}(\{y\}), y \in Y$ indexed by the small "horizontal" space "downstairs"; the shift S on the factor space Y downstairs determines how the fibres move (the shift T upstairs sends each vertical fibre $\pi^{-1}(\{y\})$ to another vertical fibre $\pi^{-1}(\{Sy\})$, but does not govern the dynamics *within* each fibre. More generally, any *factor map* (i.e. a surjective morphism) exhibits this type of behaviour[9].

Example 2.2.5 (Universal pointed dynamical system)**.** Let $\mathbf{Z} = (\mathbf{Z}, +1)$ be the dynamical system given by the integers with the standard shift $n \mapsto n+1$. Then given any other dynamical system (X, T)

[9]Another example of a factor map is the map $\pi : \mathbf{Z}/N\mathbf{Z} \to \mathbf{Z}/M\mathbf{Z}$ between two cyclic groups (with the standard shift $x \mapsto x + 1$) given by $\pi : x \mapsto x \bmod M$. This is a well-defined factor map when M is a factor of N, which may help explain the terminology. If we wanted to adhere strictly to the category theoretic philosophy, we should use *epimorphisms* rather than surjections, but we will not require this subtle distinction here.

with a distinguished point $x \in X$, the orbit map $\phi : n \mapsto T^n x$ is a morphism from $\mathbf{Z}$ to X. This allows us to lift most questions about dynamical systems (with a distinguished point x) to those for a single "universal" dynamical system, namely the integers (with distinguished point 0). One cannot pull off the same trick directly with topological dynamical systems or measure-preserving systems, because $\mathbf{Z}$ is non-compact and does not admit a shift-invariant probability measure. As we shall see later, the former difficulty can be resolved by passing to a universal compactification of the integers, namely the *StoneCěch compactification* $\beta\mathbf{Z}$ (or equivalently, the space of *ultrafilters* on the integers), though with the important caveat that this compactification is not metrisable. To resolve the second difficulty (with the assistance of a distinguished set rather than a distinguished point), see the next example.

Example 2.2.6 (Universal dynamical system with distinguished set)**.** Recall the boolean Bernoulli system $(2^{\mathbf{Z}}, U)$ (Example 2.1.6). Given any other dynamical system (X, T) with a distinguished set $A \subset X$, the *recurrence map* $\phi : X \to 2^{\mathbf{Z}}$ defined by $\phi(x) := \{n \in \mathbf{Z} : T^n x \in A\}$ is a morphism. Observe that $A = \phi^{-1}(B)$, where B is the *cylinder set* $B := \{E \in 2^{\mathbf{Z}} : 0 \in E\}$. Thus we can push forward an arbitrary dynamical system (X, T, A) with distinguished set to a universal dynamical system $(2^{\mathbf{Z}}, U, B)$. Actually one can restrict $(2^{\mathbf{Z}}, U, B)$ to the subsystem $(\phi(X), U \restriction_{\phi(X)}, B \cap \phi(X))$, which is easily seen to be shift-invariant. In the category of topological dynamical systems, the above assertions still hold (giving $2^{\mathbf{Z}}$ the *product topology*), so long as A is clopen. In the category of measure-preserving systems $(X, \mathcal{X}, \mu, T)$, the above assertions hold as long as A is measurable, $2^{\mathbf{Z}}$ is given the product σ-algebra, and the *push-forward measure* $\phi_*(\mu)$.

Now we begin our analysis of dynamical systems. When studying other mathematical objects (e.g. groups or representations), often one of the first steps in the theory is to decompose general objects into "irreducible" ones, and then hope to classify the latter. Let's see how this works for dynamical systems (X, T) and topological dynamical systems $(X, \mathcal{F}, T)$. (For measure-preserving systems, the analogous decomposition will be the *ergodic decomposition*, which we will discuss in Section 2.9.5.)

Definition 2.2.7. A *minimal dynamical system* is a system (X, T) which has no proper subsystems (Y, S). A *minimal topological dynamical system*[10] is a system $(X, \mathcal{F}, T)$ with no proper subsystems $(Y, \mathcal{G}, S)$.

For a dynamical system, it is not hard to see that for any $x \in X$, the orbit $Y = T^{\mathbf{Z}}x = \{T^n x : n \in \mathbf{Z}\}$ is a minimal system, and conversely that all minimal systems arise in this manner; in particular, every point is contained in a minimal orbit. It is also easy to see that any two minimal systems (i.e. orbits) are either disjoint or coincident. Thus every dynamical system can be uniquely decomposed into the disjoint union of minimal systems. Also, every orbit $T^{\mathbf{Z}}x$ is isomorphic to $\mathbf{Z}/\mathrm{Stab}(x)$, where $\mathrm{Stab}(x) := \{n \in \mathbf{Z} : T^n x = x\}$ is the *stabiliser group* of x. Since we know what all the subgroups of $\mathbf{Z}$, we conclude that every minimal system is either equivalent to a cyclic group shift $(\mathbf{Z}/N\mathbf{Z}, x \mapsto x+1)$ for some $N \geq 1$, or to the integer shift $(\mathbf{Z}, x \mapsto x+1)$. Thus we have completely classified all dynamical systems up to isomorphism as the arbitrary union of these minimal examples[11].

For topological dynamical systems, it is still true that any two minimal systems are either disjoint or coincident (why?), but the situation nevertheless is more complicated. First of all, orbits need not be closed (consider for instance the circle shift $(\mathbf{R}/\mathbf{Z}, x \mapsto x + \alpha)$ with α irrational). If one considers the *orbit closure* $\overline{T^{\mathbf{Z}}x}$ of a point x, then this is now a subsystem (why?), and every minimal system is the orbit closure of any of its elements (why?), but in the converse direction, not all orbit closures are minimal. Consider for instance the boolean Bernoulli system $(2^{\mathbf{Z}}, A \mapsto A-1)$ with $x = \mathbf{N} := \{0, 1, 2, \ldots\} \in 2^{\mathbf{Z}}$ being the natural numbers. Then the orbit $T^{\mathbf{Z}}x$ of x consists of all the half-lines $\{a, a+1, \ldots, \} \in 2^{\mathbf{Z}}$ for $a \in \mathbf{Z}$, but it is not closed; it has the point $\mathbf{Z} \in 2^{\mathbf{Z}}$ and the point $\emptyset \in 2^{\mathbf{Z}}$ as limit

[10] One could make the same definition for measure-preserving systems, but it tends to be a bit vacuous - given any measure preserving system that contains points of measure zero, one can make it trivially smaller by removing the orbit $T^{\mathbf{Z}}x := \{T^n x : n \in \mathbf{Z}\}$ of any point x of measure zero. One could place a topology on the space X and demand that it be compact, in which case minimality just means that the probability measure μ has full support.

[11] In the case of finite dynamical systems, the integer shift does not appear, and we have recovered the classical fact that every permutation is uniquely decomposable as the product of disjoint cycles.

points (recall that $2^{\mathbf{Z}}$ is given the product (i.e. pointwise) topology). Each of these points is an invariant point of T and thus forms its own orbit closure, which is obviously minimal[12].

Thus we see that finite dynamical systems do not quite form a perfect model for topological dynamical systems. A slightly better (but still imperfect) model would be that of *non-invertible* finite dynamical systems (X, T), in which $T : X \to X$ is now just a function rather than a permutation. Then we can still verify that all minimal orbits are given by disjoint cycles, but they no longer necessarily occupy all of X; it is quite possible for the orbit $T^{\mathbf{N}}x = \{T^n x : n \in \mathbf{N}\}$ of a point x to start outside of any of the minimal cycles, although it will eventually be absorbed in one of them.

In the above examples, the limit points of an orbit formed their own minimal orbits. In some cases, one has to pass to limits multiple times before one reaches a minimal orbit. For instance, consider the boolean Bernoulli system again, but now consider the point

$$y := \bigcup_{n=0}^{\infty} [4^n, 2 \times 4^n] = [1,2] \cup [4,8] \cup [16,32] \cup \ldots \in 2^{\mathbf{Z}}$$

where we use the notation $[N, M] := \{n \in \mathbf{Z} : N \leq n \leq M\}$. Observe that the point x defined earlier is not in the orbit $T^{\mathbf{N}}y$, but lies in the orbit closure, as it is the limit of $T^{4^n}y$. On the other hand, the orbit closure of x does not contain y. So the orbit closure of x is a subsystem of that of y, and then inside the former system one has the minimal systems $\{\mathbf{Z}\}$ and $\{\emptyset\}$. It is not hard to iterate this type of example and see that we can have quite intricate hierarchies of systems.

Exercise 2.2.1. Construct a topological dynamical system $(X, \mathcal{F}, T)$ and a sequence of orbit closures $\overline{T^{\mathbf{Z}}x_n}$ in X which form a proper nested sequence, thus

$$\overline{T^{\mathbf{Z}}x_1} \supsetneq \overline{T^{\mathbf{Z}}x_2} \supsetneq \overline{T^{\mathbf{Z}}x_3} \supsetneq \ldots$$

Hint: Take a countable family of nested Bernoulli systems, and find a way to represent each one as a orbit closure.

[12] This argument shows that x itself is not contained in any minimal system - why?

Despite this apparent complexity, we can always terminate such hierarchies of subsystems at a minimal system:

Lemma 2.2.8. *Every topological dynamical system $(X, \mathcal{F}, T)$ contains a minimal dynamical system.*

Proof. Observe that the intersection of any chain of subsystems of X is again a subsystem (here we use the *finite intersection property* of compact sets to guarantee that the intersection is non-empty, and we also use the fact that the arbitrary intersection of closed or T-invariant sets is again closed or T-invariant). The claim then follows from *Zorn's lemma*[13]. □

Exercise 2.2.2. Recall that every compact metrisable space is *second countable* and thus has a countable topological base. Suppose we are given an explicit enumeration $V_1, V_2, \ldots$ of such a base. Then find a proof of Lemma 2.2.8 which avoids the axiom of choice.

It would be nice if we could use Lemma 2.2.8 to decompose topological dynamical systems into the union of minimal subsystems, as we did in the case of non-topological dynamical systems. Unfortunately this does not work so well; the problem is that the complement of a minimal system is an open set rather than a closed set, and so we cannot cleanly separate a minimal system from its complement[14].

We will study minimal dynamical systems in detail in the next few lectures. I'll close now with some examples of minimal systems.

Example 2.2.9 (Cyclic group shift)**.** The cyclic group shift $(\mathbf{Z}/N\mathbf{Z}, x \mapsto x+1)$, where N is a positive integer, is a minimal system, and these are the only discrete minimal topological dynamical systems. More generally, if x is a periodic point of a topological dynamical system (thus $T^N x = x$ for some $N \geq 1$), then the closed orbit of x is isomorphic to a cyclic group shift and is thus minimal.

[13] We will always assume the axiom of choice throughout this course.

[14] In any case, the preceding examples already show that there can be some points in a system that are not contained in any minimal subsystem. Also, in contrast with non-invertible non-topological dynamical systems, our examples also show that a closed orbit can contain multiple minimal subsystems, so we cannot reduce to some sort of "nilpotent" system that has only one minimal system.

Example 2.2.10 (Torus shift)**.** Consider a torus shift $((\mathbf{R}/\mathbf{Z})^d, x \mapsto x + \alpha)$, where $\alpha \in \mathbf{R}^d$ is a fixed vector. It turns out that this system is minimal if and only if[15] α is *totally irrational*, which means that $n \cdot \alpha$ is not an integer for any non-zero $n \in \mathbf{Z}^d$.

Example 2.2.11 (Morse sequence)**.** Let $A = \{a, b\}$ be a two-letter alphabet, and consider the Bernoulli system $(A^{\mathbf{Z}}, T)$ formed from doubly infinite words

$$\ldots x_{-2}x_{-1}.x_0x_1x_2\ldots$$

in A with the left-shift. Now define the sequence of finite words

$$\begin{aligned} w_1 &:= a.b; \\ w_2 &:= abba.baab; \\ w_3 &:= abbabaabbaababba.baababbaabbabaab; \\ &\ldots \end{aligned}$$

by the recursive formula

$$w_1 := a.b; \quad w_{i+1} := f(w_i)$$

where $f(w)$ denotes the word formed from w by replacing each occurrence of a and b by abba and baab respectively. These words w_i converge pointwise to an infinite word

$$w = \ldots abbabaabbabbabaabbaababba.baababbaabbabaababbabaab\ldots.$$

Exercise 2.2.3. Show that w is not a periodic element of $A^{\mathbf{Z}}$, but that the orbit $\overline{T^{\mathbf{Z}}w}$ is both closed and minimal. *Hint*: find large subwords of w which appear *syndetically*, which means that the gaps between each appearance are bounded. In fact, all subwords of w appear syndetically. One can also work with a more explicit description of w involving the number of non-zero digits in the binary expansion of the index. (This set is an example of a *substitution minimal set*.)

Exercise 2.2.4. Let $(X, \mathcal{F}, T)$ and $(Y, \mathcal{G}, S)$ be topological dynamical systems. Define the *product* of these systems to be $(X \times Y, \mathcal{F} \times \mathcal{G}, T \times S)$, where $X \times Y$ is the Cartesian product, $\mathcal{F} \times \mathcal{G}$ is the product topology, and $T \times S$ is the map $(x, y) \mapsto (Tx, Sy)$. Note that there

[15] The "if" part is slightly non-trivial; see Corollary 1.4.2; but the "only if" part is easy, and is left as an exercise.

are obvious projection morphisms from this product system to the two original systems. Show that this product system is indeed a product in the sense of category theory, thus any other system that maps to the two original systems factors uniquely through the product. Establish analogous claims in the categories of dynamical systems and measure-preserving systems.

Exercise 2.2.5. Let $(X, \mathcal{F}, T)$ and $(Y, \mathcal{G}, S)$ be topological dynamical systems. Define the *disjoint union* of these systems to be $(X \uplus Y, \mathcal{F} \uplus \mathcal{G}, T \uplus S)$ where $(X \uplus Y, \mathcal{F} \uplus \mathcal{G})$ is the disjoint union of $(X, \mathcal{F})$ and $(Y, \mathcal{G})$, and $T \uplus S$ is the map which agrees with T on X and agrees with S on Y. Note that there are obvious embedding morphisms from the two original systems into the disjoint union. Show that the disjoint union is a coproduct in the sense of category theory, thus any system that is mapped to from the two origina systems factors uniquely through the disjoint union. Are analogous claims true for the categories of dynamical systems and measure-preserving systems?

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/01/10`. Thanks to Andy P. and anonymous commenters for corrections.

2.3. Minimal dynamical systems, recurrence, and the Stone-Cěch compactification

We now begin the study of *recurrence* in topological dynamical systems $(X, \mathcal{F}, T)$ - how often a non-empty open set U in X returns to intersect itself, or how often a point x in X returns to be close to itself. Not every set or point needs to return to itself; consider for instance what happens to the shift $x \mapsto x + 1$ on the compactified integers $\{-\infty\} \cup \mathbf{Z} \cup \{+\infty\}$. Nevertheless, we can always show that at least one set (from any open cover) returns to itself:

Theorem 2.3.1 (Simple recurrence in open covers)**.** *Let $(X, \mathcal{F}, T)$ be a topological dynamical system, and let $(U_\alpha)_{\alpha \in A}$ be an open cover of X. Then there exists an open set U_α in this cover such that $U_\alpha \cap T^n U_\alpha \neq \emptyset$ for infinitely many n.*

Proof. By compactness of X, we can refine the open cover to a finite subcover. Now consider an orbit $T^{\mathbf{Z}} x = \{T^n x : x \in \mathbf{Z}\}$ of some

arbitrarily chosen point $x \in X$. By the infinite pigeonhole principle, one of the sets U_α must contain an infinite number of the points $T^n x$ counting multiplicity; in other words, the recurrence set $S := \{n : T^n x \in U_\alpha\}$ is infinite. Letting n_0 be an arbitrary element of S, we thus conclude that $U_\alpha \cap T^{n_0-n} U_\alpha$ contains $T^{n_0} x$ for every $n \in S$, and the claim follows. □

Exercise 2.3.1. Conversely, use Theorem 2.3.1 to deduce the infinite pigeonhole principle (i.e. that whenever $\mathbf{Z}$ is coloured into finitely many colours, one of the colour classes is infinite). *Hint*: look at the orbit closure of c inside $A^{\mathbf{Z}}$, where A is the set of colours and $c : \mathbf{Z} \to A$ is the colouring function.)

Now we turn from recurrence of sets to recurrence of individual points, which is a somewhat more difficult, and highlights the role of minimal dynamical systems (as introduced in Section 2.2) in the theory. We will approach the subject from two (largely equivalent) approaches, the first one being the more traditional "epsilon and delta" approach, and the second using the *Stone-Čěch compactification* $\beta\mathbf{Z}$ of the integers (or equivalently, via *ultrafilters*).

Before we begin, it will be notationally convenient[16] to place a metric d on our compact metrisable space X. There are of course infinitely many metrics that one could place here, but they are all coarsely equivalent in the following sense: if d, d' are two metrics on X, then for every $\delta > 0$ there exists an $\varepsilon > 0$ such that $d'(x, y) < \delta$ whenever $d(x, y) < \varepsilon$, and similarly with the role d and d' reversed. This claim follows from the standard fact that continuous functions between compact metric spaces are uniformly continuous. Because of this equivalence, it will not actually matter for any of our results what metric we place on our spaces. For instance, we could endow a Bernoulli system $A^{\mathbf{Z}}$, where A is itself a compact metrisable space (and thus $A^{\mathbf{Z}}$ is compact by *Tychonoff's theorem*), with the metric

$$d((a_n)_{n\in\mathbf{Z}}, (b_n)_{n\in\mathbf{Z}}) := \sum_{n\in\mathbf{Z}} 2^{-|n|} d_A(a_n, b_n) \tag{2.4}$$

[16] As an exercise, the reader is encouraged to recast all the material here in a manner which does not explicitly mention a metric.

where d_A is some arbitrarily selected metric on A. Note that this metric is not shift-invariant.

Exercise 2.3.2. Show that if A contains at least two points, then the Bernoulli system $A^{\mathbf{Z}}$ (with the standard shift) cannot be endowed with a shift-invariant metric. *Hint*: find two distinct points which converge to each other under the shift map.

Fix a metric d. For each n, the shift $T^n : X \to X$ is continuous, and hence uniformly continuous since X is compact, thus for every $\delta > 0$ there exists $\varepsilon > 0$ depending on δ and n such that $d(T^n x, T^n y) < \delta$ whenever $d(x, y) < \varepsilon$. However, we caution that the T^n need not be uniformly *equicontinuous*; the quantity ε appearing above can certainly depend on n. Indeed, they need not even be equicontinuous. For instance, this will be the case for the Bernoulli shift with the metric 2.4 (why?), and more generally for any system that exhibits "mixing" or other chaotic behaviour. At the other extreme, in the case of *isometric* systems - systems in which T preserves the metric d - the shifts T^n are all isometries, and thus are clearly uniformly equicontinuous. (We will study isometric systems further in Section 2.6.)

We can now classify points x in X based on the dynamics of the orbit $T^{\mathbf{Z}}x := \{T^n x : n \in \mathbf{Z}\}$:

Definition 2.3.2 (Points in a topological dynamical system)**.** (1) x is *invariant* if $Tx = x$.

(2) x is *periodic* if $T^n x = x$ for some non-zero n.

(3) x is *almost periodic* if for every $\varepsilon > 0$, the set $\{n \in \mathbf{Z} : d(T^n x, x) < \varepsilon\}$ is *syndetic* (i.e. it has bounded gaps);

(4) x is *recurrent* if for every $\varepsilon > 0$, the set $\{n \in \mathbf{Z} : d(T^n x, x) < \varepsilon\}$ is infinite. Equivalently, there exists a sequence n_j of integers with $|n_j| \to \infty$ such that $\lim_{j\to\infty} T^{n_j} x = x$.

It is clear that every invariant point is periodic, that every periodic point is almost periodic, and every almost periodic point is recurrent. These inclusions are all strict. For instance, in the circle shift system $(\mathbf{R}/\mathbf{Z}, x \mapsto x + \alpha)$ with $\alpha \in \mathbf{R}$ irrational, it turns out that every point is almost periodic, but no point is periodic.

Exercise 2.3.3. In the boolean Bernoulli system $(2^{\mathbf{Z}}, A \mapsto A - 1)$, show that the discrete Cantor set

$$x := \bigcup_{N=1}^{\infty} \{\sum_{n=0}^{N} \epsilon_n 10^n : \epsilon_n \in \{-1, 0, +1\}\} \tag{2.5}$$

is recurrent but not almost periodic.

In a general topological dynamical system, it is quite possible to have points which are non-recurrent (as the example of the compactified integer shift already shows). But if we restrict to a *minimal* dynamical system, things get much better:

Lemma 2.3.3. *If $(X, \mathcal{F}, T)$ is a minimal topological dynamical system, then every element of X is almost periodic (and hence recurrent).*

Proof. Suppose for contradiction that we can find a point x of X which is not almost periodic. This means that we can find $\varepsilon > 0$ such that the set $\{n : d(T^n x, x) < \varepsilon\}$ is not syndetic. Thus, for any $m > 0$, we can find an n_m such that $d(T^n x, x) \geq \varepsilon$ for all $n \in [n_m - m, n_m + m]$ (say).

Since X is compact, the sequence $T^{n_m} x$ must have at least one limit point y. But then one verifies (using the continuity of the shift operators) that

$$d(T^h y, x) = \lim_{m \to \infty} d(T^{n_m + h} x, x) \geq \varepsilon \tag{2.6}$$

for all h. But this means that the orbit closure $\overline{T^{\mathbf{Z}} y}$ of y does not contain x, contradicting the minimality of X. The claim follows. □

Exercise 2.3.4. If x is a point in a topological dynamical system, show that x is almost periodic if and only if it lies in a minimal system. Because of this, almost periodic points are sometimes referred to as *minimal* points.

Combining Lemma 2.3.3 with Lemma 2.2.8, we immediately obtain the

Theorem 2.3.4 (Birkhoff recurrence theorem)**.** *Every topological dynamical system contains at least one point x which is almost periodic (and hence recurrent).*

Note that this is stronger than Theorem 2.3.1, as can be seen by considering the element U_α of the open cover which contains the almost periodic point. Indeed, we now have obtained a stronger conclusion, namely that the set of return times $\{n : T^n U_\alpha \cap U_\alpha \neq \emptyset\}$ is not only infinite, it is syndetic.

Exercise 2.3.5. State and prove a version of the Birkhoff recurrence theorem in which the map $T : X \to X$ is continuous but not assumed to be invertible. (Of course, all references to $\mathbf{Z}$ now need to be replaced with $\mathbf{N}$.)

The Birkhoff recurrence theorem does not seem particularly strong, as it only guarantees existence of a single recurrent (or almost periodic point). For general systems, this is inevitable, because it can happen that the majority of the points are non-recurrent (look at the compactified integer shift system, for instance). However, suppose the system is a group quotient $(G/\Gamma, x \mapsto gx)$. To make this a topological dynamical system, we need G to be a topological group, and Γ to be a cocompact subgroup of G (such groups are also sometimes referred to as *uniform* subgroups). Then we see that the system is a *homogeneous space*: given any two points $x, y \in G/\Gamma$, there exists a group element $h \in G$ such that $hx = y$. Thus we expect any two points in G/Γ to behave similarly to each other. Unfortunately, this does not quite work in general, because the action of h need not preserve the shift $x \mapsto gx$, as there is no reason that h commutes with g. But suppose that g is a *central* element of G, i.e. it commutes with every element of G; this is for instance the case if G is abelian. Then the action of h is now an isomorphism on the dynamical system $(G/\Gamma, x \mapsto gx)$. In particular, if $hx = y$, we see that x is almost periodic (or recurrent) if and only if y is. We thus conclude:

Theorem 2.3.5 (Kronecker type approximation theorem)**.** *Let $(G/\Gamma, x \mapsto gx)$ be a topological group quotient dynamical system such that g lies in the centre $Z(G)$ of G. Then* every *point in this system is almost periodic (and hence recurrent).*

Applying this theorem to the torus shift $((\mathbf{R}/\mathbf{Z})^d, x \mapsto x + \alpha)$, where $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbf{R}^d$ is a vector, we thus obtain that for any

$\varepsilon > 0$, the set

$$\{n \in \mathbf{Z} : \mathrm{dist}(n\alpha, \mathbf{Z}^d) < \varepsilon\} \tag{2.7}$$

is syndetic (and in particular, infinite). This should be compared with the classical Kronecker approximation theorem.

It is natural to ask what happens when g is not central. If G is a Lie group and the action of g on the Lie algebra $\mathfrak{g}$ is unipotent rather than trivial, then Theorem 2.3.5 still holds; this follows from *Ratner's theorem*, which we will discuss in Sections 2.16-2.17. But the claim is not true for all group quotients. Consider for instance the Bernoulli shift system $(X, T) = ((\mathbf{Z}/2\mathbf{Z})^{\mathbf{Z}}, T)$, which is isomorphic to the boolean Bernoulli shift system. As the previous examples have already shown, this system contains both recurrent and non-recurrent elements. On the other hand, it is intuitive that this system has a lot of symmetry, and indeed we can view it as a group quotient $(G/\Gamma, x \mapsto gx)$. Specifically, G is the *lamplighter group* $G = \mathbf{Z}/2\mathbf{Z} \wr \mathbf{Z}$. To describe this group, we observe that the group $(\mathbf{Z}/2\mathbf{Z})^{\mathbf{Z}}$ acts on X by addition, whilst the group $\mathbf{Z}$ acts on X via the shift map T. The lamplighter group $G := (\mathbf{Z}/2\mathbf{Z})^{\mathbf{Z}} \times \mathbf{Z}$ then acts by both addition and shift:

$$(a, n) : x \mapsto T^n x + a \text{ for all } (a, n) \in G. \tag{2.8}$$

In order for this to be a group action, we endow G with the multiplication law

$$(a, n)(b, m) := (a + T^n b, n + m); \tag{2.9}$$

one easily verifies that this really does make G into a group, and if we give G the product topology, it becomes a *topological group*. G clearly acts transitively on the compact space X, and so $X \equiv G/\Gamma$ for some cocompact subgroup Γ (which turns out to be isomorphic to $\mathbf{Z}$ - why?). By construction, the shift map T can be expressed using the group element $(0, 1) \in G$, and so we have turned the Bernoulli system into a group quotient. Since this system contains non-recurrent points (e.g. the indicator function of the natural numbers) we see that Theorem 2.3.5 does not hold for arbitrary group quotients.

2.3.1. The ultrafilter approach. Now we turn to a different approach to topological recurrence, which relies on compactifying the underlying group $\mathbf{Z}$ that acts on topological dynamical systems. By doing so, all the epsilon management issues (cf. Section 1.5 of *Structure and Randomness*) go away, and the subject becomes very algebraic in nature. On the other hand, some subtleties arise also; for instance, the compactified object $\beta\mathbf{Z}$ is not a group, but merely a left-continuous semigroup.

This approach is based on *ultrafilters* or (equivalently) via the *Stone-Čech compactification*. Let us recall how this compactification works:

Theorem 2.3.6 (Stone-Čech compactification)**.** *Every locally compact Hausdorff (LCH) space X can be embedded in a compact Hausdorff space βX in which X is an open dense set. (In particular, if X is already compact, then $\beta X = X$.) Furthermore, any continuous function $f : X \to Y$ between LCH spaces extends uniquely to a continuous function $\beta f : \beta X \to \beta Y$.*

Proof. (Sketch) This proof uses the intuition that βX should be the "finest" compactification of X. Recall that a compactification of a LCH space X is any compact Hausdorff space containing X as an open dense set. We say that one compactification Y of X is *finer* than another Z if there is a surjective[17] continuous map from Y to Z that is the identity on X. For instance, the two-point compactification $\{-\infty\} \cup \mathbf{Z} \cup \{+\infty\}$ of the integers is finer than the one-point compactification $\mathbf{Z} \cup \{\infty\}$. This is clearly a partial ordering; also, the *inverse limit* of any chain (totally ordered set) of compactifications can be verified (by *Tychonoff's theorem*) to still be a compactification. Hence, by *Zorn's lemma*[18], there is a maximal compactification βX. To verify the extension property for continuous functions $f : X \to Y$, note (by replacing Y with βY if necessary) that we may take Y to be compact. Let Z be the closure of the graph $X' := \{(x, f(x)) : x \in X\}$

[17] Note that as X is dense in Y, and Z is Hausdorff, this surjection is unique.

[18] There is a technical step one needs to verify to apply this lemma, namely the moduli space of compactifications of X is a set rather than a class. We leave this to the reader.

in $(\beta X) \times Y$. X' is clearly homeomorphic to X, and so Z is a compactification of X. Also, there is an obvious surjective continuous map from Z to βX; thus by maximality, this map must be a homeomorphism, thus Z is the graph of a continuous function $\beta f : \beta X \to \beta Y$, and the claim follows (the uniqueness of βf is easily established). □

Exercise 2.3.6. Let X be discrete (and thus clearly LCH), and let βX be the Stone-Cěch compactification. For any $p \in \beta X$, let $[p] \in 2^{2^X}$ be the collection of all sets $A \subset X$ such that $\beta 1_A(p) = 1$. Show that $[p]$ is an *ultrafilter*, or in other words that it obeys the following four properties:

(1) $\emptyset \notin [p]$.

(2) If $U \in [p]$ and $V \in 2^X$ are such that $U \subset V$, then $V \in [p]$.

(3) If $U, V \in [p]$, then $U \cap V \in [p]$.

(4) If $U, V \in 2^X$ are such that $U \cup V = X$, then at least one of U and V lie in [p].

Furthermore, show that the map $p \mapsto [p]$ is a homeomorphism between βX and the space of ultrafilters, which we endow with the topology induced from the product topology on $2^{2^X} \equiv \{0,1\}^{2^X}$, where we give $\{0,1\}$ the discrete topology (one can place some other topologies here also). Thus we see that in the discrete case, we can represent the Stone-Cěch compactification explicitly via ultrafilters.

It is easy to see that $\beta(g \circ f) = (\beta g) \circ (\beta f)$ whenever $f : X \to Y$ and $g : Y \to Z$ are continuous maps between LCH spaces. In the language of category theory, we thus see that β is a *covariant functor* from the category of LCH spaces to the category of compact Hausdorff spaces[19]

Exercise 2.3.7. Let X and Y be two LCH spaces. Show that the *disjoint union* $(\beta X) \uplus (\beta Y)$ of βX and βY is isomorphic to $\beta(X \uplus Y)$. (Indeed, this isomorphism is a *natural isomorphism*.) In the language of category theory, this means that β preserves coproducts[20].

[19]The above theorem does not explicitly define βX, but it is not hard to see that this compactification is unique up to homeomorphism, so the exact form of βX is somewhat moot. However, it is possible to create an ultrafilter-based description of βX for general LCH spaces X, though we will not do so here.

[20]Unfortunately, β does not preserve products, which leads to various subtleties, such as the non-commutativity of the compactification of commutative groups.

Note that if $f : X \to Y$ is continuous, then $\beta f : \beta X \to \beta Y$ is continuous also; since X is dense in βX, we conclude that[21]

$$\beta f(p) = \lim_{x \to p} f(x) \tag{2.10}$$

for all $p \in \beta X$, where x is constrained to lie in X. In particular, the limit on the right exists for any continuous $f : X \to Y$, and thus if X is discrete, it exists for any (!) function $f : X \to Y$. Each p can then be viewed as a recipe for taking limits of arbitrary functions in a consistent fashion (although different p's can give different limits, of course). It is this ability to take limits without needing to check for convergence and without running into contradictions that makes the Stone-Cěch compactification a useful tool here[22].

The integers $\mathbf{Z}$ are discrete, and thus are clearly LCH. Thus we may form the compactification $\beta\mathbf{Z}$. The addition operation $+ : \mathbf{Z} \times \mathbf{Z} \to \mathbf{Z}$ can then be extended to $\beta\mathbf{Z}$ by the plausible-looking formula

$$p + q := \lim_{n \to p} \lim_{m \to q} n + m \tag{2.11}$$

for all $p, q \in \beta\mathbf{Z}$, where n, m range in the integers $\mathbf{Z}$. Note that the double limit is guaranteed to exist by (2.10). Equivalently, we have

$$\lim_{l \to p+q} f(l) = \lim_{n \to p} \lim_{m \to q} f(n+m) \tag{2.12}$$

for all functions $f : \mathbf{Z} \to X$ into an LCH space X; one can derive (2.12) from (2.11) by applying $\beta f : \beta\mathbf{Z} \to \beta X$ to both sides of (2.11) and using (2.10) and the continuity of βf repeatedly. This addition operation clearly extends that of $\mathbf{Z}$ and is associative, thus we have turned $\beta\mathbf{Z}$ into a semigroup. We caution however that this semigroup is not commutative, due to the usual difficulty that double limits in (2.11) cannot be exchanged. (We will prove non-commutativity shortly.) For similar reasons, $\beta\mathbf{Z}$ is not a group; the obvious attempt to define a negation operation $-p := \lim_{n \to p} -n$ is well-defined, but does not actually invert addition. The operation $(p, q) \mapsto p + q$ is continuous in p for fixed q (why?), but is not necessarily continuous in q for fixed p - again, due to the exchange of limits problem. Thus

[21] Here and in the sequel, limits such as $\lim_{x \to p}$ are interpreted in the usual topological sense, thus (2.10) means that for every neighbourhood V of $\beta f(p)$, there exists a neighbourhood U of p such that $f(x) \in V$ for all $x \in U$.

[22] See also Section 1.5 of *Structure and Randomness* for further discussion.

$\beta\mathbf{Z}$ is merely a left-continuous semigroup. If however p is an integer, then the first limit in (2.11) disappears, and one easily shows that $q \mapsto q+p$ is continuous in this case (and for similar reasons one also recovers commutativity, $q+p = p+q$).

Exercise 2.3.8. Let us endow the two-point compactification $\{-\infty\} \cup \mathbf{Z} \cup \{+\infty\}$ with the semigroup structure $+$ in which $x + (+\infty) = +\infty$ and $x + (-\infty) = -\infty$ for all $x \in \{-\infty\} \cup \mathbf{Z} \cup \{+\infty\}$ (compare with (2.11)). Show that there is a unique continuous map $\pi : \beta\mathbf{Z} \to \mathbf{Z} \cup \{-\infty\} \cup \{+\infty\}$ which is the identity on $\mathbf{Z}$, and that this map is a surjective semigroup homomorphism. Using this homomorphism, conclude:

(1) $\beta\mathbf{Z}$ is not commutative. Furthermore, show that the centre $Z(\beta\mathbf{Z}) := \{p \in \beta\mathbf{Z} : p+q = q+p \text{ for all } q \in \beta\mathbf{Z}\}$ is exactly equal to $\mathbf{Z}$.

(2) Show that if $p, q \in \beta\mathbf{Z}$ are such that $p+q \in \mathbf{Z}$, then $p, q \in \mathbf{Z}$. ("Once you go to infinity, you can never return.") Conclude in particular that $\beta\mathbf{Z}$ is not a group[23].

Remark 2.3.7. More generally, we can take any LCH left-continuous semigroup S and compactify it to obtain a compact Hausdorff left-continuous semigroup βS. Observe that if $f : S \to S'$ is a homomorphism between two LCH left-continuous semigroups, then $\beta f : \beta S \to \beta S'$ is also a homomorphism. Thus, from the viewpoint of category theory, β can be viewed as a covariant functor from the category of LCH left-continuous semigroups to the category of CH left-continuous semigroups.

The left-continuous non-commutative semigroup structure of $\beta\mathbf{Z}$ may appear to be terribly weak when compared against the jointly continuous commutative group structure of $\mathbf{Z}$, but $\beta\mathbf{Z}$ has a decisive trump card over $\mathbf{Z}$: it is *compact*. We will see the power of compactness a little later in this lecture.

A topological dynamical system $(X, \mathcal{F}, T)$ yields an action $n \mapsto T^n$ of the integers $\mathbf{Z}$. But we can automatically extend this action to

[23] Note that this conclusion could already be obtained using the coarser one-point compactification $\mathbf{Z} \cup \{\infty\}$ of the integers.

an action $p \mapsto T^p$ of the compactified integers $\beta\mathbf{Z}$ by the formula

$$T^p x := \lim_{n \to p} T^n x. \tag{2.13}$$

(Note that X is already compact, so that the limit in (2.13) stays in X.) One easily checks from (2.12) that this is indeed an action of $\beta\mathbf{Z}$ (thus $T^p T^q = T^{p+q}$ for all $p, q \in \mathbf{Z}$). The map $T^p x$ is continuous in p by construction; however we caution that it is no longer continuous in x (it's the exchange-of-limits problem once more!). Indeed, the map $T^p : X \to X$ can be quite nasty from an analytic viewpoint; for instance, it is possible for this map to not be Borel measurable[24]. But as we shall see, the *algebraic* properties of T^p are very good, and suffice for applications to recurrence, because once one has compactified the underlying semigroup $\beta\mathbf{Z}$, the need for point-set topology (and for all the epsilons that come with it) mostly disappears. For instance, we can now replace orbit closures by orbits:

Lemma 2.3.8. *Let $(X, \mathcal{F}, T)$ be a topological dynamical system, and let $x \in X$. Then*

$$\overline{T^{\mathbf{Z}}(x)} = T^{\beta\mathbf{Z}} x := \{T^p x : p \in \beta\mathbf{Z}\}.$$

Proof. Since $\beta\mathbf{Z}$ is compact, $T^{\beta\mathbf{Z}} x$ is compact also. Since $\mathbf{Z}$ is dense in $\beta\mathbf{Z}$, $T^{\mathbf{Z}} x$ is dense in $T^{\beta\mathbf{Z}} x$. The claim follows. □

From (2.13) we see that T^p is some sort of "limiting shift" operation. To get some intuition, let us consider the compactified integer shift $(\{-\infty\} \cup \mathbf{Z} \cup \{+\infty\}, x \mapsto x+1)$, and look at the orbit of the point 0. If one only shifts by integers $n \in \mathbf{Z}$, then $T^n 0$ can range across the region $\mathbf{Z}$ in the system but cannot reach $-\infty$ or $+\infty$. But now let $p \in \beta\mathbf{Z} \backslash \mathbf{Z}$ be any limit point of the positive integers $\mathbf{Z}^+$ (note that at least one such limit point must exist, since $\mathbf{Z}^+$ is not compact. Indeed, in the language of Exercise 2.3.8, the set of all such limit points is $\pi^{-1}(+\infty)$.) Then from (2.13) we see that $T^p 0 = +\infty$. Similarly, if $q \in \beta\mathbf{Z} \backslash \mathbf{Z}$ is a limit point of the negative integers $\mathbf{Z}^-$ then $T^q 0 = -\infty$. Now, since $+\infty$ invariant, we have $T^q(+\infty) = +\infty$ by (2.13) again, and thus $T^q T^p 0 = +\infty$, while $T^p T^q 0 = -\infty$. In particular, we see that $p + q \neq q + p$, demonstrating non-commutativity in $\beta\mathbf{Z}$ (again,

[24] This is the price one pays for introducing beasts generated by the axiom of choice into one's mathematical ecosystem.

compare with Exercise 2.3.8). Informally, the problem here is that in (2.11), n+m will go to $+\infty$ if we let m go to $+\infty$ first and then $n \to -\infty$ next, but if we take $n \to -\infty$ first and then $m \to +\infty$ next, $n+m$ instead goes to $-\infty$.

Exercise 2.3.9. Let $A \subset \mathbf{Z}$ be a set of integers.

(1) Show that βA can be canonically identified with the closure of A in $\beta\mathbf{Z}$, in which case βA becomes a *clopen subset* of $\beta\mathbf{Z}$.

(2) Show that A is infinite if and only if $\beta A \not\subset \mathbf{Z}$.

(3) Show that A is syndetic if and only if $\beta A \cap (\beta\mathbf{Z} + p) \neq \emptyset$ for every $p \in \beta\mathbf{Z}$. (Since βA is clopen, this condition is also equivalent to requiring $\beta A \cap (\mathbf{Z} + p) \neq \emptyset$ for every $p \in \beta\mathbf{Z}$.)

(4) A set of integers A is said to be *thick* if it contains arbitrarily long intervals $[a_n, a_n+n]$; thus syndetic and thick sets always intersect each other. Show that A is thick if and only if there exists $p \in \beta\mathbf{Z}$ such that $\beta\mathbf{Z}+p \subset \beta A$. (Again, this condition is equivalent to requiring $\mathbf{Z} + p \subset \beta A$ for some p.)

Recall that a system is *minimal* if and only if it is the orbit closure of every point in that system. We thus have a purely algebraic description of minimality:

Corollary 2.3.9. *Let $(X, \mathcal{F}, T)$ be a topological dynamical system. Then X is minimal if and only if the action of $\beta\mathbf{Z}$ is transitive; thus for every $x, y \in \mathbf{Z}$ there exists $p \in \beta\mathbf{Z}$ such that $T^p x = y$.*

One also has purely algebraic descriptions of almost periodicity and recurrence:

Exercise 2.3.10. Let $(X, \mathcal{F}, T)$ be a topological dynamical system, and let x be a point in X.

(1) Show that x is almost periodic if and and only if for every $p \in \beta\mathbf{Z}$ there exists $q \in \beta\mathbf{Z}$ such that $T^q T^p x = x$. (In particular, Lemma 2.3.3 is now an immediate consequence of Corollary 2.3.9.)

(2) Show that x is recurrent if and only if there exists $p \in \beta\mathbf{Z} \backslash \mathbf{Z}$ such that $T^p x = x$.

Note that $\beta\mathbf{Z}$ acts on itself $\beta\mathbf{Z}$ by addition, $p : q \mapsto p + q$, with the action being continuous when p is an integer. Thus one can view $\beta\mathbf{Z}$ itself as a topological dynamical system, except with the caveat that $\beta\mathbf{Z}$ is not metrisable or even first countable (see Exercise 2.3.13). Nevertheless, it is still useful to think of $\beta\mathbf{Z}$ as behaving like a topological dynamical system. For instance:

Definition 2.3.10. An element $p \in \beta\mathbf{Z}$ is said to be *minimal* or *almost periodic* if for every $q \in \beta\mathbf{Z}$ there exists $r \in \beta\mathbf{Z}$ such that $r + q + p = p$.

Equivalently, p is minimal if $\beta\mathbf{Z} + p$ is a minimal left-ideal of $\beta\mathbf{Z}$, which explains the terminology.

Exercise 2.3.11. Show that for every $p \in \beta\mathbf{Z}$ there exists $q \in \beta\mathbf{Z}$ such that $q + p$ is minimal. *Hint*: adapt the proof of Lemma 2.2.8. Also, show that if p is minimal, then $q + p$ and $p + q$ are also minimal for any $q \in \beta\mathbf{Z}$. This shows that minimal elements of $\beta\mathbf{Z}$ exist in abundance. However, observe from Exercise 2.3.6 that no integer can be minimal.

Exercise 2.3.12. Show that if $p \in \beta\mathbf{Z}$ is minimal, and x is a point in a topological dynamical system $(X, \mathcal{F}, T)$, then $T^p x$ is almost periodic. Conversely, show that x is almost periodic if and only if $x = T^p x$ for some minimal p. This gives an alternate (and more "algebraic") proof of the Birkhoff recurrence theorem.

Exercise 2.3.13. Show that no element of $\beta\mathbf{Z} \backslash \mathbf{Z}$ can be written as a limit of a sequence in $\mathbf{Z}$. *Hint*: if a sequence $n_j \in \mathbf{Z}$ converged to a limit $p \in \beta\mathbf{Z}$, one must have $\beta f(p) = \lim_{j \to \infty} f(n_j)$ for all functions $f : \mathbf{Z} \to K$ mapping into a compact Hausdorff space K. Conclude in particular that $\beta\mathbf{Z}$ is not metrisable, first countable, or sequentially compact.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/01/13`. Thanks to Richard, R.A., Eric, Liu Xiao Chuan, S.P., and Sean for corrections.

2.4. Multiple recurrence

In Section 2.3, we established single recurrence properties for both open sets and for sequences inside a topological dynamical system $(X, \mathcal{F}, T)$. In this lecture, we generalise these results to multiple recurrence. More precisely, we shall show

Theorem 2.4.1 (Multiple recurrence in open covers)**.** *Let $(X, \mathcal{F}, T)$ be a topological dynamical system, and let $(U_\alpha)_{\alpha \in A}$ be an open cover of X. Then there exists U_α such that for every $k \geq 1$, we have $U_\alpha \cap T^{-r} U_\alpha \cap \ldots \cap T^{-(k-1)r} U_\alpha \neq \emptyset$ for infinitely many r.*

Note that this theorem includes Theorem 2.3.1 as the special case $k = 2$. This theorem is also equivalent to the following well-known combinatorial result:

Theorem 2.4.2 (van der Waerden's theorem)**.** [**vdW1927**] *Suppose the integers $\mathbf{Z}$ are finitely coloured. Then one of the colour classes contains arbitrarily long arithmetic progressions.*

Exercise 2.4.1. Show that Theorem 2.4.1 and Theorem 2.4.2 are equivalent.

Exercise 2.4.2. Show that Theorem 2.4.2 fails if "arbitrarily long" is replaced by "infinitely long". Deduce that a similar strengthening of Theorem 2.4.1 also fails.

Exercise 2.4.3. Use Theorem 2.4.2 to deduce a finitary version: given any positive integers m and k, there exists an integer N such that whenever $\{1, \ldots, N\}$ is coloured into m colour classes, one of the colour classes contains an arithmetic progression of length k. *Hint*: use a "compactness and contradiction" argument, as in Section 1.3 of *Structure and Randomness*.

We also have a stronger version of Theorem 2.4.1:

Theorem 2.4.3 (Multiple Birkhoff recurrence theorem)**.** *Let $(X, \mathcal{F}, T)$ be a topological dynamical system. Then for any $k \geq 1$ there exists a point $x \in X$ and a sequence $r_j \to \infty$ of integers such that $T^{ir_j} x \to x$ as $j \to \infty$ for all $0 \leq i \leq k-1$.*

These results already have some application to equidistribution of explicit sequences. Here is a simple example (which is also a consequence of *Weyl's polynomial equidistribution theorem*, Theorem 2.6.26):

Corollary 2.4.4. *Let α be a real number. Then there exists a sequence $r_j \to \infty$ of integers such that $\mathrm{dist}(r_j^2\alpha, \mathbf{Z}) \to 0$ as $j \to \infty$.*

Proof. Consider the skew shift system $X = (\mathbf{R}/\mathbf{Z})^2$ with $T(x,y) := (x+\alpha, y+x)$. By Theorem 2.4.3, there exists $(x,y) \in X$ and a sequence $n_j \to \infty$ such that $T^{r_j}(x,y)$ and $T^{2r_j}(x,y)$ both convege to (x,y). If we then use the easily verified identity

$$(x,y) - 2T^{r_j}(x,y) + T^{2r_j}(x,y) = (0, r_j^2\alpha) \tag{2.14}$$

we obtain the claim. □

Exercise 2.4.4. Use Theorem 2.4.1 or Theorem 2.4.2 in place of Theorem 2.4.3 to give an alternate derivation of Corollary 2.4.4.

Exercise 2.4.5. Prove Theorem 1.4.1.

As in Section 2.3, we will give both a traditional topological proof and an ultrafilter-based proof of Theorem 2.4.1 and Theorem 2.4.3; the reader is invited to see how the various proofs are ultimately equivalent to each other.

2.4.1. Topological proof of van der Waerden. We begin by giving a topological proof of Theorem 2.4.1, due to Furstenberg and Weiss[**FuWe1978**], which is secretly a translation of van der Waerden's original "colour focusing" combinatorial proof of Theorem 2.4.2 into the dynamical setting. To prove Theorem 2.4.1, it suffices to show the following slightly weaker statement:

Theorem 2.4.5. *Let $(X, \mathcal{F}, T)$ be a topological dynamical system, and let $(U_\alpha)_{\alpha \in A}$ be an open cover of X. Then for every $k \geq 1$ there exists an open set U_α which contains an arithmetic progression $x, T^r x, T^{2r} x, \ldots, T^{(k-1)r} x$ for some $x \in X$ and $r > 0$.*

To see how Theorem 2.4.5 implies Theorem 2.4.1, first observe from compactness that we can take the open cover to be a finite cover. Then by the infinite pigeonhole principle, it suffices to establish

Theorem 2.4.1 for each $k \geq 1$ separately. For each such k, Theorem 2.4.5 gives a single arithmetic progression $x, T^r x, \ldots, T^{(k-1)r} x$ inside one of the U_α. By replacing the system (X, T) with the product system $(X \times \mathbf{Z}/N\mathbf{Z}, (x, m) \mapsto (Tx, m+1))$ for some large N and replacing the open cover $(U_\alpha)_{\alpha \in A}$ of X with the open cover $(U_\alpha \times \{m\})_{\alpha \in A, m \in \mathbf{Z}/N\mathbf{Z}}$ of $X \times \mathbf{Z}/N\mathbf{Z}$, one can make the spacing r in the arithmetic progression larger than any specified integer N. Thus by another application of the infinite pigeonhole principle, one of the U_α contains arithmetic progressions with arbitrarily large step r, and the claim follows.

Now we need to prove Theorem 2.4.5. By Lemma 2.2.8 to establish this theorem for minimal dynamical systems. We will need to note that for minimal systems, Theorem 2.4.5 automatically implies the following stronger-looking statement:

Theorem 2.4.6. *Let $(X, \mathcal{F}, T)$ be a minimal topological dynamical system, let U be a non-empty open set in X, and let $k \geq 1$. Then U contains an arithmetic progression $x, T^r x, \ldots T^{(k-1)r} x$ for some $x \in X$ and $r \geq 1$.*

Indeed, the deduction of Theorem 2.4.6 from Theorem 2.4.5 is immediate from the following useful fact (cf. Lemma 2.3.3):

Lemma 2.4.7. *Let $(X, \mathcal{F}, T)$ be a minimal topological dynamical system, and let U be a non-empty open set in X. Then X can be covered by a finite number of translates $T^n U$ of U.*

Proof. The set $X \backslash \bigcup_{n \in \mathbf{Z}} T^n U$ is a proper closed invariant subset of X, which must therefore be empty since X is minimal. The claim then follows from the compactness of X. □

Remark 2.4.8. Of course, the claim is highly false for non-minimal systems; consider for instance the case when T is the identity. More generally, if X is non-minimal, consider an open set U which is the complement of a proper subsystem of X.

Now we need to prove Theorem 2.4.5. We do this by induction on k. The case $k = 1$ is trivial, so suppose $k \geq 2$ and the claim has already been shown for $k - 1$. By the above discussion, we see that Theorem 2.4.6 is also true for $k - 1$.

Now fix a minimal system $(X, \mathcal{F}, T)$ and an open cover $(U_\alpha)_{\alpha \in A}$, which we can take to be finite. We need to show that one of the U_α contains an arithmetic progression $x, T^r x, \ldots, T^{(k-1)r} x$ of length k. To do this, we first need an auxiliary construction.

Lemma 2.4.9 (Construction of colour focusing sequence)**.** *Let the notation and assumptions be as above. Then for any $J \geq 0$ there exists a sequence $x_0, \ldots, x_J$ of points in X, a sequence $U_{\alpha_0}, \ldots, U_{\alpha_J}$ of sets in the open cover (not necessarily distinct), and a sequence $r_1, \ldots, r_J$ of positive integers such that $T^{i(r_{a+1}+\ldots+r_b)} x_b \in U_{\alpha_a}$ for all $0 \leq a \leq b \leq J$ and $1 \leq i \leq k-1$.*

Proof. We induct on J. The case $J = 0$ is trivial. Now suppose inductively that $J \geq 1$, and that we have already constructed $x_0, \ldots, x_{J-1}, U_{\alpha_0}, \ldots, U_{\alpha_{J-1}}$, and $r_1, \ldots, r_{J-1}$ with the required properties. Now let V be a suitably small neighbourhood of x_{J-1} (depending on all the above data) to be chosen later. By Theorem 2.4.6 for $k-1$, V contains an arithmetic progression $y, T^{r_J} y, \ldots, T^{(k-2)r_J} y$ of length $k-1$. If one sets $x_J := T^{-r_J} y$, and lets U_{α_J} be an arbitrary set in the open cover containing x_J, then we observe that

(2.15)
$$T^{i(r_{a+1}+\ldots+r_J)} x_J = T^{i(r_{a+1}+\ldots+r_{J-1})}(T^{(i-1)r_J} y) \in T^{i(r_{a+1}+\ldots+r_{J-1})}(V)$$

for all $0 \leq a < J$ and $1 \leq i \leq k-1$. If V is a sufficiently small neighbourhood of x_{J-1}, we thus see (from the continuity of the $T^{i(r_{a+1}+\ldots+r_{J-1})}$ that we verify all the required properties needed to close the induction. □

We apply the above lemma with J equal to the number of sets in the open cover. By the pigeonhole principle, we can thus find $0 \leq a < b \leq J$ such that $U_{\alpha_a} = U_{\alpha_b}$. If we then set $x := x_b$ and $r := r_{a+1} + \ldots + r_b$ we obtain Theorem 2.4.5 as required.

Remark 2.4.10. It is instructive to compare the $k = 2$ case of the above arguments with the proof of Theorem 2.3.1. (For a comparison of this type of proof with the more classical combinatorial proof, see [**Ta2007**].)

2.4.2. Ultrafilter proof of van der Waerden. We now give a translation of the above proof into the language of *ultrafilters* (or

more precisely, the language of *Stone-Cěch compactifications*). This language may look a little strange, but it will be convenient when we study more general colouring theorems in the next lecture. As before, we will prove Theorem 2.4.5 instead of Theorem 2.4.1 (thus we only need to find one progression, rather than infinitely many). The key proposition is

Proposition 2.4.11 (Ultrafilter version of van der Waerden)**.** *Let p be a minimal element of $\beta\mathbf{Z}$. Then for any $k \geq 1$ there exists $q \in \beta(\mathbf{Z} \times \mathbf{N})$ such that*

$$\lim_{(n,r)\to q} n + ir + p = p \text{ for all } 0 \leq i \leq k-1. \tag{2.16}$$

Suppose for the moment that this proposition is true. Applying it with some minimal element p of $\beta\mathbf{Z}$ (which must exist, thanks to Exercise 2.3.11), we obtain $q \in \beta(\mathbf{Z} \times \mathbf{N})$ obeying (2.16). If we let $x := T^p y$ for some arbitrary $y \in X$, we thus obtain

$$\lim_{(n,r)\to q} T^{n+ir} x = x \text{ for all } 0 \leq i \leq k-1. \tag{2.17}$$

If we let U_α be an element of the open cover that contains x, we thus see that $T^{n+ir}x \in U_\alpha$ for all $0 \leq i \leq k-1$ and all $(n,r) \in \mathbf{Z} \times \mathbf{N}$ which lie in a sufficiently small neighbourhood of q. Since a LCH space is always dense in its Stone-Cěch compactification, the space of all (n,r) with this property is non-empty, and Theorem 2.4.5 follows.

Proof of Proposition 2.4.11. We induct on k. The case $k = 1$ is trivial (one could take e.g. $q = (0,1)$, so suppose $k > 1$ and that the claim has already been proven for $k-1$. Then we can find $q' \in \beta(\mathbf{Z} \times \mathbf{N})$ such that

$$\lim_{(n,r)\to q'} n + ir + p = p \tag{2.18}$$

for all $0 \leq i \leq k-2$.

Now consider the expression

$$p_{i,a,b} := \lim_{(n_1,r_1)\to q'} \ldots \lim_{(n_b,r_b)\to q'} i(r_{a+1} + \ldots + r_b) + m_b + p \tag{2.19}$$

for any $1 \leq a \leq b$ and $1 \leq i \leq k-1$, where

$$m_b := \sum_{i=1}^{b} n_i - r_i. \tag{2.20}$$

Applying (2.18) to the (n_b, r_b) limit in (2.19), we obtain the recursion $p_{i,a,b} = p_{i,a,b-1}$ for all $b > a$. Iterating this, we conclude that

$$p_{i,a,b} = p_{i,a,a} = p_{0,a,a} \tag{2.21}$$

for all $1 \leq i \leq k-1$. For $i = 0$, (2.21) need not hold, but instead we have the easily verified identity

$$p_{0,a,b} = p_{0,b,b}. \tag{2.22}$$

Now let $p_* \in \beta\mathbf{Z}\backslash\mathbf{Z}$ be arbitrary (one could pick $p_* := p$, for instance) and define $p' := \lim_{a\to p_*} p_{0,a,a} = \lim_{b\to p_*} p_{0,b,b}$. Observe from (2.19) that all the $p_{i,a,b}$ lie in the closed set $\beta\mathbf{Z} + p$, and so p' does also. Since p is minimal, there must exist $p'' \in \beta\mathbf{Z}$ such that $p = p'' + p'$. Expanding this out using (2.21) or (2.22), we conclude that

$$\lim_{h\to p''} \lim_{a\to p_*} \lim_{b\to p_*} h + p_{i,a,b} = p \tag{2.23}$$

for all $0 \leq i \leq k-1$. Applying (2.19), we conclude

$$\lim_{h\to p''} \lim_{a\to p_*} \lim_{b\to p_*} \lim_{(n_1,r_1)\to q'} \ldots \lim_{(n_b,r_b)\to q'} n + ir + p = p \tag{2.24}$$

where $n := h + m_b$ and $r := r_{a+1} + \ldots + r_b$. Now, define $q \in \beta(\mathbf{Z}\times\mathbf{N})$ to be the limit

$$q := \lim_{h\to p''} \lim_{a\to p_*} \lim_{b\to p_*} \lim_{(n_1,r_1)\to q'} \ldots \lim_{(n_b,r_b)\to q'} (n, r) \tag{2.25}$$

then we obtain Proposition 2.4.11 as desired. □

Exercise 2.4.6. Strengthen Proposition 2.4.11 by adding the additional conclusion $\lim_{(n,r)\to q} r \notin \mathbf{N}$. Using this stronger version, deduce Theorem 2.4.1 directly without using the trick of multiplying X with a cyclic shift system that was used to deduce Theorem 2.4.1 from Theorem 2.4.5.

Theorem 2.4.1 can be generalised to multiple commuting shifts:

Theorem 2.4.12 (Multiple recurrence in open covers)**.** *Let $(X, \mathcal{F})$ be a compact topological space, and let $T_1, \ldots, T_k : X \to X$ be commuting homeomorphisms. Let $(U_\alpha)_{\alpha\in A}$ be an open cover of X. Then there exists U_α such that $T_1^{-r}U_\alpha \cap \ldots \cap T_k^{-r}U_\alpha \neq \emptyset$ for infinitely many r.*

Exercise 2.4.7. By adapting one of the above arguments, prove Theorem 2.4.12.

Exercise 2.4.8. Use Theorem 2.4.12 to establish the following the *multidimensional van der Waerden theorem* (due to Gallai): if a lattice $\mathbf{Z}^d$ is finitely coloured, and $v_1, \ldots, v_d \in \mathbf{Z}^d$, then one of the colour classes contains a pattern of the form $n + rv_1, \ldots, n + rv_d$ for some $n \in \mathbf{Z}^d$ and some non-zero r.

Exercise 2.4.9. Show that Theorem 2.4.12 can fail, even for $k = 3$ and $T_1 = \mathrm{id}$, if the shift maps T_j are not assumed to commute. *Hint*: First show that in the free group F_2 on two generators a, b, and any word $w \in F_2$ and non-zero integer r, the three words $w, a^n w, b^n w$ cannot all begin with the same generator after reduction. This can be used to disprove a non-commutative multidimensional van der Waerden theorem, which can turn be used to disprove a non-commutative version of Theorem 2.4.12.

2.4.3. Proof of multiple Birkhoff. We now use van der Waerden's theorem and an additional Baire category argument to deduce Theorem 2.4.3 from Theorem 2.4.1. The key new ingredient is

Lemma 2.4.13 (Semicontinuous functions are usually continuous)**.** *Let (X, d) be a metric space, and let $F : X \to \mathbf{R}$ be semicontinuous. Then the set of points x where F is discontinuous is a set of the first category (i.e. a countable union of nowhere dense sets). In particular, by the* Baire category theorem*, if X is complete and non-empty, then F is continuous at at least one point.*

Proof. Without loss of generality we can take F to be upper semicontinuous. Suppose F is discontinuous at some point x. Then, by upper continuity, there exists a rational number q such that

$$\liminf_{y \to x} F(y) < q \leq F(x). \tag{2.26}$$

In other words, x lies in the boundary of the closed set $\{x : F(x) \geq q\}$. But boundaries of closed sets are always nowhere dense, and the claim follows. □

Now we prove Theorem 2.4.3. Without loss of generality we can take X to be minimal. Let us place a metric d on the space X. Define

the function $F : X \to \mathbf{R}^+$ by the formula

$$F(x) := \inf_{n \geq 1} \sup_{1 \leq i \leq k-1} d(T^{in}x, x). \tag{2.27}$$

It will suffice to show that $F(x) = 0$ for at least one x (notice that if the infimum is actually attained at zero for some n, then x is a periodic point and the claim is obvious). Suppose for contradiction that F is always positive. Observe that F is upper semicontinuous, and so by Lemma 2.4.13 there exists a point of continuity of F. In particular there exists a non-empty open set U such that F is bounded away from zero.

By uniform continuity of T^n, we see that if F is bounded away from zero on U, it is also bounded away from zero on $T^n V$ for any n (though the bound from below depends on n). Applying Lemma 2.4.7, we conclude that F is bounded away from zero on all of X, thus there exists $\varepsilon > 0$ such that $F(x) > \varepsilon$ for all $x \in X$. But this contradicts Theorem 2.4.1 (or Theorem 2.4.5), using the balls of radius $\varepsilon/2$ as the open cover. This contradiction completes the proof of Theorem 2.4.3.

Exercise 2.4.10. Generalise Theorem 2.4.3 to the case in which T is merely assumed to be continuous, rather than be a homeomorphism. *Hint*: let $\tilde{X} \subset X^{\mathbf{Z}}$ denote the space of all sequences $(x_n)_{n \in \mathbf{Z}}$ with $x_{n+1} = Tx_n$ for all n, with the topology induced from the product space $X^{\mathbf{Z}}$. Use a limiting argument to show that $\tilde{X}$ is non-empty. Then turn $\tilde{X}$ into a topological dynamical system and apply Theorem 2.4.3.

Exercise 2.4.11. Generalise Theorem 2.4.3 to multiple commuting shifts (analogously to how Theorem 2.4.12 generalises Theorem 2.4.1).

Exercise 2.4.12. Combine Exercises 2.4.10 and 2.4.11 by obtaining a generalisation of Theorem 2.4.3 to multiple non-invertible commuting shifts.

Exercise 2.4.13. Let $(X, \mathcal{F}, T)$ be a minimal topological dynamical system, and let $k \geq 1$. Call a point x in X *k-fold recurrent* if there exists a sequence $n_j \to \infty$ such that $T^{in_j}x \to x$ for all $0 \leq i \leq k-1$. Show that the set of k-fold recurrent points in X is *residual* (i.e. the

complement is of the first category). In particular, the set of k-fold recurrent points is dense.

Exercise 2.4.14. In the boolean Bernoulli system $(2^{\mathbf{Z}}, A \mapsto A + 1)$, show that the set A consisting of all non-zero integers which are divisible by 2 an even number of times is almost periodic. Conclude that there exists a minimal topological dynamical system $(X, \mathcal{F}, T)$ such that not every point in X is 3-fold recurrent (in the sense of the previous exercise). (Compare this with the arguments in the previous lecture, which imply that every point in X is 2-fold recurrent.)

Exercise 2.4.15. Suppose that a sequence of continuous functions $f_n : X \to \mathbf{R}$ on a metric space converges pointwise everywhere to another function $f : X \to \mathbf{R}$. Show that f is continuous on a residual set.

Exercise 2.4.16. Let $(X, \mathcal{F}, T)$ be a minimal topological dynamical system, and let $f : X \to \mathbf{R}$ be a function which is T-invariant, thus $Tf = f$. Show that if f is continuous at even one point x_0, then it has to be constant. *Hint*: x_0 is in the orbit closure of every point in X.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/01/15`. Thanks to Nilay and an anonymous commenter for corrections. Ed Dean (answering a question of Richard Borcherds) pointed out the recent paper [**Ge2008**] (building on the earlier paper [**Gi1987**]) that uses proof mining techniques to convert the topological dynamics proof of van der Waerden's theorem into a quantitative argument that gives essentially the same bounds as the classical combinatorial proof of that theorem.

2.5. Other topological recurrence results

In this lecture, we use topological dynamics methods to prove some other Ramsey-type theorems, and more specifically the polynomial van der Waerden theorem, the hypergraph Ramsey theorem, Hindman's theorem, and the Hales-Jewett theorem. In proving these statements, I have decided to focus on the ultrafilter-based proofs, rather than the combinatorial or topological proofs, though of course these styles of proof are also available for each of the above theorems.

2.5.1. The polynomial van der Waerden theorem. We first prove a significant generalisation of van der Waerden's theorem (Theorem 2.4.2):

Theorem 2.5.1. *(Polynomial van der Waerden theorem). Let $(P_1, \ldots, P_k)$ be a tuple of be integer-valued polynomials $P_1, \ldots, P_k : \mathbf{Z} \to \mathbf{Z}$ (or* tuple *for short) with $P_1(0) = \ldots = P_k(0)$. Then whenever the integers are finitely coloured, one of the colour classes will contain a pattern of the form $n + P_1(r), \ldots, n + P_k(r)$ for some $n \in \mathbf{Z}$ and $r \in \mathbf{N}$.*

This result is due to Bergelson and Leibman [**BeLe1996**], who proved it using "epsilon and delta" topological dynamical methods. A combinatorial proof was obtained more recently in [**Wa2000**]. In these notes, I will translate the Bergelson-Leibman argument to the ultrafilter setting.

Note that the case $P_j(r) := (j-1)r$ recovers the ordinary van der Waerden theorem. But the result is significantly stronger; it implies for instance that one of the colour classes contains arbitrarily many shifted geometric progressions $n+r, n+r^2, \ldots, n+r^k$, which does not obviously follow from the van der Waerden theorem. The result here only claims a single monochromatic pattern $n + P_1(r), \ldots, n + P_k(r)$, but it is not hard to amplify this theorem to show that at least one colour class contains infinitely many such patterns.

Remark 2.5.2. The theorem can fail if the hypothesis $P_1(0) = \ldots = P_k(0)$ is dropped; consider for instance the case $k = 2$, $P_1(r) = 0$, $P_2(r) = 2r + 1$, and with the integers partitioned (or coloured) into the odd and even integers. More generally, the theorem fails whenever there exists a modulus N such that the polynomials $P_1, \ldots, P_k$ are never simultaneously equal modulo N. This turns out to be the only obstruction; this is a somewhat difficult recent result of Bergelson, Leibman, and Lesigne[**BeLeLe2007**].

Exercise 2.5.1. Show that the polynomial $P(r) := (r^2 - 2)(r^2 - 3)(r^2 - 6)(r^2 - 7)(r^3 - 3)$ has a root modulo N for every positive integer N, but has no root in the integers. Thus we see that the Bergelson-Leibman-Lesigne result is stronger than the polynomial van der Waerden theorem; it does not seem possible to directly use the

latter to conclude that in every finite colouring of the integers, one of the classes contains the pattern $n, n + P(r)$.

Here are the topological dynamics and ultrafilter versions of the above theorem.

Theorem 2.5.3 (Polynomial van der Waerden theorem, topological dynamics version)**.** *Let $(P_1, \ldots, P_k)$ be a tuple with $P_1(0) = \ldots = P_k(0)$. Let $(U_\alpha)_{\alpha \in A}$ be an open cover of a topological dynamical system $(X, \mathcal{F}, T)$. Then there exists a set U_α in this cover such that $T^{P_1(r)}U \cap \ldots \cap T^{P_k(r)}U \neq \emptyset$ for at least one $r > 0$.*

Theorem 2.5.4 (Polynomial van der Waerden theorem, ultrafilter version)**.** *Let $(P_1, \ldots, P_k)$ be a tuple with $P_1(0) = \ldots = P_k(0)$, and let $p \in \beta\mathbf{Z}$ be a minimal ultrafilter. Then there exists $q \in \beta(\mathbf{Z} \times \mathbf{N})$ such that*

$$\lim_{(n,r) \to q} n + P_i(r) + p = p \text{ for all} 1 \leq i \leq k. \tag{2.28}$$

Exercise 2.5.2. Show that Theorem 2.5.1 and Theorem 2.5.3 are equivalent, and that Theorem 2.5.4 implies Theorem 2.5.3 (or Theorem 2.5.1). (For the converse implication, see Exercise 2.5.21.)

As in Section 2.4, we shall prove Theorem 2.5.4 by induction. However, the induction will be more complicated than just inducting on the number k of polynomials involved, or on the degree of these polynomials, but will instead involve a more complicated measure of the "complexity" of the polynomials being measured. Let us say that a tuple $(P_1, \ldots, P_k)$ *obeys the vdW property* if the conclusion of Theorem 2.5.4 is true for this tuple. Thus for instance, from Proposition 2.4.11 we know that any tuple of *linear* polynomials which vanish at the origin will obey the vdW property.

Our goal is to show that every tuple of polynomials which simultaneously vanish at the origin has the vdW property. The strategy will be to reduce from any given tuple to a collection of "simpler" tuples. We first begin with an easy observation, that one can always shift one of the polynomials to be zero:

Lemma 2.5.5. *(Translation invariance) Let Q be any integer-valued polynomial. Then a tuple $(P_1, \ldots, P_k)$ obeys the vdW property if and only if $(P_1 - Q, \ldots, P_k - Q)$ has the vdW property.*

Proof. Let $p \in \beta\mathbf{Z}$ be minimal. If $(P_1 - Q, \ldots, P_k - Q)$ has the vdW property, then we can find $q \in \beta(\mathbf{Z} \times \mathbf{N})$ such that

$$\lim_{(n,r)\to q} n + P_i(r) - Q(r) + p = p \text{ for all } 1 \leq i \leq k. \tag{2.29}$$

If we then define $q' := \lim_{(n,r)\to q}(n - Q(r), r) \in \beta(\mathbf{Z} \times \mathbf{N})$ one easily verifies that (2.28) holds (with q replaced by q'), and the claim holds. The converse implication is similar. □

Now we come to the key inductive step.

Lemma 2.5.6 (Inductive step)**.** *Let $(P_0, P_1, \ldots, P_k)$ be a tuple with $P_0 = 0$, and let Q be another integer-valued polynomial. Suppose that for every finite set of integers $h_1, \ldots, h_m$, the tuple $(P_i(\cdot + h_j) - P_i(h_j) - Q(\cdot))_{1\leq i\leq k; 1\leq j\leq m}$ has the vdW property. Then $(0, P_1, \ldots, P_k)$ also has the vdW property.*

Proof. This will be a reprise of the proof of Proposition 2.4.11. Given any finite number of pairs $(n_1, r_1), \ldots, (n_{b-1}, r_{b-1}) \in \mathbf{Z} \times \mathbf{N}$ with $b \geq 1$, we see from hypothesis that there exists $q_b \in \beta(\mathbf{Z} \times \mathbf{N})$ (depending on these pairs) such that

$$\lim_{(n_b,r_b)\to q_b} n_b + P_i(r_{a+1} + \ldots + r_b) - P_i(r_{a+1} + \ldots + r_{b-1}) - Q(r_b) + p = p \tag{2.30}$$

for all $0 \leq a < b$.

Now, for every $0 \leq a \leq b$ and $0 \leq i \leq k$, consider the expression $p_{a,b,i} \in \beta\mathbf{Z} + p$ defined by

$$p_{a,b,i} := \lim_{(n_1,r_1)\to q_1} \ldots \lim_{(n_b,r_b)\to q_b} P_i(r_{a+1} + \ldots + r_b) + m_b + p, \tag{2.31}$$

where $q_1, \ldots, q_b$ are defined recursively as above and

$$m_b := \sum_{i=1}^{b} n_i - Q(r_i) \tag{2.32}$$

From (2.30) we see that

$$p_{a,b,i} = p_{a,b-1,i} \tag{2.33}$$

for all $0 \leq a < b$ and $1 \leq i \leq k$, and thus

$$p_{a,b,i} = p_{a,a,i} = p_{a,a,0} \tag{2.34}$$

in this case. For $i = 0$, we have the slightly different identity

$$p_{a,b,0} = p_{b,b,0}. \tag{2.35}$$

We let $p_* \in \beta\mathbf{Z}/\mathbf{Z}$ be arbitrary, and set $p' := \lim_{a\to p_*} p_{a,a,0} = \lim_{b\to p_*} p_{b,b,0} \in \beta\mathbf{Z} + p$. By the minimality of p, we can find $p'' \in \beta\mathbf{Z}$ such that $p'' + p' = p$. We thus have

$$\lim_{h\to p''} \lim_{a\to p_*} \lim_{b\to p_*} h + p_{a,b,i} = p \tag{2.36}$$

for all $0 \leq i \leq k$. If one then sets

$$q := \lim_{h\to p''} \lim_{a\to p_*} \lim_{b\to p_*} \lim_{(n_1,r_1)\to q_1} \dots \lim_{(n_b,r_b)\to q_b} (n, r) \tag{2.37}$$

where $n := h + m_b$ and $r := r_{a+1} + \dots + r_b$, one easily verifies (2.28) as required. □

Let's see how this lemma is used in practice. Suppose we wish to show that the tuple $(0, r^2)$ has the vdW property (where we use r to denote the independent variable). Applying Lemma 2.5.6 with $Q(r) := r^2$, we reduce to showing that the tuples $((r + h_1)^2 - h_1^2 - r^2, \dots, (r + h_m)^2 - h_m^2 - r_m^2)$ have the vdW property for all finite collections $h_1, \dots, h_m$ of integers. But observe that all the polynomials in these tuples are linear polynomials that vanish at the origin. By the ordinary van der Waerden theorem, these tuples all have the vdW property, and so $(0, r^2)$ has the vdW property also.

A similar argument shows that the tuple $(0, r^2 + P_1(r), \dots, r^2 + P_k(r))$ has the vdW property whenever $P_1, \dots, P_k$ are linear polynomials that vanish at the origin. Applying Lemma 2.5.5, we see that $(Q_1(r), r^2 + P_1(r), \dots, r^2 + P_k(r))$ obeys the vdW property when Q_1 is also linear and vanishing at the origin.

Now, let us consider a tuple $(Q_1(r), Q_2(r), r^2 + P_1(r), \dots, r^2 + P_k(r))$ where Q_2 is also a linear polynomial that vanishes at the origin. The vdW property for this tuple follows from the previously established vdW properties by first applying Lemma 2.5.5 to reduce to the case $Q_1 = 0$, and then applying Lemma 2.5.6 with $Q = Q_2$. Continuing in this fashion, we see that a tuple $(Q_1(r), \dots, Q_l(r), r^2 + P_1(r), \dots, r^2 + P_k(r))$ will also obey the vdW property for any linear $Q_1, \dots, Q_l, P_1, \dots, P_k$ that vanish at the origin, for any k and l.

Now the vdW property for the tuple $(0, r^2, 2r^2)$ follows from the previously established cases and Lemma 2.5.6 with $Q(r) = r^2$.

Remark 2.5.7. It is possible to continue this inductive procedure (known as *PET induction*; the PET stands, variously, for "polynomial ergodic theorem" or "polynomial exhaustion theorem"); this is carried out in Exercise 2.5.3 below.

Exercise 2.5.3. Define the *top order monomial* of a non-zero polynomial $P(r) = a_d r^d + \ldots + a_0$ with $a_d \neq 0$ to be $a_d r^d$. Define the top order monomials of a tuple $(0, P_1, \ldots, P_k)$ to be the set of top order monomials of the $P_1, \ldots, P_k$, not counting multiplicity; for instance, the top order monomials of $(0, r^2, r^2 + r, 2r^2, 2r^2 + r)$ are $\{r^2, 2r^2\}$. Define the *weight vector* of a tuple $(P_1, \ldots, P_k)$ relative to one of its members P_i to be the infinite vector $(w_1, w_2, \ldots) \in \mathbf{Z}^{\mathbf{N}}_{\geq 0}$, where each w_d denotes the number of monomials of degree d in the top order monomials of $(P_1 - P_i, \ldots, P_k - P_i)$. Thus for instance, the tuple $(0, r^2, r^2 + r, 2r^2, 2r^2 + r)$ has weight vector $(0, 2, 0, \ldots)$ with respect to 0, but weight vector $(1, 2, 0, \ldots)$ with respect to (say) r^2. Let us say that one weight vector $(w_1, w_2, \ldots)$ is larger than another $(w'_1, w'_2, \ldots)$ if there exists $d \geq 1$ such that $w_d > w'_d$ and $w_i = w'_i$ for all $i > d$.

(1) Show that the space of all weight vectors is a well-ordered set.

(2) Show that if $(0, P_1, \ldots, P_k)$ is a tuple with $k \geq 1$ and P_1 nonlinear, and $h_1, \ldots, h_m$ are integers with $m \geq 1$, then the weight vector of $(P_i(\cdot + h_j) - P_i(h_j))_{1 \leq i \leq k; 1 \leq j \leq m}$ with respect to $P_1(\cdot + h_1)$ is strictly smaller than the weight vector of $(0, P_1, \ldots, P_k)$ with respect to P_1.

(3) Using the previous two claims, Lemma 2.5.5, and Lemma 2.5.6, deduce Theorem 2.5.4.

Exercise 2.5.4. Find a direct proof of Theorem 2.5.3 analogous to the "epsilon and delta" proof of Theorem 2.5.8 from the previous lecture. (You can look up [**BeLe1996**] if you're stuck.)

Exercise 2.5.5. Let $P_1, \ldots, P_k : \mathbf{Z} \to \mathbf{Z}^d$ be vector-valued polynomials (thus each of the d components of each of the P_i is a polynomial) which all vanish at the origin. Show that if $\mathbf{Z}^d$ is finitely

coloured, then one of the colour classes contains a pattern of the form $n + P_1(r), \ldots, n + P_k(r)$ for some $n \in \mathbf{Z}^d$ and $r \in \mathbf{N}$.

Exercise 2.5.6. Show that for any polynomial sequence $P : \mathbf{Z} \to (\mathbf{R}/\mathbf{Z})^d$ taking values in a torus, there exists integers $n_j \to \infty$ such that $P(n_j)$ converges to $P(0)$. (One can also tweak the argument to make the n_j converge to positive infinity, by the "doubling up" trick of replacing $P(n)$ with $(P(n), P(-n))$.) On the other hand, show that this claim can fail with exponential sequences such as $P(n) := 10^n\alpha \bmod 1 \in \mathbf{R}/\mathbf{Z}$ for certain values of α. Thus we see that polynomials have better recurrence properties than exponentials.

2.5.2. Ramsey's theorem. Given any finite palette K of colours, a vertex set V, and an integer $k \geq 1$, define a *K-coloured hypergraph* $G = (V, E)$ of order k on V to be a function $E : \binom{V}{k} \to K$, where $\binom{V}{k} := \{e \subset V : |e| = k\}$ denotes the k-element subsets of V. Thus for instance a hypergraph of order 1 is a vertex colouring, a hypergraph of order 2 is an edge-coloured complete graph, and so forth. We say that a hypergraph G is *monochromatic* if the edge colouring function E is constant. If W is a subset of V, we refer to the hypergraph $G \downharpoonright_W := (W, E \downharpoonright_{\binom{W}{k}})$ as a *subhypergraph* of G.

We will now prove the following result:

Theorem 2.5.8 (Hypergraph Ramsey theorem)**.** *Let K be a finite set, let $k \geq 1$, and let $G = (V, E)$ be a K-coloured hypergraph of order k on a countably infinite vertex set V. Then G contains arbitrarily large finite monochromatic subhypergraphs.*

Remark 2.5.9. There is a stronger statement known, namely that G contains an infinitely large monochromatic subhypergraph, but we will not prove this statement, known as the *infinite hypergraph Ramsey theorem*. In the case $k = 1$, these statements are the pigeonhole principle and infinite pigeonhole principle respectively, and are compared in Section 1.3 of *Structure and Randomness*.

Exercise 2.5.7. Show that Theorem 2.5.8 implies a finitary analogue: given any finite K and positive integers k, m, there exists N such that every K-coloured hyeprgraph of order k on $\{1, \ldots, N\}$ contains a monochromatic subhypergraph on m vertices. *Hint*: as

in Exercise 2.5.4, one should use a compactness and contradiction argument (as in Section 1.3 of *Structure and Randomness*).

It is not immediately obvious, but Theorem 2.5.8 is a statement about a topological dynamical system, albeit one in which the underlying group is not the integers $\mathbf{Z}$, but rather the symmetric group $\mathrm{Sym}_0(V)$, defined as the group of bijections from V to itself which are the identity outside of a finite set. More precisely, we have

Theorem 2.5.10 (Hypergraph Ramsey theorem, topological dynamics version)**.** *Let V be a countably infinite set, and let W be a finite subset of V, thus $\mathrm{Sym}_0(W) \times \mathrm{Sym}_0(V\backslash W)$ is a subgroup of $\mathrm{Sym}_0(V)$. Let $(X, \mathcal{F}, T)$ be a $\mathrm{Sym}_0(V)$-topological dynamical system, thus $(X, \mathcal{F})$ is compact metrisable and $T : \sigma \mapsto T^\sigma$ is an action of $\mathrm{Sym}_0(V)$ on X via homeomorphisms. Let $(U_\alpha)_{\alpha\in A}$ be an open cover of X, such that each U_α is $\mathrm{Sym}_0(W) \times \mathrm{Sym}_0(V\backslash W)$-invariant. Then there exists an element U_α of this cover such that for every finite set $\Gamma \subset \mathrm{Sym}_0(V)$ there exists a group element $\sigma \in \mathrm{Sym}_0(V)$ such that $\bigcap_{\gamma\in\Gamma}(T^{\gamma\sigma})^{-1}(U_\alpha) \neq \emptyset$ (i.e. there exists $x \in X$ such that $T^{\gamma\sigma}x \in U_\alpha$ for all $\gamma \in \Gamma$).*

This claim should be compared with Theorem 2.5.3 or Theorem 2.4.1.

Exercise 2.5.8. Show that Theorem 2.5.8 and Theorem 2.5.10 are equivalent. *Hint*: At some point, you will need the use the fact that the quotient space $\mathrm{Sym}_0(V)/(\mathrm{Sym}_0(W) \times \mathrm{Sym}_0(V\backslash W))$ is isomorphic to $\binom{V}{|W|}$.

As before, though, we shall only illustrate the ultrafilter approach to Ramsey's theorem, leaving the other approaches to exercises. Here, we will not work on the compactified integers $\beta\mathbf{Z}$, but rather on the compactified[25] permutations $\beta\mathrm{Sym}_0(V)$. This is a semigroup with the usual multiplication law

$$pq := \lim_{\sigma\to p} \lim_{\rho\to q} \sigma\rho. \tag{2.38}$$

[25] We will view $\mathrm{Sym}_0(V)$ here as a discrete group; one could also give this group the topology inherited from the product topology on V^V, leading to a slightly coarser (and thus less powerful) compactification, though one which is still sufficient for the arguments here.

Let us say that $p \in \beta\mathrm{Sym}_0(V)$ is *minimal* if $\beta\mathrm{Sym}_0(V)p$ is a minimal left-ideal of $\beta\mathrm{Sym}_0(V)$. One can show (by repeating Exercise 2.3.11) that every left ideal $\beta\mathrm{Sym}_0(V)p$ contains at least one minimal element; in particular, minimal elements exist.

Note that if W is a k-element subset of V, then there is an image map $\pi_W : \mathrm{Sym}_0(V) \to \binom{V}{k}$ which maps a permutation σ to its inverse image $\sigma^{-1}(W)$ of W. We can compactify this to a map[26] $\beta\pi_W : \beta\mathrm{Sym}_0(V) \to \beta\binom{V}{k}$. We can now formulate the ultrafilter version of Ramsey's theorem:

Theorem 2.5.11 (Hypergraph Ramsey theorem, ultrafilter version)**.** *Let V be countably infinite, and let $p \in \beta\mathrm{Sym}_0(V)$ be minimal. Then for every finite set W, $\beta\pi_W$ is constant on $\beta\mathrm{Sym}_0(V)p$, thus $\beta\pi_W(qp) = \beta\pi_W(p)$ for all $q \in \beta\mathrm{Sym}_0(V)$.*

This result should be compared with Proposition 2.4.11 (or Theorem 2.5.4).

Exercise 2.5.9. Show that Theorem 2.5.11 implies both Theorem 2.5.8 and Theorem 2.5.10.

Proof of Theorem 2.5.11. By relabeling we may assume $V = \{1, 2, 3, \ldots\}$ and $W = \{1, \ldots, k\}$ for some k.

Given any integers $1 \leq a < i_1 < i_2 < \ldots < i_a$, let $\sigma_{i_1,\ldots,i_a} \in \mathrm{Sym}_0(V)$ denote the permutation that swaps j with i_j for all $1 \leq j \leq a$, but leaves all other integers unchanged. We select some non-principal ultrafilter $p_* := \beta V \backslash V$ and define the sequence $p_1, p_2, \ldots \in \beta\mathrm{Sym}_0(V)$ by the formula

$$p_a := \lim_{i_1 \to p_*} \ldots \lim_{i_a \to p_*} \sigma_{i_1,\ldots,i_a} p. \tag{2.39}$$

(Note that the condition $a < i_1 < \ldots < i_a$ will be asymptotically true thanks to the choice of limits here.)

Let $a \geq k$, and let $\alpha \in \mathrm{Sym}_0(V)$ be the a permutation which is the identity outside of $\{1, \ldots, a\}$. Then we have the identity

$$\pi_W(\alpha\sigma_{i_1,\ldots,i_a}\rho) = \pi_W(\sigma_{i_{j_1},\ldots,i_{j_k}}\rho) \tag{2.40}$$

[26] Caution: $\beta\binom{V}{k}$ is *not* the same thing as $\binom{\beta V}{k}$; for instance the latter is not even compact.

for every $\rho \in \mathrm{Sym}_0(V)$, where $j_1 < \ldots < j_k$ are the elements of $\alpha^{-1}(\{1,\ldots,k\})$ in order. Taking limits as $\rho \to p$, and then inserting the resulting formula into (2.39), we conclude (after discarding the trivial limits and relabeling the rest) that

$$\beta\pi_W(\alpha p_a) = \lim_{i_1 \to p_*} \ldots \lim_{i_k \to p_*} \beta\pi_W(\sigma_{i_1,\ldots,i_k} p) \tag{2.41}$$

and in particular that $\beta\pi_W(\alpha p_a)$ is independent of α (if α is the identity outside of $\{1,\ldots,j\}$. Now let $p' := \lim_{a \to p_*} p_a$, then we have $\beta\pi_W(\alpha p')$ independent of p' for all $\alpha \in \mathrm{Sym}_0(V)$. Taking limits we conclude that $\beta\pi_W$ is constant on $(\beta\mathrm{Sym}_0(V))p'$. But from construction we see that p' lies in the closed minimal ideal $(\beta\mathrm{Sym}_0(V))p$, thus $(\beta\mathrm{Sym}_0(V))p' = (\beta\mathrm{Sym}_0(V))p$. The claim follows. □

Exercise 2.5.10. Establish Theorem 2.5.8 directly by a combinatorial argument without recourse to topological dynamics or ultrafilters. (If you are stuck, I recommend reading the classic text [**GrRoSp1980**].)

Exercise 2.5.11. Establish Theorem 2.5.10 directly by a topological dynamics argument, using combinatorial arguments for the $k = 1$ case but then proceeding by induction afterwards (as in the proof of Theorem 2.4.5).

Remark 2.5.12. More generally, one can interpret the theory of graphs and hypergraphs on a vertex set V through the lens of dynamics of $\mathrm{Sym}_0(V)$ actions; I learned this perspective from Balazs Szegedy.

2.5.3. Idempotent ultrafilters and Hindman's theorem. Thus far, we have been using ultrafilter technology rather lightly, and indeed all of the arguments so far can be converted relatively easily to the topological dynamics formalism, or even a purely combinatorial formalism, with only a moderate amount of effort. But now we will exploit some deeper properties of ultrafilters, which are more difficult to replicate in other settings. In particular, we introduce the notion of an *idempotent* ultrafilter.

Definition 2.5.13 (Idempotent)**.** Let $(S,\cdot)$ be a discrete semigroup, and let βS be given the usual semigroup operation $\cdot$. An element $p \in \beta S$ is *idempotent* if $p \cdot p = p$. (We of course define idempotence analogously if the group operation on S is denoted by $+$ instead of $\cdot$.)

Of course, 0 is idempotent, but the remarkable fact is that many other idempotents exist as well. The key tool for creating this is

Lemma 2.5.14 (Ellis-Nakamura lemma)**.** [**El1958**] *Let S be a discrete semigroup, and let K be a compact non-empty sub-semigroup of βS. Then K contains at least one idempotent.*

Proof. A simple application of Zorn's lemma shows that K contains a compact non-empty sub-semigroup K' which is minimal with respect to set inclusion. We claim that every element of K' is idempotent. To see this, let p be an arbitrary element of K'. Then observe that $K'p$ is a compact non-empty sub-semigroup of K' and must therefore be equal to K'; in particular, $p \in K'p$. (Note that semigroups need not contain an identity.) In particular, the stabiliser $K'' := \{q \in K' : qp = p\}$ is non-empty. But one easily observes that this stabiliser is also a compact sub-semigroup of K', and so $K'' = K'$. In particular, p must stabilise itself, i.e. it is idempotent. □

Remark 2.5.15. *A posteriori*, this results shows that the minimal non-empty sub-semigroups K' are in fact just the singleton sets consisting of idempotents. But one cannot really see this without first deriving all of Lemma 2.5.14.

Idempotence turns out to be particularly powerful when combined with minimality, and to this end we observe the following corollary of the above lemma:

Corollary 2.5.16. *Let S be a discrete semigroup. For every $p \in \beta S$, there exists $q \in (\beta S)p$ which is both minimal and idempotent.*

Proof. By Exercise 2.3.11, there exists $r \in (\beta S)p$ which is minimal. It is then easy to see that every element of $(\beta S)r$ is minimal. Since $(\beta S)r \subset (\beta S)p$ is a compact non-empty sub-semigroup of βS, the claim now follows from Lemma 2.5.14. □

Remark 2.5.17. Somewhat amusingly, minimal idempotent ultrafilters require *three* distinct applications of Zorn's lemma to construct: one to define the compactified space βS, one to locate a minimal left-ideal, and one to locate an idempotent inside that ideal! It seems particularly challenging therefore to define civilised substitutes for this tool which do not explicitly use the axiom of choice.

What can we do with minimal idempotent ultrafilters? One particularly striking example is *Hindman's theorem*[**Hi1974**]. Given any set A of positive integers, define $FS(A)$ to be the set of all finite sums $\sum_{n\in B} n$ from A, where B ranges over all finite non-empty subsets of A. (For instance, if $A = \{1, 2, 4, \ldots\}$ are the powers of 2, then $FS(A) = \mathbf{N}$.)

Theorem 2.5.18 (Hindman's theorem)**.** *Suppose that the natural numbers* $\mathbf{N}$ *are finitely coloured. Then one of the colour classes contains a set of the form* $FS(A)$ *for some infinite set* A*.*

Remark 2.5.19. Theorem 2.5.18 implies *Folkman's theorem*[**Fo1970**], which has the same hypothesis but concludes that one of the colour classes contains sets of the form $FS(A)$ for arbitrarily large but finite sets A. In the converse direction, it does not seem possible to easily deduce Hindman's theorem from Folkman's theorem.

Exercise 2.5.12. Folkman's theorem in turn implies *Schur's theorem*[**Sc1916**], which asserts that if the natural numbers are finitely coloured, one of the colour classes contains a set of the form $FS(\{x, y\}) = \{x, y, x + y\}$ for some x, y (compare with the $k = 3$ case of van der Waerden's theorem). Using the Cayley graph construction, deduce Schur's theorem from Ramsey's theorem (the $k = 2$ case of Theorem 2.5.8). Thus we see that there are some connections between the various Ramsey-type theorems discussed here.

Proof of Theorem 2.5.18. By Corollary 2.5.16, we can find a minimal idempotent element p in $\beta\mathbf{N}$; note that as no element of $\mathbf{N}$ is minimal (cf. Exercise 2.3.8), we know that $p \notin \mathbf{N}$. Let $c : \mathbf{N} \to \{1, \ldots, m\}$ denote the given colouring function, then $\beta c(p)$ is a colour in $\{1, \ldots, m\}$. Since

$$\lim_{n\to p} \beta c(n) = \beta c(p) \tag{2.42}$$

and

$$\lim_{n\to p} \beta c(n + p) = \beta c(p + p) = \beta c(p) \tag{2.43}$$

we may find a positive integer n_1 such that $\beta c(n_1) = \beta c(n_1 + p) = \beta c(p)$. Now from (2.42), (2.43) and the similar calculations

$$\lim_{n\to p} \beta c(n_1 + n) = \beta c(n_1 + p) = \beta c(p) \tag{2.44}$$

and

$$\lim_{n\to p} \beta c(n_1+n+p) = \beta c(n_1+p+p) = \beta c(n_1+p) = \beta c(p) \tag{2.45}$$

we can find an integer $n_2 > n_1$ such that $\beta c(n_2) = \beta c(n_2+p) = \beta c(n_2+n_1) = \beta c(n_2+n_1+p) = \beta c(p)$, thus $\beta c(m) = \beta c(m+p) = \beta c(p)$ for all $m \in FS(\{n_1, n_2\})$. Continuing inductively in this fashion, one can find $n_1 < n_2 < n_3 < \ldots$ such that $\beta c(m) = \beta c(m+p) = \beta c(p)$ for all $m \in FS(\{n_1, \ldots, n_k\})$ and all k. If we set $A := \{n_1, n_2, \ldots\}$, the claim follows. □

Remark 2.5.20. Purely combinatorial (and quite succinct) proofs of Hindman's theorem exist - see for instance the one in [**GrRoSp1980**] - but they generally rely on some *ad hoc* trickery. Here, the trickery has been encapsulated into the existence of minimal idempotent ultrafilters, which can be reused in other contexts (for instance, we will use it to prove the Hales-Jewett theorem below).

Exercise 2.5.13. Define an *IP-set* to be a set of positive integers which contains a subset of the form $FS(A)$ for some infinite A. Show that if an IP-set S is finitely coloured, then one of its colour classes is also an IP-set. *Hint*: S contains $FS(A)$ for some infinite $A = \{a_1, a_2, a_3, \ldots\}$. Show that the set $\bigcap_{n=1}^{\infty} \beta FS(\{a_n, a_{n+1}, \ldots\})$ is a compact non-empty semigroup and thus contains a minimal idempotent ultrafilter p. Use this p to repeat the proof of Theorem 2.5.18.

2.5.4. The Hales-Jewett theorem. Given a finite alphabet A, let $A^{<\omega}$ be the free semigroup generated by A, i.e. the set of all finite non-empty words using the alphabet A, with concatenation as the group operation. (E.g. if $A = \{a, b, c\}$, then $A^{<\omega}$ contains words such as abc and cbb, with $abc \cdot cbb = abccbb$.) If we add another letter $*$ to A (the "wildcard" letter), we create a larger semigroup $(A \cup \{*\})^{<\omega}$ (e.g. containing words such as $ab**c*$). We of course assume that $*$ was not already present in A. Given any letter $x \in A$, we have a semigroup homomorphism $\pi_x : (A \cup \{*\})^{<\omega} \to A^{<\omega}$ which substitutes every occurrence of the wildcard $*$ with x and leaves all other letters unchanged. (For instance, $\pi_a(ab**c*) = abaaca$.) Define a *combinatorial line* in $A^{<\omega}$ to be any set of the form $\{\pi_x(v) : x \in A\}$ for some $v \in (A \cup \{*\})^{<\omega} \backslash A^{<\omega}$. For instance, if $A = \{a, b, c\}$, then

$\{abaaca, abbbcb, abcccc\}$ is a combinatorial line, generated by the word $v = ab * *c*$.

We shall prove the following fundamental theorem.

Theorem 2.5.21 (Hales-Jewett theorem)**.** [**HaJe1963**] *Let A be a finite alphabet. If $A^{<\omega}$ is finitely coloured, then one of the colour classes contains a combinatorial line.*

Exercise 2.5.14. Show that the Hales-Jewett theorem has the following equivalent formulation: for every finite alphabet A and any $m \geq 1$ there exists N such that if A^N is partitioned into m classes, then one of the classes contains a combinatorial line.

Exercise 2.5.15. Assume the Hales-Jewett theorem. In this exercise we compare the strength of this theorem against other Ramsey-type theorems.

(1) Deduce van der Waerden's theorem (Theorem 2.4.2). *Hint*: the base k representation of the non-negative natural numbers provides a map from $\{0, \ldots, k-1\}^{<\omega}$ to $\mathbf{Z}_{\geq 0}$.

(2) Deduce the multidimensional van der Waerden's theorem of Gallai (Exercise 2.4.8).

(3) Deduce the *syndetic van der Waerden theorem* of Furstenberg[**Fu1977**] if the integers are finitely coloured and k is a positive integer, then there are infinitely many monochromatic arithmetic progressions $n, n+r, \ldots, n+(k-1)r$ of length k, and furthermore the set of all the step sizes r which appear in such progressions is syndetic (i.e. it has bounded caps). *Hint*: argue by contradiction, assuming that the set of all step sizes has arbitrarily long gaps, and use the Hales-Jewett theorem in a manner adapted to these gaps. (For an additional challenge, show that there exists a *single* colour class whose progressions of length k have spacings in a syndetic set for every k.)

(4) Deduce the *IP-van der Waerden theorem*: If the integers are finitely coloured, k is a positive integer, and S is an IP-set (see Exercise 2.5.13), show that there are infinitely many monochromatic arithmetic progressions whose step size lies

in S. (For an additional challenge, show that one of the classes has the property that for every k, the spacings of the k-term progressions in that class forms an *IP^*-set*, i.e. it has non-empty intersection with every IP-set. There is an even stronger topological dynamics version of this statement, due to Furstenberg and Weiss[**FuWe1978**], which I will not describe here.)

(5) For any $d \geq 1$, define a *d-dimensional combinatorial subspace* of $A^{<\omega}$ to be a set of the form $\{\pi_{x_1,\ldots,x_d}(v) : x_1, \ldots, x_d \in A\}$, where $v \in (A \cup \{*_1, \ldots, *_d\})^{<\omega}$ is a word containing at least one copy of each of the d wildcards $*_1, \ldots, *_d$, and $\pi_{x_1,\ldots,x_d} : (A \cup \{*_1, \ldots, *_d\})^{<\omega} \to A^{<\omega}$ is the homomorphism that substitutes each wildcard $*_j$ with x_j. Show that if $A^{<\omega}$ is finitely coloured, then one of the colour classes contains arbitrarily high-dimensional combinatorial subspaces.

(6) Let F be a finite field. If the vector space $\lim_{n\to\infty} F^n$ (the inverse limit of the finite vector spaces F^n) is finitely coloured, show that one of the colour classes contains arbitrarily high-dimensional affine subspaces over F. (This *geometric Ramsey theorem* is due to [**GrLeRo1972**].)

We now give an ultrafilter-based proof of the Hales-Jewett theorem due to Blass[**Bl1993**]. As usual, the first step is to obtain a statement involving ultrafilters rather than colourings:

Proposition 2.5.22 (Hales-Jewett theorem, ultrafilter version)**.** *Let A be a finite alphabet, and let p be a minimal idempotent element of the semigroup $\beta(A^{<\omega})$. Then there exists $q \in \beta(A \cup \{*\})^{<\omega}) \backslash \beta(A^{<\omega})$ such that $\beta\pi_x(q) = p$ for all $x \in A$.*

Exercise 2.5.16. Deduce Theorem 2.5.21 from Proposition 2.5.22.

To prove Proposition 2.5.22, we need a variant of Corollary 2.5.16. If $(S, \cdot)$ is a discrete semigroup and p and q are two idempotents in βS, let us write $p \prec q$ if we have $pq = qp = p$.

Exercise 2.5.17. Show that $\prec$ is a partial ordering on the idempotents of βS, and that an idempotent is minimal in βS if and only if it is minimal with respect to $\prec$.

Lemma 2.5.23. *Let S be a discrete semigroup, and let p be an idempotent in βS. Then there exists a minimal idempotent q in βS such that $q \prec p$.*

Proof. By Exercise 2.3.11 (generalised to arbitrary discrete semigroups S), $(\beta S)p$ contains a minimal left-ideal $(\beta S)r$. By Lemma 2.5.14, $(\beta S)r$ contains an idempotent s. Since $s \in (\beta S)p$ and p is idempotent, we conclude $sp = s$. If we then set $q := ps$, we easily check that q is idempotent, that $q \prec p$, and (since q lies in the minimal left-ideal $(\beta S)r$) it is minimal. The claim follows. □

Proof of Proposition 2.5.22. Since p is an idempotent element of $\beta(A^{<\omega})$, it is also an idempotent element of $\beta(A \cup \{*\})^{<\omega}$. It need not be minimal in that semigroup, though. However, by Lemma 2.5.23, we can find a minimal idempotent q in $\beta(A \cup \{*\})^{<\omega}$ such that $q \prec p$.

Now let $x \in A$. Since $\pi_x : (A \cup \{*\})^{<\omega} \to A^{<\omega}$ is a homomorphism, $\beta\pi_x : \beta(A \cup \{*\})^{<\omega} \to \beta A^{<\omega}$ is also a homomorphism (why?). Since q is idempotent and $q \prec p$ (note that these are both purely *algebraic* statements), we conclude that $\beta\pi_x(q)$ is idempotent and $\beta\pi_x(q) \prec \beta\pi_x(p)$. But $\beta\pi_x(p) = p$ is minimal in $\beta A^{<\omega}$, hence by Exercise 2.5.17, we have $\beta\pi_x(q) = p$. The claim follows. □

Exercise 2.5.18. Adapt the above proof to give an alternate proof of the ultrafilter version of van der Waerden's theorem (Proposition 2.5.22) which relies on idempotence rather than on induction on k. (If you are stuck, read the proof of [**Gl2003**, Proposition 1.55].)

Remark 2.5.24. Several of the above Ramsey-type theorems can be unified. For instance, the polynomial van der Waerden theorem and the Hales-Jewett theorem have been unified into the polynomial Hales-Jewett theorem of Bergelson and Leibman[**BeLe1999**] (see also [**Wa2000**]). This type of Ramsey theory is still an active subject, and we do not yet have a comprehensive and systematic theory (or a "universal" Ramsey theorem) that encompasses all known examples.

Exercise 2.5.19. Let X be an at most countable set (with the discrete topology), and let $\mathcal{F}$ be a family of subsets of X. Show that the following two statements are equivalent:

(1) Whenever X is finitely coloured, one of the colour classes contains a subset in $\mathcal{F}$.

(2) There exists $p \in \beta X$ such that every neighbourhood of p contains a subset in $\mathcal{F}$.

Exercise 2.5.20. Let X, Y be at most countable sets with the discrete topology, and let $f_1, \ldots, f_k : Y \to X$ be a finite collection of functions. Show that the following two statements are equivalent:

(1) Whenever X is finitely coloured, one of the colour classes contains a set $\{f_1(y), \ldots, f_k(y)\}$ for some $y \in Y$.

(2) There exists $q \in \beta Y$ such that $\beta f_1(q) = \ldots = \beta f_k(q)$.

Hint: look at the closure of $\{(f_1(y), \ldots, f_k(y)) : y \in Y\}$ in $(\beta X)^k$.

Exercise 2.5.21. Using the previous exercise, deduce Theorem 2.5.4 from Theorem 2.5.1, and deduce Theorem 2.5.11 from Theorem 2.5.8.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/01/21`. Thanks to Yury, Liu Xiao Chuan, Nilay and an anonymous commenter for corrections.

2.6. Isometric systems and isometric extensions

In this lecture, we move away from recurrence, and instead focus on the *structure* of topological dynamical systems. One remarkable feature of this subject is that starting from fairly "soft" notions of structure, such as topological structure, one can extract much more "hard" or "rigid" notions of structure, such as *geometric* or *algebraic* structure. The key concept needed to capture this structure is that of an *isometric system*, or more generally an *isometric extension*, which we shall discuss in this lecture. As an application of this theory we characterise the distribution of polynomial sequences in torii (a baby case of a variant of Ratner's theorem due to [**Gr1961**], which we will cover in Section 2.16).

2.6.1. Isometric systems. We begin with a key definition.

Definition 2.6.1 (Equicontinuous and isometric systems)**.** Let $(X, \mathcal{F}, T)$ be a topological dynamical system.

(1) We say that the system is *isometric* if there exists a metric d on X such that the shift maps $T^n : X \to X$ are all isometries, thus $d(T^n x, T^n y) = d(x, y)$ for all n and all x, y. (Of course, once T is an isometry, all powers T^n are automatically isometries also, so it suffices to check the $n = 1$ case.)

(2) We say that the system is *equicontinuous* if there exists a metric d on X such that the shift maps $T^n : X \to X$ form a uniformly equicontinuous family, thus for every $\varepsilon > 0$ there exists $\delta > 0$ such that $d(T^n x, T^n y) \leq \varepsilon$ whenever n, x, y are such that $d(x, y) \leq \delta$. (As X is compact, equicontinuity and uniform equicontinuity are equivalent concepts.)

Example 2.6.2. The circle shift $x \mapsto x + \alpha$ on $\mathbf{R}/\mathbf{Z}$ is both isometric and equicontinuous. On the other hand, the Bernoulli shift on $\{0, 1\}^{\mathbf{Z}}$ is neither isometric nor equicontinuous (why?).

Example 2.6.3. Any finite dynamical system is both isometric and equicontinuous (as one can see by using the discrete metric).

Since all metrics are essentially equivalent on compact spaces, we see that the choice of metric is not actually important when checking equicontinuity, but it seems to be more important when checking for isometry. Nevertheless, there is actually no distinction between the two properties:

Exercise 2.6.1. Show that a topological dynamical system is isometric if and only if it is equicontinuous. *Hint*: one direction is obvious. For the other, if T^n is a uniformly equicontinuous family with respect to a metric d, consider the modified metric $\tilde{d}(x, y) := \sup_n d(T^n x, T^n y)$.

Remark 2.6.4. From this exercise we see that we can upgrade *topological* structure (equicontinuity) to *geometric* structure (isometry). The motif of studying topology through geometry pervades modern topology; witness for instance Perelman's proof of the Poincaré conjecture (Chapter 3).

Exercise 2.6.2 (Ultrafilter characterisation of equicontinuity)**.** Let $(X, \mathcal{F}, T)$ be a topological dynamical system. Show that X is equicontinuous if and only if the maps $T^p : X \to X$ are homeomorphisms for every $p \in \beta\mathbf{Z}$.

Now we upgrade the geometric structure of isometry to the *algebraic* structure of being a compact abelian group action.

Definition 2.6.5 (Kronecker system)**.** A topological dynamical system $(X, \mathcal{F}, T)$ is said to be a *Kronecker system* if it is isomorphic to a system of the form $(K, \mathcal{K}, S)$, where $(K, +, \mathcal{K})$ is a compact abelian metrisable topological group[27], and $S : x \mapsto x + \alpha$ is a group rotation for some $\alpha \in K$.

Example 2.6.6. The circle rotation system is a Kronecker system, as is the standard shift $x \mapsto x + 1$ on a cyclic group $\mathbf{Z}/N\mathbf{Z}$. Any product of Kronecker systems is again a Kronecker system.

Let us first observe that a Kronecker system is equicontinuous (and hence isometric). Indeed, the compactness of the topological group K (and the joint continuity of the addition law $+ : K \times K \to K$) easily ensures that the group rotations $g : x \mapsto x + g$ are uniformly equicontinuous as $g \in K$ varies. Since the shifts $T^n : x \mapsto x + n\alpha$ are all group rotations, the claim follows.

On the other hand, not every equicontinuous or isometric system is Kronecker. Consider for instance a finite dynamical system which is the disjoint union of two cyclic shifts of distinct order; it is not hard to see that this is not a Kronecker system. Nevertheless, it clearly contains Kronecker systems within it. Indeed, we have

Proposition 2.6.7. *Every* minimal *equicontinuous (or isometric) system $(X, \mathcal{F}, T)$ is a Kronecker system, i.e. isomorphic to an abelian group rotation $(K, \mathcal{K}, x \mapsto x+\alpha)$. Furthermore, the orbit $\{n\alpha : n \in \mathbf{Z}\}$ is dense in K.*

Proof. By Exercise 2.6.1, we may assume that the system is isometric, thus we can find a metric d such that all the shift maps T^n are

[27] A *topological group* is a group with a topology, such that the group operations $x \mapsto x^{-1}$ and $(x, y) \mapsto xy$ are continuous.

isometries. We view the T^n as lying inside the space $C(X \to X)$ of continuous maps from X to itself, endowed with the uniform topology. Let $G \subset C(X \to X)$ be the closure of the maps $\{T^n : n \in \mathbf{Z}\}$. One easily verifies that G is a closed metrisable topological group of isometries in $C(X \to X)$; from the Arzelá-Ascoli theorem we see that G is compact. Also, since T^n and T^m commute for every n and m, we see upon taking limits that G is abelian.

Now let $x \in X$ be an arbitrary point. Then we see that the image $\{f(x) : f \in G\}$ of G under the evaluation map $f \mapsto f(x)$ is a compact non-empty invariant subset of X, and thus equal to all of X by minimality. If we then define the stabiliser $\Gamma := \{f \in G : f(x) = x\}$, we see that Γ is a closed (hence compact) subgroup of the abelian group G. Since $X = \{f(x) : f \in G\}$, we thus see that there is a continuous bijection $f\Gamma \mapsto f(x)$ from the quotient group $K := G/\Gamma$ (with the quotient topology) to X. Since both spaces here are compact Hausdorff, this map is a homeomorphism. This map is thus an isomorphism of topological dynamical systems between the Kronecker system K (with the group rotation given by $\alpha := T \bmod \Gamma \in G/\Gamma$) and X. Since K is a compact metrisable (thanks to Hausdorff distance) topological group, the claim follows (relabeling the group operation as $+$). Note that the density of $\{n\alpha : n \in \mathbf{Z}\}$ in K is clear from construction. □

Remark 2.6.8. Once one knows that X is homeomorphic to a Kronecker system with $\{n\alpha : n \in \mathbf{Z}\}$ dense, one can *a posteriori* return to the proof and conclude that the stabiliser Γ is trivial. But I do not see a way to establish that fact directly. In any case, when we move to isometric extensions below, the analogue of the stabiliser Γ can certainly be non-trivial.

To get from minimal isometric systems to non-minimal isometric systems, we can use

Proposition 2.6.9. *Any isometric system $(X, \mathcal{F}, T)$ can be partitioned as the union of disjoint minimal isometric systems.*

Proof. Since minimal systems are automatically disjoint, it suffices to show that every point $x \in X$ is contained in a minimal dynamical system, or equivalently that the orbit closure $\overline{T^{\mathbf{Z}}x}$ is minimal. If this

is not the case, then there exists $y \in \overline{T^{\mathbf{Z}}x}$ such that x does not lie in the orbit closure of y. But by definition of orbit closure, we can find a sequence n_j such that $T^{n_j}x$ converges to y. By the isometry property, this implies that $T^{-n_j}y$ converges to x, and so x is indeed in the orbit closure of y, a contradiction. □

Thus every equicontinuous or isometric system can be expressed as a union of disjoint Kronecker systems.

We can use the algebraic structure of isometric systems to obtain much quicker (and slightly stronger) proofs of various recurrence theorems. For instance, we can give a short proof of (a slight strengthening of) the multiple Birkhoff recurrence theorem (Theorem 2.4.3) as follows:

Proposition 2.6.10 (Multiple Birkhoff for isometric systems)**.** *Let $(X, \mathcal{F}, T)$ be an isometric system. Then for every $x \in X$ there exists a sequence $n_j \to \infty$ such that $T^{kn_j}x \to x$ for every integer k.*

Proof. By Proposition 2.6.9 followed by Proposition 2.6.7, it suffices to check this for Kronecker systems $(K, \mathcal{K}, x \mapsto x + \alpha)$ in which $\{n\alpha : n \in \mathbf{Z}\}$ is dense in K. But then we can find a sequence n_j such that $n_j\alpha \to 0$ in K, and thus (since K is a topological group) $kn_j\alpha \to 0$ in K for all k. The claim follows. □

The above argument illustrates one of the reasons why it is desirable to have an algebraic structural theory of various types of dynamical systems; it makes it much easier to answer many interesting questions regarding such systems, such as those involving recurrence.

2.6.2. The Kronecker factor. We have seen isometric systems are basically Kronecker systems (or unions thereof). Of course, not all systems are isometric. However, it turns out that every system contains a maximal isometric *factor*. Recall that a factor of a topological dynamical system $(X, \mathcal{F}, T)$ is a surjective morphism $\pi : X \to Y$ from X to another topological dynamical system $(Y, \mathcal{G}, S)$. (We shall sometimes abuse notation and refer to $\pi : X \to Y$ as the factor, when it is really the quadruplet $(\pi, Y, \mathcal{G}, S)$.) We say that one factor $\pi : X \to Y$ *refines* or is *finer than* another factor $\pi' : X \to Y'$ if

we can factorise $\pi' = f \circ \pi$ for some continuous map $f : Y \to Y'$. (Note from surjectivity that this map, if it exists, is unique.) We say that two factors are *equivalent* if they refine each other. Observe that modulo equivalence, refinement is a partial ordering on factors.

Example 2.6.11. The identity factor $\mathrm{id} : X \to X$ is finer than any other factor of X, which in turn is finer than the trivial factor $\mathrm{pt} : X \to \mathrm{pt}$ that maps to a point.

Exercise 2.6.3. Show that any factor of a minimal topological dynamical system is again minimal.

We note two useful operations on factors. Firstly, given two factors $\pi : X \to Y = Y$ and $\pi' : X \to Y'$, one can define their *join* $\pi \vee \pi' : X \to Y \vee Y'$, where $Y \vee Y' := \{(\pi(x), \pi'(x)) : x \in X\} \subset Y \times Y'$ is the compact subspace of the product system $Y \times Y'$, and $\pi \vee \pi' : X \to Y \vee Y'$ is the surjective morphism $\pi \vee \pi' : x \mapsto (\pi(x), \pi'(x))$. One can verify that $\pi \vee \pi'$ is the least common refinement of π and π', hence the name.

Secondly, given a chain $(\pi_\alpha)_{\alpha \in A}$ of factors $\pi_\alpha : X \to Y_\alpha$ (thus π_α refines π_β for all $\alpha > \beta$), one can form their *inverse limit* $\pi = \lim_{\leftarrow}(\pi_\alpha)_{\alpha \in A} : X \to Y = \lim_{\leftarrow}(Y_\alpha)_{\alpha \in A}$ by first letting $f_{\alpha\beta} : Y_\alpha \to Y_\beta$ be the factoring maps for all $\alpha > \beta$, observing that $f_{\beta\gamma} \circ f_{\alpha\beta} = f_{\alpha\gamma}$ for all $\alpha > \beta > \gamma$, and then defining $Y \subset \prod_\alpha Y_\alpha$ to be the compact subspace of the product system $\prod_\alpha Y_\alpha$ defined as

$$Y := \{(y_\alpha)_{\alpha \in A} : f_{\alpha\beta}(y_\alpha) = y_\beta \text{ whenever } \alpha > \beta\} \tag{2.46}$$

and then setting $\pi : x \mapsto (\pi_\alpha(x))_{\alpha \in A}$. One easily verifies that π is indeed a factor of X, and it is the least upper bound of the π_α.

Next, we observe that these operations interact well with the isometry property:

Exercise 2.6.4. Let $\pi : X \to Y$ and $\pi' : X \to Y'$ be two factors such that Y and Y' are both isometric. Then $\pi \vee \pi' : X \to Y \vee Y'$ is also isometric.

Lemma 2.6.12. *Let $(\pi_\alpha)_{\alpha \in A}$ be a totally ordered set of factors $\pi_\alpha : X \to Y_\alpha$ with $Y_\alpha = (Y_\alpha, \mathcal{G}_\alpha, S_\alpha)$ isometric. Then the inverse limit $\pi : X \to Y$ of the π_α is such that Y is also isometric.*

Proof. Observe that we have factor maps $f_\alpha : Y \to Y_\alpha$ which are surjective morphisms, which themselves factor as $f_\beta = f_{\alpha\beta} \circ f_\alpha$ for $\alpha > \beta$ and some surjective morphisms $f_{\alpha\beta} : Y_\alpha \to Y_\beta$. Let us fix some metric d on Y. For each $\alpha \in A$, consider the compact subset $\Delta_\alpha := \{(y, y') \in Y \times Y : f_\alpha(y) = f_\alpha(y')\}$ of $Y \times Y$. These sets decrease as α increases, and their intersection is the diagonal $\{(y, y) : y \in Y\}$ (why?). Applying the *finite intersection property* in the compact sets $\{(y, y') \cap \Delta_\alpha : d(y, y') \geq \varepsilon\}$, we conclude that for every $\varepsilon > 0$ there exists α such that $d(y, y') < \varepsilon$ whenever $f_\alpha(y) = f_\alpha(y')$.

Now suppose for contradiction that Y is not isometric, and hence not uniformly equicontinuous. Then there exists a sequences $y_j, y'_j \in Y$ with $d(y_j, y'_j) \to 0$, an $\varepsilon > 0$, and a sequence n_j of integers such that $d(S^{n_j} y_j, S^{n_j} y'_j) > \varepsilon$. By compactness we may assume that y_j, y'_j both converge to the same point. But by the preceding discussion, we can find $\alpha \in A$ such that $d(y, y') < \varepsilon/4$ whenever $f_\alpha(y) = f_\alpha(y')$. In other words, for any z in Y_α, the fibre $f_\alpha^{-1}(\{z\})$ has diameter at most $\varepsilon/4$.

Now let $z_j := f_\alpha(y_j)$ and $z'_j := f_\alpha(y'_j)$. Then z_j and z'_j converge to the same point z in Y_α, and so by equicontinuity of Y_α, $d(S_\alpha^{n_j} z_j, S_\alpha^{n_j} z'_j)$ goes to zero. By compactness and passing to a subsequence we can assume that $S_\alpha^{n_j} z_j$ and $S_\alpha^{n_j} z'_j$ both converge to some point z_* in Y_α. On the other hand, from the preceding discussion and the triangle inequality, we see that the fibres $f_\alpha^{-1}(\{S_\alpha^{n_j} z_j\})$ and $f_\alpha^{-1}(\{S_\alpha^{n_j} z'_j\})$ are separated by a distance at least $\varepsilon/2$ in Y. On the other hand, the distance between $f_\alpha^{-1}(\{S_\alpha^{n_j} z_j\})$ and $f_\alpha^{-1}(\{z_*\})$ must go to zero as $j \to \infty$ (as a simple sequential compactness argument shows), and similarly the distance between $f_\alpha^{-1}(\{S_\alpha^{n_j} z'_j\})$ and $f_\alpha^{-1}(\{z_*\})$ goes to zero. Since $f_\alpha^{-1}(\{z_*\})$ has diameter at most $\varepsilon/4$, we obtain a contradiction. The claim follows. □

Combining Exercise 2.6.2 and Lemma 2.6.12 with Zorn's lemma (and noting that with the trivial factor $\mathrm{pt} : X \to \mathrm{pt}$, the image pt is clearly isometric) we obtain

Corollary 2.6.13 (Existence of maximal isometric factor)**.** *For every topological dynamical system $(X, \mathcal{F}, T)$ there is a factor $\pi : X \to Y$ with Y isometric, and which is maximal with respect to refinement*

among all such factors with this property. This factor is unique up to equivalence.

By Proposition 2.6.7 and Exercise 2.6.3, the maximal isometric factor of a minimal system is a Kronecker system, and we refer to it as the*Kronecker factor* of that minimal system X.

Exercise 2.6.5 (Explicit description of Kronecker factor)**.** Let $(X, \mathcal{F}, T)$ be a minimal topological dynamical system, and let $Q \subset X \times X$ be the set

$$Q := \bigcap_V \overline{(T \times T)^{\mathbf{Z}}(V)} \tag{2.47}$$

where V ranges over all open neighbourhoods of the diagonal $\{(x, x) : x \in X\}$ of $X \times X$, and $T \times T : (x, y) \mapsto (Tx, Ty)$ is the product shift. Let $\sim$ be the finest equivalence relation on X such that the set $R_\sim := \{(x, y) \in X \times X : x \sim y\}$ is closed and contains Q. (The existence and uniqueness of $\sim$ can be established by intersecting $R_\sim$ over all candidates $\sim$ together.) Show that the projection map $\pi : X \to X/\sim$ to the equivalence classes of $\sim$ (with the quotient topology) is (up to isomorphism) the Kronecker factor of X.

The Kronecker factor is also closely related to the concept of an eigenfunction. We say that a continuous function $f : X \to \mathbf{C}$ is an *eigenfunction* of a topological dynamical system $(X, \mathcal{F}, T)$ if it is not identically zero and we have $Tf = \lambda f$ for some $\lambda \in \mathbf{C}$, which we refer to as an *eigenvalue* for T.

Exercise 2.6.6. Let $(X, \mathcal{F}, T)$ be a minimal topological dynamical system.

(1) Show that if λ is an eigenvalue for T, then λ lies in the unit circle $S^1 := \{z \in \mathbf{C} : |z| = 1\}$, and furthermore there exists a unimodular eigenfunction $g : X \to S^1$ with this eigenvalue. *Hint*: the zero set of an eigenfunction is a closed shift-invariant subset of X.

(2) Show that for every eigenvalue λ, the eigenspace $\{f \in C(X) : Tf = \lambda f\}$ is one-dimensional, i.e. all eigenvalues have geometric multiplicity 1. *Hint*: first establish this in the case $\lambda = 1$.

(3) If $g : X \to S^1$ is a unimodular eigenfunction with non-trivial eigenvalue $\lambda \neq 1$, show that $g : X \to g(X)$ is an isometric factor of X, where $g(X) \subset S^1$ is given the shift $z \mapsto \overline{\lambda} z$. Conclude in particular that $g = c\chi \circ \pi$, where $\pi : X \to K$ is the Kronecker factor, $\chi : K \to S^1$ is a character of K, and c is a constant. Conversely, show that all functions of the form $c\chi \circ \pi$ are eigenfunctions[28].

We will see eigenfunctions (and various generalisations of the eigenfunction concept) playing a decisive role in the structure theory of measure-preserving systems, which we will get to in a few lectures.

2.6.3. Isometric extensions. To cover more general systems than just the isometric systems, we need the more flexible concept of an *isometric extension*.

Definition 2.6.14 (Extensions)**.** If $\pi : X \to Y = (Y, \mathcal{G}, S)$ is a factor of $(X, \mathcal{F}, T)$, we say that $(X, \mathcal{F}, T)$ is an *extension* of $(Y, \mathcal{G}, S)$, and refer to $\pi : X \to Y$ as the *projection map* or *factor map*. We refer to the (compact) spaces $\pi^{-1}(\{y\})$ for $y \in Y$ as the *fibres* of this extension.

Example 2.6.15. The skew shift (Example 2.2.4) is an extension of the circle shift, with the fibres being the "vertical" circles. All systems are extensions of a point, and (somewhat trivially) are also extensions of themselves.

Definition 2.6.16 (Isometric extensions)**.** Let $(X, \mathcal{F}, T)$ be an extension of a topological dynamical system $(Y, \mathcal{G}, S)$ with projection map $\pi : X \to Y$. We say that this extension is *isometric* if there exists a metric $d_y : \pi^{-1}(\{y\}) \times \pi^{-1}(\{y\}) \to \mathbf{R}^+$ on each fiber $\pi^{-1}(\{y\})$ with the following properties:

(a) (Isometry) For every $y \in Y$ and $x, x' \in \pi^{-1}(\{y\})$, we have $d_{Sy}(Tx, Tx') = d_y(x, x')$.

(b) (Continuity) The function $d : \bigcup_{y \in Y} \pi^{-1}(\{y\}) \times \pi^{-1}(\{y\}) \to \mathbf{R}^+$ formed by gluing together all the d_y is continuous (where

[28] From this, it is possible to reconstruct the Kronecker factor canonically from the eigenfunctions of X; we leave the details to the reader.

we view the domain as a compact subspace $\{(x, x') \in X \times X : \pi(x) = \pi(x')\}$ of $X \times X$).

(c) (Isometry, again) For any $y, y' \in Y$, the metric spaces $(\pi^{-1}(\{y\}), d_y)$ and $(\pi^{-1}(\{y'\}), d_{y'})$ are isometric.

Example 2.6.17. The skew shift is an isometric extension of the circle shift, where we give each fibre the standard metric.

Example 2.6.18. A topological dynamical system is an isometric extension of a point if and only if it is isometric.

Exercise 2.6.7. If X is minimal, show that properties (a), (b) in Definition 2.6.16 automatically imply property (c). Furthermore, in this case show that the isometry group $\mathrm{Isom}(\pi^{-1}(\{y\}))$ of any fibre acts transitively on that fibre. Show however that property (c) can fail even when properties (a) and (b) hold if X is not assumed to be minimal.

Exercise 2.6.8 (Topological characterisation of isometric extensions)**.** Let $(X, \mathcal{F}, T)$ be a extension of a minimal topological dynamical system $(Y, \mathcal{G}, S)$ with factor map $\pi : X \to Y$, and let d be a metric on X. Show that X is an isometric extension if and only if the shift maps T^n are uniformly equicontinuous relative to π in the sense that for every $\varepsilon > 0$ there exists $\delta > 0$ such that every $x, y \in X$ with $\pi(x) = \pi(y)$ and $d(x, y) < \delta$, we have $d(T^n x, T^n y) < \varepsilon$ for all n.

An important subclass of isometric extensions are the *group extensions*. Recall that an *automorphism* of a topological dynamical system is an isomorphism of that system to itself, i.e. a homeomorphism that commutes with the shift.

Definition 2.6.19 (Group extensions)**.** Let $(X, \mathcal{F}, T)$ be a topological dynamical system. Suppose that we have a compact group G of automorphisms of X (where we endow G with the uniform topology). Then the quotient space $Y := G\backslash X = \{Gx : x \in X\}$ is also a compact metrisable space, and one easily sees that the projection map $\pi : X \mapsto Y$ is a factor map. We refer to X as a *group extension* of Y (or of any other system isomorphic to Y). We refer to G as the *structure group* of the extension. We say that the group extension is an *abelian group extension* if G is abelian.

Example 2.6.20 (Cocycle extensions)**.** If G is a compact topological metrisable group, $(Y, \mathcal{G}, S)$ is a topological dynamical system, and a continuous map $\sigma : Y \to G$, then we define the *cocycle extension* $X = Y \times_\sigma G$ to be the product space $Y \times G$ with the shift $T : (y, \zeta) \mapsto (Sy, \sigma(y)\zeta)$, and with the factor map $\pi : (y, \zeta) \mapsto y$. One easily verifies that X is a group extension of Y with structure group G. The converse is not quite true for topological reasons; not every G-bundle can be globally trivialised, although one can still describe general group extensions by patching together cocycle extensions on local trivalisations.

Example 2.6.21. The skew shift is a cocycle extension (and hence group extension) $Y \times_\sigma (\mathbf{R}/\mathbf{Z})$ of the circle shift Y, with $\sigma(y) := y$ being the identity map. Any Kronecker system is an abelian group extension of a point.

Exercise 2.6.9. Show that every group extension is an isometric extension. *Hint*: the group G acts equicontinuously on itself, and thus isometrically on itself by choosing the right metric, as in Exercise 2.6.1.

Exercise 2.6.10. Let $(Y, \mathcal{G}, S)$ be a topological dynamical system, and G a compact topological metrisable group. We say that two cocycles $\sigma, \sigma' : Y \to G$ are *cohomologous* if we have $\sigma'(y) = \phi(Sy)\sigma(y)\phi(y)^{-1}$ for some continuous map $\phi : Y \to G$. Show that if σ, σ' are cohomologous, then the cocycle extensions $Y \times_\sigma G$ and $Y \times_{\sigma'} G$ are isomorphic. Understanding exactly which cocycles are cohomologous to each other is a major topic of study in dynamical systems (though not one which we will pursue here).

In view of Proposition 2.6.7 and Exercise 2.6.15, it is reasonable to ask whether every minimal isometric extension is a group extension. The answer is no (though actually constructing a counterexample is a little tricky). The reason is that we can form intermediate systems between a system $Y = G\backslash X$ and a group extension X of that system by quotienting out a subgroup. Indeed, if H is a closed subgroup of the structure group G, then $H\backslash X$ is a factor of X and an isometric extension of $G\backslash X$, but need not be a group extension of $G\backslash X$

(basically because G/H need not be a group). But this is the only obstruction to obtaining an analogue of Proposition 2.6.7:

Lemma 2.6.22. *Suppose that X is an isometric extension of another topological dynamical system Y with projection map $\pi : X \to Y$. Suppose also that X is minimal. Then there exists a group extension Z of Y with structure group G (thus $Y \equiv G\backslash Z$) and a closed subgroup H of G such that X is isomorphic to $H\backslash Z$, and π is (after applying the isomorphisms) the projection map from $H\backslash Z$ to $G\backslash Z$; thus we have the commutative diagram*

$$\begin{array}{ccc} Z & \to & X = H\backslash Z \\ & \searrow & \downarrow \\ & & Y = G\backslash Z \end{array} . \tag{2.48}$$

Proof. For each $y \in Y$, let V_y be the metric space $\pi^{-1}(\{y\})$ with the metric d_y given by Definition 2.6.19. Thus for any integer n and any $y \in Y$, T^n is an isometry from V_y to $V_{S^n y}$; taking limits, we see for any $p \in \beta\mathbf{Z}$ that T^p is an isometry from V_y to $V_{S^p y}$. Also, the T^p clearly commute with the shift T.

Fix a point $y_0 \in Y$, and set $G := \mathrm{Isom}(V_{y_0})$.

Let W be the space of all pairs (y, f) where $y \in Y$ and f is an isometry from V_{y_0} to V_y. This is a compact metrisable space with a shift $U : (y, f) \mapsto (Sy, T \circ f)$ and an action $g : (y, f) \mapsto (y, f \circ g^{-1})$ of G that commutes with U. We let Z be the orbit closure in W of the G-orbit $\{y_0\} \times G$ under the shift U. If we fix a point $x_0 \in V_{y_0}$, then Z projects onto X by the map $f \mapsto f(y_0)$, and onto Y by the map $(y, f) \mapsto y$; these maps of course commute with the projection $\pi : x \mapsto \pi(x)$ from X to Y. Because X is minimal (and thus equal to all of its orbit closures), one sees that all of these projections are surjective morphisms, thus Z extends both Y and X. Also, one verifies that Z is a group extension over Y with structure group G, and a group extension over X with structure group given by the stabiliser $H := \{g \in G : gx_0 = x_0\}$. The claim follows. □

Exercise 2.6.11. Show that if an minimal extension $\pi : X \to Y$ is finite, then it is automatically an abelian group extension. *Hint*: recall from Section 2.2 that minimal finite systems are equivalent to shifts on a cyclic group.

An important feature of isometric or group extensions is that they tend to preserve recurrence properties of the system. We will see this phenomenon prominently when we turn to the ergodic theory analogue of isometric extensions, but for now let us give a simple illustrative result in this direction:

Proposition 2.6.23. *Let $(X, \mathcal{F}, T)$ be an isometric extension of $(Y, \mathcal{G}, S)$ with factor map $\pi : X \to Y$, and let y be a recurrent point of Y (see Definition 2.3.2 for a definition). Then every point x in the fibre $\pi^{-1}(\{y\})$ is a recurrent point in X.*

Proof. It will be convenient to use ultrafilters. In view of Lemma 2.6.22, it suffices to prove the claim for group extensions (note that recurrence is preserved under morphisms). Since y is recurrent, there exists $p \in \beta\mathbf{Z}\backslash\mathbf{Z}$ such that $S^p y = y$ (see Exercise 2.3.10). Thus $\pi(T^p x) = \pi(x)$. Since $Y = G\backslash X$, this implies that $T^p x = gx$ for some $g \in G$. We can iterate this (recalling that G commutes with T) to conclude that $T^{np} x = g^n x$ for all positive integers n. But by considering the action of g on G, we know (from Theorem 2.3.4) that we have $g^{n_j} h \to h$ for some $h \in G$ and $n_j \to +\infty$; canceling the h, and then applying to x, we conclude that $g^{n_j} x \to x$, and thus $T^{n_j p} x \to x$. If we write $q := \lim_{j \to r} n_j p$ for some $r \in \beta\mathbf{N}\backslash\mathbf{N}$, we conclude that $T^q x = x$ and so x is recurrent as desired. □

2.6.4. Application: distribution of polynomial sequences in torii. Now we apply the above theory to the following specific problem:

Problem 2.6.24. Let $P : \mathbf{Z} \to (\mathbf{R}/\mathbf{Z})^d$ be a polynomial sequence in a d-dimensional torus, thus $P(n) = \sum_{j=0}^k c_j n^j$ for some $c_0, \ldots, c_k \in (\mathbf{R}/\mathbf{Z})^d$. Compute the orbit closure $\overline{P(\mathbf{Z})} = \overline{\{P(n) : n \in \mathbf{Z}\}}$.

(We will be vague here about what "compute" means.)

Example 2.6.25. Is the orbit $\{(\sqrt{2}n \bmod 1, \sqrt{3}n^2 \bmod 1) : n \in \mathbf{Z}\}$ dense in the two-dimensional torus $(\mathbf{R}/\mathbf{Z})^2$?

The answer should of course depend on the polynomial P; for instance if P is constant then the orbit closure is clearly a point. Similarly, if the polynomial P has a constraint of the form $m \cdot P = c$ for

some non-zero $m \in \mathbf{Z}^d$ and $c \in \mathbf{R}/\mathbf{Z}$, then the orbit closure is clearly going to be contained inside the proper subset $\{x \in (\mathbf{R}/\mathbf{Z})^d : m \cdot x = c\}$ of the torus. For instance, $\{(\sqrt{2}n^2 \bmod 1, 2\sqrt{2}n^2 \bmod 1) : n \in \mathbf{Z}\}$ is clearly not dense in the two-dimensional torus, as it is contained in the closed one-dimensional subtorus $\{(x, 2x) : x \in \mathbf{R}/\mathbf{Z}\}$.

In the above example, it is clear that the problem of computing the orbit closure of $(\sqrt{2}n^2 \bmod 1, 2\sqrt{2}n^2 \bmod 1)$ reduces to computing the orbit closure of $(\sqrt{2}n^2 \bmod 1)$. More generally, if a polynomial $P : \mathbf{Z} \to (\mathbf{R}/\mathbf{Z})^d$ obeys a constraint $m \cdot P = c$ for some non-zero *irreducible* $m \in \mathbf{Z}^d$ (i.e. m does not factor as $m = qm'$ for some $q > 1$ and $m' \in \mathbf{Z}^d$, or equivalently that the greatest common divisor of the coefficients of m is 1), then some elementary number theory shows that the set $\{x \in (\mathbf{R}/\mathbf{Z})^d : m \cdot x = c\}$ is isomorphic (after an invertible affine transformation with integer coefficients on the torus) to the standard subtorus $(\mathbf{R}/\mathbf{Z})^{d-1}$).

Exercise 2.6.12. Prove the above claim. *Hint*: the Euclidean algorithm may come in handy.

Because of this, we see that whenever we have a constraint of the form $m \cdot P = c$ with m irreducible, we can reduce Problem 2.6.24 to an instance of Problem 2.6.24 with one lower dimension. What about if m is not irreducible? A typical example of this would be when[29] $P(n) := (\sqrt{2}n^2, 2\sqrt{2}n^2 + \frac{1}{2}n)$. Here, we have the constraint $(-4, 2) \cdot P(n) = 0$, which constrains P to the union of two one-dimensional torii, rather than a single one-dimensional torus. But we can eliminate this multiplicity by the trick of working with the odd and even components $\{P(2n+1) : n \in \mathbf{Z}\}$ and $\{P(2n) : n \in \mathbf{Z}\}$ respectively. One observes that each component obeys an irreducible constraint, namely $(-2, 1) \cdot P(2n) = 0$ and $(-2, 1) \cdot P(2n+1) = \frac{1}{2}$ respectively, and so by the preceding discussion, the problem of computing the orbit closures for each of these components reduces to that of computing an orbit closure in a torus of one lower dimension.

Exercise 2.6.13. More generally, show that whenever P obeys a constraint $m \cdot P(n) = c$ with m not necessarily irreducible, then there

[29] I'm going to drop the "mod 1" terms to remove clutter.

exists an integer $q \geq 1$ such that the orbits $\{P(qn+r) : n \in \mathbf{Z}\}$ obey a constraint $m' \cdot P(qn+r) = c_r$ with m' irreducible.

From Exercises 2.6.12 and 2.6.13, we see that every time we have a constraint of the form $m \cdot P(n) = c$ for some non-zero m, we can reduce Problem 2.6.24 to one or more copies of Problem 2.6.24 in one lower dimension. So, without loss of generality (and by inducting on dimension) we may assume that no such constraint exists. (We will see this "induction on dimension" type of argument also in Section 2.16, when we study Ratner-type theorems in more detail.)

Now that all the "obvious" restrictions on the orbit have been removed, one might now expect $P(n)$ to be uniformly distributed throughout the torus. Happily, this is indeed the case (at least at the topological level):

Theorem 2.6.26 (Equidistribution theorem)**.** *Let $P : \mathbf{Z} \to (\mathbf{R}/\mathbf{Z})^d$ be a polynomial sequence which does not obey any constraint of the form $m \cdot P(n) = c$ with $m \in \mathbf{Z}^d$ non-zero. Then the orbit $P(\mathbf{Z})$ is dense in $(\mathbf{R}/\mathbf{Z})^d$ (i.e. the orbit closure is the whole torus).*

Remark 2.6.27. The recurrence theorems we have already encountered (e.g. Corollary 2.4.4 or Theorem 2.5.1) do not seem to directly establish this result, instead giving the weaker result that every element in $P(\mathbf{Z})$ is a limit point.

Exercise 2.6.14. Assuming Theorem 2.6.26, show that the answer to Problem 2.6.24 is always "a finite union of subtorii", regardless of what the coefficients of P are.

Theorem 2.6.26 can be proven using Weyl's theory of equidistribution (Theorem 1.4.1), which is based on bounds on exponential sums; but we shall instead use a topological dynamics argument based on some ideas of Furstenberg[**Fu1981**]. Amusingly, this argument will use some *global* topology (specifically, *winding numbers*) and not just *local* (point-set) topology.

To begin proving this theorem, let us first consider the linear one-dimensional case, in which one considers the orbit closure of $\{n\alpha+\beta : n \in \mathbf{Z}\}$ for some $\alpha, \beta \in \mathbf{R}/\mathbf{Z}$. The constant term β only affects this closure by a translation and we can ignore it. One then easily

checks that the orbit closure $\overline{\{n\alpha : n \in \mathbf{Z}\}}$ is a closed subgroup of $\mathbf{R}/\mathbf{Z}$. Fortunately, we have a classification of these objects:

Lemma 2.6.28. *Let H be a closed subgroup of $\mathbf{R}/\mathbf{Z}$. Then either $H = \mathbf{R}/\mathbf{Z}$, or H is a cyclic group of the form $H = \{x \in \mathbf{R}/\mathbf{Z} : Nx = 0\}$ for some $N \geq 1$.*

Proof. If H is not all of $\mathbf{R}/\mathbf{Z}$, then its complement, being a non-empty open set, is the union of disjoint open intervals. Let x be the boundary of one of these intervals, then x lies in the closed set H, Translating the group H by x, we conclude that 0 is also the boundary of one of these intervals. Since $H = -H$, we thus see that 0 is an isolated point in H, If we then let y be the closest non-zero element of H to the origin (the case when $H = \{0\}$ can of course be checked separately), we check (using the Euclidean algorithm) that y generates H, and the claim easily follows. □

Exercise 2.6.15. Using the above lemma, prove Theorem 2.6.26 in the case when $d = 1$ and P is linear.

Exercise 2.6.16. Obtain another proof of Lemma 2.6.28 using Fourier analysis and the fact that the only non-trivial subgroups of $\mathbf{Z}$ (the Pontryagin dual of $\mathbf{R}/\mathbf{Z}$) are the groups $N \cdot \mathbf{Z}$ for $N \geq 1$.

Now we consider the linear case in higher dimensions. The key lemma is

Lemma 2.6.29. *Let H be a closed subgroup of $(\mathbf{R}/\mathbf{Z})^d$ for some $d \geq 1$ such that $\pi(H) = (\mathbf{R}/\mathbf{Z})^{d-1}$, where $\pi : (\mathbf{R}/\mathbf{Z})^d \to (\mathbf{R}/\mathbf{Z})^{d-1}$ is the canonical projection. Then either $H = (\mathbf{R}/\mathbf{Z})^d$ or $H = \{x \in (\mathbf{R}/\mathbf{Z})^d : m \cdot x = 0\}$ for some $m \in \mathbf{Z}^d$ with final coefficient non-zero.*

Proof. The fibre $H \cap \pi^{-1}(\{0\})$ is isomorphic to a closed subgroup of $\mathbf{R}/\mathbf{Z}$, so we can apply Lemma 2.6.28. If this subgroup is full, then it is not hard to see that $H = (\mathbf{R}/\mathbf{Z})^d$, so suppose instead that $H \cap \pi^{-1}(\{0\})$ is isomorphic to the cyclic group of order N. We then apply the homomorphism $f_N : (x_1, \ldots, x_d) \to (x_1, \ldots, x_{d-1}, Nx_d)$, and observe that $H_N := f_N(H)$ is a closed subgroup of $(\mathbf{R}/\mathbf{Z})^d$ whose fibres are a point, i.e. H_N is a graph $\{(x, \phi(x)) : x \in (\mathbf{R}/\mathbf{Z})^{d-1}\}$ for some $\phi : (\mathbf{R}/\mathbf{Z})^{d-1} \to \mathbf{R}/\mathbf{Z}$. Observe that the projection map

$(x, \phi(x)) \mapsto x$ is a continuous bijection from the compact Hausdorff space H_N to the compact Hausdorff space $(\mathbf{R}/\mathbf{Z})^{d-1}$, and is thus a homeomorphism; in particular, ϕ is continuous. Also, since H_N is a group, ϕ must be a homomorphism. It is then a standard exercise to conclude that ϕ is linear, and therefore takes the form $(x_1, \ldots, x_{d-1}) \mapsto m_1 x_1 + \ldots + m_{d-1} x_{d-1}$ for some integers $m_1, \ldots, m_{d-1}$. The claim then follows by some routine algebra. □

Exercise 2.6.17. Using the above lemma, prove Theorem 2.6.26 in the case when d is arbitrary and P is linear.

We now turn to the polynomial case. The basic idea is to reexpress $P(n)$ in terms of the orbit $T^n x$ of some topological dynamical system on a torus. We have already seen this happen with the skew shift $((\mathbf{R}/\mathbf{Z})^2, (x, y) \mapsto (x + \alpha, y + x))$, where the orbits $T^n x$ exhibit quadratic behaviour in n. More generally, an iterated skew shift such as

$$((\mathbf{R}/\mathbf{Z})^d, (x_1, \ldots, x_d) \mapsto (x_1 + \alpha, x_2 + x_1, \ldots, x_d + x_{d-1})) \tag{2.49}$$

generates orbits $T^n x$ whose final coefficient contains degree d terms such as $\frac{n(n-1)\ldots(n-d+1)}{d!}\alpha$. What we would like to do is find criteria under which we could demonstrate that systems such as (2.49) are *minimal*; this would mean that every orbit closure in that system is dense, which would clearly be relevant for proving results such as Theorem 2.6.26.

To do this, we will exploit the fact that systems such as (2.49) can be built as towers of isometric extensions; for instance, the system (2.49) is an isometric extension over the same system (2.49) associated to $d - 1$ (which, in the case $d = 1$, is simply a point). Now, isometric extensions don't always preserve minimality; for instance, if one takes a trivial cocycle extension $Y \times_0 G$ then the system is certainly non-minimal, as every horizontal slice $Y \times \{g\}$ of that system is a subsystem. More generally, any cocycle extension which is cohomologous to the trivial cocycle (see Exercise 2.6.10) will not be minimal. However, it turns out that if one has a topological obstruction to triviality, then minimality is preserved. We will formulate this fact using the machinery of *winding numbers*. Recall that every continuous map $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ has a *winding number* $[f] \in \mathbf{Z}$, which

can be defined as the unique integer such that f is homotopic to the linear map $x \mapsto [f]x$. Note that the map $f \mapsto [f]$ is linear, and also that $[f]$ is unchanged if one continuously deforms f.

We now give a variant of a lemma of Furstenberg [**Fu1981**].

Lemma 2.6.30. *Let $(Y, \mathcal{G}, S)$ be a minimal topological dynamical system. Let $\sigma : Y \to (\mathbf{R}/\mathbf{Z})^d$ be a cocycle such that for every non-zero $m \in \mathbf{Z}^d$ there exists a loop $\gamma : \mathbf{R}/\mathbf{Z} \to Y$ such that $S \circ \gamma$ is homotopic to γ and $[m \cdot \sigma \circ \gamma] \neq 0$. Then $Y \times_\sigma \mathbf{R}/\mathbf{Z}$ is also minimal.*

Proof. We induct on d. The case $d = 0$ is trivial, so suppose $d \geq 1$ and the claim has already been proven for $d - 1$. Suppose for contradiction that $Y \times_\sigma (\mathbf{R}/\mathbf{Z})^d$ contains a proper minimal subsystem Z. Then $\pi(Z)$ is a subsystem of Y, and must therefore equal all of Y, by minimality of Y. Now we use the action of $(\mathbf{R}/\mathbf{Z})^d$ on $Y \times_\sigma (\mathbf{R}/\mathbf{Z})^d$, which commutes with the shift $T : (y, \zeta) \mapsto (Sy, \sigma(y) + \zeta)$. For every $\theta \in (\mathbf{R}/\mathbf{Z})^d$, we see that $\theta + Z$ is also a minimal subsystem, and so is either equal to Z or disjoint from Z. If we let $H := \{\theta \in (\mathbf{R}/\mathbf{Z})^d : \theta + Z = Z\}$, we conclude that H is a closed subgroup of $(\mathbf{R}/\mathbf{Z})^d$.

