Lecture notes in Ergodic Theory & Dynamical Systems

Lewis Bowen[*]
University of Texas at Austin

April 14, 2015

Contents

1 The $\times 2$ map 4
1.1 Coding 4

2 Rotations 5
2.1 Group Translations 5
2.2 Group Endomorphisms 5
2.3 Equidistribution 5

3 Van der Waerden's Theorem 6

4 Quadratic maps (following Devaney) 6

5 Smale's Horseshoe 7

6 Symbolic dynamics (following Katok-Hasselblatt) 9
6.1 Topological Markov chains 9
6.2 Current research in symbolic dynamics 10

7 Toral flows 11

8 Toral automorphisms 11

9 Topological entropy 13
9.1 Isometries 15
9.2 Topological Markov chains 15
9.3 Toral automorphisms 16

---

[*] supported in part by NSF grant DMS-0968762, NSF CAREER Award DMS-0954606 and BSF grant 2008274

10 Equicontinuous and distal flows **16**
10.1 Distal maps 17

11 Multiple recurrence and Ramsey theory **18**
11.1 Further results 20

12 Examples of measure-preserving actions and basic notions **20**

13 Ergodic Theorems **22**

14 Applications of the ergodic theorem **22**

15 Invariant measures **23**

16 Unique ergodicity **24**
16.1 Topological models 25

17 Pure point spectrum **25**

18 Weak mixing **26**
18.1 Notable results related to weak mixing 29

19 Spectral theory **29**
19.1 The spectral theorem for normal operators 30
19.2 Examples 30
19.3 Notable results/questions on spectral theory 31

20 Ergodic decomposition theorem **31**

21 Entropy theory (rough) **32**

22 Entropy Theory **32**

23 Relative entropy **33**

24 Kolmogorov's Theorem **36**
24.1 Markov Chains 38

25 The Shannon-McMillan-Breiman Theorem **39**

26 The variational principle **41**

27 Ornstein theory **44**
27.1 Joinings... (rough draft) 44
27.2 Rohklin's Lemma 45
27.3 Sinai's Theorem 45

27.3.1 θ . . . 47
27.3.2 ϵ_2 . . . 47
27.3.3 α and $\mathbf{q}$ . . . 47
27.3.4 n . . . 47
27.3.5 γ_2 . . . 48
27.3.6 F . . . 48
27.4 The partition upgrade procedure . . . 49
27.4.1 δ_1 . . . 49
27.4.2 θ . . . 50
27.4.3 ϵ_2 . . . 50
27.4.4 α . . . 51
27.4.5 n . . . 51
27.4.6 F . . . 52

28 Sofic entropy theory **52**

29 Kingman's subadditive ergodic theorem **52**

30 Oseledec's Multiplicative Ergodic Theorem **55**
30.1 Example . . . 55
30.2 A bit of linear algebra . . . 56
30.3 Multilinear algebra . . . 56
30.4 Sketch of the proof of the multiplicative ergodic theorem . . . 57
30.5 Entropy and Lyapunov exponents . . . 57

31 Hyperbolic dynamics **60**
31.1 ϵ-orbits . . . 61
31.2 Invariant cones . . . 65
31.3 Local stable/unstable manifolds . . . 67
31.4 Inclination lemma . . . 72
31.5 Horseshoes . . . 72
31.6 Local product structure . . . 73
31.7 Axiom A and structural stability . . . 74

32 Interval Exchange Transformations (IETs) **75**
32.1 Rough notes . . . 75
32.2 introduction . . . 76
32.3 Entropy . . . 76
32.4 Invariant measures . . . 77

1 The $\times 2$ map

Identify the circle $\mathbb{T}$ with $[0,1)$. Let $T : [0,1) \to [0,1)$ be the map $T(x) = 2x \mod 1$. We will try to understand the behaviour of **forward orbits**. These are sequences of the form $x, T(x), T^2(x), \ldots$. For example, x is said to have **period** n if $T^n x = x$ and there does not exist $m < n$ with $T^m x = x$.

How many elements have period n?

Before answering this, consider the special case $n = 1$. A **fixed point** is an element $x \in [0,1)$ with $Tx = x$. Note 0 is the only fixed point. An **eventually fixed point** is a point x such that there is some n such that $T^n x$ is a fixed point. What are the eventually fixed points?

The map T exhibits **sensitive dependence on initial conditions** (this means that if for any $x \in \mathbb{T}$ there is a $\delta > 0$ such that if x and y are δ-close then there is some $n \in \mathbb{N}$ such that $T^n x$ and $T^n y$ are at least δ-far apart.

To motivate this property consider the problem of trying to predict the weather in a deterministic universe. We assume that we know the evolution equations for our universe exactly. Therefore if we know the state of the universe exactly at any particular moment then we can predict its behaviour in all future times. However, in reality we do not know the state of the universe exactly. There is always some error in measurement. So suppose $x \in [0,1)$ represents the true state of the universe but because of our clumsiness or whatever, we measure y and $|x - y| < \epsilon$. So given y, what is:

$$\max\{|T^n x - T^n y| : \ |x - y| < \epsilon\}?$$

A special case occurs when $y = x + 1/2^n$. In this case, $T^{n-1} y = x + 1/2$. The maximum distance between any two points in $[0,1)$ is $1/2$. So: if $2^{-n} < \epsilon$ then after $n - 1$ steps, we have no information about the true state state of the system. In other words, it takes only $-\lfloor \log_2(\epsilon) \rfloor - 1$ steps to lose all information about system.

1.1 Coding

To better understand this system, consider writing numbers $x \in [0,1)$ in binary. So $x = \sum_{i=1}^{\infty} x_i 2^{-i}$ where $x_i \in \{0,1\}$ are the bits in the binary expansion of x. Then $Tx = \sum_{i=1}^{\infty} x_{i+1} 2^{-i}$. In other words, $(Tx)_i = x_{i+1}$.

Let's make this more precise. Consider the space $\Sigma_2 = \{0,1\}^{\mathbb{N}}$ with the product topology. Let $\sigma : \Sigma_2 \to \Sigma_2$ denote the shift map $\sigma(x)_i = x_{i+1}$. Observe this is a homeomorphism. Define $S : \Sigma_2 \to [0,1)$ by $S(x) = \sum_{i=1}^{\infty} x_i 2^{-i}$. This map is continuous but it is not injective. Why?

Observe that $S\sigma = TS$. This map is what we call a **semi-conjugacy** or **factor map**: it is continuous and satisfies $S\sigma = TS$ (so it 'conjugates' the dynamics). A semi-conjugacy with continuous inverse is a called a **topological conjugacy**. One general goal of dynamical systems theory is to classify homeomorphisms up to topological conjugacy and semi-conjugacy.

Exercise 1. How many points of period n are there in (Σ_2, σ)? How many are there in $(T, [0,1))$? For each system: are the periodic points dense?

2 Rotations

Let $\alpha \in [0,1)$ and consider the map $R : [0,1) \to [0,1)$ defined by $R(x) = x + \alpha \mod 1$.

Exercise 2. Find all periodic points of R. Does R exhibit sensitive dependence to initial conditions?

This system has a very different nature than the ×2 map. It is **isometric** (it preserves the usual metric on the circle).

We say a system is **minimal** if all orbits are dense.

Exercise 3. If α is irrational then $(R, \mathbb{T})$ is minimal.

Minimality is an interesting concept so let's take a moment to explore it.

Exercise 4. If X is a compact Hausdorff space and $T : X \to X$ is a homeomorphism then there exists a closed T-invariant subspace Y such that (T, Y) is minimal.

Exercise 5 (Birkhoff's Recurrence Theorem). If X is a compact Hausdorff space and $T : X \to X$ is a homeomorphism then there exists a sequence $\{n_i\}$ of integers such that $T^{n_i}x \to x$ as $i \to \infty$.

2.1 Group Translations

Let X denote a compact metrizable group and $g \in G$. Let $T_g : X \to X$ be the translation $T_g(x) = gx$.

Exercise 6. Show (X, T_g) is isometric. Is (X, T_g) necessarily minimal?

A homeomorphism $T : X \to X$ is **topologically transitive** if there is a dense forward orbit. So: if T is minimal then it is topologically transitive.

Exercise 7. Suppose X is a compact metrizable group and $g \in X$. If T_g is topologically transitive then X is abelian.

2.2 Group Endomorphisms

The ×2 map is not a group translation, it is a group endomorphism. Is it minimal?

More generally, if X is a compact group and $\alpha : X \to X$ is an endomorphism we may consider the system (X, α). For example, consider the torus $\mathbb{T}^2$ and α the map $\alpha(x,y) = (x, x+y)$. Is this map minimal? Does it exhibit sensitive dependence to initial conditions? How about $\beta(x,y) = (2x+y, x+y)$? This might be difficult to answer right now; we will answer it later.

2.3 Equidistribution

Are the orbits of R_α equidistributed? How about the orbits of the ×2 map? We will answer these questions later.

3 Van der Waerden's Theorem

I plan to follow Pollicott-Yuri's exposition or Petersen's.

4 Quadratic maps (following Devaney)

It is frequently of interest to understand how dynamical properties change when the map T is perturbed (in say that C_0 topology in the space of self-maps). Here we will consider an interesting and simple example: the 1-parameter family of maps on the real line defined by $F_\mu(x) = \mu x(1-x)$ where $\mu > 0$ (the case $\mu < 0$ is similar). The omega-limit set of a point x is the set of all $y \in \mathbb{R}$ such that there exist integers $n_1 < n_2 < n_3 < \cdots$ such that $F^{n_i}x \to y$. This is a closed subset, denoted by $\omega(y)$. (Similarly, we can define the α-limit set).

The first general question we ask: given $x \in \mathbb{R}$ what can be said about $\omega(x)$?

To begin: observe that if $x \notin [0,1]$ then $\lim_{n\to\infty} F_\mu^n(x) = -\infty$. This can be seen, for example, by drawing phase portrait.

Incidentally, this leads us to a new concept: we say that $p \in \mathbb{R}$ is **nonwandering** if for every open set $O \ni p$, there exists $q \in O$ and $n > 0$ such that $F_\mu^n q \in O$. Informally, there exist nearby points that recur or return to a neighborhood of p. The set $\Lambda \subset \mathbb{R}$ of all nonwandering points is closed.

We have just proven that $\Lambda \subset [0,1]$. What else can be said about Λ? ...

Before getting to that:

- $F_\mu(0) = F_\mu(1) = 0$;
- $F_\mu(p_\mu) = p_\mu$ where $p_\mu = \frac{\mu-1}{\mu}$;
- $F'_\mu(x) = \mu(1-2x)$ so $F'_\mu(1/2) = 0$ is the only critical point and F_μ is increasing for $x < 1/2$ and decreasing for $x > 1/2$; (F is **unimodal**);

If $0 < \mu < 1$ then 0 is an attracting fixed point and p_μ is repelling. Moreover, for any x, $\omega(x) = \{0\}, \{p_\mu\}$ or $\{-\infty\}$. More precisely, there is a point $\bar{p}_\mu$ such that $\bar{p}_\mu > 0$ and $F(\bar{p}_\mu) = p_\mu < 0$. If $x \in (p_\mu, \bar{p}_\mu)$ then $\omega(x) = \{0\}$. If $x \in \{p_\mu, \bar{p}_\mu\}$ then $\omega(x) = \{p_\mu\}$ and if $x \notin [p_\mu, \bar{p}_\mu]$ then $\omega(x) = \{-\infty\}$.

If $\mu = 1$ then 0 is attracting on one side and repelling on the other. It is the only fixed point. If $x \in [0,1]$ then $\omega(x) = \{0\}$. Otherwise $\omega(x) = \{-\infty\}$.

If $1 < \mu < 3$ then 0 is a repelling fixed point and p_μ is attracting (because $|F'(0)| > 1$ and $|F'(p_\mu)| < 1$). One can show that for any $x \in (0,1]$, $\omega(x) = \{p_\mu\}$ and $x \notin [0,1]$ then $\omega(x) = \{-\infty\}$. A graphical analysis makes this convincing.

In summary, the dynamics of F_μ for $0 < \mu < 3$ is fairly "tame". This tameness continuous past $\mu = 3$ for awhile although the dynamics does become slightly more complicated. If $\mu > 3$ then there is a periodic point of period 2. In fact, the larger μ gets the more periodic points it has. Let us leave the region $3 \le \mu \le 4$ to be a bit mysterious for now and focus on $\mu > 4$.

When $\mu > 4$, the graph of F pokes out above the square $[0,1] \times [0,1]$. In other words, the maximum of F (which is $F(1/2) = \mu/4$) in $[0,1]$ is larger than 1. So there exists an open interval $A_0 \subset [0,1]$ such that $F(A_0) \subset (1,\infty)$. For any $x \in A_0$, $\omega(x) = \{-\infty\}$.

Let I_0, I_1 denote the (closed) subintervals of $[0,1]$ in the complement of A_0. Note that F maps each of these onto $[0,1]$. In particular, there exist open subintervals in each of I_0, I_1 that map to A_0. Let A_1 be the union of these two open intervals. Then $I_0 - A_1$ (for example) consists of two subintervals I_{01} and I_{11}. Each of these are mapped by F onto either I_0 or I_1. So there exist subintervals of I_{01} and I_{11} that are mapped into A_1 which is then mapped into A_0, etc.

We now have an intuitive picture of $\Lambda := \{x \in [0,1] : \ F^n(x) \in [0,1] \forall n\}$. Indeed,

$$\Lambda = \cap_{n=1}^{\infty} F^{-n}(I_0 \cup I_1)$$

and this is a decreasing intersection. This gives us an idea: define $S : \Lambda \to \Sigma_2$ by $S(x)_n = i$ if $F^n(x) \in I_i$.

Note $S(Fx)_n = i$ iff $F^{n+1}(x) \in I_i$ iff $S(x)_{n+1} = i$. So $S(Fx) = \sigma S(x)$ and S conjugates F with the shift.

Because F is continuous, S is continuous. So S is a semi-conjugacy with a subshift. In fact, it is a conjugacy. Proving this is a bit difficult in general, so we make the simplifying assumption that $\mu > 2 + \sqrt{5}$. This assumption implies that $F'(x) > 1$ for all $x \in I_0 \cup I_1$ and therefore there is a $\lambda > 1$ such that $F'(x) > \lambda$ for all $x \in \Lambda$.

Let's show S is injective. Let $x, y \in \Lambda$ and suppose $S(x) = S(y)$. In particular, this implies either $x, y \in I_0$ or $x, y \in I_1$. A bit of reflection shows that the interval between x and y is contained in Λ (in fact, $S^{-1}(p)$ must be an interval for any $p \in \Sigma_2$). Since F is monotone on I_0 and monotone on I_1 and $F' \geq \lambda$, we have that $|F^n x - F^n y| \geq \lambda^n |x - y|$. If $x \neq y$ then we obtain a contradiction since the $\lambda^n |x - y| > 1$ for large enough n but $F^n x, F^n y \in [0,1]$.

As an exercise, you can verify that S is onto.

We have shown that F restricted to its nonwandering set is topologically conjugate to the 2-shift. In particular, it has periodic points of all orders. It is not minimal but it is topologically transitive.

Definition 1. A set $\Lambda \subset \mathbb{R}$ is a **repelling hyperbolic set** for a map $f : \mathbb{R} \to \mathbb{R}$ if Λ is compact, f-invariant and for some N $|(f^n)'(x)| > 1$ for all $n > N$ and all $x \in \Lambda$.

5 Smale's Horseshoe

Definition 2. Given a diffeomorphism $f : M \to M$ of a Riemannian manifold, a subset $\Lambda \subset M$ is a **hyperbolic set** if it is invariant, compact and there are constants $0 < \lambda < 1, C > 0$ and for every $x \in \Lambda$ a splitting $T_x M = E^s_x \oplus E^u_x$ of the tangent space such that

- $df_x E^s_x = E^s_{fx}$, $df_x E^u_x = E^u_{fx}$
- for every $v \in E^s_x$ and $n \geq 0$, $\|df^n_x(v)\| \leq C\lambda^n \|v\|$,

- for every $v \in E^u_x$ and $n \geq 0$, $\|df^{-1}n_x(v)\| \leq C\lambda^n\|v\|$.

For example, the Λ we discussed in the section on the quadratic family is hyperbolic. Here we will discuss the famous example of Smale's horseshoe.

This involves a lot of drawing which is a bit inconvenient to do on my laptop (sorry). First, draw a unit square and flank it with two semi-circles on its vertical sides. This is a "stadium". Choose some $\delta > 0$. Shrink the stadium by δ in the vertical direction, expand the square by $1/\delta$ in the horizontal direction (leaving the ends the same size), then bend it into a horseshoe shape and place it back in the stadium. Call the resulting map f.

We want to know; for $x \in S$ ($S =$ the stadium), what can be said about $\lim_n f^n(x)$?

The two semicircular regions are contracted by f. The first one, call it D_1, is mapped into itself. The contraction mapping principle implies there is a unique point $p \in D_1$ such that for any $x \in D_1$, $\lim_n f^n x = p$. Since $f(D_2) \subset D_1$, the same is true of points in D_2. In fact, f also maps a middle vertical strip into D_2 which is then mapped into D_1. So the same is true of the middle vertical strip.

There are two vertical rectangles V_0, V_1 such that $f(V_i) \subset [0,1]^2$. By the way, we can assume that f restricted to $V_0 \cup V_1$ is linear. Notice that f maps V_0 into a horizontal rectangle H_0 and V_1 into a horiztonal rectange H_1. Now if $f^2(x)$ is in the unit square, then $f(x) \in V_0 \cup V_1$ which means x is contained in one of four vertical rectangles. It's easy to see the pattern now: there is a Cantor set C such that $f^n(x)$ is in the unit square for all $n \geq 0$ iff $x \in C \times [0,1] =: \Lambda_+$.

Similar reasoning shows that if $f^{-n}x$ is well-defined and in the unit square for all $n \geq 0$ iff $x \in [0,1] \times C$ (assuming an obvious symmetry). Also observe that, since we assumed f is linear of $V_0 \cup V_1$, $|\partial f/\partial x| = 1/\delta$ for all $x \in C \times [0,1]$ whereas $|\partial f/\partial y| = \delta$ for all $x \in [0,1] \times C =: \Lambda_-$.

Let $\Lambda = \Lambda_- \cap \Lambda_+$ and define $S : \Lambda \to \Sigma_2 := \{0,1\}^{\mathbb{Z}}$ by $S(x)_i = \beta$ if $f^i x \in V_\beta$. This gives a topological conjugacy. It's probably not obvious why this is 1-1 right now. The other parts, continuity and surjectivity are easy. We will go over injectivity in a moment.

Definition 3. Points x, y are **forward asymptotic** if $\lim_{n\to\infty} |f^n x - f^n y| = 0$ and **backwards asymptotic** if $\lim_{n\to\infty} |f^{-n}x - f^{-n}y| = 0$. The **stable set of** x, denoted $W^s(x)$, is the set of all points forward asymptotic with x. The **unstable set of** x, denoted $W^u(x)$ is the set of all point backwards asymptotic to x.

Suppose $x, y \in \Lambda_+$ are on the same vertical line. Then they are forward asymptotic. This is because f shrinks everything by δ in the vertical direction. Similarly, if $x, y \in \Lambda_-$ are on the same horizontal line then they are backwards asymptotic.

This explains why S is 1-1: if $Sx = Sy$ then since f is contracting in the vertical direction, we must have $|\pi_2(x) - \pi_2(y)| \leq \delta^n |\pi_2(f^{-n}x) - \pi_2(f^{-n}y)|$ for every n, which implies $\pi_2(x) = \pi_2(y)$ (where π_2 is projection to the second coordinate). A similar argument holds in the horizontal direction; so f is 1-1.

Exercise: $x, y \in \Lambda$ are forward asymptotic iff Sx, Sy have the same forward tail (and similarly in the backwards direction).

Exercise: x, y are forward asymptotic iff there is some $n \geq 0$ such that $f^n x, f^n y$ are on the same vertical line. So $W^s(x) = \cup_n f^{-n}(l_x \cap \Lambda)$ where l_x is the vertical line containing x.

6 Symbolic dynamics (following Katok-Hasselblatt)

Let $\Omega_N = \{0, \ldots, N-1\}^{\mathbb{Z}}$ and $\Omega_N^R = \{0, \ldots, N-1\}^{\mathbb{N}}$. We let $\sigma : \Omega_N \to \Omega_N$ denote the shift map $\sigma(\omega)_i = \omega_{i+1}$. The shift map on Ω_N^R is defined similarly.

Lemma 6.1. *Periodic points for the shifts σ_N, σ_N^R are dense in Ω_N and Ω_N^R respectively. The number of periodic points of period n is N^n in both cases. Both transformations are topologically mixing (this means for any open subsets U, V, there exists N such that $n > N$ implies $\sigma^n(U) \cap V$ is nonempty).*

A **subshift** is any closed shift-invariant subspace of Ω_N (or Ω_N^R) and symbolic dynamics in the study of such subshifts. They are used in information theory, probability theory and to "code" other dynamical systems (as we have already seen with ×2 map and the quadratic family). They also provide a rich source of examples.

Exercise 8. It is not difficult to see that Ω_N (for example) is not minimal for the shift. Can you construct a subshift (with infinite cardinality) that is minimal?

6.1 Topological Markov chains

Let $A = (a_{ij})$ be an $N \times N$ matrix with $a_{ij} \in \{0, 1\}$ for all i, j. Let Ω_A be the set of all $\omega \in \Omega_N$ such that $a_{\omega_n, \omega_{n+1}} = 1$ for all n. Ω_A is called a **topological Markov chain**. It is an example of a subshift of finite type (a concept which is only slightly more general).

We say A is **irreducible** if for every i, j there is some n such that $(a^n)_{ij} > 0$.

Exercise 9. If A is irreducible then periodic points are dense in Ω_A. Moreover, the shift σ restricted to Ω_A is topologically mixing.

Sometimes is helps to consider the directed graph G with vertex set $\{0, 1, \ldots, N-1\}$ and edges (i, j) for $a_{ij} = 1$. Then Ω_A is just the set of bi-infinite directed paths in G.

Lemma 6.2. *a_{ij}^n is the number of directed paths in G that start at i, end at j and have length n.*

Corollary 6.3. *The number of points x in Ω_A with $\sigma^n x = x$ is $tr(A^n)$.*

Theorem 6.4 (Perron-Frobenius)**.** *Suppose L is an $N \times N$ matrix with nonnegative entries and, for some integer $n > 0$, L^n has all positive entries. Then there exists a unique eigenvalue λ of L which admits a eigenvector with all positive entries. Moreover, λ is positive, simple and if μ is any other eigenvalue of L then $|\mu| < \lambda$.*

Corollary 6.5. *If P_n is the number of $x \in \Omega_A$ with $\sigma^n x = x$ then $P_n = \lambda^n + O(\mu^n)$ where $\lambda > 1$ is the PF-eigenvalue of A and $0 < \mu < \lambda$.*

Proof sketch of the PF theorem. Let $P \subset \mathbb{R}^N$ be the positive orthant and $\sigma \subset P$ the set of all $x \in \mathbb{R}^N$ with $\sum_i x_i = 1$. This is the convex hull of the usual basis $e_1, \ldots, e_N$. Let $T : \sigma \to \sigma$ be the map $Tx = Lx/\|Lx\|_1$. This is well-defined since L has all nonnegative entries.

Consider $\cap_n T^n\sigma$. This is a decreasing intersection. It is invariant under T, closed and convex. We will show that it contains only a single point. First we show it has at most N extreme points. Indeed, let $x \in \cap_n T^n\sigma$. Then for every n, $x \in T^n\sigma$. Since $T^n\sigma$ is the convex hull of the vectors $T^n e_i$, $x = \sum_{i=1}^N \lambda_i^{(n)} T^n e_i$ for some $\lambda_i^{(n)} \ge 0$ with $\sum_i \lambda_i^{(n)} = 1$. There exists a subsequence $n_j \to \infty$ such that $\lim_j T^{n_j} e_i =: p_i$ exists for each i and $\lambda_i := \lim_j \lambda_i^{(n_j)}$ also exists. So $p_i \in \cap_n T^n\sigma$ for all i and $x = \sum_i \lambda_i p_i$. Thus we have shown that $\cap_n T^n\sigma$ is contained in the convex hull of the p_i's. This proves the claim: the p_i's must be the extreme points of $\cap_n T^n\sigma$. We want to prove that all of the p_i's are equal to each other.

There is some $m > 0$ such that T^m fixes every extreme point of $\cap_n T^n\sigma$. The extreme points are therefore eigenvectors of L^m. They have all positive coordinates because L^n has all positive entries.

Suppose $x, y \in \cap_n T^n\sigma$ both have eigenvalue $\lambda > 0$. Since these have all positive entries, there is some linear combination $tx - sy$ that has all nonnegative entries, at least one zero entry and at least one positive entry (assuming $x \neq y$). But then $L^{nm}(tx - sy) = \lambda^n(tx - sy)$ stays in the positive orthant but still has one zero entry. This contradicts the assumption that L^n has all positive entries.

Suppose $x, y \in \cap_n T^n\sigma$, x has eigenvalue λ and y has eigenvalue μ with $0 < \mu < \lambda$. Then $y - \epsilon x \in P$ if $\epsilon > 0$ is small enough. But then $L^{nm}(y - \epsilon x) = \mu^n y - \lambda^n \epsilon x$ goes outside P if n is sufficiently large contradicting that L leaves P invariant.

We now know that $\cap_n T^n\sigma$ contains a single point v which must be an eigenvector of L with some eigenvalue $\lambda > 0$. We must show that if μ is any other eigenvalue of L then $|\mu| < \lambda$.

If $|\mu| > \lambda$ and μ is real then after replacing L with L^2 we may assume $\mu > \lambda$. Then if x is an eigenvector with eigenvalue μ then x must lie outside P. On the other hand $v + \epsilon x \in P$ if ϵ is small enough. But $L^n(v + \epsilon x) = \lambda^n v + \mu^n \epsilon x$ lies outside P if n is large enough contradicting $LP \subset P$.

The cases when μ is complex or $|\mu| = \lambda$ are similar and left as exercises.

□

6.2 Current research in symbolic dynamics

1. Hidden Markov chains (see work of Mike Boyle, Karl Peterson)
2. low complexity subshifts (see work of Van Cyr, Bryna Kra)
3. topological entropy of $\mathbb{Z}^d$-shifts (Mike Hochman and Tom Meyerovitch)
4. symbolic extensions of systems (Downarowicz et al)
5. automorphism groups of symbolic systems (is $\mathrm{Aut}(\Sigma_2)$ isomorphic to $\mathrm{Aut}(\Sigma_3)$?)
6. the topological conjugacy problem: it is currently unknown whether topological conjugacy of top Markov chains is decidable. There are a number of computable invariants which often distinguish different subshifts but they are not known to be complete.

Let A, B be 0-1 matrices. We say they are lag-1 shift equivalent (SE) if exists a pairs of matrices R, S with entries in $\mathbb{Z}_{\geq 0}$ such that $A = RS, B = SR$. The transitive closure of this relation is called strong shift equivalence (SSE). Bob Williams proved in his 1973 paper that A and B are SSE iff the shifts σ_A, σ_B are top conjugate.

A, B are shift equivalent (SE) if there exist matrices R, S over $\mathbb{Z}_{\geq 0}$ and a positive integer l (called the lag) such that $A^l = RS, B^l = SR, AR = AB$ amd $SA = BS$. Bob showed that if A, B are SE then σ_A^n is top conj to σ_B^n for all but finitely many n. He gave an incorrect proof that SE implies SSE. This latter claim was thoroughly disproven by Kim-Roush in 1999. Kim-Roush also obtained (in earlier work) an algorithm for detecting SE. Unfortunately, there is no algorithm known for deciding SSE.

7 Toral flows

Let $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ denote the 2-dimensional torus. Given a unit vector $v \in \mathbb{R}^2$ we consider the flow $T_t : \mathbb{T}^2 \to \mathbb{T}^2$ given by $T_t(x, y) = (x, y) + tv \mod \mathbb{Z}^2$.

To study the dynamics of this map, we will consider an appropriately chosen subset $S \subset \mathbb{T}^2$ and the return time map $T^S : S \to S$ defined by $T^S z = T_t z$ where $t > 0$ is the smallest number such that $T_t z \in S$. To be useful, we would like S to intersect every orbit in a discrete set.

For simplicity assume $v = (v_1, v_2)$ and $v_1, v_2 \neq 0$. Let S be the circle $x = 0$. Then $T^S = T_{v_2/v_1}$ and $T^S(0, y) = (0, (v_2/v_1)y)$. So T^S is naturally conjugate to rotation by v_2/v_1. In particular, if v_2/v_1 is rational then every orbit is closed (in fact, each is a circle). If v_2/v_1 is irrational then each orbit is dense in S. Replacing S with $S + (x, 0)$ for an arbitrary x and applying the same argument we see that if v_2/v_1 is irrational then every orbit is dense so the flow is **minimal**. This can be generalized to arbitrary dimensions.

8 Toral automorphisms

Recall that $SL(2, \mathbb{Z})$ is the group of 2×2 matrices with integer entries and determinant 1. If $A \in SL(2, \mathbb{Z})$ then $A\mathbb{Z}^2 = \mathbb{Z}^2$. Thus A descends to a map of the quotient space $\mathbb{R}^2/\mathbb{Z}^2$ which we denote by $\mathbb{T}^2$, the 2-dimensional torus.

We say A is **hyperbolic** if it does not have an eigenvalue on the unit circle (in the complex plane). The dynamics of A on the 2-torus are especially interesting in this case. Let assume for now on that A is hyperbolic with eigenvalues $\lambda_1 > 1 > \lambda_2 = \lambda_1^{-1} > 0$. Let $T : \mathbb{T}^2 \to \mathbb{T}^2$ denote the induced map.

