Contents

9/9

0 Analysis review

- Let X be a set. A subset of the power set of X is called a σ **algebra** if it's closed under complement and countable union.
- If X admits a topology, we call the **Borel set** of X the elements in the minimal σ-algebra that contains all open sets.
- A measure on a σ **algebra** is a function from the σ algebra to $\mathbb{R}_+ \cup \{\infty\}$, such that it is additive under countable disjoint union. If $\mu(X) < \infty$ we call it a **finite measure** and if $\mu(X) = 1$ we call it a **probability**

measure. In dynamics we will mostly focus on finite and probability measures.

- A Radon measure is a measure on Borel set that is "compatible" with the topology. In the topological spaces we will discuss for this semester, which are **separable complete metric spaces**, any **locally finite measure** (for every point there is a neighborhood where the measure is finite) is Radon.
- If X is locally compact Hausdorff space, then there is a one-one correspondence between Radon measures on Borel sets and **positive linear functionals** on $C_c(X)$. Here a positive linear functional is a linear map from $C_c(X)$ to $\mathbb{R}$, such that it sends non negative functions to non negative numbers. This is called the **Rietz representation theorem**.
- Given a measure μ on a σ algebra E on X, we can extend μ to a larger σ algebra such that the subset of any measure 0 set is measure 0. This is called a **completion**. In particular, for any μ which is a Radon measure, we call a set A is μ-measurable iff it lies in the σ-algebra for its completion.
- To define Lebesgue measure on $\mathbb{R}$ one can use one of the following two equivalent approaches:
  - Apply Rietz representation theorem to Riemann integral, get a Radon measure on the Borel set. Then complete it, one gets the Lebesgue measure.
  - Define the "outer measure" λ of a set as the infimum of the sum of lengths of open intervals that can cover this set. If a subset E of $\mathbb{R}$ satisfies that for any other subset A, $\lambda(A) = \lambda(A \cap E) + \lambda(A \backslash E)$, then we call E Lebesgue measurable, and the Lebesgue measure of E is $\lambda(E)$.

  The equivalence of these two approaches can be found in measure theory textbooks.
- A function is measurable iff the preimage of any open set is measurable. A non negative measurable function is integrable if it can be bounded from above by a countable linear combinations of characteristic functions $\sum_i c_i \chi_{A_i}$ for $c_i > 0$, and $\sum_i c_i \mu(A_i) < \infty$. The infimum among these $\sum_i c_i \mu(A_i)$ is called the integral of this non negative measurable functions. Now the L^p space under μ is defined as functions on X such that the p-th power of their absolute value is integrable, and two functions that differ on a measure 0 set are identified. For $1 \leq p \leq \infty$, the L^p space is a complete metric space under the L^p distance:

  $$d(f, g) = \|f - g\|_p, \|f\|_p = (\int |f|^p d\mu)^{1/p}$$

  The L^2 norm is induced by an inner product, hence it is also a **Hilbert space**.

1 Introduction and Examples

Instructor: Chenxi Wu, he/him, email: cwu367@wisc.edu, Office Hours: Van Vleck 517, 2-3pm W, 1:30-2:30pm Th.

1.1 Basic objects and questions in dynamics

Basic object: Let X be a set, with a topology, metric, or measure.

- A self-map: $f : X \to X$
- A flow: $g : \mathbb{R} \times X \to X$, such that $g(0, x) = x$, $g(s, g(t, x)) = g(s + t, x)$.
- More generally, a group or semigroup action on X.

Basic Questions:

- What is the "long term behavior" when one iterates a map, or let the flow go for a long time? (ω limit set, how often the orbit would stay in a subset of X ("equidistribution"), etc.)
- The existence and uniqueness of invariant measure and invariant sets of certain properties.
- Classification up to conjugation and semiconjugation.

Some definitions for map:

Definition 1.1. *Let X be a topological space, $f : X \to X$ be a map.*

- *The* **(forward) orbit** *of f starting at $p \in X$ is the set $\{f^{\circ n}(p) : n \in \mathbb{N}\}$. Similarly one can define backward orbits.*
- *An orbit is called* **periodic** *if $f^{\circ n}(p) = p$ for some $n > 0$.*
- *The* ω**-limit set** *of an orbit $\{f^{\circ n}(p) : n \in \mathbb{N}\}$ is the set of points in X, such that for any neighborhood U, for any $M > 0$, there is some $n > M$ such that $f^{\circ n}(p) \in U$.*
- *A (Radon) measure μ on X is called* **invariant** *under a map f if $\mu(f^{-1}(A)) = \mu(A)$ for any measurable set A. A subset A is called* **(forward) invariant** *if $f(A) \subset A$.*
- *A map $f : X \to X$ and a map $f' : Y \to Y$ are said to be* **conjugate** *if there is a bijection $h : X \to Y$, such that $h \circ f = f' \circ h$. If h is just a surjection then this is called a* **semiconjugation***.*

Analogous definitions for flow:

Definition 1.2. *Let X be a topological space, $g : \mathbb{R} \times X \to X$ be a flow on X.*

- *The* **(forward) orbit** *of g starting at $p \in X$ is the set $\{g(t,p) : t > 0\}$. Similarly one can define backward orbits.*

- *An orbit is* **periodic** *if $g(t,p) = p$ for some $t > 0$.*

- *The* **ω-limit set** *of an orbit starting at p is the set of points in X, such that for any neighborhood U, for any $M > 0$, there is some $t > M$ such that $g(t,p) \in U$.*

- *A (Radon) measure μ on X is called* **invariant** *under a flow g if $\mu(\{g(-t,a) : a \in A\}) = \mu(A)$ for any measurable set A and any $t > 0$. A subset A is called* **(forward) invariant** *if $a \in A$ implies $g(t,a) \in A$ for all $t > 0$.*

- *A flow $g : \mathbb{R}\times X \to X$ and a map $g' : \mathbb{R}\times Y \to Y$ are said to be* **conjugate** *if there is some bijection $h : X \to Y$, such that $h(g(t,x)) = g'(t,h(x))$ for all t, x. They are called* **semiconjugate** *if h is just a surjection.*

Some examples of self maps and flows:

Example 1.3.