We now claim that the projection of H to $(\mathbf{R}/\mathbf{Z})^{d-1}$ must be all of $(\mathbf{R}/\mathbf{Z})^{d-1}$. For if this were not the case, we could project Z down to $Y \times_{\sigma'} (\mathbf{R}/\mathbf{Z})^{d-1}$, where $\sigma' : Y \to (\mathbf{R}/\mathbf{Z})^{d-1}$ is the projection of σ, and obtain a proper subsystem of that extension. But by induction hypothesis we see that $Y \times_{\sigma'} (\mathbf{R}/\mathbf{Z})^{d-1}$ is minimal, a contradiction, thus proving the claim.

We can now apply Lemma 2.6.29. If H is all of $(\mathbf{R}/\mathbf{Z})^d$ then Z is all of $Y \times_\sigma (\mathbf{R}/\mathbf{Z})^d$, a contradiction. Thus we have $H = \{\zeta \in (\mathbf{R}/\mathbf{Z})^d : m \cdot \zeta = 0\}$ for some non-zero $m \in \mathbf{Z}^d$, and thus Z must take the form

$$Z = \{(y, \zeta) \in Y \times_\sigma (\mathbf{R}/\mathbf{Z})^d : m \cdot \zeta = \phi(y)\} \tag{2.50}$$

for some $\phi : Y \to \mathbf{R}/\mathbf{Z}$. Arguing as in the proof of Lemma 2.6.29 we can show that Y is homeomorphic to the image of Z under the map $(y, \zeta) \mapsto (y, m \cdot \zeta)$ and so ϕ must be continuous. Since Z is shift-invariant, we must have the equation

$$\phi(Sy) = \phi(y) + m \cdot \sigma(y). \tag{2.51}$$

We apply this for y in the loop γ associated to m by hypothesis, and take degrees to conclude

$$[\phi \circ S \circ \gamma] = [\phi \circ \gamma] + [m \cdot \sigma \circ \gamma]. \tag{2.52}$$

But as $S \circ \gamma$ is homotopic to γ, we have $[\phi \circ S \circ \gamma] = [\phi \circ \gamma]$ and thus $[m \cdot \sigma \circ \gamma] = 0$, contradicting the hypothesis. □

Exercise 2.6.18. Using the above lemma and an induction on d, show that the system (2.49) is minimal whenever α is irrational. (The key, of course, is to make a good choice for the loop γ that makes all computations easy.)

Exercise 2.6.19. More generally, show that the product of any finite number of systems of the form (2.49) remains minimal, as long as the numbers α that generate each factor system are linearly independent with respect to each other and to 1 over the rationals $\mathbf{Q}$.

It is now possible to deduce Theorem 2.6.26 from Exercise 2.6.19 and a little bit of linear algebra. We sketch the ideas as follows. Firstly we take all the non-constant coefficients that appear in P and look at the space they span, together with 1, over the rationals $\mathbf{Q}$. This is a finite-dimensional space, and so has a basis containing 1 which is linearly independent over $\mathbf{Q}$. The non-constant coefficients of P are rational linear combinations of elements of this basis; by dividing the basis elements by some suitable integer (and using the trick of passing from $P(n)$ to $P(qn+r)$ if necessary) we can ensure that the coefficients of P are in fact integer linear combinations of basis elements. This allows us to write P as an affine-linear combination (with integer coefficients) of the coefficients of an orbit in the type of product system considered in Exercise 2.6.19. If this affine transformation has full rank, then we are done; otherwise, the affine transformation maps to some subspace of the torus of the form $\{x : m \cdot x = c\}$, contradicting the hypothesis on P. Theorem 2.6.26 follows.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/01/24`. Thanks to Nilay, mmailliw/william, Zaher Hani, Sugata, and Liu Xiao Chuan for corrections.

2.7. Structural theory of topological dynamical systems

In our final lecture on topological dynamics, we discuss a remarkable theorem of Furstenberg[**Fu1963**] that classifies a major type of topological dynamical system - *distal* systems - in terms of highly structured (from an algebraic point of view) systems, namely towers of isometric extensions. This theorem is also a model for an important analogous result in ergodic theory, the *Furstenberg-Zimmer structure theorem*, which we will turn to in a few lectures. We will not be able to prove Furstenberg's structure theorem for distal systems here in full, but we hope to illustrate some of the key points and ideas.

2.7.1. Distal systems. Furstenberg's theorem concerns a significant generalisation of the equicontinuous (or isometric) systems, namely the *distal* systems.

Definition 2.7.1 (Distal systems)**.** Let $(X, \mathcal{F}, T)$ be a topological dynamical system, and let d be an arbitrary metric on X (it is not important which one one picks here). We say that two points x, y in X are *proximal* if we have $\liminf_{n\to\infty} d(T^n x, T^n y) = 0$. We say that X is distal if no two distinct points $x \neq y$ in X are proximal, or equivalently if for every distinct x, y there exists $\varepsilon > 0$ such that $d(T^n x, T^n y) \geq \varepsilon$ for all n.

It is obvious that every isometric or equicontinuous system is distal, but the converse is not true, as the following example shows:

Example 2.7.2. If $\alpha \in \mathbf{R}$, then the skew shift $((\mathbf{R}/\mathbf{Z})^2, (x,y) \mapsto (x+\alpha, y+x))$ turns out to be not equicontinuous; indeed, if we start with a pair of nearby points $(0,0), (0,1/2n)$ for some large n and apply T^n, one ends up with $(n\alpha, \frac{n(n-1)}{2}\alpha)$ and $(\alpha, \frac{n(n-1)}{2}\alpha + \frac{1}{2})$, thus demonstrating failure of equicontinuity. On the other hand, the system is still distal: given any pair of distinct points $(x,y), (x',y')$, either $x \neq x'$ (in which case the horizontal separation between $T^n(x,y)$ and $T^n(x',y')$ is bounded from below) or $x = x'$ (in which case the vertical separation is bounded from below).

Exercise 2.7.1. Show that any non-trivial Bernoulli system $\Omega^{\mathbf{Z}}$ is not distal.

Distal systems interact nicely with the action $p \mapsto T^p$ of the compactified integers $\beta\mathbf{Z}$:

Exercise 2.7.2. Let $(X, \mathcal{F}, T)$ be a topological dynamical system.

(1) Show that two points x, y in X are proximal if and only if $T^p x = T^p y$ for some $p \in \beta\mathbf{Z}$.

(2) Show that X is distal if and only if all the maps T^p for $p \in \beta\mathbf{Z}$ are injective.

(3) If X is distal, show that $T^p = \mathrm{id}$ whenever $p \in \beta\mathbf{Z}$ is idempotent. *Hint*: use part 2.

(4) If X is distal, show that the set of transformations $G := \{T^p : p \in \beta\mathbf{Z}\}$ on X forms a group G, known as the *Ellis group* of X. *Hint*: use part 3, together with Lemma 2.5.14.

(5) Show that G is a compact subset of X^X (with the product topology), and that G acts transitively on X if and only if X is minimal.

Exercise 2.7.3. Show that an inverse limit of a totally ordered set $(Y_\alpha)_{\alpha \in A}$ of distal factors is still distal. (This turns out to be slightly easier than Lemma 2.6.12.)

Exercise 2.7.4. Show that every topological dynamical system has a maximal distal factor. *Hint*: repeat the proof of Corollary 2.6.13.

Exercise 2.7.5. Show that any distal system can be partitioned into disjoint minimal distal systems. *Hint*: One can of course adapt the proof of Proposition 2.6.9; but there is a slicker way to do it by exploiting the Ellis group.

Note that the skew shift system, while not isometric, does have a non-trivial isometric factor, namely the circle shift $(\mathbf{R}/\mathbf{Z}, x \mapsto x + \alpha)$ with the projection map $\pi : (x, y) \mapsto x$. It turns out that this phenomenon is general:

Theorem 2.7.3 (Baby Furstenberg structure theorem)**.** *Let $(X, \mathcal{F}, T)$ be minimal, distal and non-trivial (i.e. not a point). Then X has a non-trivial isometric factor $\pi : X \to Y$.*

This result - a toy case of Furstenberg's full structure theorem - is already rather difficult to establish. We will not give Furstenberg's

original proof here (though see Exercise 2.7.13 below), but will at least sketch how the factor $\pi : X \to Y$ is constructed. A key object in the construction is the symmetric function $F : X \times X \to \mathbf{R}^+$ defined by the formula

$$F(x,y) := \inf_{n \in \mathbf{Z}} d(T^n x, T^n y). \tag{2.53}$$

Example 2.7.4. We again consider the skew shift $((\mathbf{R}/\mathbf{Z})^2, (x,y) \mapsto (x+\alpha, y+x))$ with α irrational. For sake of concreteness let us choose the taxicab metric $d((x,y),(x',y')) := \|x-x'\|_{\mathbf{R}/\mathbf{Z}} + \|y-y'\|_{\mathbf{R}/\mathbf{Z}}$, where $\|x\|_{\mathbf{R}/\mathbf{Z}}$ is the distance from x to the integers. Then one can check that $F((x,y),(x',y'))$ is equal to $\|x-x'\|_{\mathbf{R}/\mathbf{Z}}$ when $x-x'$ is irrational, and equal to $\|x-x'\|_{\mathbf{R}/\mathbf{Z}} + \frac{1}{q}\|q(y-y')\|_{\mathbf{R}/\mathbf{Z}}$ when $x-x'$ is rational, where q is the least positive integer such that $q(x-x')$ is an integer. Thus F is highly discontinuous, but it is at least upper semi-continuous in each of its two variables[30].

Exercise 2.7.6. Let G be the Ellis group of a minimal distal system X.

(1) For any $x, y \in X$, show that $F(x,y) = \inf_{g \in G} d(gx, gy)$. In particular, $F(gx, gy) = F(x,y)$ for all $g \in G$.

(2) For any $x, y \in X$, show that the set $\{(gx, gy) : g \in G\}$ is a minimal subsystem of $X \times X$ (with the product shift $(x,y) \mapsto (Tx, Ty)$). Conclude in particular that if $F(x,y) < a$, then the set $\{n \in \mathbf{Z} : d(T^n x, T^n y) < a)\}$ is syndetic.

(3) If $x, y \in X$ and $a > 0$ is such that $F(x,y) < a$, show that there exists ε such that $F(x,z) < a$ whenever $F(y,z) < \varepsilon$.

(4) Let $X_F = (X, \mathcal{F}_F)$ be the space X whose topology is generated by the basic open sets $U_{a,x} := \{y \in X : F(x,y) < a\}$. (That this is a base follows from 3.) Equivalently, X_F is equipped with the weakest topology on which F is upper semi-continuous in each variable. Show that X_F is a weaker topological space than X (i.e. the identity map from X to X_F is continuous); in particular, X_F is compact. Also show that all the maps in G are homeomorphisms on X_F.

[30] Actually, the upper semi-continuity of F holds for arbitrary topological dynamical systems, since F is the infimum of continuous functions.

If the space X_F defined in Exercise 2.7.6 were Hausdorff, then the system $(X_F, \mathcal{F}_F, T)$ would be equicontinuous, by Exercise 2.6.2. Unfortunately, X_F is not Hausdorff in general. However, it turns out that we can "quotient out" the non-Hausdorff nature of X_F. Define the equivalence relation $\sim$ on X_F by declaring $x \sim y$ if we have $F(x,z) = F(y,z)$ for all z outside of a set of the first category in X. This is clearly an equivalence relation, and so we can create the quotient space $Y := X_F/\sim$; since X embeds into X_F we thus have a factor map $\pi : X \to Y$. It is a deep fact (which we will not prove here) that this quotient space is non-trivial and Hausdorff, and that $\sim$ is preserved by the shift T and even by the Ellis group G (thus if $x \sim y$ and $g \in G$ then $gx \sim gy$). Because of this, G continues to act on Y homeomorphically, and so by Exercise 2.6.2, $\pi : X \to Y$ is a non-trivial isometric factor of X as desired.

Exercise 2.7.7. Show that in the case of the skew shift (Example 2.7.4), this construction recovers the factor that was discussed just before Theorem 2.7.3. (The trickiness of this exercise should already give you some idea of the difficulty level of Theorem 2.7.3.)

2.7.2. The Furstenberg structure theorem for distal systems. We have already noted that isometric systems are distal systems. More generally, we have

Exercise 2.7.8. Show that an isometric extension of a distal system is still distal. *Hint*: Example 2.7.2 is a good model case.

Thus, for instance, the iterated skew shifts that appear in (2.49) are distal. Also, recall from Exercise 2.7.7 that the inverse limit of distal systems is again distal. It turns out that these are the *only* ways to generate distal systems, in the following sense:

Theorem 2.7.5 (Furstenberg's structure theorem for distal systems)**.** [**Fu1963**] *Let $(X, \mathcal{F}, T)$ be a distal system. Then there exists an ordinal α and a factor Y_β for every $\beta \leq \alpha$ with the following properties:*

(1) *$Y_\emptyset$ is a point.*

(2) *For every successor ordinal $\beta + 1 \leq \alpha$, $Y_{\beta+1}$ is an isometric extension of Y_β.*

(3) *For every limit ordinal $\beta \leq \alpha$, Y_β is an inverse limit of the Y_γ for $\gamma < \beta$.*

(4) *Y_α is equal to X.*

The collection of factors $(Y_\beta)_{\beta \leq \alpha}$ is sometimes known as a "Furstenberg tower".

Theorem 2.7.5 follows by applying Zorn's lemma with the following key proposition:

Proposition 2.7.6 (Key inductive step)**.** *Let $(X, \mathcal{F}, T)$ be a distal system, and let Y be a proper factor of X (i.e. the factor map is not an isomorphism). Then there exists another factor Z of X which is a proper isometric extension of Y.*

Note that Theorem 2.7.3 is the special case of Proposition 2.7.6 when Y is a point. Indeed, Proposition 2.7.6 is proven in the same way as Theorem 2.7.3, but with several additional technicalities which I will not discuss here; see [**Fu1963**] for details.

Exercise 2.7.9. Deduce Theorem 2.7.5 from Proposition 2.7.6 and Zorn's lemma.

Remark 2.7.7. It is known that in Theorem 2.7.5, one can take the ordinal α to be countable, and conversely that for every countable ordinal α, there exists a system whose smallest Furstenberg tower has height α; see [**BeFo1996**].

Remark 2.7.8. Several generalisations and extensions of Furstenberg's structure theorem are known, but they are somewhat technical to state and will not be detailed here; see [**Gl2000**] for a discussion.

2.7.3. Weak mixing and isometric factors. We have seen that distal systems always contain non-trivial isometric factors. What about more general systems? It turns out that there is in fact a nice dichotomy between systems with non-trivial isometric factors, and those without.

Definition 2.7.9 (Topological transitivity)**.** A topological dynamical system $(X, \mathcal{F}, T)$ is *topologically transitive* if, for every pair U, V of non-empty open sets, there exists an integer n such that $T^n U \cap V \neq \emptyset$.

Exercise 2.7.10. Show that a topological dynamical system is topologically transitive if and only if it is equal to the orbit closure of one of its points[31].

Exercise 2.7.11. Show that any factor of a topologically transitive system is again topologically transitive.

Definition 2.7.10 (Topological weak mixing)**.** A topological dynamical system $(X, \mathcal{F}, T)$ is *topologically weakly mixing* if the product system $X \times X$ is topologically transitive.

Exercise 2.7.12. A system is said to be *topologically mixing* if for every pair U, V of non-empty open sets, one has $T^n U \cap V \neq \emptyset$ for all sufficiently large n. Show that topological mixing implies topological weak mixing. (The converse is false, but actually constructing a counterexample is somewhat tricky.)

Example 2.7.11. No circle shift $(\mathbf{R}/\mathbf{Z}, x \mapsto x + \alpha)$ is topologically weak mixing (or topologically mixing), even though such shifts are minimal (and hence transitive) when α is irrational. On the other hand, any Bernoulli shift is easily seen to be topologically mixing (and hence topologically weak mixing).

We have the following dichotomy, first proven in [**KeRo1969**] (using ideas from [**Fu1963**]):

Theorem 2.7.12 (Dichotomy between structure and randomness)**.** [**KeRo1969**] *Let $(X, \mathcal{F}, T)$ be a minimal topological dynamical system. Then exactly one of the following statements is true:*

(1) *(Structure) X has a non-trivial isometric factor.*

(2) *(Randomness) X is topologically weakly mixing.*

Remark 2.7.13. Combining this with Exercise 2.6.6, we obtain an equivalent formulation of this theorem: a minimal system is topologically weakly mixing if and only if it has no non-trivial eigenfunctions.

[31] Compare this with minimal systems, which is the orbit closure of *any* of its points. Thus minimality is stronger than topological transitivity; for instance, the compactified integers $\{-\infty\} \cup \mathbf{Z} \cup \{+\infty\}$ with the usual shift is topologically transitive but not minimal.

Proof. We first prove the easy direction: that if X has a non-trivial isometric factor, then it is not topologically weakly mixing. In view of Exercise 2.7.11, it suffices to prove this when X itself is isometric. Let x, x' be two distinct points of Y, let r denote the distance between x and x' with respect to the metric that makes X isometric, and let B and B' be the open balls of radius $r/10$ centred at x and x' respectively. As X is isometric, we see for any integer n that $T^n B$ cannot intersect both B and B', or equivalently that $(T \times T)^n (B \times B)$ cannot intersect $B \times B'$. Thus X is not topologically transitive as desired.

Now we prove the difficult direction: if X is not topologically weakly mixing, then it has a non-trivial isometric factor. For this we use an argument from [**BlHoMa2000**], based on the earlier work [**McM1978**]. By Definition 2.7.10, there exist open non-empty sets U, V in $X \times X$ such that $(T \times T)^n U \cap V = \emptyset$ for all n. If we thus set $K := \overline{\bigcup_n (T \times T)^n U}$, we see that K is a compact proper $T \times T$-invariant subset of $X \times X$ with non-empty interior. On the other hand, the projection of K to either factor of $X \times X$ is a non-empty compact invariant subset of X and thus must be all of X.

We need to somehow use K to build an isometric factor of X. For this, we shall move from the topological dynamics setting to that of the ergodic theory setting. By Corollary 2.7.17 in the appendix, X admits an invariant Borel measure μ. The support of μ is a non-empty closed invariant subset of X, and is thus equal to all of X by minimality.

The space $L^1(X, \mu)$ is a metric space, with an isometric shift map $Tf := f \circ T^{-1}$. We define the map $\pi : X \to L^1(X, \mu)$ by the formula

$$\pi(x) : y \mapsto 1_K(x, y) \tag{2.54}$$

for all $x \in X$, where 1_K is the indicator function of K. Because K has non-empty interior and non-empty exterior, and because μ has full support, it is not hard to show that π is non-constant. By the T-invariance of W, it also preserves the shift T. So if we can show that π is continuous, we see that $\pi(X)$ will be a non-trivial isometric factor of X and we will be done.

Let us first consider the scalar function $f(x) := \int_X 1_K(x,y)\, d\mu(y)$. From the dominated convergence theorem and the fact that K is closed, we see that f is upper semi-continuous, and continuous at at least one point, thanks to Lemma 2.4.13. On the other hand, since K is $T \times T$-invariant and μ is T-invariant, we see that f is T-invariant. Applying Exercise 2.4.16 we see that f is constant. On the other hand, as K is closed we have $\limsup_{x \to x_0} 1_K(x,y) \leq 1_K(x_0,y)$ for any $x_0 \in X$, and so by dominated convergence again we see that $1_K(x,\cdot)$ converges in L^1 to zero outside of the support of $1_K(x_0,\cdot)$. Combining this with the constancy of f we conclude that $1_K(x,\cdot)$ converges to $1_K(x_0,\cdot)$ in L^1 on all of X, and thus π is continuous as required. □

Remark 2.7.14. Note how the measure-theoretic structure was used to obtain metric structure, by passing from the measure space (X,μ) to the metric space $L^1(X,\mu)$. This again shows that one can sometimes upgrade weak notions of structure (such as topological or measure-theoretic structure) to strong notions (such as geometric or algebraic structure).

Exercise 2.7.13. Use Theorem 2.7.12 to prove Theorem 2.7.3. *Hint*: use Exercise 2.7.10.

Remark 2.7.15. It would be very convenient if one had a relative version of Theorem 2.7.12, namely that if X is an extension of Y, then X is either relatively topologically weakly mixing with respect to Y (which means that the relative product $X \times_Y X := \{(x,x') \in X \times X : \pi(x) = \pi(x')\}$ is topologically transitive), or else X has a factor Z which is a non-trivial isometric extension of Y; among other things, this would have given a new proof of Theorem 2.7.5, and in fact establish a somewhat stronger structural theorem. Unfortunately, this relative version fails; a counterexample (based on the Morse sequence, Example 2.2.11) can be found in [**Gl2003**, Exercise 1.19.3]. Nevertheless, the analogue of this claim does hold true in the measure-theoretic setting, as we shall see in Section 2.12.

2.7.4. Appendix: sequential compactness of Borel probability measures. We now recall some standard facts from measure theory about Borel probability measures on a compact metrisable space

X. Recall that a sequence of such measures μ_n converges in the *vague topology* to another μ if we have $\int_X f\ d\mu_n \to \int_X f\ d\mu$ for all $f \in C(X)$.

Lemma 2.7.16 (Vague sequential compactness)**.** *The space* $\Pr(X)$ *of Borel probability measures on* X *is sequentially compact in the vague topology.*

Proof. The Riesz representation theorem identifies $\Pr(X)$ with the dual of $C(X)$. From the Stone-Weierstrass theorem we know that $C(X)$ is separable. The claim then follows from the usual Arzelá-Ascoli diagonalisation argument. □

Corollary 2.7.17 (Krylov-Bogolubov theorem)**.** *Let* $(X, \mathcal{F}, T)$ *be a topological dynamical system. Then there exists a* T*-invariant probability measure* μ *on* X*.*

Proof. Pick any point $x_0 \in X$ and consider the finite probability measures

$$\mu_N := \frac{1}{N} \sum_{n=1}^{N} \delta_{T^n x_0} \tag{2.55}$$

where δ_x is the Dirac mass at x. By Lemma 2.7.16, some subsequence μ_{N_j} converges in the vague topology to another Borel probability measure μ. Since we have

$$\int Tf\ d\mu_N = \int f\ d\mu_N + O_f(1/N) \tag{2.56}$$

for all bounded continuous f, we conclude on taking vague limits and using the Riesz representation theorem that μ is T-invariant as required. □

Remark 2.7.18. Note that Corollary 2.7.17, like many other results obtained via compactness methods, guarantees existence of an invariant measure but not uniqueness (this latter property is known as *unique ergodicity*). Even for minimal systems, it is possible for uniqueness to fail, although actually constructing an example is tricky (see for instance [**Fu1961**]). However, as already observed in the proof of Theorem 2.7.12, any invariant measure on a minimal topological dynamical system must be *full* (i.e. its support must be the whole space).

Exercise 2.7.14. Show that any topological dynamical system which is uniquely ergodic is necessarily minimal.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/01/28`.

2.8. The mean ergodic theorem

We now leave topological dynamics, and begin our study of *measure-preserving systems* $(X, \mathcal{X}, \mu, T)$, i.e. a probability space $(X, \mathcal{X}, \mu)$ together with a probability space isomorphism $T : (X, \mathcal{X}, \mu) \to (X, \mathcal{X}, \mu)$ (thus $T : X \to X$ is invertible, with T and T^{-1} both being measurable, and $\mu(T^n E) = \mu(E)$ for all $E \in \mathcal{X}$ and all n). For various technical reasons it is convenient to restrict to the case when the σ-algebra $\mathcal{X}$ is separable, i.e. countably generated. One reason for this is as follows:

Exercise 2.8.1. Let $(X, \mathcal{X}, \mu)$ be a probability space with $\mathcal{X}$ separable. Then the Banach spaces $L^p(X, \mathcal{X}, \mu)$ are separable (i.e. have a countable dense subset) for every $1 \leq p < \infty$; in particular, the Hilbert space $L^2(X, \mathcal{X}, \mu)$ is separable. Show that the claim can fail for $p = \infty$. (We allow the L^p spaces to be either real or complex valued, unless otherwise specified.)

Remark 2.8.1. In practice, the requirement that $\mathcal{X}$ be separable is not particularly onerous. For instance, if one is studying the recurrence properties of a function $f : X \to \mathbf{R}$ on a non-separable measure-preserving system $(X, \mathcal{X}, \mu, T)$, one can restrict $\mathcal{X}$ to the separable sub-σ-algebra $\mathcal{X}'$ generated by the level sets $\{x \in X : T^n f(x) > q\}$ for integer n and rational q, thus passing to a separable measure-preserving system $(X, \mathcal{X}', \mu, T)$ on which f is still measurable. Thus we see that in many cases of interest, we can immediately reduce to the separable case. (In particular, for many of the theorems in this course, the hypothesis of separability can be dropped, though we won't bother to specify for which ones this is the case.)

We are interested in the recurrence properties of sets $E \in \mathcal{X}$ or functions $f \in L^p(X, \mathcal{X}, \mu)$. The simplest such recurrence theorem is

Theorem 2.8.2 (Poincaré recurrence theorem)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $E \in \mathcal{X}$ be a set of positive*

measure. Then $\limsup_{n\to+\infty} \mu(E \cap T^n E) \geq \mu(E)^2$. *In particular,* $E \cap T^n E$ *has positive measure (and is thus non-empty) for infinitely many* n.

Remark 2.8.3. This theorem should be compared with Theorem 2.3.1.

Proof. For any integer $N > 1$, observe that $\int_X \sum_{n=1}^N 1_{T^n E}\, d\mu = N\mu(E)$, and thus by Cauchy-Schwarz

$$\int_X (\sum_{n=1}^N 1_{T^n E})^2\, d\mu \geq N^2 \mu(E)^2. \tag{2.57}$$

The left-hand side of (2.57) can be rearranged as

$$\sum_{n=1}^N \sum_{m=1}^N \mu(T^n E \cap T^m E). \tag{2.58}$$

On the other hand, $\mu(T^n E \cap T^m E) = \mu(E \cap T^{m-n} E)$. From this one easily obtains the asymptotic

$$(2.58) \leq (\limsup_{n\to\infty} \mu(E \cap T^n E) + o(1))N^2, \tag{2.59}$$

where $o(1)$ denotes an expression which goes to zero as N goes to infinity. Combining (2.57), (2.58), (2.59) and taking limits as $N \to +\infty$ we obtain

$$\limsup_{n\to\infty} \mu(E \cap T^n E) \geq \mu(E)^2 \tag{2.60}$$

By shift-invariance we have $\mu(E \cap T^{-n} E) = \mu(E \cap T^n E)$, and the claim follows. □

Remark 2.8.4. In classical physics, the evolution of a physical system in a compact phase space is given by a (continuous-time) measure-preserving system (this is Hamilton's equations of motion combined with Liouville's theorem). The Poincaré recurrence theorem then has the following unintuitive consequence: every collection E of states of positive measure, no matter how small, must eventually return to overlap itself given sufficient time. For instance, if one were to burn a piece of paper in a closed system, then there exist arbitrarily small perturbations of the initial conditions such that, if one waits long enough, the piece of paper will eventually reassemble (modulo

arbitrarily small error)! This seems to contradict the second law of thermodynamics, but the reason for the discrepancy is because the time required for the recurrence theorem to take effect is inversely proportional to the measure of the set E, which in physical situations is exponentially small in the number of degrees of freedom (which is already typically quite large, e.g. of the order of the Avogadro constant). This gives more than enough[32] opportunity for *Maxwell's demon* to come into play to reverse the increase of entropy. The more sophisticated recurrence theorems we will see later have much poorer quantitative bounds still, so much so that they basically have no direct significance for any physical dynamical system with many relevant degrees of freedom.

Exercise 2.8.2. Prove the following generalisation of the Poincaré recurrence theorem: if $(X, \mathcal{X}, \mu, T)$ is a measure-preserving system and $f \in L^1(X, \mathcal{X}, \mu)$ is non-negative, then $\limsup_{n\to+\infty} \int_X fT^n f \geq (\int_X f \, d\mu)^2$.

Exercise 2.8.3. Give examples to show that the quantity $\mu(X)^2$ in the conclusion of Theorem 2.8.2 cannot be replaced by any smaller quantity in general, regardless of the actual value of $\mu(X)$. *Hint*: use a Bernoulli system example.

Exercise 2.8.4. Using the pigeonhole principle instead of the Cauchy-Schwarz inequality (and in particular, the statement that if $\mu(E_1) + \ldots + \mu(E_n) > 1$, then the sets $E_1, \ldots, E_n$ cannot all be disjoint), prove the weaker statement that for any set E of positive measure in a measure-preserving system, the set $E \cap T^n E$ is non-empty for infinitely many n. (This exercise illustrates the general point that the Cauchy-Schwarz inequality can be viewed as a quantitative strengthening of the pigeonhole principle.)

For this section and the next we shall study several variants of the Poincaré recurrence theorem. We begin by looking at the mean ergodic theorem, which studies the limiting behaviour of the ergodic averages $\frac{1}{N}\sum_{n=1}^N T^n f$ in various L^p spaces, and in particular in L^2.

[32] This can be viewed as a manifestation of the *curse of dimensionality*.

2.8.1. Hilbert space formulation. We begin with the Hilbert space formulation of the mean ergodic theorem, due to von Neumann.

Theorem 2.8.5 (Von Neumann ergodic theorem)**.** *Let $U : H \to H$ be a unitary operator on a separable Hilbert space H, Then for every $v \in H$ we have*

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n=0}^{N-1} U^n v = \pi(v), \tag{2.61}$$

where $\pi : H \to H^U$ is the orthogonal projection from H to the closed subspace and let $H^U := \{v \in H : Uv = v\}$ consisting of the U-invariant vectors.

Proof. We give the slick (but not particularly illuminating) proof of von Neumann. It is clear that (2.61) holds if v is already invariant (i.e. $v \in H^U$). Next, let W denote the (possibly non-closed) space $W := \{Uw - w : w \in H\}$. If $Uw - w$ lies in W and v lies in H^U, then by unitarity

$$\langle Uw - w, v\rangle = \langle w, U^{-1}v\rangle - \langle w, v\rangle = \langle w, v\rangle - \langle w, v\rangle = 0 \tag{2.62}$$

and thus W is orthogonal to H^U. In particular $\pi(Uw-w) = 0$. From the telescoping identity

$$\frac{1}{N} \sum_{n=0}^{N-1} U^n(Uw - w) = \frac{1}{N}(U^N w - w) \tag{2.63}$$

we conclude that (2.61) also holds if $v \in W$; by linearity we conclude that (2.61) holds for all v in $H^U + \overline{W}$. A standard limiting argument (using the fact that the linear transformations $v \mapsto \pi(v)$ and $v \mapsto \frac{1}{N}\sum_{n=0}^{N-1} U^n v$ are bounded on H, uniformly in n) then shows that (2.61) holds for v in the closure $\overline{H^U + W}$.

To conclude, it suffices to show that the closed space $\overline{H^U + W}$ is all of H, Suppose for contradiction that this is not the case. Then there exists a non-zero vector w which is orthogonal to all of $\overline{H^U + W}$. In particular, w is orthogonal to $Uw - w$. Applying the easily verified identity $\|Uw-w\|^2 = -2\mathrm{Re}\langle Uw-w, w\rangle$ (related to the parallelogram law) we conclude that $Uw = w$, thus w lies in H^U. This implies that w is orthogonal to itself and is thus zero, a contradiction. □

On a measure-preserving system $(X, \mathcal{X}, \mu, T)$, the shift map $f \mapsto Tf$ is a unitary transformation on the separable Hilbert space $L^2(X, \mathcal{X}, \mu)$. We conclude

Corollary 2.8.6 (Mean ergodic theorem)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $f \in L^2(X, \mathcal{X}, \mu)$. Then we have $\frac{1}{N}\sum_{n=1}^{N} T^n f$ converges in $L^2(X, \mathcal{X}, \mu)$ norm to $\pi(f)$, where $\pi(f) : L^2(X, \mathcal{X}, \mu) \to L^2(X, \mathcal{X}, \mu)^T$ is the orthogonal projection to the space $\{f \in L^2(X, \mathcal{X}, \mu) : Tf = f\}$ consists of the shift-invariant functions in $L^2(X, \mathcal{X}, \mu)$.*

Example 2.8.7 (Finite case)**.** Suppose that $(X, \mathcal{X}, \mu, T)$ is a finite measure-preserving system, with $\mathcal{X}$ discrete and μ the uniform probability measure. Then T is a permutation on X and thus decomposes as the direct sum of disjoint cycles (possibly including trivial cycles of length 1). Then the shift-invariant functions are precisely those functions which are constant on each of these cycles, and the map $f \mapsto \pi(f)$ replaces a function $f : X \to \mathbf{C}$ with its average value on each of these cycles. It is then an instructive exercise to verify the mean ergodic theorem by hand in this case.

Exercise 2.8.5. With the notation and assumptions of Corollary 2.8.6, show that the limit $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \int_X T^n f \overline{f}\ d\mu$ exists, is real, and is greater than or equal to $|\int_X f|^2$. *Hint*: the constant function 1 lies in $L^2(X, \mathcal{X}, \mu)^T$.) Note that this is stronger than the conclusion of Exercise 2.8.2.

Let us now give some other proofs of the von Neumann ergodic theorem. We first give a proof using the spectral theorem for unitary operators. This theorem asserts (among other things) that a unitary operator $U : H \to H$ can be expressed in the form $U = \int_{S^1} \lambda\ d\mu(\lambda)$, where $S^1 := \{z \in \mathbf{C} : |z| = 1\}$ is the unit circle and μ is a projection-valued Borel measure on the circle. More generally, we have

$$U^n = \int_{S^1} \lambda^n\ d\mu(\lambda) \tag{2.64}$$

and so for any vector v in H and any positive integer N we have

$$\frac{1}{N}\sum_{n=0}^{N-1} U^n v = \int_{S^1} \frac{1}{N}\sum_{n=0}^{N-1} \lambda^n\ d\mu(\lambda) v. \tag{2.65}$$

We separate off the $\lambda = 1$ portion of this integral. For $\lambda \neq 1$, we have the geometric series formula

$$\frac{1}{N}\sum_{n=0}^{N-1} \lambda^n = \frac{1}{N}\frac{\lambda^N - 1}{\lambda - 1} \tag{2.66}$$

(compare with (2.63)), thus we can rewrite (2.65) as

$$\mu(\{1\})v + \int_{S^1\setminus\{1\}} \frac{1}{N}\frac{\lambda^N - 1}{\lambda - 1}\, d\mu(\lambda)v. \tag{2.67}$$

Now observe (using (2.66)) that $\frac{1}{N}\frac{\lambda^N-\lambda}{\lambda-1}$ is bounded in magnitude by 1 and converges to zero as $N \to \infty$ for any fixed $\lambda \neq 1$. Applying the dominated convergence theorem (which requires a little bit of justification in this vector-valued case), we see that the second term in (2.67) goes to zero as $N \to \infty$. So we see that (2.65) converges to $\mu(\{1\})v$. But $\mu(\{1\})$ is just the orthogonal projection to the eigenspace of U with eigenvalue 1, i.e. the space H^U, thus recovering the von Neumann ergodic theorem[33].

Remark 2.8.8. The above argument in fact shows that the rate of convergence in the von Neumann ergodic theorem is controlled by the spectral gap of U - i.e. how well-separated the trivial component $\{1\}$ of the spectrum is from the rest of the spectrum. This is one of the reasons why results on spectral gaps of various operators are highly prized.

We now give another proof of Theorem 2.8.5, based on the *energy decrement method*; this proof is significantly lengthier, but is particularly well suited for conversion to finitary quantitative settings. For any positive integer N, define the averaging operators $A_N := \frac{1}{N}\sum_{n=0}^{N-1} U^n$; by the triangle inequality we see that $\|A_N v\| \leq \|v\|$ for all v. Now we observe

Lemma 2.8.9 (Lack of uniformity implies energy decrement)**.** *Suppose $\|A_N v\| \geq \varepsilon$. Then $\|v - A_N^* A_N v\|^2 \leq \|v\|^2 - \varepsilon^2$.*

[33]It is instructive to use spectral theory to interpret von Neumann's proof of this theorem and see how it relates to the argument just given.

Proof. This follows from the identity

$$\|v - A_N^* A_N v\|^2 = \|v\|^2 - 2\|A_N v\|^2 + \|A_N^* A_N v\|^2 \tag{2.68}$$

and the fact that A_N^* has operator norm at most 1. □

We now iterate this to obtain

Proposition 2.8.10 (Koopman-von Neumann type theorem)**.** *Let v be a unit vector, let $\varepsilon > 0$, and let $1 < N_1 < N_2 < \ldots < N_J$ be a sequence of integers with $J > 1/\varepsilon^2 + 2$. Then there exists $1 \leq j < J$ and a decomposition $v = s + r$ where $\|Us - s\| = O(J\frac{1}{N_{j+1}})$ and $\|A_N r\| \leq \varepsilon$ for all $N \geq N_j$.*

Remark 2.8.11. The letters s, r stand for "structured" and "random" (or "residual") respectively. For more on decompositions into structured and random components, see [**Ta2007b**].

Proof. We perform the following algorithm:

(1) Initialise $j := J - 1$, $s := 0$, and $r := v$.

(2) If $\|A_N r\| \leq \varepsilon$ for all $N \geq N_j$ then `STOP`. If instead $\|A_N r\| > \varepsilon$ for some $N \geq N_j$, observe from Lemma 2.8.9 that $\|r - A_N^* A_N r\|^2 \leq \|r\|^2 - \varepsilon^2$.

(3) Replace r with $r - A_N^* A_N r$, replace s with $s + A_N^* A_N r$, and replace j with $j - 1$. Then return to Step 2.

Observe that this procedure must terminate in at most $1/\varepsilon^2$ steps (since the energy $\|r\|^2$ starts at 1, drops by at least ε^2 at each stage, and cannot go below zero). In particular, j stays positive. Observe also that r always has norm at most 1, and thus $\|(U - I)A_N^* A_N r\| = O(1/N)$ at any given stage of the algorithm. From this and the triangle inequality one easily verifies the required claims. □

Corollary 2.8.12 (Partial von Neumann ergodic theorem)**.** *For any vector v, the averages $A_N v$ form a Cauchy sequence in H,*

Proof. Without loss of generality we can take v to be a unit vector. Suppose for contradiction that $A_N v$ was not Cauchy. Then one could find $\varepsilon > 0$ and $1 < N_1 < M_1 < N_2 < M_2 < \ldots$ such that $\|A_{N_j} v - A_{M_j} v\| \geq 5\varepsilon$ (say) for all j. By sparsifying the sequence if necessary

we can assume that N_{j+1} is large compared to N_j, M_j and ε. Now we apply Proposition 2.8.10 to find $j = O_\varepsilon(1)$ and a decomposition $v = s + r$ such that $\|Us - s\| = O_\varepsilon(1/N_{j+1})$ and $\|A_{N_j} r\|, \|A_{M_j} r\| \leq \varepsilon$. If N_{j+1} is large enough depending on N_j, M_j, ε, we thus have $\|A_{N_j} s - s\|, \|A_{M_j} s - s\| \leq \varepsilon$, and thus by the triangle inequality, $\|A_{N_j} v - A_{M_j} v\| \leq 4\varepsilon$, a contradiction. □

Remark 2.8.13. This result looks weaker than Theorem 2.8.5, but the argument is much more robust; for instance, one can modify it to establish convergence of multiple averages such as $\frac{1}{N}\sum_{n=1}^{N} T_1^n f_1 T_2^n f_2 T_3^n f_3$ in L^p norms for commuting shifts T_1, T_2, T_3; see [**Ta2008**]. Further quantitative analysis of the mean ergodic theorem can be found in [**AvGeTo2008**].

Corollary 2.8.12 can be used to recover Theorem 2.8.5 in its full strength, by combining it with a weak form of Theorem 2.8.5:

Proposition 2.8.14 (Weak von Neumann ergodic theorem)**.** *The conclusion* (2.61) *of Theorem 2.8.5 holds in the weak topology.*

Proof. The averages $A_N v$ lie in a bounded subset of the separable Hilbert space H, and are thus sequentially precompact in the weak topology by the Banach-Alaoglu theorem. Thus, if (2.61) fails, then there exists a subsequence $A_{N_j} v$ which converges in the weak topology to some limit w other than $\pi(v)$. By telescoping series we see that $\|UA_{N_j} v - A_{N_j} v\| \leq 2\|v\|/N_j$, and so on taking limits we see that $\|Uw - w\| = 0$, i.e. $w \in H^U$. On the other hand, if y is any vector in H^U, then $A_{N_j}^* y = y$, and thus on taking inner products with v we obtain $\langle y, A_{N_j} v\rangle = \langle y, v\rangle$. Taking limits we obtain $\langle y, w\rangle = \langle y, v\rangle$, i.e. $v - w$ is orthogonal to H^U. These facts imply that $w = \pi(v)$, giving the desired contradiction. □

2.8.2. Conditional expectation. We now turn away from the abstract Hilbert approach to the ergodic theorem (which is excellent for proving the mean ergodic theorem, but not flexible enough to handle more general ergodic theorems) and turn to a more measure-theoretic dynamics approach, based on manipulating the four components $X, \mathcal{X}, \mu, T$ of the underlying system separately, rather than working with the single object $L^2(X, \mathcal{X}, \mu)$ (with the unitary shift

T). In particular it is useful to replace the σ-algebra $\mathcal{X}$ by a sub-σ-algebra $\mathcal{X}' \subset \mathcal{X}$, thus reducing the number of measurable functions. This creates an isometric embedding of Hilbert spaces

$$L^2(X, \mathcal{X}', \mu) \subset L^2(X, \mathcal{X}, \mu) \tag{2.69}$$

and so the former space is a closed subspace of the latter. In particular, we have an orthogonal projection $\mathbf{E}(\cdot|\mathcal{X}') : L^2(X, \mathcal{X}, \mu) \to L^2(X, \mathcal{X}', \mu)$, which can be viewed as the adjoint of the inclusion (2.69). In other words, for any $f \in L^2(X, \mathcal{X}, \mu)$, $\mathbf{E}(f|\mathcal{X}')$ is the unique[34] element of $L^2(X, \mathcal{X}', \mu)$ such that

$$\int_X \mathbf{E}(f|\mathcal{X}')\overline{g}\, d\mu = \int_X f\overline{g}\, d\mu \tag{2.70}$$

for all $g \in L^2(X, \mathcal{X}', \mu)$.

Example 2.8.15 (Finite case)**.** Let X be a finite set, thus $\mathcal{X}$ can be viewed as a partition of X, and $\mathcal{X}' \subset \mathcal{X}$ is a coarser partition of X. To avoid degeneracies, assume that every point in X has positive measure with respect to μ. Then an element f of $L^2(X, \mathcal{X}, \mu)$ is just a function $f : X \to \mathbf{C}$ which is constant on each atom of $\mathcal{X}$. Similarly for $L^2(X, \mathcal{X}', \mu)$. The conditional expectation $\mathbf{E}(f|\mathcal{X}')$ is then the function whose value on each atom A of $\mathcal{X}'$ is equal to the average value $\frac{1}{\mu(A)} \int_A f\, d\mu$ on that atom. (What needs to be changed here if some points have zero measure?)

We leave the following standard properties of conditional expectation as an exercise.

Exercise 2.8.6. Let $(X, \mathcal{X}, \mu)$ be a probability space, and let $\mathcal{X}'$ be a sub-σ-algebra. Let $f \in L^2(X, \mathcal{X}, \mu)$.

(1) The operator $f \mapsto \mathbf{E}(f|\mathcal{X}')$ is a bounded self-adjoint projection on $L^2(X, \mathcal{X}, \mu)$. It maps real functions to real functions, it preserves constant functions (and more generally preserves $\mathcal{X}'$-valued functions), and commutes with complex conjugation.

[34] A reminder: when dealing with L^p spaces, we identify any two functions which agree μ-almost everywhere. Thus, technically speaking, elements of L^p spaces are not actually functions, but rather equivalence classes of functions.

(2) If f is non-negative, then $\mathbf{E}(f|\mathcal{X}')$ is non-negative (up to sets of measure zero, of course). More generally, we have a *comparison principle*: if f, g are real-valued and $f \leq g$ pointwise a. e., then $\mathbf{E}(f|\mathcal{X}') \leq \mathbf{E}(g|\mathcal{X}')$ a.e. Similarly, we have the *triangle inequality* $|\mathbf{E}(f|\mathcal{X}')| \leq \mathbf{E}(|f||\mathcal{X}')$ a.e..

(3) (Module property) If $g \in L^\infty(X, \mathcal{X}', \mu)$, then $\mathbf{E}(fg|\mathcal{X}') = \mathbf{E}(f|\mathcal{X}')g$ a.e..

(4) (Contraction) If $f \in L^2(X, \mathcal{X}, \mu) \cap L^p(X, \mathcal{X}, \mu)$ for some $1 \leq p \leq \infty$, then $\|\mathbf{E}(f|\mathcal{X}')\|_{L^p} \leq \|f\|_{L^p}$. *Hint*: do the $p = 1$ and $p = \infty$ cases first. (This implies in particular that conditional expectation has a unique continuous extension to $L^p(X, \mathcal{X}, \mu)$ for $1 \leq p \leq \infty$; the $p = \infty$ case is exceptional, but note that L^∞ is contained in L^2 since μ is finite.)

For applications to ergodic theory, we will only be interested in taking conditional expectations with respect to a *shift-invariant* sub-σ-algebra $\mathcal{X}'$, thus T and T^{-1} preserve $\mathcal{X}'$. In that case T preserves $L^2(X, \mathcal{X}', \mu)$, and thus T commutes with conditional expectation, or in other words that

$$\mathbf{E}(T^n f|\mathcal{X}') = T^n \mathbf{E}(f|\mathcal{X}') \tag{2.71}$$

a.e. for all $f \in L^2(X, \mathcal{X}, \mu)$ and all n.

Now we connect conditional expectation to the mean ergodic theorem. Let $\mathcal{X}^T := \{E \in \mathcal{X} : TE = E \text{ a.e.}\}$ be the set of essentially shift-invariant sets. One easily verifies that this is a shift-invariant sub-σ-algebra of $\mathcal{X}$.

Exercise 2.8.7. Show that if E lies in $\mathcal{X}^T$, then there exists a set $F \in \mathcal{X}$ which is genuinely invariant ($TF = F$) and which differs from E only by a set of measure zero. Thus it does not matter whether we deal with shift-invariance or essential shift-invariance here. (More generally, it will not make any significant difference if we modify any of the sets in our σ-algebras by null sets.)

The relevance of this algebra to the mean ergodic theorem arises from the following identity:

Exercise 2.8.8. Show that $L^2(X, \mathcal{X}, \mu)^T = L^2(X, \mathcal{X}^T, \mu)$.

As a corollary of this and Corollary 2.8.6, we have

Corollary 2.8.16 (Mean ergodic theorem, again)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system. Then for any $f \in L^2(X, \mathcal{X}, \mu)$, the averages $\frac{1}{N} \sum_{n=0}^{N-1} T^n f$ converge in L^2 norm to $\mathbf{E}(f|\mathcal{X}^T)$.*

Exercise 2.8.9. Show that Corollary 2.8.12 continues to hold if L^2 is replaced throughout by L^p for any $1 \leq p < \infty$. *Hint*: for the case $p < 2$, use that L^2 is dense in L^p. For the case $p > 2$, use that L^∞ is dense in L^p. What happens when $p = \infty$?

Let us now give another proof of Corollary 2.8.16 (leading to a fourth proof of the mean ergodic theorem). The key here will be the decomposition[35] $f = f_{U^\perp} + f_U$, where $f_{U^\perp} := \mathbf{E}(f|\mathcal{X}^T)$ is the "structured" part of f (at least as far as the mean ergodic theorem is concerned) and $f_U := f - f_{U^\perp}$ is the "random" part. As $f_{U^\perp}$ is shift-invariant, we clearly have

$$\frac{1}{N} \sum_{n=0}^{N-1} T^n f_{U^\perp} = f_{U^\perp} \tag{2.72}$$

so it suffices to show that

$$\| \frac{1}{N} \sum_{n=0}^{N-1} T^n f_U \|_{L^2}^2 \to 0 \tag{2.73}$$

as $N \to \infty$. But we can expand out the left-hand side (using the unitarity of T) as

$$\langle F_N, f_U \rangle := \int_X F_N \overline{f_U}\, d\mu \tag{2.74}$$

where F_N is the *dual function* of f_U, defined as

$$F_N := \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} T^{n-m} f_U. \tag{2.75}$$

Now, from the triangle inequality we know that the sequence of dual functions F_N is uniformly bounded in L^2 norm, and so by Cauchy-Schwarz we know that the inner products $\langle F_N, f_U \rangle$ are bounded.

[35] The subscripts $U^\perp, U$ stand for "anti-uniform" and "uniform" respectively; this notation is not standard.

If they converge to zero, we are done; otherwise, by the Bolzano-Weierstrass theorem, we have $\langle F_{N_j}, f_U \rangle \to c$ for some subsequence N_j and some non-zero c. (One could also use ultrafilters instead of subsequences here if desired, it makes little difference to the argument.) By the *Banach-Alaoglu theorem* (or more precisely, the sequential version of this in the separable case), there is a further subsequence $F_{N'_j}$ which converges *weakly* (or equivalently in this Hilbert space case, in the *weak-** sense) to some limit $F_\infty \in L^2(X, \mathcal{X}, \mu)$. Since c is non-zero, F_∞ must also be non-zero. On the other hand, from telescoping series one easily computes that $\|TF_N - F_N\|_{L^2}$ decays like O(1/N) as $N \to \infty$, so on taking limits we have $TF_\infty - F_\infty = 0$. In other words, F_∞ lies in $L^2(X, \mathcal{X}^T, \mu)$.

On the other hand, by construction of f_U we have $\mathbf{E}(f_U|\mathcal{X}^T) = 0$. From (2.71) and linearity we conclude that $\mathbf{E}(F_N|\mathcal{X}^T) = 0$ for all N, so on taking limits we have $\mathbf{E}(F_\infty|\mathcal{X}^T) = 0$. But since F_∞ is already in $L^2(X, \mathcal{X}^T, \mu)$, we conclude $F_\infty = 0$, a contradiction.

Remark 2.8.17. The above argument is lengthier than some of the other proofs of the mean ergodic theorem, but it turns out to be fairly robust; it demonstrates (using the compactness properties of certain "dual functions") that a function f_U with sufficiently strong "mixing" properties (in this case, we require that $\mathbf{E}(f_U|\mathcal{X}^T) = 0$) will cancel itself out when taking suitable ergodic averages, thus reducing the study of averages of f to the study of averages of $f_U = \mathbf{E}(f|\mathcal{X}^T)$. In the modern jargon, this means that $\mathcal{X}^T$ is (the σ-algebra induced by) a characteristic factor of the ergodic average $f \mapsto \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N T^n f$. We will see further examples of characteristic factors for other averages later in this course.

Exercise 2.8.10. Let $(\Gamma, \cdot)$ be a countably infinite discrete group. A *Følner sequence* is a sequence of increasing finite non-empty sets F_n in Γ with $\bigcup_n F_n = \Gamma$ with the property that for any given finite set $S \subset \Gamma$, we have $|(F_n \cdot S)\Delta F_n|/|F_n| \to 0$ as $n \to \infty$, where $F_n \cdot S := \{fs : f \in F_n, s \in S\}$ is the product set of F_n and S, $|F_n|$ denotes the cardinality of F_n, and Δ denotes *symmetric difference*. (For instance, in the case $\Gamma = \mathbf{Z}$, the sequence $F_n := \{-n, \ldots, n\}$ is a Følner sequence.) If Γ acts (on the left) in a measure-preserving manner on a probability space $(X, \mathcal{X}, \mu)$, and $f \in L^2(X, \mathcal{X}, \mu)$, show

that $\frac{1}{|F_n|}\sum_{\gamma\in F_n} f\circ\gamma^{-1}$ converges in L^2 to $\mathbf{E}(f|\mathcal{X}^\Gamma)$, where $\mathcal{X}^\Gamma$ is the collection of all measurable sets which are Γ-invariant modulo null sets, and $f\circ\gamma^{-1}$ is the function $x\mapsto f(\gamma^{-1}x)$.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/01/30`. Thanks to Lior Silberman, Pedro Lauridsen Ribeiro, Orr, mmailliw/william, Sugata, and Liu Xiao Chuan for corrections.

2.9. Ergodicity

We continue our study of basic ergodic theorems, establishing the maximal and pointwise ergodic theorems of Birkhoff. Using these theorems, we can then give several equivalent notions of the fundamental concept of *ergodicity*, which (roughly speaking) plays the role in measure-preserving dynamics that minimality plays in topological dynamics. A general measure-preserving system is not necessarily ergodic, but we shall introduce the *ergodic decomposition*, which allows one to express any non-ergodic measure as an average of ergodic measures (generalising the decomposition of a permutation into disjoint cycles).

2.9.1. The maximal ergodic theorem. Just as we derived the mean ergodic theorem from the more abstract von Neumann ergodic theorem in Section 2.8, we shall derive the maximal ergodic theorem from the following abstract maximal inequality.

Theorem 2.9.1 (Dunford-Schwartz maximal inequality)**.** *Let $(X,\mathcal{X},\mu)$ be a probability space, and let $P: L^1(X,\mathcal{X},\mu)\to L^1(X,\mathcal{X},\mu)$ be a linear operator with $P1=1$ and $P^*1=1$ (i.e. $\int_X Pf\,d\mu=\int_X f\,d\mu$ for all $f\in L^1(X,\mathcal{X},\mu)$). Assume also that P maps non-negative functions to non-negative functions. Then the maximal function $Mf:=\sup_{N>0}\frac{1}{N}\sum_{n=1}^N P^n f$ obeys the inequality*

$$\lambda\mu(\{Mf>\lambda\})\leq\int_{Mf>\lambda} f\,d\mu \tag{2.76}$$

for any $\lambda\in\mathbf{R}$.

Proof. We can rewrite (2.76) as

$$\int_{Mf-\lambda>0} (f-\lambda)\ d\mu \geq 0. \tag{2.77}$$

Since $Mf - \lambda = M(f-\lambda)$, we thus see (by replacing f with $f-\lambda$) that we can reduce to proving (2.77) in the case $\lambda = 0$.

For every $m \geq 1$, consider the modified maximal function $F_m := \sup_{0\leq N\leq m} \sum_{n=0}^{N-1} P^n f$. Observe that $Mf(x) > 0$ if and only if $F_m(x) > 0$ for all sufficiently large m. By the dominated convergence theorem, it thus suffices to show that

$$\int_{F_m>0} f\ d\mu \geq 0 \tag{2.78}$$

for all m. But observe from definition of F_m (and the positivity preserving nature of P) that we have the pointwise recursive inequality

$$F_m(x) \leq F_{m+1}(x) = \max(0, f + PF_m(x)). \tag{2.79}$$

Integrating this on the region $F_m > 0$ and using the non-negativity of F_m, we obtain

$$\int_X F_m\ d\mu \leq \int_{F_m>0} f + \int_X PF_m\ d\mu. \tag{2.80}$$

Since $F_m \in L^1(X, \mathcal{X}, \mu)$ and $P^*1 = 1$, the claim follows. □

Applying this in the case when P is a shift operator, and replacing f by $|f|$, we obtain

Corollary 2.9.2 (Maximal ergodic theorem)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system. Then for any $f \in L^1(X, \mathcal{X}, \mu)$ and $\lambda > 0$ one has*

$$\mu(\{\sup_N \frac{1}{N} \sum_{n=0}^{N-1} |T^n f| > \lambda\}) \leq \frac{1}{\lambda} \|f\|_{L^1(X,\mathcal{X},\mu)}. \tag{2.81}$$

Note that this inequality implies Markov's inequality

$$\mu(\{|f| > \lambda\}) \leq \frac{1}{\lambda} \int_X |f|\ d\mu. \tag{2.82}$$

as a special case. Applying the real interpolation method, one also easily deduces the maximal inequality

$$\| \sup_N \frac{1}{N} \sum_{n=0}^{N-1} |T^n f| \|_{L^p(X,\mathcal{X},\mu)} \leq C_p \|f\|_{L^p(X,\mathcal{X},\mu)} \tag{2.83}$$

for all $1 < p \leq \infty$, where the constant C_p depends on p (it blows up like $O(1/(p-1))$ in the limit $p \to 1$).

Exercise 2.9.1 (Rising sun inequality)**.** If $f \in l^1(\mathbf{Z})$, and $f^*(m) := \sup_N \frac{1}{N} \sum_{n=0}^{N-1} f(m+n)$, establish the *rising sun inequality*

$$\lambda |\{m \in \mathbf{Z} : f^*(m) > \lambda\}| \leq \sum_{m \in \mathbf{Z}} f(m) \tag{2.84}$$

for any $\lambda > 0$. *Hint*: one can either adapt the proof of Theorem 2.9.1, or else partition the set appearing in (2.84) into disjoint intervals. The latter proof also leads to a proof of Corollary 2.9.2 which avoids the Dunford-Schwartz trick of introducing the functions F_m. The terminology "rising sun" comes from seeing how these intervals interact with the graph of the partial sums of f, which resembles the shadows cast on a hilly terrain by a rising sun.

Exercise 2.9.2 (Transference principle)**.** Show that Corollary 2.9.2 can be deduced directly from (2.84). *Hint*: given $f \in L^1(X, \mathcal{X}, \mu)$, apply (2.84) to the functions $f_x(n) := T^n f(x)$ for each $x \in X$ (truncating the integers to a finite set if necessary), and then integrate in x using Fubini's theorem. (This is an example of a *transference principle* between maximal inequalities on $\mathbf{Z}$ and maximal inequalities on measure-preserving systems.)

Exercise 2.9.3 (Stein-Stromberg maximal inequality)**.** [**StSt1983**] Derive a continuous version of the Dunford-Schwartz maximal inequality, in which the operators P^n are replaced by a semigroup P_t acting on both L^1 and L^∞, in which the underlying measure space is only assumed to be σ-finite rather than a probability space, and the averages $\frac{1}{N} \sum_{n=0}^{N-1} P^n$ are replaced by $\frac{1}{T} \int_0^T P^t \, dt$. Apply this continuous version with $P_t := e^{t\Delta}$ equal to the heat operator on $\mathbf{R}^d$ for

$d \geq 1$ to deduce the *Stein-Stromberg maximal inequality*[36]

(2.85)
$$m(\{x \in \mathbf{R}^d : \sup_{R>0} \frac{1}{m(B(x,R))} \int_{B(x,r)} |f|\, dm > \lambda\}) \leq \frac{Cd}{\lambda} \|f\|_{L^1(\mathbf{R}^d, dm)}$$

for all $\lambda > 0$ and $f \in L^1(\mathbf{R}^d, dm)$, where m is Lebesgue measure, $B(x,R)$ is the Euclidean ball of radius R centred at x, and the constant C is absolute (independent of d).

Remark 2.9.3. The study of maximal inequalities in ergodic theory is, of course, a subject in itself; a classical reference is [**St1970**].

2.9.2. The pointwise ergodic theorem. Using the maximal ergodic theorem and a standard limiting argument we can now deduce

Theorem 2.9.4 (Pointwise ergodic theorem)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $f \in L^1(X, \mathcal{X}, \mu)$. Then for μ-almost every $x \in X$, $\frac{1}{N} \sum_{n=0}^{N-1} T^n f(x)$ converges to $\mathbf{E}(f|\mathcal{X}^T)(x)$.*

Proof. By subtracting $\mathbf{E}(f|\mathcal{X}^T)$ from f if necessary, it suffices to show that

$$\limsup_{N\to\infty} |\frac{1}{N} \sum_{n=0}^{N-1} T^n f(x)| = 0 \tag{2.86}$$

a.e. whenever $\mathbf{E}(f|\mathcal{X}^T) = 0$. By telescoping series, (2.86) is already true when f takes the form $f = Tg - g$ for some $g \in L^\infty(X, \mathcal{X}, \mu)$. So by the arguments used to prove Theorem 2.8.5, we have already established the claim for a dense class of functions f in $L^2(X, \mathcal{X}, \mu)$ with $\mathbf{E}(f|\mathcal{X}^T) = 0$, and thus also for a dense class of functions in $L^1(X, \mathcal{X}, \mu)$ with $\mathbf{E}(f|\mathcal{X}^T) = 0$ (since the latter space is dense in the former, and the L^2 norm controls the L^1 norm by the Cauchy-Schwarz inequality).

Now we use a standard limiting argument. Let $f \in L^1(X, \mathcal{X}, \mu)$ with $\mathbf{E}(f|\mathcal{X}^T) = 0$. Then we can find a sequence f_j in the above

[36] This improves upon the *Hardy-Littlewood maximal inequality*, which gives the same estimate but with Cd replaced by C^d. It is an open question whether the dependence on d can be removed entirely; the estimate (2.85) is still the best known in high dimension. For $d = 1$, the best constant C is known to be $\frac{11+\sqrt{61}}{12} = 1.567\ldots$, a result of Melas[**Me2003**].

dense class which converges in L^1 to f. For almost every x, we thus have

$$\lim_{N\to\infty} |\frac{1}{N}\sum_{n=0}^{N-1} T^n f_j(x)| = 0 \tag{2.87}$$

for all j, and so by the triangle inequality we have

$$\limsup_{N\to\infty} |\frac{1}{N}\sum_{n=0}^{N-1} T^n f(x)| \leq \sup_N \frac{1}{N}\sum_{n=0}^{N-1} T^n |f - f_j|(x). \tag{2.88}$$

But by Corollary 2.9.2 we see that the right-hand side of (2.88) converges to zero in measure as $j \to \infty$. Since the left-hand side does not depend on j, it must vanish almost everywhere, as required. □

Remark 2.9.5. More generally, one can derive a pointwise convergence result on a class of rough functions by first establishing convergence for a dense subclass of functions, and then establishing a maximal inequality which is strong enough to allow one to take limits and establish pointwise convergence for all functions in the larger class. Conversely, principles such as Stein's maximal principle[**St1961**] indicate that in many cases this is in some sense the *only* way to establish such pointwise convergence results for rough functions.

Remark 2.9.6. Using the dominated convergence theorem (starting first with bounded functions f in order to get the domination), one can deduce the mean ergodic theorem from the pointwise ergodic theorem. But the converse is significantly more difficult; pointwise convergence for various ergodic averages is often a much harder result to establish than the corresponding norm convergence result (in particular, many of the techniques discussed in this course appear to be of sharply limited utility for pointwise convergence problems), and many questions in this area remain open.

Exercise 2.9.4 (Lebesgue differentiation theorem)**.** Let $f \in L^1(\mathbf{R}^d, dm)$ with Lebesgue measure dm. Show that for almost every $x \in \mathbf{R}^d$, we have $\lim_{r\to 0^+} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|\, dx = 0$, and in particular that $\lim_{r\to 0^+} \frac{1}{m(B(x,r))} \int_{B(x,r)} f(y)\, dx = f(x)$.

2.9.3. Ergodicity. Combining the mean ergodic theorem with the pointwise ergodic theorem (and with Exercises 2.8.7, 2.8.8) we obtain

Theorem 2.9.7 (Characterisations of ergodicity)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system. Then the following are equivalent:*

(1) *Any set $E \in \mathcal{X}$ which is invariant (thus $TE = E$) has either full measure $\mu(E) = 1$ or zero measure $\mu(E) = 0$.*

(2) *Any set $E \in \mathcal{X}$ which is almost invariant (thus TE differs from E by a null set) has either full measure or zero measure.*

(3) *Any measurable function f with $Tf = f$ a.e. is constant a.e.*

(4) *For any $1 < p < \infty$ and $f \in L^p(X, \mathcal{X}, \mu)$, the averages $\frac{1}{N}\sum_{n=0}^{N} T^n f$ converge in L^p norm to $\int_X f$.*

(5) *For any two $f, g \in L^\infty(X, \mathcal{X}, \mu)$, we have $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int_X (T^n f)g$ $(\int_X f\ d\mu)(\int_X g\ d\mu)$.*

(6) *For any two measurable sets E and F, we have $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(T^n$ $F) = \mu(E)\mu(F)$.*

(7) *For any $f \in L^1(X, \mathcal{X}, \mu)$, the averages $\frac{1}{N}\sum_{n=0}^{N} T^n f$ converge pointwise almost everywhere to $\int_X f\ d\mu$.*

A measure-preserving system with any (and hence all) of the above properties is said to be *ergodic*.

Remark 2.9.8. Strictly speaking, ergodicity is a property that applies to a measure-preserving system $(X, \mathcal{X}, \mu, T)$. However, we shall sometimes abuse notation and apply the adjective "ergodic" to a single component of a system, such as the measure μ or the shift T, when the other three components of the system are clear from context.

Here are some simple examples of ergodicity:

Example 2.9.9. If X is finite with uniform measure, then a shift map $T : X \to X$ is ergodic if and only if it is a cycle.

Example 2.9.10. If a shift T is ergodic, then so is T^{-1}. However, from Example 2.9.9 we see that it is not necessarily true that T^n is ergodic for all n (this latter property is also known as *total ergodicity*).

Exercise 2.9.5. Show that the circle shift $(\mathbf{R}/\mathbf{Z}, x \mapsto x+\alpha)$ (with the usual Lebesgue measure) is ergodic if and only if α is irrational. *Hint*: analyse the equation $Tf = f$ for (say) $f \in L^2(X, \mathcal{X}, \mu)$ using Fourier analysis. Another way to proceed is to use the Lebesgue density theorem (or Lebesgue differentiation theorem) combined with Exercise 2.6.15.

Exercise 2.9.6. Let $(\Omega, \mathcal{B}, \mu)$ be a probability space. Show that the Bernoulli shift on the product system $(\Omega^{\mathbf{Z}}, \mathcal{B}^{\mathbf{Z}}, \mu^{\mathbf{Z}})$ is ergodic. *Hint*: first establish property 6 of Theorem 2.9.7 when E and F each depend on only finitely many of the coordinates of $\Omega^{\mathbf{Z}}$.

Exercise 2.9.7. Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system. Show that if λ is an eigenvalue of $T : L^2(X, \mathcal{X}, \mu) \to L^2(X, \mathcal{X}, \mu)$, then $|\lambda| = 1$, the eigenspace $\{f \in L^2(X, \mathcal{X}, \mu) : Tf = \lambda f\}$ is one-dimensional, and that every eigenfunction f has constant magnitude $|f|$ a.e.. Show that the eigenspaces are orthogonal to each other in $L^2(X, \mathcal{X}, \mu)$, and the set of all eigenvalues of T forms an at most countable subgroup of the unit circle S^1.

Now we give a less trivial example of an ergodic system.

Proposition 2.9.11. *(Ergodicity of skew shift) Let $\alpha \in \mathbf{R}$ be irrational. Then the skew shift $((\mathbf{R}/\mathbf{Z})^2, (x,y) \mapsto (x+\alpha, y+x))$ is ergodic.*

Proof. Write the skew shift system as $(X, \mathcal{X}, \mu, T)$. To simplify the notation we shall omit the phrase "almost everywhere" in what follows.

We use an argument of Parry[**Pa1969**]. If the system is not ergodic, then we can find a non-constant $f \in L^2(X, \mathcal{X}, \mu)$ such that $Tf = f$. Next, we use Fourier analysis to write $f = \sum_m f_m$, where $f_m(x,y) := \int_{\mathbf{R}/\mathbf{Z}} f(x, y+\theta) e^{-2\pi i m\theta}\, d\theta$. Since f is T-invariant, and the vertical rotations $(x,y) \mapsto (x, y+\theta)$ commute with T, we see that the f_m are also T-invariant. The function f_0 depends only on the x variable, and so is constant by Exercise 2.9.5. So it suffices to show that f_m is zero for all non-zero m.

Fix m. We can factorise $f_m(x,y) = F_m(x) e^{2\pi i m y}$. The T-invariance of f_m now implies that $F_m(x+\alpha) = e^{-2\pi i m x} F_m(x)$. If

we then define $F_{m,\theta} := F_m(x+\theta)\overline{F_m(x)}$ for $\theta \in \mathbf{R}$, we see that $F_{m,\theta}(x+\alpha) = e^{-2\pi i m\theta} F_{m,\theta}(x)$, thus $F_{m,\theta}$ is an eigenfunction of the circle shift with eigenvalue $e^{-2\pi i m\theta}$. But this implies (by Exercise 2.9.7) that $F_{m,\theta}$ is orthogonal to $F_{m,0}$ for θ close to zero. Taking limits we see that $F_{m,0}$ is orthogonal to itself and must vanish; this implies that F_m and hence f_m vanish as well, as desired. □

Exercise 2.9.8. Show that for any irrational α and any $d \geq 1$, the iterated skew shift system $(\mathbf{R}/\mathbf{Z}^d, (x_1, \ldots, x_d) \to (x_1 + \alpha, x_2 + x_1, \ldots, x_d + x_{d-1}))$ is ergodic.

2.9.4. Generic points. Now let us suppose that we have a topological measure preserving system $(X, \mathcal{F}, \mu, T)$, i.e. a measure-preserving system $(X, \mathcal{X}, \mu, T)$ which is also a topological dynamical system $(X, \mathcal{F}, T)$, with $\mathcal{X}$ the Borel σ-algebra of T. Then we have the space $C(X)$ of continuous (real or complex-valued) functions on X, which is dense inside $L^2(X)$. From the Stone-Weierstrass theorem we also see that $C(X)$ is separable.

Definition 2.9.12. Let $(X, \mathcal{X}, \mu)$ be a probability space. A sequence $x_1, x_2, x_3, \ldots$ in X is said to be *uniformly distributed* with respect to μ if we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_X f \, d\mu \tag{2.89}$$

for all $f \in C(X)$. A point x in X is said to be *generic* if the forward orbit $x, Tx, T^2x, \ldots$ is uniformly distributed.

Exercise 2.9.9. Let $(X, \mathcal{F}, \mu)$ be a compact metrisable space with a Borel probability measure μ, and let $x_1, x_2, \ldots$ be a sequence in X. Show that this sequence is uniformly distributed if and only if $\lim_{N\to\infty} \frac{1}{N} |\{1 \leq i \leq N : x_i \in U\}| = \mu(U)$ for all open sets U in X.

From Theorem 2.9.7 and the separability of $C(X)$ we obtain

Proposition 2.9.13. *A topological measure-preserving system is ergodic if and only if almost every point is generic.*

A topological measure-preserving system is said to be uniquely ergodic if *every* point is generic. The following exercise explains the terminology:

Exercise 2.9.10. Show that a topological measure-preserving system $(X, \mathcal{F}, \mu, T)$ is uniquely ergodic if and only if the only T-invariant Borel probability measure on T is μ. *Hint*: use Lemma 2.7.16. Because of this fact, one can sensibly define what it means for a topological dynamical system $(X, \mathcal{F}, T)$ to be uniquely ergodic, namely that it has a unique T-invariant Borel probability measure.

It is not always the case that an ergodic system is uniquely ergodic. For instance, in the Bernoulli system $\{0,1\}^{\mathbf{Z}}$ (with uniform measure on $\{0,1\}$, say), the point $0^{\mathbf{Z}}$ is not generic. However, for more algebraic systems, it turns out that ergodicity and unique ergodicity are largely equivalent. We illustrate this with the circle and skew shifts:

Exercise 2.9.11. Show that the circle shift $(\mathbf{R}/\mathbf{Z}, x \mapsto x + \alpha)$ (with the usual Lebesgue measure) is uniquely ergodic if and only if α is irrational. *Hint*: first show in the circle shift system that any translate of a generic point is generic.

Proposition 2.9.14 (Unique ergodicity of skew shift)**.** *Let $\alpha \in \mathbf{R}$ be irrational. Then the skew shift $((\mathbf{R}/\mathbf{Z})^2, (x,y) \mapsto (x+\alpha, y+x))$ is uniquely ergodic.*

Proof. We use an argument of Furstenberg[**Fu1981**]. We again write the skew shift as $(X, \mathcal{X}, \mu, T)$. Suppose this system was not uniquely ergodic, then by Exercise 2.9.10 there is another shift-invariant Borel probability measure $\mu' \neq \mu$. If we push μ and μ' down to the circle shift system $(\mathbf{R}/\mathbf{Z}, x \mapsto x + \alpha)$ by the projection map $(x,y) \mapsto x$, then by Exercises 2.9.10, 2.9.11 we must get the same measure. Thus μ and μ' must agree on any set of the form $A \times (\mathbf{R}/\mathbf{Z})$.

Let E denote the points in X which are generic with respect to μ; note that this set is Borel measurable. By Proposition 2.9.13, this set has full measure in μ. Also, since the vertical rotations $(x,y) \mapsto (x, y+\theta)$ commute with T and preserve μ, we see that E must be invariant under such rotations; thus they are of the form $A \times (\mathbf{R}/\mathbf{Z})$ for some A. By the preceding discussion, we conclude that E also has full measure in μ'. But then (by the pointwise or mean ergodic theorem for $(X, \mathcal{X}, \mu', T)$) we conclude that $\mathbf{E}_{\mu'}(f|\mathcal{X}^T) = \int_X f \, d\mu$ μ'-almost everywhere for every continuous f, and thus on integrating

with respect to μ' we obtain $\int_X f\ d\mu' = \int_X f\ d\mu$ for every continuous f. But then by the Riesz representation theorem we have $\mu = \mu'$, a contradiction. □

Corollary 2.9.15. *If $\alpha \in \mathbf{R}$ is irrational, then the sequence $(\alpha n^2 \bmod 1)_{n\in\mathbf{N}}$ is uniformly distributed in $\mathbf{R}/\mathbf{Z}$ (with respect to uniform measure).*

Exercise 2.9.12. Show that the systems considered in Exercise 2.9.8 are uniquely ergodic. Conclude that the exponent 2 in Corollary 2.9.15 can be replaced by any positive integer d.

Note that the topological dynamics theory developed in Section 2.6 only establishes the weaker statement that the above sequence is dense in $\mathbf{R}/\mathbf{Z}$ rather than uniformly distributed. More generally, it seems that ergodic theory methods can prove topological dynamics results, but not vice versa. Here is another simple example of the same phenomenon:

Exercise 2.9.13. Show that a uniquely ergodic topological dynamical system is necessarily minimal. (The converse is not necessarily true, as already mentioned in Remark 2.7.18.)

2.9.5. The ergodic decomposition. Just as not every topological dynamical system is minimal, not every measure-preserving system is ergodic. Nevertheless, there is an important decomposition that allows one to represent non-ergodic measures as averages of ergodic measures. One can already see this in the finite case, when X is a finite set with the discrete σ-algebra, and $T : X \to X$ is a permutation on X, which can be decomposed as the disjoint union of cycles on a partition $X = C_1 \cup \ldots \cup C_m$ of X. In this case, all shift-invariant probability measures take the form

$$\mu = \sum_{j=1}^{m} \alpha_j \mu_j \tag{2.90}$$

where μ_j is the uniform probability measure on the cycle C_j, and α_j are non-negative constants adding up to 1. Each of the μ_j are ergodic, but no non-trivial linear combination of these measures is ergodic. Thus we see in the finite case that every shift-invariant

measure can be uniquely expressed as a convex combination of ergodic measures.

It turns out that a similar decomposition is available in general, at least if the underlying measure space is a compact topological space (or more generally, a Radon space). This is because of the following general theorem from measure theory.

Definition 2.9.16 (Probability kernel)**.** Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces. A *probability kernel* $y \mapsto \mu_y$ is an assignment of a probability measure μ_y on X to each $y \in Y$ in such a way that the map $y \mapsto \int_X f \, d\mu_y$ is measurable for every bounded measurable $f : X \mapsto \mathbf{C}$.

Example 2.9.17. Every measurable map $\phi : Y \to X$ induces a probability kernel $y \mapsto \delta_{\phi(y)}$. Every probability measure on X can be viewed as a probability kernel from a point to X. If $y \mapsto \mu_y$ and $x \mapsto \nu_x$ are two probability kernels from Y to X and from X to Z respectively, their composition $x \mapsto (\mu \circ \nu)_x := \int_X \mu_y \, d\nu_x(y)$ is also a probability kernel, where $\int_X \mu_y \, d\nu_x(y)$ is the measure that assigns $\int_X \mu_y(E) \, d\nu_x(y)$ to any measurable set E in Z. Thus one can view the class of measurable spaces and their probability kernels as a category, which includes the class of measurable spaces and their measurable maps as a subcategory.

Definition 2.9.18 (Regular space)**.** A measurable space $(X, \mathcal{X})$ is said to be *regular* if there exists a compact metrisable topology $\mathcal{F}$ on X for which $\mathcal{X}$ is the Borel σ-algebra.

Example 2.9.19. Every topological measure-preserving system is regular.

Remark 2.9.20. Measurable spaces $(X, \mathcal{X})$ in which $\mathcal{X}$ is the Borel σ-algebra of a topological space generated by a separable complete metric space (i.e. a Polish space) are known as *standard Borel spaces*. It is a non-trivial theorem from descriptive set theory that up to measurable isomorphism, there are only three types of standard Borel spaces: finite discrete spaces, countable discrete spaces, and the unit interval [0,1] with the usual Borel σ-algebra. From this one can see that regular spaces are the same as standard Borel spaces, though we will not need this fact here.

Theorem 2.9.21 (Disintegration theorem)**.** *Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be probability spaces, with $(X, \mathcal{X})$ regular. Let $\pi : X \to Y$ be a morphism (thus $\nu = \pi_\# \mu$). Then there exists a probability kernel $y \mapsto \mu_y$ such that*

$$\int_X f(g \circ \pi)\ d\mu = \int_Y (\int_X f\ d\mu_y) g(y)\ d\nu(y) \tag{2.91}$$

for any bounded measurable $f : X \to \mathbf{C}$ and $g : Y \to \mathbf{C}$. Also, for any such g, we have

$$g \circ \pi = g(y), \mu_y - a.e. \tag{2.92}$$

for ν-a.e. y.

Furthermore, this probability kernel is unique up to ν-almost everywhere equivalence, in the sense that if $y \mapsto \mu'_y$ is another probability kernel with the same properties, then $\mu_y = \mu'_y$ for ν-almost every y.

We refer to the probability kernel $y \mapsto \mu_y$ generated by the above theorem as the *disintegration* of μ relative to the factor map π.

Proof. We begin by proving uniqueness. Suppose we have two probability kernels $y \mapsto \mu_y, y \mapsto \mu'_y$ with the above properties. Then on subtraction we have

$$\int_Y (\int_X f\ d(\mu_y - \mu'_y)) g(y)\ d\nu(y) = 0 \tag{2.93}$$

for all bounded measurable $f : X \to \mathbf{C}$, $g : Y \to \mathbf{C}$. Specialising to $f = 1_E$ for some measurable set $E \in \mathcal{X}$, we conclude that $\mu_y(E) = \mu'_y(E)$ for ν-almost every y. Since $\mathcal{X}$ is regular, it is separable and we conclude that $\mu_y = \mu'_y$ for ν-almost every y, as required.

Now we prove existence. The pullback map $\pi^\# : L^2(Y, \mathcal{Y}, \nu) \to L^2(X, \mathcal{X}, \mu)$ defined by $g \mapsto g \circ \pi$ has an adjoint $\pi_\# : L^2(X, \mathcal{X}, \mu) \to L^2(Y, \mathcal{Y}, \nu)$, thus

$$\int_X f(g \circ \pi)\ d\mu = \int_Y (\pi_\# f) g\ d\nu \tag{2.94}$$

for all $f \in L^2(X, \mathcal{X}, \mu)$ and $g \in L^2(Y, \mathcal{Y}, \nu)$. It is easy to see from duality that we have $\|\pi_\# f\|_{L^\infty(Y,\mathcal{Y},\nu)} \leq \|f\|_{C(X)}$ for all $f \in C(X)$ (where we select a compact metrisable topology that generates the regular σ-algebra $\mathcal{X}$). Recall that $\pi_\# f$ is not quite a measurable function, but is instead an equivalence class of measurable functions

modulo ν-almost everywhere equivalence. Since $C(X)$ is separable, we find a measurable representative $\tilde{\pi}_\# f : Y \to \mathbf{C}$ of $\pi_\# f$ to every $f \in C(X)$ which varies linearly with f, and is such that $|\tilde{\pi}_\# f(y)| \leq \|f\|_{C(X)}$ for all y outside of a set E of ν-measure zero and for all $f \in C(X)$. For all such y, we can then apply the Riesz representation theorem to obtain a Borel probability measure μ_y such that

$$\tilde{\pi}_\# f(y) = \int_X f \, d\mu_y \tag{2.95}$$

for all such y. We set μ_y equal to some arbitrarily fixed Borel probability measure for $y \in E$. We then observe that the required properties (including the measurability of $y \mapsto \int_X f \, d\mu_y$) are already obeyed for $f \in C(X)$. To generalise this to bounded measurable f, observe that the class $\mathcal{C}$ of f obeying the required properties is closed under dominated pointwise convergence, and so contains the indicator functions of open sets (by *Urysohn's lemma*). Applying dominated pointwise convergence again, together with linearity, we see that the sets whose indicator functions lie in $\mathcal{C}$ form a σ-algebra and so contain all Borel sets. Thus all simple measurable functions lie in $\mathcal{C}$, and on taking uniform limits we obtain the claim.

Finally, we prove (2.92). From two applications of (2.91) we have

$$\int_Y (\int_X f(g \circ \pi) \, d\mu_y) h(y) \, d\nu(y) = \int_Y (\int_X fg(y) \, d\mu_y) h(y) \, d\nu(y) \tag{2.96}$$

for all bounded measurable $f : X \to \mathbf{C}$ and $h : Y \to \mathbf{C}$. The claim follows (using the separability of the space of all f). □

Exercise 2.9.14. Let the notation and assumptions be as in Theorem 2.9.21. Suppose that $\mathcal{Y}$ is also regular, and that the map $\pi : X \to Y$ is continuous with respect to some compact metrisable topologies that generate $\mathcal{X}$ and $\mathcal{Y}$ respectively. Then show that for ν-almost every y, the probability measure ν_y is supported in $\pi^{-1}(\{y\})$.

Proposition 2.9.22 (Ergodic decomposition)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a regular measure-preserving system. Let $(Y, \mathcal{Y}, \nu, S)$ be the system defined by $Y := X$, $\mathcal{Y} := \mathcal{X}^T$, $\nu := \mu \restriction_{\mathcal{Y}}$, and $S := T$, and let $\pi : X \to Y$ be the identity map. Let $y \mapsto \mu_y$ be the disintegration of μ with respect to the factor map π. Then for ν-almost every y, the measure μ_y is T-invariant and ergodic.*

Proof. Observe from the T-invariance $\mu = T_\# \mu$ of μ (and of $\mathcal{X}^T$) that the probability kernel $y \mapsto T_\# \mu_y$ would also be a disintegration of μ. Thus we have $\mu_y = T_\# \mu_y$ for ν-almost every y.

Now we show the ergodicity. As the space of bounded measurable $f : X \to \mathbf{C}$ is separable, it suffices by Theorem 2.9.7 and a limiting argument to show that for any fixed such f, the averages $\frac{1}{N} \sum_{n=1}^N T^n f$ converge pointwise μ_y-a.e. to $\int_X f \, d\mu_y$ for ν-a.e. y.

From the pointwise ergodic theorem, we already know that $\frac{1}{N} \sum_{n=1}^N T^n f$ converges to $\mathbf{E}(f|\mathcal{X}^T)$ outside of a set of μ-measure zero. By (2.91), this set also has μ_y-measure zero for ν-almost every y. Thus it will suffice to show that $\mathbf{E}(f|\mathcal{X}^T)$ is μ_y-a.e. equal to $\int_X f \, d\mu_y$ for ν-a.e. y. Now observe that $\mathbf{E}(f|\mathcal{X}^T)(x) = \pi_\# f(\pi(x))$, so the claim follows from (2.92) and (2.95). □

Exercise 2.9.15. Let $(X, \mathcal{X})$ be a separable measurable space, and let T be bimeasurable bijection $T : X \to X$. Let $M(\mathcal{X})$ denote the Banach space of all finite measures on $\mathcal{X}$ with the total variation norm. Let $\Pr(\mathcal{X})^T \subset M(\mathcal{X})$ denote the collection of probability measures on $\mathcal{X}$ which are T-invariant. Show that this is a closed convex subset of $M(\mathcal{X})$, and the extreme points of $\Pr(\mathcal{X})^T$ are precisely the ergodic probability measures (which also form a closed subset of $M(\mathcal{X})$). (This allows one to prove a variant of Proposition 2.9.22 using Choquet's theorem.)

Exercise 2.9.16. Show that a topological measure-preserving system $(X, \mathcal{F}, T, \mu)$ is uniquely ergodic if and only if the only ergodic shift-invariant Borel probability measure on X is μ.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/02/04`. Thanks to Lior Silberman and Liu Xiao Chuan for corrections.

2.10. The Furstenberg correspondence principle

In this lecture, we describe the simple but fundamental *Furstenberg correspondence principle* which connects the[37] "soft analysis" subject of ergodic theory (in particular, recurrence theorems) with the "hard

[37] See Section 1.3 of *Structure and Randomness* for a discussion of the relationship between soft and hard analysis.

analysis" subject of combinatorial number theory (or more generally with results of "density Ramsey theory" type). Rather than try to set up the most general and abstract version of this principle, we shall instead study the canonical example of this principle in action, namely the equating of the *Furstenberg multiple recurrence theorem* with Szemerédi's theorem on arithmetic progressions.

In [**Sz1975**], Szemerédi established the following theorem, which had been conjectured by Erdős and Turán[**ErTu1936**]:

Theorem 2.10.1 (Szemerédi's theorem)**.** [**Sz1975**] *Let $k \geq 1$ be an integer, and let A be a set of integers of positive upper density, thus $\limsup_{N\to\infty} \frac{1}{2N+1}|A \cap \{-N, \ldots, N\}| > 0$. Then A contains a non-trivial arithmetic progression $n, n+r, \ldots, n+(k-1)r$ of length k. (By "non-trivial" we mean that $r \neq 0$.) Or more succinctly: every set of integers of positive upper density contains arbitrarily long arithmetic progressions.*

Remark 2.10.2. This theorem is trivial for $k = 1$ and $k = 2$. The first non-trivial case is $k = 3$, which was proven in [**Ro1953**] and will be discussed in Section 2.12.4. The $k = 4$ case was also established earlier in [**Sz1969**].

In [**Fu1977**], Furstenberg gave another proof of Szemerédi's theorem, by establishing the following equivalent statement:

Theorem 2.10.3 (Furstenberg multiple recurrence theorem)**.** *Let $k \geq 1$ be an integer, let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let E be a set of positive measure. Then there exists $r > 0$ such that $E \cap T^{-r}E \cap \ldots \cap T^{-(k-1)r}E$ is non-empty.*

Remark 2.10.4. The negative signs here can be easily removed because T is invertible, but I have placed them here for consistency with some later results involving non-invertible transformations, in which the negative sign becomes important.

Exercise 2.10.1. Prove that Theorem 2.10.3 is equivalent to the apparently stronger theorem in which "is non-empty" is replaced by "has positive measure", and "there exists $r > 0$" is replaced by "there exist infinitely many $r > 0$".

Note that the $k = 1$ case of Theorem 2.10.3 is trivial, while the $k = 2$ case follows from the Poincaré recurrence theorem (Theorem 2.8.2). We will prove the higher k cases of this theorem in Sections 2.11-2.15. In this one, we will explain why, for any fixed k, Theorem 2.10.1 and Theorem 2.10.3 are equivalent.

Let us first give the easy implication that Theorem 2.10.1 implies Theorem 2.10.3. This follows immediately from

Lemma 2.10.5. *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let E be a set of positive measure. Then there exists a point x in X such that the recurrence set $\{n \in \mathbf{Z} : T^n x \in E\}$ has positive upper density.*

Indeed, from Lemma 2.10.5 and Theorem 2.10.1, we obtain a point x for which the set $\{n \in \mathbf{Z} : T^n x \in E\}$ contains an arithmetic progression of length k and some step r, which implies that $E \cap T^r E \cap \ldots \cap T^{(k-1)r} E$ is non-empty.

Proof of Lemma 2.10.5. Observe (from the shift-invariance of μ) that

$$\int_X \frac{1}{2N+1} \sum_{n=-N}^{N} 1_{T^n E} \, d\mu = \mu(E). \tag{2.97}$$

On the other hand, the integrand is at most 1. We conclude that for each N, the set $A_N := \{x : \frac{1}{2N+1} \sum_{n=-N}^{N} 1_{T^n E}(x) \geq \mu(E)/2\}$ must have measure at least $\mu(E)/2$. This implies that the function $\sum_N 1_{A_N}$ is not absolutely integrable even after excluding an arbitrary set of measure up to $\mu(E)/4$, which implies that $\sum_N 1_{A_N}$ is not finite a.e., and the claim follows (cf. the proof of the Borel-Cantelli lemma, Lemma 1.5.5). □

Now we show how Theorem 2.10.3 implies Theorem 2.10.1. If we could pretend that "upper density" was a probability measure on the integers, then this implication would be immediate by applying Theorem 2.10.3 to the dynamical system $(\mathbf{Z}, n \mapsto n + 1)$. Of course, we know that the integers do not admit a shift-invariant probability measure (and upper density is not even additive, let alone a probability measure). So this does not work directly. Instead, we need to

first lift from the integers to a more abstract universal space and use a standard "compactness and contradiction" argument in order to be able to build the desired probability measure properly.

More precisely, let A be as in Theorem 2.10.1. Consider the topological boolean Bernoulli dynamical system $2^{\mathbf{Z}}$ with the product topology and the shift $T : B \mapsto B + 1$. The set A can be viewed as a point in this system, and the orbit closure $X := \overline{\{A + n : n \in \mathbf{Z}\}}$ of that point becomes a subsystem of that Bernoulli system, with the relative topology.

Suppose for contradiction that A contains no non-trivial progressions of length k, thus $A \cap A + r \cap \ldots \cap A + (k-1)r = \emptyset$ for all $r > 0$. Then, if we define the *cylinder set* $E := \{B \in X : 0 \in B\}$ to be the collection of all points in X which (viewed as sets of integers) contain 0, we see (after unpacking all the definitions) that $E \cap T^r E \cap \ldots T^{(k-1)r} E = \emptyset$ for all $r > 0$.

In order to apply Theorem 2.10.3 and obtain the desired contradiction, we need to find a shift-invariant Borel probability measure μ on X which assigns a positive measure to E.

For each integer N, consider the measure μ_N which assigns a mass of $\frac{1}{2N+1}$ to the points $T^{-n}A$ in X for $-N \leq n \leq N$, and no mass to the rest of X. Then we see that $\mu_N(E) = \frac{1}{2N+1}|A \cap \{-N, \ldots, N\}|$. Thus, since A has positive upper density, there exists some sequence N_j going to infinity such that $\liminf_{j\to\infty} \mu_{N_j}(E) > 0$. On the other hand, by vague sequential compactness (Lemma 2.7.16) we know that some subsequence of μ_{N_j} converges in the *vague topology* to a probability measure μ, which then assigns a positive measure to the (clopen) set E. As the μ_{N_j} are asymptotically shift invariant, we see that μ is invariant also (as in the proof of Corollary 2.7.17). As μ now has all the required properties, we have completed the deduction of Theorem 2.10.1 from Theorem 2.10.3.

Exercise 2.10.2. Show that Theorem 2.10.3 in fact implies a seemingly stronger version of Theorem 2.10.1, in which the conclusion becomes the assertion that the set $\{n : n, n+r, \ldots, n+(k-1)r \in A\}$ has positive upper density for infinitely many r.

Exercise 2.10.3. Show that Theorem 2.10.1 in fact implies a seemingly stronger version of Theorem 2.10.3: If $E_1, E_2, E_3, \ldots$ are sets in a probability space with uniformly positive measure (i.e. $\inf_n \mu(E_n) > 0$), then for any k there exists positive integers n, r such that $\mu(E_n \cap E_{n+r} \cap \ldots \cap E_{n+(k-1)r}) > 0$.

2.10.1. Varnavides type theorems. As observed in [**BeHoMcCPa2000**], a similar "compactness and contradiction" argument (combined with a preliminary averaging-over-dilations trick of Varnavides[**Va1959**]) allows us to use Theorem 2.10.3 to imply the following apparently stronger statement:

Theorem 2.10.6. *(Uniform Furstenberg multiple recurrence theorem) Let $k \geq 1$ be an integer and $\delta > 0$. Then for any measure-preserving system $(X, \mathcal{X}, \mu, T)$ and any measurable set E with $\mu(E) \geq \delta$ we have*

$$\frac{1}{N} \sum_{r=0}^{N-1} \mu(E \cap T^r E \cap \ldots \cap T^{(k-1)r} E) \geq c(k, \delta) \tag{2.98}$$

for all $N \geq 1$, where $c(k, \delta) > 0$ is a positive quantity which depends only on k and δ (i.e. it is uniform over all choices of system and of the set E with measure at least δ).

Exercise 2.10.4. Assuming Theorem 2.10.6, show that[38] if N is sufficiently large depending on k and δ, then any subset of $\{1, \ldots, N\}$ with cardinality at least δN will contain at least $c'(k, \delta)N^2$ non-trivial arithmetic progressions of length k, for some $c'(k, \delta) > 0$. Conclude in particular that Theorem 2.10.6 implies Theorem 2.10.1.

It is clear that Theorem 2.10.6 implies Theorem 2.10.3; let us now establish the converse. We first use an averaging argument of Varnavides to reduce Theorem 2.10.6 to a weaker statement, in which the conclusion (2.98) is not asserted to hold for all N, but instead one asserts that

$$\frac{1}{N_0} \sum_{r=1}^{N_0-1} \mu(E \cap T^r E \cap \ldots \cap T^{(k-1)r} E) \geq c(k, \delta) \tag{2.99}$$

[38] This result for $k = 3$ was first established in [**Va1959**] via an averaging argument from Roth's theorem.

is true for some $N_0 = N_0(k,\delta) > 0$ depending only on k and δ (note that the r=0 term in (2.99) has been dropped, otherwise the claim is trivial). To see why one can recover (2.98) from (2.99), observe by replacing the shift T with a power T^a that we can amplify (2.99) to

$$\frac{1}{N_0} \sum_{r=1}^{N_0-1} \mu(E \cap T^{ar}E \cap \ldots \cap T^{(k-1)ar}E) \geq c(k,\delta) \tag{2.100}$$

for all a. Averaging (2.100) over $1 \leq a \leq N$ we easily conclude (2.98).

It remains to prove that (2.100) holds under the hypotheses of Theorem 2.10.6. Our next reduction is to observe that for it suffices to perform this task for the boolean Bernoulli system $X_0 := 2^{\mathbf{Z}}$ with the cylinder set $E_0 := \{B \in X_0 : 0 \in B\}$ as before. To see this, recall from Example 2.2.6 that there is a morphism $\phi : X \to X_0$ from any measure-preserving system $(X, \mathcal{X}, \mu, T)$ with a distinguished set E to the system X_0 with the product σ-algebra $\mathcal{X}_0$, the usual shift T_0, and the set E_0, and with the push-forward measure $\mu_0 := \phi_\# \mu$. Specifically, ϕ sends any point x in X to its recurrence set $\phi(x) := \{n \in \mathbf{Z} : T^n x \in E\}$. Using this morphism it is not difficult to show that the claim (2.98) for $(X, \mathcal{X}, \mu, T)$ and E would follow from the same claim for $(X_0, \mathcal{X}_0, \mu_0, T_0)$ and E_0.

We still need to prove (2.99) for the boolean system. The point is that by lifting to this universal setting, the dynamical system $(X, \mathcal{X}, T)$ and the set E have been canonically fixed; the only remaining parameter is the probability measure μ. But now we can exploit vague sequential compactness again as follows.

Suppose for contradiction that Theorem 2.10.6 failed for the boolean system. Then by carefully negating all the quantifiers, we can find $\delta > 0$ such that for any N_0 there is a sequence of shift-invariant probability measures μ_j on X with $\mu_j(E) \geq \delta$,

$$\frac{1}{N_0} \sum_{r=1}^{N_0-1} \mu_j(E \cap T^r E \cap \ldots \cap T^{(k-1)r}E) \to 0 \tag{2.101}$$

as $j \to \infty$. Note that if (2.101) holds for one value of N_0, then it also holds for all smaller values of N_0. A standard diagonalisation argument then allows us to build a sequence μ_j as above, but which obeys (2.101) for *all* $N_0 \geq 1$.

Now we are finally in a good position to apply vague sequential compactness. By passing to a subsequence if necessary, we may assume that μ_j converges vaguely to a limit μ, which is a shift-invariant probability measure. In particular we have $\mu(E) \geq \delta > 0$, while from (2.101) we see that

$$\frac{1}{N_0} \sum_{r=1}^{N_0-1} \mu(E \cap T^r E \cap \ldots \cap T^{(k-1)r} E) = 0 \tag{2.102}$$

for all $N_0 \geq 1$; thus the sets $E \cap T^r E \cap \ldots \cap T^{(k-1)r} E$ all have zero measure for $r > 0$. But this contradicts Theorem 2.10.3 (and Exercise 2.10.1). This completes the deduction of Theorem 2.10.6 from Theorem 2.10.3.

2.10.2. Other recurrence theorems and their combinatorial counterparts. The Furstenberg correspondence principle can be extended to relate several other recurrence theorems to their combinatorial analogues. We give some representative examples here (without proofs). Firstly, there is a multidimensional version of Szemerédi's theorem (compare with Exercise 2.4.8):

Theorem 2.10.7 (Multidimensional Szemerédi theorem)**.** [**FuKa1979**] *Let $d \geq 1$, let $v_1, \ldots, v_k \in \mathbf{Z}^d$, and let $A \subset \mathbf{Z}^d$ be a set of positive upper Banach density (which means that $\limsup_{N\to\infty} |A \cap B_N|/|B_N| > 0$, where $B_N := \{-N, \ldots, N\}^d$). Then A contains a pattern of the form $n + rv_1, \ldots, n + rv_k$ for some $n \in \mathbf{Z}^d$ and $r > 0$.*

Note that Theorem 2.10.1 corresponds to the special case when $d = 1$ and $v_i = i - 1$.

This theorem was first proven by Furstenberg and Katznelson [**FuKa1979**], who deduced it via the correspondence principle from the following generalisation of Theorem 2.10.3:

Theorem 2.10.8 (Recurrence for multiple commuting shifts)**.** [**FuKa1979**] *Let $k \geq 1$ be an integer, let $(X, \mathcal{X}, \mu)$ be a probability space, let $T_1, \ldots, T_k : X \to X$ be measure-preserving bimeasurable maps which commute with each other, and let E be a set of positive measure. Then there exists $r > 0$ such that $T_1^r E \cap T_2^r E \cap \ldots \cap T_k^r E$ is non-empty.*

Exercise 2.10.5. Show that Theorem 2.10.7 and Theorem 2.10.8 are equivalent.

Exercise 2.10.6. State an analogue of Theorem 2.10.6 for multiple commuting shifts, and prove that it is equivalent to Theorem 2.10.8.

There is also a polynomial version of these theorems (cf. Theorem 2.5.1), which we will also state in general dimension:

Theorem 2.10.9 (Multidimensional polynomial Szemerédi theorem)**.** [**BeLe1996**] *Let $d \geq 1$, let $P_1, \ldots, P_k : \mathbf{Z} \to \mathbf{Z}^d$ be polynomials with $P_1(0) = \ldots = P_k(0) = 0$, and let $A \subset \mathbf{Z}^d$ be a set of positive upper Banach density. Then A contains a pattern of the form $n + P_1(r), \ldots, n + P_k(r)$ for some $n \in \mathbf{Z}^d$ and $r > 0$.*

This theorem was established by Bergelson and Leibman[**BeLe1996**], who deduced it from

Theorem 2.10.10 (Polynomial recurrence for multiple commuting shifts)**.** [**BeLe1996**] *Let k, $(X, \mathcal{X}, \mu)$, $T_1, \ldots, T_k : X \to X$, E be as in Theorem 2.10.8, and let $P_1, \ldots, P_k$ be as in Theorem 2.10.9. Then there exists $r > 0$ such that $T^{-P_1(r)}E \cap T^{-P_2(r)}E \cap \ldots \cap T^{-P_k(r)}E$ is non-empty, where we adopt the convention $T^{(a_1,\ldots,a_k)} := T_1^{a_1} \ldots T_k^{a_k}$ (thus we are making the action of $\mathbf{Z}^d$ on X explicit).*

Exercise 2.10.7. Show that Theorem 2.10.9 and Theorem 2.10.10 are equivalent.

Exercise 2.10.8. State an analogue of Theorem 2.10.6 for polynomial recurrence for multiple commuting shifts, and prove that it is equivalent to Theorem 2.10.10. *Hint*: first establish this in the case that each of the P_j are monomials, in which case there is enough dilation symmetry to use the Varnavides averaging trick. Interestingly, if one only restricts attention to one-dimensional systems $k = 1$, it does not seem possible to deduce the uniform polynomial recurrence theorem from the non-uniform polynomial recurrence theorem, thus indicating that the averaging trick is less universal in its applicability than the correspondence principle.

In the above theorems, the underlying action was given by either the integer group $\mathbf{Z}$ or the lattice group $\mathbf{Z}^d$. It is not too difficult to

generalise these results to the semigroups $\mathbf{N}$ and $\mathbf{N}^d$ (thus dropping the assumption that the shift maps are invertible), by using a trick similar to that used in Exercise 2.4.10, or by using the correspondence principle back and forth a few times. A bit more surprisingly, it is possible to extend these results to even weaker objects than semigroups. To describe this we need some more notation.

Define a *partial semigroup* $(G, \cdot)$ to be a set G together with a partially defined multiplication operation $\cdot : \Omega \to G$ for some subset $\Omega \subset G \times G$, which is associative in the sense that whenever $(a \cdot b) \cdot c$ is defined, then $a \cdot (b \cdot c)$ is defined and equal to $(a \cdot b) \cdot c$, and vice versa. A good example of a partial semigroup is the finite subsets $\binom{S}{<\omega} := \{A \subset S : |A| < \infty\}$ of a fixed set S, where the multiplication operation $A \cdot B$ is disjoint union, or more precisely $A \cdot \mathrm{B} := A \cup B$ when A and B are disjoint, and $A \cdot B$ is undefined otherwise.

Remark 2.10.11. One can extend a partial semigroup to be a genuine semigroup by adjoining a new element err to G, and redefining multiplication $a \cdot b$ to equal err if it was previously undefined (or if one of a or b was already equal to err). However, we will avoid using this trick here, as it tends to complicate the notation a little.

One can take Cartesian products of partial semigroups in the obvious manner to obtain more partial semigroups. In particular, we have the partial semigroup $\binom{\mathbf{N}}{<\omega}^d$ for any $d \geq 1$, defined as the collection of d-tuples $(A_1, \ldots, A_d)$ of finite sets of natural numbers (not necessarily disjoint), with the partial semigroup law $(A_1, \ldots, A_d) \cdot (B_1, \ldots, B_d) := (A_1 \cup B_1, \ldots, A_d \cup B_d)$ whenever A_i and B_i are disjoint for each $1 \leq i \leq d$.

If $(X, \mathcal{X}, \mu)$ is a probability space and $(G, \cdot)$ is a partial semigroup, we define a *measure-preserving action* of G on X to be an assignment of a measure-preserving transformation $T^g : X \to X$ (not necessarily invertible) to each $g \in G$, such that $T^{g \cdot h} = T^g T^h$ whenever $g \cdot h$ is defined.

An action T of $\binom{\mathbf{N}}{<\omega}$ on X is known as an *IP system* on X; it is generated by a countable number $T_1, T_2, \ldots$ of commuting measure-preserving transformations, with $T^A := \prod_{i \in A} T^i$. (Admittedly, it is possible that the action of the empty set is not necessarily the identity,

but this turns out to have a negligible impact on matters.) An action T of $\binom{\mathbf{N}}{<\omega}^d$ is then a collection of d simultaneously commuting IP systems.

In [**FuKa1985**], Furstenberg and Katznelson showed the following generalisation of Theorem 2.10.8:

Theorem 2.10.12 (IP multiple recurrence theorem)**.** *Let T be an action of $\binom{\mathbf{N}}{<\omega}^d$ on a probability space $(X, \mathcal{X}, \mu)$. Then there exists a non-empty set $A \in \binom{\mathbf{N}}{<\omega}$ such that $E \cap (T^{A_1})^{-1}(E) \cap \ldots \cap (T^{A_d})^{-1}(E)$ is non-empty, where $A_i := (\emptyset, \ldots, \emptyset, A, \emptyset, \ldots, \emptyset)$ is the group element which equals A in the i^{th} position and is the empty set otherwise.*

This theorem has a number of combinatorial consequences[39], such as the following strengthening of Szemerédi's theorem:

Theorem 2.10.13 (IP Szemerédi theorem)**.** [**FuKa1985**] *Let A be a set of integers of positive upper density, let $k \geq 1$, and let $B \subset \mathbf{N}$ be infinite. Then A contains an arithmetic progression $n, n+r, \ldots, n+(k-1)r$ of length k in which r lies in $FS(B)$, the set of finite sums of B (cf. Theorem 2.5.18).*

Exercise 2.10.9. Deduce Theorem 2.10.13 from Theorem 2.10.12.

Exercise 2.10.10. Using Theorem 2.10.13, show that for any k, and any set of integers A of positive upper density, the set of steps r which occur in the arithmetic progressions in A of length k is syndetic.

Exercise 2.10.11. Using Theorem 2.10.12, show that if $\mathbf{F}$ is a finite field, and $\mathbf{F}^{<\omega} := \bigcup_{n=0}^{\infty} \mathbf{F}^n$ is the canonical vector space over $\mathbf{F}$ spanned (in the algebraic sense) by a countably infinite number of basis vectors, show that any subset A of $\mathbf{F}^{<\omega}$ of positive upper Banach density (which means that $\limsup_{n\to\infty} |A \cap \mathbf{F}^n|/|\mathbf{F}^n| > 0$ contains affine subspaces of arbitrarily high dimension.

The IP recurrence theorem is already very powerful, but even stronger theorems are known. For instance, in [**FuKa1991**], Furstenberg and Katznelson established the following deep strengthening of

[39] There is also a multidimensional version of this theorem, but it requires a fair amount of notation to state properly.

the Hales-Jewett theorem (Theorem 2.5.21), as well as of Exercise 2.10.11 above:

Theorem 2.10.14 (Density Hales-Jewett theorem)**.** [**FuKa1991**] *Let A be a finite alphabet. If E is a subset of $A^{<\omega}$ of positive upper Banach density, then E contains a combinatorial line.*

This theorem was deduced (via an advanced form of the correspondence principle) by a somewhat complicated recurrence theorem which we will not state here; rather than the action of a group, semigroup, or partial semigroup, one instead works with an ensemble of sets (as in Exercise 2.10.3), and furthermore one regularises the system of the probability space and set ensemble (which can collectively be viewed as a random process) to be what Furstenberg and Katznelson call a *strongly stationary process*, which (very) roughly means that the statistics of this process look "the same" when restricted to any combinatorial subspace of a fixed dimension.

Remark 2.10.15. Similar correspondence principles can be established connecting property testing results for graphs and hypergraphs to the measure theory of exchangeable measures: see [**Ta2007c**], [**AuTa2008**], [**AvGeTo2008**], [**Ta2008**]. Finally, we have implicitly been using a similar correspondence principle between topological dynamics and colouring Ramsey theorems in Sections 2.3, 2.4, 2.5.

Remark 2.10.16. The Furstenberg correspondence principle also comes tantalisingly close to deducing my theorem with Ben Green [**GrTa2008**] that the primes contain arbitrarily long arithmetic progressions from Szemerédi's theorem. More precisely, they show that any subset A of a *genuinely* random set of integers with logarithmic-type density B, with A having positive *relative* upper density with respect to B, contains arbitrarily long arithmetic progressions; see [**Ta**]. Unfortunately, the almost primes are not known to quite obey enough "correlation conditions" to behave sufficiently pseudorandomly that these arguments apply to the primes, though perhaps there is still a "softer" way to prove our theorem than the way we did it (see the recent papers [**Go2008**], [**ReTrTuVa2008**] for some progress in this direction).

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/02/10`. Thanks to Liu Xiao Chuan for corrections.

2.11. Compact systems

The primary objective of this lecture and the next few will be to give a proof of the Furstenberg recurrence theorem (Theorem 2.10.3). Along the way we will develop a structural theory for measure-preserving systems.

The basic strategy of Furstenberg's proof is to first prove the recurrence theorems for very simple systems - either those with "almost periodic" (or *compact*) dynamics or with "weakly mixing" dynamics. These cases are quite easy, but don't manage to cover all the cases. To go further, we need to consider various combinations of these systems. For instance, by viewing a general system as an extension of the maximal compact factor, we will be able to prove Roth's theorem (which is equivalent to the $k = 3$ form of the Furstenberg recurrence theorem). To handle the general case, we need to consider compact extensions of compact factors, compact extensions of compact extensions of compact factors, etc., as well as weakly mixing extensions of all the previously mentioned factors.

In this lecture, we will consider those measure-preserving systems $(X, \mathcal{X}, \mu, T)$ which are *compact* or *almost periodic*. These systems are analogous to the equicontinuous or isometric systems in topological dynamics discussed in Section 2.6, and as with those systems, we will be able to characterise such systems (or more precisely, the ergodic ones) algebraically as Kronecker systems, though this is not strictly necessary for the proof of the recurrence theorem.

2.11.1. Almost periodic functions. We begin with a basic definition.

Definition 2.11.1. Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system. A function $f \in L^2(X, \mathcal{X}, \mu)$ is *almost periodic* if the orbit closure $\overline{\{T^n f : n \in \mathbf{Z}\}}$ is compact in $L^2(X, \mathcal{X}, \mu)$.

Example 2.11.2. If f is periodic (i.e. $T^n f = f$ for some $n > 0$) then it is clearly almost periodic. In particular, any shift-invariant function (such as a constant function) is almost periodic.

Example 2.11.3. In the circle shift system $(\mathbf{R}/\mathbf{Z}, x \mapsto x+\alpha)$, every function $f \in L^2(\mathbf{R}/\mathbf{Z})$ is almost periodic, because the orbit closure lies inside the set $\{f(\cdot+\theta) : \theta \in \mathbf{R}/\mathbf{Z}\}$, which is the continuous image of a circle $\mathbf{R}/\mathbf{Z}$ and therefore compact.

Exercise 2.11.1. Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $f \in L^2(X, \mathcal{X}, \mu)$. Show that f is almost periodic in the ergodic theory sense (i.e. Definition 2.11.1 above) if and only if it is almost periodic in the topological dynamical systems sense (see Section 2.3), i.e. if the sets $\{n \in \mathbf{Z} : \|T^n f - f\|_{L^2(X,\mathcal{X},\mu)} \leq \varepsilon\}$ are syndetic for every $\varepsilon > 0$. *Hint*: if f is almost periodic in the ergodic theory sense, show that the orbit closure is an isometric system and thus a Kronecker system, at which point Theorem 2.3.5 can be applied. For the converse implication, use the *Heine-Borel theorem* and the isometric nature of T on L^2.

Exercise 2.11.2. Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system. Show that the space of almost periodic functions in $L^2(X, \mathcal{X}, \mu)$ is a closed shift-invariant subspace which is also closed under the pointwise operations $f, g \mapsto \max(f, g)$ and $f, g \mapsto \min(f, g)$. Similarly, show that the space of almost periodic functions in $L^\infty(X, \mathcal{X}, \mu)$ is a closed subspace which is also an algebra (closed under products) as well as closed under max and min.

Exercise 2.11.3. Show that in any Bernoulli system $\Omega^{\mathbf{Z}}$, the only almost periodic functions are the constants. *Hint*: first show that if $f \in L^2(X, \mathcal{X}, \mu)$ has mean zero, then $\lim_{n\to\infty} \int_X f T^n f \, d\mu = 0$, by first considering elementary functions.

Let us now recall the Furstenberg multiple recurrence theorem, which we now phrase in terms of functions rather than sets:

Theorem 2.11.4 (Furstenberg multiple recurrence theorem)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, let $k \geq 1$, and let $f \in L^\infty(X, \mathcal{X}, \mu)$ be a non-negative function with $\int_X f \, d\mu > 0$. Then we*

have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{r=0}^{N-1} \int_X f T^r f \dots T^{(k-1)r} f > 0.$$

Exercise 2.11.4. Show that Theorem 2.11.4 is equivalent to Theorem 2.10.3.

We can now quickly establish this theorem in the almost periodic case:

Proposition 2.11.5. *Theorem 2.11.4 is true whenever f is almost periodic.*

Proof. Without loss of generality we may assume that f is bounded a.e. by 1. Let $\varepsilon > 0$ be chosen later. Recall from Exercise 2.11.1 that $T^n f$ lies within ε of f in the L^2 topology for a syndetic set of n. For all such n, one also has $\|T^{(j+1)n} f - T^{jn} f\|_{L^2(X,\mathcal{X},\mu)} \leq \varepsilon$ for all j, since T acts isometrically. By the triangle inequality, we conclude that $T^{jn} f$ lies within $O_k(\varepsilon)$ of f in L^2 for $0 \leq j \leq k$. On the other hand, from Hölder's inequality we see that on the unit ball of $L^\infty(X, \mathcal{X}, \mu)$ with the L^2 topology, pointwise multiplication is Lipschitz. Applying this fact repeatedly, we conclude that for n in this syndetic set, $f T^n f \dots T^{(k-1)n} f$ lies within $O_k(\varepsilon)$ in L^2 of f^k. In particular,

$$\int_X f T^n f \dots T^{(k-1)n} f \, d\mu = \int_X f^k \, d\mu + O_k(\varepsilon). \tag{2.103}$$

On the other hand, since $\int f \, d\mu > 0$, we must have $\int_X f^k \, d\mu > 0$. Choosing ε sufficiently small, we thus see that the left-hand side of (2.103) is uniformly bounded away from zero in a syndetic set, and the conclusion of Theorem 2.11.4 follows. □

Remark 2.11.6. Because f lives in a Kronecker system, one can also obtain the above result using various multiple recurrence theorems from topological dynamics, such as Proposition 2.6.10 or the Birkhoff multiple recurrence theorem (Theorem 2.3.4), though to get the full strength of the results, one needs to use either syndetic van der Waerden theorem, see part 3 of Exercise 2.5.15, or the Varnavides averaging trick from the Section 2.10.1. We leave the details to the reader.

2.11.2. Kronecker systems and Haar measure. We have seen how nice almost periodic functions are. Motivated by this, we define

Definition 2.11.7. A measure-preserving system $(X, \mathcal{X}, \mu, T)$ is said to be *compact* if every function in $L^2(X, \mathcal{X}, \mu)$ is almost periodic.

Thus, for instance, by Example 2.11.3, the circle shift system is compact, but from Exercise 2.11.3, no non-trivial Bernoulli system is compact. From Proposition 2.11.5 we know that the Furstenberg recurrence theorem is true for compact systems.

One source of compact systems comes from *Kronecker systems*, as introduced in Definition 2.6.5. As such systems are topological rather than measure-theoretic, we will need to endow them with a canonical measure - *Haar measure* - first.

Let G be a compact metrisable topological group (not necessarily abelian). Without an ambient measure, we cannot yet define the convolution $f * g$ of two continuous functions $f, g \in C(G)$. However, we can define the convolution $\mu * f$ of a finite Borel measure μ on G and a continuous function $f \in C(G)$ to be the function

$$\mu * f(x) := \int_G f(y^{-1}x)\ d\mu(y), \tag{2.104}$$

which (by the uniform continuity of f) is easily seen to be another continuous function. We similarly define

$$f * \mu(x) := \int_G f(xy^{-1})\ d\mu(y). \tag{2.105}$$

Also, one can define the convolution $\mu * \nu$ of two finite Borel measures to be the finite Borel measure defined as

$$\mu * \nu(E) := \int_G \nu(y^{-1} \cdot E)\ d\mu(y) \tag{2.106}$$

for all Borel sets E. For instance, the convolution $\delta_x * \delta_y$ of two Dirac masses is another Dirac mass δ_{xy}. Fubini's theorem tells us that the convolution of two finite measures is another finite measure. Convolution is also bilinear and associative (thus $(\mu * \nu) * \rho = \mu * (\nu * \rho)$, $f * (\mu * \nu) = (f * \mu) * \nu$, $(\mu * f) * \nu = \mu * (f * \nu)$, and $(\mu * \nu) * f = \mu * (\nu * f)$ for measures μ, ν, ρ and continuous f); in particular, left-convolution and right-convolution commute. Also observe that the convolution of two Borel probability measures is again a Borel probability measure.

Convolution also has a powerful *smoothing effect* that can upgrade weak convergence to strong convergence. Specifically, if μ_n converges in the vague sense to μ, and f is continuous, then an easy application of compactness of the underlying group G reveals that $\mu_n * f$ converges in the uniform sense to $\mu * f$.

Let us say that a number c is a *left-mean* (resp. *right-mean*) of a continuous function $f \in C(G)$ if there exists a probability measure μ such that $\mu * f$ (resp. $f * \mu_n$) is equal to a constant c. For compact metrisable groups G, this mean is well defined:

Lemma 2.11.8 (Existence and uniqueness of mean)**.** *Let G be a compact metrisable topological group, and let $f \in C(G)$. Then there exists a unique constant c which is both a left-mean and right-mean of f.*

Proof. Without loss of generality we can take f to be real-valued. Let us first show that there exists a left-mean. Define the oscillation of a real-valued continuous function to be the difference between its maximum and minimum. By the vague sequential compactness of probability measures (Lemma 2.7.16), one can find a probability measure μ which minimises the oscillation of $\mu * f$. If this oscillation is zero, we are done. If the oscillation is non-zero, then (using the compactness of the group and the transitivity of the group action) it is not hard to find a finite number of left rotations of $\mu * f$ whose average has strictly smaller oscillation than that of $\mu * f$ (basically by rotating the places where $\mu * f$ is near its maximum to cover where it is near its mimum). Thus we have a finitely supported probability ν with $\nu * \mu * f$ having smaller oscillation than $\mu * f$, a contradiction. We thus see that a left-mean exists. Similarly, a right-mean exists. But since left-convolution commutes with right-convolution, we see that all left-means are equal to all right-means, and the claim follows. □

The map $f \mapsto c$ from a continuous function to its mean is a bounded non-negative linear functional on $C(G)$ which preserves constants, and thus by the Riesz representation theorem is given by a unique probability measure μ; since left- and right-convolution commute, we see that this measure is both left- and right- invariant. Conversely, given any such measure μ we easily see (again using the

commutativity of left-and right-convolution) that $f * \mu = \mu * f = c$. We have thus shown

Corollary 2.11.9 (Existence and uniqueness of Haar measure)**.** *If G is a compact metrisable topological group, then there exists a unique Borel probability measure μ on G which is both left and right invariant.*

In particular, every topological Kronecker system $(K, x \mapsto x+\alpha)$ can be canonically converted into a measure-preserving system, which is then compact by the same argument used to establish Example 2.11.3. (Actually this observation works for non-abelian Kronecker systems as well as abelian ones.)

Remark 2.11.10. One can also build left- and right- Haar measures for locally compact groups; these measures are locally finite Radon measures rather than Borel probability measures, and are unique up to constants; however it is no longer the case that such measures are necessarily equal to each other except in special cases, such as when the group is abelian or compact. These measures play an important role in the harmonic analysis and representation theory of such groups, but we will not discuss these topics further here.

2.11.3. Classification of compact systems. We have just seen that every Kronecker system is a compact system. The converse is not quite true; consider for instance the disjoint union of two Kronecker systems from different groups (with the probability measure being split, say, $50-50$, between the two components). The situation is similar to that in Section 2.6, in which every Kronecker system was equicontinuous and isometric, but the converse only held under the additional assumption of minimality. There is a similar situation here, but first we need to define the notion of *equivalence* of two measure-preserving systems.

Define an *abstract measure-preserving system* $(\mathcal{X}, \mu, T)$ to be an abstract separable σ-algebra $\mathcal{X}$ (i.e. a Boolean algebra in which every countable sequence has both a supremum and an infimum), together with an abstract probability measure[40] $\mu : \mathcal{X} \to [0,1]$ and an abstract invertible shift $T : \mathcal{X} \to \mathcal{X}$ which preserves the measure μ

[40] An abstract measure space $(\mathcal{X}, \mu)$ is sometimes also known as a *measure algebra*.

(but does not necessarily come from an invertible map $T : X \to X$ on some ambient space). There is an obvious notion of a morphism $\Phi : (\mathcal{X}, \mu, T) \to (\mathcal{Y}, \nu, S)$ between abstract measure-preserving systems, in which $\Phi : \mathcal{Y} \to \mathcal{X}$ (note the contravariance) is a σ-algebra homomorphism with $\nu = \mu \circ \Phi$ and $S \circ \Phi = \Phi \circ T$. This makes the class of abstract measure-preserving systems into a category. In particular we have a notion of two abstract measure-preserving systems being isomorphic.

Example 2.11.11. Let $(X, \mathcal{X}, \mu, T)$ be a skew shift $(y, z) \mapsto (y + \alpha, z + y)$ and let $(Y, \mathcal{Y}, \nu, S)$ be the underlying circle shift $y \mapsto y + \alpha$. These systems are of course non-isomorphic, although there is a factor map $\pi : X \to Y$ which is a morphism. If however we consider the σ-algebra $\pi^\#(\mathcal{Y}) \subset \mathcal{X}$ (which are the Cartesian products of horizontal Borel sets with the vertical circle $\mathbf{R}/\mathbf{Z}$), we see that π induces an isomorphism between the abstract measure-preserving systems $(\pi^\#(\mathcal{Y}), \mu, T)$ and $(\mathcal{Y}, \nu, S)$.

Given a concrete measure-preserving system $(X, \mathcal{X}, \mu, T)$, we can define its *abstraction* $(\mathcal{X}/\sim, \mu, T)$, where $\sim$ is the equivalence relation of almost everywhere equivalence modulo μ. In category theoretic language, abstraction is a covariant functor from the category of concrete measure-preserving systems to the category of abstract measure-preserving systems. We say that two concrete measure-preserving systems are *equivalent* if their abstractions are isomorphic. Thus for instance, in Example 2.11.11 above, $(X, \pi^\#(\mathcal{Y}), \mu, T)$ and $(Y, \mathcal{Y}, \nu, S)$ are equivalent; there is no concrete isomorphism between these two systems, but once one abstracts away the underlying sets X and Y, we can recover an equivalence. As another example, we see that if we add or remove a null set to a measure-preserving system, we obtain an abstractly equivalent measure-preserving system.

Remark 2.11.12. Up to null sets, we can also identify an abstract measure-preserving system $(\mathcal{X}, \mu, T)$ with its commutative *von Neumann algebra* $L^\infty(\mathcal{X}, \mu)$ (which acts on the Hilbert space $L^2(\mathcal{X}, \mu)$ by pointwise multiplication), together with an automorphism T of that algebra; conversely, one can recover the algebra $\mathcal{X}$ as the idempotents 1_E of the von Neumann algebra, and the measure $\mu(E)$ of a set being the trace of the idempotent 1_E. A significant portion of ergodic

theory can in fact be rephrased in terms of von Neumann algebras (which, in particular, naturally suggests a non-commutative generalisation of the subject), although we will not adopt this perspective here.

Many results and notions about concrete measure-preserving systems $(X, \mathcal{X}, \mu, T)$ can be rephrased to not require knowledge of the underlying space X (and to be stable under modification by null sets), and so can be converted to statements about abstract measure-preserving systems; for instance, the Furstenberg recurrence theorem is of this form once one replaces "non-empty" with "positive measure" (see Exercise 2.10.1). The notion of ergodicity is also of this form. In particular, such results and notions automatically become preserved under equivalence. In view of this, the following classification result is of interest:

Theorem 2.11.13 (Classification of ergodic compact systems)**.** *Every ergodic compact system is equivalent to an (abelian) Kronecker system.*

To prove this theorem, it is convenient to use a harmonic analysis approach. Define an *eigenfunction* of a measure-preserving system $(X, \mathcal{X}, \mu, T)$ to be a bounded measurable function f, not a.e. zero, such that $Tf = \lambda f$ a.e..

Let $\mathcal{Z}_1 \subset \mathcal{X}$ denote the σ-algebra generated by all the eigenfunctions. Note that this contains $\mathcal{Z}_0 := \mathcal{X}^T$, which is the σ-algebra generated by the eigenfunctions with eigenvalue 1. We have the following fundamental result:

Proposition 2.11.14 (Description of the almost periodic functions)**.** *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic measure-preserving system, and let $f \in L^2(X, \mathcal{X}, \mu)$. Then f is almost periodic if and only if it lies in $L^2(X, \mathcal{Z}_1, \mu)$, i.e. if it is $\mathcal{Z}_1$-measurable (note that $\mathcal{Z}_1$ contains all null sets of $\mathcal{X}$).*

Remark 2.11.15. One can view $(X, \mathcal{Z}_1, \mu, T)$ as the maximal compact factor of $(X, \mathcal{X}, \mu, T)$, in much the same way that $(X, \mathcal{Z}_0, \mu, T)$ is the maximal factor on which the system is essentially trivial (every function is essentially invariant).

Proof. It is clear that every eigenfunction is almost periodic. From repeated application of Exercise 2.11.2 we conclude that the indicator of any set in $\mathcal{Z}_1$ is also almost periodic, and thus (by more applications of Exercise 2.11.2) every function in $L^2(X, \mathcal{Z}_1, \mu)$ is almost periodic.

Conversely, suppose $f \in L^2(X, \mathcal{X}, \mu)$ is almost periodic. Then the orbit closure $Y_f \subset L^2(X, \mathcal{X}, \mu)$ of f is an isometric system; the orbit of f is clearly dense in Y_f, and thus by isometry the orbit of every other point is also dense. Thus Y_f is minimal, and therefore Kronecker by Proposition 2.6.7; thus we have an isomorphism $\phi : K \to Y_f$ from a group rotation $(K, x \mapsto x+\alpha)$ to Y_f. By rotating if necessary we may assume that $\phi(0) = f$.

By Corollary 2.11.9, K comes with an invariant probability measure ν. The theory of Fourier analysis on compact abelian groups then says that $L^2(K, \nu)$ is spanned by an (orthonormal) basis of characters χ. In particular, the Dirac mass at 0 (the group identity of K) can be expressed as the weak limit of finite linear combinations of such characters.

Now we need to move this information back to X. For this we use the operator $S : L^2(K, \nu) \to L^2(X, \mathcal{X}, \mu)$ defined by $Sh := \int_K \phi(y)h(y)\ d\nu(y)$; one checks from Minkowski's integral inequality that this is a bounded linear map. Because ϕ is a morphism, and each character is an eigenfunction of the group rotation $x \mapsto x + \alpha$, one easily checks that the image $S\chi$ of a character χ is an eigenfunction. Since the image of the Dirac mass is (formally) just f, we thus conclude that f is the weak limit[41] of finite linear combinations of characters. In particular, f is equivalent a.e. to a $\mathcal{Z}_1$-measurable function, as desired. □

Exercise 2.11.5 (Spectral description of Kronecker factor)**.** Show that the product of two eigenfunctions is again an eigenfunction. Using this and Proposition 2.11.14, conclude that $L^2(X, \mathcal{Z}_1, \mu)$ is in fact equal to $\mathbf{H}_{pp}$, the closed subspace of the Hilbert space $\mathbf{H} := L^2(X, \mathcal{X}, \mu)$ generated by the eigenfunctions of the shift operator T.

Exercise 2.11.6. Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $f \in L^2(X, \mathcal{X}, \mu)$. We say that f is quasiperiodic if the orbit

[41] One can in fact use compactness and continuity to make this a strong limit, but this is not necessary here.

$\{T^n f : n \in \mathbf{Z}\}$ lies in a finite-dimensional space. Show that a function is quasiperiodic if and only if it is a finite linear combination of eigenfunctions. Deduce that a function is almost periodic if and only if it is the limit in L^2 of quasiperiodic functions.

Exercise 2.11.7. The purpose of this exercise is to show how abstract measure-preserving systems, and the morphisms between them, can be satisfactorily modeled by concrete systems and morphisms.

(1) Let $(\mathcal{X}, \mu, T)$ be an abstract measure-preserving system. Show that there exists a concrete regular measure-preserving system $(X', \mathcal{X}', \mu', T')$ which is equivalent to $(\mathcal{X}, \mu, T)$ (thus after omitting X' and quotienting out both σ-algebras by null sets, the two resulting abstract measure-preserving systems are isomorphic); the notion of regularity was defined in Definition 2.9.18. *Hint*: take a countable shift-invariant family of sets that generate $\mathcal{X}$ (thus T acts on this space by permutation), and use this to create a σ-algebra morphism from $\mathcal{X}$ to $\mathcal{X}'$, the product σ-algebra of some boolean space $X' := 2^{\mathbf{Z}}$, endowed with a permutation action T'.

(2) Let $\phi : (\mathcal{X}, \mu, T) \to (\mathcal{Y}, \nu, S)$ be an abstract morphism. Show that there exist regular measure-preserving systems $(X', \mathcal{X}', \mu', T')$ and $(Y', \mathcal{Y}', \nu', S')$ equivalent to $(\mathcal{X}, \mu, T)$ and $(\mathcal{Y}, \nu, S)$, together with a concrete morphism $\phi' : X' \to Y'$, such that obvious commuting square connecting the abstract σ-algebras $\mathcal{X}, \mathcal{Y}, \mathcal{X}', \mathcal{Y}'$ quotiented out by null sets does indeed commute.

Remark 2.11.16. Exercise 2.11.7 (and various related results) show that the distinction between concrete and abstract measure-preserving systems are very minor in practice. There are however other areas of mathematics in which taking an abstract or "point-less" approach by deleting (or at least downplaying) the underlying space can lead to non-trivial generalisations or refinements of the original concrete concept, for instance when moving from varieties to schemes.

Proof of Theorem 2.11.13. Note that if f is an eigenfunction then $T|f| = |f|$, and so (if the system is ergodic) $|f|$ is a.e. constant (which implies also that the eigenvalue lies on the unit circle). In particular,

any eigenfunction is invertible. The quotient of two eigenfunctions of the same eigenvalue is then T-invariant and thus constant a.e. by ergodicity, which shows that all eigenspaces have geometric multiplicity 1 modulo null sets. As T is unitary, any eigenfunctions of different eigenvalues are orthogonal to each other; as $L^2(X, \mathcal{X}, \mu)$ is separable, we conclude that the number of eigenfunctions (up to constants and a.e. equivalence) is at most countable.

Let $(\phi_n)_{n\in A}$ be a collection of representative eigenfunctions for some at most countable index set A with eigenvalues λ_n; we can normalise $|\phi_n| = 1$ a.e.. By modifying each eigenfunction on a set of measure zero (cf. Exercise 2.8.7) we can assume that $T\phi_n = \lambda_n\phi_n$ and $|\phi_n| = 1$ *everywhere* rather than just almost everywhere. Then the map $\Phi : x \mapsto (\log \phi_n(x))_{n\in A}$ is a morphism from $(X, \mathcal{X}, \mu, T)$ to the torus $(\mathbf{R}/\mathbf{Z})^A$ with the product σ-algebra $\mathcal{B}$, the push-forward measure $\Phi_\# \mu$, and the shift $x \mapsto x+\alpha$, where $\alpha := (\log \lambda_n)_{n\in A}$. From Proposition 2.11.14 we see that every measurable set in $\mathcal{X}$ differs by a null set from a set in the pullback σ-algebra $\Phi^\#(\mathcal{B})$. From this it is not hard to see that $(X, \mathcal{X}, \mu, T)$ is equivalent to the system $(\mathbf{R}/\mathbf{Z})^A, \mathcal{B}, \Phi_\# \mu, x \mapsto x + \alpha)$.

Now, $\mathbf{R}/\mathbf{Z})^A$ is a compact metrisable space. The orbit closure K of α inside this space is thus also compact metrisable. The support of $\Phi_\# \mu$ is shift-invariant and thus K-invariant; but from the ergodicity of μ we conclude that the support must in fact be a single translate of K. In particular, $\Phi_\# \mu$ is just a translate of Haar measure on K. From this one easily concludes that $(\mathbf{R}/\mathbf{Z})^A, \mathcal{B}, \Phi_\# \mu, x \mapsto x + \alpha)$ is equivalent to the Kronecker system $(K, x \mapsto x + \alpha)$ with the Borel σ-algebra and Haar measure, and the claim follows. □

Exercise 2.11.8. Let $(X, \mathcal{X}, \mu, T)$ be a compact system which is not necessarily ergodic, and let $y \mapsto \mu_y$ be the ergodic decomposition of μ relative to the projection $\pi : (X, \mathcal{X}, \mu, T) \to (Y, \mathcal{Y}, \nu, S)$ given by Proposition 2.9.22. Show that $(X, \mathcal{X}, \mu_y, T)$ is a compact ergodic system for ν-almost every y. From this and Theorem 2.11.13, we conclude that every compact system can be disintegrated into ergodic Kronecker systems (cf. the discussion after Proposition 2.6.9).

Remark 2.11.17. We comment here on finitary versions of the above concepts. Consider the cyclic group system $(\mathbf{Z}/N\mathbf{Z}, x \mapsto x + 1)$ with

the discrete σ-algebra and uniform probability measure. Strictly speaking, every function on this system is periodic with period N and thus almost periodic, and so this is a compact system. But suppose we consider N as a large parameter going to infinity (in which case one can view these systems, together with some function $f = f_N$ on these systems, "converging" to some infinite system with some limit function f, as in the derivation of Theorem 2.10.6 from Theorem 2.10.3. Then we would be interested in *uniform* control on the almost periodicity of the function or the compactness of the system, i.e. quantitative bounds involving expressions such as $O(1)$ which are bounded uniformly in N. With such a perspective, the analogue of a quasiperiodic function (see Exercise 2.11.6) is a function $f : \mathbf{Z}/N\mathbf{Z} \to \mathbf{C}$ which is a linear combination of at most $O(1)$ characters (i.e. its Fourier transform is non-zero at only $O(1)$ frequencies), whilst an almost periodic function f is one which is approximable in L^2 by quasiperiodic functions, thus for every $\varepsilon > 0$ one can find a function with only $O_\varepsilon(1)$ frequencies which lies within ε of f in L^2 norm. Most functions on $\mathbf{Z}/N\mathbf{Z}$ for large N are not like this, and so the cyclic shift system is not compact in the asymptotic limit $N \to \infty$; however if one coarsens the underlying σ-algebra significantly one can recover compactness, though unfortunately one has to replace exact shift-invariance by approximate shift-invariance when one does so. For instance if one considers a σ-algebra $\mathcal{B}$ generated by a bounded (O(1)) number of Bohr sets $\{n \in \mathbf{Z}/N\mathbf{Z} : \|\frac{\xi n}{N} - a\|_{\mathbf{R}/\mathbf{Z}} \leq \varepsilon\}$, then $\mathcal{B}$ is no longer shift-invariant in general, but all the functions which are measurable with respect to this algebra are uniformly almost periodic in the above sense. For some further developments of these sorts of "quantitative ergodic theory" ideas, see [**GrTa2008**], [**GrTa2009a**], [**GrTa2006**], [**Ta2006**], [**Ta2006b**], [**GrTa2009b**], [**Ta2008**].

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/02/11`. Thanks to Emmanuel Kowalski and Liu Xiao Chuan for corrections.

As was pointed out to me anonymously, Theorem 2.12.26 was essentially established by von Neumann and Halmos (more precisely, they showed that any ergodic system in which the spectrum of the shift map is purely discrete is equivalent to a Kronecker system).

It is also possible to construct the Kronecker system explicitly via Pontryagin duality.

2.12. Weakly mixing systems

In Section 2.11, we studied the recurrence properties of compact systems, which are systems in which all measurable functions exhibit almost periodicity - they almost return completely to themselves after repeated shifting. Now, we consider the opposite extreme of *mixing systems* - those in which all measurable functions (of mean zero) exhibit *mixing*[42] - they become orthogonal to themselves after repeated shifting.

We shall see that for weakly mixing systems, averages such as $\frac{1}{N}\sum_{n=0}^{N-1} T^n f \dots T^{(k-1)n} f$ can be computed very explicitly (in fact, this average converges to the constant $(\int_X f\, d\mu)^{k-1}$). More generally, we shall see that weakly mixing components of a system tend to average themselves out and thus become irrelevant when studying many types of ergodic averages. Our main tool here will be the humble Cauchy-Schwarz inequality, and in particular a certain consequence of it, known as the *van der Corput lemma*.

As one application of this theory, we will be able to establish Roth's theorem[**Ro1953**] (the $k = 3$ case of Szemerédi's theorem).

2.12.1. Mixing functions. Much as compact systems were characterised by their abundance of almost periodic functions, we will characterise mixing systems by their abundance of mixing functions (this is not standard terminology). To define and motivate this concept, it will be convenient to introduce a weak notion of convergence (this notation is also not standard):

Definition 2.12.1 (Cesáro convergence)**.** A sequence c_n in a normed vector space is said to *converge in the Cesáro sense* to a limit c if the averages $\frac{1}{N}\sum_{n=0}^{N-1} c_n$ converge strongly to c, in which case we write $C\!-\!\lim_{n\to\infty} c_n = c$. We also write $C\!-\!\sup_{n\to\infty} c_n := \limsup_{N\to\infty} \|\frac{1}{N}\sum_{n=0}^{N-1} c_n\|$ (thus $C\!-\!\lim_{n\to\infty} c_n = 0$ if and only if $C\!-\!\sup_{n\to\infty} c_n = 0$).

[42] Actually, there are two different types of mixing, *strong mixing* and *weak mixing*, depending on whether the orthogonality occurs individually or on the average; it is the latter concept which is of more importance to the task of establishing the Furstenberg recurrence theorem.

Example 2.12.2. The sequence $0, 1, 0, 1, \ldots$ has a Cesáro limit of $1/2$.

Exercise 2.12.1. Let c_n be a bounded sequence of *non-negative* numbers. Show that the following three statements are equivalent:

(1) $C-\lim_{n\to\infty} c_n = 0$.

(2) $C-\lim_{n\to\infty} |c_n|^2 = 0$.

(3) c_n converges to zero in density[43].

Which of the implications between 1, 2, 3 remain valid if c_n is not bounded?

Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $f \in L^2(X, \mathcal{X}, \mu)$ be a function. We consider the *correlation coefficients* $\langle T^n f, f\rangle := \int_X T^n f \overline{f}\, d\mu$ as n goes to infinity. Note that we have the symmetry $\langle T^n f, f\rangle = \overline{\langle T^{-n} f, f\rangle}$, so we only need to consider the case when n is positive. The mean ergodic theorem (Corollary 2.8.16) tells us the Cesáro behaviour of these coefficients. Indeed, we have

(2.107)
$$C-\lim_{n\to\infty} \langle T^n f, f\rangle = \langle \mathbf{E}(f|\mathcal{X}^T), f\rangle = \|\mathbf{E}(f|\mathcal{X}^T)\|^2_{L^2(X,\mathcal{X},\mu)}$$

where $\mathcal{X}^T$ is the σ-algebra of essentially shift-invariant sets. In particular, if the system is ergodic, and f has mean zero (i.e. $\int_X f\, d\mu = 0$), then we have

$$C-\lim_{n\to\infty} \langle T^n f, f\rangle = 0, \tag{2.108}$$

thus the correlation coefficients go to zero in the Cesáro sense. However, this does not necessarily imply that these coefficients go to zero pointwise. For instance, consider a circle shift system $(\mathbf{R}/\mathbf{Z}, x \mapsto x + \alpha)$ with α irrational (and with uniform measure), thus this system is ergodic by Exercise 2.9.5. Then the function $f(x) := e^{2\pi i x}$ has mean zero, but one easily computes that $\langle T^n f, f\rangle = e^{2\pi i n\alpha}$. The coefficients $e^{2\pi i n\alpha}$ converge in the Cesáro sense to zero, but have magnitude 1 and thus do not converge to zero pointwise.

---

[43] We say c_n *converges in density* to c if for any $\varepsilon > 0$, the set $\{n \in \mathbf{N} : |c_n - c| > \varepsilon\}$ has upper density zero.

Definition 2.12.3 (Mixing)**.** Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system. A function $f \in L^2(X, \mathcal{X}, \mu)$ is *strongly mixing* if $\lim_{n\to\infty} \langle T^n f, f\rangle = 0$, and *weakly mixing* if $C\text{-}\lim_{n\to\infty} |\langle T^n f, f\rangle| = 0$.

Remark 2.12.4. Clearly strong mixing implies weak mixing. From (2.107) we also see that if f is weakly mixing, then $\mathbf{E}(f|\mathcal{X}^T)$ must vanish a.e..

Exercise 2.12.2. Show that if f is both almost periodic and weakly mixing, then it must be 0 almost everywhere. In particular, in a compact system, the only weakly mixing function is 0 (up to a.e. equivalence).

Exercise 2.12.3. In any Bernoulli system $\Omega^{\mathbf{Z}}$ with the product σ-algebra and a product measure, and the standard shift, show that any function of mean zero is strongly mixing. *Hint*: first do this for functions that depend on only finitely many of the variables.

Exercise 2.12.4. Consider a skew shift system $((\mathbf{R}/\mathbf{Z})^2, (x, y) \mapsto (x+\alpha, y+x))$ with the usual Lebesgue measure and Borel σ-algebra, and with α irrational. Show that the function $f(x, y) := e^{2\pi i x}$ is neither strongly mixing nor weakly mixing, but that the function $g(x, y) := e^{2\pi i y}$ is both strongly mixing and weakly mixing.

Exercise 2.12.5. Let $X := \mathbf{C}^{\mathbf{Z}}$ be given the product Borel σ-algebra $\mathcal{X}$ and the shift $T : (z_n)_{n\in\mathbf{Z}} \to (z_{n+1})_{n\in\mathbf{Z}}$. For each $d \geq 1$, let μ_d be the probability distribution in X of the random sequence $(z_n)_{n\in\mathbf{Z}}$ given by the rule

$$z_n := \frac{1}{2^{d/2}} \sum_{\omega_1,\ldots,\omega_d \in \{0,1\}} w_{\omega_1,\ldots,\omega_d} e^{2\pi i \sum_{j=1}^d \omega_j n/100^j}, \tag{2.109}$$

where the $w_{\omega_1,\ldots,\omega_d}$ are iid standard complex Gaussians (thus each w has probability distribution $e^{-\pi|w|^2}\, dw$). Show that each μ_d is shift invariant. If μ is a vague limit point of the sequence μ_d, and $f : X \to \mathbf{C}$ is the function defined as $f((z_n)_{n\in\mathbf{Z}}) := \mathrm{sgn}(\mathrm{Re} z_0)$, show that f is weakly mixing but not strongly mixing (and more specifically, that $\langle T^{100^j} f, f\rangle$ stays bounded away from zero) with respect to the system $(X, \mathcal{X}, \mu, T)$.

Remark 2.12.5. Exercise 2.12.5 illustrates an important point, namely that *stationary processes* yield a rich source of measure-preserving

systems (indeed the two notions are almost equivalent in some sense, especially after one distinguishes a specific function f on the measure-preserving system). However, we will not adopt this more probabilistic perspective to ergodic theory here.

Remark 2.12.6. We briefly discuss the finitary analogue of the weak mixing concept in the context of functions $f : \mathbf{Z}/N\mathbf{Z} \to \mathbf{C}$ on a large cyclic group $\mathbf{Z}/N\mathbf{Z}$ with the usual shift $x \mapsto x+1$. Then one can compute

$$C-\lim_{n\to\infty} |\langle T^n f, f\rangle|^2 = \sum_{\xi \in \mathbf{Z}/N\mathbf{Z}} |\hat{f}(\xi)|^4 \tag{2.110}$$

where $\hat{f}(\xi) := \frac{1}{N}\sum_{x\in\mathbf{Z}/N\mathbf{Z}} f(x) e^{-2\pi i x\xi/N}$ are the Fourier coefficients of f. Comparing this against the Plancherel identity $\|f\|_{L^2}^2 = \sum_{\xi\in\mathbf{Z}/N\mathbf{Z}} |\hat{f}(\xi)|^2$ we thus see that a function f bounded in L^2 norm should be considered "weakly mixing" if it has no large Fourier coefficients. Contrast this with Remark 2.11.17.

Now let us see some consequences of the weak mixing property. We need the following lemma, which gives a useful criterion as to whether a sequence of bounded vectors in a Hilbert space converges in the Cesáro sense to zero.

Lemma 2.12.7 (van der Corput lemma)**.** *Let $v_1, v_2, v_3, \ldots$ be a bounded sequence of vectors in a Hilbert space H, If*

$$C-\lim_{h\to\infty} C-\sup_{n\to\infty} \langle v_n, v_{n+h}\rangle = 0 \tag{2.111}$$

then $C-\lim_{n\to\infty} v_n = 0$.

Informally, this lemma asserts that if each vector in a bounded sequence tends to be orthogonal to nearby elements in that sequence, then the vectors will converge to zero in the Cesáro sense. This formulation of the lemma is essentially the version in [**Be1987**], except that we have made the minor change of replacing one of the Cesáro limits with a Cesáro supremum.

Proof. We can normalise so that $\|v_n\| \leq 1$ for all n. In particular, we have $v_n = O(1)$, where $O(1)$ denotes a vector of bounded magnitude.