Lemma 8.1. *A point $p = [x, y] \in \mathbb{T}^2$ is periodic iff x, y are rational. Moreover, the number of points p with $T^n p = p$ is $|\lambda_1^n + \lambda_2^n - 2|$.*

Proof. If x, y are rational then p is periodic is easy.

If $[x, y]$ is periodic then $A^n(x, y)^T = (x, y)^T + (n_1, n_2)^T$ for some integers n_1, n_2. Because A is hyperbolic, $A^n - I$ is invertible. Also $(x, y)^T = (A^n - I)^{-1}(n_1, n_2)^T$. But $(A^n - I)^{-1}$ has rational entries. So (x, y) are rational.

For the second statement, consider the map $A^n - I$ applied to the unit square $[0,1)^2 \subset \mathbb{R}^2$. $T^n p = p$ if and only if $(A^n - I)(x,y)^T \in (x,y)^T + \mathbb{Z}^2$. So the number of such point equals the $|(A^n - I)[0,1)^2 \cap \mathbb{Z}^2|$. Note $(A^n - I)[0,1)^2$ is a rhomboid. So this cardinality is exactly the area of $(A^n - I)[0,1)^2$ which is the absolute value of the determinant of $A^n - I$. (More detail: we obtain a tiling by copies of this rhomboid....)

The eigenvalues of $A^n - I$ are $\lambda_1^n - 1$ and $\lambda_2^n - 1$. So the determinant is $(\lambda_1^n - 1)(\lambda_2^n - 1)$. Since $\lambda_2 = \lambda_1^{-1}$, the result follows. □

Lemma 8.2. *Given $[x,y] \in \mathbb{T}^2$, let $W^s[x,y]$ be the set of points forward asymptotic to it. Let $W^u[x,y]$ be the set of points backwards asymptotic to it. Then $W^s[x,y]$ is the image of the line that passes through (x,y) and is parallel to the eigenline of A with corresponding eigenvector $\lambda_2 = \lambda_1^{-1}$. Similarly $W^u[x,y]$ is the image of the line that passes through (x,y) and is parallel to the eigenline of A with corresponding eigenvector λ_1. These sets are dense in $\mathbb{T}^2$. In fact $W^s[x,y] \cap W^u[x,y]$ is dense.*

A Markov partition of the torus is collection $R_1, \ldots, R_n$ of rectangles satisfying

- R_i is the closure of its interior (so it's not allowed to "border itself")
- the interiors of R_i and R_j do not overlap (if $i \neq j$)
- (for the purposes of this example, the last condition is:) the boundary of R_i decomposes as a union of stable and unstable boundaries with the stable boundary contained in an eigenline corresponding to λ_2 and unstable contained in an eigenline corresponding to λ_1. if $\partial^s R_i$ denote the stable boundary then $f(\partial^s R_i) \subset \cup_j \partial^s R_j$ and $\partial^u R_i \subset \cup_j f(\partial^u R_j)$.

This latter condition implies a Markov condition: if $f^n R_i \cap R_j \neq \emptyset$ and $f^m R_j \cap R_k \neq \emptyset$ then $f^{n+m} R_i \cap R_k \neq \emptyset$.

Given a Markov partition we let A be the matrix $A_{ij} = 1$ if $f(R_i) \cap R_j \neq \emptyset$ and $A_{ij} = 0$ otherwise.

Then we obtain a semi-conjugacy $S : \Sigma_A \to \mathbb{T}^2$ by: $S(x) = y$ where y is the unique point satisfying $f^n y \in R_{x_n}$ for all $n \in \mathbb{Z}$. The Markov condition above implies that there is a point y satisfying this condition. The contracting and expanding implies that it is unique.

To be concrete, suppose

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

This matrix has eigenvalues

$$\lambda_1 = \frac{3+\sqrt{5}}{2} > 1, \lambda_2 = \frac{3-\sqrt{5}}{2} = \lambda_1^{-1} < 1.$$

The corresponding eigenlines are $y = \frac{\sqrt{5}-1}{2}x$ and $y = \frac{-\sqrt{5}-1}{2}x$.

Observe that A preserves the family of lines $y = \frac{\sqrt{5}-1}{2}x + b$. That is, it takes any line in this family to another line in this family. Moreover, it expands these lines by a factor of λ_1. So if w, z are in one of these lines then $|Aw - Az| = \lambda_1 |w - z|$. These lines descend to the torus and the same is true on the torus if we interpret distance as "distance along the line in this family containing these points". Because $\frac{\sqrt{5}-1}{2}$ is irrational, any line in this family projects to a dense subset of the torus (by results of the previos subsection).

Similar statements hold for the family of lines $y = \frac{-\sqrt{5}-1}{2}x + b$.

The examples above are called **Anosov diffeomorphisms**: the whole manifold $\mathbb{T}^2$ is a hyperbolic set for T. It's an open problem to characterize smooth manifolds that admit Anosov diffeomorphisms. All known examples are on tori or nilmanifolds.

9 Topological entropy

Intuition: suppose $f : X \to X$ is a continuous map of a compact metric space (X, ρ). We would like to count how many orbits f has *up to a scale*. Of course, there are infinitely many orbits. But in a realistic scenario, we won't be able distinguish between two orbits that are very close. So we count how many orbits we can distinguish. In fact, we take the exponential rate of growth of the number of finite partial orbits that we are able to distinguish at scale $\epsilon > 0$; this is the topological entropy.

For later applications it is convenient to allow ρ to be a pseudo-metric; that is we allow $d(x, y) = 0$ even if $x \neq y$. A subset $S \subset X$ is

- (ρ, ϵ)-separated if $\rho(s, t) > \epsilon$ for $s \neq t \in S$;
- (ρ, ϵ)-spanning if for every $x \in X$ $\exists s \in S$ with $\rho(x, s) < \epsilon$.

Let ρ_n be the pseudo-metric:

$$\rho_n(x, y) = \max\{\rho(f^k x, f^k y) : \ 0 \leq k \leq n-1\}.$$

Let

- $span(\rho_n^f, \epsilon)$ be the minimum cardinality of a ρ_n, ϵ-spanning set
- $sep(\rho_n^f, \epsilon)$ be the maximum card. of a (ρ_n, ϵ)-separated set
- $cov(\rho_n^f, \epsilon)$ be the min card of a covering of X by sets of ρ_n diameter $< \epsilon$

Lemma 9.1. *$cov(n, 2\epsilon, f) \leq span(\rho_n^f, \epsilon) \leq sep(\rho_n^f, \epsilon) \leq cov(\rho_n^f, \epsilon)$.*

$$cov(m+n, \epsilon, f) \leq cov(m, \epsilon, f) cov(\rho_n^f, \epsilon).$$

Proof. $cov(n, 2\epsilon, f) \leq span(\rho_n^f, \epsilon)$: use a minimal card spanning set and compactness.

$span(\rho_n^f, \epsilon) \leq sep(\rho_n^f, \epsilon)$: take a max card separated set. If its' not spanning, add another a point. Contradiction.

$sep(\rho_n^f, \epsilon) \leq cov(\rho_n^f, \epsilon)$: take a max card separated set A and a cover $\mathcal{C}$. Define a function $\mathcal{P} : A \to \mathcal{C}$ so that $a \in \mathcal{P}(a)$. This is injective.

Let $\mathcal{C}_k$ be a $(k, \epsilon, f) - cover$. Then $\mathcal{C}_m \vee f^{-m}\mathcal{C}_n$ is a $(m+n, \epsilon, f)$-cover.

□

$h_{top}(f, \rho) := \lim_{\epsilon \searrow 0} \lim_n \frac{1}{n} \log cov(\rho^f_n, \epsilon)$ is the topological entropy. The limit above really exists because of subadditivity.

By the previous inequalities, we can replace *cov* with *span* or *sep* in the above.

We say ρ is dynamically generating if for every $x \neq y$ there exists n such that $\rho(f^n x, f^n y) \neq 0$.

Theorem 9.2. *If ρ_1, ρ_2 are dynamically generating continuous pseudometrics then $h_{top}(f, \rho_1) = h_{top}(f, \rho_2)$.*

The common number above is called the **topological entropy**. It is an invariant of topological conjugacy.

Lemma 9.3. *Let ρ be dynamically generated and continuous. Define d by*

$$d(x, y) = \sum_{n=0}^{\infty} 2^{-n} \rho(f^n x, f^n y).$$

Then d is a continuous metric and $h_{top}(f, \rho) = h_{top}(f, d)$.

Proof. Let $\epsilon > 0$. Since $d \geq \rho$, any (ρ_n, ϵ)-separated set is also (d_n, ϵ)-separated. So $sep(\rho^f_n, \epsilon) \leq sep(d^f_n, \epsilon)$ which implies $h_{top}(f, \rho) \leq h_{top}(f, d)$.

Let $A \subset X$ be (d_n, ϵ)-separated. So if $a \neq b \in A$ then $d_n(a, b) > \epsilon$ which means there is some m with $0 \leq m < n$ with $d(f^m a, f^m b) > \epsilon$ which means

$$\sum_{k=0}^{\infty} 2^{-k} \rho(f^{k+m} x, f^{k+m} y) > \epsilon.$$

If D is the ρ-diameter of X then for any K

$$\epsilon < \sum_{k=0}^{\infty} 2^{-k} \rho(f^{k+m} x, f^{k+m} y) \leq 2^{-K} D + \sum_{k=0}^{K} 2^{-k} \rho(f^{k+m} x, f^{k+m} y).$$

In particular, if we choose K so that $2^{-K} D < \epsilon/2$ then we have

$$\epsilon/2 < \sum_{k=0}^{K} 2^{-k} \rho(f^{k+m} x, f^{k+m} y).$$

So there exists some number l with $0 \leq l \leq K + n$ such that $\rho(f^l x, f^l y) > \epsilon/4$. So A is $(\rho_{K+n}, \epsilon/4)$-separated. Thus

$$sep(d^f_n, \epsilon) \leq sep(\rho^f_{n+K}, \epsilon).$$

$$\begin{aligned} h_{top}(d, f) &= \lim_{\epsilon \searrow 0} \lim_{n} 1/n \log sep(d^f_n, \epsilon) \\ &\leq \lim_{\epsilon \searrow 0} \lim_{n} 1/n \log sep(\rho^f_{n+K}, \epsilon) \\ &= h_{top}(\rho, f). \end{aligned}$$

□

Proof of Theorem 9.2. By previous lemma, we may assume ρ_1, ρ_2 are metrics. For $\epsilon > 0$ let $\delta(\epsilon) = \sup\{\rho_2(x,y) : \ \rho_1(x,y) \le \epsilon\}$. By compactness $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Note $cov((\rho_2)^f_{n+K}, \delta(\epsilon)) \le cov((\rho_1)^f_{n+K}, \epsilon)$. This implies $h_{top}(f, \rho_2) \le h_{top}(f, \rho_1)$. □

Lemma 9.4.
- *Topological entropy is monotone under factor maps.*
- $h_{top}(f \times g) = h_{top}(f) + h_{top}(g)$ *(where $f : X \to X, g : Y \to Y$).*

Proof. The first statement is clear. Let d_X, d_Y be metrics on X, Y. Define ρ on $X \times Y$ by $\rho((x_1,y_1),(x_2,y_2)) = \max(d_X(x_1,x_2), d_Y(y_1,y_2))$. If $\mathcal{C}$ is an (f,n,ϵ)-cover of X and $\mathcal{D}$ is a (g,n,ϵ)-cover of Y then $\mathcal{C} \times \mathcal{D}$ is a $(f \times g, n, \epsilon)$-cover of $X \times Y$. this implies $h_{top}(f \times g) \le h_{top}(f) + h_{top}(g)$.

If $A \subset X$ is (n,ϵ)-separated and $B \subset Y$ is (n,ϵ)-separated then $A \times B$ is (n,ϵ)-separated. So $sep(n,\epsilon, f \times g) \ge sep(\rho^f_n, \epsilon) sep(\rho^g_n, \epsilon)$ which implies $h_{top}(f \times g) \ge h_{top}(f) + h_{top}(g)$. □

9.1 Isometries

have zero entropy because $d_n = d$ for all n. In particular, irrational rotations have zero entropy. More generally, if $\{f^n\}_{n \ge 0}$ is equicontinuous, then f has zero entropy.

9.2 Topological Markov chains

First example: consider Σ_r with the pseudometric $\rho(x,y) = 1$ if $x_0 \ne y_0$ and $\rho(x,y) = 0$ otherwise. We observe that the max card of an (n,ϵ)-separated subset (if $\epsilon < 1$) is r^n. So $h_{top}(\sigma_r) = \log r$.

The same argument shows:

Lemma 9.5. *The topological entropy of a subshift $X \subset \Sigma_r$ is equal to*

$$\lim_{n \to \infty} \frac{1}{n} \log \#\textit{projection of } X \textit{ to the first } n \textit{ coordinates}$$

Let's apply this to a topological Markov chain determined by a matrix A. The number of admissible words of length $n+1$ is

$$\sum_{i,j} A^n_{ij}.$$

In particular, this is at least as large as the trace of A^n, which we showed grows like $\lambda^n + O(\mu^n)$. So: we have $h_{top}(X_A) \ge \log \lambda$.

On the other hand, we could write the above as

$$\sum_{i,j} \|A^n e_i\|_1$$

where e_i are the standard basis vectors. We have $\|A^n e_i\|_1 \le \lambda^n$ for all i, so this proves $h_{top}(X_A) = \log \lambda$.

9.3 Toral automorphisms

Let $A \in SL(2,\mathbb{Z})$ have eigenvalues λ, λ^{-1} with $|\lambda| > 1$. Let $T : \mathbb{T}^2 \to \mathbb{T}^2$ denote the induced map.

Theorem 9.6. $h_{top}(T) = \log|\lambda|$.

Proof. Let v_1, v_2 be unit eigenvectors for λ, λ^{-1}. Given a vector $w \in \mathbb{R}^2$, write $w = w_1 v_1 + w_2 v_2$ and set $\|w\| = \max(|w_1|, |w_2|)$. This is a norm on $\mathbb{R}^2$. In particular, it gives a distance d on $\mathbb{R}^2$. Let $d_n(x,y) = \max_{0 \le k < n} d(A^k x, A^k y)$. Then ϵ-balls in d_n are Euclidean rhomboids with side lengths $2\lambda^{-n}\epsilon$ (in the v_1-direction) and 2ϵ. They have area $4C\lambda^{-n}\epsilon^2$ where C is a constant depending on the angle of v_1 with v_2.

We push down to obtain a metric, also denoted by d_n, on $\mathbb{T}^2$. Using area, we obtain that $cov(T,n,\epsilon) \ge 1/(4C\lambda^{-n}\epsilon^2)$ which implies $h_{top}(T) \ge \lambda$. On the other hand, we can built a covering by ϵ-balls using at most $100/(4C\lambda^{-n}\epsilon^2)$ balls. This shows the opposite inequality. $\square$

More generally, for $A \in SL(n,\mathbb{Z})$ the entropy of $T_A : \mathbb{T}^n \to \mathbb{T}^n$ is $\log|\lambda_1 \cdots \lambda_m|$ where $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of A (with multiplicity) satisfying $|\lambda_i| > 1$. This assume A is hyperbolic.

10 Equicontinuous and distal flows

Let (X,d) be a compact metric space. A homeomorphism $f : X \to X$ is **equicontinuous** if the family $\{f^n : n \in \mathbb{Z}\}$ is equicontinuous. This means that for every $\epsilon > 0$ there exists a $\delta > 0$ such that if $d(x,y) < \delta$ then $d(f^n x, f^n y) < \epsilon$ *for every* n. We will prove a structure theorem for equicontinuous homeomorphisms (in the literature, these are often refered to as equicontiuous flows even though they are homeomorphisms). First,

Lemma 10.1. *$f : X \to X$ is equicontinuous iff there exists a continuous metric d' on X such that f is an isometry.*

Proof. If f is equicontinuous, then let $d'(x,y) = \sup_{n\in\mathbb{Z}} d(f^n x, f^n y)$. This is a metric and f acts by isometries. $\square$

A **Kronecker system** consists of a compact group G, an element $g_0 \in G$ and the left multiplication map $L_{g_0} : G \to G$ given by $L_{g_0}(g) = g_0 g$. We further assume L_{g_0} is minimal. This implies G is abelian.

Theorem 10.2. *$f : X \to X$ is equicontinuous iff it is topologically conjugate to a Kronecker system.*

Proof. It's easy to see that Kronecker systems are equicontinuous. So suppose f is equicontinuous. By the lemma, we may assume f is an isometry. Consider the group $Isom(X)$ of isometries of X endowed with the uniform topology (so $\mathcal{P}_n \to \mathcal{P}$ iff $\sup_x d(\mathcal{P}_n x, \mathcal{P} x) \to 0$ as $n \to \infty$.)

We view $\{f^n : n \in \mathbb{Z}\} \subset Isom(X)$. Let G denote the closure of $\{f^n : n \in \mathbb{Z}\}$. This is a group. By the Ascoli-Arzela Theorem, it is compact. It is also abelian.

Fix $x_0 \in X$. Then Gx_0 is compact and f-invariant. By minimality, $Gx_0 = X$. Let $K < G$ denote the stablizer of x_0. This is a closed subgroup. So G/K admits the quotient topology (same as the Hausdorff topology on closed subsets of G). The map $G/K \to X$, $gK \mapsto gx_0$ is a bijection; in fact a homeomorphism that conjugates L_f on G/K to f on X. $\square$

More generally, if $T : X \to X$ is any minimal homeomorphism then there exists a unique maximal equicontinuous factor $f : Y \to Y$. This is often called the **Kronecker factor**.

10.1 Distal maps

A homeomorphism $f : X \to X$ is **distal** if for every pair of distinct points $x \neq y$, $\inf_{n\in\mathbb{Z}} d(f^n x, f^n y) > 0$. Clearly, equicontinuity implies distality. But the converse is not true: consider the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ and the map $f(x, y) = (x + \alpha, x + y) \pmod 1$ where α is irrational. Why is this distal? Suppose $p_1 = (x_1, y_1), p_2 = (x_2, y_2)$ are distinct points. If $x_1 \neq x_2$ then $d(f^n p_1, f^n p_2) \geq |x_1 - x_2| > 0$ for all n. If $x_1 = x_2$ then $d(f^n p_1, f^n p_2) \geq |y_1 - y_2|$ for all n. This proves it.

But this map is not equicontinuous. Let $\epsilon > 0$ and $\delta > 0$. Let $p = (0, 0)$ and $q = (\delta, 0)$. Then $d(p, q) = \delta$ but

$$d(f^n p, f^n q) = d((n\alpha, (n-1)\alpha), (\delta + n\alpha, n\delta + (n-1)\alpha))$$

this is large if $n\delta$ is close to $1/2 \mod 1$ which is certainly possible if δ is irrational (or even if δ is rational but small).

So distality is more general than equicontinuity.

Definition 4. Suppose $f : X \to X$ factors onto $g : Y \to Y$ via a map $\pi : X \to Y$. Then f is called an **extension** of g. It is called an *isometric extension* if $d(fx, fx') = d(x, x')$ for every $x, x' \in X$ with $\pi(x) = \pi(x')$. Note that the above example is an isometric extension of an isometry (consider projecting onto the first factor).

Theorem 10.3 (Furstenberg, 1963)**.** *If $f : X \to X$ is a minimal distal homeomorphism then there exists a countable ordinal α and for every ordinal $\beta \leq \alpha$ a homeomorphism $f_\beta : X_\beta \to X_\beta$ such that:*

- *$f_0 : X_0 \to X_0$ is the trivial map on a one-point space*
- *$f_\alpha : X_\alpha \to X_\alpha$ is $f : X \to X$*
- *for any $\beta < \alpha$, $f_{\beta+1}$ is an isometric extension of f_β*
- *if $\beta \leq \alpha$ is a limit ordinal then $f_\beta : X_\beta \to X_\beta$ is the inverse limit of the systems $\{f_\theta : X_\theta \to X_\theta\}_{\theta<\beta}$.*

The **distal rank** of a minimal distal system is the least ordinal α that such that one can find systems as above. Beleznay and Foreman (1995) showed that for every countable ordinal α there exists a minimal distal system whose distal rank is exactly α.

11 Multiple recurrence and Ramsey theory

Theorem 11.1 (Van der Waerden)**.** *If $\mathbb{N} = \sqcup_{i=1}^{K} S_i$ is a disjoint union then one of the S_i's contains arbitrarily long finite arithmetic progressions.*

Theorem 11.2 (Furstenberg's multiple recurrence theorem)**.** *Let $T : X \to X$ be a homeomorphism of a compact metric space (X, d). Let $L \in \mathbb{N}$ and $\epsilon > 0$. Then there exists $x \in X$ and $k \in \mathbb{N}$ such that*

$$d(x, T^{kl}x) < \epsilon \quad \forall 0 \le l \le L.$$

Proof of Van der Waerden from Fursterberg. Let $\omega = (\omega_1, \omega_2, \ldots)$ denote the sequence $\omega_i = j$ if $i \in S_j$.

Let X denote the set of all sequences $x \in \Sigma_K = K^{\mathbb{Z}}$ such that x locally looks like ω. To be precise, this means that for every $n \in \mathbb{Z}$, $I \in \mathbb{N}$ there is some $m \in \mathbb{N}$ such that $x_{n+i} = \omega_{m+i}$ for all $0 \le i \le I$.

Let $\sigma : \Sigma_K \to \Sigma_K$ denote the usual shift map. Notice that X is closed and shift-invariant, so can restrict the shift σ to X.

We apply Furstenberg's multiple recurrence theorem to (X, σ) with $\epsilon < 1$. The choice of ϵ means that $d(x, y) < \epsilon$ implies $x_0 = y_0$. So for every $L \in \mathbb{N}$ there exists $x \in X$ and $k \in \mathbb{N}$ such that $x_0 = (\sigma^{kl}x)_0 = x_{kl}$ for every $0 \le l \le L$. By choice of X, there exists $m \in \mathbb{N}$ such that $\omega_m = \omega_{m+kl}$ for every $0 \le l \le L$. But this means that $m, m+k, m+2k, \ldots, m+Lk \in S_i$ where $\omega_m = i$. ☐

We will prove a result (due to Furstenberg-Weiss, 1978) that is apriori stronger than Furstenberg's multiple recurrence theorem. The main point is that we can replace powers of T with an arbitrary finite collection of commuting abelian groups. We will formalize this in a way that makes the proof easy.

Let G be an abelian group. An **IP-system** in G is basically a homomorphism from $\oplus_{\mathbb{N}}\mathbb{Z}$ into G. To make this definition easy to use, we say that an IP-system is a collection of elements $\{T_S\} \subset G$ indexed by finite subsets of $\mathbb{N}$ such that $T_S = T_{s_1} \cdots T_{s_n}$ if $S = \{s_1, \ldots, s_n\}$. IP stands for "infinite-dimensional parallelopiped".

Theorem 11.3 (Furstenberg-Weiss, 1978)**.** *Let G be an abelian group acting minimally on a compact space X. Then for every open $U \subset X$, finite $\alpha \subset \mathbb{N}$, IP-systems $T^{(1)}, \ldots, T^{(L)}$ in G there exists finite $\beta \subset \mathbb{N}$ such that $\alpha < \beta$ and*

$$U \cap T^{(1)}_{\beta} U \cap \cdots T^{(L)}_{\beta} U \neq \emptyset.$$

Proof of Furstenberg's multiple recurrence theorem from Furstenberg-Weiss. After passing to a minimal subspace, we may assume $T : X \to X$ is minimal. Let $x \in X$ be arbitrary and U be the ϵ-neighborhood of x.

If $S = \{s_1, \ldots, s_n\} \subset \mathbb{N}$ is finite then let $T^{(k)}_S = T^{ks_1} \cdots T^{ks_n}$ for $1 \le k \le L$. We now apply Furstenberg-Weiss to conclude. ☐

Proof of Furstenberg-Weiss. Because X is compact there is a finite set $G_* \subset G$ such that $X = \cup_{g \in G_*} gU$.

We argue by induction on L. First we handle the base case $L = 1$ and let $T = T^{(1)}$. Set $V_0 = U$. For $k > 0$, choose $g_k \in G_*$ so that

$$V_k = T_k V_{k-1} \cap g_k U \neq \emptyset.$$

Note

$$T_k^{-1} V_k \subset V_{k-1} \subset g_{k-1} U.$$

By the pigeonhole principle, we can choose $p < q$ with $\alpha < \beta := \{p+1, \dots, q\}$ and $g_0 \in G_*$ such that $V_p \cup V_q \subset g_0 U$.

$$T_\beta^{-1}(g_0^{-1} V_q) = g_0^{-1} T_{p+1}^{-1} \cdots T_q^{-1}(V_q) \subset g_0^{-1} V_p \subset U.$$

Thus $U \cap T_\beta(U) \supset g_0^{-1} V_q$ is nonempty.

Now we assume the theorem is true for any L IP-systems and prove it for $L + 1$.

Claim. There exist nonempty open sets $V_k \subset X$ and an increasing sequence $\alpha_k \in 2^{\mathbb{N}}_{fin}$ such that

- $\alpha_k > \alpha$ for all k
- $V_0 = U$, $\cup_{j=1}^{L+1} (T^{(j)}_{\alpha_k})^{-1}(V_k) \subset V_{k-1}$
- $V_k \subset g_k U$ for some $g_k \in G_*$.

Proof of claim. Apply the inductive hypothesis to $(T^{(L+1)})^{-1}T^{(1)}, \dots (T^{(L+1)})^{-1}T^{(L)}$ to obtain the existence of $\alpha_1 > \alpha$ satisfying

$$U \cap (T^{(L+1)})^{-1} T^{(1)}_{\alpha_1} U \cap \cdots (T^{(L+1)})^{-1} T^{(L)}_{\alpha_1} U \neq \emptyset.$$

Applying $T^{(L+1)}_{\alpha_1}$ we obtain:

$$T^{(1)}_{\alpha_1} U \cap \cdots T^{(L+1)}_{\alpha_1} U \neq \emptyset.$$

Choose $g_1 \in G_*$ and set

$$V_1 := g_1 U \cap T^{(1)}_{\alpha_1} U \cap \cdots T^{(L+1)}_{\alpha_1} U \neq \emptyset.$$

If V_{k-1} and α_{k-1} have been constructed apply the inductive hypothesis to V_{k-1} and

$$(T^{(L+1)})^{-1}T^{(1)}, \dots (T^{(L+1)})^{-1}T^{(L)}$$

to obtain $\alpha_k > \alpha_{k-1}$ such that

$$V_{k-1} \cap (T^{(L+1)})^{-1} T^{(1)}_{\alpha_1} V_{k-1} \cap \cdots \cap (T^{(L+1)})^{-1} T^{(L)}_{\alpha_1} V_{k-1} \neq \emptyset.$$

As before, apply $T^{(L+1)}_{\alpha_k}$ to obtain

$$T^{(1)}_{\alpha_k} V_{k-1} \cap \cdots \cap T^{(L+1)}_{\alpha_k} V_{k-1} \neq \emptyset.$$

Now choose $g_k \in G_*$ and set

$$V_k := g_k U \cap T^{(1)}_{\alpha_k} V_{k-1} \cap \cdots T^{(L+1)}_{\alpha_k} V_{k-1} \neq \emptyset.$$

This proves the claim. □

By the pigeonhole principle there is some $g_0 \in G_*$ such that $V_k \subset g_0 U$ for infinitely many k. In particular, there are arbitrarily large $p < q$ such that $V_p \cup V_q \subset g_0 U$. Set $\beta = \alpha_{p+1} \cup \cdots \cup \alpha_q$. For every $1 \leq j \leq L+1$,

$$(T^{(j)}_\beta)^{-1} g_0^{-1} V_q = g_0^{-1} (T^{(j)}_{\alpha_{p+1}})^{-1} \cdots (T^{(j)}_{\alpha_q})^{-1} V_q \subset g_0^{-1} V_p \subset U.$$

So $g_0^{-1} V_q \subset U \cap T^{(j)}_\beta U$. Since this is true for every j, we've proven the theorem.

□

11.1 Further results

Theorem 11.4. *Let $A \subset \mathbb{Z}^d$ be finite and $\mathbb{Z}^d = \sqcup_{j=1}^K S_j$ be a partition. Then there exists $k \in \{1, \ldots, K\}$ such that S_k contains a homothetic copy of A. That is: $a + bA \subset S_k$ for some $a \in \mathbb{Z}^d, b \in \mathbb{N}$. In fact, there is a k that contains infinitely many homothetic copies of A.*

Theorem 11.5. *Let F denote a finite field and $F_\infty = \oplus_{\mathbb{N}} F$. Let $F_\infty = \sqcup_{j=1}^K S_j$ be a partition. Then there exists a $k \in \{1, \ldots, K\}$ such that S_k contains arbitrarily large (finite) affine subspaces.*

12 Examples of measure-preserving actions and basic notions

Theorem 12.1. *Let $T : (X, \mu) \to (X, \mu)$ be a measure-preserving transformation of a probability space. The following are equivalent:*

1. *there exists a set $A \subset X$ with $0 < \mu(A) < 1$ such that $T^{-1}A = A$*
2. *there exists a nonconstant function $f \in L^1(X, \mu)$ such that $f(Tx) = f(x)$ for a.e. x*
3. *there exist T-invariant Borel probability measures $\mu_1 \neq \mu_2$ and $0 < t < 1$ such that $\mu = t\mu_1 + (1-t)\mu_2$.*

Proof. (1) implies (2): let $f = \chi_A$ the characteristic function of A.