- *(Circle rotation) Let $X = \mathbb{R}/\mathbb{Z}$. let $f : X \to X$ be $x \mapsto x + a$.*

- *(Real quadratic map) Let $X = [0,1]$, let $f : X \to X$ be $x \mapsto 4x(1-x)$.*

- *(General quadratic dynamics) Let X be a field k (e.g. $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Q}_p$ etc.), $c \in k$, let $f : X \to X$ be $x \mapsto x^2 - c$.*

- *Let X be a smooth manifold, Y a vector field on X (a smooth section of the tangent bundle TX), then we can define the flow induced by Y as $g : \mathbb{R} \times X \to X$ such that $\partial_t g(t,x) = Y|_{g(t,x)}$.*

- *(Geodesic flow) Let X be a Riemannian manifold. TX be the tangent bundle over X. We can define a flow on TX as $g : \mathbb{R} \times TX \to TX$, such that $\gamma(t) = \pi \circ g(t,q)$ is a geodesic with constant speed, and $\gamma'(0) = q$.*

- *(Gradient flow) Let X be a smooth manifold, V a smooth function on X, we can consider the flow defined by the gradient field of V, called the gradient flow.*

- *(Shift map) Let $X = \{0,1\}^{\mathbb{Z}}$ be the set of functions from $\mathbb{Z}$ to $\{0,1\}$, let $f : X \to X$ be $g \mapsto g \circ \sigma$, where $\sigma(n) = n+1$.*

- *(Random walk on groups) Let G be a group that acts on a space Y, let $X = G^{\mathbb{N}} \times Y$, let $f : X \mapsto X$ be $((g_0, g_1, \dots), y) \mapsto ((g_1, \dots), g_0(y))$.*

Applications of dynamics on other area of math:

- Qualitative studies of ODEs and PDEs: classical mechanics, differential geometry, geometric analysis

- Dynamics of gradient flow: optimization, deep learning, Morse theory etc.

- Dynamical methods in number theory: homognuous dynamics, Diophantine approximation, continued fraction etc.
- Random walk on graph, group and more general spaces: connection with probability, stochastic processes, geometric group theory etc.
- ...

Main topics we will cover for this semester:

- Ergodic theorems. Possibly some Ratner's theory.
- Hyperbolic dynamics and their structural stability.
- Other topics like KAM, thermodynamics formalism etc, if time permits.

1.2 Properties of some simple dynamical systems

1.2.1 Circle rotation

Let $S^1 = \mathbb{R}/\mathbb{Z}$. Let $r_a : S^1 \to S^1$ be $x \mapsto x + a$. When a is rational we call it a **rational rotation**, otherwise we call it an irrational rotation.

Proposition 1.4. *All orbits of rational rotation are periodic (hence the ω-limit set are themselves). All orbits of irrational rotation are not periodic and also dense (hence the ω-limit set are S^1).*

Proof. The case for rational rotation is obvious. The case for irrational rotation is left as an exercise (and is also followed from Proposition 1.6. □

Proposition 1.5. *If μ is a finite Radon measure invariant under irrational rotation, then μ is a multiple of the Lebesgue measure.*

Proof. By Rietz representation theorem (which we will review later), we only need to show that for any continuous function g, $\int g d\mu = \mu(S^1) \int g dx$.

Invariance under r_a implies that

$$\int g d\mu = \int g \circ r_a d\mu = \int g \circ r_a \circ r_a d\mu = \dots$$

Hence

$$\int g d\mu = \int \frac{1}{n}(g + g \circ r_a + g \circ r_a \circ r_a + \dots g \circ r_a^{\circ(n-1)}) d\mu$$

By Stone-Weierstrass, for any $\epsilon > 0$, there is some p which is a linear combination of $\sin(2k\pi x)$ and $\cos(2k\pi x)$, such that $|g - p| < \epsilon$. By computation,

$$\frac{1}{n}(\sin(2k\pi x) + \sin(2k\pi(x + a)) + \cdots + \sin(2k\pi(x + (n-1)a)))$$
$$= \frac{1}{2n} \frac{\cos(2k\pi x - ak\pi) - \cos(2k\pi x + k(2n-1)a\pi)}{\sin(ka\pi)}$$

Hence

$$\int \sin(2k\pi x)d\mu = \lim_{n\to\infty}\int \frac{1}{n}(\sin(2k\pi x)+\sin(2k\pi(x+a))+\cdots+\sin(2k\pi(x+(n-1)a)))d\mu$$

$$= 0 = \mu(S^1)\int \sin(2k\pi x)dx$$

Similarly one can show $\int \cos(2k\pi x)d\mu = \mu(S^1)\int \cos(2k\pi x)dx$ (note that when $k = 0$, $\cos(2k\pi x) = 1$). Hence, we have

$$|\int gd\mu - \mu(S^1)\int gdx| \le |\int (g-p)d\mu + \int pd\mu - \mu(S^1)\int pdx - \mu(S^1)\int (g-p)dx| \le 2\epsilon\mu(S^1)$$

Now let $\epsilon \to 0$, we get $d\mu = \mu(S^1)dx$. □

The above property shows that irrational rotation is **uniquely ergodic**. Hence, by ergodic theorems which we will learn later, we know that:

Proposition 1.6. *Under irrational rotation, any Lebesgue-measurable invariant subset of S^1 is of zero measure or full measure.*

Proof. Suppose A is such a subset whose Lebesgue measure is non-zero, then $\mu'(B) = \mu_{Lebesgue}(B \cap A)$ is an invariant measure, hence it must be a multiple of Lebesgue measure. This implies that A is of full Lebesgue measure. □

Proposition 1.7. *Let A be any Lebesgue measurable subset of S^1, then for almost every x, the limit*

$$\lim_{n\to\infty}\frac{1}{n}\|\{k \in \{1, 2, \dots, n\} : r_a^{\circ k}(x) \in A\}\|$$

is the Lebesgue measure of A.