For any h and $N \geq 1$, we thus have the telescoping identity

$$\frac{1}{N}\sum_{n=0}^{N-1} v_{n+h} = \frac{1}{N}\sum_{n=0}^{N-1} v_n + O(|h|/N); \tag{2.112}$$

averaging this over all h from 0 to $H-1$ for some $H \geq 1$, we obtain

$$\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{H}\sum_{h=0}^{H-1} v_{n+h} = \frac{1}{N}\sum_{n=0}^{N-1} v_n + O(H/N); \tag{2.113}$$

by the triangle inequality we thus have

$$\|\frac{1}{N}\sum_{n=0}^{N-1} v_n\| \leq \frac{1}{N}\sum_{n=0}^{N-1}\|\frac{1}{H}\sum_{h=0}^{H-1} v_{n+h}\| + O(H/N) \tag{2.114}$$

where the $O()$ terms are now scalars rather than vectors. We square this (using the crude inequality $(a+b)^2 \leq 2a^2 + 2b^2$) and apply Cauchy-Schwarz to obtain

$$\|\frac{1}{N}\sum_{n=0}^{N-1} v_n\|^2 \leq O(\frac{1}{N}\sum_{n=0}^{N-1}\|\frac{1}{H}\sum_{h=0}^{H-1} v_{n+h}\|^2) + O(H^2/N^2) \tag{2.115}$$

which we rearrange as

$$\|\frac{1}{N}\sum_{n=0}^{N-1} v_n\|^2 \leq O(\frac{1}{H^2}\sum_{0\leq h,h'<H}\frac{1}{N}\sum_{n=0}^{N-1}\langle v_{n+h}, v_{n+h'}\rangle) + O(H^2/N^2). \tag{2.116}$$

We take limits as $N \to \infty$ (keeping H fixed for now) to conclude

$$\limsup_{N\to\infty}\|\frac{1}{N}\sum_{n=0}^{N-1} v_n\|^2 \leq O(\frac{1}{H^2}\sum_{0\leq h,h'<H} C\text{-}\sup_{n\to\infty}\langle v_{n+h}, v_{n+h'}\rangle). \tag{2.117}$$

Another telescoping argument (and symmetry) gives us

$$C\text{-}\sup_{n\to\infty}\langle v_{n+h}, v_{n+h'}\rangle = C\text{-}\sup_{n\to\infty}\langle v_{n+|h-h'|}, v_n\rangle \tag{2.118}$$

and so

$$\limsup_{N\to\infty}\|\frac{1}{N}\sum_{n=0}^{N-1} v_n\|^2 \leq O(\frac{1}{H}\sum_{0\leq h<H} C\text{-}\sup_{n\to\infty}\langle v_{n+h}, v_n\rangle). \tag{2.119}$$

Taking limits as $H \to \infty$ and using (2.111) we obtain the claim. □

Exercise 2.12.6. Let $P : \mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ be a polynomial with at least one irrational non-constant coefficient. Using Lemma 2.12.7 (in the scalar case $H = \mathbf{C}$) and an induction on degree, show that $C-\lim_{n\to\infty} e^{2\pi i P(n)} = 0$. Conclude that the sequence $(P(n))_{n\in\mathbf{N}}$ is uniformly distributed with respect to uniform measure (see Definition 2.9.12 for a definition of uniform distribution).

Exercise 2.12.7. Using Exercise 2.12.6, give another proof of Theorem 2.6.26.

We now apply the van der Corput lemma to weakly mixing functions.

Corollary 2.12.8. *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $f \in L^2(X, \mathcal{X}, \mu)$ be weakly mixing. Then for any $g \in L^2(X, \mathcal{X}, \mu)$ we have $C-\lim_{n\to\infty} |\langle T^n f, g\rangle| = 0$ and $C-\lim_{n\to\infty} |\langle f, T^n g\rangle| = 0$.*

Proof. We just prove the first claim, as the second claim is similar. By Exercise 2.12.1, it suffices to show that

$$\frac{1}{N}\sum_{n=0}^{N-1} |\langle T^n f, g\rangle|^2 \to 0 \tag{2.120}$$

as $N \to \infty$. The left-hand side can be rewritten as

$$\langle \frac{1}{N}\sum_{n=0}^{N-1} \langle g, T^n f\rangle T^n f, g\rangle \tag{2.121}$$

so by Cauchy-Schwarz it suffices to show that

$$C-\lim_{N\to\infty} \langle g, T^n f\rangle T^n f = 0. \tag{2.122}$$

Applying the van der Corput lemma and discarding the bounded coefficients $\langle g, T^n f\rangle$, it suffices to show that

$$C-\lim_{H\to\infty} C-\sup_{n\to\infty} |\langle T^{n+h} f, T^n f\rangle| = 0. \tag{2.123}$$

But $\langle T^{n+h} f, T^n f\rangle = \langle T^h f, f\rangle$, and the claim now follows from the weakly mixing nature of f. □

2.12.2. Weakly mixing systems. Now we consider systems which are full of mixing functions.

Definition 2.12.9. (Mixing systems) A measure-preserving system $(X, \mathcal{X}, \mu, T)$ is *weakly mixing* (resp. *strongly mixing*) if every function $f \in L^2(X, \mathcal{X}, \mu)$ with mean zero is weakly mixing (resp. strongly mixing).

Example 2.12.10. From Exercise 2.12.2, we know that any system with a non-trivial Kronecker factor is not weakly mixing (and thus not strongly mixing). On the other hand, from Exercise 2.12.3, we know that any Bernoulli system is strongly mixing (and thus weakly mixing also). From Remark 1 we see that any strongly or weakly mixing system must be ergodic.

Exercise 2.12.8. Show that the system in Exercise 2.12.5 is weakly mixing but not strongly mixing.

Here is another characterisation of weak mixing:

Exercise 2.12.9. Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. Show that the following are equivalent:

(1) $(X, \mathcal{X}, \mu, T)$ is weakly mixing.

(2) For every $f, g \in L^2(X, \mathcal{X}, \mu)$, $\langle T^n f, g\rangle$ converges in density to $(\int_X f \, d\mu)(\int_X \overline{g} \, d\mu)$. (See Exercise 2.12.1 for a definition of convergence in density.)

(3) For any measurable E, F, $\mu(T^n E \cap F)$ converges in density to $\mu(E)\mu(F)$.

(4) The product system $(X \times X, \mathcal{X} \times \mathcal{X}, \mu \times \mu, T \times T)$ is ergodic.

Hint: To equate 1 and 2, use the decomposition $f = (f - \int_X f \, d\mu) + \int_X f \, d\mu$ of a function into its mean and mean-free components. To equate 2 and 4, use the fact that the space $L^2(X \times X, \mathcal{X} \times \mathcal{X}, \mu \times \mu)$ is spanned (in the topological vector space sense) by tensor products $(x, y) \mapsto f(x)g(y)$ with $f, g \in L^2(X, \mathcal{X}, \mu)$.

Exercise 2.12.10. Show that the equivalences between 1, 2, 3 in Exercise 2.12.9 remain if "weak mixing" and "converges in density" are replaced by "strong mixing" and "converges" respectively.

Exercise 2.12.11. Let $(X, \mathcal{F}, T)$ be any minimal topological system with Borel σ-algebra $\mathcal{B}$, and let μ be a shift invariant Borel probability measure. Show that if $(X, \mathcal{B}, \mu, T)$ is weakly mixing (resp. strongly mixing), then $(X, \mathcal{F}, T)$ is topologically weakly mixing (resp. topologically mixing), as defined in Definition 2.12.9 and Exercise 2.7.12.

Exercise 2.12.12. If $(X, \mathcal{X}, \mu, T)$ is weakly mixing, show that $(X, \mathcal{X}, \mu, T^n)$ is weakly mixing for any non-zero n.

Exercise 2.12.13. Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. Show that the following are equivalent:

(1) $(X, \mathcal{X}, \mu, T)$ is weakly mixing.

(2) Whenever $(Y, \mathcal{Y}, \nu, S)$ is ergodic, the product system $(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu, T \times S)$ is ergodic.

Hint: To obtain 1 from 2, use Exercise 2.12.9. To obtain 2 from 1, repeat the *methods* used to prove Exercise 2.12.9.

Exercise 2.12.14. Show that the product of two weakly mixing systems is again weakly mixing. *Hint*: use Exercises 2.12.9 and 2.12.13.

Now we come to an important type of observation for the purposes of establishing the Furstenberg recurrence theorem: in weakly mixing systems, functions of mean zero are negligible as far as multiple averages are concerned.

Proposition 2.12.11. *Let $a_1, \ldots, a_k \in \mathbf{Z}$ be distinct non-zero integers for some $k \geq 1$. Let $(X, \mathcal{X}, \mu, T)$ be weakly mixing, and let $f_1, \ldots, f_k \in L^\infty(X, \mathcal{X}, \mu)$ be such that at least one of $f_1, \ldots, f_k$ has mean zero. Then we have*

$$C\text{-}\lim_{n\to\infty} T^{a_1 n} f_1 \ldots T^{a_k n} f_k = 0 \tag{2.124}$$

in $L^2(X, \mathcal{X}, \mu)$.

Proof. We induct on k. When $k = 1$ the claim follows from the mean ergodic theorem and Exercise 2.12.12 (recall from Example 2.12.10 that all weakly mixing systems are ergodic).

Now let $k \geq 2$ and suppose that the claim has already been proven for $k-1$. Without loss of generality we may assume that it is f_1 which

has mean zero. Applying the van der Corput lemma (Lemma 2.12.7), it suffices to show that

$$C-\sup_{n\to\infty}\langle T^{a_1(n+h)}f_1\dots T^{a_k(n+h)}f_k, T^{a_1 n}f_1\dots T^{a_k n}f_k\rangle \tag{2.125}$$

converges in density to zero as $h\to\infty$. But the left-hand side can be rearranged as

$$C-\sup_{n\to\infty}\int_X T^{(a_1-a_k)n}f_{1,h}\dots T^{(a_{k-1}-a_k)n}f_{k-1,h}f_{k,h}\,d\mu \tag{2.126}$$

where $f_{j,h}:=T^{a_j h}f_j\overline{f_j}$. Applying Cauchy-Schwarz, it suffices to show that

$$C-\sup_{n\to\infty}T^{(a_1-a_k)n}f_{1,h}\dots T^{(a_{k-1}-a_k)n}f_{k-1,h} \tag{2.127}$$

converges in density to zero as $h\to\infty$.

Since $(X,\mathcal{X},\mu,T)$ is weakly mixing, the mean-zero function f_1 is weakly mixing, and so the mean of $f_{1,h}$ goes to zero in density as $h\to\infty$. As all functions are assumed to be bounded, we can thus subtract the mean from $f_{1,h}$ in (2.127) without affecting the desired conclusion, leaving behind the mean-zero component $f_{1,h}-\int_X f_{1,h}\,d\mu$. But then the contribution of this expression to (2.127) vanishes by the induction hypothesis. □

Remark 2.12.12. The key point here was that functions f of mean zero were weakly mixing and thus had the property that $T^h f\overline{f}$ almost had mean zero, and were thus almost weakly mixing. One could iterate this further to investigate the behaviour of "higher derivatives" of f such as $T^{h+h'}f\overline{T^h f T^{h'}f}f$. Pursuing this analysis further leads to the Gowers-Host-Kra seminorms[**HoKr2005**], which are closely related to the Gowers uniformity norms[**Go2001**] in additive combinatorics.

Corollary 2.12.13. *Let $a_1,\dots,a_k\in\mathbf{Z}$ be distinct integers for some $k\geq 1$, let $(X,\mathcal{X},\mu,T)$ be a weakly mixing system, and let $f_1,\dots,f_k\in L^\infty(X,\mathcal{X},\mu)$. Then $\int_X T^{a_1 n}f_1\dots T^{a_k n}f_k\,d\mu$ converges in the Cesáro sense to $(\int_X f_1\,d\mu)\dots(\int_X f_k\,d\mu)$.*

Note in particular that this establishes the Furstenberg recurrence theorem (Theorem 2.11.4) in the case of weakly mixing systems.

Proof. We again induct on k. The $k = 1$ case is trivial, so suppose $k > 1$ and the claim has already been proven for $k - 1$. If any of the functions f_j is constant then the claim follows from the induction hypothesis, so we may subtract off the mean from each function and suppose that all functions have mean zero. By shift-invariance we may also fix a_k (say) to be zero. The claim now follows from Proposition 2.12.11 and Cauchy-Schwarz. □

Exercise 2.12.15. Show that the Cesáro convergence in Corollary 2.12.8 can be strengthened to convergence in density. *Hint*: first reduce to the mean zero case, then apply Exercise 2.12.14 to work with the product system instead.

Exercise 2.12.16. Let $(X, \mathcal{X}, \mu, T)$ be a weakly mixing system, and let $f \in L^\infty(X, \mathcal{X}, \mu)$ have mean zero. Show that $T^{n^2} f$ converges in the Cesáro sense in $L^2(X, \mathcal{X}, \mu)$ to zero. *Hint*: use van der Corput and Proposition 2.12.11 or Corollary 2.12.13.)

Exercise 2.12.17. Show that Corollary 2.12.13 continues to hold if the linear polynomials $a_1 n, \ldots, a_k n$ are replaced by arbitrary polynomials $P_1(n), \ldots, P_k(n)$ from the integers to the integers, so long as the difference between any two of these polynomials is non-constant. *Hint*: you will need the "PET induction" machinery from Exercise 2.5.3. This result was first established in [**Be1987**].

2.12.3. Hilbert-Schmidt operators. We have now established the Furstenberg recurrence theorem for two distinct types of systems: compact systems and weakly mixing systems. From Example 2.12.10 we know that these systems are indeed quite distinct from each other. Here is another indication of "distinctness":

Exercise 2.12.18. In any measure-preserving system $(X, \mathcal{X}, \mu, T)$, show that almost periodic functions and weakly mixing functions are always orthogonal to each other.

On the other hand, there are certainly systems which are neither weakly mixing nor compact (e.g. the skew shift). But we have the following important dichotomy (cf. Theorem 2.7.12):

Theorem 2.12.14. *Suppose that $(X, \mathcal{X}, \mu, T)$ is a measure-preserving system. Then exactly one of the following statements is true:*

(1) *(Structure)* $(X, \mathcal{X}, \mu, T)$ *has a non-trivial compact factor*[44].

(2) *(Randomness)* $(X, \mathcal{X}, \mu, T)$ *is weakly mixing.*

In Example 2.12.10 we have already shown that 1 and 2 cannot be both true; the tricky part is to show that lack of weak mixing implies a non-trivial compact factor.

In order to prove this result, we recall some standard results about Hilbert-Schmidt operators on a separable[45] Hilbert space. We begin by recalling the notion of tensor product of two Hilbert spaces:

Proposition 2.12.15. *Let H, H' be two separable Hilbert spaces. Then there exists another separable Hilbert space $H \otimes H'$ and a bilinear tensor product map $\otimes : H \times H' \to H \otimes H'$ such that*

$$\langle v \otimes v', w \otimes w' \rangle_{H \otimes H'} = \langle v, w \rangle_H \langle v', w' \rangle_{H'} \tag{2.128}$$

for all $v, w \in H$ and $v', w' \in H'$. Furthermore, the tensor products $(e_n \otimes e'_{n'})_{n \in A, n' \in A'}$ between any orthonormal bases $(e_n)_{n \in A}$, $(e'_{n'})_{n' \in A'}$ of H and H' respectively, form an orthonormal basis of $H \otimes H'$.

It is easy to see that $H \otimes H'$ is unique up to isomorphism, and so we shall abuse notation slightly and refer to $H \otimes H'$ as **the** tensor product of H and H'.

Proof. Take any orthonormal bases $(e_n)_{n \in A}$ and $(e'_{n'})_{n' \in A'}$ of H and H' respectively, and let $H \otimes H'$ be the Hilbert space generated by declaring the formal quantities $e_n \otimes e'_{n'}$ to be an orthonormal basis. If one then defines

$$(\sum_n c_n e_n) \otimes (\sum_{n'} c'_{n'} e_{n'}) := \sum_n \sum_{n'} c_n c'_{n'} e_n \otimes e_{n'} \tag{2.129}$$

for all square-summable sequences c_n and $c'_{n'}$, one easily verifies that $\otimes$ is indeed a bilinear map that obeys (2.128). in particular, if $(f_m)_{m \in B}$ and $(f'_{m'})_{m' \in B'}$ are some other orthonormal bases of H, H' respectively, then from (2.128) $(f_m \otimes f'_{m'})_{m \in B, m' \in B'}$ is an orthonormal

[44] [In ergodic theory, a *factor* of a measure-preserving system is simply a morphism from that system to some other measure-preserving system. Unlike the case with topological dynamics, we do not need to assume surjectivity of the morphism, since in the measure-theoretic setting, the image of a morphism always has full measure.

[45] As usual, the hypothesis of separability is not absolutely essential, but is convenient to assume throughout; for instance, it assures that orthonormal bases always exist and are at most countable.

set, and one can approximate any element $e_n \otimes e'_{n'}$ in the original orthonormal basis to arbitrary accuracy by linear combinations from this orthonormal set, and so this set is in fact an orthonormal basis as required. □

Example 2.12.16. The tensor product of $L^2(X, \mathcal{X}, \mu)$ and $L^2(Y, \mathcal{Y}, \nu)$ is $L^2(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu)$, with the tensor product operation $f \otimes g(x, y) := f(x)g(y)$. The tensor product of $\mathbf{C}^m$ and $\mathbf{C}^n$ is $\mathbf{C}^{n \times m}$, which can be thought of as the Hilbert space of $n \times m$ (or $m \times n$) matrices, with the inner product $\langle A, B \rangle := \mathrm{tr}(AB^\dagger) = \mathrm{tr}(A^\dagger B)$.

Given a Hilbert space H, define its *complex conjugate* $\overline{H}$ to be the same set as H, but with the conjugated scalar multiplication structure $z, v \mapsto \overline{z}v$ and the conjugated inner product $\langle z, w \rangle_{\overline{H}} := \overline{\langle z, w \rangle_H} = \langle w, z \rangle_H$, but with all other structures unchanged. This is also a Hilbert space[46].

Example 2.12.17. The conjugation map $f \mapsto \overline{f}$ is a Hilbert space isometry between the Hilbert space $L^2(X, \mathcal{X}, \mu)$ and its complex conjugate.

Every element $K \in \overline{H} \otimes H'$ induces a bounded linear operator $T_K : H \to H'$, defined via duality by the formula

$$\langle T_K v, v' \rangle_{H'} := \langle K, v \otimes v' \rangle \tag{2.130}$$

for all $v \in H, v' \in H'$. We refer to K as the *kernel* of T_K. Any operator $T = T_K$ that arises in this manner is called a *Hilbert-Schmidt operator* from H to H'. The Hilbert space structure on the space $\overline{H} \otimes H'$ of kernels induces an analogous Hilbert space structure on the Hilbert-Schmidt operators, leading to the Hilbert-Schmidt norm $\|T\|_{HS}$ and inner product $\langle S, T \rangle_{HS}$ for such operators. Here are some other characterisations of this concept:

Exercise 2.12.19. Let H, H' be Hilbert spaces with orthonormal bases $(e_n)_{n \in A}$ and $(e'_{n'})_{n' \in A'}$ respectively, and let $T : H \to H'$ be a bounded linear operator. Show that the following are equivalent:

(1) T is a Hilbert-Schmidt operator.

[46] Of course, for real Hilbert spaces rather than complex, the notion of complex conjugation is trivial.

(2) $\sum_{n\in A} \|Te_n\|_{H'}^2 < \infty$.
(3) $\sum_{n\in A} \sum_{n'\in A'} |\langle Te_n, e'_{n'}\rangle_{H'}|^2 < \infty$.

Also, show that if $T, S : H \to H'$ are Hilbert-Schmidt operators, then

$$\langle T, S\rangle_{HS} = \sum_{n\in A} \langle Te_n, Se_n\rangle_{H'} \tag{2.131}$$

and

$$\|T\|_{HS}^2 = \sum_{n\in A} \|Te_n\|_{H'}^2 = \sum_{n\in A} \sum_{n'\in A'} |\langle Te_n, e'_{n'}\rangle_{H'}|^2 \tag{2.132}$$

As one consequence of the above exercise, we see that the Hilbert-Schmidt norm controls the operator norm, thus $\|Tv\| \leq \|T\|_{HS}\|v\|$ for all vectors v.

Remark 2.12.18. From this exercise and Fatou's lemma, we see in particular that the limit (in either the norm, strong or weak operator topologies) of a sequence of Hilbert-Schmidt operators with uniformly bounded Hilbert-Schmidt norm, is still Hilbert-Schmidt. We also see that the composition of a Hilbert-Schmidt operator with a bounded operator is still Hilbert-Schmidt (thus the Hilbert-Schmidt operators can be viewed as a closed two-sided ideal in the space of bounded operators).

Example 2.12.19. An operator $T : L^2(X, \mathcal{X}, \mu) \to L^2(Y, \mathcal{Y}, \nu)$ is Hilbert-Schmidt if and only if it takes the form $Tf(y) := \int_X K(x, y)f(x)\, d\mu(x)$ for some kernel $K \in L^2(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu)$, in which case the Hilbert-Schmidt norm is $\|K\|_{L^2(X\times Y, \mathcal{X}\times\mathcal{Y}, \mu\times\nu)}$. The Hilbert-Schmidt inner product is defined similarly.

Example 2.12.20. The identity operator on an infinite-dimensional Hilbert space is never Hilbert-Schmidt, despite being bounded. On the other hand, every finite rank operator is Hilbert-Schmidt.

One of the key properties of Hilbert-Schmidt operators which will be relevant to us is the following.

Lemma 2.12.21. *If $T : H \to H'$ is Hilbert-Schmidt, then it is compact (i.e. the image of any bounded set is precompact).*

Proof. Let $\varepsilon > 0$ be arbitrary. By Exercise 2.12.19 and monotone convergence, we can find a finite orthonormal set $e_1, \ldots, e_N$ such that $\sum_{n=1}^N \|Te_n\|_{H'}^2 \geq \|T\|_{HS}^2 - \varepsilon^2$, and in particular that $\|Te_{n+1}\|_{H'} \leq \varepsilon$ for any e_{n+1} orthogonal to $e_1, \ldots, e_n$. As a consequence, the image of the unit ball of H under T lies within ε of the image of the unit ball of the finite-dimensional space $\mathrm{span}(e_1, \ldots, e_N)$. This image is therefore totally bounded and thus precompact. □

The following exercise may help illuminate the distinction between bounded operators, Hilbert-Schmidt operators, and compact operators:

Exercise 2.12.20. Let λ_n be a sequence of complex numbers, and consider the diagonal operator $T : (z_n)_{n\in\mathbf{N}} \mapsto (\lambda_n z_n)_{n\in\mathbf{N}}$ on $l^2(\mathbf{N})$.

(1) Show that T is a well-defined bounded linear operator on $l^2(\mathbf{N})$ if and only if the sequence (λ_n) is bounded.

(2) Show that T is Hilbert-Schmidt if and only if the sequence (λ_n) is square-summable.

(3) Show that T is compact if and only if the sequence (λ_n) goes to zero as $n \to \infty$.

Now we apply the above theory to establish Theorem 2.12.14. Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $f \in L^2(X, \mathcal{X}, \mu)$. The rank one operators $g \mapsto \langle g, T^n f\rangle T^n f$ can easily be verified to have a Hilbert-Schmidt norm of $\|f\|_{L^2}^2$, and so by the triangle inequality, their averages $S_{f,N} : g \mapsto \frac{1}{N}\sum_{n=0}^{N-1} \langle g, T^n f\rangle T^n f$ have a Hilbert-Schmidt norm of at most $\|f\|_{L^2}^2$. On the other hand, from the identity

$$\langle S_{f,N} g, h\rangle = \frac{1}{N}\sum_{n=0}^{N-1} \langle g \otimes \overline{h}, (T\otimes T)^n (f \otimes \overline{f})\rangle \tag{2.133}$$

and the mean ergodic theorem (applied to the product space) we see that $S_{f,N}$ converges in the weak operator topology[47] to some limit S_f, which is then also Hilbert-Schmidt by Remark 2.12.18, and

[47] Actually, $S_{f,N}$ converges to S_f in the Hilbert-Schmidt norm, and thus also in the operator norm and in the strong topology: this is another application of the mean ergodic theorem, which we leave as an exercise. Since each of the $S_{f,N}$ is clearly finite rank, this gives a direct proof of the compactness of S_f.

thus compact by Lemma 2.12.21. Also, it is easy to see that S_f is self-adjoint and commutes with T. As a consequence, we conclude that for any $g \in L^2(X, \mathcal{X}, \mu)$, the image $S_f g$ is almost periodic (since $\{T^n S_f g : n \in \mathbf{Z}\} = S_f\{T^n g : n \in \mathbf{Z}\}$ is the image of a bounded set by the compact operator S_f and therefore precompact).

On the other hand, observe that

$$\langle S_f f, f\rangle = C - \lim_{n\to\infty} |\langle T^n f, f\rangle|^2. \tag{2.134}$$

Thus by Definition 2.12.3 (and Exercise 2.12.1), we see that $\langle S_f f, f\rangle \neq 0$ whenever f is not weakly mixing. In particular, f is not orthogonal to the almost periodic function $S_f f$. From this and Exercise 2.12.18, we have thus shown

Proposition 2.12.22 (Dichotomy between structure and randomness)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system. A function $f \in L^2(X, \mathcal{X}, \mu)$ is weakly mixing if and only if it is orthogonal to all almost periodic functions (or equivalently, orthogonal to all eigenfunctions).*

Remark 2.12.23. Interestingly, essentially the same result appears in the spectral and scattering theory of linear Schrödinger equations, which in that context is known as the "RAGE theorem" [**Ru1969**], [**AmGe1973**], [**En1978**].

Remark 2.12.24. The finitary analogue of the expression $S_f f$ is the *dual function* (of order 2) of f (the dual function of order 1 was briefly discussed in Section 2.8. If we are working on $\mathbf{Z}/N\mathbf{Z}$ with the usual shift, then S_f can be viewed as a Fourier multiplier which multiplies the Fourier coefficient at ξ by $|\hat{f}(\xi)|^2$; informally, S_f filters out all the low amplitude frequencies of f, leaving only a handful of high-amplitude frequencies.

Recall from Proposition 2.12.15 and Exercise 2.12.5 of Lecture 11 that a function $f \in L^2(X, \mathcal{X}, \mu)$ is almost periodic if and only if it is $\mathcal{Z}_1$-measurable, or if it lies in the pure point component $\mathbf{H}_{pp}$ of the shift operator T. We thus have

Corollary 2.12.25 (Koopman-von Neumann theorem)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $f \in L^2(X, \mathcal{X}, \mu)$. Let $\mathcal{Z}_1$ be the σ-algebra generated by the eigenfunctions of T.*

(1) *f is almost periodic if and only if $f \in L^2(X, \mathcal{Z}_1, \mu)$ if and only if $f \in \mathbf{H}_{pp}$.*

(2) *f is weakly mixing if and only if $\mathbf{E}(f|\mathcal{Z}_1) = 0$ a.e. if and only if $f \in \mathbf{H}_c = \mathbf{H}_{sc} + \mathbf{H}_{ac}$ (corresponding to the continuous spectrum of T).*

(3) *In general, f has a unique decomposition $f = f_{U^\perp} + f_U$ into an almost periodic function $f_{U^\perp}$ and a weakly mixing function f_U. Indeed, $f_{U^\perp} = \mathbf{E}(f|\mathcal{Z}_1)$ and $f_U = f - \mathbf{E}(f|\mathcal{Z}_1)$.*

Theorem 2.12.14 follows immediately from Corollary 2.12.25. Indeed, if a system is not weakly mixing, then by the above Corollary we see that $\mathcal{Z}_1$ is non-trivial, and the identity map from $(X, \mathcal{X}, \mu, T)$ to $(X, \mathcal{Z}_1, \mu, T)$ yields a non-trivial compact factor.

2.12.4. Roth's theorem. As a quick application of the above machinery we give a proof of Roth's theorem. We first need a variant of Corollary 2.12.8, which is proven by much the same means:

Exercise 2.12.21. Let $(X, \mathcal{X}, \mu, T)$ be an ergodic measure-preserving system, let a_1, a_2, a_3 be distinct integers, and let $f_1, f_2, f_3 \in L^\infty(X, \mathcal{X}, \mu)$ with at least one of f_1, f_2, f_3 weakly mixing. Show that $C - \lim_{n\to\infty} \int_X T^{a_1 n} f_1 T^a$ 0.

Theorem 2.12.26 (Roth's theorem)**.** *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic measure-preserving system, and let $f \in L^\infty(X, \mathcal{X}, \mu)$ be non-negative with $\int_X f \, d\mu > 0$. Then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f T^n f T^{2n} f \, d\mu > 0. \tag{2.135}$$

Proof. We decompose $f = f_{U^\perp} + f_U$ as in Corollary 2.12.25. The contribution of f_U is negligible by Exercise 2.12.21, so it suffices to show that

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f_{U^\perp} T^n f_{U^\perp} T^{2n} f_{U^\perp} \, d\mu > 0. \tag{2.136}$$

But as $f_{U^\perp}$ is almost periodic, the claim follows from Proposition 2.11.5. □

One can then immediately establish the $k = 3$ case of Furstenberg's theorem (Theorem 2.10.3) by combining the above result with the ergodic decomposition (Proposition 2.9.22). The $k = 3$ case of Szemerédi's theorem (i.e. Roth's theorem) then follows from the Furstenberg correspondence principle (see Section 2.10).

Exercise 2.12.22. Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $f \in L^2(X, \mathcal{X}, \mu)$ be non-negative. Show that for every $\varepsilon > 0$, one has $\langle T^n f, f\rangle \geq \int_X (f\ d\mu)^2 - \varepsilon$ for infinitely many n. *Hint*: first show this when f is almost periodic, and then use Corollary 2.12.8 and Corollary 2.12.25 to prove the general case.) This is a simplified version of the *Khintchine recurrence theorem*, which asserts that the set of such n is not only infinite, but is also syndetic. Analogues of the Khintchine recurrence theorem hold for double recurrence but not for triple recurrence; see [**BeHoKr2005**] for details.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/02/21`. Thanks to Liu Xiao Chuan for corrections.

2.13. Compact extensions

In Section 2.11, we studied *compact* measure-preserving systems - those systems $(X, \mathcal{X}, \mu, T)$ in which every function $f \in L^2(X, \mathcal{X}, \mu)$ was almost periodic, which meant that their orbit $\{T^n f : n \in \mathbf{Z}\}$ was *precompact* in the $L^2(X, \mathcal{X}, \mu)$ topology. Among other things, we were able to easily establish the Furstenberg recurrence theorem (Theorem 2.11.4) for such systems.

In this section, we generalise these results to a "relative" or "conditional" setting, in which we study systems which are compact relative to some factor $(Y, \mathcal{Y}, \nu, S)$ of $(X, \mathcal{X}, \mu, T)$. Such systems are to compact systems as isometric extensions are to isometric systems in topological dynamics. The main result we establish here is that the Furstenberg recurrence theorem holds for such compact extensions whenever the theorem holds for the base. The proof is essentially the same as in the compact case; the main new trick is to not to work in the Hilbert spaces $L^2(X, \mathcal{X}, \mu)$ over the complex numbers, but rather in the *Hilbert module*[48] $L^2(X, \mathcal{X}, \mu|Y, \mathcal{Y}, \nu)$ over the (commutative)

[48]Modules are to rings as vector spaces are to fields.

von Neumann algebra $L^\infty(Y, \mathcal{Y}, \nu)$. Because of the compact nature of the extension, it turns out that results from topological dynamics (and in particular, van der Waerden's theorem) can be exploited to good effect in this argument[49].

2.13.1. Hilbert modules. Let $X = (X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $\pi : X \to Y$ be a factor map, i.e. a morphism from X to another system $Y = (Y, \mathcal{Y}, \nu, S)$. The algebra $L^\infty(Y)$ can be viewed (using π) as a subalgebra of $L^\infty(X)$; indeed, it is isomorphic to $L^\infty(X, \pi^\#(\mathcal{Y}), \mu)$, where $\pi^\#(\mathcal{Y}) := \{\pi^{-1}(E) : E \in \mathcal{Y}\}$ is the pullback of $\mathcal{Y}$ by π.

Example 2.13.1. Throughout these notes we shall use the *skew shift* as our running example. Thus, in this example, $X = (\mathbf{R}/\mathbf{Z})^2$ with shift $T : (y, z) \mapsto (y + \alpha, z + y)$ for some fixed α (which can be either rational or irrational), $Y = \mathbf{R}/\mathbf{Z}$ with shift $S : y \mapsto y + \alpha$, with factor map $\pi : (y, z) \mapsto y$. In this case, $L^\infty(Y)$ can be thought of (modulo equivalence on null sets, of course) as the space of bounded functions on $(\mathbf{R}/\mathbf{Z})^2$ which depend only on the first variable.

Example 2.13.2. Another (rather trivial) example is when the factor system Y is simply a point. In this case, $L^\infty(Y)$ is the space of constants and can be identified with $\mathbf{C}$. At the opposite extreme, another example is when Y is equal to X. It is instructive to see how all of the concepts behave in each of these two extreme cases, as well as the typical intermediate case presented in Example 2.13.1.

The idea here will be to try to "relativise" the machinery of Hilbert spaces over $\mathbf{C}$ to that of Hilbert modules over $L^\infty(Y)$. Roughly speaking, all concepts which used to be complex or real-valued (e.g. inner products, norms, coefficients, etc.) will now take values in the algebra $L^\infty(Y)$. The following table depicts the various concepts that will be relativised:

[49]Note: this operator-algebraic approach is not the only way to understand these extensions; one can also proceed by disintegrating μ into fibre measures μ_y for almost every $y \in Y$ and working fibre by fibre. We will discuss the connection between the two approaches below.

Absolute / unconditional	Relative / cond
Constants $\mathbf{C}$	Factor-measura
Expectation $\mathbf{E}f = \int_X f \, d\mu \in \mathbf{C}$	Conditional ex
Inner product $\langle f, g \rangle_{L^2(X)} = \mathbf{E} f \overline{g}$	Conditional in
Hilbert space $L^2(X)$	Hilbert module
Finite-dimensional subspace $\{\sum_{j=1}^d c_j f_j : c_1, \ldots, c_d \in \mathbf{C}\}$	Finitely genera
Almost periodic function	Conditionally
Compact system	Compact exten
Hilbert-Schmidt operator	Conditionally
Weakly mixing function	Conditionally
Weakly mixing system	Weakly mixing

Remark 2.13.3. In information-theoretic terms, one can view Y as representing all the observables in the system X that have already been "measured" in some sense, so that it is now permissible to allow one's "constants" to depend on that data, and only study the remaining information present in X conditioning on the observed values in Y. Note though that once we activate the shift map T, the data in Y will similarly shift (by S), and so the various fibres of π can interact with each other in a non-trivial manner, so one should take some caution in applying information-theoretic intuition to this setting.

We have already seen that the factor Y induces a sub-σ-algebra $\pi^\#(\mathcal{Y})$ of $\mathcal{X}$. We therefore have a conditional expectation map $f \mapsto \mathbf{E}(f|Y)$ defined for all absolutely integrable f by the formula

$$\mathbf{E}(f|Y) := \mathbf{E}(f|\pi^\#(\mathcal{Y})). \tag{2.137}$$

In general, this expectation only lies in $L^1(Y)$, though for the functions we shall eventually study, the expectation will always lie in $L^\infty(Y)$ when needed.

As stated in the table, conditional expectation will play the role in the conditional setting that the unconditional expectation $\mathbf{E}f = \int_X f \, d\mu$ plays in the unconditional setting. Note though that the conditional expectation takes values in the algebra $L^\infty(Y)$ rather than in the complex numbers. We recall that conditional expectation is linear over this algebra, thus

$$\mathbf{E}(cf + dg|Y) = c\mathbf{E}(f|Y) + d\mathbf{E}(g|Y) \tag{2.138}$$

for all absolutely integrable f, g and all $c, d \in L^\infty(Y)$.

Example 2.13.4. Continuing Example 2.13.1, we see that for any absolutely integrable f on $(\mathbf{R}/\mathbf{Z})^2$, we have $\mathbf{E}(f|Y)(y,z) = \int_{\mathbf{R}/\mathbf{Z}} f(y,z')\,dz'$ almost everywhere.

Let $L^2(X|Y)$ be the space of all $f \in L^2(X, \mathcal{X}, \mu)$ such that the conditional norm

$$\|f\|_{L^2(X|Y)} := \mathbf{E}(|f|^2|Y)^{1/2} \tag{2.139}$$

lies in $L^\infty(Y)$ (rather than merely in $L^2(Y)$, which it does automatically). Thus for instance we have the inclusions

$$L^\infty(X) \subset L^2(X|Y) \subset L^2(X). \tag{2.140}$$

The space $L^2(X|Y)$ is easily seen to be a vector space over $\mathbf{C}$, and moreover (thanks to (2.138)) is a module over $L^\infty(Y)$.

Exercise 2.13.1. If we introduce the inner product

$$\langle f, g\rangle_{L^2(X|Y)} := \mathbf{E}(f\overline{g}|Y) \tag{2.141}$$

(which, initially, is only in $L^1(Y)$), establish the pointwise Cauchy-Schwarz inequality

$$|\langle f, g\rangle_{L^2(X|Y)}| \leq \|f\|_{L^2(X|Y)}\|g\|_{L^2(X|Y)} \tag{2.142}$$

almost everywhere. In particular, the inner product lies in $L^\infty(Y)$. *Hint*: repeat the standard proof of the Cauchy-Schwarz inequality verbatim, but with $L^\infty(Y)$ playing the role of the constants $\mathbf{C}$.

Example 2.13.5. Continuing Examples 1 and 3, $L^2(X|Y)$ consists (modulo null set equivalence) of all measurable functions $f(y,z)$ such that $\|f\|_{L^2(X|Y)} = (\int_{\mathbf{R}/\mathbf{Z}} |f(y,z)|^2\,dz)^{1/2}$ is bounded a.e. in y, with the relative inner product

$$\langle f, g\rangle_{L^2(X|Y)}(y) := \int_{\mathbf{R}/\mathbf{Z}} f(y,z)\overline{g(y,z)}\,dz \tag{2.143}$$

defined a.e. in y. Observe in this case that the relative Cauchy-Schwarz inequality (2.142) follows easily from the standard Cauchy-Schwarz inequality.

Exercise 2.13.2. Show that the function $f \mapsto \| \|f\|_{L^2(X|Y)} \|_{L^\infty(Y)}$ is a norm on $L^2(X|Y)$, which turns that space into a Banach space[50]. *Hint*: you may need to "relativise" one of the standard proofs that $L^2(X)$ is complete. You may also want to start with the skew shift example to build some intuition.

As π is a morphism, one can easily check the intertwining relationship

$$\mathbf{E}(T^n f|Y) = S^n \mathbf{E}(f|Y) \tag{2.144}$$

for all $f \in L^1(X)$ and integers n. As a consequence we see that the map T (and all of its powers) preserves the space $L^2(X|Y)$, and furthermore is conditionally unitary in the sense that

$$\langle T^n f, T^n g \rangle_{L^2(X|Y)} = S^n \langle f, g \rangle_{L^2(X|Y)} \tag{2.145}$$

for all $f, g \in L^2(X|Y)$ and integers n.

In the Hilbert space $L^2(X)$ one can create finite dimensional subspaces $\{c_1 f_1 + \ldots + c_d f_d : c_1, \ldots, c_d \in \mathbf{C}\}$ for any $f_1, \ldots, f_d \in L^2(X)$. Inside such subspaces we can create the bounded finite-dimensional *zonotopes* $\{c_1 f_1 + \ldots + c_d f_d : c_1, \ldots, c_d \in \mathbf{C}, |c_1|, \ldots, |c_d| \leq 1\}$. Observe (from the Heine-Borel theorem) that a subset E of $L^2(X)$ is precompact if and only if it can be approximated by finite-dimensional zonotopes in the sense that for every $\varepsilon > 0$, there exists a finite-dimensional zonotope Z of $L^2(X)$ such that E lies within the ε neighbourhood of Z.

Remark 2.13.6. There is nothing special about zonotopes here; just about any family of bounded finite-dimensional objects would suffice for this purpose. In fact, it seems to be slightly better (for the purposes of quantitative analysis, and in particular in controlling the dependence on dimension d) to work instead with octahedra, in which the constraint $|c_1|, \ldots, |c_d| \leq 1$ is replaced by $|c_1| + \ldots + |c_d| = 1$; this perspective is used for instance in [**Ta2006**].

Inspired by this, let us make some definitions. A *finitely generated module* of $L^2(X|Y)$ is any submodule of $L^2(X|Y)$ of the form $\{c_1 f_1 + \ldots + c_d f_d : c_1, \ldots, c_d \in L^\infty(Y)\}$, where $f_1, \ldots, f_d \in L^2(X|Y)$. Inside

[50] Because of this completeness, we refer to $L^2(X|Y)$ as a *Hilbert module* over $L^\infty(Y)$.

such a module we can define a *finitely generated module zonotope* $\{c_1 f_1 + \ldots + c_d f_d : c_1, \ldots, c_d \in L^\infty(Y); \|c_1\|_{L^\infty(Y)}, \ldots, \|c_d\|_{L^\infty(Y)} \leq 1\}$.

Definition 2.13.7.

- A subset E of $L^2(X|Y)$ is said to be *conditionally precompact* if for every $\varepsilon > 0$, there exists a finitely generated module zonotope Z of $L^2(X|Y)$ such that E lies within the ε-neighbourhood of Z (using the norm from Exercise 2.13.2).
- A function $f \in L^2(X|Y)$ is said to be *conditionally almost periodic* if its orbit $\{T^n f : n \in \mathbf{Z}\}$ is conditionally precompact.
- A function $f \in L^2(X|Y)$ is said to be *conditionally almost periodic in measure* if every $\varepsilon > 0$ there exists a set E in Y of measure at most ε such that $f 1_{E^c}$ is conditionally almost periodic.
- The system X is said to be a compact extension of Y if every function in $L^2(X|Y)$ is conditionally almost periodic in measure.

Example 2.13.8. Any bounded subset of $L^\infty(Y)$ is conditionally precompact (though note that it need not be precompact in the topological sense, using the topology from Exercise 2.13.2). In particular, every function in $L^\infty(Y)$ is conditionally almost periodic.

Example 2.13.9. Every system is a compact extension of itself. A system is a compact extension of a point if and only if it is a compact system.

Example 2.13.10. Consider the skew shift (Examples 2.13.1, 2.13.4, 2.13.5), and consider the orbit of the function $f(y,z) := e^{2\pi i z}$. A computation shows that

$$T^n f(y,z) = e^{2\pi i \frac{-n(-n-1)}{2}\alpha} e^{-2\pi i n y} f \tag{2.146}$$

which reveals (for α irrational) that f is not almost periodic in the unconditional sense. However, observe that all the shifts $T^n f$ lie in the zonotope $\{cf : c \in L^\infty(Y), \|c\|_{L^\infty(Y)} \leq 1\}$ generated by a single generator f, and so f is *conditionally* almost periodic.

Exercise 2.13.3. Consider the skew shift (Examples 1,3,4,7). Show that a sequence $f_n \in L^\infty(X)$ is conditionally precompact if and only if the sequences $f_n(y,\cdot) \in L^\infty(\mathbf{R}/\mathbf{Z})$ are precompact in $L^2(\mathbf{R}/\mathbf{Z})$ (with the usual Lebesgue measure) for almost every y.

Exercise 2.13.4. Show that the space of conditionally almost periodic functions in $L^2(X|Y)$ is a shift-invariant $L^\infty(Y)$ module, i.e. it is closed under addition, under multiplication by elements of $L^\infty(Y)$, and under powers T^n of the shift operator.

Exercise 2.13.5. Consider the skew shift (Examples 2.13.1, 2.13.4, 2.13.5, 2.13.10 and Exercise 2.13.3) with α irrational, and let $f \in L^2(X|Y)$ be the function defined by setting $f(y,z) := e^{2\pi i n z}$ whenever $n \geq 1$ and $y \in (1/(n+1), 1/n]$. Show that f is conditionally almost periodic in measure, but not conditionally almost periodic. Thus the two notions can be distinct even for bounded functions (a subtlety that does not arise in the unconditional setting).

Exercise 2.13.6. Let $\mathcal{Z}_{X|Y}$ denote the collection of all measurable sets E in X such that 1_E is conditionally almost periodic in measure. Show that $\mathcal{Z}_{X|Y}$ is a shift-invariant sub-σ-algebra of $\mathcal{X}$ that contains $\pi^\# \mathcal{Y}$, and that a function $f \in L^2(X|Y)$ is conditionally almost periodic in measure if and only if it is $\mathcal{Z}$-measurable. (In particular, $(X, \mathcal{Z}_{X|Y}, \mu, T)$ is the maximal compact extension of Y.) *Hint*: you may need to truncate the generators $f_1, \ldots, f_d$ of various module zonotopes to be in $L^\infty(X)$ rather than $L^2(X|Y)$.

Exercise 2.13.7. Show that the skew shift (Examples 2.13.1, 2.13.4, 2.13.5, 2.13.10 and Exercises 2.13.3, 2.13.5) is a compact extension of the circle shift. *Hint*: Use Example 2.13.10 and Exercise 2.13.6. Alternatively, approximate a function on the skew torus by its vertical Fourier expansions. For each fixed horizontal coordinate y, the partial sums of these vertical Fourier series converge (in the vertical L^2 sense) to the original function, pointwise in y. Now apply *Egorov's theorem*.

Exercise 2.13.8. Show that each of the iterated skew shifts (Exercise 2.9.8) are compact extensions of the preceding skew shift.

Exercise 2.13.9. Let $(Y, \mathcal{Y}, \nu, S)$ be a measure-preserving system, let G be a compact metrisable group with a closed subgroup H, let

$\sigma : Y \to G$ be measurable, and let $Y \times_\sigma G/H$ be the extension of Y with underlying space $Y \times G/H$, with measure equal to the product of ν and Haar measure, and shift map $T : (y, \zeta) \mapsto (Sy, \sigma(y)\zeta)$. Show that $Y \times_\sigma G/H$ is a compact extension of Y.

2.13.2. Multiple recurrence for compact extensions. Let us say that a measure-preserving system $(X, \mathcal{X}, \mu, T)$ obeys the *uniform multiple recurrence* (UMR) property if the conclusion of the Furstenberg multiple recurrence theorem holds for this system, thus for all $k \geq 1$ and all non-negative $f \in L^\infty(X)$ with $\int_X f \, d\mu > 0$, we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f T^n f \dots T^{(k-1)n} f \, d\mu > 0. \tag{2.147}$$

Thus in Section 2.11 we showed that all compact systems obey UMR, and in Section 2.12 we showed that all weakly mixing systems obey UMR. The Furstenberg multiple recurrence theorem asserts, of course, that *all* measure-preserving systems obey UMR.

We now establish an important further step (and, in many ways, the *key* step) towards proving that theorem:

Theorem 2.13.11. *Suppose that $X = (X, \mathcal{X}, \mu, T)$ is a compact extension of $Y = (Y, \mathcal{Y}, \nu, S)$. If Y obeys UMR, then so does X.*

Note that the converse implication is trivial: if a system obeys UMR, then all of its factors automatically do also.

Proof. Fix $k \geq 1$, and fix a non-negative function $f \in L^\infty(X)$ with $\int_X f \, d\mu > 0$. Our objective is to show that (2.147) holds. As X is a compact extension, f is conditionally almost periodic in measure; by definition (and uniform integrability), this implies that f can be bounded from below by another conditionally almost periodic function which is non-negative with positive mean. Thus we may assume without loss of generality that f is conditionally almost periodic.

We may normalise $\|f\|_{L^\infty(X)} = 1$ and $\int_X f \, d\mu = \delta$ for some $0 < \delta < 1$. The reader may wish to follow this proof using the skew shift example as a guiding model.

Let $\varepsilon > 0$ be a small number (depending on k and δ) to be chosen later. If we set $E := \{y \in Y : \mathbf{E}(f|Y) > \delta/2\}$, then E must have measure at least $\delta/2$.

Since f is almost periodic, we can find a finitely generated module zonotope $\{c_1 f_1 + \ldots + c_d f_d : \|c_1\|_{L^\infty(Y)}, \ldots, \|c_d\|_{L^\infty(Y)} \leq 1\}$ whose ε-neighbourhood contains the orbit of f. In other words, we have an identity of the form

$$T^n f = c_{1,n} f_1 + \ldots + c_{d,n} f_d + e_n \tag{2.148}$$

for all n, where $c_{1,n}, \ldots, c_{d,n} \in L^\infty(Y)$ with norm at most 1, and $e_n \in L^2(X,Y)$ is an error with $\|e_n\|_{L^2(X|Y)} = O(\varepsilon)$ almost everywhere.

By splitting into real and imaginary parts (and doubling d if necessary) we may assume that the $c_{j,n}$ are real-valued. By further duplication we can also assume that $\|f_i\|_{L^2(X|Y)} \leq 1$ for each i. By rounding off $c_{j,n}(y)$ to the nearest multiple of ε/d for each y (and absorbing the error into the e_n term) we may assume that $c_{j,n}(y)$ is always a multiple of ε/d. Thus each $c_{j,n}$ only takes on $O_{\varepsilon,d}(1)$ values.

Let K be a large integer (depending on k, d, δ, ε) to be chosen later. Since the factor space Y obeys UMR, and E has positive measure in Y, we know that

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_Y 1_E T^n 1_E \ldots T^{(K-1)n} 1_E \, d\nu > 0. \tag{2.149}$$

In other words, there exists a constant $c > 0$ such that

$$\nu(\Omega_n) > c \tag{2.150}$$

for a set of n of positive lower density, where Ω_n is the set

$$\Omega_n := E \cap T^n E \cap \ldots \cap T^{(K-1)n} E. \tag{2.151}$$

Let n be as above. By definition of Ω_n and E (and (2.145)), we see that

$$\mathbf{E}(T^{an} f|Y)(y) \geq \delta/2 \tag{2.152}$$

for all $y \in \Omega_n$ and $0 \leq a < K$. Meanwhile, from (2.148) we have

$$\|T^{an} f - c_{1,an} f_1 - \ldots - c_{d,an} f_d\|_{L^2(X|Y)}(y) = O(\varepsilon) \tag{2.153}$$

for all $y \in \Omega_n$ and $0 \leq a < K$.

Fix y. For each $0 \leq a < K$, the d-tuple $\vec{c}_{an}(y) := (c_{1,an}(y), \ldots, c_{d,an}(y))$ ranges over a set of cardinality $O_{d,\varepsilon}(1)$. One can view this as a colouring of $\{0, \ldots, K-1\}$ into $O_{d,\varepsilon}(1)$ colours. Applying van der Waerden's theorem (Exercise 2.4.3), we can thus find (if K is sufficiently large depending on d, ε, k) an arithmetic progression $a(y), a(y)+r(y), \ldots, a(y)+(k-1)r(y)$ in $\{0, \ldots, K-1\}$ for each y such that

$$\vec{c}_{a(y)n}(y) = \vec{c}_{(a(y)+r(y))n}(y) = \ldots = \vec{c}_{(a(y)+(k-1)r(y))n}(y). \tag{2.154}$$

The quantities $a(y)$ and $r(y)$ can of course be chosen to be measurable in y. By the pigeonhole principle, we can thus find a subset Ω'_n of Ω_n of measure at least $\sigma > 0$ for some σ depending on c, K, d, ε but independent of n, and an arithmetic progression $a, a+r, \ldots, a+(k-1)r$ in $\{0, \ldots, K-1\}$ such that

$$\vec{c}_{an}(y) = \vec{c}_{(a+r)n}(y) = \ldots = \vec{c}_{(a+(k-1)r)n}(y) \tag{2.155}$$

for all $y \in \Omega'_n$. (The quantities a and r can still depend on n, but this will not be of concern to us.)

Fix these values of a, r. From (2.153), (2.155) and the triangle inequality we see that

$$\|T^{(a+jr)n} f - T^{an} f\|_{L^2(X|Y)}(y) = O(\varepsilon) \tag{2.156}$$

for all $1 \leq j \leq k$ and $y \in \Omega'_n$. Recalling that f was normalised to have $L^\infty(X)$ norm 1, it is then not hard to conclude (by induction on k and the relative Cauchy-Schwarz inequality) that

$$\|T^{an} f T^{(a+r)n} f \ldots T^{(a+(k-1)r)n} f - (T^{an} f)^k\|_{L^2(X|Y)}(y) = O_k(\varepsilon) \tag{2.157}$$

and thus (by another application of relative Cauchy-Schwarz)

$$\mathbf{E}(T^{an} f T^{(a+r)n} f \ldots T^{(a+(k-1)r)n} f)(y) \geq \mathbf{E}((T^{an} f)^k|Y)(y) - O_k(\varepsilon). \tag{2.158}$$

But from (2.151), (2.152) and relative Cauchy-Schwarz again we have

$$\mathbf{E}(T^{an} f|Y)(y) \geq \delta/2 - O(\varepsilon) \tag{2.159}$$

and so by several more applications of relative Cauchy-Schwarz we have

$$\mathbf{E}((T^{an} f)^k|Y)(y) \geq c(k, \delta) > 0 \tag{2.160}$$

for some positive quantity $c(k,\delta)$ (if ε is sufficiently small depending on k, δ). From (2.158), (2.160) we conclude that

$$\mathbf{E}(T^{an} f T^{(a+r)n} f \dots T^{(a+(k-1)r)n} f)(y) \geq c(k,\delta)/2 \tag{2.161}$$

for $y \in \Omega'_n$, again if ε is small enough. Integrating this in y and using the shift-invariance we conclude that

$$\int_X f T^{nr} f \dots T^{(k-1)nr} f \, d\mu \geq c(k,\delta)\sigma/2. \tag{2.162}$$

The quantity r depends on n, but ranges between 1 and $K-1$, and so (by the non-negativity of f)

$$\sum_{s=1}^{K-1} \int_X f T^{ns} f \dots T^{(k-1)ns} f \, d\mu \geq c(k,\delta)\sigma/2 \tag{2.163}$$

for a set of n of positive lower density. Averaging this for n from 1 to N (say) one obtains (2.147) as desired. □

Thus for instance we have now established UMR for the skew shift as well as higher iterates of that shift, thanks to Exercises 2.13.7 and 2.145.

Remark 2.13.12. One can avoid the use of Hilbert modules, etc. by instead appealing to the theory of disintegration of measures (Theorem 2.9.21). We sketch the details as follows. First, one has to restrict attention to those spaces X which are regular, though an inspection of the Furstenberg correspondence principle (Section 2.10) shows that this is in fact automatic for the purposes of such tasks as proving Szemerédi's theorem. Once one disintegrates μ with respect to ν, the situation now resembles the concrete example of the skew shift, with the fibre measures μ_y playing the role of integration along vertical fibers $\{(y,z) : z \in \mathbf{R}/\mathbf{Z}\}$. It is then not difficult (and somewhat instructive) to convert the above proof to one using norms such as $L^2(X, \mathcal{X}, mu_y)$ rather than the module norm $L^2(X|Y)$. We leave the details to the reader (who can also get them from [**Fu1981**]).

Remark 2.13.13. It is an intriguing question as to whether there is any interesting non-commutative extension of the above theory, in which the underlying von Neumann algebra $L^\infty(Y, \mathcal{Y}, \nu)$ is replaced by a non-commutative von Neumann algebra. While some of the

theory seems to extend relatively easily, there does appear to be some genuine difficulties with other parts of the theory, particularly those involving multiple products such as $fT^n fT^{2n} f$.

Remark 2.13.14. Just as ergodic compact systems can be described as group rotation systems (Kronecker systems), it turns out that ergodic compact extensions can be described as (inverse limits of) group quotient extensions, somewhat analogously to Lemma 2.6.22. Roughly speaking, the idea is to first use some spectral theory to approximate conditionally almost periodic functions by conditionally *quasiperiodic* functions - those functions whose orbit lies on a finitely generated module zonotope (as opposed to merely being close to one). One can then use the generators of that zonotope as a basis from which to build the group quotient extension, and then use some further trickery to make the group consistent across all fibres. The precise machinery for this is known as *Mackey theory*; it is of particular importance in the deeper structural theory of dynamical systems, but we will not describe it in detail here, instead referring the reader to the papers of Furstenberg[**Fu1977**] and Zimmer[**Zi1976**].

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/02/27`. Thanks to Liu Xiao Chuan for corrections.

2.14. Weakly mixing extensions

Having studied compact extensions in Section 2.13, we now consider the opposite type of extension, namely that of a *weakly mixing extension*. Just as compact extensions are "relative" versions of compact systems (see Section 2.11), weakly mixing extensions are "relative" versions of weakly mixing systems (see Section 2.12), in which the underlying algebra of scalars $\mathbf{C}$ is replaced by $L^\infty(Y)$. As in the case of unconditionally weakly mixing systems, we will be able to use the van der Corput lemma to neglect "conditionally weakly mixing" functions, thus allowing us to lift the uniform multiple recurrence property (UMR) from a system to any weakly mixing extension of that system.

To finish the proof of the Furstenberg recurrence theorem requires two more steps. One is a relative version of the dichotomy between mixing and compactness: if a system is not weakly mixing relative

to some factor, then that factor has a non-trivial compact extension. This will be accomplished using the theory of conditional Hilbert-Schmidt operators in this lecture. Finally, we need the (easy) result that the UMR property is preserved under limits of chains; this will be accomplished in the next lecture.

2.14.1. Conditionally weakly mixing functions. Recall that in a measure-preserving system $X = (X, \mathcal{X}, \mu, T)$, a function $f \in L^2(X) = L^2(X, \mathcal{X}, \mu)$ is said to be *weakly mixing* if the squared inner products $|\langle T^n f, f\rangle_X|^2 := (\int_X T^n f \overline{f}\ d\mu)^2$ converge in the Cesáro sense, thus

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |\int_X T^n f \overline{f}\ d\mu|^2 = 0. \tag{2.164}$$

Now let $Y = (Y, \mathcal{Y}, \nu, S)$ be a factor of X, so that $L^\infty(Y)$ can be viewed as a subspace of $L^\infty(X)$. Recall that we have the conditional inner product $\langle f, g\rangle_{X|Y} := \mathbf{E}(f\overline{g}|Y)$ and the Hilbert module $L^2(X|Y)$ of functions f for which $\langle f, f\rangle_{X|Y}$ lies in $L^\infty(Y)$. We shall say that a function $f \in L^2(X|Y)$ is *conditionally weakly mixing* relative to Y if the L^2 norms $\|\langle T^n f, f\rangle_{X|Y}\|^2_{L^2(Y)}$ converge to zero in the Cesáro sense, thus

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X |\mathbf{E}(T^n f \overline{f}|Y)|^2\ d\mu = 0. \tag{2.165}$$

Example 2.14.1. If $X = Y \times Z$ is a product system of the factor space $Y = (Y, \mathcal{Y}, \nu, S)$ and another system $Z = (Z, \mathcal{Z}, \rho, R)$, then a function $f(y, z) = f(z)$ of the vertical variable $z \in Z$ is weakly mixing relative to Y if and only if $f(z)$ is weakly mixing in Z.

Much of the theory of weakly mixing systems extends easily to the conditionally weakly mixing case. For instance:

Exercise 2.14.1. By adapting the proof of Corollary 2.12.13, show that if $f \in L^2(X|Y)$ is conditionally weakly mixing and $g \in L^2(X|Y)$, then $\|\langle T^n f, g\rangle_{X|Y}\|^2_{L^2(Y)}$ and $\|\langle f, T^n g\rangle_{X|Y}\|^2_{L^2(Y)}$ converge to zero in the Cesáro sense. *Hint*: you will need to show that expressions such as $\langle g, T^n f\rangle_{X|Y} T^n f$ converge in $L^2(X)$ in the Cesáro sense. Apply the van der Corput lemma and use the fact that $\langle g, T^n f\rangle_{X|Y}$ are uniformly bounded in $L^\infty(Y)$ by conditional Cauchy-Schwarz.

Exercise 2.14.2. Show that the space of conditionally weakly mixing functions in $L^2(X|Y)$ is a module over $L^\infty(Y)$ (i.e. it is closed under addition and multiplication by the "scalars" $L^\infty(Y)$), which is also shift-invariant and topologically closed in the topology of $L^2(X|Y)$ (see Exercise 2.13.2).

Let us now see the first link between conditional weak mixing and conditional almost periodicity (cf. Exercise 2.12.18):

Lemma 2.14.2. *If $f \in L^2(X|Y)$ is conditionally weakly mixing and $g \in L^2(X|Y)$ is conditionally almost periodic, then $\langle f, g\rangle_{X|Y} = 0$ a.e.*

Proof. Since $\langle f, g\rangle_{X|Y} = T^{-n}\langle T^n f, T^n g\rangle_{X|Y}$, it will suffice to show that

$$C\text{-}\sup_{n\to\infty} |\langle T^n f, T^n g\rangle_{X|Y}|_{L^2(Y)} = 0. \tag{2.166}$$

Let $\varepsilon > 0$ be arbitrary. As g is conditionally almost periodic, one can find a finitely generated module zonotope $\{c_1 f_1 + \ldots + c_d f_d : \|c_1\|_{L^\infty(Y)}, \ldots, \|c_d\|_{L^\infty(Y)} \leq 1\}$ with $f_1, \ldots, f_d \in L^2(X|Y)$ such that all the shifts $T^n g$ lie within ε (in $L^2(X|Y)$) of this zonotope. Thus (by conditional Cauchy-Schwarz) we have

$$\|\langle T^n f, T^n g\rangle_{X|Y}\|_{L^2(Y)} = \|\langle T^n f, c_{1,n} f_1 + \ldots + c_{d,n} f_d\rangle_{X|Y}\|_{L^2(Y)} + O(\varepsilon) \tag{2.167}$$

for all n and some $c_{1,n}, \ldots, c_{d,n} \in L^\infty(Y)$ with norm at most 1. We can pull these constants out of the conditional inner product and bound the left-hand side of (2.167) by

$$\|\langle T^n f, f_1\rangle\|_{L^2(Y)} + \ldots + \|\langle T^n f, f_1\rangle\|_{L^2(Y)} + O(\varepsilon). \tag{2.168}$$

By Exercise 2.14.1, the Cesáro supremum of (2.168) is at most $O(\varepsilon)$. Since ε is arbitrary, the claim (2.166) follows. □

Since all functions in $L^\infty(Y)$ are conditionally almost periodic, we conclude that every conditionally weakly mixing function f is orthogonal to $L^\infty(Y)$, or equivalently that $\mathbf{E}(f|Y) = 0$ a.e. Let us say that f has *relative mean zero* if the latter holds.

Definition 2.14.3. A system X is a *weakly mixing extension* of a factor Y if every $f \in L^2(X|Y)$ with relative mean zero is relatively weakly mixing.

Exercise 2.14.3. Show that a product $X = Y \times Z$ of a system Y with a weakly mixing system Z is always a weakly mixing extension of Y.

Remark 2.14.4. If X is regular, then we can disintegrate the measure μ as an average $\mu = \int_Y \mu_y d\nu(y)$, see Theorem 2.9.21. It is then possible to construct a relative product system $X \times_Y X$, which is the product system $X \times X$ but with the measure $\mu \times_\nu \mu := \int_Y \mu_y \times \mu_y d\nu(y)$ instead of $\mu \times \mu$. It can then be shown (cf. Exercise 2.12.9) that X is a weakly mixing extension of Y if and only if $X \times_Y X$ is ergodic; see for instance [**Fu1981**] for details. However, in these notes we shall focus instead on the more abstract operator-algebraic approach which avoids the use of disintegrations.

Now we show that the uniform multiple recurrence property (UMR) from Section 2.13 is preserved under weakly mixing extensions (cf. Theorem 2.13.11).

Theorem 2.14.5. *Suppose that $X = (X, \mathcal{X}, \mu, T)$ is a weakly mixing extension of $Y = (Y, \mathcal{Y}, \nu, S)$. If Y obeys UMR, then so does X.*

The proof of this theorem rests on the following analogue of Proposition 2.12.11:

Proposition 2.14.6. *Let $a_1, \ldots, a_k \in \mathbf{Z}$ be distinct integers for some $k \geq 1$. Let $X = (X, \mathcal{X}, \mu, T)$ is a weakly mixing extension of $Y = (Y, \mathcal{Y}, \nu, S)$, and let $f_1, \ldots, f_k \in L^\infty(X)$ be such that at least one of $f_1, \ldots, f_k$ has relative mean zero. Then*

$$C\text{-}\lim_{n\to\infty} T^{a_1 n} f_1 \ldots T^{a_k n} f_k = 0 \tag{2.169}$$

in $L^2(X, \mathcal{X}, \mu)$.

Exercise 2.14.4. Prove Proposition 2.14.6. *Hint*: modify (or "relativise") the proof of Proposition 2.12.11.

Corollary 2.14.7. *Let $a_1, \ldots, a_k \in \mathbf{Z}$ be distinct integers for some $k \geq 1$. Let $X = (X, \mathcal{X}, \mu, T)$ is a weakly mixing extension of $Y = (Y, \mathcal{Y}, \nu, S)$, and let $f_1, \ldots, f_k \in L^\infty(X)$. Then*

(2.170)

$$C\text{-}\lim_{n\to\infty} \int_X T^{a_1 n} f_1 \ldots T^{a_k n} f_k \, d\mu - \int_X T^{a_1 n} \mathbf{E}(f_1|Y) \ldots T^{a_k n} \mathbf{E}(f_k|Y) \, d\mu = 0.$$

Exercise 2.14.5. Prove Corollary 2.14.7. *Hint*: adapt the proof of Corollary 2.12.13.

Proof of Theorem 2.14.5. Let $f \in L^\infty(X)$ be non-negative with positive mean. Then $\mathbf{E}(f|Y) \in L^\infty(Y)$ is also non-negative with positive mean. Since Y obeys UMR, we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{E}(f|Y)T^n\mathbf{E}(f|Y)\dots T^{(k-1)n}\mathbf{E}(f|Y) > 0. \tag{2.171}$$

Applying Corollary 2.14.7 we see that the same statement holds with $\mathbf{E}(f|Y)$ replaced by f, and the claim follows. □

Remark 2.14.8. As the above proof shows, Corollary 2.14.7 lets us replace functions in the weakly mixing extension X by their expectations in Y for the purposes of computing k-fold averages. In the notation of [**FuWe1996**], Corollary 2.14.7 asserts that Y is a *characteristic factor* of X for the average (2.170). The deeper structural theory of such characteristic factors (and in particular, on the minimal characteristic factor for any given average) is an active and difficult area of research, with surprising connections with Lie group actions (and in particular with flows on nilmanifolds), as well as the theory of inverse problems in additive combinatorics (and in particular to inverse theorems for the Gowers norms); see for instance [**Kr2006**] for a survey of recent developments. The concept of a characteristic factor (or more precisely, finitary analogues of this concept) also is fundamental in my work with Ben Green[**GrTa2008**] on primes in arithmetic progression.

2.14.2. The dichotomy between structure and randomness. The remainder of this lecture is devoted to proving the following "relative" generalisation of Theorem 2.12.14, and which is a fundamental ingredient in the proof of the Furstenberg recurrence theorem:

Theorem 2.14.9. *Suppose that $X = (X, \mathcal{X}, \mu, T)$ is an extension of a system $Y = (Y, \mathcal{Y}, \nu, S)$. Then exactly one of the following statements is true:*

(1) *(Structure) X has a factor Z which is a non-trivial compact extension of Y.*

(2) *(Randomness) X is a weakly mixing extension of Y.*

As in Section 2.12, the key to proving this theorem is to show

Proposition 2.14.10. *Suppose that $X = (X, \mathcal{X}, \mu, T)$ is an extension of a system $Y = (Y, \mathcal{Y}, \nu, S)$. Then a function $f \in L^2(X|Y)$ is relatively weakly mixing if and only if $\langle f, g\rangle_{X|Y} = 0$ a.e. for all relatively almost periodic g.*

The "only if" part of this proposition is Lemma 2.14.2; the harder part is the "if" part, which we will prove shortly. But for now, let us see why Proposition 2.14.10 implies Theorem 2.14.9.

From Lemma 2.14.2, we already know that no non-trivial function can be simultaneously conditionally weakly mixing and conditionally almost periodic, which shows that cases 1 and 2 of Theorem 2.14.9 cannot simultaneously hold. To finish the proof of Theorem 2.14.9, suppose that X is not a weakly mixing extension of Y, thus there exists a function $f \in L^2(X|Y)$ of relative mean zero which is not weakly mixing. By Proposition 2.14.10, there must exist a relatively almost periodic $g \in L^2(X|Y)$ such that $\langle f, g\rangle_{X|Y}$ does not vanish a.e.. Since f is orthogonal to all functions in $L^\infty(Y)$, we conclude that g is *not* in $L^\infty(Y)$, thus we have a single relatively almost periodic function. From Exercise 2.13.6, this shows that the maximal compact extension of Y is non-trivial, and the claim follows.

It thus suffices to prove the "if" part of Proposition 2.14.10; thus we need to show that every non-conditionally-weakly-mixing function correlates with some conditionally almost periodic function. But observe that if $f \in L^2(X|Y)$ is not conditionally weakly mixing, then by definition we have

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |\mathbf{E}(T^n f \overline{f}|Y)|^2_{L^2(Y)} > 0. \tag{2.172}$$

We can rearrange this as

$$\limsup_{N\to\infty} \langle S_{f,N} f, f\rangle_X > 0 \tag{2.173}$$

where $S_{f,N} : L^2(X|Y) \to L^2(X|Y)$ is the operator

$$S_{f,N}g := \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{E}(g\overline{T^n f}|Y)T^n f. \tag{2.174}$$

To prove Proposition 2.14.10, it thus suffices (by weak compactness) to show that

Proposition 2.14.11 (Dual functions are almost periodic)**.** *Suppose that $X = (X, \mathcal{X}, \mu, T)$ is an extension of a system $Y = (Y, \mathcal{Y}, \nu, S)$, and let $f \in L^2(X|Y)$. Let S_f be any limit point of $S_{f,N}$ in the weak operator technology. Then $S_f f$ is relatively almost periodic.*

Remark 2.14.12. By applying the mean ergodic theorem to the dynamical system $X \times_Y X$, one can show that the sequence D_N is in fact convergent in the weak or strong operator topologies (at least when X is regular). But to avoid some technicalities we shall present an argument that does not rely on existence of a strong limit.

As one might expect from the experience with unconditional weak mixing, the proof of Proposition 2.14.11 relies on the theory of conditionally Hilbert-Schmidt operators on $L^2(X|Y)$. We give here a definition of such operators which is suited for our needs.

Definition 2.14.13. Let X, Y be as above. A *sub-orthonormal set* in $L^2(X|Y)$ is any at most countable sequence $e_\alpha \in L^2(X|Y)$ such that $\langle e_\alpha, e_\beta\rangle_{X|Y} = 0$ a.e. for all $\alpha \neq \beta$ and $\langle e_\alpha, e_\alpha\rangle_{X|Y} \leq 1$ a.e. for all α. A linear operator $A : L^2(X|Y) \to L^2(X|Y)$ is said to be a *conditionally Hilbert-Schmidt operator* if we have the module property

$$A(cf) = cAf \text{ for all } c \in L^\infty(Y) \tag{2.175}$$

and the bound

$$\sum_\alpha \sum_\beta |\langle Ae_\alpha, f_\beta\rangle_{X,Y}|^2 \leq C^2 \text{ a.e.} \tag{2.176}$$

for all sub-orthonormal sets $\{e_\alpha\}$, $\{f_\beta\}$ and some constant $C > 0$; the best such C is called the *(uniform) conditional Hilbert-Schmidt norm* $\|\|A\|_{HS(X|Y)}\|_{L^\infty(Y)}$ of A.

Remark 2.14.14. As in Section 2.12, one can also set up the concept of a tensor product of two Hilbert modules, and use that to

define conditionally Hilbert-Schmidt operators in a way which does not require sub-orthonormal sets. But we will not need to do so here. One can also define a pointwise conditional Hilbert-Schmidt norm $\|A\|_{HS(X|Y)}(y)$ for each $y \in Y$, but we will not need this concept.

Example 2.14.15. Suppose Y is just a finite set (with the discrete σ-algebra), then X splits into finitely many fibres $\pi^{-1}(\{y\})$ with the conditional measures μ_y, and $L^2(X|Y)$ can be direct sum (with the l^∞ norm) of the Hilbert spaces $L^2(\mu_y)$. A conditional Hilbert-Schmidt operator is then equivalent to a family of Hilbert-Schmidt operators $A_y : L^2(\mu_y) \to L^2(\mu_y)$ for each y, with the A_y uniformly bounded in Hilbert-Schmidt norm.

Example 2.14.16. In the skew shift example $X = (\mathbf{R}/\mathbf{Z})^2 = \{(y,z) : y,z \in \mathbf{R}/\mathbf{Z}\}$, $Y = (\mathbf{R}/\mathbf{Z})$, one can show that an operator A is conditionally Hilbert-Schmidt if and only if it takes the form $Af(y,z) = \int_{\mathbf{R}/\mathbf{Z}} K_y(z,z')f(y,z')\,dz'$ a.e. for all $f \in L^2(X|Y)$, with $\|\|A\|_{HS(X|Y)}\|_{L^\infty(Y)} = \sup_y(\int_{\mathbf{R}/\mathbf{Z}} \int_{\mathbf{R}/\mathbf{Z}} |K_y(z,z')|^2 dzdz')^{1/2}$ finite.

Exercise 2.14.6. Let $f_1, f_2 \in L^2(X|Y)$ with $\|f_1\|_{L^2(X|Y)}, \|f_2\|_{L^2(X|Y)} \leq 1$ a.e.. Show that the rank one operator $g \mapsto \langle g, f_1\rangle_{X|Y} f_2$ is conditionally Hilbert-Schmidt with norm at most 1.

Observe from (2.174) that the $S_{f,N}$ are averages of rank one operators arising from the functions $T^n f$, and so by Exercise 2.14.6 and the triangle inequality we see that the $S_{f,N}$ are uniformly conditionally Hilbert-Schmidt. Taking weak limits using (2.176) (and Fatou's lemma) we conclude that S_f is also conditionally Hilbert-Schmidt.

Next, we observe from the telescoping identity that for every h, $T^h S_{f,N} - S_{f,N} T^h$ converges to zero in the weak operator topology (and even in the operator norm topology) as $N \to \infty$; taking limits, we see that S_f commutes with T. To show that $S_f f$ is conditionally almost periodic, it thus suffices to show the following analogue of Lemma 2.12.21:

Lemma 2.14.17. *Let $A : L^2(X|Y) \to L^2(X|Y)$ be a conditionally Hilbert-Schmidt operator. Then the image of the unit ball of $L^2(X|Y)$ under A is conditionally precompact.*

Proof. We shall prove this lemma by establishing a sort of conditional *singular value decomposition* for A. We can normalise A to have uniform conditional Hilbert-Schmidt norm 1. We fix $\varepsilon > 0$, and we will also need an integer k and a small quantity $\delta > 0$ depending on ε to be chosen later.

We first consider the quantities $|\langle Ae_1, f_1\rangle_{X|Y}|^2$ where e_1, f_1 ranges over all sub-orthonormal sets of cardinality 1. On the one hand, these quantities are bounded pointwise by 1, thanks to (2.176). On the other hand, observe that if $|\langle Ae_1, f_1\rangle_{X|Y}|^2$ and $|\langle Ae'_1, f'_1\rangle_{X|Y}|^2$ are of the above form, then so is the join $\max(|\langle Ae_1, f_1\rangle_{X|Y}|^2, |\langle Ae'_1, f'_1\rangle_{X|Y}|^2)$, as can be seen by taking $tildee_1 := e_1 1_E + e'_1 1_{E^c}$ and $tildef_1 := f_1 1_E + f'_1 1_{E^c}$, where E is the set where $|\langle Ae_1, f_1\rangle_{X|Y}|^2$ exceeds $|\langle Ae'_1, f'_1\rangle_{X|Y}|^2$. By using a maximising sequence for the quantity $\int_Y |\langle Ae, f\rangle_{X|Y}|^2\, d\nu$ and applying joins repeatedly, we can thus (on taking limits) find a pair e_1, f_1 which is near-optimal in the sense that $|\langle Ae_1, f_1\rangle_{X|Y}|^2 \geq (1-\delta)|\langle Ae'_1, f'_1\rangle_{X|Y}|^2$ a.e. for all competitors e'_1, f'_1.

Now fix e_1, f_1, and consider the quantity $|\langle Ae_2, f_2\rangle_{X|Y}|^2$, where $\{e_1, e_2\}$ and $\{f_1, f_2\}$ are sub-orthonormal sets. By arguing as before we can find an e_2, f_2 which is near optimal in the sense that $|\langle Ae_2, f_2\rangle_{X|Y}|^2 \geq (1-\delta)|\langle Ae'_2, f'_2\rangle_{X|Y}|^2$ a.e. for all competitors e'_2, f'_2.

We continue in this fashion k times to obtain sub-orthonormal sets $\{e_1, \ldots, e_k\}$ and $\{f_1, \ldots, f_k\}$ with the property that $|\langle Ae_i, f_i\rangle_{X|Y}|^2 \geq (1-\delta)|\langle Ae'_i, f'_i\rangle_{X|Y}|^2$ whenever $\{e_1, \ldots, e_{i-1}, e'_i\}, \{f_1, \ldots, f_{i-1}, f'_i\}$ are sub-orthonormal sets. On the other hand, from (2.176) we know that $\sum_i |\langle Ae_i, f_i\rangle_{X|Y}|^2 \leq 1$. From these two facts we soon conclude that $|\langle Ae, f\rangle_{X|Y}|^2 \leq 1/k + O_k(\delta)$ a.e. whenever $\{e_1, \ldots, e_k, e\}$ and $\{f_1, \ldots, f_k, f\}$ are sub-orthonormal. If k, δ are chosen appropriately we obtain $|\langle Ae, f\rangle_{X|Y}| \leq \varepsilon$ a.e. Thus (by duality) A maps the unit ball of the orthogonal complement of the span of $\{e_1, \ldots, e_k\}$ to the ε-neighbourhood of the span of $\{f_1, \ldots, f_k\}$ (with notions such as orthogonality, span, and neighbourhood being defined conditionally of course, using the $L^\infty(Y)$-Hilbert module structure of $L^2(X|Y)$). From this it is not hard to establish the desired precompactness. □

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/03/02`. Thanks to Lior Silberman, Orr, and Liu Xiao Chuan for corrections.

2.15. The Furstenberg-Zimmer structure theorem and the Furstenberg recurrence theorem

In this lecture - the final one on general measure-preserving dynamics - we put together the results from the past few lectures to establish the Furstenberg-Zimmer structure theorem for measure-preserving systems, and then use this to finish the proof of the Furstenberg recurrence theorem.

2.15.1. The Furstenberg-Zimmer structure theorem. Let $X = (X, \mathcal{X}, \mu, T)$ be a measure-preserving system, and let $Y = (Y, \mathcal{Y}, \nu, S)$ be a factor. In Theorem 2.14.9, we showed that if X was not a weakly mixing extension of Y, then we could find a non-trivial compact extension Z of Y (thus $L^2(Z)$ is a non-trivial superspace of $L^2(Y)$). Combining this with Zorn's lemma (and starting with the trivial factor $Y = \text{pt}$), one obtains

Theorem 2.15.1 (Furstenberg-Zimmer structure theorem)**.** [**Fu1977**],[**Zi1976**] *Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system. Then there exists an ordinal α and a factor $Y_\beta = (Y_\beta, \mathcal{Y}_\beta, \nu_\beta, S_\beta)$ for every $\beta \leq \alpha$ with the following properties:*

(1) *$Y_\emptyset$ is a point.*

(2) *For every successor ordinal $\beta + 1 \leq \alpha$, $Y_{\beta+1}$ is a compact extension of Y_β.*

(3) *For every limit ordinal $\beta \leq \alpha$, Y_β is the inverse limit of the Y_γ for the $\gamma < \beta$, in the sense that $L^2(Y_\beta)$ is the closure of $\bigcup_{\gamma<\beta} L^2(Y_\gamma)$.*

(4) *X is a weakly mixing extension of Y_α.*

Remark 2.15.2. This theorem should be compared with Furstenberg's structure theorem for distal systems in topological dynamics (Theorem 2.7.5). Indeed, in analogy to that theorem, the factors Y_β are known as *distal measure-preserving systems*.

Exercise 2.15.1. Deduce Theorem 2.15.1 from Theorem 2.14.9.

Remark 2.15.3. Since the Hilbert spaces $L^2(Y_\beta)$ are increasing inside the separable Hilbert space $L^2(X)$, it is not hard to see that the

ordinal α must be at most countable. Conversely, in [**BeFo1996**] it was shown that every countable ordinal can appear as the minimal length of a Furstenberg tower of a given system. Thus, in some sense, the complexity of a system can be as great as any countable ordinal. This is because the structure theorem roots out every last trace of structure from the system, so much so that every remaining function orthogonal to the final factor $L^2(Y_\alpha)$ is weakly mixing. But in many applications one does not need so much weak mixing; for instance to establish k-fold recurrence for a function f, it would be enough to obtain weak mixing control on just a few combinations of f (such as $T^h f \overline{f}$), as we already saw in the proof of Roth's theorem in Section 2.12. In fact, it is not hard to show that to prove Furstenberg's recurrence theorem for a fixed k, one only needs to analyse the first $k-2$ steps of the Furstenberg tower. As one consequence of this, it is possible to avoid the use of Zorn's lemma (and the axiom of choice) in the proof of the recurrence theorem.

Remark 2.15.4. Analogues of the structure theorem exist for other actions, such as the action of $\mathbf{Z}^d$ on a measure space (which can equivalently be viewed as the action of d commuting shifts $T_1, \ldots, T_d : X \to X$). There is a new feature in this case, though: instead of having a tower of purely compact extensions, followed by one weakly mixing extension at the end, one instead has a tower of hybrid extensions (known as primitive extensions), each one of which is compact along one subgroup of $\mathbf{Z}^d$ and weakly mixing along a complementary subgroup. See for instance [**Fu1981**] for details.

2.15.2. The Furstenberg recurrence theorem. The Furstenberg recurrence theorem asserts that every measure-preserving system $(X, \mathcal{X}, \mu, T)$ has the uniform multiple recurrence (UMR) property, thus

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f T^n f \ldots T^{(k-1)n} f \, d\mu > 0 \tag{2.177}$$

whenever $k \geq 1$ and $f \in L^\infty(X)$ is non-negative with positive mean. The UMR property is trivially true for a point, and we have already shown that UMR is preserved by compact extensions (Theorem 2.13.11) and by weakly mixing extensions (Theorem 2.14.5). The former result lets us climb the successor ordinal steps of the tower in

Theorem 2.15.1, while the latter lets us jump from the final distal system Y_α to X. But to clinch the proof of the recurrence theorem, we also need to deal with the limit ordinals. More precisely, we need to prove

Theorem 2.15.5. *(Limits of chains) Let $(Y_\beta)_{\beta\in B}$ be a totally ordered family of factors of a measure-preserving system X (thus $L^2(Y_\beta)$ is increasing with β, and let Y be the inverse limit of the Y_β. If each of the Y_β obeys the UMR, then Y does also.*

With this theorem, the Furstenberg recurrence theorem (Theorem 2.11.4) follows from the previous theorems and transfinite induction.

The main difficulty in establishing Theorem 2.15.5 is that while each Y_β obeys the UMR separately, we do not know that this property holds uniformly in β. The main new observation needed to establish the theorem is that there is another way to leverage the UMR from a factor to an extension... if the support of the function f is sufficiently "dense". We motivate this by first considering the unconditional case.

Proposition 2.15.6. *(UMR for densely supported functions) Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system, let $k \geq 1$ be an integer, and let $f \in L^\infty(X)$ be a non-negative function whose support $\{x : f(x) > 0\}$ has measure greater than $1 - 1/k$. Then* (2.177) *holds.*

Proof. By monotone convergence, we can find $\varepsilon > 0$ such that $f(x) > \varepsilon$ for all x outside of a set E of measure at most $1/k - \varepsilon$. For any n, this implies that $f(x)T^n f(x) \dots T^{(k-1)n} f(x) > \varepsilon^k$ for all x outside of the set $E \cup T^n E \cup \dots \cup T^{(k-1)n} E$, which has measure at most $1 - k\varepsilon$. In particular we see that

$$\int_X f T^n f \dots T^{(k-1)n} f \, d\mu > k\varepsilon^{k+1} \tag{2.178}$$

for all n, and the claim follows. □

As with the other components of the proof of the recurrence theorem, we will need to upgrade the above proposition to a "relative" version:

Proposition 2.15.7. *(UMR for relatively densely supported functions) Let $(X, \mathcal{X}, \mu, T)$ be an extension of a factor $(Y, \mathcal{Y}, \nu, S)$ with the UMR, let $k \geq 1$ be an integer, and let $f \in L^\infty(X)$ be a non-negative function whose support $\Omega := \{x : f(x) > 0\}$ is such that the set $\{y \in Y : \mathbf{E}(1_\Omega|Y) > 1 - 1/k\}$ has positive measure in Y. Then* (2.177) *holds.*

Proof. By monotone convergence again, we can find $\varepsilon > 0$ such that the set $E := \{x : f(x) > \varepsilon\}$ is such that the set $F := \{y \in Y : \mathbf{E}(1_E|Y) > 1 - 1/k + \varepsilon\}$ has positive measure. Since Y has the UMR, this implies that (2.177) holds for 1_F. In other words, there exists $c > 0$ such that

$$\nu(F \cap T^n F \cap \ldots \cap T^{(k-1)n} F) > c \tag{2.179}$$

for all n in a set of positive lower density.

Now we turn to f. We have the pointwise lower bound $f(x) \geq \varepsilon 1_E(x)$, and so

$$f T^n f \ldots T^{(k-1)n} f(x) \geq \varepsilon^k 1_{E \cap T^n E \cap \ldots \cap T^{(k-1)n} E}(x). \tag{2.180}$$

We have the crude lower bound

$$1_{E \cap T^n E \cap \ldots \cap T^{(k-1)n} E}(x) \geq 1 - \sum_{j=0}^{k-1} 1_{T^{jn} E^c}(x); \tag{2.181}$$

inserting this into (2.180) and taking conditional expectations, we conclude

$$\mathbf{E}(f T^n f \ldots T^{(k-1)n} f|Y)(y) \geq \varepsilon^k (1 - \sum_{j=0}^{k-1} \mathbf{E}(1_{T^{jn} E^c}|Y)(y)) \tag{2.182}$$

a.e. On the other hand, we have

$$\mathbf{E}(1_{T^{jn} E^c}|Y) = 1 - \mathbf{E}(1_{T^{jn} E}|Y) = 1 - T^{jn} \mathbf{E}(1_E|Y). \tag{2.183}$$

By definition of F, we thus see that if y lies in $F \cap T^n F \cap \ldots \cap T^{(k-1)n} F$, then

$$\mathbf{E}(f T^n f \ldots T^{(k-1)n} f|Y)(y) \geq \varepsilon^k \times k\varepsilon. \tag{2.184}$$

Integrating this and using (2.179), we obtain

$$\int_X f T^n f \ldots T^{(k-1)n} f \, d\mu \geq c\varepsilon^k \times k\varepsilon \tag{2.185}$$

for all n in a set of positive lower density, and (2.177) follows. □

Proof of Theorem 2.15.5. Let $f \in L^\infty(Y)$ be non-negative with positive mean $\int_X f\ d\mu = c > 0$; we may normalise f to be bounded by 1. Since Y is the inverse limit of the Y_β, we see that the orthogonal projections $\mathbf{E}(f|Y_\beta)$ converge in $L^2(X)$ norm to $\mathbf{E}(f|Y) = f$. Thus, for any ε, we can find β such that

$$\|f - \mathbf{E}(f|Y_\beta)\|_{L^2(X)} \leq \varepsilon. \tag{2.186}$$

Now $\mathbf{E}(f|Y_\beta)$ has the same mean c as f, and is also bounded by 1. Thus the set $E := \{y : \mathbf{E}(f|Y_\beta)(y) \geq c/2\}$ must have measure at least c/2 in Y_β. Now if $\Omega := \{x : f(x) > 0\}$, then we have the pointwise bound

$$|f - \mathbf{E}(f|Y_\beta)| \geq \frac{c}{2} 1_{\Omega^c} 1_E; \tag{2.187}$$

squaring this and taking conditional expectations we obtain

$$\mathbf{E}(|f - \mathbf{E}(f|Y_\beta)|^2|Y_\beta)(y) \geq \frac{c^2}{4}(1 - \mathbf{E}(1_\Omega|Y_\beta)(y))1_E(y), \tag{2.188}$$

and so by (2.186) and Markov's inequality we see that $1-\mathbf{E}(1_\Omega|Y_\beta)(y)1_E(y) < 1/k$ on a set of measure $O_c(\varepsilon^2)$. Choosing ε sufficiently small depending on c, we conclude (from the lower bound $\mu(E) \geq c/2$) that $\mathbf{E}(1_\Omega|Y_\beta)(y) > 1 - 1/k$ on a set of positive measure. The claim now follows from Proposition 2.15.7. □

The proof of the Furstenberg recurrence theorem (and thus Szemerédi's theorem) is finally complete.

Remark 2.15.8. The same type of argument yields many further recurrence theorems, and thus (by the correspondence principle) many combinatorial results also. For instance, in [**Fu1977**] it was noted that the above arguments allow one to strengthen (2.177) to

$$\liminf_{N\to\infty} \inf_M \frac{1}{N} \sum_{n=M}^{M+N-1} \int_X f T^n f \dots T^{(k-1)n} f\ d\mu > 0, \tag{2.189}$$

which allows one to conclude that in a set A of positive upper density, the set of n for which $A \cap (A+n) \cap \dots \cap (A+(k-1)n)$ has positive upper density is syndetic for every k. One can also extend the argument to higher dimensions, and to polynomial recurrence without too

many changes in the structure of the proof. But some more serious modifications to the argument are needed for other recurrence results involving IP systems or Hales-Jewett type results; see Section 2.10 for more discussion.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/03/05`. Thanks to Lior Silberman, Nate Chandler, and Liu Xiao Chuan for corrections.

2.16. A Ratner-type theorem for nilmanifolds

This section and the next will be on *Ratner's theorems* on equidistribution of orbits on homogeneous spaces (see also Section 1.11 of *Structure and Randomness* for an introduction to this family of results). Here, I will discuss two special cases of Ratner-type theorems. In this lecture, I will talk about Ratner-type theorems for discrete actions (of the integers $\mathbf{Z}$) on nilmanifolds; this case is much simpler than the general case, because there is a simple criterion in the nilmanifold case to test whether any given orbit is equidistributed or not. Ben Green and I had need recently[**GrTa2009c**] to develop quantitative versions of such theorems for a number-theoretic application. In Section 2.17, I will discuss Ratner-type theorems for actions of $SL_2(\mathbf{R})$, which is simpler in a different way (due to the semisimplicity of $SL_2(\mathbf{R})$, and lack of compact factors).

2.16.1. Nilpotent groups. Before we can get to Ratner-type theorems for nilmanifolds, we will need to set up some basic theory for these nilmanifolds. We begin with a quick review of the concept of a nilpotent group - a generalisation of that of an abelian group. Our discussion here will be purely algebraic (no manifolds, topology, or dynamics will appear at this stage).

Definition 2.16.1 (Commutators)**.** Let G be a (multiplicative) group. For any two elements g,h in G, we define the commutator $[g,h]$ to be $[g,h] := g^{-1}h^{-1}gh$ (thus g and h commute if and only if the commutator is trivial). If H and K are subgroups of G, we define the *commutator* $[H,K]$ to be the group generated by all the commutators $\{[h,k] : h \in H, k \in K\}$.

For future reference we record some trivial identities regarding commutators:

$$gh = hg[g,h] = [g^{-1}, h^{-1}]hg \tag{2.190}$$

$$h^{-1}gh = g[g,h] = [h, g^{-1}]g \tag{2.191}$$

$$[g,h]^{-1} = [h,g]. \tag{2.192}$$

Exercise 2.16.1. Let H, K be subgroups of a group G.

(1) Show that $[H,K] = [K,H]$.
(2) Show that H is abelian if and only if $[H,H]$ is trivial.
(3) Show that H is central if and only if $[H,G]$ is trivial.
(4) Show that H is normal if and only if $[H,G] \subset H$.
(5) Show that $[H,G]$ is always normal.
(6) If $L \lhd H, K$ is a normal subgroup of both H and K, show that $[H,K]/([H,K] \cap L) \equiv [H/L, K/L]$.
(7) Let HK be the group generated by $H \cup K$. Show that $[H,K]$ is a normal subgroup of HK, and when one *quotients* by this subgroup, $H/[H,K]$ and $K/[H,K]$ become abelian.

Exercise 2.16.2. Let G be a group. Show that the group $G/[G,G]$ is abelian, and is the *universal* abelianisation of G in the sense that every *homomorphism* $\phi : G \to H$ from G to an abelian group H can be uniquely factored as $\phi = \tilde{\phi} \circ \pi$, where $\pi : G \to G/[G,G]$ is the quotient map and $\tilde{\phi} : G/[G,G] \to H$ is a homomorphism.

Definition 2.16.2 (Nilpotency)**.** Given any group G, define the *lower central series*

$$G = G_0 = G_1 \rhd G_2 \rhd G_3 \rhd \dots \tag{2.193}$$

by setting $G_0, G_1 := G$ and $G_{i+1} := [G_i, G]$ for $i \geq 1$. We say that G is *nilpotent* of step s if G_{s+1} is trivial (and G_s is non-trivial).

Examples 2.16.3. A group is nilpotent of step 0 if and only if it is trivial. It is nilpotent of step 1 if and only if it is non-trivial and abelian. Any subgroup or homomorphic image of a nilpotent group of step s is nilpotent of step at most s. The direct product of two nilpotent groups is again nilpotent, but the semi-direct product of

nilpotent groups is merely solvable in general. If G is any group, then G/G_{s+1} is nilpotent of step at most s.

Example 2.16.4. Let $n \geq 1$ be an integer, and let

$$U_n(\mathbf{R}) = \begin{pmatrix} 1 & \mathbf{R} & \dots & \mathbf{R} \\ 0 & 1 & \dots & \mathbf{R} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \tag{2.194}$$

be the group of all upper-triangular $n \times n$ real matrices with 1s on the diagonal (i.e. the group of unipotent upper-triangular matrices). Then $U_n(\mathbf{R})$ is nilpotent of step n. Similarly if $\mathbf{R}$ is replaced by other fields.

Exercise 2.16.3. Let G be an arbitrary group.

(1) Show that each element G_i of the lower central series is a *characteristic subgroup* of G, i.e. $\phi(G_i) = G_i$ for all automorphisms[51] $\phi : G \to G$.

(2) Show the *filtration property* $[G_i, G_j] \subset G_{i+j}$ for all $i, j \geq 0$. *Hint*: induct on i+j; then, holding i+j fixed, quotient by G_{i+j}, and induct on (say) i. Note that once one quotients by G_{i+j}, all elements of $[G_{i-1}, G_j]$ are central (by the first induction hypothesis), while G_{i-1} commutes with $[G, G_j]$ (by the second induction hypothesis). Use these facts to show that all the generators of $[G, G_{i-1}]$ commute with G_j.

Exercise 2.16.4. Let G be a nilpotent group of step 2. Establish the identity

$$g^n h^n = (gh)^n [g,h]^{\binom{n}{2}} \tag{2.195}$$

for any integer n and any $g, h \in G$, where $\binom{n}{2} := \frac{n(n-1)}{2}$. (This can be viewed as a discrete version of the first two terms of the *Baker-Campbell-Hausdorff formula*. Conclude in particular that the space of *Hall-Petresco sequences* $n \mapsto g_0 g_1^n g_2^{\binom{n}{2}}$, where $g_i \in G_i$ for $i = 0, 1, 2$, is a group under pointwise multiplication (this group is known as the *Hall-Petresco group* of G). There is an analogous identity (and an

[51] Specialising to *inner automorphisms*, we see in particular that the G_i are all normal subgroups of G.

analogous group) for nilpotent groups of higher step; see for instance [**Le1998**] for details. The Hall-Petresco group is rather useful for understanding multiple recurrence and polynomial behaviour in nilmanifolds; we will not discuss this in detail, but see Exercise 2.16.5 below for a hint as to the connection.

Exercise 2.16.5 (Arithmetic progressions in nilspaces are constrained)**.** Let G be a nilpotent group of step $s \leq 2$, and consider two arithmetic progressions $x, gx, \ldots g^{s+1}x$ and $y, hy, \ldots, h^{s+1}y$ of length $s+2$ in G, where $x, y \in X$ and $g, h \in G$. Show that if these progressions agree in the first $s+1$ places (thus $g^i x = h^i y$ for all $i = 0, \ldots, s$) then they also agree in the last place. *Hint*: the only tricky case is $s = 2$. For this, either use direct algebraic computation, or experiment with the group of Hall-Petresco sequences from the previous exercise. The claim is in fact true for general s; see e.g. [**GrTa2009d**]

Remark 2.16.5. By Exercise 2.16.3.2, the lower central series is a filtration with respect to the commutator operation $g, h \mapsto [g, h]$. Conversely, if G admits a filtration $G = G_{(0)} = G_{(1)} \geq \ldots$ with $[G_{(i)}, G_{(j)}] \subset G_{(i+j)}$ and $G_{(j)}$ trivial for $j > s$, then it is nilpotent of step at most s. It is sometimes convenient for inductive purposes to work with filtrations rather than the lower central series (which is the "minimal" filtration available to a group G); see for instance [**GrTa2009c**].

Remark 2.16.6. Let G be a nilpotent group of step s. Then $[G, G_s] = G_{s+1}$ is trivial and so G_s is central (by Exercise 2.16.1), thus abelian and normal. By another application of Exercise 2.16.1, we see that G/G_s is nilpotent of step $s-1$. Thus we see that any nilpotent group G of step s is an *extension* of a nilpotent group G/G_s of step $s-1$, in the sense that we have a short exact sequence

$$0 \to G_s \to G \to G/G_s \to 0 \tag{2.196}$$

where the kernel G_s is abelian. Conversely, every abelian extension of an $s-1$-step nilpotent group is nilpotent of step at most s. In principle, this gives a recursive description of s-step nilpotent groups as an s-fold iterated tower of abelian extensions of the trivial group. Unfortunately, while abelian groups are of course very well understood, abelian *extensions* are a little inconvenient to work with algebraically;

the sequence (2.196) is not quite enough, for instance, to assert that G is a semi-direct product of G_s and G/G_s (this would require some means of embedding G/G_s back into G, which is not available in general). One can identify G (using the axiom of choice) with a product set $G/G_s \times G_s$ with a group law $(g,n) \cdot (h,m) = (gh, nm\phi(g,h))$, where $\phi : G/G_s \times G/G_s \to G_s$ is a map obeying various cocycle-type identities, but the algebraic structure of ϕ is not particularly easy to exploit. Nevertheless, this recursive tower of extensions seems to be well suited for understanding the *dynamical* structure of nilpotent groups and their quotients, as opposed to their *algebraic* structure (cf. our use of recursive towers of extensions in our previous lectures in dynamical systems and ergodic theory).

In our applications we will not be working with nilpotent groups G directly, but rather with their *homogeneous spaces* X, i.e. spaces with a transitive left-action of G. (Later we will also add some topological structure to these objects, but let us work in a purely algebraic setting for now.) Such spaces can be identified with group quotients $X \equiv G/\Gamma$ where $\Gamma \leq G$ is the *stabiliser* $\Gamma = \{g \in G : gx = x\}$ of some point x in X. (By the transitivity of the action, all stabilisers are conjugate to each other.) It is important to note that in general, Γ is not normal, and so X is not a group; it has a left-action of G but not right-action of G. Note though that any central subgroup of G acts on either the left or the right.

Now let G be s-step nilpotent, and let us temporarily refer to $X = G/\Gamma$ as an *s-step nilspace*. Then G_s acts on the right in a manner that commutes with the left-action of G. If we set $\Gamma_s := G_s \cap \Gamma \lhd G_s$, we see that the right-action of Γ_s on G/Γ is trivial; thus we in fact have a right-action of the abelian group $T_s := G_s/\Gamma_s$. (In our applications, T_s will be a torus.) This action can be easily verified to be free. If we let $\overline{X} := X/T_s$ be the quotient space, then we can view X as a principal T_s-bundle over $\overline{X}$. It is not hard to see that $\overline{X} \equiv \pi(G)/\pi(\Gamma)$, where $\pi : G \to G/G_s$ is the quotient map. Observe that $\pi(G)$ is nilpotent of step $s-1$, and $\pi(\Gamma)$ is a subgroup. Thus we have expressed an arbitrary s-step nilspace as a principal bundle (by some abelian group) over an $s-1$-step nilspace, and so s-step nilspaces can be

viewed as towers of abelian principal bundles, just as s-step nilpotent groups can be viewed as towers of abelian extensions.

2.16.2. Nilmanifolds. It is now time to put some topological structure (and in particular, Lie structure) on our nilpotent groups and nilspaces.

Definition 2.16.7 (Nilmanifolds)**.** An *s-step nilmanifold* is a nilspace G/Γ, where G is a finite-dimensional Lie group which is nilpotent of step s, and Γ is a discrete subgroup which is *cocompact* or *uniform* in the sense that the quotient G/Γ is compact.

Remark 2.16.8. In the literature, it is sometimes assumed that the nilmanifold G/Γ is connected, and that the group G is connected, or at least that its group $\pi_0(G) := G/G^\circ$ of connected components ($G^\circ \lhd G$ being the identity component of G) is finitely generated (one can often easily reduce to this case in applications). It is also convenient to assume that G° is simply connected (again, one can usually reduce to this case in applications, by passing to the universal cover of G° if necessary), as this implies (by the *Baker-Campbell-Hausdorff formula*) that the nilpotent Lie group G° is *exponential*, i.e. the exponential map $\exp : \mathfrak{g} \to G^\circ$ is a homeomorphism.

Example 2.16.9 (Skew torus)**.** If we define

$$G := \begin{pmatrix} 1 & \mathbf{R} & \mathbf{R} \\ 0 & 1 & \mathbf{Z} \\ 0 & 0 & 1 \end{pmatrix}; \quad \Gamma := \begin{pmatrix} 1 & \mathbf{Z} & \mathbf{Z} \\ 0 & 1 & \mathbf{Z} \\ 0 & 0 & 1 \end{pmatrix} \tag{2.197}$$

(thus G consists of the upper-triangular unipotent matrices whose middle right entry is an integer, and Γ is the subgroup in which all entries are integers) then G/Γ is a 2-step nilmanifold. If we write

$$[x,y] := \begin{pmatrix} 1 & x & y \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Gamma \tag{2.198}$$

then we see that G/Γ is isomorphic to the square $\{[x,y] : 0 \leq x,y \leq 1\}$ with the identifications $[x,1] \equiv [x,0]$ and $[0,y] := [1,y+x \bmod 1]$. (Topologically, this is homeomorphic to the ordinary 2-torus $(\mathbf{R}/\mathbf{Z})^2$, but the skewness will manifest itself when we do dynamics.)

Example 2.16.10 (Heisenberg nilmanifold)**.** If we set

$$G := \begin{pmatrix} 1 & \mathbf{R} & \mathbf{R} \\ 0 & 1 & \mathbf{R} \\ 0 & 0 & 1 \end{pmatrix}; \quad \Gamma := \begin{pmatrix} 1 & \mathbf{Z} & \mathbf{Z} \\ 0 & 1 & \mathbf{Z} \\ 0 & 0 & 1 \end{pmatrix} \tag{2.199}$$

then G/Γ is a 2-step nilmanifold. It can be viewed as a three-dimensional cube with the faces identified in a somewhat skew fashion, similarly to the skew torus in Example 2.16.9.

Let $\mathfrak{g}$ be the Lie algebra of G. Every element g of G acts linearly on $\mathfrak{g}$ by conjugation. Since G is nilpotent, it is not hard to see (by considering the iterated commutators of g with an infinitesimal perturbation of the identity) that this linear action is unipotent, and in particular has determinant 1. Thus, any constant volume form on this Lie algebra will be preserved by conjugation, which by basic differential geometry allows us to create a volume form (and hence a measure) on G which is invariant under both left and right translation; this Haar measure is clearly unique up to scalar multiplication. (In other words, nilpotent Lie groups are unimodular.) Restricting this measure to a fundamental domain of G/Γ and then descending to the nilmanifold we obtain a left-invariant Haar measure, which (by compactness) we can normalise to be a Borel probability measure. (Because of the existence of a left-invariant probability measure μ on G/Γ, we refer to the discrete subgroup Γ of G as a lattice.) One can show that this left-invariant Borel probability measure is unique.

Definition 2.16.11 (Nilsystem)**.** An *s-step nilsystem (or nilflow)* is a topological measure-preserving system (i.e. both a topological dynamical system and a measure-preserving system) with underlying space G/Γ a s-step nilmanifold (with the Borel σ-algebra and left-invariant probability measure), with a shift T of the form $T : x \mapsto gx$ for some $g \in G$.

Example 2.16.12. The Kronecker systems $x \mapsto x + \alpha$ on compact abelian Lie groups are 1-step nilsystems.

Example 2.16.13. The skew shift system $(x, y) \mapsto (x + \alpha, y + x)$ on the torus $(\mathbf{R}/\mathbf{Z})^2$ can be identified with a nilflow on the skew torus (Example 2.16.9), after identifying (x,y) with [x,y] and using

the group element

$$g := \begin{pmatrix} 1 & \alpha & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.200}$$

to create the flow.

Example 2.16.14. Consider the Heisenberg nilmanifold (Example 2.16.10) with a flow generated by a group element

$$g := \begin{pmatrix} 1 & \gamma & \beta \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix} \tag{2.201}$$

for some real numbers α, β, γ. If we identify

$$[x, y, z] := \begin{pmatrix} 1 & z & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} \Gamma \tag{2.202}$$

then one can verify that

(2.203)

$$T^n : [x, y, z] \mapsto [\{x+n\alpha\}, y+n\beta+\frac{n(n+1)}{2}\alpha\gamma-\lfloor x+n\alpha\rfloor(z+n\gamma) \bmod 1, z+n\gamma \text{ mo}$$

where $\lfloor\rfloor$ and $\{\}$ are the integer part and fractional part functions respectively. Thus we see that orbits in this nilsystem are vaguely quadratic in n, but for the presence of the not-quite-linear operators $\lfloor\rfloor$ and $\{\}$. (These expressions are known as *bracket polynomials*, and are intimately related to the theory of nilsystems.)

Given that we have already seen that nilspaces of step s are principal abelian bundles of nilspaces of step $s-1$, it should be unsurprising that nilsystems of step s are abelian extensions of nilsystems of step $s-1$. But in order to ensure that topological structure is preserved correctly, we do need to verify one point:

Lemma 2.16.15. *Let G/Γ be an s-step nilmanifold, with G connected and simply connected. Then $\Gamma_s := G_s \cap \Gamma$ is a discrete cocompact subgroup of G_s. In particular, $T_s := G_s/\Gamma_s$ is a compact connected abelian Lie group (in other words, it is a torus).*

Proof. Recall that G is exponential and thus identifiable with its Lie algebra $\}$. The commutators G_i can be similarly identified with the Lie algebra commutators $\}_i$; in particular, the G_i are all connected, simply connected Lie groups.

The key point to verify is the cocompact nature of Γ_s in G_s; all other claims are straightforward. We first work in the abelianisation G/G_2, which is identifiable with its Lie algebra and thus isomorphic to a vector space. The image of Γ under the quotient map $G \to G/G_2$ is a cocompact subgroup of this vector space; in particular, it contains a basis of this space. This implies that Γ contains an "abelianised" basis $e_1, \ldots, e_d$ of G in the sense that every element of G can be expressed in the form $e_1^{t_1} \ldots e_d^{t_d}$ modulo an element of the normal subgroup G_2 for some real numbers $t_1, \ldots, t_d$, where we take advantage of the exponential nature of G to define real exponentiation $g^t := \exp(t \log(g))$. Taking commutators s times (which eliminates all the "modulo G_2" errors), we then see that G_s is generated by expressions of the form $[e_{i_1}, [e_{i_2}, [\ldots, e_{i_s}] \ldots]^t$ for $i_1, \ldots, i_s \in \{1, \ldots, d\}$ and real t. Observe that these expressions lie in Γ_s if t is an integer. As G_s is abelian, we conclude that each element in G_s can be expressed as an element of Γ_s, times a bounded number of elements of the form $[e_{i_1}, [e_{i_2}, [\ldots, e_{i_s}] \ldots]^t$ with $0 \leq t < 1$. From this we conclude that the quotient map $G_s \mapsto G_s/\Gamma_s$ is already surjective on some bounded set, which we can take to be compact, and so G_s/Γ_s is compact as required. □

As a consequence of this lemma, we see that if $X = G/\Gamma$ is an s-step nilmanifold with G connected and simply connected, then X/T_s is an $s-1$-step nilmanifold (with G still connected and simply connected), and that X is a principal T_s-bundle over X/T_s in the topological sense as well as in the purely algebraic sense. One consequence of this is that every s-step nilsystem (with G connected and simply connected) can be viewed as a *toral extension* (i.e. a group extension by a torus) of an $s-1$-step nilsystem (again with G connected and simply connected). Thus for instance the skew shift system (Example 2.16.13) is a circle extension of a circle shift, while the Heisenberg nilsystem (Example 2.16.14) is a circle extension of an abelian 2-torus shift.

Remark 2.16.16. One should caution though that the converse of the above statement is not necessarily true; an extension $X \times_\phi T$ of an $s-1$-step nilsystem X by a torus T using a cocycle $\phi : X \to T$ need not be isomorphic to an s-step nilsystem (the cocycle ϕ has to obey an an additional equation (or more precisely, a system of equations when $s > 2$), known as the *Conze-Lesigne equation*, before this is the case. See for instance [**Zi2007**] for further discussion.

Exercise 2.16.6. Show that Lemma 2.16.15 continues to hold if we relax the condition that G is connected and simply connected, to instead require that G/Γ is connected, that G/G° is finitely generated, and that G° is simply connected.

Exercise 2.16.7. Show that Lemma 2.16.15 continues to hold if G_s and Γ_s are replaced by G_i and $\Gamma_i = G_i \cap \Gamma$ for any $0 \leq i \leq s$. In particular, setting $i = 2$, we obtain a projection map $\pi : X \to X_2$ from X to the Kronecker nilmanifold $X_2 = (G/G_2)/(\Gamma G_2/G_2)$.

Remark 2.16.17. One can take the structural theory of nilmanifolds much further, in particular developing the theory of *Mal'cev bases* (of which the elements $e_1, \ldots, e_d$ used to prove Lemma 2.16.15 were a very crude prototype). See the foundational paper[**Ma1951**] of Mal'cev for details, as well as the later paper [**Le2005**] which addresses the case in which G is not necessarily connected.

2.16.3. A criterion for ergodicity. We now give a useful criterion to determine when a given nilsystem is ergodic.

Theorem 2.16.18. *Let $(X, T) = (G/\Gamma, x \mapsto gx)$ be an s-step nilsystem with G connected and simply connected, and let (X_2, T_2) be the underlying Kronecker factor, as defined in Exercise 2.16.7. Then X is ergodic if and only if X_2 is ergodic.*

This result was first proven in [**Gr1961**], using spectral theory methods. We will use an argument of Parry[**Pa1969**] (and adapted in [**Le2005**]), relying on "vertical" Fourier analysis and topological arguments, which we have already used for the skew shift in Proposition 2.9.11. An alternate proof also appears in Section 1.4.

Proof. If X is ergodic, then the factor X_2 is certainly ergodic. To prove the converse implication, we induct on s. The case $s \leq 1$ is

trivial, so suppose $s > 1$ and the claim has already been proven for $s-1$. Then if X_2 is ergodic, we already know from induction hypothesis that X/T_s is ergodic. Suppose for contradiction that X is not ergodic, then we can find a non-constant shift-invariant function on X. Using Fourier analysis (or representation theory) of the vertical torus T_s as in Proposition 2.9.11, we may thus find a non-constant shift-invariant function f which has a *single vertical frequency* χ in the sense that one has $f(g_s x) = \chi(g_s) f(x)$ for all $x \in X$, $g_s \in G_s$, and some character $\chi : G_s \to S^1$. If the character χ is trivial, then f descends to a non-constant shift-invariant function on X/T_s, contradicting the ergodicity there, so we may assume that χ is non-trivial. Also, $|f|$ descends to a shift-invariant function on X/T_s and is thus constant by ergodicity; by normalising we may assume $|f| = 1$.

Now let $g_{s-1} \in G_{s-1}$, and consider the function $F_{g_{s-1}}(x) := f(g_{s-1}x)\overline{f(x)}$. As G_s is central, we see that $F_{g_{s-1}}$ is G_s-invariant and thus descends to X/T_s. Furthermore, as f is shift-invariant (so $f(gx) = f(x)$), and $[g_{s-1}, g] \in G_s$, some computation reveals that $F_{g_{s-1}}$ is an eigenfunction:

$$F_{g_{s-1}}(gx) = \chi([g_{s-1}, g]) F_{g_{s-1}}(x). \tag{2.204}$$

In particular, if $\chi([g_{s-1}, g]) \neq 1$, then $F_{g_{s-1}}$ must have mean zero. On the other hand, by continuity (and the fact that $|f| = 1$) we know that $F_{g_{s-1}}$ has non-zero mean for g_{s-1} close enough to the identity. We conclude that $\chi([g_{s-1}, g]) = 1$ for all g_{s-1} close to the identity; as the map $g_{s-1} \mapsto \chi([g_{s-1}, g])$ is a homomorphism, we conclude in fact that $\chi([g_{s-1}, g]) = 1$ for all g_{s-1}. In particular, from (2.204) and ergodicity we see that $F_{g_{s-1}}$ is constant, and so $f(g_{s-1}x) = c(g_{s-1}) f(x)$ for some $c(g_{s-1}) \in S^1$.

Now let $h \in G$ be arbitrary. Observe that

$$\begin{aligned}\int_G f(hg_{s-1}x)\overline{f(x)}\,d\mu &= \int_G f(hy)\overline{f(g_{s-1}^{-1}y)}\,d\mu \\ &= c(g_{s-1})\int_G f(hy)\overline{f(y)}\,d\mu \\ &= \int_G f(g_{s-1}hy)\overline{f(y)}\,d\mu \\ &= \chi([g_{s-1},h])\int_G f(hg_{s-1}y)\overline{f(y)}\,d\mu. \end{aligned} \tag{2.205}$$

For h and g_{s-1} close enough to the identity, the integral is non-zero, and we conclude that $\chi([g_{s-1},h]) = 1$ in this case. The map $(g_{s-1},h) \mapsto \chi([g_{s-1},h])$ is a homomorphism in each variable and so is constant. Since $G_s = [G_{s-1},G]$, we conclude that χ is trivial, a contradiction. □

Remark 2.16.19. The hypothesis that G is connected and simply connected can be dropped; see [**Le2005**] for details.

One pleasant fact about nilsystems, as compared with arbitrary dynamical systems, is that ergodicity can automatically be upgraded to unique ergodicity:

Theorem 2.16.20. *Let (X,T) be an ergodic nilsystem. Then (X,T) is also uniquely ergodic. Equivalently, for every $x \in X$, the orbit $(T^n x)_{n\in\mathbf{Z}}$ is equidistributed.*

Exercise 2.16.8. By inducting on step and adapting the proof of Proposition 2.9.14, prove Theorem 2.16.20.

2.16.4. A Ratner-type theorem. A *subnilsystem* of a nilsystem $(X,T) = (G/\Gamma,T)$ is a compact subsystem (Y,S) which is of the form $Y = Hx$ for some $x \in X$ and some closed subgroup $H \leq G$. One easily verifies that a subnilsystem is indeed a nilsystem.

From the above theorems we quickly obtain

Corollary 2.16.21 (Dichotomy between structure and randomness)**.** *Let (X,T) be a nilsystem with group G connected and simply connected, and let $x \in X$. Then exactly one of the following statements is true:*

(1) *The orbit* $(T^n x)_{n\in\mathbf{Z}}$ *is equidistributed.*

(2) *The orbit* $(T^n x)_{n\in\mathbf{Z}}$ *is contained in a proper subnilsystem* (Y, S) *with group* H *connected and simply connected, and with dimension strictly smaller than that of* G*.*

Proof. It is clear that 1. and 2. cannot both be true. Now suppose that 1. is false. By Theorem 2.16.20, this means that (X, T) is not ergodic; by Theorem 2.16.18, this implies that the Kronecker system (X_2, T_2) is not ergodic. Expanding functions on $X_2 \equiv G/G_2$ into characters and using Fourier analysis, we conclude that there is a non-trivial character $\chi : G/G_2 \to S^1$ which is T_2-invariant. If we let $\pi : G \to G/G_2$ be the canonical projection, then $\chi : G \to S^1$ is a continuous homomorphism, and the kernel H is a closed connected subgroup of G of strictly lower dimension. Furthermore, Hx is equal to a level set of χ and is thus compact. Since χ is T_2 invariant, we see that $T^n x \in Hx$ for all n, and the claim follows. □

Iterating this corollary, we obtain

Corollary 2.16.22 (Ratner-type theorem for nilmanifolds)**.** *Let* (X, T) *be a nilsystem with group* G *connected and simply connected, and let* $x \in X$*. Then the orbit* $(T^n x)_{n\in\mathbf{Z}}$ *is equidistributed in some subnilmanifold* (Y, S) *of* (X, T)*. (In particular, this orbit is dense in* Y*.) Furthermore,* $Y = Hx$ *for some closed connected subgroup* H *of* G*.*

Remark 2.16.23. Analogous claims also hold when G is not assumed to be connected or simply connected, and if the orbit $(T^n x)_{n\in\mathbf{Z}}$ is replaced with a polynomial orbit $(T^{p(n)} x)_{n\in\mathbf{Z}}$; see [**Le2005**], [**Le2005b**]. In a different direction, such discrete Ratner-type theorems have been extended to other unipotent actions on finite volume homogeneous spaces by Shah[**Sh1996**]. Quantitative versions of this theorem have also been obtained by Ben Green and myself[**GrTa2009c**].

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/03/09`. Thanks to Jordi-Lluis Figueras Romero for corrections.

2.17. A Ratner-type theorem for $SL_2(R)$ orbits

In this final section of this chapter, we establish a Ratner-type theorem for actions of the special linear group $SL_2(\mathbf{R})$ on homogeneous spaces. More precisely, we show:

Theorem 2.17.1. *Let G be a Lie group, let $\Gamma < G$ be a discrete subgroup, and let $H \leq G$ be a subgroup isomorphic to $SL_2(\mathbf{R})$. Let μ be an H-invariant probability measure on G/Γ which is ergodic with respect to H (i.e. all H-invariant sets either have full measure or zero measure). Then μ is* homogeneous *in the sense that there exists a closed connected subgroup $H \leq L \leq G$ and a closed orbit $Lx \subset G/\Gamma$ such that μ is L-invariant and supported on Lx.*

This result is a special case of a more general theorem of Ratner, which addresses the case when H is generated by elements which act unipotently on the Lie algebra $\mathfrak{g}$ by conjugation, and when G/Γ has finite volume. To prove this theorem we shall follow an argument of Einsiedler[**Ei2006**], which uses many of the same ingredients used in Ratner's arguments but in a simplified setting (in particular, taking advantage of the fact that H is semisimple with no non-trivial compact factors). These arguments have since been extended and made quantitative in [**EiMaVe2007**].

2.17.1. Representation theory of $SL_2(\mathbf{R})$. Theorem 2.17.1 concerns the action of $H \equiv SL_2(\mathbf{R})$ on a homogeneous space G/Γ. Before we are able to tackle this result, we must first understand the linear actions of $H \equiv SL_2(\mathbf{R})$ on real or complex vector spaces - in other words, we need to understand the representation theory of the Lie group $SL_2(\mathbf{R})$ (and its associated Lie algebra $\mathfrak{sl}_2(\mathbf{R})$).

Of course, this theory is very well understood, and by using the machinery of weight spaces, raising and lowering operators, etc. one can completely classify all the finite-dimensional representations of $SL_2(\mathbf{R})$; in fact, all such representations are isomorphic to direct sums of *symmetric powers* of the standard representation of $SL_2(\mathbf{R})$ on $\mathbf{R}^2$. This classification quickly yields all the necessary facts we will need here. However, we will use only a minimal amount of this

machinery here, to obtain as direct and elementary a proof of the results we need as possible.

The first fact we will need is that finite-dimensional representations of $SL_2(\mathbf{R})$ are completely reducible:

Lemma 2.17.2 (Complete reducibility)**.** *Let $SL_2(\mathbf{R})$ act linearly (and smoothly) on a finite-dimensional real vector space V, and let W be a $SL_2(\mathbf{R})$-invariant subspace of V. Then there exists a complementary subspace W' to W which is also $SL_2(\mathbf{R})$-invariant (thus V is isomorphic to the direct sum of W and W').*

Proof. We will use *Weyl's unitary trick* to create the complement W', but in order to invoke this trick, we first need to pass from the non-compact group $SL_2(\mathbf{R})$ to a compact counterpart. This is done in several stages.

First, we linearise the action of the Lie group $SL_2(\mathbf{R})$ by differentiating to create a corresponding linear action of the Lie algebra $\mathfrak{sl}_2(\mathbf{R})$ in the usual manner.

Next, we complexify the action. Let $V^{\mathbf{C}} := V \otimes \mathbf{C}$ and $W^{\mathbf{C}} := W \otimes \mathbf{C}$ be the complexifications of V and W respectively. Then the complexified Lie algebra $\mathfrak{sl}_2(\mathbf{C})$ acts on both $V^{\mathbf{C}}$ and $W^{\mathbf{C}}$, and in particular the *special unitary* Lie algebra $\mathfrak{su}_2(\mathbf{C})$ does also.

Since the *special unitary group*

$$SU_2(\mathbf{C}) = \left\{ \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} : \alpha, \beta \in \mathbf{C}; |\alpha|^2 + |\beta|^2 = 1 \right\} \tag{2.206}$$

is topologically equivalent to the 3-sphere S^3 and is thus simply connected, a standard homotopy argument allows one[52] to exponentiate the $\mathfrak{su}_2(\mathbf{C})$ action to create a $SU_2(\mathbf{C})$ action, thus creating the desired compact action.

Now we can apply the unitary trick. Take any Hermitian form $\langle , \rangle$ on $V^{\mathbf{C}}$. This form need not be preserved by the $SU_2(\mathbf{C})$ action, but if one defines the averaged form

$$\langle u, v \rangle_{SU_2} := \int_{SU_2(\mathbf{C})} \langle gu, gv \rangle \, dg \tag{2.207}$$

[52]This trick is not restricted to $\mathfrak{sl}_2(\mathbf{R})$, but can be generalised to other semisimple Lie algebras using the *Cartan decomposition*.

where dg is Haar measure on the compact Lie group $SU_2(\mathbf{C})$, then we see that $\langle , \rangle_{SU_2}$ is a Hermitian form which is $SU_2(\mathbf{C})$-invariant; thus this form endows $V^{\mathbf{C}}$ with a Hilbert space structure with respect to which the $SU_2(\mathbf{C})$-action is unitary. If we then define $(W')^{\mathbf{C}}$ to be the orthogonal complement of $W^{\mathbf{C}}$ in this Hilbert space, then this vector space is invariant under the $SU_2(\mathbf{C})$ action, and thus (by differentiation) by the $\mathfrak{su}_2(\mathbf{C})$ action. But observe that $\mathfrak{su}_2(\mathbf{C})$ and $\mathfrak{sl}_2(\mathbf{R})$ have the same complex span (namely, $\mathfrak{sl}_2(\mathbf{C})$); thus the complex vector space $(W')^{\mathbf{C}}$ is also $\mathfrak{sl}_2(\mathbf{R})$-invariant.

The last thing to do is to undo the complexification. If we let W' be the space of real parts of vectors in $(W')^{\mathbf{C}}$ which are real modulo $W^{\mathbf{C}}$, then one easily verifies that W' is $\mathfrak{sl}_2(\mathbf{R})$-invariant (hence $SL_2(\mathbf{R})$-invariant, by exponentiation) and is a complementary subspace to W, as required. □

Remark 2.17.3. We can of course iterate the above lemma and conclude that every finite-dimensional representation of $SL_2(\mathbf{R})$ is the direct sum of irreducible representations, which explains the term "complete reducibility". Complete reducibility of finite-dimensional representations of a Lie algebra (over a field of characteristic zero) is equivalent to that Lie algebra being *semisimple*. The situation is slightly more complicated for Lie groups, though, if such groups are not simply connected.

An important role in our analysis will be played by the one-parameter unipotent subgroup $U := \{u^t : t \in \mathbf{R}\}$ of $SL_2(\mathbf{R})$, where

$$u^t := \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}. \tag{2.208}$$

Clearly, the elements of U are unipotent when acting on $\mathbf{R}^2$. It turns out that they are unipotent when acting on all other finite-dimensional representations also:

Lemma 2.17.4. *Suppose that $SL_2(\mathbf{R})$ acts on a finite-dimensional real or complex vector space V. Then the action of any element of U on V is unipotent.*

Proof. By complexifying V if necessary we may assume that V is complex. The action of the Lie group $SL_2(\mathbf{R})$ induces a Lie algebra

homomorphism $\rho : \mathfrak{sl}_2(\mathbf{R}) \to \mathrm{End}(V)$. To show that the action of U is unipotent, it suffices to show that $\rho(\log u)$ is nilpotent, where

$$\log u = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{2.209}$$

is the infinitesimal generator of U. To show this, we will exploit the fact that $\log u$ induces a *raising operator*. We introduce the diagonal subgroup $D := \{d^t : t \in \mathbf{R}\}$ of $SL_2(\mathbf{R})$, where

$$d^t := \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}. \tag{2.210}$$

This group has infinitesimal generator

$$\log d = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.211}$$

Observe that $[\log d, \log u] = 2 \log u$, and thus (since ρ is a Lie algebra homomorphism)

$$[\rho(\log d), \rho(\log u)] = 2\rho(\log u). \tag{2.212}$$

We can rewrite this as

$$(\rho(\log d) - \lambda - 2)\rho(\log u) = \rho(\log u)(\rho(\log d) - \lambda) \tag{2.213}$$

for any $\lambda \in \mathbf{C}$, which on iteration implies that

$$(\rho(\log d) - \lambda - 2r)^m \rho(\log u)^r = \rho(\log u)^r (\rho(\log d) - \lambda)^m \tag{2.214}$$

for any non-negative integers m, r. But this implies that $\rho(\log u)^r$ raises generalised eigenvectors of $\rho(\log d)$ of eigenvalue λ to generalised eigenvectors of $\rho(\log d)$ of eigenvalue $\lambda + 2m$. But as V is finite dimensional, there are only finitely many eigenvalues of $\rho(\log d)$, and so $\rho(\log u)$ is nilpotent on each of the generalised eigenvectors of $\rho(\log d)$. By the Jordan normal form (see Section 1.13 of *Structure and Randomness*), these generalised eigenvectors span V, and we are done. □

Exercise 2.17.1. By carrying the above analysis further (and also working with the adjoint of U to create lowering operators) show (for complex V) that $\rho(\log d)$ is diagonalisable, and the eigenvalues are all integers. For an additional challenge: deduce from this that the representation is isomorphic to a direct sum of the representations of $SL_2(\mathbf{R})$ on the symmetric tensor powers $\mathrm{Sym}^k(\mathbf{R}^2)$ of $\mathbf{R}^2$ (or, if

you wish, on the space of homogeneous polynomials of degree k on 2 variables).

The group U is merely a subgroup of the group $SL_2(\mathbf{R})$, so it is not a priori evident that any vector (in a space that $SL_2(\mathbf{R})$ acts on) which is U-invariant, is also $SL_2(\mathbf{R})$-invariant. But, thanks to the highly non-commutative nature of $SL_2(\mathbf{R})$, this turns out to be the case, even in infinite dimensions, once one restricts attention to *continuous unitary* actions:

Lemma 2.17.5 (Mautner phenomenon)**.** *Let $\rho : SL_2(\mathbf{R}) \to U(V)$ be a continuous unitary action on a Hilbert space V (possibly infinite dimensional). Then any vector $v \in V$ which is fixed by U, is also fixed by $SL_2(\mathbf{R})$.*

Proof. We use an argument of Margulis. We may of course take v to be non-zero. Let $\varepsilon > 0$ be a small number. Then even though the matrix $w^\varepsilon := \begin{pmatrix} 1 & 0 \\ \varepsilon & 1 \end{pmatrix}$ is very close to the identity, the double orbit $Uw^\varepsilon U$ can stray very far away from U. Indeed, from the algebraic identity

$$\begin{pmatrix} e^t & 0 \\ \varepsilon & e^{-t} \end{pmatrix} = u^{(e^t-1)/\varepsilon} w^\varepsilon u^{(e^{-t}-1)/\varepsilon} \tag{2.215}$$

which is valid for any $t \in \mathbf{R}$, we see that this double orbit in fact comes very close to the diagonal group D. Applying (2.215) to the U-invariant vector v and taking inner products with v, we conclude from unitarity that

$$\left\langle \rho(\begin{pmatrix} e^t & 0 \\ \varepsilon & e^{-t} \end{pmatrix})v, v \right\rangle = \langle \rho(w^\varepsilon)v, v\rangle. \tag{2.216}$$

Taking limits as $\varepsilon \to 0$ (taking advantage of the continuity of ρ) we conclude that $\langle \rho(d^t)v, v\rangle = \langle v, v\rangle$. Since $\rho(d^t)v$ has the same length as v, we conclude from the converse Cauchy-Schwarz inequality that $\rho(d^t)v = v$, i.e. that v is D-invariant. Since U and D generate $SL_2(\mathbf{R})$, the claim follows. □

Remark 2.17.6. The key fact about U being used here is that its Lie algebra is not trapped inside any proper ideal of $sl_2(\mathbf{R})$, which, in turn, follows from the fact that this Lie algebra is simple. One

can do the same thing for semisimple Lie algebras provided that the unipotent group U is non-degenerate in the sense that it has non-trivial projection onto each simple factor.

This phenomenon has an immediate dynamical corollary:

Corollary 2.17.7 (Moore ergodic theorem)**.** *Suppose that $SL_2(\mathbf{R})$ acts in a measure-preserving fashion on a probability space $(X, \mathcal{X}, \mu)$. If this action is ergodic with respect to $SL_2(\mathbf{R})$, then it is also ergodic with respect to U.*

Proof. Apply Lemma 2.17.5 to $L^2(X, \mathcal{X}, \mu)$. □

2.17.2. Proof of Theorem 2.17.1. Having completed our representation-theoretic preliminaries, we are now ready to begin the proof of Theorem 2.17.1. The key is to prove the following dichotomy:

Proposition 2.17.8 (Lack of concentration implies additional symmetry)**.** *Let G, H, μ, Γ be as in Theorem 2.17.1. Suppose there exists a closed connected subgroup $H \leq L \leq G$ such that μ is L-invariant. Then exactly one of the following statements hold:*

(1) *(Concentration) μ is supported on a closed orbit Lx of L.*

(2) *(Additional symmetry) There exists a closed connected subgroup $L < L' \leq G$ such that μ is L'-invariant.*

Iterating this proposition (noting that the dimension of L' is strictly greater than that of L) we will obtain Theorem 2.17.1. So it suffices to establish the proposition.

We first observe that the ergodicity allows us to obtain the concentration conclusion (2.206) as soon as μ assigns any non-zero mass to an orbit of L:

Lemma 2.17.9. *Let the notation and assumptions be as in Proposition 2.17.8. Suppose that $\mu(Lx_0) > 0$ for some x_0. Then Lx_0 is closed and μ is supported on Lx_0.*

Proof. Since Lx_0 is H-invariant and μ is H-ergodic, the set Lx_0 must either have full measure or zero measure. It cannot have zero measure by hypothesis, thus $\mu(Lx_0) = 1$. Thus, if we show that

Lx_0 is closed, we automatically have that μ is supported on Lx_0. As G/Γ is a homogeneous space, we may assume without loss of generality (conjugating L if necessary) that x_0 is at the origin, then $Lx_0 \equiv L/(\Gamma \cap L)$. The measure μ on this set can then be pulled back to a measure m on L by the formula

$$\int_L f(g)\ dm(g) = \int_{L/(\Gamma\cap L)} \sum_{g \in x(\Gamma \cap L)} f(g)\ d\mu(x). \tag{2.217}$$

By construction, m is left L-invariant (i.e. a left Haar measure) and right $(\Gamma \cap L)$-invariant. From uniqueness of left Haar measure up to constants, we see that for any g in L there is a constant $c(g) > 0$ such that $m(Eg) = c(g)m(E)$ for all measurable E. It is not hard to see that $c : L \to \mathbf{R}^+$ is a character, i.e. it is continuous and multiplicative, thus $c(gh) = c(g)c(h)$ for all g, h in L. Also, it is the identity on $(\Gamma \cap L)$ and thus descends to a continuous function on $L/(\Gamma \cap L)$. Since μ is L-invariant, we have

$$\int_{L/(\Gamma\cap L)} c(g)\ d\mu(g) = \int_{L/(\Gamma\cap L)} c(hg)\ d\mu(g) = \int_{L/(\Gamma\cap L)} c(h)c(g)\ d\mu(g) \tag{2.218}$$

for all h in L, and thus c is identically 1 (i.e. L is unimodular). Thus m is right-invariant, which implies that μ obeys the right-invariance property $\mu(Kx_0) = \mu(Kgx_0)$ for any g in L and any sufficiently small compact set $K \subset L$ (small enough to fit inside a single fundamental domain of $L/(\Gamma \cap L)$).

Recall that $\mu(Lx_0) = 1$. By partitioning L into countably many small sets as above, we can thus find a small compact set $K \subset L$ such that $\mu(Kx_0) > 0$. Now consider a maximal set of disjoint translates $Kg_1x_0, Kg_2x_0, \ldots, Kg_kx_0$ of Kx_0; since all of these sets have the same positive measure, such a maximal set exists and is finite. Then for any g in L, Kgx_0 must intersect one of the sets Kg_ix_0, which implies that $Lx_0 = \bigcup_{i=1}^k K^{-1}Kg_ix_0$. But the right-hand side is compact, and so Lx_0 is closed as desired. □

We return to the proof of Proposition 2.17.8. In view of Lemma 2.17.9, we may assume that μ is totally non-concentrated on L-orbits in the sense that

$$\mu(Lx) = 0 \text{ for all} x \in G/\Gamma. \tag{2.219}$$

In particular, for μ-almost every x and y, y does not lie in the orbit Lx of x and vice versa; informally, the group elements in G that are used to move from x to y should be somehow "transverse" to L. On the other hand, we are given that μ is ergodic with respect to H, and thus (by Corollary 2.17.7) ergodic with respect to U. This implies (cf. Proposition 2.9.13) that μ-almost every point x in G/Γ is *generic* (with respect to U) in the sense that

$$\int_{G/\Gamma} f \, d\mu = \lim_{T\to+\infty} \frac{1}{T} \int_0^T f(u^t x) \, dt. \tag{2.220}$$

for all continuous compactly supported $f : G/\Gamma \to \mathbf{R}$.

Exercise 2.17.2. Prove this claim. *Hint*: obtain continuous analogues of the theory from Sections 2.8, 2.9.

The equation (2.220) (and the *Riesz representation theorem*) lets us describe the measure μ in terms of the U-orbit of a generic point. On the other hand, from (2.219) and the ensuing discussion we see that any two generic points are likely to be separated from each other by some group element "transverse" to L. It is the interaction between these two facts which is going to generate the additional symmetry needed for Proposition 2.17.8. We illustrate this with a model case, in which the group element centralises U:

Proposition 2.17.10. *(central case). Let the notation and assumptions be as in Proposition 2.17.8. Suppose that x, y are generic points such that $y = gx$ for some $g \in G$ that centralises U (i.e. it commutes with every element of u). Then μ is invariant under the action of g.*

Proof. Let $f : G/\Gamma \to \mathbf{R}$ be continuous and compactly supported. Applying (2.220) with x replaced by $y = gx$ we obtain

$$\int_{G/\Gamma} f \, d\mu = \lim_{T\to+\infty} \frac{1}{T} \int_0^T f(u^t g x) \, dt. \tag{2.221}$$

Commuting g with u^t and using (2.220) again, we conclude

$$\int_{G/\Gamma} f \, d\mu = \int_{G/\Gamma} f(gy) \, d\mu(y) \tag{2.222}$$

and the claim follows from the Riesz representation theorem. □

Of course, we don't just want invariance under one group element g; we want a whole group L' of symmetries for which one has invariance. But it is not hard to leverage the former to the latter, provided one has enough group elements:

Lemma 2.17.11. *Let the notation and assumptions be as in Proposition 2.17.8. Suppose one has a sequence g_n of group elements tending to the identity, such that the action of each of the g_n preserve μ, and such that none of the g_n lie in L. Then there exists a closed connected subgroup $L < L' \leq G$ such that μ is L-invariant.*

Proof. Let S be the stabiliser of μ, i.e. the set of all group elements g whose action preserves μ. This is clearly a closed subgroup of G which contains L. If we let L' be the identity connected component of S, then L' is a closed connected subgroup containing L which will contain g_n for all sufficiently large n, and in particular is not equal to L. The claim follows. □

From Proposition 2.17.10 and Lemma 2.17.11 we see that we are done if we can find pairs $x_n, y_n = g_n x_n$ of nearby generic points with g_n going to the identity such that $g_n \notin L$ and that g_n centralises U. Now we need to consider the non-central case; thus suppose for instance that we have two generic points x, $y = gx$ in which g is close to the identity but does not centralise U. The key observation here is that we can use the U-invariance of the situation to pull x and y slowly apart from each other. More precisely, since x and y are generic, we observe that $u^t x$ and $u^t y$ are also generic for any t, and that these two points differ by the conjugated group element $g^t := u^t g u^{-t}$. Taking logarithms (which are well-defined as long as g^t stays close to the identity), we can write

$$\log(g^t) = u^t \log(g) u^{-t} = \exp(t \mathrm{ad}(\log u)) \log(g) \tag{2.223}$$

where ad is the *adjoint representation*. From Lemma 2.17.4, we know that $\mathrm{ad}(\log u) : \mathfrak{g} \to \mathfrak{g}$ is nilpotent, and so (by Taylor expansion of the exponential) $\log(g^t)$ depends polynomially on t. In particular, if g does not centralise U, then $\log(g^t)$ is non-constant and thus must diverge to infinity as $t \to +\infty$. In particular, given some small ball B around the origin in $\mathfrak{g}$ (with respect to some arbitrary norm), then

whenever $\log g$ lies inside B around the origin and is not central, there must be a first time $t = t_g$ such that $\log g^{t_g}$ reaches the boundary ∂B of this ball. We write $g^* := g^{t_g} \in \partial B$ for the location of g when it escapes. We now have the following variant of Proposition 2.17.10:

Proposition 2.17.12 (Non-central case)**.** *Let the notation and assumptions be as in Proposition 2.17.8. Suppose that $x_n, y_n \in G$ are generic points such that $y_n = g_n x_n$ for some $g_n \in G$ which do not centralise u, but such that g_n converge to the identity (in particular, $g_n \in B$ for all sufficiently large n). Suppose furthermore that x_n, y_n are* uniformly generic *in the sense that for any continuous compactly supported $f : G/\Gamma \to \mathbf{R}$, the convergence of* (2.220) *(with x replaced by x_n or y_n) is uniform in n. Then μ is invariant under the action of any limit point $g^* \in \partial B$ of the g_n^*.*

Proof. By passing to a subsequence if necessary we may assume that g_n^* converges to g^*. For each sufficiently large n, we write $T_n := t_{g_n}$, thus $g_n^t \in B$ for all $0 \leq t \leq T_n$, and $g_n^{T_n} = g_n^*$. We rescale this by defining the functions $h_n : [0,1] \to B$ by $h_n(s) := g_n^{sT_n}$. From the unipotent nature of U, these functions are polynomial (with bounded degree), and also bounded (as they live in B), and are thus equicontinuous (since all norms are equivalent on finite dimensional spaces). Thus, by the Arzelá-Ascoli theorem, we can assume (after passing to another subsequence) that h_n is uniformly convergent to some limit f, which is another polynomial. Since we already have $h_n(2.206) = g_n^*$ converging to g_*, this implies that for any $\varepsilon > 0$ there exists $\delta > 0$ such that $h_n(s) = g_* + O(\varepsilon)$ for all $1 - \delta \leq s \leq 1$ and all sufficiently large n. In other words, we have

$$u^t g_n u^{-t} = g_* + O(\varepsilon) \tag{2.224}$$

for sufficiently large n, whenever $(1-\delta)T_n \leq t \leq T_n$.

This is good enough to apply a variant of the Proposition 2.17.10 argument. Namely, if $f : G/\Gamma \to \mathbf{R}$ is continuous and compactly supported, then by uniform genericity we have for T sufficiently large that

$$\int_{G/\Gamma} f \, d\mu = \frac{1}{\delta T} \int_{(1-\delta)T}^{T} f(u^t y_n) \, dt + O(\varepsilon) \tag{2.225}$$

for all n. Applying (2.224) we can write $u^t y_n = g_* u^t x_n + O(\varepsilon)$ on the support of f, and so by uniform continuity of f

$$\int_{G/\Gamma} f \, d\mu = \frac{1}{\delta T} \int_{(1-\delta)T}^{T} f(g_* u^t x_n) \, dt + o(1) \tag{2.226}$$

where $o(1)$ goes to zero as $\varepsilon \to 0$, uniformly in n. Using (2.220) again and then letting $\varepsilon \to 0$, we obtain the g_*-invariance of μ as desired. □

Now we have all the ingredients to prove Proposition 2.17.8, and thus Theorem 2.17.1.

Proof of Proposition 2.17.8. We know that μ-almost every point is generic. Applying Egorov's theorem, we can find sets $E \subset G/\Gamma$ of measure arbitrarily close to 1 (e.g. $\mu(E) \geq 0.9$) on which the points are uniformly generic.

Now let V be a small neighbourhood the origin in L. Observe from the Fubini-Tonelli theorem that

$$\int_X \frac{1}{m(V)} \int_V 1_E(x) 1_E(gx) \; dm(g) d\mu(x) \geq 2\mu(E) - 1 \geq 0.8 \tag{2.227}$$

where m is the Haar measure on the unimodular group L, from which one can find a set $E' \subset E$ of positive measure such that $m(\{g \in V : gx \in E\}) = 0.7m(V)$ for all $x \in E'$; one can view E' as "points of density" of E in some approximate sense (and with regard to the L action).

Since E' has positive measure, and using (2.219), it is not hard to find sequences $x_n, y_n \in E'$ with $y_n \notin Lx_n$ for any n and with $\mathrm{dist}(x_n, y_n) \to 0$ (using some reasonable metric on G/Γ).

Exercise 2.17.3. Verify this. *Hint*: G/Γ can be covered by countably many balls of a fixed radius.

Next, recall that $H \equiv SL_2(\mathbf{R})$ acts by conjugation on the Lie algebra $\mathfrak{g}$ of G, and also leaves the Lie algebra $\mathfrak{l} \subset \mathfrak{g}$ of L invariant. By Lemma 2.17.2, this implies there is a complementary subspace W of $\mathfrak{l}$ in $\mathfrak{g}$ which is also H-invariant (and in particular, U-invariant). From the inverse function theorem, we conclude that for any group element g in G sufficiently close to the identity, we can factor $g = \exp(w)l$

where $l \in L$ is also close to the identity, and $w \in W$ is small (in fact this factorisation is unique). We let $\pi_L : g \mapsto l$ be the map from g to l; this is well-defined and smooth near the identity.

Let n be sufficiently large, and write $y_n = g_n x_n$ where g_n goes to the identity as n goes to infinity. Pick $l_n \in V$ at random (using the measure m conditioned to V). Using the inverse function theorem and continuity, we see that the random variable $\pi_L(l_n g_n)$ is supported in a small neighbourhood of V, and that its distribution converges to the uniform distribution of V (in, say, total variation norm) as $n \to \infty$. In particular, we see that $y'_n := l_n y_n \in E$ with probability at least 0.7 and $x'_n := \pi_L(l_n g_n) x_n \in E$ with probability at least 0.6 (say) if n is large enough. In particular we can find an $l_n \in V$ such that y'_n, x'_n both lie in E. Also by construction we see that $y'_n = \exp(w_n) x'_n$ for some $w_n \in W$; since $y_n \notin \mathrm{L} x_n$, we see that w_n is non-zero. On the other hand, since W is transverse to $\mathfrak{l}$ and the distance between x_n, y_n go to zero, we see that w_n goes to zero.

There are now two cases. If $\exp(w_n)$ centralises U for infinitely many n, then from Proposition 2.17.10 followed by Lemma 2.17.11 we obtain conclusion 2 of Proposition 2.17.8 as required. Otherwise, we may pass to a subsequence and assume that none of the $\exp(w_n)$ centralise U. Since W is preserved by U, we see that the group elements $\exp(w_n)^*$ also lie in $\exp(K)$ for some compact set K in W, and also on the boundary of B. This space is compact, and so by Proposition 2.17.12 we see that μ is invariant under some group element $g \in \exp(K) \cap \partial B$, which cannot lie in L. Since the ball B can be chosen arbitrarily small, we can thus apply Lemma 2.17.11 to again obtain conclusion 2 of Proposition 2.17.8 as required. □

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/03/15`.

Chapter 3

The Poincaré conjecture

3.1. Riemannian manifolds and curvature

In this preliminary section, I will quickly review the basic notions of infinitesimal[1] Riemannian geometry, and in particular defining the Riemann, Ricci, and scalar curvatures of a Riemannian manifold. This is a review only, in particular omitting any leisurely discussion of examples or motivation for Riemannian geometry; I will have to refer you to a textbook on the subject for a more complete treatment.

3.1.1. Smooth manifolds. Riemannian geometry takes place on *smooth manifolds*[2] M of some dimension $d = 0, 1, 2, \ldots$. Recall that a d-dimensional manifold (or *d-manifold* for short) M consists of the following structures[3]:

- A topological space M (which for technical reasons we assume to be Hausdorff and second countable);
- An *atlas* of *charts* $\phi_\alpha : U_\alpha \to V_\alpha$, which are homeomorphisms from open sets U_α in M to open sets V_α in $\mathbf{R}^d$, such that the U_α cover M.

We say that the manifold M is *smooth* if the charts ϕ_α define a consistent smooth structure, in the sense that the maps $\phi_\alpha \circ \phi_\beta^{-1}$ is smooth (i.e. infinitely differentiable) on $\phi_\beta(U_\alpha \cap U_\beta)$ for every α, β. One can then assert that a function $f : M \to X$ from M to another space with a smooth structure (e.g. $\mathbf{R}$ or $\mathbf{C}$) is smooth if $f \circ \phi_\alpha^{-1}$ is smooth on V_α for every α; a smooth map with an inverse which is also smooth is known as a *diffeomorphism*. The space of all smooth functions $f : M \to \mathbf{R}$ is denoted $C^\infty(M)$; this is a topological algebra over the reals. More generally, we have the algebra $C^\infty(U)$ for any open subset of M.