(2) implies (3): By decomposing f into positive and negative parts, we see that, wlog, we may assume $f \geq 0$. After replacing f with $\min(f, C)$ for some constant C, we see that, wlog, we may assume $f \in L^\infty$. After scaling f we may assume $0 \leq f \leq 1$. Define μ_1, μ_2 by

$$\mu_1(B) = \|f\|_1^{-1} \int_B f \, d\mu, \quad \mu_2(B) = \|1-f\|_1^{-1} \int_B 1 - f \, d\mu$$

for any Borel $B \subset X$. These are T-invariant measures because f is T-invariant. Clearly,

$$\|f\|_1\mu_1 + \|1-f\|_1\mu_2 = \mu.$$

This proves (3).

(3) implies (1): Consider $\mu_1 - \mu_2$. This is a signed measure. By Hahn decomposition there exists a partition $X = P \sqcup N$ of X such that $\mu_1 - \mu_2$ is positive on P and negative on N. Because $\mu_1 \neq \mu_2$ we must have $0 < \mu(P) < 1$. Clearly P is a T-invariant. □

Definition 5. Let $T : (X,\mu) \to (X,\mu)$ be a measure-preserving transformation of a probability space. T is **ergodic** if the conditions in the previous Theorem do not hold. We also say that μ is **ergodic**. One should probably say T is ergodic wrt μ or μ is T-ergodic (since this notion depends on both T and μ) but it's fairly common to simply say T is ergodic or μ is ergodic if the measure or transformation is understood.

Example 1: rotations of the circle. Let $R_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ denote $R_\alpha(x) = x + \alpha \mod 1$. Let μ denote Lebesgue/Haar measure on the circle. This measure is invariant under any rotation. It is ergodic iff α is rational. One direction is easy, the other is a consequence of Fourier analysis.

Definition 6. Let $T : X \to X$ be pmp. T is **mixing** if for every measurable $A, B \subset X$,

$$\lim_{n\to\infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

Equivalently, T is mixing if

$$\lim_{n\to\infty} \int f(T^n x) g(x)\ d\mu(x) = \int f(x)\ d\mu(x) \int g(x)\ d\mu(x)$$

for every bounded measurable f, g. Of course, mixing implies ergodicity.

Example 2: Let (K,κ) be a standard probability space and consider the product measure space $(K,\kappa)^{\mathbb{Z}}$. The measure $\kappa^{\mathbb{Z}}$ is invariant under the shift action $\sigma : K^{\mathbb{Z}} \to K^{\mathbb{Z}}$ given by $\sigma(x)_i = x_{i+1}$. This action is mixing. This fact is easily seen to be true for cylinder sets. It is therefore true for finite unions of cylinder sets. We can approximate any subset by finite unions of cylinder sets.

Example 3: the times 2 map on the circle is mixing.

Example 4: hyperbolic toral automorphisms are mixing. (via Fourier analysis).

Extra: mixing is also called **strong mixing** (to distinguish it from weak mixing which will be discussed later) and **2-mixing**. An automorphisms $T : X \to X$ is k**-mixing** (where $k \geq 2$ is an integer) if for any collection of measurable sets $A_1, \ldots A_k \subset X$ we have

$$\lim_{n_1,n_2,\ldots,n_k\to\infty} \mu(T^{-n_1}A_1 \cap \cdots \cap T^{-n_k}A_k) = \prod_{i=1}^{k} \mu(A_i)$$

where the limit is taken as each $n_i \to \infty$ and $|n_i - n_j| \to \infty$ for all $i \neq j$. The automorphism is **mixing of all orders** if it is k-mixing for every k.

Open problem: does 2-mixing imply mixing of all orders?

13 Ergodic Theorems

Theorem 13.1 (von Neumann and Birkhoff's ergodic theorems)**.** *Let $T : X \to X$ be measure-preserving and $f \in L^1(X, \mu)$. Consider the* **ergodic averages**

$$f_n(x) := \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \in L^1(X, \mu).$$

Then $f_n \to$ in L^1 and pointwise a.e. where is the conditional expectation of f on the sigma-algebra of T-invariant sets. In particular, if T is ergodic then equals the constant $\int f \, d\mu$ a.e.

von Neumann proved convergence in L^1; Birkhoff proved pointwise convergence.

Lemma 13.2. *Let $f \in L^\infty(X)$. Then $f - f \circ T$ satisfying the ergodic theorem.*

Lemma 13.3. *The set of functions of the form $\{f - f \circ T : \ f \in L^\infty\}$ is dense in L^2.*

Lemma 13.4. *The mean ergodic theorem.*

Let $Mf(x) = \sup_n |f|_n(x)$. This is called the Hardy-Littlewood maximal operator. It is a nonlinear operator. However:

Theorem 13.5 (Maximal ergodic theorem)**.** *For any $f \in L^1$ and any $t > 0$,*

$$\mu(\{x \in X : \ Mf(x) > t\}) \leq C\|f\|_1/t$$

for some absolute constant $C > 0$.

(This is basically a covering argument).

We can now prove the pointwise ergodic theorem: observe that the set $\mathcal{S}$ of all functions $f \in L^1$ satisfying the theorem is dense by the previous lemmas. It is also closed by the maximal ergodic theorem. This does it.

14 Applications of the ergodic theorem

A number $x \in [0, 1)$ is **normal** if its decimal expansion $x = 0.i_1 i_2 \cdots$ is such that for every finite sequence $(j_1, \ldots, j_k) \in \{0, 1, \ldots, 9\}^k$,

$$\lim_{n\to\infty} \frac{\#\{0 \leq m \leq n : \ i_m = j_1, i_{m+1} = j_2, \ldots, i_{m+k-1} = j_k\}}{n} = 10^{-k}.$$

By the pointwise ergodic theorem (applied to the times 10 map on the circle), Lebesgue-a.e. x is normal.

OPEN: is $\pi - 3$ normal?

continued fraction expansions....

15 Invariant measures

If $T : X \to X$ is a continuous map of a topological space X then we say that a measure μ on X is **invariant** (or T-invariant) if $\mu(T^{-1}A) = \mu(A)$ for every Borel set $A \subset X$.

Most of ergodic theory is the study of measure-preserving transformations (sometime we also study measure-class preserving systems that do not preserve the measure). So there are two paths we could walk down here: we could study measure-preserving transformations of a measure space that has no topology on it or we could assume we are given a continuous map $T : X \to X$ of a topological space and study the invariant measures on X as a way of understanding the dynamics of T. We will take the second path now and the first path later on in these notes.

The first questions we ask are: does an invariant measure exist? if so is it unique? What does the set of all invariant measures 'look like'? (in other words, what sort of structure does it have?)

This last question is the key to answering the other questions. Let us assume that X is a compact Hausdorff space. Let $\mathcal{P}(X)$ denote the set of all Borel probability measures on X. We can think of $\mathcal{P}(X)$ as a subset of $C(X)^*$ which is the Banach dual of the space of continuous functions on X. As such it inherits the weak* topology. To be concrete, this means that a sequence μ_n in $\mathcal{P}(X)$ converges to a measure μ_∞ (wrt weak* topology) iff for every continuous function $f : X \to \mathbb{C}$, $\int f \, d\mu_n \to \int f \, d\mu_\infty$.

The Banach-Alaoglu Theorem states that the unit ball B in $C(X)^*$ is compact with respect to the weak* topology. The proof goes like this: we embed B into D^S (where $D \subset \mathbb{C}$ is the unit disk, $S \subset C(X)$ is the unit sphere) by $b \mapsto (f \mapsto b(f))$. By Tychonoff's Theorem, the latter space is compact (wrt the product topology). We check that this map is a homeomorphism onto its image which is a closed subspace.

Note that $\mathcal{P}(X) \subset C(X)^*$ is closed (this follows from the Riesz-Markov Theorem). So it's also weak* compact. Let $\mathcal{P}(X,T) \subset \mathcal{P}(X)$ be the invariant measures.

Theorem 15.1. *$\mathcal{P}(X,T)$ is nonempty.*

Proof. This follows from a more general result called the Markov-Kakutani fixed point theorem: if $A \subset V$ is a compact convex subset of a locally convex topological vector space (in our case $A = \mathcal{P}(X)$ and $V = C(X)^*$) and $f : A \to A$ is an affine homeomorphism, then there exists a fixed point. (Actually, the Markov-Kakutani theorem is a bit more general; we can replace f with any family of commuting affine homeomorphisms. The proof is essentially the same).

Let $x \in A$ and define $x_n = \frac{1}{n}\sum_{i=0}^{n-1} f^i x$. By compactness there exists a subsequence $\{n_i\}$ such that $x_\infty := \lim_{i\to\infty} x_{n_i}$ exists. By continuity,

$$f(x_\infty) = \lim_{i\to\infty} f(x_{n_i}) = \lim_{i\to\infty} x_{n_i} + (1/n_i)(f^{n_i}x - x) = x_\infty + \lim_{i\to\infty}(1/n_i)(f^{n_i}x - x).$$

If $\Phi \in V^*$ then Φ restricted to A is bounded (since A is compact). So

$$\Phi(\lim_{i\to\infty}(1/n_i)(f^{n_i}x - x)) = \lim_{i\to\infty} \Phi((1/n_i)(f^{n_i}x - x)) = 0.$$

Since this is true for every Φ, we must have $\lim_{i\to\infty}(1/n_i)(f^{n_i}x - x) = 0$ which implies x_∞ is a fixed point. $\square$

Remark: we have used here the Banach-Alaoglu Theorem which relies on Tychonoff's Theorem which uses the Axiom of Choice. So this proof is completely non-constructive. I don't think this can be helped.

16 Unique ergodicity

Theorem 16.1. *Let X be a compact metrizable space and $f : X \to X$ a homeomorphism. TFAE*

1. *there exists a unique f-invariant Borel probability measure on X,*
2. *for every continuous function $\mathcal{P} : X \to \mathbb{C}$ there is a constant $C_\mathcal{P}$ such that for every $x \in X$,*
$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{P}(f^i x) = C_\mathcal{P},$$
3. *there is a dense subset $\mathcal{D} \subset C(X)$ such that for every $\mathcal{P} \in \mathcal{D}$ there is a constant $C_\mathcal{P}$ such that for every $x \in X$,*
$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{P}(f^i x) = C_\mathcal{P}.$$

Proof. (2 implies 1): the main $\mathcal{P} \mapsto C_\mathcal{P}$ is a bounded positive linear functional on the Banach space $C(X)$. By the Riesz-Markov Theorem, there is a (unique) Borel probability measure μ such that $\int \mathcal{P}\, d\mu = C_\mathcal{P}$ for every $\mathcal{P} \in C(X)$. Clearly μ is f-invariant since $C_{\mathcal{P}\circ f} = C_\mathcal{P}$. Suppose ν is some other f-invariant Borel probability measure. By the ergodic theorem, we must have $\int \mathcal{P}\, d\mu = \int \mathcal{P}\, d\nu$ for every continuous $\mathcal{P}$. By the uniqueness part of the Riesz-Markov Theorem, $\mu = \nu$.

(1 implies 2): Let δ_x denote the Dirac probability measure concentrated on x. As in the proof of the Markov-Kakutani fixed point theorem any weak* limit point μ of

$$\frac{1}{n}\sum_{i=0}^{n-1} \delta_{f^i x}$$

is f-invariant. Since there is only one f-invariant measure, it must be that $\frac{1}{n}\sum_{i=0}^{n-1} \delta_{f^i x}$ converges weak* to μ. This implies (2).

(2 is equivalent to 3): by continuity.

$\square$

We say f is **uniquely ergodic** if the above conditions hold. For example: irrational rotations of the circle are uniquely ergodic. This can be proven using Fourier analysis (and #3 above). The shift on Σ_2 is not uniquely ergodic. Hyperbolic toral automorphisms are also not uniquely ergodic (however, there are very few ergodic measures on the torus). The horocyclic flow on a hyperbolic surface is uniquely ergodic (by Furstenberg). The geodesic flow is not.

16.1 Topological models

Definition 7. A **(compact) topological model** for a pmp system $T : (X, \mu) \to (X, \mu)$ consists of a compact Hausdorff space Y, a Borel probability measure ν on Y, and a measure-preserving homeomorphism $\Phi : Y \to Y$ such that f is measurably conjugate to Φ.

Lemma 16.2. *Any pmp system on a standard probability space admits a topological model.*

Proof. (An easy proof) Wlog we may assume $X = [0, 1]$. Let $Y = [0, 1]^{\mathbb{Z}}$ with the product topology and shift action. Let $S : [0, 1] \to [0, 1]^{\mathbb{Z}}$ be the map $S(x)_n = T^n x$. Let $\nu = S_* \mu$.

(a C* algebra proof) ...

□

Theorem 16.3 (Jewett-Krieger)**.** *Any ergodic pmp system admits a uniquely ergodic topological model.*

This result was quite shocking when it was first proven. Many researchers thought the opposite would be true (consider Bernoulli shifts for example).

17 Pure point spectrum

Let $T : (X, \mu) \to (X, \mu)$ be measure-preserving. The **Koopman operator** $U_T : L^2(X) \to L^2(X)$ is defined by $U_T(f) = f \circ T$. If T is invertible then U_T is unitary (which means it preserves the inner product).

A function $f \in L^2(X)$ is an *eigenfunction* if $U_T f = \alpha f$ for some $\alpha \in \mathbb{C}$, called the **eigenvalue**.

Lemma 17.1. *Suppose T is ergodic. If f is a nonconstant eigenfunction of U_T then $|\alpha| = 1$ and $|f|$ is a.e. constant. Moreover the set of all eigenvalues forms a countable subgroup of the circle and each eigenvalue is simple.*

Proof. Note that $|f|$ is also an eigenfunction with eigenvalue $|\alpha|$. Since U_T is an isometry, $|\alpha| = 1$. Since T is ergodic, $|f|$ is a.e. constant. If f, g are eigenfunctions with eigenvalue α then f/g has eigenvalue 1. By ergodicity, f/g must be constant a.e.

Note is an eigenfunction with eigenvalue $\bar{\alpha} = \alpha^{-1}$. If f, g are eigenfunctions with eigenvalues α, β then fg is an eigenfunction with eigenvalue $\alpha\beta$. □

Definition 8. We say T has **pure point spectrum** if there is an ON basis in $L^2(X)$ consisting of eigenfunctions for U_T. For example, irrational rotations of the circle have pure point spectrum.

For another example: if T is mixing then U_T has no eigenfunctions. We will strengthen this observation later.

Theorem 17.2 (Halmos-von Neumann)**.** *If T is ergodic and has pure point spectrum then there is a compact abelian group G an element $g_0 \in G$ such that T is measurably conjugate to $L_{g_0} : G \to G$ wrt Haar measure on G. Moreover any two ergodic pmp transformations with pure point spectrum are measurably conjugate iff they have the same set of eigenvalues.*

Proof. Let K be the subgroup of the circle generated by the eigenvalues of U_T. For each $k \in K$, let $f_k \in L^2(X)$ be a corresponding eigenfunction. After scaling if necessary, we may assume $|f_k| = 1$ a.e.

Note that $\{f_k\}$ is an ON basis. To see this, observe that if $k \neq l, k, l \in K$ then

$$\langle f_k, f_l \rangle = \langle U_T f_k, U_T f_l \rangle = k\bar{l}\langle f_k, f_l \rangle$$

implies $f_k \perp f_l$. So they are orthonormal. They form a basis since eigenvalues are simple and T has pure point spectrum.

For $k, l \in K$, $f_k f_l$ has eigenvalue kl. So wlog we may choose the f_k's so that $f_k f_l = f_{kl}$ and $f_{k^{-1}} = f_k^{-1}$. In other words, the map $k \mapsto f_k$ is a group isomorphism. We can also assume that the f_k are defined everywhere and satisfy $f_k \circ T = k f_k$ everywhere (by removing a set of measure zero if necessary).

For $x \in X$, define $\mathcal{P}_x : K \to \mathbb{C}$ by $\mathcal{P}_x(k) = f_k(x)$. Note that $\mathcal{P}_x(kl) = f_{kl}(x) = f_k(x)f_l(x) = \mathcal{P}_x(k)\mathcal{P}_x(l)$. Thus $\mathcal{P}_x$ is a homomorphism from K into the unit circle. Let $G = \mathrm{Hom}(K, S^1)$. We have shown $\mathcal{P}_x \in G$ and $\mathcal{P}$ gives us a map $\Phi : X \to G$ by $\Phi(x) = \mathcal{P}_x$.

$\mathcal{P}_{Tx}(k) = f_k(Tx) = k f_k(x) = k\mathcal{P}_x(k)$. So if $g_0 \in G$ is the element $g_0(k) = k$ then $\Phi T = L_{g_0}\Phi$.

Because K is canonically isomorphic with $\mathrm{Hom}(G, S^1)$, we can likewise find an ON basis for $L^2(G)$ indexed by K. A routine computation shows that $\Phi^* : L^2(G) \to L^2(X)$ (defined by $\Phi^*(\psi) = \psi\Phi$) maps this ON basis bijectively to the one for X. This implies Φ is measure-preserving and invertible.

□

18 Weak mixing

(borrowed from Petersen's book)

Definition 9. Let $T \in \mathrm{Aut}(X, \mu)$. T is

- **ergodic** if $\forall A, B \subset X$, $\lim_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N-1} \mu(T^{-i}A \cap B) = \mu(A)\mu(B)$
- **weakly mixing** if $\forall A, B \subset X$, $\lim_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0$

- **strongly mixing** if $\forall A, B \subset X$, $\lim_{N\to\infty} |\mu(T^{-N}A \cap B) - \mu(A)\mu(B)| = 0$

Theorem 18.1. *Let $T \in \mathrm{Aut}(X, \mu)$. TFAE*

1. *T is weakly mixing*
2. $\lim_n \frac{1}{n}\sum_{k=0}^{n-1} |\langle U_T^n f, g\rangle - (\int f)\overline{(\int g)}| = 0$ *for every $f, g \in L^2(X)$*
3. *given $A, B \subset X$ there is a set $J \subset \mathbb{N}$ with upper Banach density equal to 0 such that*
$$\lim_{n\to\infty, n\neq J} |\mu(T^n A \cap B) - \mu(A)\mu(B)| = 0.$$
4. *$T \times T$ is weakly mixing*
5. *$T \times S$ is ergodic for every ergodic pmp S*
6. *$T \times T$ is ergodic*
7. *T has no nonconstant eigenfunctions.*

(1 is equivalent to 2): this is just an approximation argument (we can approximate any function in L^2 by simple functions)

To prove (1 and 3 are equivalent) we need:

Lemma 18.2 (Koopman-von Neumann)**.** *Let $f : \mathbb{N} \to [0, \infty)$ be bounded. Then* $\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n} f(m) = 0$ *iff there exists a set $E \subset \mathbb{N}$ of density zero with* $\lim_{n\to\infty, n\notin E} f(n) = 0$.

Proof. If there exists such a set E then it is clear.

So suppose $\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n} f(m) = 0$. After scaling f if necessary, we may assume wlog that $f \leq 1$. Let $E_m = \{k \in \mathbb{N} : f(k) > 1/m\}$. Each E_m has density zero. Choose $0 = i_0 < i_1 < i_2 < ...$ such that
$$\frac{1}{n}\sum_{k=0}^{n} \chi_{E_m}(k) < 1/m$$
for $n \geq i_{m-1}$. Define
$$E = \bigcup_m E_m \cap (i_{m-1}, i_m].$$
Notice $f(k) < 1/m$ if $k \notin E$ and $k \in (i_{m-1}, i_m]$. So $\lim_{n\to\infty, n\notin E} f(n) = 0$.

Note if $i_{m-1} < n \leq i_m$ then
$$\frac{1}{n}\sum_{k=0}^{n-1} \chi_E(k) \leq \frac{1}{n}\sum_{k=0}^{n-1} \chi_{E_m}(k) < 1/m.$$

So E has density zero.

□

It's now clear that (1 and 3 are equivalent).

(3 implies 4) is a straightforward computation using the fact that the union of two sets of density zero is a set with density zero.

(4 implies 5) is straightforward (using the definition of ergodicity above).

(5 impies 6) is immediate.

(6 implies 3) Prove

$$\lim_n \frac{1}{n}\sum_{k=0}^{n-1}[\mu(T^kA\cap B)-\mu(A)\mu(B)]^2=0.$$

The Koopman-von Neumann Lemma now implies (3).

(6 implies 7): If f is an eigenfunction of T then let $g(x,y)=f(x)(\mathbf{y})$ be a function on $X\times X$. g has eigenvalue 1. Since $T\times T$ is ergodic, g is constant a.e. This implies f is constant a.e.

(7 implies 1): Let $f\in L^2_0(X)$, $g\in L^2(X)$. Let $E_{f,g}$ be the corresponding spectral measure. Let E be the projection valued measure. So $E_{f,g}(A)=\langle E(A)f,g\rangle$.

Claim: Since T has no nonconstant eigenfunctions, $E_{f,g}$ has no atoms.

To see this $\lambda_0\in\mathbb{C}$ and observe that

$$U_TE(\{\lambda_0\})f=\int_{\{\lambda_0\}}\lambda dE(\lambda)f=\lambda_0E(\{\lambda_0\})f$$

is an eigenvector with eigenvalue λ_0. Since there are no nonconstant eigenfunctions and since f is orthogonal to the constants, $E(\{\lambda_0\})f=0$. It follows that $E_{f,g}(\{\lambda_0\})=\langle E(\{\lambda_0\})f,g\rangle=0$.

Now we compute:

$$\begin{aligned}\frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^if,g\rangle|^2 &= \frac{1}{n}\sum_{i=0}^{n-1}|\int\lambda^i\ dE_{f,g}(\lambda)|^2\\ &= \frac{1}{n}\sum_{i=0}^{n-1}\left(\int\lambda^i\ dE_{f,g}(\lambda)\right)\overline{\left(\int\lambda^i\ dE_{f,g}(\lambda)\right)}\\ &= \frac{1}{n}\sum_{i=0}^{n-1}\iint\lambda^i\overline{\zeta}^i\ dE_{f,g}(\lambda)dE_{f,g}(\zeta)\\ &= \iint\frac{1}{n}\sum_{i=0}^{n-1}\lambda^i\overline{\zeta}^i\ dE_{f,g}(\lambda)dE_{f,g}(\zeta)\\ &= \frac{1}{n}\iint\frac{1-(\lambda\overline{\zeta})^n}{1-\lambda\overline{\zeta}}\ dE_{f,g}(\lambda)dE_{f,g}(\zeta).\end{aligned}$$

We used the claim in the last line. We observe that the integrand is bounded. So we can apply the bounded convergence theorem to obtain that this integral tends to zero as $n\to\infty$. The result now follows from the Koopman-von Neumann Lemma.

18.1 Notable results related to weak mixing

Halmos proved that weak mixing is generic in $\mathrm{Aut}(X,\mu)$. By contrast, mixing is meager. So there exist transformations that are weak mixing but not mixing.

There is a recent paper in the Annals (by Avila and Forni) showing that 'most' IETs are weakly mixing. It is known that they are not mixing.

A key step in Furstenberg's proof of Szemeredi's Theorem is the special case in which T is weakly mixing.

19 Spectral theory

Let $T:(X,\mu)\to(X,\mu)$ be measure-preserving. The **Koopman operator** $U_T: L^2(X)\to L^2(X)$ is defined by $U_T(f)=f\circ T$. If T is invertible then U_T is unitary (which means it preserves the inner product).

We say two measure-preserving transformations T_1, T_2 are **spectrally isomorphic** if there is an isometry $\Phi: L^2(X_1)\to L^2(X_2)$ such that $\Phi U_{T_1} = U_{T_2}\Phi$. In other words, if their Koopman operators are unitarily equivalent.

Intuitively, we think of Hilbert space as a kind of infinite-dimensional Euclidean space. In finite-dimensional Euclidean space an invertible linear map is characterized up to unitary equivalence by its eigenvalues. So we might expect something of the sort here. The main difference is that, well, there might not be any eigenvectors in the usual sense and instead of a set of numbers we'll get a measure.

We say operators $U_1: H_1\to H_1, U_2: H_2\to H_2$ (on Hilbert spaces H_1, H_2) are **unitarily equivalent** if there is a linear isometry $\Phi: H_1\to H_2$ such that $\Phi U_1 = U_2\Phi$.

The **spectrum** of an operator $A: H\to H$ is the set of all $\lambda\in\mathbb{C}$ such that $A-\lambda I$ is does not have a bounded inverse. It is denoted by $\sigma(A)$. For example, suppose H has an ON basis $\{e_i\}_{i\in\mathbb{N}}$ and A is defined by $Ae_i=\lambda_i e_i$ where λ_i's are some complex numbers with $\lambda_i\to 0$ as $i\to\infty$. If $\lambda_i\neq 0$ for all i then A has trivial kernel but A does not have a bounded inverse.

Here's another example: suppose $\mathcal{P}\in L^\infty(X,\mu)$ where (X,μ) is a probability space. Define $M_{\mathcal{P}}: L^2(X,\mu)\to L^2(X,\mu)$ by $M_{\mathcal{P}}(f)(z)=\mathcal{P}(z)f(z)$. This is a **multiplication operator**. Let's assume that μ has no atoms.

Lemma 19.1. *$M_{\mathcal{P}}$ has a nonzero eigenvector with eigenvalue λ if and only if $\mu(\mathcal{P}^{-1}(\lambda))>0$. However, $\sigma(M_{\mathcal{P}})$ is the essential range of $\mathcal{P}$. In other words, $t\in\sigma(M_{\mathcal{P}})$ if and only if: for every $\epsilon>0$,*

$$\mu(\{x\in X:\ \mathcal{P}(x)\in[t-\epsilon,t+\epsilon]\})>0.$$

An example that will come up later: let $\mathcal{P}:\mathbb{T}\to\mathbb{T}$ be the identity map and consider $M_{\mathcal{P}}: L^2(\mathbb{T})\to L^2(\mathbb{T})$ (with respect to Lebesgue measure). Then $\sigma(M_{\mathcal{P}})=\mathbb{T}$ but $M_{\mathcal{P}}$ has no eigenvectors.

19.1 The spectral theorem for normal operators

An operator is normal if it commutes with its adjoint. Let $U : H \to H$ be normal.

1. There exists a finite measure space (X, μ) and a function $\mathcal{P} \in L^\infty(X, \mu)$ such that U is unitarily equivalent to $M_{\mathcal{P}}$.

2. There are Borel measures $\sigma_1 >> \sigma_2 >> \cdots$ on $\mathbb{C}$ such that U is unitarily equivalent to the operator S on $\oplus_i L^2(\mathbb{C}, \sigma_i)$ defined by $S(f_1(z_1), f_2(z_2), \ldots) = (z_1 f(z_1), z_2 f(z_2), \ldots)$. The measures σ_i are uniquely determined up to absolute continuity-equivalence.

3. There is a unique projection-valued measure E on $\mathbb{C}$ such that $U = \int \lambda \, dE(\lambda)$. The support of E is $\sigma(U)$. If $F : \sigma(U) \to \mathbb{C}$ is any bounded measurable function then we can define
$$F(U) = \int F(\lambda) \, dE(\lambda).$$
For example,
$$U^k = \int \lambda^k \, dE(\lambda).$$
For any $v, w \in H$, we obtain a complex-valued measure $E_{v,w}$ on $\mathbb{C}$ by $E_{v,w}(A) = \langle E(A)v, w \rangle$. This measure is sometimes denoted as $d\langle E(\lambda)v, w\rangle$. We observe that
$$\langle U^k v, w \rangle = \int \lambda^k \, d\langle E(\lambda)v, w\rangle.$$

In our case, we apply the spectral theorem to U_T. Because T is measure-preserving, U_T is unitary (so $U_T^{-1} = U_T^*$) and therefore normal. This also implies that $\sigma(T)$ is contained in the unit circle.

19.2 Examples

Consider $L^2(\mathbb{T}, \text{Leb})$. For $n \in \mathbb{Z}$, let $f_n \in L^2(\mathbb{T})$ be the function $f_n(z) = z^n$. By Fourier analysis, the f_n's form an ON-basis for $L^2(\mathbb{T})$. Let $U : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ be the operator $(Uf)(z) = zf(z)$. Note $Uf_n = f_{n+1}$. So U is unitarily equivalent to the shift operator on $\ell^2(\mathbb{Z})$. Note also that the spectrum of U is the circle, the spectral measure is Lebesgue measure. (This is almost a tautology). It has simple spectrum (this means the multiplicity function is $\equiv 1$).

Consider the Bernoulli shift σ on $(\{0,1\}, u_2)^{\mathbb{Z}}$. If $S \subset \mathbb{Z}$ is finite, let $f_S(x) = \prod_{s \in S} (-1)^{x_s}$. Then $\{f_S\}$ forms an ON basis of $L^2(\{0,1\}^{\mathbb{Z}}, u_2^{\mathbb{Z}})$ showing that this Bernoulli has Lebesgue spectrum of infinite multiplicity.

Now consider the Bernoulli shift $\sigma : (K, \kappa)^{\mathbb{Z}} \to (K, \kappa)^{\mathbb{Z}}$. Let $\{f_j\}_{j \in \mathbb{N}}$ be an ON basis for $L^2(K, \kappa)$. If $W \subset \mathbb{Z}$ is finite and $\psi : W \to \mathbb{N}$ then define $\Phi_\psi \in L^2((K, \kappa)^{\mathbb{Z}})$ by $\Phi_\psi(x) = \prod_{w \in W} f_{\psi(w)}(x(w))$.

Observe that the Φ_ψ's form an ON basis for $L^2((K, \kappa)^{\mathbb{Z}})$ that is permuted by the shift action. It follows that U_σ is unitarily equivalent to a subrepresentation of the countable sum of regular representations. It therefore has Lebesgue spectrum of infinite multiplicity.

19.3 Notable results/questions on spectral theory

It is, in general, a wide-open problem to classify all all possibly operators of the form U_T up to unitary equivalence.