Proof. This is the Birkhoff ergodic theorem which we will prove later. □

Remark 1.8. *For the above two propositions, one can also prove them via Fourier analysis without the use of the ergodic theorems.*

The invariant sets and invariant measures for rational rotation is simpler and will be left as an exercise.

9/14

1.2.2 Increasing Homeomorphisms on an interval

Theorem 1.9. *Let $I = [0, 1]$, $f : I \to I$ a continuous bijection, $f(0) = 0$, $f(1) = 1$. Then every point $x \in I$ is either a fixed point, or the forward and backward orbit of x both converge monotonously to the closest fixed points to x from both sides.*

Proof. If $f(x) > x$, then $f(f(x)) > f(x)$, hence $f^n(x)$ will be an increasing sequence. Suppose it converges to some x^*, then there can not be any other fixed points y between x and x^* as otherwise there would be some $f^n(x) \leq y$ and $f(f^n(x)) > y = f(y)$ which is a contradiction. The case when $f(x) < x$ is similar. □

Definition 1.10. *Let $f : X \mapsto X$, a point p in X is called a wandering point if there is a neighborhood U of p such that $f^n(U) \cap U = \emptyset$ for all $n > 0$.*

In the case of this interval homeomorphism, the wandering set are the open set $\{x : f(x) \neq x\}$.

Theorem 1.11. *Two increasing homeomorphisms $f : I \to I$ and $g : I \to I$ are conjugate with one another by an increasing homeomorphism, iff there is some homeomorphism h from I to itself that sets $\{x : f(x) > x\}$ to $\{x : g(x) > x\}$, $\{x : f(x) = x\}$ to $\{x : g(x) = x\}$, $\{x : f(x) < x\}$ to $\{x : g(x) = x\}$.*

Proof. If f and g are conjugate by an increasing homeomorphism, we can just let the h to be this homeomorphism then it satisfies the conditions in the theorem. To show the other direction, suppose f is a map satisfying the given assumptions, let's define the conjugation map h_1. If x is a fixed point of f, let $h_1(x) = h(x)$. If not, suppose $f(x) > x$, then define h_1 as sending $[x, f(x)]$ to $[h(x), g(h(x))]$. Now for any $n \in \mathbb{Z}$, let h_1 sends $[f^{\circ n}(x), f^{\circ n+1}(x)]$ to $[g^{\circ n}(h(x)), g^{\circ n+1}(h(x))]$ by $h_1(y) = g^{\circ n} h_1 f^{\circ -n}(y)$. The case for $f(x) < x$ is analogous. □

Theorem 1.12. *Any finite invariant measure μ can not have support on wandering set of a homeomorphism.*

Proof. Let x be a point in the wandering set, let U be a neighborhood of x such that $f^{\circ n}(U) \cap U$ are all empty. Then it follows that $f^{\circ n}(U) \cap f^{\circ m}(U) = \emptyset$ if $n \neq m$. Hence $\mu(U) = 0$, because if not, $\mu(\cup_n f^{\circ n}(U))$ would be infinite. □

As an application, all invariant measures on I are those supported on the set of fixed points.

1.2.3 Homeomorphisms on a circle

Let $S^1 = \mathbb{R}/\mathbb{Z}$. Let $f : S^1 \to S^1$ be an increasing homeomorphism. By topology, we know that there is an increasing homeomorphism $F : \mathbb{R} \to \mathbb{R}$ which is a lift of f, i.e. $F(x + n) = F(x) + n$ for any $n \in \mathbb{Z}$, and $f(x + \mathbb{Z}) = F(x) + \mathbb{Z}$.

Definition 1.13. *Let $F : \mathbb{R} \to \mathbb{R}$ be a lift of f. Let $a \in \mathbb{R}$, then the number*

$$\lim_{n \to \infty} \frac{F^{\circ n}(a) - a}{n}$$

mod $\mathbb{Z}$ is called the **rotation number** *of f.*

Theorem 1.14.
- *The limit exists.*

- *The limit doesn't depend on the choice of a. If one replace F with another lift, the limit differ at most by an integer. Hence rotation number is well defined in $\mathbb{R}/\mathbb{Z}$.*
- *The rotation number is invariant under conjugation via orientation preserving homeomorphisms.*
- *The rotation number is rational iff there are points with periodic orbit.*

Proof. (Sketch)

- Firstly we prove that for any $a, b \in \mathbb{R}$,

$$|(F^{\circ n}(a) - a) - (F^{\circ n}(b) - b)| \leq 1$$

This is because there is some $m \in \mathbb{Z}$ such that $a + m \leq b \leq a + m + 1$, hence

$$F^{\circ n}(a) - a + (a + m - b) = F^{\circ n}(a + m) - b \leq F^{\circ n}(b) - b$$

$$\leq F^{\circ n}(a + m + 1) - b = F^{\circ n}(a) - a + (a + m + 1 - b)$$

Now, given $\epsilon > 0$, let n be large enough that $1/n < \epsilon/4$. Let $k >> 1$, $0 \leq m < n$, then

$$\frac{F^{\circ m+kn}(a) - a}{m + kn} = \frac{(F^{\circ m+kn}(a) - F^{\circ kn}(a)) + \sum_{i=1}^{k}(F^{\circ kn}(a) - F^{\circ (k-1)n}(a))}{m + kn}$$

Because $F^{\circ m+kn}(a) - F^{\circ kn}(a)$ is bounded, and $F^{\circ kn}(a) - F^{\circ (k-1)n}(a)$ differs from $F^{\circ n}(a) - a$ by at most 1, as $k \to \infty$, no matter which m one chooses, the number will get close to $[\frac{F^{\circ n}(a)-a}{n} - \epsilon/4, \frac{F^{\circ n}(a)-a}{n} + \epsilon/4]$. Now let K be such that for any $0 \leq m < n$, any $k > K$,

$$\frac{F^{\circ m+kn}(a) - a}{m + kn} \in [\frac{F^{\circ n}(a) - a}{n} - \epsilon/2, \frac{F^{\circ n}(a) - a}{n} + \epsilon/2]$$

This implies that for any N, N' greater than Kn,

$$|\frac{F^{\circ N}(a) - a}{N} - \frac{F^{\circ N'}(a) - a}{N'}| \leq \epsilon$$

Hence the limit exists.