Remark 3.1.1. The most intuitive way to view manifolds is from an *extrinsic* viewpoint: as subsets of some larger-dimensional space

[1] The more "global" aspects of Riemannian geometry, for instance concerning the relationship between distance, curvature, injectivity radius, and volume, will be discussed later in this chapter.

[2] Unless otherwise stated, all manifolds are assumed to be without boundary.

[3] It is possible to view smooth manifolds more abstractly (and in a fully coordinate-independent fashion) by using the *structure sheaf* of algebras $C^\infty(U)$ to define the smooth structure, rather than the atlas of charts, but we will not need to take this perspective here.

(e.g. viewing curves as subsets of the plane, surfaces as subsets of a Euclidean space such as $\mathbf{R}^3$). While every smooth manifold can be viewed this way (thanks to the *Whitney embedding theorem*), we will in fact not use the extrinsic perspective at all in this course! Instead, we will rely exclusively on the intrinsic perspective - by studying the various structures on a smooth manifold M purely in terms of objects that can be defined in terms of the atlas. In fact, once we set up the most basic such structure - the *tangent bundle* - we will often not use the atlas directly at all (thus working in a "coordinate-free" fashion). However, the "local coordinates" provided by the charts in an atlas will be useful for computations at various junctures.

Remark 3.1.2. It is a surprising and unintuitive fact that a single topological manifold can have two distinct smooth structures which are not diffeomorphic to each other! This is most famously the case for 7-spheres S^7, giving rise to *exotic spheres*. However, in the case of 3-manifolds - which is the focus of this course - all smooth structures are diffeomorphic (a result of Munkres[**Mu1960**] and Whitehead[**Wh1961**]; see also Smale[**Sm1961**] for higher-dimensional variants), and so this subtlety need not concern us.

Remark 3.1.3. As $C^\infty(M)$ is commutative, we will multiply by functions in this space on the left or on the right interchangeably. In *noncommutative geometry*, this algebra is replaced by a noncommutative algebra, and one has to take substantially more care with the order of multiplication, but we will not use noncommutative geometry here.

We will be interested in various vector bundles over a smooth manifold M. A *vector bundle* V is a collection of (real) vector spaces V_x of a fixed dimension k (the *fibres* of the bundle) associated to each point $x \in M$, whose disjoint union $V = \biguplus_{x \in M} V_x$ can itself be given the structure of a smooth ($d+k$-dimensional) manifold, in such a way that for all sufficiently small neighbourhoods U of any given point x, the set $\biguplus_{x \in U} V_x$ has a *trivialisation*, i.e. there is a diffeomorphism between $\biguplus_{x \in U} V_x$ and $U \times \mathbf{R}^k$, with each fibre V_x being identified in a linearly isomorphic way with the vector space $\{x\} \times \mathbf{R}^k \equiv \mathbf{R}^k$. A (global) *section* of a vector bundle V is a smooth map $f : M \to V$ such that $f(x) \in V_x$ for every $x \in V$. The space of all

sections is denoted $\Gamma(V)$; it is a vector space over $\mathbf{R}$, and furthermore is a module over $C^\infty(M)$. We will sometimes also be interested in *local* sections $f : U \to V$ on some open subset U of M; the space of such sections (which form a module over $C^\infty(U)$) will be denoted $\Gamma(U \to V)$. All of the discussion below on the global manifold M can be easily adapted to local open sets U in this manifold (indeed, one can interpret U itself as a manifold); as all our computations will be entirely local (and because of the ready availability of smooth cutoff functions), the theory on M and the theory on U will be completely compatible.

Example 3.1.4. The space $C^\infty(M)$ can be canonically identified with the space of sections $\Gamma(M \times \mathbf{R})$ of the trivial line bundle $M \times \mathbf{R}$.

In Riemannian geometry, the most fundamental vector bundle over a manifold M is the *tangent bundle* TM, defined by letting the *tangent space* $T_x M$ at a point $x \in M$ be the space of all *tangent vectors* in M at x. A tangent vector $v \in T_x M$ can be defined as a vector which can be expressed as the (formal) derivative $v = \gamma'(0)$ of some smooth curve $\gamma : (-\varepsilon, \varepsilon) \to M$ which passes through x at time zero, thus $\gamma(0) = x$. One can express these tangent vectors concretely by using any chart that covers x.

To be somewhat informal, given any point $x \in M$ and tangent vector $v \in T_x M$, one can define a trajectory of points $x + tv + O(t^2) \in M$ for all "infinitesimal" t, which is only defined up to an error of $O(t^2)$ (as measured, for instance, in some coordinate chart), but whose derivative at $t = 0$ is equal to v. Thus, while the global manifold M need not have any reasonable notion of vector addition, we do have this infinitesimal notion of translation by a tangent vector which is well-defined up to second-order errors.

Given a tangent vector $v \in T_x M$ and a smooth function $f \in C^\infty(M)$, we can define the *directional derivative* $\nabla_v f(x)$ by the formula

$$\nabla_v f(x) := \lim_{t \to 0} \frac{f(x + vt + O(t^2)) - f(x)}{t} \tag{3.1}$$

(or, a bit more formally, $\nabla_v f(x) = \frac{d}{dt} f(\gamma(t))$ for any curve $\gamma : (-\varepsilon, \varepsilon) \to M$ with $\gamma(0) = x$ and $\gamma'(0) = v$). This is a linear functional

on $C^\infty(M)$ which annihilates constants and obeys the *Leibniz rule*

$$\nabla_v(fg)(x) = f(x)\nabla_v g(x) + \nabla_v f(x) g(x). \tag{3.2}$$

Conversely, one can *define* the tangent space $T_x V$ to be the space of all linear functionals on $C^\infty(M)$ with the above two properties, though we will not need to do so here.

A section $X \in \Gamma(TM)$ of M is known as a *vector field*; it assigns a tangent vector $X(x) \in T_x M$ to each point $x \in M$. A vector field X determines a *first-order differential operator* $\nabla_X : C^\infty(M) \to C^\infty(M)$, defined by setting $\nabla_X f(x) := \nabla_{X(x)} f(x)$. From (3.2), we see that ∇_X is a *derivation*, i.e. it is linear over $\mathbf{R}$ and obeys the Leibniz rule

$$\nabla_X(fg) = f\nabla_X g + (\nabla_X f) g. \tag{3.3}$$

Conversely, one can easily show that every derivation on $C^\infty(M)$ arises uniquely in this manner. This provides a convenient means to define new types of vector fields. For example, if X and Y are two vector fields, one can easily see (from (3.3)) that the commutator $[\nabla_X, \nabla_Y] := \nabla_X \nabla_Y - \nabla_Y \nabla_X$ is also a derivation, and must thus be given by another vector field $[X, Y]$, thus

$$\nabla_X \nabla_Y f - \nabla_Y \nabla_X f - \nabla_{[X,Y]} f = 0 \tag{3.4}$$

for all vector fields X, Y and all scalar fields f.

Example 3.1.5. Suppose we have a local coordinate chart $\phi : U \to V \subset \mathbf{R}^d$. The standard first-order differential operators $\frac{d}{dx^1}, \ldots, \frac{d}{dx^d}$ induced by the coordinates $x^1, \ldots, x^d$ on $\mathbf{R}^d$ can be viewed as vector fields, and pulled back via ϕ to vector fields $\phi^* \frac{d}{dx^1}, \ldots, \phi^* \frac{d}{dx^d}$ on U. These in fact form a *frame* for U since they span the tangent space at every point. Since $\frac{d}{dx^i}$ and $\frac{d}{dx^j}$ commute in $\mathbf{R}^d$, we see that $[\phi^* \frac{d}{dx^i}, \phi^* \frac{d}{dx^j}] = 0$.

Exercise 3.1.1. Show that the map $(X, Y) \mapsto [X, Y]$ endows the space $\Gamma(TM)$ of vector fields with the structure of an abstract Lie algebra. Also establish the Leibniz rule

$$[X, fY] = (\nabla_X f) Y + f[X, Y] \tag{3.5}$$

for all $X, Y \in \Gamma(TM)$ and $f \in C^\infty(M)$.

Various operations on finite-dimensional vector spaces generalise easily to vector bundles. For instance, every finite-dimensional vector space V has a dual V^*, and similarly every vector bundle V also has a dual bundle V^*, whose fibres V_x^* are the dual to the fibres V_x of V; one can also view V^* as the space of $C^\infty(M)$-linear functionals from V to $C^\infty(M)$. Similarly, given two vector bundles V, W over M, one can define the *direct sum* $V \oplus W$, the *tensor product* $V \otimes W$, the space $\mathrm{Hom}(V, W)$ of fibre-wise linear transformations from V to W, the *symmetric powers* $\mathrm{Sym}^k(V)$ and *exterior powers* $\bigwedge^k(V)$, and so forth. The construction of all of these concepts is straightforward but rather tedious, and will be omitted here.

Applying these constructions to the tangent bundle TM, one gets a variety of useful bundles for doing Riemannian geometry:

- The bundle $T^*M := (TM)^*$ is the *cotangent bundle*; elements of T_x^*M are *cotangent vectors*.
- Sections of $\bigwedge^k(T^*M)$ are known as *k-forms*.
- Sections of $(TM)^{\otimes k} \otimes (T^*M)^{\otimes l}$ are known as *rank (k,l) tensor fields*, and individual elements of this bundle are *rank (k,l) tensors*. Many tensors of interest obey various symmetry or antisymmetry properties[4], for instance k-forms are totally anti-symmetric rank $(0,k)$ tensors.

It is convenient to use *abstract index notation*, denoting rank (k,l) tensor fields using k superscripted Greek indices and l subscripted Greek indices, thus for instance $\mathrm{Riem} = \mathrm{Riem}^\delta_{\alpha\beta\gamma}$ denotes a rank $(1,3)$ tensor. One should think of these indices as placeholders; if one chooses a *frame* $(e_a)_{a\in A}$ for the tangent bundle (i.e. a collection of vector fields which form a basis for the tangent space at every point), which induces the associated *dual frame* $(e^a)_{a\in A}$ for the cotangent bundle, then this notation can be viewed as describing the coefficients of the tensor in terms of the basis generated by such frames, thus for

[4]To fully enumerate the various symmetry properties available to tensors is a task essentially equivalent to understanding the finite-dimensional representation theory of the permutation group; this is a beautiful and important subject, but will not be discussed here.

instance

$$\text{Riem} = \sum_{a,b,c,d \in A} \text{Riem}^d_{abc} e^a \otimes e^b \otimes e^c \otimes e_d. \tag{3.6}$$

But it is perhaps better to view a tensor such as $\text{Riem}^\delta_{\alpha\beta\gamma}$ as existing independently of any choice of frame, in which case the labels $\alpha, \beta, \gamma, \delta$ are abstract placeholders.

Example 3.1.6. We continue Example 3.1.5. A local coordinate chart $\phi : U \to \mathbf{R}^d$ generates a (local) frame $e_a := \phi^* \frac{d}{dx^a}$ with an associated dual frame $e^a := \phi^*(dx^a)$. These frames can be slightly easier to work with for computations than general frames, because we automatically have $[e_a, e_b] = 0$ as already noted in Example 3.1.5. On the other hand, it is often convenient to work in frames that don't come from coordinate charts in order to obtain other good properties; in particular, it is very convenient to work in *orthonormal* frames, which are usually unavailable if one restricts attention to frames arising from coordinate charts.

We use the usual (and very handy) *Einstein summation convention*: repeated indices (with each repeated index appearing exactly once as a superscript and once as a subscript) are implicitly summed over a choice of frame (the exact choice is not important). For instance, the rank $(0,4)$ tensor $X_{\alpha\beta\sigma\mu} := \text{Riem}^\delta_{\alpha\beta\gamma} \text{Riem}^\gamma_{\delta\sigma\mu}$ is defined to be the tensor which is given by the formula

$$X_{absm} = \sum_{g,d \in A} \text{Riem}^d_{abg} \text{Riem}^g_{dsm} \tag{3.7}$$

for any choice of frame $(e_a)_{a \in A}$ (one can easily verify that this definition is independent of the choice of frame). We will also apply this summation convention when the Greek labels are replaced with concrete counterparts arising from a frame, thus for instance we can now abbreviate (3.6) as

$$\text{Riem} = \text{Riem}^d_{abc} e^a \otimes e^b \otimes e^c \otimes e_d \tag{3.8}$$

3.1.2. Connections. We have seen that vector fields $X \in \Gamma(TM)$ allow us to differentiate scalar functions $f \in C^\infty(M)$ to obtain a

differentiated function $\nabla_X f$. Furthermore, this concept obeys the Leibniz rule (3.3), and is linear over $C^\infty(M)$ in X, or in other words

$$\nabla_{gX+Y} f = g\nabla_X f + \nabla_Y f \tag{3.9}$$

for all $g \in C^\infty(M)$ and $X, Y \in \Gamma(TM)$. As a consequence, one can interpret $X \mapsto \nabla_X f$ as a $C^\infty(M)$-linear functional on $\Gamma(TM)$, which is identified with a section $df \in \Gamma(T^*M)$ of the cotangent bundle, thus $\nabla_X f = df(X)$.

Now suppose one wants to differentiate $\nabla_X f$, where $f \in \Gamma(V)$ is now a section of a bundle V. It turns out that there is now more than one good notion of differentiation. Each such notion can be formalised by the concept of an (linear) connection:

Definition 3.1.7. A *connection* ∇ on a bundle V is an assignment of a section $\nabla_X f \in \Gamma(V)$ (the *covariant derivative* of f in the direction X via the connection ∇) to each vector field $X \in \Gamma(TM)$ and section $f \in \Gamma(V)$, in such a way that $(f, X) \mapsto \nabla_X f$ is bilinear in f and X, that the Leibniz rule (3.3) is obeyed for $f \in C^\infty(M)$ and $g \in \Gamma(V)$ (or vice versa), and the linearity rule (3.9) is obeyed for all $g \in C^\infty(M)$ and $X, Y \in \Gamma(TM)$.

If $f \in \Gamma(V)$ is such that $\nabla_X f = 0$ for all vector fields X, we say that f is *parallel* to the connection ∇.

A connection on the tangent bundle TM is known as an *affine connection*.

Remark 3.1.8. Informally, a connection assigns an infinitesimal linear isomorphism $\phi_v : V_x \to V_{x+v}$ (the *parallel transport map*) to each infinitesimal tangent vector $v \in V$, in a manner which is linear in v for fixed x. The connection between this informal definition and the above formal one is given by the formula $\nabla_X f(x) = \lim_{t\to 0} \frac{\phi_{tX(x)}^{-1}(f(x+tX(x))) - f(x)}{t}$. One can make this informal definition more precise (e.g. using non-standard analysis, as in Section 1.5 of *Structure and Randomness*) but we will not do so here. An alternate definition of a connection is as a complementary subbundle to the *vertical bundle* $\biguplus_{x\in M} TV_x$ in TV, known as a *horizontal bundle*, obeying some additional linearity conditions in the vertical variable.

Once one has a connection on a bundle V, one automatically can define a connection on the dual bundle V^* and more generally on tensor powers $V^{\otimes k} \otimes (V^*)^{\otimes l}$, by enforcing all possible instances of the Leibniz rule[5], e.g.

$$\nabla_X(f_\gamma^{\alpha\beta} g_\delta^\gamma) = (\nabla_X f_\gamma^{\alpha\beta}) g_\delta^\gamma + f_\gamma^{\alpha\beta} \nabla_X g_\delta^\gamma \tag{3.10}$$

for all rank $(2,1)$ tensors f and rank (1,1) tensors g. In particular, any connection on the tangent bundle (which is the case of importance in Riemannian geometry) naturally induces a connection on the cotangent bundle and the bundle of rank (k,l) tensors.

Here it is important to note that the indices are abstract, rather than corresponding to some frame: for instance, if ∇ is a connection on the tangent bundle TM, then after choosing a frame $(e_a)_{a\in A}$, it is usually *not* the case that the coefficient $(\nabla_X f)^a$ of a vector field $f \in \Gamma(M)$ at a is equal to the derivative $\nabla_X(f^a)$ of that component of f. Instead, one has a relationship of the form

$$(\nabla_X f)^a = \nabla_X(f^a) + \Gamma_{bc}^a X^b f^c \tag{3.11}$$

where for each a, b, c, the *Christoffel symbol* $\Gamma_{bc}^a := e^a(\nabla_{e_b} e_c)$ of the connection relative to the frame $(e_a)_{a\in A}$ is a smooth function on M. It is important to note that Christoffel symbols are *not* tensors, because the expression $\Gamma_{bc}^a e_a \otimes e^b \otimes e^c$ turns out to depend on the choice of frame.

Using the Leibnitz rule repeatedly, it is not hard to use (3.11) to give a formula for the components of derivatives of other tensors, e.g.

$$(\nabla_X \omega)_a = \nabla_X(\omega_a) - \Gamma_{ba}^c X^b \omega_c \tag{3.12}$$

for any 1-form ω,

$$(\nabla_X g)_{ab} = \nabla_X(g_{ab}) - \Gamma_{da}^c X^d g_{cb} - \Gamma_{db}^c X^d g_{ac} \tag{3.13}$$

for any rank $(0,2)$ tensor g, and so forth.

We have remarked that Christoffel symbols are not tensors. On the other hand, because $\nabla_X f$ is linear in X, we can legitimately define a tensor field $\nabla_\alpha f$, which is a section of $T^*M \otimes V \equiv \mathrm{Hom}(TM, V)$,

[5]It is a straightforward but tedious task to verify that all the Leibniz rules are consistent with each other, and that (3.10) and its relatives uniquely define a connection on every tensor power of V.

thus $\nabla_X f = X^\alpha \nabla_\alpha f$. It is also possible to express the *difference* of two connections as a tensor:

Exercise 3.1.2. Let ∇, ∇' be two connections on TM. Show that there exists a unique rank $(1,2)$ tensor $\Gamma^\alpha_{\beta\gamma} = \nabla' - \nabla$ such that

$$\nabla'_\beta f^\alpha - \nabla_\beta f^\alpha = \Gamma^\alpha_{\beta\gamma} f^\gamma \tag{3.14}$$

for all vector fields f^α. Now interpret the Christoffel symbol Γ^a_{bc} of a connection ∇ on TM relative to a frame $e = (e_a)_{a\in A}$ as the difference $\nabla - \nabla^{(e)}$ of that connection with the *flat connection* $\nabla^{(e)}$ induced by the trivialisation of the tangent bundle induced by that frame.

Let ∇ be a connection on TM. We say that this connection is *torsion-free* if we have the pleasant identity

$$\nabla_\alpha \nabla_\beta f = \nabla_\beta \nabla_\alpha f \tag{3.15}$$

(cf. *Clairaut's theorem* from several variable calculus) for all scalar fields $f \in C^\infty(M)$, or in other words that the *Hessian* $\mathrm{Hess}(f)_{\alpha\beta} := \nabla_\alpha \nabla_\beta f$ of f is a symmetric rank $(0,2)$ tensor.

Exercise 3.1.3. Show that ∇ is torsion-free if and only if

$$[X,Y]^\alpha = X^\beta \nabla_\beta Y^\alpha - Y^\beta \nabla_\beta X^\alpha \tag{3.16}$$

for all vector fields X, Y (or in coordinate-free notation, $[X,Y] = \nabla_X Y - \nabla_Y X$).

Remark 3.1.9. Roughly speaking, the torsion-free connections are those which have a good notion of an infinitesimal *parallelogram* with corners $x, x+tv+O(t^2), x+tw+O(t^2), x+tv+tw+O(t^2)$ for some infinitesimal t, such that each edge is the parallel transport of the opposing edge to error[6] $O(t^3)$.

It would be nice if (3.15) extended to tensor fields f. This is true for flat connections, but false in general. The defect in (3.15) for such fields is measured by the *curvature tensor* $R \in \Gamma(\mathrm{Hom}(\bigwedge^2 TM, \mathrm{Hom}(TM,TM)))$ of the connection ∇, defined by the formula

$$\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z =: R(X,Y)Z \tag{3.17}$$

[6]Without the torsion-free hypothesis, the error is merely $O(t^2)$.

for all vector fields X, Y, Z (cf. (3.4)). One easily sees that R is indeed a section of $\mathrm{Hom}(\bigwedge^2 TM, \mathrm{Hom}(TM, TM))$ and can thus be viewed as a rank $(1,3)$ tensor.

Exercise 3.1.4. If ∇ is a torsion-free connection on TM, and $R^\delta_{\alpha\beta\gamma}$ is the tensor form of the curvature R, defined by requiring that

$$(R(X,Y)Z)^\delta = R^\delta_{\alpha\beta\gamma} X^\alpha Y^\beta Z^\gamma, \tag{3.18}$$

then show that

$$\nabla_\alpha \nabla_\beta X^\delta - \nabla_\beta \nabla_\alpha X^\delta = R^\delta_{\alpha\beta\gamma} X^\gamma \tag{3.19}$$

for all vector fields X^δ. What is the analogue of (3.19) if X^δ is replaced by a rank (k,l) tensor?

Connections describe a way to transport tensors as one moves from point to point in the manifold. There is another way to transport tensors, which is induced by diffeomorphisms $\phi : M \to M$ of the base manifold; this transportation procedure maps points $x \in M$ to points $\phi(x) \in M$, maps tangent vectors $v \in T_x M$ to tangent vectors $\phi_*(v) \in T_{\phi(x)} M$ (defined by requiring that the chain rule $\frac{d}{dt}(\gamma \circ \phi) = \phi_*(\frac{d}{dt}\gamma)$ hold for all curves γ) and then maps other tensors in the unique manner consistent with the tensor operations (e.g. $\phi_*(v \otimes w) = \phi_*(v) \otimes \phi_*(w)$). This procedure is important for describing symmetries of tensor fields (consider, for instance, what it means for the vector field $(y, -x)$ in $\mathbf{R}^2$ to be invariant under rotations around the origin). To relate this diffeomorphism transport to infinitesimal differential geometry, though, we have to look at an *infinitesimal* diffeomorphism, which we can view as the derivative $\frac{d}{dt}\phi_t|_{t=0}$ of a smoothly varying family ϕ_t of diffeomorphisms, with ϕ_0 equal to the identity. By chasing all the definitions we see that $\frac{d}{dt}\phi_t|_{t=0}$ is just a vector field X. The infinitesimal rate of change $\frac{d}{dt}\phi_*(v)|_{t=0}$ of a tensor field v under this diffeomorphism is known as the *Lie derivative* $\mathcal{L}_X v$ of v with respect to the vector field X (it does not depend on any aspect of ϕ other than its infinitesimal vector field). On scalars f, it agrees with directional derivative

$$\mathcal{L}_X f = \nabla_X f, \tag{3.20}$$

while on vector fields Y, it agrees with the commutator:

$$\mathcal{L}_X Y = [X, Y], \tag{3.21}$$

and its action on all other tensors can be given by the Leibniz rule (as is the case for connections). It should be emphasised, though, that the Lie derivative is *not* a connection, because it is not linear (over $C^\infty(M)$) in X; $\mathcal{L}_{fX} w \neq f \mathcal{L}_X w$ in general.

3.1.3. Riemannian manifolds and curvature tensors. We now specialise our attention from smooth manifolds to our main topic of interest, namely *Riemannian manifolds*. Informally, a Riemannian manifold is a manifold equipped with notions of length, angle, area, etc. which are infinitesimally isomorphic at every point to the corresponding notions in Euclidean space. In Euclidean space, all these geometric notions can be defined in terms of a positive definite inner product, and Riemannian manifolds are similarly founded on a positive definite *Riemannian metric*.

Definition 3.1.10. A *Riemannian manifold* (M, g) is a smooth manifold M, together with a *Riemannian metric* $g = g_{\alpha\beta}$ on M, i.e. a section of $\mathrm{Sym}^2(T^*M)$ which is positive definite in the sense that $g(v, w) := \langle v, w \rangle_{g(x)} := g_{\alpha\beta}(x) v^\alpha w^\beta$ is a positive-definite inner product on $T_x M$ for every point x.

We now use the metric g to build several other tensors of interest. Firstly, we have the inverse metric $g^{-1} = g^{\alpha\beta}$, which is the unique rank $(2,0)$ tensor that inverts the $(0,2)$ tensor g in the sense that $g^{\alpha\beta} g_{\beta\gamma} = g_{\gamma\beta} g^{\beta\alpha} = \delta^\alpha_\gamma$ is the identity section of $\mathrm{Hom}(TM, TM)$; this tensor is also symmetric and positive-definite. One can use these tensors to raise and lower the indices of other tensors; for instance, given a rank $(0,2)$ tensor $\pi_{\alpha\beta}$, one can define the rank $(1,1)$ tensors $\pi_\alpha^{\ \beta} = g^{\beta\gamma} \pi_{\alpha\gamma}$ and $\pi^\beta_{\ \alpha} = g^{\beta\gamma} \pi_{\gamma\alpha}$ and the rank $(2,0)$ tensor $\pi^{\alpha\beta} := g^{\alpha\gamma} g^{\beta\delta} \pi_{\gamma\delta}$. We will generally only use these conventions when there is enough symmetry that there is no danger of ambiguity.

Remark 3.1.11. All Riemannian manifolds can be viewed extrinsically (locally, at least) as subsets of a Euclidean space, thanks to the famous *Nash embedding theorem*. But we will not need this extrinsic viewpoint in this course.

After the metric, the next fundamental object in Riemannian geometry is the *Levi-Civita connection*.

Theorem 3.1.12 (Fundamental theorem of Riemannian geometry)**.** *Let (M, g) be a Riemannian manifold. Then there exists a unique affine connection ∇ (which is known as the* Levi-Civita connection*) which is torsion-free and* respects *the metric g in the sense that $\nabla g = 0$.*

Exercise 3.1.5. Prove this theorem. *Hint*: one can either

- Use abstract index notation and study expressions such as $\nabla_\alpha X^\beta$;
- Use coordinate-free notation and study expressions such as $g(\nabla_X Y, Z)$; or
- Use local coordinates (e.g. use a frame $e_a := \phi^* \frac{d}{dx^a}$ arising from a chart ϕ as in Example 3.1.5) and work with the Christoffel symbols Γ^a_{bc}.

It is instructive to do this exercise in all three possible ways in order to appreciate the equivalence (and relative advantages and disadvantages) between these three perspectives.

Geometrically, the condition $\nabla g = 0$ asserts that parallel transport by the Levi-Civita connection is an isometry. At a computational level, it means (in conjunction with the Leibnitz rule) that covariant differentiation using the Levi-Civita connection commutes with the raising and lowering operations, for instance given a vector field X^α we have

$$(\nabla_\alpha X)_\beta = g_{\beta\gamma} \nabla_\alpha X^\gamma = \nabla_\alpha (g_{\beta\gamma} X^\gamma) = \nabla_\alpha (X_\beta) \tag{3.22}$$

and so we may safely use raising and lowering operations in the presence of Levi-Civita covariant derivatives without much risk of serious error. We can also raise and lower the covariant derivative itself, defining

$$\nabla^\alpha := g^{\alpha\beta} \nabla_\beta = \nabla_\beta g^{\alpha\beta}. \tag{3.23}$$

This leads to the *covariant Laplacian* (or *Bochner Laplacian*)

$$\Delta := \nabla_\alpha \nabla^\alpha = \nabla^\alpha \nabla_\alpha = g^{\alpha\beta} \nabla_\alpha \nabla_\beta \tag{3.24}$$

defined on all tensor fields (for instance, when applied to scalar fields it becomes the *trace* of the Hessian, and is known as the *Laplace-Beltrami operator*). When applied to non-scalar fields, the covariant Laplacian differs slightly from the *Hodge Laplacian* (or *Laplace-de Rham operator*) $d^*d + dd^*$ by a lower order term which is given by the *Weitzenböck identity*.

As discussed earlier, all connections on TM have a curvature tensor in $\mathrm{Hom}(\bigwedge^2 TM, \mathrm{Hom}(TM, TM))$. The curvature of the Levi-Civita connection is known as the *Riemann curvature tensor* $\mathrm{Riem} = \mathrm{Riem}^\delta_{\alpha\beta\gamma}$, thus

$$\nabla_\alpha \nabla_\beta X^\delta - \nabla_\beta \nabla_\alpha X^\delta = \mathrm{Riem}^\delta_{\alpha\beta\gamma} X^\gamma. \tag{3.25}$$

One can also write Riem in co-ordinate free notation by defining $\mathrm{Riem}(X,Y)Z$ for vector fields X, Y, Z by the formula

$$\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z = \mathrm{Riem}(X,Y)Z \tag{3.26}$$

or equivalently as $(\mathrm{Riem}(X,Y)Z)^\delta = \mathrm{Riem}^\delta_{\alpha\beta\gamma} X^\alpha Y^\beta Z^\gamma$.

Because ∇ respects g, one eventually deduces from (3.25) and the Leibniz rule that $\mathrm{Riem}^\delta_{\alpha\beta\gamma}$ is skew-adjoint in the γ, δ indices:

$$\mathrm{Riem}^\delta_{\alpha\beta\gamma} = -g_{\gamma\mu} g^{\delta\sigma} \mathrm{Riem}^\mu_{\alpha\beta\sigma}. \tag{3.27}$$

It is also clearly skew-symmetric in the α, β indices. Also, from the analogue of (3.25) for 1-forms, i.e.

$$\nabla_\alpha \nabla_\beta \omega_\delta - \nabla_\beta \nabla_\alpha \omega_\delta = -\mathrm{Riem}^\gamma_{\alpha\beta\delta} \omega_\gamma \tag{3.28}$$

and the torsion-free nature of the connection, we have

$$\nabla_\alpha \nabla_\beta \nabla_\delta f - \nabla_\beta \nabla_\alpha \nabla_\delta f = -\mathrm{Riem}^\gamma_{\alpha\beta\delta} \nabla_\gamma f \tag{3.29}$$

for all scalar fields f. Cyclically summing this in α, β, δ we obtain the *first Bianchi identity*

$$\mathrm{Riem}^\gamma_{\alpha\beta\delta} + \mathrm{Riem}^\gamma_{\beta\delta\alpha} + \mathrm{Riem}^\gamma_{\delta\alpha\beta} = 0. \tag{3.30}$$

Exercise 3.1.6. Show that the above three symmetries of Riem imply that Riem is a self-adjoint section of $\mathrm{Hom}(\bigwedge^2 TM, \bigwedge^2 TM)$, and that these conditions are in fact equivalent in three and fewer dimensions. What happens in four dimensions?

Exercise 3.1.7. By differentiating (3.25) and cyclically summing, establish the *second Bianchi identity*

$$\nabla_\mu \mathrm{Riem}^\gamma_{\alpha\beta\delta} + \nabla_\beta \mathrm{Riem}^\gamma_{\mu\alpha\delta} + \nabla_\alpha \mathrm{Riem}^\gamma_{\beta\mu\delta} = 0. \tag{3.31}$$

Exercise 3.1.8. Show that a Riemannian manifold (M, g) is locally isomorphic (as Riemannian manifolds) to Euclidean space if and only if the Riemann curvature tensor vanishes. *Hint*: one direction is easy. For the other direction, the quickest way is to apply the *Frobenius theorem* to obtain a local trivialisation of the tangent bundle which is flat with respect to the Levi-Civita connection.

This illustrates the point that the Riemann curvature captures all the local obstructions that prevent a Riemannian manifold from being flat. (Compare this situation with the superficially similar subject of symplectic geometry, in which *Darboux's theorem* guarantees that there are no local obstructions whatsoever to a symplectic manifold (M, ω) being flat.)

The Riemann curvature measures the "infinitesimal monodromy" of parallel transport. For our applications we will need to study a slightly different curvature, the *Ricci curvature* $\mathrm{Ric}_{\alpha\beta}$, which measures how much the volume-radius relationship on infinitesimal sectors has been distorted from the Euclidean one[7]. It is defined as the trace of the Riemannian tensor, or more precisely as[8]

$$\mathrm{Ric}_{\alpha\beta} := \mathrm{Riem}^\gamma_{\gamma\alpha\beta}. \tag{3.32}$$

We also write $\mathrm{Ric}(X, Y)$ for $\mathrm{Ric}_{\alpha\beta} X^\alpha Y^\beta$ when X, Y are vector fields. The symmetries of Riem easily imply that Ric is a symmetric rank $(2, 0)$ tensor - just like the metric g! This observation[9] will of course be vital for defining Ricci flow later.

---

[7] This will not be obvious presently, as we have not yet defined the volume measure dg on a Riemannian manifold, but will be made clearer later.

[8] One could also contract other indices than these, but due to the various symmetry properties of the Riemann tensor, one ends up with essentially the same tensor as a consequence.

[9] This observation, as well as a similar observation for the *stress-energy tensor*, was also decisive in leading Einstein to the equations of general relativity, but that's a whole other story.

We can take the trace of the Ricci tensor to form the *scalar curvature*

$$R := g^{\alpha\beta}\mathrm{Ric}_{\alpha\beta} = g^{\alpha\beta}\mathrm{Riem}^{\gamma}_{\gamma\alpha\beta}; \tag{3.33}$$

up to normalisations, R can also be viewed as the trace of the Riemann tensor (viewed as a section of $\mathrm{Hom}(\bigwedge^2 TM, \bigwedge^2 TM)$). The scalar curvature measures how the relationship of volume of infinitesimal balls to their radius is distorted by the geometry.

The relationship between the Riemannian, Ricci, and scalar curvatures depends on the dimension:

(1) In one dimension, all three curvatures vanish; there are no degrees of freedom.

(2) In two dimensions, the Riemannian and Ricci curvatures are just multiples of the scalar curvature (by some tensor depending algebraically on the metric); there is only one degree of freedom.

(3) In three dimensions, the Riemann tensor is a linear combination of the Ricci curvature (see also Exercise 3.1.8 below). On the other hand, the scalar curvature does not control Ricci (or Riemann); the Ricci tensor contains an additional trace-free component. (However, once we start evolving by Ricci flow, we shall see that the *Hamilton-Ivey pinching phenomenon* will allow us to use the scalar curvature to mostly control Ricci and hence Riemann near singularities; see Section ???.)

(4) In four and higher dimensions, the Riemann tensor is not fully controlled by the Ricci curvature; there is an additional component to the Riemann tensor, namely the *Weyl tensor*. Similarly, the Ricci curvature is not fully controlled by the scalar curvature.

Exercise 3.1.9. (Ricci controls Riemann in three dimensions) In three dimensions, suppose that the (necessarily real) eigenvalues of the Riemann curvature at a point x (viewed as an element of $\mathrm{Hom}(\bigwedge^2 TM, \bigwedge^2 TM$ are λ, μ, ν. Show that the eigenvalues of the Ricci curvature at x

(viewed as an element of $\mathrm{Hom}(TM, TM)$ are $\lambda+\mu, \mu+\nu, \nu+\lambda$. Conclude in particular that

$$\|\mathrm{Riem}\|_g = O(\|\mathrm{Ric}\|_g) \tag{3.34}$$

where we endow the (fibres of the) spaces $\mathrm{Hom}(\bigwedge^2 TM, \bigwedge^2 TM)$ and $\mathrm{Hom}(TM, TM)$ with the Hilbert (or Hilbert-Schmidt) structure induced by the metric g.

Remark 3.1.13. The fact that Ricci controls Riemann in three dimensions, without itself degenerating into scalar curvature or zero, seems to explain why Ricci flow is especially powerful in three dimensions; it is still useful, but harder to work with, in two dimensions, useless in one dimension, and too weak to fully control the geometry in four and higher dimensions. It seems to me that the special nature of three dimensions stems from the fact that it is the unique number of dimensions in which 2-forms (which are naturally associated with curvature) are Hodge dual to vector fields (as opposed to scalars, or to higher-rank tensors); this is the same special feature of three dimensions which gives us the cross product (as opposed to the more general wedge product).

Because of the variety of curvatures, there are various notions of what it means for a manifold to have "non-negative curvature" at some point.

Definition 3.1.14. Let x be a point on a Riemannian manifold (M,g). We say that x has

(I) *non-negative scalar curvature* if $R(x) \geq 0$;

(II) *non-negative Ricci curvature* if $\mathrm{Ric}(x) \geq 0$ as a quadratic form on TM, i.e. $\mathrm{Ric}_{\alpha\beta}(x)v^\alpha v^\beta \geq 0$ for all vectors $v \in T_xM$;

(III) *non-negative sectional curvature* if $g(\mathrm{Riem}(x)(X,Y)X,Y)(x) = \mathrm{Riem}^\delta_{\alpha\beta\gamma}(x)X_\alpha Y_\beta X^\gamma Y_\delta \geq 0$ for all vectors $X, Y \in T_xM$;

(IV) *non-negative Riemann curvature* if $\mathrm{Riem}(x) \geq 0$ as a quadratic form on $\bigwedge^2 TM$, thus $\mathrm{Riem}^\delta_{\alpha\beta\gamma}(x)\omega^{\alpha\beta}(x)\omega^\gamma{}_\delta(x) \geq 0$ for all two-forms ω.

It is not hard to show that, in arbitrary dimension, (IV) implies (III) implies (II) implies (I). In one dimension, these conditions are

vacuously true; in two dimensions; these conditions are all equivalent; and in three dimensions, non-negative Riemann curvature is equivalent to non-negative sectional curvature (because every 2-form is the wedge product of two one-forms in this case) but these conditions are otherwise distinct. In four and higher dimensions all of these conditions are distinct. One can also define the analogous notions of positive curvature (or negative curvature, or non-positive curvature) in the usual manner.

Remark 3.1.15. Geometrically, positive scalar curvature means that infinitesimal balls have slightly less volume than in the Euclidean case; positive Ricci curvature means that infinitesimal sectors have slightly less volume than in the Euclidean case; and positive sectional curvature means that all infinitesimally geodesic two-dimensional surfaces have positive mean curvature. I don't know of a geometrically simple way to describe positive Riemann curvature.

We now give a "cartoon" or "schematic" description of these curvatures when viewed in some local coordinate system ϕ, using the associated frame $e_a := \phi^* \frac{d}{dx^a}$ as in Example 3.1.5 to express all tensors as arrays of numbers. Writing $g_{ab} = O(g)$, we thus schematically have the following relationships:

(1) The Christoffel symbols Γ^a_{bc} are schematically of the form $O(g^{-1}\partial g)$. Thus a covariant derivative $\nabla_a w$ of a tensor w looks schematically like $O(\partial w + g^{-1}(\partial g)w)$, and the Laplacian Δw looks like $O(g^{-1}\partial^2 w + g^{-2}(\partial g)\partial w + g^{-2}(\partial^2 g)w + g^{-3}(\partial g)^2 w)$.

(2) The Riemann curvature tensor Riem^d_{abc} and the Ricci curvature tensor Ric_{ab} schematically take the form $O(g^{-1}\partial^2 g + g^{-2}(\partial g)^2)$.

(3) The scalar curvature R schematically takes the form $O(g^{-2}\partial^2 g + g^{-3}(\partial g)^2)$. (Thus the scalar curvature has the same scaling as the Laplacian.)

Remark 3.1.16. Note how in all of these expressions, the "number of derivatives" and "number of g's" stays fixed among all terms in a given expression. This can be viewed as an example of *dimensional analysis* in action, and is useful for catching errors in manipulations with these

sorts of expressions. From a more representation-theoretic viewpoint, what is going on is that all of the above expressions have constant weight with respect to the joint (commuting) actions of the dilation operation $x^i \mapsto \lambda x^i$ on the underlying coordinate chart (which essentially controls the number of derivatives ∂ that appear) and the homogeneity operation $g \mapsto cg$ (which, naturally enough, controls the number of g's that appear).

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/03/26`. Thanks to Pedro Lauridsen Ribeiro, David Speyer, Kestutis Cesnavicius, dsilvestre, Michael Kinyon, Arbieto, Weiqi Gao, Mohammad, JC, BD, "nobody", and anonymous commenters for corrections.

3.2. Flows on Riemannian manifolds

In this section, we introduce *flows* $t \mapsto (M(t), g(t))$ on Riemannian manifolds (M, g), which are recipes for describing smooth deformations of such manifolds over time, and derive the basic *first variation formulae* for how various structures on such manifolds (e.g. curvature, length, volume) change[10] by such flows. We then specialise to the case of Ricci flow (together with some close relatives of this flow, such as renormalised Ricci flow, or Ricci flow composed with a diffeomorphism flow). We also discuss the "de Turck trick" that modifies the Ricci flow into a nonlinear parabolic equation, for the purposes of establishing local existence and uniqueness of that flow.

For the purposes of this chapter, we are not interested in just a single Riemannian manifold (M, g), but rather a one-parameter family of such manifolds $t \mapsto (M(t), g(t))$, parameterised by a "time" parameter t. The manifold $(M(t), g(t))$ at time t is going to determine the manifold $(M(t+dt), g(t+dt))$ at an infinitesimal time $t+dt$ into the future, according to some prescribed evolution equation (e.g. Ricci flow). In order to do this rigorously, we will need to "differentiate" a manifold flow $t \mapsto (M(t), g(t))$ with respect to time.

There are at least two ways to do this. The simplest is to restrict to the case in which the underlying manifold $M = M(t)$ is fixed (as a

[10] One can view these formulae as describing the relationship between two "infinitesimally close" Riemannian manifolds.

smooth manifold), so that only the metric $g = g(t)$ varies in time. As g takes values as sections in a vector bundle, there is then no difficulty in defining time derivatives $\dot{g}(t) = \frac{d}{dt}g(t)$ in the usual manner:

$$\frac{d}{dt}g(t) := \lim_{dt \to 0} \frac{g(t+dt) - g(t)}{dt}. \tag{3.35}$$

We can of course similarly define the time derivative of any other tensor field by the same formula.

The one drawback of the above simple approach is that it forces the topology of the underlying manifold M to stay constant. A more general approach is to view each d-dimensional manifold M(t) as a slice of a $d+1$-dimensional "spacetime" manifold $\mathbf{M}$ (possibly with boundary or singularities). This spacetime is (usually) equipped with a *time coordinate* $t : \mathbf{M} \to \mathbf{R}$, as well as a *time vector field* $\partial_t \in \Gamma(T\mathbf{M})$ which obeys the transversality condition $\partial_t t = 1$. The level sets of the time coordinate t then determine the sets M(t), which (assuming non-degeneracy of t) are smooth d-dimensional manifolds which collectively have a tangent bundle $\ker(dt) \subset T\mathbf{M}$ which is a d-dimensional subbundle of the $d+1$-dimensional tangent bundle $T\mathbf{M}$ of $\mathbf{M}$. The metrics g(t) can then be viewed collectively as a section $\mathbf{g}$ of $(\ker(dt)^*)^{\otimes 2}$. The analogue of the time derivative $\frac{d}{dt}g(t)$ is then the Lie derivative $\mathcal{L}_{\partial_t}\mathbf{g}$. One can then define other Riemannian structures (e.g. Levi-Civita connections, curvatures, etc.) and differentiate those in a similar manner.

The former approach is of course a special case of the latter, in which $\mathbf{M} = M \times I$ for some time interval $I \subset \mathbf{R}$ with the obvious time coordinate and time vector field. The advantage of the latter approach is that it can be extended (with some technicalities) into situations in which the topology changes (though this may cause the time coordinate to become degenerate at some point, thus forcing the time vector field to develop a singularity). This leads to concepts such as *generalised Ricci flow*, which we will not discuss here, though it is an important part of the definition of *Ricci flow with surgery* (see Chapters 3.8 and 14 of [**MoTi2007**]). Instead, we focus exclusively for now on the former viewpoint, in which $M = M(t)$ does not depend on time.

Suppose we have a smooth flow $(M, g(t))$ of metrics on a fixed background manifold M. The rate of change of the metric $g_{\alpha\beta}(t)$ is given by $\dot{g}_{\alpha\beta}(t)$. By the chain rule, this implies that any other expression that depends on this metric, such as the curvatures $\mathrm{Riem}^{\delta}_{\alpha\beta\gamma}(t)$, $\mathrm{Ric}_{\alpha\beta}(t)$, $R(t)$, should have a rate of change that depends linearly on $\dot{g}_{\alpha\beta}(t)$. We now compute exactly what these rates of change are. In principle, this can be done by writing everything explicitly using local coordinates and applying the chain rule, but we will try to keep things as coordinate-free as possible as it seems to cut down the computation slightly.

To abbreviate notation, we shall omit the explicit time dependence in what follows, e.g. abbreviating $g(t)$ to just g. We shall call a tensor field w *time-independent* or *static* if it does not depend on t, or equivalently that $\dot{w} = 0$.

From differentiating the identity

$$g^{\alpha\beta} g_{\beta\gamma} = \delta^{\alpha}_{\gamma} \tag{3.36}$$

we obtain the variation formula[11]

$$\frac{d}{dt} g^{\alpha\beta} = -g^{\alpha\gamma} g^{\beta\delta} \dot{g}_{\gamma\delta}. \tag{3.37}$$

Next, we compute how covariant differentiation deforms with respect to time. For a scalar function f, the derivative $\nabla_\alpha f \equiv df$ does not involve the metric, and so the rate of change formula is simple:

$$\frac{d}{dt} \nabla_\alpha f = \nabla_\alpha \dot{f}. \tag{3.38}$$

In particular, if f is static, then so is $\nabla_\alpha f$.

Now we take a static vector field X^β. From (3.38) and the product rule we see that the expression $\frac{d}{dt} \nabla_\alpha X^\beta$ is linear over $C^\infty(M)$ (interpreted as the space of static scalar fields). Thus we must have

$$\frac{d}{dt} \nabla_\alpha X^\beta = \dot{\Gamma}^{\beta}_{\alpha\gamma} X^\gamma \tag{3.39}$$

[11] Here is a place where the raising and lowering conventions can be confusing if applied blindly!

for some rank $(1,2)$ tensor $\dot{\Gamma}^{\beta}_{\alpha\gamma}$. From the Leibnitz rule and (3.38) we can obtain similar formulae for other tensors, e.g.

$$\frac{d}{dt}\nabla_\alpha \omega_\beta = -\dot{\Gamma}^{\gamma}_{\alpha\beta}\omega_\gamma \tag{3.40}$$

for any static one-form ω_β.

What is $\dot{\Gamma}^{\gamma}_{\alpha\beta}$? Well, we can work it out from the properties of the Levi-Civita connection. Differentiating the torsion-free identity

$$\nabla_\alpha \nabla_\beta f = \nabla_\beta \nabla_\alpha f \tag{3.41}$$

for static scalar fields f using (3.38), (3.40), we conclude the symmetry $\dot{\Gamma}^{\gamma}_{\alpha\beta} = \dot{\Gamma}^{\gamma}_{\beta\alpha}$. Similarly, differentiating the respect-of-metric identity $\nabla_\alpha g_{\beta\gamma} = 0$ we conclude that

$$-\dot{\Gamma}^{\delta}_{\alpha\beta} g_{\delta\gamma} - \dot{\Gamma}^{\delta}_{\alpha\gamma} g_{\beta\delta} + \nabla_\alpha \dot{g}_{\beta\gamma} = 0. \tag{3.42}$$

These two facts allow us to solve for $\dot{\Gamma}^{\gamma}_{\alpha\beta}$:

$$\dot{\Gamma}^{\gamma}_{\alpha\beta} = \frac{1}{2} g^{\gamma\delta}(\nabla_\alpha \dot{g}_{\beta\delta} + \nabla_\beta \dot{g}_{\alpha\delta} - \nabla_\delta \dot{g}_{\alpha\beta}) \tag{3.43}$$

(compare with the usual formula for the Christoffel symbols in local coordinates, see e.g. (3.287)).

Now we turn to curvature tensors. We have the identity

$$\nabla_\alpha \nabla_\beta X^\gamma - \nabla_\beta \nabla_\alpha X^\gamma = \mathrm{Riem}^{\gamma}_{\alpha\beta\delta} X^\delta \tag{3.44}$$

for any static vector field X. Taking the time derivative of this using (3.39), (3.40), etc. we obtain

$$\begin{aligned} &- \dot{\Gamma}^{\delta}_{\alpha\beta}\nabla_\delta X^\gamma + \dot{\Gamma}^{\gamma}_{\alpha\delta}\nabla_\beta X^\delta + \nabla_\alpha \dot{\Gamma}^{\gamma}_{\beta\delta} X^\delta \\ &- \dot{\Gamma}^{\delta}_{\beta\alpha}\nabla_\delta X^\gamma - \dot{\Gamma}^{\gamma}_{\beta\delta}\nabla_\alpha X^\delta + \nabla_\beta \dot{\Gamma}^{\gamma}_{\alpha\delta} X^\delta \\ &= \dot{\mathrm{Riem}}^{\gamma}_{\alpha\beta\delta} X^\delta \end{aligned} \tag{3.45}$$

which eventually simplifies to

$$\dot{\mathrm{Riem}}^{\gamma}_{\alpha\beta\delta} = \nabla_\alpha \dot{\Gamma}^{\gamma}_{\beta\delta} - \nabla_\beta \dot{\Gamma}^{\gamma}_{\alpha\delta}. \tag{3.46}$$

(one can view this as a linearisation of the usual formula for the Riemann curvature tensor in terms of Christoffel symbols). Combining

(3.43) and (3.46), and using the fact that the Levi-Civita connection respects the metric, we thus have

(3.47)
$$\dot{\mathrm{Riem}}^{\gamma}_{\alpha\beta\delta} = \frac{1}{2}g^{\gamma\sigma}(\nabla_\alpha\nabla_\delta\dot{g}_{\beta\sigma} - \nabla_\alpha\nabla_\sigma\dot{g}_{\delta\beta} - \nabla_\beta\nabla_\delta\dot{g}_{\alpha\sigma} + \nabla_\beta\nabla_\sigma\dot{g}_{\delta\alpha} - \mathrm{Riem}^{\mu}_{\alpha\beta\delta}\dot{g}_{\mu\sigma} - \mathrm{Riem}^{\mu}_{\alpha\beta\sigma}\dot{g}_{\delta\mu}).$$

Exercise 3.2.1. Show that (3.47) is consistent with the antisymmetry properties of the Riemann tensor, and with the Bianchi identities, as presented in Section 3.1.

Taking traces, we obtain a variation formula for the Ricci tensor,

(3.48)
$$\dot{\mathrm{Ric}}_{\alpha\beta} = -\frac{1}{2}\Delta_L\dot{g}_{\alpha\beta} - \frac{1}{2}\nabla_\alpha\nabla_\beta\mathrm{tr}(\dot{g}) - \frac{1}{2}\nabla_\alpha\nabla^\gamma\dot{g}_{\beta\gamma} - \frac{1}{2}\nabla_\beta\nabla^\gamma\dot{g}_{\alpha\gamma},$$

where $\mathrm{tr}(\pi) := g^{\alpha\beta}\pi_{\alpha\beta}$ is the trace, and the *Lichnerowicz Laplacian* (or *Hodge-de Rham Laplacian*) Δ_L on symmetric rank $(0,2)$ tensors $\pi_{\alpha\beta}$ is defined by the formula

$$\Delta_L\pi_{\alpha\beta} := \Delta\pi_{\alpha\beta} + 2\mathrm{Riem}^{\delta}_{\alpha\gamma\beta}\pi_{\gamma\delta} - \mathrm{Ric}^{\gamma}_{\alpha}\pi_{\gamma\beta} - \mathrm{Ric}^{\gamma}_{\beta}\pi_{\gamma\alpha} \tag{3.49}$$

and $\Delta\pi_{\alpha\beta} = \nabla_\gamma\nabla^\gamma\pi_{\alpha\beta}$ is the usual connection Laplacian. Taking traces once again, one obtains a variation formula for the scalar curvature:

$$\dot{R} = -\mathrm{Ric}^{\alpha\beta}\dot{g}_{\alpha\beta} - \Delta\mathrm{tr}(\dot{g}) + \nabla^\alpha\nabla^\beta\dot{g}_{\alpha\beta}. \tag{3.50}$$

Exercise 3.2.2. Verify the derivation[12] of (3.48) and (3.50).

We will also need to understand how deformation of the metric affects two other quantities, length and volume. The *length* $L(\gamma)$ of a curve $\gamma : [a,b] \to M$ in a Riemannian manifold (M,g) is given by the formula

$$L(\gamma) := \int_a^b g(\gamma'(u),\gamma'(u))^{1/2}\,du =: \int_\gamma ds. \tag{3.51}$$

where $ds = \gamma_* g(\gamma'(u),\gamma'(u))^{1/2}\,du$ is the measure on the curve γ induced by the metric.

[12] I wonder if there are more direct derivations of (3.48) and (3.50) that do not require one to go through so many computations. One can use (3.58) and (3.2.2) below as consistency checks for these formulae, but this does not quite seem sufficient.

Exercise 3.2.3. If γ varies smoothly in time (but with static endpoints $\gamma(a), \gamma(b)$, show that

$$\frac{d}{dt} L(\gamma) = \frac{1}{2} \int_\gamma \dot{g}(S,S)\ ds - \int_\gamma g(\nabla_S S, V)\ ds \tag{3.52}$$

where at every point $x = \gamma(u)$ of the curve, $S = \gamma'(u)/g(\gamma'(u),\gamma'(u))^{1/2}$ is the unit tangent, and $V = \dot{\gamma}(u)$ is the variation field[13].

The distance between two points x, y on a manifold is defined as $d(x,y) := \inf L(\gamma)$, where γ ranges over all curves from x to y. For smooth connected manifolds, it is not hard to show (e.g. by using a reduction to the unit speed case, followed by a minimising sequence argument and the Arzelá-Ascoli theorem, combined with some local theory of short geodesics to ensure C^1 regularity of the limiting curve) that this infimum is actually attained[14] for some minimising *geodesic* γ, which is then a critical point for $L(\gamma)$. From (3.52) we conclude that such geodesics must obey the equation $\nabla_S S = 0$ (thus the unit tangent vector parallel transports itself). We also conclude that

$$\frac{d}{dt} d(x,y) = \inf \frac{1}{2} \int_\gamma \dot{g}(S,S)\ ds \tag{3.53}$$

where the infimum is over all the minimising geodesics from x to y. Thus, a positive $\dot{g}$ (in the sense of quadratic forms) will increase distances between two marked points, while a negative $\dot{g}$ will decrease it.

Next, we look at the evolution of the volume measure $d\mu = d\mu(t)$. This measure is defined using any frame $(e_a)_{1\leq a\leq d}$ and dual frame $(e^a)_{1\leq a\leq d}$ as

$$d\mu := \sqrt{\det g}\, \left|e^1 \wedge \ldots \wedge e^d\right| \tag{3.54}$$

where $\det g$ is the determinant of the matrix with components $g_{ab} = g(e_a, e_b)$ (one can check that this measure is defined independently of the choice of frame). Intuitively, this measure is the unique measure such that an infinitesimal cube whose sides are orthogonal vectors of infinitesimal length r, will have volume $r^d + O(r^{d+1})$. It is not hard

[13] Strictly speaking, one needs to work on the pullback tensor bundles on $[a,b]$ rather than M in order to make the formulae in (3.52) well defined.

[14] However, this infimum need not be unique if x, y are far apart.

to show (using coordinates, and the variation formula $\frac{d}{dt}\det(A) = \mathrm{tr}(A^{-1}\dot{A})\det(A)$ for the determinant) that one has

$$\frac{d}{dt}d\mu = \frac{1}{2}\mathrm{tr}(\dot{g})\ d\mu. \tag{3.55}$$

Thus, a positive trace for $\dot{g}$ implies volume expansion, and a negative trace implies volume contraction. This is broadly consistent with how length is affected by metric distortion, as discussed previously.

3.2.1. Dilations. Now we specialise to some specific flows $(M, g(t))$ of a Riemannian metric on a fixed background manifold M. The simplest such flow (besides the trivial flow $g(t) = g(0)$, of course) is that of a dilation

$$g(t) := A(t)g(0) \tag{3.56}$$

where $A(t) > 0$ is a positive scalar with $A(0) = 1$. The flow here is given by

$$\dot{g}(t) = a(t)g(t) \tag{3.57}$$

where $a(t) := \frac{\dot{A}(t)}{A(t)} = \frac{d}{dt}\log A(t)$ is the *logarithmic derivative of* A (or equivalently, $A(t) = \exp(\int_0^t a(t')\ dt')$). In this case our variation formulas become very simple:

$$\begin{aligned}
\frac{d}{dt}g^{\alpha\beta} &= -ag^{\alpha\beta} \\
\dot{\Gamma}^{\gamma}_{\alpha\beta} &= 0 \\
\dot{\mathrm{Riem}}^{\delta}_{\alpha\beta\gamma} &= 0 \\
\dot{\mathrm{Ric}}_{\alpha\beta} &= 0 \\
\dot{R} &= -aR \\
\frac{d}{dt}d(x,y) &= \frac{1}{2}ad(x,y) \\
\frac{d}{dt}d\mu &= \frac{d}{2}a\ d\mu;
\end{aligned} \tag{3.58}$$

note that these formulae are consistent with (3.56) and the scaling heuristics at the end of the Section 3.1. In particular, a positive value of a means that length and volume are increasing, and a negative value means that length and volume are decreasing.

3.2.2. Diffeomorphisms. Another basic flow comes from smoothly varying one-parameter families of diffeomorphisms $\phi(t) : M \to M$ with $\phi(0)$ equal to the identity. This induces a flow

$$g(t) := \phi(t)^* g(0) \tag{3.59}$$

Infinitesimally, this flow is given by the *Lie derivative*

$$\dot{g}(t) = \mathcal{L}_{X(t)} g(t) \tag{3.60}$$

where $X(t) := \phi^*(t)\dot{\phi}(t)$ is the vector field representing the infinitesimal[15] diffeomorphism at time t. The quantity $\pi_{\alpha\beta} := \mathcal{L}_X g_{\alpha\beta}$ is known as the *deformation tensor*[16] of X, and it is a short exercise to verify the identity

$$\pi_{\alpha\beta} = \nabla_\alpha X_\beta + \nabla_\beta X_\alpha. \tag{3.61}$$

It is clear from diffeomorphism invariance that all tensors[17] deform via the Lie derivative:

$$\begin{aligned} \frac{d}{dt} g^{\alpha\beta} &= \mathcal{L}_X g^{\alpha\beta} \\ \dot{\mathrm{Riem}}^{\delta}_{\alpha\beta\gamma} &= \mathcal{L}_X \mathrm{Riem}^{\delta}_{\alpha\beta\gamma} \\ \dot{\mathrm{Ric}}_{\alpha\beta} &= \mathcal{L}_X \mathrm{Ric}_{\alpha\beta} \\ \dot{R} &= \mathcal{L}_X R. \end{aligned} \tag{3.62}$$

Exercise 3.2.4. Establish the *first variation formula* $\frac{d}{dt} d(x,y) = \inf g(X(y), S(y)) - g(X(x), S(x))$, where the infimum ranges over all minimal geodesics from x to y (which in particular determine the unit tangent vector S at x and at y).

Remark 3.2.1. As observed by Kazdan[**Ka1981**], one can compare the identities (3.2.2) with the variation formulae (3.46), (3.48), (3.50) to provide an alternate derivation of the Bianchi identities.

---

[15](One can use *Picard's existence theorem* to recover ϕ from X, though one has to solve an ODE for this and so the formula is not fully explicit.

[16]Informally, this tensor measures the obstruction to K being an infinitesimal symmetry, or *Killing vector field*.

[17]The formula for $\dot{\Gamma}^{\gamma}_{\alpha\beta}$ does not have such a nice representation, since $\Gamma^{\gamma}_{\alpha\beta}$ is not a tensor.

Applying (3.55), (3.61) we see that variation of the volume measure $d\mu$ is given by

$$\frac{d}{dt} d\mu = \mathrm{div}(X)\ d\mu \tag{3.63}$$

where $\mathrm{div}(X) := \nabla_\alpha X^\alpha$ is the divergence of X. On the other hand, for compact manifolds M at least, diffeomorphisms preserve the total volume $\mathrm{Vol}(M) := \int_M\ d\mu$. We thus conclude *Stokes' theorem*

$$\int_M \mathrm{div}(X)\ d\mu = 0 \tag{3.64}$$

on compact manifolds for arbitrary smooth vector fields X. It is not difficult to extend this to non-compact manifolds in the case when X is compactly supported. From (3.64) and the product rule we also obtain the *integration by parts formula*

$$\int_M f \nabla_\alpha X^\alpha\ d\mu = -\int_M (\nabla_\alpha f) X^\alpha\ d\mu. \tag{3.65}$$

As one particular special case of (3.65), we observe that the Laplacian on $C^\infty(M)$ is formally *self-adjoint*.

3.2.3. Ricci flow. Finally, we come to the main focus of this entire course, namely *Ricci flow*. A one-parameter family of metrics $g(t)$ on a smooth manifold M for all time t in an interval I is said to obey *Ricci flow* if we have

$$\frac{d}{dt} g(t) = -2\mathrm{Ric}(t). \tag{3.66}$$

Note that this equation makes tensorial sense since g and Ric are both symmetric rank 2 tensors. The factor of 2 here is just a notational convenience and is not terribly important, but the minus sign $-$ is crucial (at least, if one wants to solve Ricci flow *forwards*[18] in time).

In the preceding examples of dilation flow and diffeomorphism flow, it was easy to get from the infinitesimal evolution to the global evolution, either by using an integrating factor or by solving some ODEs. The situation for Ricci flow turns out to be significantly less trivial (and indeed, resolving the global existence problem properly

[18]Note that Ricci flow, like all other parabolic flows (of which the *heat equation* is the model example), is not time-reversible - solvability forwards in time does not imply solvability backwards in time!

is a large part of the proof of the Poincaré conjecture). Nevertheless, we do have the following relatively easy result:

Theorem 3.2.2 (Local existence)**.** *If M is compact and $g(0)$ is a smooth Riemannian metric on M, then there exists a time $T > 0$, and a unique Ricci flow $t \mapsto g(t)$ with initial metric $g(0)$ on the time interval $t \in [0, T)$.*

This theorem was first proven by Hamilton[**Ha1982**] using the *Nash-Moser iteration* method, and then a simplified proof given by de Turck[**DeT1983**]. We will not prove Theorem 3.2.2 here, but we will shortly indicate the main trick of de Turck used to reduce the problem to a standard local existence problem for nonlinear parabolic PDE.

Solutions have various names depending on their interval I of existence (or *lifespan*):

(1) A solution is *ancient* if I has $-\infty$ as a left endpoint.

(2) A solution is *immortal* if I has $+\infty$ as a right-endpoint.

(3) A solution is *global* if it is both ancient and immortal, thus $I = \mathbf{R}$.

The ancient solutions will play a particularly important role in our analysis later in this course, when we rescale (or *blow up*) the time variable (and the metric) as we approach a singularity of the Ricci flow, and then look at the asymptotic limiting profile of these rescaled solutions. It is a routine matter to compute the variations of various tensors under the Ricci flow:

$$\begin{aligned} \frac{d}{dt} g^{\alpha\beta} &= 2\mathrm{Ric}^{\alpha\beta} \\ \dot{R} &= \Delta R + 2|\mathrm{Ric}|^2 \\ \dot{\mathrm{Ric}}_{\alpha\beta} &= \Delta_L \mathrm{Ric}_{\alpha\beta} \\ &= \Delta \mathrm{Ric}_{\alpha\beta} + 2\mathrm{Ric}_\delta^\gamma \mathrm{Riem}^\delta_{\alpha\gamma\beta} - 2\mathrm{Ric}_{\alpha\gamma}\mathrm{Ric}^\gamma_\beta \\ \dot{\mathrm{Riem}} &= \Delta \mathrm{Riem} + \mathcal{O}(g^{-1}\mathrm{Riem}^2) \end{aligned} \tag{3.67}$$

where $\mathcal{O}(g^{-1}\mathrm{Riem}^2)$ is a moderately complicated combination of the tensors g^{-1}, Riem, and Riem that I will not write down explicitly here. In particular, we see that all of the curvature tensors obey some sort of tensor nonlinear heat equation. Parabolic theory then suggests

that these tensors will behave for short times much like solutions to the linear heat equation (for instance, they should become smoother over time, and they should obey various maximum principles). We will see various manifestations of this principle later in this course.

We also have variation formulae for length and volume:

$$\frac{d}{dt} d(x,y) = -\sup \int_\gamma \mathrm{Ric}(S,S)\ ds \tag{3.68}$$

$$\frac{d}{dt} d\mu = -R\ d\mu. \tag{3.69}$$

Thus Ricci flow tends to enlarge length and volume in regions of negative curvature, and reduce length and volume in regions of positive curvature.

3.2.4. Modifying Ricci flow. Ricci flow (3.66) combines well with the dilation flows (3.57) and diffeomorphism flows (3.2.2), thanks to the dilation symmetry and diffeomorphism invariance of Ricci flow[19]. For instance, if $g(t)$ solves Ricci flow and we set $\tilde{g}(s) := A(s) g(t(s))$ for some reparameterised time $s = s(t)$ and some scalar $A = A(s) > 0$, then the Ricci curvature here is $\tilde{\mathrm{Ric}}(s) = \mathrm{Ric}(t(s))$. We then see from the chain rule that $\tilde{g}$ obeys the equation

$$\frac{d}{ds}\tilde{g}(s) = -2A(s)\frac{dt}{ds}\tilde{\mathrm{Ric}}(s) + a(s)\tilde{g}(s) \tag{3.70}$$

where a is the logarithmic derivative of A. If we normalise the time reparameterisation by requiring $\frac{dt}{ds} = 1/A(s)$, we thus see that $\tilde{g}$ obeys *normalised Ricci flow*

$$\frac{d}{ds}\tilde{g} = -2\tilde{\mathrm{Ric}} + a\tilde{g}(s) \tag{3.71}$$

which can be viewed as a combination of (3.66) and (3.57). Conversely, it is not difficult to reverse these steps and transform a solution to (3.71) for some a into a solution of Ricci flow by reparameterising time and renormalising the metric by a scalar. Normalised Ricci flow is useful for studying singularities, as it can "blow up" the interesting portion of the dynamics to keep it at unit scale, instead of cascading to finer and finer scales as is usual when approaching a

[19] It can even be combined with these two flows simultaneously, although we will not need such a unified flow here.

singularity. The parameter a is at one's disposal to set; for instance, one could choose a to normalise the volume of M to be constant, or perhaps to normalise the maximum scalar curvature $\|R\|_\infty$ to be constant[20]. Setting $a = 0$, we observe in particular that the solution space to Ricci flow enjoys the *scaling symmetry*

$$g(t) \mapsto \lambda g(\frac{t}{\lambda}) \tag{3.72}$$

for any $\lambda > 0$. Thus, if we enlarge a manifold M by $\sqrt{\lambda}$ (or equivalently, if we fix M but make the metric g λ times as large), then Ricci flow will become slower by a factor of λ, and conversely if we shrink a manifold by $\sqrt{\lambda}$ then Ricci flow speeds up by λ. Thus, as a first approximation, big manifolds tend to evolve slowly under Ricci flow, and small ones tend to evolve quickly. Similarly, Ricci flow combines well with diffeomorphisms. If $g(t)$ solves Ricci flow and $\phi(t) : M \to M$ is a smoothly varying family of diffeomorphisms, then we can define a modified Ricci flow $\tilde{g}(t) := \phi(t)^* g(t)$ (cf. (3.59)). As Ricci curvature is intrinsic, this new metric has curvature $\tilde{\mathrm{Ric}}(t) = \phi(t)^* \mathrm{Ric}(t)$. It is then not hard to see that $\tilde{g}$ evolves by the flow

$$\frac{d}{dt}\tilde{g} = -2\tilde{\mathrm{Ric}} + \mathcal{L}_X \tilde{g} \tag{3.73}$$

where $X(t) := \phi^*(t)\dot{\phi}(t)$ are the vector fields that direct the flow ϕ as before. Note that (3.73) is a combination of (3.66) and (3.2.2). Conversely, given a solution to a modified Ricci flow (3.73) for some smoothly time-varying vector field X, one can convert it back to a Ricci flow by solving for the diffeomorphisms ϕ and then using them as a change of variables. The modified flows (3.73) (with various choices of vector field X) arise in a number of contexts. For instance, they are useful for studying *gradient Ricci solitons*, which will be an important special solution to Ricci flow that we will encounter later. Also, modified Ricci flow is an excellent tool for assisting the proof of local existence (Theorem 3.2.2), because it can be used (via the "de Turck trick") to "gauge away" some nasty non-parabolic components in Ricci flow, leaving behind a nicely parabolic non-linear PDE known as *Ricci-de Turck flow* which is straightforward to solve.

[20] Of course, only one quantity at a time can be normalised to be constant, since one only has one free parameter to set.

To explain this, let us first write the Ricci flow equation (3.66) "in coordinates" in order to attempt to solve it as a nonlinear PDE[21].

The traditional way to express Ricci flow in coordinates is, of course, to use local coordinate charts, but let us present a slightly different way to do this, relying on an arbitrarily chosen *background metric* $\overline{g}$ on M which does not depend on time[22]. This gives us a background connection $\overline{\nabla}$, background curvature tensors $\overline{\mathrm{Riem}}, \overline{\mathrm{Ric}}, \overline{R}$, and so forth. One can then express the evolving metric in terms of the background by a variety of formulae. For instance, the evolving connection ∇ can be expressed in terms of the background connection $\overline{\nabla}$ by the formula

$$\nabla_\alpha X^\beta = \overline{\nabla}_\alpha X^\beta + \Gamma^\beta_{\alpha\gamma} X^\gamma \tag{3.74}$$

where the Christoffel symbol $\Gamma^\beta_{\alpha\gamma}$ is given by

$$\Gamma^\beta_{\alpha\gamma} = \frac{1}{2} g^{\beta\delta}(\overline{\nabla}_\alpha g_{\gamma\delta} + \overline{\nabla}_\gamma g_{\alpha\delta} - \overline{\nabla}_\delta g_{\alpha\gamma}). \tag{3.75}$$

Exercise 3.2.5. Verify (3.74) and (3.75). Then use these formulae to give an alternate derivation of (3.39) and (3.43).

From (3.74) and the definition of Riemann curvature one concludes that

$$\begin{aligned} \mathrm{Riem}^\delta_{\alpha\beta\gamma} &= \overline{\mathrm{Riem}}^\delta_{\alpha\beta\gamma} + \overline{\nabla}_\alpha \Gamma^\delta_{\beta\gamma} - \overline{\nabla}_\beta \Gamma^\delta_{\alpha\gamma} \\ &+ \Gamma^\delta_{\alpha\mu}\Gamma^\mu_{\beta\gamma} - \Gamma^\delta_{\beta\mu}\Gamma^\mu_{\alpha\gamma}. \end{aligned} \tag{3.76}$$

Contracting this, we conclude

$$\mathrm{Ric}_{\alpha\beta} = \overline{\mathrm{Ric}}_{\alpha\beta} + \overline{\nabla}_\delta \Gamma^\delta_{\alpha\beta} - \overline{\nabla}_\alpha \Gamma^\delta_{\delta\beta} + \mathcal{O}(\Gamma^2). \tag{3.77}$$

Inserting (3.75) and only keeping careful track of the top order terms, we can eventually rewrite (3.77) as

$$\overline{\mathrm{Ric}}_{\alpha\beta} - \frac{1}{2} g^{\gamma\delta} \overline{\nabla}_\gamma \overline{\nabla}_\delta g_{\alpha\beta} + \frac{1}{2} \mathcal{L}_X g_{\alpha\beta} + \mathcal{O}(g^{-2} \overline{\nabla} g \overline{\nabla} g) \tag{3.78}$$

[21] The current state of the art of PDE existence theory does not cope all that well with the coordinate-independent frameworks which are embraced by differential geometers; in order to demonstrate existence of just about any equation, one usually has to break the covariance of the situation, and pick some coordinate system to work with. On the other hand, for particularly geometric equations, such as Ricci flow, there are often some special coordinate systems that one can pick that will simplify the PDE analysis enormously. See Section 1.10 for further discussion.

[22] For instance, one could pick $\overline{g} = g(0)$ to be the initial metric, although we do not need to do so.

where X is the vector field

$$X^\alpha := g^{\beta\gamma}\Gamma^\alpha_{\beta\gamma}. \tag{3.79}$$

Exercise 3.2.6. Show that the expression (3.78) for the Ricci curvature can be used to imply (3.48). Conversely, use (3.48) to recover (3.78) without performing an excessive amount of explicit computation. *Hint*: first show that the Ricci tensor can be crudely expressed as $\overline{\mathrm{Ric}} + \mathcal{O}(g^{-1}\overline{\nabla}^2 g) + \mathcal{O}(g^{-2}\overline{\nabla} g \overline{\nabla} g)$.

Thus, if we happen to have a solution g to modified Ricci flow (3.73) with the vector field X given by (3.79), then (3.73) simplifies to the *Ricci-de Turck flow*

$$\frac{d}{dt}g = g^{\gamma\delta}\overline{\nabla}_\gamma \overline{\nabla}_\delta g - 2\overline{\mathrm{Ric}} + \mathcal{O}(g^{-2}\overline{\nabla} g \overline{\nabla} g). \tag{3.80}$$

Conversely, it is not too difficult to reverse these steps and convert a solution to Ricci-de Turck flow to a solution to Ricci flow.

The equation (3.80) is a quasilinear parabolic evolution equation on g (which we think of now as evolving on a fixed background Riemannian manifold $(M, \overline{g})$, and one can establish local existence for (3.80) by a variety of methods. From this and the preceding remarks one can eventually establish Theorem 3.2.2, although we will not do so in detail here.

Remark 3.2.3. One particular way to establish existence for Ricci-de Turck flow (and probably not the most efficient) is sketched as follows. If one writes $g = \overline{g} + h$, then one can recast (3.80) as a heat equation against the fixed background metric that takes the form

$$\frac{d}{dt}h - \overline{\Delta} h = \mathcal{O}(h\overline{\nabla}^2 h) + F(h, \overline{\nabla} h) \tag{3.81}$$

for some smooth function F depending on the background (assuming that h is small in L^∞ norm so that one can compute the inverse $g^{-1} = (\overline{g}+h)^{-1}$ smoothly). The essentially semilinear equation (3.81) can be solved (for initial data small and smooth, and on small time intervals) on a compact manifold M by, say, the Picard iteration method, based on estimates such as the energy inequality

$$\|u\|_{L_t^\infty H_x^k(I\times M)} + \|u\|_{L_t^2 H_x^{k+1}(I\times M)} \ll_{I,k} \|u(0)\|_{H_x^k(M)} + \|F\|_{L_t^2 H_x^{k-1}(I\times M)} \tag{3.82}$$

for some suitably large integer k ($k > d/2 + 1$ will do), and with implied constants depending on the background metric, whenever u is a tensor that solves the heat equation $\frac{d}{dt}u + \Delta u = F$. This energy estimate can be easily established by integration by parts. To expand in a little more detail: the Picard iteration method proceeds by constructing iterative approximations $h^{(n)}$ to a solution h of (3.81) by solving a sequence of inhomogeneous heat equations

$$\frac{d}{dt}h^{(n)} - \overline{\Delta}h^{(n)} = \mathcal{O}(h^{(n-1)}\overline{\nabla}^2 h^{(n-1)}) + F(h^{(n-1)}, \overline{\nabla h^{(n-1)}}) \tag{3.83}$$

starting from $h^{(0)} = 0$ (say). The main task is to show that the sequence $h^{(n)} - h^{(n-1)}$ converges rapidly to zero in a suitable function space, such as $C_t^0 H_x^k \cap L_t^2 H_x^{k+1}$. This can be done by applying (3.82) with $u = h^{(n)} - h^{(n-1)}$ or $u = h^{(n)}$, and also using some product estimates in Sobolev spaces that are ultimately based on the Sobolev embedding theorem.

There is the still the issue of how to establish existence for the *linear* heat equation on tensors, but this can be done by functional calculus (once one establishes that Δ is a genuinely self-adjoint operator), or by making a reasonably accurate parametrix for the heat kernel. One (minor) advantage of this Picard iteration based approach is that it allows one to establish uniqueness and continuous dependence on initial data as well as just existence, and to show that the nonlinear solution obeys similar estimates (locally in time) to that of the linear heat equation. But uniqueness and continuity will not be necessary for the arguments in this course, and the estimates we need can always be established *a posteriori* by energy inequalities anyway.

Remark 3.2.4. The diffeomorphisms needed to convert solutions to Ricci-de Turck flow (3.80) back to solutions of Ricci flow (3.66) themselves obey a pleasant evolution equation; in fact, they evolve by harmonic map heat flow from the fixed domain $(M, \overline{g})$ to the target $(M, g(t))$. See [**ChKn2004**, Chapter 3.4] for further discussion. More generally, it seems that harmonic maps (and harmonic map heat flow, and harmonic coordinates) often provide natural coordinate systems that make various geometric PDE analytically tractable. On the other hand, for *geometric* arguments it seems better to work with the original Ricci flow; the de Turck diffeomorphisms seem to obscure

many of the delicate monotonicity properties that are essential to the deeper understanding of Ricci flow, and are also not completely covariant as they rely on an arbitrary choice of background metric $\overline{g}$.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/03/28`. Thanks to Dylan Thurston, Stefan, Dan, and Mohammad for corrections.

3.3. The Ricci flow approach to the Poincaré conjecture

In order to motivate the lengthy and detailed analysis of Ricci flow that will occupy the rest of this chapter, I will spend this section giving a high-level overview of Perelman's Ricci flow-based proof of the Poincaré conjecture, and in particular how that conjecture is reduced to verifying a number of (highly non-trivial) facts about Ricci flow.

At the risk of belaboring the obvious, here is the statement of that conjecture:

Theorem 3.3.1. *(Poincaré conjecture) Let M be a compact 3-manifold which is* simply connected *(i.e. it is connected, and every loop is contractible to a point). Then M is homeomorphic to a 3-sphere S^3.*

I will take it for granted that this result is of interest; see e.g. [**Mi2003**], [**Mo2007**], [**Mi2006**] for background and motivation for this conjecture. Perelman's methods also extend to establish further generalisations of the Poincaré conjecture, most notably *Thurston's geometrisation conjecture*, but I will focus this chapter just on the Poincaré conjecture. (On the other hand, the geometrisation conjecture will be rather visibly lurking beneath the surface in the discussion of this section.)

3.3.1. Examples of compact 3-manifolds. Before we get to the Ricci flow approach to the Poincaré conjecture, we will need to discuss some examples of compact 3-manifolds. Here (as in the statement of the Poincaré conjecture) we will work in the topological category, so our manifolds are *a priori* not endowed with a smooth structure or a Riemannian structure, and with two manifolds considered equivalent if they are homeomorphic. As mentioned in Section 3.1, in three

dimensions it is not difficult (once one has the triangulation theorem of Whitehead[**Wh1961**] and Munkres[**Mu1960**]) to move from the topological category back to the smooth or Riemannian category or vice versa, so one should not be too concerned about changes in category here.

The most basic example of a compact 3-manifold is the *sphere* S^3, which is easiest to define extrinsically as the unit sphere in $\mathbf{R}^4$, but can also be defined intrinsically as the one-point compactification of $\mathbf{R}^3$ (via the stereographic projection, for instance). Using the latter description, it is easy to see that the sphere is simply connected (note that in two and higher dimensions one can always perturb a loop to avoid a specific point, such as the point at infinity).

When we view S^3 as the unit sphere in $\mathbf{R}^4$, it acquires a transitive action of the special orthogonal group $SO(4)$, whose stabiliser is equivalent to $SO(3)$, thus we have a third important description of S^3, namely as the homogeneous space $SO(4)/SO(3)$. Now suppose one has a finite subgroup Γ of $SO(4)$ whose action on S^3 is free. Then one can quotient S^3 by Γ to create a new space

$$\Gamma\backslash S^3 \equiv \Gamma\backslash SO(4)/SO(3) \equiv \{\Gamma x : x \in S^3\}, \tag{3.84}$$

which remains a manifold since the action is free[23]. Such manifolds are known as *spherical 3-manifolds*. If the action of Γ is not completely trivial, then this new manifold $\Gamma\backslash S^3$ is topologically inequivalent to the original sphere S^3. The easiest way to see this is to observe that $\Gamma\backslash S^3$ is not simply connected. Indeed, as the action is not trivial, we can find $g \in \Gamma$ and $x \in S^3$ such that $gx \neq x$. Then a path from x to gx in S^3 descends to a closed loop on $\Gamma\backslash S^3$ which cannot be contracted to a point (basically because the orbit Γx of x in S^3 is discrete), and so $\Gamma\backslash S^3$ cannot be simply connected.

Remark 3.3.2. The above argument in fact shows that the fundamental group $\pi_1(\Gamma\backslash S^3)$ of $\Gamma\backslash S^3$ is just Γ; this is ultimately because S^3 is the *universal covering space* for $\Gamma\backslash S^3$. Conversely, Perelman's arguments can be used to show that spherical 3-manifolds are the

[23]If the action had some isolated fixed points, then the quotient space would merely be an *orbifold*.

only compact 3-manifolds with finite fundamental group (this conjecture, known as the *elliptisation conjecture*, is also a corollary of the geometrisation conjecture).

The most well-known example of a spherical 3-manifold (other than S^3 itself) is real projective space $\mathcal{RP}^3$, which is equivalent to the quotient $\Gamma\backslash S^3$ of S^3 by the two-element group $\{+1, -1\} \subset SO(4)$. Other examples of spherical space forms include *lens spaces* (in which Γ is a cyclic group), as well as a handful of other spaces in which Γ is essentially the symmetry group of a regular polytope (the four-dimensional analogue of the classical *Platonic solids*). An interesting example of the latter is the *Poincaré homology sphere*, which has the same homology groups as the sphere but is not homeomorphic to it.

The unit sphere S^3 (with the usual smooth structure) has a natural Riemannian metric g on it, which can be viewed either as the one induced from the Euclidean metric on the ambient space $\mathbf{R}^4$ (by restricting the tangent spaces of the latter to the former), or the one induced from the Lie group $SO(4)$, which is in turn induced by the *Killing form* on the Lie algebra $\mathfrak{so}(4)$. This metric has constant sectional curvature $+1$, which means that[24]

$$g(\mathrm{Riem}(u,v)u,v) = +1 \tag{3.85}$$

whenever $x \in S^3$ and $u, v \in T_x M$ are orthonormal vectors.

The metric g is invariant under the action of the rotation group $SO(4)$, and so it also descends to provide a Riemannian metric on every spherical 3-manifold of constant curvature $+1$ (and thus also constant positive Ricci and scalar curvature). Such Riemannian manifolds are known as *spherical space forms*. Conversely, it is not difficult to show that any compact connected 3-manifold M of constant curvature $+1$ arises in this manner; this is basically because (3.85) ensures that there is an infinitesimal action of the Lie algebra $\mathfrak{so}(4)$ on the orthonormal frame bundle of M, which then extends to an action of $SO(4)$ which is transitive (thanks to the connectedness of M) and has stabiliser equal to some finite extension of $SO(3)$

[24] To put it another way, Riem is $+1$ times the identity section of $\mathrm{Hom}(\bigwedge^2 TM, \bigwedge^2 TM)$.

(which is the structure group of the orthonormal frame bundle). Since $S^3 \equiv SO(4)/SO(3)$, the claim follows.

Remark 3.3.3. The spherical space forms are one of the eight *Thurston geometries* (or *model geometries*) that arise in the geometrisation conjecture, namely the *spherical* or *elliptic* geometries. (These eight geometries are also closely related, though not identical to, the clsasical classification of Bianchi of three-dimensional Lie algebras into nine families.)

Two more examples of 3-manifolds arise by considering S^2-bundles over S^1, or equivalently, $S^2 \times [0,1]$ with the two spheres $S^2 \times \{0\}$ and $S^2 \times \{1\}$ identified. Some basic degree theory (or even just winding number theory) shows that up to continuous deformation, there are only two such identifications; the orientation-preserving one (which is equivalent to the identity map, or a rotation map) and the orientation-reversing one (which is equivalent to a reflection map). The first such identification leads to the orientable S^2-bundle $S^2 \times S^1$, and the second identification leads to the non-orientable S^2-bundle (which is a 3-manifold analogue of the Klein bottle). Both of these manifolds can be viewed as quotients of the cylinder $S^2 \times \mathbf{R}$ by an action of $\mathbf{Z}$. More precisely, $S^2 \times \mathbf{R}$ has an obvious transitive action of $O(3) \times \mathbf{R}$ on it given by the formula $(U,s)(\omega,t) := (U\omega, s+t)$; the stabiliser subgroup is $O(2) \times \{0\}$, thus

$$S^2 \times \mathbf{R} \equiv (O(3) \times \mathbf{R})/(O(2) \times \{0\}) \tag{3.86}$$

Every group element $(U,s) \in O(3) \times \mathbf{R}$ with $s \neq 0$ generates a discrete subgroup $\Gamma = \{(U^n, ns) : n \in \mathbf{Z}\}$, and the quotient space

$$\Gamma\backslash S^2 \times \mathbf{R} \equiv \Gamma\backslash(O(3) \times \mathbf{R})/(O(2) \times \{0\}) \tag{3.87}$$

is homeomorphic to either the orientable or non-orientable S^2-bundle over S^1, depending whether $\det(U)$ is equal to $+1$ (i.e. $U \in SO(3)$ is rotation) or to -1 (i.e. $U \notin SO(3)$ is a reflection).

Let M be one of the above S^2-bundles over S^1. The projection map from M to S^1 induces a homomorphism from the fundamental group $\pi_1(M)$ to the fundamental group $\pi_1(S^1) \equiv \mathbf{Z}$.

Exercise 3.3.1. Show that this map is in fact bijective, thus $\pi_1(M) \equiv \mathbf{Z}$. In particular these manifolds are not homeomorphic to the spherical 3-manifolds.

There is an obvious Riemannian metric g to place on $S^2 \times \mathbf{R}$, being the direct sum of the standard metrics on S^2 and $\mathbf{R}$ respectively; it can also be defined in terms of the *Killing form* on the Lie algebra $\mathfrak{so}(3) \times \mathbf{R}$. As this metric is $O(3) \times \mathbf{R}$-invariant, it descends to both the oriented and non-oriented S^2-bundles over S^1. This metric is not of constant sectional curvature (the sectional curvature is positive on planes transverse to the axis of the cylinder, but vanish on planes parallel to that axis). But it has non-negative Riemann, sectional, Ricci curvature and positive scalar curvature.

Remark 3.3.4. The geometries coming from $S^2 \times \mathbf{R}$ are another of the eight Thurston geometries. These are the only two of the eight geometries that have some positive curvature in them; it is because of this that these two geometries can get extinguished in finite time by Ricci flow (remember, this flow compresses positively curved geometries and expands negatively curved ones). Very roughly speaking, it is these two geometries that show up in the finite time analysis of Ricci flow (in which the time variable t is bounded), whereas the other six geometries (being flat or negatively curved) only show up in the asymptotic analysis of Ricci flow (in the limit $t \to \infty$).

One can form further 3-manifolds out of the ones already discussed by the procedure of taking connected sums. Recall that the *connected sum* $M \# M'$ of two connected 3-manifolds M, M' is formed by choosing small 3-balls B, B' in M and M' respectively; if these balls are small enough, they are homeomorphic to the Euclidean ball B^3. Remove the interior of these two balls, leaving behind two boundaries ∂B and $\partial B'$ which are homeomorphic to S^2, and then identify these two boundaries together to create a new connected 3-manifold.

Remark 3.3.5. It is not hard to see that the location of the small balls B, B' is not relevant, since the connectedness of M and M' easily allows one to "slide" these balls around. There is however a subtle technicality regarding the identification map between ∂B and $\partial B'$. A bit of degree theory shows that up to homotopy, there are only two

possible identifications, one of which reverses the orientation of the other; for instance, any homeomorphism of the unit sphere S^2 to itself can be continuously deformed (while remaining a homeomorphism) to either a rotation or a reflection. If one of the manifolds M, M' is non-orientable then either choice of identification gives an equivalent connected sum up to homeomorphism, as one can slide one of the balls around the non-orientable manifold to return to the original location with the reversed orientation. Similarly, if both manifolds M, M' are orientable and one of them has an orientation-reversing homeomorphism, then again the two connected sums are homeomorphic. But there are some orientable 3-manifolds which lack orientation-reversing homeomorphisms (e.g. some lens spaces are of this type), and so there are cases in which the connected sum operation of two orientable manifolds M, M' is ambiguous. However, if one selects one of the two available orientations on M and M' (thus upgrading these orient*able* manifolds to orient*ed* manifolds), then one can define an *oriented connected sum* by asking that the oriented structures on M and M' are compatible upon gluing, thus yielding an oriented connected manifold $M\#M'$. This operation is then unambiguous up to homeomorphism. So, if we adopt the convention that all manifolds are equipped with an orientation if they are orientable, and are (of course) not equipped with an orientation if they are non-orientable, then we have a well-defined connected sum[25].

Once one addresses the technical issues raised in Remark 3.3.5, one can show that the connected sum operation is well-defined, commutative, and associative up to homeomorphism. There is also a strong relationship between the topology of a connected sum and that of its components:

Exercise 3.3.2. Let M, M' be connected manifolds of the same dimension.

(1) Show that $M\#M'$ is compact if and only if M and M' are both compact.

[25]Another way to avoid this issue is to lift non-oriented manifolds up to their oriented double cover whenever necessary.

(2) Show that $M\#M'$ is orientable if and only if M and M' are both orientable.

(3) Show that $M\#M'$ is simply connected if and only if M and M' are both simply connected.

The sphere also plays a special role, as the identity for the connected sum operation:

Exercise 3.3.3. Let M be a connected manifold, and let S be a sphere of the same dimension. Show that $M\#S$ (or $S\#M$) is homeomorphic to M.

This property uniquely defines the sphere topologically; if there was another connected manifold S' of the same dimension as a sphere S which was also an identity for the connected sum, then a consideration of $S\#S'$ shows that these two manifolds must be homeomorphic.

Exercise 3.3.4. Let M be a connected manifold. Suppose one connects M to itself by taking two small disjoint balls B and B' inside M, removing the interiors of these balls, and identifying the boundaries ∂B and $\partial B'$. Show that the resulting manifold is homeomorphic to the connected sum of M with an S^2 bundle over S^1.

Remark 3.3.6. For this remark, let us restrict attention to compact connected oriented 3-manifolds; modulo homeomorphisms, this is a commutative associative monoid with respect to connected sum, with identity S^3. A manifold is said to *non-trivial* if it is not the identity, and *prime* if it is non-trivial but not representable as the sum of two non-trivial manifolds. It then turns out that there is a *prime decomposition theorem* for such manifolds, analogous to the fundamental theorem of arithmetic: every such manifold is expressible as the connected sum of finitely many prime manifolds, and that this decomposition is unique up to rearrangement. However, useful as this decomposition is, it turns out that one does not actually need the prime decomposition theorem to prove the Poincaré conjecture[26].

[26]In fact, it is remarkable how little actual topology is needed to prove what is manifestly a topological conjecture; almost the entire proof of Perelman is conducted instead in the arena of differential geometry (and more specifically, Riemannian geometry) and partial differential equations.

Recall that of the spherical space forms and S^2-bundles over S^1 mentioned above, the sphere S^3 was the only one which was simply connected. From Exercises 3.3.2 and 3.3.3 we thus have

Corollary 3.3.7 (Poincaré conjecture for positively curved Thurston geometries)**.** *Let M be a simply connected 3-manifold which is the connected sum of finitely many spherical space forms and S^2-bundles over S^1. Then M is homeomorphic to the sphere S^3.*

Remark 3.3.8. One can establish similar results for combinations of any of the eight Thurston geometries (where we now allow "combination" to not just include the connected sum operation which glues along spheres S^2, but also more complicated joining operations which glue along tori T^2). Because of this, it is not difficult to show that the *geometrisation conjecture* implies the Poincaré conjecture. But the former conjecture is a significantly stronger and richer conjecture than the latter; it classifies *all* compact 3-manifolds, not just the simply connected ones.

3.3.2. Perelman's theorems and the Poincaré conjecture. We now have enough background to state the main results of Perelman; the rest of the chapter will be devoted to proving as much of these results as possible.

In [**Ha1982**], Hamilton realised that Ricci flow was an exceptionally promising tool for uniformising the geometry of a Riemmanian manifold, to the point where its topology became recognisable. The first evidence he established towards this phenomenon is the following *rounding theorem*: if a compact 3-manifold (M, g) has everywhere positive Ricci curvature, then the Ricci flow $(M(t), g(t))$ with this initial data develops a singularity in finite time t_*. Furthermore, as one approaches this singularity, the Ricci curvature becomes increasingly uniform, and more precisely that $\frac{3}{\overline{R}(t)}\mathrm{Ric}_{\alpha\beta}(t,x)$ converges uniformly to the metric $g_{\alpha\beta}(t,x)$ as $t \to t_*$, where $\overline{R}(t)$ is the average scalar curvature on $(M(t), g(t))$. (One can also show that $\overline{R}(t) \to +\infty$ as $t \to t_*$.) Once the Ricci curvature is sufficiently uniform, one can apply tools from Riemannian geometry such as the *sphere theorem* to deduce that the original manifold M was in fact homeomorphic to a

spherical space form (and in particular, if M is simply connected, is homeomorphic to a sphere S^3).

Unfortunately, for more general Riemannian 3-manifolds, Ricci flow was only able to partially uniformise[27] the geometry before the appearance of the first singularity. In order to address this issue, Hamilton[**Ha1997**] introduced (in the context of 4-manifolds) a notion of surgery on manifolds at each development of a singularity in order to continue the Ricci flow (but possibly with a topology change after each surgery). These and other results then led to Hamilton's program to develop a systematic theory of Ricci flow with surgery that would be able to uniformise and then recognise the topology of various Riemannian manifolds, particularly 3-manifolds. However, prior to Perelman's work, there was insufficient understanding of even the first singularity of Ricci flow to carry out this program, unless additional curvature assumptions were placed on the initial manifold.

The most important of Perelman's results in this direction is the following global existence result for a certain modification of Ricci flow.

Theorem 3.3.9 (Global existence of Ricci flow with surgery)**.** [**Pe2002**], [**Pe2003**] *Let (M, g) be a compact Riemannian 3-manifold, such that M does not contain any embedded copy of the real projective plane $\mathcal{RP}^2$ with trivial normal bundle. Then there exists a* Ricci flow with surgery *$t \mapsto (M(t), g(t))$ which assigns a compact Riemannian manifold $(M(t), g(t))$ to each time $t \in [0, +\infty)$, as well as a closed set $T \subset (0, +\infty)$ of* surgery times*, with the following properties:*

(1) *(Initial data) $M(0) = M$ and $g(0) = g$.*

(2) *(Ricci flow) If I is any connected component of $[0, +\infty) \backslash T$ (and is therefore an interval), and t_I is the left-endpoint of I, then $t \mapsto (M(t), g(t))$ is a Ricci flow on $\{t_I\} \cup I$, as defined in the Section 3.2 (in particular, $M(t)$ is constant on this interval).*

[27] However, work of Hamilton[**Ha1988**] and Chow[**Ch1991**] did show that Ricci flow does accomplish this task for 2-manifolds. These methods can give a new proof of the classical *uniformisation theorem*: see [**ChLuTi2006**].

(3) *(Topological compatibility) If $t \in T$, and $\varepsilon > 0$ is sufficiently small, then each connected component of $M_{t-\varepsilon}$ is homeomorphic to the connected sum of finitely many connected components of M_t, together with a finite number of spherical space-forms, $\mathcal{RP}^3 \# \mathcal{RP}^3$, and S^2 bundle over S^1. Furthermore, each connected component of M_t is used in the connected sum decomposition of exactly one component of $M_{t-\varepsilon}$.*

(4) *(Geometric compatibility) For each $t \in T$, the metric $g(t)$ on $M(t)$ is related to a certain limit of the metrics $g(t-\varepsilon)$ on $M(t-\varepsilon)$ as $\varepsilon \to 0$ by a certain* surgery procedure *which we will state precisely much later, see Section ???.*

The key here is item 3, which depends crucially on the structural analysis of Ricci flow as it approaches a singularity. The precise definition of surgery in item 4 is highly technical (and differs in a number of ways from Hamilton's version of the concept), but fortunately we do not need to know exactly what it is for topological applications such as the Poincaré conjecture, which only require item 3. (We do need to understand surgery in order to prove Theorems 3.3.12 and 3.3.13 below, though.)

The condition that M does not contain any embedded copy of $\mathcal{RP}^2$ with trivial normal bundle is a technical one, but is not a significant obstacle for proving the Poincaré conjecture, thanks to the following topological lemma:

Lemma 3.3.10. *Let M be a simply connected 3-manifold. Then M does not contain any embedded copy of $\mathcal{RP}^2$ with trivial normal bundle.*

Proof. Suppose for contradiction that M contained an embedded copy Σ of $\mathcal{RP}^2$ with trivial normal bundle. Then one can find a loop γ in Σ whose normal bundle in Σ is non-trivial[28]. Since Σ has trivial

[28]Indeed, one can find a loop whose neighbourhood is a Möbius strip; this is easiest seen by viewing $\mathcal{RP}^2$ topologically as the unit square with diametrically opposing points identified, and then taking γ to be a horizontal or vertical line through the centre of this square.

normal bundle in M, we conclude that γ has non-trivial normal bundle in M. But then γ cannot be contracted to a point, contradicting the hypothesis that M is simply connected. □

Remark 3.3.11. The above argument in fact shows that no orientable manifold can contain an embedded $\mathcal{RP}^2$ with trivial normal bundle; note that all simply connected manifolds are automatically orientable. The argument can also be modified to show that a simply connected manifold cannot contain any embedded copy of $\mathcal{RP}^2$ at all (regardless of whether the normal bundle is trivial).

To prove the Poincaré conjecture, we need to combine Theorem 3.3.9 with two other results about Ricci flow with surgery. The first is relatively easy:

Theorem 3.3.12 (Discrete surgery times)**.** *Let $t \to (M(t), g(t))$ be a Ricci flow with surgery with no embedded $\mathcal{RP}^2$ with trivial normal bundle. Then the set T of surgery times is discrete. In particular, any compact time interval only contains a finite number of surgeries.*

This theorem is basically proven by obtaining a lower bound on how much volume is removed by each surgery, combined with an upper bound on how much the volume can grow during the Ricci flow stage of the process. We have isolated Theorem 3.3.12 from Theorem 3.3.9 to highlight the importance of the former, but in practice the two results are proven simultaneously (since the geometric and topological compatibility in Theorem 3.3.9 would become more difficult to formulate properly if the surgery times were allowed to accumulate).

The second result is more non-trivial, though it is still significantly easier to prove than Theorem 3.3.9:

Theorem 3.3.13. *(Finite time extinction) Let (M, g) be a compact 3-manifold which is simply connected, and let $t \mapsto (M(t), g(t))$ be an associated Ricci flow with surgery. Then M(t) is empty for all sufficiently large times t.*

This result is analogous to finite time blowup results in nonlinear evolution equations. It is established by constructing a nonnegative quantity W(t) depending on the geometry $(M(t), g(t))$ at

time t, which decreases in such a manner that it must vanish in finite time, at which point one can show that the manifold becomes empty[29]. There are two known candidates for this quantity, one due to Perelman[**Pe2003b**] (based on a min-max functional over loops), and one due to Colding and Minicozzi[**CoMi2005**] (based on minimal spheres). Both quantities are known to be strong enough to establish Theorem 3.3.13.

Once one has Theorems 3.3.9, 3.3.12, and 3.3.13 in hand, the proof of the Poincaré conjecture is easy:

Proof of Theorem 3.3.1. Let M be a compact, simply connected 3-manifold. By an old result of Moise[**Mo1952**], every 3-manifold can be triangulated and so can easily be endowed with a smooth structure[30] Using a standard partition of unity argument, one can then create a smooth Riemannian metric g on M.

Theorem 3.3.9 (and Lemma 3.3.10) gives us a Ricci flow with surgery $t \mapsto (M(t), g(t))$ with surgery with initial data (M, g). By Theorem 3.3.13, there is some finite time t_* after which the manifolds $M(t)$ are empty. By Theorem 3.3.12, the number of surgeries up to that time are finite. By item 3. of Theorem 3.3.9 and working backwards from time t_* to time 0, we conclude that $M(0) = M$ is the connected sum of finitely many spherical space forms, copies of $\mathcal{RP}^3 \# \mathcal{RP}^3$, and S^2 bundles over S^1. Actually, since $\mathcal{RP}^3$ is already a spherical space form, we can absorb the second case into the first. The claim now follows from Corollary 3.3.7. □

Remark 3.3.14. Perelman's arguments in fact show a stronger version of Theorem 3.3.13: the finite time extinction occurs not only for simply connected manifolds, but more generally for any compact 3-manifold M whose fundamental group $\pi_1(M)$ is a free product of finite groups and infinite cyclic groups. The above argument then shows that any such manifold is diffeomorphic to the connected sum of finitely many space forms and S^2 bundles over S^1 (i.e. it is made up of the positively curved Thurston geometries). Conversely, it can

[29] From topological compatibility it is clear that if the manifold is empty at time t, it is empty for all subsequent times.

[30] In the converse direction, the results of Munkres[**Mu1960**] and Whitehead[**Wh1961**], show that this smooth structure is unique.

be shown that any such connected sum has a fundamental group of the above form. In particular this gives a topological necessary and sufficient condition for finite time extinction. One corollary of this is the *spherical space form conjecture* (or *elliptisation conjecture*): any compact 3-manifold with finite fundamental group is diffeomorphic to a spherical space form. See [**MoTi2007**] for details.

Remark 3.3.15. Theorems 3.3.9 and 3.3.12 also form the basis of the proof of the geometrisation conjecture. However, an additional ingredient is also needed, namely an analysis of the behaviour of the solutions to Ricci flow with surgery in the asymptotic limit $t \to \infty$. Also, in order to avoid dealing with all of the other Thurston geometries , a substantial amount of existing theory concerning geometrisation is first used to topologically simplify the manifold before applying Ricci flow (for instance, one works only with prime manifolds); see [**KlLo2006**], [**CaZh2006**], [**MoTi2008**] for details.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/04/01`. Thanks to Richard Kent, and anonymous commenters for corrections and references.

Richard Borcherds pointed out that in the smooth category, the connected sum (of two spheres, say) can have a rather large number of inequivalent realisations, due to the presence of exotic diffeomorphism classes on the sphere, although this issue only arises in dimensions larger than three.

3.4. The maximum principle, and the pinching phenomenon

We now begin the study of (smooth) solutions $t \mapsto (M(t), g(t))$ to the Ricci flow equation

$$\frac{d}{dt} g_{\alpha\beta} = -2\mathrm{Ric}_{\alpha\beta}, \tag{3.88}$$

particularly for compact manifolds in three dimensions. Our first basic tool will be the *maximum principle* for parabolic equations, which we will use to bound (sub-)solutions to nonlinear parabolic PDE by (super-)solutions, and vice versa. Because the various curvatures $\mathrm{Riem}^{\delta}_{\alpha\beta\gamma}$, $\mathrm{Ric}_{\alpha\beta}$, R of a manifold undergoing Ricci flow do

indeed obey nonlinear parabolic PDE (see (3.2.3)), we will be able to obtain some important lower bounds on curvature, and in particular establishes that the curvature is either bounded, or else that the positive components of the curvature dominate the negative components. This latter phenomenon, known as the *Hamilton-Ivey pinching phenomenon*, is particularly important when studying singularities of Ricci flow, as it means that the geometry of such singularities is almost completely dominated by regions of non-negative (and often quite high) curvature.

3.4.1. The maximum principle. In freshman calculus, one learns that if a smooth function $u : [a,b] \to \mathbf{R}$ has a local minimum at an interior point x_0, then the first derivative $u'(x_0)$ vanishes and the second derivative $u''(x_0)$ is non-negative. This implies a higher-dimensional version: if U is an open domain in $\mathbf{R}^d$ and $u : U \to \mathbf{R}$ has a local minimum at some $x_0 \in U$, then $\nabla u(x_0) = 0$ and $\Delta u(x_0) \geq 0$. Geometrically, the Laplacian $\Delta u(x_0)$ measures the extent to which u at x_0 dips below the average value of u near x_0, which explains why the Laplacian is non-negative at local minima.

The same phenomenon occurs on Riemannian manifolds:

Lemma 3.4.1. *Let (M,g) be a d-dimensional Riemannian manifold, and let $u : M \to \mathbf{R}$ be a C^2 function that has a local minimum at a point $x_0 \in M$. Then $\nabla_\alpha u(x_0) = 0$ and $\Delta u(x_0) \geq 0$.*

Proof. The vanishing $\nabla_\alpha u(x_0) = 0$ of the first derivative is clear, so we turn to the second derivative estimate. We let $e_1, \ldots, e_d$ be a (local) orthonormal frame of M. Then (by the Leibniz rule)

$$\Delta u = e_a^\alpha e_a^\beta \nabla_\alpha \nabla_\beta u = \nabla_{e_a} \nabla_{e_a} u - \nabla_{\nabla_{e_a} e_a} u. \tag{3.89}$$

Since u has vanishing first derivative at x_0, we conclude that

$$\Delta u(x_0) = \nabla_{e_a} \nabla_{e_a} u(x_0). \tag{3.90}$$

But as u has a local minimum at x_0, it also has a local minimum on the geodesic through x_0 with velocity e_a. From one-dimensional calculus we conclude that $\nabla_{e_a} \nabla_{e_a} u(x_0)$ is non-negative for each a, and the claim follows. □

For applications to nonlinear parabolic PDE, we need a time-dependent version of this fact, in which the function u and the metric g also vary with time. It is also convenient to consider work not with one function u, but with a pair u, v, and to consider relative local minima of u with respect to v (i.e. local minima of $u - v$).

Lemma 3.4.2 (Dichotomy)**.** *Let $t \mapsto (M, g(t))$ be a smooth flow of compact Riemannian manifolds on a time interval $[0, T]$. Let $u, v : [0, T] \times M \to \mathbf{R}$ be C^2 functions such that $u(0, x) \geq v(0, x)$ for all $x \in M$. Let $A \in \mathbf{R}$. Then exactly one of the following is true:*

(1) *$u(t, x) \geq v(t, x)$ for all $(t, x) \in [0, T] \times M$.*

(2) *There exists $(t, x) \in (0, T] \times M$ such that*

$$\begin{aligned} u(t,x) &< v(t,x) \\ \nabla_\alpha u(t,x) &= \nabla_\alpha v(t,x) \\ \Delta_{g(t)} u(t,x) &\geq \Delta_{g(t)} v(t,x) \\ \frac{d}{dt} u(t,x) &\leq \frac{d}{dt} v(t,x) - A(v(t,x) - u(t,x)) \end{aligned} \tag{3.91}$$

where $\Delta_{g(t)}$ is the Laplacian with respect to the metric $g(t)$.

Proof. By replacing u, v with $u - v, 0$ respectively we may assume that $v = 0$. If we then replace $u(t, x)$ by $e^{At} u(t, x)$ we may also assume that $A = 0$.

Clearly 1. and 2. cannot both hold. If 1. fails, then there exists $\varepsilon > 0$ such that $u(t, x) \leq -\varepsilon$ for some $(t, x) \in [0, T] \times M$. Let t be the first time for which this occurs, and let $x \in M$ be a point such that $u(t, x) = -\varepsilon$. Then $t > 0$. Also x is a local minimum of u(t) and thus $\nabla_\alpha u(t, x) = 0$ and $\Delta_{g(t)} u(t, x) \geq 0$ by Lemma 3.4.1. Also, since $u(t', x) > -\varepsilon$ for all $t' < t$ we have $\frac{d}{dt} u(t, x) \leq 0$. The claim follows. □

This gives us our first version of the parabolic maximum principle.

Corollary 3.4.3 (Supersolutions dominate subsolutions)**.** *Let the assumptions be as in Lemma 3.4.2. Suppose also that we have the supersolution property*

$$\frac{d}{dt} u(t,x) \geq \Delta_{g(t)} u(t,x) + \nabla_{X(t)} u(t,x) + F(t, u(t,x)) \tag{3.92}$$

and the subsolution property

$$\frac{d}{dt}v(t,x) \leq \Delta_{g(t)}v(t,x) + \nabla_{X(t)}v(t,x) + F(t,v(t,x)) \tag{3.93}$$

for all $(t,x) \in [0,T] \times M$ *where for each time* t, $X(t)$ *is a vector field, and* $F(t) : \mathbf{R} \to \mathbf{R}$ *is a Lipschitz function of constant less than* A. *Then* $u(t,x) \geq v(t,x)$ *for all* $0 \leq t \leq T$.

Proof. If we subtract (3.92) from (3.93) and use the Lipschitz nature of F we obtain

$$\frac{d}{dt}(u-v)(t,x) \geq \Delta_{g(t)}(u-v)(t,x) + \nabla_{X(t)}(u-v)(t,x) - A'|u(t,x)-v(t,x)| \tag{3.94}$$

for some $A' < A$. But this is inconsistent with the set of equations (3.91). The claim then follows immediately from Lemma 3.4.2. □

In our applications, the subsolution $v(t,x) = v(t)$ will in fact be independent of x, and so is really an ODE subsolution rather than a PDE subsolution:

$$\frac{d}{dt}v(t) \leq F(t,v(t)) \tag{3.95}$$

Thus the parabolic maximum principle allows us to lower bound PDE supersolutions by ODE subsolutions, as long as we have a bound at time zero.

The above maximum principle is already very useful for scalar solutions (or supersolutions) $u : [0,T] \times M \to \mathbf{R}$ to scalar nonlinear parabolic PDE, but we will in fact need a more general version of this principle for vector-valued solutions $u : [0,T] \mapsto \Gamma(V)$ to nonlinear parabolic PDE, where V is a vector bundle[31] over M, equipped with some connection ∇.

We will need some more notation. Let us say that a subset K of a tensor bundle V is *fibrewise convex* if the fiber $K_x := K \cap V_x$ over each point $x \in M$ is a convex subset of the vector space V_x. We say that a subset K of a vector bundle V is *parallel* to the connection ∇

[31] In practice, V will be derived from the tangent bundle, and ∇ will be derived from the Levi-Civita connection.

if for any vector field X on M, the induced vector field ∇_X preserves K (i.e. K is preserved by parallel transport).

We have a tensor variant of Lemma 3.4.1:

Lemma 3.4.4. *Let (M, g) be a d-dimensional Riemannian manifold, let V be a vector bundle over M with a connection ∇, and let K be a closed, fibrewise convex subset of V which is parallel with respect to the connection. Let $u \in \Gamma(V)$ be a section such that $u(x) \in \partial K_x$ at some point $x \in M$, and $u(y) \in K_y$ for all y in a neighbourhood of x (thus u in some sense "attains a local maximum" at x with respect to K). Then every directional derivative $\nabla_X u(x) \in T_x M$ of u at x is a tangent vector*[32] *to K_x at $u(x)$, and the Laplacian $\nabla^\alpha \nabla_\alpha u(x) \in T_x M$ is an inward or tangential pointing vector to K_x at $u(x)$ (i.e. it lives in the closed convex cone of $K_x - u(x)$). Here the space $T^*M \times V$ that ∇u is a section of is equipped with the direct sum of the Levi-Civita connection and the connection on V; by abuse of notation, we refer to all of these connections as ∇.*

Note that Lemma 3.4.1 corresponds to the special case when $V = M \times \mathbf{R}$ and $K = M \times [a, +\infty)$ for some a.

Proof. We begin with the claim concerning the first derivatives $\nabla_X u(x)$. One can restrict attention from M to (a local piece of) the one-dimensional geodesic through x with velocity $X(x)$, thus essentially reducing matters to the case $d = 1$. Any one-dimensional connection can be locally trivialised (this is essentially the *Picard existence* theorem for ODE) and so we may take M to be a small interval $(-\varepsilon, \varepsilon)$ (with x now being identified with 0), take V to be the trivial bundle $M \times V_0$, and take ∇ to be the trivial connection. The set K can then be identified with $M \times K_0$, and u can be viewed as a smooth function from $(-\varepsilon, \varepsilon)$ to K_0 that attains the boundary of K_0 at 0. It is then clear that the first derivative of u at 0 is tangent to K_0 at $u(0)$.

[32] A vector v is said to be an *inward pointing vector* at the boundary x of some convex set B if there is some conic neighbourhood of the ray of direction v emenating from x that is contained in B, and an *outward pointing vector* if there is a conic neighbourhood that lies outside of B. It is a *tangent vector* if it is neither inward pointing or outward pointing.

Now we turn to the second derivatives. As in the proof of Lemma 3.4.1, we introduce an orthonormal frame e_a and express the Laplacian in terms of this frame via the Leibniz rule as in (3.89). The first derivative terms are already tangential, so it suffices by convexity to show that $\nabla_{e_a}\nabla_{e_a}u(x)$ is tangential or inward pointing for each $a = 1, \ldots, d$ separately. But for fixed a, we can reduce to the one-dimensional setting considered previously by restricting to the geodesic through x with velocity $e_a(x)$ as before, so that once again u is now a smooth function from $(-\varepsilon, \varepsilon)$ to K_0 which attains a boundary value of K_0 at 0. In particular, if $\{v \in V_0 : \lambda(v) \leq c\}$ is a supporting halfspace for K_0 at $u(0)$ for some linear functional $\lambda : V_0 \to \mathbf{R}$, then the scalar function $x \mapsto \lambda(u(x)) \cdot w$ attains a maximum at 0 and thus has non-negative second derivative. The claim follows. □

As a consequence we can establish a rather general and powerful tensor maximum principle of Hamilton [**Ha1997**]:

Proposition 3.4.5 (Hamilton's maximum principle)**.** [**Ha1997**] *Let $t \mapsto (M, g(t))$ be a smooth flow of compact Riemannian manifolds on a time interval $[0, T]$. Let V be a vector bundle over M with connection ∇, and let $u : [0, T] \mapsto \Gamma(V)$ be a smoothly varying family of sections that obeys the nonlinear PDE*

$$\frac{d}{dt}u(t, x) = \nabla^\alpha \nabla_\alpha u + F(t, x, u) \tag{3.96}$$

where for each $(t, x) \in [0, T] \times M$, $F(t, x) : V_x \to V_x$ is a locally Lipschitz function (using the metric on V_x induced by g) which is continuous in t,x with uniformly bounded Lipschitz constant in the 1-neighbourhood (say) of K_x. For each time $t \in [0, T]$, let $K(t) \subset V$ a closed fibrewise convex parallel set varying continuously in t. We assume that K is preserved by F in the sense that for each $(t, x) \in [0, T) \times M$ and each boundary point $v \in \partial K_x(t) \subset V_x$, the spacetime vector $(1, F(t, x, v)) \in \mathbf{R} \times V_x$ is an inward or tangential vector to the spacetime body $K_x := \{(t', v') : t' \in [0, T], v' \in K_x(t')\}$ at the boundary point (t, v). Suppose also that $u(0, x) \in K_x(0)$ for all $x \in M$. Then $u(t, x) \in K_x(t)$ for all $(t, x) \in [0, T] \times M$.

Proof. By continuity in time, it suffices to prove the claim in $[0, T) \times M$ rather than $[0, T] \times M$.

Let us first give an "almost proof" of the claim, and then explain how to modify this to an actual proof. Suppose the claim failed; then $u(t,x)$ must exit $K_x(t)$ for some $(t,x) \in [0,T) \times M$. If we let t be the first time at which this occurs, then $t > 0$ and there exists $x \in M$ such that $u(t,x) \in \partial K_x(t)$, and $u(t,y) \in K_y(t)$ for all other $y \in M$. By Lemma 3.4.4, this implies that $\nabla^\alpha \nabla_\alpha u(t,x)$ is a tangential or inward pointing vector to $K_x(t)$ at $u(t,x)$. Also, since $u(t',x) \in K_x(t')$ for all $t' < t$, we see that $(1, \frac{d}{dt}u(t,x))$ is a tangential or outward pointing vector of K_x at $(t, u(t,x))$. From (3.96) we conclude that $(1, F(u,t,x))$ is also a tangential or outward pointing vector. This *almost* contradicts the hypothesis, except that it is still possible that $(1, F(u,t,x))$ is tangential.

To modify this, what we do is that we enlarge the set K slightly. Let A be a large number (essentially this is the bound on the local Lipschitz constant on F) $\varepsilon > 0$ be small. For each $(t,x) \in [0,T] \times M$, let $K_x^{(\varepsilon,A)}(t)$ be the εe^{At}-neighbourhood of $K_x(t)$ in V_x. If ε is small enough compared to A, this new set $K_x^{(\varepsilon,A)}(t)$ lives in the 1-neighbourhood of the old set $K_x(t)$. If A is sufficiently large compared to the local Lipschitz constant of F, then (by the growth of the exponential function e^{At}, and the hypotheses on F) the vector $(1, F(t,x,u))$ will now always be inward pointing, and not just tangential or inward pointing, to the spacetime body $K^{(\varepsilon,A)}$ whenever (t,x,u) is at a boundary point of this body. This allows us to use the previous arguments with $K_x(t)$ replaced by $K^{(\varepsilon,A)}$ throughout, to show that $u(t,x)$ cannot escape $K^{(\varepsilon,A)}$ if A is large enough. Sending $\varepsilon \to 0$ we obtain the claim. □

Remark 3.4.6. One can easily also add a drift term $\nabla_{X(t)} u(t,x)$ to (3.96), as in Corollary 3.4.3, though we will not need to do so here. With some more effort, one could start defining notions of "tensor supersolutions" and "tensor subsolutions", which take values as fibrewise convex sets rather than sections, to try to obtain a true tensor generalisation of Corollary 3.4.3, but this becomes very technical and we will not need to use such generalisations here.

Remark 3.4.7. The above maximum principles are known as *weak* maximum principles: starting from an assumption of non-negativity (or similar closed bounds) at time zero, they ensure non-negativity

(or closed bounds) at later times. Later on we shall also need *strong* maximum principles, in which one additionally assumes *positivity* at some initial point at time zero, and that the manifold is connected, and concludes positivity *everywhere* at later times[33]. Actually, it is the contrapositive of these strong maximum principles which will be of use to us, as they allow one to use vanishing of some key curvature at one point in spacetime to deduce vanishing of curvatures at many other points in spacetime also, which in particular will lead to some very important *splitting theorems* that will arise in the arguments later.

3.4.2. Applications of the maximum principle. We now apply the maximum principle (in both its scalar and tensor forms) to solutions of the Ricci flow (3.88) on some time interval $[0, T]$. The simplest application of these principles arises from exploiting the equation

$$\frac{d}{dt}R = \Delta R + 2|\mathrm{Ric}|^2 \tag{3.97}$$

for the scalar curvature (see (3.2.3)).

Remark 3.4.8. Intuitively, the two components on the RHS of (3.97) can be interpreted as follows. The dissipative term ΔR reflects the fact that a point in M with much higher (resp. lower) curvature than its neighbours (or more precisely, than the average curvature of its neighbours) will tend to revert to the mean, because the Ricci flow (3.88) will strongly contract the metric at regions of particularly high curvature (resp. strongly expand the metric at regions of particularly low curvature); one may visualise Ricci flow on a very pointed cigar, or a highly curved saddle, to try to see what is going on. The nonlinear term $2|\mathrm{Ric}|^2$ reflects the fact that if one is in a positive curvature region (e.g. a region behaving like a sphere), then the metric will contract under Ricci flow, thus increasing the curvature to be even more positive; conversely, if one is in a negative curvature region (such as a region behaving like a saddle), then the metric will expand, thus weakening the negativity of curvature. Note that in both cases

[33]This can be viewed as a substantial generalisation of the fact that the heat kernel on a connected manifold is everywhere strictly positive, or more informally that Brownian motion has a positive probability of hitting any given non-empty open region of the manifold.

the curvature is trending upwards, which is consistent with the non-negativity of $2|\mathrm{Ric}|^2$.

Remark 3.4.9. Another source of intuition can come from *Einstein metrics*, which are those metrics with the property that $\mathrm{Ric}_{\alpha\beta} = kg_{\alpha\beta}$ for some constant k; in particular we have constant scalar curvature $R = kd$, where d is the dimension. It is not hard to show (using (3.58)) that the Ricci flow for such metrics is given explicitly by the formulae

$$\begin{aligned} g_{\alpha\beta}(t) &= (1-2kt)g_{\alpha\beta}(0) \\ g^{\alpha\beta}(t) &= \frac{1}{1-2kt}g^{\alpha\beta}(0) \\ \mathrm{Ric}_{\alpha\beta}(t) &= \mathrm{Ric}_{\alpha\beta}(0) \\ R(t) &= \frac{1}{1-2kt}R(0) = \frac{1}{1-2kt}kd. \end{aligned} \tag{3.98}$$

Of course, this is completely consistent with (3.97). Note that if k is positive (which occurs for instance in manifolds of constant positive sectional curvature, such as the sphere and its quotients) then a singularity develops at time $1/2k$, in which the diameter of the manifold has shrunk to zero and the curvature has become infinitely positive. In contrast, if k is negative (which occurs for manifolds of constant negative sectional curvature, such as hyperbolic space) the metric expands, becomes increasingly flat over time and does not develop singularities.

Since R is the trace of the self-adjoint tensor $\mathrm{Ric}_{\alpha\beta}$, one has the decomposition

$$|\mathrm{Ric}|^2 = \frac{1}{d}R^2 + |\mathrm{Ric}^0|^2, \tag{3.99}$$

where $\mathrm{Ric}^0_{\alpha\beta} := \mathrm{Ric}_{\alpha\beta} - \frac{1}{d}Rg_{\alpha\beta}$ is the traceless component of the Ricci tensor. We conclude that R is a supersolution to a nonlinear parabolic PDE:

$$\frac{d}{dt}R \geq \Delta R + \frac{2}{d}R^2. \tag{3.100}$$

For each time t, let $R_{\min}(t)$ denote the minimum value of the scalar curvature. We thus conclude

Proposition 3.4.10. *(Lower bounds on scalar curvature). Let $(M, g(t))$ be a Ricci flow on a compact d-dimensional manifold on some time interval $[0, T]$. Then for every $t \in [0, T]$, we have*

$$R_{\min}(t) \geq \frac{R_{\min}(0)}{1 - \frac{2t}{d} R_{\min}(0)}. \tag{3.101}$$

In particular, if $R \geq c$ at time zero for some $c \in \mathbf{R}$, then $R \geq c$ for all subsequent times for which the flow exists; and if furthermore c is positive, then the flow cannot be extended beyond time $\frac{d}{2c}$.

From Remark 3.4.1 we see that for Einstein metrics, (3.101) is obeyed with equality, so that (3.101) can be quite sharp.

Exercise 3.4.1. Use Corollary 3.4.3 to deduce Proposition 3.4.10.

Proposition 3.4.10 asserts that while the scalar curvature can become extremely large and positive as time increases, it cannot become extremely large and negative. One quick corollary of this is

Corollary 3.4.11 (Upper bound on volume growth)**.** *Let $(M, g(t))$ be a Ricci flow on a compact d-dimensional manifold on some time interval $[0, T]$, such that we have the pointwise lower bound $R \geq c$ at time zero. Then we have*

$$\mathrm{Vol}(M, g(t)) \leq e^{-2ct} \mathrm{Vol}(M, g(0)) \tag{3.102}$$

for all $0 \leq t \leq T$.

Proof. From the variation formula (3.69) for the volume measure $d\mu$ we have

$$\frac{d}{dt} \mathrm{Vol}(M, g(t)) = -\int_M 2R(t, x)\, d\mu_{g(t)}(x). \tag{3.103}$$

By Proposition 3.4.10, R is bounded from below by c, leading to the inequality $\frac{d}{dt}\mathrm{Vol}(M, g(t)) \leq -2c\mathrm{Vol}(M, g(t))$. The claim now follows from *Gronwall's inequality.* □

Exercise 3.4.2. Strengthen the bound (3.4.11) to

$$\mathrm{Vol}(M, g(t)) \leq (1 - \frac{2ct}{d})^d \mathrm{Vol}(M, g(0)) \tag{3.104}$$

and show that this inequality is sharp for Einstein metrics. Note that this improved bound demonstrates rather visibly that when $c > 0$, some singularity must develop at or before time $d/2c$.

We now turn to applications of the tensor maximum principle. It is natural to apply this principle to the equation for the Riemann tensor,

$$\frac{d}{dt}\mathrm{Riem}_{\alpha\beta} = \Delta \mathrm{Riem}_{\alpha\beta} + \mathcal{O}(\mathrm{Riem}^2) \tag{3.105}$$

(see (3.2.3)). In principle, this expression is of the required form (3.96), but the nonlinearity $\mathcal{O}(\mathrm{Riem}^2)$, while explicit, is rather messy to work with. It is convenient to simplify (3.105) further by viewing things in a certain evolving orthonormal frame. For ease of notation, let us assume that the compact manifold $M = M(0)$ is *parallelisable*[34], so that it enjoys a global orthonormal frame $e_1(0), \ldots, e_d(0) \in \Gamma(TM(0))$ for the metric $g(0)$. This orthonormal frame induces a linear identification between the tangent bundle $TM(0)$ and the trivial bundle $M \times \mathbf{R}^d$, with $e_1(0), \ldots, e_d(0)$ being identified with the standard basis sections of the trivial bundle. The metric $g_{\alpha\beta}(0)$ is then identified with the Euclidean section $\eta_{\alpha\beta} := e^a_\alpha e^a_\beta \in \mathrm{Sym}^2(M \times \mathbf{R}^d)$ (which is giving the fibres of $M \times \mathbf{R}^d$ a Euclidean structure). Note that this is NOT directly a metric on M, since $M \times \mathbf{R}^d$ is distinct from the tangent bundle TM, but the orthonormal frame provides an identification between the section η and the metric $g(0)$. Now we start the Ricci flow, creating a family of new metrics $g(t)$ for $t \in [0, T]$. There is no reason why the frame $e_1(0), \ldots, e_d(0)$ should remain orthonormal in these new metrics. However, if we evolve the frame by the equation

$$\frac{d}{dt} e^\alpha_a := \mathrm{Ric}^{\alpha\beta}(e_a)_\beta \tag{3.106}$$

(which, by Picard's existence theorem for ODE, exists for all $t \in [0, T]$) then an easy computation using (3.88) (and Gronwall's inequality) reveals that $e_1(t), \ldots, e_d(t)$ remain orthonormal with respect to $g(t)$.

Exercise 3.4.3. Prove this. *Hint*: differentiate $g_{\alpha\beta} e^\alpha_a e^\alpha_b$ in time and use (3.88), (3.4.2).

[34] To handle the general case, one could work locally, or pass to a covering space, and/or replace the trivial bundle $M \times \mathbf{R}^d$ appearing below by a non-trivial bundle and eliminate explicit mention of the orthonormal frame altogether; we leave the details to the interested reader. In three dimensions, every orientable manifold is parallelisable, so it is even easier to reduce to the parallelisable case in that setting.

The frame $e_1(t), \ldots, e_d(t)$ can be used to identify the tangent manifold $TM(t)$ at time t with the trivial bundle $M \times \mathbf{R}^d$, which identifies $g(t)$ with η. In particular, the Levi-Civita connection $\nabla_{g(t)}$ can be identified with a connection $\nabla(t)$ on $M \times \mathbf{R}^d$ to which η is parallel (thus parallel transport by $\nabla(t)$ proceeds by rotations). Similarly, we can identify the Riemann tensor $\mathrm{Riem}(t) \in \mathrm{Hom}(\bigwedge^2 T^*M, \bigwedge^2 T^*M)$ at that time with a tensor $\mathcal{T}(t) \in M \times \mathrm{Hom}(\bigwedge^2 \mathbf{R}^d, \bigwedge^2 \mathbf{R}^d)$. Using the natural identification between $\bigwedge^2 \mathbf{R}^d$ and the Lie algebra $\mathfrak{so}(d)$, one can thus view $\mathcal{T}(t)$ as a section of $M \times \mathrm{Hom}(\mathfrak{so}(d), \mathfrak{so}(d))$. Actually, since the Riemann tensor is self-adjoint, $\mathcal{T}(t,x) : \mathfrak{so}(d) \to \mathfrak{so}(d)$ is self-adjoint also (using the *Killing form* on $\mathfrak{so}(d)$).

After some significant algebraic computation, the equation (3.105) can be revealed to take the form

$$\frac{d}{dt}\mathcal{T} = \nabla^\gamma \nabla_\gamma \mathcal{T} + \mathcal{T}^2 + \mathcal{T}^\# \tag{3.107}$$

where the connection $\nabla = \nabla(t)$ has been extended from $M \times \mathbf{R}^d$ to $M \times \mathrm{Hom}(\mathfrak{so}(d), \mathfrak{so}(d))$ in the usual manner, $\mathcal{T}^2$ is the usual square of $\mathcal{T}$ (viewed as a linear operator from $\mathfrak{so}(d)$ to itself), and $\mathcal{T}^\#$ is the *Lie algebra square* of $\mathcal{T}$, defined by the formula

$$\langle \mathcal{T}^\# X, Y \rangle := \mathrm{tr}(\mathcal{T}(\mathrm{ad}X)\mathcal{T}(\mathrm{ad}Y)) \tag{3.108}$$

for all $X, Y \in \mathfrak{so}(d)$ where $\mathrm{ad}X : Y \to [X,Y]$ is the usual adjoint operator and $\langle X, Y \rangle = \mathrm{tr}(\mathrm{ad}X \mathrm{ad}Y)$ is the *Killing form*. One easily verifies that if $\mathcal{T}$ is self-adjoint, then so[35] are $\mathcal{T}^2$ and $\mathcal{T}^\#$.

Exercise 3.4.4. Show that (3.108) implies (3.97).

If $\mathcal{T}$ is positive semi-definite (which is equivalent to the Riemann tensor being non-negative), then it is easy to see that $\mathcal{T}^2 + \mathcal{T}^\#$ are also. Since the space $\mathcal{P}$ of positive semi-definite self-adjoint elements of $\mathrm{Hom}(\mathfrak{so}(d), \mathfrak{so}(d))$ forms a closed convex cone which is invariant under the action of $SO(d)$ (and in particular, $M \times \mathcal{P}$ is parallel with respect to the connections $\nabla_{g(t)}$), one can then apply the tensor maximum principle to conclude

[35] Curiously, in four and higher dimensions the Bianchi identity that $\mathcal{T}$ will satisfy if it comes from the Riemann tensor is not preserved by either $\mathcal{T}^2$ or $\mathcal{T}^\#$, but it is preserved by their sum $\mathcal{T}^2 + \mathcal{T}^\#$.

Proposition 3.4.12 (Non-negative Riemann curvature is preserved)**.** *Let $(M, g(t))$ be a Ricci flow on a compact d-dimensional manifold on some time interval $[0,T]$. Suppose that the Riemann curvature is everywhere non-negative at time zero. Then the Riemann curvature is everywhere non-negative for all times $t \in [0,T]$.*

Remark 3.4.13. Strictly speaking, there is an issue because the nonlinearity $\mathcal{T} \mapsto \mathcal{T}^2 + \mathcal{T}^\#$ is only locally Lipschitz rather than globally Lipschitz. But as we are assuming that the manifold is compact and the metrics vary smoothly, $\mathcal{T}$ is already bounded, and so one can truncate the nonlinearity by brute force outside of these bounds to ensure global Lipschitz bounds. We shall take advantage of this trick again below without further comment.

Now we specialise to three dimensions, in which the situation simplifies substantially, because $\mathfrak{so}(3) \equiv \bigwedge^2 \mathbf{R}^3$ can be identified with $\mathbf{R}^3$ by Hodge duality. If the self-adjoint map $\mathcal{T} : \mathbf{R}^3 \to \mathbf{R}^3$ is diagonalised as $\mathrm{diag}(\lambda, \mu, \nu)$ in some orthonormal frame, then we have $\mathcal{T}^2 = \mathrm{diag}(\lambda^2, \mu^2, \nu^2)$ and $\mathcal{T}^\# = \mathrm{diag}(\mu\nu, \lambda\nu, \lambda\mu)$. Also, if $\mathcal{T}$ was representing the Riemann tensor, then the Ricci curvature in the same frame can be computed to be $\mathrm{diag}(\mu+\nu, \lambda+\nu, \lambda+\mu)$, and so the scalar curvature is $2(\lambda+\mu+\nu)$.

Heuristically, the tensor maximum principle predicts that the evolution of the equation (3.107) should be somehow "controlled" by the evolution of the ODE

$$\frac{d}{dt}(\lambda, \mu, \nu) = F(\lambda, \mu, \nu) \tag{3.109}$$

where $F(\lambda, \mu, \nu) := (\lambda^2 + \mu\nu, \mu^2 + \lambda\nu, \nu^2 + \lambda\mu)$. It seems difficult to formulate this heuristic rigorously in complete generality (the main problem being that the convexity requirements of the maximum principle ultimately translate to rather significant constraints on what types of properties of the eigenvalues λ, μ, ν one can study with this principle). However, we can do so in two important special cases. We begin with the simpler one.

Proposition 3.4.14 (Non-negative Ricci curvature is preserved in three dimensions)**.** *Let $(M, g(t))$ be a Ricci flow on a compact 3-dimensional manifold on some time interval $[0,T]$. Suppose that the*

Ricci curvature is everywhere non-negative at time zero. Then the Ricci curvature is everywhere non-negative for all times $t \in [0,T]$.

Proof. By the previous discussion, having non-negative Ricci curvature is equivalent to having all sums of pairs $\lambda+\mu, \mu+\nu, \nu+\lambda$ of $\mathcal{T}$ non-negative. Equivalently, this is asserting that the partial traces $\mathrm{tr}(\mathcal{T}|_V)$ of $\mathcal{T}$ on any two-dimensional subspace of V is non-negative. If we let $K = K(t) \subset M \times \mathrm{Hom}(\mathbf{R}^3, \mathbf{R}^3)$ denote all the pairs $(x, \mathcal{T})$ for which this is true, we see that K is closed, convex, and parallel with respect to the connections $\nabla(t)$, since parallel transport by these connections acts on $\mathrm{Hom}(\mathbf{R}^3, \mathbf{R}^3)$ by orthogonal conjugation. Elementary algebraic computation also reveals that if the triplet (λ, μ, ν) has the property that the sum of any two elements is non-negative, then the same is true of $F(\lambda, \mu, \nu)$. From this we see that the hypotheses of Proposition 3.4.5 are satisfied, and the claim follows. □

Remark 3.4.15. This claim is special to three (and lower) dimensions; it fails for four and higher dimensions. Similarly, in three dimensions, since non-negative Riemann curvature is equivalent to non-negative sectional curvature, we see from Proposition 3.4.12 that the latter is also preserved by three-dimensional Ricci flow. However, this claim also fails in four and higher dimensions.

Results such as Proposition 3.4.12 and Proposition 3.4.14 are of course useful if one has an initial assumption of non-negative curvature. But for our applications, we need to understand what is going on for manifolds which may have combinations of both positive and negative curvature at various points and in various directions. The bound on scalar curvature given by Proposition 3.4.10 is helpful in this regard, but it only partially controls the situation (in terms of the eigenvalues λ, μ, ν, it offers a lower bound on $\lambda+\mu+\nu$, but not on λ, μ, ν individually). It turns out that one cannot completely establish a unilateral lower bound on the individual curvatures λ, μ, ν, but one can at least show that if one of these curvatures is large and negative, then one of the others must be extremely large and positive, and so in regions of high curvature, the positive curvature components dominate. This important phenomenon for Ricci flow is known as *Hamilton-Ivey pinching*, and is formalised as follows:

Theorem 3.4.16 (Hamilton-Ivey pinching phenomenon)**.** *Let $(M, g(t))$ be a Ricci flow on a compact 3-dimensional manifold on some time interval $[0,T]$. Suppose that the least eigenvalue $\nu(t,x)$ of the Riemann curvature tensor is bounded below by -1 at times $t = 0$ and all $x \in M$. Then, at all spacetime points $(t,x) \in [0,T] \times M$, we have the scalar curvature bound*

$$R \geq \frac{-6}{4t+1} \tag{3.110}$$

and furthermore whenever one has negative curvature in the sense that $\nu(t,x) < 0$, then one also has the pinching bound

$$R \geq 2|\nu|(\log|\nu| + \log(1+t) - 3). \tag{3.111}$$

Exercise 3.4.5. With the assumptions of Theorem 3.4.16, use (3.110) and (3.111) to establish the lower bound

$$(1+t)\nu \geq -C\frac{100 + (1+t)R}{\log(100 + (1+t)R)} \tag{3.112}$$

for all $(t,x) \in [0,T] \times M$ and some absolute constant C (note that $100+(1+t)R > 1$, thanks to (3.110). Conclude in particular that the scalar curvature controls the Riemann and Ricci tensors in the sense that we have the pointwise bounds

$$|\mathrm{Ric}|_g, |\mathrm{Riem}|_g \leq C(100 + (1+t)R) \tag{3.113}$$

for another absolute constant C.

Proof of Theorem 3.4.16. Since $R = 2(\lambda + \mu + \nu)$ and the least eigenvalue ν is at least -1 at time zero, we have $R \geq -6$ at time zero. The claim (3.110) then follows immediately from Proposition 3.4.10.

The proof of (3.111) requires more work. Starting with the tensor $\mathcal{T}$ and its eigenvalues $\lambda \geq \mu \geq \nu$, we define the trace $S := \lambda+\mu+\nu = \frac{1}{2}R$ and the quantity $X := \max(-\nu, 0)$. We write $f_t(x) := x(\log x + \log(1+t)-3)$ and let Ω_t be the set of all pairs (x,s) such that $s \geq \frac{-3}{1+t}$ and such that $s \geq f_t(x)$ if $x > \frac{1}{1+t}$. (For $x < \frac{1}{1+t}$, the only constraint we place on s is that $s \geq \frac{-3}{1+t}$. Elementary calculus shows that Ω_t is a convex set, and furthermore is left-monotone in the sense that if $(x,s) \in \Omega_t$ and $x' < x$, then $(x',s) \in \Omega_t$. Because trace is a linear functional and the least eigenvalue ν is a convex functional, it is not hard to then see that the set $K(t) := \{(x, \mathcal{T}) : (X,S) \in \Omega_t\} \subset$

$M \times \mathrm{Hom}(\mathbf{R}^3, \mathbf{R}^3)$ is closed and fibrewise convex. Also, since parallel transport on the connections $\nabla(t)$ acts by orthogonal conjugation, $K(t)$ is also parallel.

The initial conditions easily ensure that $\mathcal{T}$ lies in $K(0)$ at time zero (since $X \leq 1$ and $S \geq -3$ in this case). Similarly, the conclusion (3.111) follows easily from the claim that $\mathcal{T}$ lies in $K(t)$ at all later times t (note that in the case $X \leq \frac{1}{1+t}$, one can use the trivial bound $S \geq -3X$ to establish the claim, rather than by exploiting the inclusion $\mathcal{T} \in K(t)$). So to finish the proof, it suffices by Proposition 3.4.5 to show that K is preserved by the ODE (3.109). This can be accomplished by a (rather tedious) elementary calculation, the key point being that if (λ, μ, ν) solve (3.109) with $\lambda \geq \mu \geq \nu$ and X, S are defined as before, then one has the inequality

$$\frac{d}{dt}\left(\frac{S}{X} - \log X\right) \geq X \tag{3.114}$$

whenever $X > 0$.

The set $K(t)$ can be viewed as the region in which either $X \leq \frac{1}{1+t}$, or $X > \frac{1}{1+t}$ and $\frac{S}{X} - \log X \geq \log(1+t) - 3$, and then (3.114) implies that this region is preserved by the ODE. □

Remark 3.4.17. One can informally see how (3.109) is forcing some sort of pinching towards positive curvature as follows. In order for pinching not to occur, one needs ν to be large and negative, and λ to be of order $O(|\nu|)$ in magnitude. Given the lower bounds on the scalar curvature, this in fact forces λ to be positive and comparable to $|\nu|$ in magnitude. Now if μ is also positive, then the equation $\frac{d}{dt}\nu = \nu^2 + \lambda\mu$ rapidly causes ν to be less negative, while the equation $\frac{d}{dt}\lambda = \lambda^2 + \mu\nu$ can cause λ to decrease, but not as rapidly as ν is increasing, thus the geometry does not become more pinched. If instead μ is negative, then ν can become more negative, but now λ will increase faster than ν is decreasing, thus increasing the pinching towards positive (consider e.g. the case when $\nu = \mu = -\lambda/2$).

Remark 3.4.18. There are further applications of the tensor maximum principle to Ricci flow. One notable one is *Hamilton's rounding theorem*[**Ha1982**], which asserts that if the Ricci curvature of a compact 3-manifold is strictly positive at time zero, then not only does

a singularity develop in finite time (by Proposition 3.4.10), but the geometry becomes increasingly round in the sense that the ratio between the largest and smallest eigenvalues of this curvature go to 1 as one approaches the singularity. In fact, the rescaled limit of the geometry here has constant positive sectional curvature and is thus either a sphere or a spherical space form.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/04/04`. Thanks to Paul Smith, Danny Calegari, Dan, Muhammad, and an anonymous commenter for corrections and references.

3.5. Finite time extinction of the second homotopy group

Recall from Section 3.3 that one of the key pillars of the proof of that conjecture is the finite time extinction result (see Theorem 3.3.13), which asserted that if a compact Riemannian 3-manifold (M, g) was initially simply connected, then after a finite amount of time evolving via Ricci flow with surgery, the manifold will be empty.

In this lecture and the next few, we will describe some of the key ideas used to prove this theorem. We will not be able to completely establish this theorem at present, because we do not have a full definition of "surgery", but we will be able to establish some partial results, and indicate (in informal terms) how to cope with the additional technicalities caused by the surgery procedure. (See also Section 3.19 for further discussion.)

The proof of finite time extinction proceeds in several stages. The first stage, which was already accomplished in the previous lecture (in the absence of surgery, at least), is to establish lower bounds on the least scalar curvature $R_{\min}$. The next stage, which we discuss in this lecture, is to show that the second *homotopy group* $\pi_2(M)$ of the manifold must become extinct in finite time, thus all *immersed copies* of the 2-sphere S^2 in M(t) for sufficiently large t must be contractible to a point. The third stage is to show that the third homotopy group $\pi_3(M)$ also becomes extinct so that all immersed copies of the 3-sphere S^3 in M are similarly contractible. The final stage, which uses homology theory, is to show that a non-empty 3-manifold cannot

have $\pi_1(M), \pi_2(M), \pi_3(M)$ simultaneously trivial, thus yielding the desired claim[36].

More precisely, in this section we will discuss (most of) the proof of

Theorem 3.5.1 (Finite time extinction of $\pi_2(M)$)**.** *Let $t \mapsto (M(t), g(t))$ be a Ricci flow with surgery on compact 3-manifolds with $t \in [0, +\infty)$, with $M(0)$ containing no embedded copy of $\mathcal{RP}^2$ with trivial normal bundle. Then for all sufficiently large t, $\pi_2(M(t))$ is trivial (or more precisely, every connected component of $M(t)$ has trivial π_2).*

The technical assumption about having no copy of $\mathcal{RP}^2$ with trivial normal bundle is needed solely in order to apply the known existence theory for Ricci flow with surgery (see Theorem 3.3.9).

The intuition for this result is as follows. From the *Gauss-Bonnet theorem* (and the fact that the *Euler characteristic* $\chi(S^2) = V - E + F = 2$ of the sphere is positive), we know that 2-spheres tend to have positive (Gaussian) curvature on the average, which should make them shrink under Ricci flow[37]. On the other hand, the presence of negative scalar curvature can counteract this by expanding these spheres. But the lower bounds on scalar curvature tell us that the negativity of scalar curvature becomes weakened over time, and it turns out that the shrinkage caused by the Gauss-Bonnet theorem eventually dominates and sends the area of all minimal immersed 2-spheres into zero, at which point one can conclude the triviality of $\pi_2(M)$ by the Sacks-Uhlenbeck theory[**SaUh1981**] of minimal 2-spheres.

The arguments here are drawn from [**MoTi2007**] and [**CoMi2005**]. The idea of using minimal surfaces to force disappearance of various topological structures under Ricci flow originates with Hamilton[**Ha1999**] (who used 2-torii instead of 2-spheres, but the idea is broadly the same).

---

[36]Note that a simply connected manifold has trivial $\pi_1(M)$ by definition; also, from Exercise 3.3.2 we see that all components of M remain simply connected even after surgery.

[37]Here I am conflating Gaussian curvature with Ricci curvature; however, by restricting to a special class of 2-spheres, namely *minimal surfaces*, one can connect the two notions of curvature to each other (and to scalar curvature) quite nicely, as we shall see.

3.5.1. Curvature on surfaces. We have seen how Riemannian manifolds (M, g) have various notions of curvature: Riemannian curvature Riem, Ricci curvature Ric, and scalar curvature R. These are *intrinsic* notions of curvature: they depend only on the manifold M (and its metric g), and not how this manifold is embedded (if it is embedded at all) in some larger space. However, there are some important *extrinsic* notions of curvature as well, which describe how an immersed manifold Σ is curved inside its ambient space M. In particular, we will recall the *Gauss curvature* K, *principal curvatures* λ_1, λ_2, and *mean curvature* H of a surface (i.e. a 2-dimensional manifold) Σ inside a 3-manifold[38] (M, g). We will also recall the standard fact that the mean curvature H vanishes whenever the surface is a minimal surface.

Let Σ be an immersed 2-surface in a Riemannian 3-manifold (M, g). All our computations here will be local, in the neighbourhood of some point x_0 in Σ (and thus in M; in particular we can pretend that the immersed manifold Σ is in fact embedded as a submanifold of M. If we let h be the restriction of the metric g to Σ (restricting TM to $T\Sigma$, etc.) then of course (Σ, h) is a Riemannian 2-manifold.

It is convenient to pick a unit normal vector field $n \in \Gamma(TM)$, thus n has norm 1 and is orthogonal to $T\Sigma$ at every point in Σ. It is only the value of n on the submanifold Σ which is important, but we will arbitrarily extend n smoothly to all of M so that we can take advantage of vector field operations on the ambient space. There is a choice of sign for n (e.g. if Σ bounded a three-dimensional region, we could pick either the outward or inward normal), which can lead to an ambiguity in sign in the principal and mean curvatures, but it will not affect the sign of the Gauss curvature.

Let $\nabla = \nabla^{(M)}$ be the Levi-Civita connection on M, and let X, Y be two vector fields which are tangential to Σ, thus $X(x), Y(x) \in T_x\Sigma$ for all $x \in \Sigma$. Then the covariant derivative $\nabla^{(M)}_X Y$ need not be tangential to Σ, but we can decompose

$$\nabla^{(M)}_X Y = \nabla^{(\Sigma)}_X Y + \Pi(X, Y)n, \tag{3.115}$$

[38]These notions can also be defined for other dimensions, but we will focus exclusively on the case of surfaces inside 3-manifolds.

where $\Pi(X,Y)n$ is the component of $\nabla^{(M)}_X Y$ parallel to n, and $\nabla^{(\Sigma)}_X Y$ is the component which is orthogonal to n (and in particular lies in $T\Sigma$) on Σ.