Observe that specifying the spectral isomorphism class of U_T is equivalent to specify σ_1 up to equivalence (called the maximal spectral type) and the sets $A_i \subset \sigma(T)$ defined by $A_i = \{x \in \mathbb{C} : \frac{d\sigma_i}{d\sigma_1}(x) > 0\}$. Let $M(T) = \sum_{i=1}^{\infty} \chi_{A_i}$. This is the spectral multipicity function. U_T is determined (up to unitary equivalence) by its spectral type and multiplicity function.

This leads to some natural questions: what is the essential range of $M(T)$ for ergodic T? It is shown by Kwiatkowski-Lemancyk (1995) and Ageev (2001) that all subsets of $\mathbb{N} \cup \{\infty\}$ containing $\{1\}$ occur.

Another question in this area: consider U_T restricted to $L_0^2(X)$. Assume ergodicity. What is the essential range of $M(T)$ now? Ageev showed that $\{n\}$ is possible (for any $n > 1$) in 2005.

An open problem, due to Banach, asks whether U_T restricted to $L_0^2(X)$ can be such that σ_1 is equivalent to Lebesgue measure on the circle and $M(T) = 1$. This might be the oldest open problem in ergodic theory. (In this case, T is said to have simple Lebesgue spectrum in the orthogonal complement of the constants).

20 Ergodic decomposition theorem

Corollary 20.1. *Let $T : X \to X$ be a continuous map of a compact metric space. A measure $\mu \in \mathcal{P}(X, T)$ is ergodic if and only if it is extreme.*

Lemma 20.2. *The extreme point of any convex closed subset A of a locally convex topological vector space V are a G_δ. In particular, it is Borel.*

Theorem 20.3 (Ergodic Decomposition Theorem)**.** *Let $T : X \to X$ be a Borel transformation of a standard Borel space. Let μ be a Borel invariant probability measure. Then there exists a unique measure ω on the Borel space of ergodic T-invariant measures such that*

$$\mu = \int \nu \, d\omega(\nu).$$

Proof. The Krein-Milman Theorem states that every closed convex set is the convex hull of its extreme points. This proves existence.

I'll use Sarig's notes (pages 42-43) to finish the proof. □

One consequence of this theorem: suppose $T : X \to X$ is a homeomorphism of a compact metric space. Let $\mathcal{P}_T(X)$ be the space of T-invariant Borel probability measures. Then $\mathcal{P}_T(X)$ is a **simplex**: it is a closed convex subspace of a locally convex topological vector space with the property that for every $\mu \in \mathcal{P}_T(X)$ there exists a unique measure κ on the set of extreme points of $\mathcal{P}_T(X)$ such that $\mu = \int \nu \, d\kappa(\nu)$.

A curious example: a simplex is **Poulsen** if its extreme points are dense. It turns out that there is only one (separable) Poulsen simplex up to affine homeomorphism. Moreover, its extreme points are homeomorphic to ℓ^2. Consider the set $\mathcal{P}_\sigma(\{0,1\}^{\mathbb{Z}})$ of shift-invariant Borel probability measures on $\{0,1\}^{\mathbb{Z}}$. Show that this is a Poulsen simplex. Hint: show that in fact the shift-invariant measures that are supported on single finite orbits of the shift are already dense.

21 Entropy theory (rough)

outline:

information theory, Kolmogorov-Sinai Theorem, examples: Bernoulli shifts, Markov chains, rotations, more?

Shannon-McMillan Theorem,

basic properties: factors, direct products, erg. decomposition,

Variational principle

when is there a unique measure of maximal entropy?

a touch of Ornstein theory

22 Entropy Theory

Suppose (X,μ) is a standard probability space and $x \in X$ is chosen at random. Let $E \subset X$. If we are told that $x \in E$ then intuitively we have gained a certain amount of information. Let us try to make this intuitive concept precise.

Let $I(E)$ be the amount of information we gain by being told that $x \in E$. Suppose that $E, F \subset X$ are *independent events*. This means that $\mu(E \cap F) = \mu(E)\mu(F)$. For example, if $X = [0,1]^2$ and E is determined by the first coordinate while F is determined by the second coordinate. In this case, intuitively, we should have $I(E \cap F) = I(E) + I(F)$.

Now $I(E)$ should depend only on $\mu(E)$. So we can consider it to be a function from the interval $[0,1]$ to $[0,+\infty]$. The condition above implies that $I(ts) = I(t) + I(s)$ for any $t, s \in [0,1]$. We should also have $I(1) = 0$. So I should be the logarithmic function $I(t) = -\log_b(t)$ for some base $b > 1$. Information theorists take $b = 2$ and mathematicians tend to use $b = e$. In practice, it matters very little since it is easy to convert one to the other.

Let $\mathcal{P}$ be a partition of X into finitely many or countably many measurable subsets. For $x \in X$, let $\mathcal{P}(x) \in \mathcal{P}$ be the partition element containing x. So $\mathcal{P}(x) \subset X$. Let $I_{\mathcal{P}} : X \to \mathbb{R}$ be the information function defined by $I_{\mathcal{P}}(x) = -\log \mu(\mathcal{P}(x))$ is the amount of information gained by learning which part of $\mathcal{P}$ contains x.

The *Shannon entropy* of $\mathcal{P}$ is the average

$$H_\mu(\mathcal{P}) = \int I_{\mathcal{P}}(x)\ d\mu(x) = -\sum_{P \in \mathcal{P}} \mu(P) \log \mu(P).$$

...discussion about $\binom{N}{p_1N, p_2N, \ldots, p_kN} \sim e^{NH(p_1,\ldots,p_k)}$...

The join of two partitions...

Let $T \in \mathrm{Aut}(X,\mu)$.

The *entropy rate* of $\mathcal{P}$ is defined by

$$h(T,\mathcal{P}) := h_\mu(T,\mathcal{P}) := \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}\right).$$

The *Kolmogorov-Sinai entropy* of the action $T \curvearrowright (X,\mathcal{B},\mu)$ is $h(T) = h_\mu(T) := \sup h(T,\mathcal{P})$ where the supremum is over all finite partitions $\mathcal{P}$. It is obviously an invariant of measure-conjugacy. This fact by itself is not particularly enlightening; we need tools to compute the entropy. The first such tool is presented is Kolmogorov-Sinai's Theorem 24.1 below. To prove it, we need to know about relative entropy.

23 Relative entropy

Let $\mathcal{F}$ be a Borel sigma-algebra. An important case to keep in mind is when $\mathcal{F}$ is generated by a finite partition (in which case $\mathcal{F}$ is finite). $I_{\mathcal{P}}(x|\mathcal{F})$ will be the amount of information that we gain by learning which part of $\mathcal{P}$ contains x *given* that we already know which subsets of $\mathcal{F}$ contain x. More formally,

$$I_{\mathcal{P}}(x|\mathcal{F}) = -\log \mathbb{E}[\chi_{\mathcal{P}(x)}|\mathcal{F}](x)$$

where $\chi_{\mathcal{P}(x)}$ is the characteristic function of x and $\mathbb{E}[\chi_{\mathcal{P}(x)}|\mathcal{F}]$ is its conditional expectation on $\mathcal{F}$. In particular, if $\mathcal{F}$ is generated by a finite or countable partition $\mathcal{Q}$ then

$$I_{\mathcal{P}}(x|\mathcal{F}) = -\log \frac{\mu(\mathcal{P}(x)\cap\mathcal{Q}(x))}{\mu(\mathcal{Q}(x))}.$$

In this case we also write $I_{\mathcal{P}}(x|\mathcal{F}) = I_{\mathcal{P}}(x|\mathcal{Q})$. In general, we will not distinguish between partitions and the sigma-algebras they generate when it is not particularly important to do so.

The Shannon entropy of $\mathcal{P}$ *relative* to $\mathcal{F}$ is $H(\mathcal{P}|\mathcal{F}) = \int I_{\mathcal{P}}(x|\mathcal{F})\, d\mu(x)$. In the special case $\mathcal{F} = \sigma(\mathcal{Q})$,

$$\begin{aligned} H(\mathcal{P}|\mathcal{Q}) &= -\sum_{P\in\mathcal{P}}\sum_{Q\in\mathcal{Q}} \mu(P\cap Q)\log\left(\frac{\mu(P\cap Q)}{\mu(Q)}\right) \\ &= H(\mathcal{P}\vee\mathcal{Q}) - H(\mathcal{Q}). \end{aligned}$$

We will first obtain a few basic formulas concerning limits and joins of partitions. We will need:

Theorem 23.1 (Martingale Convergence Theorem)**.** *Suppose* $(\Omega,\mathcal{B},\mu)$ *is a standard probability space (where* $\mathcal{B}$ *is the sigma-algebra of measurable sets). Let* $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \mathcal{B}$ *be an increasing sequence of sigma-sub-algebras. Suppose we have random variables* $X_i : \Omega \to \mathbb{C}$ *such that* X_i *is* $\mathcal{F}_i$*-measurable and* $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \ge X_n$ *a.e.. (In this case* $\{X_n\}$ *is called a submartingale). If* $\sup_n \mathbb{E}[|X_n|] < \infty$ *then* X_n *converges pointwise a.e. and in* L^1 *as* $n\to\infty$*.*

(The martingale convergence theorem is proven in pages 103-107 of Petersen).

(I used Glasner's book to make the notes below)

Lemma 23.2 (Chung). *If* $\{\mathcal{F}_i\}_{i=1}^\infty$ *is an increasing sequence of sigma-algebras contained in* $\mathcal{B}$, $\mathcal{P}$ *is a partition of* X *with finite Shannon entropy and* $f(x) = \sup_n I_\mathcal{P}(x|\mathcal{F}_n)$ *then* f *is an* L^1*-function.*

Proof. Fix $t > 0$ and set

$$B = \{x \in X : \ f(x) > t\}.$$

Note that $f(x) > t$ if and only if $\mathbb{E}[1_{\mathcal{P}(x)}|\mathcal{F}_n](x) < e^{-t}$ for some n. Fix $P \in \mathcal{P}$. Define

$$B_n = \{x \in X : \ \mathbb{E}[1_{\mathcal{P}(x)}|\mathcal{F}_k](x) \ge e^{-t} \forall 1 \le k < n, \text{ but } \mathbb{E}[1_{\mathcal{P}(x)}|\mathcal{F}_n](x) < e^{-t}\}.$$

So $B \cap P$ is the disjoint union of $B_n \cap P$ over all $n = 1, 2, \ldots$.

Since $B_n \in \mathcal{F}_n$,

$$\mu(B_n \cap P) = \int_{B_n} 1_P \ d\mu = \int_{B_n} \mathbb{E}[1_{\mathcal{P}(x)}|\mathcal{F}_n](x) \ d\mu(x) < e^{-t}\mu(B_n).$$

Summing over all n we obtain $\mu(B \cap P) \le e^{-t}$. So

$$\mu(B) \le \sum_{P \in \mathcal{P}} \min(\mu(P), e^{-t}).$$

Now:

$$\begin{aligned}
\int f(x) \ d\mu &= \int \int_0^{f(x)} 1 \ dt d\mu(x) \\
&\le \int_0^\infty \mu(\{x \in X : \ f(x) > t\}) \ dt \\
&\le \int_0^\infty \sum_{P \in \mathcal{P}} \min(\mu(P), e^{-t}) \ dt \\
&= \sum_{P \in \mathcal{P}} \int_0^\infty \min(\mu(P), e^{-t}) \ dt \\
&= \sum_{P \in \mathcal{P}} \int_0^{-\log \mu(P)} \mu(P) \ dt + \int_{-\log(\mu(P))}^\infty e^{-t} \ dt \\
&= H_\mu(\mathcal{P}) + 1.
\end{aligned}$$

□

Given two sigma-algebras $\mathcal{F}_1, \mathcal{F}_2$ we let $\mathcal{F}_1 \vee \mathcal{F}_2$ denote the smallest sigma-algebra containing both.

Lemma 23.3. *If* $\{\mathcal{F}_i\}_{i=1}^{\infty}$ *is an increasing sequence of sigma-algebras contained in* $\mathcal{B}$ *and* $\mathcal{F}_\infty = \bigvee_{i=1}^{\infty} \mathcal{F}_i$ *then* $I_{\mathcal{P}}(\cdot|\mathcal{F}_n)$ *converges pointwise a.e. to* $I_{\mathcal{P}}(\cdot|\mathcal{F}_\infty)$ *as* $n \to \infty$*. Thus*

$$\lim_{n\to\infty} H(\mathcal{P}|\mathcal{F}_n) = H(\mathcal{P}|\mathcal{F}_\infty).$$

Proof. Let $Y \subset X$ be measurable. Notice that $\mathbb{E}[\mathbb{E}[1_Y|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[1_Y|\mathcal{F}_n]$ a.e. for every n. This means that $\{\mathbb{E}[1_Y|\mathcal{F}_n]\}_{n=1}^{\infty}$ is a martingale with respect to $\{\mathcal{F}_n\}_{n=1}^{\infty}$. By the Martingale Convergence Theorem, $\mathbb{E}[1_Y|\mathcal{F}_n]$ converges pointwise a.e. to $\mathbb{E}[1_Y|\mathcal{F}_\infty]$ as $n \to \infty$. So $I_{\mathcal{P}}(\cdot|\mathcal{F}_n)$ converges pointwise a.e. to $I_{\mathcal{P}}(\cdot|\mathcal{F}_\infty)$ as $n \to \infty$.

By the previous lemma, $I_{\mathcal{P}}(\cdot|\mathcal{F}_n)$ is dominated by an L^1-function. So the Dominated Convergence Theorem implies $\lim_{n\to\infty} H(\mathcal{P}|\mathcal{F}_n) = H(\mathcal{P}|\mathcal{F}_\infty)$.

□

Lemma 23.4. *For any* $\mathcal{P}, \mathcal{Q}, \mathcal{F}$

1. $0 \le H(\mathcal{P}|\mathcal{F}) \le H(\mathcal{P})$,
2. $H(\mathcal{P} \vee \mathcal{Q}) \le H(\mathcal{P}) + H(\mathcal{Q})$.

Proof. Using the previous lemma, we can reduce to the case in which $\mathcal{F}$ is generated by a finite partition $\mathcal{Q}$.

Define $\eta : [0,1] \to \mathbb{R}$ by $\eta(x) = -x\log(x)$. Then η is concave. In particular, if $t_i, p_i \in [0,1]$ (for $i = 1...n$) and $\sum p_i = 1$ then $\eta(\sum_{i=1}^n p_i t_i) \ge \sum_i p_i \eta(t_i)$. Because η is concave,

$$\begin{aligned}
H(\mathcal{P}|\mathcal{Q}) &:= \int I_{\mathcal{P}}(x|\mathcal{Q})\, d\mu(x) \\
&= -\sum_{P\in\mathcal{P}}\sum_{Q\in\mathcal{Q}} \mu(P\cap Q)\log\left(\frac{\mu(P\cap Q)}{\mu(Q)}\right) \\
&= -\sum_{P\in\mathcal{P}}\sum_{Q\in\mathcal{Q}} \mu(Q)\frac{\mu(P\cap Q)}{\mu(Q)}\log\left(\frac{\mu(P\cap Q)}{\mu(Q)}\right) \\
&= \sum_{P\in\mathcal{P}}\sum_{Q\in\mathcal{Q}} \mu(Q)\eta\left(\frac{\mu(P\cap Q)}{\mu(Q)}\right) \\
&\le \sum_{P\in\mathcal{P}} \eta\left(\sum_{Q\in\mathcal{Q}} \mu(Q)\frac{\mu(P\cap Q)}{\mu(Q)}\right) \\
&\le \sum_{P\in\mathcal{P}} \eta\left(\mu(P)\right) = H(\mathcal{P}).
\end{aligned}$$

Next,

$$H(\mathcal{P}\vee\mathcal{Q}) = H(\mathcal{Q}) + H(\mathcal{P}|\mathcal{Q}) \le H(\mathcal{Q}) + H(\mathcal{P}).$$

□

Exercise 10. Prove the relative formulas:

1. $H(\mathcal{P}|\mathcal{Q} \vee \mathcal{F}) = H(\mathcal{P} \vee \mathcal{Q}|\mathcal{F}) - H(\mathcal{Q}|\mathcal{F})$.
2. $0 \leq H(\mathcal{P}|\mathcal{F}_1 \vee \mathcal{F}_2) \leq H(\mathcal{P}|\mathcal{F}_2)$,
3. $H(\mathcal{P} \vee \mathcal{Q}|\mathcal{F}) \leq H(\mathcal{P}|\mathcal{F}) + H(\mathcal{Q}|\mathcal{F})$.

Exercise 11. How do we know the limit in the definition of $h(T, \mathcal{P})$ exists?

1. Let $\mathcal{P}^{(n)} = \bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}$. Show that $H(\mathcal{P}^{(n+m)}) \leq H(\mathcal{P}^{(n)}) + H(\mathcal{P}^{(m)})$.
2. A sequence $\{h_n\}_{n=1}^{\infty}$ is *sub-additive* if $h_{n+m} \leq h_n + h_m$ for all $n, m \geq 0$. Show that if $\{h_n\}_{n=1}^{\infty}$ is sub-additive then $\lim_{n\to\infty} \frac{h_n}{n}$ exists.
3. Conclude that $h(T, \mathcal{P})$ is well-defined.

24 Kolmogorov's Theorem

Definition 10. We say a sub-sigma-algebra $\mathcal{F}$ is **T-invariant** if $T^{-1}A \in \mathcal{F}$ for every $A \in \mathcal{F}$. If T is invertible then we require also require $TA \in \mathcal{F}$ for every $A \in \mathcal{F}$.

Definition 11. Let $\sigma(T, \mathcal{P})$ be the smallest T-invariant sigma-algebra containing $\mathcal{P}$. We say $\mathcal{P}$ is **generating** for $T \in \mathrm{Aut}(X, \mu)$ if $\sigma(T, \mathcal{P})$ is the sigma-algebra of all measurable sets (up to sets of measure zero).

Examples: Bernoulli shifts.

What this has to do with factor maps.

Given a partition $\mathcal{P}$ we obtain a factor onto a symbolic system via $\pi : X \to \mathcal{P}^{\mathbb{Z}}$ where $\pi(x)_i = P$ if $T^i x \in P$. Then $\mathcal{P}$ is generating for T iff π is a measure-conjugacy.

Theorem 24.1 (Kolmogorov-Sinai Theorem (1959))**.** *Suppose $\mathcal{P}, \mathcal{Q}$ are two partitions and $\sigma(T, \mathcal{P}) \subset \sigma(T, \mathcal{Q})$. Then $h(T, \mathcal{P}) \leq h(T, \mathcal{Q})$. In particular, if $\mathcal{Q}$ is generating then $h(\mathcal{Q}) = h(T)$.*

For $n < m$ let $\mathcal{P}_n^m = \bigvee_{i=n}^{m} T^{-i}\mathcal{P}$. We allow $m = \infty$, in which case we interpret $\mathcal{P}_n^m$ to be a sigma-algebra.

Theorem 24.2. *Let $T \in \mathrm{Aut}(X, \mu)$ and $\mathcal{P}$ be a partition with finite Shannon entropy. Then*

$$h(T, \mathcal{P}) \quad = \quad H\left(\mathcal{P}|\mathcal{P}_1^{\infty}\right).$$

Proof.

$$H(\mathcal{P}_0^n) = H(\mathcal{P}_0|\mathcal{P}_1^n) + H(\mathcal{P}_1^n).$$

Because μ is T-invariant, $H(T^{-1}\mathcal{R}) = H(\mathcal{R})$ for any partition $\mathcal{R}$. So $H(\mathcal{P}_1^n) = H(\mathcal{P}_0^{n-1})$. An inductive argument now implies

$$H(\mathcal{P}_0^n) = H(\mathcal{P}) + \sum_{k=1}^{n} H(\mathcal{P}|\mathcal{P}_1^k).$$

The statement now follows from Lemma 23.3. $\square$

Corollary 24.3. *For any $n > 0$, $h(T, \mathcal{P}) = h(T, \mathcal{P}_0^n)$.*

Proof.

$$h(T, \mathcal{P}_0^n) = H(\mathcal{P}_0^n | (\mathcal{P}_0^n)_1^\infty) = H(\mathcal{P}_0^n | \mathcal{P}_1^\infty) = H(\mathcal{P} | \mathcal{P}_1^\infty) = h(T, \mathcal{P}).$$

□

Lemma 24.4. *$h(T, \mathcal{P}) \leq h(T, \mathcal{Q}) + H(\mathcal{P}|\mathcal{Q})$.*

Proof.

$$\begin{aligned}
H\big(\mathcal{P}_0^n\big) &\leq H\big(\mathcal{P}_0^n \vee \mathcal{Q}_0^n\big) \\
&= H\big(\mathcal{Q}_0^n\big) + H\big(\mathcal{P}_0^n | \mathcal{Q}_0^n\big) \\
&\leq H\big(\mathcal{Q}_0^n\big) + \sum_{i=0}^{n} H\big(T^{-i}\mathcal{P} | \mathcal{Q}_0^n\big) \\
&\leq H\big(\mathcal{Q}_0^n\big) + \sum_{i=0}^{n} H\big(T^{-i}\mathcal{P} | T^{-i}\mathcal{Q}\big) \\
&= H\big(\mathcal{Q}_0^n\big) + (n+1)H(\mathcal{P}|\mathcal{Q}).
\end{aligned}$$

Divide both sides by $n+1$ and take the limit as $n \to \infty$ to finish. □

Proof of Theorem 24.1. By lemma 24.4 and then lemma 24.3

$$\begin{aligned}
h(T, \mathcal{P}) &\leq h\big(T, \mathcal{Q}_0^n\big) + H\big(\mathcal{P} | \mathcal{Q}_0^n\big) \\
&= h(T, \mathcal{Q}) + H\big(\mathcal{P} | \mathcal{Q}_0^n\big).
\end{aligned}$$

The last term tends $H(\mathcal{P}|\mathcal{Q}_0^\infty)$ by Lemma 23.3. The same inequality holds if we replace $\mathcal{P}$ with $T^{-n}\mathcal{P}$. So

$$h(T, \mathcal{P}) = h(T, T^{-n}\mathcal{P}) \leq h(T, \mathcal{Q}) + H\big(T^{-n}\mathcal{P} | \mathcal{Q}_0^\infty\big).$$

The last term tends to zero as $n \to \infty$ as long as $\mathcal{Q}$ is generating. This implies the theorem. □

Let (K, κ) be a probability space. If κ is supported on a countable set $K' \subset K$ then define

$$H(K, \kappa) := -\sum_{k \in K'} \kappa(k) \log \kappa(k).$$

Otherwise, let $H(K, \kappa) = +\infty$.

Theorem 24.5 (Kolmogorov, 1958)**.** *The Bernoulli shift $\mathbb{Z} \curvearrowright^{\sigma} (K, \kappa)^{\mathbb{Z}}$ with base space (K, κ) has entropy $h(\sigma) = H(K, \kappa)$.*

Proof. Let $\mathcal{P}$ be the time 0 partition: $\mathcal{P} = \{P_k : k \in K\}$, $P_k = \{x \in K^{\mathbb{Z}} : x(0) = k\}$. This is generating. Therefore, $h(\sigma) = h(\sigma, \mathcal{P})$. An exercise verifies:

$$H\left(\mathcal{P}_0^n\right) = (n+1)H(K, \kappa)$$

for every n. Divide by $n+1$ and take the limit as $n \to \infty$ to obtain the result. $\square$

It follows that if two Bernoulli shifts $\mathbb{Z} \curvearrowright (K, \kappa)^{\mathbb{Z}}$ and $\mathbb{Z} \curvearrowright (L, \lambda)^{\mathbb{Z}}$ are isomorphic then $H(K, \kappa) = H(L, \lambda)$. It a major result due to Ornstein (1970) that the converse to the above theorem is true: if (K, κ), (L, λ) are two probability spaces with $H(K, \kappa) = H(L, \lambda)$ then the Bernoulli shifts $\mathbb{Z} \curvearrowright (K, \kappa)^{\mathbb{Z}}$ and $\mathbb{Z} \curvearrowright (L, \lambda)^{\mathbb{Z}}$ are isomorphic.

24.1 Markov Chains

Let A be a finite set and $p : A \times A \to [0,1]$ a transition kernel (so $\sum_b p(a,b) = 1$ for any a). Let $\pi : A \to [0,1]$ be a stationary vector for p. So $\sum_{a \in A} \pi(a)p(a,b) = \pi(b)$ for every $b \in A$. The *Markov Chain* associated to p and π is dynamical system $\mathbb{Z} \curvearrowright^{\sigma} (A^{\mathbb{Z}}, \mu)$ where μ is the shift-invariant probability measure defined by: for any $a_0, a_1, a_2, \ldots, a_n \in A$,

$$\mu(\{x \in A^{\mathbb{Z}} : x(i) = a_i \text{ for } i = 0, \ldots, n\}) = \pi(a_0) \prod_{i=0}^{n-1} p(a_i, a_{i+1}).$$

Let $\mathcal{P} : A^{\mathbb{Z}} \to A$ be the time 0 map: $\mathcal{P}(x) = x(0)$. This is a generating observable. So

$$h_\mu(\sigma) = h(\sigma, \mathcal{P}) = H\left(\mathcal{P} | \bigvee_{i=1}^{\infty} \mathcal{P} \circ \sigma^{-i}\right).$$

However, because this is a Markov Chain,

$$H\left(\mathcal{P} | \bigvee_{i=1}^{\infty} \mathcal{P} \circ \sigma^{-i}\right) = H(\mathcal{P} | \mathcal{P} \circ \sigma^{-1}).$$

Indeed, this is implied by the following fact: if $\mathcal{F}$ is the sigma-algebra generated by $\bigvee_{i=1}^{\infty} \mathcal{P} \circ \sigma^{-i}$ then $I_{\mathcal{P}}(\cdot|\mathcal{F}) = I_{\mathcal{P}}(\cdot|\mathcal{P} \circ \sigma^{-1})$. In other words, the amount of information you gain from $\mathcal{P}$, given that you know $\bigvee_{i=1}^{\infty} \mathcal{P} \circ \sigma^{-i}$ is the same as the amount of information you gain from $\mathcal{P}$ given that you know $\mathcal{P} \circ \sigma^{-1}$.

Now,

$$\begin{aligned}
H(\mathcal{P} | \mathcal{P} \circ \sigma^{-1}) &= -\sum_{a,b \in A} \mu(\mathcal{P}^{-1}(b)) \log \left(\frac{\mu(\mathcal{P}^{-1}(a) \cap (\mathcal{P} \circ \sigma^{-1})^{-1}(b))}{\mu(\mathcal{P}^{-1}(b))} \right) \\
&= -\sum_{a,b \in A} \pi(b) \log \left(\frac{\pi(b)p(b,a)}{\pi(b)} \right) \\
&= -\sum_{a,b \in A} \pi(b) \log \left(p(b,a) \right).
\end{aligned}$$

The point: it is very easy to compute the entropy of a Markov Chain from its stationary vector and transition matrix. This has a more general utility: sometimes it is possible to compare a given dynamical system with a Markov chain and use that to estimate its entropy.

25 The Shannon-McMillan-Breiman Theorem

(much borrowed from Petersen's book)

Theorem 25.1 (SMB Theorem)**.** *Let $T \in \mathrm{Aut}(X, \mu)$ be ergodic and $\mathcal{P}$ a countable measurable partition of X with $H_\mu(\mathcal{P}) < \infty$. Let $\mathcal{P}_0^n = \bigvee_{i=0}^n T^{-i}\mathcal{P}$. Then for a.e. $x \in X$, the atom of $\mathcal{P}_0^n$ containing x has measure $2^{-h(T,\mathcal{P})n+o(n)}$. I.e.*

$$\frac{1}{n+1} I_{\mathcal{P}_0^n}(x) \to h(T, \mathcal{P})$$

pointwise a.e. as $n \to \infty$.

For $n \geq 1$, let $f_n(x) = I_{\mathcal{P}}(x|\mathcal{P}_1^n)$, $f^*(x) = \sup_n f_n(x)$.

Lemma 25.2 (Maximal inequality)**.** *For each $\lambda > 0$ and $P \in \mathcal{P}$,*

$$\mu(\{x \in P : \ f^*x > \lambda\}) < 2^{-\lambda}.$$

Proof. Let $f_n^P(x) = -\log \mathbb{E}[1_P|\mathcal{P}_1^n](x)$ and

$$B_n^P = \{x \in P : \ f_i^P(x) \leq \lambda \forall i < n \text{ and } f_i^P(x) > \lambda\}.$$

This set is $\mathcal{P}_1^n$-measurable. So

$$\begin{aligned} \mu(P \cap B_n^P) &= \int_{B_n^P} 1_P \, d\mu = \int_{B_n^P} \mathbb{E}[1_P|\mathcal{P}_1^n](x) \, d\mu \\ &= \int_{B_n^P} 2^{-f_n^P(x)} \, d\mu \leq 2^{-\lambda}\mu(B_n^P). \end{aligned}$$

So

$$\mu(\{x \in P : \ f^*x > \lambda\}) = \sum_{k=1}^{\infty} \mu(B_k^P \cap P) \geq \sum_{k=1}^{\infty} 2^{-\lambda}\mu(B_n^P) \geq 2^{-\lambda}.$$

□

Lemma 25.3. $f^* \in L^1$.