- This follows from the fact that $|(F^{\circ n}(a) - a) - (F^{\circ n}(b) - b)| \leq 1$ for any $a, b \in \mathbb{R}$, and the fact that two different lifts of f differ at most by an integer.
- Lift the congugation map to $\mathbb{R}$.

- If there are points with periodic orbit, use them as the a to calculate the rotation number. If the rotation number of f is rational, there is a multiple of f whose rotation number is zero. Hence we only need to show that a map with rotation number zero must have a fixed point. Let G be the lift of such a map such that $\lim_{n\to\infty}\frac{G^{\circ n}(a)-a}{n}=0$. Suppose $G(x)\not\approx x$ for all x, then $G(x)-x$ must be bounded away from 0 due to the compactness of S^1, hence there would be a contradiction.

$\square$

Theorem 1.15. *(Poincare classification)Increasing homeomorphisms on a circle must be of one of the following types:*

- *Has rational rotational number.*
  - *Every point is periodic. In this case, it is conjugate to a rational rotation.*
  - *Has more than one periodic orbits, and the forward and backward orbits of all other points approach two different periodic orbits.*
  - *Has one periodic orbit, the forward and backward orbits of all other points approach this orbit.*
- *Has irrational rotation number.*
  - *Every orbit is dense. In this case, it is conjugate to an irrational rotation.*
  - *The non wandering set is a cantor set and every orbit there is dense, while orbits in the wandering set approaches this cantor set. In this case, it is semiconjugate to an irrational rotation by collapsing the wandering intervals.*

Proof. (Sketch) The rational case is due to the analysis of increasing interval map. For the irrational case, if the orbit is not dense, we can construct the semiconjugacy as follows: take some $x\in\mathbb{R}$, let c be the rotation number corresponding to lift F, define $H: F^n(x)+m=cn+m$. This H defines a conjugation from $B=\{f^n(x) \mod \mathbb{Z}\}$ to S^1. Now extend it continuously to the closure of B, then to the whole S^1 via setting function on complementary intervals as constants. It is now easy to see that the intervals where H is constant are the wandering sets. $\square$

Theorem 1.16. *(Denjoy's theorem)If f is C^2, then the last case in the classification above is impossible.*

If C^2 is replaced by C^1 then there are counter-examples.

You can find the proof and the example in our reference or by Googling.

1.2.4 Circle doubling map

Let $X = \mathbb{R}/\mathbb{Z}$, f be $x \mapsto 2x$. Then there is a semiconjugation between this dynamical system and $g : Y = \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$, $g(d_0 d_1 d_2 d_3 \dots) = (d_1 d_2 d_3 \dots)$, by binary expansion of real numbers.

As a consequence, we have:

- There are orbits of f of any length.
- There is a measure 0 set Z such that $x \in X \backslash Z$ then the orbit of x is dense.

Furthermore, f has **structural stability**:

Theorem 1.17. *If ϵ is a small $\mathbb{Z}$-periodic function whose derivative is very small, then f and $f + \epsilon$ are conjugate via an orientation preserving homeomorphism.*

Proof. (sketch) Let $G(x) = 2x + \epsilon(x)$, $H_0(x) = x$, $H_{i+1}(x) = \frac{1}{2} H_i(G(x))$. It is easy to see that $\sup |H_{i+1} - H_i| \leq \frac{1}{2} \sup |H_i - H_{i-1}|$, hence for any $n > m > 0$,

$$\sup |H_n - H_m| \leq \sum_{k=m}^{n-1} \sup |H_{k+1} - H_k| \leq \sum_{k=m}^{n-1} 2^{-(k-1)} \sup |H_2 - H_1|$$

which goes to 0 as m, n goes to infinity. This implies that as $n \to \infty$, H_n converges uniformly to a continuous function on $\mathbb{R}$, which we call H. One can prove by induction that $H_n(x + k) = H_n(x) + k$ for any $k \in \mathbb{Z}$, hence $H(x + k) = H(x) + k$, H induces a continuous map from S^1 to S^1. It's also evident that H is a semiconjugacy between $x \mapsto 2x$ and $x \mapsto 2x + \epsilon$. The fact that it is a conjugacy can be shown using techniques discussed later in the semester, e.g. kneading theory. □

It is easy to see that circle rotation does not have structural stability, while circle doubling does not have unique ergodicity. The interval homeomorphisms are not unique ergodic, some are structurally stable some aren't.

1.3 Relationship between maps and flows

- Given a map $f : X \to X$, if f is bijection, we can define the **suspension flow** on the mapping torus $X \times \mathbb{R}/\sim$, $(x, t + 1) \sim (f(x), t)$, and the flow being $g_t(p, s) = (p, s + t)$.
- Given a flow g_t on X, let $Y \subset X$ be a subset, we can define the **first return map** on Y as $p \mapsto g_s(p)$, where $s > 0$ is the smallest positive number such that $g_s(p) \in Y$. Usually we let X be a manifold, Y be a submanifold, such that the flow lines are all transversal to Y.
- Given a flow g on X, we can also consider the **time t map**: $g_t : X \to X$, $g_t(x) = g(t, x)$.

2 Ergodicity and Mixing

Definition 2.1. *Let $T : X \to X$ be a map preserving finite measure μ. μ is called* **ergodic** *if any μ-measurable (forward and backward) invariant subset (i.e. A such that $T(A) = T^{-1}(A) = A$) is either of zero measure or of full measure. If X is a (complete metrizable, seperable, locally compact) topological space, and there is only one T invariant probability Radon measure, then we call f to be* **uniquely ergodic**.