Exercise 3.5.1. Show that $\nabla^{(\Sigma)}$ is the Levi-Civita connection on (Σ, h), and that

$$\Pi(X,Y) = -g(\nabla^{(M)}_X n, Y) \tag{3.116}$$

on Σ. *Hint*: for the latter, compute the quantity $\nabla_X g(n,Y)$ in two different ways. Conclude that Π can be identified with a symmetric rank $(0,2)$ tensor (known as the *second fundamental form*) on Σ, which (up to sign) is independent of the choice of normal n.

Exercise 3.5.2. Using (3.115), deduce the *Gauss equation*[39]

$$g(\mathrm{Riem}^{(M)}(X,Y)Z,W) = g(\mathrm{Riem}^{(\Sigma)}(X,Y)Z,W) + \Pi(X,W)\Pi(Y,Z) - \Pi(X,Z)\Pi \tag{3.117}$$

on Σ, whenever X, Y, Z, W are vector fields that are tangent to Σ, and $\mathrm{Riem}^{(M)}$ and $\mathrm{Riem}^{(\Sigma)}$ are the Riemann curvature tensors of (M,g) and (Σ,h) respectively.

At any point $x \in \Sigma$, the second fundamental form $\Pi(x)$ can be viewed as a symmetric bilinear form on the two-dimensional space $T_x\Sigma$, which thus has two real eigenvalues $\lambda_1 \geq \lambda_2$, known as the *principal curvatures* of Σ (as embedded in M) in x. The normalised trace $H := \frac{1}{2}\mathrm{tr}(\Pi) = \frac{1}{2}(\lambda_1 + \lambda_2)$ of the second fundamental form is known as the *mean curvature*. Meanwhile, the *Gauss curvature* $K = K(x)$ at a point $x \in \Sigma$ is defined[40] as equal to half the scalar curvature of Σ: $K = \frac{1}{2}R^{(\Sigma)}$.

Exercise 3.5.3. Using Exercise 3.5.2, establish the identity

$$K = \det(\Pi) + K_M = \lambda_1\lambda_2 + K_M \tag{3.118}$$

where K_M is the *sectional curvature* of $T\Sigma$ in M, defined at a point x by the formula $K_M = g(\mathrm{Riem}^{(M)}(X,Y)X,Y)$ where X, Y are an

[39] One could of course write (3.117) in abstract index notation, but we have chosen not to do so to avoid confusion between the two bundles TM and $T\Sigma$ that are implicitly in play here.

[40] In particular, this manifestly demonstrates that the Gauss curvature K is intrinsic; this fact, combined with Exercise 3.5.3 below, is essentially the famous *theorema egregium* of Gauss.

orthonormal basis of $T\Sigma$ at x. In particular, if $M = (\mathbf{R}^3, \eta)$ is Euclidean space, then the Gauss curvature is just the product of the two principal curvatures (or equivalently, the determinant of the second fundamental form).

From (3.118) and the arithmetic mean-geometric mean inequality, we obtain in particular the following relationship between Gauss, mean, and sectional curvature:

$$K \leq H^2 + K_M. \tag{3.119}$$

Next, we now recall a special case of the Gauss-Bonnet theorem.

Proposition 3.5.2 (Gauss-Bonnet theorem for S^2)**.** *Let (Σ, h) be an immersion of the sphere S^2, and let $K := \frac{1}{2}R$ be the Gauss curvature. Then $\int_\Sigma K\, d\mu = 4\pi$, where μ is the volume measure (or area measure) associated to h.*

Proof. We use a flow-based argument. Since Gauss curvature is intrinsic, we may pull back and assume that Σ is in fact equal to S^2, but with some generic Riemannian metric which we shall call h_0, which may differ from the standard Riemannian metric on S^2, which we shall call h_1. We can flow from h_0 to h_1 by the linear flow $h(t) := (1-t)h_0 + th_1$ (say); note that this is a smooth flow on Riemannian metrics. Our task is to show that $\int_{S^2} R\, d\mu = 8\pi$ at time zero. By equations (3.50), (3.55), we have

$$\frac{d}{dt}\int_{S^2} R\, d\mu = \int_{S^2} (-\mathrm{Ric}^{\alpha\beta}\dot{h}_{\alpha\beta} - \Delta \mathrm{tr}(\dot{h}_{\alpha\beta}) + \nabla^\alpha \nabla^\beta \dot{h}_{\alpha\beta} + \frac{1}{2} R \mathrm{tr}(\dot{h}_{\alpha\beta}))\, d\mu. \tag{3.120}$$

The contribution of the second and third terms vanish thanks to Stokes' theorem (3.64). And in two dimensions, the Bianchi identities force the Ricci curvature $\mathrm{Ric}^{\alpha\beta}$ to be conformal, i.e. it is equal to $\frac{1}{2}Rh^{\alpha\beta}$. Thus the right-hand side of (3.120) vanishes completely, and so by the fundamental theorem of calculus, the value of $\int_{S^2} R\, d\mu$ at time 0 is equal to that at time 1. The claim then follows from the standard facts that S^2 with the usual metric has area 4π and constant scalar curvature $+2$ (or Gauss curvature $+1$). □

From this and (3.119) we conclude that

$$\int_\Sigma K_M + H^2 \, d\mu \geq 4\pi \tag{3.121}$$

for any immersed copy of S^2. Thus we can start lower bounding sectional curvatures on the average, as soon as we figure out how to deal with the mean curvature H.

To do this, we now specialise to immersed spheres Σ which are *minimal*; they have minimal area $\int_\Sigma \, d\mu$ with respect to smooth deformations. The following proposition is very well known:

Proposition 3.5.3. *Let Σ be a minimal immersed surface. Then the mean curvature H of Σ is identically zero.*

Proof. Let us consider a local perturbation of Σ. Working in local coordinates as before, we choose a unit normal field n, and flow Σ using the velocity field $Z := fn$, where f is a localised scalar function. This has the effect of deforming the metric h on Σ at the rate $\dot{h} = \mathcal{L}_Z g$, where $\mathcal{L}_Z$ is the Lie derivative along the vector field Z. By (3.55), the area of Σ will thus change under this deformation at the rate

$$\frac{d}{dt} \int_\Sigma \, d\mu = \int_\Sigma \frac{1}{2} \mathrm{tr}_h(\mathcal{L}_Z g) \, d\mu. \tag{3.122}$$

On the other hand, as Σ is minimal, the left-hand side vanishes. Also, using (3.61), we have

$$\mathrm{tr}_h(\mathcal{L}_Z g) = 2\nabla_\alpha Z_\beta (X^\alpha X^\beta + Y^\alpha Y^\beta) \tag{3.123}$$

where X, Y is an orthonormal frame of Σ (we can work locally, so as to avoid the topological obstruction of the *hairy ball theorem*). Expanding out $Z_\beta = f n_\beta$ and recalling that n is orthogonal to X and Y, some calculation using (3.116) allows us to express (3.123) as

$$-2f(\mathrm{II}(X,X) + \mathrm{II}(Y,Y)) = -4fH. \tag{3.124}$$

Putting all this together, we conclude that $\int_\Sigma fH \, d\mu = 0$ for all local perturbations f, which implies that H vanishes identically. □

It is an instructive exercise to try to convince oneself of the validity of Proposition 3.5.3 by pure geometric intuition regarding curvature and area.

From (3.121) and Proposition 3.5.3 we conclude a lower bound

$$\int_\Sigma K_\Sigma \, d\mu \geq 4\pi \tag{3.125}$$

for the integrated sectional curvature of a minimal immersed 2-sphere Σ in a 3-manifold M.

3.5.2. Minimal immersed spheres and Ricci flow. Now let (M, g) be a compact 3-manifold with a non-trivial second homotopy group $\pi_2(M)$. Thus there exist immersions $f : S^2 \to M$ which cannot be contracted to a point. It is a theorem of Sacks and Uhlenbeck[**SaUh1981**] that the area of such incontractible immersions cannot be arbitrarily small (for fixed M, g), and so if one defines $W_2(M)$ to be the infimum of the areas of all incontractible immersed spheres, then $W_2(M)$ is strictly positive.

It is a result of Meeks and Yau[**MeYa1980**] that the infimum here is actually attained, which would mean that there is a incontractible minimal immersed 2-sphere $f : S^2 \to M$ which has area exactly $W_2(M)$. However, it suffices for our purposes to use a simpler result that an incontractible minimal 2-sphere $f : S^2 \to M$ of area exactly $W_2(M)$ exists which is a *branched* immersion rather than an immersion, which roughly speaking means that there are a finite number of points in S^2 where the function f behaves like an embedding of the power function $z \mapsto z^n$ in the neighbourhood of the complex origin. See [**MoTi2007**, Lemma 18.10] for details[41]. For simplicity we shall ignore the effects of branching here; basically, branch points increase the integrated Gauss curvature in the Gauss-Bonnet theorem, but this effect turns out to have a favourable sign and is thus ultimately harmless.

Now suppose that $t \mapsto (M, g(t))$ is a Ricci flow for t in some time interval I. Suppose that t lies in I but is not the right endpoint of I. Then we have an incontractible minimal 2-sphere $f : S^2 \to M$ of area $W_2(M(t))$ which is a branched immersion; we will suppose that it is an immersion for simplicity. Let us now see how the area $\int_{f(S^2)} d\mu$

[41]Roughly, one needs to regularise the energy functional to obtain the *Palais-Smale condition*, then take limits to obtain a weak *harmonic map*, using a somewhat crude surgery argument to show that bubbling does not occur in the minimum area limit.

of $f(S^2)$ changes under Ricci flow. Using the variation formula (3.55) for the 2-dimensional measure $d\mu$, specialised to Ricci flow, we have

$$\frac{d}{dt}\int_{f(S^2)} d\mu = -\int_{f(S^2)} \mathrm{tr}_h(\mathrm{Ric}^{(M)})d\mu \tag{3.126}$$

where $\mathrm{Ric}^{(M)}$ is the Ricci curvature of the 3-manifold M, and h is the 2-dimensional metric formed by restricting g to $f(S^2)$. We now apply the following identity:

Exercise 3.5.4. Show that $\mathrm{tr}_h(\mathrm{Ric}^{(M)}) = K_{f(S^2)} + \frac{1}{2}R$, where $K_{f(S^2)}$ is the sectional curvature of $f(S^2)$ and R is the scalar curvature of M. *Hint*: use two tangent vectors of $f(S^2)$ and one normal vector to build an orthonormal basis, and write the Ricci and scalar curvatures in terms of sectional curvatures.

Inserting this identity into (3.126) and using (3.125), as well as the lower bound $R \geq R_{\min}$, we conclude that

$$\frac{d}{dt}\int_{f(S^2)} d\mu \leq -4\pi - \frac{1}{2}R_{\min}\int_{f(S^2)} d\mu; \tag{3.127}$$

by definition of $W_2(M(t))$, we thus conclude the ordinary differential inequality

$$\frac{d}{dt}W_2(M(t)) \leq -4\pi - \frac{1}{2}R_{\min}W_2(M(t)) \tag{3.128}$$

in the sense of forward difference quotients.

This is already enough to obtain a weak version of Theorem 3.5.1:

Theorem 3.5.4 (Non-trivial $\pi_2(M)$ implies finite time singularity)**.** *Let $t \mapsto (M, g(t))$ be a Ricci flow on a time interval $[0, T)$ for a compact 3-manifold with $\pi_2(M)$ non-trivial. Then T must be finite.*

Proof. At time zero, the minimal scalar curvature $R_{\min}(0)$ is of course finite. By rescaling if necessary we may assume $R_{\min}(0) \geq -1$ (say). Then Proposition 3.4.10 implies that $R_{\min}(t) \geq -3/(3+2t)$, and so from (3.128) we have

$$\frac{d}{dt}W_2(M(t)) \leq -4\pi + \frac{3}{6+4t}W_2(M(t)). \tag{3.129}$$

This can be rewritten (by the usual method of *integrating factors*) as

$$\frac{d}{dt}\big((6+4t)^{-3/4}W_2(M(t))\big) \leq -4\pi(6+4t)^{-3/4}. \tag{3.130}$$

Now, the expression $4\pi(6+4t)^{-3/4}$ is divergent when integrated from zero to infinity, while the expression $(6+4t)^{-3/4}W_2(M(t))$ is finite and non-negative. These two facts contradict each other if T is infinite, and so T is finite as claimed. □

Remark 3.5.5. This argument in fact gives an explicit upper bound for the time of development of the first singularity, in terms of the minimal Ricci curvature at time zero and minimal area of an immersed sphere at time zero.

We now briefly discuss how the same arguments can be extended to tackle Ricci flow with surgery, though this discussion will have to be somewhat informal since we have not yet fully defined what surgery is. The basic idea is to ensure that the inequality (3.128) persists through surgery. In a little more detail, the argument proceeds as follows:

(1) The first step is to clarify the topological nature of the surgery. It turns out that at each surgery time t, the manifold $M(t)$ can be obtained (in the topological category) from $M(t-)$ by finding a collection of disjoint 2-spheres in $M(t-)$, performing surgery on each 2-sphere to replace it with a pair of disks, then removing all but finitely many of the connected components that are created as a consequence.

(2) At any given time t, let $s(t)$ denote the maximal number of embedded 2-spheres one can place in $M(t)$ which are *homotopically essential* in the sense that none of these spheres can be contracted to a point, or deformed to any other sphere. It is possible to use homological arguments and van Kampen's theorem[**vKa1933**] to show that $s(t)$ is always finite.

(3) By homotopy theory, one can show that every time a surgery involves at least one homotopically essential sphere, the quantity $s(t)$ decreases by at least one. Thus, after a finite

number of surgeries, all spheres involved in surgery are contractible to a point. By shifting the time variable if necessary, we may thus assume that the above claim is true for all times $t \geq 0$.

(4) Once all spheres involved in surgery are contractible, one can show that whenever surgery is applied to a connected manifold, either the manifold is removed completely, or one of the post-surgery components is homotopy equivalent to the original manifold, and the rest are homotopy spheres. In particular, if a connected manifold has non-trivial π_2 before surgery, then it is either removed by surgery, or one of the post-surgery components has the same π_2; and if a connected manifold has trivial π_2 then all post-surgery components do also. Thus if Theorem 3.5.1 fails, one can find a "path of components" through the Ricci flow with surgery with non-trivial π_2 for all time. We now restrict attention to this path of components, which by abuse of notation we shall continue to call $M(t)$ at each time t.

(5) Using the geometric properties of the surgery and standard limiting arguments, we can show that if $R_{\min}$ is non-positive before surgery, then it cannot decrease as a consequence of surgery (thus $R_{\min}(t) \geq \lim_{t' \to t^-} R_{\min}(t')$, and similarly if $R_{\min}$ is non-negative before surgery, then it stays non-negative after surgery (here we adopt the convention that $R_{\min} = +\infty$ when the manifold is empty). These facts are ultimately because surgery is only performed in regions of high positive curvature. From this, one can conclude (assuming the initial normalisation $R_{\min}(0) \geq -1$ that the bound $R_{\min}(t) \geq -3/(3+2t)$ persists even after surgery.

(6) Finally, using the geometric properties of the surgery and standard limiting arguments, one can show that $W_2(M(t))$ has no upward jump discontinuity at surgery times t in the sense that $W_2(M(t)) \leq \liminf_{t' \to t^-} W_2(M(t'))$. This allows us to repeat the proof of Theorem 3.5.4 and obtain the desired contradiction to prove Theorem 3.5.1.

Further details can be found in [**MoTi2007**, Section 18.12].

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/04/11`. Thanks to wenwen, and Andy Sanders for corrections and references.

3.6. Finite time extinction of the third homotopy group, I

In Section 3.5, we saw that Ricci flow with surgery ensures that the second homotopy group $\pi_2(M)$ became extinct in finite time (assuming, as stated in the above erratum, that there is no embedded $\mathcal{RP}^2$ with trivial normal bundle). It turns out that the same assertion is true for the third homotopy group, at least in the simply connected case[42]:

Theorem 3.6.1 (Finite time extinction of $\pi_3(M)$)**.** *Let $t \mapsto (M(t), g(t))$ be a Ricci flow with surgery on compact 3-manifolds with $t \in [0, +\infty)$, with $M(0)$ simply connected. Then for all sufficiently large t, $\pi_3(M(t))$ is trivial (or more precisely, every connected component of $M(t)$ has trivial π_3).*

Suppose we apply Ricci flow with surgery to a compact simply connected Riemannian 3-manifold (M, g) (which, by Lemma 3.3.10, has no embedded $\mathcal{RP}^2$ with trivial normal bundle). From the above theorem, as well as Theorem 3.5.1, we know that all components of $M(t)$ eventually have trivial π_2 and π_3 for all sufficiently large t. Also, since M is initially simply connected, we see from Exercise 3.3.2, as well as Theorem 3.3.9.1, that all components of $M(t)$ also have trivial π_1. The finite time extinction result (Theorem 3.3.13) then follows immediately from Theorem 3.6.1 and the following topological result, combined with the following topological observation:

Lemma 3.6.2. *Let M be a compact non-empty connected 3-manifold. Then it is not possible for $\pi_1(M)$, $\pi_2(M)$, and $\pi_3(M)$ to simultaneously be trivial.*

This lemma follows immediately from the *Hurewicz theorem*, but for sake of self-containedness we shall give a proof of it in this section.

[42] It seems likely that this theorem should also be true if one merely assumes that $M(0)$ contains no embedded copy of $\mathcal{RP}^2$ with trivial bundle, as opposed to $M(0)$ being simply connected, but I will be conservative and only state Theorem 3.6.1 with this stronger hypothesis, as this is all that is necessary for proving the Poincaré conjecture.

There are two known approaches to establishing Theorem 3.6.1; one due to Colding and Minicozzi[**CoMi2005**], and one due to Perelman[**Pe2003**]. The former is conceptually simpler, but requires a certain technical concentration-compactness type property for a min-max functional which has only been established recently[**CoMi2007**]. This approach will be the focus of this section, while the latter approach of Perelman, which has also been rigorously shown to imply finite time extinction, will be the focus of the next section.

3.6.1. A little algebraic topology. We begin by proving Lemma 3.6.2. We need to recall some (very basic) singular homology theory (over the integers $\mathbf{Z}$). Fix a compact manifold M, and let k be a non-negative integer. Recall that a *singular k-chain* (or *k-chain* for short) is a formal integer-linear combination of k-dimensional singular simplices $\sigma(\Delta_k)$ in M, where Δ_k is the standard k-simplex and $\sigma : \Delta_k \to M$ is a continuous map. There is a boundary map ∂ taking k-chains to $(k-1)$-chains, defined by mapping $\sigma(\Delta_k)$ to an alternating sum of restrictions of σ to the $(k-1)$-dimensional boundary simplices of Δ_k, and then extending by linearity. A k-chain is said to be a *k-cycle* if its boundary vanishes, and is a *k-boundary* if it is the boundary of a $(k+1)$-chain. One easily verifies that $\partial^2 = 0$, and so every k-boundary is a k-cycle. We say that M has *trivial k^{th} homology group $H_k(M)$* if the converse is true, i.e. every k-cycle is a k-boundary.

Our main tool for proving Lemma 3.6.2 is

Proposition 3.6.3 (Baby Hurewicz theorem)**.** *Let M be a triangulated connected compact manifold such that the fundamental groups $\pi_1(M), \ldots, \pi_k(M)$ all vanish for some $k \geq 1$. Then M has trivial j^{th} homology group $H_j(M)$ for every $1 \leq j \leq k$.*

Proof. Because of all the vanishing fundamental groups, one can show by induction on j that for any integer $1 \leq j \leq k$, any singular complex in M involving singular simplices of dimension at most j can be continuously deformed to a point (while preserving all boundary relationships between the singular simplices in that complex). As a consequence, every j-cycle, being the combination of singular simplices in a singular complex involving singular simplices of dimension

at most j, can be expressed as the boundary of a $(j+1)$-chain, and the claim follows. □

Remark 3.6.4. The full Hurewicz theorem asserts some further relationships between homotopy groups and homology groups, in particular that under the assumptions of Proposition 3.6.3, that the *Hurewicz homomorphism* from $\pi_{k+1}(M)$ to $H_{k+1}(M)$ is in fact an isomorphism (and, in the $k=0$ case, that $H_1(M)$ is canonically isomorphic to the abelianisation of $\pi_1(M)$). However, we do not need this slightly more advanced result here.

Now we can quickly prove Lemma 3.6.2.

Proof of Lemma 3.6.2. Suppose for contradiction that we have a non-empty connected compact 3-manifold M with $\pi_1(M), \pi_2(M), \pi_3(M)$ all trivial. Since M is simply connected, it is orientable (as all loops are contractible, there can be no obstruction to extending an orientation at one point to the rest of the manifold). Also, it is a classical result[43] of Moise[**Mo1952**] that every 3-manifold can be triangulated. Using a consistent orientation on M, we may therefore build a 3-cycle on M consisting of the sum of oriented 3-simplices with disjoint interiors that cover M (i.e. a *fundamental class*), thus the net multiplicity of this cycle at any point in M is odd. On the other hand, the net multiplicity of any 3-boundary at any point can be seen to necessarily be even. Thus we have found a 3-cycle which is not a 3-boundary, which contradicts Proposition 3.6.3. □

Remark 3.6.5. Using (slightly) more advanced tools from algebraic topology, one can in fact say a lot more about the homology and homotopy groups of connected and simply connected compact 3-manifolds M. Firstly, since $\pi_1(M)$ is trivial, one sees from the full Hurewicz theorem that $H_1(M)$ is also trivial. Also, as M is connected, $H_0(M) \equiv \mathbf{Z}$. From orientability (which comes from simple connectedness) and triangularisability we have *Poincaré duality*, which implies that the cohomology group $H^2(M)$ is trivial and

[43] One can avoid the use of Moise's theorem here by working in the category of smooth manifolds, or by using more of the basic theory of singular homology. Also, the use of orientability can be avoided by working with homologies over $\mathbf{Z}/2\mathbf{Z}$ rather than over $\mathbf{Z}$.

$H^3(M) \equiv \mathbf{Z}$, which by the *universal coefficient theorem* for cohomology implies that $H_2(M)$ is trivial and $H_3(M) \equiv \mathbf{Z}$. Of course, being 3-dimensional, all higher homology groups vanish, and so M is a *homology sphere*. On the other hand, by orientability, we can find a map from M to S^3 that takes a fundamental class of M to a fundamental class of S^3, by taking a small ball in M and contracting everything else to a point; this map is thus an isomorphism on each homology group. Using the *relative Hurewicz theorem* (and the simply connected nature of M and S^3) we conclude that this map is also an isomorphism on each homotopy group, and thus by *Whitehead's theorem*, the map is a homotopy equivalence, thus M is a *homotopy sphere*. Thus, to complete the proof of the Poincaré conjecture, it suffices to show that every compact 3-manifold which is a homotopy sphere is also homeomorphic to a sphere. Unfortunately this observation does not seem to significantly simplify the proof of that conjecture[44], although it does allow one at least to get the extinction of π_2 from the previous lecture "for free" in the simply connected case.

3.6.2. The Colding-Minicozzi approach to π_3 extinction. We now sketch the Colding-Minicozzi approach towards proving Theorem 3.6.1. Our discussion here will not be fully rigorous; further details can be found in [**CoMi2005**], [**CoMi2007**].

In the previous lecture, we obtained the differential inequality

$$\frac{d}{dt}\int_{f(S^2)} d\mu \leq -4\pi - \frac{1}{2}R_{\min}\int_{f(S^2)} d\mu; \tag{3.131}$$

for any minimal immersed 2-sphere $f(S^2)$ in a Ricci flow $t \mapsto (M(t), g(t))$. The inequality also holds for the slightly larger class of minimal 2-spheres that are *branched* immersions rather than just immersions; furthermore, an inspection of the proof reveals that the surface does not actually have to be a local area minimiser, but merely needs to have zero mean curvature (i.e. to be a critical point for the area functional, rather than a local minimum). The inequality (3.131) was a key ingredient in the proof of finite time extinction of the second homotopy group $\pi_2(M(t))$.

[44]Note also that there are homology 3-spheres that are not homeomorphic to the 3-sphere, such as the *Poincaré homology sphere*; thus homology theory is not sufficient by itself to resolve this conjecture.

The Colding-Minicozzi approach seeks to exploit the same inequality (3.131) to also prove finite time extinction of $\pi_3(M(t))$. It is not immediately obvious how to do this, since $\pi_3()$ involves immersed 3-spheres $f(S^3)$ in M, whereas (3.131) involves immersed 2-spheres $f(S^2)$. However, one can view the 3-sphere as a loop of 2-spheres with fixed base point; indeed if one starts with the cylinder $[0,1] \times S^2$ and identifies $\{0,1\} \times S^2 \cup [0,1] \times \{N\}$ to a single point, where N is a single point in S^2, one obtains a (topological) 3-sphere. Because of this, any immersed 3-sphere $f(S^3)$ is swept out by a loop $s \mapsto f_s$ of immersed 2-spheres $f_s(S^2)$ for $0 < s < 1$ with fixed base point $f_s(N) = p$, with f_s varying continuously in t for $0 \leq s \leq 1$, and $f_0 = f_1 \equiv p$ being the trivial map.

Suppose that we have a Ricci flow in which M is connected and $\pi_3(M)$ is non-trivial; then we have at least one immersed 3-sphere $f(S^3)$ which is not contractible to a point. We then define the functional $W_3(t)$ by the min-max formula

$$W_3(t) := \inf_f \sup_{0 \leq s \leq 1} \int_{f_s(S^2)} d\mu \tag{3.132}$$

where f ranges over all incontractible immersed 3-spheres, and μ is the volume element of $f_s(S^2)$ with respect to the restriction of the ambient metric $g(t)$.

It can be shown (see e.g. [**Jo1991**, page 125]) that $W_3(t)$ is strictly positive; in other words, if the area of each 2-sphere in a loop of immersed 2-spheres is sufficiently small, then the whole loop is contractible to a point.

Suppose for the moment that the infimum in (3.132) was actually attained, thus there exists an incontractible immersed 3-sphere f whose maximum value of $\int_{f_s(S^2)} d\mu$ is exactly $W_3(t)$. Applying mean curvature flow for a short time[45] to reduce the area of any sphere which does not already have vanishing mean curvature, we may assume without loss of generality that the maximum value is only attained when $f_s(S^2)$ has zero mean curvature. If we then use

[45]To make this rigorous, one either has to prove a local well-posedness result for mean curvature flow, or else to use a cruder version of this flow, for instance deforming f a small amount along a vector field which points in the same direction as the mean curvature. We omit the details.

(3.131), we thus (formally, at least) obtain the differential inequality

$$\frac{d}{dt}W_3(t) \leq -4\pi - \frac{1}{2}R_{\min}W_3(t) \tag{3.133}$$

much as in Section 3.5. Arguing as in that lecture, we obtain a contradiction if the Ricci flow persists without developing singularities for a sufficiently long time.

A similar analysis can also be performed when the infimum in (3.132) is not actually attained, in which case one has a minimising sequence of loops of 2-spheres whose width approaches $W_3(t)$. This sequence can be analysed by Sacks-Uhlenbeck theory (together with some later analysis of bubbling due to Siu and Yau) and a finite number of minimal 2-spheres extracted as a certain "limit" of the above sequence, although as in the previous lecture, these 2-spheres need only branched immersions rather than immersions. From this one can establish (3.133) (in a suitably weak sense) in the general case in which the infimum in (3.132) is not necessarily attained, assuming that one can show that all the 2-spheres with area close to $W_3(t)$ that appear in a loop in the minimising sequence are close to the union of the limiting minimal 2-spheres in a certain technical sense; see [**CoMi2005**] (and the references therein) for details. This property (which is a sort of concentration-compactness type property for the min-max functional (3.132), which is a partial substitute for the failure of the Palais-Smale condition for this functional) was recently established [**CoMi2007**], using the theory of harmonic maps.

There is also the issue of how to deal with surgery. This follows the same lines that were briefly (and incompletely) sketched out in the previous lecture. Namely, one first observes that after finitely many surgeries, all remaining surgeries are along 2-spheres that are homotopically trivial (i.e. contractible to a point). Because of this, one can show that any incontractible 3-sphere will, after surgery, lead to at least one incontractible 3-sphere on one of the components of the post-surgery manifold. Furthermore, it turns out that there is a distance-decreasing property of surgery which can be used to show that $W_3(t)$ does not increase at any surgery time. We will discuss these sorts of issues in a bit more detail in the next section, when we turn to the Perelman approach to π_3 extinction.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/04/15`. Thanks to Greg Kuperberg, Reid Barton for corrections and references.

Greg Kuperberg pointed out a more elementary proof of Lemma 3.6.2 in the category of triangulated 3-dimensional manifolds; in fact one can show the slightly stronger statement that if $\pi_1(M)$ and $\pi_2(M)$ both vanish, then M is homotopy equivalent to S^3. To see this, consider the more general method of building a manifold by gluing together polyhedra face to face. Then one can always reduce the number of 3-cells to 1 by knocking out walls; ultimately, one ends up with a single polyhedron to itself. Since $\pi_1(M)$ and $\pi_2(M)$ both vanish, you can homotope the identity map on M to collapse its 2-skeleton to a point. This collapse of M is plainly homeomorphic to S^3. So you immediately get pair of maps $f : M \to S^3$ and $g : S^3 \to M$ such that the composition gf is homotopic to the identity. The other composition is also homotopic to the identity because it has degree 1, and the claim follows.

3.7. Finite time extinction of the third homotopy group, II

In this section we discuss Perelman's original approach to finite time extinction of the third homotopy group (Theorem 3.6.1), which, as previously discussed, can be combined with the finite time extinction of the second homotopy group to imply finite time extinction of the entire Ricci flow with surgery for any compact simply connected Riemannian 3-manifold, i.e. Theorem 3.3.13.

3.7.1. Minimal disks. In Section 3.5, we studied *minimal immersed spheres* $f : S^2 \to M$ into a three-manifold, and how their area varied with respect to Ricci flow. This area variation formula was used to establish π_2 extinction, and was also used in the Colding-Minicozzi approach to π_3 extinction (see Section 3.6). The Perelman approach is similar, but is based upon minimal disks rather than minimal 2-spheres, which we will define as Lipschitz immersed maps $f : D^2 \to M$ from the unit disk D to M which are smooth on the interior of the disk, and with mean curvature zero on the interior of the disk.

For simplicity let us restrict attention to 3-manifolds (M, g) which are simply connected (this case is, of course, our main concern in this course). Then every loop $\gamma : S^1 \to M$ spans at least one disk. Let $A(\gamma, g)$ denote the minimal area of all such spanning disks. From the work of Morrey[**Mo1948**] and Hildebrandt[**Hi1969**] on *Plateau's problem in Riemannian manifolds*, it is known that this area is in fact attained by a minimal disk[46] whose boundary traces out γ. One can think of $A(\gamma, g)$ as the two-dimensional generalisation of the distance function $d(x, y)$ between two points x, y (which one can think of a map from S^0 to M). For instance, we have the following first variation formula for $A(\gamma, g)$ analogous to that for the distance function.

Lemma 3.7.1. *(First variation formula) Let $\gamma : S^1 \to M$ be a loop in a 3-manifold (M, g), and let $f : D^2 \to M$ be a minimal-area disk spanning γ, thus $\int_{f(D^2)} d\mu = A(\gamma, g)$. Let $t \mapsto \gamma_t$ be a smooth deformation of γ with $\gamma_0 = \gamma$. Then we have*

$$\frac{d}{dt} A(\gamma_t, g)|_{t=0} \leq \int_\gamma g(N, \frac{d}{dt}\gamma_t|_{t=0})\ ds \tag{3.134}$$

where ds is the length element and n is the outward normal vector to $f(D^2)$ on the boundary γ.

Proof. First suppose that $\frac{d}{dt}\gamma_t|_{t=0}$ is orthogonal to the disk $f(D^2)$. Then one can deform the disk $f(D^2)$ to span γ_t for infinitesimally non-zero times t by flowing the disk along a vector field normal to that disk. Since $f(D^2)$ is minimal, it has mean curvature zero, and so the first variation of the area in this case is zero by the calculation used to prove Proposition 3.5.3. Since the area of this deformed disk is an upper bound for $A(\gamma_t, g)$, this proves (3.134) in this case.

In the case when $\frac{d}{dt}\gamma_t|_{t=0}$ is tangential to $f(D^2)$, the claim is clear simply by modifying the disk $f(D^2)$ at the boundary to accommodate the change in γ_t with respect to the time parameter t. The general case then follows by combining the above two arguments. □

Now we let the manifold evolve by Ricci flow, and obtain a similar variation formula:

[46] The fact that this disk is immersed was established in [**GuLe1973**], [**HaSi1985**].

Corollary 3.7.2 (First variation formula with Ricci flow)**.** *Let $t \mapsto (M, g(t))$ be a Ricci flow, and for each time t let $\gamma_t : S^1 \to M$ be a loop in a 3-manifold (M, g) smoothly varying in t, and let $f_t : D^2 \to M$ be a minimal-area disk spanning γ_t. Then we have*

$$\begin{aligned}\frac{d}{dt}A(\gamma_t, g(t)) \leq &-\int_{f_t(D^2)} K_{f_t(D^2)}\ d\mu - \frac{1}{2}R_{\min}A(\gamma_t, g(t)) \\ &+ \int_{\gamma_t} g(N_t, \frac{d}{dt}\gamma_t)\ ds\end{aligned} \tag{3.135}$$

where $K_{f_t(D^2)}$ is the Gauss curvature of $f_t(D^2)$.

Proof. This follows from the chain rule and the computations used to derive (3.127). □

To deal with the Gauss curvature term, we need an analogue of the Gauss-Bonnet theorem for disks. Fortunately, we have such a result:

Proposition 3.7.3 (Gauss-Bonnet for disks)**.** *Let $f : D^2 \to M$ be an immersed disk with boundary γ. Then we have*

$$\int_{f(D^2)} K_{f(D^2)}\ d\mu + \int_\gamma k_{\gamma, f(D^2)}\ ds = 2\pi \tag{3.136}$$

where[47] *$k_{\gamma,f(D^2)} = -g(\nabla_T T, N)$ is the signed curvature of the curve γ relative to the disk $f(D^2)$.*

Proof. We use another flow argument. All quantities here are intrinsic and so we may pull back to the unit disk D^2. Our task is now to show that

$$\int_{D^2} K\ d\mu - \int_{S^1} g(\nabla_T T, N)\ ds = 2\pi \tag{3.137}$$

for all metrics (D^2, g) on the unit disk. (Note that the vectors T, N will depend on g.) By the argument used to prove Proposition 3.5.2, the left-hand side is invariant under any *compactly supported* perturbation of the metric g, so we may assume that the metric is Euclidean on some neighbourhood of the origin.

[47] Here T is the unit tangent vector to γ, oriented in either direction; we are also abusing notation slightly by pulling back the Levi-Civita connection on TM to the pullback bundle S^1 in order to define $\nabla_T T$ properly.

We express D^2 in polar coordinates (r, θ). It will then suffice to establish the $R = 1$ case of the identity

(3.138)
$$\int_0^R \int_0^{2\pi} K(r,\theta)(\partial_r \wedge \partial_\theta)\ d\theta dr - \int_0^{2\pi} (\nabla_{\partial_\theta} T \wedge T)(R,\theta)\ d\theta = 2\pi$$

where we use the metric g to identify the 2-form $\partial_r \wedge \partial_\theta$ with a scalar, and T and N on (R, θ) are the tangent and outward normal vectors to the circle $\{r = R\}$ (the orientation of T is not relevant, but let us fix it as anticlockwise for sake of discussion, with the orientation chosen so that $N \wedge T$ is positive). Note that we are heavily relying here on the two-dimensionality of the situation!

Because the metric is Euclidean near the origin, (3.138) is true for R close to zero. Thus by the fundamental theorem of calculus, it suffices to verify the identity

$$\int_0^{2\pi} K(r,\theta)(\partial_r \wedge \partial_\theta)\ d\theta - \int_0^{2\pi} \partial_r(\nabla_{\partial_\theta} T \wedge T)(R,\theta)\ d\theta = 0. \tag{3.139}$$

for all $0 < r < 1$. But as the Levi-Civita connection respects the metric (and all constructions arising from that metric, such as the identification of 2-forms with scalars) we have

$$\partial_r(\nabla_{\partial_\theta} T \wedge T) = (\nabla_{\partial_r} \nabla_{\partial_\theta} T \wedge T) + (\nabla_{\partial_\theta} T \wedge \nabla_{\partial_r} T). \tag{3.140}$$

Since T is always a unit vector, $\nabla_{\partial_\theta} T$ and $\nabla_{\partial_r} T$ must be orthogonal to T and thus (in this two-dimensional setting) must be parallel, so the last term in (3.140) vanishes. An integration by parts then shows that $(\nabla_{\partial_\theta} \nabla_{\partial_r} T \wedge T$ has vanishing integral. Finally, from the Bianchi identities that allow one to express Riemann curvature of 2-manifolds in terms of Gauss curvature, we have

$$(\nabla_{\partial_r} \nabla_{\partial_\theta} - \nabla_{\partial_\theta} \nabla_{\partial r}) T \wedge T = K(\partial_r \wedge \partial_\theta) \tag{3.141}$$

and the claim follows. □

Exercise 3.7.1. Use Proposition 3.7.3 to reprove Proposition 3.5.2.

We can now combine Corollary 3.7.2 and Proposition 3.7.3 as follows. We say that a family of curves $\gamma_t : S^1 \to M$ is undergoing *curve-shortening flow* if we have

$$\frac{\partial}{\partial t} \gamma_t = \nabla_T T \tag{3.142}$$

where T is the tangent vector to γ_t.

Exercise 3.7.2. If γ_t undergoes curve-shortening flow, show that

$$\frac{d}{dt}\int_{\gamma_t} ds = -\frac{1}{2}\int_{\gamma_t} |\nabla_T T|_g^2 \, ds \tag{3.143}$$

which may help explain the terminology "curve-shortening flow".

Corollary 3.7.4 (First variation formula with Ricci flow and curve shortening flow)**.** *Let $t \mapsto (M, g(t))$ be a Ricci flow, and for each time t let $\gamma_t : S^1 \to M$ be a loop in a 3-manifold (M, g) undergoing curve shortening flow, and let $f_t : D^2 \to M$ be a minimal-area disk spanning γ_t. Then we have*

$$\frac{d}{dt} A(\gamma_t, g(t)) \leq -2\pi + \frac{1}{2} R_{\min} A(\gamma_t, g(t)). \tag{3.144}$$

3.7.2. Perelman's width functional. We now begin a non-rigorous discussion of Perelman's width functional, and how it is used to derive finite time π_3 extinction. There is a significant analytical difficulty regarding singularities in curve shortening flow, but we will address this issue later.

To simplify the exposition slightly, we will restrict attention to compact 3-manifolds whose components are all simply connected, and take advantage of Remark 3.6.5, although one can avoid use of this remark (and extend the analysis here to slightly more general manifolds, namely those with fundamental group a direct sum of cyclic groups and finite groups, and which contain no embedded $\mathcal{RP}^2$ with trivial normal bundle) by using the π_2 extinction theory from Section 3.5.

By this remark, all connected components of such manifolds are homotopy spheres, and in particular have trivial π_2 and π_3 isomorphic to the integers; thus every map $f : S^3 \to M$ has a degree $\deg(f) \in \mathbf{Z}$. This degree only fixed up to sign, so we shall work primarily with the magnitude $|\deg(f)|$ of this degree.

Let M be one of these connected components, and fix a base point $x_0 \in M$.

We can identify[48] S^3 with the space $S^2 \times S^1/(\{p_2\} \times S^1) \cup (S^2 \times \{p_1\})$, where p_1, p_2 are base points of S^1, S^2 respectively, thus we contract $\{p_2\} \times S^1$ and $S^2 \times \{p_1\}$ to a single point p_3. Thus, any map $f : S^3 \to M$ with $f(p_3) = x_0$ can be viewed as a family $\gamma_\omega : S^1 \to M$ of loops with fixed base point $\gamma_\omega(p_1) = x_0$ for $\omega \in S^2$, such that γ_ω varies continuously in ω and is identically equal to x_0 when $\omega = p_2$.

A little more generally, define a *loop family* $\gamma = (\gamma_\omega)_{\omega \in S^2}$ to be a family[49] $\gamma_\omega : S^1 \to M$ of loops parameterised continuously by $\omega \in S^2$, such that $\gamma_{p_2} \equiv x_0$. Thus we see that every map $f : S^3 \to M$ with $f(p_3) = x_0$ generates a loop family. The converse is not quite true, because we are not requiring the loops γ_ω in a loop family to have fixed base point (i.e. we do not require $\gamma_\omega(p_1) = x_0$ for all ω, only for $\omega = p_2$). However, as $\pi_2(M)$ is trivial, the 2-sphere $\omega \mapsto \gamma_\omega(p_1)$ is contractible, and so every loop family is homotopic to a loop family associated to a map $f : S^3 \to M$, and so in particular can be assigned a degree $|\deg(\gamma)|$. This degree is well-defined and stable under deformations:

Exercise 3.7.3. Show that each loop family γ is associated to a unique degree magnitude, no matter how one chooses to contract the 2-sphere $\omega \mapsto \gamma_\omega(p_1)$. Also, show that if a loop family γ can be continuously deformed to another loop family $\tilde{\gamma}$ while staying within the class of loop families, then both loop families have the same degree . Conclude that the space of homotopy classes $\pi_2(\Lambda M, x_0)$ of loop families can be canonically identified with $\pi_3(M) \equiv \mathbf{Z}$.

Exercise 3.7.4. Show that for any $d \geq 1$, the quotient $S^d \times S^1/\mathrm{pt} \times S^1$ is homotopy equivalent to the *wedge sum* $S^d \vee S^{d+1}$, and then use this to give another proof of Exercise 3.7.3. *Hint*: [50] first show that both spaces are homotopy equivalent to a sphere S^{d+1} with a disk D^d glued to it (identifying the boundary ∂D^d of the disk with some copy of S^{d-1} in S^{d+1}. The case $d = 1$ might be easiest to visualise.

---

[48] To see why these spaces are topologically isomorphic, use the standard identification of n-sphere S^n with an n-cube $[0,1]^n$ with the entire boundary identified with a point p_n.

[49] To put it another way, a loop family is a continuous map from S^2 to the *loop space* ΛM which has the constant loop x_0 as base point; equivalently, a loop family is a continuous map from $S^2 \times S^1/\{p_2\} \times S^1$ which maps $\{p_2\} \times S^1$ to x_0.

[50] Thanks to Kenny Maples, Peter Petersen, and Paul Smith for this hint.

Given a loop family $\gamma = (\gamma_\omega)_{\omega \in S^2}$, define the *width* $\tilde{W}_3(\gamma)$ of this family to be the quantity

$$\tilde{W}_3(\gamma) := \sup_{\omega \in S^2} A(\gamma_\omega) \tag{3.145}$$

and then for every non-negative $\xi \in \mathbf{Z}^+$, define the width $\tilde{W}_3(\xi)$ to be the quantity

$$\tilde{W}_3(\xi) := \inf_{\gamma : |\deg(\gamma)| = \xi} \tilde{W}_3(\gamma) \tag{3.146}$$

(this is an inf of a sup of an inf!). We can define this concept for non-empty disconnected manifolds M also, by taking the infimum across all components and all choices of base point.

I do not know if $\tilde{W}_3(\xi)$ is always positive when M is non-empty and ξ is positive (or equivalently, that if one has a loop family in which each loop is spanned by a disk of small area, that the entire loop family is contractible to a point). However, one can at least say that if γ is a loop family associated to a non-trivial degree ξ, then the length $\int_{\gamma_\omega} ds$ of at least one of the loops γ_ω is bounded away from zero by some constant depending only on $M = (M, g)$, because if instead all loops had small length, then they could be contracted to a point, thus degenerating the loop family to an image of S^2, which is contractible since we are assuming $\pi_2(M)$ to be trivial. This lower bound on length is important for technical reasons (which we are mostly suppressing here).

Let us temporarily pretend, though, that at some point in time during a Ricci flow $t \mapsto (M, g(t))$, that $\tilde{W}_3(\xi) = \tilde{W}_3(\xi, t)$ is positive for some positive ξ, and that the infimum in (3.146) is attained by a smooth loop family γ, thus $A(\gamma_\omega, g(t))$ attains a maximum value of $\tilde{W}_3(\xi, t)$ for some $\omega \in S^2$.

We now run the Ricci flow, while simultaneously deforming each loop γ_ω in the loop family by curve-shortening flow (local existence for the latter flow is a result of Gage and Hamilton[**GaHa1986**]). Applying (3.144), we conclude that

$$\frac{d}{dt} \tilde{W}_3(\gamma) \leq -2\pi - \frac{1}{2} R_{\min} \tilde{W}_3(\gamma) \tag{3.147}$$

(in the sense of forward difference quotients), and thus by assumption on γ that

$$\frac{d}{dt}\tilde{W}_3(\xi) \leq -2\pi - \frac{1}{2}R_{\min}\tilde{W}_3(\xi). \tag{3.148}$$

Now we investigate what happens when a surgery occurs. It turns out that whenever a component of a pre-surgery manifold is disconnected into components of a post-surgery manifold, that there exist degree 1 (or -1) maps from the pre-surgery components to each of the post-surgery components (recall that all components are homotopy spheres, and in particular the 2-spheres that one performs surgery on are automatically contractible). Furthermore, these maps can be chosen to have Lipschitz constant less than $1+\eta$ for any fixed $\eta > 0$, thus they are almost contractions. (We will discuss this fact later in this course, when we define surgery properly.) Because of this, we can convert any loop family on the pre-surgery component to a loop family on the post-surgery component which has the same degree magnitude and which has only slightly larger width at worst. Because of this, we can conclude that $\tilde{W}_3(\xi)$ does not increase during surgery.

By arguing as in Section 3.5 we now conclude (using (3.148) and lower bounds in $R_{\min}$) that either the manifold becomes totally extinct or that $\tilde{W}_3(\xi)$ becomes negative. The latter is absurd, and so we obtain the required finite time extinction (indeed, we have shown extinction not just of π_3 here, but of the entire manifold).

3.7.3. Ramps. The above argument had one significant gap in it; it assumed that the infimum in (3.146) was always attained. In practice, this is not necessarily the case, and so the best one can do is find loop families γ for each time t with homotopy class ξ whose width is within ε of the minimal width $\tilde{W}_3(\xi, t)$, for any small $\varepsilon > 0$. One can try to run the above arguments with this near-minimiser γ in place of an exact minimiser, but in order to do so, it is necessary to ensure that the curve-shortening flow, when applied to γ, exists for a period of time that is bounded from below uniformly in ε.

Unfortunately, the local existence theory of [**GaHa1986**] (see also [**AlGr1992**]) only guarantees such a uniform lower bound on time of existence when the curvature magnitude $\kappa := |\nabla_T T|_g$ of these

curves is uniformly bounded from above[51]. And, in general, such curvature bounds are not available[52].

To resolve this moderately serious technical obstacle, Perelman employed the use of *ramps*, following the work of [**AlGr1992**] (see also a related argument in [**EcHu1991**]). The basic idea is to give all the loops an upward "slope" that is bounded from below, which (in conjunction with the maximum principle) will prevent singularities from forming. In order to create this upward slope, it is necessary to increase the dimension of the ambient manifold M by one, working with[53] $M \times S^1_\lambda$ instead of M, where $S^1_\lambda = \mathbf{R}/\lambda\mathbf{Z}$ is the circle of length λ for some small $\lambda > 0$.

We now turn to the details. We first develop some general variation formulae and estimates for a curve-shortening flow $t \mapsto \gamma_t$ in a time-varying Riemannian manifold $(M, g(t))$ of arbitrary dimension. As before, we let T denote the unit tangent vector along γ_t. We write $H := \nabla_T T = \frac{\partial}{\partial t}\gamma$ for the curvature vector, which is of course also the rate of change of the curve under curve shortening flow, and write $k := |H|_g$ for the curvature.

Write x for the variable parameterising the loop $\gamma_t : S^1 \to M$, and write $X := \frac{\partial}{\partial x}\gamma_t$ for the spatial velocity vector for this loop, thus X is a scalar multiple of the tangent vector T. Here and in the sequel we abuse notation by identifying connections on the tangent bundle TM with connections on pullback bundles.

Exercise 3.7.5 (Commutativity of X and H)**.** Show that $\nabla_X H = \nabla_H X$. *Hint*: first show that $\nabla_H \nabla_X F = \mathrm{Hess}(F)(H, X) + dF(\nabla_H X)$ for any scalar function $F : M \to \mathbf{R}$, and similarly with the roles of H and X reversed. Now use the torsion-free nature of the Levi-Civita connection and duality.

We now record a variation formula for the squared speed $g(X, X)$.

---

[51] Indeed, by considering what curve-shortening flow does to small circles in Euclidean space, it is clear that one cannot hope to obtain uniform lifespan bounds without such a curvature bound.

[52] For instance, as one approaches the minimal value of $\tilde{W}_3(\xi, t)$, the curves may begin to develop cusps or folds (i.e. they cease to be immersed).

[53] Amusingly, this idea of attaching some tightly rolled up dimensions to space also appears in string theory, though I doubt that there is any connection here.

Lemma 3.7.5. *For fixed x, we have*

$$\frac{\partial}{\partial t} g(X,X) = -2\mathrm{Ric}(X,X) - 2k^2 g(X,X). \tag{3.149}$$

Proof. From the chain rule we have

$$\frac{\partial}{\partial t} g(X,X) = (\frac{\partial}{\partial t} g)(X,X) + 2g(\nabla_H X, X). \tag{3.150}$$

The first term on the right-hand side of (3.150) is $-2\mathrm{Ric}(X,X)$ by the Ricci flow equation. On the other hand, H is orthogonal to T (as T is a unit vector), and so $g(X,H) = 0$. From this and Exercise 3.7.5 we have

$$g(\nabla_H X, X) = g(\nabla_X H, X) = -g(H, \nabla_X X). \tag{3.151}$$

Writing $X = g(X,X)^{1/2} T$, and again using that H is orthogonal to T, we have

$$g(H, \nabla_X X) = g(X,X) g(H, \nabla_T T) = g(X,X) g(H,H). \tag{3.152}$$

Since $g(H,H) = k^2$, the claim follows. □

Corollary 3.7.6. *We have $[H,T] = (k^2 + \mathrm{Ric}(T,T))T$.*

Proof. We already know that $[H,X]$ vanishes. Expressing $X = g(X,X)^{1/2}T$ and using the previous lemma (writing $\frac{\partial}{\partial t} g(X,X)$ as $\nabla_H g(X,X)$), the corollary follows after a brief computation. □

We can now derive a heat equation for the curvature (vaguely reminiscent of a Bochner-type identity):

Lemma 3.7.7 (First variation of squared curvature)**.** *We have*

$$\frac{\partial}{\partial t} k^2 = \nabla_T \nabla_T (k^2) - 2g(\pi(\nabla_T H), \pi(\nabla_T H)) + 2k^4 + O(k^2) \tag{3.153}$$

where π is the projection to the orthogonal complement of X, and the implied constants in the $O()$ terms depend only on the Riemannian manifold (M,g(t)) (and in particular on bounds on the Riemann curvature tensor).

Proof. We write $k^2 = g(H,H)$. By the chain rule and the Ricci flow equation we have

$$\frac{\partial}{\partial t} g(H,H) = -2\mathrm{Ric}(H,H) + 2g(\nabla_H H, H). \tag{3.154}$$

The first term on the right-hand side is $O(k^2)$ which is acceptable. As for the second term, we expand

$$\nabla_H H = \nabla_H \nabla_T T = \nabla_T \nabla_H T + \nabla_{[H,T]} T + O(k). \tag{3.155}$$

The O(k) term gives a contribution of $O(k^2)$ to (3.154) which is acceptable. By Corollary 3.7.6, we have

$$\nabla_{[H,T]} T = (k^2 + O(1)) \nabla_T T = k^2 H + O(k) \tag{3.156}$$

which gives a contribution of $2k^4 + O(k^2)$ to (3.154). Finally, we deal with the top-order term $\nabla_T \nabla_H T$. We express $\nabla_H T = \nabla_T H + [H, T]$. Applying Corollary 3.7.6 (and the orthogonality of T and H), we have

$$g(\nabla_T [H, T], H) = (k^2 + O(1)) g(\nabla_T T, H) = k^4 + O(k^2) \tag{3.157}$$

whereas from the Leibniz rule we have

$$g(\nabla_T \nabla_T H, H) = \frac{1}{2} \nabla_T \nabla_T g(H, H) - g(\nabla_T H, \nabla_T H). \tag{3.158}$$

Since H is orthogonal to T, we have

$$g(\nabla_T H, T) = -g(H, \nabla_T T) = -g(H, H) = -k^2 \tag{3.159}$$

and so by Pythagoras

$$g(\nabla_T H, \nabla_T H) = g(\pi(\nabla_T H), \pi(\nabla_T H)) - k^4. \tag{3.160}$$

Substituting (3.160) into (3.7.3), and combining this with (3.157) to calculate the net contribution of $\nabla_T \nabla_H T$, we obtain (3.153) as desired. □

Corollary 3.7.8 (First variation of curvature)**.** *We have*

$$\frac{\partial}{\partial t} k \leq \nabla_T \nabla_T k + k^3 + O(k) \tag{3.161}$$

Proof. Expanding out (3.153) using the product rule and comparing with (3.161), we see that it suffices to show that

$$(\nabla_T \nabla_T k)^2 \leq g(\pi(\nabla_T H), \pi(\nabla_T H)). \tag{3.162}$$

But if we differentiate the identity $k^2 = g(H, H)$ along T, we obtain

$$k \nabla_T k = g(\nabla_T H, H) = g(\pi(\nabla_T H), H) \tag{3.163}$$

(since H is orthogonal to T) and the claim now follows from Cauchy-Schwarz. □

Note that by combining this corollary with the maximum principle (Corollary 3.4.3) we can get upper bounds[54] on k for short times based on upper bounds for k at time zero. Unfortunately, the nonlinear term k^3 on the right-hand side has an unfavourable sign and can generate finite time blowup.

The situation is much improved, however, for a special class of loops known as *ramps*. These curves take values not in an arbitrary manifold M, but in a product manifold $M \times S^1_\lambda$ (with the product Riemannian metric). The point here is that we have a vertical unit vector field U on this manifold (corresponding to infinitesimal rotation of the S^1_λ factor) which is completely parallel to the Levi-Civita connection: $\nabla_\alpha U = 0$. Define a ramp to be a curve $\gamma : S^1 \to M \times S^1_\lambda$ whose unit tangent vector T is always upward sloping in the sense that $g(T, U) > 0$ on all of γ (thus the ramp must "wrap around" the vertical fibre S^1_λ at least once, in order to return to its starting point). In particular, since γ is compact, we have a uniform lower bound $g(T, U) \geq c$ for some $c > 0$. Write $u := g(T, U)$ for the evolution of such a ramp under curve shortening flow, thus one can view u as a function of t and x. On the one hand, we have the trivial pointwise bound

$$|u| \leq 1 \tag{3.164}$$

from Cauchy-Schwarz. On the other hand, we have an evolution equation for u:

Proposition 3.7.9 (First variation of u)**.** *We have*

$$\frac{\partial}{\partial t} u = \nabla_T \nabla_T u + (k^2 + O(1))u \tag{3.165}$$

Proof. Differentiating $u = g(T, U)$ by the chain rule as before (using the fact that U is parallel to the connection, as well as the Ricci flow equation) we have

$$\frac{\partial}{\partial t} u = -2\mathrm{Ric}(T, U) + g(\nabla_H T, U). \tag{3.166}$$

[54] By using energy estimates, one can also control higher derivatives of k, obtaining the usual parabolic type estimates as a consequence; such estimates are important for the analysis here but we will omit them.

Since U is parallel to the connection, it is annihilated by any commutator $\mathrm{Riem}(X,Y) = [\nabla_X, \nabla_Y]$ and thus $\mathrm{Ric}(T,U) = 0$. Writing $\nabla_H T = [H,T] + \nabla_T H = [H,T] + \nabla_T \nabla_T T$ and using Corollary 3.7.6, we conclude

$$\frac{\partial}{\partial t} u = (k^2 + O(1))u + g(\nabla_T \nabla_T T, U). \tag{3.167}$$

Since U is parallel to the connection, we can write $g(\nabla_T \nabla_T T, U) = \nabla_T \nabla_T g(T,U)$, and the claim follows from the definition of u. □

As a particular corollary of (3.165), we have the inequality

$$\frac{\partial}{\partial t} u \geq \nabla_T \nabla_T u - O(|u|). \tag{3.168}$$

Using the maximum principle (Corollary 3.4.3), and the assumption that u is initially bounded away from zero, we conclude that u continues to be bounded away from zero for all time t for which the curve-shortening flow exists (though this bound can deteriorate exponentially fast in t). In particular, u is positive and the curve continues to be a ramp. Furthermore, by applying the quotient rule to (3.161) and (3.165), one obtains after some calculation the differential inequality

$$\frac{\partial}{\partial t} f \leq \nabla_T \nabla_T f + \frac{2\nabla_T u}{u} \nabla_T f + O(f) \tag{3.169}$$

for the quantity $f := k/u$. Applying the maximum principle again, and noting that f is initially bounded at time zero, we conclude that f is bounded for all time for which the solution exists (though again, the bound can deteriorate exponentially in t). Combining this with the trivial bound (3.164), we conclude that the curvature k is bounded for any period of time on which the solution exists, with the bound deteriorating exponentially in t. Combining this with the local existence theory (see [**AlGr1992**]), which asserts that the curve shortening flow can be continued whenever the curvature remains bounded, we conclude that curve shortening flows for ramps persist globally in time.

Of course, in our applications to Ricci flow, the curves γ_ω that we are applying curve shortening flow to are not ramps; they live in M rather than $M \times S^1_\lambda$. To address this, one has to embed M in $M \times S^1_\lambda$ for some small λ and approximate each γ_ω by a ramp that wraps

around $M \times S^1_\lambda$ exactly once. One then flows the ramps by curve shortening flow, and works with the minimal spanning areas $A(\gamma)$ of these evolved ramps (rather than working with the curve shortening flow applied directly to the original curves).

There are of course many technical obstacles to this strategy. One of them is that one needs to show that small changes in the ramp γ do not significantly affect the area $A(\gamma)$ of the minimal spanning disk. To achieve this, one needs to show that if two ramps γ_1, γ_2 are initially close in the sense that there is an annulus connecting them of small area, then they stay close (in the same sense) for any bounded period of time under curve shortening flow. This can be accomplished by using a first variation formula for area of minimal annuli which is similar to Corollary 3.7.4. There are several other technical difficulties of an analytical nature to resolve; see [**MoTi2007**, Chapter 19] for full details.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/04/18`.

3.8. Rescaling of Ricci flows and κ-noncollapsing

We now set aside our discussion of the finite time extinction results for Ricci flow with surgery (Theorem 3.3.13), and turn instead to the main portion of Perelman's argument, which is to establish the global existence result for Ricci flow with surgery (Theorem 3.3.9), as well as the discreteness of the surgery times (Theorem 3.3.12).

As mentioned in Section 3.2, *local* existence of the Ricci flow is a fairly standard application of nonlinear parabolic theory, once one uses de Turck's trick to transform Ricci flow into an explicitly parabolic equation. The trouble is, of course, that Ricci flow can and does develop singularities (indeed, we have just spent several sectionsshowing that singularities must inevitably develop when certain topological hypotheses (e.g. simple connectedness) or geometric hypotheses (e.g. positive scalar curvature) occur). In principle, one can use surgery to remove the most singular parts of the manifold at every singularity time and then restart the Ricci flow, but in order to do this one needs

some rather precise[55] control on the geometry and topology of these singular regions.

In order to analyse these singularities, Hamilton and then Perelman employed the standard nonlinear PDE technique[56] of "blowing up" the singularity using the scaling symmetry, and then exploiting as much "compactness" as is available in order to extract an "asymptotic profile" of that singularity from a sequence of such blowups, which had better properties than the original Ricci flow. A sufficiently good classification of all the possible asymptotic profiles will, in principle, lead to enough structural properties on general singularities to Ricci flow that one can see how to perform surgery in a manner which controls both the geometry and the topology.

However, in order to carry out this program it is necessary to obtain geometric control on the Ricci flow which does not deteriorate when one blows up the solution; in the jargon of nonlinear PDE, we need to obtain bounds on some quantity which is both *coercive* (it bounds the geometry) and either *critical* (it is essentially invariant under rescaling) or *subcritical* (it becomes more powerful when one blows up the solution) with respect to the scaling symmetry. The discovery of controlled quantities for Ricci flow which were simultaneously coercive and critical was Perelman's first major breakthrough in the subject (previously known controlled quantities were either supercritical or only partially coercive); it made it possible[57], at least *in principle*, to analyse general singularities of Ricci flow and thus to begin the surgery program discussed above. The mere existence of such a quantity does not by any means establish global existence of Ricci flow with surgery immediately, but it does give one a non-trivial starting point from which one can hope to make progress.

---

[55] In particular, there are some hypothetical bad singularity scenarios which cannot be easily removed by surgery, due to topological obstructions; a major difficulty in the Perelman program is to show that such scenarios in fact cannot occur in a Ricci flow.

[56] The PDE notion of a blowing up a solution around a singularity, by the way, is vaguely analogous to the algebraic geometry notion of *blowing up* a variety around a singularity, though the two notions are certainly not identical.

[57] In contrast, the main reason why questions such as Navier-Stokes global regularity are so difficult is that no controlled quantity which is both coercive and critical or subcritical is known; see Section 3.4 of *Structure and Randomness*.

To be a more precise, recall from Section 3.2 that the Ricci flow equation $\frac{d}{dt}g = -2\mathrm{Ric}$, in any spatial dimension d, has two basic symmetries (besides the geometric symmetry of diffeomorphism invariance); it has the obvious time-translation symmetry $g(t) \mapsto g(t - t_0)$ (keeping the manifold M fixed), but it also has the scaling symmetry

$$g(t) \mapsto \lambda^2 g(\frac{t}{\lambda^2}) \tag{3.170}$$

for any $\lambda > 0$ (again keeping M fixed as a topological manifold). When applied with $\lambda < 1$, this scaling shrinks all lengths on the manifold M by a factor λ (recall that the length $|v|_g$ of a tangent vector v is given by the *square root* of $g(v, v)$), and also speeds up the flow of time by a factor $1/\lambda^2$; conversely, when applied with $\lambda > 0$, the scaling expands all lengths by a factor λ, and slows down the flow of time by $1/\lambda$.

Suppose now that one has a Ricci flow $t \mapsto (M, g(t))$ which becomes singular at some time $T > 0$. To analyse the behaviour of the flow as one approaches the singular time T, one picks a sequence of times $t_n \to T^-$ approaching T from below, a sequence of marked points $x_n \in M(t_n) = M$ on the manifold, and a sequence of length scales $L_n > 0$ which go to zero as $n \to \infty$. One then considers the blown up Ricci flows $t \mapsto (M^{(n)}, g^{(n)}(t))$, where $M^{(n)}$ is equal to M as a topological manifold (with x_n as a marked point or "origin" O), and $g^{(n)}(t)$ is the flow of metrics given by the formula

$$g^{(n)}(t) := \frac{1}{L_n^2} g(t_n + L_n^2 t). \tag{3.171}$$

Thus the flow $t \mapsto (M^{(n)}, g^{(n)}(t))$ represents a renormalised flow in which the time t_n has been redesignated as the temporal origin 0, the point x_n has been redesignated as the spatial origin O, and the length scale L_n has been redesignated as the unit length scale (and the time scale L_n^2 has been redesignated as the unit time scale). Thus the behaviour of the rescaled flow $t \mapsto (M^{(n)}, g^{(n)}(t))$at unit scales of space and time around the spacetime origin (thus $t = O(1)$ and $x \in B(O, O(1))$) correspond to the behaviour of the original flow $t \mapsto (M, g(T))$ at spatial scale L_n and time scale L_n^2 around the spacetime point (t_n, x_n), thus $t = t_n + O(L_n^2)$ and $x \in B(x_n, O(L_n))$.

Because the original Ricci flow existed on the time interval $0 \leq t < T$, the rescaled Ricci flow will exist on the time interval $-\frac{t_n}{L_n^2} \leq t < \frac{T-t_n}{L_n^2}$. In particular, in the limit $n \to \infty$ (leaving aside for the moment the question of what "limit" means precisely here), these Ricci flows become increasingly *ancient*, in that they will have existed on the entire past time interval $-\infty < t \leq 0$ in the limit.

The strategy is now to show that these renormalised Ricci flows $t \mapsto (M^{(n)}, g^{(n)})$ (with the marked origin O) exhibit enough "compactness" that there exists a subsequence of such flows which converge to some asymptotic limiting profile $t \mapsto (M^{(\infty)}, g^{(\infty)})$ in some sense[58]. If the notion of convergence is strong enough, then we will be able to conclude that this limiting profile of Ricci flows is also a Ricci flow[59]. This limiting Ricci flow has better properties than the renormalised flows; for instance, while the renormalised flows are almost ancient, the limiting flow actually *is* an ancient solution. Also, while the Hamilton-Ivey pinching phenomenon from Section 3.4 suggests that the renormalised flows have mostly non-negative curvature, the limiting flow will have everywhere non-negative curvature (provided that the points (t_n, x_n) and scales L_n are chosen properly; we will return to this "point-picking" issue later in this chapter).

If one was able to classify all possible asymptotic profiles to Ricci flow, this would yield quite a bit of information on singularities to such flows, by the standard and general nonlinear PDE method of *compactness and contradiction*. This method, roughly speaking, runs as follows. Suppose we want to claim that whenever one is sufficiently close to a singularity, some scale-invariant property P eventually occurs[60]. To prove this, we argue by contradiction, assuming we can find a Ricci flow $t \mapsto (M, g(t))$ in which P fails on a sequence of points in spacetime that approach the singularity, and on some sequence of

[58] We will define the precise notion of convergence of such flows later, in Section ???, but *pointed Gromov-Hausdorff convergence* is a good first approximation of the convergence concept to keep in mind for now.

[59] Actually, due to the parabolic smoothing effects of Ricci flow, we will be able to automatically upgrade weak notions of convergence to strong ones, and so this step is in fact rather easy.

[60] In our specific application, P is roughly speaking going to assert that the geometry and topology of high-curvature regions can be classified as belonging to one of a short list of possible "canonical neighbourhood" types, all of which turn out to be amenable to surgery.

scales going to zero. We then rescale the flow to create a sequence of rescaled Ricci flows $t \mapsto (M^{(n)}, g^{(n)}(t))$ as discussed above, each of which exhibits failure of P at unit scales near the origin (here we use the hypothesis that P is scale-invariant). Now, we use compactness to find a subsequence of flows converging to an asymptotic profile $t \mapsto (M^{(\infty)}, g^{(\infty)}(t))$. If the convergence is strong enough, the asymptotic profile will also exhibit failure of P. But now one simply goes through the list of all possible profiles in one's classification and verifies that each of them obeys P; and one is done.

Unfortunately, just knowing that a Ricci flow is ancient and has everywhere non-negative curvature does not seem enough, by itself, to obtain a full classification of asymptotic profiles (though one can definitely say some non-trivial statements about ancient Ricci flows with non-negative curvature, most notably the *Li-Yau-Hamilton inequality*, which we will discuss in Section ???). To proceed further, one needs further control on asymptotic profiles $t \mapsto (M^{(\infty)}, g^{(\infty)}(t))$. The only reasonable way to obtain such control is to obtain control on the rescaled flows $t \mapsto (M^{(n)}, g^{(n)}(t))$ which is uniform in n. While some control of this sort can be established merely by choosing the points (t_n, x_n) and scales L_n in a clever manner, there is a limit as to what one can accomplish just by point-picking alone (especially if one is interested in establishing properties P that apply to quite general regions of spacetime and general scales, rather than specific, hand-picked regions and scales). To really get good control on the rescaled flows $t \mapsto (M^{(n)}, g^{(n)}(t))$, one needs to obtain control on the original flow $t \mapsto (M, g(t))$ which does not deteriorate when one passes from the original flow to the rescaled flow.

One can express what "does not deteriorate" means more precisely using the language of *dimensional analysis*, or more precisely using the concepts of subcriticality, criticality, and supercriticality from nonlinear PDE. Suppose we have some (non-negative) scalar[61] quantity $F(M, g(\cdot))$ that measures some aspect of a flow $t \mapsto (M, g(t))$. In many situations, this quantity has some specific *dimension* k, in the

[61] Dimensional analysis becomes trickier when considering tensor-valued quantities, though in practice one can use the magnitude of such quantities as a scalar-valued proxy for these tensor-valued objects; see [**Ta2**] for some further discussion.

sense that one has a scaling relationship

$$F(M, \lambda^2 g(\tfrac{\cdot}{\lambda^2})) = \lambda^k F(M, g(\cdot)) \tag{3.172}$$

that measures how that quantity changes under the rescaling (3.170). In dimensional analysis language, (3.172) asserts that F has the units length^k.

Assuming that F is also invariant under time translation (and under changes of spatial origin), (3.172) implies that

$$F(M^{(n)}, g^{(n)}(\cdot)) = L_n^{-k} F(M, g(\cdot)). \tag{3.173}$$

Thus, if F is *critical* or *dimensionless* (which means that $k = 0$) or *subcritical* (which means that $k < 0$), any upper bound on F for the original Ricci flow $t \mapsto (M, g(t))$ will imply uniform bounds on the rescaled flows $t \mapsto (M^{(n)}, g^{(n)}(t))$, and thus (assuming the convergence is strong enough, and F has some good continuity properties) on the asymptotic profile $t \mapsto (M^{(\infty)}, g^{(\infty)}(t))$. In the subcritical case, F should in fact now vanish in the limit. On the other hand, if F is *supercritical* (which means that $k > 0$) then no information about the asymptotic profile $t \mapsto (M^{(\infty)}, g^{(\infty)}(t))$ is obtained.

In order for control of $F(M^{(\infty)}, g^{(\infty)}(\cdot))$ to be truly useful, we would like the quantity F to be *coercive*. This term is not precisely defined (though it is somewhat analogous to the notion of a *proper map*), but coercivity basically means that upper bounds on $F(M, g(\cdot))$ translate to some upper bounds on various norms or similar quantities measuring the "size" of $(M, g(\cdot))$, and (hopefully) to then obtain useful bounds on the topology and geometry of $(M, g(\cdot))$.

Let us give some examples of various such quantities F for Ricci flow. We begin with some supercritical quantities:

(1) Any length-type quantity, e.g. the diameter $\mathrm{diam}(M)$ of the manifold, or the *injectivity radius*, has dimension 1 and is thus supercritical.

(2) The various widths $W_2(t), W_3(t), \tilde{W}_3(t)$ of 3-dimensional Ricci flows from the previous lectures, which were based on areas of minimal surfaces, have dimension 2 and are also supercritical. Thus the various bounds we have on these quantities

from Sections 3.5, 3.6, 3.7 do not directly tell us anything about asymptotic profiles.

(3) The volume $\int_M d\mu$ of 3-manifolds has dimension 3 and is thus also supercritical. Thus upper bounds on volume, such as Corollary 3.4.11, do not directly tell us anything about asymptotic profiles (though they are useful for other tasks, most notably for ensuring that surgery times are discrete, see Theorem 3.3.12).

As for subcritical quantities, one notable one is the minimal scalar curvature $R_{\min}$. One can check (cf. the dimensional analysis at the end of Section 3.1) that scalar curvature has dimension -2 and is thus subcritical. The quantity $F(M, g(\cdot)) := \sup_t \max(-R_{\min}, 0)$, that measures the maximal amount of negative scalar curvature present in a Ricci flow, is then bounded (by the maximum principle, see Proposition 3.4.10), and so by the previous discussion will vanish for asymptotic profiles; in other words, asymptotic profiles always have non-negative scalar curvature. Unfortunately, this quantity is only partially coercive; it prevents scalar curvature from becoming arbitrarily large and negative, but does not prevent scalar curvature from becoming arbitrarily large and positive[62]. So this quantity does say something non-trivial about asymptotic profiles, but is insufficient by itself to fully control such profiles.

In the next lecture we shall see that the least eigenvalue $\lambda_1(-4\Delta + R)$ of the modified Laplace-Beltrami operator, which can be viewed as an analytic analogue of the geometric quantity $R_{\min}$ related to Poincaré inequalities, also enjoys a monotonicity property (which is connected to a certain gradient flow interpretation of (modified) Ricci flow); like $R_{\min}$, the least eigenvalue has dimension -2 and is thus also subcritical, but again it is not fully coercive, as it only prevents scalar curvature from becoming too negative.

[62] Also, it is possible for other curvatures, such as Ricci and Riemann curvatures, to be large even while the scalar curvature is small or even zero.

So far we have not discussed any critical quantities[63]. One way to create critical quantities is to somehow combine subcritical and supercritical examples together. Here is one simple example, due to Hamilton[**Ha1999**]:

Exercise 3.8.1. Show that the quantity[64] $\max(-R_{\min}(t)V(t)^{2/d}, 0)$ is critical (scale-invariant) and monotone non-increasing in time under d-dimensional Ricci flow, where $V = \int_M d\mu(t)$ denotes the volume of $(M, g(t))$ at time t.

In the next few lectures, we will see two more advanced versions of critical controlled quantities of an analytic nature, the *Perelman entropy* (a scale-invariant version of the minimal eigenvalue $\lambda_1(-4\Delta + R)$, which is to log-Sobolev inequalities as the latter quantity is to Poincaré inequalities) and the *Perelman reduced volume* (which measures how heat-type kernels on Ricci flows compare against heat kernels on Euclidean space). These quantities were both introduced in [**Pe2002**]. The key feature of these new critical quantities, which distinguishes them from previously known examples, is that they are now *coercive*: they provide a crucial scale-invariant geometric control on a flow $t \mapsto (M, g(t))$, which is now known as *κ-noncollapsing*. This control, which describes a relationship between the supercritical quantities of length and volume and the subcritical quantities of curvature, will be discussed next.

3.8.1. Length, volume, curvature, and collapsing. Let p be a point in a d-dimensional complete Riemannian manifold (M, g) (we make no assumptions on the dimension d here). We will establish[65] here some basic results in *comparison geometry*, which seeks to understand the relationship between the Riemann curvature Riem of the manifold M, and various geometric quantities of M such as the

[63] One can create some trivial examples of critical quantities, such as the dimension $\dim(M)$ or topological quantities such as $\pi_1(M)$, but these are not obviously coercive (the topological coercivity of the latter quantity being, of course, precisely the Poincaré conjecture that we are trying to prove!).

[64] This quantity can be used, for instance, to show that Ricci flow admits no "breather" solutions, i.e. non-constant periodic solutions; see the discussion in [**Pe2002**]. Unfortunately, as with previous examples, it is not fully coercive.

[65] This is only a brief introduction; see e.g. [**Pe2006**, Chapters 6, 9, 10] for a detiled treatment.

volume of balls and the injectivity radius, especially when compared against model geometries such as the sphere and hyperbolic space.

Of course, in the case of Euclidean space $\mathbf{R}^d$ with the Euclidean metric, the Riemann curvature is identically zero, and the volume of $B(p,r)$ is $c_d r^d$ for some explicit constant $c_d := \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)} > 0$ depending only on dimension. For Riemannian manifolds, it is easy to see that the volume of $B(p,r)$ is $(1+o(1))c_d r^d$ in the limit $r \to 0$; for more precise asymptotics, see Exercises 3.8.7, 3.8.8 below. One of the most effective tools to study these questions comes from *normal coordinates*, or more precisely from the *exponential map* $\exp_p : T_pM \to M$ from the tangent space T_pM to M, defined by setting $\exp_p(v)$ to be the value of $\gamma(1)$, where $\gamma : [0,1] \to M$ is the unique constant-speed geodesic with $\gamma(0) = p$ and $\gamma'(0) = v$. By the *Hopf-Rinow theorem*, M is complete (in the metric sense) if and only if the exponential map is defined on all of T_pM. Henceforth we will always assume M to be complete. The ball $B(p,r)$ of radius $r > 0$ in M centred at p is then the image under the exponential map of the ball $B_{T_pM}(0,r)$ of the tangent space of the same radius (using the metric $g(p)$, of course):

$$B(p,r) = \exp_p(B_{T_pM}(0,r)). \tag{3.174}$$

Thus we can study the balls centred at p by using the exponential map to pull back to the tangent space T_pM and analysing the geometry there. Two radii become relevant for this approach:

(1) The *injectivity radius* at p is the supremum of all radii r such that $\exp_p$ is injective on $B_{T_pM}(0,r)$.

(2) The *conjugate radius* at p is the supremum of all radii r such that $\exp_p$ is an *immersion* on $B_{T_pM}(0,r)$ (i.e. its gradient has full rank at every point in $B_{T_pM}(0,r)$).

In many situations, these two radii are equal, but there are cases in which the injectivity radius is smaller. In fact the injectivity radius is always less than or equal to the conjugate radius; see Exercise 3.8.4 below.

Example 3.8.1 (Sphere)**.** Let $K > 0$, and let $M = \frac{1}{\sqrt{K}} S^d := \{(x_1, \ldots, x_{d+1}) \in \mathbf{R}^{d+1} : x_1^2 + \ldots + x_{d+1}^2 = 1/K\}$ be the sphere of radius $1/\sqrt{K}$, with the metric induced from the metric $ds^2 =$

$dx_1^2 + \ldots + dx_{d+1}^2$ of Euclidean space $\mathbf{R}^{d+1}$. Then at every point p of M, the injectivity radius and conjugate radius are both equal to $\pi/\sqrt{K}$, which is also the diameter of the manifold. Note also that this manifold has constant sectional curvature K.

Example 3.8.2 (Hyperbolic space)**.** Let $K > 0$, and let $M = \frac{1}{\sqrt{K}} H^d := \{(t, x_1, \ldots, x_d) \in \mathbf{R}^{1+d} : x_1^2 + \ldots x_d - t^2 = 1/K; t > 0\} \subset \mathbf{R}^{1+d}$ be hyperbolic space of hyperbolic radius $1/\sqrt{K}$, with the metric induced from the metric $ds^2 = dx_1^2 + \ldots + dx_d^2 - dt^2$ of Minkowski space. Then at any point p in M, e.g. $p = (1, 0)$, the injectivity radius, conjugate radius, and diameter are infinite. This manifold has constant sectional curvature $-K$.

Example 3.8.3 (Torus)**.** Let $r > 0$, and let $M = (\mathbf{R}/r\mathbf{Z})^d$ be the d-torus which is the product of d circles of length r. Then for any point p in M, the injectivity radius is $r/2$ and the conjugate radius is infinite. Here the sectional curvature is of course 0 everywhere.

The metric g on M induces a pullback metric on T_pM, which by abuse of notation we shall also call g. This metric can degenerate once one passes the conjugate radius, but let us ignore this issue for the time being. On T_pM, we have the radial variable r (defined as the magnitude of a tangent vector with respect to $g(p)$), and the radial vector field ∂_r (defined as the dual vector field to r using polar coordinates), which is smooth away from the origin.

In Euclidean space, the vector field ∂_r is the gradient of r. Happily, the same fact is true for more general Riemannian manifolds:

Lemma 3.8.4 (Gauss lemma)**.** (1) *Away from the origin, we have* $|\partial_r|_g = 1$ *and* $\nabla_{\partial_r} \partial_r = 0$.

(2) *Away from the origin,* ∂_r *is the gradient* $\mathrm{grad} r$ *of* r *with respect to the metric* g*, thus* $(\partial_r)^\alpha = \nabla^\alpha r$.

Exercise 3.8.2. Prove Lemma 3.8.4. *Hint*: part 1 follows from the geodesic flow equation $\nabla_{\dot\gamma} \dot\gamma = 0$. For part 2, one way to proceed is to establish the ODE

$$\nabla_{\partial_r}(\partial_r - \mathrm{grad} r)^\alpha = (\nabla^\alpha (\partial_r)_\beta)(\partial_r - \mathrm{grad} r)^\beta \tag{3.175}$$

and then apply Gronwall's inequality.

Lemma 3.8.4 gives some important relationships between the radial vector field ∂_r and the Hessian $\mathrm{Hess}(r)_{\alpha\beta} := \nabla_\alpha \nabla_\beta r = \nabla_\alpha (\partial_r)_\beta$ (which can be viewed as the second fundamental form of the spheres centred at p):

Exercise 3.8.3. Away from the origin, obtain the deformation formula

$$\mathcal{L}_{\partial_r} g = 2\mathrm{Hess}(r) \tag{3.176}$$

and the *Riccati-type equation*

$$\nabla_{\partial_r} \mathrm{Hess}_{\alpha\beta} + \mathrm{Hess}_{\alpha\beta}\mathrm{Hess}^\beta_\gamma = \mathrm{Riem}^\delta_{\alpha\gamma\beta}(\partial_r)^\gamma (\partial_r)_\delta. \tag{3.177}$$

Also, show that $\mathrm{Hess}_{\alpha\beta}$ has ∂_r as a null eigenvector.

Exercise 3.8.4. Show that the injectivity radius r_i of a point p cannot exceed the conjugacy radius r_c. *Hint*: there are several ways to establish this. Here is one: suppose for contradiction that $r_i > r_c$, thus $r_i > (1+\varepsilon) r_c$ for some small $\varepsilon > 0$. Let $v \in T_p M$ be a vector of magnitude at most r_c. Observe that the function $d(p,x) + d(\exp_p((1+\varepsilon)v), x)$ achieves a global minimum at $\exp_p(v)$ whenever and so has non-negative Hessian. Use this to obtain a lower bound on $\mathrm{Hess}(r)$ on $B(p, r_c)$, and combine this with Exercise 3.8.3 to show that the exponential map is in fact immersed on a neighbourhood of $B(p, r_c)$, a contradiction. Another approach is based on Klingenberg's inequality (see Lemma 3.8.11 below), while a third approach is based on the second variation formula for the energy of a geodesic.

Let us now impose the bound that all sectional curvatures are bounded by some $K > 0$ on a ball $B(p, r_0)$, thus

$$|g(\mathrm{Riem}(X,Y)X,Y)| \leq K \tag{3.178}$$

for all orthonormal tangent vectors X, Y at any point in $B(p, r_0)$. From Example 3.8.1 we know that the exponential map can become singular past the radius $\pi/\sqrt{K}$, so let us also assume that

$$r_0 \leq \pi/\sqrt{K}. \tag{3.179}$$

Note that the sectional curvature bound also implies a Ricci curvature bound $|\mathrm{Ric}(X,X)| \leq (d-1)K$ for all unit tangent vectors based in $B(p, r_0)$.

From (3.178) and (3.179) we see that $|\mathrm{Riem}|_g = O_d(r_0^{-2})$ on the ball $B(p, r_0)$. When this latter property occurs, let us informally say that M has *bounded normalised curvature* at scale r_0 at p. Our analysis here can thus be interpreted as a study of the volume of balls (and of related quantities, such as the injectivity radius) under assumptions of bounded normalised curvature.

Remark 3.8.5. If one wishes, one can rescale to normalise K (or r_0) to equal 1, although this does not significantly simplify the computations that follow below.

Using (3.177), one can obtain sharp upper and lower bounds for $\mathrm{Hess}(r)$:

Exercise 3.8.5 (Comparison estimates for $\mathrm{Hess}(r)$)**.** Assume that (3.178) and (3.179) hold. At any non-zero point in $B_{T_pM}(0, r_0)$, let $\lambda_{\min} \leq \lambda_{\max}$ be the least and greatest eigenvalues of $\mathrm{Hess}(r)$ on the orthogonal complement of ∂_r. Use (3.177) to establish the differential inequalities

$$\nabla_{\partial_r} \lambda_{\max} + \lambda_{\max}^2 \leq K \tag{3.180}$$

and

$$\nabla_{\partial_r} \lambda_{\min} + \lambda_{\min}^2 \geq -K \tag{3.181}$$

for $0 < r < r_0$ and also establish the infinitesimal bound

$$\lambda_{\min}, \lambda_{\max} = \frac{1}{r} + O(r) \tag{3.182}$$

for all sufficiently small positive r. From (3.180), (3.181), (3.182), conclude the bounds

$$\sqrt{K} \coth(\sqrt{K} r) \leq \lambda_{\min} \leq \lambda_{\max} \leq \sqrt{K} \cot(\sqrt{K} r) \tag{3.183}$$

and in particular that

$$(d-1)\sqrt{K} \coth(\sqrt{K} r) \leq \Delta r \leq (d-1)\sqrt{K} \cot(\sqrt{K} r). \tag{3.184}$$

Using (3.176) and (3.183), deduce the bound

$$dr^2 + \frac{\sin^2(\sqrt{K} r)}{K} d\theta^2 \leq dg^2 \leq dr^2 + \frac{\sinh^2(\sqrt{K} r)}{K} d\theta^2 \tag{3.185}$$

where (r, θ) are the usual Euclidean polar coordinates on T_pM, thus the Euclidean metric (induced by $g(p)$) is given by $ds^2 = dr^2 + r^2 d\theta^2$).

Remark 3.8.6. Each of the above bounds are attained by either the sphere of constant sectional curvature $+K$ (Example 3.8.1) or the hyperbolic space of constant sectional curvature $-K$ (Example 3.8.2). More generally, one should think of these two examples as the two extreme geometries obeying the assumption (3.178). In the limit $K = 0$ one recovers the formulae for Euclidean space $\mathbf{R}^d$ or for the torus (Example 3.8.3).

Exercise 3.8.6 (Bounded curvature implies lower bound on conjugacy radius)**.** Using Exercise 3.8.3, show that if (3.178) and (3.179) hold, then the conjugacy radius of p is at least r_0.

Remark 3.8.7. A converse of sorts to Exercise 3.8.6 is provided by *Myers' theorem* (Exercise 3.10.2), which asserts that if $\mathrm{Ric} \geq (d-1)K$, then the diameter of M is at most $\pi/\sqrt{K}$. Another result in a somewhat similar spirit is the *1/4-pinched sphere theorem*. The Ricatti-type equations and inequalities developed above play a key role in the proof of such theorems.

Now we relate the Hessian of r to the volume metric $d\mu$ and the Laplacian Δr:

Exercise 3.8.7. Away from the origin, obtain the deformation formula

$$\mathcal{L}_{\partial_r} d\mu = (\Delta r) d\mu \tag{3.186}$$

and the Riccati-type inequality

$$\nabla_{\partial_r} \Delta r + \frac{1}{d-1}(\Delta r)^2 \leq \nabla_{\partial_r} \Delta r + |\mathrm{Hess}(r)|^2 = -\mathrm{Ric}(\partial_r, \partial_r). \tag{3.187}$$

Exercise 3.8.8 (Absolute volume comparison)**.** Assume (3.178) and (3.179). Using Exercises 5 and 7, show that the volume of $B_{T_pM}(0, r_0)$ is maximised in the case of hyperbolic space (Example 3.8.2) and minimised in the case of the sphere (Example 3.8.1). In particular, if $r_0 \ll \sqrt{K}$, conclude that the volume of $B_{T_pM}(0, r_0)$ is comparable to r_0^d, with the comparability constants depending only on d and on the implied constant in the $O()$ notation.

Remark 3.8.8. In later sections we will need a relative variant of this comparison inequality, known as the *Bishop-Gromov comparison inequality* (Lemma 3.10.1), which will assert that certain ratios between volumes of balls are monotone in the radius r_0.

Exercise 3.8.9. Show that the volume of $B(p,r)$ is $(c_d - \frac{R(p)}{6(d+2)}r^2 + O(r^4))r^d$ for sufficiently small r, where $R(p)$ is the scalar curvature at p. Thus we see that scalar curvature distorts the infinitesimal volume growth of balls. Develop a similar interpretation of the Ricci curvature $\mathrm{Ric}(p)(v,v)$ as the volume distortion of infinitesimal sectors with apex p and direction v.

If r_0 is less than the injectivity radius, we see from (3.174) that $B(p,r_0)$ has the same volume as $B_{T_pM}(0,r_0)$. From Exercise 3.8.8, we thus conclude that

$$\mathrm{Vol}(B(p,r_0)) \sim_d r_0^d \tag{3.188}$$

whenever (3.178) holds, and r_0 is less than both $O(1/\sqrt{K})$ and the injectivity radius of p.

What happens if r_0 exceeds the injectivity radius? We still obtain the upper bound in (3.187), but can lose the lower bound, as can already be seen by considering the torus example (Example 3.8.3) with the injectivity radius r small. Thus we see that failure of injectivity can lead to collapse in the volume of balls.

A deep result of Cheeger[**Ch1970**] shows that in fact injectivity failure *always* collapses the volume of balls (assuming bounded normalised curvature), or equivalently that non-collapsing of volume is equivalent to a lower bound on the injectivity radius:

Theorem 3.8.9 (Cheeger's lemma)**.** *Suppose that* $|\mathrm{Riem}|_g \leq Cr_0^{-2}$ *on* $B(p,r_0)$ *and that* $\mathrm{Vol}(B(p,r_0)) \geq \delta r_0^d$ *for some* $\delta > 0$*. Then the injectivity radius of* p *is at least* $c(C,\delta,d)r_0$ *for some* $c(C,\delta,d) > 0$ *depending only on* C,δ,d*.*

Remark 3.8.10. This lemma is closely related to the *Cheeger finiteness theorem*, which asserts that the number of possible topologies for the ball $B(p,r_0)$ under the assumptions of Theorem 3.8.9 is finite, as well as *Gromov's compactness theorem*, which essentially asserts that the metrics on these balls form a compact set in a certain topology.

We will not discuss the proof of Cheeger's lemma here. Cheeger's original proof relies on the following inequality which is also of interest:

Lemma 3.8.11 (Klingenberg's inequality)**.** *Assume the conjugacy radius is at least r_0. Then exactly one of the following holds:*

(1) *The injectivity radius of p is at least r_0.*

(2) *There exists a non-trivial geodesic starting and ending at p of length less than $2r_0$.*

Proof. (Sketch) It is clear that 1. and 2. cannot both be true. Now suppose that the injectivity radius r is strictly less than r_0, then there exist two distinct geodesic rays γ_1, γ_2 from p to another point q, one of length r and the other of length at most r. On the other hand, by hypothesis the exponential map is an immersion on $B(x, r_0)$. From the inverse function theorem (and Lemma 3.8.4) we can then perturb the rays γ_1, γ_2 from p to have lengths slightly less than r but still ending up at the same point (thus contradicting the definition of r), unless γ_1, γ_2 have length exactly r and have equal and opposite tangent vectors at q. But then we have formed a geodesic path from p to p of length 2r, and the claim follows. □

Exercise 3.8.10. Show that the injectivity radius of p is equal to the minimum of the conjugacy radius of p, and half the length of the shortest non-trivial geodesic path from p to itself (or $+\infty$ if no such path exists).

Exercise 3.8.11. Let M be a compact manifold whose sectional curvatures are all bounded in magnitude by K. Show that if the injectivity radius r of M (defined as the infimum of the injectivity radii of every point p in M) is less than $\pi/\sqrt{K}$, then there exists a closed geodesic loop of length exactly $2r$.

Let us informally say that a Riemannian manifold M is *non-collapsed* at scale r_0 at a point p if $B(p, r_0)$ has volume $\gtrsim_d r_0^d$. The above discussion then says that, under the assumption of bounded normalised curvature at scale r_0 at p, that non-collapsing is equivalent to a lower bound of $\gtrsim_d r_0$ on the injectivity radius, which is in turn equivalent to a lower bound of $\gtrsim_d r_0$ on the length of any non-trivial

geodesic paths from p to itself. Thus we see that the non-collapsing property is quite coercive; it implies some important control on the local geometry of the Riemannian manifold.

Example 3.8.12. The sphere (Example 3.8.1) of dimension 2 and higher and hyperbolic space (Example 3.8.2) are non-collapsed at every point and scale for which one has bounded normalised curvature. (For the sphere, volume collapses at scales bigger than the diameter of the sphere, but one no longer has bounded normalised curvature in this regime.) Similarly for Euclidean space, or for products of any of these three examples. On the other hand, the torus (Example 3.8.3) (or the sphere of dimension 1) is collapsed at large scales even though one still retains normalised bounded curvature. Similarly for the cylinder $S^1 \times \mathbf{R}$.

Now we adapt this concept to Ricci flows. The following definition is fundamental to Perelman's arguments:

Definition 3.8.13 (κ-collapsing)**.** Let $t \mapsto (M, g(t))$ be a d-dimensional Ricci flow, and let $\kappa > 0$. We say that the Ricci flow is *κ-collapsed* at a point (t_0, x_0) in spacetime at scale r_0 if the following statements hold:

(1) (Bounded normalised curvature) We have $|\mathrm{Riem}(t,x)|_g \leq r_0^{-2}$ for all (t,x) the spacetime cylinder $[t_0 - r_0^2, t_0] \times B_{g(t_0)}(x_0, r_0)$ (in particular, we assume that the lifespan of the Ricci flow includes the time interval $[t_0 - r_0^2, t_0]$);

(2) (Collapsed volume) At time t_0, the ball $B_{g(t_0)}(x_0, r_0)$ has volume at most κr_0^d.

Otherwise, we say that the Ricci flow is *κ-noncollapsed* at this point and scale.

Remark 3.8.14. One should view κ here as a small dimensionless quantity, in order to make the notion of κ-noncollapsing scale-invariant.

It turns out that Perelman's critical quantities are controlled enough, and coercive enough, to establish κ-noncollapsing at non-zero times assuming some noncollapsing at time zero. There are

many ways to formulate this important non-collapsing result; here is one typical phrasing.

Theorem 3.8.15 (Perelman's non-collapsing theorem, first version)**.** *Let $t \mapsto (M, g(t))$ be a Ricci flow on compact 3-manifolds on a time interval $[0, T_0]$ such that at time zero, we have the normalised non-collapsing hypotheses $|\mathrm{Riem}(p)|_g \leq 1$ and $\mathrm{Vol}(B(p,1)) \geq \omega$ for all $p \in M$, where $\omega > 0$ is fixed. Then the Ricci flow is κ-noncollapsed for all $(t_0, x_0) \in [0, T_0] \times M$ and all scales $0 < r_0 < \sqrt{t_0}$, where $\kappa > 0$ depends only on ω and T_0.*

Note that the conclusion here is scale-invariant and will therefore persist to asymptotic profiles $(M^{(\infty)}, g^{(\infty)})$ as discussed in the beginning of this section.

Remark 3.8.16. Actually, to establish the global existence results for Ricci flow with surgery, we will need to extend Definition 3.8.13 and Theorem 3.8.15 to Ricci flows with surgery; we shall return to this point later in this chapter.

Remark 3.8.17. This non-collapsing theorem in fact holds in all dimensions, not just 3, but of course many other aspects of our analysis will only work in three dimensions.

The next few sections will be devoted to the proof of Theorem 3.8.15, and then we will discuss how Theorem 3.8.15 can be used to analyse asymptotic profiles near a Ricci flow singularity.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/04/20`. Thanks to Pedro Lauridsen Ribiero and Dan for corrections.

3.9. Ricci flow as a gradient flow, log-Sobolev inequalities, and Perelman entropy

It is well known that the *heat equation*

$$\dot{f} = \Delta f \tag{3.189}$$

on a compact Riemannian manifold (M, g) (with metric g static, i.e. independent of time), where $f : [0, T] \times M \to \mathbf{R}$ is a scalar field, can

be interpreted as the gradient flow for the *Dirichlet energy functional*

$$E(f) := \frac{1}{2}\int_M |\nabla f|_g^2 \, d\mu \tag{3.190}$$

using the inner product $\langle f_1, f_2\rangle_\mu := \int_M f_1 f_2 \, d\mu$ associated to the volume measure $d\mu$. Indeed, if we evolve f in time at some arbitrary rate $\dot{f}$, a simple application of integration by parts (equation (3.65)) gives

$$\frac{d}{dt}E(f) = -\int_M (\Delta f)\dot{f} \, d\mu = \langle -\Delta f, \dot{f}\rangle_\mu \tag{3.191}$$

from which we see that (3.189) is indeed the gradient flow for (3.191) with respect to the inner product. In particular, if f solves the heat equation (3.189), we see that the Dirichlet energy is decreasing in time:

$$\frac{d}{dt}E(f) = -\int_M |\Delta f|^2 \, d\mu. \tag{3.192}$$

Thus we see that by representing the PDE (3.189) as a gradient flow, we automatically gain a controlled quantity of the evolution, namely the energy functional that is generating the gradient flow. This representation also strongly suggests (though does not quite prove) that solutions of (3.189) should eventually converge to stationary points of the Dirichlet energy (3.190), which by (3.191) are just the harmonic functions (i.e. the functions f with $\Delta f = 0$).

As one very quick application of the gradient flow interpretation, we can assert that the only periodic (or "breather") solutions to the heat equation (3.189) are the harmonic functions (which, in fact, must be constant if M is compact, thanks to the maximum principle). Indeed, if a solution f was periodic, then the monotone functional E must be constant, which by (3.192) implies that f is harmonic as claimed.

It would therefore be desirable to represent Ricci flow as a gradient flow also, in order to gain a new controlled quantity, and also to gain some hints as to what the asymptotic behaviour of Ricci flows should be. It turns out that one cannot quite do this directly (there is an obstruction caused by *gradient steady solitons*, of which we shall say more later); but Perelman nevertheless observed that one *can*

interpret Ricci flow as gradient flow if one first quotients out the diffeomorphism invariance of the flow. In fact, there are infinitely many such gradient flow interpretations available. This fact already allows one to rule out "breather" solutions to Ricci flow, and also reveals some information about how *Poincaré's inequality* deforms under this flow.

The energy functionals associated to the above interpretations are subcritical (in fact, they are much like $R_{\min}$) but they are not coercive; Poincaré's inequality holds both in collapsed and non-collapsed geometries, and so these functionals are not excluding the former. However, Perelman discovered a perturbation of these functionals associated to a deeper inequality, the *log-Sobolev inequality* (first introduced by Gross[**Gr1975**] in Euclidean space). This inequality *is* sensitive to volume collapsing at a given scale. Furthermore, by optimising over the scale parameter, the controlled quantity (now known as the *Perelman entropy*) becomes scale-invariant and prevents collapsing at any scale - precisely what is needed to carry out the first phase of the strategy outlined in Section 3.8 to establish global existence of Ricci flow with surgery.

The material here is loosely based on [**Pe2002**], [**KlLo2006**], and [**Mu2006**].

3.9.1. Ricci flow as gradient flow. We would like to represent Ricci flow

$$\dot{g} = -2\mathrm{Ric} \tag{3.193}$$

as a gradient flow of some functional (with respect to some inner product, or at least with respect to some Riemannian metric on the space of all metrics g). We will assume that all quantities are smooth and that the manifold is either compact or that all expressions being integrated are rapidly decreasing at infinity (so no boundary terms etc. arise from integration by parts).

To do this, our starting point will be the first variation formula (3.50) for the scalar curvature R for an arbitrary instantaneous deformation $\dot{g}$ of the metric g:

$$\dot{R} = -\mathrm{Ric}^{\alpha\beta}\dot{g}_{\alpha\beta} - \Delta\mathrm{tr}(\dot{g}) + \nabla^\alpha\nabla^\beta\dot{g}_{\alpha\beta}. \tag{3.194}$$

We can integrate in M to eliminate the latter two terms on the right-hand side (by Stokes theorem (3.64)) to get

$$\int_M \dot{R}\ d\mu = -\int_M \mathrm{Ric}^{\alpha\beta}\dot{g}_{\alpha\beta}\ d\mu. \tag{3.195}$$

This looks rather promising; it suggests that if we introduce the *Einstein-Hilbert functional*

$$H(M,g) := \int_M R\ d\mu \tag{3.196}$$

then the Ricci flow (3.193) might be interpretable as a gradient flow for $-2H$.

Unfortunately, there is a problem because R is not the only time-dependent quantity in the right-hand side of (3.196); the volume measure $d\mu$ also evolves in time by the formula

$$\frac{d}{dt}d\mu = \frac{1}{2}\mathrm{tr}(\dot{g})\ d\mu \tag{3.197}$$

(see (3.55)). Thus, from the product rule, the true variation of the Einstein-Hilbert functional is given by the formula

$$\frac{d}{dt}H(M,g) = \int_M (-\mathrm{Ric}^{\alpha\beta} + \frac{1}{2}Rg^{\alpha\beta})\dot{g}_{\alpha\beta}\ d\mu. \tag{3.198}$$

So the gradient flow of -2H (using the inner product associated to $d\mu$) is not Ricci flow, but is instead a rather strange flow

$$\dot{g} = -2\mathrm{Ric} + Rg = -2G \tag{3.199}$$

where $G := \mathrm{Ric} - \frac{1}{2}R$ is the *Einstein tensor*. This flow does not have any particularly nice properties in general (it is not parabolic in three and higher dimensions, even after applying the de Turck trick from Section 3.2). On the other hand, in two dimensions the right-hand side of (3.198) vanishes and $H(M,g)$ becomes invariant under deformations (we have already exploited this fact to prove Proposition 3.5.2). More generally, we recover see from (3.198) the fact (well known in general relativity) that the (formal) stationary points of the Einstein-Hilbert functional are precisely the solutions of the *vacuum Einstein equations* $G = 0$ (or equivalently, $\mathrm{Ric} = 0$ in any dimension other than 2).

We see that the variation of the measure $d\mu$ in time is causing us some difficulty. To fix this problem, let us take the (rather non-geometric looking) step of replacing this evolving measure $d\mu$ by some static measure dm which we select in advance, and consider instead the variation of the functional $\int_M R\,dm$ with respect to some arbitrary perturbation $\dot{g}$. Now that m is static, we can apply (3.194) to get

$$\frac{d}{dt}\int_M R\,dm = \int_M (-\mathrm{Ric}^{\alpha\beta}\dot{g}_{\alpha\beta} - \Delta \mathrm{tr}(\dot{g}) + \nabla^\alpha\nabla^\beta\dot{g}_{\alpha\beta})\,dm. \tag{3.200}$$

Previously, we used Stokes' theorem to eliminate the latter two terms on the right-hand side to leave us with the one term $\int_M \mathrm{Ric}^{\alpha\beta}\dot{g}_{\alpha\beta}\,dm$ that we do want. Unfortunately, Stokes' theorem only applies for the volume measure $d\mu$, not for our static measure dm! In order to apply Stokes' theorem, we must therefore convert the static measure back to volume measure. The Radon-Nikodym derivative $\frac{d\mu}{dm}$ of the two measures should be some positive function, which we shall denote by e^f for some scalar (and time-varying) function $f : M \to \mathbf{R}$, thus

$$dm = e^{-f}d\mu. \tag{3.201}$$

Inserting (3.201) into (3.200), integrating by parts using the volume measure $d\mu$, and then using (3.201) again to convert back to the static measure dm, we see after a little calculation that

$$\int_M \Delta \mathrm{tr}(\dot{g})\,dm = \int_M (|\nabla f|_g^2 - \Delta f)\mathrm{tr}(\dot{g})\,dm \tag{3.202}$$

and similarly

$$\int_M \nabla^\alpha\nabla^\beta\dot{g}_{\alpha\beta}\,dm = \int_M ((\nabla^\alpha f)(\nabla^\beta g) - \nabla^\alpha\nabla^\beta f)\dot{g}_{\alpha\beta}\,dm \tag{3.203}$$

and so we can express the right-hand side of (3.200) as

$$\langle -\mathrm{Ric}^{\alpha\beta} - (|\nabla f|_g^2 - \Delta f)g^{\alpha\beta} + (\nabla^\alpha f)(\nabla^\beta f) - \nabla^\alpha\nabla^\beta f, \dot{g}_{\alpha\beta}\rangle_m. \tag{3.204}$$

This looks rather unpleasant; we managed to eradicate the scalar curvature term $\frac{1}{2}R$ that was present in the variation in (3.198), but at the cost of introducing four new terms involving f. But to deal with this, first observe from differentiating (3.201) and using (3.197) and the static nature of dm that we know the first variation of f:

$$\dot{f} = \frac{1}{2}\mathrm{tr}(\dot{g}). \tag{3.205}$$

So the term $\langle \Delta f g^{\alpha\beta}, \dot{g}_{\alpha\beta} \rangle_m$ that appears in (3.204) can be rewritten as $2\int_M (\Delta f)\dot{f}\ dm$. Now this term looks familiar... in fact, it essentially the variation (3.191) of the Dirichlet energy functional for the measure dm! This suggests that we may be able to simplify (3.204) if we modify our functional $\int_M R\ dm$ by adding some multiple of the Dirichlet functional $E := \frac{1}{2}\int_M |\nabla f|_g^2\ dm$.

One cannot apply (3.191) directly, though, because (a) g is evolving in time, rather than static, and also (b) dm is not the volume measure for g. But we have all the equations to deal with this, and one can compute the first variation of E:

Exercise 3.9.1. Show that

$$\frac{d}{dt}E = -\frac{1}{2}\langle \Delta f g^{\alpha\beta} - |\nabla f|_g^2 g^{\alpha\beta} + (\nabla f)^\alpha (\nabla f)^\beta, \dot{g}_{\alpha\beta}\rangle_m. \tag{3.206}$$

Hint: expand out $|\nabla f|_g^2 = g^{\alpha\beta}(\nabla_\alpha f)(\nabla_\beta f)$ and use (3.37).

If we thus define the functional

$$\mathcal{F}_m(M,g) := \int_M (R + |\nabla f|^2)\ dm \tag{3.207}$$

we see from (3.204), (3.206) that we get a lot of cancellation, ending up with

$$\frac{d}{dt}\mathcal{F}_m(M,g) = -\langle \mathrm{Ric}^{\alpha\beta} + \nabla^\alpha \nabla^\beta f, \dot{g}_{\alpha\beta}\rangle_m. \tag{3.208}$$

Thus the gradient flow of $-2\mathcal{F}_m(M,g)$ with respect to the inner product $\langle h, k\rangle_m := \int_M h^{\alpha\beta}k_{\alpha\beta}\ dm$ on symmetric two-forms (or more precisely, on the tangent space of such forms at g) is given by

$$\dot{g}_{\alpha\beta} = -2\mathrm{Ric}_{\alpha\beta} - 2\nabla_\alpha \nabla_\beta f. \tag{3.209}$$

From (3.205) we see that f now evolves by a backward heat equation

$$\dot{f} = -\Delta f - R. \tag{3.210}$$

With this flow, we see that $\mathcal{F}_m$ is monotone increasing, with

$$\frac{d}{dt}\mathcal{F}_m = 2\int_M |\mathrm{Ric} + \mathrm{Hess}(f)|^2\ dm. \tag{3.211}$$

The equation (3.209) is *almost* Ricci flow (3.193), but with one additional term associated with f. But we can observe (using (3.61)) that $2\nabla_\alpha \nabla_\beta f = \mathcal{L}_{\nabla f} g_{\alpha\beta}$ is just the Lie derivative of g in the direction of

the gradient vector field $\nabla^\gamma f$. Thus we see that (3.211) is a modified Ricci flow (3.72), which is conjugate to genuine Ricci flow by a diffeomorphism as discussed in that lecture. Thus while we have not established Ricci flow as a gradient flow directly, we have managed to find a whole family of gradient flows (parameterised by a choice of static measure dm, or equivalently by a choice of potential function f evolving by (3.205)) which are equivalent to Ricci flow modulo diffeomorphism[66]. As remarked in [**Pe2002**], one can view f as a kind of gauge function for the Ricci flow.

Example 3.9.1. If (M, g) is a Euclidean space $M = \mathbf{R}^d$ with the contracted Euclidean metric $g = \frac{\tau}{t_0}\eta$ for times $0 \leq t < t_0$, where $\tau := t_0 - t$ and η is the standard metric, with dm equal to the Gaussian measure $\frac{1}{(4\pi t_0)^{d/2}} e^{-|x|^2/4t_0}\, dx$ (thus $f(t,x) = \frac{|x|^2}{4t_0} + \frac{d}{2}\log(4\pi\tau)$), then g, f solve (3.209), (3.210). (One has to be a bit careful here because M is non-compact, of course.)

We can of course conjugate away the infinitesimal diffeomorphism given by the vector field ∇f, which converts the system (3.209), (3.210) to the system

$$\dot{g} = -2\mathrm{Ric}; \quad \dot{f} = -\Delta f + |\nabla f|_g^2 - R \tag{3.212}$$

(here we use the fact that $\mathcal{L}_{\nabla f} f = |\nabla f|_g^2$), which is Ricci flow coupled with a nonlinear backwards heat equation[67] for the potential f). The non-linear backwards heat equation equation for f can be linearised by setting $u := e^{-f}$, in which case it becomes the *adjoint heat equation*

$$\dot{u} = -\Delta u + Ru. \tag{3.213}$$

Exercise 3.9.2. Writing $dm := u d\mu$, show that (3.213) is equivalent to the equation

$$\frac{d}{dt} dm = -\Delta dm \tag{3.214}$$

[66]Indeed, by placing an appropriate Riemannian structure on the moduli space of metrics modulo diffeomorphism, one can express Ricci flow modulo diffeomorphism as a true (formal) gradient flow; see [**KlLo2006**, Section 9].

[67]Note that the equation for f is not always solvable forwards in time for any non-zero amount of time, but we can always solve it instantaneously at any fixed time, which is good enough for first variation analysis.

where dm is viewed as a d-form for the purposes of applying the Laplacian. Thus the adjoint heat equation can be viewed as the backwards heat equation for d-forms.

Example 3.9.2. If (M, g) is a static Euclidean space $M = \mathbf{R}^d$ and $f(t,x) = \frac{|x|^2}{4\tau} + \frac{d}{2}\log(4\pi\tau)$ with $\tau = t_0 - t$ and the time variable t is restricted to be less than t_0, then g, f solve (3.212), and $dm = e^{-f}d\mu$ is the Gaussian measure $\frac{1}{(4\pi\tau)^{d/2}}e^{-|x|^2/4\tau}\,dx$, which solves the backwards heat equation. Note that this is the conjugated version of Example 3.9.1. Again, one needs to take care because M is non-compact.

By performing this conjugation, the measure m is no longer static, and we reflect this by changing the notation a little to

$$\mathcal{F}(M,g,f) := \mathcal{F}_{e^{-f}\mu}(M,g) = \int_M (|\nabla f|^2 + R)e^{-f}\,d\mu. \tag{3.215}$$

The relationship between $\mathcal{F}$ and the flow (3.212) is analogous to that between $\mathcal{F}_m$ and (3.209), (3.210). For instance, we have the following analogue of (3.211):

Exercise 3.9.3. If g, f solve (3.212), show that

$$\frac{d}{dt}\mathcal{F}(M,g,f) = 2\int_M |\mathrm{Ric} + \mathrm{Hess}(f)|^2 e^{-f}\,d\mu. \tag{3.216}$$

Thus $\mathcal{F}(M,g,f)$ is monotone non-decreasing in time. We would like to use this to develop a controlled quantity for Ricci flow, but we need to eliminate f. This can be accomplished by taking an infimum, defining

$$\lambda(M,g) := \inf_{f:\int_M e^{-f}\,d\mu=1} \mathcal{F}(M,g,f). \tag{3.217}$$

The normalisation $\int_M e^{-f}\,d\mu = 1$ (which makes dm a probability measure) is needed to ensure a meaningful infimum; note that this normalisation is preserved by the flow (3.212) since dm is only moved around by diffeomorphisms. This quantity has an interpretation as the best constant in a Poincaré inequality:

Exercise 3.9.4. Show that $\lambda(M,g)$ is the least number for which one has the inequality

$$\int_M 4|\nabla u|_g^2 + R|u|^2 \, d\mu \geq \lambda(M,g) \int_M |u|^2 \, d\mu \tag{3.218}$$

for all u in the *Sobolev space* $H^1(M)$. *Hint*: reduce to the case when u is positive and smooth and then make the substitution $u = e^{-f/2}$. Conclude in particular that $\lambda(M,g)$ is finite, that it is the least eigenvalue of the self-adjoint modified Laplacian $-4\Delta + 4R$, and lies between $R_{\min}$ and the average scalar curvature $\overline{R} := \int_M R \, d\mu / \int_M \, d\mu$.

A variational argument (using the standard fact that $H^1(M)$ embeds compactly into $L^2(M)$) shows that equality in (3.218) is attained by some strictly positive $u = e^{-f/2}$ with norm $\int_M |u|^2 \, d\mu = 1$, and so the infimum in (3.217) is also attained for some f. Applying the flow (3.212) instantaneously at a given time, we conclude (formally[68], at least) that we have the monotonicity formula

$$\frac{d}{dt}\lambda(M,g) = 2\int_M |\mathrm{Ric} + \mathrm{Hess}(f)|^2 e^{-f} \, d\mu \tag{3.219}$$

for any solution to Ricci flow (3.193), where f is the extremiser for (3.217) (note that this extremiser f need not evolve via (3.215)).

This monotonicity is similar to the monotonicity of $R_{\min}$. For instance, the functional $\lambda(M,g)$ has a dimension of -2 in the sense of the previous lecture, which is the same as $R_{\min}$. As further evidence of similarity, we have:

Exercise 3.9.5. Show that $\frac{d}{dt}\lambda(M,g) \geq \frac{2}{d}\lambda(M,g)^2$, and use this to conclude an analogue of Proposition 3.4.5 for $\lambda(M,g)$. In particular conclude that Ricci flow must develop a finite time singularity if $\lambda(M,g)$ is positive.

Exercise 3.9.6. If (M,g) is a Ricci flow which is a *steady breather* in the sense that it is periodic modulo isometries (thus $(M,g(t))$ is isometries to $(M,g(0))$ for some $t > 0$), show that at time zero we have

$$\mathrm{Ric} = -\mathrm{Hess}(f) = -\frac{1}{2}\mathcal{L}_{\nabla f} g \tag{3.220}$$

[68] One can in fact make this formula rigorous whenever the Ricci flow is smooth and M is compact, but we will not detail this here.

for some $f : M \to \mathbf{R}$. Conclude that $g(t) = \exp(t\nabla f)^* g(0)$, thus $(M, g(t))$ simply evolves by diffeomorphism by the gradient field f. (For this you may need to use the uniqueness of the initial value problem for Ricci flow.) In other words, all steady breathers are *gradient steady solitons*.

Remark 3.9.3. One can apply a similar argument to deal with compact *expanding breathers* (in which $(M, g(t))$ is isometric to a larger dilate of $(M, g(0))$ for some $t > 0$ by normalising $\lambda(M, g)$ by a power of the volume as in Exercise 3.8.1, concluding that such breathers are necessarily gradient expanding solitons with

$$\mathrm{Ric} = -\mathrm{Hess}(f) - \frac{g}{2\sigma} \tag{3.221}$$

at time zero for some potential f and some $\sigma > 0$; see [**Pe2002**] or [**KlLo2006**, Section 7] for details. With a little more work[69] (using the maximum principle) one can in fact show that f is constant, and so the only compact expanding breathers are Einstein manifolds. This normalisation of $\lambda(M, g)$ is also closely related to the *Yamabe invariant* of M; see [**Ko2006**] for further discussion.

Example 3.9.4. Any Ricci-flat manifold (i.e. $\mathrm{Ric} = 0$) is of course a gradient steady soliton with $f = 0$. A more non-trivial example is given by *Hamilton's cigar soliton* (also known as *Witten's black hole*), which is the two-dimensional manifold $M = \mathbf{R}^2$ with the conformal metric $dg^2 = \frac{dx^2+dy^2}{1+x^2+y^2}$ and gradient function $f := \log\sqrt{1+x^2+y^2}$; we leave the verification of the gradient shrinking property (3.220) as an exercise.

Remark 3.9.5. If Ricci flow was a gradient flow for a functional which was geometric (or more precisely, invariant under diffeomorphism), then this flow could not deform a metric by any non-trivial diffeomorphism (since this is a stationary direction for this functional, rather than a steepest descent). Thus the existence of non-trivial gradient steady solitons, such as the cigar soliton, explains why Ricci flow

[69] This result can also be established using Exercise 3.8.1 directly, as follows from work of Hamilton[**Ha1999**].

cannot be directly expressed[70] as a gradient flow without introducing a non-geometric object such as the reference measure dm or the potential function f.

Exercise 3.9.7. If (M, g) is a gradient steady soliton with potential f, show that $R + \Delta f = 0$, $|\nabla f|^2 + R = \text{const}$, and $\dot{f} = |\nabla f|^2$. *Hint*: to prove the second identity, differentiate (3.220) and use the second Bianchi identity (Exercise 3.1.7.) Use the maximum principle to then conclude that the only compact gradient steady solitons are the Ricci-flat manifolds.

3.9.2. Nash entropy. Let us return to our analysis of the functional $\mathcal{F}_m(M, g)$, in which $dm = e^{-f}\, d\mu$ was fixed and g evolved by the modified Ricci flow (3.209) (which forced f to evolve by the backwards heat equation (3.210)). We then obtained the monotonicity formula (3.211). We shall normalise dm to be a probability measure.

We can squeeze a little bit more out of this formula - in particular, making it scale invariant - by introducing the *Nash entropy*

$$N_m(M, g) := \int \log \frac{dm}{d\mu}\, dm = -\int f\, dm \tag{3.222}$$

which is the *relative entropy*[71] of $d\mu$ with respect to the background measure dm. From (3.210) and one integration by parts (using (3.201), of course) we know how this entropy changes with time:

$$\frac{d}{dt} N_m(M, g) = \int (|\nabla f|^2 + R)\, dm = \mathcal{F}_m(M, g). \tag{3.223}$$

To exploit this identity, let us first consider the case of gradient shrinking solitons:

Exercise 3.9.8. Suppose that a Riemannian manifold $(M, g) = (M, g(0))$ verifies an equation of the form

$$\text{Ric} = -\text{Hess}(f) + \frac{1}{2\tau} g \tag{3.224}$$

for some function f and some $\tau > 0$. Show that this equation is preserved for times $0 \leq t < \tau(0)$ if g evolves by Ricci flow, if τ

[70] See also [**Mu2006**, Proposition 1.7] for a different way of seeing that Ricci flow is not a pure gradient flow.

[71] Some further relations and analogies between the functionals described here and notions of entropy from statistical mechanics are discussed in [**Pe2002**].

evolves by $\dot{\tau} = -1$ (i.e. $\tau(t) = \tau(0) - t$), and $\partial_t f = |\nabla f|_g^2$, and that $g(t) = \frac{\tau(t)}{\tau(0)} \exp(t\nabla f) g(0)$ for all $0 \leq t < \tau(0)$. Such solutions are known as *gradient shrinking solitons*; they combine Ricci flow with the diffeomorphism and scaling flows from Section 3.2. Note that any positively curved Einstein manifold, such as the sphere, will be a gradient shrinking soliton (with $f = 0$). Example 3.9.1 also shows that Euclidean space can also be viewed as a gradient shrinking soliton.

If we are to find a scale-invariant (and diffeomorphism-invariant) monotone quantity for Ricci flow, it had better be constant on the gradient shrinking solitons. In analogy with (3.211), we would therefore like the variation of this monotone quantity with respect to Ricci flow to look something like

$$2 \int_M |\mathrm{Ric} + \mathrm{Hess}(f) - \frac{1}{2\tau} g|_g^2 \, dm \tag{3.225}$$

where τ is a backwards time variable, i.e. some quantity decreasing at the constant rate

$$\dot{\tau} = -1. \tag{3.226}$$

But the scaling is wrong; time has dimension 2 with respect to the Ricci flow scaling (3.170), and so the dimension of a variation of a scale-invariant quantity should be -2, while the expression (3.225) has dimension[72] -4. So actually we should be looking at

$$2\tau \int_M |\mathrm{Ric} + \mathrm{Hess}(f) - \frac{1}{2\tau} g|_g^2 \, dm. \tag{3.227}$$

To find a functional whose derivative is (3.227), we expand the integrand as

$$|\mathrm{Ric} + \mathrm{Hess}(f) - \frac{1}{2\tau} g|_g^2 = |\mathrm{Ric} + \mathrm{Hess}(f)|_g^2 - \frac{1}{\tau}(R + \Delta f) + \frac{d}{4\tau^2}. \tag{3.228}$$

Using (3.223) and the normalisation $\int_M \, dg = 1$, we can thus express (3.227) as

$$\tau \frac{d}{dt} \mathcal{F}_m(M, g) - 2\mathcal{F}_m(M, g) + \frac{d}{2\tau}. \tag{3.229}$$

[72] Note that f should be dimensionless (up to logarithms), τ has the same dimension of time, i.e. 2, and $\int_M dm = 1$ is of course dimensionless.

Using (3.223) and (3.226), we can express this as a total derivative:

$$\frac{d}{dt}(\tau \mathcal{F}_m(M,g) - N_m(M,g) + \frac{d}{2}\log\tau). \tag{3.230}$$

Thus the quantity in parentheses is monotone increasing in time under Ricci flow (and with f, τ evolving by (3.210), (3.226)). In analogy with Example 3.9.1, we rewrite the potential function f as

$$f = \tilde{f} + \frac{d}{2}\log(4\pi\tau) \tag{3.231}$$

then $\tilde{f}$ obeys a slight variant of (3.210), namely

$$\frac{d}{dt}\tilde{f} = -\Delta f - R + \frac{d}{2\tau} \tag{3.232}$$

and is related to the fixed measure m by the formula

$$dm = (4\pi\tau)^{-d/2} e^{-\tilde{f}}\, d\mu. \tag{3.233}$$

The equality between (3.227) and (3.230) now becomes

$$\frac{d}{dt}\mathcal{W}_m(M,g,\tau) = 2\tau \int_M |\mathrm{Ric} + \mathrm{Hess}(\tilde{f}) - \frac{1}{2\tau}g|_g^2\, dm \tag{3.234}$$

where

$$\mathcal{W}_m(M,g,\tau) := \int_M [\tau(R + |\nabla\tilde{f}|^2) + \tilde{f} - d]\, dm. \tag{3.235}$$

The $-d$ term here is harmless (since m is fixed), and is in place to normalise this expression to vanish in the Euclidean case (Example 3.9.1, where now $\tilde{f}(t,x) = |x|^2/4t_0$).

As before, it is convenient to conjugate away the diffeomorphism by ∇f to recover a pure Ricci flow. Define the *Perelman entropy* $\mathcal{W}(M,g,f,\tau)$ of a manifold (M,g), a scalar function $f : M \to \mathbf{R}$, and a positive real $\tau > 0$, by

$$\mathcal{W}(M,g,f,\tau) = \int_M [\tau(R + |\nabla f|^2) + f - d](4\pi\tau)^{-d/2} e^{-f}\, d\mu. \tag{3.236}$$

Note that this quantity has dimension 0 (if f is viewed as dimensionless, and τ given the dimension 2).

Exercise 3.9.9. Suppose that g evolves by Ricci flow (3.193), f evolves by the nonlinear backward heat equation

$$\dot{f} = -\Delta f + |\nabla f|^2 - R + \frac{d}{2\tau}, \tag{3.237}$$

and τ evolves by (3.226). Show that

(3.238)
$$\frac{d}{dt}\mathcal{W}(M,g,f,\tau) = 2\tau \int_M |\mathrm{Ric} + \mathrm{Hess}(f) - \frac{1}{2\tau}g|_g^2 (4\pi\tau)^{-d/2} e^{-f}\ d\mu.$$

If we write $u := (4\pi\tau)^{-d/2}e^{-f}$, show that (3.237) is also equivalent to the adjoint heat equation

$$\dot{u} = -\Delta u + Ru. \tag{3.239}$$

We have thus obtained a scale-invariant monotonicity formula, albeit one which depends on two additional time-varying parameters, f and τ. To eliminate them, the obvious thing to do is to just take the infimum over all f and τ; but we need to be sure that the infimum exists at all. This will be studied next.

3.9.3. Connection to the log-Sobolev inequality. We have just established the monotonicity formula (3.238) whenever g evolves by Ricci flow (3.193) and f, τ evolve by (3.237), (3.226). Let us now temporarily specialise to the case when (M,g) is a static Euclidean space $\mathbf{R}^d$ (which of course obeys Ricci flow), and $\tau = -t$ (which of course obeys (3.226)), and now restrict to negative times $t < 0$. Now all curvatures R, Ric vanish, thus for instance by (3.239) we see that $u = (4\pi\tau)^{-d/2}e^{-f}$ obeys the free backwards heat equation $\dot{u} = -\Delta u$. We will normalise $dm = u\ d\mu$ to be a probability measure, thus $\int_{\mathbf{R}^d} u\ dx = 1$.

Example 3.9.6. The key example to keep in mind here is $f(t,x) = |x|^2/4\tau$, in which case u becomes the backwards heat kernel $u(t,x) = (4\pi\tau)^{-d/2}e^{-|x|^2/4\tau}$.

We can now re-express the functional (3.236) in terms of u as

$$\mathcal{W}(M,g,f,\tau) = \int_{\mathbf{R}^d} (\tau\frac{|\nabla u|^2}{u} - u\log u)\ dx - \frac{d}{2}\log(4\pi\tau) - d. \tag{3.240}$$

One easily verifies by direct calculation that this expression vanishes in the model case of Example 3.9.6. For more general u, we know that this quantity is monotone increasing in time, and so

$$\mathcal{W}(M,g,f,\tau)(t) \geq \lim_{t\to-\infty} \mathcal{W}(M,g,f,\tau)(t). \tag{3.241}$$

Now suppose u is some non-negative test function $u_0(x)$ at time zero with total mass 1, then from the fundamental solution for the backwards heat equation we have

(3.242)
$$u(t,x) = \frac{1}{(4\pi\tau)^{d/2}} \int_{\mathbf{R}^d} e^{-|x-y|^2/4\tau} u_0(y)\ dy = \frac{1}{(4\pi\tau)^{d/2}} \tilde{u}(t, x/\sqrt{\tau})$$

where $\tilde{u}$ is the renormalised solution

$$\tilde{u}(t,x) := \int_{\mathbf{R}^d} e^{-|x-(y/\sqrt{\tau})|^2/4} u_0(y)\ dy. \tag{3.243}$$

Observe that $\tilde{u}(t,x)$ converges pointwise to $e^{-|x|^2/4}$ as $t \to -\infty$ for fixed x. Thus in some renormalised sense this general solution is converging to the model solution in Example 3.9.4 in the limit $t \to -\infty$.

We can rewrite the functional (3.240) after some calculation as

$$\mathcal{W}(M,g,f,\tau) = \int_{\mathbf{R}^d} [\tau \frac{|\nabla \tilde{u}|^2}{\tilde{u}^2} - \log \tilde{u}](-4\pi)^{-d/2} \tilde{u}\ dx - d. \tag{3.244}$$

One can check that $\nabla \tilde{u}$ is converging pointwise to $\nabla e^{-|x|^2/4}$. A careful application of dominated convergence then shows that in the limit $t \to -\infty$, (3.244) converges to the value attained in Example 3.9.4, i.e. zero. By the monotonicity formula, we have thus demonstrated that

$$\mathcal{W}(M,g,f,\tau) \geq 0 \tag{3.245}$$

for all times $-\infty < t < 0$. Writing $u = \phi^2$ and rearranging (3.240), we conclude the *log-Sobolev inequality*

$$2 \int_{\mathbf{R}^d} \phi^2 \log \phi\ dx \leq 4\tau \int_{\mathbf{R}^d} |\nabla \phi|^2\ dx - \frac{d}{2} \log(4\pi\tau) - d \tag{3.246}$$

valid whenever $\tau > 0$ and $\int_{\mathbf{R}^d} \phi^2\ dx = 1$.

Exercise 3.9.10. By letting $dm := (4\pi\tau)^{-d/2} e^{-|x|^2/4\tau} dx$ be standard Gaussian measure and writing $u dx = F^2 dm$, deduce the original log-Sobolev inequality[73]

$$\int_{\mathbf{R}^d} F^2 \log F^2\ dm \leq \frac{1}{\tau} \int_{\mathbf{R}^d} |\nabla F|^2 dm \tag{3.247}$$

[73] One key feature of this inequality, as compared to more traditional Sobolev inequalities, is that it is almost completely independent of the dimension d.

of Gross[**Gr1975**], valid whenever $\tau > 0$ and $\int_{\mathbf{R}^d} F^2 \, dm = 1$.

Remark 3.9.7. We have seen how knowledge of the heat kernel can lead to log-Sobolev inequalities, by evolving by the (backwards) heat flow (this is an example of the *semigroup method* for proving inequalities). This connection can in fact be reversed, using log-Sobolev inequalities to deduce information about heat kernels. Heat kernels can in turn be used to deduce ordinary Sobolev estimates, which then imply log-Sobolev estimates by convexity inequalities such as Hölder's inequality, thus showing that all these phenomena are morally equivalent. There is a vast literature on these subjects (and other related topics, such as hypercontractivity); so much so that there are not only multiple surveys on the subject, but even a survey of all the surveys[**Gr2006**]!

We now return to the case of general Ricci flows (not just the Euclidean one).

Exercise 3.9.11. Let (M, g) be a compact Riemannian manifold, and let $\tau > 0$. Using the Euclidean log-Sobolev inequality (3.239), show that we have a lower bound of the form $\mathcal{W}(M, g, f, \tau) \geq -C(M, g, \tau)$ for all functions f with $\int (4\pi\tau)^{-d/2} e^{-f} \, d\mu = 1$. Show in fact that $C(M, g, \tau)$ can be chosen to depend only on τ, the dimension, an upper bound for the magnitude of the Riemann curvature, and a lower bound for the injectivity radius. Using a rescaling and compactness argument, show also that we can take $C(M, g, \tau) \to 0$ as $\tau \to 0$; details can be found in [**Pe2002**, Section 3.1].

We can now define the quantity $\mu(M, g, \tau)$ to be the infimum of $\mathcal{W}(M, g, f, \tau)$ for all functions f with $\int (4\pi\tau)^{-d/2} e^{-f} \, d\mu = 1$; thus $\mu(M, g, \tau)$ is non-decreasing if we evolve τ by (3.226). Thus we have obtained a one-parameter family of dimensionless monotone quantities (recalling that τ has dimension 2 with respect to scaling).

Remark 3.9.8. One can interpret $\mu(M, g, \tau)$ as a nonlinear analogue of the eigenvalue $\lambda(M, g)$. Indeed, just as $\lambda(M, g)$ is the least number λ for which one can solve the linear eigenfunction equation

$$(4\Delta + R)\Phi = \lambda\Phi \tag{3.248}$$

subject to the constraint $\int_M \Phi^2 = 1$, $\mu(M, g, \tau)$ is the least number μ for which one can solve the nonlinear eigenfunction equation

$$\tau(4\Delta + R)\Phi = 2\Phi \log \Phi + (\mu + d)\Phi \tag{3.249}$$

subject to the constraints $\Phi > 0$ and $\int_M (4\pi\tau)^{-d/2}\Phi^2 \, d\mu = 1$. In particular we expect $\mu(M, g, \tau)$ to behave roughly like $\tau\lambda(M, g)$ in the limit $\tau \to \infty$.

Exercise 3.9.12. Show that the only shrinking breathers (those in which $(M, g(t))$ is isometric to a contraction of $(M, g(0))$ for some $t > 0$) are the gradient shrinking solitons.

3.9.4. Non-collapsing. We now relate log-Sobolev inequalities (i.e. lower bounds on $\mu(M, g, \tau)$) to non-collapsing. We first note that by substituting $(4\pi\tau)^{-d/2}e^{-f} = \phi^2$ into (3.236) as in the Euclidean case, that we have the log-Sobolev inequality

$$\begin{aligned}\int_M \phi^2 \log \phi^2 \, d\mu \leq 4\tau \int_M |\nabla\phi|_g^2 \, d\mu \\ + \tau \int_M R|\phi|^2 \, d\mu - \frac{d}{2}\log(4\pi\tau) - d - \mu(M, g, \tau)\end{aligned} \tag{3.250}$$

whenever ϕ is non-negative with $\int_M \phi^2 \, d\mu = 1$.

To use this, suppose we have a ball $B = B(p, \sqrt{\tau})$ which has bounded normalised curvature, so in particular $R = O(\tau^{-1})$ on this ball. On the other hand, if ϕ is supported on B with L^2 mass 1, then from *Jensen's inequality* we have

$$\int_M \phi^2 \log \phi^2 \, d\mu \geq \log \frac{1}{\mathrm{Vol}(B)} \tag{3.251}$$

and we thus conclude from (3.9.4) that

$$\log \frac{\tau^{d/2}}{\mathrm{Vol}} \leq 4\tau \int_M |\nabla\phi|_g^2 + O(1) - \mu(M, g, \tau). \tag{3.252}$$

If we let $\phi(x) := c\psi(d(x, p)/\sqrt{\tau})$, where ψ is a bump function that equals 1 on $[-1/2, 1/2]$ and is supported on $[-1, 1]$ (thus $\phi = c$ on the ball $B_{1/2} := B(p, \sqrt{\tau}/2)$, and $c \leq 1/\mathrm{Vol}(B_{1/2})^{1/2}$ is the normalisation constant needed to ensure that ϕ has L^2 mass one, then

$\nabla\phi = O(c/\sqrt{\tau})$ on this ball, and so we conclude

$$\log \frac{\tau^{d/2}}{\mathrm{Vol}(B)} \leq O(\mathrm{Vol}(B)/\mathrm{Vol}(B_{1/2})) - \mu(M,g,\tau). \tag{3.253}$$

At this point we need to invoke the relative Bishop-Gromov inequality (see Lemma 3.10.1 below) from comparison geometry, which among other things ensures that $\mathrm{Vol}(B) = O(\mathrm{Vol}(B_{1/2})$ under the assumption of bounded normalised curvature. Indeed, from equations (3.203) and (3.205) from the previous lecture we see that $\mathcal{L}_{\partial_r}\, d\mu = O(1/r)\, d\mu$ inside the ball of radius $1/\sqrt{\tau}$, from which the claim easily follows within[74] the radius of injectivity.

Using this inequality, we thus conclude that

$$\mathrm{Vol}(B) \gg \tau^{d/2} \exp(\mu(M,g,\tau)). \tag{3.254}$$

Thus a lower bound on $\mu(M,g,\tau)$ enforces non-collapsing of volume at scale τ.

Exercise 3.9.13. Use (3.254), Exercise 3.9.11 and the monotonicity properties of $\mu(M,g,\tau)$ to establish κ-noncollapsing of Ricci flows (Theorem 3.8.15).

Remark 3.9.9. This argument in fact establishes a stronger form of non-collapsing, in which in order to get non-collapsing at time t_0 and scale r_0, one only needs bounded normalised curvature at time t_0 (instead of on the time interval $[t_0 - r_0^2, t_0]$). It also works in arbitrary dimension. The second proof of non-collapsing that we will give, based on the Perelman reduced volume instead of Perelman entropy, needs the spacetime bounded normalised curvature assumption but also works in arbitrary dimension.

Remark 3.9.10. The parameter κ in the above result, which measures the quality of the non-collapsing, will deteriorate with time T. This is because the decay of τ from (3.226) entails that in order to get non-collapsing of the manifold at time t_0 and scale r_0, one needs some non-collapsing at time zero and scale $\sqrt{r_0 + t_0^2}$. Of course, since the manifold is initially compact, one always has some non-collapsing

[74] To generalise the inequality beyond this region, one simply works on the region inside the *cut locus*, which is star-shaped around the origin in T_pM.

at each scale, but the quantitative constants associated to this non-collapsing will deteriorate as the scale increases, which will happen when T increases. Fortunately (and especially in view of our finite time extinction results) we only need to analyse Ricci flow on compact (though potentially rather large) time intervals $[0, T]$.

Remark 3.9.11. It was recently shown[**Zh2007**] that the monotonicity properties of the quantities $\mu(M, g, \tau)$ also hold for Ricci flows with surgery. This can be used to replace all applications of Perelman reduced volume in the existing proof of the Poincaré conjecture in the literature by Perelman (as well as in the expositions of [**KlLo2006**], [**CaZh2006**], and [**MoTi2007**]) by Perelman entropy; see [**Zh2008**]. However, we shall mostly follow the original arguments of Perelman in this course.

Remark 3.9.12. The above entropy functionals are also useful for studying the forward or backward heat equation on a static Riemannian manifold (M, g) (basically, one keeps the heat-type equations for u or f but now replace Ricci flow by the trivial flow $\dot{g} = 0$). However, some sign assumptions on curvature are now needed to recover the same type of monotonicity results. See [**Ni2004**] for details.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/04/24`. Thanks to Américo Tavares, and Dan for corrections.

3.10. Comparison geometry, the high-dimensional limit, and Perelman reduced volume

We now turn to Perelman's second scale-invariant monotone quantity for Ricci flow, now known as the *Perelman reduced volume*. We saw in the previous lecture that the monotonicity for Perelman entropy was ultimately derived (after some twists and turns) from the monotonicity of a potential under gradient flow. In this lecture, we will show (at a heuristic level only) how the monotonicity of Perelman's reduced volume can also be "derived", in a formal sense, from another source of monotonicity, namely the relative Bishop-Gromov inequality (Lemma 3.10.1) in comparison geometry, which has already been mentioned in previous lectures. Interestingly, in order to obtain this

connection, one must first reinterpret parabolic flows such as Ricci flow as the limit of a certain high-dimensional Riemannian manifold as the dimension becomes infinite; this is part of a more general philosophy that parabolic theory is in some sense an infinite-dimensional limit of elliptic theory. Our treatment here is a (liberally reinterpreted) version of [**Pe2002**, Section 6].

In the next few lectures we shall give a rigorous proof of this monotonicity, without using the infinite-dimensional limit and instead[75] using results related to the Li-Yau-Hamilton Harnack inequality.

3.10.1. The Bishop-Gromov inequality. Let p be a point in a complete d-dimensional Riemannian manifold (M, g). As noted in Section 3.8, we can use the exponential map to pull back M and g to the tangent space T_pM, which is also equipped with the radial variable r and the radial vector field $\partial_r = \mathrm{grad}(r)$. From Exercise 3.8.7, we have the transport equation

$$\mathcal{L}_{\partial_r} d\mu = (\Delta r)\, d\mu \tag{3.255}$$

for the volume measure $d\mu$, and a transport inequality

$$\nabla_{\partial_r} \Delta r + \frac{1}{d-1}(\Delta r)^2 \leq \nabla_{\partial_r} \Delta r + |\mathrm{Hess}(r)|^2 = -\mathrm{Ric}(\partial_r, \partial_r) \tag{3.256}$$

for the Laplcian Δr which appears in (3.255). In particular, if we assume the lower bound

$$\mathrm{Ric} \geq (d-1)Kg \tag{3.257}$$

for Ricci curvature in a ball $B(p, r_0)$ for some real number K, then from the Gauss lemma (Lemma 3.8.4) we have

$$\nabla_{\partial_r} \Delta r + \frac{1}{d-1}(\Delta r)^2 \leq -(d-1)K. \tag{3.258}$$

Also, from an expansion around the origin (see e.g. (3.182) or (3.184)) we have

$$\Delta r = \frac{d-1}{r} + O(r) \tag{3.259}$$

[75] There are several other approaches to understanding Perelman's reduced volume, such as Lott's formulation[**Lo2008**] based on optimal transport, but we will restrict attention in this chapter to the methods that are in [**Pe2002**].

for small r. In principle, (3.258) and (3.259) lead to upper bounds on Δr, which when combined with (3.255) lead to upper bounds on $d\mu$, which in turn lead to upper bounds on $B(p, r_0)$. One can of course just go ahead and compute these bounds, but one computation-free way to proceed is to introduce the model geometry (M_K, g_K), defined as

(1) the standard round sphere $\sqrt{K} \cdot S^d$ of radius $\sqrt{K}$ (and thus constant sectional curvature K) if $K > 0$ (Example 3.8.1);

(2) the standard hyperbolic space $\sqrt{-K} \cdot H^d$ of constant sectional curvature K if $K < 0$ (Example 3.8.2); or

(3) the standard Euclidean space $\mathbf{R}^d$ if $K = 0$.

As all of these spaces are homogeneous (in fact, they are *symmetric spaces*), the choice of origin p in this model geometry is irrelevant. Observe that the orthogonal group $O(d)$ acts isometrically on each of these spaces, with the orbits being the spheres centred at p. This implies that at any point q not equal to p, Hess(r) is invariant under conjugation by the stabiliser of that group on q, which easily implies that it is diagonal on the tangent space to the sphere (i.e. to the orthogonal complement of ∂_r). From this we see that for this model geometry, the inequality in (3.256) is in fact an equality. Since the model geometry also has constant sectional curvature K (which implies equality in (3.257)), we thus see that one has equality in (3.258) for this model geometry as well. From this we can conclude:

Lemma 3.10.1 (Relative Bishop-Gromov inequality)**.** *With the assumptions as above, the volume ratio* $\mathrm{Vol}_{M,g}(B_{M,g}(p,r))/\mathrm{Vol}_{M_K,g_K}(B_{M_K,g_K}(p,r$ *is a non-increasing function of* r *as* $0 < r < r_0$.

Exercise 3.10.1. Prove Lemma 3.10.1. *Hint*: One can avoid all issues with non-injectivity by working inside the *cut locus* of p, which determines a star-shaped region in T_pM. In the positive curvature case $K > 0$, the model geometry M_K has a finite radius of injectivity, but observe that we may without loss of generality reduce to the case when r_0 is less than or equal to that radius (or one can invoke *Myers' theorem*, see Exercise 3.10.2 below). To prove the monotonicity of ratios of volumes of balls, it may be convenient to first achieve the analogous claim for ratios of volumes of spheres, and then use the

Gauss lemma (Lemma 3.8.4) and the fundamental theorem of calculus to pass from spheres to balls.

Exercise 3.10.2. Prove *Myers' theorem*: if a Riemannian manifold obeys (3.257) everywhere for some $K > 0$, then the diameter of the manifold is at most $\pi/\sqrt{K}$. *Hint*: in the model geometry, the sphere of radius r collapses to a point when r approaches $\pi/\sqrt{K}$.

Remark 3.10.2. Lemma 3.10.1 implies the volume comparison result $\mathrm{Vol}(B(p,r))/\mathrm{Vol}(B(p,r/2)) = O(1)$ whenever one has bounded normalised curvature, which was used in Section 3.9; indeed, thanks to the above inequality, it suffices to prove the claim for model geometries.

Setting $K = 0$, we obtain

Corollary 3.10.3. *Let (M,g) be a complete d-dimensional Riemannian manifold of non-negative Ricci curvature, and let p be a point in M. Then $\mathrm{Vol}(B(p,r))/r^d$ is a non-increasing function of r.*

Let us refer to the quantity $\mathrm{Vol}(B(p,r))/r^d$ as the *Bishop-Gromov reduced volume* at the point p and the scale r; thus we see that this quantity is dimensionless (i.e. invariant under scaling of the manifold and of r), and non-increasing in r when one has non-negative Ricci curvature (and in particular, for Ricci-flat manifolds).

Exercise 3.10.3. Use the Bishop-Gromov inequality to state and prove a rigorous version of the following informal claim: if a Riemannian manifold is non-collapsed at a point p at one scale $r_0 > 0$ (as defined in Section 3.8), then it is also non-collapsed at all larger scales $r_1 > r_0$.

3.10.2. Parabolic theory as infinite-dimensional elliptic theory. We now come to an interesting (but still mostly heuristic) correspondence principle between elliptic theory and parabolic theory, with the latter being viewed as an infinite-dimensional limit of the former, in a manner somewhat analogous to that of the *central limit theorem* in probability. To get some idea of what I mean by this correspondence, consider the following (extremely incomplete, non-rigorous, inaccurate, and imprecise) dictionary:

Elliptic	Parabolic
Riemannian manifold (M, g) Complete manifold	Riemannian flow $t \mapsto (M, g(t$ Ancient flow of complete ma
Spatial origin 0 Elliptic scaling $x \mapsto \lambda x$	Spacetime origin $(0, 0)$ Parabolic scaling $(t, x) \mapsto (\lambda$
Laplace equation $\Delta u = 0$ Ricci flat manifold $\mathrm{Ric} = 0$	Heat equation $-\partial_t u + \Delta u =$ Ricci flow $\partial_t g = -2\mathrm{Ric}$
Mean value principle $u(0) = \int_{S^{d-1}} u(r\omega)\, d\mu(\omega)$ Normalised measure on the sphere $r \cdot S^{d-1}$	Fundamental solution $u(0,0) = \frac{1}{(4\pi\tau)^{d/2}} \int_{\mathbf{R}^d} e^{-\lvert x\rvert^2/4}$ Heat kernel $\frac{1}{(4\pi\tau)^{d/2}} e^{-\lvert x\rvert^2/4\tau}$
Maximum principle	Maximum principle
Ball of radius $O(r)$ around spatial origin Radial variable $r = \lvert x\rvert$	Cylinder of radius $O(r)$ and extending backwards in time $\lvert x\rvert$ or $\sqrt{-t} = \sqrt{\tau}$
Bishop-Gromov reduced volume	Perelman reduced volume

Remark 3.10.4. Of course, we have not defined Perelman reduced volume yet, but the point is that the monotonicity of Perelman reduced volume for Ricci flow is supposed to be the parabolic analogue of the monotonicity of Bishop-Gromov reduced volume for Ricci-flat manifolds. Note that one has two competing notions of the parabolic radial variable, —x— and $\sqrt{\tau}$, where $\tau := -t$ is the backwards time variable; the ratio between these two competitors is essentially the *Perelman reduced length*, which does not really have a good analogue in the elliptic theory (except perhaps in the "latitude" variable one gets when decomposing a sphere into cylindrical coordinates).

It is well known that elliptic theory can be viewed as the static (i.e. steady state) special case of parabolic theory, but here we want to discuss a rather different connection between the two theories that goes in the opposite direction, in which we view parabolic theory as a limiting case of elliptic theory as the dimension d goes to infinity.