Proof.

$$\begin{aligned}
\int f^* \, d\mu &= \sum_{P\in\mathcal{P}} \int_0^\infty \mu(\{x \in P: \; f^*(x) > \lambda\}) \, d\lambda \\
&\leq \sum_{P\in\mathcal{P}} \int_0^\infty \min\{2^{-\lambda}, \mu(P)\} \, d\lambda \\
&= \sum_{P\in\mathcal{P}} \int_0^{-\log\mu(P)} \mu(P) \, d\lambda + \int_{-\log\mu(P)}^\infty 2^{-\lambda} \, d\lambda \\
&= H_\mu(\mathcal{P}) + \sum_{P\in\mathcal{P}} \mu(P)/\log 2 < \infty.
\end{aligned}$$

□

Proof of the SMB Theorem. Observe that for any partitions $\mathcal{Q}, \mathcal{R}$,

$$I_{\mathcal{Q}\vee\mathcal{R}}(x) = I_{\mathcal{Q}}(x) + I_{\mathcal{R}}(x|\mathcal{Q})$$

$$I_{T^{-1}\mathcal{Q}}(x) = I_{\mathcal{Q}}(Tx).$$

Therefore,

$$I_{\mathcal{P}_0^n}(x) = I_{\mathcal{P}_1^n}(x) + I_{\mathcal{P}}(x|\mathcal{P}_1^n) = I_{\mathcal{P}_0^{n-1}}(Tx) + f_n(x).$$

By an inductive argument,

$$I_{\mathcal{P}_0^n}(x) = \sum_{k=0}^n f_{n-k}(T^k x)$$

where $f_0(x) = I_{T^{-1}\mathcal{P}}(x)$.

Let $f(x) = I_{\mathcal{P}}(x|\mathcal{P}_1^\infty)$. By the previous lemma and the Martingale Convergence Theorem, $f_n \to f$ pointwise a.e. and in L^1.

So

$$\begin{aligned}
\frac{1}{n+1} I_{\mathcal{P}_0^n}(x) &= \frac{1}{n+1} \sum_{k=0}^n f_{n-k}(T^k x) \\
&= \frac{1}{n+1} \sum_{k=0}^n [f_{n-k}(T^k x) - f(T^k x)] + f(T^k x).
\end{aligned}$$

By the ergodic theorem,

$$\frac{1}{n+1} \sum_{k=0}^n f(T^k x) \to \int f(x) \, d\mu = H(\mathcal{P}|\mathcal{P}_1^\infty) = h(\mathcal{P}, T).$$

So it suffice to show

$$\frac{1}{n+1} \sum_{k=0}^n [f_{n-k}(T^k x) - f(T^k x)] \to 0.$$

Let

$$F_N = \sup_{k \geq N} |f_k - f|.$$

Then

$$\begin{aligned} & \frac{1}{n+1} \sum_{k=0}^{n} |f_{n-k}(T^k x) - f(T^k x)| \\ = \; & \frac{1}{n+1} \sum_{k=0}^{n-N} |f_{n-k}(T^k x) - f(T^k x)| + \frac{1}{n+1} \sum_{k=n-N+1}^{n} |f_{n-k}(T^k x) - f(T^k x)| \\ \leq \; & \frac{1}{n+1} \sum_{k=0}^{n-N} F_N(T^k x) + \frac{1}{n+1} \sum_{k=n-N+1}^{n} |f_{n-k}(T^k x) - f(T^k x)|. \end{aligned}$$

Fix N and let $n \to \infty$. Because $|f_{n-k}(T^k x) - f(T^k x)| \leq f^*(T^k x) + f(T^k x) \in L^1$ the second term tends to zero. Also $0 \leq F_N \leq f^* + f$. So we apply the ergodic theorem to show that the first term tends to

$$\int F_N \, d\mu.$$

The dominated convergence theorem and the fact that $F_N \to 0$ pointwise a.e. imply that this tends to zero as $N \to \infty$.

□

26 The variational principle

This follows Misiurewicz's proof and is based on the exposition in Brin-Stuck.

Theorem 26.1. *Let (X, d) be a compact metric space and $f : X \to X$ a homeomorphism. Then*

$$h_{top}(f) = \sup_{\mu} h_\mu(f)$$

where the supremum is over all f-invariant Borel probability measures μ on X.

We'll need:

Lemma 26.2. *For any n, $h_\mu(f^n) = |n| h_\mu(f)$ and $h_{top}(f^n) = |n| h_{top}(f)$.*

Proof. We'll just prove the first statement for $n > 0$; the rest is similar. Let $\mathcal{P}$ be a partition. Then

$$h_\mu(f^n, \mathcal{P}_0^{n-1}) = \lim_{k \to \infty} \frac{1}{k+1} H_\mu(\mathcal{P}_0^{nk}) = n h_\mu(f, \mathcal{P}).$$

The result follows by taking the sup over all $\mathcal{P}$. □

First we obtain the lower bound:

Lemma 26.3. *With f, X, d as in the theorem, if μ is any f-invariant probability measure on X then $h_{top}(f) \geq h_\mu(f)$.*

Proof. Let $\mathcal{P}$ be a finite partition. For each $P \in \mathcal{P}$, let $K_P \subset P$ be a compact set; chosen so that if $\mathcal{Q}$ is the smallest partition of X containing K_P for all P then $H_\mu(\mathcal{P}|\mathcal{Q}) < 1$. So

$$h(f, \mathcal{P}) \leq h(f, \mathcal{Q}) + H(\mathcal{P}|\mathcal{Q}) \leq h(f, \mathcal{Q}) + 1.$$

Let $\mathcal{O}$ be the open covering $\mathcal{O} = \{X - K_P : \ P \in \mathcal{P}\}$. Note $|\mathcal{Q}_0^{n-1}| \leq 2^n |\mathcal{O}_0^{n-1}|$ (since each set of $\mathcal{O}$ intersects at most 2 sets in $\mathcal{Q}$). Therefore,

$$H_\mu(\mathcal{Q}_0^{n-1}) \leq \log|\mathcal{Q}_0^{n-1}| \leq n\log(2) + \log|\mathcal{O}_0^{n-1}|.$$

Let δ_0 be a Lebesgue number for $\mathcal{O}$ (wrt d). Then δ_0 is also a Lebesgue number for $\mathcal{O}_0^{n-1}$ wrt d_n^f.

$\mathcal{O}$ has no proper subcoverings and neither does $\mathcal{O}_0^{n-1}$. So for every $C \in \mathcal{O}_0^{n-1}$, we let $x_C \in C$ be an element not contained in any other set of $\mathcal{O}_0^{n-1}$. Then $S_n = \{x_C\}$ is (d_n^f, δ_0)-separated. So $\mathrm{sep(n, \delta_0, f)} \geq |\mathcal{O}_0^{\mathrm{n}-1}|$ and

$$\begin{aligned} h(f, \delta_0) &= \limsup_{n\to\infty} \frac{1}{n}\log(\mathrm{sep(n, \delta_0, f)}) \geq \limsup_{\mathrm{n}\to\infty} \frac{1}{\mathrm{n}} \mathrm{H}_\mu(\mathcal{Q}_0^{\mathrm{n}-1}) - \log(2) \\ &\geq h(\mathcal{Q}, f) - \log(2) \geq h(\mathcal{P}, f) - \log(2) - 1. \end{aligned}$$

Now replace f with f^n using the previous lemma to conclude this lemma. ☐

Next we obtain the upper bound.

Given a subset $Y \subset X$, let $\partial Y = \bar{Y} \cap \overline{X - Y}$. It will be useful to have the following elementary facts:

Lemma 26.4 (Concavity of Shannon entropy)**.** *For any probability measures μ, ν and partition $\mathcal{P}$ with finite Shannon entropy and $0 \leq t \leq 1$,*

$$H_{t\mu+(1-t)\nu}(\mathcal{P}) \geq tH_\mu(\mathcal{P}) + (1-t)H_\nu(\mathcal{P}).$$

Lemma 26.5. *Assume X is infinite and μ is a Borel probability measure on X.*

1. *For any $\delta > 0$ and $x \in X$, $\exists \delta' \in (0, \delta)$ such that $\mu(\partial B(x, \delta')) = 0$.*
2. *For any $\delta > 0$ there is a finite measurable partition of X such that its parts all have μ-measure zero boundary and diameter bounded by δ.*
3. *If μ_n converges to μ in the weak* topology, and $\mu(\partial Y) = 0$ then $\lim_n \mu_n(Y) = \mu(Y)$.*

Proof of Theorem 26.1. It suffices to obtain the upper bound. Let

- $\epsilon > 0$,
- $S_n \subset X$ be an (d_n^f, ϵ)-separated subset,

- $\nu_n \in \mathcal{P}(X)$ be uniformly distributed on S_n,
- $\mu_n := \frac{1}{n}\sum_{k=0}^{n-1} f_*^i \nu_n$,
- μ be any weak* accumulation point of $\{\mu_n\}$,
- $\mathcal{P}$ be a partition such that its parts all have μ-measure zero boundary and diameter bounded by ϵ.
- Fix $0 \le k < q < n$, let $a_k = \lceil (n-k)/q \rceil$,
- $J = [k, a_k q - 1] \cap \mathbb{Z}$, $J' = [0, n-1] \cap \mathbb{Z} - J$.

If $C \in \mathcal{P}_0^{n-1}$, then C contains at most one element of S_n. So $\nu_n(C) \in \{0, 1/|S_n|\}$. So $H_{\nu_n}(\mathcal{P}_0^{n-1}) = \log|S_n|$. Thus

$$\begin{aligned} \log|S_n| &= H_{\nu_n}(\mathcal{P}_0^{n-1}) \le H_{\nu_n}(\mathcal{P}^J) + H_{\nu_n}(\mathcal{P}^{J'}) \\ &\le \sum_{r=0}^{a_k-1} H_{\nu_n}(f^{-(rq+k)}\mathcal{P}_0^{q-1}) + 2qH(\mathcal{P}) \\ &\le \sum_{r=0}^{a_k-1} H_{f_*^{rq+k}\nu_n}(\mathcal{P}_0^{q-1}) + 2qH(\mathcal{P}). \end{aligned}$$

So

$$\begin{aligned} (q/n)\log|S_n| &\le \sum_{k=0}^{q-1} \frac{1}{n} \sum_{r=0}^{a_k-1} H_{f_*^{rq+k}\nu_n}(\mathcal{P}_0^{q-1}) + 2q^2 H(\mathcal{P})/n \\ &\le H_{\mu_n}(\mathcal{P}_0^{q-1}) + 2q^2 H(\mathcal{P})/n \end{aligned}$$

where concavity was used in the last step. By the Portmanteau Theorem, $H_{\mu_n}(\mathcal{P}_0^{q-1)} \to H_\mu(\mathcal{P}_0^{q-1})$. Thus

$$\limsup_{n\to\infty} \frac{1}{n}\log|S_n| \le (1/q)H_\mu(\mathcal{P}_0^{q-1}).$$

Now we let $q \to \infty$ to obtain

$$\limsup_{n\to\infty} \frac{1}{n}\log|S_n| \le h(f,\mathcal{P}) \le h_\mu(f).$$

Since S_n is arbitrary, we have $h(f,\epsilon) \le h_\mu(f)$ which implies the result (since ϵ is arbitrary).

□

There does not always exist a measure of maximal entropy. However, one does exist if the action is expansive: this means there exists $\delta > 0$ such that if $x, y \in X$ are distinct then there is an n such that $d(f^n x, f^n y) > \delta$. You can prove that expansitivity does not depend on the choice of metric. There are weaker forms of expansitivity (called h-expansive and asymptotic h-expansive) which imply that entropy is an upper semi-continuous function on the space of invariant measures and therefore, there exists a measure of a maximal entropy. On the other hand, there are example of homeomorphisms for which no measure of maximal entropy exists

27 Ornstein theory

There are a number of results that today fall under the heading of "Ornstein theory". These include

Sinai's Theorem (1965): if $T \in \mathrm{Aut}(X,\mu)$ is ergodic then T factors onto every Bernoulli shift with entropy $\le h(T)$

Ornstein's Theorem (1970): two Bernoulli shifts are isomorphic iff they have the same entropy. Moreover,

Krieger's Theorem (?): If $T \in\in \mathrm{Aut}(X,\mu)$ is ergodic and $\log(n) > h(T)$ then there exists a partition $\mathcal{P}$ of X with $|\mathcal{P}| \le n$ that is generating with respect to T.

Ornstein (?): There exists a measure-preserving action of $\mathbb{R}$ such that for every $t > 0$, the induced action of $t\mathbb{Z}$ is Bernoulli. This is called the Bernoulli flow. It is unique up to measure-conjugacy and time-change.

Moreover: the following automorphisms and flows are measurably conjugate to Bernoulli: factors of Bernoulli shifts, mixing Markov chains, the geodesic flow on a closed hyperbolic manifold, hyperbolic toral automorphisms.

Given all of these examples, you might ask: which automorphisms are not measurably conjugate to Bernoulli? Of course, zero entropy automorphisms are not. Also automorphisms which admit nontrivial zero entropy factors are not. Any automorphism that does not have a nontrivial zero entropy factor is called CPE (completely positive entropy). Ornstein also proved that non-Bernoulli CPE automorphisms exist; but it is still nontrivial to construct them.

In what follows, I will outline the proof of Sinai's Theorem and Ornstein's Theorem. I'll restrict attention to symbolic processes. To be precise, a symbolic process consists of a sequence $\{X_i\}_{i\in\mathbb{Z}}$ of random variables such that each X_i takes values in a finite set A. Moreover we require that the process is **stationary** which means that the law of $\{X_i\}_{i\in\mathbb{Z}}$ is the same as the law of $\{X_{i+n}\}_{i\in\mathbb{Z}}$ for any $n \in \mathbb{Z}$. Equivalently, a symbolic process consists of a shift-invariant probability measure μ on $A^{\mathbb{Z}}$ (and $\{X_i\}_{i\in\mathbb{Z}}$ is random with law μ).

27.1 Joinings... (rough draft)

Now suppose $T \in \mathrm{Aut}(X,\mu)$ and $S \in \mathrm{Aut}(Y,\nu)$. A **joining** of T and S is a $T \times S$-invariant measure λ on $X \times Y$ whose marginals are μ and ν respectively. Let $J(\mu,\nu)$ be the set of all joinings of μ and ν. If we assume that X and Y are endowed with a compact Hausdorff topologies, then we can put the weak* topology on $J(\mu,\nu)$. It can be shown that this topology is actually independent of the choice of topologies on X and Y. So $J(\mu,\nu)$ is a compact convex space. Assuming T and S are ergodic, $J(\mu,\nu)$ is a simplex (by the ergodic decomposition theorem).

Let $J^{erg}(\mu,\nu)$ be the subspace of ergodic joinings. This is a G_δ-subset of $J(\mu,\nu)$.

Examples: $\mu \times \nu$ is the **product joining**. If there is a factor map $\phi : X \to Y$ (such that $\phi T = S\phi$) and $\Phi(x) = (x,\phi(x)) \in X \times Y$ then $\Phi_*\mu$ is a **factor joining**.

Theorem 27.1. *Suppose T is ergodic, S is Bernoulli and $h(T) \ge h(S)$. Then the set of all factor joinings in $J^{erg}(\mu,\nu)$ is a dense G_δ. In particular, there exists one.*

[I'm not sure who to attribute this theorem to. It is certainly based on Ornstein theory but it is not how Ornstein originally proved his results. This statement can be found in Glasner's book, Rudolph's book or Downarowicz' book where it is referred to as the 'Residual Sinai Theorem'].

This statement clearly implies both Sinai's and Ornstein's Theorem.

27.2 Rohklin's Lemma

Here we'll just concentrate on Sinai's Theorem. We'll need the following fundamental result:

Lemma 27.2 (Rohklin's Lemma)**.** *Let* $T \in \mathrm{Aut}(X, \mu)$ *be ergodic and assume* μ *is purely nonatomic. Then for every* $\epsilon > 0$ *and* $n \in \mathbb{N}$ *there exists a subset* $F \subset X$ *such that* $F, TF, \dots, T^n F$ *are pairwise disjoint and* $\mu(\cup_{i=0}^{n} T^i F) > 1 - \epsilon$*.*

Remark 1. This lemma is also true if T is non-ergodic. However, in that case we must assume that T is aperiodic. This means, for a.e. $x \in X$ for every $n > 0$, $T^n x \neq x$.

Remark 2. This simple lemma is a key element of many results in ergodic theory aside from Ornstein theory. For example, it plays a major role in Halmos' proof that weakly mixing transformations form a dense G_δ subset of $\mathrm{Aut}(X, \mu)$ and the Jewett-Krieger theorem on the existence of uniquely ergodic models.

Proof. Let $A_0 \subset X$ be any set with $0 < \mu(A_0) < \epsilon/n$. By ergodicity, $\mu(\cup_{k=0}^{\infty} T^k A_0) = 1$. Define A_k for $k \geq 1$ by

$$A_k = T^k A_0 - \cup_{i=0}^{k-1} T^i A_0.$$

Note $A_0, A_1, \dots$ are pairwise disjoint and $T^k A_0 \subset \cup_{i=0}^{k} A_i$. Now let $F = \cup_{k \geq 1} A_{kn}$. It is easy to check that F satisfies the lemma. □

Lemma 27.3 (Slightly enhanced Rohklin Lemma)**.** *Same hypotheses. In addition suppose we are given a set* $A_0 \subset X$ *with* $\mu(A_0) < \epsilon/n$ *then we can choose* F *satisfying the conclusion of Rohlin's Lemma with* $F \cap A_0 = \emptyset$*.*

27.3 Sinai's Theorem

(This loosely follows Tim Austin's notes which are roughly prepared from Smorodinsky's lecture notes).

We assume $T \in \mathrm{Aut}(X, \mu)$ is ergodic and $\mathbf{p} = (p_1, \dots, p_k)$ is a probability vector with $h(T) = H(\mathbf{p})$. We will prove there exists a factor map $\Phi_\infty : X \to [k]^{\mathbb{Z}}$ where $[k] = \{1, \dots, k\}$ such that $\Phi_{\infty *}\mu = \mathbf{p}^{\mathbb{Z}}$. Note that any such factor map is determined by its time 0 component. So we are most interested in 'constructing' $\phi_\infty : X \to [k]$ (such that $\Phi_\infty(x)_0 = \phi_\infty(x)$ is the required factor).

Note that any map $\phi : X \to [k]$ determines a partition on X. So $h(T, \phi)$ and $H(\phi)$ are well-defined.

We will construct ϕ_∞ as a limit of maps $\phi_n : X \to [k]$. More precisely: it suffices to construct maps $\phi_n : X \to [k]$ satisfying

1. (entropy convergence) $\lim_{n\to\infty} h(T, \phi_n) = h(T)$.
2. (convergence of the marginal distribution, or weak* convergence) $\lim_{n\to\infty} \phi_{n*}\mu = \mathbf{p}$.
3. (convergence of factors) $\sum_{n=1}^{\infty} \mu(\{x \in X : \ \phi_n(x)_0 \neq \phi_{n+1}(x)_0\}) < \infty$.

[By the way, if $\mathcal{P}, \mathcal{Q}$ are partitions with finite Shannon entropy, then $H_\mu(\mathcal{P} \vee \mathcal{Q}) = H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q})$ if and only if $\mathcal{P}$ and $\mathcal{Q}$ are independent (by strict concavity of $t \mapsto -t\log(t)$). Otherwise we have $<$. So the above conditions really do imply that the limit ϕ_∞ is independent of $\phi_\infty \circ T$... and therefore the factor measure $\Phi_{\infty*}\mu$ is Bernoulli]

So it suffices to prove the following key lemma:

Proposition 27.4 (Partition upgrade procedure)**.** *Given any $\epsilon_1 > 0$ there exists $\delta_1 > 0$ such that the following holds. If $\phi_1 : X \to [k]$ satisfies*

1. *(law of ϕ_1 close enough to $\bar{p}$)*
$$\|\mu \circ \phi_1^{-1} - \bar{p}\| < \delta_1$$
2. *(ϕ_1 is entropically efficient enough)*
$$h_\mu(T, \phi_1) > H_\mu(\phi_1) - \delta_1$$

then for any $0 < \delta_2 < \delta_1$ there exists $\phi_2 : X \to [k]$ such that

1. *(law of ϕ_2 is very close enough to $\bar{p}$)*
$$\|\mu \circ \phi_2^{-1} - \bar{p}\| < \delta_2$$
2. *(ϕ_2 is very entropically efficient)*
$$h_\mu(T, \phi_2) > H_\mu(\phi_2) - \delta_2$$
3. *(ϕ_2 is close to ϕ_1):*
$$\mu(\{x \in X : \ \phi_2(x) \neq \phi_1(x)\}) < \epsilon_1.$$

Clearly this proposition implies the existence of $\{\phi_n\}$ satisfying the previous argument and therefore implies Sinai's Theorem.

Note that if $\epsilon_1 > 1$ then the third conclusion is vacuous and most of the hypotheses become irrelevant. However, the statement is still nontrivial. We will prove this weaker statement first and then show how to modify the argument to obtain the proposition.

There are essentially no canonical choices in the proof. So it's a bit messy.

To be clear: let $\delta_2 > 0$ be arbitrary. Our immediate goal is to prove the existence of ϕ_2 satisfying conclusions (1) and (2) above.

27.3.1 θ

We can assume (after replacing k with $k+1$ if necessary) wlog that $\bar{p}$ is not the uniform distribution.

Let $K \in \mathbb{N}$ and choose $\theta : X \to [K]$ so that $h_\mu(T, \theta) > h_\mu(T) - \delta_2/100$.

27.3.2 ϵ_2

If $\mathcal{Q}, \mathcal{P}$ are partitions, we write $\mathcal{Q} \leq_\epsilon \mathcal{P}$ if for every $Q \in \mathcal{Q}$ there exist sets $P_1, \ldots, P_k \in \mathcal{P}$ such that

$$\mu(Q \bigtriangleup \cup_{i=1}^k P_i) < \epsilon.$$

Exercise 12. For any $\epsilon > 0$ and partition $\mathcal{Q}$ there exists $\delta > 0$ such that if $\mathcal{P}$ is a partition with $\mathcal{Q} \leq_\delta \mathcal{P}$ then $h(T, \mathcal{P}) > h(T, \mathcal{Q}) - \epsilon$.

Hint: in the proof of the Kolmogorov-Sinai Theorem we showed that $h(T, \mathcal{P}) \geq h(T, \mathcal{Q}) - H(\mathcal{P}|\mathcal{Q})$.

Now choose $\epsilon_2 > 0$ small enough so that if $\mathcal{P}$ is any partition with $\theta \leq_{\epsilon_2} \mathcal{P}$ then

$$h(T, \mathcal{P}) > h(T, \theta) - \delta_2/100 > h(T) - \delta_2/50.$$

We can also require that $\epsilon_2 < \delta_2/100$.

27.3.3 α and $\mathbf{q}$

Choose a probability vector $\mathbf{q}$ on $[k]$ with $H(\mathbf{q}) > H(\mathbf{p})$ and $\|\mathbf{q} - \mathbf{p}\| < \delta_2/100$. This is possible because $\mathbf{p}$ is not uniform. Let $0 < \alpha \leq H(\mathbf{q}) - H(\mathbf{p})$.

27.3.4 n

Lemma 27.5. *There exists N such that if $n \geq N$ then*

$$\mu(\{x \in X : \ \mu(\theta_0^{n-1}(x)) > \exp(-nH(\mathbf{p}) - n\alpha/3)\}) > 1 - \epsilon_2/200$$

and

$$\mathbf{q}^{\mathbb{Z}}(\{x \in [k]^{\mathbb{Z}} : \ q_{x_0} \cdots q_{x_{n-1}} < \exp(-nH(\mathbf{q}) + n\alpha/3)\}) > 1 - \epsilon_2/400$$

and

$$\mathbf{q}^{\mathbb{Z}}(\{x \in [k]^{\mathbb{Z}} : \ \|\mathbf{q} - \pi_0^{n-1}(x)\| < \delta_2/100\}) > 1 - \epsilon_2/400$$

where $\pi_0^{n-1}(x)$ is the probability vector on $[k]$ given by

$$\pi_0^{n-1}(x)_i = \frac{1}{n}\#\{0 \leq j \leq n-1 : \ x_j = i\}.$$

Proof. This follows from the SMB Theorem and the law of large numbers. It uses that $h(T, \theta) \leq H(\mathbf{p})$. □

From now on, fix $n \geq N$ as above.

27.3.5 γ_2

Let $\mathcal{G}_1 \subset \theta_0^{n-1}$ be those cells with measure $> \exp(-nH(\mathbf{p}) - n\alpha/3)$.

Let $\mathcal{G}_2 \subset [k]^n$ be those words with $\mathbf{q}^n$-measure $< \exp(-nH(\mathbf{q}) + n\alpha/3)$ whose empirical distribution is $\delta_2/100$ close to $\mathbf{q}$. By the previous lemma, $\mu(\cup\mathcal{G}_1) > 1 - \epsilon_2/200$ and $\mathbf{q}^n(\mathcal{G}_2) > 1 - \epsilon_2/200$.

Note that

$$|\mathcal{G}_1| \leq \exp(nH(\mathbf{p}) + n\alpha/3) \leq \exp(nH(\mathbf{q}) - n\alpha/3) \leq |\mathcal{G}_2|.$$

So there is an injective map $\gamma_2 : \mathcal{G}_1 \to \mathcal{G}_2$.

27.3.6 F

By the Rohlin Lemma, there exists a subset $F \subset X$ such that $F, TF, \dots, T^{n-1}F$ are pairwise disjoint and

$$\mu(\cup_{i=0}^{n-1} T^i F) > 1 - \delta_2/100.$$

For the moment, suppose that $F \subset \cup\mathcal{G}_1$. Then define $\phi_2 : X \to [k+1]$ by

- if $x \in F$ then define $(\phi_2(x), \phi_2(Tx), \dots, \phi_2(T^{n-1}x)) = \gamma_2(\theta_0^{n-1}x)$.
- the above statement actually defines $\phi_2(x)$ whenever $x \in \cup_{i=0}^{n-1} T^i(F)$. This is well-defined because $F, TF, \dots, T^{n-1}F$ are pairwise disjoint.
- $\phi_2(x) = k + 1$ otherwise.

Observe that

$$\|\phi_{2*}\mu - \mathbf{p}\| \leq \|\phi_{2*}\mu - \mathbf{q}\| + \|\mathbf{q} - \mathbf{p}\| \leq \delta_2/100 + \frac{\delta_2/100}{1 - \delta_2/100} < \delta_2.$$

Also note that $(\phi_2)_0^{n-1} \geq_{\epsilon_2} \theta$ (because γ_2 is injective). So the choice of ϵ_2 implies $h(T, \phi_2) \geq h(T, \theta) - \delta_2/100 \geq h(T) - \delta_2$. Of course, the fact that we have expanded $[k]$ to $[k+1]$ plays no essential role we could have defined $\phi_2(x)$ arbitrarily in the third bullet above; but it helps to understand why $(\phi_2)_0^{n-1} \geq_{\epsilon_2} \theta$.

What do we do if F is not contained in $\cup\mathcal{G}_1$? Let $B = X - \cup\mathcal{G}_1$. There must exist some j with $0 \leq j \leq n-1$ such that

$$\mu(T^j F \cap B) \leq \mu(B)/n \leq \epsilon_2/200n.$$

Now define $\phi_2 : X \to [k+1]$ by

- if $x \in T^j F$ then define $(\phi_2(x), \phi_2(Tx), \dots, \phi_2(T^{n-1}x)) = \gamma_2(\theta_0^{n-1}x)$.
- the above statement actually defines $\phi_2(x)$ whenever $x \in \cup_{i=0}^{n-1} T^{i+j}(F)$. This makes sense because $T^j F \cap T^k F = \emptyset$ if $j < k \leq j + n - 1$.

- $\phi_2(x) = k+1$ otherwise.

Observe that

$$\|\phi_{2*}\mu - \mathbf{p}\| \le \|\phi_{2*}\mu - \mathbf{q}\| + \|\mathbf{q} - \mathbf{p}\| \le \delta_2/100 + \frac{\delta_2/100}{1-\delta_2/100} + \epsilon_2/200 < \delta_2.$$

Also note that $(\phi_2)_0^{n-1} \ge_{\epsilon_2} \theta$ (because γ_2 is injective). So the choice of ϵ_2 implies $h(T, \phi_2) \ge h(T, \theta) - \delta_2/100 \ge h(T) - \delta_2$.

27.4 The partition upgrade procedure

Here we prove the partition upgrade proposition. The proof essentially follows the previous with a few changes.

27.4.1 δ_1

We can assume (after replacing k with $k+1$ if necessary) wlog that $\bar{p}$ is not the uniform distribution.

Definition 12. Let $T \in \mathrm{Aut}(X, \mu)$ and $S \in \mathrm{Aut}(Y, \nu)$. A **joining** of T and S is a $T \times S$-invariant Borel probability measure η on $X \times Y$ with marginals μ and ν.

Definition 13. Let μ, ν be shift-invariant Borel probabiliity measures on $[k]^{\mathbb{Z}}$ (for some k). The $\bar{d}$-distance between μ and ν is

$$\bar{d}(\mu, \nu) := \inf \eta\{(x,y) :\ x_0 \ne y_0\}$$

where the inf is over all joinings η of μ and ν.

Definition 14. A shift-invariant measure μ on $[k]^{\mathbb{Z}}$ is **finitely determined** if: for every $\epsilon > 0$ there exists $\delta > 0$ such that if ν is any shift-invariant measure on $[k]^{\mathbb{Z}}$ satisfying

- $|h_\mu(\sigma) - h_\nu(\sigma)| < \delta$
- the 1-dimensional marginals of μ and ν are δ-close

then $\bar{d}(\mu, \nu) < \epsilon$.

One of the key ingredients in Ornstein theory is:

Lemma 27.6. *Product measures $\mu = \mu_0^{\mathbb{Z}}$ are finitely determined.*

In fact as a consequence of Ornstein theory one can show that μ is finitely determined iff the automorphism $\sigma \in \mathrm{Aut}([k]^{\mathbb{Z}}, \mu)$ is isomorphic to a Bernoulli shift. We will not prove this.