Proposition 2.2. *If f is uniquely ergodic, then the invariant probability measure is an ergodic measure.*

Proof. If not, then let A be a measurable invariant set which is of neither zero nor full measure, then $\mu'(B) = \mu(B \cap A)$ is invariant under T. However by unique ergodicity μ' must be a multiple of μ, a contradiction. □

9/21

2.1 Mean ergodic theorem

Let T be a map on X, preserving a finite measure μ. The space of L^2-integrable functions (up to a measure zero set) $L^2_\mu(X)$ is a vector space with a complete metric $d(f, g) = \|f - g\| = \sqrt{\int |f - g|^2 d\mu}$, which is induced by an inner product $(f, g) = \int fg d\mu$, hence is called a real **Hilbert space**. Let V be a map on L^2_μ induced by T as $Vf = f \circ T$. It is easy to see that $(Vf, Vg) = (f, g)$, i.e. V is called a "unitary" operator.

There are some properties of Hilbert space (complete inner product space) that are generalizations of the properties of Euclidean spaces which we learned in linear algebra:

- A Hilbert space is "self dual": in other words, any continuous linear function from H to $\mathbb{R}$ can be written in the form $x \mapsto (x, y)$ for a unique value $y \in H$.
- This implies the existence of **conjugate operators**: let $Q : H \to H$ be a continuous linear map, then we can define another continuous linear map Q^* by $(Q^*(x), y) = (x, Q(y))$.
- If $H' \subset H$ is a closed subspace, then given any $x \in H$, there is a unique point $x' \in H'$ whose distance to x is minimized, called the **orthogonal projection of x on H'**. The map $x \mapsto x'$ is a continuous operator from H to H' called **orthogonal projection**.
- If $H' \subset H$ is a closed subspace, the **orthogonal complement** of H' is defined as
$$H'^{\perp} = \{x \in H : \forall y \in H', (x, y) = 0\}$$
$$((H')^{\perp})^{\perp} = H'$$

Theorem 2.3. *(Einsiedler-Ward, Theorem 2.21) Let T be a map on X preserving probability measure μ, I be the closed subspace of $L^2_\mu(X)$ consisting of T-invariant functions. Then for any $f \in L^2_\mu(X)$, $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$ converges under L^2_μ to the orthogonal projection of f onto I.*

Proof. Let $B = \{g \circ T - g\}$. Firstly we show that $I^\perp = \overline{B}$. This is equivalent to showing $\overline{B}^\perp = I$, in other words, for any $g \circ T - g \in B$, $(f, g \circ T - g) = 0$ iff $f \circ T = f$.

If $f \circ T = f$, then

$$(f, g \circ T - g) = (f, g \circ T) - (f, g) = (f \circ T, g \circ T) - (f, g) = 0$$

On the other hand, if for any g, $(f, g \circ T - g) = 0$, then let V be $V(f) = f \circ T$, we have

$$0 = (f, g \circ T - g) = (f, Vg) - (f, g) = (V^* f - f, g)$$

So $f = V^* f$.

Now

$$(f - Vf, f - Vf) = (f, f) + (Vf, Vf) - 2(f, Vf) = 2(f, f) - 2(V^* f, f) = 0$$

Hence $f = Vf$. This proved that $I^\perp = \overline{B}$.

Now let $f \in L^2_\mu(X)$, f_1 the orthogonal projection of f on I, then $f = f_1 + f_2$, where $f_2 \in \overline{B}$. So given any $\epsilon > 0$, there is some g such that $\|f_2 - (g \circ T - g)\| < \epsilon$. Hence

$$\|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k - f_1\| = \|\frac{1}{n} \cdot n f_1 + \frac{1}{n}\sum_{k=0}^{n-1} f_2 \circ T^k - f_1\|$$

$$= \|\frac{1}{n}\sum_{k=0}^{n-1} f_2 \circ T^k\|$$

$$\leq \frac{1}{n}(\sum_{k=0}^{n-1} \|(f_2 - (g \circ T - g)) \circ T^k\| + \|g\| + \|g \circ T^n\|)$$

$$\leq \epsilon + \frac{2\|g\|}{n}$$

Which, as $n \to \infty$, converges to ϵ. Because ϵ can be arbitrarily small, the theorem is proved. □

Remark 2.4. *If μ is ergodic, i.e. $A = T^{-1}(A)$ implies $\mu(A) = 0$ or 1, then the orthogonal projection to I will be $f \mapsto \int_X f d\mu$. Hence $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$ converges under L^2_μ to $\int_X f d\mu$.*

Corollary 2.5. *If f is in $L^1_\mu(X)$, $\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k$ converges under L^1_μ to a T-invariant function.*

The proof is by approximating L^1 functions with L^2 functions.

Remark 2.6. *The mean ergodic theorem can also be proved using Gelfand's spectral theory of unitary operators, which tells us that the speed of convergence can be estimated via "spectral gap".*

9/23

2.2 Pointwise ergodic theorem

Theorem 2.7. *(Katok-Hasselblatt, Theorem 4.1.2) Let T be a map on X preserving probability measure μ, $f \in L^1_\mu(X)$, $\overline{f}$ is the L^1_μ limit of $\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k$, then for μ a.e. $x \in X$,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) = \overline{f}(x)$$

Proof. Define

$$F_k(x) = \max\{0, g(x), g(x)+g(T(x)), \dots, \sum_{i=0}^{k} g\circ T^i(x)\}$$

$$A = \{x \in X : \lim_{k\to\infty} F_k = \infty\}$$

Firstly we show that A is T-invariant: if $x \in A$, it implies that

$$\limsup_{k\to\infty}(\sum_{i=0}^{k} g\circ T^i(x)) = \infty$$

Hence

$$\limsup_{k\to\infty}(\sum_{i=1}^{k} g\circ T^i(x)) = \infty$$

Hence $T(x) \in A$.