To motivate how this works, let us begin with a smooth ancient solution $u : (-\infty, 0] \times \mathbf{R}^d \to \mathbf{R}$ to the Euclidean heat equation

$$-\partial_t u + \Delta_x u = 0 \tag{3.260}$$

and ask how to convert it to a high-dimensional solution to the Laplace equation. At first glance this looks unreasonable: the Laplacian only contains second order derivative terms, but we have to somehow generate the first-order derivative ∂_t out of this. The trick is to use polar coordinates. Recall that if we parameterise a Euclidean variable $y \in \mathbf{R}^N$ away from the origin as $y = r\omega$ for $r > 0$ and $\omega \in S^{N-1}$, then the Laplacian $\Delta_y f$ of a function $f : \mathbf{R}^N \to \mathbf{R}$ can be expressed by the classical formula

$$\Delta_y f = \partial_{rr} f + \frac{N-1}{r}\partial_r f + \frac{1}{r^2}\Delta_\omega f \tag{3.261}$$

where Δ_ω is the Laplace-Beltrami operator on the sphere. In particular, if f is a radial or spherically symmetric function (so by abuse of notation we write $f(y) = f(r)$), we have

$$\Delta_y f = \partial_{rr} f + \frac{N-1}{r}\partial_r f. \tag{3.262}$$

Now if we look at the high-dimensional limit $N \to \infty$ (noting that f, being radial, is well defined in every dimension), we see that the first order term $\frac{N-1}{r}\partial_r f$ dominates, despite the fact that Δ_y is a second order operator. To clarify this domination (and to bring into view the operator $-\partial_t$ appearing in (3.260)), let us make the change of variables

$$t = -\tau = -\frac{r^2}{2N} = -\frac{y_1^2 + \ldots + y_N^2}{2N} \tag{3.263}$$

(thus $\tau = -t$ is the average of the squared coordinates $y_1^2, \ldots, y_N^2$). A quick application of the chain rule then yields

$$\Delta_y f = -\frac{t}{2N}\partial_{tt} f - \partial_t f \tag{3.264}$$

(one can also see this by writing $f(y) = \tilde{f}(t) = \tilde{f}(-(y_1^2+\ldots+y_N^2)/2N)$ and applying the Laplacian operator Δ_y directly). If we restrict attention to the region of $\mathbf{R}^N$ where all the coordinates y_i are $O(1)$, so $r^2 = O(N)$ and $\tau = -t = O(1)$, and fix $\tilde{f}$ while letting N go off to infinity, we thus see that $\Delta_y f$ converges to $-\partial_t f$ (with errors that are $O(1/N)$).

Returning back to our ancient solution $u : (-\infty, 0] \times \mathbf{R}^d \to \mathbf{R}$ to the heat equation (3.260), it is now clear how to express this solution

as a high-dimensional nearly harmonic function: if we define the high-dimensional lift $u^{(N)} : \mathbf{R}^N \times \mathbf{R}^d \to \mathbf{R}$ of u to the $N+d$-dimensional Euclidean space $\mathbf{R}^N \times \mathbf{R}^d := \{(y,x) : y \in \mathbf{R}^N, x \in \mathbf{R}^d\}$ for some large N by using the change of variables (3.263), i.e.

$$u^{(N)}(y,x) := u(t,x) = u\left(-\frac{y_1^2 + \ldots + y_N^2}{2N}, x\right) \tag{3.265}$$

then we see from (3.264) and (3.260) that $u^{(N)}$ is nearly harmonic as claimed; indeed we have

$$\Delta_{y,x} u^{(N)} = \frac{r^2}{4N^2}\partial_{tt} u - \partial_t u + \Delta_x u = O(1/N) \tag{3.266}$$

in the region $y_i = O(1), x = O(1)$, which implies as before that

$$y_i = O(1), x = O(1), r^2 = O(N), \tau = -t = O(1). \tag{3.267}$$

Remark 3.10.5. Writing y in polar coordinates as $y = r\omega$, the metric ds^2 on $\mathbf{R}^N \times \mathbf{R}^d$ can be expressed as

$$ds^2 = dr^2 + r^2 d\omega^2 + dx^2 = \frac{N}{2\tau} d\tau^2 + \tau d\omega_{1/2N}^2 + dx^2 \tag{3.268}$$

where $d\omega_{1/2N}^2$ is the metric on the sphere S^N of constant curvature $1/2N$. This polar coordinate expression is essentially the first equation in [**Pe2002**, Section 6] (in the Euclidean case), but I have found that the Cartesian coordinate approach can be more illuminating at times.

Remark 3.10.6. The formula (3.263) seems closely related to *Itô's formula* $dt = (dB)^2$ from stochastic calculus, combined perhaps with the central limit theorem, though I was not able to make this connection absolutely precise. Note that for reasons of duality, stochastic calculus tends to involve the backwards heat equation rather than the forwards heat equation (see e.g. the Black-Scholes formula, Section 1.6), which seems to explain why the minus sign in (3.263) is not present in Itô's formula.

To illustrate how this correspondence could be used, let us heuristically derive the classical formula

$$u(0,0) = \frac{1}{(4\pi\tau)^{d/2}} \int_{\mathbf{R}^d} e^{-|x|^2/4\tau} u(-\tau, x)\, dx \tag{3.269}$$

for solutions $u : (-\infty, 0] \times \mathbf{R}^d \to \mathbf{R}$ to the heat equation (3.260) from the classical *mean value principle*

$$u^{(N)}(0,0) = \frac{1}{\mathrm{mes}(r \cdot S^{N+d-1})} \int_{S^{N+d-1}} u^{(N)}(r\omega)\, d\omega \tag{3.270}$$

for harmonic functions $u^{(N)} : \mathbf{R}^N \times \mathbf{R}^d \to \mathbf{R}$. Actually, it will be slightly simpler to use the mean value principle for balls rather than spheres,

$$u^{(N)}(0,0) = \frac{1}{\mathrm{Vol}(B^{N+d}(0,r_0))} \int_{|y|^2+|x|^2 \leq r_0^2} u^{(N)}(y,x)\, dydx, \tag{3.271}$$

though in high dimensions there is actually very little difference between balls and spheres (the bulk of the volume of a high-dimensional ball is concentrated near its boundary, which is a sphere).

Let u and $u^{(N)}$ be as in (3.260) and (3.265). From (3.266) we see that $u^{(N)}$ is almost harmonic; let us be non-rigorous and pretend that $u^{(N)}$ is close enough to harmonic that the formula (3.271) remains accurate for this function. We write the volume of the ball $B^{N+d}(0,r_0)$ as $C_{N,d} r_0^{N+d}$ for some constant $C_{N,d}$. As for the integrand in (3.271), we use polar coordinates $y = r\omega$, $dy = r^{N-1} dr d\omega$ and rewrite (3.271) as

$$c_{N,d} r_0^{-N-d} \int_{\mathbf{R}^d} \int_{0 \leq r \leq \sqrt{r_0^2-|x|^2}} u(-r^2/2N, x) r^{N-1} dr dx \tag{3.272}$$

for some other constant $c_{N,d} > 0$. In view of (3.263), it is natural to write $r_0^2 = 2N\tau$ for some $\tau > 0$, and in view of (3.267) it is natural to work in the regime in which $x = O(1), \tau = O(1)$, and $r_0^2 = O(N)$. Because r^{N-1} is so rapidly increasing when N is large, the bulk of the inner integral is concentrated at its endpoint (cf. our previous remark about high-dimensional balls concentrating near their boundary), and so we expect

$$\begin{aligned} &\int_{0 \leq r \leq \sqrt{r_0^2-|x|^2}} u(-r^2/2N, x) r^{N-1} dr \\ &\quad \approx \frac{1}{N} (\sqrt{r_0^2 - |x|^2})^N u(-(r_0^2 - |x|^2)/2N, x). \end{aligned} \tag{3.273}$$

Since r_0 is so much larger than $|x|$ in our regime of interest, we can heuristically approximate $u(-(r_0^2 - |x|^2)/2N, x)$ by $u(-r_0^2/2N, x) =$

$u(-\tau, x)$. Also, by Taylor approximation we have

$$(\sqrt{r_0^2 - |x|^2})^N \approx r_0^{N/2} \exp(-\frac{N|x|^2}{2r_0^2}). \tag{3.274}$$

Putting all this together, and substituting $r_0^2 = 2N\tau$, we heuristically conclude

$$u(0,0) \approx \frac{\tilde{c}_{N,d}}{\tau^{d/2}} \int_{\mathbf{R}^d} e^{-|x|^2/4\tau} u(-\tau, x)\, dx \tag{3.275}$$

for some other constant $\tilde{c}_{N,d} > 0$. Taking limits as $N \to \infty$ we heuristically obtain (3.269) up to a constant.

Exercise 3.10.4. Work through the calculations more carefully (but still heuristically), using *Stirling's approximation* $\Gamma(n+1) \approx (2\pi n)^{1/2} n^n e^{-n}$ to the Gamma function, together with the classical formulae $\mathrm{mes}(S^{n-1}) = 2\pi^{n/2}/\Gamma(n/2)$, $\mathrm{Vol}(B^n) = \mathrm{mes}(S^{n-1})/n$ for the volume of balls and spheres, to verify that one does indeed get the right constant of $\frac{1}{(4\pi)^{d/2}}$ in (3.269) at the end of the day (as one must).

Now let us perform a variant of the above computations which is more closely related to the monotonicity of Perelman's reduced volume. The Euclidean space $\mathbf{R}^N \times \mathbf{R}^d$ is of course Ricci-flat, and so from Corollary 3.10.3 we know that the Bishop-Gromov reduced volume

$$r_0^{-N-d} \int_{|y|^2+|x|^2 \leq r_0^2} dydx \tag{3.276}$$

is non-decreasing[76] in r_0 (and thus non-decreasing in τ). Repeating all the above computations (but with u and $u^{(N)}$ replaced by 1) we thus heuristically conclude that the quantity

$$\frac{1}{\tau^{d/2}} \int_{\mathbf{R}^d} e^{-|x|^2/4\tau}\, dx \tag{3.277}$$

is also non-decreasing[77] in τ. The quantity (3.10.2) is precisely the *Perelman reduced volume* of Euclidean space $\mathbf{R}^d$ (which we view as a

[76] Of course, being Euclidean, (3.276) is equal to a constant $C_{N,d}$; but let us ignore this fact (which we have already used in our heuristic derivation of (3.269)) for now.

[77] Indeed, this quantity is equal to $(4\pi)^{d/2}$ for all τ.

trivial example of an ancient Ricci flow) at the spacetime origin (0,0) and backwards time parameter τ.

3.10.3. From Ricci flow to Ricci flat manifolds. We have seen how ancient solutions to the heat equation on a Euclidean spacetime can be viewed as (approximately) harmonic functions on an "infinitely high dimensional" Euclidean space. Now we would like to analogously view ancient solutions to a heat equation on a flow $t \mapsto (M, g(t))$ of Riemannian manifolds as harmonic functions on an "infinitely high dimensional" Riemannian manifold, and similarly to view ancient Ricci flows as infinite dimensional infinitely high dimensional Ricci-flat manifolds.

Let's begin with the former task. Starting with an ancient flow $t \mapsto (M, g(t))$ of d-dimensional Riemannian metrics for $t \in (-\infty, 0]$ (which we will not assume to be a Ricci flow just yet) and a large integer N, we can consider the $N + d$-dimesional manifold $M^{(N)} := \mathbf{R}^N \times M = \{(y, x) : y \in \mathbf{R}^N, x \in M\}$. As a first attempt to mimic the situation in the Euclidean case, it is natural to endow $M^{(N)}$ with the Riemannian metric $g^{(N)}$ given by the formula

$$(dg^{(N)})^2 = dy^2 + dg(t)^2 \tag{3.278}$$

where t is given by the formula (3.263). In terms of local coordinates, if we use the indices a, b, c to denote the d indices for the x variable and i, j, k to denote the N indices for the y variable, we have

$$g^{(N)}_{ab} = g_{ab}(t); g^{(N)}_{ai} = g^{(N)}_{ia} = 0; g^{(N)}_{ij} = \delta_{ij} \tag{3.279}$$

where δ is the Kronecker delta. From this we see that the volume measure $d\mu^{(N)}$ on $M^{(N)}$ is given by

$$d\mu^{(N)} = d\mu(t)dy \tag{3.280}$$

and the Dirichlet form

$$\begin{aligned} E^{(N)}(u, v) &:= \int_{M^{(N)}} g^{(N)}(\nabla^{(N)} u, \nabla^{(N)} v)\ d\mu^{(N)} \\ &= -\int_{M^{(N)}} \Delta^{(N)} uv\ d\mu^{(N)} \\ &= -\int_{M^{(N)}} u\Delta^{(N)} v\ d\mu^{(N)} \end{aligned} \tag{3.281}$$

for this Riemannian manifold is given by

(3.282)
$$E^{(N)}(u,v) = \int_{\mathbf{R}^N} \int_{M(t)} \nabla_y u \cdot \nabla_y v + g(t)(\nabla_{x,g(t)} u, \nabla_{x,g(t)} v)\ d\mu(t) dy,$$

where $\nabla_{x,g(t)} u$ is the gradient of u in the x variable using the metric $g(t)$. We can then integrate by parts to compute the Laplacian $\Delta^{(N)} u$. Recalling from (3.55) that $d\mu(t)$ varies in t by the formula

$$\frac{d}{dt} d\mu(t) = \frac{1}{2} \mathrm{tr}(\dot{g}) d\mu(t) \tag{3.283}$$

and using (3.263) and the chain rule, we see that

$$\Delta^{(N)} u^{(N)} = \Delta_y u^{(N)} + \Delta_{x,g(t)} u^{(N)} - \frac{r}{2N} \mathrm{tr}(\dot{g}) \partial_r u^{(N)} \tag{3.284}$$

where $\Delta_{x,g(t)}$ is the Laplace-Beltrami operator in the x variable using the metric $g(t)$. If we specialise to radial functions

$$u^{(N)}(y,x) = u(t,x) \tag{3.285}$$

and use (3.264) and the chain rule, we can rewrite (3.283) as

$$-\frac{t}{2N} \partial_{tt} u - \partial_t u + \Delta_{x,g(t)} u + \frac{t}{N} \mathrm{tr}(\dot{g}) \partial_t u \tag{3.286}$$

Thus we see that if u solves the heat equation $u_t = \Delta_{g(t)} u$, then its lift $u^{(N)} : M^{(N)} \to \mathbf{R}$ is approximately harmonic in the sense that $\Delta^{(N)} u^{(N)} = O(1/N)$ in the region where $-t = \tau = O(1)$ and x is confined to a compact region of space.

Remark 3.10.7. The $\frac{t}{N} \mathrm{tr}(\dot{g}) \partial_t u$ term in (3.286) is somewhat annoying; we will later tweak the metric (3.278) in order to remove it (at the cost of other, more acceptable, terms.

Now let us see whether Ricci flows $t \mapsto (M, g(t))$ lift to approximately Ricci-flat manifolds $M^{(N)}$. We begin by computing the *Christoffel symbols* $(\Gamma^{(N)})^\gamma_{\alpha\beta}$ in local coordinates, where α, β, γ refer to the $N+d$ combined indices coming from the indices a on M and the indices i on $\mathbf{R}^N$. We recall the standard formula

$$\Gamma^\gamma_{\alpha\beta} = \frac{1}{2} g^{\gamma\delta} (\partial_\alpha g_{\beta\delta} + \partial_\beta g_{\alpha\delta} - \partial_\delta g_{\alpha\beta}) \tag{3.287}$$

for the Christoffel symbols of a general Riemannian manifold in local coordinates. Specialising to the metric (3.10.3), some computation

reveals that

$$\begin{aligned}
(\Gamma^{(N)})^i_{jk} &= 0 \\
(\Gamma^{(N)})^i_{ja} = (\Gamma^{(N)})^i_{aj} &= (\Gamma^{(N)})^a_{ij} = 0 \\
(\Gamma^{(N)})^i_{ab} &= \frac{y_i}{2N}\dot{g}_{ab} \\
(\Gamma^{(N)})^a_{ib} &= (\Gamma^{(N)})^a_{bi} = -\frac{y_i}{2N}g^{ac}\dot{g}_{cb} \\
(\Gamma^{(N)})^a_{bc} &= \Gamma^a_{bc}.
\end{aligned} \tag{3.288}$$

Now the Ricci curvature $\mathrm{Ric}_{\alpha\beta}$ can be computed from the Christoffel symbols by the standard formula

$$\mathrm{Ric}_{\alpha\beta} = \partial_\gamma \Gamma^\gamma_{\alpha\beta} - \partial_\beta \Gamma^\gamma_{\alpha\gamma} + \Gamma^\gamma_{\alpha\beta}\Gamma^\mu_{\gamma\mu} - \Gamma^\mu_{\alpha\gamma}\Gamma^\gamma_{\beta\mu}. \tag{3.289}$$

If we apply this formula we obtain (after some computation)

$$\begin{aligned}
\mathrm{Ric}^{(N)}_{ij} &= \frac{\delta_{ij}}{2N}\mathrm{tr}(\dot{g}) + O(1/N^2) \\
\mathrm{Ric}^{(N)}_{ia} &= O(1/N) \\
\mathrm{Ric}^{(N)}_{ab} &= \mathrm{Ric}_{ab} + \frac{1}{2}\dot{g}_{ab} + O(1/N).
\end{aligned} \tag{3.290}$$

We thus see that if the original flow $t \mapsto (M, g(t))$ obeys the Ricci flow equation $\dot{g} = -2\mathrm{Ric}$, then the lifted manifold $(M^{(N)}, \mu^{(N)})$ is nearly Ricci flat in the sense that all components of the Ricci curvature tensor are $O(1/N)$ (in the region $t = O(1)$). In fact the above estimates show that the Ricci curvature tensor is also O(1/N) in the operator norm sense and $O(1/\sqrt{N})$ in the *Hilbert-Schmidt* (or *Frobenius*) sense.

It turns out that this approximation is not quite good enough for applications to Ricci flow, mainly because the $\frac{\delta_{ij}}{2N}\mathrm{tr}(\dot{g}) = -R\delta_{ij}/N$ term in (3.290) gives a significant contribution to the trace of the Ricci tensor $\mathrm{Ric}^{(N)}$ (i.e. the scalar curvature $R^{(N)}$), even in the limit $N \to \infty$. It turns out however that one can eliminate this problem by adding a correction term to the metric (3.278) involving the scalar curvature. More precisely, given an ancient Ricci flow $t \mapsto (M, g(t))$, define the modified metric $\tilde{g}^{(N)}$ by the formula

$$(d\tilde{g}^{(N)})^2 = dy^2 + \frac{r^2}{N^2}R(t)dr^2 + dg(t)^2 \tag{3.291}$$

where of course $dr = \sum_{i=1}^N \frac{y_i}{r} dy_i$ is the derivative of the radial variable r, and $R(t,x)$ is the scalar curvature of $g(t)$ at x. In coordinates, we have

$$\tilde{g}^{(N)}_{ij} = \delta_{ij} + \frac{y_i y_j}{N^2} R(t); \quad \tilde{g}^{(N)}_{ia} = 0; \quad \tilde{g}^{(N)}_{ab} = g_{ab}. \tag{3.292}$$

Exercise 3.10.5. Let $t \mapsto (M, g(t))$ be a smooth ancient Ricci flow on $(-\infty, 0]$, and let $\tilde{g}^{(N)}$ be defined by (3.291). Show that in the region where $y_i = O(1)$ (so $-t = \tau = O(1)$) and x ranges in a compact set, the Christoffel symbols $(\tilde{\Gamma}^{(N)})^\gamma_{\alpha\beta}$ take the form

$$\begin{aligned}
(\tilde{\Gamma}^{(N)})^i_{jk} &= \frac{\delta_{jk}}{N^2} R y_i + O(1/N^3) \\
(\tilde{\Gamma}^{(N)})^i_{ja}, (\tilde{\Gamma}^{(N)})^i_{aj}, (\tilde{\Gamma}^{(N)})^a_{ij} &= O(1/N^2) \\
(\tilde{\Gamma}^{(N)})^i_{ab} &= \frac{y_i}{2N} \dot{g}_{ab} + O(1/N^2) \\
(\tilde{\Gamma}^{(N)})^a_{ib} = (\tilde{\Gamma}^{(N)})^a_{bi} &= -\frac{y_i}{2N} g^{ac} \dot{g}_{cb} + O(1/N^2) \\
(\tilde{\Gamma}^{(N)})^a_{bc} &= \Gamma^a_{bc}.
\end{aligned} \tag{3.293}$$

and the Ricci curvature $\widetilde{\mathrm{Ric}}^{(N)}_{\alpha\beta}$ takes the form

$$\begin{aligned}
\widetilde{\mathrm{Ric}}^{(N)}_{ij} &= O(1/N^2) \\
\widetilde{\mathrm{Ric}}^{(N)}_{ia} &= O(1/N) \\
\widetilde{\mathrm{Ric}}^{(N)}_{ab} &= O(1/N).
\end{aligned} \tag{3.294}$$

In particular, $\widetilde{\mathrm{Ric}}^{(N)}$ has norm $O(1/\sqrt{N})$ in the trace (or *nuclear*) norm (and hence in the Hilbert-Schmidt/Frobenius and operator norms).

Exercise 3.10.6. Let the assumptions and notation be as in Exercise 3.10.5, let $u : (-\infty, 0] \times M \to \mathbf{R}$ be a smooth function, and let $u^{(N)}$ be as in (3.285). Show that the Laplacian $\tilde{\Delta}^{(N)}$ associated to $\tilde{g}^{(N)}$ obeys a similar formula to (3.286), but with the $\frac{r}{2N} \mathrm{tr}(\dot{g}) \partial_r u^{(N)}$ term replaced by terms which are $O(1/N^2)$ when t, x are bounded.

3.10.4. Perelman's reduced length and reduced volume. In the previous discussion, we have converted a Ricci flow $t \mapsto (M, g(t))$ to a Riemannian manifold $(M^{(N)}, \tilde{g}^{(N)})$ of much higher dimension

which is almost Ricci flat. Let us adopt the heuristic that this latter manifold is sufficiently close to being Ricci flat that the Bishop-Gromov inequality (Corollary 3.10.3) holds (at least in the asymptotic limit $N \to \infty$), thus the Bishop-Gromov reduced volume $r_0^{-N-d} B_{\tilde{g}^{(N)}}((0,x_0),r_0)$ should heuristically be non-increasing in r_0, where we fix a spatial origin $x_0 \in M$. In order to exploit the above heuristic, we first need to understand the distance function on $(\tilde{M}, g(t))$. Let $(y_1, x_1) = (r_1\omega_1, x_1)$ be a point in $\tilde{M} = \mathbf{R}^N \times M$, and consider a length-minimising geodesic $\gamma^{(N)} : [0,\tau_1] \to \tilde{M}$ from $(0,x_0)$ to $(r_1\omega_1, x_1)$, where we have normalised the length τ_1 of the parameter interval by the formula $\tau_1 = r_1^2/2N$. Observe that the metric (3.291) can be rewritten in polar coordinates (after substituting $-t = \tau = r^2/2N$) as

$$(d\tilde{g}^{(N)})^2 = (\frac{N}{2\tau} + R)d\tau^2 + 2N\tau d\omega^2 + dg(-\tau)^2 \tag{3.295}$$

(which is essentially the first formula in [**Pe2002**, Section 6]). Note that the angular variable ω only influences the second term in this metric and not the other two. Because of this, one sees that the geodesic $\gamma^{(N)}$ must keep ω constant in order to be length-minimising (i.e. $\omega = \omega_1$ for the duration of the geodesic). Turning next to the τ variable, we then see that for N large enough, the geodesic $\gamma^{(N)}$ should increase τ continuously from 0 to τ_1 (as the $\frac{N}{2\tau}$ term in (3.295) will severely penalise any backtracking. After a reparameterisation we may in fact assume that τ increases at constant speed, thus we have

$$\gamma^{(N)}(\tau) = (\sqrt{2N\tau}\omega_1, \gamma(\tau)) \tag{3.296}$$

for some path $\gamma : [0,\tau_1] \to M$ from x_0 to x_1. Using (3.295), the length of this geodesic is

$$\int_0^{\tau_1} \sqrt{\frac{N}{2\tau} + R + |\gamma'(\tau)|^2_{g(-\tau)}}\, d\tau \tag{3.297}$$

which by Taylor expansion is equal to

$$\sqrt{2N\tau_1} + \frac{1}{\sqrt{2N}}\mathcal{L}(\gamma) + O(N^{-3/2}) \tag{3.298}$$

where the $\mathcal{L}$-*length* of γ is defined as

$$\mathcal{L}(\gamma) := \int_0^{\tau_1} \sqrt{\tau}(R + |\gamma'(\tau)|^2_{g(-\tau)})\, d\tau. \tag{3.299}$$

Note that this quantity is independent of N. Thus, heuristically, geodesics in $\mathcal{M}$ from $(0, x_0)$ to $(r_1\omega_1, x_1)$ should (approximately) minimise the $\mathcal{L}$-length. If we define $L_{(0,x_0)}(-\tau_1, x_1)$ to be the infimum of $\mathcal{L}(\gamma)$ over all paths $\gamma : [0, \tau_1] \to M$ from x_0 to x_1, we thus obtain the heuristic approximation

$$d_{\tilde{g}^{(N)}}((0, x_0), (r_1\omega_1, x_1)) = \sqrt{2N\tau_1} + \frac{1}{\sqrt{2N}} L_{(0,x_0)}(-\tau_1, x_1). \tag{3.300}$$

Exercise 3.10.7. When M is the Euclidean space $\mathbf{R}^d$ (with the trivial Ricci flow, of course), show that $L_{(0,x_0)}(-\tau_1, x_1) = |x_1 - x_0|^2/2\sqrt{\tau_1}$, and the minimiser is given by $\gamma(\tau) = x_0 + \sqrt{\frac{\tau}{\tau_1}}(x_1 - x_0)$.

From (3.300) we see that the ball in $(M^{(N)}, \tilde{g}^{(N)})$ of radius $r_0 = \sqrt{2N\tau_0}$ centred at $(0, x_0)$ (where, as always, we are in the regime $\tau_0 = O(1)$, so $r_0^2 = O(N)$) should heuristically take the form

$$\{(r_1\omega_1, x_1) : L_{(0,x_0)}(-\tau_1, x_1) \leq 2N(\sqrt{\tau_0} - \sqrt{\tau_1})\}. \tag{3.301}$$

If we make the plausible assumption that $L_{(0,x_0)}(-\tau, x)$ varies smoothly in τ, then (3.301) is heuristically close (when N is large) to

$$\{(r_1\omega_1, x_1) : L_{(0,x_0)}(-\tau_0, x_1) \leq 2N(\sqrt{\tau_0} - \sqrt{\tau_1})\} \tag{3.302}$$

or equivalently

$$\{(r_1\omega_1, x_1) : r_1 \leq r_0 - \sqrt{2N} L_{(0,x_0)}(-\tau_0, x_1)\}. \tag{3.303}$$

Now, the volume measure of (3.291) is of the form $(1+O(1/N))dyd\mu(t)$, and so the volume of (3.303) is approximately

$$C_N \int_M \int_0^{r_0 - \sqrt{2N} L_{(0,x_0)}(-\tau_0, x_1)} r_1^{N-1}\, dr_1 d\mu(t_1)(x_1). \tag{3.304}$$

(Note there is a slight abuse of notation since t_1 depends on r_1, but it will soon be clear that this abuse is harmless.) When N is large, the inner integral is dominated by its right endpoint as before, and so (3.304) is approximately

$$\frac{1}{N} C_N \int_M (r_0 - \sqrt{2N} L_{(0,x_0)}(-\tau_0, x))^N d\mu(t_0)(x). \tag{3.305}$$

We can Taylor expand this to be approximately

$$\frac{1}{N} C_N r_0^N \int_M \exp(-l_{(0,x_0)}(-\tau_0, x))\, d\mu(t_0)(x) \tag{3.306}$$

where the *Perelman reduced length* $l_{(0,x_0)}(-\tau_0, x)$ is defined as

$$l_{(0,x_0)}(-\tau, x) := \frac{L_{(0,x_0)}(-\tau, x)}{2\sqrt{\tau}} = \frac{\sqrt{2N} L_{(0,x_0)}(-\tau, x)}{2r_0} \tag{3.307}$$

Example 3.10.8. Continuing the Euclidean example of Exercise 3.10.7, we have $l_{(0,x_0)}(-\tau, x) = |x - x_0|^2/4\tau$, which is the familiar exponent in the fundamental solution (3.269). This is, of course, not a coincidence.

From (3.306) we thus heuristically conclude that the Bishop-Gromov reduced volume of $(M^{(N)}, \tilde{g}^{(N)})$ at $(0, x_0)$ and at radius $r_0 = \sqrt{2N\tau_0}$ is approximately equal to a constant multiple of $\tilde{V}_{(0,x_0)}(-\tau)$, where the *Perelman reduced volume* $\tilde{V}_{(0,x_0)}(-\tau)$ is defined as

$$\tilde{V}_{(0,x_0)}(-\tau) := \int_M \tau^{-d/2} \exp(-l_{(0,x_0)}(-\tau, x))\, d\mu(-\tau)(x). \tag{3.308}$$

Example 3.10.9. Again continuing the Euclidean example, the reduced volume in Euclidean space (with the trivial Ricci flow) is always $(4\pi)^{d/2}$.

Formally applying Corollary 3.10.3, we are thus led to

Conjecture 3.10.10 (Monotonicity of Perelman reduced volume)**.** *Let $t \mapsto (M, g(t))$ be a Ricci flow on $[-T, 0]$, and let $x_0 \in M_0$. Then the quantity $\tilde{V}_{(0,x_0)}(-\tau)$ for $0 < \tau \leq T$ is monotone non-increasing in τ.*

Remark 3.10.11. Note here we are not taking the Ricci flow to be ancient; this would correspond to the manifold $M^{(N)}$ being replaced by an incomplete manifold, of radius about $\sqrt{2NT}$. However, because of the restriction $\tau \leq T$, the above heuristic arguments never "encounter" the lack of completeness, and so it is reasonable to expect that the conjecture will continue to hold in the non-ancient case. This is of course an essential point for our applications, since the Ricci flows we study are not assumed to be ancient.

Remark 3.10.12. At an crude heuristic level, the Perelman reduced volume $\tilde{V}_{(0,x_0)}(-\tau)$ is roughly like $\mathrm{Vol}_{g(-\tau)}(-\tau, O(\sqrt{\tau}))/\tau^{d/2}$ (since, in view of Exercise 3.10.7, we expect $l_{(0,x_0)}(-\tau, x)$ to behave like $d(x_0, x)^2/\tau$, especially in regions of bounded normalised curvature, where we are deliberately vague about exactly what metric we using to define d). This heuristic suggests that Conjecture 3.10.10 should be able to establish the non-collapsing result we want (Theorem 3.8.15). This will be made more rigorous in subsequent lectures. For now, we observe that the Perelman reduced length and reduced volume are dimensionless (just as the Bishop-Gromov reduced volume is), which as discussed in Section 3.8 is basically a necessary condition in order for this quantity to force non-collapsing of the geometry.

A rigorous proof of Conjecture 3.10.10 that follows the above high-dimensional comparison geometry heuristic argument was only recently obtained[78], in [**CaTo2008**]. Nevertheless, it is possible to prove Conjecture 3.10.10 by other means, and in particular by developing parabolic analogues of all the comparison geometry machinery that is used to prove the Bishop-Gromov inequality (and in particular, developing a theory of $\mathcal{L}$-geodesics analogous to the "elliptic" theory of geodesics on a Riemannian manifold. This will be the focus of the next few sections.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/04/20`. Pedro Lauridsen Ribiero pointed out the intriguing similarity between the ideas of treating parabolic equations as the high-dimensional limit of an elliptic equation, and treating the fundamental solution to a parabolic equation as the scaling limit of random walks.

3.11. Variation of L-geodesics, and monotonicity of Perelman reduced volume

Having completed a heuristic derivation of the monotonicity of Perelman reduced volume (Conjecture 3.10.10), we now turn to a rigorous proof. Whereas in Section 3.10 we derived this monotonicity

[78] For further results in the direction of formalising the dictionary between elliptic and parabolic equations, see [**Pe2002**, Section 6], [**CaZh2006**, Section 3.1], [**ChCh1995**], [**ChCh1996**].

by converting a parabolic spacetime to a high-dimensional Riemannian manifold, and then formally applying tools such as the Bishop-Gromov inequality (Corollary 3.10.3) to that setting, our approach here shall take the opposite tack, finding parabolic analogues of the *proof* of the elliptic Bishop-Gromov inequality, in particular obtaining analogues of the classical first and second variation formulae for geodesics, in which the notion of length is replaced by the notion of $\mathcal{L}$-length introduced in Section 3.10. The material here is primarily based on [**Pe2002**], [**Mu2006**], but detailed treatments also appear in [**Ye2008**], [**KlLo2006**], [**MoTi2007**], [**CaZh2006**].

3.11.1. Reduction to a pointwise inequality. Recall that the Bishop-Gromov inequality (Corollary 3.10.3) states (among other things) that if a d-dimensional complete Riemannian manifold (M, g) is Ricci-flat (or more generally, has non-negative Ricci curvature), and x_0 is any point in M, then the Bishop-Gromov reduced volume $\mathrm{Vol}(B(x_0, r))/r^d$ is a non-increasing function of r. In fact one can obtain the slightly sharper result that $\mathrm{Area}(S(x_0, r))/r^{d-1}$ is a non-increasing function of r, where $S(x_0, r)$ is the sphere of radius r centred at x_0. From the basic formula $\mathcal{L}_{\partial r} d\mu = (\Delta r)\ d\mu$ (see (3.255)) and the Gauss lemma (Lemma 3.8.4), one readily obtains the identity

$$\frac{d}{dr}\mathrm{Area}(S(x_0, r)) = \int_{S(x_0,r)} \Delta r\ dS \tag{3.309}$$

where dS is the area element. The monotonicity of $\mathrm{Area}(S(x_0, r))/r^{d-1}$ then follows (formally, at least) from the pointwise inequality

$$\Delta r \leq \frac{d-1}{r} \tag{3.310}$$

which we will derive shortly (at least for the portion of the manifold inside the cut locus) as a consequence of the first and second variation formulae for geodesics[79]. Observe that (3.310) is an equality when (M, g) is a Euclidean space $\mathbf{R}^d$. It turns out that the monotonicity of Perelman reduced volume for Ricci flows can similarly be reduced to a pointwise inequality, in which the Laplacian Δ is replaced by a heat operator, and the radial variable r is replaced by the

[79] In Section 3.10, the inequality (3.310) was derived from a transport inequality for Δr, but we will take a slightly different tack here.

Perelman reduced length. More precisely, given an ancient Ricci flow $t \mapsto (M, g(t))$ for $t \in (-\infty, 0]$, a time $-\tau$, and two points $x_0, x \in M$, recall that the reduced length $l_{(0,x_0)}(-\tau, x)$ is defined as

$$l_{(0,x_0)}(-\tau, x) := \frac{1}{2\sqrt{\tau}} \inf_{\gamma} \mathcal{L}(\gamma) \tag{3.311}$$

where the $\mathcal{L}$-*length* $\mathcal{L}(\gamma)$ of a curve $\gamma : [0, \tau_1] \to M$ from x_0 to x_1 is defined as

$$\mathcal{L}(\gamma) = \int_0^{\tau_1} \sqrt{\tau}(R + |X|^2_{g(-\tau)})\, d\tau, \tag{3.312}$$

where we adopt the shorthand $X := \partial_\tau \gamma$, and that Conjecture 3.10.10 asserts that the Perelman reduced volume

$$\tilde{V}_{(0,x_0)}(-\tau) = \int_M \tau^{-d/2} \exp(-l_{(0,x_0)}(-\tau, x))\, d\mu_{g(-\tau)}(x) \tag{3.313}$$

is non-increasing in τ for Ricci flows. If we differentiate (3.313) in τ, using the variation formula $\frac{d}{d\tau} d\mu = R\, d\mu$, we easily verify that the monotonicity of (3.313) will follow (assuming $l_{(0,x_0)}$ is sufficiently smooth, and that either M is compact, or $l_{(0,x_0)}$ grows sufficiently quickly at infinity) from the pointwise inequality

$$\partial_\tau l_{(0,x_0)} - \Delta_{g(-\tau)} l_{(0,x_0)} + |\nabla l|^2_{g(-\tau)} - R + \frac{d}{2\tau} \geq 0 \tag{3.314}$$

which should be viewed as a parabolic analogue to (3.310).

Exercise 3.11.1. Verify that (3.314) is an equality in the case of the (trivial) Ricci flow on Euclidean space, using Example 3.10.8. (This is of course consistent with Example 3.10.9.)

Exercise 3.11.2. Show that (3.314) is equivalent to the assertion[80] that the function $v(-\tau, x) := (4\pi\tau)^{-d/2} \exp(-l_{(0,x_0)}(-\tau, x))$ is a subsolution of the adjoint heat equation, or more precisely that $\partial_t v - \Delta v + Rv \leq 0$. Note that this fact implies the monotonicity of Perelman reduced volume (cf. Exercise 3.9.2).

[80] It seems that the elliptic analogue of this fact is the assertion that the Newton-type potential $1/r^{d-2}$ is subharmonic away from the origin for Ricci flat manifolds of dimension three or larger , which is a claim which is easily seen to be equivalent to (3.310) thanks to the Gauss lemma, Lemma 3.8.4.

So to prove monotonicity of the Perelman reduced volume, the main task[81] will be to establish the pointwise inequality (3.314).

We will perform a minor simplification: by using the rescaling symmetry $g(t,x) \mapsto \lambda^2 g(\frac{t}{\lambda^2})$ (and noting the unsurprising fact that (3.314) is dimensionally consistent) we can normalise $\tau_1 = 1$.

3.11.2. First and second variation formulae for $\mathcal{L}$-geodesics. To establish (3.314), we of course need some *variation formulae* that compute the first and second derivatives of the reduced length function $l_{(0,x_0)}$. To motivate these formulae, let us first recall the more classical variation formulae that give the first and second derivatives of the metric function $d(x_0,x)$ on a Riemannian manifold (M,g), which in particular can be used to derive (3.310) when the Ricci curvature is non-negative.

We recall that the distance $d(x_0,x)$ can be defined by the energy-minimisation formula

$$\frac{1}{2}d(x_0,x)^2 = \inf_\gamma E(\gamma) \tag{3.315}$$

where $\gamma : [0,1] \to M$ ranges over all C^1 curves from x_0 to x, and the *Dirichlet energy* $E(\gamma)$ of the curve is given by the formula

$$E(\gamma) = \frac{1}{2}\int_0^1 |X|_g^2 \, dt \tag{3.316}$$

where we write $X := \partial_t \gamma$. It is known that this infimum is always attained by some geodesic γ; we shall assume this implicitly in the computations which follow.

Now suppose that we deform such a curve γ with respect to a real parameter $s \in (-\varepsilon,\varepsilon)$, thus $\gamma : (s,t) \mapsto \gamma(s,t)$ is now a function on the two-dimensional parameter space $(\varepsilon,\varepsilon)\times[0,1]$. The first variation here can be computed as

$$\frac{d}{ds}E(\gamma) = \int_0^1 g(\nabla_X X, X)\, dt \tag{3.317}$$

where ∇_X is the pullback of the Levi-Civita connection on M with respect to γ applied in the direction ∂_t; here we of course use that g

[81] There are some additional technical issues, mainly concerning the parabolic counterpart of the cut locus, which we will also have to address, but we will work formally for now, and deal with these analytical matters later.

is parallel with respect to this connection. The torsion-free nature of this connection gives us the identity

$$\nabla_Y X = \nabla_X Y \tag{3.318}$$

where $Y = \partial_s \gamma$ is the infinitesimal variation, and ∇_Y is the pullback of the Levi-Civita connection applied in the direction ∂_s (cf. Exercise 3.7.5). An integration by parts (again using the parallel nature of g) then gives the first variation formula

$$\frac{d}{ds} E(\gamma) = g(Y, X)|_{t=0}^1 - \int_0^1 g(Y, \nabla_X X)\, dt. \tag{3.319}$$

If we fix the endpoints of γ to be $\gamma(s,0) = x_0$ and $\gamma(s,1) = x_1$, then the first term on the right-hand side of (3.319) vanishes. If we consider arbitrary infinitesimal variations Y of γ with fixed endpoints, we thus conclude that in order to be a minimiser for (3.315), that γ must obey the *geodesic flow equation*

$$\nabla_X X = 0. \tag{3.320}$$

One consequence of this is that the speed $|X|_g$ of such a minimiser must be constant, and from (3.315) we then conclude

$$|X|_g = d(x_0, x). \tag{3.321}$$

If we then vary a geodesic $\gamma(0, \cdot)$ with the initial endpoint $\gamma(s,0)$ fixed at x_0 and the final endpoint $\gamma(s,1) = x(s)$ variable, the variation formula (3.319) gives

$$\frac{d}{ds} E(\gamma) = g(x'(s), X(s,1)) \tag{3.322}$$

which, if we insert this back into (3.315) and use (3.321), gives

$$\frac{d}{ds} d(x_0, x) \leq g(x'(s), X(s,1)/|X(s,1)|_g) \tag{3.323}$$

which is a (one-sided) version of the Gauss lemma (Lemma 3.8.4). If one is inside the *cut locus*, then the metric function is smooth, and one can then replace the inequality with an equality by considering variations both forwards and backwards in the s variable, recovering the full Gauss lemma. In particular, we conclude in this case that $\nabla d(x_0, x)$ is a unit vector.

Now we consider the second variation $\frac{d^2}{ds^2}E(\gamma)$ of the energy, when γ is already a geodesic. For simplicity we assume that γ evolves geodesically in the s direction, thus[82]

$$\nabla_Y Y = 0. \tag{3.324}$$

Differentiating (3.317) once more we obtain

$$\frac{d^2}{ds^2}E(\gamma) = \int_0^1 g(\nabla_Y \nabla_Y X, X) + |\nabla_Y X|_g^2 \, dt. \tag{3.325}$$

Using (3.318), (3.324), and the definition of curvature, we have

$$\begin{aligned} \nabla_Y \nabla_Y X &= \nabla_Y \nabla_X Y \\ &= \nabla_X \nabla_Y Y + \mathrm{Riem}(Y, X)Y \\ &= -\mathrm{Riem}(X, Y)Y \end{aligned} \tag{3.326}$$

and thus (by one further application of (3.318))

$$\frac{d^2}{ds^2}E(\gamma) = \int_0^1 |\nabla_X Y|_g^2 - g(\mathrm{Riem}(X, Y)Y, X) \, dt. \tag{3.327}$$

Now let us fix the initial endpoint $\gamma(s, 0) = x_0$ and let the other endpoint $\gamma(s, 1) = x(s)$ vary, thus $\partial_s \gamma$ equals 0 at time $t = 0$ and equals $x'(s)$ at time $t = 1$. From Cauchy-Schwarz we conclude

$$\int_0^1 |\nabla_X Y|_g^2 \, dt \leq \int_0^1 (\partial_t |Y|_g)^2 \, dt \leq (\int_0^1 \partial_t |Y|_g \, dt)^2 = |x'(s)|^2. \tag{3.328}$$

Actually, we can attain equality here by choosing the vector field Y appropriately:

Exercise 3.11.3. If we set $Y := tv$, where v is the parallel transport of $x'(s)$ along X, or more precisely the vector field that solves the ODE

$$\nabla_X v = 0; v(s, 1) = x'(s) \tag{3.329}$$

show that all the inequalities in (3.328) are obeyed with equality.

[82] Actually, since γ is already a geodesic and thus is stationary with respect to perturbations that respect the endpoints, the values of $\nabla_Y Y$ away from endpoints - which represents a second-order perturbation respecting the endpoints - will have no ultimate effect on the second variation of $E(\gamma)$. Nevertheless it is convenient to assume (3.324) to avoid a few routine additional calculations.

For such a vector field, we conclude that

$$\frac{d^2}{ds^2}E(\gamma) = |x'(s)|^2 - \int_0^1 -t^2 g(\mathrm{Riem}(X,v)v,X)\ dt. \tag{3.330}$$

From this formula (and the first variation formula) we conclude that

$$\frac{d^2}{ds^2}\frac{1}{2}d(x_0,x)^2 \leq |x'(s)|^2 - \int_0^1 t^2 g(\mathrm{Riem}(X,v)v,X)\ dt. \tag{3.331}$$

Now let $x'(0)$ vary over an orthonormal basis of the tangent space of $x(0)$; by (3.329) we see that v determines an orthonormal frame for $s = 0$ and $0 \leq t \leq 1$. Summing (3.331) over this basis (and using the formula for the Laplacian in normal coordinates) we conclude that

$$\Delta\frac{1}{2}d(x_0,x)^2 \leq d - \int_0^1 t^2 \mathrm{Ric}(X,X)\ dt. \tag{3.332}$$

In particular, for manifolds of non-negative Ricci curvature we have

$$\Delta\frac{1}{2}d(x_0,x)^2 \leq d \tag{3.333}$$

from which (3.310) easily follows from the Gauss lemma, Lemma 3.8.4. (Observe that (3.333) is obeyed with equality in the Euclidean case.)

Now we develop analogous variational formulae for $\mathcal{L}$-length (and reduced length) on a Ricci flow. We shall work formally for now, assuming that all infima are actually attained and that all quantities are as smooth as necessary for the analysis that follows to work; we then discuss later how to justify all of these assumptions. As mentioned earlier, we normalise $\tau_1 = 1$.

Let us take a path $\gamma : [0,1] \to M$ and vary it with respect to some additional parameter s as before. Differentiating (3.312), we obtain

$$\frac{d}{ds}\mathcal{L}(\gamma) = \int_0^1 \sqrt{\tau}(\nabla_Y R + 2g(X, \nabla_Y X))\ d\tau \tag{3.334}$$

where $X := \partial_\tau \gamma$ and $Y := \partial_s \gamma$. On the other hand, if we have a Ricci flow $\partial_\tau g = 2\mathrm{Ric}$, we see that

$$\partial_\tau g(X,Y) = g(\nabla_X X, Y) + g(X, \nabla_X Y) + 2\mathrm{Ric}(X,Y); \tag{3.335}$$

placing this into (3.334) and using the fundamental theorem of calculus, we can express the right-hand side of (3.334) as

$$2g(X,Y)(\tau_1) - 2\int_0^1 \sqrt{\tau} g(Y, G(X))\, d\tau \tag{3.336}$$

where $G(X)$ is the vector field

$$G := \nabla_X X - \frac{1}{2}\nabla R + \frac{1}{2\tau} X + 2\mathrm{Ric}(X,\cdot)^*. \tag{3.337}$$

Here $\mathrm{Ric}(X,\cdot)^*$ is the vector field $(\mathrm{Ric}(X,\cdot)^*)^\alpha = g^{\alpha\beta}\mathrm{Ric}_{\gamma\beta}X^\gamma$, or equivalently it is the vector field Z such that $\mathrm{Ric}(X,W) = g(Z,W)$ for all vector fields W.

Note that G does not depend on Y. From this we see that in order for γ to be a minimiser of $\mathcal{L}(\gamma)$ with the endpoints fixed, we must have $G(X) = 0$, which is the parabolic analogue of the geodesic flow equation (3.324).

Example 3.11.1. In the case of the trivial Euclidean flow, the minimal $\mathcal{L}$-path from $(0,x_0)$ to $(-1,x_1)$ takes the form $\gamma(\tau) = x_0 + v\sqrt{\tau}$ where $v := x_1 - x_0$, in which case $X = \frac{v}{2\sqrt{\tau}}$. It is not hard to verify that $G = 0$ in this case.

Arguing as in the elliptic case, we conclude (assuming the existence of a unique minimiser, and the local smoothness of reduced length) the first variation formula

$$\partial_s l_{(0,x_0)}(-1,x_1) = g(X, \partial_s x_1)(1) \tag{3.338}$$

or equivalently

$$\nabla l_{(0,x_0)}(-1,x_1) = X(1). \tag{3.339}$$

Example 3.11.2. Continuing Example 3.10.13, note that $l_{(0,x_0)}(-1,x_1) = |x_1 - x_0|^2/4$ and $\partial_\tau \gamma = (x_1 - x_0)/2$, which is of course consistent with (3.338).

Having computed the spatial derivative of the reduced length, we turn to the time derivative. The simplest way to compute this is to observe that any partial segment of an $\mathcal{L}$-minimising path must again

be a $\mathcal{L}$-minimising path. From (3.312) and the fundamental theorem of calculus we have

$$\frac{d}{d\tau_1}\mathcal{L}(\gamma)|_{\tau_1=1} = R + |X|_g^2 \tag{3.340}$$

where we vary γ in τ_1 by truncation; by (3.311) and the above discussion we conclude

$$\frac{d}{d\tau_1}(2\sqrt{\tau_1} l_{(0,x_0)}(-\tau_1, x_1))|_{\tau_1=1} = (R + |X|_g^2) \tag{3.341}$$

where (τ_1, x_1) varies along γ (in particular, $\partial_{\tau_1} x_1 = X$). Applying the product and chain rules, we can expand the left-hand side of (3.341) as

(3.342)
$$l_{(0,x_0)}(-1, x_1) + 2\partial_{\tau_1} l_{(0,x_0)}(-\tau_1, x_1)|_{\tau_1=-1} + 2g(\nabla l_{(0,x_0)}(-1, x_1), X);$$

using (3.339), we conclude that

(3.343)
$$\partial_{\tau_1} l_{(0,x_0)}(-\tau_1, x_1)|_{\tau_1=1} = \frac{1}{2}(R + |X|_g^2) - \frac{1}{2} l_{(0,x_0)}(-1, x_1) - |X|_g^2.$$

Now we turn to the second spatial variation of the reduced length. Let γ be a $\mathcal{L}$-minimiser, so that $G = 0$. Differentiating (3.334) again, we obtain

(3.344)
$$\frac{d^2}{ds^2}\mathcal{L}(\gamma) = \int_0^1 \sqrt{\tau}(\nabla_Y \nabla_Y R + 2|\nabla_Y X|^2 + 2g(X, \nabla_Y \nabla_Y X))\, d\tau.$$

As in the elliptic case, it is convenient to assume that we have a geodesic variation (3.324). In that case, we again have (3.326), and we also have $\nabla_Y \nabla_Y R = \mathrm{Hess}(R)(Y, Y)$. Using (3.318), we thus express (3.344) as

$$\int_0^1 \sqrt{\tau}(\mathrm{Hess}(R)(Y, Y) + 2|\nabla_X Y|^2 - 2g(\mathrm{Riem}(X, Y)Y, X))\, d\tau. \tag{3.345}$$

As before, we optimise this in Y. Because the metric g now changes in time by Ricci flow, one has to modify the prescription in Exercise 3.10.10 slightly. More precisely, we now set $Y := \sqrt{\tau} v$, where v solves the following variant of (3.329),

$$\nabla_X v = -\mathrm{Ric}(v, \cdot)^*; v(s, 1) = x'(s). \tag{3.346}$$

The point of doing this is that the ODE is orthogonal; the length of v is preserved along X, as is the inner product between any two such v's (cf. (3.4.2)). A brief computation then shows that

$$\nabla_X Y = \frac{1}{2\sqrt{\tau}} v - \sqrt{\tau}\mathrm{Ric}(v,\cdot)^* \tag{3.347}$$

and hence

$$|\nabla_X Y|_g^2 = \frac{1}{4\tau}|x'(s)|_g^2 + \tau|\mathrm{Ric}(v,\cdot)|^2 - \mathrm{Ric}(v,v). \tag{3.348}$$

Putting all of this into (3.345), we now see that the second variation (3.344) is equal to

$$\begin{aligned}\int_0^1 \tau^{3/2}\mathrm{Hess}(R)(v,v) + \frac{1}{2\tau^{1/2}}|x'(s)|_g^2 + 2\tau^{3/2}|\mathrm{Ric}(v,\cdot)|^2 \\ - 2\tau^{1/2}\mathrm{Ric}(v,v) - 2\tau^{3/2} g(\mathrm{Riem}(X,v)v,X)\, d\tau.\end{aligned} \tag{3.349}$$

We now let $x'(0)$ range over an orthonormal basis of $x(0)$, which leads to v being an orthonormal frame at every point (0,t). Summing over (3.349) and also using (3.311), we conclude that

(3.350)

$$\Delta l_{(0,x_0)}(-1,x_1) \leq \int_0^1 \frac{\tau^{3/2}}{2}\Delta R + \frac{d}{4\tau^{1/2}} + \tau^{3/2}|\mathrm{Ric}|_g^2 - \tau^{1/2}R - \tau^{3/2}\mathrm{Ric}(X,X)\, d\tau.$$

Now we simplify the right-hand side of (3.350). The second term is of course elementary:

$$\int_0^1 \frac{d}{4\tau^{1/2}}\, d\tau = \frac{d}{2} \tag{3.351}$$

and this is consistent with the Euclidean case (in which $\Delta l_{(0,x_0)}$ is exactly $\frac{d}{2}$ when $\tau_1 = 1$, and all curvature terms vanish). To simplify the remaining terms, we recall the variation formula

$$-\partial_\tau R = \Delta R + 2|\mathrm{Ric}|_g^2 \tag{3.352}$$

for the scalar curvature (see (3.2.3)); by the chain rule, we thus have the total derivative formula

$$\frac{d}{d\tau}R = -\Delta R - 2|\mathrm{Ric}|_g^2 + \nabla_X R. \tag{3.353}$$

Inserting (3.351), (3.353) into (3.350) and integrating by parts, we express the right-hand side of (3.350) as

$$\frac{d}{2} - \frac{1}{2}R + \int_0^1 \frac{\tau^{3/2}}{2}\nabla_X R - \frac{\tau^{1/2}}{4}R - \tau^{3/2}\mathrm{Ric}(X,X)\ d\tau. \tag{3.354}$$

To simplify this further, recall that the quantity G defined in (3.337) vanishes. This (and the fact that g evolves by Ricci flow $\partial_\tau g = 2\mathrm{Ric}$) allows one to compute the variation of $\tau|X|_g^2$:

$$\partial_\tau(\tau|X|_g^2) = \tau\partial_X R - 2\tau\mathrm{Ric}(X,X). \tag{3.355}$$

Inserting this into (3.354) and integrating by parts, one can rewrite (3.354) as

$$\frac{d}{2} - \frac{1}{2}R + \frac{1}{2}|X|_g^2 - \frac{1}{4}\int_0^1 \sqrt{\tau}(R + |X|^2)\ d\tau \tag{3.356}$$

and so by (3.311) we obtain the inequality

$$\Delta l_{(0,x_0)}(-1,x_1) \leq \frac{d}{2} - \frac{1}{2}R + \frac{1}{2}|X|_g^2 - \frac{1}{2}l_{(0,x_0)}(-1,x_1). \tag{3.357}$$

Combining (3.339), (3.343), and (3.357) we obtain (3.314) as desired.

3.11.3. Analytical issues. We now discuss in broad terms the analytical issues that one must address in order to make the above arguments rigorous. We first review the classical elliptic theory (i.e. the theory of geodesics in a Riemannian manifold) before turning to Perelman's parabolic theory of $\mathcal{L}$-geodesics in a flow of Riemannian metrics.

In a complete Riemannian manifold, a geodesic $\gamma : [0,1] \to M$ from a fixed point $\gamma(0) = x_0$ to some other point $\gamma(1) = x_1$ has a well-defined initial velocity vector $\gamma'(0) = X(0)$, and conversely each initial velocity vector $v = X(0) \in T_{x_0}M$ determines a unique geodesic with an endpoint $x_1 = \exp_{x_0}(v)$, thus defining the *exponential map* based at x_0. One can show (from standard ODE theory) that this exponential map is smooth (with the derivative of this map controlled by *Jacobi fields*). Also, if M is connected, then any two points can be joined by a geodesic, and the exponential map is onto. However, there can be vectors v for which this map degenerates (i.e. its derivative ceases to be invertible) - these correspond to the *conjugate points* of x_0 in M.

Define the *injectivity region* of x_0 to be the set of all x_1 for which there is a unique minimising geodesic from x_0 to x_1, and that the exponential map is not degenerate along this geodesic (in particular, x_0 and x_1 are not conjugate points). An analysis of Jacobi fields reveals that the injectivity region is open, that the distance function is smooth in this region (except at the origin), and that all the computations given above for the distance function can be justified. So it remains to understand what happens on the complement of the injectivity region, known as the *cut locus*. Points on the cut locus are either conjugate points to x_0, or are else places where minimising geodesics are not unique, which (by a variant of the Gauss lemma) forces the distance function to be non-differentiable at these points. The former type of points form a set of measure zero, thanks to *Sard's theorem*, whereas the latter set of points also form a set of measure zero, thanks to *Radamacher's differentiation theorem* and the Lipschitz nature of the distance function (i.e. the triangle inequality). Thus the injectivity region has full measure. While this does mean that pointwise inequalities such as (3.310) now hold almost everywhere, this is unfortunately not quite enough[83] to ensure that (3.310) holds in the sense of distributions, which is what one really needs in order to fully justify results such as the Bishop-Gromov inequality. Fortunately, one can address this technical issue by constructing barrier functions to the radius function r at every point x_1, i.e. C^2 functions $u = u_\varepsilon$ for each $\varepsilon > 0$ which upper bound r near x_1 (and match r exactly at x_1, and which obeys the inequality (3.310) at x_1 up to a loss of ε. Such functions can be constructed at any x_1, even those in the cut locus, by perturbing the origin x_0 by an epsilon, and then one can use these barrier functions[84] to justify (3.310) in the sense of distributions.

[83] Indeed, by considering simple examples such as the unit circle, we see that the distribution Δr can in fact contain some negative singular measures, although one should note that this does not actually contradict (3.310) due to the favourable sign of these singular components.

[84] As far as I can tell, these arguments controlling the distance function outside of the injectivity region originate with a paper of Calabi[**Ca1958**].

From this one can rigorously justify the Bishop-Gromov inequality for all radii, even those exceeding the radius of injectivity. Analogues of the above assertions hold for the monotonicity of Perelman reduced volume on flows on compact Ricci flows (and more generally for Ricci flows of complete manifolds of bounded curvature). For instance, one can show (using compactness arguments in various weighted Sobolev spaces) that, as long as the manifold M is connected[85], a minimiser to (3.311) always exists, and is attained by an $\mathcal{L}$-geodesic (defined as a curve $\gamma : [0, \tau_1] \to M$ for which the G quantity defined in (3.337) vanishes). Such geodesics turn out to have a well-defined "initial velocity" $v := \lim_{\tau \to 0} \sqrt{\tau} X(\tau)$, as can be seen by working out the ODE for the quantity $\sqrt{\tau} X(\tau)$ (it is also convenient to reparameterise in terms of the variable $r := \sqrt{\tau}$ to remove any apparent singularity at $\tau = 0$). This leads to an $\mathcal{L}$-exponential map $\mathcal{L}\exp_{(0,x_0),\tau_1} : T_{x_0} M \to M$ for any fixed time $-\tau_1$, which is smooth. The derivative of this map is controlled by $\mathcal{L}$-Jacobi fields, which are close analogues of their elliptic counterparts, and which lead to the notion of a $\mathcal{L}$-conjugate point x_1 to $(0, x_0)$ at the fixed time $-\tau_1$. One can then define the injectivity domain and cut locus as before (again for a fixed time $-\tau_1$), and show as before that the former region has full measure. This lets one rigorously derive (3.314) almost everywhere (especially after noting that any segment of a minimising $\mathcal{L}$-geodesic without conjugate points is again a minimising $\mathcal{L}$-geodesic without conjugate points, thus establishing that the injectivity region is in some sense "star-shaped"), but again one needs to justify (3.314) in the sense of distributions in order to derive the monotonicity of Perelman reduced volume. This can again be done by use of barrier functions, perturbing the base point $(0, x_0)$ both spatially and also backwards in time by an epsilon. The details of this become rather technical; see for instance [**Ye2008**], [**KlLo2006**], for details.

Thus far we have only discussed how reduced length and reduced volume behave on smooth Ricci flows of compact manifolds. Of course, to fully establish the global existence of Ricci flow with surgery, one also needs to build an analogous theory for Ricci flows with surgery. Here there turns out to be significant new technical

[85] One can easily reduce to the connected case, since the reduced length is clearly infinite when x_0 and x_1 lie on distinct connected components.

difficulties, basically because one has to restrict attention to paths γ which avoid all regions in which surgery is taking place. This creates some "holes" in the region of integration for the reduced volume, as in some cases the minimising path between two points in spacetime goes through a surgery region. Fortunately it turns out that (very roughly speaking) these holes only occur when the reduced length (or a somewhat technical modification thereof) is rather large, which means that the holes do not significantly impact lower bounds on this reduced volume, which is what is needed to establish κ-noncollapsing. We will discuss these points in Section ???.

In order to control ancient κ-noncollapsing solutions, which are complete but not necessarily compact, one also needs to extend the above theory to complete non-compact manifolds. It turns out that this can be done as long as one has uniform bounds on curvature; a key task here is to establish that the reduced length $l_{(0,x_0)}(-\tau_1, x_1)$ behaves roughly like $d(x_0, x_1)^2/4\tau_1$ (which is basically what it is in the Euclidean case) as x_1 goes to infinity, which allows the integrand in the definition of reduced volume to have enough decay to justify all computations. The technical details here can be found in several places [**Ye2008**], [**KlLo2006**], [**MoTi2007**], [**CaZh2006**].

Remark 3.11.3. A theory analogous to Perelman's theory above was worked out earlier by Li and Yau[**LiYa1986**], but with the Ricci flow replaced by a static manifold with a lower bound on Ricci curvature, and with a time-dependent potential attached to the Laplacian.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/05/09`. Thanks to Richard Borcherds for corrections.

3.12. κ-noncollapsing via Perelman reduced volume

Having established the monotonicity of the Perelman reduced volume in Section **??** (after first heuristically justifying this monotonicity in Section 3.10), we now show how this can be used to establish κ-noncollapsing of Ricci flows, thus giving a second proof of Theorem 3.8.15. Of course, we already proved (a stronger version) of this theorem already in Section 3.9, using the Perelman entropy, but

this second proof is also important, because the reduced volume is a more localised quantity (due to the weight $e^{-l_{(0,x_0)}}$ in its definition and so one can in fact establish *local* versions of the non-collapsing theorem which turn out to be important when we study ancient κ-noncollapsing solutions later in Perelman's proof, because such solutions need not be compact and so cannot be controlled by global quantities (such as the Perelman entropy).

The route to κ-noncollapsing via reduced volume proceeds by the following scheme:

(3.358) Non-collapsing at time $t = 0$

$$\Downarrow$$

(3.359) Large reduced volume at time $t = 0$

$$\Downarrow$$

(3.360) Large reduced volume at later times t

$$\Downarrow$$

(3.361) Non-collapsing at later times t.

The implication (3.359) $\implies$ (3.360) is the monotonicity of Perelman reduced volume. In this lecture we discuss the other two implications (3.358) $\implies$ (3.359), and (3.360) $\implies$ (3.361)). Our arguments here are based on [**Pe2002**], [**KlLo2006**], [**MoTi2007**], though the material in [**MoTi2007**] differs in some key respects from the other two texts. A closely related presentation of these topics also appears in the paper of [**CaZh2006**].

3.12.1. Definitions. Let us first recall our definitions. Previously we defined *Perelman reduced length* and *reduced volume* for ancient flows $t \mapsto (M, g(t))$ for $t \in (-\infty, 0]$, centred at a point $(0, x_0)$ on the final time slice $t = 0$, but one can also define these quantities for flows on the time interval $[0, T]$ and for points $(t_0, x_0) \in [0, T] \times M$ as follows. We introduce the backward time variable $\tau := t_0 - t$. Given any path $\gamma : [0, \tau_1] \to M$, we define its *length*

$$L(\gamma) := \int_0^{\tau_1} \sqrt{\tau}(R + |\dot\gamma(\tau)|_g^2)\, d\tau \tag{3.362}$$

and for any (t_1, x_1) with $0 \leq t_1 < t_0$, with $\tau_1 := t_0 - t_1$, we define the *reduced length*

$$l_{(t_0,x_0)}(t_1, x_1) := \frac{1}{2\sqrt{\tau_1}} \inf_{\gamma} L(\gamma) \tag{3.363}$$

where $\gamma : [0, \tau_1] \to M$ ranges over all C^1 paths from x_0 to x_1 (which can also be viewed as trajectories in the spacetime manifold $[0, T] \times M$ from (t_0, x_0) to (t_1, x_1)). The *reduced volume*[86] is then defined as

$$\tilde{V}_{(t_0,x_0)}(\tau_1) := \frac{1}{\tau_1^{d/2}} \int_M e^{-l_{(t_0,x_0)}(t_1,x_1)} \, d\mu_{t_1}(x_1). \tag{3.364}$$

The arguments of Section **??** show that if $t \mapsto (M, g(t))$ is a Ricci flow, then the reduced volume is a non-increasing function of τ_1 for fixed (t_0, x_0). In particular, the reduced volume at later times t_1 is bounded from below by the reduced volume at time 0 (which is the implication (3.359) $\implies$ (3.360)).

3.12.2. Heuristic analysis. In the case of the trivial Euclidean flow, the reduced length is given by the formula

$$l_{(t_0,x_0)}(t_1, x_1) = \frac{|x_1 - x_0|^2}{4\tau_1} = \frac{|x_1 - x_0|^2}{4(t_1 - t_0)} \tag{3.365}$$

with the minimising geodesic given by the formula

$$\gamma(\tau) = x_0 + 2v\sqrt{\tau} \text{ with } v := \frac{x_1 - x_0}{2\sqrt{\tau_1}} \tag{3.366}$$

Here, we briefly argue why we expect heuristically to have a similar relationship

$$l_{(t_0,x_0)}(t_1, x_1) \approx \frac{d_{g(t_1)}(x_0, x_1)^2}{\tau_1} + O(1) \tag{3.367}$$

for the reduced length on more general Ricci flows, under an assumption of bounded normalised curvature.

Specifically, suppose that we have a normalised curvature bound $|\mathrm{Riem}|_g = O(1/\tau_1)$. Then we have $\dot{g} = -2\mathrm{Ric} = O(g/\tau_1)$, and so over the time scale τ_1, we see that the metric only changes by a multiplicative constant. If we ignore such constants for now, we see

[86] Note: some authors normalise the reduced volume by using $(4\pi\tau_1)^{d/2}$ instead of $\tau_1^{d/2}$, in order to give Euclidean space a reduced volume of 1, but this makes no essential difference to the analysis.

that the distance function $d_{g(t)}(x,y)$ does not change much over the time interval of interest.

Let γ be a minimising $\mathcal{L}$-geodesic from (t_0,x_0) to (t_1,x_1). This path has to traverse a distance roughly $d_{g(t_1)}(x_0,x_1)$ in time τ_1, and so its speed $|\dot\gamma|_g$ should be at least $d_{g(t_1)}(x_0,x_1)/\tau_1$. Also, the scalar curvature R should be $O(1/\tau_1)$ by the bounded normalised curvature assumption. Putting all this into (3.362) and (3.363) we heuristically obtain (3.367).

From (3.367), we expect the expression $e^{-l_{(t_0,x_0)}(t_1,x_1)}$ to be comparable to 1 when x_1 is inside the ball $B_{g(t_1)}(x_0,O(\sqrt{t_1}))$, and to be exponentially small outside of this ball. Using (3.364), we thus obtain a heuristic approximation for the Perelman reduced volume:

$$\tilde V_{(t_0,x_0)}(\tau_1) \approx \mathrm{Vol}_{g(t_1)}(x_0,\sqrt{\tau_1})/\tau_1^{d/2}. \tag{3.368}$$

Thus the Perelman reduced volume $\tilde V_{(t_0,x_0)}(\tau_1)$ is heuristically equivalent to the Bishop-Gromov reduced volume at (x_1,t_1) at scale τ_1. Since the latter measures non-collapsing, we heuristically obtain the implications (3.358) $\implies$ (3.359) and (3.360) $\implies$ (3.361).

3.12.3. From non-collapsing to lower bounds on reduced volume. Now we discuss implications of the form (3.358) $\implies$ (3.359) in more detail. Specifically, we show

Proposition 3.12.1. *Let $t\mapsto (M,g(t))$ be a d-dimensional Ricci flow on a complete manifold M for $t\in[0,T]$ such that we have the normalised initial conditions $|\mathrm{Riem}(0,x)|_g \le 1$ and $\mathrm{Vol}_{g(0)}(B_{g(0)}(x,1)) \ge \omega$ at time t=0 for some $\omega>0$ and all x (so in particular, the geometry is non-collapsed at scale 1 at all points at time zero). Then we have $\tilde V_{(t_0,x_0)}(t_0)\ge c$ for some $c=c(d,\omega,T)>0$ and all $(t_0,x_0)\in (0,T)\times M$.*

The main task in proving implications of the form (3.358) $\implies$ (3.359) is to show the existence of some large ball at time zero on which $l=l_{(t_0,x_0)}$ is bounded from above.

Turning to the specific proposition above, we first observe that we can reduce to the large time case $t_0\ge 1$. Indeed, if $0<t_0<1$, then we can rescale the Ricci flow until $t_0=1$ (this increases T, but we can simply truncate T to compensate for this). This rescaling reduces the

size of the initial Riemann curvature, and the volume of balls of unit radius are still bounded from below thanks to the Bishop-Gromov inequality (Corollary 3.10.3).

The next observation we need is that the control on the geometry at time zero persists for a short amount of additional time:

Lemma 3.12.2 (Local persistence of controlled geometry)**.** *Let the hypotheses be as in Proposition 3.11.1. Then there exists an absolute constant $c > 0$ (depending only on d) such that $|\mathrm{Riem}(t,x)|_g \leq 2$ for all times $0 \leq t \leq c$ and $x \in M$. Also we have $\mathrm{Vol}_{g(t)}(B_{g(t)}(x,1)) \geq \omega'$ for all $0 \leq t \leq c$ and $x \in M$, and some $\omega' > 0$ depending only on ω, d.*

Proof. We recall from (3.2.3) the nonlinear heat equation

$$\partial_t \mathrm{Riem} = \Delta \mathrm{Riem} + \mathcal{O}(g^{-1}\mathrm{Riem}^2) \tag{3.369}$$

for the Riemann curvature tensor Riem under Ricci flow. The bound on Riemann curvature can then obtained by an application of Hamilton's maximum principle (Proposition 3.4.5); we leave this as an exercise to the reader[87]. As in the heuristic discussion, the bounds on the Riemann curvature (and hence the Ricci curvature) show that the metric g and the distance function $d_{g(t)}(x,y)$ only change by at most a multiplicative constant; this also implies that the volume measure only changes by a multiplicative constant as well. From this we see that the lower bound on the volume of unit balls at time zero implies a lower bound on the volume of balls of radius $O(1)$ at times $0 \leq t \leq c$; one can then get back to balls of radius 1 by invoking the Bishop-Gromov inequality (Corollary 3.10.3). □

The next task is to find a point $y \in M$ such that the reduced length from (t_0, x_0) to $(0, y)$ is small, since this should force y (and the points close to y) to give a large contribution to the reduced volume. In the Euclidean case, one would just take $y = x_0$ (see (3.365)), but this does not necessarily work for general Ricci flows:

[87] Technically, one needs to first generalise the maximum principle from compact manifolds to complete manifolds of bounded curvature. This can be done using barrier functions, but it is somewhat technically involved: see [**CCGGIIKLLN2008**, Chapter 12].

note from (3.362), (3.363) that the reduced length from (t_0, x_0) to (t_1, x_0) could in principle be as large as

$$\frac{1}{2\sqrt{t_0 - t_1}} \int_0^{t_0 - t_1} \sqrt{\tau} R(t_0 - \tau, x_0)\, d\tau, \tag{3.370}$$

which could be quite large if the scalar curvature becomes large and positive (which is certainly within the realm of possibility, especially if one is approaching a singularity).

Fortunately, we can use the parabolic properties of the reduced length $l = l_{(t_0,x_0)}$, combined with the maximum principle, to locate a good point y with the required properties. From (3.339), (3.343), (3.357), and some rescaling and time translation, we obtain[88] the identities and inequalities

$$\nabla l = X \tag{3.371}$$

$$\partial_\tau l = \frac{1}{2} R - \frac{1}{2} |X|_g^2 - \frac{1}{2\tau} l \tag{3.372}$$

$$\Delta l \leq \frac{d}{2\tau} + \frac{1}{2} |X|_g^2 - \frac{1}{2} R - \frac{1}{2\tau} l, \tag{3.373}$$

where $X = \gamma'(\tau)$ is the final velocity vector of the minimising $\mathcal{L}$-geodesic from (t_0, x_0) to (t_1, x_1). From (3.372), (3.373) we obtain in particular that l is a supersolution of a heat equation:

$$\partial_t l \geq \Delta l + \frac{l - (d/2)}{\tau}. \tag{3.374}$$

Note that (3.374) holds with equality in the Euclidean case (3.365).

From the maximum principle (Corollary 3.4.3), we see that if we have the uniform lower bound $l \geq d/2$ at some time $0 \leq t < t_0$, then this bound will persist for all times between t and t_0. On the other hand, by using the upper bound (3.369) for $l(t_1, x_0)$ we see that the bound $l \geq d/2$ breaks down for times t sufficiently close to t_0. We therefore conclude that $\inf_{x \in M} l(t, x) < d/2$ for all $0 \leq t < t_0$. In particular we can find a point y such that

$$l(c, y) < d/2, \tag{3.375}$$

[88] We only derived (3.371)-(3.373) rigorously inside the domain of injectivity, but as discussed in Section **??**, one can establish the above inequalities in the sense of distributions on the whole manifold M.

where c is the small constant in Lemma 3.11.2. Given the bounded geometry control in Lemma 3.11.2 (and in particular the fact that $g(t)$ is comparable to $g(0)$ for $0 \leq t \leq c$), it is thus not hard to see (by concatenating the minimising path from $(0, x_0)$ to (c, y) with a geodesic segment (in the $g(0)$ metric) from (c, y) to $(0, y')$) that

$$l(0, y') \leq C \text{ for } y' \in B_{g(0)}(y, c') \tag{3.376}$$

for some $C, c' > 0$ depending only on d. The hypotheses on the geometry of $g(0)$, combined with the Bishop-Gromov inequality (Corollary 3.10.3), give a uniform lower bound for the volume of $B_{g(0)}(y, c')$, and Proposition 3.11.1 now follows directly from the definition (3.364) of reduced volume.

3.12.4. From lower bounds on reduced volume to non-collapsing. Now we consider the reverse type of implication (3.360) $\implies$ (3.361) from those just discussed. Here, the task is reversed; rather than establishing *upper* bounds on l on a ball of radius comparable to one, the main challenge is now to establish *lower* bounds (of the form $l \geq -O(1)$) on l on such a ball, as well as some growth bounds on l away from this ball.

We begin by formally stating the result of the form (3.360) $\implies$ (3.361) that we shall establish.

Proposition 3.12.3. *Let $t \mapsto (M, g(t))$ be a d-dimensional Ricci flow on a complete manifold M for $t \in [0, T]$, and let $0 \leq t_0 - r_0^2 \leq t_0 \leq T$ and $x_0 \in M$ be such that $|\mathrm{Riem}(t, x)|_g \leq r_0^{-2}$ for $x \in B_{g(t_0)}(x_0, r_0)$ and $t \in [t_0 - r_0^2, t_0]$, and such that $\tilde{V}_{(t_0, x_0)}(\tau) \geq \delta$ for some $\delta > 0$ and all $0 < \tau < r_0^2$. Then one has $\mathrm{Vol}_{g(t_0)}(B_{g(t_0)}(x_0, r_0)) \geq c$ for some c depending only on d and δ.*

Exercise 3.12.1. Use Proposition 3.11.1, Proposition 3.11.3, and the monotonicity of Perelman reduced volume to deduce Theorem 3.8.15.

We now prove Proposition 3.11.3. We first observe by time translation (and by removing the portion of the Ricci flow below $t_0 - r_0^2$ that we may normalise $t_0 - r_0^2 = 0$, and then by scaling we may normalise $t_0 = 1$. Thus we now have a Ricci flow on $[0, 1]$ with $|\mathrm{Riem}(t, x)|_g \leq 1$

on $[0,1] \times B_{g(1)}(x_0, 1)$ and

$$\tilde{V}_{(1,x_0)}(\tau) = \int_M e^{-l(\tau,x)}\, d\mu_{g(\tau)}(x) \geq \delta \tag{3.377}$$

for all $0 < \tau \leq 1$, where $l = l_{(1,x_0)}$ is the reduced length function. Our task is to show that $\mathrm{Vol}_{g(1)}(B_{g(1)}(x_0, 1))$ is bounded away from zero.

We first observe (as in Lemma 3.11.2) that the metrics $g(t)$ for $0 \leq t \leq 1$ are all comparable to each other up to multiplicative constants on $B_{g(1)}(x_0, 1)$, and so the balls in these metrics also differ only up to multiplicative constants.

Next, we would like to localise the reduced volume (3.377) to the ball $B_{g(1)}(x_0, 1)$ (since this is the only place where we really control the geometry). To do this it is convenient to work in the parabolic counterpart of normal coordinates around $(1, x_0)$ and exploit the pointwise version of the Perelman reduced volume monotonicity. To motivate this, recall from the pointwise inequality

$$\mathcal{L}_{\partial r} d\mu \leq \frac{d-1}{r}\, d\mu \tag{3.378}$$

that we had the Bishop-Gromov inequality

$$\partial_r r^{-(d-1)} \int_{S(x_0,r)} dS \leq 0 \tag{3.379}$$

where $S(x_0, r)$ is the sphere of radius r centred at x_0 with area element dS. Indeed, we can[89] rewrite the left-hand side of (21') as

$$\partial_r r^{-(d-1)} \int_{S^{d-1}} J_r(\omega)\, d\omega \tag{3.380}$$

where S^{d-1} is the standard sphere with the standard area element $d\omega$, and J_r is the Jacobian of the exponential map $\omega \mapsto \exp_{x_0}(r\omega)$; in the Euclidean case, $J_r(\omega) = r^{d-1}$. The inequality (3.378) (when combined with the Gauss lemma, Lemma 3.8.4) is equivalent to the pointwise inequality

$$\partial_r r^{-(d-1)} J_r(\omega) \leq 0 \tag{3.381}$$

[89] Actually, once the radius r exceeds the injectivity radius, one has to restrict to the portion of S^{d-1} that has not yet encountered the cut locus, but let us ignore this technical issue for now.

which certainly implies (3.380), but also implies the stronger fact that the Bishop-Gromov inequality can be localised to arbitrary sectors in the sense that $r^{-(d-1)} \int_\Omega J_r(\omega)\ d\omega$ (which can be viewed as the Bishop-Gromov reduced volume of the sector $\{\exp_{x_0}(r\omega) : \omega \in \Omega\}$) is non-increasing in r.

Now we develop parabolic analogues of the above observations. Recall from Section **??** that we have an $\mathcal{L}$-exponential map $\mathcal{L}\exp_{(1,x_0),\tau_1} : T_{x_0}M \to M$ for $0 \leq \tau_1 \leq 1$ that sends a tangent vector v to $\gamma(\tau_1)$, where $\gamma : [0,\tau] \to M$ is the unique $\mathcal{L}$-geodesic starting at x_0 with initial condition $v = \lim_{\tau\to 0} \sqrt{\tau} X(\tau) = \lim_{\tau\to 0} \sqrt{\tau}\gamma'(\tau)$. In the Euclidean case, this map is given by the formula

$$\mathcal{L}\exp_{(1,x_0),\tau_1}(v) = x_0 + 2(x_1 - x_0)\sqrt{\tau_1} \tag{3.382}$$

as can be seen from (3.366). We can[90] then rewrite the reduced volume $\tilde{V}_{(1,x_0)}(\tau)$ in terms of "normal coordinates" as

$$\tilde{V}_{(1,x_0)}(\tau) = \tau^{-d/2} \int_{\mathbf{R}^d} e^{-l(\mathcal{L}\exp_{(1,x_0),\tau}(v))} J_\tau(v)\ dv \tag{3.383}$$

where J_τ is the Jacobian of the map $v \mapsto \mathcal{L}\exp_{(1,x_0),\tau_1}(v)$.

In Section **??** we saw that the monotonicity of Perelman reduced volume followed from the pointwise inequality

$$\partial_\tau l - \Delta l + |\nabla l|_g^2 - R + \frac{d}{2\tau} \geq 0 \tag{3.384}$$

which of course also follows from (3.371)-(3.373).

Exercise 3.12.2. Use (3.371), (3.384), and the identity

$$\partial_\tau \mathcal{L}\exp_{(1,x_0),\tau}(v) = X \tag{3.385}$$

(which basically follows from the fact that any segment of a minimising $\mathcal{L}$-geodesic is again a $\mathcal{L}$-geodesic) to derive the pointwise inequality

$$\partial_\tau \tau^{-d/2} e^{-l(\mathcal{L}\exp_{(1,x_0),\tau}(v))} J_\tau(v) \leq 0. \tag{3.386}$$

Remark 3.12.4. Exercise 3.11.2 reproves the monotonicity of Perelman reduced volume (3.383), but also proves a stronger local version of this monotonicity in which the region of integration $\mathbb{R}^d$ is replaced

[90] Again, one has to restrict $\mathbf{R}^d$ to the portion of the tangent manifold lies inside the injectivity domain, but this domain turns out to be non-increasing in τ (for much the same reason that the region inside the cut locus of a point in a Riemannian manifold is star-shaped) and so this effect works in our favour as far as monotonicity is concerned.

by an arbitrary region Ω (intersected with the injectivity region, as mentioned earlier).

In the Euclidean case, a computation using (3.365) and (3.382) shows that $l(\mathcal{L}\exp_{(1,x_0),\tau}(v)) = |v|^2$ and $J_\tau(v) = 2^n\tau^{d/2}$. Also, one can use some basic analysis arguments to show that in the limit $\tau \to 0$, the expressions in (3.383) converge pointwise to their Euclidean counterparts. As a consequence we obtain the pointwise domination

$$\tau^{-d/2}e^{-l(\mathcal{L}\exp_{(1,x_0),\tau}(v))}J_\tau(v) \leq 2^{n/2}e^{-|v|^2} \tag{3.387}$$

for any v and any $0 < \tau < 1$. As a consequence, the far part of (3.383) (corresponding to "fast" geodesics) is negligible: we have

$$\tau^{-d/2}\int_{|v|>C} e^{-l(\mathcal{L}\exp_{(1,x_0),\tau}(v))}J_\tau(v)\,dv \leq \delta/2 \tag{3.388}$$

for some C depending only on d and δ. From this and the hypothesis (3.376) we thus obtain lower bounds on *local* Perelman reduced volume, or more precisely that

$$\tau^{-d/2}\int_{|v|\leq C} e^{-l(\mathcal{L}\exp_{(1,x_0),\tau}(v))}J_\tau(v)\,dv \geq \delta/2 \tag{3.389}$$

for all $0 < \tau \leq 1$. Now, we have bounded curvature on the cylinder $[0,1]\times B_{g(1)}(x_0,1)$. Using the heat equation (3.369) and standard parabolic regularity estimates, we thus conclude that any first[91] derivatives of the curvature are also bounded on the cylinder $[1/2,1]\times B_{g(1)}(x_0,1/2)$. In particular we have $\nabla R = O(1)$ in this cylinder. Thus the equation $G = 0$ for an $\mathcal{L}$-geodesic (where G was defined in (3.337)) becomes

$$\nabla_\tau X + \frac{1}{2\tau}X = O(1) + O(|X|) \tag{3.390}$$

or equivalently that

$$\nabla_\tau(\sqrt{\tau}X) = O(\sqrt{\tau}) + O(\sqrt{\tau}|X|) \tag{3.391}$$

as long as the geodesic stays inside this smaller cylinder. From this and Gronwall's inequality one easily verifies that for sufficiently small $0 < \tau < 1/2$ (depending on C, d), the exponential map $\mathcal{L}\exp_{(1,x_0),\tau}(v)$ does not exit the cylinder $[1/2,1]\times B_{g(1)}(x_0,1/2)$ for $|v| \leq C$. On the

[91] In fact, all higher derivatives are controlled as well; see [**Sh1989**] for full details.

other hand, at time τ, we see from (3.362), (3.363) and the bounds on curvature in this cylinder that the reduced length l of the associated $\mathcal{L}$-geodesic is bounded below by some constant depending on τ, C, d. We thus see (from the change of variables formula) that the left-hand side of (3.389) is bounded above by $O_{\tau,C,d}(\mathrm{Vol}_{g(1-\tau)}(B_{g(1)}(x_0, 1/2)))$. Choosing τ to be a small number depending on C, d, we thus conclude from (3.389) that the volume of $B_{g(1)}(x_0, 1)$ with respect to $g(1-\tau)$ (and hence $g(1)$, by comparability of metrics) is bounded from below by some constant depending on C and d, and thus ultimately on δ and d, giving Proposition 3.11.3 as desired.

3.12.5. Extensions. The pointwise nature of the monotonicity of Perelman reduced volume allows one to derive local versions of the non-collapsing result, in which one only needs a portion of the geometry to be non-collapsed at the initial time. A typical version of such a local noncollapsing result reads as follows.

Theorem 3.12.5 (Perelman's non-collapsing theorem, second version)**.** *Let $t \mapsto (M, g(t))$ be a d-dimensional Ricci flow on the time interval $[0, r_0^2]$, and suppose that one has the bounded normalised curvature condition $|\mathrm{Riem}|_g \leq r_0^{-2}$ on a cylinder $[0, r_0^2] \times B_{g(0)}(x_0, r_0)$ for some $x_0 \in M$. Suppose also that we have the volume lower bound $\mathrm{Vol}_{g(0)}(B_{g(0)}(x_0, r_0)) \geq c r_0^d$ for some $c > 0$. Then for any $A > 0$, the Ricci flow is κ-noncollapsed at (r_0^2, x) for any $x \in B_{g(r_0^2)}(x_0, Ar_0)$ and at any scale $0 < r < r_0$, for some κ depending only on d, c, A.*

The novelty here is that the geometry is controlled in a cylinder, rather than on the initial time slice, but one gets to conclude κ-noncollapsing at points some distance away from the cylinder. In view of Lemma 3.11.2, we see that this result is more or less a strengthening of the previous κ-noncollapsing theorem.

This theorem (or more precisely, a generalisation of it involving Ricci flow with surgery) is used in the original argument of Perelman[**Pe2002**] (and then in the later treatments by [**KlLo2006**] and [**CaZh2006**]) in order to deal with the long-time behaviour of Ricci flow with surgery, which is needed for the geometrisation conjecture. For proving the Poincaré conjecture, though, one has finite time extinction, and it turns out that the above theorem is not needed for the proof of that

conjecture (for instance, it does not appear in [**MoTi2007**]). Nevertheless I will sketch how the above theorem is proven below, since there are one or two interesting technical tricks that get used in the argument.

The proof of Theorem 3.11.5 is, unsurprisingly, a modification of the previous arguments. The implications (3.359) $\implies$ (3.360) and (3.360) $\implies$ (3.361) are basically unchanged, but one needs to replace Proposition 3.11.1 by the following variant.

Proposition 3.12.6. *Let the hypotheses be as in Theorem 3.11.5. Then for any $x \in B_{g(r_0^2)}(x_0, Ar_0)$ one has $\tilde{V}_{(r_0^2,x)}(r_0^2) \geq c'$ for some $c' > 0$ depending on A, c, d.*

We sketch the proof of Proposition 3.11.6. It is convenient to rescale so that $r_0 = 1$. In view of the non-collapsed nature of the geometry in $B_{g(0)}(x_0, 1)$, it suffices to establish a lower bound of the form $l_{(1,x)}(0, z) \geq -C$ for all $z \in B_{g(0)}(x_0, 1/2)$ for some $C > 0$ depending on A, c, d. Actually, because of the bounded geometry in the cylinder, it suffices to show that $l_{(1,x)}(1/2, y) \geq -C'$ for just one point $z \in B_{g(1/2)}(x_0, 1/10)$ for some $C' > 0$ depending on A, c, d, since one can join $(1/2, y)$ by a geodesic to $(1, z)$ much as in the proof of Proposition 3.11.1.

The task is now analogous to that of finding a point y that obeyed the relation (3.375), so we expect the heat equation (3.374) to again play a role. We do not need the sharp bound of $d/2$ which occurs in (3.375); on the other hand, y is now constrained to lie in a ball, which defeats a direct application of the maximum principle. To fix this one has to multiply the reduced length l by a penalising weight to force the minimum to lie in the desired ball at time $1/2$, and then rapidly relax this weight as one moves from time $1/2$ to time 1 so that it incorporates the point x at time 1. It turns out the maximum principle can then be applied with a suitable choice of weights, as long as one knows that the distance function $r(t, y) = d_{g(t)}(x_0, y)$ is a supersolution to a heat equation, and more precisely that $\partial_t r - \Delta r \geq -C$ when r is bounded away from the origin. But this can be established by the first and second variation formulae for the distance function, and in particular using the non-negativity of the second variation for minimising geodesics. Details can be found

in [**Pe2002**, Section 8], [**KlLo2006**, Sections 26-27], or [**CaZh2006**, Section 3.4].

Remark 3.12.7. One can also interpret the above analysis in terms of heat kernels, and using (3.384) instead of (3.374). The former inequality is equivalent to the assertion that the function $v := (4\pi\tau)^{-d/2}e^{-l}$ is a subsolution of the adjoint heat equation: $\partial_t v + \Delta v - Rv \leq 0$. As $t \to 1$, v approaches a Dirac mass at x (indeed, v asymptotically resembles the Euclidean backwards heat kernel from $(1, x_0)$) and the task is to obtain upper bounds on v at some point on a ball $B_{g(1/2)}(x_0, 1/10)$ at time $1/2$. This is basically equivalent to establishing lower bounds of Gaussian type for the fundamental solution of the adjoint heat equation at some point in $B_{g(1/2)}(x_0, 1/10)$. Similar analysis in the case of a static manifold with potential (and a lower bound on Ricci curvature) was carried out somewhat earlier in [**LiYa1986**].

As mentioned previously, in order to apply the non-collapsing result beyond the first surgery time, it is necessary to develop analogues of the above theory for Ricci flows with surgery. This turns out to be remarkably technical, but the main ideas at least are fairly clear. Firstly, one has to delete all $\mathcal{L}$-geodesics which pass through surgery regions when defining the Perelman reduced volume; such curves are called "inadmissible". Note that if $(1, x_0)$ is in a surgery region to begin with, then every curve is inadmissible but in this case the geometry can be controlled directly from the surgery theory. As it turns out, one can similarly deal with the case when $(1, x_0)$ has extremely high curvature because one can control the geometry of such regions. So we can easily eliminate these bad cases.

Because of the pointwise nature of the monotonicity formula for reduced volume, this restriction of admissibility does not affect the "(3.359) $\implies$ (3.360)" stage of the argument. The "(3.360) $\implies$ (3.361)" step is also largely unaffected, since removing inadmissible components of the reduced volume only serves to strengthen the hypothesis (3.360). But significant new technical difficulties arise in the "(3.358) $\implies$ (3.359)" portion of the argument, when one has to argue that not too much of the reduced volume has been deleted by all the various surgeries that take place between time $t = 0$ and time

$t = 1$. In particular, we still need to find a point y obeying (3.375) (or something very much like (3.375)) which is admissible. To do this, the basic idea is to establish that inadmissible curves have large reduced length (and so removing them will not impact the search for a solution to (3.375)). For technical reasons it is better to restrict attention to *barely admissible* curves - curves which just touch the border of the surgery region, but do not actually enter it. In this case it is possible to use the geometric control of the surgery regions to give some non-trivial lower bounds on the reduced length of such curves, although there are still significant technical issues to resolve beyond this. We will return to this point in Section ???.

3.12.6. Epilogue: a connection between Perelman entropy and Perelman reduced volume. We have shown two routes towards establishing κ-non-collapsing of Ricci flows, one using the (parameterised) Perelman entropies

(3.392)
$$\mu(g(t),\tau) := \inf\{\int_M (\tau(|\nabla f|^2 + R) + f - d)(4\pi\tau)^{-d/2} e^{-f}\ d\mu : \int_M (4\pi\tau)^{-d/2} e^{-f}\ d\mu = 1\}$$

and one using the reduced volumes $\tilde{V}_{(0,x_0)}$ mentioned above. Actually, the two quantities are related to each other (this is hinted at in [**Pe2002**, Section 9]); very roughly speaking, the potential function f in the theory of Perelman entropy plays the same role that reduced length l does in the theory of Perelman volume. Indeed, using (3.11.6) and shifting f by a constant if necessary, we have the log-Sobolev inequality

(3.393)
$$\int_M (\tau(|\nabla f|^2 + R) + f - d)(4\pi\tau)^{-d/2} e^{-f}\ d\mu \geq [\mu(g(t),\tau) - \log \int_M (4\pi\tau)^{-d/2} e^{-f}\ d\mu] \int_M (4\pi\tau)^{-d/2} e^{-f}\ d\mu.$$

An integration by parts reveals that we can replace the $|\nabla f|^2$ on the left-hand side by Δf, and hence one can also replace this quantity by $2\Delta f - |\nabla f|^2$.

We now apply this inequality with $\tau := t_0 - t$ and $f = l_{(t_0,x_0)}$ for some spacetime point (t_0, x_0) in the Ricci flow. Using (3.371), (3.373) we see that

$$2\Delta f - |\nabla f|^2 \leq \frac{d}{2\tau} - R - \frac{1}{\tau} f \tag{3.394}$$

and thus the left-hand side of (3.393) is non-positive. Using (3.364) we thus conclude a simple relationship between entropy and reduced volume[92]:

$$\mu(g(t_0 - \tau), \tau) \leq \log \frac{\tilde{V}_{(t_0,x_0)}(\tau)}{(4\pi)^{d/2}}. \tag{3.395}$$

Thus the Perelman entropy can be viewed as a global analogue of the Perelman reduced volume, in which we allow the base point x_0 to vary;thus it measures the global non-collapsing nature of the manifold, as opposed to the local nature; we already saw this in Section 3.9. Compare in particular (3.254) with the heuristic (3.368) using (3.395).

There are other connections between entropy and reduced volume; compare for instance the flow equation (3.237) for the potential f with (3.384). The adjoint heat equation $\partial_t u + \Delta u - Ru = 0$ also makes essentially the same appearance in both theories. See [**Pe2002**, Section 9] for further discussion.

Remark 3.12.8. As remarked above, the flow equation for f can be viewed as a pointwise versions of the entropy monotonicity formula, which in principle leads to localised monotonicity formulae for the Perelman entropy; some analysis in this direction appears in [**Pe2002**, Section 9]. But I do not know if these localised entropy formulae can substitute to give a different proof of Theorem 3.11.5.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/05/14`. Thanks to Sylvain Maillot and Dan for corrections.

[92] As usual, we have equality in physical space; this inequality also reinforces the suggestion that one normalise the reduced volume by an additional factor of $1/(4\pi)^{d/2}$.

3.13. High curvature regions of Ricci flow and κ-solutions

In previous sections, we have established (modulo some technical details) two significant components of the proof of the Poincaré conjecture: finite time extinction of Ricci flow with surgery (Theorem 3.3.13), and a κ-noncollapsing of Ricci flows with surgery (which, except for the surgery part, is Theorem 3.8.15). Now we come to the heart of the entire argument: the topological and geometric control of the high curvature regions of a Ricci flow, which is absolutely essential[93] in order for one to define surgery on these regions in order to move the flow past singularities. This control is intimately tied to the study of a special type of Ricci flow, the *κ-solutions* to the Ricci flow equation; we will be able to use compactness arguments (as well as the κ-noncollapsing results already obtained) to deduce control of high curvature regions of arbitrary Ricci flows from similar control of κ-solutions. A secondary compactness argument lets us obtain that control of κ-solutions from control of an even more special type of solution, the *gradient shrinking solitons* that we already encountered in Section 3.9.

The next few sections will be devoted to the analysis of κ-solutions, culminating in Perelman's topological and geometric classification (or near-classification) of such solutions (which in particular leads to the *canonical neighbourhood theorem* for these solutions, which we will briefly discuss below). In this lecture we shall formally define the notion of a κ-solution, and indicate informally why control of such solutions should lead to control of high curvature regions of Ricci flows. We'll also outline the various types of results that we will prove about κ-solutions.

Our treatment here is based primarily on [**MoTi2007**].

[93]Even once one has this control of high curvature regions, the proof of the Poincaré conjecture is still not finished; there is significant work required to properly define the surgery procedure, and then one has to show that the surgeries do not accumulate in time, and also do not disrupt the various monotonicity formulae that we are using to deduce finite time extinction, κ-noncollapsing, etc. But the control of high curvature regions is arguably the largest single task one has to establish in the entire proof.

3.13.1. Definition of a κ-solution. We fix a small number $\kappa > 0$ (basically the parameter that comes out of the non-collapsing theorem). Here is the formal definition of a κ-solution:

Definition 3.13.1 (κ-solutions)**.** A *κ-solution* is a Ricci flow $t \mapsto (M, g(t))$ which is

(1) *Ancient*, in the sense that t ranges on the interval $(-\infty, 0]$;

(2) *Complete and connected* (i.e. (M,g(t)) is complete and connected for every t);

(3) *Non-negative Riemann curvature*, i.e. Riem : $\bigwedge^2 TM \to \bigwedge^2 TM$ is positive semidefinite at all points in spacetime;

(4) *Bounded curvature*, thus $\sup_{(t,x)\in(-\infty,0]\times M} |\mathrm{Riem}|_g < +\infty$;

(5) *κ-noncollapsed* (see Definition 3.8.13) at every point (t_0, x_0) in spacetime and at every scale $r_0 > 0$;

(6) *Non-flat*, i.e. the curvature is non-zero at at least one point in spacetime.

This laundry list of properties arises because they are the properties that we are able to directly establish on limits of rescaled Ricci flows; see below.

Remark 3.13.2. If a d-dimensional Riemann manifold is both flat (thus Riem $= 0$) and non-collapsed at every scale, then (by Cheeger's lemma, Theorem 3.8.9) its injectivity radius is infinite, and by normal coordinates the manifold is isometric to Euclidean space $\mathbf{R}^d$. Thus the non-flat condition is only excluding the *trivial Ricci flow* $M = \mathbf{R}^d$ with the standard (and static) metric. The non-flat condition tells us that the (scalar, say) curvature is positive in at least one point of spacetime, but we will shortly be able to use the strong maximum principle to conclude in fact that the curvature is positive everywhere.

Remark 3.13.3. In three dimensions, the condition of non-negative RIemann curvature is equivalent to that of non-negative sectional curvature; see the discussion in Section 3.1. In any dimension, the conditions of non-negative bounded Riemann curvature imply that R and Ric are non-negative, and that $|\mathrm{Riem}|_g, |\mathrm{Ric}|_g = O(R)$ and $R = O_d(1)$. Thus as far as magnitude is concerned, the Riemann and Ricci curvatures of κ-solutions are controlled by the scalar curvature.

Now we discuss examples (and non-examples) of κ-solutions.

Example 3.13.4. Every gradient shrinking soliton or gradient steady soliton (M, g) (see Section 3.9) gives an ancient flow. This flow will be a κ-solution for sufficiently small κ if the Einstein manifold (M, g) is complete, connected, non-collapsed at every scale, and is not Euclidean space. For instance, the round sphere S^d with the standard metric is a gradient shrinking solution and will generate a κ-solution for any $d \geq 2$ and sufficiently small $\kappa > 0$, which we will refer to as the *shrinking round sphere* κ-solution.

Exercise 3.13.1. Show that the Cartesian product of two κ-solutions is again a κ-solution (with a smaller value of κ), as is the Cartesian product of a κ-solution. Thus for instance the product $S^2 \times \mathbf{R}$ of the shrinking round 2-sphere and the Euclidean line is a κ-solution, which we refer to as the *shrinking round* 3*-cylinder* $S^2 \times \mathbf{R}$.

Example 3.13.5. In one dimension, there are no κ-solutions, as every manifold is flat; in particular, the 1-sphere (i.e. a circle) is *not* a κ-solution (it is flat and also collapsed at large scales). In two dimensions, the shrinking round 2-sphere S^2 is κ-solution, as discussed above. We can quotient this by the obvious $\mathbf{Z}/2$ action to also get a shrinking round projective plane $\mathcal{RP}^2$ as a κ-solution. But we shall show in later lectures that if we restrict attention to oriented manifolds, then the shrinking round 2-sphere is the only 2-dimensional κ-solutions; this result is due to Hamilton, see e.g. [**ChKn2004**, Chapter 5]. For instance, the 2-cylinder $S^1 \times \mathbf{R}$ is not a κ-solution (it is both flat and collapsed at large scales). The cigar soliton (Example 3.9.4) also fails to be a κ-solution due to it being collapsed at large scales.

Example 3.13.6. In three dimensions, we begin to get significantly more variety amongst κ-solutions. We have the round shrinking 3-sphere S^3, but also all the quotients S^3/Γ of such round spheres by free finite group actions (including the projective space $\mathcal{RP}^3$, but with many other examples. We refer to these examples as *round shrinking* 3*-spherical space forms*. We have also seen the shrinking round cylinder $S^2 \times \mathbf{R}$; there are also finite quotients of this example such as shrinking round projective cylinder $\mathcal{RP}^2 \times \mathbf{R}$, or the quotient

of the cylinder by the orientation-preserving free involution $(\omega, z) \mapsto (-\omega, -z)$. We refer to these examples as the *unoriented and oriented quotients of the shrinking round* 3*-cylinder* respectively. The oriented quotient can be viewed as a half-cylinder $S^2 \times [1, +\infty)$ capped off with a punctured $\mathcal{RP}^3$ (and the whole manifold is in fact homeomorphic to a punctured $\mathcal{RP}^3$).

Example 3.13.7. One can also imagine perturbations of the shrinking solutions mentioned above. For instance, one could imagine non-round versions of the shrinking S^2 or shrinking $\mathcal{RP}^3$ example, in which the manifold has sectional curvature which is still positive but not constant. We shall informally refer to such solutions as *C-components* (we will define this term formally later, and explain the role of the parameter C). Similarly one could imagine variants of the oriented quotient of the shrinking round cylinder, which are approximately round half-cylinders $S^2 \times [1, +\infty)$ capped off with what is topologically either a punctured $\mathcal{RP}^3$ or punctured S^3 (i.e. with something homeomorphic to a ball); a 3-dimensional variant of a cigar soliton would fall into this category (such solitons have been constructed in [**Br2004**], [**Ca1996**]). We informally refer to such solutions as *C-capped strong ε-tubes* (we will define this term precisely later). One can also consider *doubly C-capped strong ε-tubes*, in which an approximately round finite cylinder $S^2 \times [-T, T]$ is capped off at both ends by either a punctured $\mathcal{RP}^3$ or punctured S^3; such manifolds then become homeomorphic to either S^3 or $\mathcal{RP}^3$. (Note we need to cap off any ends that show up in order to keep the manifold M complete.)

An important theorem of Perelman shows that these examples of κ-solutions are in fact the only ones:

Theorem 3.13.8 (Perelman classification theorem, imprecise version)**.** *Every* 3*-dimensional κ-solution takes on one of the following forms at time zero (after isometry and rescaling, if necessary):*

(1) *A shrinking round* 3*-sphere S^3 (or shrinking round spherical space form S^3/Γ);*

(2) *A shrinking round* 3*-cylinder $S^2 \times \mathbf{R}$, the quotient $\mathcal{RP}^2 \times \mathbf{R}$, or one of its quotients (either oriented or unoriented);*

(3) *A C-component;*

(4) *A C-capped strong ε-tube;*

(5) *A doubly C-capped strong ε-tube.*

We will make this theorem more precise in later sections (see also [**MoTi2007**, Chapter 9]).

Remark 3.13.9. At very large scales, Theorem 3.12.8 implies that an ancient solution at time zero either looks 0-dimensional (because the manifold was compact, as in the case of a sphere, spherical space form, C-component, or doubly C-capped strong ε-tube) or 1-dimensional, resembling a line (in the case of the cylinder) or half-line (for C-capped strong ε-tube). Oversimplifying somewhat, this 0- or 1-dimensionality of the three-dimensional κ-solutions is the main reason why surgery is even possible; if Ricci flow singularities could look 2-dimensional (such as $S^1 \times \mathbf{R}^2$, or as the product of the cigar soliton and a line) or 3-dimensional (as in $\mathbf{R}^3$) then it is not clear at all how to define a surgery procedure to excise the singularity. The point is that all the potential candidates for singularity that look 2-dimensional or 3-dimensional at large scales (after rescaling) are either flat or collapsed (or do not have bounded nonnegative curvature), and so are not κ-solutions. The unoriented quotiented cylinder $\mathcal{RP}^2 \times \mathbf{R}$ also causes difficulties with surgery (despite being only one-dimensional at large scales), because it is hard to cap off such a cylinder in a manner which is well-behaved with respect to Ricci flow; however if we assume that the original manifold M contains no embedded copy of $\mathcal{RP}^2 \times \mathbf{R}$ (which is for instance the case if the manifold is oriented, and in particular if it is simply connected) then this case does not occur.

Remark 3.13.10. In four and higher dimensions, things look much worse; consider for instance the product of a shrinking round S^2 with the trivial plane $\mathbf{R}^2$. This is a κ-solution but has a two-dimensional large-scale structure, and so there is no obvious way to remove singularities of this shape by surgery. It may be that in order to have analogues of Perelman's theory in higher dimensions one has to make much stronger topological or geometric assumptions on the manifold.

Note however that four-dimensional Ricci flows with surgery were already considered in [**Ha1986**] (with a rather different definition of surgery, however).

The classification theorem lets one understand the geometry of neighbourhoods of any given point in a κ-solution. Let us make the following imprecise definitions (which, again, will be made precise in later lectures):

Definition 3.13.11 (Canonical neighbourhoods, informal version)**.** Let (M, g) be a complete connected 3-manifold, let x be a point in M, and let U be an open neighbourhood of x. We normalise the scalar curvature at x to be 1.

(1) We say that U is an *ε-neck* if it is close (in a smooth topology) to a round cylinder $S^2 \times (-R, R)$, with x well in the middle of of this cylinder;

(2) We say that U is a *C-component* if U is diffeomorphic to S^3 or $\mathcal{RP}^3$ (in particular, it must be all of M) with sectional curvatures bounded above and below by positive constants, and with diameter comparable to 1.

(3) We say that U is *ε-round* if it is close (in a smooth topology) to a round sphere S^3 or spherical space form S^3/Γ (i.e. it is close to a constant curvature manifold).

(4) We say that U is a *(C, ε)-cap* if it consists of an ε-neck together with a cap at one end, where the cap is homeomorphic to either an open 3-ball or a punctured $\mathcal{RP}^3$ and obeys similar bounds as a C-component, and that x is contained inside the cap. (For technical reasons one also needs some derivative bounds on curvature, but we omit them here.)

(5) We say that U is a *canonical neighbourhood* of x if it is one of the above four types.

When the scalar curvature is some other positive number than 1, we can generalise the above definition by rescaling the metric to have curvature 1.

Using Theorem 3.12.8 (and defining all terms precisely), one can easily show the following important statement:

Corollary 3.13.12 (Canonical neighbourhood theorem for κ-solitons, informal version)**.** *Every point in a 3-dimensional κ-solution that does not contain an embedded copy of $\mathcal{RP}^2$ with trivial normal bundle is contained in a canonical neighbourhood.*

The next few sections will be devoted to establishing precise versions of Theorem 3.12.8, Definition 3.12.11, and Corollary 3.12.12.

3.13.2. High curvature regions of Ricci flows. Corollary 3.12.12 is an assertion about κ-solutions only, but it implies an important property about more general[94] Ricci flows:

Theorem 3.13.13 (Canonical neighbourhood for Ricci flows, informal version)**.** *Let $t \mapsto (M, g)$ be a Ricci flow of compact 3-manifolds on a time interval $[0, T)$, without any embedded copy of $\mathcal{RP}^2$ with trivial normal bundle. Then every point $(t, x) \in [0, T) \times M$ with sufficiently large scalar curvature is contained in a canonical neighbourhood.*

The importance of this theorem lies in the fact that all the singular regions that need surgery will have large scalar curvature, and Theorem 3.12.13 provides the crucial topological and geometric control in order to perform surgery on these regions[95].

Theorem 3.12.13 is deduced from Corollary 3.12.12 and a significant number of additional arguments. The strategy is to use a compactness-and-contradiction argument. As a very crude first approximation, the proof goes as follows:

(1) Suppose for contradiction that Theorem 3.12.13 failed. Then one could find a sequence $(t_n, x_n) \in [0, T) \times M$ of points with $R(t_n, x_n) \to +\infty$ which were not contained in canonical neighbourhoods.

[94] Actually, as with many other components of this proof, we actually need a generalisation of this result for Ricci flow with surgery, but we will address this (non-trivial) complication later.

[95] This is a significant oversimplification, as one has to also study certain "horns" that appear at the singular time in order to find a particularly good place to perform surgery, but we will postpone discussion of this major additional issue later in this chapter.

(2) M, being compact, has finitely many components; by restricting attention to a subsequence of points if necessary, we can take M to be connected.

(3) On any compact time interval $[0,t]\times M$, the scalar curvature is necessarily bounded, and thus $t_n \to T$. As a consequence, if we define the rescaled Ricci flows $g^{(n)}(t) = \frac{1}{L_n^2} g(t_n + L_n^2 t)$, where $L_n := R(t_n, x_n)^{-1/2}$ is the natural length scale associated to the scalar curvature at (t_n, x_n), then these flows will become increasingly ancient. Note that in the limit (which we will not define rigorously yet, but think of a *pointed Gromov-Hausdorff limit* for now), the increasingly large manifolds $(M, g^{(n)}(t))$ may cease to be compact, but will remain complete.

(4) Because of the Hamilton-Ivey pinching phenomenon (Theorem 3.4.16), we expect the rescaled flows $t \mapsto (M, g^{(n)}(t))$ to have non-negative Ricci curvature in the limit (and hence non-negative Riemann curvature also, as we are in three dimensions).

(5) If we can pick the points (t_n, x_n) suitably (so that the scalar curvature $R(t_n, x_n)$ is larger than or comparable to the scalar curvatures at other nearby points), then we should be able to ensure that the rescaled flows $t \mapsto (M, g^{(n)}(t))$ have bounded curvature in the limit.

(6) Since κ-noncollapsing is invariant under rescaling, the noncollapsing theorem (Theorem 3.8.15) should ensure that the rescaled flows remain κ-noncollapsed in the limit.

(7) Since the rescaled scalar curvature at the base point x_n of $(M, g^{(n)})$ is equal to 1 by construction, any limiting flow will be non-flat.

(8) Various compactness theorems (of Gromov, Hamilton, and Perelman) exploiting the non-collapsed, bounded curvature, and parabolic nature of the rescaled Ricci flows now allows one to extract a limiting flow $(M^{(\infty)}, g^{(\infty)})$. This limit is initially in a fairly weak sense, but one can use parabolic theory to upgrade the convergence to quite a strong (and

smooth) convergence. In particular, the limit of the Ricci flows will remain a Ricci flow.

(9) Applying Steps 2-8, we see that the limiting flow $(M^{(\infty)}, g^{(\infty)})$ is a κ-solution.

(10) Applying Corollary 3.12.12, we conclude that every point in the limiting flow lies inside a canonical neighbourhood. Using the strong nature of the convergence (and the scale-invariant nature of canonical neighbourhoods), we deduce that the points (t_n, x_n) also lie in canonical neighbourhoods for sufficiently large n, a contradiction.

There are some non-trivial technical difficulties in executing the above scheme, especially in Step 5 and Step 8. Step 8 will require some compactness theorems for κ-solutions which we will deduce in later lectures. For Step 5, the problem is that the points (t_n, x_n) that we are trying to place inside canonical neighbourhoods have large curvature, but they may be adjacent to other points of significantly higher curvature, so that the limiting flow $(M^{(\infty)}, g^{(\infty)})$ ends up having unbounded curvature. To get around this, Perelman established Theorem 3.12.13 by a downwards induction argument on the curvature, first establishing the result for extremely high curvature, then for slightly less extreme curvature, and so forth. The point is that with such an induction hypothesis, any potentially bad adjacent points of really high curvature will be safely tucked away in a canonical neighbourhood of high curvature, which in turn is connected to another canonical neighbourhood of high curvature, and so forth; some basic topological and geometric analysis then eventually lets us conclude that this bad point must in fact be quite far from the base point (t_n, x_n) (much further away than the natural length scale L_n, in particular), so that it does not show up in the limiting flow $(M^{(\infty)}, g^{(\infty)})$. We will discuss these issues in more detail in later lectures.

3.13.3. Benchmarks in controlling κ-solutions. As mentioned earlier, the next few lectures will be focused on controlling κ-solutions. It turns out that the various properties in Definition 3.12.1 interact very well with each other, and give remarkably precise control on

these solutions. In this section we state (without proofs) some of the results we will establish concerning such solutions.

Proposition 3.13.14 (Consequences of Hamilton's Harnack inequality)**.** *Let $t \mapsto (M, g(t))$ be a κ-solution. Then $R(t,x)$ is a non-decreasing function of time. Furthermore, for any $(t_0, x_0) \in (-\infty, 0] \times M$, we have the pointwise inequalities*

$$|\nabla l_{(t_0,x_0)}|^2 + R \leq \frac{3 l_{(t_0,x_0)}}{\tau} \tag{3.396}$$

and

$$-2\frac{l_{(t_0,x_0)}}{\tau} \leq \frac{\partial l_{(t_0,x_0)}}{\partial \tau} \leq \frac{l_{(t_0,x_0)}}{\tau} \tag{3.397}$$

on $(-\infty, t_0) \times M$, where of course $\tau := t_0 - t$ is the backwards time variable.

These inequalities follow from an important Harnack inequality[**Ha1993**] of Hamilton (also related to the earlier paper [**LiYa1986**]) that we will discuss in the next lecture. These results rely primarily on the ancient and non-negatively curved nature of κ-solutions, as well as the Ricci flow equation $\dot{g} = -2\mathrm{Ric}$ of course.

Now one can handle the two-dimensional case:

Proposition 3.13.15 (Classification of 2-dimensional κ-solutions)**.** *The only two-dimensional κ-solutions are the round shrinking 2-spheres.*

This proposition relies on first studying a certain asymptotic limit of the κ-solution, known as the *asymptotic soliton*, to be defined later. One shows that this asymptotic limit is a round shrinking 2-sphere, which implies that the original κ-solution is asymptotically a round shrinking 2-sphere. One can then invoke Hamilton's rounding theorem[**Ha1982**] to finish the claim.

Turning now to three dimensions, the first important result that the curvature R decays slower at infinity than what scaling naively predicts.

Proposition 3.13.16 (Asymptotic curvature)**.** *Let $t \mapsto (M, g(t))$ be a 3-dimensional κ solution which is not compact. Then for any time $t \in (-\infty, 0)$ and any base point $p \in M$, we have* $\limsup_{x \to \infty} R(t,x) d_{g(t)}(x,p)^2 = +\infty$.

The proof of Proposition 3.12.16 is based on another compactness-and-contradiction argument which also heavily exploits some splitting theorems in Riemannian geometry, as well as the following version of the *soul theorem* of Cheeger and Gromoll [**ChGr1972**], first proven by Perelman:

Theorem 3.13.17 (Perelman's soul theorem)**.** [**Pe1994**] *Every complete non-compact d-dimensional manifold with non-negative sectional curvatures, and with strictly positive curvatures at at least one point, is diffeomorphic to $\mathbf{R}^d$.*

The increasing curvature at infinity can be used to show that the volume does not grow as fast at infinity as scaling predicts:

Proposition 3.13.18 (Asymptotic volume collapse)**.** *Let $t \mapsto (M, g(t))$ be a 3-dimensional κ solution which is not compact. Then for any time $t \in (-\infty, 0)$ and any base point $p \in M$, we have* $\limsup_{r\to+\infty} \mathrm{Vol}_{g(t)}(B_{g(t)})($ 0*.*

Note that Proposition 3.12.18 does not contradict the non-collapsed nature of the flow, since one does not expect bounded normalised curvature at extremely large scales. Proposition 3.12.18 morally follows from Bishop-Gromov comparison geometry theory, but its proof in fact uses yet another compactness-and-contradiction argument combined with splitting theory.

An important variant of Proposition 3.12.18 and Proposition 3.12.16 (and yet another compactness-and-contradiction argument) states that on any ball $B_{g(0)}(p, r)$ at time zero on which the volume is large (e.g. larger than νr^3 for some $\nu > 0$), one has bounded normalised curvature, thus $R = O_\nu(1/r^2)$ on this ball. This fact helps us deduce

Theorem 3.13.19 (Perelman compactness theorem, informal version)**.** *The space of all pointed κ-solutions (allowing $\kappa > 0$ to range over the positive real numbers) is compact (in a suitable topology) after normalising the scalar curvature at the base point to be* 1*.*

One corollary of this compactness is that there is in fact a universal $\kappa_0 > 0$ such that every κ-solution is a κ_0-solution. (Indeed, the proof of this universality is one of the key steps in the proof of the

above theorem.) This theorem is proven by establishing some uniform curvature bounds on κ-solutions which come from the previous volume analysis.

The proof of Theorem 3.12.8 (and thus Corollary 3.12.12) follows from this compactness once one can classify the asymptotic solitons mentioned earlier. This task in turn requires many of the techniques already mentioned, together with some variational analysis of the gradient curves of the potential function f that controls the geometry of the soliton.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/05/16`. Thanks to Anton Fonarev for corrections.

3.14. Li-Yau-Hamilton Harnack inequalities and κ-solutions

We now turn to the theory of *parabolic Harnack inequalities*, which control the variation over space and time of solutions to the scalar heat equation

$$u_t = \Delta u \tag{3.398}$$

which are bounded and non-negative, and (more pertinently to our applications) of the curvature of Ricci flows

$$g_t = -2\mathrm{Ric} \tag{3.399}$$

whose Riemann curvature Riem or Ricci curvature Ric is bounded and non-negative. For instance, the classical[96] parabolic Harnack inequality of Moser[**Mo1964**] asserts, among other things, that one has a bound of the form

$$u(t_1, x_1) \leq C(t_1, x_1, t_0, x_0, T_-, T_+, M) u(t_0, x_0) \tag{3.400}$$

whenever $u : [T_-, T_+] \times M \to \mathbf{R}^+$ is a bounded non-negative solution to (3.398) on a complete static Riemannian manifold M of bounded curvature, $(t_1, x_1), (t_0, x_0) \in [T_-, T_+] \times M$ are spacetime points with $t_1 < t_0$, and $C(t_1, x_1, t_0, x_0, T_-, T_+, M)$ is a constant which is uniformly bounded for fixed t_1, t_0, T_-, T_+, M when x_1, x_0 range over a

[96] The even more classical elliptic Harnack inequality gives (3.398) in the steady state case, i.e. for bounded non-negative harmonic functions.

compact set. In terms of heat kernels, one can view (3.398) as an assertion that the heat kernel associated to (t_0, x_0) dominates (up to multiplicative constants) the heat kernel at (t_1, x_1).

The classical proofs of the parabolic Harnack inequality do not give particularly sharp bounds on the constant $C(t_1, x_1, t_0, x_0, T_-, T_+, M)$. Such sharp bounds were obtained in [**LiYa1986**], especially in the case of the scalar heat equation (3.398) in the case of static manifolds of non-negative Ricci curvature, using Bochner-type identities and the scalar maximum principle. In fact, a stronger differential version of (3.400) was obtained which implied (3.400) by an integration along spacetime curves (closely analogous to the $\mathcal{L}$-geodesics considered in earlier lectures). These bounds were particularly strong in the case of ancient solutions (in which one can send $T_- \to -\infty$). Subsequently, Hamilton[**Ha1993**] applied his tensor-valued maximum principle together with some remarkably delicate tensor algebra manipulations to obtain Harnack inequalities of Li-Yau type for solutions to the Ricci flow (3.399) with bounded non-negative Riemannian curvature. In particular, this inequality applies to the κ-solutions introduced in Section 3.12.

In this section, we shall discuss all of these inequalities (although we will not give the full details for the proof of Hamilton's Harnack inequality, as the computations are quite involved), and derive several important consequences of that inequality for κ-solutions. The material here is based on several sources, including [**Ev1998**], [**Mu2006**], [**MoTi2007**], [**CaZh2006**], and of course the primary source papers mentioned in this section.

3.14.1. Scalar parabolic Harnack inequalities. Before we turn to the inequalities for Ricci flows (which are our main interest), we first consider the simpler case of scalar non-negative bounded solutions $u : [T_-, T_+] \times M \to \mathbf{R}^+$ to the heat equation (3.398) on a static complete smooth Riemannian manifold (M, g). This case will not actually be used in our applications but serve as an important motivating example of the method. Our basic tools will be the scalar maximum principle and the following identity.

Exercise 3.14.1. Let $f : M \to \mathbf{R}$ be a smooth function. Establish the *Bochner formula*

$$\Delta|\nabla f|_g^2 = 2\nabla_{\nabla f}\Delta f + 2|\mathrm{Hess}(f)|_g^2 + 2\mathrm{Ric}(\nabla f, \nabla f). \tag{3.401}$$

Hint: use abstract index notation, and use the torsion-free nature of the connection, combined with the definitions of Riemann and Ricci curvature.

This leads to the following consequence:

Exercise 3.14.2. Let $u : [T_-, T_+] \times M \to \mathbf{R}^+$ be a strictly positive solution to (3.398), and let $f := \log u$. Establish the nonlinear heat equation identities

$$f_t = \Delta f + \nabla_{\nabla f} f = \Delta f + |\nabla f|_g^2 \tag{3.402}$$

$$\partial_t(\Delta f) = \Delta(\Delta f) + \nabla_{\nabla f}\Delta f + 2|\mathrm{Hess} f|_g^2 + 2\mathrm{Ric}(\nabla f, \nabla f) \tag{3.403}$$

$$\partial_t(|\nabla f|^2) = \Delta(|\nabla f|^2) + \nabla_{\nabla f}(|\nabla f|^2) - 2|\mathrm{Hess} f|_g^2 + 2\mathrm{Ric}(\nabla f, \nabla f) \tag{3.404}$$

Now we can state the Li-Yau Harnack inequality.

Proposition 3.14.1 (Li-Yau Harnack inequality)**.** *Let M be a smooth compact d-dimensional Riemannian manifold with non-negative Ricci curvature, and let $u : [T_-, T_+] \times M \to \mathbf{R}^+$ be a strictly positive smooth solution to* (3.398)*. Then for every $(t, x) \in (T_-, T_+] \times M$, we have*

$$\frac{\partial_t u}{u} - \frac{|\nabla u|^2}{u^2} + \frac{d}{2(t - T_-)} \geq 0. \tag{3.405}$$

Proof. By adding an epsilon to u if necessary (and then sending epsilon back to zero at the end of the argument) we may assume[97] that $u \geq \varepsilon$ for some $\varepsilon > 0$. Write $f := \log u$ and $F := \Delta f$. From Cauchy-Schwarz we have $|\mathrm{Hess} f|_g^2 \geq \frac{1}{d} F^2$, and so from (3.403) we see that F is a supersolution to a nonlinear heat equation:

$$F_t \geq \Delta F + \nabla_{\nabla f} F + \frac{2}{d}|F|^2. \tag{3.406}$$

[97] We shall use this trick frequently in the sequel and refer to it as the *epsilon-regularisation trick*.

On the other hand, $-\frac{d}{2(t-T_-)}$ is a sub-solution to the same equation, and the hypothesis that u is smooth and bounded below by ε (together with the compactness of M) implies that F dominates $-\frac{d}{2(t-T_-)}$ at times close to T_-. Applying the scalar maximum principle (Corollary 3.4.3) we conclude that $F \geq -\frac{d}{2(t-T_-)}$. The claim (3.405) now follows from (3.402) and the chain rule. □

Remark 3.14.2. One can extend this inequality to the case when M is not compact, but is instead complete with bounded curvature, as long as one now adds the hypothesis that u is bounded (which was automatic in the compact case). The basic idea used to modify the proof is to multiply u by a suitable weight that grows at infinity to force the minimum value of F to lie in a compact set so that the maximum principle arguments can still be applied; we omit the standard details.

Remark 3.14.3. Observe that when $M = \mathbf{R}^d$ is Euclidean and u is the fundamental solution $u(t,x) = \frac{1}{(4\pi(t-T_-))^{d/2}} e^{-|x-x_0|^2/4(t-T_-)}$ for some $x_0 \in \mathbf{R}^d$, that (3.405) becomes an equality.

For strictly positive ancient solutions $u : (-\infty, 0] \times M \to \mathbf{R}^+$ to (3.398) on a compact manifold of non-negative Ricci curvature, one can send T_- to negative infinity, we conclude from (3.405) that

$$\frac{\partial_t u}{u} \geq \frac{|\nabla u|^2}{u^2}. \tag{3.407}$$

In particular we see that $\partial_t u \geq 0$; thus non-negative ancient solutions to the linear heat equation on compact manifolds of non-negative Ricci curvature are non-decreasing[98] in time.

One can linearise the inequality (3.407) in u, obtaining the assertion that

$$\partial_t u - \nabla_X u + \frac{1}{4}|X|_g^2 u \geq 0 \tag{3.408}$$

for any vector field X. Indeed (3.407) and (3.408) are easily seen to be equivalent by the Cauchy-Schwarz inequality. One advantage of

[98] Actually, it turns out that the only such solutions are in fact constant, but we will shortly generalise this assertion to less trivial situations.

the formulation (3.408) is that it also holds true when u is merely non-negative, as opposed to strictly positive u, by the epsilon-regularisation trick. In terms of $f = \log u$, (3.408) can also be expressed as

$$\partial_t f - \nabla_X f + \frac{1}{4}|X|_g^2 \geq 0 \tag{3.409}$$

although now one needs u to be strictly positive for (3.13.1) to make sense.

Now let $(t_0, x_0), (t_1, x_1) \in (-\infty, 0] \times M$ be points in spacetime with $t_1 < t_0$, let $\tau_1 := t_0 - t_1$, and let $\gamma : [0, \tau_1] \to M$ be a path from x_0 to x_1. From the fundamental theorem of calculus and the chain rule we have

$$f(t_1, x_1) - f(t_0, x_0) = \int_0^{\tau_1} -\partial_t f + \nabla_X f \; d\tau \tag{3.410}$$

where $X := \gamma'(\tau)$ and the integrand is evaluated at $(t_0 - \tau, \gamma(\tau))$. Applying (3.13.1) and then exponentiating we obtain the Harnack inequality

$$u(t_1, x_1) \leq \exp(\frac{1}{4}\int_0^{\tau_1} |X|^2 \; d\tau) u(t_0, x_0) \tag{3.411}$$

which can be extended from strictly positive solutions u to non-negative solutions u by the epsilon-regularisation trick[99]. By choosing γ to be the constant-speed minimising geodesic from x_0 to x_1, we thus conclude that

$$u(t_1, x_1) \leq \exp(d(x_0, x_1)^2/4\tau_1) u(t_0, x_0). \tag{3.412}$$

Remark 3.14.4. Specialising to the case when u is a static harmonic function and sending $\tau_1 \to \infty$ (and using Remark 3.13.2), we recover a variant of *Liouville's theorem*: a bounded harmonic function on a Riemannian manifold of bounded non-negative curvature is constant.

Exercise 3.14.3. If the non-negative solution u to (3.398) is not ancient, but is only restricted to a time interval $[T_-, T_+]$, show that one still has the variant

$$u(t_1, x_1) \leq (\frac{t_2 - T_-}{t_1 - T_-})^{d/2} \exp(d(x_0, x_1)^2/4\tau_1) u(t_0, x_0). \tag{3.413}$$

[99] Observe the similarity here with the $\mathcal{L}$-geodesic theory from Section **??**.

Exercise 3.14.4. If the non-negative solution u to (3.398) is restricted to a time interval $[T_-, T_+]$, and one no longer assumes that the Ricci curvature is non-negative (but it will still be bounded, since M is compact), establish the Harnack inequality (3.400) for some $C(t_1, x_1, t_0, x_0, T_-, T_+, M)$. *Hint*: repeat the above arguments but with F replaced by $F_\kappa := \Delta f + \kappa |\nabla f|^2$ for some small κ. Show (using (3.403), (3.404)) that if κ is small enough, then F_κ obeys an inequality similar to (3.13.1) but with an additional factor of $-O_\kappa(|F_\kappa|)$ on the right-hand side.

Exercise 3.14.5. Establish the *strong maximum principle*: if M is compact and $u : [T_-, T_+] \times M \to \mathbf{R}^+$ is a non-negative solution to (3.398) which is not identically zero, then it is strictly positive for times $t \in (T_-, T_+]$ (or equivalently, if u vanishes at even one point in $(T_-, T_+] \times M$, then it is identically zero).

Exercise 3.14.6. Generalise the strong maximum principle to the case when u is a supersolution $u_t \geq \Delta u$ to the heat equation rather than a solution. Also generalise it to the case when the metric g is not static, but instead varies smoothly in time. (For an additional challenge, generalise further to the case when M is complete, the metric has uniformly bounded Riemann curvature, u is bounded, and one also has a drift term $\nabla_X u$ on the right-hand side of the equation for some bounded X.)

Exercise 3.14.7. Using the final generalisation of Exercise 3.13.6, as well as the evolution equation (3.2.3) for scalar curvature, show that the scalar curvature of a κ solution is strictly positive at every point in spacetime. (We will prove stronger versions of this fact later in this section.)

Further variants and applications of these scalar Harnack inequalities can be found in [**LiYa1986**].

3.14.2. Parabolic Harnack inequalities for the Ricci flow. Now we turn from the scalar equation (3.398) to the Ricci flow equation (3.399), which one could think of as a kind of tensor-valued quasilinear heat equation (by de Turck's trick, see Section 3.2). To begin with let us first consider the simple two-dimensional case $d = 2$. In

this case the Bianchi identities make the Riemann, Ricci, and scalar curvatures are all essentially equivalent (see Section 3.1); in particular one has the identity

$$\mathrm{Ric} = \frac{1}{2} R g \tag{3.414}$$

in the two-dimensional case. In particular, the heat equation (3.2.3) for scalar curvature simplifies to

$$\partial_t R = \Delta R + R^2 \tag{3.415}$$

in this case; compare this with (3.398).

Suppose that the scalar curvature R is strictly positive. Setting $f := \log R$, one has an analogue of (3.402):

$$\partial_t f = \Delta f + \nabla_f f + R. \tag{3.416}$$

Exercise 3.14.8. If we set $F := \partial_t f - \nabla_f f = \Delta f + R$, show the following analogue of (3.13.1):

$$\partial_t F \geq \Delta F + 2\nabla_{\nabla f} F + F^2. \tag{3.417}$$

Hint: you will need to first derive the identity $\partial_t \Delta v = \Delta \partial_t v + R\Delta v$ for arbitrary smooth v. Conclude that if M is compact and R is strictly positive on the time interval $[T_-, T_+]$, then $F \geq -1/(t - T_-)$, and thus conclude *Hamilton's Harnack inequality for surfaces*:

$$\partial_t R - \frac{|\nabla R|_g^2}{R} + \frac{R}{t - T_-} \geq 0. \tag{3.418}$$

Extend this inequality to the case when R is merely non-negative rather than strictly positive by setting f equal to $\log(R + \varepsilon)$ rather than $\log R$ and then setting ε to zero[100].

For ancient two-dimensional solutions with non-negative curvature, we thus conclude from the Harnack inequality (3.418) that R (and $f = \log R$) obeys the same bounds (3.407), (3.408), (3.13.1) that scalar solutions u did previously. In particular R is non-decreasing in time, and more generally

$$\partial_t R - \nabla_X R + \frac{1}{4}|X|_g^2 R \geq 0 \tag{3.419}$$

[100]This is how one performs the epsilon regularisation trick for Ricci flow: by modifying the logarithm function by an epsilon, rather than the solution.

for any X. We can also obtain an analogue of (3.13.1). Also, observe from the assumption of non-negative curvature and (3.399) that the metric is non-increasing with time, and so we can also deduce an analogue of (3.412):

$$R(t_1, x_1) \leq \exp(d_{g(t_1)}(x_0, x_1)^2/4\tau_1)R(t_0, x_0). \tag{3.420}$$

With a non-trivial amount of effort, one can extend Hamilton's Harnack inequality to higher dimensions. One cannot argue solely using the scalar curvature R, because the equation $\partial_t R = \Delta R + 2|\mathrm{Ric}|^2$ for that curvature also involves the Ricci tensor, which thus also needs to be controlled. What is worse, one cannot argue solely using the Ricci tensor either, because the equation

$$\partial_t \mathrm{Ric}_{\alpha\beta} = \Delta \mathrm{Ric}_{\alpha\beta} + 2\mathrm{Ric}^{\gamma}_{\delta}\mathrm{Riem}^{\delta}_{\alpha\gamma\beta} - 2\mathrm{Ric}_{\alpha\gamma}\mathrm{Ric}^{\gamma}_{\beta} \tag{3.421}$$

for the evolution of that curvature involves the Riemann tensor. To proceed, one in fact has to deal with the equation for the full Riemann tensor,

$$\partial_t \mathrm{Riem} = \Delta \mathrm{Riem} + \mathcal{O}(g^{-1}\mathrm{Riem}^2) \tag{3.422}$$

where $\mathcal{O}(g^{-1}\mathrm{Riem}^2)$ is an explicit but rather complicated quadratic expression in the Riemann curvature. This expression simplifies when using a moving orthonormal frame, as was done in Section 3.4, to the form

$$\partial_t \mathcal{T} = \Delta \mathcal{T} + \mathcal{T}^2 + \mathcal{T}^{\#}. \tag{3.423}$$

By using (3.423) and many tensor calculations, one can (eventually) establish a (rather complicated) analogue of the (3.419) for $\mathcal{T}$, and hence for Riem and then (after taking some traces) to R. In particular, we have

Theorem 3.14.5 (Hamilton's Harnack inequality for ancient Ricci flows)**.** *Let $t \mapsto (M, g(t))$ be a complete ancient Ricci flow with non-negative bounded Riemann curvature*[101]*. Then we have the pointwise inequality*

$$\partial_t R - \nabla_X R + \frac{1}{2}\mathrm{Ric}(X, X) \geq 0 \tag{3.424}$$

for any vector field X.

[101]Note in particular that all κ-solutions are of this form.

Note that in the two-dimensional case, (3.424) collapses to (3.13.2) thanks to (3.414).

The proof of (3.424) is remarkably delicate (in particular, going through the tensor curvature equation (3.423)), but ultimately follows broadly similar lines to the previous arguments (i.e. Bochner-type identities, Cauchy-Schwarz type inequalities, and tensor maximum principles). For technical reasons it is also convenient to carry auxiliary tensor fields such as the vector field X appearing in (3.424) throughout the argument. We refer the reader to [**Ha1993**] for details[102].

Exercise 3.14.9. Suppose that (M, g) solves the gradient steady soliton equation $\mathrm{Ric} + \mathrm{Hess}(f) = 0$ for some smooth f. Using the Bianchi identity $\nabla_\alpha R = 2\nabla^\beta \mathrm{Ric}_{\alpha\beta}$, establish the identity

$$\nabla_\alpha R = 2\mathrm{Ric}_{\alpha\beta}\nabla^\beta f \tag{3.425}$$

(note this identity also holds for gradient shrinking or expanding solitons) and then by taking divergences and using the Bianchi identity again, establish that

$$\Delta R + 2|\mathrm{Ric}|^2 = \nabla_\alpha R \nabla^\alpha f. \tag{3.426}$$

Conclude that (3.424) is an identity in this case when one sets $X^\alpha := 2\nabla^\alpha f$.

3.14.3. Applications of the Harnack inequality. Now we develop some applications of the Harnack inequality for κ-solutions. One easy application follows by setting X equal to 0, giving the pointwise monotonicity of the scalar curvature in time:

$$\partial_t R \geq 0. \tag{3.427}$$

Another application is to obtain a slightly weakened version of (3.420) (with the 4 in the denominator replaced by 2):

Exercise 3.14.10. Show that one has $\mathrm{Ric}(X, X) \leq \frac{1}{2} R|X|^2$ whenever one has non-negative Riemann curvature. Using this and (3.424),

[102] There are alternate proofs, such as the one in [**ChCh1995**] using a metric closely related to the high-dimensional metrics considered in Section 3.10, but all of the proofs I know of require a significant amount of calculation.

show that

$$R(t_1, x_1) \leq \exp(d_{g(t_1)}(x_0, x_1)^2/2\tau_1) R(t_0, x_0). \tag{3.428}$$

for all κ-solutions and all spacetime points $(t_0, x_0), (t_1, x_1)$ with $t_1 < t_0$.

Now we use the Harnack inequality to obtain some further control on the reduced length function $l_{(t_0,x_0)}(t_1, x_1)$. Recall that this quantity takes the form

$$l_{(t_0,x_0)}(t_1, x_1) = \frac{1}{\sqrt{2\tau_1}} \int_0^{\tau_1} \sqrt{\tau}(|X|_g^2 + R)\, d\tau \tag{3.429}$$

where $X := \gamma'$ and $\gamma : [0, \tau_1] \to M$ is a minimising $\mathcal{L}$-geodesic, which in particular means that it obeys the $\mathcal{L}$-geodesic equation

$$\nabla_X X - \frac{1}{2}\nabla_X R + \frac{1}{2\tau} X + 2\mathrm{Ric}(X, \cdot)^* = 0 \tag{3.430}$$

(see (3.337)). Using (3.430) and the chain rule, we can compute the *total derivative* $\frac{d}{d\tau}|X|^2$ along the path γ as

$$\frac{d}{d\tau}|X|^2 = \nabla_X R - \frac{1}{\tau}|X|^2 - 2\mathrm{Ric}(X, X). \tag{3.431}$$

On the other hand, the Harnack inequality (3.424) (with X replaced by $2X$) lets us bound the total derivative of R:

$$\frac{d}{d\tau} R \leq -\nabla_X R + 2\mathrm{Ric}(X, X). \tag{3.432}$$

We add (3.431) and (3.432) and rearrange to obtain

$$\frac{d}{d\tau}[\tau^{3/2}(|X|^2 + R)] \leq \frac{3}{2}\sqrt{\tau}(|X|^2 + R) - \tau^{1/2}|X|^2. \tag{3.433}$$

We (somewhat crudely) discard the non-negative $\tau^{1/2}|X|^2$ term and integrate in τ using (3.429) to obtain

$$\tau_1^{3/2}(|X(\tau_1)|^2 + R) \leq \frac{3}{2} 2\tau^{1/2} l(t_1, x_1) \tag{3.434}$$

where we abbreviate $l_{(t_0,x_0)}$ as l. Using the first variation formulae for reduced length (see equations (3.371), (3.372)), as well as the nonnegativity of R (and hence of l), we obtain the useful inequalities

$$0 \leq |\nabla l|^2 + R \leq \frac{3l}{\tau} \tag{3.435}$$

and

$$-\frac{2l}{\tau} \leq \partial_\tau l \leq \frac{l}{\tau}. \tag{3.436}$$

Informally, this means that at any given point (t_1, x_1) to the past of (t_0, x_0), l is roughly constant at spatial scales $\sqrt{\tau}$ and at temporal scales τ. Furthermore, if l is bounded, then one has bounded normalised curvature at such scales.

3.14.4. A splitting theorem. Our final application of these ideas (or more precisely, of the strong maximum principle) will be to establish a dichotomy[**Ha1986**] for 3-dimensional κ-solutions: either their Ricci curvature is strictly positive, or the solution splits locally as the product of a line with a two-dimensional solution.

Proposition 3.14.6. [**Ha1986**] *Let $t \mapsto (M, g(t))$ be a three-dimensional κ-solution. Suppose that the Ricci tensor has a zero eigenvalue at some point (t_0, x_0). Then on the slab $(-\infty, t_0) \times M$, the Ricci flow locally splits as the product of a two-dimensional Ricci flow and a line.*

Proof. The first stage is to show that the Ricci tensor has a zero eigenvalue on all of $(-\infty, t_0) \times M$. Let $0 \leq \nu \leq \mu \leq \lambda$ denote the three eigenvalues of the Riemann tensor $\mathcal{T}$ as viewed in an orthonormal frame (as in Section 3.4), thus a zero eigenvalue of the Ricci tensor is equivalent to $\nu + \mu = 0$. Suppose for contradiction that at some time $t_1 < t_0$, this quantity is not identically zero, thus we can find some non-negative scalar function $h(t_1, \cdot) : M \to \mathbf{R}^+$, not identically zero, such that $\nu + \mu \geq h$ at time t_1. We then extend h by the heat equation, so by the strong maximum principle h is strictly positive for all times after t_1. From the convexity of the functional $\nu + \mu$ (which one can view as the minimal trace of $\mathcal{T}$ over two-dimensional subspaces), we see that the set $\{(\mathcal{T}, h) : \nu + \mu \geq h\}$ cuts out a fibrewise convex parallel subset of a suitable vector bundle over $[t_1, t_0] \times M$ (in the sense of the tensor maximum principle, Proposition 3.4.5), which one can easily check to be preserved under the ODE associated to the simultaneous evolution of (3.423) and the scalar heat equation for h.

Applying the tensor maximum principle we conclude that $\nu + \mu \geq h$ for all times in $[t_1, t_0]$, and in particular that $\nu + \mu$ is non-zero at

(t_0, x_0), a contradiction. Thus the Ricci curvature must have a zero eigenvalue on all of $(-\infty, t_0) \times M$, thus $\nu = \mu = 0$ on this slab. On the other hand, from Exercise 3.13.7 we must have $\lambda > 0$ throughout this slab.

The symmetric rank 2 tensor $\mathcal{T}$ thus has rank 1 at every point, and thus locally can be expressed in the form $\mathcal{T} = av \otimes v$ for some smooth non-zero scalar a and a unit vector field v. (If M was orientable, one could extend this vector field to be global). The equation (3.423) then becomes

$$\begin{aligned} a_t v \otimes v + av \otimes v_t + av_t \otimes v &= (\Delta a + a^2) v \otimes v \\ &\quad + (\nabla^\alpha a)(v \otimes \nabla_\alpha v + \nabla_\alpha v \otimes v) + 2a \nabla_\alpha v \otimes \nabla^\alpha v. \end{aligned} \tag{3.437}$$

Since v is a unit vector field, the vector fields $\nabla_X v$ are orthogonal to v for every v. Thus we can restrict to the component of (3.437) that is completely orthogonal to v, and conclude (since a is nonzero) that $\nabla_\alpha v \otimes \nabla^\alpha v = 0$. If we then inspect the component of (3.437) which is partially orthogonal to v, we also learn that $v_t = 0$. Expressing the left-hand side in an orthonormal basis as the sum of rank one positive semi-definite matrices, we easily conclude that $\nabla_\alpha v = 0$, i.e. v is parallel to the connection. This implies that the dual one-form $v^* \in \Gamma(T^*M)$ is closed and hence locally exact; thus v is locally the gradient of some potential function f. From this we easily see that the flow locally splits as the product of a two-dimensional flow (on a level set of f) and a line (the flow lines of v), and then it is easy to verify that the two-dimensional flow is a Ricci flow, as claimed. □

Remark 3.14.7. One cannot always extend this local splitting to a global one, due to topological obstructions; consider for instance the oriented round shrinking cylinder quotient (Example 3.12.6). One could also imagine the product of a round shrinking S^2 and a static circle S^1, in which the null eigenvector of the Ricci tensor splits off as a circle rather than a line; but this is not a κ-solution because it becomes collapsed at large scales in the distant past.

Remark 3.14.8. The above splitting analysis can be carried out in any dimension; for instance, one can show that the rank of the Riemann tensor is a constant for any ancient solution with bounded

non-negative Riemann curvature. For this and further splitting results in this case, see [**Ha1986**].

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/05/19`.

3.15. Stationary points of Perelman entropy or reduced volume are gradient shrinking solitons

We continue our study of κ-solutions. In Section 3.13 we primarily exploited the non-negative curvature of such solutions; in this lecture and the next, we primarily exploit the ancient nature of these solutions, together with the finer analysis of the two scale-invariant monotone quantities we possess (Perelman entropy and Perelman reduced volume) to obtain a important scaling limit of κ-solutions, the *asymptotic gradient shrinking soliton* of such a solution.

The main idea here is to exploit what I have called the *infinite convergence principle* (Section 1.3 of *Structure and Randomness*): that every bounded monotone sequence converges. In the context of κ-solutions, we can apply this principle to either of our monotone quantities: the *Perelman entropy*

(3.438)
$$\mu(g(t),\tau) := \inf\{\mathcal{W}(M,g(t),f,\tau) : \int_M (4\pi\tau)^{-d/2} e^{-f}\ d\mu = 1\}$$

where $\tau := -t$ is the backwards time variable and

(3.439)
$$\mathcal{W}(M,g(t),f,\tau) := \int_M (\tau(|\nabla f|^2 + R) + f - d)(4\pi\tau)^{-d/2} e^{-f}\ d\mu,$$

or the*Perelman reduced volume*

$$\tilde{V}_{(0,x_0)}(-\tau) := \tau^{-d/2} \int_M e^{-l_{(0,x_0)}(-\tau,x)}\ d\mu(x) \tag{3.440}$$

where $x_0 \in M$ is a fixed base point. As pointed out in Section 3.11, these quantities are related, and both are non-increasing in τ. The reduced volume starts off at $(4\pi)^{d/2}$ when $\tau = 0$, and so it must it approach some asymptotic limit[103] $0 \leq \tilde{V}_{(0,x_0)}(-\infty) \leq (4\pi)^{d/2}$ as

[103]We will later see that this limit is *strictly* between 0 and $(4\pi)^{d/2}$.

$\tau \to -\infty$. On the other hand, the reduced volume is invariant under the scaling

$$g^{(\lambda)}(t) := \frac{1}{\lambda^2} g(\lambda^2 t), \tag{3.441}$$

in the sense that

$$\tilde{V}^{(\lambda)}_{(0,x_0)}(-\tau) = \tilde{V}_{(0,x_0)}(-\lambda^2 \tau). \tag{3.442}$$

Thus, as we send $\lambda \to \infty$, the reduced volumes of the rescaled flows $t \mapsto (M, g^{(\lambda)}(t))$ (which are also κ-solutions) converge pointwise to a constant $\tilde{V}_{(0,x_0)}(-\infty)$.

Suppose that we could somehow "take a limit" of the flows $t \mapsto (M, g^{(\lambda)}(t))$ (or perhaps a subsequence of such flows) and obtain some limiting flow $t \mapsto (M^{(\infty)}, g^{(\infty)}(t))$. *Formally*, such a flow would then have a constant reduced volume of $\tilde{V}_{(0,x_0)}(-\infty)$. On the other hand, the reduced volume is monotone. If we could have a criterion as to when the reduced volume became stationary, we could thus classify all possible limiting flows $t \mapsto (M^{(\infty)}, g^{(\infty)}(t))$, and thus obtain information about the asymptotic behaviour of κ-solutions (at least along a subsequence of scales going to infinity).