Proof sketch. Let

$$\bar{d}_n(\mu,\nu) = \inf \frac{1}{n}\sum_{i=0}^{n-1} \eta_n(\{(x,y):\ x_i \ne y_i\})$$

where the inf is over all probability measures η_n on $[k]^n$ with marginals μ_0^{n-1} and ν_0^{n-1}. Clearly $\bar{d}_n(\mu,\nu) \to \bar{d}(\mu,\nu)$ so it suffices to show that for every n, $\bar{d}_n(\mu,\nu) < \epsilon$ (under the assumption that ν is sufficiently close to μ).

Once we decide on δ, the way that the joining η_n is built is as follows. First we build η_1. This is a measure on $[k]\times[k]$. We choose η_1 to realize $\bar{d}_1(\mu,\nu)$.

Assuming η_{n-1} has been chosen, we choose η_n as follows. Let $(x,y) \in [k]^{n-1}\times[k]^{n-1}$ be random with law η_{n-1}. So y has law ν_0^{n-1}. Let $y' \in [k]^n$ be random with law equal to ν_0^n conditioned on y (so y' projects to y). Let ν_y be the law of the last coordinate of y'. This is a probability measure on $[k]$. If we have choose $\delta > 0$ sufficiently small then it can be shown that ν_y is close to μ_0 (for all y in a set of measure $> 1-\epsilon/2$). Therefore there exists a joining η'_y of μ_0 and ν_y such that

$$\eta'_y(\{(t,s)\in[k]\times[k]:\ t\ne s\}) < \epsilon/2.$$

Let (t,s) be random with law η'_y and let η_n be the law of (xt,ys). This works. □

Lemma 27.7. *For any $\epsilon_1 > 0$ there exists $\delta_1 > 0$, a probability $\mathbf{q}$ on $[k]$ and N such that if*

$$\|\mu\circ\phi_1^{-1} - \bar{p}\| < \delta_1$$

$$h_\mu(T,\phi_1) > H_\mu(\phi_1) - \delta_1$$

and $n \ge N$ and $F\subset X$ is a set with $F, TF, \ldots, T^{n-1}F$ disjoint and $\mu(\cup_i T^iF) > 1-\delta_1$ then there exists a map $\psi : F \to [k]^n$ satisfying

1. $\psi_*(\mu \restriction F) = \mathbf{q}^n$
2. *for a.e.* $x\in F$, $\frac{1}{n}\#\{0\le j\le n-1:\ \phi_1(T^jx)\ne\psi(x)_j\}) < \epsilon_1/100$
3. $H(\mathbf{q}) > H(\mathbf{p})$.

27.4.2 θ

Let $K\in\mathbb{N}$ and choose $\theta : X\to[K]$ so that $h_\mu(T,\theta) > h_\mu(T) - \delta_2/100$ AND $\theta \ge \phi_1$.

27.4.3 ϵ_2

As before, choose $\epsilon_2 > 0$ small enough so that if $\mathcal{P}$ is any partition with $\theta \le_{\epsilon_2} \mathcal{P}$ then

$$h(T,\mathcal{P}) > h(T,\theta) - \delta_2/100 > h(T) - \delta_2/50.$$

We also require $\epsilon_2 < \delta_2/100$.

27.4.4 α

Let $0 < \alpha < H(\mathbf{q}) - H(\mathbf{p})$.

For $\omega \in [k]^n$, let $\pi(\omega)$ be the probability measure on $[k]$ defined by

$$\pi(\omega)_i = \frac{1}{n}\#\{0 \le j \le n-1 : \ \omega_j = i\}.$$

This is the empirical distribution of ω.

Lemma 27.8. *There exists N such that if $n > N$ then the following holds. Let $\mathcal{G}_1$ be the set of atoms of $(\phi_1)_0^{n-1}$ with measure $> \exp(-nh(T,\phi_1) - n\alpha/3)$. Let $\mathcal{G}_2 \subset [k]^n$ be the set of words ω satisfying*

$$\mathbf{q}^n(\omega) < \exp(-nH(\mathbf{q}) + n\alpha/3)$$
$$\|\mathbf{q} - \pi(\omega)\| < \epsilon_2/100.$$

There exists a $\exp(n[H(\mathbf{q}) - h(T,\phi_1)] - 2n\alpha/3)$-to-1 map $\gamma_2' : \mathcal{G}_1 \to \mathcal{G}_2$ such that for every $\omega \in \mathcal{G}_1$,

$$\|\pi(\omega) - \pi(\gamma_2'(\omega))\| < \epsilon_1/100.$$

Proof. It suffices to verify the conditions of the Marriage Lemma. That is, we need to show that if $S \subset \mathcal{G}_1$ and $N(S) \subset \mathcal{G}_2$ is the set of all ω such that there exists $\omega' \in S$ such that $\|\pi(\omega') - \pi(\omega)\| < \epsilon_1/100$ then $|N(S)| \ge \exp(n[H(\mathbf{q}) - h(T,\phi_1)] - 2n\alpha/3)|S|$.

Indeed, Lemma ** above implies $\mu(\cup S) \le \mathbf{q}^n(\cup N(S))$. Also

$$|S| \le \mu(\cup S)\exp(nh(T,\phi_1) + n\alpha/3)$$
$$|N(S)| \ge \mathbf{q}^n(\cup N(S))\exp(nH(\mathbf{q}) - n\alpha/3).$$

So

$$|N(S)| \ge \exp(n[H(\mathbf{q}) - h(T,\phi_1)] - 2n\alpha/3)|S|.$$

□

27.4.5 n

Lemma 27.9. *There exists N such that if $n \ge N$ then*

$$\mu(\{x \in X : \ \mu(\theta_0^{n-1}(x)|(\phi_1)_0^{n-1}) > \exp(-n[H(\mathbf{p}) - H(T,\phi_1)] - n\alpha/3)\}) > 1 - \epsilon_2/200$$

Proof. This follows from the relative SMB Theorem. □

Let $\mathcal{G}_0$ be the set of atoms A of θ_0^{n-1} with $\mu(A|(\phi_1)_0^{n-1}) > \exp(-n[H(\mathbf{p}) - H(T,\phi_1)] - n\alpha/3)$. So for each atom $B \in \mathcal{G}_1$ there exist at most $\exp(n[H(\mathbf{p}) - H(T,\phi_1)] + n\alpha/3)$ atoms $A \in \mathcal{G}_0$ with $A \subset B$. Since

$$\exp(n[H(\mathbf{p}) - H(T,\phi_1)] + n\alpha/3) \le \exp(n[H(\mathbf{q}) - h(T,\phi_1)] - 2n\alpha/3)$$

there exist an injective map $\gamma_2 : \mathcal{G}_0 \to \mathcal{G}_2$ such that for all $A \in \mathcal{G}_0$ and a.e. $x \in A$,

$$\frac{1}{n}\sum_{i=0}^{n-1} \mu(\{x \in X : \ \phi_1(T^i x) \ne \gamma_2(x)_i\}) < \epsilon_1/10.$$

27.4.6 F

The rest is exactly as before.

28 Sofic entropy theory

Outline:

1. Benjamini-Schramm convergence
2. What is a sofic group?
3. soficity is closed under many operations; all linear groups are sofic. It is unknown whether all groups are sofic.
4. Topological entropy for sofic groups via metrics. Then via pseudometrics.
5. The amenable case. Dependence on sofic approximation. Example: free group G acting on 2-point space nontrivially.
6. Topological entropy in the special case of a subshift $X \subset [k]^G$.
7. Gottshalks' conjecture and direct finiteness.
8. Measure-theoretic entropy via metrics (and pseudometrics).
9. The variational principle.
10. Classification of Bernoulli shifts.
11. the amenable case and dependence of sofic approximation.
12. formulas: direct product, subgroup, ergodic decomposition, ...
13. Open problems: extend Ornstein theory, especially Sinai's Theorem. Dependence on sofic approximation. Classify Markov chains over a free group. Krieger's generator theorem.

29 Kingman's subadditive ergodic theorem

I plan to loosely follow Sarig's notes. You can google Omri Sarig Ergodic Theory to obtain his lecture notes. I'm also using Krengel's book "Ergodic Theorems".

Let $T \in \mathrm{End}(X, \mu)$. A function $c : \mathbb{N} \times X \to \mathbb{R}$ is a **subadditive cocycle** if

$$c(m+n, x) \leq c(m, T^n x) + c(n, x).$$

(It is a cocycle if we have equality above).

Example 1 (Random walks)*.* Let G be a countable group and μ a probability measure on G. Consider the shift action $\sigma : G^{\mathbb{N}} \to G^{\mathbb{N}}$, $\sigma(x)_n = x_{n+1}$. This preserves the product measure μ^G. Note that if $x \in G^{\mathbb{N}}$ is random with law μ^G then $n \mapsto x_1 x_2 \cdots x_n$ is a random walk on G.

There are several quantities related to this random walk that we might be interested in. For example, suppose d is a left-invariant metric on G. We may be interested in $c_1(n,x) := d(e, x_1 \cdots x_n)$. This is a sub-additive cocycle.

Another example: we may be interested in

$$c_2(n,x) = \#\{x_1 \cdots x_m :\ 1 \le m \le n\}$$

the size of the range of the random walk. This is also a sub-additive cocycle.

Another example: we may be interested in $c_3(n,x) := -\log \mu^n(\{x_1 \cdots x_n\})$ where μ^n is the law of $x_1 \cdots x_n$. This is the amount of information we gain by learning where the random walker is at time n.

Theorem 29.1. *Suppose c is subadditive and $c(1,\cdot) \in L^1(X,\mu)$.*

Then

$$f(x) := \lim_{n\to\infty} \frac{1}{n} c(n,x)$$

exists a.e. and $f \circ T = f$.

Example 2*.* If the random walk has finite first moment (this means $\sum_{g\in G} d(e,g)\mu(g) < \infty$) then $c_1(1,\cdot) \in L^1$. So the Theorem implies that

$$\lim_{n\to\infty} \frac{d(e, x_1 \cdots x_n)}{n}$$

exists for a.e. x. This is called the **speed** of the random walk.

If $H(\mu) < \infty$ then $c_3(1,\cdot) \in L^1$. In this case

$$\lim_{n\to\infty} \frac{-\log \mu^n(\{x_1 \cdots x_n\})}{n} = \lim_{n\to\infty} H(\mu^n)/n$$

exists for a.e. x. This is called the **entropy** of the random walk. It is not the same as the Kolmogorov-Sinai entropy of the shift action on $(G^{\mathbb{N}}, \mu^{\mathbb{N}})$. For example, the entropy of any random walk on $\mathbb{Z}^d$ is zero.

Proof. Wlog we may assume $c(n,x) \le 0$ for all n and a.e. x. Reason: let $c'(n,x) = c(n,x) - c(1,x) - c(1,Tx) - \cdots - c(1,T^{n-1}x)$. Note $c'(n,x) \le 0$ and $c'(n,x)$ satisfies the theorem iff $c(n,x)$ does (b/c of Birkhoff's ergodic theorem).

Let $\bar{f}(x) := \liminf_{n\to\infty} \frac{1}{n} c(n,x)$. Apriori this may be $-\infty$. Because $c(n+1,x) \le c(n,x) + c(1,Tx)$, we have $\bar{f} \le \bar{f} \circ T$ a.e. This implies $\bar{f} = \bar{f} \circ T$ (since $\int \bar{f} = \int \bar{f} \circ T$ since T is measure-preserving).

For $m > 0$, let $\bar{f}_M(x) = \max(\bar{f}(x), -M)$. It suffices to show $\limsup_n c(n,x)/n \le \bar{f}_M$.

Let X_0 be the set of $x \in X$ such that $\bar{f}(x) = \bar{f}(T^k x)$ for all $k \geq 1$. This set has measure 1.

$$\mathrm{Bad(N, M, \epsilon) = \{x \in X_0 : \ c(l, x)/l > \bar{f}_M(x) + \epsilon \ \forall 1 \leq l \leq N\}.}$$

Fix $x_0 \in X_0$. We estimate $c(n, x_0)$ as follows. Write

$$\{x_0, Tx_0, \ldots, T^n x_0\} = \sqcup_{i=1}^{g} \{T^{\tau_i} x_0, \cdots, T^{\tau_i + l_i} x_0\}$$

$$\sqcup \{T^j x_0 \in \mathrm{Bad(N, M, \epsilon), 0 \leq j \leq n - N\} \cup \{T^k x_0 : \ n - N < k \leq n, T^k x_0 \notin Bad(N, M, \epsilon)\}.}$$

In the above decomposition we assume that the orbit seqments $\{T^{\tau_i} x_0, \cdots, T^{\tau_i + l_i} x_0\}$ have been chosen so that $c(l_i, T^{\tau_i} x_0) \leq \bar{f}_M(x) + \epsilon$ and these segments are maximal with this condition. So we really partitioned this orbit segment.

By the subadditivity condition and nonpositivity of c,

$$\begin{aligned} c(n, x_0) &\leq \sum_{i=1}^{g} c(l_i, T^{\tau_i} x_0) + \sum_{j : T^j x_0 \in \mathrm{Bad(N,M,\epsilon)}} c(1, T^j x_0) + \sum_k c(1, T^k x_0) \\ &\leq \sum_{i=1}^{g} c(l_i, T^{\tau_i} x_0) \\ &\leq (l_1 + \cdots + l_g)(\bar{f}_M(x) + \epsilon). \end{aligned}$$

Now

$$n \leq (l_1 + \cdots + l_g) + \#\{1 \leq j \leq n - N : T^j x_0 \in \mathrm{Bad(N, M, \epsilon)}\} + \mathrm{N}.$$

By Birkhoff's ergodic theorem,

$$\lim_{n \to \infty} \frac{l_1 + \cdots + l_g}{n} = 1 - \mathbb{E}[1_{\mathrm{Bad(N,M,\epsilon)}} | \mathfrak{Inv}](x)$$

for a.e. x. So we obtain

$$\limsup_n c(n, x)/n \leq (\bar{f}_M(x) + \epsilon)(1 - \mathbb{E}[1_{\mathrm{Bad(N,M,\epsilon)}} | \mathfrak{Inv}](x))$$

for a.e. x.

Let $N \to \infty$ to finish the proof of the first statement.

□

Proposition 29.2. *If T is ergodic in the theorem above then f equals the constant $\inf_n \frac{1}{n} \int c(n, x) \, d\mu(x)$ a.e.*

More applications: first-passage percolation, cocycles into $GL(n, \mathbb{R})$.

Theorem 29.3 (Furstenberg-Kesten)**.** *Let $T \in \mathrm{End}(X, \mu)$, $c : \mathbb{N} \times X \to GL(d, \mathbb{R})$ be a cocycle. (So $c(n + m, x) = c(n, T^m x) c(m, x)$). If $\log \|c(1, \cdot)\| \in L^1(X, \mu)$ then*

$$\lambda(x) := \lim_{n \to \infty} \frac{1}{n} \log \|c(n, x)\|$$

exists a.e. Moreover $\lambda = \lambda \circ T$ and if T is ergodic then $\lambda = \inf_n \frac{1}{n} \int \log \|c(n, x)\| \, d\mu(x)$.

This theorem predates Kingman's Subadditive Ergodic Theorem.

30 Oseledec's Multiplicative Ergodic Theorem

Theorem 30.1 (Oseledec's Multiplicative Ergodic Theorem)**.** *Let* $T \in \mathrm{End}(X, \mu)$*,* $c : \mathbb{N} \times X \to \mathrm{Mat(d,\mathbb{R})}$ *be a cocycle. (So* $c(n+m,x) = c(n, T^m x)c(m,x)$*). If* $\log \|c(1,\cdot)^{\pm 1}\| \in L^1(X,\mu)$ *then*

$$\Lambda(x) := \lim_{n\to\infty} \left[c(n,x)^* c(n,x)\right]^{1/2n}$$

exists a.e. and

$$\lim_{n\to\infty} \frac{1}{n} \|(c(n,x)\Lambda(x)^{-1})^{\pm 1}\| = 0.$$

Moreover, if $\exp(\lambda_1(x)) < \cdots < \exp(\lambda_s(x))$ *are the distinct eigenvalues of* $\Lambda(x)$ *with corresponding eigenspaces* $E_1, \ldots, E_s$ *and if* $V_j = \oplus_{i\le j} E_i$ *then*

1. *for* $v \in V_j \setminus V_{j-1}$*,*
$$\lim_{n\to\infty} \frac{1}{n} \log \|c(n,x)v\| = \lambda_j.$$
2. *the functions* $x \mapsto \lambda_i(x)$ *and* $x \mapsto \dim(E_i(x))$ *are* T*-invariant.*

The λ_i's above are called the **Lyapunov exponents**.

30.1 Example

Suppose X is a smooth Riemannian compact manifold and T is a diffeomorphism. Let $f : TX \to \mathbb{R}^d$ be a measurable map such that f restricted to a.e. tangent space $T_x(X)$ is a linear isomorphism and an isometry with respect to the Riemannian metric. Define $c : \mathbb{N} \times X \to GL(d, \mathbb{R})$ by

$$c(1,x) = f \circ dT \circ (f \restriction T_x(X))^{-1}$$

and $c(n,x) = c(1, T^{n-1}x)c(1,T^{n-2}x)\cdots c(1,x)$ for $n > 1$. By compactness, this cocycle satisfies the conditions of Oseledec's Theorem.

You might imagine that the Lyapunov exponents depend on the choice of f or the choice of Riemannian metric. Under mild conditions, this is not true. The reason is that if f_1, f_2 are two such maps (possibly corresponding to two Riemannian metrics) with corresponding cocycles c_1, c_2 then

$$c_2(n,x) = \phi(T^n x)c_1(n,x)\phi^{-1}(x)$$

where $\phi : X \to GL(d,\mathbb{R})$ is the map

$$\phi(x) = f_2 \circ (f_1 \restriction T_x(X))^{-1}.$$

Because f_1, f_2 are isometries and X is compact, $\|\phi\| \in L^\infty$ implies c_1, c_2 have the same Lyapunov exponents.

We will sketch the proof. But first,...

30.2 A bit of linear algebra

A matrix A is **positive semi-definite** if for every vector v, $\langle Av, v\rangle \geq 0$. The **adjoint** of A is a matrix A^* defined by $\langle Av, w\rangle = \langle v, A^*w\rangle$ for all v, w. Observe that A^*A and AA^* are positive semi-definite and symmetric for any matrix A. Any positive semi-definite matrix A (over the reals) can be written as

$$A = C^{-1}DC$$

where D is diagonal and C is orthogonal. We can even choose C so that $0 \leq D_{11} \leq D_{22} \leq \cdots \leq D_{nn}$.

The Polar Decomposition Theorem says that any real $n \times n$ matrix A can be written as

$$A = (AA^*)^{1/2}C' = C''(A^*A)^{1/2}$$

the Multiplicative Ergodic Theorem states

$$c(n, x) \approx C'''_{n,x}\Lambda^n$$

for some orthogonal matrix $C'''_{n,x}$. Since $C'''_{n,x}$ is an isometry, this means that for any vector v,

$$\|c(n, x)v\| \approx \|\Lambda^n v\|.$$

For example, if v is an eigenvector of Λ with eigenvalue $\exp(\lambda_i)$ then

$$\|c(n, x)v\| \approx \exp(n\lambda_i)\|v\|.$$

If v is a linear combination of eigenvectors then $\|c(n, x)v\| \approx \exp(n\lambda_i)\|v_i\|$ where $\exp(\lambda_i)$ is the largest eigenvalue appearing in the decomposition and v_i is the projection of v to the corresponding eigenspace.

30.3 Multilinear algebra

Let V denote a finite-dimensional real vector space and $\bigvee_k V$ the k-th exterior power of V. This is a vector space consisting of all real linear combinations of expressions of the form $v_1 \vee \cdots \vee v_k$ with $v_i \in V$ subject to the equations

$$v_1 \vee \cdots \vee (v_i + v_i') \vee \cdots v_k = v_1 \vee \cdots \vee v_i \vee \cdots v_k + v_1 \vee \cdots \vee v_i' \vee \cdots v_k$$

$$v_1 \vee \cdots \vee cv_i \vee \cdots v_k = cv_1 \vee \cdots \vee v_i \vee \cdots v_k$$

for any permutation π of $\{1, \ldots, k\}$,

$$v_{\pi 1} \vee \cdots \vee v_{\pi k} = (-1)^{\mathrm{sign}(\pi)} v_1 \vee \cdots \vee v_k.$$

From a basis of V one can construct, in a canonical way, a basis for $\bigvee_k V$.

Now suppose $A \in \mathrm{Mat(V)}$. Then we define $A^{\vee k} \in \mathrm{Mat}(\bigvee_{\mathrm{k}} \mathrm{V})$ by

$$A^{\vee k}(v_1 \vee \cdots \vee v_k) = Av_1 \vee \cdots \vee Av_k.$$

Lemma 30.2. *$\|A^{\vee k}\| = \lambda_1 \cdots \lambda_k$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of $(A^*A)^{1/2}$ in decreasing order with multiplicities.*

30.4 Sketch of the proof of the multiplicative ergodic theorem

Part 1

Let $\eta_1(n,x) \geq \cdots \geq \eta_d(n,x) \geq 0$ be the eigenvalues of $A(n,x) := \sqrt{c(n,x)^* c(n,x)}$ with multiplicities in decreasing order.

Let $g_j(n,x) = \sum_{i \leq j} \log \eta_i(n,x) = \log \|A(n,x)^{\vee k}\|$.

It can be shown that $\log \|A(n,\cdot)^{\vee k}\| \in L^1(X,\mu)$ (see Sarig's notes). So the subadditive ergodic theorem implies $\lim_{n\to\infty} \frac{1}{n} g_j(n,x)$ exists which implies $\frac{1}{n}\log \eta_i(n,x) =: \eta_i(x)$ exists (and is T-invariant).

Part 2

Let $\lambda_1(x) < \cdots < \lambda_{s(x)}(x)$ be a list of the distinct values of $\{\eta_i(x)\}_i$. Define $\phi : \{1, \dots d\} \to \{1, \dots, s(x)\}$ so that $\eta_i(x) = \lambda_{\phi(i)}(x)$. Let $U_j(n,x)$ be the sum of the eigenspaces of $A(n,x)$ corresponding to the eigenvalues $\eta_i(n,x)$ with $\phi(i) = j$.

Proposition 30.3. *For a.e. x, $U_j(n,x)$ converges to a subspace $U_j(x)$ as $n \to \infty$.*

Proposition 30.4. *The theorem holds with $\Lambda(x)$ defined to be dilation by $\lambda_j(x)$ on $U_j(x)$.*

30.5 Entropy and Lyapunov exponents

Theorem 30.5. *Let X be a C^∞ compact manifold, $f \in \mathrm{End}(X,\mu)$ be a C^1-map. There exists*

1. *a Borel subset $X_0 \subset X$ of full μ-measure*
2. $-\infty \leq \lambda_x^{(1)} < \lambda_x^{(2)} < \cdots < \lambda_x^{s(x)}$
3. *for each $x \in X_0$, a strictly increasing sequence $0 = V_x^{(0)} < V_x^{(1)} < \cdots < V_x^{s(x)} = T_x(X)$ such that for $r = 1 \dots s(x)$,*

   $$\lim_{n\to\infty} \frac{1}{n} \log \|df_x^n v\| = \lambda_x^{(r)}$$

   for $v \in V_x^{(r)} \setminus V_x^{(r-1)}$.
4. *The $V_x^{(r)}, \lambda_x^{(r)}$ are uniquely defined (independent of the choice of Riemannian metric used to define $\|\cdot\|$) and vary measurably with x.*
5. *$\lambda_x^{(r)}$ is f-invariant and $df_x V_x^{(r)} = V_{fx}^{(r)}$*

Moreover,

$$h_\mu(f) \leq \int \sum_{r:\ \lambda_x^{(r)}>0} [\dim(V_x^{(r)}) - \dim(V_x^{(r-1)})]\lambda_x^{(r)}\, d\mu(x).$$

Finally, if f is C^2 then we have equality.

The first statement follows essentially from the Multiplicative Ergodic Theorem. The entropy inequality is due to Margulis and Ruelle. The equality in the case of C^2 maps is due to Pesin.

We now sketch the proof of the inequality $\le$ using Barreira-Pesin's book as a guide.

Using the ergodic decomposition theorem, we may assume wlog that μ is ergodic. So we may write $\lambda_i = \lambda_x^{(i)}$ and $k_i = \dim(V_x^{(i)}) - \dim(V_x^{(i-1)})$. We aim to prove $h_\mu(f) \le \sum_{r:\ \lambda_r > 0} k_r \lambda_r$.

Since X is compact, or each $m \in \mathbb{N}$ there exists $r_m > 0$ such that for every $0 < r < r_m$, $y \in X$, $x \in B(y,r)$,

$$\frac{1}{2} df_x^m(\exp_x^{-1}(B(y,r))) \subset \exp_{f^m x}^{-1} f^m(B(y,r)) \subset 2 df_x^m(\exp_x^{-1}(B(y,r))). \tag{1}$$

Lemma 30.6. *For any $m \in \mathbb{N}, \epsilon > 0$ there exists a partition $\mathcal{P}$ of X satisfying*

1. $h_\mu(f^m, \mathcal{P}) \ge h_\mu(f^m) - \epsilon$
2. *for each $P \in \mathcal{P}$ there exists $x \in P$ and r, r' with $0 < r < 2r' \le r_m/20$ such that*

$$B(x, r') \subset P \subset B(x, r).$$

Proof. Use perturbed Dirichlet domains an $r_m/1000$-separated subset of X of maximum cardinality. □

Next observe that

$$\begin{aligned} h_\mu(f^m, \mathcal{P}) &= H_\mu(\mathcal{P} | \bigvee_{i=1}^{\infty} f^{im}\mathcal{P}) \\ &\le H_\mu(\mathcal{P} | f^m \mathcal{P}) \\ &= -\sum_{P,Q \in \mathcal{P}} \mu(P \cap f^m Q) \log\left(\frac{\mu(P \cap f^m Q)}{\mu(f^m Q)}\right) \\ &= \sum_{P,Q \in \mathcal{P}} \mu(f^m Q) \eta\left(\frac{\mu(P \cap f^m Q)}{\mu(f^m Q)}\right) \\ &= \sum_{Q \in \mathcal{P}} \mu(f^m Q) H_{\mu \restriction f^m Q}(\mathcal{P} \restriction f^m Q) \\ &\le \sum_{Q \in \mathcal{P}} \mu(f^m Q) \log(\#\{P \in \mathcal{P} :\ P \cap f^m Q \ne \emptyset\}) \end{aligned}$$

where $\eta(x) = -x \log x$.

Let $R_m(\epsilon)$ be the set of all $x \in X$ such that for all $k > m$,

$$\Big| \frac{1}{k} \log \|df_x^k v\| - \lambda_r \Big| < \epsilon$$

for all $v \in V_x^{(r)} - V_x^{(r-1)}$. Note $\cup_m R_m(\epsilon)$ has measure 1 in X.

Lemma 30.7. *If $Q \in \mathcal{P}$ is such that $f^m Q \cap R_m \neq \emptyset$ there is a constant $C > 0$ such that*

$$\log \#\{P \in \mathcal{P} : \ P \cap f^m Q \neq \emptyset\} \leq C + \epsilon m + m \sum_{r:\ \lambda_r > 0} k_r \lambda_r .$$

Proof. Let

- $x \in Q \cap f^{-m} R_m$,
- $B = \mathrm{Ball(x, 2diam(Q))}$,
- $\tilde{B}_0 = df_x^m(\exp_x^{-1} B) \subset T_{f^m x}(X)$,
- $B_0 = \exp_{f^m x}(\tilde{B}_0)$,
- $B_1 = \{y \in X : \ d(y, B_0) < \text{max.diam}(\mathcal{P})\}$.

Because $f^m Q \subset B_0$

$$\begin{aligned} \#\{P \in \mathcal{P} : \ P \cap f^m Q \neq \emptyset\} &\leq \#\{P \in \mathcal{P} : \ P \cap B_0 \neq \emptyset\} \\ &\leq \#\{P \in \mathcal{P} : \ P \subset B_1\} \leq \mathrm{vol(B_1) min.vol}(\mathcal{P})^{-1}. \end{aligned}$$

Note $\exp_x^{-1} B$ is a round ball. So $\tilde{B}_0$ is an ellipsoid. Up to a bounded factor, $\mathrm{vol}(B_1)$ is bounded by the volume of $\tilde{B}_0$ which is bounded by the products of the lengths of its axes. Because $f^m x \in R_m$, we can estimate these lengths to obtain:

$$\begin{aligned} \mathrm{vol(B_1)} &\leq K e^{m\epsilon} \mathrm{diam(B)^p} \prod_{\mathrm{i:}\ \lambda_\mathrm{i} > 0} \exp(\mathrm{m}(\lambda_\mathrm{i} + \epsilon)\mathrm{k_i}) \\ &\leq K e^{m\epsilon} \mathrm{diam}(\mathcal{P})^\mathrm{p} \prod_{\mathrm{i:}\ \lambda_\mathrm{i} > 0} \exp(\mathrm{m}(\lambda_\mathrm{i} + \epsilon)\mathrm{k_i}) \end{aligned}$$

which implies the lemma.