Next, we show that for any $x \in A$, $0 \le F_{k+1}(x) - F_k(T(x)) \le f(x)$, and when k is sufficiently large $F_{k+1}(x) - F_k(T(x)) = g(x)$. To show this, firstly note that

$$F_{k+1}(x) = \max\{0, g(x), g(x)+g(T(x)), \dots, g(x)+g(T(x))+\dots+g(T^{k+1}(x))\}$$

$$F_k(T(x)) = \max\{0, g(T(x)), \dots, g(T(x))+\dots+g(T^{k+1}(x))\}$$

So

$$0 \le F_{k+1}(x) - F_k(T(x)) \le f(x)$$

Because $x \in A$, when k is sufficiently large $F_{k+1}(x) \neq 0$, hence for such k,

$$F_{k+1}(x) = \max\{g(x), g(x) + g(T(x)), \dots, g(x) + g(T(x)) + \cdots + g(T^{k+1}(x))\}$$

$$= g(x) + \max\{0, g(T(x)), \dots, g(T(x)) + \cdots + g(T^{k+1}(x))\} = g(x) + F_k(T(x))$$

Hence we have:

$$0 \leq \lim_{k\to\infty} \int_A (F_{k+1} - F_k)(x) d\mu = \int_A g d\mu$$

The equal sign is due to dominant convergence theorem.

Now for any $\epsilon > 0$, let $g = \overline{f} - f - \epsilon$, then

$$\int_A \overline{f} - f - \epsilon = \int_A (\overline{f}(x) + \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x)) d\mu - \epsilon\mu(A) = -\epsilon\mu(A)$$

The last equal sign is because $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$ converges to $\overline{f}$ under L^1_μ. Hence $\mu(A) = 0$.

Now let $\epsilon = 1/m$ and A_m be the corresponding A, then $\bigcup_m A_m$ is a set of zero measure, and in the complement of this set we have

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \geq \overline{f}$$

Similarly one can show a.e. on X, we have

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \leq \overline{f}$$

which together prove the theorem. □

Example 2.8. *Let $X = \mathbb{R}/\mathbb{Z}$, $c \notin \mathbb{Q}$, $T : x \mapsto x + c$. Then by setting f to be a characteristic function, we can prove Proposition 1.7.*

Example 2.9. *Let $X = \mathbb{R}/\mathbb{Z}$, $T : x \mapsto 2x$. Then we will show that the Lebesgue measure is ergodic as below: let A be a T-invariant set, then the characteristic function χ_A is T-invariant. Because χ_A is in $L^2(X)$, by Fourier analysis we have $\chi_A = \sum_{n=0}^{\infty} a_n \cos(2n\pi x) + \sum_{n=1}^{\infty} b_n \sin(2n\pi x)$ in the L^2 sense, and $\sum_n a_n^2 + \sum_n b_n^2 < \infty$. Now $\chi_A = \chi_A \cdot T$ implies that $a_n = a_{2n}$, $b_n = b_{2n}$, hence a_0 is the only possible coefficient to be non-zero, i.e. χ_A is a.e. 0 or 1.*

Now apply Birkhoff ergodic theorem we get a.e. x,

$$\lim_{n\to\infty} \frac{\|\{k \in \{0, 1, 2, \dots, n-1\} : T^k(x) \in A\}\|}{n} = \mu(A)$$

Where μ is the Lebesgue measure.

Now consider a countable sequence of sets A_m which goes through all intervals of rational end points and have length more than 0. Then Birkhoff ergodic theorem gives an alternative proof that the orbit of a.e. point under circle doubling is dense.

9/28

2.3 Existence of ergodic measures

For this section, we assume that X is compact, locally compact, Hausdorff, separably and completely metrizable.

Definition 2.10. *Let V be a topological vector space. The dual space V^* is the set of continous linear functionals on X.*

Definition 2.11. *We define the weak-$*$ topology on V^* as the coarsest topology where all maps of the form $y \mapsto y(x)$ is continuous, here $y \in V^*$, $x \in V$.*

Let M be the set of Radon probability measures on X. M is a convex closed subset of the dual of the space of continuous functions $C(X)$ by Rietz representation theorem. Let N be the set of elements in M that are T-invariant.

Proposition 2.12. *A point in N correspond to an ergodic measure iff it can not be written as the non-trivial linear combination of two other points in N.*

Proof. If μ is not ergodic, there is some $A \subset X$ such that $\mu(A)$ and $\mu(X\backslash A)$ are both non-zero. Let $\mu_1(B) = \mu(B \cap A)/\mu(A)$, $\mu_2(B) = \mu(B\backslash A)/\mu(X\backslash A)$, then

$$\mu = \mu(A)\mu_1 + \mu(X\backslash A)\mu_2$$

If $\mu = t\mu_1 + (1-t)\mu_2$, let f be a bounded continuous function whose integral on μ_1 and μ_2 are different, then consider $\overline{f} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i$, we see that f is T-invariant and has different integral under μ_1 and μ_2 as well, hence it can not be a.e. constant, i.e. μ can not be ergodic. □

In other words, a basis of this topology consists of finite intersections of the sets $\{y \in X^* : y(x) \in I\}$ for some open interval I.