We will carry out this program more formally in the next lecture, in which we define the concept of an *asymptotic gradient-shrinking soliton* of a κ-solution.

In this section, we content ourselves with a key step in this program, namely to characterise when the Perelman entropy or Perelman reduced volume becomes stationary; this requires us to revisit the theory we have built up in the last few section. It turns out that, roughly speaking, this only happens when the solution is a gradient shrinking soliton, thus at any given time $-\tau$ one has an equation of the form $\mathrm{Ric} + \mathrm{Hess}(f) = \lambda g$ for some $f : M \to \mathbf{R}$ and $\lambda > 0$. Our computations here will be somewhat formal in nature; we will make them more rigorous in the next lecture. The material here is largely based on [**MoTi2007**], [**Pe2002**]. Closely related treatments also appear in [**KlLo2006**], [**CaZh2006**].

3.15.1. Stationarity of the Perelman entropy. We begin with a discussion of the Perelman entropy, which is simpler than the Perelman reduced volume but which will serve as a model for the latter.

To simplify the exposition we shall argue at a formal level, assuming all integrals converge, that all functions are smooth, all infima are actually attained, etc.

In Exercise 3.9.9, we already saw that if $f : (-\infty, 0] \times M \to \mathbf{R}$ solves the nonlinear backwards heat equation

$$f_\tau = \Delta f - |\nabla f|_g^2 + R - \frac{d}{2\tau} \tag{3.443}$$

then the quantity $\mathcal{W}(M, g(t), f, \tau)$ obeyed the monotonicity formula

$$\frac{d}{d\tau}\mathcal{W}(M, g(t), f, \tau) = -\int_M H \, d\mu \tag{3.444}$$

where H is the non-negative quantity

$$H := 2\tau|\mathrm{Ric} + \mathrm{Hess}(f) - \frac{1}{2\tau}g|^2(4\pi\tau)^{-d/2}e^{-f}. \tag{3.445}$$

In terms of the function $u := (4\pi\tau)^{-d/2}e^{-f}$, we also recall that (3.443) can be rewritten as the adjoint heat equation $u_\tau = \Delta u + Ru$. In particular, we see that if $\mathcal{W}(M, g(t), f, \tau)$ is ever stationary at some time τ, then the solution must obey the gradient shrinking soliton equation

$$\mathrm{Ric} + \mathrm{Hess}(f) = \frac{1}{2\tau}g \tag{3.446}$$

at that time τ. Using the uniqueness properties of Ricci flow (and of the backwards heat equation), one can then show that (3.446) persists for all subsequent times. Formally at least, this argument also shows that the Perelman reduced entropy $\mu(M, g, \tau)$ can only be stationary on gradient shrinking solitons.

Let us analyse the monotonicity formula (3.444) further. If we write

(3.447)
$$v := (\tau(|\nabla f|^2 + R) + f - d)(4\pi\tau)^{-d/2}e^{-f} = (\tau(|\nabla f|^2 + R) + f - d)u$$

then (3.444) asserts that

$$\frac{d}{d\tau}\int_M v \, d\mu = -\int_M H \, d\mu. \tag{3.448}$$

Since $\frac{d}{d\tau}d\mu = R \, d\mu$, we thus see that $\partial_\tau v$ must equal $-H - Rv$ plus a quantity which integrates to zero (i.e. a divergence). Given this, and

given the fact that u (which is a close relative to v) obeys the adjoint heat equation), the following fact is then not so surprising:

Exercise 3.15.1. With the above assumptions, show that v obeys the forced adjoint heat equation

$$v_\tau = \Delta v - Rv - H. \tag{3.449}$$

3.15.2. Stationarity in the Bishop-Gromov reduced volume. Before we turn to the monotonicity of the Perelman reduced volume, we first consider the simpler model case of the Bishop-Gromov reduced volume (Corollary 3.10.3). An inspection of the proof of that result reveals that the key point was to establish the pointwise inequality

$$\Delta r \leq \frac{d-1}{r} \tag{3.450}$$

on a manifold (M, g) of non-negative Ricci curvature $\mathrm{Ric} \geq 0$, where $r := d(x, x_0)$ for some fixed origin x_0. To simplify the exposition let us assume we are inside the injectivity radius, and away from the origin, to avoid any issues with lack of smoothness. We gave a proof of (3.450) using the second variation formula

$$\frac{d^2}{ds^2} E(\gamma)|_{s=0} = \int_0^1 |\nabla_X Y|_g^2 - g(\mathrm{Riem}(X,Y)Y,X)\ dt \tag{3.451}$$

whenever $\gamma : (-\varepsilon, \varepsilon) \times [0,1] \to M$ is a geodesic at $s = 0$, with $X = \partial_t \gamma$ and $Y := \partial_s \gamma$; (see (3.327)). From this (and the first variation formula) we obtain the inequality

$$\mathrm{Hess}(r)(v,v) \leq \int_0^1 |\nabla_X Y|_g^2 - g(\mathrm{Riem}(X,Y)Y,X)\ dt \tag{3.452}$$

for any vector field Y along the minimising geodesic from x_0 to x that equals 0 at $t = 0$ and equals v at $t = 1$.

Of course, the only way that (3.452) can be an equality is if Y minimises the right-hand side subject to the constraints just mentioned. A standard calculus of variations computation lets one extract the *Euler-Lagrange equation* for this variational problem:

Exercise 3.15.2. Show that if (3.452) is obeyed with equality, then Y must obey the *Jacobi equation*

$$\nabla_X \nabla_X Y + \mathrm{Riem}(Y, X)X = 0. \tag{3.453}$$

Vector fields obeying (3.453) are known as *Jacobi fields*.

Recall from Section **??** that the inequality (3.450) was derived by applying (3.452) for v in an arbitrary orthonormal frame, and with $Y(t) := tv$, where v was extended by parallel transport along γ (thus $\nabla_X v = 0$). Thus, in order for (3.450) to be obeyed with equality, the fields $Y(t) = tv$ must be a Jacobi field for each v. Applying (3.453), and noting that $X = \partial_r$, we conclude that we must have

$$\mathrm{Riem}(\cdot, \partial_r)\partial_r = 0 \tag{3.454}$$

along γ in order for (3.450) to be obeyed with equality. The converse is also true:

Exercise 3.15.3. Establish the identity

$$\nabla_{\partial r}\mathrm{Hess}(r)_{\alpha\beta} + \mathrm{Hess}(r)_{\alpha\gamma}\mathrm{Hess}(r)^{\gamma}_{\beta} = -\mathrm{Riem}_{\alpha\gamma\delta\beta}(\partial_r)^{\gamma}(\partial_r)^{\delta} \tag{3.455}$$

in the injectivity region, and conclude (3.450) is true with equality whenever (3.454) holds along the minimising geodesic γ.

As a consequence of the above analysis, we see that the Bishop-Gromov reduced volume can only be stationary on a sphere when (3.454) holds on the ball within that sphere.

We can also use the theory of Jacobi fields to get a more precise formula for $\mathrm{Hess}(r)$ (and hence Δr). The key observation is that the Jacobi equation (3.453) can be written as the linearisation

$$\nabla_Y(\nabla_X X) = 0 \tag{3.456}$$

of the geodesic equation $\nabla_X X = 0$. This is ultimately unsurprising, since the geodesic equation and the Jacobi equation come from the Euler-Lagrange equations for the energy functional and a quantity related to a variation of the energy functional. But it allows us (at least inside the injectivity region, which also turns out (again, unsurprisingly) to be the region where the boundary value problem for the Jacobi equation always has unique solutions), to view Jacobi fields as the infinitesimal deformation field of geodesics.

Now let $\gamma : (-\varepsilon, \varepsilon) \times [0,1] \to M$ be a family of geodesics $\gamma_s : [0,1] \to M$ from x_0 to $x(s)$, so that $\nabla_X X = 0$ and so (by (3.456)) Y is a Jacobi field[104] for each s with $Y(s,0) = 0$ and $Y(s,1) = v(s) := x'(s)$. The first variation formula (i.e. the Gauss lemma $\nabla r = X(1)$, see Lemma 3.8.4) then gives

$$\nabla_v r = g(X(\cdot,1), v) \tag{3.457}$$

and differentiating this again gives

$$\nabla_v \nabla_v r = g(\nabla_v X(\cdot,1), v) + g(X(\cdot,1), \nabla_v v). \tag{3.458}$$

Expanding out the left-hand side by the product rule and using (3.457) and the torsion-free identity $\nabla_Y X = \nabla_X Y$ we conclude the second variation formula

$$\mathrm{Hess}(r)(v,v) = g(\nabla_X Y(1), v) \tag{3.459}$$

whenever Y is a Jacobi field along the minimal geodesic γ from x_0 to x with $Y(0) = 0$ and $Y(1) = v$, and whenever one is inside the injectivity region.

Exercise 3.15.4. Let Y be a Jacobi field with $Y(0) = 0$ and $Y(1) = v$, and suppose one is inside the injectivity region. Use (3.459) and (3.453) to show that (3.452) in fact holds with equality, thus providing a converse to Exercise 3.14.2. *Hint*: apply the fundamental theorem of calculus to the right-hand side of (3.459).

3.15.3. Constancy of the Perelman reduced volume. We can obtain parabolic analogues of the above elliptic arguments to conclude when the Perelman reduced volume is stationary. Again, let us argue formally and assume that we are working inside the injectivity domain from a point $(0, x_0)$.

Write $l = l_{(0,x_0)}$. Recall from Section **??** that the proof of monotonicity of reduced volume relied on the inequality

$$\partial_\tau l - \Delta l + |\nabla l|^2 - R + \frac{d}{2\tau} \geq 0 \tag{3.460}$$

[104] In general one no longer expects to have Y be geodesic in the s direction, i.e. $\nabla_Y Y$ need not be zero, but this will not concern us.

x which in turn followed from the three equalities and estimates

$$\nabla l = X \tag{3.461}$$

$$\partial_\tau l = \frac{1}{2}R - \frac{1}{2}|X|_g^2 - \frac{1}{2\tau}l \tag{3.462}$$

$$\Delta l \leq \frac{d}{2\tau} + \frac{1}{2}|X|_g^2 - \frac{1}{2}R - \frac{1}{2\tau}l. \tag{3.463}$$

Thus, in order for the reduced volume to be stationary at some time $t = -\tau_1$, one must have (3.460) (or equivalently, (3.463)) holding with equality throughout M at this time.

It is convenient to normalise $\tau_1 = 1$. Recall from Section **??** that the proof of (3.463) proceeded via the second variation formula

$$\begin{aligned}\frac{d^2}{ds^2}\mathcal{L}(\gamma) = \int_0^1 \sqrt{\tau}(&\mathrm{Hess}(R)(Y,Y) \\ &+ 2|\nabla_X Y|^2 - 2g(\mathrm{Riem}(X,Y)Y,X))\ d\tau\end{aligned} \tag{3.464}$$

applied to the vector field $Y := \sqrt{\tau}v$, where v obeys the ODE

$$\nabla_X v = -\mathrm{Ric}(v,\cdot)^*; v(s,1) = x'(s). \tag{3.465}$$

As in the elliptic case, equality in (3.463) can only hold if Y obeys the Euler-Lagrange equation for the right-hand side of (3.460), which can be computed to be

$$\begin{aligned}\nabla_X\nabla_X Y + \mathrm{Riem}(Y,X)X - \frac{1}{2}\nabla_Y(\nabla R) \\ + \frac{1}{2\tau}\nabla_X Y + 2(\nabla_Y \mathrm{Ric})(X,\cdot)^* + 2\mathrm{Ric}(\nabla_X Y,\cdot)^* = 0.\end{aligned} \tag{3.466}$$

Solutions of (3.466) are known as *$\mathcal{L}$-Jacobi fields*. As in the elliptic case, this equation can be rewritten as the linearisation

$$\nabla_Y G(X) = 0 \tag{3.467}$$

of the $\mathcal{L}$-geodesic equation $G(X) = 0$, where

$$G(X) := \nabla_X X - \frac{1}{2}\nabla R + \frac{1}{2\tau}X + 2\mathrm{Ric}(X,\cdot)^* \tag{3.468}$$

was introduced in Section **??**.

Exercise 3.15.5. Verify (3.466) and (3.467).

If $\gamma : (-\varepsilon, \varepsilon) \times [0, \tau_1] \to M$ is now a smooth family of minimising $\mathcal{L}$-geodesics from $(0, x_0)$ to $(-\tau_1, x_1(s))$, then the variation field $Y = \partial_s X$ is an $\mathcal{L}$-Jacobi field by (3.467) (and conversely, inside the region of injectivity, any Jacobi field on a minimising geodesic can be extended locally to such a smooth family. The first variation formula (3.461) gives

$$\nabla_v l = g(X, v) \tag{3.469}$$

where $v(s) := x_1'(s) = Y(s, 1)$, and so on differentiating again and arguing as in the elliptic case we obtain

$$\mathrm{Hess}(l)(v, v) = g(\nabla_X Y(1), v) \tag{3.470}$$

whenever Y is an $\mathcal{L}$-Jacobi field with $Y(0) = 0$ and $Y(1) = v$.

Exercise 3.15.6. Show (using (3.468) and the fundamental theorem of calculus, as in Exercise 3.14.4) that (3.470) is equal to (3.464).

Now we return to our analysis of when the reduced volume is stationary at $\tau = 1$. We had found in that case that the vector field $Y := \sqrt{\tau} v$, where v solved (3.465), must be a Jacobi field. Combining this with (3.470) we conclude that

$$\mathrm{Hess}(l)(v, v) = \frac{1}{2}|v|^2 - \mathrm{Ric}(v, v) \tag{3.471}$$

for any v, or in other words that

$$\mathrm{Ric} + \mathrm{Hess}(l) = \frac{1}{2} g. \tag{3.472}$$

This is for time $\tau = 1$; rescaling the above analysis gives more generally that

$$\mathrm{Ric} + \mathrm{Hess}(l) = \frac{1}{2\tau} g. \tag{3.473}$$

We thus conclude (formally[105], at least) that whenever the reduced volume is stationary, then the manifold is a gradient shrinking soliton (at that instant in time, at least) with potential function given by the reduced length.

[105] The computation is only formal at present, because we have not addressed the issue of what to do on the $\mathcal{L}$-cut locus.

Exercise 3.15.7. If (3.463) is obeyed with equality, show that the function $f := l$ obeys (3.443) and that $\mathcal{W}(M, g(t), f, \tau) = 0$ (cf. the computations in Section 3.11.6). From this and (3.444), deduce another (formal) proof of (3.473) whenever the reduced volume is stationary on an open time interval.

Remark 3.15.1. We have just seen that in the case of stationary reduced volume, the function f that appears in the entropy functional can be taken to be equal to the reduced length l. In general, one can take f to be a function bounded from above by the reduced length; see [**Pe2002**, Corollary 9.5].

3.15.4. Ricci flows of maximal reduced volume. Recall that the reduced volume $\tilde{V}_{(0,x_0)}(-\tau)$ is equal to $(4\pi)^{d/2}$ in the case of Euclidean space, and converges to this value in the limit $\tau \to 0$ in the case of complete Ricci flows of bounded curvature (this can be shown by an analysis of the $\mathcal{L}$-exponential map for small values of τ, as discussed in Section 3.11). From this and the monotonicity of reduced volume we conclude that

$$\tilde{V}_{(0,x_0)}(-\tau) \leq (4\pi)^{d/2} \tag{3.474}$$

for all such flows. We now characterise when equality occurs:

Theorem 3.15.2. *Suppose that $t \mapsto (M, g(t))$ is a connected Ricci flow of bounded curvature on $[-\tau_1, 0]$ for some $\tau_1 > 0$, such that* (3.474) *is obeyed with equality at the initial time $-\tau_1$ for some point $x_0 \in M$. Then M is Euclidean.*

Proof. We give a sketch here only; full details can be found in [**MoTi2007**, Proposition 7.27]. An inspection of the proof of monotonicity of reduced volume (especially as viewed through the $\mathcal{L}$-exponential map, as in Section 3.11) reveals that the domain of injectivity $\Omega \subset T_{x_0}M$ of the exponential map must have full measure, otherwise there will be a loss of reduced volume. The previous analysis then reveals that the equation (3.469) must hold outside of the cut locus; as l is Lipschitz and the manifold is smooth, one can then take limits and conclude that (3.469) holds globally (and so l is in fact smooth).

Combining (3.469) with the Ricci flow equation we obtain

$$\frac{d}{dt}g = \mathcal{L}_{\nabla l}g - \frac{1}{\tau}g, \tag{3.475}$$

thus the metric is shrinking and also deforming by a vector field. In particular this gives an analogous equation for the magnitude $|\mathrm{Riem}|_g^2$ of curvature (see equations (3.459), (3.463)):

$$\frac{d}{dt}|\mathrm{Riem}|_g^2 = \nabla_{\nabla l}|\mathrm{Riem}|_g^2 + \frac{1}{\tau}|\mathrm{Riem}|_g^2. \tag{3.476}$$

A maximum principle argument (which of course works in the absence of the dissipation term) then shows that if $\sup_x |\mathrm{Riem}|_g^2$ is strictly positive at one time, then it blows up as $\tau \to 0$ (like $1/\tau$, in fact), which is absurd; and so this supremum must always be zero. In other words, the manifold is flat, and is therefore the quotient of $\mathbf{R}^d$ by some discrete subgroup. But as the exponential map is almost always in the injectivity domain, this subgroup must be trivial, and the claim follows. □

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/05/21`.

3.16. Geometric limits of Ricci flows, and asymptotic gradient shrinking solitons

We now begin using the theory established in the last two lectures to rigorously extract an asymptotic gradient shrinking soliton from the scaling limit of any given κ-solution. This will require a number of new tools, including the notion of a *geometric limit* of pointed Ricci flows $t \mapsto (M, g(t), p)$, which can be viewed as the analogue of the *Gromov-Hausdorff limit* in the category of smooth Riemannian flows. A key result here is *Hamilton's compactness theorem*[**Ha1995**]: a sequence of complete pointed non-collapsed Ricci flows with uniform bounds on curvature will have a subsequence which converges geometrically to another Ricci flow. This result, which one can view as an analogue of the *Arzelá-Ascoli theorem* for Ricci flows, relies on some parabolic regularity estimates for Ricci flow due to Shi[**Sh1989**].

Next, we use the estimates on reduced length from the Harnack inequality analysis in Section 3.13 to locate some good regions of spacetime of a κ-solution in which to do the asymptotic analysis.

Rescaling these regions and applying Hamilton's compactness theorem (relying heavily here on the κ-noncollapsed nature of such solutions) we extract a limit. Formally, the reduced volume is now constant and so Section 3.14 suggests that this limit is a gradient soliton; however, some care is required to make this argument rigorous. In the next section we shall study such solitons, which will then reveal important information about the original κ-solution.

Our treatment here is primarily based on [**MoTi2007**], [**Ye2008**]; other treatments can be found in [**Pe2002**], [**KlLo2006**], [**CaZh2006**], [**ChLuNi2006**]. See also the foundational papers of [**Sh1989**], [**Ha1995**].

3.16.1. Geometric limits. To develop the theory of geometric limits for pointed Ricci flows $t \mapsto (M, g(t), p)$, we begin by studying such limits in the simpler context of pointed Riemannian manifolds (M, g, p), i.e. a Riemannian manifold (M, g) together with a point $p \in M$, which we shall call the *origin* or distinguished point of the manifold. To simplify the discussion, let us restrict attention to complete Riemannian manifolds (though for later analysis we will eventually have to deal with incomplete manifolds).

Definition 3.16.1 (Geometric limits)**.** A sequence (M_n, g_n, p_n) of pointed d-dimensional connected complete Riemannian manifolds is said to converge geometrically to another pointed d-dimensional connected complete Riemannian manifold $(M_\infty, g_\infty, p_\infty)$ if there exists a sequence $V_1 \subset V_2 \subset \ldots$ of connected neighbourhoods of p_∞ increasing to M_∞ (i.e. $\bigcup_n V_n = M_\infty$) and a sequence of smooth embeddings $\phi_n : V_n \to M_n$ mapping p_∞ to p_n such that

(1) The closure of each V_n is compact and contained in V_{n+1} (note that this implies that every compact subset of M_∞ will be contained in V_n for sufficiently large n);

(2) The pullback metric $\phi_n^* g_n$ converges in the $C^\infty_{loc}(M_\infty)$ topology to g_∞ (i.e. all derivatives of the metric converge uniformly on compact sets).

Example 3.16.2. The pointed round d-sphere of radius R converges geometrically to the pointed Euclidean space $\mathbf{R}^d$ as $R \to \infty$. Note how this example shows that the geometric limit of compact manifolds can be non-compact.

Example 3.16.3. If (M, g) is Hamilton's cigar (Example 3.9.4), and p_n is a sequence on M tending to infinity, then (M, g, p_n) converges geometrically to the pointed round 2-cylinder.

Example 3.16.4. The d-torus of length $1/n$ does not converge to a geometric limit as $n \to \infty$, despite being flat. More generally, the sequence needs to be locally uniformly non-collapsed in order to have a geometric limit.

Exercise 3.16.1. Show that the geometric limit $(M_\infty, g_\infty, p_\infty)$ of a sequence (M_n, g_n, p_n), if it exists, is unique up to (pointed) isometry.

Geometric limits, as their name suggests, tend to preserve all (local) "geometric" or "intrinsic" information about the manifold, although *global* information of this type can be lost. Here is a typical example:

Exercise 3.16.2. Suppose that (M_n, g_n, p_n) converges geometrically to $(M_\infty, g_\infty, p_\infty)$. Show that $\mathrm{Vol}_{g_\infty}(B_{g_\infty}(p_\infty, r)) = \lim_{n\to\infty} \mathrm{Vol}_{g_n}(B_{g_n}(p_n, r))$ for every $0 < r < \infty$, and that we have the Fatou-type inequality $\mathrm{Vol}_{g_\infty}(M_\infty) \leq \liminf_{n\to\infty} \mathrm{Vol}_{g_n}(M_n)$. Give an example to show that the latter inequality can be strict.

Here is the basic compactness theorem for such limits.

Theorem 3.16.5 (Compactness theorem)**.** *Let (M_n, g_n, p_n) be a sequence of connected complete Riemannian d-dimensional manifolds. Assume that*

(1) *(Uniform bounds on curvature and derivatives) For all $k, r_0 \geq 0$, one has the pointwise bound $|\nabla^k \mathrm{Riem}_n|_{g_n} \leq C_{k,r_0}$ on the ball $B_n(p_n, r_0)$ for all sufficiently large n and some constant C_{k,r_0}.*

(2) *(Uniform non-collapsing) For every $r_0 > 0$ there exists $\delta, \kappa > 0$ such that $\mathrm{Vol}_n(x, r) \geq \kappa r^d$ for all $x \in B_n(p_n, r_0)$ and $0 < r \leq \delta$, and all sufficiently large n.*

Then, after passing to a subsequence if necessary, the sequence (M_n, g_n, p_n) has a geometric limit.

Proof. (Sketch) Let $r_0 > 0$ be an arbitrary radius. From Cheeger's lemma (Theorem 3.8.9) and hypothesis 2, we know that the injectivity radius on $B_n(p_n, 2r_0)$ is bounded from below by some small $\delta > 0$ for sufficiently large n. Also, from the curvature bounds and Bishop-Gromov comparison geometry (Lemma 3.10.1) we know that the volume of $B_n(p_n, 2r_0)$ is uniformly bounded from above for sufficiently large n.

Now find a maximal $\delta/4$-net $x_{n,1}, \ldots, x_{n,k}$ of $B_n(p_n, r_0)$, thus the balls $B_n(x_{n,1}, \delta/8), \ldots, B_n(x_{n,k}, \delta/8)$ are disjoint and the balls $B_n(x_{n,1}, \delta/4), \ldots, B_n(x_{n,k}, \delta/4)$ cover $B_n(p_n, r_0)$. Volume counting shows that k is bounded for all sufficiently large n; by passing to a subsequence we may assume that it is constant. Similarly we may assume that all the distances $d_n(x_{n,i}, x_{n,j})$ converge to a limit. Using the exponential map and some arbitrary identification of tangent spaces with $\mathbf{R}^d$, we can identify each ball $B_n(x_{n,i}, \delta/2)$ with the standard Euclidean ball of radius $\delta/2$. Any pair $x_{n,i}, x_{n,j}$ of separation less than $\delta/2$ induces a smooth transition map from the Euclidean ball of radius $\delta/2$ into some subset of $\mathbf{R}^d$, which can be shown by comparison geometry to be uniformly bounded in C^∞ norms; applying the (C^∞ version of the) Arzelá-Ascoli theorem we may thus pass to a subsequence and assume that all these transition maps converge in C^∞ to a limit. It is then a routine matter to glue together all the limit transition maps to fashion an incomplete manifold to which the balls $B_n(p_0, r_0)$ converge geometrically (up to errors of $O(\delta)$ at the boundary). Furthermore, as one increases r_0, one can show (by a modification of Exercise 3.15.1) that these limits are compatible. Now letting r_0 go to infinity (and using the usual diagonalisation trick on all the subsequences obtained), and then gluing together all the incomplete limits obtained, one can create the full geometric limit. □

Remark 3.16.6. One could use ultrafilters here in place of subsequences (cf. Section 1.5 of *Structure and Randomness*), but this does not significantly affect any of the arguments.

Now we turn to geometric limits of pointed Ricci flows (Ricci flows $t \mapsto (M, g(t))$ with a specified origin $p \in M$).

Definition 3.16.7. Let $t \mapsto (M_n, g_n(t), p_n)$ be a sequence of pointed d-dimensional complete connected Ricci flows, each on its own time interval I_n. We say that a pointed d-dimensional complete connected Ricci flow $t \mapsto (M_\infty, g_\infty(t), p_\infty)$ on a time interval I is a *geometric limit* of this sequence if

(1) Every compact subinterval of I is contained in I_n for all sufficiently large n.

(2) There exists neighbourhoods V_n of p_∞ as in Definition 3.15.1, compact time intervals $J_n \subset I$ increasing to I, and smooth embeddings $\phi_n : V_n \to M_n$ preserving the origin such that the pullback of the flow g_n to $J_n \times V_n$ converges in spacetime C^∞_{loc} to g_∞.

Exercise 3.16.3. Show that if a sequence of κ-noncollapsed Ricci flows (with a uniform value of κ) converges geometrically to another Ricci flow, then the limit flow is also κ-noncollapsed.

Now we present Hamilton's compactness theorem for Ricci flows, which requires less regularity hypotheses than Theorem 3.15.5 due to the parabolic smoothing effects of Ricci flow (as captured by Shi's estimates, see Theorem 3.15.13).

Theorem 3.16.8 (Hamilton compactness theorem)**.** [**Ha1995**] *Let $t \mapsto (M_n, g_n(t), p_n)$ and I_n be as in Definition 3.15.7, and let I be an open interval obeying hypothesis 1 of that definition. Let $t_0 \in I$ be a time. Suppose that*

(1) *For every compact subinterval J of I containing t_0 and every $r > 0$, one has the curvature bound $|\mathrm{Riem}_n|_{g_n} \leq K$ on the cylinder $J \times B_{g_n(t_0)}(p_n, r)$ for some $K = K(J, r)$ and all sufficiently large n; and*

(2) *One has the non-collapsing bound $\mathrm{Vol}_{g_n(t_0)}(B_{g_n(t_0)}(p_n, r)) \geq \kappa r^d$ for some $r > 0$ and $\kappa > 0$, and all sufficiently large n.*

Then some subsequence of $t \mapsto (M_n, g_n(t), p_n)$ converges geometrically to a limit $t \mapsto (M_\infty, g_\infty(t), p_\infty)$ on I.

Proof. By Shi's estimates (Theorem 3.15.13) we can upgrade the bound on curvature in hypothesis 1 to bounds on derivatives of curvature. Indeed, these estimates imply that for any J, r as in that hypothesis, and any $k \geq 0$, we have $|\nabla^k \mathrm{Riem}_n|_{g_n} \leq K$ for some $K = K(J, r, k)$ and sufficiently large n.

Now we restrict to the time slice $t = t_0$ and apply Theorem 3.15.5. Passing to a subsequence, we can assume that $(M_n, g_n(t_0), p_n)$ converges geometrically to a limit $(M_\infty, g_\infty(t_0), p_\infty)$.

For any radius r and any compact J in I containing t_0, we can pull back the flow $t \mapsto (M_n, g_n(t), p_n)$ to a (spatially incomplete) flow $t \mapsto (B_{g_\infty(t_0)}(p_\infty, r), \tilde{g}_n(t), p_\infty)$ on the cylinder $J \times B_{g_\infty(t_0)}(p_\infty, r)$ for sufficiently large n. By construction, $\tilde{g}_n(t_0)$ converges in $C^\infty_{\mathrm{loc}}(B_{g_\infty(t_0)}(p_\infty, r))$ norm to $g_\infty(t_0)$; in particular, it is uniformly bounded in each of the seminorms of this space. Also, each $t \mapsto \tilde{g}_n(t)$ is a Ricci flow with uniform bounds on any derivative of curvature for sufficiently large n.

Exercise 3.16.4. Using these facts, show that the sequence of flows $t \mapsto \tilde{g}_n$ is uniformly bounded in each of the seminorms of $C^\infty_{\mathrm{loc}}(J \times B_{g_\infty(t_0)}(p_\infty, r))$ for each fixed J, r, and for n sufficiently large.

By using the Arzelá-Ascoli theorem as before, we may thus pass to a further subsequence and assume that $t \mapsto \tilde{g}_n(t)$ converges in $C^\infty_{\mathrm{loc}}(J \times B_{g_\infty(t_0)}(p_\infty, r))$ to a limiting flow $t \mapsto g_\infty(t)$. Clearly this limit is a Ricci flow. Letting $r \to \infty$ and pasting together the resulting limits one obtains[106] the desired geometric limit. □

3.16.2. Locating an asymptotic gradient shrinking soliton. We now return to the study of κ-solutions $t \mapsto (M, g(t))$. We pick an arbitrary point $x_0 \in M$ and consider the reduced length function $l = l_{(0,x_0)}$. Recall[107] from (3.375) that we had

$$\inf_{x \in M} l(t, x) < d/2 \tag{3.477}$$

[106] One has to verify that every geodesic in $(M_\infty, g_\infty(t), p_\infty)$ starting from p_∞ can be extended to any desired length, thus establishing completeness by the *Hopf-Rinow theorem*, but this is easy to establish given all the uniform bounds on the metric and curvature, and their derivatives.

[107] This bound was obtained from the parabolic inequality $\partial_\tau l \geq \Delta l + \frac{l - (d/2)}{\tau}$ and the maximum principle.

for every $t < 0$. Thus we can find a sequence $(-\tau_n, x_n) \in (-\infty, 0] \times M$ with $\tau_n \to \infty$ such that

$$l(-\tau_n, x_n) = O(1). \tag{3.478}$$

Now recall that as a consequence of Hamilton's Harnack inequality, we have the pointwise estimates

$$0 \leq |\nabla l|^2 + R \leq \frac{3l}{\tau} \tag{3.479}$$

and

$$-\frac{2l}{\tau} \leq \partial_t l \leq \frac{l}{\tau} \tag{3.480}$$

(see equations (3.435), (3.436)). From these bounds and Gronwall's inequality, one easily sees that we can extend (3.478) to say that

$$l(-\tau, x) = O_r(1) \tag{3.481}$$

for any $(-\tau, x)$ in the cylinder $[-\tau_n/r, -r\tau_n] \times B_{g(-\tau')}(x_n, r\sqrt{\tau_n})$ and any $r \geq 1$ and $\tau_n/r \leq \tau' \leq r\tau_n$. Applying (3.479) once more, together with the hypothesis of non-negative curvature more, we also obtain bounded normalised curvature on this cylinder:

$$|\mathrm{Riem}(-\tau, x)|_g = O_r(\tau^{-1}). \tag{3.482}$$

If we thus introduce the rescaled flow $t \mapsto (M_n, g_n(t), p_n)$ by setting $M_n := M$, $p_n := x_n$, and $g_n(t) := t_n g(tt_n)$, we see that these flows obey hypothesis 1 of Theorem 3.15.8. Also, since the original κ-solutions are κ-noncollapsed, so are their rescalings, which (in conjunction with hypothesis 1) gives us hypothesis 2. We can thus invoke Theorem 3.15.8 and assume (after passing to a subsequence) that the rescaled flows converge geometrically to an ancient Ricci flow $t \mapsto (M_\infty, g_\infty(t), p_\infty)$ on the time interval $t \in (-\infty, 0)$. From Exercise 3.15.3 we see that this limit is also κ-noncollapsed. Since the rescaled flows have non-negative curvature, the limit flow has non-negative curvature also[108].

Let $l_n : (-\infty, 0) \times M_n \to \mathbf{R}$ be the rescaled length function, thus $l_n(t, x) := l(tt_n, x)$. From (3.481) we see that l_n is uniformly bounded

[108]Note however that we do not expect in general that $(M_\infty, g_\infty(t))$ has bounded curvature (for instance, if the original κ-solution was a round shrinking sphere terminating at the unit radius sphere, the limit object would be a round shrinking sphere terminating at a point). In particular we do not expect (M_∞, g_∞) to be a κ-solution.

on compact subsets of $(-\infty, 0) \times M_\infty$ for n sufficiently large (where we identify compact subsets of M_∞ with subsets of M_n for n large enough). By the rescaled versions of (3.480) and (3.481) we also see that $|\nabla l_n|_{g_\infty}, |\partial_t l_n|$ is also uniformly bounded on such compact sets for sufficiently large n; thus the l_n are uniformly Lipschitz on each compact set. Applying the Arzelá-Ascoli theorem and passing to a subsequence, we may thus assume that the l_n converge uniformly on compact sets to some limit l_∞, which is then locally Lipschitz.

Remark 3.16.9. We do not attempt to interpret l_∞ as a reduced length function arising from some point at time $t = 0$; indeed we expect the limiting flow to develop a singularity at this time.

We know that the reduced volume $\int_M \tau^{-d/2} e^{-l}\, d\mu$ is non-increasing in τ and ranges between 0 and $(4\pi)^{d/2}$, and so converges to a limit $\tilde{V}(-\infty)$ between 0 and $(4\pi)^{d/2}$. This limit cannot equal $(4\pi)^{d/2}$ since this would mean that the κ-solution is flat (by Theorem 3.14.2), which is absurd. The limit cannot be zero either, since the bounds (3.481) and the non-collapsing ensure a uniform lower bound on the reduced volume. By rescaling, we conclude that

$$\int_{M_n} \tau^{-d/2} e^{-l_n}\, d\mu_n \to \tilde{V}(-\infty) \tag{3.483}$$

for each fixed $\tau > 0$.

Let us now argue informally, and then return to make the argument rigorous later. Formally taking limits in (3.483), we conclude that

$$\int_{M_\infty} \tau^{-d/2} e^{-l_\infty}\, d\mu_\infty = \tilde{V}(-\infty). \tag{3.484}$$

On the other hand, from the proof of the monotonicity of reduced volume from Section **??** we have (formally, at least)

$$\partial_\tau l - \Delta l + |\nabla l|_g^2 - R + \frac{d}{2\tau} \geq 0 \tag{3.485}$$

and hence by rescaling

$$\partial_\tau l_n - \Delta l_n + |\nabla l_n|_{g_n}^2 - R + \frac{d}{2\tau} \geq 0. \tag{3.486}$$

Formally taking limits, we obtain

$$\partial_\tau l_\infty - \Delta l_\infty + |\nabla l_\infty|^2_{g_\infty} - R_\infty + \frac{d}{2\tau} \geq 0. \tag{3.487}$$

We can rewrite this as the assertion that $\tau^{-d/2}e^{-l_\infty}$ is a subsolution of the backwards heat equation:

$$(\partial_\tau - \Delta_{g_\infty} - R_\infty)(\tau^{-d/2}e^{-l_\infty}) \leq 0. \tag{3.488}$$

This (formally) implies that the left-hand side of (3.484) is nonincreasing in τ. On the other hand, this quantity is constant in τ; and so (3.488) must be obeyed with equality, and thus

$$\partial_\tau l_\infty - \Delta l_\infty + |\nabla l_\infty|^2 - R_\infty + \frac{d}{2\tau} = 0. \tag{3.489}$$

Also, recall from Section **??** that

$$\partial_\tau l = \frac{1}{2}R - \frac{1}{2}|\nabla l|^2_g - \frac{1}{2\tau}l. \tag{3.490}$$

Rescaling and taking limits, we formally conclude that the same is true for l_∞;

$$\partial_\tau l_\infty = \frac{1}{2}R_\infty - \frac{1}{2}|\nabla l_\infty|^2_{g_\infty} - \frac{1}{2\tau}l_\infty. \tag{3.491}$$

From (3.490) and (3.491) we obtain that the Perelman $\mathcal{W}$-functional

$$\mathcal{W}(M_\infty, g_\infty(t), l_\infty, \tau) = \int_{M_\infty} (\tau(|\nabla l_\infty|^2 + R_\infty) + l_\infty - d)(4\pi\tau)^{-d/2}e^{-l_\infty}\, d\mu_\infty \tag{3.492}$$

vanishes (cf. Section 3.11.6). In particular, it is constant. On the other hand, by (3.489) and the monotonicity formula for this functional (see Exercise 3.9.9) we have

$$\frac{\partial}{\partial\tau}\mathcal{W}(M_\infty, g_\infty(t), l_\infty, \tau) = -\int_M 2\tau|\mathrm{Ric}_\infty + \mathrm{Hess}(l_\infty) - \frac{1}{2\tau}g_\infty|^2_{g_\infty}(4\pi\tau)^{-d/2}e^{-l_\infty}\, d\mu_\infty. \tag{3.493}$$

Combining this with the vanishing of (3.492) we thus conclude that

$$\mathrm{Ric}_\infty + \mathrm{Hess}(l_\infty) - \frac{1}{2\tau}g_\infty = 0 \tag{3.494}$$

and thus $t \mapsto (M_\infty, g_\infty(t))$ is a gradient shrinking soliton as desired.

3.16.3. Making the argument rigorous I. Spatial localisation. Now we turn to the (surprisingly delicate) task of justifying the steps from (3.483) to (3.494).

The first task is to deduce (3.484) from (3.483). From the dominated convergence theorem it is not difficult to show that

$$\int_{B_n(p_n,r)} \tau^{-d/2} e^{-l_n}\, d\mu_n \to \int_{B_\infty(p_\infty,r)} \tau^{-d/2} e^{-l_\infty}\, d\mu_\infty \tag{3.495}$$

for any fixed τ and r; the difficulty is to prevent the escape of mass[109] of e^{-l_n} to spatial infinity.

In order to prevent such an escape, one needs a lower bound[110] on $l_n(-\tau, x)$ when $d_{g_n(-\tau)}(x_n, x)$ is large. The problem is equivalent to that of upper bounding $d_{g_n(-\tau)}(x_n, x)$ in terms of $l_n(-\tau, x)$. To do this we need some control on quantities related to the distance function at extremely large distances. Remarkably, such bounds are possible. We begin with a lemma from [**Pe2002**] (related to an earlier argument in [**Ha1993b**]).

Lemma 3.16.10. *Let (M, g) be a d-dimensional Riemannian manifold, let $x, y \in M$, and let $r > 0$. Suppose that $\mathrm{Ric} \leq K$ on the balls $B(x, r)$ and $B(y, r)$. Then for any minimising geodesic γ connecting x and y, we have $\int_\gamma \mathrm{Ric}(X, X) \ll_d Kr + r^{-1}$, where $X := \gamma'$ is the velocity field.*

Proof. We may assume that $d(x, y) \geq 2r$, since the claim is trivial otherwise. We recall from (3.327) the second variation formula

$$\frac{d^2}{ds^2} E(\gamma) = \int_\gamma |\nabla_X Y|^2 - g(\mathrm{Riem}(X, Y)X, Y) \tag{3.496}$$

whenever one deforms a geodesic γ along a vector field Y. Since γ is minimising, the left-hand side of (3.496) is non-negative when Y vanishes at the endpoints of γ. Now let v be any unit vector at x, transported by parallel transport along γ. Setting $Y(t)$ to equal tv/r

[109] Fatou's lemma will tell us that the left-hand side of (3.484) is less than or equal to the right, but this is not enough for our application.

[110] Note that estimates such as (3.479), (3.480) only provide upper bounds on $l_n(-\tau, x)$.

when $0 \leq t \leq r$, equal to v when $r \leq t \leq d(x,y) - r$, and equal to $(d(x,y) - t)v/r$ when $d(x,y) - r \leq t \leq d(x,y)$, we conclude that

$$\int_\gamma g(\mathrm{Riem}(X,Y)X,Y) \ll r^{-1}. \tag{3.497}$$

Letting v vary over an orthonormal frame and summing, we soon obtain the claim. □

The above lemma, combined with the Ricci flow equation, gives an upper bound as to how rapidly the distance function can grow as one goes backwards in time.

Corollary 3.16.11. *Let $t \mapsto (M, g(t))$ be a d-dimensional Ricci flow, let $x, y \in M$, let t be a time, and let $r > 0$. Suppose that* $\mathrm{Ric} \leq K$ *on $B_{g(t)}(x,r)$ and on $B_{g(t)}(y,r)$. Then $\frac{d}{d\tau} d_{g(t)}(x,y) \ll_d Kr + r^{-1}$ (in the sense of forward difference quotients).*

Using this estimate, we can now obtain a bound on distance in terms of reduced length.

Proposition 3.16.12. *Let $t \mapsto (M, g(t))$ be a d-dimensional κ-solution, let $x_0, p, p' \in M$, and $\tau_1 > 0$. Then*

$$\frac{d_{g(-\tau_1)}(p,p')^2}{\tau_1} \ll_d 1 + l_{(0,x_0)}(-\tau_1, p) + l_{(0,x_0)}(-\tau_1, p'). \tag{3.498}$$

Proof. We use an argument of Ye[**Ye2008**]. Write A for the expression inside the $O_d()$ on the right-hand side, and let $\gamma, \gamma' : [0, \tau_1] \to M$ be minimising $\mathcal{L}$-geodesics from x_0 to p, p' respectively. By the fundamental theorem of calculus, we have

$$d_{g(-\tau_1)}(p,p') = \int_0^{\tau_1} \frac{d}{d\tau} d_{g(-\tau)}(\gamma(-\tau), \gamma'(-\tau))\, d\tau. \tag{3.499}$$

Using (3.479) and the $\mathcal{L}$-Gauss lemma $\nabla l = X$ we see that γ, γ' move at speed $O(A^{1/2}/\tau^{1/2})$, and that all curvature tensors are $O(A/\tau)$ in a $O(\tau^{1/2}/A^{1/2})$-neighbourhood of either curve. Applying Corollary 3.15.11, the chain rule, and the Gauss lemma (Lemma 3.8.4) we conclude that

$$\frac{d}{d\tau} d_{g(-\tau)}(\gamma(-\tau), \gamma'(-\tau)) \ll_d A^{1/2}/\tau^{1/2}; \tag{3.500}$$

inserting this into (3.499) we obtain the claim. □

Combining this proposition with (3.481) and rescaling we see that we have a bound of the form

$$l_n(-\tau, x) \geq c d_{g_n(-\tau)}(p_n, x)^2/\tau - O_d(1) \tag{3.501}$$

for all x and some $c = c_d > 0$; taking limits we also obtain

$$l_\infty(-\tau, x) \geq c d_{g_\infty(-\tau)}(p_\infty, x)^2/\tau - O_d(1). \tag{3.502}$$

On the other hand, from the Bishop-Gromov inequality we know that balls of radius r in either $(M_n, g_n(-\tau))$ or $(M_\infty, g_\infty(-\tau))$ have volume $O_d(r^d)$. These facts are enough to establish that the portion of (3.483) or (3.484) outside of the ball of radius r decays exponentially fast in r, uniformly in n, and this allows us to take limits in (3.495) as $r \to \infty$ to deduce (3.484) from (3.483).

3.16.4. Making the argument rigorous II. Parabolic inequality for l_∞. The next major task in making the previous arguments rigorous is to justify the passage from (3.486) to (3.487). First of all, because of the $\mathcal{L}$-cut locus, (3.486) is only valid in the sense of distributions. We would like to take limits and conclude that (3.487) holds in the sense of distributions as well. There is no difficulty taking limits with the linear terms $\partial_\tau l_n - \Delta l_n$ in (3.486), or in the zeroth order terms $-R_n + \frac{d}{\tau}$; the only problem is in justifying the limit from $|\nabla l_n|^2_{g_n}$ to $|\nabla l_\infty|^2_{g_n}$. We know that the l_n are uniformly locally Lipschitz, and converge locally uniformly to l_∞; but this is unfortunately not enough to ensure that $|\nabla l_n|^2_{g_n}$ converges in the sense of distributions to $|\nabla l_\infty|^2_{g_n}$, due to possible high frequency oscillations in l_n. To give a toy counterexample, the one-dimensional functions $l_n(x) := \frac{1}{n}\sin(nx)$ are uniformly Lipschitz and converge uniformly to zero, but $|\frac{d}{dx} l_n|^2 = \cos^2(nx)$ converges in the distributional sense to $\frac{1}{2}$ rather than zero.

Since $\nabla l_n - \nabla l_\infty$ is bounded and converges distributionally to zero, it will be locally asymptotically orthogonal l_∞. From this and Pythagoras' theorem we obtain

$$\lim_{n\to\infty} |\nabla l_n|^2_{g_n} = |\nabla l_\infty|^2_g + \lim_{n\to\infty} |\nabla l_\infty - \nabla l_n|^2_{g_n} \tag{3.503}$$

in the sense of distributions, where we pass to a subsequence in order to make the limits on both sides exist[111]. The task is now to show that there is not enough oscillation to cause the second term on the right-hand side to be non-vanishing.

To do this, we observe that (3.486) provides an upper bound on $\Delta_{g_n} l_n$; indeed on any fixed compact set in $(-\infty, 0) \times M_\infty$, we have $\Delta_{g_n} l_n \leq O(1)$. This one-sided bound on the Laplacian is enough to rule out the oscillation problem. Indeed, as l_n converges locally uniformly to l_∞, we see that

$$\limsup_{n\to\infty} \int_{M_\infty} \phi(l_\infty - l_n + \varepsilon_n)\Delta_{g_n} l_n \, d\mu_n \leq 0 \tag{3.504}$$

for any non-negative bump function ϕ and $\varepsilon_n \to 0$ chosen so that $l_\infty - l_n + \varepsilon_n \geq 0$ on the support of ϕ. Integrating by parts and disposing of a lower order term, we conclude that

$$\limsup_{n\to\infty} \int_{M_\infty} \phi \langle \nabla(l_n - l_\infty), \nabla l_n \rangle_{g_n} \, d\mu_n \leq 0. \tag{3.505}$$

On the other hand, since $\nabla(l_n - l_\infty)$ is bounded converges weakly to zero, one has

$$\limsup_{n\to\infty} \int_{M} \phi \langle \nabla(l_n - l_\infty), \nabla l_\infty \rangle_{g_\infty} \, d\mu_\infty \to 0. \tag{3.506}$$

One can easily replace g_∞ and μ_∞ here by g_n and μ_n. Combining (3.505) and (3.506) we conclude that the second term on the RHS of (3.503) is non-positive in the sense of distributions. But it is clearly also non-negative, and so it vanishes as required.

This gives (3.487); as a by-product of the argument we have also established the useful fact

$$\lim_{n\to\infty} |\nabla l_n|^2_{g_n} = |\nabla l_\infty|^2_g \tag{3.507}$$

in the sense of distributions. Combining this with the growth bounds (3.501), (3.502) on l_n and l_∞ from the previous section (which give exponential decay bounds on e^{-l_n}, e^{-l_∞} and their first derivatives),

[111] Note that g_n converges locally uniformly to g and so there is no difficulty passing back and forth between those metrics.

it is not too difficult[112] to then justify the remaining steps (3.488)-(3.494) of the argument rigorously; see [**MoTi2007**, Section 9.2] for full details.

3.16.5. The asymptotic gradient shrinking soliton is not flat. Finally, we show that the asymptotic gradient shrinking soliton $t \mapsto (M_\infty, g_\infty(t))$ is non-trivial in the sense that its curvature is not identically zero at some time. For if the curvature did vanish everywhere at time t, then the equation (3.494) simplifies to $\mathrm{Hess}(l_\infty) = \frac{1}{2\tau} g_\infty$. On the other hand, being flat, $(M_\infty, g_\infty(t))$ is the quotient of Euclidean space $\mathbf{R}^d$ by some discrete subgroup. Lifting l_∞ up to this space, we thus see that f is quadratic, and more precisely is equal to $|x|^2/4\tau$ plus an affine-linear function. Thus f has no periodicity whatsoever and so the above-mentioned discrete subgroup is trivial. If we now apply (3.484) we see that $\tilde{V}(-\tau) = (4\pi)^{d/2}$. But on the other hand, as the original κ-solution was not flat, its reduced volume was strictly less than $(4\pi)^{d/2}$ by Theorem 3.14.2, a contradiction. Thus the asymptotic gradient soliton is not flat.

3.16.6. Appendix: Shi's derivative estimates. The purpose of this appendix is to prove the following estimate of Shi[**Sh1989**].

Theorem 3.16.13. [**Sh1989**] *Suppose that $t \mapsto (M, g(t))$ is a complete d-dimensional Ricci flow on the time interval $[0,T]$, and that on the cylinder $[0,T] \times B_{g(0)}(x_0, r_0)$ one has the pointwise curvature bound $|\mathrm{Riem}|_g \leq K$. Then on any slightly smaller cylinder $(0,T] \times B_{g(0)}(x_0, (1-\varepsilon)r_0)$ one has the curvature bounds $|\nabla^k \mathrm{Riem}|_g = O(t^{-k/2})$ for any $k \geq 0$, where the implied constant depends on $d, T, r_0, K, \varepsilon, k$.*

Proof. (Sketch) We induct on k. The case $k = 0$ is trivial, so suppose that $k \geq 1$ and that the claim has already been proven for all smaller values of k. We allow all implied constants in the O() notation to depend on $d, T, r_0, K, \varepsilon, k$. We refer to $[0,T] \times B_{g(0)}(x_0, r_0)$ and $(0,T] \times B_{g(0)}(x_0, (1-\varepsilon)r_0)$ as the "large cylinder" and "small cylinder" respectively.

[112] Note that once one reaches (3.489), one has a nonlinear heat equation for l_∞, and it is not difficult to use the smoothing effects of the heat kernel to then show that the locally Lipschitz function l_∞ is in fact smooth.

We make some reductions. It is easy to see that we can take r_0 and T to be small.

Since $|\dot{g}|_g = 2|\mathrm{Ric}| = O(1)$ on the cylinder, we see that the metric at later times of the large cylinder is comparable to the initial metric up to multiplicative constants. The curvature bound tells us that if r_0 is small, then we are inside the conjugacy radius; pulling back under the exponential map, we may thus assume that the exponential map from x_0 is injective on the large cylinder. Let $r = d_{g(t)}(x_0, x)$ be the *time-varying* radial coordinate; observe that the annulus $\{(1 - 2\varepsilon/3)r_0 \leq r \leq (1 - \varepsilon/3)r_0\}$ will be contained between the large cylinder and small cylinder for T small enough.

Exercise 3.16.5. Show that if r_0 and T are small enough, then $|\mathrm{Hess}(r)|_g = O(1/r)$ on the large cylinder.

Let $\eta = \eta(r)$ be a smooth non-negative radial cutoff to the large cylinder that equals 1 on the small cylinder. From the above exercise, the Gauss lemma (Lemma 3.8.4), and the chain rule, we see that $|\nabla\eta|_g, |\partial_t\eta|, |\mathrm{Hess}\eta|_g, \Delta\eta = O(1)$.

Now we study the heat equation obeyed by the "energy densities" $|\nabla^m \mathrm{Riem}|_g^2$ for various m.

Exercise 3.16.6 (Bochner-Weitzenböck type estimate)**.** For any $m \geq 0$, show that

(3.508)
$$(\partial_t - \Delta)|\nabla^m \mathrm{Riem}|_g^2 \ll \sum_{j=0}^{m} |\nabla^j \mathrm{Riem}|_g |\nabla^{m-j}\mathrm{Riem}|_g |\nabla^m \mathrm{Riem}|_g - 2|\nabla^{m+1}\mathrm{Riem}|_g^2.$$

Hint: start with the equation $\partial_t \mathrm{Riem} = \Delta\mathrm{Riem} + \mathcal{O}(g^{-1}\mathrm{Riem}^2)$ and use the product rule and the definition of curvature repeatedly.

From this exercise and the induction hypothesis we see that

(3.509)
$$(\partial_t - \Delta)[\eta^{2m+2}t^m|\nabla^m \mathrm{Riem}|_g^2] \leq O(1) + O(\eta^{2m}t^{m-1}|\nabla^m \mathrm{Riem}|_g^2) - 2\eta^{2m+2}t^m|\nabla^{m+1}\mathrm{Riem}|_g^2$$

for all $0 \leq m \leq k$, with the understanding that the second term on the right-hand side is absent when $m = 0$. Telescoping this, we can thus find an expression

$$E := \sum_{m=0}^{k} C^{-m} \eta^{2m+2} t^m |\nabla^m \mathrm{Riem}|_g^2 \tag{3.510}$$

for some sufficiently large positive constant C, which obeys the heat equation $(\partial_t - \Delta)E \leq O(1)$. Also, by hypothesis we have E=O(1) at time zero. Applying the maximum principle, we obtain the claim. □

Exercise 3.16.7. Suppose that in the hypotheses of Shi's theorem that we also have $|\nabla^j \mathrm{Riem}| = O(1)$ for $0 \leq j \leq m$ on the large cylinder at time zero. Conclude that we have $|\nabla^j \mathrm{Riem}| = O(1 + t^{-(j-m)/2})$ on the small cylinder for all j.

Exercise 3.16.8. Let (M, g) be a smooth compact manifold, and let $u : [0, T] \times M \to \mathbf{R}$ be a bounded solution to the heat equation $\partial_t u = \Delta u$ which obeys a pointwise bound $|u(0)| \leq K$ at time zero. Establish the bounds $|\nabla^k u|_g = O(t^{-k/2})$ on the spacetime $[0, T] \times M$ and all $k \geq 0$, where the implied constant depends on (M, g), K, T, and k.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/05/27`. Thanks to Paul Smith and Dan for corrections.

3.17. Classification of asymptotic gradient shrinking solitons

In Section 3.15, we showed that every κ-solution generated at least one asymptotic gradient shrinking soliton $t \mapsto (M, g(t))$. This soliton is known to have the following properties:

(1) It is ancient: t ranges over $(-\infty, 0)$.

(2) It is a Ricci flow.

(3) M is complete and connected.

(4) The Riemann curvature is non-negative (though it could theoretically be unbounded).

(5) $\frac{dR}{dt}$ is non-negative.

(6) M is κ-noncollapsed.

(7) M is not flat.

(8) It obeys the gradient shrinking soliton equation

$$\mathrm{Ric} + \mathrm{Hess}(f) = \frac{1}{2\tau} g \tag{3.511}$$

for some smooth f.

The main result of this section is to classify all such solutions in low dimension:

Theorem 3.17.1 (Classification of asymptotic gradient shrinking solitons)**.** *Let $t \mapsto (M, g(t))$ be as above, and suppose that the dimension d is at most 3. Then one of the following is true (up to isometry and rescaling):*

(1) *$d = 2, 3$ and M is a round shrinking spherical space form (i.e. a round shrinking S^2, S^3, $\mathcal{RP}^2$, or S^3/Γ for some finite group Γ acting freely on S^3).*

(2) *$d = 3$ and M is the round shrinking cylinder $S^2 \times \mathbf{R}$ or the oriented or unoriented quotient of this cylinder by an involution.*

The case $d = 2$ of this theorem is due to [**Ha1988**]; the compact $d = 3$ case is due to [**Iv1993**]; and the full $d = 3$ case was sketched out in [**Pe2002**]. In higher dimension, partial results towards the full classification (and also relaxing many of the hypotheses 1-8) have been established in [**PeWy2007**], [**NiWa2007**], [**Na2007**]; these papers also give alternate proofs of Perelman's classification.

To prove this theorem, we induct on dimension. In 1 dimension, all manifolds are flat and so the claim is trivial. We will thus take $d = 2$ or $d = 3$, and assume that the result has already been established for dimension $d - 1$. We will then split into several cases:

(1) Case 1: Ricci curvature has a zero eigenvector at some point. In this case we can use Hamilton's splitting theorem to reduce the dimension by one, at which point we can use the induction hypothesis.

(2) Case 2: Manifold noncompact, and Ricci curvature is positive and unbounded. In this case we can take a further geometric limit (using some Toponogov theory on the asymptotics of rays in a positively curved manifold) which is a round cylinder (or quotient thereof), and also a gradient steady soliton. One can easily rule out such an object by studying the potential function of that soliton on a closed loop.

(3) Case 3: Manifold noncompact, and Ricci curvature is positive and bounded. Here we shall follow the gradient curves of f using some identities arising from the gradient shrinking soliton equation to get a contradiction.

(4) Case 4: Manifold compact, and curvature positive. Here we shall use Hamilton's rounding theorem[**Ha1982**] to show that one is a round shrinking sphere or spherical space form.

We will follow the treatment in [**MoTi2007**] of Perelman's argument[**Pe2002**]; see also [**KlLo2006**], [**CaZh2006**], [**ChLuNi2006**] for other treatments of this argument.

3.17.1. Case 1: Ricci curvature degenerates at some point. This case cannot happen in two dimensions. Indeed, since the Ricci curvature is conformal in this case, the only way that the Ricci curvature can degenerate is if the scalar curvature vanishes also. But then the strong maximum principle (Exercise 3.13.7) forces the gradient shrinking soliton to be flat at all sufficiently early times (and hence at all times), a contradiction[113].

So now suppose that we are in three dimensions with bounded Ricci curvature, and a point where the Ricci curvature vanishes. Then by Hamilton's splitting theorem (Proposition 3.13.6) the gradient shrinking soliton locally splits into the product of a two-dimensional flow and a line (for sufficiently early times, at least), with the Ricci curvature being degenerate along these lines that foliate the flow[114]. In particular, from (3.511) we see that $\mathrm{Hess}(f)$ is constant and strictly

[113] It turns out that this application of strong maximum principle can be extended to cover the case in which one does not have bounded curvature.

[114] Again, one has to extend the strong maximum principle argument to cover the case of unbounded curvature, but this can be done.

positive along these lines; in other words, f is strictly convex (and quadratic) along these lines. As a consequence, the lines cannot loop back upon themselves.

By lifting to a double cover if necessary, we can find a global unit vector field X along these lines, thus $\mathrm{Ric}(X,\cdot) = 0$ and $\nabla X = 0$. If we set $F := \nabla_X f$, we conclude from (3.511) that $\nabla F = X/2\tau$, thus the level sets of F have X as a unit normal. Thus, at any fixed time, we use F to *globally* split the manifold M (or a double cover thereof) as the product of a line and a two-dimensional manifold (given by the level sets of F). Applying the induction hypothesis, we conclude that M (or a double cover) is a product of a line and a round shrinking S^2 or $\mathcal{RP}^2$ (as these are the only two-dimensional spherical space forms), at which point we end up in alternative 2 of Theorem 3.16.1. (We initially establish this fact only for sufficiently early times, but then by uniqueness of Ricci flow one obtains it for late times also.)

Remark 3.17.2. We can also proceed here using the global splitting theorem from [**Ha1986**, Lemma 9.1].

3.17.2. Case 2: Manifold non-compact, curvature positive and unbounded. Now we handle the case in which M is non-compact (and in particular has a meaningful notion of convergence to spatial infinity) with Ricci curvature strictly positive and unbounded. In particular one has a sequence of points $x_n \to \infty$ in M such that

$$R(x_n) d(x_0, x_n)^2 \to \infty \tag{3.512}$$

at some time (which we can normalise to $t = -1$), where we arbitrarily pick an origin $x_0 \in M$. Thus the curvature is not decaying as fast as $1/d(x_0,\cdot)^2$ at infinity, and may even be unbounded. Henceforth we normalise t as $t = -1$ and write g for $g(-1)$.

The basic idea here is to look at the rescaled pointed manifolds $(M_n, g_n, p_n) := (M, R(x_n)^{1/2} g, x_n)$ and extract a limit in which the original base point x_0 has now been sent off to infinity (thanks to (3.16.2)). There is a technical obstacle to doing this, though, which is that the rescaled manifolds have bounded curvature at x_n (indeed, it has been normalised to equal 1) but might have unbounded curvature at nearby points y_n with respect to the rescaled metric (i.e. points y_n within distance $O(R(x_n)^{-1/2}) = o(d(x_n, x_0))$ in the original

metric) because such points may have significantly higher curvature than x_n (e.g. $R(y_n) \geq 4R(x_n)$). But it is easy to resolve this: simply pick y_n instead of x_n. Now y_n may itself be close to another point of even higher curvature, but we can then move that point instead. We can continue in this manner, moving in a geometrically decreasing sequence of distances, until we stop (which we must, since the manifold is smooth and so curvature is locally bounded). The precise result of this "point-picking argument", originally due to Hamilton, that we will need is as follows:

Exercise 3.17.1 (Point picking lemma)**.** Assuming that (3.16.2) holds for some sequence $x_n \to \infty$, show that there exists another sequence $y_n \to \infty$ also obeying (3.16.2), and such that for any $A > 1$, and for all n sufficiently large depending on A, we have $R(z_n) \leq 4R(y_n)$ for all $z_n \in B(y_n, AR(y_n)^{-1/2})$. If the original manifold had unbounded curvature, show that we can also ensure that $R(y_n) \to \infty$.

We now let y_n be as above, and consider the rescaled manifolds $(M_n, g_n, p_n) := (M, R(y_n)^{1/2} g, y_n)$. Using Hamilton's compactness theorem (Theorem 3.15.8) we may assume that these manifolds converge geometrically to a limit $(M_\infty, g_\infty, p_\infty)$ of nonnegative Riemann curvature whose scalar curvature is at most 4 (and is equal to 1 at p_∞); in particular the limit has bounded curvature. From the analogue of (3.16.2) for y_n we have $d_{g_n}(x_0, p_n) \to \infty$, and so x_0 has "escaped to infinity" in the limit M_∞ (this shows in particular that M_∞ is non-compact).

Let $r_n := d(x_0, y_n)$, thus $r_n \to \infty$. By refining this sequence we may assume that we have rapid growth in the sense that $r_n = o(r_{n+1})$. Let $x_0 y_n$ be a minimising geodesic from x_0 to y_n; by compactness we may assume that the direction of $x_0 y_n$ at x_0 is convergent. In particular, the angle subtended between $x_0 y_n$ and $x_0 y_{n+1}$ is $o(1)$. If we let $y_n y_{n+1}$ be a minimising geodesic from y_n to y_{n+1}, we thus see from the triangle inequality and the cosine rule (Lemma 3.16.6) that

$$d(y_n, y_{n+1}) = r_{n+1} - r_n + o(r_n). \tag{3.513}$$

Using the cosine rule again, we see that the angle subtended between $x_0 y_n$ and $y_n y_{n+1}$ is $\pi - o(1)$. Using relative Toponogov comparison (Exercise 3.16.3) we see that the rays $x_0 y_n$ and $y_n y_{n+1}$ asymptotically

form a minimising geodesic, in the sense that $d(z, y_n) + d(y_n, w) = d(z, w) + o(1)$ for any z, w at a bounded distance away from y_n on $x_0 y_n$ and $y_n y_{n+1}$ respectively. From this, we see in the limit $(M_\infty, g_\infty, p_\infty)$ that there exists a minimising geodesic line through p_∞. But by the Cheeger-Gromoll splitting theorem (Theorem 3.16.8) we see that M_∞ splits into the product of a line and a manifold Σ of one dimension less. This cannot happen in the two-dimensional case d=2, since Σ becomes one-dimensional and thus flat, and M_∞ has non-zero curvature at p_∞ (indeed, its scalar curvature is equal to 1). So we can now assume $d = 3$.

We have only taken limits at time $t = -1$. But we can use Hamilton's compactness theorem (Theorem 3.15.8) again (using the property $\partial_t R \geq 0$) and extend M_∞ to a Ricci flow backwards in time from $t = -1$; this is a limit of rescaled versions of (M, g, y_n) by $R(y_n)^{1/2}$. Since M was originally a gradient shrinking soliton, and $R(y_n)$ is going to infinity, the limit $(M_\infty, g_\infty, p_\infty)$ can be shown to be a gradient steady soliton: $\mathrm{Ric}_\infty + \mathrm{Hess}(f_\infty) = 0$ for some f_∞.

Since M_∞ had bounded curvature at time $t = -1$, it had bounded curvature for all previous times also. Since the Ricci curvature is vanishing along one direction, we can now apply the Case 1 argument and show that M_∞ is the product of a line and a round shrinking S^2 or $\mathcal{RP}^2$. In particular, M_∞ contains closed geodesic loops γ on which the Ricci curvature $\mathrm{Ric}(X, X)$ is strictly positive. From the gradient steady equation, this means that f_∞ is strictly concave on this loop, which is absurd. Thus this situation does not occur.

Remark 3.17.3. In [**MoTi2007**], the contradiction was obtained using the soul theorem (Theorem 3.12.17), and a rather non-trivial result asserting that complete manifolds of non-negative sectional curvature cannot contain arbitrarily small necks, but the above argument seems to be somewhat shorter. An even simpler argument (avoiding the use of the splitting theorem altogether) was given in [**Na2007**], based on the observation (from (3.511)) that the normalised gradient vector field $\nabla \mathrm{f}/|\nabla f|$ of the potential function becomes increasingly parallel to the connection if $|\nabla f|$ goes to infinity. We thank Peter Petersen for pointing out Naber's argument to us.

3.17.3. Case 3: M noncompact, curvature positive and bounded. Now we assume that M is compact, with Ricci curvature strictly positive but also bounded. By Lemma 3.15.10, we conclude in particular that

$$\int_\gamma \mathrm{Ric}(X,X)\ ds \leq C \tag{3.514}$$

for some C and all minimising geodesics (thus the Ricci curvature must decay along long geodesics). On the other hand, along such a geodesic, we see from (3.511) that

$$\frac{d^2}{ds^2} f(\gamma(s)) = \frac{1}{2} - \mathrm{Ric}(X,X). \tag{3.515}$$

From (3.16.3) and (3.515) we see that $\nabla_X f(\gamma(s))$ increases like $s/2$ as $s \to \infty$. Similarly, if E is any vector field orthogonal to X and transported by parallel transport along γ, an application of Cauchy-Schwarz, (3.16.3), and the bounded curvature hypothesis gives

$$|\int_\gamma \mathrm{Ric}(X,E)|\ ds \leq C'|\gamma|^{1/2} \tag{3.516}$$

while (3.511) gives

$$\frac{d}{ds} \nabla_E f(\gamma s)) = -\mathrm{Ric}(X,E) \tag{3.517}$$

and so $\nabla_E f(\gamma(s))$ grows like at most $O(s^{1/2})$ as $s \to \infty$. These bounds ensure that f goes to $+\infty$ at infinity (in particular, it is a *proper* function), and that there exist curves following the gradient ∇f of f which go to infinity.

On the other hand, using the identity

$$\nabla_\alpha R = 2\mathrm{Ric}_{\alpha\beta} \nabla^\beta f \tag{3.518}$$

(see (3.425)) we see that $\nabla_{\nabla f} R > 0$, thus R is increasing along gradient flow curves. In particular, $R(\infty) := \limsup_{x\to\infty} R(x)$ is strictly positive (and finite, since curvature is bounded).

As a consequence, we can repeat the point-picking arguments from Case 2 and extract a sequence of points $y_n \to \infty$ for which (M, g, x_n) converges geometrically to a limit $(M_\infty, g_\infty, p_\infty)$, which has scalar curvature $R(\infty)$ at p_∞. Since M is a gradient shrinking soliton on $(-\infty, 0)$, one can show that M_∞ is also. By repeating

the Case 2 analysis one can show that M_∞ is also a round shrinking $\mathbf{R} \times S^2$ or $\mathbf{R} \times \mathcal{RP}^2$. Since these solitons have scalar curvature 1 at time -1, we thus have $R(\infty) = 1$.

For sake of argument let us take M to be the round shrinking cylinder $\mathbf{R} \times S^2$; the other case is similar but with all areas divided by a factor of two[115].

Now we return to the original gradient shrinking soliton M. Since R is strictly increasing along gradient flow curves, we conclude that $R < 1$ near infinity. Since M has non-negative Riemann curvature, this implies $\mathrm{Ric} < \frac{1}{2}g$ near infinity. From (3.511) this implies that f is strictly convex (i.e. $\mathrm{Hess}(f) > 0$) near infinity. Thus the level sets of f have increasing area. On the other hand, on any region of M that approaches M_∞ (e.g. in the neighbourhoods of y_n) one easily sees (e.g. from (3.511), or from the analysis from Case 2) that the level sets of f converge to the sections S^2 of the cylinder, which have area 8π (note we are normalising the scalar curvature here to be 1, rather than the sectional curvature, which is $1/2$). Thus the level sets Σ of f have area strictly less than 8π.

On the other hand, from the Gauss-Codazzi formula (3.118), the Gaussian curvature K of Σ is given by the formula

$$K = K_M + \det(\mathrm{II}) \tag{3.519}$$

where K_M is the sectional curvature of Σ, and $\mathrm{II} = \frac{\mathrm{Hess}(f)|_\Sigma}{|\nabla f|}$ is the second fundamental form. Applying (3.511) we eventually compute

$$2K \leq R - 2\mathrm{Ric}(n,n) - \frac{(1 - R + \mathrm{Ric}(n,n))^2}{2|\nabla f|^2}. \tag{3.520}$$

Following the gradient flow lines of f, we see from previous analysis that $|\nabla f|$ goes to infinity (while curvature stays bounded and strictly positive), and so it is not hard to see that the right-hand side must be strictly less than 1 near infinity. But this means that $\int_\Sigma K < 4\pi$, contradicting the Gauss-Bonnet formula (Proposition 3.5.2). Thus Case 3 cannot in fact occur.

[115] One can also eliminate this case by appealing to the soul theorem (Theorem 3.12.17), or by adding an additional hypothesis throughout the argument that the manifolds being studied do not contain embedded $\mathcal{RP}^2$'s with trivial normal bundle.

3.17.4. Case 4: M compact, strictly positive curvature. Let us first deal with the two-dimensional case. Here one could use Hamilton's results[**Ha1988**] on Ricci flow for surfaces to show that this gradient shrinking soliton must be a round shrinking S^2 or $\mathcal{RP}^2$, but we give here an argument adapted from [**ChKn2004**]. It relies on the following identity, that provides an additional global constraint on the curvature R beyond that provided by the Gauss-Bonnet theorem:

Lemma 3.17.4 (Kazhdan-Warner type identity)**.** *Let (M, g) be a compact surface, and let X be a conformal Killing vector field (thus $\mathcal{L}_X g$ is a scalar multiple of g). Then $\int_M R\mathrm{div}X \; d\mu = 0$.*

Proof. When M has constant curvature, the claim is clear by integration by parts. On the other hand, by the *uniformisation theorem*, any metric g can be conformally deformed to a constant curvature metric. Note also from definition that a conformal Killing vector field remains conformal after any conformal change of metric. Thus it suffices to show that $\int_M R\mathrm{div}X \; d\mu$ is constant under any conformal change $\dot{g} = ug$ of g, keeping X static.

From the variation formulae from Section 3.2, we have $\dot{R} = -Ru - \Delta u$ and $\dot{d\mu} = u \; d\mu$. Inserting these formulae and integrating by parts to isolate u, we see that it suffices to show that $\nabla_\alpha(X^\alpha R) + \Delta(\nabla_\alpha X^\alpha) = 0$. On the other hand, since $\mathcal{L}_X g_{\alpha\beta} = \nabla_\alpha X_\beta + \nabla_\beta X_\alpha$ is conformal, we have the identity $\nabla_\alpha X_\beta + \nabla_\beta X_\alpha = (\nabla^\gamma X_\gamma) g_{\alpha\beta}$. Taking divergences of this identity twice and rearranging derivatives repeatedly, we eventually obtain this claim. □

Remark 3.17.5. This identity is closely related to one in [**KaWa1974**]. I do not know of any proof of the Kazhdan-Warner identity that does not require the uniformisation theorem; the result seems to have an irreducibly "global" nature to it.

Now we apply this lemma to the vector field ∇f, which is conformal thanks to (3.511). We conclude that $\int_M R\Delta f \; d\mu = 0$. On the other hand, from the trace of (3.511) we have $R - 1/\tau = \Delta f$. Integrating this against Δf we conclude that $\int_M |\Delta f|^2 \; d\mu = 0$, thus f is harmonic; and so $R = 1/\tau$. M is now constant curvature and is therefore either a round shrinking S^2 or $\mathcal{RP}^2$ as required.

Now we turn to three dimensions. The result in this case follows immediately from Hamilton's rounding theorem[**Ha1982**], but we will take advantage of the gradient shrinking soliton structure to extract just the key components of that theorem here. Let $\lambda \geq \mu \geq \nu \geq 0$ denote the eigenvalues of the Riemann curvature. Note that as the Ricci curvature is positive, $\mu + \nu$ is strictly greater than zero.

The quantity $(\mu + \nu)/\lambda$ ranges between 0 and 2 and reaches a minimum value δ at some point x. If we rewrite things in terms of the tensor $\mathcal{T}$ from Section 3.4, the gradient shrinking soliton structure means that

$$\frac{1}{\tau}\mathcal{T} = \Delta\mathcal{T} + \mathcal{L}_{\nabla f}\mathcal{T} + \mathcal{T}^2 + \mathcal{T}^{\#}. \tag{3.521}$$

But the region $\{\mathcal{T} : \nu \geq 0; \mu+\nu \geq \delta\lambda\}$ is fibrewise convex and parallel, and at x, $\frac{1}{\tau}\mathcal{T}$ and $\mathcal{L}_{\nabla f}\mathcal{T}$ are tangential to this region and $\Delta\mathcal{T}$ is tangential or inward. On the other hand, a computation shows that $\mathcal{T}^2+\mathcal{T}^{\#}$ is strictly inward unless $\delta = 2$, in which case it is tangential. So we must have $\delta = 2$, which implies that $\lambda = \mu = \nu$. In other words, the Ricci tensor is conformal: $\mathrm{Ric} = \frac{1}{3}Rg$. Comparing this with the Bianchi identity $\nabla_\alpha R = 2\nabla^\beta \mathrm{Ric}_{\alpha\beta}$ (see (3.31)) we conclude that $\nabla R = 0$, and thus $\nabla \mathrm{Ric} = 0$. Thus M has constant sectional curvature and is therefore a round shrinking spherical space form, as required.

The proof of Theorem 3.16.1 is now complete.

3.17.5. Appendix: Toponogov theory. Roughly speaking, Toponogov comparison theory[**To1959**] is to triangle geometry as Bishop-Gromov theory is to volumes of balls: in both cases, lower bounds on curvature are used to bound the geometry of Riemannian manifolds by model geometries such as Euclidean space. This theory links modern Riemannian geometry with the more classical approach to curved space (or non-Euclidean geometries) which often proceeded via analysing the angles formed by a triangle. The material here is loosely drawn from [**Pe2006**].

Lemma 3.17.6 (Toponogov cosine rule)**.** *Let (M, g) be a complete Riemannian manifold of non-negative sectional curvature, and let x_0, x_1, x_2 be three distinct points in M. Let θ be the angle formed*

at x_1 by the minimising geodesics from x_0, x_2 to x_1. Then
(3.522)
$$d(x_2,x_0)^2 \leq d(x_1,x_0)^2 + d(x_2,x_1)^2 - 2d(x_1,x_0)d(x_2,x_1)\cos\theta.$$

Of course, when M is flat we have equality in (3.522), by the classical cosine rule.

Proof. Let f be the function $f(x) := \frac{1}{2}d(x,x_0)^2$, and let $\gamma : [0, d(x_2,x_1)] \to M$ be the unit speed geodesic from x_1 to x_2. Our task is to show that

$$f(\gamma(t)) \leq f(\gamma(0)) + \frac{1}{2}t^2 - td(x_1,x_0)\cos\theta \tag{3.523}$$

for $t = d(x_2,x_1)$. From the Gauss lemma (Lemma 3.8.4) we know that $\frac{d}{dt}f(\gamma(t))|_{t=0} \leq -d(x_1,x_0)\cos\theta$. On the other hand, from the second variation formula (3.331) for distance and the non-negative sectional curvature assumption we have[116] $\frac{d^2}{dt^2}f(\gamma(t)) \leq 1$. The claim follows. □

There is an appealing reformulation of this lemma. Define a *triangle* to be three points A, B, C connected by three minimising geodesics AB, BC, CA.

Exercise 3.17.2 (Positive curvature increases angles)**.** Let ABC be a triangle in a Riemannian manifold of non-negative sectional curvature, and let $A'B'C'$ be a triangle in Euclidean space with the same side lengths as ABC. Show that the angle subtended at A is larger than or equal to that subtended at A' (and similarly of course for B and B', and C and C'). In particular, the sum of the angles of ABC is at least π.

There is also a relative version of this result:

Exercise 3.17.3 (Relative Toponogov comparison)**.** Let the notation and assumptions be as in the previous exercise. Let X, Y be points on AB, AC respectively, and let X', Y' be the corresponding points on $A'B'$ and $A'C'$. Show that the length of XY is greater than or equal to the length of $X'Y'$. *Hint*: it suffices to do this in the case $X = B$ (or $Y = C$), since the general case follows by two applications

[116] Actually one has to justify this in a suitable barrier sense when one is in the cut locus, but let us ignore this issue here for simplicity.

of this special case. Now repeat the argument used to prove Lemma 3.16.6.

Remark 3.17.7. Similar statements hold when one assumes that the sectional curvatures are bounded below by some number K other than zero. In this case, one replaces Euclidean space with the model geometry of constant curvature K, much as in the discussion of the Bishop-Gromov inequality in Section 3.10. See [**Pe2006**] for details.

3.17.6. The Cheeger-Gromoll splitting theorem. When a manifold has positive curvature, it is difficult for long geodesics to be minimising; see for example Myers' theorem (Exercise 3.10.2) for one instance of this phenomenon. Another important example of this is the *Cheeger-Gromoll splitting theorem*:

Theorem 3.17.8 (Splitting theorem)**.** [**ChGr1971**] *Let (M, g) be a complete Riemannian manifold of nonnegative Ricci curvature that contains a minimising geodesic line $\gamma : \mathbf{R} \to M$. Then M splits as the product of $\mathbf{R}$ with a manifold of one lower dimension.*

Remark 3.17.9. If one strengthens the non-negative Ricci curvature assumption to non-negative sectional curvature, this follows from [**To1964**]; if one strengthens further to have a uniform positive lower bound on sectional curvature, then this follows from Myers' theorem (Exercise 3.10.2).

Proof. We can parameterise γ to be unit speed. Consider the *Busemann functions* $B_+, B_- : M \to \mathbf{R}$ defined by

$$B_\pm(x) := \lim_{t \to \pm\infty} d(\gamma(t), x) - t. \tag{3.524}$$

One can show that the limits exist (because, by the triangle inequality, the expressions in the limits are bounded and monotone), and that B_+, B_- are both Lipschitz. From the non-negative curvature we have the upper bound $\mathrm{Hess}(r)(v, v) \leq 1/r$ for any distance function $r = d(x, x_0)$ (see (3.331)); applying this with $x_0 = \gamma(t)$ and letting $t \to \pm\infty$ we obtain the concavity $\mathrm{Hess}(B_\pm) \leq 0$. In particular, $B_+ + B_-$ is concave. On the other hand, from the triangle inequality we see that $B_+ + B_-$ is non-negative and vanishes on γ. Applying the (elliptic) strong maximum principle (which can be viewed as the static case

of the parabolic strong maximum principle, Exercise 3.13.5, though in the static case the bounded curvature hypothesis is not needed) we conclude that $B_+ + B_-$ vanishes identically. Since B_+ and B_- were both concave, they now must flat in the sense that $\mathrm{Hess}(B_+) = \mathrm{Hess}(B_-) = 0$. In particular they are smooth, and the gradient vector field $X := \nabla B_+$ is parallel to the Levi-Civita connection. On the other hand, by applying the Gauss lemma (Lemma 3.8.4) carefully we see that X is a unit vector field. Thus X splits M into a line and the level sets of B_+ (cf. Proposition 3.13.6) as desired. □

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/05/30`.

3.18. The structure of κ-solutions

Having classified all asymptotic gradient shrinking solitons in three and fewer dimensions in Section 3.16, we now use this classification, combined with extensive use of compactness and contradiction arguments, as well as the comparison geometry of complete Riemannian manifolds of non-negative curvature, to understand the structure of κ-solutions in these dimensions, with the aim being to state and prove precise versions of Theorem 3.12.8 and Corollary 3.12.12.

The arguments are particularly simple when the asymptotic gradient shrinking soliton is compact; in this case, the rounding theorems of Hamilton[**Ha1982**] show that the κ-solution is a (time-shifted) round shrinking spherical space form. This already classifies κ-solutions completely in two dimensions; the only remaining case is the three-dimensional case when the asymptotic gradient soliton is a round shrinking cylinder (or a quotient thereof by an involution). To proceed further, one has to show that the κ-solution exhibits significant amounts of curvature, and in particular that one does not have bounded normalised curvature at infinity. This curvature (combined with comparison geometry tools such as the Bishop-Gromov inequality) will cause asymptotic volume collapse of the κ-solution at infinity. These facts lead to the fundamental *Perelman compactness theorem* for κ-solutions, which then provides enough geometric control on such solutions that one can establish the structural theorems mentioned earlier.

The treatment here is a (slightly simplified) version of the arguments in [**MoTi2007**], which is based in turn on [**Pe2002**] and [**KlLo2006**] (see also [**CaZh2006**] for a slightly different treatment of this theory).

3.18.1. The compact soliton case. As we saw in Section 3.15, every κ-solution $t \mapsto (M, g(t))$ has at least one asymptotic gradient shrinking soliton $t \mapsto (M_\infty, g_\infty(t))$ associated to it. Suppose we are in the case in which at least one of these asymptotic gradient shrinking solitons is compact; by Theorem 3.16.1, this means that this soliton is a round shrinking spherical space form. Since this soliton is the geometric limit of a rescaled sequence of M, this implies that M is homeomorphic to M_∞ and, along a sequence of times $t_n \to \infty$, converges geometrically after rescaling to a round spherical space form. Thus M is asymptotically round as $t \to -\infty$.

One can now apply Hamilton's rounding theorems in two[**Ha1988**] and three[**Ha1982**] dimensions to conclude that M is in fact perfectly round. In the case of two dimensions this can be done by a variety of methods; let me sketch one way, using Perelman entropy; this is not the most elementary way to proceed but allows us to quickly utilise a lot of the theory we have built up. First we can lift M up to be S^2 instead of the quotient $\mathcal{RP}^2$. Then we observe from the Gauss-Bonnet theorem (Proposition 3.5.2) that $\int_M R\, d\mu = 4\pi$, and hence by the volume variation formula (3.69) the volume $\int_M d\mu$ is decreasing in time at a constant rate -4π. Let us shift time so that the volume is in fact equal to $4\pi\tau$, and consider the Perelman entropy $\mu(M, g(t), \tau)$ defined in Section 3.9. Testing this entropy with $f := 0$) we obtain an upper bound $\mu(M, g(t), \tau) \leq -4\pi$. On the other hand, on the sequence of times $t_n \to -\infty$, $(M, g(t))$ is smoothly approaching a round sphere, on which the entropy can be shown to be *exactly* -4π by the log-Sobolev inequality for the sphere (which can be proven in a similar way to the log-Sobolev inequality for Euclidean space in Section 3.9). Thus one can soon show that $\mu(M, g(t_n), \tau_n) \to -4\pi$. On the other hand, this entropy is non-increasing in τ; thus $\mu(M, g(t), \tau)$ is constant. Applying the results from Section 3.14 we conclude that this time-shifted manifold M is itself a gradient shrinking soliton, and thus is round by the results of Section 3.15.

Exercise 3.18.1. In this exercise we give an alternate way to establish the roundness of M in two dimensions, using a slightly different notion of "entropy". Firstly, observe that under conformal change of metric $g = ah$ on a surface, one has $d\mu_g = a d\mu_h$, $\Delta_g = \frac{1}{a}\Delta_h$, and $R_g = \frac{1}{a}(R_h - \Delta_h \log a)$. If we then express $g = ah$ where h is the metric on S^2 of constant curvature $+1$, show that the Ricci flow equation becomes $\partial_t a = \Delta \log a - 1$, and in particular that the volume $\int_M a \, d\mu_h$ is decreasing at constant rate 4π. If we time shift so that $\int_M a \, d\mu_h = 4\pi\tau$, show that the *relative entropy* $\int_M \frac{a}{\tau} \log \frac{a}{\tau} \, d\mu_h$ is non-decreasing in τ, and converges to 0 along τ_n (here one needs a stability result for the uniformisation theorem). From this and the converse to *Jensen's inequality*, conclude that a is constant at every time, which gives the rounding[117].

In two dimensions, we saw in Section 3.16 that the only gradient shrinking soliton was the round shrinking sphere. We have thus shown the following classification of κ-solutions in two dimensions:

Proposition 3.18.1. *The only two-dimensional κ-solutions are time translates of the round shrinking S^2 and $\mathcal{RP}^2$.*

For three dimensions, we can argue as in Case 4 of Section 3.16. Write $\lambda \geq \mu \geq \nu$ for the eigenvalues of the curvature tensor. At the times t_n, we have $(\mu + \nu)/\lambda \geq 2 - \delta_n$ for some $\delta_n \to 0$. Applying the tensor maximum principle (Proposition 3.4.5) and the analysis from Case 4 of Section 3.16, we thus see that $(\mu + \nu)/\lambda \geq 2 - \delta_n$ for all times $t \geq t_n$; sending n to infinity we conclude that $(\mu + \nu)/\lambda \geq 2$ for all times, and so curvature is conformal. Using the Bianchi identity as in Case 4 of Section 3.16, we conclude that the manifold is round.

3.18.2. The case of a vanishing curvature. Now we deal with the case in which there is a vanishing curvature:

Proposition 3.18.2. *Let $t \mapsto (M, g(t))$ be a 3-dimensional κ-solution for which the Ricci curvature has a null eigenvector at some point in spacetime. Then M is a time-shifted round shrinking cylinder, or the oriented or unoriented quotient of that cylinder by an involution.*

[117] For more proofs of the rounding theorem, for instance using the *Hamilton entropy* $\int_M R \log R \, d\mu$, see [**ChKn2004**].

Proof. If the Ricci curvature vanishes at any point, then by Hamilton's splitting theorem (Proposition 3.13.6) the flow splits (locally, at least) as a line and a two-dimensional flow. Passing to a double cover if necessary, we see that the flow is the product of a two-dimensional Ricci flow and either a line or a circle. The two-dimensional flow is itself a κ-solution and is thus a round shrinking S^2 or $\mathcal{RP}^2$. Checking all the cases and eliminating those which are not κ-noncollapsed we obtain the claim. □

3.18.3. Asymptotic volume collapse. Our next structural result on κ-solutions is

Proposition 3.18.3 (Asymptotic collapse of Bishop-Gromov reduced volume)**.** *Let $(M, g(t))$ be a κ-solution of dimension 3. Then for any time t and $r \in M$, $\lim_{r\to\infty} \mathrm{Vol}(B_{g(t)}(p,r))/r^3 \to 0$.*

Proof. We first observe, by inspecting all the possibilities from Theorem 3.16.1, that the claim is already true of all 3-dimensional asymptotic gradient shrinking solitons. We apply this to a gradient shrinking soliton for M and conclude that for any $\varepsilon > 0$ there exists arbitrarily negative times t_n, points x_n and radii r_n such that $B_{g(t_n)}(x_n, r_n)/r_n^3 \leq \varepsilon$. Applying the Bishop-Gromov comparison inequality (Lemma 3.10.1) we conclude that $\lim_{r\to\infty} B_{g(t_n)}(x_n, r)/r^3 \leq \varepsilon$. By the triangle inequality this implies that $\lim_{r\to\infty} B_{g(t_n)}(p, r)/r^3 \leq \varepsilon$.

Now we need to move from time t_n to time t; since t_n is arbitrarily negative we can assume $t \geq t_n$. Recall from Lemma 3.15.10 and the bounded curvature hypothesis that $\int_\gamma \mathrm{Ric}(X, X)$ is bounded for all times and all geodesics γ. Plugging this into the Ricci flow equation, we see that $\frac{d}{dt} d_{g(t)}(x, y)$ is also bounded (in the sense of forward difference quotients) for all times and all geodesics. In particular we have the additive distance fluctuation estimate $d_{g(t)}(p, x) = d_{g(t_n)}(p, x) + O(|t_n - t|)$, where the error is bounded even as $d_{g(t_n)}(p, x)$ or $d_{g(t)}(p, x)$ goes to infinity. Also, from equation (3.69) we know that the volume measure $d\mu$ is decreasing over time. From this we conclude that $\lim_{r\to\infty} B_{g(t)}(p, r)/r^3 \leq \varepsilon$. Since ε is arbitrary, the claim follows. □

We have a corollary:

Corollary 3.18.4. *Let $(M, g(t))$ be a non-compact*[118] *κ-solution of dimension 3. Then for any time t and point $p \in M$ we have* $\limsup_{x\to\infty} R(x)d(p,$ $+\infty$.

Proof. By time shifting we may take $t = 0$. Suppose for contradiction that $\limsup_{x\to\infty} R(x)d(p,x)^2$ is finite, thus $R(x) = O(1/d(p,x)^2)$ at time $t = 0$, and thus at all previous times since $\partial_t R \geq 0$ (see (3.427)). From the non-negativity of the curvature we obtain the similar upper bounds on the Riemann curvature. From the κ-noncollapsed nature of M we may thus conclude that $\mathrm{Vol}B(x, cd(p,x))/d(p,x)^3$ is bounded away from zero for some small $c > 0$. But this contradicts Proposition 3.17.3. □

Remark 3.18.5. In other treatments of this argument (e.g. in [**MoTi2007**]), Corollary 3.17.4 is established first (using the Topogonov theory from Section 3.16) and then used to derive Proposition 3.17.3. The two approaches are essentially just permutations of each other, but the arguments above seem to be slightly simpler (in particular, the theory of the *Tits cone* is avoided).

By combining Proposition 3.17.3 with another compactness argument, we obtain an important relationship:

Corollary 3.18.6 (Volume noncollapsing implies curvature bound)**.** *Let $t \mapsto (M, g(t))$ be a 3-dimensional κ-solution, and let $B(x_0, r)$ be a ball at time zero with volume at least νr^3. Then for every $A > 0$ we have a bound $R(x) = O_{\kappa,\nu,A}(r^{-2})$ for all x in $B(x_0, Ar)$.*

This result can be viewed as a converse to the κ-noncollapsing property (bounded curvature implies volume noncollapsing). A key point here is that the bound depends only on κ, ν, A and not on the κ-solution itself; this uniformity will be a crucial ingredient in the Perelman compactness theorem below.

Proof. Since $B(x_0, r)$ is contained in $B(x, (A + 1)r)$, it suffices to establish the claim when $x = x_0$. By replacing r with Ar if necessary we may normalise $A = 1$; we may also rescale $R(x_0) = 1$. Suppose

[118] Of course, the claim is vacuous for compact solutions.

the claim failed, then there exists a sequence of pointed κ-solutions $t \mapsto (M_n, g_n(t), x_n)$ with $R_n(0, x_n) = 1$ and balls $B_{g_n(0)}(x_n, r_n)$ with $r_n \to \infty$ whose volume is bounded below by νr_n^3 for some $\nu > 0$. Using the point picking argument (Exercise 3.16.1) we can also ensure that for each r, we have $R_n(0, x) \leq 4$ on $B_{g_n(0)}(0, r)$ if n is sufficiently large depending on r. Using the monotonicity $\partial_t R \geq 0$ and Hamilton's compactness theorem (Theorem 3.15.8) we may may thus pass to a subsequence and assume that the flows $t \mapsto (M_n, g_n(t), x_n)$ converge geometrically to a limit $t \mapsto (M_\infty, g_\infty(t), x_\infty)$, which one easily verifies to be a κ-solution whose asymptotic volume at time zero is bounded below by ν. But this contradicts Proposition 3.17.3. □

3.18.4. The Perelman compactness theorem. Corollary 3.17.6 leads to another important bound:

Proposition 3.18.7 (Bounded curvature at bounded distance)**.** *Let $\kappa > 0$, and let $t \mapsto (M, g(t))$ be a three-dimensional κ-solution. Then at time zero, for every $x_0 \in M$ and $A > 0$ we have $R(x) = O_{\kappa,A}(R(x_0))$ on $B(x_0, AR(x_0)^{-1/2})$.*

Proof. If the claim failed, then there will be an $A > 0$ sequence $t \mapsto (M_n, g_n(t), x_n)$ of pointed κ-solutions and $y_n \in B_{g_n(0)}(x_n, R_n(x_n)^{-1/2})$ and $R_n(y_n)/R_n(x_n) \to \infty$. Applying Corollary 3.17.6 in the contrapositive we conclude that $\mathrm{Vol}_{g_n(0)}(B_{g_n(0)}(x_n, R_n(x_n)^{-1/2})/R_n(x_n)^{-3/2} = o(1)$. By the Bishop-Gromov inequality (Lemma 3.10.1), we can thus find a radius $r_n = o(R_n(x_n)^{-1/2})$ such that $\mathrm{Vol}_{g_n(0)}(B_{g_n(0)}(x_n, r_n)/r_n^3 = \omega_3/2$ (say), where $\omega_3 := \frac{4}{3}\pi$ is the volume of the Euclidean 3-ball. By rescaling we may normalise $r_n = 1$, thus $R_n(x_n) = o(1)$. By Corollary 3.17.6 we now have $R_n(x) = O_{\kappa,A}(1)$ on $B_{g_n(0)}(x_n, A)$ for every $A > 0$. We may thus use monotonicity $\partial_t R_n \geq 0$ and Hamilton compactness as before to extract a limiting solution $t \mapsto (M_\infty, g_\infty(t), x_\infty)$ with $R_\infty(0, x_\infty) = 0$ and with $B_{g_\infty(0)}(x_\infty, 1) = \omega_3/2$. But then by the strong maximum principle (see Exercise 3.13.7), M_∞ must be flat; since it is κ-non-collapsed, it must be $\mathbf{R}^3$. But then we have $B_{g_\infty(0)}(x_\infty, 1) = \omega_3$, a contradiction. □

Exercise 3.18.2. Use Proposition 3.17.7 to improve the lim sup in Corollary 3.17.4 to a lim inf.

This in turn gives a fundamental compactness theorem.

Theorem 3.18.8 (Perelman compactness theorem)**.** *Let $\kappa > 0$, and let $t \mapsto (M_n, g_n(t), p_n)$ be a sequence of three-dimensional κ-solutions, normalised so that $R_n(0, p_n) = 1$. Then after passing to a subsequence, these solutions converge geometrically to another κ-solution $t \mapsto (M_\infty, g_\infty(t), p_\infty)$.*

Proof. By Proposition 3.17.7, we have $R_n(0, x) = O_A(1)$ on $B_{g_n(0)}(p_n, A)$ for every $A > 0$. Using monotonicity $\partial_t R_n \geq 0$ and Hamilton compactness as before, the claim follows. □

3.18.5. Universal noncollapsing. The Perelman compactness theorem requires κ to be fixed. However, the theorem can be largely extended to allow for variable κ by the following proposition.

Proposition 3.18.9 (Universal κ)**.** *There exists a universal $\kappa_0 > 0$ such that every 3-dimensional κ-solution which is not round, is in fact a κ_0-solution (no matter how small $\kappa > 0$ is).*

The reason one needs to exclude the round case is that sphere quotients S^3/Γ can be arbitrarily collapsed if one takes Γ to be large (e.g. consider the action of the n^{th} roots of unity on the unit ball of $\mathbf{C}^2$ (which is of course identifiable with S^3) for n large).

Proof. By time shifting it suffices to show κ_0-noncollapsing at time zero at at some spatial origin x_0, which we now fix.

Let $t \mapsto (M, g(t))$ be a κ-solution. By Proposition 3.17.1, M is non-compact, which means that any asymptotic gradient shrinking soliton must also be non-compact. By Theorem 3.16.1, all asymptotic gradient shrinking solitons are thus round shrinking cylinders, or the oriented or unoriented quotient of such a cylinder.

Let $l = l_{(0,x_0)}$ be the reduced length function from $(0, x_0)$. Recall from Section 3.15 that one can find a sequence of points (t_n, x_n) with $t_n \to -\infty$ with $l = O_A(1)$ and $R = O_A(t_n^{-1})$ on any cylinder $[At_n, t_n/A] \times B_{g(t_n)}(x_n, At_n^{1/2})$, whose rescalings by t_n converge geometrically to an asymptotic gradient shrinking soliton (and thus to a round cylinder or quotient thereof), and the bound $O_A(1)$ does not depend on κ. A computation shows that these round cylinders or

quotients are κ_0'-noncollapsed for some universal $\kappa_0' > 0$, and so the cylinders $[At_n, t_n/A] \times B_{g(t_n)}(x_n, At_n^{1/2})$ are similarly κ_0''-noncollapsed (for some slightly smaller but universal κ_0''). From the bounds on l and R, this implies that reduced volume at time t_n is bounded from below by a constant independent of κ. Using monotonicity of reduced volume, we thus have this lower bound for all times. The arguments in Section 3.11 then give κ_0-noncollapsing for some other universal $\kappa_0 > 0$. □

Here is one useful corollary of Perelman compactness and universality:

Corollary 3.18.10 (Universal derivative bounds)**.** *Let $t \mapsto (M, g(t))$ be a three-dimensional κ-solution. Then we have the pointwise bounds $|\partial_t^k \nabla^m \mathrm{Riem}| = O_{k,m}(R^{1+m/2+k})$ for all $m, k \geq 0$. In particular we have $|\partial_t^k \nabla^m R| = O_{k,m}(R^{1+m/2+k})$.*

Proof. The claim is clear for the round shrinking solitons (which we can lift up to live on the sphere S^3), so we may assume that the κ-solution is not round. By Proposition 3.17.9, we may then replace κ by a universal κ_0. We may then time shift so that $t = 0$ and rescale so that $R(0, x) = 1$. If the claim failed, then we could find a sequence $t \mapsto (M_n, g_n(t), x_n)$ of pointed κ-solutions with $R_n(0, x_n) = 1$, but such that some derivative of the curvature goes to infinity at this point. But this contradicts Theorem 3.17.8. □

Here is another useful consequence:

Exercise 3.18.3. Let $t \mapsto (M_n, g_n(t))$ be a sequence of three-dimensional κ_n-solutions, and let $x_n, y_n \in M_n$ and $t_n \leq 0$. If $R_n(t_n, x_n) d_{g(t_n)}(x_n, y_n)^2 \to \infty$, show that $R_n(t_n, y_n) d_{g(t_n)}(x_n, y_n)^2 \to \infty$. (Note that this generalises Corollary 3.17.4 or Exercise 3.17.2.) *Hint*: the claim is trivial in the round case, so assume non-roundness; then apply universality and compactness.

3.18.6. Global structure of κ-solutions. Roughly speaking, the above theory tells us that the geometry around any point (t, x) in a 3-dimensional κ-solutions has only bounded complexity if we only move $O(R(t, x)^{-1/2})$ in space and $O(R(t, x)^{-1})$ in time. This is about

as good a control on the local geometry of such solutions as we can hope for[119]; we now turn to the global geometry.

Let us begin with non-compact 3-dimensional κ-solutions. A key point is that if such solutions are not already round cylinders (or quotients thereof), they must mostly resemble such cylinders.

Definition 3.18.11 (Necks)**.** Let $\varepsilon > 0$. An *ε-neck* in a Riemannian 3-manifold (M, g) centred at a point $x \in M$ is a diffeomorphism $\phi : S^2 \times (-\frac{1}{\varepsilon}, \frac{1}{\varepsilon}) \to M$ from a long cylinder to M, such that the normalised pullback metric $R(x)\phi^* g$ lies within ε of the standard round metric on the cylinder in the $C^{\lfloor 1/\varepsilon \rfloor}$ topology, where we require of course that R(x) ¿ 0. The number $R(x)^{-1/2}$ is called the *width scale* of the neck, and $R(x)^{-1/2}/\varepsilon$ is the *length scale*.

Clearly, the notion of a ε-neck is a scale-invariant concept. Note that if a sequence of pointed manifolds (M_n, g_n, x_n) is converging geometrically (after rescaling) to a round cylinder $S^2 \times \mathbf{R}$, then for any $\varepsilon > 0$, x_n will be in the centre of an ε-neck for sufficiently large n. Since round cylinders appear prominently as geometric limits, it is then not surprising that κ-solutions, particularly non-compact ones, tend to be awash in ε-necks. For instance, we have

Proposition 3.18.12. *For every $\varepsilon > 0$ there exists an $A > 0$ such that whenever (t, x) is a point in a 3-dimensional non-compact κ-solution of strictly positive curvature and $\gamma : [0, +\infty) \to M$ is a unit speed minimising geodesic from x to infinity (such things can easily be shown to exist by compactness arguments) at time t, then every point in $\gamma([AR(x)^{-1/2}, +\infty))$ lies in the centre of an ε-neck at time t.*

Proof. By time shifting we can take $t = 0$. Suppose the claim is not the case, then we have a sequence $t \mapsto (M_n, g_n(t), x_n)$ of pointed 3-dimensional non-compact κ-solutions of strictly positive curvature and y_n on a minimising geodesic from x_n to infinity such that $d_n(x_n, y_n)^2 R_n(x_n) \to \infty$ and y_n is not the centre of a ε-neck at time zero. By Exercise 3.17.3 we thus have $d_n(x_n, y_n)^2 R_n(y_n) \to \infty$. Let us now rescale so that $R(y_n) = 1$. Since the M_n are non-compact,

[119] It is unlikely that the space of 3-dimensional κ-solutions is finite dimensional, as it is in the 2-dimensional case; see for instance [**Pe2003**, Example 1.4] for what is probably an infinite-dimensional family of κ-solutions.

they are non-round and so by Proposition 3.17.9 we can take κ to be universal, at which point by Perelman compactness (Theorem 3.17.8) we can pass to a subsequence and assume that $t \mapsto (M_n, g_n(t), y_n)$ is converging to a limit $t \mapsto (M_\infty, g_\infty(t), y_\infty)$, which is also a κ-solution. Since $d_n(x_n, y_n) \to \infty$, we see that the limit manifold contains a minimising geodesic line through y_∞, and hence by the Cheeger-Gromoll splitting theorem (Theorem 3.16.8) M_∞ must split into the product of a line and a positively curved manifold. By Proposition 3.17.2, we conclude that M_∞ is either a cylinder $S^2 \times \mathbf{R}$ or a projective cylinder $\mathcal{RP}^2 \times \mathbf{R}$.

The latter can be ruled out by topological considerations; a positively curved complete non-compact 3-manifold M_n is homeomorphic to $\mathbf{R}^3$ by the soul theorem ((Theorem 3.12.17), and so does not contain any embedded $\mathcal{RP}^2$ with trivial normal bundle[120]. So M_∞ is a round cylinder, and thus y_n is the centre of an ε-neck, a contradiction, and the claim follows. □

There is a variant of Proposition 3.17.12 that works in the compact case also:

Proposition 3.18.13. *For every $\varepsilon > 0$ there exists an $A > 0$ such that whenever $(t, x), (t, y)$ are points in a 3-dimensional κ-solution (either compact or noncompact) then at time t, any point on the minimising geodesic between x and y at a distance at least $AR(x)^{-1/2}$ from x and $AR(y)^{-1/2}$ from y, lies in the centre of an ε-neck at time t.*

Proof. We can repeat the proof of Proposition 3.17.12. The one non-trivial task is the topological one, namely to show that M does not contain an embedded $\mathcal{RP}^2$ with trivial normal bundle in the compact case (the non-compact case already being covered in Proposition 3.17.12). But M is compact and has strictly positive curvature (thanks to Proposition 3.17.2) and so by Hamilton's rounding theorem[**Ha1982**], is diffeomorphic to a spherical space form S^3/Γ for some finite Γ; in particular the fundamental group $\pi_1(M) \equiv \Gamma$ is finite. On the other hand, an embedded $\mathcal{RP}^2$ with trivial normal

[120] In any event, for applications to the Poincaré conjecture one can always assume that no such embedded projective plane exists in any manifold being studied.

bundle cannot separate M (as its Euler characteristic is 1) and so a closed loop in M can have a non-trivial intersection number with such a projective plane (using the normal bundle to give a sign to each intersection), leading to a non-trivial homomorphism from $\pi_1(M)$ to $\mathbf{Z}$, contradicting the finiteness of the fundamental group[121]. □

Informally, the above proposition shows that any two sufficiently far apart points in a compact κ-solution will be separated almost entirely by ε-necks. Since the only way that necks can be glued together is by forming a tube, one can then show the following two corollaries:

Corollary 3.18.14 (Description of non-compact positively curved κ-solutions)**.** *For every $\varepsilon > 0$ there exists $A > 0$ such that for every non-compact 3-dimensional positively curved κ-solution $t \mapsto (M, g(t))$ and time t there exists a point $p \in M$ such that at time t*

(1) *Every point outside of $B(p, AR(p)^{-1/2})$ lies in an ε-neck (and in particular, the exterior of this ball is topologically a half-infinite cylinder $S^2 \times [0, +\infty)$); and*

(2) *Inside the ball $B(p, AR(p)^{-1/2})$ (which is topologically a standard 3-ball by a version of the soul theorem, Theorem 3.12.17) all sectional curvatures are comparable to $R(p)$ modulo constants C depending only on ε, and the volume of the ball is comparable to $R(p)^{-3/2}$ modulo similar constants C.*

(The control inside the ball is coming from results such as Corollary 3.17.10, as well as the non-collapsed nature of M.)

In the language of [**MoTi2007**][122], we have described non-compact positively curved 3-dimensional κ-solutions as *C-capped ε-tubes*. Combined with Proposition 3.17.2, we now have a satisfactory description of non-compact κ-solutions: they are either round cylinders (and thus doubly infinite ε-tubes), oriented quotients of round cylinders (and

[121] An alternate argument would be to use Perelman compactness to extract a non-compact (but positively curved) limiting κ-solution from a sequence of increasingly long compact κ-solutions. Proposition 3.17.12 prohibits the limiting solutions from asymptotically looking like $\mathcal{RP}^2 \times \mathbf{R}$, and so the long compact solutions cannot have such projective necks either.

[122] Actually, more is proven in [**MoTi2007**]; one controls the time evolution of the necks and not just individual time slices, leading to the notion of a *strong ε-neck*. See [**MoTi2007**, Section 9.8] for details, as well as a precise definition of the C-capped ε-tubes.

thus a half-infinite ε-tube capped off by a punctured $\mathcal{RP}^3$), oriented quotients of round cylinders (and thus containing an $\mathcal{RP}^2$ with trivial normal bundle), or a half-infinite ε-tube capped off by a 3-ball.

For compact κ-solutions, we have something similar:

Proposition 3.18.15 (Characterisation of large compact κ-solutions)**.** *For every $\varepsilon > 0$ there exists $A, A' > 0$ such that if $t \mapsto (M, g)$ is a compact 3-dimensional κ-solution with $\operatorname{diam}(M) \geq A'(\sup R)^{-1/2}$ at some time t, then $(M, g(t))$ can be partitioned into an ε-tube (roughly speaking, a region in which every point lies in the middle of an ε-neck, and bordered on both ends by an S^2) and two (C, ε)-caps (roughly speaking, two regions diffeomorphic to either a 3-ball or punctured $\mathcal{RP}^3$, bounded by an S^2, in which the sectional curvatures are comparable to a scalar R, the diameter is comparable to $R^{-1/2}$, and volume comparable to $R^{-3/2}$). See* [**MoTi2007**, Section 9.8] *for precise definitions.*

The topological characterisation of the caps (that they are either 3-balls or punctured $\mathcal{RP}^3$s) follows from the corresponding characterisations of the caps in the non-compact case, followed by a compactness argument. Note that the round compact manifolds have diameter $O(R^{-1/2})$, where R is the constant curvature, and thus are not covered by the above Proposition.

By considering the various topologies for the caps, we see from basic topology then tells us that the manifolds in this case are homeomorphic to either S^3 or $\mathcal{RP}^3$, or $\mathcal{RP}^3 \# \mathcal{RP}^3$. The latter has infinite fundamental group, though, and thus not homeomorphic to a spherical space form; thus it cannot actually arise since Hamilton's rounding theorem[**Ha1982**] asserts that all compact manifolds of positive curvature are homeomorphic to spherical space forms.

Finally, we turn to small compact non-round κ-solutions.

Proposition 3.18.16 (Characterisation of small compact κ-solutions)**.** *Let $C > 0$, and let $t \mapsto (M, g)$ be a compact 3-dimensional κ-solution with $\operatorname{diam}(M) \leq C(\sup R)^{-1/2}$ at some time t which is not round, then all sectional curvatures are comparable up to constants depending on C, the diameter is comparable to $(\sup R)^{-1/2}$ up to similar*

constants, the volume is comparable to $(\sup R)^{-3/2}$*, and the manifold is topologically either* S^3 *or* $\mathcal{RP}^3$*.*

Proof. The diameter, curvature, and volume bounds follow from the compactness theory. To get the topological type, observe from the treatment of the compact soliton case that as M is not round, the asymptotic gradient shrinking soliton is non-compact, and thus must be a cylinder or one of its quotients. In particular this implies that as one goes back in time, the manifold M must eventually become large in the sense of Proposition 3.17.15. Since the manifolds in that proposition were topologically either S^3 or $\mathcal{RP}^3$, the same is true here. □

Putting all of the above results together, we obtain Proposition 3.12.14 (modulo some imprecision in the definitions which I have decided not to detail here).

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/06/02`.

3.19. The structure of high-curvature regions of Ricci flow

Having characterised the structure of κ-solutions, we now use them to describe the structure of high curvature regions of Ricci flow, as promised back in Section 3.12, in particular controlling their geometry and topology to the extent that surgery will be applied, which we will discuss in the next (and final) section of this chapter.

The material here is drawn largely from [**MoTi2007**], [**Pe2002**], [**Pe2003**]; see also [**KlLo2006**], [**CaZh2006**] for closely related material.

3.19.1. Canonical neighbourhoods. Let us formally define the notions of a canonical neighbourhood that were introduced in Section 3.12. They are associated with the various local geometries that are possible for three-dimensional κ-solutions. The first type of neighbourhood is related to the round spherical space forms S^3/Γ.

Definition 3.19.1 (ε-round)**.** Let $\varepsilon > 0$. A compact connected 3-manifold (M, g) is *ε-round* if one can identify M with a spherical

space form S^3/Γ with the constant curvature metric h such that some multiple of g lies within ε of h in the $C^{\lfloor 1/\varepsilon \rfloor}$ topology.

Note that if a sequence of manifolds (M_n, g_n, p_n), after rescaling, is converging geometrically to a spherical space form, then for any $\varepsilon > 0$ such manifolds will be ε-round for sufficiently large n.

The next type of canonical neighbourhood is associated to the small compact manifolds from Proposition 3.17.16.

Definition 3.19.2 (C-component)**.** Let $C > 0$. A *C-component* is[123] a connected 3-manifold (M, g) homeomorphic to S^3 or $\mathcal{RP}^3$, such that after rescaling the metric by a constant, the sectional curvatures, diameter, and volume are bounded between 1/C and C, and first and second derivatives of the curvature are also bounded by C.

Thus, for instance, every compact κ-solution which is small but not round in the sense of Proposition 3.17.16 will be a C-component for some C. Also observe that if a sequence of manifolds converge geometrically to a C-component, then these manifolds will be (say) $2C$-components once one is sufficiently far along the sequence.

The remaining canonical neighbourhoods are incomplete, corresponding to portions of non-compact (or compact but large) κ-solutions. One of them is the ε-neck defined in Definition 3.17.11. The other is that of a cap.

Definition 3.19.3 ((C, ε)-cap)**.** Let $C, \varepsilon > 0$. A (C, ε)-cap $(N \cup Y, g)$ is an incomplete 3-manifold that is the union of an ε-neck N with an incomplete core Y along one of the boundaries S^2 of the neck N. The core is homeomorphic to $\mathbf{R}^3$ or a punctured $\mathcal{RP}^3$, and has boundary S^2 equal to the above boundary of N. Furthermore, after rescaling g by a constant, the sectional curvatures, diameter, volume in the core are bounded between $1/C$ and C, and the zeroth, first and second derivatives of curvature in the cap are bounded above by C.

Definition 3.19.4 (Canonical neighbourhood)**.** Let $C, \varepsilon > 0$. We say that a point x in a 3-manifold (M, g) (possibly disconnected) has a *(C, ε)-canonical neighbourhood* if one of the following is true:

[123]We have deviated slightly here from the definition in [**MoTi2007**] by adding control of first and second derivatives for minor technical reasons.

(1) x lies in an ε-round component of M.

(2) x lies in a C-component of M.

(3) x is the centre of an ε-neck in M.

(4) x lies in the core of a (C, ε)-cap in M.

We remark that if a sequence of pointed manifolds (M_n, g_n, x_n) converges to a limit $(M_\infty, g_\infty, x_\infty)$, and x_∞ has a (C, ε)-neighbourhood of M_∞, then for sufficiently large n, x_n has a $(2C, 2\varepsilon)$-neighbourhood (say) of M_n. Also observe from construction that the property of having a canonical neighbourhood is scale-invariant.

Exercise 3.19.1 (First derivatives of curvature)**.** Show that if ε is sufficiently small, and x has a (C, ε)-canonical neighbourhood, then $R(x)$ is positive, $\nabla R(x) = O_C(R(x)^{3/2})$, and $\partial_t R(x) = O_C(R(x)^2)$.

From the theory in Section 3.17 we have

Proposition 3.19.5. *For every $\varepsilon > 0$ there exists $C > 0$ such that every point in a 3-dimensional κ-solution at any given time will have a (C, ε)-canonical neighbourhood, unless it is a round shrinking $\mathbf{R} \times \mathcal{RP}^2$.*

Note that C and ε are independent of κ; this is thanks to the universality property (Proposition 3.17.9).

Remark 3.19.6. For technical reasons, one actually needs a slightly stronger version of this proposition, in which any canonical neighbourhood which is an ε-neck is extended backwards to some extent in time in a manner that preserves the neck structure (leading to the notion of a *strong ε-neck* and *strong canonical neighbourhood*); see [**MoTi2007**, Chapter 9.8]. For similar reasons, the results below actually need to be stated for strong canonical neighbourhoods. We will ignore these minor technicalities.

The objective of this section is to establish the analogue of Proposition 3.18.5 for high-curvature regions of arbitrary Ricci flows:

Theorem 3.19.7 (Structure of high-curvature regions)**.** *For every $\varepsilon > 0$ there exists $C > 0$ such that the following holds. Let $t \mapsto (M, g)$ be a three-dimensional compact Ricci flow on some time interval $[0, T)$*

with no embedded $\mathcal{RP}^2$ with trivial normal bundle. Then there exists $K > 0$ such for every time $t \in [0, T)$, every $x \in (M, g(t))$ with $R(x) \geq K$ has a (C, ε)-canonical neighbourhood.

This theorem will then allow us to perform surgery on Ricci flows, as we will discuss in Section 3.19.

Morally speaking, Theorem 3.18.7 follows from Proposition 3.18.5 by rescaling and compactness arguments, but there is a rather delicate issue involved, namely to gain enough control on curvature at points in spacetime both near (and far) from the chosen point (t, x) that the Hamilton compactness theorem can be applied.

3.19.2. Overview of proof. We begin with some reductions. We can of course take M to be connected. Fix ε, and take C sufficiently large depending on ε (but not depending on any other parameters). We first observe that it suffices to prove the theorem for closed intervals $[0, T]$ rather than half-open ones, as long as the bounds on K depend only on an upper bound T_0 on T and the initial metric $g(0)$ and not on T itself (in particular, one cannot just use the trivial fact that R will be bounded on any compact subset of spacetime such as $[0, T] \times M$.) Once one does this, one sees that Theorem 3.18.7 is now true for some enormous K that depends on T; the task is to get a uniform K that depends only on the initial metric $g(0)$ and on an upper bound T_0 for T.

Perelman's argument proceeds by a downward induction on K; assume that K is large (depending on g(0) and T_0), and that Theorem 3.18.7 has already been established for $4K$ (say); and then establish the claim for K. By the previous discussion, this conditional result will in fact imply the full theorem.

By rescaling we may assume that $g(0)$ has normalised initial conditions (curvature bounded in magnitude by 1, volume of unit balls bounded below by some positive constant ω). We will now show that the conditional version of Theorem 3.18.7 holds for K sufficiently large depending only on ω and T_0.

Suppose this were not the case. Then there would be a ω and a T_0, and a sequence $t \mapsto (M_n, g_n(t), x_n)$ of pointed Ricci flows on $[0, T_n]$ (not containing any embedded $\mathcal{RP}^2$ with trivial normal bundle) for

some $0 \leq T_n \leq T_0$ with normalised initial conditions with constant ω, times $t_n \in [0, T_n]$, and scalars $K_n \to \infty$ such that every point in $[0, T_n] \times M_n$ of scalar curvature at least $4K_n$ has a (C, ε)-canonical neighbourhood, but that $R_n(t_n, x_n) \geq K_n$ but does not have a (C, ε)-canonical neighbourhood (in particular, $R_n(t_n, x_n) < 4K_n$). We want to extract a contradiction from this.

From the local theory (Lemma 3.11.2) we know that the curvature is bounded for short times (t less than a universal constant depending only on ω), so t_n must be bounded uniformly from below.

As usual, we define the rescaled pointed flows $t \mapsto (\tilde{M}_n, \tilde{g}_n(t), \tilde{x}_n)$ by $\tilde{M}_n := M_n$, $\tilde{x}_n := x_n$, and $\tilde{g}_n(t) := K_n^2 g_n(t_n + K_n^{-2} t)$. Thus these flows are increasingly ancient and have scalar curvature between 1 and 4 at the origin $(0, \tilde{x}_n)$. Also, any point in these flows of curvature at least 4 is contained in a canonical neighbourhood.

By Perelman's non-collapsing theorem (Theorem 3.8.15), we know that the flows $t \mapsto (M_n, g_n(t))$ flow is κ-noncollapsed at all scales less than 1 (say) for some κ depending only on ω; by rescaling, the rescaled flows $t \mapsto (\tilde{M}_n, \tilde{g}_n(t))$ are then κ-noncollapsed at all scales less than $1/o(1)$.

Meanwhile, from the Hamilton-Ivey pinching theorem (Theorem 3.4.16) we have $R_n \geq -O(1)$ and $\mathrm{Riem}_n \geq -o(R_n)$ whenever $R_n \to \infty$. Rescaling this, we obtain $\tilde{R}_n \geq -o(1)$ and $\widetilde{\mathrm{Riem}}_n \geq -o(1 + |\tilde{R}_n|)$.

Suppose we were able to prove the following statement.

Proposition 3.19.8 (Asymptotically globally bounded normalised curvature)**.** *For any $A, \tau > 0$ we have a bound $\tilde{R}_n(t, x) = O_{C,\varepsilon}(1)$ for all $x \in B_{\tilde{g}_n(0)}(\tilde{x}_n, A)$ and $t \in [-\tau, 0]$, if n is sufficiently large depending on A, τ.*

From this and the κ-noncollapsing, we see that the Hamilton compactness theorem (Theorem 3.15.8) applies, and after passing to a subsequence we see that the pointed flows $t \mapsto (\tilde{M}_n, \tilde{g}_n(t), \tilde{x}_n)$ converges geometrically to a Ricci flow $t \mapsto (M_\infty, g_\infty, x_\infty)$ which has bounded scalar curvature on $[-\tau, 0] \times M_\infty$ for each $\tau > 0$, and is automatically connected, complete, and ancient, and without an embedded $\mathcal{RP}^2$ with trivial normal bundle. From pinching we also see that we have non-negative sectional curvature; from the κ-noncollapsing of

the flows $t \mapsto (\tilde{M}_n, \tilde{g}_n(t), \tilde{x}_n)$ we have κ-noncollapsing of the limiting flow $t \mapsto (M_\infty, g_\infty, x_\infty)$. From Hamilton's Harnack inequality (cf. Section 3.13) we can show $\partial_t R \geq 0$, and so we in fact have globally bounded curvature. Finally, since $\tilde{R}_n(0, \tilde{x}_n)$ is bounded between 1 and 4, so is $R_\infty(0, x_\infty)$; thus the flow is not flat. Putting all this together, we conclude that $t \mapsto (M_\infty, g_\infty, x_\infty)$ is a κ-solution (see Definition 3.12.1). From Proposition 3.18.5, $(0, x_\infty)$ has a $(C/2, \varepsilon/2)$-canonical neighbourhood in M_∞ (if C is chosen large enough depending on ε); thus $(0, \tilde{x}_n)$ will have a (C, ε)-canonical neighbourhood in $\tilde{M}_n$ for large enough n, and so by rescaling $(0, x_n)$ has a (C, ε)-canonical neighbourhood in M_n, contradicting the hypothesis, and we are done.

So it remains to prove Proposition 3.18.8. If we had the luxury of picking (t_n, x_n) to be a point which had maximal curvature amongst all other points in $[0, t_n] \times M$, then this proposition would be automatic. However, we do not have this luxury (roughly speaking, this would only let us get canonical neighbourhoods for the "highest curvature region" of the Ricci flow, leaving aside the "second highest curvature region", "third highest curvature region", etc., unprepared for surgery). So one has to work significantly harder to achieve this aim.

3.19.3. Bounded curvature at bounded distance. A key step in the execution of Proposition 3.18.8 is the following partial result, in which the bound on curvature is allowed to depend on A, and for which one cannot go backwards in time.

Proposition 3.19.9. *(Bounded curvature at bounded distance) For any $A > 0$ we have a bound $\tilde{R}_n(0, x) = O_{C,\varepsilon,A}(1)$ for all $x \in B_{\tilde{g}_n(0)}(\tilde{x}_n, A)$, if n is large enough depending on A.*

This partial result is already rather tricky; we sketch the proof as follows (full details can be found in [**MoTi2007**, Chapter 10], [**KlLo2006**, Section 51], or [**CaZh2006**, Section 7.1]). If this result failed, then we have a sequence $\tilde{y}_n$ with $d_{\tilde{g}_n(0)}(x_n, y_n)$ bounded and $\tilde{R}_n(0, \tilde{y}_n) \to \infty$, thus one can move a bounded distance along a minimising geodesic from $\tilde{x}_n$ (which has curvature between 1 and 4) to $\tilde{y}_n$ and reach a point of arbitrarily high curvature. On the other hand, we know that every point of curvature at least 4 has a canonical

neighbourhood. Thus there is a bounded length minimising geodesic in $(\tilde{M}_n, \tilde{g}_n(0))$ that goes entirely through canonical neighbourhoods, starts with scalar curvature 4, and ends up with arbitrarily high curvature, with curvature staying 4 or greater throughout this process. This cannot happen if the canonical neighbourhoods are ε-round or C-components (since these neighhourhoods are already complete and curvatures are comparable to each other on the entire neighbourhood), so this geodesic can only go through ε-necks and (C, ε)-caps. One can also rule out the latter possibility (a long geodesic path that goes through the core of a (C, ε)-cap can easily be shown to not be minimising); thus the geodesic is simply going through a tube of ε-necks, with the width of these necks starting off being comparable to 1 and ending up being arbitrarily small. It turns out that by using a version of Hamilton's compactness theorem for incomplete Ricci flows, one can take a limit, which at time zero is a tube (topologically $[0, 1] \times S^2$) of non-negative curvature in which the curvature has become infinite at one end. Also, thanks to time derivative control on the curvature (see Exercise 3.18.1), the tube can be extended a little bit backwards in time as an incomplete Ricci flow (though the amount to which one can do this shrinks to zero as one approaches the infinite curvature end of the tube).

One can show that as one approaches the infinite curvature end of the cylinder and rescales, the cylinder increasingly resembles a cone. (For instance, one can use the bound $\int_\gamma \mathrm{Ric}(X, X) = O(1)$ from Lemma 3.15.10, where γ are geodesics emanating from the infinite curvature end, to establish this sort of thing.) By taking another limit one can then get an incomplete Ricci flow which at time zero is a cone. Because curvature is bounded away from zero, this cone is not flat. At this point, a version of Hamilton's splitting theorem (Proposition 3.13.6) for incomplete flows asserts that the manifold locally splits as the product of a line and a two-dimensional manifold. But non-flat cones cannot split like this, a contradiction. This establishes Proposition 3.18.9.

Remark 3.19.10. More generally, this argument can be used to show that if $\tilde{R}_n(t, x)$ is bounded by some $L \geq 1$, then $\tilde{R}_n(t, y)$ is bounded by $O_{A,C,\varepsilon}(L)$ for all $y \in B_{\tilde{g}_n(t)}(x, AL^{-1/2})$.

3.19.4. Bounded curvature at all distances. Now we extend Proposition 3.18.9 by making the bound global in A:

Proposition 3.19.11 (Bounded curvature at all distances)**.** *For any $A > 0$ we have a bound $\tilde{R}_n(0,x) = O_{C,\varepsilon}(1)$ for all $x \in B_{\tilde{g}_n(0)}(\tilde{x}_n, A)$, if n is large enough depending on A.*

We sketch a proof as follows. From Proposition 3.18.9 and compactness (taking advantage of non-collapsing, of course) we already know (passing to a subsequence if necessary) that $(\tilde{M}_n, \tilde{g}_n(0), x_n)$ converges to some limit $(\tilde{M}_\infty, \tilde{g}_\infty(0), x_\infty)$ which has non-negative curvature; it can also be extended a little bit backwards in time as an incomplete Ricci flow. Also, every point in this limit of curvature greater than 4 has a canonical neighbourhood. Our task is to basically to show that $(\tilde{M}_\infty, \tilde{g}_\infty(0))$ has bounded curvature. If this is not the case, then there are points of arbitrarily high curvature, which must be contained in either ε-necks or (C,ε)-caps. We conclude that there exist arbitrarily narrow ε-necks. One can then show that the manifold had strictly positive curvature, since otherwise by Hamilton's splitting theorem the manifold would split locally into a product of a two-dimensional manifold and a line, which can be shown to be incompatible with having arbitrarily narrow necks.

At this point one uses a general result that complete manifolds of strictly positive curvature cannot have arbitrarily narrow necks. We sketch the proof as follows. Clearly we may assume the manifold is non-compact, and hence by the soul theorem ((Theorem 3.12.17) is diffeomorphic to $\mathbf{R}^3$. This implies that every neck in fact separates the manifold into a compact part and a non-compact part. In fact, one can show that if p is a soul for the manifold (see [**Pe1994**]), then there is a minimising geodesic $\gamma : [0,+\infty) \to M$ from p to infinity that passes through all the necks. But if one then considers the Busemann function $B(y) := \lim_{s\to\infty} d(y,\gamma(s)) - s$, one can show that the gradient field ∇B is a unit vector which is within $O(\varepsilon)$ to parallel to the necks. This, combined with Stokes theorem, tells us that the area of the level sets of B inside a neck (which, up to errors of $O(\varepsilon)$, are basically slices of that neck) does not fluctuate by more than ε, even as one compares very distant necks together. But this

contradicts the assumption that there are arbitrarily small necks. (For full details see [**MoTi2007**, Proposition 2.19].)

3.19.5. Bounded curvature at all times. Now we need to extend Proposition 3.18.11 backwards in time. The time derivative bound on curvature (Exercise 3.18.1) lets us extend backwards by some fixed amount of time, but at the cost of potentially increasing the curvature, and we cannot simply iterate this (much as one cannot iterate a local existence result for a PDE to obtain a global one without some sort of *a priori* bound on whatever is controlling the time of existence). But what Exercise 3.18.1 does let us do, is reduce matters to establishing an *a priori* bound:

Proposition 3.19.12 (A priori bound)**.** *Let $\tau > 0$, and suppose $\tilde{R}_n$ is uniformly bounded on $[-\tau, 0] \times \tilde{M}_n$ for all sufficiently large n. Then in fact we can bound $\tilde{R}_n$ on these slabs by a universal bound $O_{C,\varepsilon,\tau}(1)$ (not depending on the previous universal bound).*

Indeed, Exercise 3.18.1 then lets us extend the uniform bounds a little bit to the past of τ, and one can continue this procedure indefinitely to establish Proposition 3.18.8.

We sketch the proof as follows. We allow all implied constants to depend on C, ε, τ for brevity. The bounds are already enough to give a non-ancient limiting flow $t \mapsto (M_\infty, g_\infty(t), x_\infty)$ on $[-\tau, 0]$ which is complete, connected, and non-negative curvature which is bounded at all times (but with an unspecified bound), and bounded at time zero by $O(1)$. Also, every point with curvature greater than 4 is known to have a canonical neighbourhood. The challenge is now to propagate the quantitative curvature bounds backwards in time, to replace the qualitative bound.

In the case of an ancient flow of non-negative curvature, Hamilton's Harnack inequality (3.427) gives $\partial_t R \geq 0$, which automatically does this propagation for us. We are however non-ancient here, and the Harnack inequality in this setting only gives a bound of the form $\partial_t R \geq R/(t+\tau)$. This can be integrated to give $R(t, x) = O(1/(t+\tau))$, thus our bounds blow up as we approach τ. However, this is at least enough to get good control on distances; in particular, using Corollary

3.15.11 we see that

$$-O(\frac{1}{\sqrt{t+\tau}}) \leq \frac{d}{dt} d_{g(t)}(x,y) \leq 0 \tag{3.525}$$

for all $x, y \in M_\infty$. Fortunately, the left-hand side here is absolutely integrable, and so we obtain a useful global distance comparison estimate:

$$d_{g(0)}(x,y) - O(1) \leq d_{g(t)}(x,y) \leq d_{g(0)}(x,y). \tag{3.526}$$

To use this, pick a large curvature $L \geq 1$, then a much larger radius r, then an extremely large curvature L'. Now suppose for contradiction that we have a point (t,x) in $[\tau, 0] \times M$ of curvature larger than L'. This point is then contained in a canonical neighbourhood. This neighbourhood cannot be compact (i.e. an ε-round or C-component), since that would mean that the minimal scalar curvature $R_{\min}$ was comparable to L' at time t, which by monotonicity of $R_{\min}$ (Proposition 3.4.10) would mean that the scalar curvature is comparable to L' at time 0, contradicting the boundedness of curvature there. This argument in fact shows that all large curvature regions are contained in either ε-necks or (C, ε)-caps.

Consider the ball $B_{g(t)}(x, r)$. From Remark 3.18.10 we see (if L' is large enough) that the curvature is larger than L on this ball, and so this ball consists entirely of necks and caps of width at most $O(L^{-1/2})$. From this it is not hard to see that the volume of this ball at time t is $O(L^{-1/2} r)$. On the other hand, there must be at least one point y on the boundary of this ball, since otherwise $R_{\min}$ would be at least L, which as noted before is not possible.

Applying (3.526) (and noting that Ricci flow reduces volume when there is non-negative curvature, by (3.69)) we conclude that $B_{g(t)}(x, r - O(1))$ also has volume $O(L^{-1/2} r)$. On the other hand, we know that there is a point y at distance r from x at time t, thus y at distance $r - O(1)$ from x at time 0. Thus (by the triangle inequality, and dividing the geodesic from x to y at time zero into unit length segments) $B_{g(0)}(x, r)$ contains $\sim r$ disjoint balls of radius $1/2$ (say). By the non-collapsing and curvature bounds at time zero, this forces $B_{g(0)}(x, r)$ to have volume at least comparable to r, a contradiction. This proves Proposition 3.18.12 and thus Theorem 3.18.7.

Remark 3.19.13. Perelman (and the authors who follow him) uses a slight variant of this argument, using the soul theorem (Theorem 3.12.17) to fashion a small S^2 in a narrow neck that separates two widely distant points at time t, which then evolves to a small S^2 separating two widely distant points at time zero (here we use (3.526)). But this leads to the desired contradiction due to the bounded curvature at that time.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/06/04`.

3.20. The structure of Ricci flow at the singular time, surgery, and the Poincaré conjecture

In the previous lecture, we studied high curvature regions of Ricci flows $t \mapsto (M, g(t))$ on some time interval $[0, T)$, and concluded that (as long as a mild topological condition was obeyed) they all had canonical neighbourhoods. This is enough control to now study the limits of such flows as one approaches the singularity time T. It turns out that one can subdivide the manifold M into a *continuing region* C in which the geometry remains well behaved (for instance, the curvature does not blow up, and in fact converges smoothly to an (incomplete) limit), and a *disappearing region* D, whose topology is well controlled[124]. This allows one (at the topological level, at least) to perform surgery on the interface Σ, removing the disappearing region D and replacing them with a finite number of "caps" homeomorphic to the 3-ball B^3. The relationship between the topology of the post-surgery manifold and pre-surgery manifold is as is described way back in Section 3.3.

However, once surgery is completed, one needs to restart the Ricci flow process, at which point further singularities can occur. In order to apply surgery to these further singularities, we need to check that all the properties we have been exploiting about Ricci flows - notably the Hamilton-Ivey pinching property, the κ-noncollapsing property, and the existence of canonical neighbourhoods for every point of high curvature - persist even in the presence of a large number of surgeries

[124] For instance, the interface Σ between C and D will be a finite union of disjoint surfaces homeomorphic to S^2.

(indeed, with the way the constants are structured, all quantitative bounds on a fixed time interval $[0,T]$ have to be uniform in the number of surgery times, although we will of course need the set of such times to be discrete). To ensure that surgeries do not disrupt any of these properties, it turns out that one has to perform these surgeries deep in certain ε-horns of the Ricci flow at the singular time, in which the geometry is extremely close to being cylindrical (in particular, it should be a δ-neck and not just a ε-neck, where the surgery control parameter δ is much smaller than ε; selection of this parameter can get a little tricky if one wants to evolve Ricci flow with surgery indefinitely, although for the purposes of the Poincaré conjecture the situation is simpler as there is a fixed upper bound on the time for which one needs to evolve the flow). Furthermore, the geometry of the manifolds one glues in to replace the disappearing regions has to be carefully chosen (in particular, it has to not disrupt the pinching condition, and the geometry of these glued in regions has to resemble a (C,ε)-cap for a significant amount of (rescaled) time). The construction of the "standard solution" needed to achieve all these properties is somewhat delicate, although we will not discuss this issue much here.

In this section we shall present these issues from a high-level perspective; due to lack of space we will not cover the finer details of the surgery procedure. More detailed versions of the material here can be found in [**Pe2003**], [**KlLo2006**], [**MoTi2007**], [**CaZh2006**], [**BeBeBoMaPo2008**].

3.20.1. Ricci flow at the singular time. Suppose we have a compact 3-dimensional Ricci flow $t \mapsto (M, g(t))$ on the time interval $[0,T)$ without any embedded $\mathcal{RP}^2$ with trivial normal bundle; for simplicity we can take M to be connected (otherwise we simply treat each of the finite number of connected components of M separately). We are interested in the extent to which we can define a limiting geometry $g(T)$ on M (or on some subset of M) at the final time T, and to work out the topological structure of the portions of M for which such a limit cannot be defined.

From Theorem 3.18.7, we know that any point $(t,x) \in [0,T) \times M$ for which the curvature $R(t,x)$ exceeds a certain threshold K, will lie

in a canonical neighbourhood[125]. One consequence of this is that one has the pointwise bounds

$$\nabla R = O(R^{3/2}); \quad R_t = O(R^2) \tag{3.527}$$

whenever $R \geq K$. Also recall from the maximum principle that we have $R \geq -O(1)$ throughout.

These simple regularity properties of the scalar curvature R are already enough to classify the limiting behaviour of $R(t,x)$ as $t \to T$ for each fixed x:

Exercise 3.20.1. Using (3.527), show that for every $x \in M$ there are either two possibilities: either $R(t,x)$ remains bounded as $t \to T$ (with a bound that can depend on x), or that $R(t,x)$ goes to infinity as $t \to T$, and in the latter case we even have the stronger statement $\lim_{t\to T}(T-t)R(t,x) > c$ for some c depending only on the implied constant in (3.527). If we let $\Omega \subset M$ be the set of x for which $R(t,x)$ remains bounded, show that Ω is open, and $R(t,\cdot)$ converges uniformly on compact subsets of Ω to some limit $R(T,\cdot)$ as $t \to T$.

The pinching property also lets us establish bounds of the form $\mathrm{Riem} = O(1+|R|)$. Using this and Shi's regularity estimates (Theorem 3.15.13) and the non-collapsing property, one can show that $(\Omega, g(t))$ converges in C^∞ on compact subsets of Ω to an incomplete limit $(\Omega, g(T))$.

Our main tasks here are to understand the geometry of the limit $(\Omega, g(T))$, and the topology of the remaining region $M\backslash\Omega$ (and how the two regions connect to each other).

If Ω is all of M, then the Ricci flow continues smoothly to time T, and we can continue onwards beyond T by the local existence theory for that flow. Now let us instead consider the other extreme case in which Ω is empty. In this case, from Exercise 3.19.1 we see that we have $R(t,x) \geq K$ for all $x \in M$, if t is sufficiently close to T. In particular, this means that *every* point in M lies in a canonical neighbourhood: an ε-round component (topologically S^3/Γ), a C-component (topologically S^3 or $\mathcal{RP}^3$), a ε-neck (topologically $[-1,1]\times S^2$), or a (C,ε)-cap (topologically a 3-ball or punctured $\mathcal{RP}^3$). If any point lies

[125]For sake of discussion we shall suppress the constants C and ε, as they will not play a major role in what follows

in the first type of canonical neighbourhoods, then M is topologically a spherical space form S^3/Γ. Similarly, if any point lies in the second type, M is either an S^3 or $\mathcal{RP}^3$ this way. So the only remaining case left is when every point lies in a neck or a cap. Since each cap contains at least one neck in it, we have at least one neck; following this neck in both directions, we either must end up with a doubly capped tube, or the tube must eventually connect back to itself. In the former case we obtain an S^3, $\mathcal{RP}^3$, or $\mathcal{RP}^3\#\mathcal{RP}^3$ (depending on whether zero, one, or two of the caps are punctured $\mathcal{RP}^3$'s rather than 3-balls); in the latter case, we get an S^2 bundle over S^1, which as discussed back in Section 3.3 comes in only two topological types, oriented and unoriented.

To summarise, if Ω is empty, then M is either a spherical space form, $\mathcal{RP}^3\#\mathcal{RP}^3$, or an S^2 bundle over S^1. In this case, the surgery procedure is simply to delete the entire manifold; this respects the topological compatibility condition required for Theorem 3.3.9. (The geometric compatibility condition is moot in this case.) In this case, the disappearing region is the whole manifold M, and the continuing region is empty.

Similar considerations occur if Ω is non-empty, but that $R(T,x) \geq 2K$ (say) for all $x \in \Omega$. So we may assume that there is at least one x for which $R(T,x) < 2K$, and thus $R(t,x) < 2K$ for all t sufficiently close to T. Thus we are guaranteed at least one point of bounded curvature in M, even at times close to the singular time. We can also assume that no canonical neighbourhood in M is an ε-round or C-component, since again in this case we could delete the entire manifold by surgery. Thus every point of curvature greater than K lies in a neck or a cap.

Because of this, it is not hard to show that every boundary point of Ω (where $R(T,x)$ becomes infinite) lies at the end of an *ε-horn*:a tube of ε-necks of curvature at least $4K$ (say) throughout, with the curvature becoming infinite at one or both ends[126] (thus the width of the necks go to infinity as one approaches the boundary of Ω). If the curvature goes to infinity at both ends, we have a *double ε-horn*;

[126]Note that if the tube is ever capped off by a (C,ε)-cap, then the curvature does not go to infinity in this tube.

otherwise, we have a *single ε-horn* will have one infinite curvature end and one end with bounded curvature.

The single ε-horns are all disjoint from each other, and their volume is bounded from below, and so they are finite in number. So the geometric picture of $(\Omega, g(T))$ is that of a (possibly infinite) number of double ε-horns, together with a finite number of additional connected incomplete manifolds, with boundary consisting of a finite number of disjoint spheres S^2, with a single ε-horn glued on to each one of these spheres.

Suppose one performs a topological surgery on each single ε-horn, by taking a sphere S^2 somewhere in the middle of each horn, removing the portion between that S^2 and the boundary, and replacing it by a 3-ball. We also remove all the double ε-horns; all the removed regions form the disappearing region of M, and the remainder is the continuing region. This creates a new compact (but possibly disconnected) manifold $M(T)$, formed by gluing finitely many 3-balls to the continuing region. To see the topological relationship between this new manifold and the previous manifold M, we move backwards in time a slight amount to an earlier time t, so that the horn is no longer singular at its boundary and instead connects to the remainder $M\backslash\Omega$ of the manifold. If t is close enough to T, then (by (3.527)), the portion of the horn between the S^2 and the boundary of the horn will still have curvature at least K, and thus every point here will lie in a neck or cap. Also, all the points in $M\backslash\Omega$, and in particular the portion of the manifold beyond the boundary of the horn, will also have curvature at least K and thus lie in a neck or cap if t is close enough to T. If we then follow this desingularised horn from the surgery sphere S^2 towards its boundary and beyond (possibly passing through any number of double ε-horns in the process), we will either discover a capped tube (which is thus topologically either a 3-ball or a punctured $\mathcal{RP}^3$), or else the tube will eventually connect to another surgery sphere, which may or may not lie in the same connected component of $M(T)$. Topologically, the first case corresponds to taking a *connected sum* of (one component of) $M(T)$ with either a sphere S^3 or a projective space $\mathcal{RP}^3$; the second case corresponds to taking a connected sum of one component of $M(T)$ with either another

component of $M(T)$, or with an S^2-bundle over S^1. Putting all this together we see that M is the connected sum of the components of $M(T)$, together with finitely many S^3's, $\mathcal{RP}^3$'s, and S^2-bundles over S^1, which again gives the topological compatibility condition required for Theorem 3.3.9.

We have thus successfully performed a single (topological) surgery. However, in doing so we have lost a lot of *quantitative* properties of the geometry, such as Hamilton-Ivey pinching, κ-noncollapsing, and the canonical neighbourhood property, which means that we cannot yet ensure that we can perform any further surgeries. To resolve this problem, we need to be more precise about the surgery process, in particular using our freedom to choose the surgery sphere as deep inside the ε-horn as we please, and to prescribe the metric on the cap that we attach to that sphere.

3.20.2. Surgery. To do surgery, a key observation of Perelman is that the geometry of the horn becomes increasingly cylindrical as one goes deeper into the horn:

Lemma 3.20.1. *Let H be a single ε-horn which has width scale comparable to r at the finite curvature end, and let $\delta > 0$. Then there exists a δ-neck of width scale comparable to h inside the horn H, where $h = h(r, \delta) > 0$ is a small quantity depending only on r and δ.*

Proof. (Sketch) Suppose this was not the case; then one could find a sequence of horns H_n of this type, and a sequence of points x_n inside these horns inside ε-necks of width scale comparable to h_n which are not inside δ-necks of this scale, where $h_n \to 0$. We can find a minimising geodesic from the finite curvature end to the infinite curvature end that goes through the neck near x_n. We then rescale (H_n, x_n) to have width 1 at x_n, and then apply the machinery from Section 3.18 to obtain a limit (H_∞, x_∞); the bounded curvature at bounded distance property (Proposition 3.18.9) shows that both the bounded curvature end and the infinite curvature end of the horn must recede to be infinitely far away from x_n in the limit, and so H_∞ becomes complete; it also has non-negative curvature, by pinching. The minimising geodesic becomes a minimising line in H_∞, and so by the Cheeger-Gromoll theorem it splits H_∞ into the product of a line

and a two-dimensional manifold (which is ε-close to a sphere). It turns out that we can continue all these manifolds backwards in time and repeat these arguments (much as in Section 3.18) to eventually give H_∞ the structure of a κ-solution; but then the vanishing curvature forces it to be a round cylinder, by Proposition 3.17.2. This implies that the rescaled H_n are eventually δ-necks, a contradiction. □

In order to successfully perform Ricci flow with surgery up to some specified time T (starting from controlled initial conditions, and as always assuming that no embedded $\mathcal{RP}^2$ with trivial normal bundle is present), we shall pick δ to be a very small number depending on T and the initial condition parameters, and perform our surgery on the δ-necks the scale h provided by Lemma 3.19.1, where r^{-2} is (essentially) the curvature threshold beyond which the canonical neighbourhood condition holds[127].

Remark 3.20.2. Thanks to finite time extinction in the simply connected case, being able to perform Ricci flow with surgery up to a preassigned finite time T is sufficient for proving the Poincaré conjecture (cf. [**Pe2003b**, Remark 1.4]). For the full geometrisation conjecture, however, one needs to perform Ricci flow with surgery for an infinite amount of time. For this, one cannot pick a single δ; instead, one has to divide the time interval into bounded intervals (e.g. dyadic intervals), and pick a different δ for each one (which depends on a number of parameters, including the curvature threshold for the canonical neighbourhood property on the previous dyadic interval). This selection of constants becomes a little subtle; see e.g. [**KlLo2006**] for further discussion.

Having located a δ-neck inside each single ε-horn, we remove the half of the neck from the centre sphere to the infinite curvature region, and smoothly interpolate in its place a copy of an (appropriately rescaled) *standard solution*. There is some choice as to how to set up this solution (much as there is some freedom when selecting a cutoff function), but roughly speaking this solution should resemble the

[127] In order to avoid a circular dependence of constants, one needs to check that even after surgery, that the curvature threshold for the canonical neighbourhood condition remains bounded even after arbitrarily many surgeries, as long as δ is chosen sufficiently small depending on this scale, on T, and on the initial conditions.

manifold formed by attaching a hemispherical cap to a round unit cylinder, except that one needs to smooth out the transition between the two portions of this solution; also, one needs to ensure that one has positive curvature throughout the standard solution in order not to disrupt the Hamilton-Ivey pinching property. It is also technically convenient to demand that this solution is spherically symmetric (at which point Ricci flow collapses to a system of two scalar equations in one spatial dimension). One can show that such standard solutions exist for unit time (just as the round unit cylinder does), and asymptotically matches the round shrinking solution at spatial infinity. As one consequence of this, one can check that all points in spacetime on the standard solution have canonical neighbourhoods; and, with some effort, one can also show that the same will be true in the spacetime vicinity of the region in which a standard solution has been inserted via surgery into a Ricci flow, as long as δ is sufficiently small. This is an essential tool to ensure that the canonical neighbourhood solution persists after multiple surgeries.