□

If $R_m = X$ then the above lemma and the previous estimate imply the main theorem. To handle the case that $R_m \neq X$, we need:

Lemma 30.8. *There exists a constant $K_1 > 0$ such that for $Q \in \mathcal{P}$*

$$\#\{P \in \mathcal{P} : \ P \cap f^m Q \neq \emptyset\} \leq K_1) \sup\{\|df_x\|^{mp} : \ x \in X\}.$$

Proof. Let B denote the volume of the union of the set of all $P \in \mathcal{P}$ such that $P \cap f^m Q \neq \emptyset$. Then

$$\#\{P \in \mathcal{P} : \ P \cap f^m Q \neq \emptyset\} \leq K_1) \leq \mathrm{vol(B)/min.vol}(\mathcal{P}).$$

Now $\mathrm{vol(B) \leq Kdiam(B)^p}$ (where p is the dimension of the manifold and K is a constant).

$$\mathrm{diam(B) \leq 2max.diam}(\mathcal{P}) + \mathrm{diam(f^m Q) \leq 8r' + \sup_x\{\|df_x\|^m\}(4r')}.$$

On the other hand

$$\mathrm{min.vol}(\mathcal{P}) \geq \mathrm{K_2 min.radius}(\mathcal{P})^\mathrm{p} \geq \mathrm{K_2 (r')^p}.$$

□

Proof sketch of Margulis-Ruelle's inequality, via Barreira-Pesin.

$$\begin{aligned}
h_\mu(f^m, \mathcal{P}) &\leq \sum_{Q, f^mQ\cap R_m \neq \emptyset} \mu(Q) \log \#\{P \in \mathcal{P} :\ P \cap f^m Q \neq \emptyset\} \\
&\quad + \sum_{Q, f^mQ\cap R_m = \emptyset} \mu(Q) \log \#\{P \in \mathcal{P} :\ P \cap f^m Q \neq \emptyset\} \\
&\leq \log(C_1) + \epsilon m + m \sum_{r:\ \lambda_r > 0} k_r \lambda_r \\
&\quad + \log(C_2) + pm \log \sup_x \{\|df_x\|\} \mu(X - R_m).
\end{aligned}$$

Since $h_\mu(f) \geq \frac{1}{m} h_\mu(f^m, \mathcal{P})$ and $\cup_m R_m = X$ (mod 0), this implies the theorem.

□

31 Hyperbolic dynamics

(This section mostly follows Brin-Stuck. It also uses Katok-Hasselblatt).

Definition 15. Let

- M be a C^1-Riemannian manifold,
- $U \subset M$ a nonempty open subset,
- $f : U \to f(U) \subset M$ a C^1-diffeomorphism.

A subset $\Lambda \subset M$ is a **hyperbolic set** if it is invariant, compact and there are constants $0 < \lambda < 1, C > 0$ and for every $x \in \Lambda$ a splitting $T_xM = E^s_x \oplus E^u_x$ of the tangent space such that

- $df_x E^s_x = E^s_{fx}$, $df_x E^u_x = E^u_{fx}$
- for every $v \in E^s_x$ and $n \geq 0$, $\|df^n_x(v)\| \leq C\lambda^n \|v\|$,
- for every $v \in E^u_x$ and $n \geq 0$, $\|df^{-1}n_x(v)\| \leq C\lambda^n \|v\|$.

Lemma 31.1. *Let Λ be a hyperbolic set for $f : U \to M$ as above.*

1. *Then the subspaces $E^s(x), E^u(x)$ vary continuously with x.*
2. *$\forall \epsilon > 0$ there is a C^1-Riemannian metric $\langle , \rangle'$ in a neighborhood of Λ with respect to which*
   - *(a) $C' = 1$, $\lambda' = \lambda + \epsilon$*
   - *(b) the subspaces $E^s(x), E^u(x)$ are ϵ-orthogonal (i.e. $v^s \in E^s(x), v^u \in E^u(x) \Rightarrow |\langle v^s, v^u\rangle'| \leq \epsilon \|v^s\|' \|v^u\|'$).*

Proof sketch. The first statement holds by compactness of Λ (if $x_i \to x$ and $w_{i,j}$ is an ON basis of $E^s(x_i)$ then after passing to a subsequence we obtain that $w_{i,j} \to w_j$ is an ON set and the conditions of hyperbolicity imply that w_j is an ON basis for $E^s(x)$).

To prove the second statement, we assume $\lambda + \epsilon < 1$. For $v^s \in E^s(x), v^u \in E^u(x)$, define

$$\|v^s\|' := \sum_{n=0}^{\infty} (\lambda+\epsilon)^{-n} \|df_x^n v^s\|$$

$$\|v^u\|' := \sum_{n=0}^{\infty} (\lambda+\epsilon)^{-n} \|df_x^{-n} v^u\|$$

$$(\|v^s + v^u\|')^2 := (\|v^s\|')^2 + (\|v^u\|')^2.$$

This defines a norm on $T_x(M)$ which defines an inner product by

$$\langle v, w\rangle' = \frac{1}{2}\left((\|v+w\|')^2 - (\|v\|')^2 - (\|w\|')^2\right).$$

This new metric satisfies the lemma except that it might not be C^1. A perturbation gives a C^1-metric satisfying the lemma.

□

31.1 ϵ-orbits

Definition 16. An **δ-orbit** of $f : U \to M$ is a finite or infinite sequence $(x_n) \subset U$ such that $d(f(x_n), x_{n+1}) < \delta$ for all n. We say that a sequence (y_n) **ϵ-shadows** (x_n) if $d(x_n, y_n) < \epsilon$ for all n.

More generally:

Definition 17. Let $h : X \to X$ be a homeomorphism of a topological space. Then a **δ-map** from h to f is a continuous map $\phi : X \to M$ such that $d(\phi(hx), f\phi(x)) < \delta$ for all x. We say that a continuous map $\psi : X \to M$ **ϵ-shadows** ϕ if $d(\phi(x), \psi(x)) < \epsilon$ for all x. These concepts generalize δ-orbits and ϵ-shadowing.

The main theorem states that if $f : U \to M$ has hyperbolic set $\Lambda \subset U$ and g is a small perturbation of f then any δ-map to g is shadowed by a unique actual semiconjugacy to g.

Theorem 31.2. *Let Λ be a hyperbolic set for $f : U \to M$. Then*

$$\exists O \subset^{open} M \text{ and } \epsilon_0, \delta_0 > 0$$

such that for every $\epsilon > 0$ there exists $\delta > 0$ such that if

- $g : O \to M$, $\mathrm{dist}_1(\mathrm{g}, \mathrm{f}) < \epsilon_0$
- $h : X \to X$ *(homeo of top space),* $\phi : X \to O$ *is a δ-map from h to g*

then there is a unique semi-conjugacy $\psi : X \to M$ (from h to g) that δ_0-shadows ϕ.

Before proving this, consider some special cases. For example, suppose $X = \{*\}$ has only one point. Then h is the identity. Let $x_0 = \phi(*)$. Then $d(gx_0, x_0) < \delta$. So the theorem implies: if g is close enough to f and $x_0 \in X$ satisfies $d(gx_0, x_0) < \delta$ then there is an actual fixed point for g close to x_0. For example, suppose x_0 is a fixed point for f. Then any perturbation of f contains has a fixed point near x_0.

Next suppose $X = \mathbb{Z}/n\mathbb{Z}$ and $h : X \to X$ is addition by 1 mod n. Then theorem says that any closed δ-orbit of g of length n is δ_0-close to an actual periodic orbit of length $\leq n$. Again this implies that periodic orbits of f are stable under perturbations.

Lastly suppose $X = \mathbb{Z}$ (we never assumed X was compact so this is OK) and h is addition by 1. Then the theorem says that any δ-orbit of g is δ_0-shadowed by an actual orbit of g. So in this sense the dynamics of f are "stable". We will see more stability results later on.

We will use:

Theorem 31.3 (Contraction Mapping Principle)**.** *Let (X, d) be a complete metric space and $F : X \to X$ a contraction: $\exists 0 < C < 1$ such that $d(Fx, Fy) \leq Cd(x, y) \forall x, y \in X$. Then F admits a unique fixed point x^*. Moreover, for any $x \in X$, $\lim_n F^n x = x^*$.*

Proof of Theorem 31.2. The idea of the proof is to use the contraction mapping principle. Roughly, we look at a large space of maps from X to M and obtain ψ as the unique fixed point of some contraction on this space of maps. More precisely, we embed M into $\mathbb{R}^N$ and then consider a space of maps from X into $\mathbb{R}^N$.

By the Whitney embedding theorem and the tubular neighborhood theorem, we can embed M into $\mathbb{R}^N$ (for some large N) so that there is some α such that if $D_\alpha(y)$ is the $N - m$ dimensional flat disk of radius α centered at $y \in M$ and perpendicular to $T_y(M)$ then $\sqcup_{y \in O} D_\alpha(y) = O_\alpha \subset \mathbb{R}^N$ is the α-open nbhd of O, a relatively compact nbhd of Λ. Let $\pi : O_\alpha \to O$ be projection along the disks. So $\pi(D_\alpha(y)) = y$.

Any map $g : O \to M$ we extend to $\tilde{g} : O_\alpha \to M$ via $\tilde{g}(x) = g(\pi(x))$. Let

- $C(X, \mathbb{R}^N)$ be the space of all continuous maps from X to $\mathbb{R}^N$ with the sup norm
- $B_\alpha \subset C(X, \mathbb{R}^N)$ the ball of radius α centered at 0
- $\Phi : B_\alpha \to C(X, \mathbb{R}^N)$,

$$\Phi(v)(x) = \tilde{g}[\phi(h^{-1}(x)) + v(h^{-1}(x))] - \phi(x), \quad v \in B_\alpha, x \in X.$$

Suppose $\psi : X \to O$ is a map and $\psi(x) = \phi(x) + v(x)$. Then ψ is equivariant iff v is a fixed point of Φ. So it would be good to show that Φ is contracting but that's not what we do. Instead we define

$$F : B_\alpha \to C(X, \mathbb{R}^N)$$
$$F(v) := (d\Phi_0 - \mathrm{Id})^{-1}(\mathrm{d}\Phi_0(\mathrm{v}) - \Phi(\mathrm{v})),$$

show that F is well-defined, that fixed points of F and of Φ are the same, and that F is contracting (if constants have been chosen appropriately).

Hyperbolicity plays a role in showing that the spectrum of $d\Phi_0$ misses a neighborhood of the unit circle so that $d\Phi_0 - \mathrm{Id}$ is invertible and F is well-defined.

The derivative of Φ is defined as usual: for $v \in B_\alpha$, $d\Phi_v : C(X, \mathbb{R}^N) \to C(X, \mathbb{R}^N)$ is given by

$$d\Phi_v(w) \quad := \quad \lim_{t \to 0} \frac{1}{t}(\Phi(v + wt) - \Phi(v))$$

So

$$\begin{aligned} d\Phi_v(w)(x) \quad &:= \quad \lim_{t \to 0} \frac{1}{t}(\tilde{g}[\phi(h^{-1}(x)) + v(h^{-1}(x)) + tw(h^{-1}(x))] - \tilde{g}[\phi(h^{-1}(x)) + v(h^{-1}(x))]) \\ &= \quad d\tilde{g}_{\phi(h^{-1}(x)) + v(h^{-1}(x))}(w(h^{-1}(x)). \end{aligned}$$

By compactness, we have

$$\|d\Phi_v(w)\| \leq L\|w\|$$

for some constant $L > 0$ depending on dg and the embedding but not on X, ϕ, h.

Again, we want to show that the spectrum of $d\Phi_0$ is separated from the unit circle. To prove this, we need to extend the distributions E^s, E^u to a neighborhood of M in $\mathbb{R}^N$ and study $d\tilde{g}$ on these distributions.

To be precise let

- $\tilde{T}_z$ be the flat plane in $\mathbb{R}^N$ that passes through z and is parallel to $T_{\pi(z)}(M)$
- choose distributions $E^s(z) \oplus E^u(z) = \tilde{T}_z$ to depend continuously on $z \in O_\alpha$ (decreasing the nbhd O_α if necessary)
- $E^\perp = T_z(\mathbb{R}^N) \ominus \tilde{T}_z$
- $P^s, P^u, P^\perp : T_z(\mathbb{R}^N) \to T_z(\mathbb{R}^N)$ denote projections to $E^s, E^u, E^\perp$ respectively
- $C, \lambda > 0$ be the constants in the definition of hyperbolicity of Λ,
- $n > 0$ large enough so that $C\lambda^n < 1/2$
- $\epsilon_0 > 0$ small enough so that if $\mathrm{dist}_1(\mathrm{f}, \mathrm{g}) < \epsilon_0$, $z \in O_\alpha$ then

$$\|P^s d\tilde{g}_z^n P^s\| \leq 1/2, \quad \|P^u d\tilde{g}_z^n P^s\| \leq 1/100,$$

$$\|P^u d\tilde{g}_z^n P^u\| \geq 2, \quad \|P^s d\tilde{g}_z^n P^u\| \leq 1/100,$$

$$d\tilde{g}_z^n P^\perp = 0.$$

So if we write $d\tilde{g}_z^n : T_z(\mathbb{R}^N) \to T_{\tilde{g}^n(z)}(\mathbb{R}^N)$ using the coordinate system $E^s \oplus E^u \oplus E^\perp$, we have

$$d\tilde{g}_z^n = \begin{pmatrix} A^{ss} & A^{su} & 0 \\ A^{us} & A^{uu} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

with $\|A^{ss}\| \leq 1/2, \|A^{su}\| < 1/100, \|A^{us}\| < 1/100$ and $\|A^{uu}\| > 2$. In particular, the spectrum of $d\tilde{g}_z^n$ is bounded away from the unit circle. This implies the same of $d\tilde{g}_z$ and since

$$d\Phi_0(w)(x) = d\tilde{g}_{\phi(h^{-1}(x))}(w(h^{-1}(x))$$

we have the same result of $d\Phi_0 : C(X, \mathbb{R}^N) \to C(X, \mathbb{R}^N)$ if $\mathrm{dist}(\phi \mathrm{h}, \tilde{\mathrm{g}}\phi)$ is small enough.

So under the above assumptions, $d\Phi_0 - \mathrm{Id}$ is invertible and there is a constant $K > 0$ such that

$$\|(d\Phi_0 - \mathrm{Id})^{-1}\| \leq \mathrm{K}.$$

This shows that F above is well-defined. Recall

$$F(v) := (d\Phi_0 - \mathrm{Id})^{-1}(\mathrm{d}\Phi_0(\mathrm{v}) - \Phi(\mathrm{v})).$$

By the usual arguments if $v, w \in C(X, \mathbb{R}^N)$ have small enough norm then

$$\|d\Phi_0(v) - \Phi(v) - d\Phi_0(w) + \Phi(w)\| \leq C \max(\|v\|, \|w\|)\|v - w\|.$$

It follows that F is contractive. So it has a unique fixed point. Moreover the fixed point depends continuously on g.

If $F(v) = v$ then

$$d\Phi_0(v) - v = d\Phi_0(v) - \Phi(v)$$

implies $v = \Phi(v)$ (and vice versa).

□

Definition 18. Let $f : M \to M$ be a continuous map. A point $x \in M$ is **nonwandering** if for every open neighborhood O of x there is some $n > 0$ such that $f^n(O) \cap O \neq \emptyset$. Let $NW(f)$, the **nonwandering set** of f be the set of all nonwandering points. Note: $NW(f)$ is closed since if $x \in M$ is wandering there is an open neighborhood $O \ni x$ of wandering points.

Corollary 31.4. *If $f : U \to M$ has hyperbolic set Λ then $NW(f) \cap \Lambda = \overline{\mathrm{Per(f)} \cap \Lambda}$.*

Proof. Clearly, $\overline{\mathrm{Per(f)} \cap \Lambda} \subset NW(f) \cap \Lambda$. On the other hand, let $x \in NW(f) \cap \Lambda$ and $\epsilon > 0$. We'll show there is a periodic point z with $d(z, x) < \epsilon$.

Let $O \supset \Lambda$, $\delta_0, \epsilon_0, \delta > 0$ be as in the previous theorem. Let $O_x \subset O$ be the $\min(\delta/2, \epsilon)$-neighborhood of x and let $n > 0$ be such that $f^n O_x \cap O_x \neq \emptyset$. Let $y \in f^n O_x \cap O_x$. Define $\phi : \mathbb{Z}/n\mathbb{Z} \to O$ by $\phi(k) = f^k y$ for $0 \leq k \leq n - 1$. Let $h : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ be the add 1 map. Since $d(\phi^n y, y) < \delta$ we have $\mathrm{dist}_0(\phi \mathrm{h}, \mathrm{g}\phi) < \delta$. So the theorem implies the existence of an equivariant map $\psi : \mathbb{Z}/n\mathbb{Z} \to O$ with $\mathrm{dist}(\psi, \phi) < \epsilon$. Let $z = \psi(n\mathbb{Z})$. This is a periodic point with $d(z, x) < \epsilon$. □

Corollary 31.5. *Let $f : U \to M$ have hyperbolic set Λ. Then for every open V with $\Lambda \subset V \subset U$ and $\epsilon > 0$, $\exists \delta > 0$ such that if $g : V \to M$, $\mathrm{dist}_1(\mathrm{g}, \mathrm{f}) < \delta$ then $\exists$ hyperbolic set $K \subset V$ for g and a homeomorphisms $\psi : \Lambda \to K$ such that $\psi f = g\psi$ and $\mathrm{dist}_0(\psi, \mathrm{Id}) < \epsilon$.*

Partial proof. We will not prove that K is hyperbolic yet; that will have to wait. By the main theorem above (taking $\Lambda = X$, $h = f \restriction \Lambda$ and ϕ the inclusion map), we obtain $\psi : \Lambda \to V$ such that $\psi f = g\psi$ and $\mathrm{dist}_0(\psi, \mathrm{Id}) < \epsilon$. By applying the main theorem again with f and g swapped (and $K = \Lambda$), we obtain $\psi' : K \to \Lambda$ with $\psi' g = f\psi'$. By uniqueness in the main theorem, we must have $\psi' = \psi^{-1}$. This proves ψ is a homeomorphism (not just a semi-conjugacy). □

Example: Λ could be a hyperbolic periodic point. In this case, the corollary says that there is a hyperbolic periodic point for g nearby. Moreover, it has the same period.

Definition 19 (Structural stability)**.** Let $f \in \mathrm{Diff}^1(\mathrm{M})$. We say f is **structurally stable** if for every $\epsilon > 0$ $\exists \delta > 0$ such that $g \in \mathrm{Diff}^1(\mathrm{M})$, $\mathrm{dist_1(f, g)} < \delta$ implies $\exists h \in \mathrm{Homeo(M)}$ s.t. $hf = gh$ and $\mathrm{dist_0(h, Id)} < \epsilon$. It is important in this definition that h is only required to be a homeomorphism rather than a C^1-map. For example, suppose f is Anosov and has a fixed point at x and g is a small perturbation of f that also has a fixed point at x. If h is differentiable then df_x and dg_x are similar matrices. But this condition is too strong to hold in general.

Corollary 31.6. *Anosov diffeomorphisms are structurally stable.*

31.2 Invariant cones

How do we prove that the set K in the corollary above is a hyperbolic set for g? This means showing that it has stable and unstable subspaces. Giving a precise description of these subspaces is too difficult in general. In fact, it is known that the local stable (and unstable) manifolds are usually not differentiable (only Hölder continuous) so we should not expect a precise formula for their tangent spaces. Instead we take a softer approach through the use of "invariant cones". This device is broadly useful for proving hyperbolicity (not just for immediate purposes).

Definition 20. Let U be an open subset of M and suppose for each $x \in U$ there is a splitting $T_xM = \tilde{E}^s(x) \oplus \tilde{E}^u(x)$ that varies continuously with x. For $v \in T_xM$, write

$$v = v^s + v^u$$

with $v^s \in \tilde{E}^s, v^u \in \tilde{E}^u$. For $\alpha > 0$, define

$$K^s_\alpha(x) = \{v \in T_x(M) : \ \|v^u\| \leq \alpha \|v^s\|\}$$

$$K^u_\alpha(x) = \{v \in T_x(M) : \ \|v^s\| \leq \alpha \|v^u\|\}.$$

These are the **stable and unstable cones of size** α. Note: it's important that we do not require that $\tilde{E}^s, \tilde{E}^u$ come from the stable/unstable decomposition inherent in the definition of hyperbolicity. We do not even need a map to define these cones.

Theorem 31.7. *Let Λ be a compact invariant set of $f : U \to M$. Let U be an open subset of M and suppose for each $x \in U$ there is a splitting $T_xM = \tilde{E}^s(x) \oplus \tilde{E}^u(x)$ that varies continuously with x. Suppose there is an $\alpha > 0$ such that the α-cones satisfy*

1. $df_x K^u_\alpha(x) \subset \mathrm{int}(\mathrm{K}^u_\alpha(\mathrm{f(x)})) \cup \{0\}$
2. $df^{-1}_{fx} K^s_\alpha(f(x)) \subset \mathrm{int}(\mathrm{K}^s_\alpha(\mathrm{x})) \cup \{0\}$
3. $\|df_x v\| < \|v\|$ *for nonzero* $v \in K^s_\alpha(x)$,
4. $\|df^{-1}_x v\| < \|v\|$ *for nonzero* $v \in K^u_\alpha(x)$.

Then Λ is a hyperbolic set for f.

Observe that the conditions above are open conditions. Hence they imply that the set K in Corollary 31.5 is hyperbolic (for appropriate $\delta > 0$) finishing its proof.

Proof. By compactness of Λ there is a constant $\lambda \in (0, 1)$ such that

$$\|df_x v\| \leq \lambda \|v\| \ v \in K^s_\alpha(x)$$

$$\|df^{-1}_x v\| \leq \lambda \|v\|, \ v \in K^u_\alpha(x).$$

For $x \in \Lambda$, the subspaces

$$E^s(x) := \bigcap_{n \geq 0} df^{-n}_{f^n x} K^s(f^n(x))$$

$$E^u(x) := \bigcap_{n \geq 0} df^{n}_{f^{-n} x} K^u(f^{-n}(x))$$

satisfy the definition of hyperbolicity with $C = 1$.

Check: let $V^u_n = df^n_{f^{-n}x} \tilde{E}^u(f^{-n}(x))$. Let V^u_∞ be a limit point of the subspaces $\{V^u_n\}_n$ (in the Grassmannian of $T_x(M)$). The intersection defining $E^u(x)$ is a decreasing intersection because of (1). So $V^u_\infty \subset E^u(x)$.

Claim 1: $V^u_\infty = E^u(x)$.

Claim 2: V^u_∞ is transverse to $\tilde{E}^s(x)$. This is true because every subspace contained in the interior of $K^u_\alpha(x)$ is transverse to every subspace contained in the interior of $K^s_\alpha(x)$.

By Claim 2, given any $v \in T_x(M)$, we can write $v = a + b$ where $a \in V^u_\infty$ and $b \in \tilde{E}^s(x)$. Suppose that $v \in E^u(x)$. Then

$$\|df^{-n}_x v\| \leq \lambda^n \|v\|$$

for all $n \geq 0$. For the same reason,

$$\|df^{-n}_x a\| \leq \lambda^n \|a\|.$$

By (2), $df^{-n}_x b \in K^s_\alpha(f^{-n}x)$. So

$$\|b\| \leq \lambda^n \|df^{-n}_x b\| \leq \lambda^n \|df^{-n} v\| + \lambda^n \|df^{-n} a\| \leq \lambda^{2n}(\|v\| + \|a\|).$$

Since this is true for all n, $\|b\| = 0$ and $v \in V^u_\infty$. This proves Claim 1.

Claim 1 finishes the theorem because it shows $E^u(x), E^s(x)$ are transverse subspaces with complementary dimensions.

□

31.3 Local stable/unstable manifolds

Suppose M is a smooth manifold and for every $x \in M$ we have a subspace $E(x) \subset T_x(M)$ that varies continuously with x. Is there a submanifold $W(x)$ in a neighborhood of x such that for each $y \in W(x)$, $T_yW = E(y)$? If so then $x \mapsto E(x)$ is said to be **integrable**. For example, every vector field is integrable. However, there are non-integrable 2-plane fields on 3-manifolds. The question we address here is: if $f : U \to M$ has hyperbolic set Λ then are $x \mapsto E^s(x)$ and $x \mapsto E^u(x)$ integrable? Actually this question isn't perfect: imagine if Λ is totally disconnected. Then we would really like to extend the distributions E^s, E^u to a neighborhood of Λ and ask whether the extensions are integrable. The answer is 'yes' as we will see. The key tool is:

Theorem 31.8 (Hadamard-Perron)**.** *Let*

- *B_δ be the ball of radius $\delta > 0$ centered at 0 in $\mathbb{R}^m$,*
- *$\{f_n\}_{n\in\mathbb{N}} : B_\delta \to \mathbb{R}^m$ be a sequence of C^1-diffeomorphisms onto their images with $f_n(0) = 0 \forall n$,*
- $0 < \lambda < 1$
- *$\mathbb{R}^m = E^s(n) \oplus E^u(n)$ a sequence of splittings satisfying*
  - *(invariance) $df_n(0)E^s(n) = E^s(n+1)$, $df_n(0)E^u(n) = E^u(n+1)$*
  - *(contracting) $\|df_n(0)v\| \leq \lambda\|v\|$ for all $v \in E^s(n)$,*
  - *(expanding) $\|df_n(0)v\| \geq \lambda^{-1}\|v\|$ for all $v \in E^u(n)$,*
  - *(angle bound) the angle between $E^s(n)$ and $E^u(n)$ is uniformly bounded away from zero,*
  - *(equicontinuity) $\{df_n(\cdot)\}_n$ is uniformly equicontinuous as functions from B_δ to $GL(m, \mathbb{R})$.*

Then there exist

- $\epsilon > 0$
- *$\{\phi_n\}_{n\in\mathbb{N}}$ a uniformly Lipschitz family of functions $\phi_n : B^s_\epsilon(n) := B_\epsilon \cap E^s(n) \to E^u(n)$ such that*
  1. *(stability)* $\mathrm{graph}(\phi_\mathrm{n}) \cap \mathrm{B}_\epsilon = \mathrm{W}^\mathrm{s}_\epsilon(\mathrm{n}) := \{\mathrm{x} \in \mathrm{B}_\epsilon : \ \|\mathrm{f}_{\mathrm{n+k}} \circ \mathrm{f}_{\mathrm{n+k-1}} \circ \cdots \circ \mathrm{f}_\mathrm{n}(\mathrm{x})\| \to_{\mathrm{k}\to\infty} 0\}$,
  2. *(invariance)* $f_n(\mathrm{graph}(\phi_\mathrm{n})) = \mathrm{graph}(\phi_{\mathrm{n+1}})$,
  3. *(contraction)* $\forall x \in \mathrm{graph}(\mathrm{f}_\mathrm{n})$, $\|f_n(x)\| \leq \lambda\|x\|$

4. *(expansion)* $\forall x \in B_\epsilon - \mathrm{graph}(\phi_n)$,
$$\|P^u_{n+1} f_n(x) - \phi_{n+1}(P^s_{n+1} f_n(x))\| > \lambda^{-1} \|P^u_n x - \phi_n(P^s_n(x))\|$$
where, for example, P^s_n *is projection onto* $E^s(n)$ *along* $E^u(n)$ *(this is not the usual orthogonal projection).*
5. *(tangent space)* ϕ_n *is differentiable at* 0 *and the tangent space to* $\mathrm{graph}(\phi_n(0))$ *at* 0 *is* $E^s(n)$
6. *(continuity)* $\{\phi_n\}$ *depends continuously on* $\{f_n\}$ *with respect to the topologies induced by*
$$d_0(\phi, \psi) = \sup_{n \in \mathbb{N}, x \in B_\epsilon} 2^{-n} |\phi_n(x) - \psi_n(x)|$$
$$d(f, g) = \sup_n 2^{-n} \mathrm{dist}_1(\mathrm{f_n}, \mathrm{g_n}).$$

Proof. For $\epsilon, L > 0$, let $\Phi(L, \epsilon)$ denote the space of all sequences $\{\phi_n\}$, $\phi_n : B^s_\epsilon(n) \to E^u(n)$ such that each ϕ_n is L-Lipschitz and $\phi_n(0) = 0$. Define a metric on $\Phi(L, \epsilon)$ by
$$d(\phi, \psi) := \sup_{n \in \mathbb{N}, x \in B_\epsilon} |\phi_n(x) - \psi_n(x)|.$$

The main idea of the proof is to define a contracting map F on $\Phi(L, \epsilon)$ (for appropriate ϵ, L) such that the unique fixed point of F satisfies the theorem.
Claim 1. $\forall L > 0\ \exists \epsilon > 0$ such that $\forall \phi \in \Phi(L, \epsilon)$, $\exists! \psi \in \Phi(L, \epsilon)$ satisfying
$$\mathrm{graph}(\psi_\mathrm{n}) \subset \mathrm{f}_\mathrm{n}^{-1}(\mathrm{graph}(\phi_{\mathrm{n}+1})).$$

Proof of Claim 1. Let
$$K^s_L(n) := \{v \in \mathbb{R}^m :\ \|v^u\| \leq L \|v^s\|\}$$
be the L-stable cone where $v = v^s + v^u$ with $v^s \in E^s(n), v^u \in E^u(n)$.

Observe that $\phi_n : B^s_\epsilon(n) \to E^u(n)$ is L-Lipschitz iff for every $x \in B^s_\epsilon(n)$,
$$\mathrm{graph}(\phi_n) \subset (x, \phi_n(x)) + K^s_L(n).$$

We observe that
$$df_n^{-1}(0) K^s_L(n+1) \subset K^s_L(n)$$
for any n. So by equicontinuity of $\{df_n\}$, there exists $\epsilon > 0$ such that
$$df_n^{-1}(x) K^s_L(n+1) \subset K^s_L(n)$$
$\forall x \in B_\epsilon, n \in \mathbb{N}$. This implies that f_n^{-1} takes graphs of L-Lipschitz functions to graphs of L-Lipschitz functions. Moreover, if $\epsilon > 0$ is small enough then $\forall \phi \in \Phi(L, \epsilon)$,
$$P^s_n f^{-1}_{n+1}(\mathrm{Id}, \phi_{\mathrm{n}+1}) : \mathrm{B}^\mathrm{s}_\epsilon(\mathrm{n}+1) \to \mathrm{E}^\mathrm{s}(\mathrm{n})$$
is expanding. So its image covers $B^s_\epsilon(n)$. □

Define

$$F : \Phi(L, \epsilon) \to \Phi(L, \epsilon)$$

by $F(\phi) = \psi$ where ψ is as in Claim 1.
Claim 2. $\exists L, \epsilon > 0$ such that F is contracting on $\Phi(L, \epsilon)$.