Proposition 2.13. *M is both compact and sequentially compact.*

Proof. Because by our original assumption X is separable and metrizable, $C(X)$ has a dense subset $\{\phi_j\}$, with range I_j. M is homeomorphic (check this as an exercise) to a closed subset of $\prod_j I_j$ with product topology by

$$\mu \mapsto (\int \phi_j d\mu)$$

Hence is compact. It is evident that the product topology in $\prod_j I_j$ are metrizable (with metric being the sum of the Euclidean metric on each component divided by $2^j|I_j|$), hence M is sequentially compact. □

Now the existence of ergodic measure follows from the proposition above:

Theorem 2.14. *(Kryloff-Bogoliouboff, Cor. 4.2 Einsiedler-Ward, Theorem 4.1.1 in Katok-Hasselblatt) N is non-empty.*

Proof. Let $x \in X$, $\int f d\mu_n = \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$. $\mu_n \in M$, so they must have a subsequence n_j that converges to some μ. Hence

$$\int f d\mu = \lim_{j\to\infty} \frac{1}{n_j}\sum_{i=0}^{n_j} f(T^i(x))$$

$$\int f \circ T d\mu = \lim_{j\to\infty} \frac{1}{n_j}\sum_{i=1}^{n_j+1} f(T^i(x))$$

And

$$|\frac{1}{n_j}\sum_{i=0}^{n_j} f(T^i(x)) - \frac{1}{n_j}\sum_{i=1}^{n_j+1} f(T^i(x))| = \frac{1}{n_j}(|f(x)| + |f(T^{n_j+1}(x))|) \to 0$$

□

Theorem 2.15. *There exists at least one ergodic probability measure in N.*

Proof. Let $N_0 = N$, N_j consists of the elements in N_{j-1} that maximizes $\int \phi_j d\mu$. Then there is a unique element in $\bigcap_j N_j$ which is an extreme point. □

Example 2.16. *When T is uniquely ergodic, the sequence in the proof of Theorem 2.3 must converge as it has a single accumulation point. Hence for any $f \in C(\mathbb{R}/\mathbb{Z})$, for any $x \in \mathbb{R}/\mathbb{Z}$,*

$$\int f dx = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)$$

9/30

The argument we used in the example above implies:

Theorem 2.17. *If μ is an ergodic probability measure on compact space X, T is continuous, then T is uniquely ergodic iff for every $x \in X$, for every $f \in C(X)$,*

$$\int f d\mu = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)$$

Proof. If μ is not uniquely ergodic, there must be another ergodic measure μ' (one can construct it via a process similar to the proof of Theorem 2.3). Let $f \in C(X)$ be such that $\int f d\mu' \neq \int f d\mu$, then Birkhoff ergodic theorem says that for μ' a.e. x,

$$\int f d\mu \neq \int f d\mu' = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x)$$

If μ is uniquely ergodic, let δ_n be $\int f d\delta_n = \frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)$. All accumulation points of δ_n under weak-* topology have to be T-invariant hence must be μ, hence the weak-* limit of δ_n is μ, hence

$$\int f d\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x)$$

□

Theorem 2.18. *Ergodic decomposition: Let E be the set of ergodic Radon probability measures for T, then any invariant measure μ has a probability measure λ defined on E, such that $\mu = \int_E v d\lambda(v)$.*

Example 2.19. *(Cor 4.22 Einsiedler-Ward)If $c \notin \mathbb{Q}$, then the map $T : (\mathbb{R}/\mathbb{Z})^3 \to (\mathbb{R}/\mathbb{Z})^3$ defined by $(x_1, x_2, x_3) \mapsto (x_1+c, x_2+x_1, x_3+x_2)$ is ergodic and uniquely ergodic.*

The proof for ergodicity can be done via Fourier analysis. To show unique ergodicity, one use the following Theorem by Furstenburg:

Theorem 2.20. *Suppose X is compact, $T_x : X \to X$ is an uniquely ergodic continuous map with the unique invariant probability measure μ. Let $T : X \times \mathbb{R}/\mathbb{Z} \to X \times \mathbb{R}/\mathbb{Z}$ be*

$$T(x, y) = (T_x(x), f(x) + y)$$

Where f is a continuous function. Let dy be the Lebesgue measure on $\mathbb{R}/\mathbb{Z}$. Then, if $\mu \times dy$ is ergodic, T must also be uniquely ergodic.

Proof. Let

$$E = \{(x, y) \in X \times \mathbb{R}/\mathbb{Z} : \forall f \in C(X \times \mathbb{R}/\mathbb{Z}), \int f d(\mu \times dy) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{j} f(T^j x)\}$$

By replacing f with $g(x, y) = f(x, y + t)$, we know that if $(x, y) \in E$, then $(x, y + t) \in E$ for all t. Hence $E = E_x \times \mathbb{R}/\mathbb{Z}$.

The separability of $C(X)$ means that there is a dense subset $\{\psi_i\}$ in $C(X)$. Birkhoff ergodic theorem shows that for each ψ_i, there is a $\mu \times dy$-measure 0 set Z_i such that $\frac{1}{n}\sum_{j=0}^{j} \psi(T^j x)$ does not converge to $\int \psi_i d(\mu \times dy)$, hence all the points that are not in the union of all Z_i must be in E. Hence E has $\mu \times dy$-measure 1. In particular, E_x must have μ-measure 1.

Now suppose T is not uniquely ergodic, define E' similarly for another ergodic probability measure μ', then the same argument above shows that $E' = E'_X \times \mathbb{R}/\mathbb{Z}$, and $\mu'(E') = 1$. Let π be the projection $X \times \mathbb{R}/\mathbb{Z} \to X$, $(x, y) \mapsto x$, then the pushforward $\pi_*(\mu')$ (defined as $\pi_*(\mu')(B) = \mu'(\pi^{-1}(B))$) must be an invariant probability measure on X, by uniquely ergodicity it must be μ. Hence $\mu(E'_X) = 1$. But $E \cap E' = \emptyset$ so E_X and E'_X must be disjoint, which is a contradiction because μ is a probability measure.

□

As a consequence, we know that for any continuous function f on $\mathbb{R}/\mathbb{Z}$, for any x, y,

$$\lim_{n\to\infty} \frac{1}{n} f(cn^2 + xn + y) = \int f dx$$

2.4 Mixing

(Section 2.7 Einsiedler-Ward)

There are two conditions stronger than ergodic, weakly mixing and mixing.