Remark 3.20.3. Suppose that M is irreducible with respect to connected sum (one can easily reduce to this case for the purposes of the Poincaré conjecture, thanks to Kneser's theorem [**Kn1929**], [**Mi1962**] on the existence of the prime decomposition). Then all surgeries must be topologically trivial, which means that every ε-horn, when viewed just before the singular time, only connects to a tube capped off with a ball. Then one can show that the surgery procedure is *almost distance decreasing* in the sense that for any $\eta > 0$, there exists a $1 + \eta$-Lipschitz diffeomorphism from the pre-surgery manifold to the post-surgery manifold. This property is useful for ensuring that various arguments for establishing finite time extinction for Ricci flows, also work for Ricci flows with surgery, as discussed for in Sections 3.6, 3.7. Even if the manifold is not irreducible, one can show that there are only finitely many surgeries that change the topology of the manifold; this can be established either using the prime decomposition, or by constructing a topological invariant (namely, the maximal number of homotopically non-trivial and homotopiclly distinct embedded 2-spheres in M) which is finite, non-negative, decreases by at least one with non-trivial surgery; see [**MoTi2007**, Section 18.2] for details.

Remark 3.20.4. The various properties listed above of the standard solution and its insertion into surgery regions are the "geometric compatibility conditions" alluded to in Section 3.3.

3.20.3. Controlling the geometry after multiple surgeries. Suppose that we have already performed a large (but finite) number of surgeries. In order to be able to continue Ricci flow with surgery, it is necessary that we maintain quantitative control on the geometry of the manifold which is uniform in the number of surgeries. Specifically, we need to extend the following existing controls on Ricci flow, to Ricci flow with surgery:

(1) Lower bounds on $R_{\min}$.

(2) Hamilton-Ivey pinching type bounds that lower bound Riem in terms of R.

(3) κ-noncollapsing of the manifold.

(4) Canonical neighbourhoods for all high curvature points in the flow.

The first two controls are quite easy to establish, because they are propagated by Ricci flow (thanks to the maximum principle), and are easily preserved by surgery (basically because 1. and 2. are primarily concerned with negative curvature, and surgery is only performed in regions of high positive curvature by construction). It is significantly trickier however to preserve 3., because the proof of κ-noncollapsing is more global, requiring the use of $\mathcal{L}$-geodesics through spacetime. The key new difficulty is that thanks to the presence of surgery, the manifold can become "parabolically disconnected"; not every point in the initial manifold $(M, g(0))$ can be reached from a future point in a later manifold $(M(t), g(t))$ by an $\mathcal{L}$-geodesic, because an intervening surgery could have removed the region of spacetime that the geodesic ought to have passed through. This forces one to introduce the notion of an *admissible curve* - curves that avoid the surgery regions completely - and *barely admissible curves*, which are admissible curves which touch the boundary of the surgery regions. Roughly speaking, the monotonicity of reduced volume now controls the κ-noncollapsing at future times in terms of the non-collapsing of the portion of the initial manifold $(M, g(0))$ which can be reached by admissible curves;

this region is bordered by points which can be reached by barely admissible curves.

Now suppose we knew that all barely admissible curves had large reduced length. Then the maximum principle argument that located points of small reduced length for Ricci flows (cf. (3.375)), would continue to work for Ricci flows with surgery, with the points located being inside the admissible region. It turns out that the arguments of Section 3.11 could then be adapted to this setting without much difficulty to establish the desired κ-noncollapsing.

It is not too difficult to show that if a path did pass through a δ-neck inside an ε-horn in which surgery was taking place, then the portion of the path near to that surgery region would have a large contribution to the reduced length (unless the starting point of the path was very close to the surgery region, but then one could verify the non-collapsing property directly, essentially due to the non-collapsed nature of the standard solution). This almost settles the problem immediately, except for the technical issue that there might be regions of negative curvature elsewhere in spacetime which could drag the reduced length back down again (note that the reduced length is not guaranteed to be non-negative!). There is a technical fix for this, defining a modified reduced length in which the curvature term R is replaced by $\max(R, 0)$ (and using the lower bounds on $R_{\min}$ to measure the discrepancy between the two notions), but we will not discuss the details here; see [**Pe2003**, Lemma 5.2] (and [**MoTi2007**, Chapter 16] for a very detailed treatment).

Remark 3.20.5. In [**Zh2007**], [**Zh2008**], Perelman's entropy is used (as in Section 3.9) to establish κ-noncollapsing for Ricci flow with surgery, using the distance-decreasing property to keep control of the entropy functional after each (topologically trivial) surgery. This simplifies this part of the argument, at least in the case of irreducible manifolds M.

Finally, one has to check that all high-curvature points of Ricci flows with surgery lie in canonical neighbourhoods, where the threshold for "high curvature" is uniform in the number of surgeries performed. Very roughly speaking, there are two cases, depending on whether there was a surgery performed near (in the spacetime sense)

such a region or not. If there was no nearby surgery, then the arguments in Section 3.18 (which are local in nature) essentially go through, exploiting heavily the κ-noncollapsing and pinching properties that we have just established. If instead there was a nearby surgery in the recent past, then one needs to approximate the geometry here by the geometry of the standard solution, for which all points have canonical neighbourhoods. See for instance [**MoTi2007**, Section 17.1] for details.

3.20.4. Surgery times do not accumulate. The very last thing one needs to do to establish the Poincaré conjecture is to establish Theorem 3.3.12, which asserts that the set of surgery times is discrete. It turns out that this is in fact rather easy to establish. One first observes that each surgery removes at least some constant amount of $c(h) > 0$ of volume from the manifold (as can be seen by looking at what happens to a single δ-neck of width roughly h under surgery; all other removals under surgery of course only decrease the volume further). On the other hand, using the volume variation formula (3.69) we have an upper bound on the growth of volume during non-surgery times:

$$\frac{d}{dt}\mathrm{Vol}(M(t)) \leq -R_{\min}(t)\mathrm{Vol}(M(t)). \tag{3.528}$$

Since we have a uniform lower bound on $R_{\min}$, this implies that volume can grow at most exponentially, and in particular can only grow by a bounded amount on any fixed time interval. Hence there can be at most finitely many surgeries on each such time interval, and we are done!

Remark 3.20.6. The number of surgeries performed in a given time interval, while finite, could be incredibly large; it depends on the length scale h of the surgery, which in turn depends on the parameter δ, which needs to be very small in order not to disrupt the κ-noncollapsing or canonical neighbourhood properties of the flow. This is why it is essential that our control of such properties is uniform with respect to the number of surgeries.

Remark 3.20.7. Note also that there is no lower bound as to how close two surgery times could be to each other; indeed, there is nothing preventing two completely unrelated surgeries from being instantaneous. However, if there are an infinite number of singularities occurring at (or very close to) a single time, what tends to happen is that the earliest surgeries Knwill not only remove the immediate singularities being formed, but will also pre-emptively eradicate a large number of potential future singularities (in particular, due to the removal of all the double ε-horns, which were not immediately singular but were threatening to become singular very shortly), thus keeping the surgery times discrete.

Notes. This lecture first appeared at `terrytao.wordpress.com/2008/06/06`. Thanks to Américo Tavares for corrections.

Chapter 4

Lectures in additive prime number theory

4.1. Structure and randomness in the prime numbers

This talk concerns the subject of *additive prime number theory* - which, roughly speaking, is the theory of additive patterns contained inside the prime numbers $\mathcal{P} = \{2, 3, 5, 7, 11, \ldots\}$. This is a very old subject in mathematics; for instance, the *twin prime conjecture*, which asserts that there are infinitely many patterns of the form $n, n+2$ in the primes, may have been considered in one form or another by Euclid (although the modern version of the conjecture probably dates to [**Br1915**], which showed the first non-trivial progress towards the problem). It remains open today, although there are some important partial results. Another well-known conjecture in the subject is the *odd Goldbach conjecture* (dating from 1742), which asserts that every odd number n greater than 5 is the sum of three primes. A famous theorem of Vinogradov[**Vi1937**] asserts that this conjecture is true for all sufficiently large n; Vinogradov's original argument did not explicitly say how large is "sufficiently large", but later authors did quantify the argument; currently, it is known[**LiWa2002**] that the odd Goldbach conjecture is true for all odd $n > 10^{1346}$. The conjecture is also known[**Sa1998**] for all odd $5 < n < 10^{20}$, by a completely different method.

In this lecture, I will present the following result of myself and Ben Green in this subject:

Theorem 4.1.1 (Green-Tao Theorem)**.** [**GrTa2008**] *The prime numbers $\mathcal{P} = \{2, 3, 5, 7, \ldots\}$ contain arbitrarily long arithmetic progressions.*

More specifically, I want to talk about three basic ingredients in the proof, and how they come together to prove the theorem:

(1) Random models for the primes;

(2) Sieve theory and almost primes;

(3) Szemerédi's theorem on arithmetic progressions.

4.1.1. Random models for the primes. One of the most fundamental results in this field is the *prime number theorem*, which asserts

that the number of primes less than any large integer N is asymptotically equal to $(1+o(1))\frac{N}{\log N}$ as N goes to infinity. This theorem is proven by using the *Euler product formula*[1]

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p (1 - \frac{1}{p^s})^{-1}$$

which relates the primes to the *Riemann zeta function* $\zeta(s)$. By combining this formula with some non-trivial facts about the Riemann zeta function (and in particular, in the zeroes of that function), one can eventually obtain the prime number theorem.

One way to view the prime number theorem is as follows: if one picks an integer n from 1 to N at random, then that number n has a probability of $\frac{1+o(1)}{\log N}$ of being prime.

With this in mind, one can propose the following heuristic "proof" of the twin prime conjecture:

(1) Take a large number N, and let n be a randomly chosen integer from 1 to N. By the prime number theorem, the event that n is prime has probability $\mathbf{P}(n \text{ is prime}) = \frac{1+o(1)}{\log N}$.

(2) By another application of the prime number theorem, the event that $n+2$ is prime also has probability $\mathbf{P}(n+2 \text{ is prime}) = \frac{1+o(1)}{\log N}$. (The shift by 2 causes some additional correction terms, but these can be easily absorbed into the $o(1)$ term.)

(3) Assuming these two events are independent, we conclude $\mathbf{P}(n, n+2 \text{ both prime}) = (\frac{1+o(1)}{\log N})^2$. In other words, the number of twin primes less than N is $(1+o(1))\frac{N}{\log^2 N}$.

(4) Since $(1+o(1))\frac{N}{\log^2 N}$ goes to infinity as N goes to infinity, there are infinitely many twin primes.

---

[1] Incidentally, this formula, if rewritten using the *geometric series formula* as

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p (1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \ldots)$$

is a restatement (via *generating functions*) of the *fundamental theorem of arithmetic*; if instead one rewrites it as

$$(1 - \frac{1}{2^s})(1 - \frac{1}{3^s})(1 - \frac{1}{5^s}) \ldots \times \sum_{n=1}^{\infty} \frac{1}{n^s} = 1$$

one can view this as a restatement of (a variant of) the *sieve of Eratosthenes*.

Unfortunately, this argument doesn't work. One way to see this is to observe that the same argument could be trivially modified to imply that there are infinitely many pairs of *adjacent* primes n, $n+1$, which is clearly absurd.

OK, so the above argument is broken; can we fix it? Well, we can try to accomodate the above objection. Why is it absurd to have infinitely many pairs of adjacent primes? Ultimately, it is because of the obvious fact that the prime numbers are all odd (with one exception). In contrast, the above argument was implicitly using a *random model* for the primes (first proposed by Cramer[**Cr1936**]) in which every integer from 1 to N - odd or even - had an equal chance of $\frac{1+o(1)}{\log N}$ of being prime; this model is clearly at odds with the parity structure of the primes. But we can repair this by replacing the above random model with a more sophisticated model in which parity is taken into account. More precisely, we observe that a randomly chosen odd number from 1 to N has a probability of $\frac{2+o(1)}{\log N}$ of being prime, while a randomly chosen even number has a probability of $\frac{0+o(1)}{\log N}$ (one can be more precise than this, of course). In the language of probability theory, we have the *conditional probabilities*

$$\mathbf{P}(n \text{ is prime}|n \text{ is odd}) = \frac{2+o(1)}{\log N}$$
$$\mathbf{P}(n \text{ is prime}|n \text{ is even}) = \frac{0+o(1)}{\log N}$$

and similarly

$$\mathbf{P}(n+2 \text{ is prime}|n \text{ is odd}) = \frac{2+o(1)}{\log N}$$
$$\mathbf{P}(n+2 \text{ is prime}|n \text{ is even}) = \frac{0+o(1)}{\log N}.$$

Now, instead of assuming that the events "n is prime" and "n+2 is prime" are absolutely independent, let us assume that they are *conditionally* independent, relative to the parity of n. Then we conclude

that

$$\mathbf{P}(n, n+2 \text{ both prime} | n \text{ is odd}) = \frac{4+o(1)}{\log^2 N}$$

$$\mathbf{P}(n, n+2 \text{ both prime} | n \text{ is even}) = \frac{0+o(1)}{\log^2 N}$$

and a little computation (using the *law of total probability*) then shows that the number of twin primes less than N is now $(2+o(1))\frac{N}{\log^2 N}$, which still goes to infinity, and so we recover the twin prime conjecture again. Or do we?

Well, the above random model is still flawed. It now correctly asserts that there are extremely few pairs n, $n+1$ of adjacent primes, but it also erroneously predicts that there are infinitely many triplets of primes of the form n, $n+2$, $n+4$, when in fact there is only one - 3, 5, 7 - since exactly one of n, $n+2$, $n+4$ must be divisible by 3. But we can refine the random model further by taking mod 3 structures into account as well as mod 2 structures. Indeed, if we partition the integers from 1 to N using both the mod 2 partition and the mod 3 partition, we obtain the six residue classes $\{1 \leq n \leq N : n = i \bmod 6\}$ for $i = 0, 1, 2, 3, 4, 5$. From the prime number theorem in arithmetic progressions (a common generalisation of the prime number theorem and *Dirichlet's theorem*) one can show that

$$\mathbf{P}(n \text{ is prime} | n = i \bmod 6) = \frac{3+o(1)}{\log N}$$

for $i = 1, 5$, and

$$\mathbf{P}(n \text{ is prime} | n = i \bmod 6) = \frac{0+o(1)}{\log N}$$

for $i = 0, 2, 3, 4$. By repeating the previous analysis, the predicted count of twin primes less than N now drops from $(2+o(1))\frac{N}{\log^2 N}$ to $(1.5+o(1))\frac{N}{\log^2 N}$.

Now, it turns out that this model is still not correct - it fails to account for the mod 5 structure of the primes. But it is not hard to incorporate that structure into this model also, which revises the twin prime count downward a bit to $(1.40625+o(1))\frac{N}{\log^2 N}$. And then the mod 7 structure also changes the predicted number of twin primes a little bit more, and so on and so forth. But one notices that as

one continues to input in all this structural information about the primes, the predicted count of twin primes begins to converge to a limit, namely $(2\Pi_2 + o(1))\frac{N}{\log^2 N} \approx 1.32\frac{N}{\log^2 N}$, where

$$\Pi_2 := \prod_{p \geq 3, \text{ prime}} (1 - \frac{1}{(p-1)^2}) = 0.66016\ldots$$

is known as the *twin prime constant*. More generally, Hardy and Littlewood proposed a general conjecture[**HaLi1923**], now known as the *Hardy-Littlewood prime k-tuples conjecture*, that predicted asymptotic counts for a general class of additive patterns in the primes; this conjecture (and further refinements) would imply the twin prime conjecture, Vinogradov's theorem, my theorem with Ben Green, and many other results and conjectures in the subject also.

Roughly speaking, these conjectures assert that apart from the "obvious" structure in the primes, arising from the prime number theorem and from the local behaviour of the primes mod 2, mod 3, etc., there are no other structural patterns in the primes, and so the primes behave "pseudorandomly" once all the obvious structures are taken into account. The conjectures are plausible, and backed up by a significant amount of numerical evidence; unfortunately, nobody knows how to enforce enough pseudorandomness in the primes to make the conjectures rigorously proven. (One cannot simply take limits of the above random models as one inputs more and more mod p information, because the $o(1)$ error terms grow rapidly and soon overwhelm the main term that one is trying to understand.) The problem is that the primes may well contain "exotic structures" or "conspiracies", beyond the obvious structures listed above, which could further distort things like the twin prime count, perhaps so much so that only finitely many twin primes remain. This seems extremely unlikely, but we can't rule it out completely yet; how can one disprove a conspiracy?

Some numerics may help illustrate what I mean by the primes becoming random after the mod 2, mod 3, etc. structures are "taken into account" (though I should caution against reading *too* much into such small-scale computations, as there are many opportunities in small data sets for random coincidences or transient phenomena to

create misleading artefacts, cf. the "law of small numbers" [**Gu1988**]). Here are the first few natural numbers, with the primes in red, the odd numbers in the starred columns, and the even numbers in dotted columns:

*	.	*	.	*	.	*	.	*	.	*	.	*	.
1	2	**3**	4	**5**	6	**7**	8	**9**	10	**11**	12	**13**	14
15	16	**17**	18	**19**	20	**21**	22	**23**	24	**25**	26	**27**	28
29	30	**31**	32	**33**	34	**35**	36	**37**	38	**39**	40	**41**	42
43	44	**45**	46	**47**	48	**49**	50	**51**	52	**53**	54	**55**	56

It is then clear that the primes have mod 2 structure; they cluster in the odd numbers (the starred columns) rather than the even numbers (the dotted columns), and are thus distributed quite non-randomly. But suppose we "zoom in" on the odd numbers, discarding the even numbers:

*	.	*	.	*	.	*	.	*	.	*	.	*	.	*	.
1	3	**5**	7	**9**	11	**13**	15	**17**	19	**21**	23	**25**	27	**29**	31
33	35	**37**	39	**41**	43	**45**	47	**49**	51	**53**	55	**57**	59	**61**	63
65	67	**69**	71	**73**	75	**77**	79	**81**	83	**85**	87	**89**	91	**93**	95

Then it seems that there is no further parity structure; the starred columns (which have numbers which are 1 mod 4) and the dotted columns (which have numbers which are 3 mod 4), seem equally likely to contain primes. (Indeed, this is a proven fact, being a special case of the prime number theorem in arithmetic progressions.) But look what happens if we highlight the mod 3 structure instead:

*	.	*	.	*	.	*	.
1	3	5	**7**	9	11	**13**	15	17	**19**	21	23
25	27	29	**31**	33	35	**37**	39	41	**43**	45	47
49	51	53	**55**	57	59	**61**	63	65	**67**	69	71
73	75	77	**79**	81	83	**85**	87	89	**91**	93	95

Then the dotted columns (whose entries are 0 mod 3) are devoid of primes other than 3 itself. But if we instead zoom into (say) the starred columns (whose entries are 1 mod 3), we eliminate the mod 3 structure, making the remaining primes more randomly distributed:

*				.	*				.
1	7	13	19	25	**31**	37	43	49	55
61	67	73	79	85	**91**	97	103	109	115
121	127	133	139	145	**151**	157	163	169	175

Now the primes exhibit obvious mod 5 structure (the dotted columns, whose entries are 0 mod 5, have no primes) but look fairly randomly distributed otherwise. Indeed, if we zoom in to (say) the 1 mod 5 residue class, which are the starred columns above, there seems to be very little structure at all:

						.
1	31	61	91	121	151	181
211	241	271	301	331	361	391
421	451	481	511	541	571	601
631	661	691	721	751	781	811

Apart from the dotted column, which has all entries divisible by 7 and thus not prime, the primes seem fairly randomly distributed. In the above examples I always zoomed into the residue class 1 mod p for $p = 2, 3, 5, \ldots$, but if one picks other residue classes (other than 0 mod p, of course), one also sees the primes become increasingly randomly distributed, with no obvious pattern within each class (and no obvious relation between pairs of classes, triplets of classes, etc.). One can view the prime k-tuples conjecture as a precise formalisation of this assertion of increasing random distribution[2].

4.1.2. Sieve theory and almost primes. We have talked about random models for the primes, which seem to be very accurate, but difficult to rigorously justify. However, there is a closely related concept to a prime, namely an *almost prime*, for which we *can* show the corresponding random models to be accurate, by the elementary yet surprisingly useful technique of *sieve theory*.

The most elementary sieve of all is the *sieve of Eratosthenes*. This sieve uncovers (or "sifts out") all the prime numbers in a given range, say between $N/2$ and N for some large number N, by starting with all the integers in this range, and then discarding all the multiples of 2, then the multiples of 3, the multiples of 5 and so forth. After all

[2]In [**GrTa2008**], we rely quite crucially on this zooming in trick to improve the pseudorandomness of the almost primes, referring to it as the "W-trick".

multiples of prime numbers less than $\sqrt{N}$ are discarded, the remaining set is precisely the set of primes between $N/2$ and N.

It is tempting to use this sieve to count patterns in primes, such as twin primes. After all, it is easy to count how many twins there are in the integers from $N/2$ to N; if one then throws out all the multiples of 2, it is still straightforward to count the number of twins remaining. Similarly if one then throws out multiples of 3, of 5, and so forth; not surprisingly, the computations here bear some resemblance to those used to predict twin prime counts from the random models just mentioned. However, as with the random models, the error terms begin to accumulate rapidly, and one loses control of these counts long before one reaches the end of the sieve. More advanced sieves (in which one does not totally exclude the multiples of small numbers, but instead adjusts the "weight" or "score" of a number upward or downward depending on its factors) can improve matters significantly, but even the best sieves still only work if one stops sieving well before the $\sqrt{N}$ mark (sieve levels such as $N^{1/4}$ are typical). (The reason for this has to do with the *parity problem*, which I will not discuss further here, but see Section 3.10 of *Structure and Randomness*.)

I like to think of the sieving process as being analogous to carving out a sculpture (analogous to the primes) from a block of granite (analogous to the integers). To begin with, one uses crude tools (such as a mallet) to remove large, simple pieces from the block; but after a while, one has to make finer and finer adjustments, replacing the mallet by a chisel, and then by a pick, removing lots and lots of very small pieces, until the final sculpture is complete. Initially, the structure is simple enough that one can easily pick out patterns (such as twins, or arithmetic progressions); but there is the (highly unlikely) possibility that the many small sets removed at the end are just distributed perversely enough to knock out all the patterns one discerns in the initial stage of the process.

Because we cannot exclude this possibility, sieve theory alone does not seem to able to count patterns in primes. However, we can stop the sieve at an earlier level. If we do so, we obtain good counts of patterns, not in primes, but in the larger set of almost primes - which, for the purposes of this talk, one can think of as being defined

as those numbers with no small factors (e.g. no factors less than N^ε for some $\varepsilon > 0$). (This is an oversimplification - one needs to use the weights mentioned above - but it will suffice for this discussion.) Then it turns out that (to oversimplify some more), everything we want to show about the primes, we can show about the almost primes. For instance, it was shown by Chen that there are infinitely many twins $n, n+2$, one of which is a prime and the other is the product of at most two primes. Similarly, given any k, one can show using sieve theory that there are infinitely many arithmetic progressions of length k, each element of which has at most $O_k(1)$ prime factors. More generally, the almost primes behave the way we expect the primes to; distributed pseudorandomly, after taking into account the obvious structures (for instance, almost primes, like the primes, tend to be almost all coprime to 2, coprime to 3, etc.)

Unfortunately, there is still a gap between finding patterns in the almost primes and finding patterns in the primes themselves, because the primes are only a subset of the almost primes. For instance, while the number of primes less than N is roughly $N/\log N$, the number of numbers less than N with no factors less than (say) $N^{1/100}$ is (very) roughly $100N/\log N$. Thus the primes only form a small fraction of the almost primes.

4.1.3. Szemerédi's theorem on arithmetic progressions. Thus far, we have discussed the general problem of how to find patterns in sets such as the primes or almost primes. This problem in general seems to be very difficult, because we do not know how structured or pseudorandom the primes are. There is however one type of pattern which is special - it necessarily shows up in just about any kind of set - structured or random. This type of set is an *arithmetic progression*. In fact, we have the following important theorem:

Theorem 4.1.2 (Szemerédi's theorem)**.** [**Sz1975**] *Let $A \subset \mathbf{Z}$ be a set of integers of positive upper density (which means that $\limsup_{N\to\infty} \frac{1}{2N+1}|A \cap \{-N,\ldots,N\}| > 0$). Then A contains arbitrarily long arithmetic progressions.*

This is a remarkable theorem: it says that if one picks any set of integers at all, so long as it is large enough to occupy a positive

fraction of all integers, that is enough to guarantee the existence of arithmetic progressions of any length inside that set. This is in contrast with just about any other pattern, such as twins: for instance, the multiples of three have positive density, but contain no twins.

There are several proofs of this difficult theorem known. It would be too technical to discuss any of them here in detail, but *very* roughly speaking, all the proofs proceed by dividing the set A into "structured" components (such as sets which are periodic) and "pseudorandom" components (which, roughly, are those components for which the random model gives accurate predictions). One can show that the structured components always generate a lot of arithmetic progressions, and that the pseudorandom components do not significantly disrupt this number of progressions.

Unfortunately, Szemerédi's theorem does not apply to the primes, because they have zero density. (There are some quantitative versions of that theorem that apply to some sets of zero density, but they are not yet strong enough to directly deal with sets as sparse as the primes.)

4.1.4. Putting it all together. To summarise: random models predict arbitrarily long progressions in the primes, but we cannot verify these models. Sieve theory does let us establish long progressions in the almost primes, but the primes are only a fraction of the almost primes. Szemerédi's theorem gives progressions, but only for sets of positive density within the integers.

To proceed, we exploit the fact that the primes have positive *relative* density inside the almost primes, by the following argument (inspired, incidentally, by Furstenberg's ergodic-theoretic proof[**Fu1977**] of Szemerédi's theorem, see Sections 2.10-2.15). We conjecture that the primes obey a certain random model, in which the only structure present being mod p structure for small p. If this is the case, then we are done. If not, it means that there is some specific obstruction to pseudorandomness in the primes, much as the mod 2 or mod 3 obstructions we discussed earlier prevented the most naive random models of the primes from being accurate. We don't know exactly

what that obstruction is, but it turns out that it is possible nevertheless to use that obstruction to modify the random model for the primes to be more accurate, much as we used the mod 2 and mod 3 information previously. We then repeat this process, locating obstructions to pseudorandomness and incorporating them into our random model, until no major obstructions remain. (This can be done after only a bounded number of steps, because one can show (with some effort) that each addition to the random model increases its "energy" by a certain amount, and one can show that the total amount of energy in the model must stay bounded.) As a consequence we know that we can model the primes accurately (at least for the purposes of counting arithmetic progressions) by *some* random model[3].

The above procedure does not give a very tractable formula for what this model is. However, because the primes are a dense subset of the almost primes, which behave like a random subset of the integers, one can show (by a "comparison principle", and oversimplifying somewhat) that the primes must then behave like a random subset of a dense subset B of the integers. But then Szemerédi's theorem applies, and shows that this set B contains plenty of progressions; a random subset of B will then also contain many progressions, and thus the primes will also.

Remark 4.1.3. A simplified version of the above argument, using game theory instead of the above ergodic theory-motivated approach, was recently obtained in [**ReTrTuVa2008**], [**Go2008**].

Notes. This talk first appeared at `terrytao.wordpress.com/2008/01/07`, and was given in the annual joint meeting of the American Mathematical Society and Mathematical Association of America in San Diego in January of 2008. A similar talk also formed the first in my series of four Marker Lectures in Penn State University in November of 2008.

[3] I have not precisely defined what I mean here by "random model", but very roughly speaking, any such model consists of a partition (or σ-*algebra*) of the integers from 1 to N into a bounded number of sets, a specified density for the primes on each such set, and an assumption that the primes behave on each set like a random set with the specified density. Readers familiar with the *Szemerédi regularity lemma* may see a parallel here.

4.2. Linear equations in primes

In the previous lecture, I focused on finding a specific type of pattern inside the prime numbers, namely that of an arithmetic progression $n, n+r, \ldots, n+(k-1)r$. The main reason why the analysis there is specific to progressions is because of its reliance on *Szemerédi's theorem*, which shows that arithmetic progressions are necessarily abundant in sufficiently "large" sets of integers. There are several other variants and generalisations of this theorem known to a few other types of patterns (e.g. polynomial progressions $n + P_1(r), \ldots, n + P_k(r)$, where $P_1, \ldots, P_k$ are polynomials from the integers to the integers with $P_1(0) = \ldots = P_k(0) = 0$ and $r \neq 0$), and in some cases the analogous results about primes are known (e.g. in [**TaZi2008**] we showed that for any given $P_1, \ldots, P_k$ as above, there are infinitely many polynomial progressions of primes).

However, for most patterns, there is no analogue of Szemerédi's theorem, and the strategy used in the previous lecture cannot be directly applied. For instance, it is certainly not true that any subset of integers with positive upper density contains any twins $n, n + 2$; the multiples of three, for instance, form a counterexample, among many others[4].

Furthermore, even in the cases when these methods do work, for instance in demonstrating for each k that there are infinitely many progressions of length k inside the primes, they do not settle the more quantitative problem of how many progressions of length k there are asymptotically in any given finite range of primes, e.g. the primes less than a number N in the asymptotic limit $N \to \infty$. This is because Szemerédi's theorem provides a lower bound for the number of progressions in a large finite set, but not a matching upper bound. (For instance, given a subset A of $\{1, \ldots, N\}$ of density $1/2$, the number of progressions of length 3 in A can be as large as $(1/4 + o(1))N^2$ (if A consists of the even integers from 1 to N, for instance) and as small as $(1/8 + o(1))N^2$ (if A is a randomly chosen set of

[4]In fact, there are so many counterexamples here, that it looks unlikely that the twin prime conjecture can be attacked by this method without a significant new idea; see Section 2.1 of *Structure of Randomness* for further discussion.

density 1/2), and can even be a little bit smaller by perturbing this example slightly.)

On the other hand, as discussed in the previous lecture, one can use standard random models for the primes to predict what the correct asymptotic for these questions should be. For instance, the number of arithmetic progressions $n, n+r, \ldots, n+(k-1)r$ of a fixed length k consisting of primes less than N should be asymptotically

$$(\frac{1}{2(k-1)}(\prod_p \beta_p) + o(1))\frac{N^2}{\log^k N} \tag{4.1}$$

where the product is over all primes p, and the quantity β_p is defined as

$$\beta_p := \frac{1}{p}(\frac{p}{p-1})^{k-1}$$

for $p \leq k$, and

$$\beta_p := (1 - \frac{k-1}{p})(\frac{p}{p-1})^{k-1}$$

for $p \geq k$.

The various terms in this complicated-looking formula can be explained as follows. The "Archimedean" factors $\frac{1}{2(k-1)}$ and N^2 come from the fact that the number of arithmetic progressions $n, n+r, \ldots, n+(k-1)r$ of *natural numbers* less than N is $(\frac{1}{2(k-1)} + o(1))N^2$. The "density" factor $\frac{1}{\log^k N}$ comes from the *prime number theorem*, which roughly speaking asserts that each of the k elements $n, n+r, \ldots, n+(k-1)r$ in a typical arithmetic progression has a $(1 + o(1))\frac{1}{\log N}$ "probability" of being prime. The "local" factors β_p measures how much bias arithmetic progressions with respect to being coprime to a fixed prime p, which is relevant for progressions of primes, since primes of course tend to be coprime to p. More precisely, β_p can be defined as the probability that a random arithmetic progression of length k has all entries coprime to p, divided by the probability that a random collection of k independent numbers are all coprime to p. It is not difficult to show that the product $\prod_p \beta_p$ converges to some finite non-zero number for each k.

Similar heuristic asymptotic formulae exist for the number of many other patterns of primes; for instance, the number of representations $N = p_1 + p_2$ of a large integer N as the sum of two primes

should be equal to $(\prod_p \beta_{p,N} + o(1))N$, where $\beta_{p,N}$ is[5] the probability that two randomly chosen numbers *conditioned* to be coprime to p sum to N modulo p, divided by the probability that two randomly chosen numbers sum to N modulo p. A more general prediction for counting linear patterns inside primes exists, and is essentially the *Hardy-Littlewood prime tuples conjecture*[**HaLi1923**]. This conjecture, which is widely believed to be true, would imply many other conjectures in the subject, such as the twin primes conjecture and the *Goldbach conjecture* (for sufficiently large even numbers). Unfortunately, the cases of the prime tuples conjecture which would have these consequences remain out of reach of current technology.

Using some elementary linear algebra, one can recast the prime tuples conjecture not as a question of finding linear patterns inside primes, but rather that of solving linear equations in which all the unknowns are required to be prime, subject to some additional linear inequalities. For instance, finding progressions of length k consisting entirely of primes less than N is essentially the same as asking for solutions to the system of equations

$$p_2 - p_1 = p_3 - p_2 = \ldots = p_k - p_{k-1}$$

and inequalities

$$0 \leq p_1, \ldots, p_k \leq N$$

where the unknowns $p_1, \ldots, p_k$ are required to be primes. More generally, one could imagine the question of asking the number of k-tuples $(p_1, \ldots, p_k)$ consisting entirely of primes which is contained in some convex set B in $\mathbf{R}^k$ of some intermediate dimension[6] $1 \leq d \leq k$, which is contained in a ball of radius $O(N)$ around the origin. For instance, in the above example B is the 2-dimensional set

$$\{(x_1, \ldots, x_k) \in \mathbf{R}^k : x_2 - x_1 = \ldots = x_k - x_{k-1}; 0 \leq x_1, \ldots, x_k \leq N\}.$$

[5]In particular, $\beta_{2,N} = 0$ for odd N, reflecting the fact that it is very difficult for an odd number to be representable as the sum of two primes. More generally, one can compute that $\beta_{p,N} = 1 + \frac{1}{p-1}$ when p divides N, and $\beta_{p,N} = 1 - \frac{1}{(p-1)^2}$ otherwise.

[6]We make the technical assumption that the linear coefficients of the equations defining the d-dimensional subspace that B lives in are independent of N; k and d are of course also assumed to be independent of N. The constant coefficients, however, are allowed to vary with N; this is the situation that comes up for instance in the Goldbach conjectures.

One can think of the problem of finding points in B as that of solving $k-d$ equations in k unknowns. One can also generalise this problem slightly by enforcing some residue constraints $x_j = a_j \bmod q_j$ on the unknowns, but we will ignore this minor extension to simplify the discussion.

The prime tuples conjecture for this problem can roughly speaking be phrased as follows. Suppose that the number if k-tuples $(n_1, \ldots, n_k)$ in B consisting of *natural numbers* is known to be

$$(\beta_\infty + o(1))N^d$$

for some constant β_∞ independent of N (one can think of this constant as the normalised volume of B). Suppose also that for any fixed prime p, the number of k-tuples in B consisting of natural numbers coprime to p is known to be[7]

$$(\beta_\infty \beta_p + o(1))(1 - \frac{1}{p})^k N^d$$

for some constant *latex* $beta_p$ independent of n. Then the prime tuples conjecture asserts that the number of k-tuples in B consisting of primes should be

$$(\beta_\infty \prod_p \beta_p + o(1)) \frac{N^d}{\log^k N}. \tag{4.2}$$

In particular, this can be shown to imply that if $\beta_\infty > 0$ (thus there are no obstructions to solving the system of equations at infinity) and if $\beta_p > 0$ for all p (thus there are no obstructions to solvability mod p for any p) then there will exist many solutions to the system of equations in primes when N is large enough.

As mentioned earlier, this conjecture remains open in several important cases, most particularly in the one-dimensional case $d = 1$. For instance, the twin prime conjecture would follow from the case $B := \{(x_1, x_2) : x_2 - x_1 = 2, 0 \leq x_1, x_2 \leq N\}$, but this case remains open. However, there has now been significant progress in the higher dimensional cases $d \geq 2$, especially when the *codimension* $k-d$ (representing the number of equations in the system) is low. Firstly, the prime number theorem settles the zero codimension case $d = k$ (and

[7] The factor $(1 - \frac{1}{p})^k$ is natural, as it represents the proportion of tuples of k natural numbers in which all the entries are coprime to p.

is pretty much the only situation in which we can handle a $d = 1$ case). The Hardy-Littlewood circle method, based on Fourier analysis, settles all "non-degenerate" cases when $d \geq \max(k-1, 2)$, where "non-degenerate" roughly speaking means that the problem does not secretly contain a $d = 1$ problem inside it as a lower-dimensional projection (e.g. $B := \{(x_1, x_2, x_3) : x_2 - x_1 = 2, 0 \leq x_1, x_2, x_3 \leq N\}$ would be degenerate; more generally, B is non-degenerate if it is not contained in any hyperplane that can be defined using at most two of the unknowns). It can also handle some cases in which the codimension $k-d$ exceeds 1 (e.g. one could take the Cartesian product of some codimension 1 examples); the precise description of what problems are within reach of this method is a little technical to state and will not be given here.

Ben Green and I were able to establish the following partial result towards this conjecture:

Theorem 4.2.1. [**GrTa2009**], [**GrTa2009c**], [**GrTa2009d**], [**GrTa2009e**], [**GrTa2009f**] *The prime tuples conjecture is true in all non-degenerate situations in which* $d \geq \max(k-2, 2)$. *If the* inverse conjecture for the Gowers norms over the integers *is true, then the prime tuples conjecture is true in all non-degenerate situations in which* $d \geq 2$.

I will say a little bit more about what the inverse conjecture for the Gowers norms is later. This theorem unfortunately does not touch the most interesting case $d = 1$ (when the patterns one is seeking only have one degree of freedom), but it does largely settle all the other cases. For instance, this theorem implies the asymptotic (4.1) for prime progressions of length k for $k \leq 4$ (the cases $k \leq 3$ were established earlier by van der Corput using the circle method), and the case of higher k would also follow from the theorem once the inverse conjecture is proven. As with the circle method, we can also unconditionally handle some cases in which d is less than $\max(k-2, 2)$, but the precise statement here is technical and will be omitted. (Details and further examples can be found in [**GrTa2009**].)

Now I would like to turn to the proof of this theorem. At first glance, the result looks like it is going to be quite complicated, due to the presence of all the different factors in the asymptotic (4.2) that one is trying to prove. However, most of the factors can be dealt

with by various standard tricks. The Archimedean factor β_∞ can be eliminated from the problem by working locally (with respect to the infinite *place*), covering B by cubes of sidelength $o(N)$. For similar reasons, the local factors β_p can be eliminated by working locally mod p (i.e. restricting to a single residue class mod p for various small p). The factors $\frac{1}{\log^k N}$, which come from the density $\frac{1}{\log N}$ of primes in the region of interest (i.e. from the *prime number theorem*), can largely be compensated for (with some effort) from the transference principle technology developed in our earlier paper on long progressions in the primes, which was discussed in the previous lecture. After all this, the problem basically boils down to the following. We have a certain subset A of the integers $\{1, \ldots, N\}$ with some density $0 < \delta < 1$ (one should think of A as being a "model" for the primes, after all the distorting structure coming from local obstructions has been stripped out; in actuality, one has to replace the set A by a weight function $f : \{1, \ldots, N\} \to [0, 1]$ of mean value δ, but let us ignore this technicality to simplify the discussion). We pick a random instance of some linear pattern inside $\{1, \ldots, N\}$ (for sake of concreteness, let us pick a random arithmetic progression $n, n + r, \ldots, n + (k - 1)r$) and ask what is the probability that all the elements of this pattern lie in A. Since A has density δ, we expect each element $n + jr$ of our random progression to have a probability[8] $\delta + o(1)$ to lie in A. Since our pattern consists of k elements, we thus expect

$$\mathbf{P}(n, n + r, \ldots, n + (k - 1)r \in A) = \delta^k + o(1). \tag{4.3}$$

Roughly speaking, the key issue in proving the theorem is to work out some "easily checkable" conditions on A that would guarantee that the heuristic (4.3) is in fact valid. One then verifies that these "easily checkable" conditions do indeed hold for the set A of interest (which is a proxy for the set of primes).

As stated before, we expect $\mathbf{P}(n + jr \in A) = \delta + o(1)$ for each $0 \leq j < k$. Thus (4.3) is asserting in some sense that the events $n + jr \in A$ are "approximately independent". This would be a reasonable assertion if A was *pseudorandom* (i.e. it behaved like a random subset of $\{1, \ldots, N\}$ of the given density δ), and is consistent

[8]This is not quite the case if A is biased to lie on one side of $\{1, \ldots, N\}$ than on the other, but it turns out that one can ignore this possibility.

with the general heuristic from number theory that we expect the prime numbers to behave randomly once all the "obvious" irregularities in distribution (in particular, irregularity modulo p for small p) has been dealt with. But if A exhibits certain types of *structure* (or at least some bias towards structure), then (4.3) can fail. For instance, suppose that A consists entirely of odd numbers. Then, if the first two elements n, $n+r$ of an arithmetic progression lie in A, they are necessarily odd, which then forces the rest of the elements of this progression to be odd. As A is concentrated entirely in these odd numbers, these elements of the progression are thus expected to have an elevated probability of lying in A, and so the left-hand side of (4.3) would be expected to significantly exceed the former once $k \geq 3$. (The asymptotic (4.3) becomes trivially true for $k < 3$.) A similar distorting effect occurs if A is not entirely contained in the odd numbers, but is merely *biased* towards them, in that odd numbers are more likely to lie in A than even numbers. In this example, the bias in A caused the number of progressions to go up from the expected number predicted by (4.3); it is also possible (but more tricky) to concoct examples in which bias in A forces the number of progressions to go down somewhat, though Szemerédi's theorem does prevent one from extinguishing these progressions completely when N is large.

Bias towards odd or even numbers is equivalent to a correlation between A and the *linear character* $\chi(n) := (-1)^n$; the algebraic constraints between the $\chi(n+jr)$, and in particular the relationship

$$\chi(n+2r)\chi(n+r)^{-2}\chi(n) = 1 \tag{4.4}$$

can be viewed as the underlying source of the distorting effect that can prevent (4.3) from holding for $k \geq 3$. The same algebraic constraint holds for any other linear character, e.g. the Fourier character $\chi(n) := e(\xi n)$ (where $e(x) := e^{2\pi i x}$) for some fixed frequency $\xi \in \mathbf{R}$, for much the same reason that two points on a line determine the rest of the line (it is also closely related to the fact that the second derivative of a linear function vanishes). Because of this, we expect (4.3) to be distorted when A *correlates* with such a character (which means that the Fourier coefficient $\sum_{n\in A} \chi(n)$ is unexpectedly large in magnitude).

It turns out that in the case $k = 3$ of progressions of length three, correlation with a linear character is the *only* source of distortion in the count (4.3). A sign of this can be seen from the identity

$$\#\{(n,r) : n, n+r, n+2r \in A\} = \int_0^1 (\sum_{n_1 \in A} e(\xi n_1)(\sum_{n_2 \in A} e(-2\xi n_2))(\sum_{n_3 \in A} e(\xi n_3))\, d\xi$$

which can be viewed as a "Fourier transform" of the algebraic identity (4.4). One can formalise this using (a slight variant of) the above identity and some other Fourier-analytic tools (in particular, the *Plancherel identity*) to conclude

Theorem 4.2.2 (Inverse theorem for length three progressions)**.** *(Informal) Let $k = 3$. Suppose that A is a subset of $\{1, \ldots, N\}$ of density δ for which* (4.3) *fails. Then A correlates with a non-trivial linear character $\chi(n) = e(\xi n)$. ("Non-trivial" basically means that χ oscillates at least once on the interval $\{1, \ldots, N\}$.)*

Applying this theorem in the contrapositive, we conclude that we can justify the asymptotic (4.3) in the $k = 3$ case as long as we can show that A does not correlate with a linear character. In the case when A is a proxy for the primes, this task essentially boils down to that of establishing non-trivial estimates for exponential sums over primes, such as

$$\sum_{p<N} e(\xi p);$$

for technical reasons it is more convenient to deal with slight variants[9] of this sum such as

$$\sum_{n=1}^{N} \Lambda(n) e(\xi n) \tag{4.5}$$

where Λ is the *von Mangoldt function*, or

$$\sum_{n=1}^{N} \mu(n) e(\xi n) \tag{4.6}$$

[9]There are various elementary identities, such as summation by parts, that allow one to express one of these sums in terms of the others. One has a lot of flexibility in here as long as one retains a factor in the sum, such as $\Lambda(n)$ or $\mu(n)$, which is somehow sensitive to the prime factorisation of n.

where μ is the *Möbius function*. The reason for using these functions instead is that they enjoy a number of very useful identities, such as

$$\sum_{n=1}^{\infty} \frac{\Lambda(n)\chi(n)}{n^s} = -\frac{L'(s,\chi)}{L(s,\chi)} \tag{4.7}$$

and

$$\sum_{n=1}^{\infty} \frac{\mu(n)\chi(n)}{n^s} = \frac{1}{L(s,\chi)} \tag{4.8}$$

for any *Dirichlet character* χ (where $L(s,\chi)$ is the *Dirichlet L-function*), and also multiplicative identities such as

$$\Lambda(n) = \sum_{d|n} \mu(d) \log \frac{n}{d} \tag{4.9}$$

and

$$\mu(n) = \sum_{n=abc} \mu(a)\mu(b) \tag{4.10}$$

To cut a long story very short, identities such as (4.7), (4.8) are useful for estimating (4.5), (4.6) respectively in the *major arc* case when ξ is rational or close to rational (with small denominator), while (variants of) identities such as (4.9) or (4.10) (in particular, certain truncated versions of (4.9) and (4.10) such as *Vaughan's identity*) are useful for estimating (4.5), (4.6) respectively in the *minor arc* case when ξ is far from a rational with small denominator (or close to a rational with large denominator). This theory was pioneered by Vinogradov (and also Hardy and Littlewood), and refined and simplified over the years with many contributions by Vaughan, Davenport, Heath-Brown, and others, with the upshot being that we now have a fairly good understanding of sums such as (4.5) and (4.6), and in particular that the sum (4.6) exhibits a strong cancellation (by a factor of $O_A(\log^{-A} N)$ for any fixed A) uniformly in ξ (i.e. we can handle both major and minor arcs with a uniform estimate). Combining this with the inverse theorem and the previous reductions, one can eventually establish the asymptotic (4.1) in the $k = 3$ case[10].

[10]This is not exactly how the original proof of (4.1) by van der Corput in this case proceeded, but both proofs use the same general ingredients and method, i.e. the Hardy-Littlewood circle method and the Vinogradov method for estimating exponential sums.

Now we turn to progressions of longer length, such as the case $k = 4$. Here, linear characters $\chi(n) = e(\xi n)$ continue to cause bias that distorts the expected asymptotic (4.3), and so it is still necessary to control sums such as (4.5) or (4.6) to prevent such bias from occuring. However, a major new difficulty arises that new sources of bias also arise. For instance, if one takes a quadratic character $\chi(n) := e(\xi n^2)$ for some ξ, then one easily verifies the identity

$$\chi(n)\chi(n+r)^{-3}\chi(n+2r)^3\chi(n+3r)^{-1} = 1 \tag{4.11}$$

which reflects the fact that the third derivative of a quadratic function (such as $n \mapsto \xi n^2$) is zero (it also reflects the fact that three points on the graph of a quadratic (i.e. a parabola) determine the entire parabola). One consequence of this is that if $\chi(n), \chi(n+r), \chi(n+2r)$ are all close to 1 (say), then $\chi(n+3r)$ will be also. This constraint between the four values of χ along an arithmetic progression suggests that if A exhibits significant correlation with χ, then the event that n+3r lies in A will be influenced in some non-trivial manner by whether n, $n+r$, and $n+2r$ already lie in A, which will lead to some distortion in (4.3). Thus one will need to update the inverse theorem by taking quadratic characters into account[11]. The most optimistic conjecture in this regard would be

Theorem 4.2.3 (Proposed inverse theorem for length four progressions)**.** *(Informal) Let $k = 4$. Suppose that A is a subset of $\{1, \dots, N\}$ of density δ for which* (4.3) *fails. Then A correlates with a non-trivial quadratic character $\chi(n) = e(\xi_2 n^2 + \xi_1 n)$.*

Unfortunately, this conjecture fails. The easiest way to see this is to consider a *bracket quadratic character*, such as $\chi(n) = e(\lfloor\sqrt{2}n\rfloor\sqrt{3}n)$, where $\lfloor\rfloor$ is the *greatest integer function*. This is not quite a quadratic character, because $\lfloor\rfloor$ is not quite a linear function. However, this function is linear "a positive fraction of the time"; if one picks x and y to be some generic real numbers, one expects $\lfloor x+y\rfloor$ to equal $\lfloor x\rfloor + \lfloor y\rfloor$ about half of the time. Because of this, we see that while the identity (4.11) certainly doesn't hold all the time for $\chi(n)$, it does hold a positive fraction of the time, and this is enough to still cause

[11] Easy examples show that it is possible for a set to correlate with a quadratic character without exhibiting any correlation with linear characters, by choosing a quadratic character with irrational coefficients.

significant bias to disrupt (4.3) if A correlates with this object. It is furthermore possible to concoct examples of sets A that correlate with bracket quadratic characters such as $e(\lfloor\sqrt{2}n\rfloor\sqrt{3}n)$ but not any linear or quadratic characters[12]. Once one throws in these bracket quadratics, it turns out that these do in fact constitute all the possible obstructions to (4.3) holding in the $k = 4$ case, as shown in [**GrTa2009d**]:

Theorem 4.2.4 (Inverse theorem for length four progressions)**.** *(Informal) Let $k = 4$. Suppose that A is a subset of $\{1, \ldots, N\}$ of density δ for which* (4.3) *fails. Then A correlates with a non-trivial bracket quadratic character $\chi(n) = e(\sum_{j=1}^{J} \lfloor\alpha_j n\rfloor \beta_n j + \xi_2 n^2 + \xi_1 n)$ for some real numbers $\alpha_j, \beta_j, \xi_1, \xi_2$ and bounded J.*

The proof of this result involves both Fourier analysis and additive combinatorics, relying heavily on ideas from a paper of Gowers[**Go1998**] on Szemerédi's theorem for progressions of length 4. It will not be discussed here.

In view of the inverse theorem, the problem of establishing the asymptotic (4.1) for length 4 progressions then reduces (by suitable generalisations of the various methods discussed previously) to that of estimating exponential sums of which

$$\sum_{n=1}^{N} \mu(n) e(\lfloor\sqrt{2}n\rfloor\sqrt{3}n)$$

is a typical example. One can begin to apply the methods of Vinogradov and Vaughan to control this type of expression. But one is soon faced with the problem of understanding the distribution of quadratic phases such as $\lfloor\sqrt{2}n\rfloor\sqrt{3}n$, and in particular to estimate exponential sums such as

$$\sum_{n=1}^{N} e(\lfloor\sqrt{2}n\rfloor\sqrt{3}n).$$

[12]The same phenomenon is not visible at the linear level; a bracket linear phase such as $e(\lfloor\sqrt{2}n\rfloor\sqrt{3})$ can be rewritten as $e(\sqrt{6}n)e(-\sqrt{3}\{\sqrt{2}n\})$, which by Fourier series can be expressed as a linear combination of linear characters $e((\sqrt{6}+k\sqrt{2})n)$ for integer k. Note that the same trick does not work for $e(\lfloor\sqrt{2}n\rfloor\sqrt{3}n)$.

This turns out to be somewhat unpleasant; the standard technology of Weyl differencing and the van der Corput lemma (see Section 1.4) eventually works[**GrTa2009e**], but does not scale well to bracket polynomials of higher degree such as $e(\lfloor\lfloor\sqrt{2}n\rfloor\sqrt{3}n\rfloor\sqrt{5}n)$, which would be necessary if one were to extend (4.1) to k beyond 4.

To resolve this, and inspired by the work in ergodic theory by [**FuWe1996**], [**HoKr2005**], [**BeLe2007**], and others, we re-interpreted bracket polynomials from a more dynamical systems perspective. To motivate this, observe that the linear character $\chi(n) = e(\alpha n)$ is closely tied to the circle rotation $T : x \mapsto x + \alpha$ on the unit circle $\mathbf{R}/\mathbf{Z}$, in that the character χ can be described as a function $\chi(n) = F(T^n x_0)$ of an orbit $(T^n x_0)_{n\in\mathbf{Z}}$ on this system, where $x_0 = 0$ is the origin and $F(x) := e(x)$ is the exponential function. In a similar spirit, a quadratic character such as $\chi(n) = e(\alpha\frac{n(n-1)}{2})$ can be expressed in terms of the *skew shift system* $(x, y) \mapsto (x + \alpha, y + x)$ on the torus $(\mathbf{R}/\mathbf{Z})^2$, being of the form $\chi(n) = F(T^n x_0)$ where $x_0 := (0, 0)$ and $F(x, y) := e(y)$. More generally, one can also express bracket quadratic polynomials such as $e(\lfloor\sqrt{2}n\rfloor\sqrt{3}n)$ in the form $\chi(n) = F(T^n x_0)$, where T is now the action of a group element $x \mapsto gx$ on a 2-step nilmanifold G/Γ, and F is some reasonable (e.g. piecewise smooth) function on this nilmanifold. (See Section 2.16 or [**BeLe2007**], [**GrTa2009c**] for details.) The relevance of 2-step nilpotent groups and nilmanifolds to length 4 progressions can be glimpsed in the identity

$$(g^n x)(g^{n+r}x)^{-3}(g^{n+2r}x)^3(g^{n+3r}x)^{-1} = 1$$

which is valid for all g, x in a 2-step nilpotent group G (compare this with (4.11)); it is an instructive exercise to prove this identity and to see how the 2-step nilpotency is used[13]. Indeed, one can reformulate the inverse theorem for length 4 progressions in an equivalent form:

Theorem 4.2.5 (Inverse theorem for length four progressions, again)**.** *(Informal) Let $k = 4$. Suppose that A is a subset of $\{1, \ldots, N\}$ of density δ for which* (4.3) *fails. Then A correlates with a non-trivial* 2-step nilsequence $\chi(n) = F(T^n x_0)$ *for some* 2*-step nilmanifold G/Γ*

[13]To make the connection more precise, one needs a variant of this constraint in which x lies in G/Γ rather than G, which is harder to state; see [**Zi2007**], [**GrTa2009d**] for details.

(of bounded "complexity"), some group rotation $T : x \mapsto gx$, some starting point $x \in G/\Gamma$, and some function $F : G/\Gamma \to \mathbf{C}$ (also of "bounded complexity"; e.g. bounded Lipschitz norm will do).

The precise formulation of the theorem is a little technical; see [**GrTa2009d**] for details. Using this theorem and all the standard machinery, the task of establishing asymptotics such as (4.1) in the $k = 4$ case now reduces to that of understanding sums such as

$$\sum_{n=1}^{N} F(T^n x_0).$$

At this point, one can start using the existing theory of equidistribution of orbits on homogeneous spaces G/Γ (of which nilmanifolds are an important example). It turns out that the existing theory is not quite quantitative enough for our purposes, and we had to develop a quantitative analogue of this theory; see [**GrTa2009c**], [**GrTa2009f**] for more discussion. Anyway, it all works, and gives asymptotics for progressions of length 4 in the primes, as well as other linear patterns of similar "complexity" (e.g. any non-degenerate system of two equations in four prime unknowns is OK). To handle higher patterns, what we need is

Conjecture 4.2.6 (Inverse conjecture for arithmetic progressions)**.** *(Informal) Let $k \geq 3$. Suppose that A is a subset of $\{1, \ldots, N\}$ of density δ for which* (4.3) *fails. Then A correlates with a non-trivial $k-2$-step nilsequence $\chi(n) = F(T^n x_0)$ for some (k-2)-step nilmanifold G/Γ (of bounded "complexity"), some group rotation $T : x \mapsto gx$, some starting point $x \in G/\Gamma$, and some function $F : G/\Gamma \to \mathbf{C}$ (also of "bounded complexity").*

This is a consequence of (and very closely related to) the inverse conjecture for the Gowers norm. It is already known for $k = 3$ and $k = 4$, and hopefully the higher k cases will be resolved in the near future, presumably using a mix of techniques from Fourier analysis, additive combinatorics, and ergodic theory. At this time of writing, this is the only remaining obstacle before we can understand the asymptotics of linear patterns in primes which genuinely involve two or more free parameters (as mentioned earlier, one-parameter problems such as

the twin prime conjecture seem well out of reach of these methods for a number of reasons, one of which is that there is definitely no analogue of the inverse conjecture for such one-parameter patterns).

Notes. This talk first appeared at `terrytao.wordpress.com/2008/11/18`, and was given as the second talk in my series of four Marker Lectures in Penn State University in November of 2008.

4.3. Small gaps between primes

In this lecture, I would like to discuss the recent progress, particularly by Goldston, Pintz, and Yıldırım, on finding small gaps $p_{n+1} - p_n$ between consecutive primes. (See also the surveys [**GoPoYi2005**], [**Gr2006**], [**So2006**]; the material here is based to some extent on these prior surveys.)

The *twin prime conjecture* can be rephrased as the assertion that $p_{n+1} - p_n$ attains the value of 2 infinitely often, where $p_1 = 2, p_2 = 3, p_3 = 5, \ldots$ are the primes. As discussed in previous lectures, this conjecture remains out of reach at present, at least with the techniques centred around counting solutions to linear equations in primes. However, there is another direction to pursue towards the twin prime conjecture which has shown significant progress recently (though, again, there appears to be a significant difficulty in pushing it all the way to the full conjecture). This is to try to show that $p_{n+1} - p_n$ is unexpectedly small for many n. Let us make this a bit more quantitative by posing the following question:

Question 4.3.1. *Let N be a large number. What is the smallest value of $p_{n+1} - p_n$, where p_n is a prime between N and $2N$?*

The twin prime conjecture (or more precisely, the quantitative form of this conjecture coming from the prime tuples conjecture) would assert that the answer to this question is 2 for all sufficiently large N.

There are various ways to get upper bounds on this question. For instance, from *Bertrand's postulate* (which can be proven by elementary means) we know that $p_{n+1} - p_n = O(N)$ for all $N \leq p_n \leq 2N$. The *prime number theorem* asserts that $p_n = (1 + o(1))n \log n$, which

gives $p_{n+1} - p_n = o(N)$; using various non-trivial facts known about the zeroes of the zeta function, one can improve this to $O(N^c)$ for various c (the best value of c known unconditionally is 0.525, see [**BaHaPi2001**]). The *Riemann hypothesis* gives a significantly more precise asymptotic formula for p_n, which ultimately leads to the bound $p_{n+1} - p_n = O(\sqrt{N} \log N)$. These bounds hold for all n in the given range, and so in fact bound the *largest* value of $p_{n+1} - p_n$, not just the smallest. As far as I know, the $O(\sqrt{N} \log N)$ bound for the largest prime gap has not been improved even after one assumes the Riemann hypothesis, though this gap is generally expected to be much smaller than this[14] In the converse direction, the best result is due to Rankin[**Ra1962**], who showed the somewhat unusual bound

$$p_{n+1} - p_n \geq c \log N \frac{(\log \log N)(\log \log \log \log N)}{(\log \log \log N)^2}$$

for some n and some absolute constant $c > 0$ (in fact one can take c arbitrarily close to $2e^\gamma$). Remarkably, this type of right-hand side appears to be a genuine limit of what current methods can achieve (Paul Erdős in fact offered \$10,000 to anyone who could improve the rate of growth of the right-hand side in N).

But for the smallest value of $p_{n+1} - p_n$, much more is known. The prime number theorem already tells us that there are $(1+o(1))N/\log N$ primes between N and $2N$, so from the *pigeonhole principle* we have

$$p_{n+1} - p_n \leq (1 + o(1)) \log N \tag{4.12}$$

for some n.

This bound should not be sharp, since this would imply that the primes are almost equally spaced by $\log N$, which is suspiciously regular behaviour for a sequence as irregular as the primes. To get some intuition as to what to expect, we turn to random models of the primes. In particular, we begin with *Cramér's random model*[**Cr1936**] for the primes, which asserts that the primes between N and $2N$ behave as if each integer in this range had an independent chance of about

[14] In particular, the old conjecture that there always exists at least one prime between two consecutive square numbers remains open, even assuming RH. Cramér conjectured[**Cr1936**] a bound of $(1 + o(1)) \log^2 N$, though it is possible that the constant 1 here may need to be revised upwards to $2e^\gamma \approx 1.1229$ where γ is the Euler-Mascheroni constant; see [**Gr1995**] for details.

$1/\log N$ of being prime. Standard probability theory then shows that the primes are distributed like a *Poisson process* of intensity $1/\log N$. In particular, if one takes an random interval I_λ in $[N, 2N]$ of length $\lambda \log N$ for some $\lambda > 0$, the number of primes $|I_\lambda \cap \mathcal{P}|$ that I_λ captures is expected to behave like a Poisson random variable of mean λ; in other words, we expect

$$\mathbf{P}(|I_\lambda \cap \mathcal{P}| = k) \approx \frac{e^{-\lambda}\lambda^k}{k!} \tag{4.13}$$

for $k = 0, 1, 2, \ldots$. (In contrast, the prime number theorem only gives the much weaker statement $\mathbf{E}|I_\lambda \cap \mathcal{P}| = \lambda + o(1)$.)

Now, as discussed in previous lectures, Cramér's random model is not a completely accurate model for the primes, because it does not reflect the fact that primes very strongly favour the odd numbers, the numbers coprime to 3, and so forth. However, it turns out that even after one corrects for these local irregularities, the predicted Poisson random variable behaviour (4.13) does not change significantly for any fixed λ (e.g. a Poisson process of intensity $2/\log N$ on the odd numbers looks much the same as a Poisson process of intensity $1/\log N$ on the natural numbers, when viewed at scales comparable to $\log N$). This computation was worked out fully by Gallagher[**Ga1976**] as a rigorous consequence of the Hardy-Littlewood prime tuples conjecture[15].

Applying (4.13) for small values of λ (but still independent of N), we see that intervals of length $\lambda \log N$ are still expected to contain two or more primes with non-zero probability, which in particular would imply that $p_{n+1} - p_n \leq \lambda \log N$ for at least one value of n. So one path to creating small gaps between primes is to show that $|I_\lambda \cap \mathcal{P}|$ can exceed 1 for as small a value of λ as one can manage.

One approach to this proceeds by controlling the *second moment*

$$\mathbf{E}|I_\lambda \cap \mathcal{P}|^2; \tag{4.14}$$

the heuristic (4.13) predicts that this second moment should be $\lambda^2 + \lambda + o(1)$ (reflecting the fact that the Poisson distribution has both mean and variance equal to λ). On the other hand, the prime number

[15]On the other hand, these corrections to the Cramér model *do* disrupt (4.13) for very large values of λ; see [**So2007**] for more discussion.

theorem gives the first moment estimate $\mathbf{E}|I_\lambda \cap \mathcal{P}| = \lambda + o(1)$. Also, if $|I_\lambda \cap \mathcal{P}|$ never exceeds 1, then the first and second moments are equal. Thus if one could get the right bound for the second moment, one would be able to show that $p_{n+1} - p_n \leq \lambda \log N$ is possible for arbitrarily small λ.

Second moments such as (4.14) are very amenable to tools from Fourier analysis or complex analysis; applying such tools, we soon see that (4.14) can be re-expressed easily in terms of zeroes to the Riemann zeta function, and one can use various standard facts (or hypotheses) about these zeroes to gain enough control on (4.14) to obtain non-trivial improvements to (4.12). This approach was pursued by many authors [**Ra1937**], [**Er1940**], [**BoDa1966**], leading to non-trivial unconditional results for any $\lambda \geq 1/2$, but it seems difficult to push the method much beyond this. It was later shown in [**GoMo1987**] that the correct asymptotic for (4.14) is essentially equivalent to the Riemann hypothesis combined with a certain statement on pair correlations between zeroes, and thus well out of reach of current technology.

Another method, introduced by Maier[**Ma1985**], is based on finding some (rare) intervals I_λ of numbers of length $\lambda \log N$ for some moderately large λ which contain significantly more primes than average value of λ; if for instance one can find such an interval with over $(k+1)\lambda$ primes in it, then from the pigeonhole principle one must be able to find a prime gap of size at most $\frac{1}{k} \log N$. The ability to do this stems from the remarkable and unintuitive fact that the regularly distributed nature of primes in long arithmetic progressions, together with the tendency of primes to avoid certain residue classes, forces the primes to be *irregularly* distributed in *short* intervals. This phenomenon (which has now been systematically studied as an "uncertainty principle" for equidistribution[**GrSo2007**]), is related to the following curious failure of naive probabilistic heuristics to correctly predict the prime number theorem. Indeed, consider the question of asking how likely it is that a randomly chosen number n between N and $2N$ is to be prime. Well, n will have about a $1 - \frac{1}{2}$ chance of being coprime to 2, a $1 - \frac{1}{3}$ chance of being coprime to 3, and so forth; the *Chinese remainder* theorem also suggests that these events behave

independently. Thus one might expect that the probability would be something like

$$\prod_{p<N}(1-\frac{1}{p}).$$

We then invoke *Mertens' theorem*, which provides the asymptotic

$$\prod_{p<N}(1-\frac{1}{p}) = (e^{-\gamma}+o(1))\frac{1}{\log N}. \tag{4.15}$$

But this is off by a factor of e^γ from what the prime number theorem says the true probability of being prime is, which is $(1+o(1))\frac{1}{\log N}$. This discrepancy reflects the difficulty in cutting off the product in primes (4.15) at the right place (for instance, the *sieve of Eratosthenes* suggests that one might want to cut off at $\sqrt{N}$ instead). At any rate, this e^γ discrepancy can be exploited to find intervals with an above-average number of primes by a "first moment" argument (known as the *Maier matrix method*) that we sketch as follows. Let w be a moderately large number, and let W be the product of all the primes less than w. If we pick a random number n between N and $2N$, then as mentioned before, the prime number theorem says that this number will be prime with probability about $\frac{1}{\log N}$. But if in addition we know that the number is coprime to W, then by the prime number theorem in arithmetic progressions this information boosts the probability of being prime to about $\prod_{p<w}(1-\frac{1}{p})^{-1}\frac{1}{\log N}$, which by (4.15) is about $e^\gamma\frac{\log w}{\log N}$.

Now we restrict attention to numbers n which are equal to a mod W for some $1 \leq a \leq w$. By the prime number theorem, about $\frac{w}{\log w}$ of the a are prime and thus coprime to W. Combining this with the previous discussion, we see that the total probability that such a number n is prime is about $\frac{1}{\log w} \times e^\gamma \frac{\log w}{\log N} = e^\gamma \frac{1}{\log N}$.

On the other hand, the set of numbers n which are equal to a mod W for some $1 \leq a \leq w$ is given by a sequence of intervals of length w. By the pigeonhole principle, we must therefore have an interval of length w on which the density of primes is at least $e^\gamma \frac{1}{\log N}$ - which is greater than the expected density by a factor of e^γ.

One can make the above arguments rigorous for certain ranges of interval length, and by combining this with the pigeonhole principle

one can eventually improve (4.12) by a factor of e^γ:

$$p_{n+1} - p_n \leq (e^{-\gamma} + o(1)) \log N.$$

This is better than the bound of $(\frac{1}{2} + o(1)) \log N$ obtained by the second moment method, but on the other hand the latter method establishes that a positive proportion of primes have small gaps; Maier's method, by its very nature, is restricted to a very sparse set of primes (note that w is much smaller than W).

In a series of papers, Goldston and Yıldırım improved the numerical constants in these results by a variety of methods including those mentioned above, as well as replacing some of the reliance on information on zeroes of the zeta function with tools from sieve theory instead. To oversimplify substantially, the latter idea is to try to control the set of primes $\mathcal{P}$ in terms of a larger set $\mathcal{A}P$ of *almost primes* - numbers with few prime factors[16]. For instance, to control the second moment $\mathbf{E}|I_\lambda \cap \mathcal{P}|^2$, one can take advantage of the Cauchy-Schwarz inequality

$$\mathbf{E}|I_\lambda \cap \mathcal{P}|^2 \geq \frac{\mathbf{E}|I_\lambda \cap \mathcal{P}||I_\lambda \cap \mathcal{A}P|}{|I_\lambda \cap \mathcal{A}P|^2}.$$

The denominator on the right-hand side involves only almost primes and can be computed easily by sieve theory methods. The numerator involves primes, but only one prime at a time; note that this quantity is roughly counting the set of pairs p, q where p is prime, q is almost prime, and p and q differ by at most $\lambda \log N$. This is in contrast to the left-hand side, which is counting pairs p, q that are both prime and differ by at most $\lambda \log N$. We do not know how to use sieve theory to count the latter type of pattern (involving more than one prime); but sieve theory is perfectly capable of counting the former type of pattern, so long as we understand the distribution of primes in relatively sparse arithmetic progressions. The standard tool for this is the *Bombieri-Vinogradov theorem*[**Bo1987**], which roughly speaking asserts that the primes from N to $2N$ are well distributed in "most" residue classes q, as long as q stays significantly smaller than

[16]Strictly speaking, one does not actually work with a set of almost primes, but rather with a weight function or sieve which is large on almost primes and small for non-almost primes, but let us ignore this important technical detail to simplify the exposition.

$\sqrt{N}$; it can be viewed as an averaged version of the *generalised Riemann hypothesis* that can be proven unconditionally[17]. Using such tools, various improvements to (4.12) were established. Finally, in [**GoPoYi2005a**] it was shown that

$$p_{n+1} - p_n \leq \lambda \log N \tag{4.16}$$

held for some n and any $\lambda > 0$ (provided N was large enough depending on λ), or equivalently that $p_{n+1} - p_n = o(\log N)$; this was later improved in [**GoPoYi2007**], to $p_{n+1} - p_n = O(\sqrt{\log N}(\log\log N)^2)$. Assuming a strong version of the Bombieri-Vinogradov theorem (in which q is allowed now to get close to N rather than to $\sqrt{N}$), known as the *Elliott-Halberstam conjecture*, this was improved further to the striking result

$$p_{n+1} - p_n \leq 16,$$

thus there are infinitely many pairs of primes which differ by at most 16. This is a remarkable "near miss" to the twin prime conjecture, though it seems clear that substantial new ideas would be needed to reduce 16 all the way to 2.

Let's now discuss some of the ideas involved. As with the previous arguments, the key idea is to find groups of integers which tend to contain more primes than average. Suppose for instance one could find a certain random distribution of integers n where[18] n, $n+2$, and $n+6$ each had a probability strictly greater than $1/3$ of being prime. By linearity of expectation, we thus see that the expected number of primes in the set $\{n, n+2, n+6\}$ exceeds 1; thus, with positive probability, there will be at least two primes in this set, which then necessarily differ by at most 6.

Now, of course, the prime number theorem tells us that for n chosen uniformly at random from N to $2N$, the probability that n, $n+2$, or $n+6$ are prime is only about $1/\log N$. So for this type of strategy to work, one would have to pick a highly non-uniform

[17] The key point here is that while it is possible for the primes to be irregular with respect to a few such small moduli, the "orthogonality" of these moduli with respect to each other makes it impossible for the primes to be simultaneously irregular with respect to many of these moduli at once.

[18] We choose these separations for our discussion because it is not possible to make three large prime numbers bunch up any closer than this; for instance, n, $n+2$, $n+4$ cannot be all be prime for $n > 3$, since at least one of these numbers must be divisible by 3.

distribution for n, in which n, $n+2$, and $n+6$ are already close to being prime already. The extreme choice would be to pick n uniformly among all choices for which n, $n+2$, and $n+6$ are simultaneously prime, but we of course don't even know that any such primes exist (this is strictly harder than the twin prime conjecture!) But what we can do instead is pick n uniformly among all choices such that n, $n+2$, and $n+6$ are *almost prime* (where we shall be vague for now about what "almost prime" means). Thanks to sieve theory, we *can* assert the existence of many numbers n of this form, and get a good count as to how many there are. Also, since the primes have positive density inside the almost primes, it is quite reasonable that the conditional probabilities

$$\mathbf{P}(n \text{ prime}|n, n+2, n+6 \text{ almost prime}),$$
$$\mathbf{P}(n+2 \text{ prime}|n, n+2, n+6 \text{ almost prime}),$$
$$\mathbf{P}(n+6 \text{ prime}|n, n+2, n+6 \text{ almost prime}),$$

are large. Indeed, using sieve theory techniques (and the Bombieri-Vinogradov theorem), we can bound each of these probabilities from below by a positive constant (plus a $o(1)$ error). Unfortunately, even if we optimise the sieve that produces the almost primes, this constant is too small (typically one gets numbers of the order of $1/20$ or so, rather than $1/3$). Assuming the Elliot-Halberstam conjecture (which allows us to raise the level of sieving substantially) yields a significant improvement, but one that still falls short of the desired goal. One can of course also add more numbers to the mix than just $n, n+2, n+6$, e.g. looking at those n for which $n+h_1, \ldots, n+h_k$ are simultaneously almost prime, for some suitably chosen $h_1, \ldots, h_k$; on the one hand this lowers the threshold of probability (currently at $1/3$) that one needs to obtain, but unfortunately this is more than canceled out by the multidimensional sieving one needs to do when restricting all of these numbers to be almost prime.

To get around this, a new idea was introduced: instead of requiring numbers such as n, $n+2$, and $n+6$ to separately be almost prime, ask instead for the *product* $n(n+2)(n+6)$ to be almost prime (for a somewhat more relaxed notion of "almost prime"). This turns out to be more efficient, as it lowers the number of summations involved in

the sieve. One has to carefully select how one defines almost prime here (it is roughly like asking for the product $(n+h_1)\dots(n+h_k)$ to have at most $2k+o(k)$ prime factors, with the $o(k)$ factor being remarkably crucial to the delicate analysis); but to cut a long story short, one can establish probability bounds of the form[19]

$$(4.17)\quad \mathbf{P}(n+h_j \text{ prime}|(n+h_1)\dots(n+h_k) \text{ almost prime}) \geq \frac{c-o(1)}{k}$$

for all $1 \leq j \leq k$ and some absolute constant $c > 0$.

As soon as one assumes any non-trivial portion of the Elliott-Halberstam conjecture, the quantity c in the above inequality can be made to exceed 1 (for k large enough), leading to the conclusion that there exist infinitely many bounded prime gaps, $p_{n+1} - p_n = O(1)$; pushing the machinery to their limit (taking $\{h_1,\dots,h_k\} = \{7,11,13,17,19,23\}$ to be the first six primes larger than 6), one obtains the bound of 16. But without this conjecture, and just using Bombieri-Vinogradov, then after optimising everything in sight, one can get c arbitrarily close to 1, but not quite exceeding 1. To compensate for this, the authors also started looking at the nearby numbers $n+h$ where h was not equal to $h_1,\dots,h_k$. Here, of course, there is no particularly good reason for $n+h$ to be prime, since it is not involved as a factor to the almost prime quantity $(n+h_1)\dots(n+h_k)$; but one can show that, for generic values of h, one has

$$\mathbf{P}(n+h \text{ prime}|(n+h_1)\dots(n+h_k) \text{ almost prime}) \geq \frac{c'+o(1)}{\log N}$$

for some $c' > 0$ (there is an additional singular series factor involving the prime factors of $h-h_j$ which can be easily dealt with, that I am suppressing here.) Thus, for any $\lambda > 0$, we expect (from linearity of expectation) that for $(n+h_1)\dots(n+h_k)$ almost prime, the expected number of $h = O(\lambda \log N)$ (including the k values $h_1,\dots,h_k$) for which $n+h$ is prime is at least

$$k(\frac{c}{k}+o(1)) + \lambda \log N \frac{c'+o(1)}{\log N} = c + c'\lambda + o(1).$$

[19] More precisely, one needs to compute sums such as $\sum_{n=1}^N \Lambda(n+h_j)\Lambda_R((n+h_1)\dots(n+h_k))^2$, where Λ is the *von Mangoldt function* and Λ_R is a *Selberg sieve*-type approximation to that function.

Since we can make c arbitrarily close to 1, the extra term $c'\lambda$ can push this expectation to exceed 1 for any choice of λ, and this leads to the desired bound[20] (4.16) for any $\lambda > 0$.

It is tempting to continue to optimise these methods to improve the various constants, and I would imagine that the bound of 16, in particular, can be lowered somewhat (still assuming the Elliott-Halberstam conjecture). But there seems to be a significant obstacle to pushing things all the way to 2. Indeed, the parity problem (see Section 3.10 of *Structure and Randomness*) tells us that for any reasonable definition of "almost prime" which is amenable to sieve theory, the primes themselves can have density at most $1/2$ in these almost primes. Since we need the density to exceed $1/k$ in order for the above argument to work, it is necessary to play with at least three numbers (e.g. n, $n+2$, $n+6$), which forces the bound on the prime gaps to be at least[21] 6. But it may be that a combination of these techniques with some substantially new ideas may push things even further.

Notes. This talk first appeared at `terrytao.wordpress.com/2008/11/19`, and was given as the third talk in my series of four Marker Lectures in Penn State University in November of 2008.

Emmanuel Kowalski pointed out that Gallagher's conditional argument[**Ga1**] can in fact be extended to give Poisson-type statistics for (say) twin primes in intervals of size $O(\log^2 N)$ in $[N, 2N]$, and also mentioned his Bourbaki exposé [**Kow2006**] on the above work.

4.4. Sieving for almost primes and expanders

In this final lecture, I discuss the recent work of Bourgain, Gamburd, and Sarnak on how *arithmetic combinatorics* and *expander graphs* were used to sieve for *almost primes* in various *algebraic sets*.

[20] The subsequent improvement to (4.16) proceeds by enlarging k substantially, and by a preliminary sieving of the small primes, but we will not discuss these technical details here.

[21] Indeed, this bound has been obtained for *semiprimes* (products of two primes) - which are subject to the same parity problem restriction as primes, but are slightly better distributed; see [**GrRoSp1980**].

In previous lectures, we considered the problem of detecting tuples of primes in various linear or convex sets; in particular, we considered the size of sets of the form $V \cap \mathcal{P}^k$, where $\mathcal{P} = \{2, 3, 5, \ldots\}$ is the set of primes, and V is some *affine subspace* of $\mathbf{R}^k$. For instance, the twin prime conjecture would correspond to the case when $k = 2$ and $V = \{(x, x+2) : x \in \mathbf{R}\}$, while Theorem 4.1.1 would correspond to the case $V = \{(x, x+r, \ldots, x+(k-1)r) : x, r \in \mathbf{R}\}$.

We refer to elements of $\mathcal{P}^k$ as *prime points*. The *prime tuples conjecture*[**HaLi1923**] implies the following qualitative criterion for when such a set of prime points should be "large":

Conjecture 4.4.1 (Qualitative prime tuples conjecture)**.** *Let V be an affine subspace of $\mathbf{R}^k$. Suppose that*

(1) *(No obstructions at infinity) For any N, $V \cap \mathbf{Z}^k_{>N}$ affinely spans all of V, where $\mathbf{Z}_{>N} := \{n \in \mathbf{Z} : n > N\}$. (In particular, $V \cap \mathbf{Z}^k_{>N}$ is non-empty.)*

(2) *(No obstructions at q) For any $q > 1$, $V \cap (\mathbf{Z}^*_q)^k$ affinely spans all of V, where $\mathbf{Z}^*_q := \{n \in \mathbf{Z} : (n, q) = 1\}$. (In particular,$V \cap (\mathbf{Z}^*_q)^k$ is non-empty.)*

Then $V \cap \mathcal{P}^k$ affinely spans all of V. (In particular, V contains at least one prime point.)

Both of the hypotheses in this conjecture are easily verified for any given V, the first by (integer) linear programming and the second by modular arithmetic. This conjecture would imply several other results and conjectures in number theory, including the twin prime conjecture and Theorem 4.1.1. Needless to say, it remains open in general (though the results mentioned in the previous lecture give partial results in the case when V is at least two-dimensional and non-degenerate).

Now we attempt to generalise the above conjecture to the setting in which V is an *algebraic variety* rather than an affine subspace. (This would cover some famous open problems in number theory, for instance the *Landau problem* that asks whether there are infinitely many primes of the form $n^2 + 1$.) The notion of a set affinely spanning V is then naturally replaced by the notion of a set being

Zariski dense in V, which means that the set is not contained in any strictly smaller subvariety of V. One could then formulate a naive generalisation of the above conjecture by replacing "affine space" and "affinely spans all of" with "algebraic variety" and "is Zariski dense in" respectively. However, the hypotheses are now no longer easy to verify; indeed, just the problem of determining whether V contains an integer point $\mathbf{Z}^k$ is essentially *Hilbert's tenth problem*, which by *Matiyasevich's theorem*[**Ma1970**] is known to be undecidable[22] for general V. Indeed, since one can encode any computable set in terms of the integer points of a variety V, it is not too difficult to see that this conjecture fails in general.

Since arbitrary algebraic varieties are far too general to have any hope of a reasonable theory, one should look for prime points in much more special sets. An important class[23] here is that of an orbit Λb in $\mathbf{Z}^k$, where b is some vector in $\mathbf{Z}^k$ and Λ is some finitely generated subgroup of $SL_k(\mathbf{Z})$. Of course one should take b to be primitive (not a multiple of any smaller vector), since one clearly will have a difficult time finding prime points in Λb otherwise.

The orbit Λb will be Zariski dense in some algebraic variety V, and is clearly a collection of integer points (though it may not cover all of $V \cap \mathbf{Z}^k$). Assuming no local obstructions at infinity or at q (which means that $\Lambda b \cap \mathbf{Z}^k_{>N}$ and $\Lambda b \cap (\mathbf{Z}^*_q)^k$) are Zariski dense in V), one could then conjecture that $\Lambda b \cap \mathbf{P}^k$ is also Zariski dense in V (which, if V is infinite, would in particular imply that the orbit Λb contains infinitely many prime points).

For simplicity let us restrict attention to the two-dimensional case $k = 2$, which is already highly non-trivial; Bourgain, Gamburd and Sarnak have recently begun to get some preliminary results in $k = 3$ but I will not discuss them here. Thus Λ is now a finitely generated subgroup of $SL_2(\mathbf{Z})$. If this subgroup is *elementary* - e.g. if it is cyclic - then the orbit Λb can be exponentially sparse (a ball of radius

[22] An amusing historical connection here: one of the first demonstrations[**DaPuRo1961**] of the undecidability of Hilbert's tenth problem was conditional on the existence of arbitrarily long progressions of primes (i.e. Theorem 4.1.1), although subsequent proofs did not need this fact.

[23] One can also consider the slightly more general set of images $F(\Lambda b)$ under a polynomial map, but for simplicity let us stick to just orbits.

R may only contain $O(\log R)$ points), and it becomes extremely difficult to do any sieving or primality detection[24]. It thus makes sense to restrict attention to non-elementary subgroups of $SL_2(\mathbf{Z})$ - groups which contain a copy of the free non-abelian group on two generators, or equivalently any group whose Zariski closure is all of SL_2 (or equivalently yet again, a group whose limit set consists of more than one point). In this situation, Bourgain, Gamburd, and Sarnak conjectured:

Conjecture 4.4.2. [**BoGaSa2006**] *Let Λ be a non-elementary subgroup of $SL_2(\mathbf{Z})$, and let b be a primitive element of $\mathbf{Z}^2$. Suppose that there are no local obstructions at infinity or at finite places q. Then $\Lambda b \cap \mathcal{P}^2$ is Zariski dense in the plane (in particular, $\Lambda b \cap \mathcal{P}^2$ is infinite).*

This conjecture remains open. However, as in the linear situation, one can make progress[25] if one replaces primes with *almost primes* - products of at most r primes for some bounded r. In particular, Bourgain, Gamburd, and Sarnak were able to show

Theorem 4.4.3. [**BoGaSa2006**] *Let Λ, b be as in Conjecture 4.4.2. Then there exists an r such that $\Lambda b \cap \mathcal{P}_r^2$ is Zariski dense in the plane, where $\mathcal{P}_r$ is the set of numbers that are the product of at most r primes.*

Several further generalisations and extensions of this result, with a similar flavour, are known, but will not be discussed here. There are a number of amusing special cases of these results, for instance one can show that there exist infinitely many *Appollonian circle packings* of the unit circle by four other mutually tangent circles, all of whose radii is the reciprocal of an almost prime, or infinitely many *Pythagorean triples* whose area is an almost prime (for a sufficiently large r in the definition of "almost prime").

---

[24] In this case, the problem becomes comparable to such notoriously difficult questions as whether there are infinitely many Mersenne primes.

[25] There are certainly prior results for nonlinear patterns in the almost primes; for instance, it is a famous result of Iwaniec[**Iw1978**] that there are infinitely many numbers of the form $n^2 + 1$ that are the product of at most two primes.

Let me now discuss some of the key ideas in the proof of this theorem. One begins by rephrasing the question in a more quantitative (or finitary) manner. In the linear case, this would be done by counting the number of points in $\Lambda b \cap \mathcal{P}_r^2$ that lie inside some large Euclidean ball, thus using the Euclidean (or Archimedean) notion of distance to localise the problem. This can also be done here, but it turns out to be more convenient to instead use the *word metric* induced by the finite generating set S of Λ (which we can take to be symmetric for convenience, thus $S = S^{-1}$). One thus looks at sets of the form $B_R b \cap \mathcal{P}_r^2$, where $B_R \subset \Lambda$ consists of all words formed by products of at most R elements of S. A major new difficulty here compared to the linear theory is the exponential growth[26] of B_R (a consequence of the non-elementary nature of Λ).

The next step is to use sieve theory. Recall the *sieve of Eratosthenes*, which expresses the set of all (large) primes as the integers, minus the multiples of two, minus the multiples of three, and so forth. Using the *inclusion-exclusion principle*, we can thus view the indicator function $1_{\mathcal{P}}$ of the primes, when restricted to an interval such as $[N, 2N]$, as equal to 1, minus the indicator function $1_{2\mathbf{Z}}$ of the even numbers, minus the indicator function $1_{3\mathbf{Z}}$ of the multiples of three, plus the indicator function $1_{6\mathbf{Z}}$ of the multiples of six, and so forth. This leads to the *Legendre sieve*

$$1_{\mathcal{P}} = \sum_d \mu(d) 1_{d\mathbf{Z}}, \tag{4.18}$$

valid in an interval $[N, 2N]$ as long as one restricts d to those integers which are products of primes less than N. Here $\mu(d)$ is the *Möbius function*.

The basic idea of sieve theory is to replace the indicator function of the primes (or almost primes) by a more general divisor sum

$$\sum_d c_d 1_{d\mathbf{Z}},$$

where the sieve weights c_d are chosen in order to optimise the final bounds in the sieve (they typically resemble "smoothed out" versions

[26]Though it is not immediately apparent, the same problem also arises if one uses Euclidean balls instead of word metric balls, due to the multiplicative rather than additive nature of the group Λ.

of the Möbius function in order that these sieves be large on the almost primes and small elsewhere). In order for the sieve to be practical, one wants to restrict d in this sum to be relatively small, for instance $d \leq N^\theta$ for some absolute constant $0 < \theta < 1$ (values such as $\theta = 1/4$ are fairly typical). The selection of the sieve weights c_d is now a well-developed science (see Section 3.10 of *Structure and Randomness* for further discussion), and Bourgain, Gamburd and Sarnak basically use off-the-shelf sieves (in particular, combinatorial sieves and the Selberg sieve) in their work. Inserting these standard sieves into the problem at hand, the task of counting almost primes in the finite set $B_R b$ then quickly reduces to the question of getting good estimates on sets such as $B_R b \cap (q\mathbf{Z})^2$ for various q. This amounts to much the same thing as asking for good equidistribution bounds for B_R modulo q, thus we project the generating set S, and the ball B_R it produces, from $SL_2(\mathbf{Z})$ to $SL_2(\mathbf{Z}_q)$. For sieving purposes it turns out to be necessary to consider all squarefree moduli q, but for simplicity we shall only discuss the (massively easier) case when q is prime.

The reduction to an equidistribution problem converts the original sieving problem to a more combinatorial one, involving the *Cayley graph* G on $SL_2(\mathbf{Z}_q)$ induced by the set S, thus two vertices $x, y \in SL_2(\mathbf{Z}_q)$ are connected by an edge in G if yx^{-1} lie in S (modulo q). The image of the ball B_R in $SL_2(\mathbf{Z}_q)$ is then the set of points one can reach in the graph G from the origin by walking on a path of length at most R. The desired equidistribution result one needs can then be viewed as a mixing result for the random walk along the graph G.

Standard graph theory then tells us that the task reduces to showing that the graphs G form a family of *expander graphs* as $q \to \infty$ (recall we are restricting q to be prime for simplicity). There are many equivalent definitions of what an expander graph is, but let us give a spectral definition that is specialised to Cayley graphs. The symmetric generating set S induces a natural measure

$$\mu := \frac{1}{|S|} \sum_{s \in S} \delta_s$$

that is the uniform distribution on S, which controls the random walk along G; note that if $f : SL_2(\mathbf{Z}_q) \to \mathbf{C}$ is a function, then the

convolution $f * \mu : SL_2(\mathbf{Z}_q) \to \mathbf{C}$ is another function, whose value at any vertex is the average value of f at all the neighbours of x. The operation $f \mapsto f * \mu$ is then a self-adjoint contraction on $l^2(SL_2(\mathbf{Z}_q))$ which leaves the function 1 invariant, so its largest eigenvalue λ_1 is equal to 1. The expander graph condition is then equivalent to the existence of a *spectral gap* $\lambda_2 \leq 1-c$ for the second largest eigenvalue, where $c > 0$ is a constant independent of q.

Of course, to have a spectral gap, one necessary condition is that λ_2 be strictly less than 1. This can be easily seen to be equivalent to the statement that G is connected, which in turn is equivalent to the statement that the projection of S to $SL_2(\mathbf{Z}_q)$ generates all of $SL_2(\mathbf{Z}_q)$. This statement can be verified to be true, either by direct consideration of all possible subgroups of $SL_2(\mathbf{Z}_q)$, or by the *strong approximation property*. However, mere connectedness is not enough to ensure that a Cayley graph is an expander family (which can be viewed as a sort of "robust" version of connectedness, which can survive the deletion of large numbers of edges). For instance, the Cayley graph of the generating set $\{-1, +1\}$ in $\mathbf{Z}/N\mathbf{Z}$ is connected, but does not form an expander family as $N \to \infty$; the second largest eigenvalue[27] is about $1 - O(1/N^2)$ only.

Obtaining the spectral gap property requires more work. When the original subgroup Λ of $SL_2(\mathbf{Z})$ is as large as a finite index subgroup (in particular, if it is a *congruence subgroup*), this gap follows from a celebrated theorem of Selberg[**Se1965**] providing a similar spectral gap for arithmetic quotients of the upper half-plane. Smaller examples (in which the index is now infinite) were first constructed in [**Sh1997**], [**Ga2002**], with the latter following the method of [**SaXu1991**]. Then in [**BoGa2008**], this method was extended using additional tools from additive combinatorics to handle all non-elementary subgroups in the case of prime q.

Let us now describe the method of proof. As mentioned briefly earlier, the existence of a spectral gap implies a strong mixing property: the iterated convolutions $\mu^{(n)} := \mu * \ldots * \mu$ of the probability

[27] One can also see that the random walk on this Cayley graph takes a long time (about $O(N^2)$ steps) before it mixes to be close to the uniform distribution; with expander graphs on a set of N vertices, mixing instead occurs in time $O(\log N)$, thanks to the spectral gap.

measure μ (which can be interpreted as the probability distribution of a random walk on n steps) converges exponentially fast to the constant distribution on $SL_2(\mathbf{Z}_q)$. Since the latter distribution has an l^2 norm of $O(q^{-3/2})$, we see in particular that for any fixed $\varepsilon > 0$, we will have

$$\|\mu^{(n)}\|_{l^2(SL_2(\mathbf{Z}_q))} = O(q^{-3/2+\varepsilon}) \tag{4.19}$$

once n is a sufficiently large multiple of $\log q$. This can also be seen explicitly from the trace formula

$$\|\mu^{(n)}\|^2_{l^2(SL_2(\mathbf{Z}_q))} = \frac{1}{|SL_2(\mathbf{Z}_q)|} \sum_j \lambda_j^n. \tag{4.20}$$

In general, this implication between spectral gap and rapid mixing (4.19) cannot be reversed; the problem is that λ_2 only directly influences one term in the summation on the right-hand side of (4.20), and so upper bounds on the left-hand side do not translate efficiently to upper bounds on λ_2. However, there is an algebraic miracle that happens in the case of groups such as $SL_2(\mathbf{Z}_q)$ that allows one to reverse the implication:

Lemma 4.4.4 (Frobenius lemma)**.** *Let q be prime. Then every non-trivial finite-dimensional unitary representation of $SL_2(\mathbf{Z}_q)$ has dimension at least (q-1)/2.*

Proof. Observe that $SL_2(\mathbf{Z}_q)$ can be generated by parabolic elements, so given a non-trivial representation $\rho : SL_2(\mathbf{Z}_q) \to U(V)$, there exists a parabolic element a whose representation $\rho(a)$ is non-trivial. By a change of basis we may take

$$a = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

On the one hand, we have $a^q = 1$ and hence $\rho(a)^q = 1$; thus all eigenvalues of $\rho(a)$ are q^{th} roots of unity. On another hand, $\rho(a)$ is non-trivial, so at least one of the eigenvalues of $\rho(a)$ differs from 1. Thirdly, conjugating a by diagonal matrices in $SL_2(\mathbf{Z}_q)$, we see that a is conjugate to a^m whenever m is a quadratic residue mod q, and so the eigenvalues of $\rho(a)$ must be stable under the operation of taking m^{th} powers. On the other hand, there are $(q-1)/2$ quadratic

residues. Putting all this together we see that $\rho(a)$ must take at least $(q-1)/2$ distinct eigenvalues, and the claim follows. □

Remark 4.4.5. For our purposes, the exact value of $(q-1)/2$ is irrelevant; any multiplicity which grows like a power of q would suffice.

Applying this lemma to the eigenspace of λ_2, we obtain

Corollary 4.4.6. *The second eigenvalue λ_2 of the operation $f \mapsto f*\mu$ appears with multiplicity at least $(q-1)/2$.*

Combining this corollary with (4.20), one can now reverse the previous implication and obtain a spectral gap $\lambda_2 \leq 1-c$ as soon as one gets a mixing estimate (4.19) for some $n = O(\log q)$ and some sufficiently small ε.

The task is now to obtain the mixing estimate (4.19). The quantity $\|\mu^{(n)}\|_{l^2(SL_2(\mathbf{Z}_q))}$ starts at 1 when $n=0$ and decreases with n. If we assume (as we may) that S generates a free group, then it is not hard to see that $\mu^{(n)}$ expands rapidly for $n \ll \log q$ (because all the words generated by S will be distinct until one encounters the "wrap-around" effect of taking residues modulo q). Using this one can get a preliminary mixing bound

$$\|\mu^{(n)}\|_{l^2(SL_2(\mathbf{Z}_q))} \leq q^{-\delta}$$

for some absolute constant $\delta > 0$ and some $n = O(\log q)$. Also, since S modulo q generates all of $SL_2(\mathbf{Z}_q)$, we know that the probability measure $\mu^{(n)}$ is not trapped inside any proper subgroup H of $SL_2(\mathbf{Z}_q)$; indeed, using the classification of subgroups of $SL_2(\mathbf{Z}_q)$ (or some general "escape from subvarieties" machinery of [**EsMoOh2005**]) one can show that

$$\mu^{(n)}(H) \leq q^{-\delta}$$

for any such subgroup, and some $n = O(\log q)$. The result now follows from iterating the following lemma, which is the heart of the argument:

Lemma 4.4.7 (l^2 flattening lemma)**.** *Let ν be a symmetric probability measure on $SL_2(\mathbf{Z}_q)$ which is a little bit dispersed in the sense that*

$$\|\nu\|_{l^2(SL_2(\mathbf{Z}_q))} \leq q^{-\delta}$$

for some $\delta > 0$, and is not concentrated in a subvariety in the sense that $\nu * \nu(H) \leq q^{-\delta}$ for any proper subgroup H of $SL_2(\mathbf{Z}_q)$. Suppose also that ν is not entirely flat in the sense that

$$\|\nu\|_{l^2(SL_2(\mathbf{Z}_q))} \geq q^{-3/2+\delta}$$

(note that the minimal l^2 norm for a probability measure is comparable to $q^{-3/2}$, attained for the uniform distribution). Then $\nu * \nu$ is "flatter" than ν in the sense that

$$\|\nu * \nu\|_{l^2(SL_2(\mathbf{Z}_q))} \leq q^{-\varepsilon} \|\nu\|_{l^2(SL_2(\mathbf{Z}_q))}$$

for some $\varepsilon > 0$ depending on δ.

In the special case when ν is the uniform distribution on some set A, the flattening lemma is very close to the following theorem of Helfgott [**He2008**]:

Theorem 4.4.8 (Product theorem)**.** *Let q be a prime. Let A be a subset of $SL_2(\mathbf{Z}_q)$ which is not too big in the sense that $|A| \leq q^{3-\delta}$ for some $\delta > 0$, and which is not contained in any proper subgroup H of $SL_2(\mathbf{Z}_q)$. Then $|A \cdot A \cdot A| \geq |A|^{1+\varepsilon}$ for some $\varepsilon > 0$ depending on δ.*

Indeed, by using some standard additive combinatorics, in particular a (non-commutative version of) the Balog-Szemerédi-Gowers lemma (which can be found for instance in [**TaVu2006**]), which connects "statistical" multiplication, such as that provided by convolution $\mu, \nu \mapsto \mu * \nu$, with "combinatorial" multiplication, coming from the product set operation $A, B \mapsto A \cdot B$, one can show that these two statements are in fact equivalent to each other.

The product theorem is a manifestation of certain "nonlinear" or "noncommutative" behaviour in the group $SL_2(\mathbf{Z}_p)$; see Section 2.3 of *Structure and Randomness* for a bit more discussion on this. For now, let me just say that Helfgott's proof on this uses a variety of algebraic and combinatorial computations exploiting the special structure of $SL_2(\mathbf{Z}_p)$ (especially how commutativity or non-commutativity of various elements in this group interact with the trace of various combinations of these elements), as well as the following sum-product estimate:

Theorem 4.4.9 (Sum-product theorem)**.** [**BoKaTa2004**], [**BoKo2003**] *Let q be prime. Let A be a subset of $\mathbf{Z}_q$ which is not too big in the sense that $|A| \leq q^{1-\delta}$ for some $\delta > 0$. Then $|A+A| + |A \cdot A| \geq |A|^{1+\varepsilon}$ for some $\varepsilon > 0$ depending only on δ.*

There are now some quite elementary proofs of this theorem, but I will not discuss them here (see e.g. [**Ta2008c**] for further discussion). I should note, though, that the bulk of the Bourgain-Gamburd-Sarnak work is preoccupied with establishing a suitable extension of this sum-product theorem to the case when q is not prime, in a manner which is uniform in the number of prime factors; this turns out to be a surprisingly difficult task.

Notes. This talk first appeared at `terrytao.wordpress.com/2008/11/20`, and was given as the final talk in my series of four Marker Lectures in Penn State University in November of 2008. Thanks to Luca Trevisan and Jonathan vos Post for corrections.

Bibliography

[Aa1997] J. Aaronson, An Introduction to Infinite Ergodic Theory, Mathematical Surveys and Monographs 50, American Mathematical Society, Providence RI 1997.

[AgKaSa2004] M, Agrawal, N. Kayal, N. Saxena, *PRIMES is in P*, Annals of Mathematics **160** (2004), no. 2, pp. 781793.

[AlSoVa2003] A. Alfonseca, F. Soria, A. Vargas, *A remark on maximal operators along directions in* $\mathbf{R}^2$, Math. Res. Lett. **10** (2003), no. 1, 41–49.

[Al1999] N. Alon, *Combinatorial Nullstellensatz*, Recent trends in combinatorics (Mátraháza, 1995). Combin. Probab. Comput. **8** (1999), no. 1-2, 7–29.

[AlGr1992] S. Altschuler, M. Grayson, *Shortening space curves and flow through singularities*, J. Differential Geom. **35** (1992), no. 2, 283–298.

[AmGe1973] W.O. Amrein, V. Georgescu, *On the characterization of bound states and scattering states in quantum mechanics*, Helv. Phys. Acta **46** (1973/74), 635–658.

[AuTa2008] T. Austin, T. Tao, *On the testability and repair of hereditary hypergraph properties*, preprint.

[AvGeTo2008] J. Avigad, P. Gerhardy, H. Towsner, *Local stability of ergodic averages*, preprint.

[BaHaPi2001] R. C. Baker, G. Harman, J. Pintz, *The difference between consecutive primes. II*, Proc. London Math. Soc. (3) **83** (2001), no. 3, 532–562.

[Ba2008] S. Ball, *On sets of points in a finite affine plane containing a line in every direction*, preprint.

[Ba1996] J. Barrionuevo, *A note on the Kakeya maximal operator*, Math. Res. Lett. **3** (1996), no. 1, 61–65.

[Ba2008] M. Bateman, *Kakeya Sets and Directional Maximal Operators in the Plane*, preprint.

[BaKa2008] M. Bateman, N. Katz, *Kakeya sets in Cantor directions*, Math. Res. Lett. **15** (2008), no. 1, 73–81.

[BeFo1996] F. Beleznay, M. Foreman, *The complexity of the collection of measure-distal transformations*, Ergodic Theory Dynam. Systems **16** (1996), no. 5, 929–962.

[BeCaTa2006] J. Bennett, A. Carbery, T. Tao, *On the multilinear restriction and Kakeya conjectures*, Acta Math. **196** (2006), no. 2, 261–302.

[Be1987] V. Bergelson, *Weakly mixing PET*, Ergodic Theory Dynam. Systems **7** (1987), no. 3, 337–349.

[BeHoKr2005] V. Bergelson, B. Host, B. Kra, *Multiple recurrence and nilsequences*. With an appendix by Imre Ruzsa. Invent. Math. **160** (2005), no. 2, 261–303.

[BeHoMcCPa2000] V. Bergelson, B. Host, R. McCutcheon, F. Parreau, *Aspects of uniformity in recurrence*, Colloq. Math. **84/85** (2000), 549–576.

[BeLe1996] V. Bergelson, A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), no. 3, 725–753.

[BeLe1999] V. Bergelson, A. Leibman, *Set-polynomials and polynomial extension of the Hales-Jewett theorem*, Ann. of Math. **150** (1999), no. 1, 33–75.

[BeLe2007] V. Bergelson, A. Leibman, *Distribution of values of bounded generalized polynomials*, Acta Math. **198** (2007), no. 2, 155–230.

[BeLeLe2007] V. Bergelson, A. Leibman, E. Lesigne, *Intersective polynomials and polynomial Szemeredi theorem*, preprint. `arxiv:0710.4862`

[BeTaZi2009] V. Bergelson, T. Tao, T. Ziegler, *An inverse theorem for the uniformity seminorms associated with the action of* F_p^∞, preprint.

[Be1995] I. Berkes, *On the almost sure central limit theorem and domains of attraction*, Probability Theory and Related Fields **102** (1995), 1–17.

[Be1919] A. Besicovitch, *Sur deux questions d'integrabilite des fonctions* J. Soc. Phys. Math. **2** (1919), 105-123.

[Be1928] A. Besicovitch, *On Kakeya's problem and a similar one*, Mathematische Zeitschrift **27** (1928), 312-320.

[BeBeBoMaPo2008] L. Bessiéres, G. Besson, M. Boileau, S. Maillot, J. Porti, *Weak collapsing and geometrisation of aspherical* 3*-manifolds*, preprint.

[BlHoMa2000] F. Blanchard, B. Host, A. Maass, *Topological complexity*, Ergodic Theory Dynam. Systems **20** (2000), no. 3, 641–662.

[Bl1993] A. Blass, *Ultrafilters: where topological dynamics = algebra = combinatorics*, Topology Proc. **18** (1993), 33–56.

[BlMa2008] A. Blokhuis and F. Mazzocca, *The Finite Field Kakeya Problem*, Building Bridges Between Mathematics and Computer Science, Bolyai Society Mathematical Studies, Vol. 19 Grötschel, Martin; Katona, Gyula O.H. (Eds.), Springer, 2008.

[Bo1987] E. Bombieri, *Le Grand Crible dans la Théorie Analytique des Nombres*, (Seconde Édition). Astérisque 18, Paris 1987.

[BoDa1966] E. Bombieri, H. Davenport, *Small differences between prime numbers*, Proc. Roy. Soc. Ser. A **293** 1966 1–18.

[Bo1991] J. Bourgain, *Besicovitch type maximal operators and applications to Fourier analysis*, Geom. Funct. Anal. **1** (1991), no. 2, 147–187.

[Bo1999] J. Bourgain, *On the dimension of Kakeya sets and related maximal inequalities*, Geom. Funct. Anal. **9** (1999), no. 2, 256–282.

[Bo2001] J. Bourgain, *Λ_p-sets in analysis: results, problems and related aspects*, Handbook of the geometry of Banach spaces, Vol. I, 195–232, North-Holland, Amsterdam, 2001.

[Bo2005] J. Bourgain, *New encounters in combinatorial number theory: from the Kakeya problem to cryptography*, Perspectives in analysis, 17–26, Math. Phys. Stud., 27, Springer, Berlin, 2005.

[BoGa2008] J. Bourgain, A. Gamburd, *Uniform expansion bounds for Cayley graphs of* $\mathrm{SL}_2(\mathbf{F}_p)$, Ann. of Math. **167** (2008), no. 2, 625–642.

[BoGaSa2006] J. Bourgain, A. Gamburd, P. Sarnak, *Sieving and expanders*, C. R. Math. Acad. Sci. Paris **343** (2006), no. 3, 155–159.

[BoKaTa2004] J. Bourgain, N. Katz, T. Tao, *A sum-product estimate in finite fields, and applications*, Geom. Funct. Anal. **14** (2004), no. 1, 27–57.

[BoKo2003] J. Bourgain, S. Konyagin, *Estimates for the number of sums and products and for exponential sums over subgroups in fields of prime order*, C. R. Math. Acad. Sci. Paris **337** (2003), no. 2, 75–80.

[Br1993] J. W. Bruce, *A really trivial proof of the Lucas-Lehmer test*, Amer. Math. Monthly **100** (1993), no. 4, 370–371.

[Br1915] V. Brun, *Über das Goldbachsche Gesetz und die Anzahl der Primzahlpaare*, Archiv for Math. Og Naturvid. B34 (1915).

[Br2004] R. Bryant, *Gradient Kahler Ricci Solitons*, preprint.

[BuZw] N. Burq, M. Zworski, *Control Theory and High Frequency Eigenfunctions*, slides available at `http://math.berkeley.edu/ zworski/bz1.pdf`

[CaTo2008] E. Cabezas-Rivas, P. Topping, *The canonical shrinking soliton associated to a Ricci flow*, preprint.

[Ca1958] E. Calabi, *An extension of E. Hopf's maximum principle with an application to Riemannian geometry*, Duke Math. J. **25** (1958), 45–56.

[Ca1996] H-D. Cao, *Existence of gradient Kähler-Ricci solitons*, Elliptic and parabolic methods in geometry (Minneapolis, MN, 1994), 1–16, A K Peters, Wellesley, MA, 1996.

[CaZh2006] H-D. Cao, X-P. Zhu, *A complete proof of the Poincaré and geometrization conjectures—application of the Hamilton-Perelman theory of the Ricci flow*, Asian J. Math. **10** (2006), no. 2, 165–492.

[CaFr1967] J.W.S. Cassels, A. Fröhlich (eds.), Algebraic number theory, London, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], 1986, Reprint of the 1967 original.

[Ch1970] J. Cheeger, *Finiteness theorems for Riemannian manifolds*, Amer. J. Math. **92** (1970) 61–74.

[ChGr1971] J. Cheeger, D. Gromoll, *The splitting theorem for manifolds of nonnegative Ricci curvature*, J. Differential Geometry **6** (1971/72), 119–128.

[ChGr1972] J. Cheeger, D. Gromoll, *On the structure of complete manifolds of nonnegative curvature*, Ann. of Math. **96** (1972) 413–443.

[Ch1966] J. R. Chen, *On the representation of a large even integer as the sum of a prime and the product of at most two primes*, Kexue Tongbao **17** (1966), 385-386.

[ChLuTi2006] X. Chen, P. Lu, G. Tian, *A note on uniformization of Riemann surfaces by Ricci flow*, Proc. Amer. Math. Soc. **134** (2006), no. 11, 3391–3393.

[Ch1991] B. Chow, *The Ricci flow on the* 2*-sphere*, J. Differential Geom. **33** (1991), no. 2, 325–334.

[ChCh1995] B. Chow, S-C. Chu, *A geometric interpretation of Hamilton's Harnack inequality for the Ricci flow*, Math. Res. Lett. **2** (1995), no. 6, 701–718.

[ChCh1996] B. Chow, S-C. Chu, *A geometric approach to the linear trace Harnack inequality for the Ricci flow*, Math. Res. Lett. **3** (1996), no. 4, 549–568.

[CCGGIIKLLN2008] B. Chow, S-C. Chu, D. Glickenstein, C. Guenther, J. Isenberg, T. Ivey, D. Knopf, P. Lu, F. Luo, L. Ni,The Ricci flow: techniques and applications. Part II. Analytic aspects. Mathematical Surveys and Monographs, 144. American Mathematical Society, Providence, RI, 2008.

[ChKn2004] B. Chow, D. Knopf, *The Ricci flow: an introduction*, Mathematical Surveys and Monographs, 110. American Mathematical Society, Providence, RI, 2004.

[ChLuNi2006] B. Chow, P. Lu, L. Ni, *Hamilton's Ricci flow*, Graduate Studies in Mathematics, 77. American Mathematical Society, Providence, RI; Science Press, New York, 2006.

[Ch1984] M. Christ, *Estimates for the k-plane transform*, Indiana Univ. Math. J. **33** (1984), no. 6, 891–910.

[CoMi1997] T. Colding, W. Minicozzi, *Harmonic functions on manifolds*, Ann. of Math. **146** (1997), no. 3, 725–747.

[CoMi2005] T. Colding, W. Minicozzi, *Estimates for the extinction time for the Ricci flow on certain 3-manifolds and a question of Perelman*, J. Amer. Math. Soc. **18** (2005), no. 3, 561–569.

[CoMi2007] T. Colding, W. Minicozzi, *Width and finite extinction time of Ricci flow*, preprint.

[Co1977] A. Cordoba, *The Kakeya maximal function and the spherical summation multipliers*, Amer. J. Math. **99** (1977), no. 1, 1–22.

[CoFe1977] A. Cordoba, R. Fefferman, *On the equivalence between the boundedness of certain classes of maximal and multiplier operators in Fourier analysis*, Proc. Nat. Acad. Sci. U.S.A. **74** (1977), no. 2, 423–425.

[CoSC1993] T. Coulhon, L. Saloff-Coste, *Isopérimétrie pour les groupes et les variétés*, Rev. Mat. Iberoamericana **9** (1993), no. 2, 293–314.

[Cr1936] Harald Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arithmetica **2** (1936), pp. 23-46.

[Cu1971] F. Cunningham Jr., *The Kakeya problem for simply connected and for star-shaped sets*, Amer. Math. Monthly **78** (1971) 114–129.

[Da1971] R. Davies, *Some remarks on the Kakeya problem*, Proc. Cambridge Philos. Soc. **69** (1971) 417–421.

[DaPuRo1961] M. Davis, H. Putnam, J. Robinson, *The decision problem for exponential Diophantine equations*, Ann. Math. **74** (1961), 425–436.

[DeT1983] D. DeTurck, *Deforming metrics in the direction of their Ricci tensors*, J. Differential Geom. **18** (1983), no. 1, 157–162.

[Dr1983] S.W. Drury, *L^p estimates for the X-ray transform*, Illinois J. Math. **27** (1983), no. 1, 125–129.

[Dv2008] Z. Dvir, *On the size of Kakeya sets in finite fields*, preprint.

[EcHu1991] K. Ecker, G. Huisken, *Interior estimates for hypersurfaces moving by mean curvature*, Invent. Math. **105** (1991), no. 3, 547–569.

[Ei2006] M. Einsiedler, *Ratner's theorem on* $\mathrm{SL}(2,\mathbf{R})$*-invariant measures*, Jahresber. Deutsch. Math.-Verein. **108** (2006), no. 3, 143–164.

[EiMaVe2007] M. Einsiedler, G. Margulis, A. Venkatesh, *Effective equidistribution for closed orbits of semisimple groups on homogeneous spaces*, preprint.

[El1958] R. Ellis, *Distal transformation groups*, Pacific J. Math. **8** (1958), 401–405.

[En1978] V. Enss, *Asymptotic completeness for quantum mechanical potential scattering. I. Short range potentials*, Comm. Math. Phys. **61** (1978), no. 3, 285–291.

[Er1940] P. Erdős, *The difference of consecutive primes*, Duke Math. J. **6** (1940). 438–441.

[ErTu1936] P. Erdős, P. Turán, *On some sequences of integers*, J. London Math. Soc. **11** (1936), 261–264.

[EsMoOh2005] A. Eskin, S. Mozes, H. Oh, *On uniform exponential growth for linear groups*, Invent. Math. **160** (2005), no. 1, 1–30.

[Ev1998] L. C. Evans, Partial differential equations. Graduate Studies in Mathematics, 19. American Mathematical Society, Providence, RI, 1998.

[Fa2000] I. Farah, *Approximate homomorphisms. II. Group homomorphisms*, Combinatorica **20** (2000), no. 1, 47–60.

[FiMa2005] D. Fisher, G. Margulis, *Almost isometric actions, property (T), and local rigidity*, Invent. Math. **162** (2005), no. 1, 19–80.

[Fo1970] J. Folkman, *Graphs with monochromatic complete subgraphs in every edge coloring*, SIAM J. Appl. Math. **18** (1970), 115-124.

[Fr1973] G. Freiman, Foundations of a structural theory of set addition. Translated from the Russian. Translations of Mathematical Monographs, Vol 37. American Mathematical Society, Providence, R. I., 1973.

[Fu1961] H. Furstenberg, *Strict ergodicity and transformation of the torus*, Amer. J. Math. **83** (1961) 573–601.

[Fu1963] H. Furstenberg, *The structure of distal flows*, Amer. J. Math. **85** (1963) 477–515.

[Fu1977] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.

[Fu1981] H. Furstenberg, *Recurrence in Ergodic theory and Combinatorial Number Theory*, Princeton University Press, Princeton NJ 1981.

[FuKa1979] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291 (1979).

[FuKa1985] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for IP-systems and combinatorial theory*, J. Analyse Math. **45** (1985), 117–168.

[FuKa1991] H. Furstenberg, Y. Katznelson, A density version of the Hales-Jewett theorem, *J. d'Analyse Math.* **57** (1991), 64–119.

[FuWe1978] H. Furstenberg, B. Weiss, *Topological dynamics and combinatorial number theory*, J. Analyse Math. 34 (1978), 61–85 (1979).

[FuWe1996] H. Furstenberg, B. Weiss, *A mean ergodic theorem for* $(1/N)\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$, Convergence in ergodic theory and probability (Columbus, OH, 1993), 193–227, Ohio State Univ. Math. Res. Inst. Publ., 5, de Gruyter, Berlin, 1996.

[GaHa1986] M. Gage, R. Hamilton, *The heat equation shrinking convex plane curves*, J. Differential Geom. **23** (1986), no. 1, 69–96.

[Ga1976] P. X. Gallagher, *On the distribution of primes in short intervals*, Mathematika **23** (1976), no. 1, 4–9.

[Ga2002] A. Gamburd, *On the spectral gap for infinite index "congruence" subgroups of* $\mathrm{SL}_2(\mathbf{Z})$, Israel J. Math. **127** (2002), 157–200.

[GeLe1993] P. Gérard, E. Leichtnam, *Ergodic properties of eigenfunctions for the Dirichlet problem*, Duke Math. J. **71** (1993), no. 2, 559–607.

[Ge2008] P. Gerhardy, *Proof Mining in Topological Dynamics*, Notre Dame J. Formal Logic **49**, (2008), 431–446.

[Gi1987] J.-Y. Girard, *Proof theory and logical complexity, Vol. I*, Bibliopolis, Naples, 1987.

[Gl2000] E. Glasner, *Structure theory as a tool in topological dynamics. Descriptive set theory and dynamical systems*, (Marseille-Luminy, 1996), 173–209, London Math. Soc. Lecture Note Ser., 277, Cambridge Univ. Press, Cambridge, 2000.

[Gl2003] E. Glasner, *Ergodic theory via joinings*, Mathematical Surveys and Monographs, 101. American Mathematical Society, Providence, RI, 2003.

[GoGrPiYi2006] D. A. Goldston, S.W. Graham, J. Pintz, C.Y. Yıldırım, *Small gaps between products of two primes*, preprint.

[GoMo1987] D. Goldston, H. Montgomery, *Pair correlation of zeros and primes in short intervals*, Analytic number theory and Diophantine problems (Stillwater, OK, 1984), 183–203, Progr. Math., 70, Birkhäuser Boston, Boston, MA, 1987.

[GoPoYi2005] D. Goldston, J. Pintz, C. Yıldırım, *The Path to Recent Progress on Small Gaps Between Primes*, preprint.

[GoPoYi2005a] D. Goldston, J. Pintz, C. Yıldırım, *Primes in Tuples I*, preprint.

[GoPoYi2007] D. Goldston, J. Pintz, C. Yıldırım, *Primes in Tuples I*, preprint.

[Go1998] W. T. Gowers, *A new proof of Szemerédi's theorem for progressions of length four*, GAFA **8** (1998), no. 3, 529–551.

[Go2001] W. T. Gowers, *A new proof of Szemerédi's Theorem,* Geom. Funct. Anal. **11** (2001), no. 3, 465–588.

[Go2008] T. Gowers, *Decompositions, approximate structure, transference, and the Hahn-Banach theorem*, preprint.

[Gr1995] A. Granville, *Harald Cramér and the distribution of prime numbers*, Harald Cramér Symposium (Stockholm, 1993). Scand. Actuar. J. 1995, no. 1, 12–28.

[Gr2005] B. Green, *Finite field models in additive combinatorics*, Surveys in Combinatorics 2005, London Math. Soc. Lecture Notes 327, 1–27.

[Gr2006] B. Green, *Three topics in additive prime number theory*, preprint.

[GrTa2006] B. Green, T. Tao, *Restriction theory of the Selberg Sieve, with applications*, Journal de Théorie des Nombres de Bordeaux 18 (2006), 137–172.

[GrTa2007] B. Green, T. Tao, *The distribution of polynomials over finite fields, with applications to the Gowers norms*, preprint.

[GrTa2008] B. Green, T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Annals Math. **167** (2008), 481-547

[GrTa2009] B. Green, T. Tao, *Linear equations in primes*, to appear, Annals of Math.

[GrTa2009a] B. Green, T. Tao, *New bounds for Szemeredi's Theorem, I: Progressions of length 4 in finite field geometries*, preprint.

[GrTa2009b] B. Green, T. Tao, *New bounds for Szemeredi's Theorem, II: A new bound for* $r_4(N)$, preprint.

[GrTa2009c] B. Green, T. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, preprint.

[GrTa2009d] B. Green, T. Tao, *An inverse theorem for the Gowers* $U^3(G)$ *norm*, preprint.

[GrTa2009e] B. Green, T. Tao, *Quadratic uniformity of the Möbius function*, preprint.

[GrTa2009f] B. Green, T. Tao, *The Möbius function is asymptotically orthogonal to nilsequences*, preprint.

[Gr1961] L. W. Green, *Spectra of nilflows*, Bull. Amer. Math. Soc. **67** (1961) 414–415.

[GrLeRo1972] R.L. Graham, K. Leeb, B.L. Rothschild, *Ramsey's theorem for a class of categories*, Advances in Math. **8** (1972), 417–433.

[GrRoSp1980] R. Graham, B. Rothschild, J.H. Spencer, Ramsey Theory, *John Wiley and Sons*, NY (1980).

[GrSo2007] A. Granville, K. Soundararajan, *An uncertainty principle for arithmetic sequences*, Ann. of Math. (2) **165** (2007), no. 2, 593–635.

[Gr1981] M. Gromov, *Groups of polynomial growth and expanding maps*, Inst. Hautes Études Sci. Publ. Math. No. **53** (1981), 53–73.

[Gr2003] M. Gromov, *Isoperimetry of waists and concentration of maps*, Geom. Funct. Anal. **13** (2003), no. 1, 178–215.

[Gr1975] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), no. 4, 1061–1083.

[Gr2006] L. Gross, *Hypercontractivity, logarithmic Sobolev inequalities, and applications: a survey of surveys*, Diffusion, quantum theory, and radically elementary mathematics, 45–73, Math. Notes, 47, Princeton Univ. Press, Princeton, NJ, 2006.

[GuLe1973] R. Gulliver, F. Lesley, *On boundary branch points of minimizing surfaces*, Arch. Rational Mech. Anal. **52** (1973), 20–25.

[Gu2008] L. Guth, *The endpoint case of the Bennett-Carbery-Tao multilinear Kakeya conjecture*, preprint.

[Gu1988] R. Guy, *The Strong Law of Small Numbers*, The American Mathematical Monthly, **95** (1988), 697–712.

[HaJe1963] A.W. Hales, R.I. Jewett, *Regularity and positional games*, Trans. Amer. Math. Soc. **106** (1963), 222–229.

[Ha1982] R. Hamilton, *Three-manifolds with positive Ricci curvature*, J. Differential Geom. **17** (1982), no. 2, 255–306.

[Ha1986] R. Hamilton, *Four-manifolds with positive curvature operator*, J. Differential Geom. **24** (1986), no. 2, 153–179.

[Ha1988] R. Hamilton, *The Ricci flow on surfaces*, Mathematics and general relativity (Santa Cruz, CA, 1986), 237–262, Contemp. Math., 71, Amer. Math. Soc., Providence, RI, 1988.

[Ha1993] R. Hamilton, *The Harnack estimate for the Ricci flow*, J. Differential Geom. **37** (1993), no. 1, 225–243.

[Ha1993b] R. Hamilton, *The formation of singularities in the Ricci flow*, Surveys in differential geometry, Vol. II (Cambridge, MA, 1993), 7–136, Int. Press, Cambridge, MA, 1995.

[Ha1995] R. Hamilton, *A compactness property for solutions of the Ricci flow*, Amer. J. Math. **117** (1995), no. 3, 545–572.

[Ha1997] R. Hamilton, *Four-manifolds with positive isotropic curvature*, Comm. Anal. Geom. **5** (1997), no. 1, 1–92.

[Ha1999] R. Hamilton, *Non-singular solutions of the Ricci flow on three-manifolds*, Comm. Anal. Geom. **7** (1999), no. 4, 695–729.

[HaSi1985] R. Hardt, L. Simon, *Area minimizing hypersurfaces with isolated singularities*, J. Reine Angew. Math. **362** (1985), 102–129.

[HaLi1923] G. H., Hardy, J. E. Littlewood, *Some Problems of 'Partitio Numerorum.' III. On the Expression of a Number as a Sum of Primes.*, Acta Math. **44** (1923), 1–70.

[Ha1973] L. D. Harmon, *The Recognition of Faces*, Scientific American **229** (1973), 71–82.

[Ha2008] A. Hassell, Ergodic billiards that are not quantum unique ergodic, preprint. With an appendix by Andrew Hassell and Luc Hillairet.

[He2008] H. A. Helfgott, *Growth and generation in* $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, Ann. of Math. **167** (2008), no. 2, 601–623.

[He1991] E. Heller, *Wavepacket dynamics and quantum chaology*, Chaos et physique quantique (Les Houches, 1989), 547–664, North-Holland, Amsterdam, 1991.

[Hi1951] G. Higman, *A finitely generated infinite simple group*, J. London Math. Soc. **26**, (1951). 61–64.

[Hi1969] S. Hildebrandt, *Boundary behavior of minimal surfaces*, Arch. Rational Mech. Anal. **35** (1969) 47–82.

[Hi1974] N. Hindman, *Finite sums from sequences within cells of a partition of* N, J. Combin. Thy. Ser. A **17** (1974), 1–11.

[Hi] J. Hirschfeld, *The nonstandard treatment of Hilbert's fifth problem*, Trans. Amer. Math. Soc. **321** (1990), no. 1, 379–400.

[HoKr2005] B. Host, B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488.

[Il1997] K. Ilinski, *The physics of finance*, Proceedings of Budapest's conference on Econophysics (July 1997)

[Iv1993] T. Ivey, *Ricci solitons on compact three-manifolds*, Differential Geom. Appl. **3** (1993), no. 4, 301–307.

[Iw1978] H. Iwaniec, *Almost-primes represented by quadratic polynomials*, Invent. Math. **47** (1978), no. 2, 171–188.

[Jo1991] J. Jost, *Two-dimensional geometric variational problems*, Pure and Applied Mathematics (New York). A Wiley-Interscience Publication. John Wiley & Sons, Ltd., Chichester, 1991.

[Ka1999] N. H. Katz, *Remarks on maximal operators over arbitrary sets of directions*, Bull. London Math. Soc. **31** (1999), no. 6, 700–710.

[Ka2005] N. H. Katz, *On arithmetic combinatorics and finite groups*, Illinois J. Math. **49** (2005), no. 1, 33–43.

[KaLaTa2000] N. Katz, I. Łaba, T. Tao, *An improved bound on the Minkowski dimension of Besicovitch sets in* R^3, Ann. of Math. (2) **152** (2000), no. 2, 383–446.

[KaTa1999] N. Katz, T. Tao, *Bounds on arithmetic projections, and applications to the Kakeya conjecture*, Math. Res. Lett. **6** (1999), no. 5-6, 625–630.

[KaTa2002] N. Katz, T. Tao, *Recent progress on the Kakeya conjecture*, Proceedings of the 6th International Conference on Harmonic Analysis and Partial Differential Equations (El Escorial, 2000). Publ. Mat. 2002, Vol. Extra, 161–179.

[KaTa200b] N. Katz, T. Tao, *New bounds for Kakeya problems*, Dedicated to the memory of Thomas H. Wolff. J. Anal. Math. **87** (2002), 231–263.

[Ka1981] J. Kazdan, *Another proof of Bianchi's identity in Riemannian geometry*, Proc. Amer. Math. Soc. **81** (1981), no. 2, 341–342.

[KaWa1974] J. Kazdan, F.W. Warner, *Curvature functions for compact 2-manifolds*, Ann. of Math. (2) **99** (1974), 14–47.

[Ke1995] M. Keane, *The essence of large numbers*, Algorithms, Fractals, and Dynamics (Okayama/Kyoto, 1992), 125-129, Plenum, New York, 1995.

[Ke1999] U. Keich, *On L^p bounds for Kakeya maximal functions and the Minkowski dimension in R^2*, Bull. London Math. Soc. **31** (1999), no. 2, 213–221.

[KeRo1969] H. Keynes, J. Robertson, *Eigenvalue theorems in topological transformation groups*, Trans. Amer. Math. Soc. **139** (1969), 359–369.

[KiVi2008] R. Killip, M. Visan, *Nonlinear Schrodinger Equations at Critical Regularity*, available at `http://www.math.uchicago.edu/ mvisan/ClayLectureNotes.pdf`

[KlRo2007] S. Klainerman, I. Rodnianski, *A Kirchoff-Sobolev parametrix for the wave equation and applications*, J. Hyperbolic Differ. Equ. **4** (2007), no. 3, 401–433.

[Kl2007] B. Kleiner, *A new proof of Gromov's theorem on groups of polynomial growth*, preprint.

[KlLo2006] B. Kleiner, J. Lott, *Notes on Perelman's papers*, preprint.

[Kn1929] H. Kneser, *Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten*, Jahresbericht der Deut. Math. Verein. **38** (1929), 248–260.

[KoSc1997] N. Korevaar, R. Schoen, *Global existence theorems for harmonic maps to non-locally compact spaces*, Comm. Anal. Geom. **5** (1997), no. 2, 333–387.

[Ko2006] D. Kotschick, *Monopole classes and Perelman's invariant of four-manifolds*, preprint.

[Kow2006] E. Kowalski, *Écarts entre nombres premiers succesifs (d'après Goldston, Pintz, Yıldırım)*, Séminaire Bourbaki, 58éme année, 2005–2006, no. 959.

[Kr2006] B. Kra, *From combinatorics to ergodic theory and back again*, International Congress of Mathematicians. Vol. III, 57–76, Eur. Math. Soc., Zürich, 2006.

[La2008] I. Łaba, *From harmonic analysis to arithmetic combinatorics*, Bull. Amer. Math. Soc. (N.S.) **45** (2008), no. 1, 77–115.

[LaTa2001] I. Łaba, T. Tao, *An improved bound for the Minkowski dimension of Besicovitch sets in medium dimension*, Geom. Funct. Anal. **11** (2001), no. 4, 773–806.

[Le1998] A. Leibman, *Polynomial sequences in groups*, J. Algebra **201** (1998), no. 1, 189–206.

[Le2002] A. Leibman, *Polynomial mappings of groups*, Israel J. Math. **129** (2002), 29–60.

[Le2005] A. Leibman, *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold*, Ergodic Theory Dynam. Systems **25** (2005), no. 1, 201–213.

[Le2005b] A. Leibman, *Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold*, Ergodic Theory Dynam. Systems **25** (2005), no. 1, 215–225.

[LiYa1986] P. Li, S-T. Yau, *On the parabolic kernel of the Schrödinger operator*, Acta Math. **156** (1986), no. 3-4, 153–201.

[LiWa2002] M-C. Liu, T. Wang, *On the Vinogradov bound in the three primes Goldbach conjecture*, Acta Arith. **105** (2002), no. 2, 133–175.

[Lo2008] J. Lott, *Optimal transport and Perelman's reduced volume*, preprint.

[LoMeSa2008] S. Lovett, R. Meshulam, A. Samorodnitsky, *Inverse conjecture for the Gowers norm is false*, STOC'08.

[Ma1985] H. Maier, *Primes in short intervals*, Michigan Math. J. **32** (1985), no. 2, 221–225.

[Ma1951] A. I. Mal'cev, On a class of homogeneous spaces. (Russian) Izvestiya Akad. Nauk. SSSR. Ser. Mat. **13**, (1949). 9–32. Translated in: Amer. Math. Soc. Translation 1951, (1951). no. 39, 33 pp.

[Ma1970] Y. Matiyasevich, *Enumerable sets are Diophantine*, (in Russian) Doklady Akademii Nauk SSSR, **191** (1970), 279–282. English translation: Soviet Mathematics. Doklady, **11** (1970), 354–358.

[Ma2003] J. Matousek, *Using the Borsuk-Ulam theorem*, Lectures on topological methods in combinatorics and geometry. Written in cooperation with Anders Björner and Günter M. Ziegler. Universitext. Springer-Verlag, Berlin, 2003.

[McM1978] D. McMahon, *Relativized weak disjointness and relatively invariant measures*, Trans. Amer. Math. Soc. **236** (1978), 225–237.

[MeYa1980] W. Meeks, S-T. Yau, *Topology of three-dimensional manifolds and the embedding problems in minimal surface theory*, Ann. of Math. (2) **112** (1980), no. 3, 441–484.

[Me2003] A. Melas, *The best constant for the centered Hardy-Littlewood maximal inequality*, Ann. of Math. (2) **157** (2003), no. 2, 647–688.

[Mi1962] J. Milnor, *A unique factorization theorem for* 3*-manifolds*, Amer. J. Math. **84** (1962), 1–7.

[Mi1968] J. Milnor, *A note on curvature and fundamental group*, J. Diff. Geom. **2** (1968), 1–7.

[Mi2003] J. Milnor, *Towards the Poincaré conjecture and the classification of 3-manifolds*, Notices Amer. Math. Soc. **50** (2003), no. 10, 1226–1233.

[Mi2006] J. Milnor, *The Poincaré conjecture. The millennium prize problems*, 71–83, Clay Math. Inst., Cambridge, MA, 2006.

[MoTa2004] G. Mockenhaupt, T. Tao, *Restriction and Kakeya phenomena for finite fields*, Duke Math. J. **121** (2004), no. 1, 35–74.

[Mo1995] N. Mok, *Harmonic forms with values in locally constant Hilbert bundles*, Proceedings of the Conference in Honor of Jean-Pierre Kahane (Orsay, 1993). J. Fourier Anal. Appl. 1995, Special Issue, 433–453.

[Mo1952] E. Moise, *Affine structures in* 3*-manifolds. V. The triangulation theorem and Hauptvermutung*, Ann. of Math. (2) **56**, (1952). 96–114.

[MoZi1955] D. Montgomery, L. Zippin, Topological transformation groups. Reprint of the 1955 original. Robert E. Krieger Publishing Co., Huntington, N.Y., 1974.

[Mo2007] J. Morgan, *The Poincaré conjecture*, International Congress of Mathematicians. Vol. I, 713–736, Eur. Math. Soc., Zürich, 2007

[MoTi2007] J. Morgan, G. Tian, *Ricci flow and the Poincaré conjecture*, Clay Mathematics Monographs, 3. American Mathematical Society, Providence, RI; Clay Mathematics Institute, Cambridge, MA, 2007.

[MoTi2008] J. Morgan, G. Tian, *Completion of the Proof of the Geometrization Conjecture*, preprint.

[Mo1948] C. Morrey, *The problem of Plateau on a Riemannian manifold*, Ann. of Math. (2) **49**, (1948). 807–851.

[Mo1964] J. Moser, *A Harnack inequality for parabolic differential equations*, Comm. Pure Appl. Math. **17** (1964) 101–134

[Mu2006] R. Müller, Differential Harnack inequalities and the Ricci flow. EMS Series of Lectures in Mathematics. European Mathematical Society (EMS), Zürich, 2006.

[Mu1960] J. Munkres, *Obstructions to the smoothing of piecewise-differentiable homeomorphisms*, Ann. of Math. (2) **72** (1960) 521–554.

[Na2007] A. Naber, *Noncompact Shrinking 4-Solitons with Nonnegative Curvature*, preprint.

[NaStWa1978] A. Nagel, E. Stein, S. Wainger, *Differentiation in lacunary directions*, Proc. Nat. Acad. Sci. U.S.A. **75** (1978), no. 3, 1060–1062.

[Ne1954] B. H. Neumann, *An essay on free products of groups with amalgamations*, Philos. Trans. Roy. Soc. London. Ser. A. 246, (1954). 503–554.

[Ni2004] L. Ni, *The entropy formula for linear heat equation*, J. Geom. Anal. **14** (2004), no. 1, 87–100.

[NiWa2007] L. Ni, N. Wollach, *On a classification of the gradient shrinking solitons*, preprint.

[Pa1969] W. Parry, *Ergodic properties of affine transformations and flows on nilmanifolds*, Amer. J. Math. **91** (1969) 757–771.

[Pe1994] G. Perelman, *Proof of the soul conjecture of Cheeger and Gromoll*, J. Differential Geom. **40** (1994), 209–212.

[Pe2002] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, preprint, `math.DG/0211159`.

[Pe2003] G. Perelman, *Ricci flow with surgery on three-manifolds*, preprint, `math.DG/0303109`.

[Pe2003b] G. Perelman, *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, preprint, `math.DG/0307245`.

[Pe2006] P. Petersen, Riemannian geometry. Second edition. Graduate Texts in Mathematics, 171. Springer, New York, 2006.

[PeWy2007] P. Petersen, W. Wylie, *On the classification of gradient Ricci solitons*, preprint.

[Ra1937] R. A. Rankin, *The difference between consecutive primes*, J. Lond. Math. Soc. **13** (1938), 242–247.

[Ra1962] R. A. Rankin, *The difference between consecutive prime numbers. V*, Proc. Edinburgh Math. Soc. **13** (1962/1963) 331–332.

[ReTrTuVa2008] O. Reingold, L. Trevisan, M. Tulsiani, S. Vadhan, *New Proofs of the Green-Tao-Ziegler Dense Model Theorem: An Exposition*, preprint.

[Ri1859] B. Riemann, *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsberichte der Berliner Akademie, 1859.

[Ro1953] K.F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 245–252.

[Ru1969] D. Ruelle, *A remark on bound states in potential-scattering theory*, Nuovo Cimento A **61** 1969 655–662.

[Ru1994] I. Z. Ruzsa, *Generalized arithmetical progressions and sumsets*, Acta Math. Hungar. **65** (1994), no. 4, 379–388.

[SaUh1981] J. Sacks, K. Uhlenbeck, *The existence of minimal immersions of 2-spheres*, Ann. of Math. (2) **113** (1981), no. 1, 1–24.

[SaSu2008] S. Saraf, M. Sudan, *Improved lower bound on the size of Kakeya sets over finite fields*, preprint.

[Sa1998] Y. Saouter, *Checking the odd Goldbach conjecture up to* 10^{20}, Math. Comp. **67** (1998), no. 222, 863–866.

[SaXu1991] P. Sarnak, X. Xue, *Bounds for multiplicities of automorphic representations*, Duke Math. J. **64** (1991), no. 1, 207–227.

[Sc1998] W. Schlag, *A geometric inequality with applications to the Kakeya problem in three dimensions*, Geom. Funct. Anal. **8** (1998), no. 3, 606–625.

[Sc1962] I. Schoenberg, *On the Besicovitch-Perron solution of the Kakeya problem*, 1962 Studies in mathematical analysis and related topics pp. 359–363 Stanford Univ. Press, Stanford, Calif.

[Sc1916] I. Schur, *Über die Kongruenz* $x^m + y^m = z^m (mod p)$, Jber. Deutsch. Math.-Verein. **25** (1916), 114–116.

[Sc1980] J. T. Schwartz, *Fast probabilistic algorithms for verification of polynomial identities*, J. ACM **27** (1980), 701–717.

[Se1965] A. Selberg, *On the estimation of Fourier coefficients of modular forms*, 1965 Proc. Sympos. Pure Math., Vol. VIII pp. 1–15 Amer. Math. Soc., Providence, R.I.

[Sh1996] N. Shah, *Invariant measures and orbit closures on homogeneous spaces for actions of subgroups generated by unipotent elements*, Lie groups and ergodic theory (Mumbai, 1996), 229–271, Tata Inst. Fund. Res. Stud. Math., 14, Tata Inst. Fund. Res., Bombay, 1998.

[Sh1997] Y. Shalom, *Expanding graphs and invariant means*, Combinatorica **17** (1997), no. 4, 555–575.

[Sh1998] Y. Shalom, *The growth of linear groups*, J. Algebra **199** (1998), no. 1, 169–174.

[Sh1989] W-X. Shi, *Deforming the metric on complete Riemannian manifolds*, J. Differential Geom. **30** (1989), no. 1, 223–301.

[Sm1961] S. Smale, *Generalized Poincaré's conjecture in dimensions greater than four*, Ann. of Math. (2) **74** (1961) 391–406.

[So2006] K. Soundararajan, *Small gaps between prime numbers: The work of Goldston-Pintz-Yıldırım*, preprint.

[So2007] K. Soundararajan, *The distribution of prime numbers*, Equidistribution in Number Theory, An Introduction, NATO Science Series II: Mathematics, Physics and Chemistry, Springer Netherlands, 2007.

[St1961] E. M. Stein, *On limits of sequences of operators*, Ann. of Math. (2) **74** 1961 140–170.

[St1970] E.M. Stein, Topics in harmonic analysis related to the Littlewood-Paley theory. Annals of Mathematics Studies, No. 63 Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo 1970

[StSt1983] E.M. Stein, J.-O. Strömberg, *Behavior of maximal functions in R^n for large n*, Ark. Mat. **21** (1983), no. 2, 259–269.

[St1969] S.A. Stepanov, *The number of points of a hyperelliptic curve over a finite prime field*, Izv. Akad. Nauk SSSR Ser. Mat. **33** (1969) 1171–1181.

[Sz1969] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104.

[Sz1975] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299–345.

[Ta] T. Tao, *An ergodic transference theorem*, unpublished. Available at `http://www.math.ucla.edu/~tao/preprints/Expository/limiting.dvi`

[Ta2] T. Tao, *Perelman's proof of the Poincaré conjecture: a nonlinear PDE perspective*, unpublished. Available at `http://arxiv.org/abs/math/0610903`

[Ta3] T. Tao, *The dyadic pigeonhole principle in harmonic analysis*, unpublished. Available at `http://www.math.ucla.edu/ tao/preprints/Expository/pigeonhole.dvi`

[Ta2001] T. Tao, *From rotating needles to stability of waves: emerging connections between combinatorics, analysis, and PDE*, Notices Amer. Math. Soc. **48** (2001), no. 3, 294–303

[Ta2005] T. Tao, *A new bound for finite field Besicovitch sets in four dimensions*, Pacific J. Math. **222** (2005), no. 2, 337–363.

[Ta2006] T. Tao, *A quantitative ergodic theory proof of Szemerédi's theorem*, Electron. J. Combin. **13** (2006) No. 99, 1–49.

[Ta2006b] T. Tao, *The Gaussian primes contain arbitrarily shaped constellations*, J. dAnalyse Mathematique **99** (2006), 109–176.

[Ta2007] T. Tao, *The ergodic and combinatorial approaches to Szemerédi's theorem*, Centre de Recerches Mathématiques, CRM Proceedings and Lecture Notes Vol. 43 (2007), 145–193.

[Ta2007b] T. Tao, *Structure and randomness in combinatorics*, Proceedings of the 48th annual symposium on Foundations of Computer Science (FOCS) 2007, 3–18

[Ta2007c] T. Tao, *A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma*, J. d'Analyse Mathematique **103** (2007), 1–45.

[Ta2008] T. Tao, *Norm convergence of multiple ergodic averages for commuting transformations*, Ergodic Theory and Dynamical Systems **28** (2008), 657–688.

[Ta2008b] T. Tao, *Global regularity of wave maps IV. Absence of stationary or self-similar solutions in the energy class*, preprint.

[Ta2008c] T. Tao, *Global regularity of wave maps V. Large data local well-posedness in the energy class*, preprint.

[Ta2008b] T. Tao, Structure and Randomness: pages from year one of a mathematical blog, American Mathematical Society, Providence RI, 2008.

[Ta2008c] T. Tao, *The sum-product phenomenon in arbitrary rings*, preprint.

[TaVaVe1998] T. Tao, A. Vargas, L. Vega, *A bilinear approach to the restriction and Kakeya conjectures*, J. Amer. Math. Soc. **11** (1998), no. 4, 967–1000.

[TaVu2006] T. Tao, V. Vu, Additive combinatorics, Cambridge University Press, Cambridge, 2006.

[TaVu2008] T. Tao, V. Vu, *Random matrices: Universality of ESDs and the circular law*, preprint.

[TaZi2008] T. Tao, T. Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Math. **201** (2008), 213305.

[Ta] A. Tarski, *Une contribution a la théorie de la mesure*, Fund. Math. 15 (1930), 42-50.

[Ti1972] J. Tits, *Free subgroups in linear groups*, J. Algebra **20** (1972) 250–270.

[To1959] V. A. Toponogov, *Riemann spaces with curvature bounded below*, Uspehi Mat. Nauk **14** (1959), 87–130.

[To1964] V. A. Toponogov, *The metric structure of Riemannian spaces of non-negative curvature containing straight lines* (Russian) Sibirsk. Mat. Ž. **5** (1964) 1358–1369.

[Ul1929] S. Ulam, *Concerning functions of sets*, Fund. Math. **14** (1929), 231-233.

[vA1942] H. van Alphen, *Generalization of a theorem of Besicovitsch*, Mathematica, Zutphen. B. **10**, (1942). 144–157.

[vdDrWi1984] L. van den Dries, A. J. Wilkie, *Gromov's theorem on groups of polynomial growth and elementary logic*, J. Algebra **89** (1984), no. 2, 349–374.

[vdW1927] B.L. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw. Arch. Wisk. **15** (1927), 212-216.

[vKa1933] E. R. van Kampen, *On the connection between the fundamental groups of some related spaces*, American Journal of Mathematics, **55** (1933), pp. 261-267.

[Va1959] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. **39** (1959), 358–360.

[Vi1937] I. M. Vinogradov, *The Method of Trigonometrical Sums in the Theory of Numbers* (Russian). Trav. Inst. Math. Stekloff **10** (1937).

[Wa2000] M. Walters, *Combinatorial proofs of the polynomial van der Waerden theorem and the polynomial Hales-Jewett theorem*, J. London Math. Soc. (2) **61** (2000), no. 1, 1–12.

[Wh1961] J.H.C. Whitehead, *Manifolds with transverse fields in euclidean space*, Ann. of Math. (2) **73** (1961) 154–212.

[Wo1968] J. Wolf, *Growth of finitely generated solvable groups and curvature of Riemannian manifolds*, J. Diff. Geom. **2** (1968), 421–446.

[Wo1995] T. Wolff, *An improved bound for Kakeya type maximal functions*, Rev. Mat. Iberoamericana **11** (1995), no. 3, 651–674.

[Wo1999] T. Wolff, *Recent work connected with the Kakeya problem*, Prospects in mathematics (Princeton, NJ, 1996), 129–162, Amer. Math. Soc., Providence, RI, 1999.

[Ye] R. Ye, lecture notes, available at `http://www.math.ucsb.edu/ỹer/ricciflow.html`.

[Ye2008] R. Ye, *On the l-function and the reduced volume of Perelman. I*, Trans. Amer. Math. Soc. **360** (2008), no. 1, 507–531.

[Ye2007] S. Yekhanin, *Towards 3-query locally decodable codes of subexponential length*, Proceedings of the thirty-ninth annual ACM symposium on Theory of computing (STOC), 2007, 266 – 274.

[Ze2004] S. Zelditch, *Note on quantum unique ergodicity*, Proc. Amer. Math. Soc. **132** (2004), 1869–1872.

[Zi2007] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, J. Amer. Math. Soc. **20** (2007), no. 1, 53–97.

[Zh2007] Q. Zhang, *A uniform Sobolev inequality under Ricci flow*, preprint.

[Zh2008] Q. Zhang, *Heat kernel bounds, ancient κ solutions and the Poincaré conjecture*, preprint.

[Zi1976] R. Zimmer, *Extensions of ergodic group actions*, Illinois J. Math. **20** (1976), no. 3, 373–409.