Proof of Claim 2. We observe that

$$df_n(0)K^u_L(n) \subset K^u_L(n+1)$$

for any n. So by equicontinuity of $\{df_n\}$, there exists $\epsilon > 0$ such that

$$df_n(x)K^u_L(n) \subset K^u_L(n+1)$$

$\forall x \in B_\epsilon, n \in \mathbb{N}$.

Choose $0 < L < 1/100$ so that $\frac{\lambda}{1-5L} < 1$.

Let $\phi, \psi \in \Phi(L, \epsilon)$. We will show that $d(F(\phi), F(\psi)) \leq Cd(\phi, \psi)$ where $0 < C < 1$ is a constant.

Let $\phi' = F(\phi), \psi' = F(\psi)$. Choose $0 < \eta$ small enough so that

$$(1 - 4L)\lambda^{-1}(d(\phi', \psi') - \eta) \geq (1 - 5L)\lambda^{-1}d(\phi', \psi').$$

$\exists n$ and $y \in B_\epsilon$ s.t.

$$d(\phi', \psi') \leq |\phi'_n(y) - \psi'_n(y)| + \eta.$$

Let $f_n(y, \psi'_n(y)) = (z, \psi_{n+1}(z))$. Then

$$d(\phi, \psi) \geq |\phi_{n+1}(z) - \psi_{n+1}(z)|.$$

Let $f_n(y, \phi'_n(y)) = (w, \phi_{n+1}(w))$. So

$$|\phi_{n+1}(z) - \psi_{n+1}(z)| \geq |\psi_{n+1}(z) - \phi_{n+1}(w)| - |\phi_{n+1}(w) - \phi_{n+1}(z)|.$$

Let c_u denote the straight-line segment from $(y, \phi'_n(y))$ to $(y, \psi'_n(y))$. Then $f_n(c_u)$ is a path from $(z, \psi_{n+1}(z))$ to $(w, \phi_{n+1}(w))$. If ϵ is small enough then the tangent vectors to $f_n(c_u)$ lie in $K^u_L(n+1)$. Therefore,

$$\mathrm{length(f_n(c_u))} \leq \mathrm{(1 + 2L)|\psi_{n+1}(z) - \phi_{n+1}(w)|}.$$

(To see this, observe that the worst case (when $\frac{\mathrm{length(f_n(c_u))}}{|\psi_{n+1}(z)-\phi_{n+1}(w)|}$ is largest) occurs when $f_n(c_u)$ is a straight line with tangent vectors on the boundary of $K^u_L(n+1)$.) Note $|w - z| \leq \mathrm{length(f(c_u))}$ and $|\phi_{n+1}(w) - \phi_{n+1}(z)| \leq L|w - z|$ which implies

$$|\phi_{n+1}(w) - \phi_{n+1}(z)| \leq L\mathrm{length(f(c_u))}.$$

So

$$d(\phi, \psi) \geq \mathrm{length(f_n(c_u))}\left[\frac{1}{\mathrm{1 + 2L}} - \mathrm{L}\right] \geq \mathrm{(1 - 4L)length(f_n(c_u))}.$$

Now c^u is parallel to $E^u(n)$. So

$$\mathrm{length(f_n(c_u))} \geq \lambda^{-1}\mathrm{length(c_u)} = \lambda^{-1}|\phi'_\mathrm{n}(\mathrm{y}) - \psi'_\mathrm{n}(\mathrm{y})| \geq \lambda^{-1}(\mathrm{d}(\phi', \psi') - \eta)$$

implies

$$d(\phi, \psi) \geq (1 - 4L)\lambda^{-1}(d(\phi', \psi') - \eta) \geq (1 - 5L)\lambda^{-1}d(\phi', \psi').$$

So

$$d(F\phi, F\psi) \leq \frac{\lambda}{1 - 5L}d(\phi, \psi).$$

□

Since F is contracting and depends continuously on f, its fixed point, ϕ depends continuously on f. This proves (6, continuity). (2, invariance) is obvious. (3, contraction) and (4, expansion) follows by equicontinuity of df_n if $\epsilon > 0$ is small enough. (1, stability) is a consequence of (3). We observe that ϕ_n is uniquely defined by (1). Also as $L, \epsilon \searrow 0$, the cone $K^s_L(n)$ tends to $E^s(n)$. Since for every $L > 0$ there is an $\epsilon > 0$ such that $\mathrm{graph}(\phi_\mathrm{n}) \cap \mathrm{B}_\epsilon \subset \mathrm{K^s_L(n)}$ this implies (5, tangent space). □

Let

- $f \in \mathrm{Diff}^1(\mathrm{M})$,
- $\Lambda \subset M$ be a hyperbolic set for f with constant $\lambda \in (0, 1)$ (assume metric is adapted)
- $\Lambda^s_\delta := \{x \in M : \ \mathrm{dist(f^n(x), \Lambda) < \delta \ \forall n \geq 0}\}$
- $\Lambda^u_\delta := \{x \in M : \ \mathrm{dist(f^{-n}(x), \Lambda) < \delta \ \forall n \geq 0}\}$.

Lemma 31.9. *For every $\epsilon > 0$ there is a $\delta > 0$ such that the distributions E^s, E^u can be extended to Λ_δ (the δ-nbhd of Λ) so that*

- *E^s is continuous on Λ^s_δ,*
- *if $x \in \Lambda_\delta \cap f(\Lambda_\delta)$ then $df_x(E^s(x)) = E^s(f(x))$*
- *$\|df_x v\| < (\lambda + \epsilon)\|v\|$ for every $x \in \Lambda_\delta, v \in E^s(x)$*

and similarly with u in place of s and in the third item df_x^{-1} in place of df_x.

Proof. Let $\tilde{E}^s, \tilde{E}^u$ be a continuous extension of E^s, E^u to a nbhd U of Λ. Let $K^{s/u}_\alpha(x)$ denote the corresponding cones.

Choose $\delta > 0$ small enough so that $\Lambda_\delta \subset U$. By continuity there exist $\alpha > 0$ such that

$$df_x K^u_\alpha(x) \subset \mathrm{int K^u_\alpha(f(x))}, \ \mathrm{df_{fx}^{-1} K^s_\alpha(fx) \subset int K^s_\alpha(x)}$$

$$\|df_x^{-1}v\| < (\lambda + \epsilon)\|v\| \quad \forall v \in K^u_\alpha(x)$$
$$\|df_x v\| < (\lambda + \epsilon)\|v\| \quad \forall v \in K^s_\alpha(x).$$

So the following limit exists:

$$E^s(x) := \lim_{n\to\infty} df^{-n}_{f^n x}(\tilde{E}^s(f^n(x))).$$

If $x \in \Lambda_\delta - \Lambda^s_\delta$ then there is a maximum integer $n = n(x) > 0$ such that $f^n x \in \Lambda_\delta$. In this case define

$$E^s(x) = df^{-n}_{f^n x}(\tilde{E}^s(f^n x)).$$

A similar construction works with u in place of s. □

Theorem 31.10 (Local Stable/Unstable Manifolds)**.** *Let f, Λ be as above. Then there exist $\epsilon, \delta > 0$ such that $\forall x^s \in \Lambda^s_\delta, x^u \in \Lambda^u_\delta$,*

- *(def'n) the* **local stable/unstable manifolds**

  $$W^s_\epsilon(x^s) = \{y \in ML\ \mathrm{dist(f^n(x^s), f^n(y))} < \epsilon\ \forall \mathrm{n} \geq 0\},$$
  $$W^u_\epsilon(x^u) = \{y \in ML\ \mathrm{dist(f^{-n}(x^u), f^{-n}(y))} < \epsilon\ \forall \mathrm{n} \geq 0\},$$

  are C^1-embedded disks,
- *(tangent spaces) $\forall y^s \in W^s_\epsilon(x^s)$, $T_{y^s}(W^s_\epsilon(x^s)) = E^s(y^s)$ (and similarly with with u in place of s)*
- *(invariance) $f(W^s_\epsilon(x^s)) \subset W^s_\epsilon(f(x^s))$, $f^{-1}(W^u_\epsilon(x^u)) \subset W^u_\epsilon(f^{-1}(x^u))$,*
- *(contraction) if $y^s, z^s \in W^s_\epsilon(x^s)$ then $d^s(f(y^s), f(z^s)) < \lambda d^s(y^s, z^s)$ where d^s is the distance along W^s_ϵ*
- *(expanstion) if $y^u, z^u \in W^u_\epsilon(x^u)$ then $d^u(f^{-1}(y^u), f^{-1}(z^u)) < \lambda d^u(y^u, z^u)$ where d^u is the distance along W^u_ϵ*
- *(cones) if $0 < \mathrm{dist(x^s, y)} < \epsilon$ and $\exp^{-1}_{x^s}(y) \in K^u_\delta(x^s)$ then*

  $$\mathrm{dist(f(x^s), f(y))} > \lambda^{-1}\mathrm{dist(y, x^s)},$$

  if $0 < \mathrm{dist(x^u, y)} < \epsilon$ and $\exp^{-1}_{x^u}(y) \in K^s_\delta(x^u)$ then

  $$\mathrm{dist(f(x^u), f(y))} < \lambda \mathrm{dist(y, x^u)},$$
- *(compatibility of inclusions) if $y^s \in W^s_\epsilon(x^s)$ then $\exists \alpha > 0$ such that $W^s_\alpha(y^s) \subset W^s_\epsilon(x^s)$, if $y^u \in W^u_\epsilon(x^u)$ then $\exists \alpha > 0$ such that $W^u_\alpha(y^u) \subset W^u_\epsilon(x^u)$.*

Proof. By compactness, there exist a collection of local coordinate charts (U_x, ψ_x) covering Λ_δ (so $\psi_x : U_x \to \mathbb{R}^m$, $\psi_x(x) = 0$, U_x is an open nbhd of x) such that the change of coordinates $\psi_x \psi_y^{-1}$ have equicontinuous first derivatives. For any point $x^s \in \Lambda^s_\delta$, let

$$f_n := \psi_{f^n x^s} \circ f \circ \psi^{-1}_{f^{n-1} x^s}$$
$$E^s(n) := d\psi_{f^n x^s}(x^s) E^s(f^n x^s)$$
$$E^u(n) := d\psi_{f^n x^s}(x^s) E^u(f^n x^s).$$

Apply the previous theorem and set $W^s_\epsilon(x) = \psi_x^{-1}(\mathrm{graph}(\phi_0))$. □

Definition 21. The **global stable and unstable manifolds** are defined by

$$W^s(x) = \{y \in M : \ d(f^n x, f^n y) \to 0 \ n \to \infty\}$$

$$W^u(x) = \{y \in M : \ d(f^{-n} x, f^{-n} y) \to 0 \ n \to \infty\}.$$

Lemma 31.11. *There is an* $\epsilon_0 > 0$ *such that for every* $0 < \epsilon < \epsilon_0$ *and* $x \in \Lambda$,

$$W^s(x) = \cup_{n \geq 0} f^{-n} W^s_\epsilon(f^n(x))$$

$$W^u(x) = \cup_{n \geq 0} f^{n} W^u_\epsilon(f^{-n}(x))$$

31.4 Inclination lemma

Theorem 31.12 (Inclination Lemma)**.** *Let*

- $p \in M$ *be a hyperbolic fixed point for* $f : U \to M$,
- $\epsilon > 0$ *be s.t.* $\lambda + \epsilon < 1$ *and* $\delta > 0$ *be small enough to satisfy Lemma 31.9*
- $q \in W^s(p)$,
- $q \in D$ *be a* C^1*-submanifold of* M *of dimension* $\dim(\mathrm{E^u(p)})$ *transverse to* $W^s(p)$ *at* q

then $\forall \eta > 0$ *there are* $n_0 \in \mathbb{N}$ *and for each* $n \geq n_0$ *a subset* $D' = D'(\eta, n) \subset D$ *diffeo to a* $\dim(\mathrm{E^u(p)})$*-ball such that* $\mathrm{dist_1(f^n(D'), B^u_\delta)} < \eta$. *In particular, for all* $\alpha > 0$ *there exists* n_0 *such that* $n > n_0$ *implies* $\forall z \in D'$, $T_{f^n(z)}(f^n D')$ *is contained in the cone* $K^u_\alpha(f^n z)$.

Proof. This follows from Lemma 31.9 and continuity of $z \mapsto df_z$. □

31.5 Horseshoes

Here we'll show that under mild conditions on a diffeomorphism $f : U \to M$, there exists a compact 'rectangle' $R \subset M$ such that $f \restriction R$ is like Smale' Horseshoe. To be precise, suppose $p \in U$ is a fixed point of f. We say that $q \in U$ is **homoclinic** (for p) if q is both forward and backwards asymptotic to p (under f). Now suppose that p is also a hyperbolic fixed point. Let $W^s(p), W^u(p)$ be the stable and unstable manifolds of p. Then q is homoclinic for p iff $q \in W^s(p) \cap W^u(p)$. We say that q is a **transverse homoclinic point** for p if $W^s(p)$ is transverse to $W^u(p)$ at q.

Our goal: show that if q is a transverse homoclinic point then there exists a horseshoe for some power f^N of f near p.

It will be convenient to work in local coordinates near p. So we write $\mathbb{R}^m = \mathbb{R}^k \times \mathbb{R}^l$ where $m = \dim(M)$, $k = \dim(E^u(p))$, $l = \dim(E^s(p))$ and $p = 0$. We consider balls V^k, V^l in $\mathbb{R}^k, \mathbb{R}^l$ respectively such that $V^k \times V^l$ is a neighborhood of $p = 0$ and such that $\{0\} \times V^l$ and $V^k \times \{0\}$ are the local stable and unstable manifolds. For $z \in V$, let $F^u(z) = V^k \times \pi^s(z)$, $F^s(z) = \pi^u(z) \times V^l$ be fibers of the local stable/unstable foliations.

Let $R = R^u \times R^s \subset V$. We say that a component C of $R \cap f(R)$ is **full** if

- $\forall z \in f^{-1}(C)$, $f(F^u(z)) \cap C$ is connected and its projection to R^u is 1-1 and onto
- $\forall z \in f^{-1}(C)$, $f^{-1}(F^s(fz) \cap C)$ is connected and its projection to R^s is 1-1 and onto

We say that a subset $R \subset V$ is a **horseshoe** for f if f is 1-1 on R, $f(R) \cap R$ has at least 2 full components Δ_0, Δ_1 such that

- $\pi^s(\Delta_0 \cup \Delta_1) \subset \mathrm{int}(\mathrm{R}^s)$
- $\pi^u(f^{-1}(\Delta_0 \cup \Delta_1)) \subset \mathrm{int}(\mathrm{R}^u)$
- $D(f \restriction f^{-1}(\Delta_0 \cup \Delta_1))$ preserves and expands an unstable cone field $K^u_\alpha()$
- $D(f^{-1} \restriction (\Delta_0 \cup \Delta_1))$ preserves and expands a stable cone field $K^s_\alpha()$.

Lemma 31.13. *If R is horseshoe for f and $\Lambda = \cup_{n \in \mathbb{Z}} f^n(\Delta_0 \cup \Delta_1)$ then Λ is a hyperbolic set for f and $f \restriction \Lambda$ is topologically conjugate to the full 2-shift. In particular, f has positive topological entropy.*

Theorem 31.14. *Let $f : U \to M$ have a hyperbolic fixed point p with transverse homoclinic point q. Then every open neighborhood of p contains a horseshoe for some power of f. Moreover, the hyperbolic invariant set Λ in this horseshoe nontrivially intersects the orbit of q.*

31.6 Local product structure

Theorem 31.15. *Let Λ be a hyperbolic set for $f : U \to M$. For every small enough $\epsilon > 0$ there exists $\delta > 0$ such that if $x, y \in \Lambda$ and $d(x,y) < \delta$ then $W^s_\epsilon(x) \cap W^u_\epsilon(y)$ is transverse and contains exactly one point denoted $[x,y]$ which depends continuously on x and y. Furthermore there exists a constant $C = C(\delta) > 0$ such that*

$$d^s(x, [x,y]) < C d(x,y)$$

$$d^u(y, [x,y]) < C d(x,y)$$

where d^s, d^u denote distance along W^s, W^u.

Proof. This is immediate from the local stable/unstable manifold theorem, the compactness of Λ and continuity considerations. □

Theorem 31.16. *Let Λ be a hyperbolic set for $f : U \to M$. TFAE*

1. *Λ is* **locally maximal**. *This means there exists an open set V with $\Lambda \subset V \subset U$ and $\Lambda = \cap_{n \in \mathbb{Z}} f^n V$.*
2. *Λ has a* **local product structure**. *This means there exists $\epsilon, \delta > 0$ such that:*
   (a) $\forall x, y \in \Lambda$, $W^s_\epsilon(x) \cap W^u_\epsilon(y) \subset \Lambda$ and $|W^s_\epsilon(x) \cap W^u_\epsilon(y)| \leq 1$

(b) if $d(x,y) < \delta$ *then* $|W^s_\epsilon(x) \cap W^u_\epsilon(y)| = 1$.

Proof. Suppose Λ is locally maximal, $x, y \in \Lambda$ and $W^s_\epsilon(x) \cap W^u_\epsilon(y) \ni z$. Then the forward orbit and the backward orbit of z both lie in the ϵ-nbhd of Λ (by the contracting and expanding conclusions of the local stable/unstable manifolds theorem). So if ϵ is small enough then the entire orbit of z lies in V which, by local maximality, implies $z \in \Lambda$. This does it (assuming ϵ, δ are sufficiently small).

Suppose Λ has a local product structure with constants $\epsilon, \delta, C > 0$. Assume $C > 1$. Let $0 < \alpha < \lambda\delta/3$. Note that if $x \in \Lambda$ and $y \in W^u_\alpha(x)$ then $f(y) \in W^u_{\delta/3}(f(x))$.

Claim 1. Let $x_0 \in \Lambda$ and $y \in W^u_\alpha(x_0)$. If $\mathrm{dist(f^ny,\Lambda) < \alpha/C}$ for every $n > 0$ then $y \in \Lambda$.

Proof of Claim 1. Let $x_n \in \Lambda$ be s.t. $d(f^n y, x_n) < \epsilon$. Note that

$$d(fx_0, x_1) \le d(fx_0, fy) + d(fy, x_1) \le \delta/3 + \alpha/C < \delta.$$

So $z_1 := [x_1, fx_0] \in \Lambda$. Note $f(y) \in W^u_\alpha(z_1)$. Similarly, $z_2 := [x_2, fx_1] \in \Lambda$ and $f^2y \in W^u_\alpha(z_2)$. By repeating this argument we construct points z_n (for $n \ge 1$) with $z_n \in \Lambda$ and $f^n y \in W^u_\alpha(z_n)$. Thus $f^{-n}z_n \to y$ as $n \to \infty$. Since Λ is closed $y \in \Lambda$. □

Similarly, **Claim 2**. Let $x_0 \in \Lambda$ and $y \in W^s_\alpha(x_0)$. If $\mathrm{dist(f^ny,\Lambda) < \alpha/C}$ for every $n < 0$ then $y \in \Lambda$.

Now suppose the entire orbit of a point $y \in M$ lies in the α-neighborhood of Λ. To finish the proof, it suffices to show $y \in \Lambda$. Using invariant cone fields we can prove that the union of the orbit of y with Λ is a hyperbolic set (assuming α is sufficiently small). So the local stable/unstable manifolds of $f^n y$ are well-defined and so the product $[\cdot,\cdot]$ is defined on $\Lambda \cup \{f^n y\}$. Choose $x_0 \in \Lambda$ with $d(x_0, y) < \alpha$. The forward semi-orbit of $p = [y, x_0]$ stays in the α-neighborhood of Λ. So Claim 1 implies $p \in \Lambda$. The backwards semi-orbit of $q = [x_0, y]$ stays in the α-nbhd of Λ. So Claim 2 implies $q \in \Lambda$. Since $y = [p, q]$ the local product structure implies $y \in \Lambda$. □

All of the examples we have seen are locally maximal (the horseshoe, hyperbolic toral automorphisms, hyperbolic periodic points, etc). It was an open question for about 30-40 years whether every hyperbolic set is contained in a locally maximal hyperbolic set. This was settled in the negative by Crovisier in 2001 and then again by Todd Fisher in 2006 who obtained a wide variety of counterexamples.

31.7 Axiom A and structural stability

A diffeomorphism $f : M \to M$ satisfies **Axiom A** if $NW(f)$ is hyperbolic and $NW(f) = \overline{Per}(f)$.

Theorem 31.17 (Smale's spectral decomposition). *if f satisfies Axiom A then there is a unique representation of $NW(f)$,*

$$NW(f) = \Lambda_1 \cup \cdots \cup \Lambda_k$$

as a disjoint union of closed locally maximal hyperbolic sets (called **basic sets***) s.t.*

1. *f is topologically transitive on each Λ_i*

2. *for each i there is a decomposition $\Lambda_i = \sqcup_{j=1}^{m_i} \Lambda_{i,j}$ into closed sets that are cyclically permuted by f and f^{m_i} is topologically mixing on each $\Lambda_{i,j}$.*

We say f satisfies the **strong transversality condition** if $W^s(x)$ intersects $W^u(y)$ transversely for all x, y at all intersection points.

Theorem 31.18 (Robbin, Robinson, Mane)**.** *A diffeomorphism f is structurally stable iff it satisfies Axiom A and the strong transversality condition.*

32 Interval Exchange Transformations (IETs)

I'll use Katok-Hasselblatt and the Cornfeld-Fomin-Sinai book. By the way, there is a free online book titled "Dynamics of Interval Exchange Transformations and Teichmüller Flows" by M. Viana.

32.1 Rough notes

Main results (from the textbooks):

1. IETs are conjugate to return time maps for certain kinds of flows on surfaces; especially translation flows.
2. IETs have zero entropy
3. IETs are never mixing
4. IETs have only a finite number of invariant measures

Recent papers of interest:

1. Veech, William A. The metric theory of interval exchange transformations. I. Generic spectral properties. Amer. J. Math. 106 (1984), no. 6, 1331–1359.
2. Weak mixing for interval exchange transformations and translation flows. A. Avila, Giovanni Forni. Annals of Mathematics 165 (2007), 637-664.
3. "Every transformation is disjoint from almost every IET" by Jon Chaika, Annals of Math. Pages 237-253 from Volume 175 (2012), Issue 1.
4. Bufetov, Alexander I. Limit theorems for translation flows. Ann. of Math. (2) 179 (2014), no. 2, 431–499.

A curious question (due to Katok): is IET amenable? does it contain free subgroups?

32.2 introduction

Definition 22. A semi-interval is an interval of the form $[a, b)$. Let $\xi = \{\Delta_1, \dots, \Delta_r\}$ be a partition of $[0, 1)$ into semi-intervals numbered from left to right and let π be a permutation of $\{1, \dots, r\}$. The corresponding **interval exchange transformation** (IET) is the map $T : [0, 1) \to [0, 1)$ satisfying

- T restricted to Δ_i is a translation
- if $T\Delta_i = \Delta'_i$ then $\Delta'_1, \dots, \Delta'_r$ is a partition of $[0, 1)$ into semi-intervals numbered from left to right.

If T cannot be realized as an IET on $< r$ regments then we say it is an IET **of precisely** r **segments**.

Example: we can identify $[0, 1)$ with the circle in the usual way. Then rotations can be realized as IETs using precisely 2 segments.

Lemma 32.1. *Let T be an IET. Then either*

- *T is* **aperiodic** *(i.e. has no periodic points) or*
- *there is a sub-interval $I \subset [0, 1)$ consisting of periodic points of T*

Theorem 32.2. *TFAE*

1. *T is aperiodic*
2. $\max \mathrm{diam} \bigvee_{\mathrm{i}=0}^{\mathrm{n}} \mathrm{T}^{\mathrm{i}}\xi \to 0$ *as $n \to \infty$*
3. *the union of the positive semi-orbits of the left endpoints of the Δ_i's is dense in $[0, 1)$.*

32.3 Entropy

Theorem 32.3. *The entropy of any IET T wrt any invariant measure is zero.*

Proof. Wlog T is aperiodic. Then ξ is a generating partition. The number of semi-intervals of $\bigvee_{i=0}^{n} T^i\xi$ is at most nr (because the number of left-endpoints is at most this). So

$$h_\mu(T) = \lim_n \frac{1}{n} H(\bigvee_{i=0}^{n} T^i\xi) \le \lim_n \frac{1}{n} \log(nr) = 0.$$

□

32.4 Invariant measures

Theorem 32.4. *Let T be an aperiodic IET on precisely r intervals. Then the number of ergodic T- invariant probability measures on $[0,1)$ is $\leq r$.*

Proof. Let μ denote a T-invariant probability measure on $[0,1)$ and $\mathcal{H} = L^2([0,1),\mu)$. We do not assume μ is ergodic.

Let $U : \mathcal{H} \to \mathcal{H}$ denote the Koopman operator associated to T.

Let $\mathcal{H}_{inv} \subset \mathcal{H}$ denote the subspace of U-invariant functions. It suffices to show $\mathcal{H}_{inv}$ has dimension $\leq r$.

For $f \in \mathcal{H}$, let $\mathcal{H}(f)$ denote the closure of the subspace spanned by $\{U^n f : n \in \mathbb{Z}\}$. Let f^{inv} denote the orthogonal projection of f onto $\mathcal{H}_{inv}$ and $f^\perp = f - f^{inv}$. By von Neumann's ergodic theorem, $f^{inv} \in \mathcal{H}(f)$. So

$$\mathcal{H}(f) = \mathcal{H}(f^\perp) \oplus \langle f^{inv} \rangle.$$

Let χ_i be the characteristic function of Δ_i. Let $\mathcal{H}' = \mathcal{H}(\chi_1) + \cdots + \mathcal{H}(\chi_r)$. (This sum is not necessarily orthogonal). We'll show that $\mathcal{H} = \mathcal{H}'$. This implies $\mathcal{H}_{inv}$ is spanned by $f_1^{inv}, \ldots, f_r^{inv}$ and is therefore $\leq r$ dimensional.

Because $\max \operatorname{diam} \bigvee_{\mathrm{i=0}}^{\mathrm{n}} \mathrm{T}^{\mathrm{i}} \xi \to 0$ as $n \to \infty$ it follows that the characteristic functions of the sets in $\operatorname{diam} \bigvee_{\mathrm{i=0}}^{\mathrm{n}} \mathrm{T}^{-\mathrm{i}} \xi$ for $n = 1, \ldots$ span a dense subspace of $\mathcal{H}$. So it suffices to prove: for any set

$$\Delta_{i_0,\ldots,i_n} := \Delta_{i_0} \cap T\Delta_{i_1} \cap \cdots \cap T^n \Delta_{i_n}$$

the associated charactistic function, denoted $\chi_{i_0,\ldots,i_n}$ lies in $\mathcal{H}'$.

We will prove this by induction on n. If $n = 0$, the statement is trivial. Assume the statement for n. We must show $\chi_{i_0,\ldots,i_{n+1}} \in \mathcal{H}'$.

Now

$$\chi_{i_0,\ldots,i_{n+1}} = \chi_{i_0} U_T^{-1} \chi_{i_1,\ldots,i_{n+1}}$$

and $\chi_{i_1,\ldots,i_{n+1}} \in \mathcal{H}'$.

Case 1. Assume $i_0 = 1$.

Case 1a. Assume $T\Delta_{i_1,\ldots,i_{n+1}} \subset \Delta_1$.

In this case $\chi_{i_0,\ldots,i_{n+1}} = U_T^{-1} \chi_{i_1,\ldots,i_{n+1}}$ so the statement follows from the inductive hypothesis applied to $\chi_{i_1,\ldots,i_{n+1}}$.

Case 1b. Assume $T\Delta_{i_1,\ldots,i_{n+1}}$ is not a subset of Δ_1.

In this case,

$$\chi_{i_0,\ldots,i_{n+1}} = \chi_{i_0} - \sum U_T^{-1} \chi_{j_1,\ldots,j_{n+1}}$$

where the sum is over all $j_1, \ldots, j_{n+1}$ such that

$$T\Delta_{j_1,\ldots,j_{n+1}} \subset \Delta_1.$$

This proves Case 1b.

Case 2. Assume $i_0 = 2$.

Case 2a. Assume $T\Delta_{i_1,\ldots,i_{n+1}} \subset \Delta_1 \cup \Delta_2$.

In this case,

$$\chi_{i_0,\dots,i_{n+1}} = U_T^{-1}\chi_{i_1,\dots,i_{n+1}} - \chi_{1,i_1,\dots,i_{n+1}}.$$

So the inductive hypothesis and Case 1 proves this case.

Case 2b. Assume $T\Delta_{i_1,\dots,i_{n+1}}$ is not a subset of $\Delta_1 \cup \Delta_2$.

In this case,

$$\chi_{i_0,\dots,i_{n+1}} = \chi_1 + \chi_2 - \sum U_T^{-1}\chi_{j_1,\dots,j_{n+1}}$$

where the sum is over all $j_1, \dots, j_{n+1}$ such that

$$T\Delta_{j_1,\dots,j_{n+1}} \subset \Delta_1 \cup \Delta_2.$$

This proves Case 2b.

The other cases are handled similarly.

$\square$