Definition 2.21. *A map T that fixed a probability measure μ is called* **mixing***, if for any μ-measurable A, B,*

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

Called **weakly mixing** *if*

$$\lim_{m\to\infty} \frac{1}{m} \sum_{n=0}^{m} |\mu(A \cap T^{-n}(B)) - \mu(A)\mu(B)| = 0$$

It is easy to see that mixing implies weakly mixing implies ergodic.

Example 2.22.
- *The irrational rotation is not weakly mixing.*
- *The circle doubling map is mixing.*

More examples on mixing and weak mixing will be seen in the next section when we discuss symbolic dynamics.

10/5, 10/7

3 Symbolic Dynamics and Entropy

3.1 Symbolic dynamics

Symbolic dynamics is the study of one sided or two sided subshifts.

Definition 3.1. *Let A be a finite set. $Y_1 = A^{\mathbb{Z}}$, $Y_2 = A^{\mathbb{N}}$, both with product topology. Define σ_1 on Y_1 as shifting index by -1, σ_2 on Y_2 as deleting the first element and shift the index of all later ones by 1. Then σ_1 is called a* **two sided shift***, σ_2 a* **one sided shift***. If X is a σ_i invariant subset of Y_i, then σ_i on X is called a* **subshift***.*

The following dynamical systems can be made semiconjugate to shifts and subshifts, which is bijective except on a measure-0 set:

Example 3.2.

- *$X = \mathbb{R}/\mathbb{Z}$, $x \mapsto kx$, $k \in \mathbb{N}$, $k > 0$.*
- *$X = [0,1]$, $f(x) = 2x$ if $x < 1/2$, $2 - 2x$ if $x \geq 1/2$.*
- *$X = [0,1]$, $f(x) = \lambda x$ if $x < 1/\lambda$, $2-\lambda x$ if $x \geq 1/\lambda$, where $\lambda = (\sqrt{5}+1)/2$.*
- *(baker's map) $X = [0,1] \times [0,1]$. $(x,y) \mapsto (x/2, 2y)$ if $y < 1/2$, $(x/2 + 1/2, 2y - 1)$ if $y \geq 1/2$.*
- *$X = \mathbb{R} \times \mathbb{R}/\mathbb{Z} \times \mathbb{Z}$, $(x,y) \mapsto (x+y, x)$.*

The subshifts semoiconjugate to these dynamical systems are those consisting of sequences of the form below:

Definition 3.3. *Let D_j be a subset of $\{1, 2, \ldots, m\}$ for every $j \in \{1, 2, \ldots, m\}$, then a subset $\{1, 2, \ldots, m\}^{\mathbb{N}}$ or $\{1, 2, \ldots, m\}^{\mathbb{Z}}$ defined by $\{(a_j) : a_{j+1} \in D_{a_j}\}$ is called a* **topological Markov chain***.*

The strategy is to divide the set X into subsets that are disjoint except for possibly measure 0 regions, and encode every point p by looking at which subset $f^n(p)$ belongs to (which we call the **itineraries** of point p). This is called a **Markov decomposition**.

Existence of invariant measure on topological Markov chain can be obtained via the Perron-Forbenious theorem below:

Theorem 3.4. *(Perron-Frobenious theorem) Let A be a non-negative matrix such that there is some k where A^k has no zero entry, then A has a unique largest positive eigenvalue λ, and all other eigenvalues have norm smaller than λ. The corresponding eigenvector is positive, and there are no other positive eigenvectors.*

Proof. Let P be the set of non negative vectors in $\mathbb{R}^n$ whose coordinates add up to 1, let L be the affine subspace containing P, let T be the map induced by A, then $T(A)$ is convex closed and contained in P. Hence, there must be a fixed point of T by Brouwe fixd point theorem (continuous maps from compact convex set to itself have fixed points). Suppose there are more than one, i.e. there are $x_1 > 0$ and $x_2 > 0$ such that $A(x_1) = \lambda x_1$, $A(x_2) = \lambda x_2$, if $\lambda_1 = \lambda_2$ then T fixes $span\{x_1, x_2\} \cap P$, which contradicts with the fact that it sends P to its interior. Hence without loss of generality $\lambda_1 > \lambda_2$, now pick $c >> 1$ such that $cx_2 - x_1 > 0$, apply A^n for $n >> 1$, there is a contradiction as well. Hence the positive eigenvector must be unique. A similar argument (look at the subspace spanned by this eigenvector, another eigenvector and its complex conjugate) can be used to show that there can not be any other eigenvalue that is larger. □

10/12

Remark 3.5. *The followings are equivalent:*

- $A^k > 0$ *for some* $k > 0$.
- *For any* $k \geq 1$, A^k *has no invariant subspace spanned by standard basis vectors.*
- *The corresponding directed graph is strongly connected with the gcd of lengths of loops* 1.

Definition 3.6. *Let* $A = \{1, 2, \ldots, m\}$, $S_G \subset A^{\mathbb{N}}$ *be a subshift defined by strongly connected directed subgraph* G *whose lengths of loops has gcd* 1.

- *Given any finite word* $w = w_0 w_1 \ldots w_{n-1}$, *the set* $\Pi_w = \{w' \in S_G : w'_i = w_i \forall i < n\}$ *is called a* **cylinder set**.
- *Let* $M = [M_{ij}]$ *be a* $|A| \times |A|$ *matrix where the* i, j*-th entry is* 1 *iff there is an edge from the* i*-th vertex to the* j*-th vertex in* G, *and* 0 *if otherwise. Let* x *be the Perron-Frobenious eigenvector of* M *such that* $\sum_i x_i = 1$. *Let* λ *be the Perron-Forbenious eigenvalue. Let* M' *be*

$$M' = diag(\frac{1}{\lambda x_i}) A diag(x_i) = [\frac{M_{ij} x_j}{\lambda x_i}]$$

Let x' *be the Perron-Frobenious eigenvector of* M'^T *such that* $\sum_i x'_i = 1$. *Because*

$$M' \begin{bmatrix} 1 \\ \ldots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ \ldots \\ 1 \end{bmatrix}$$

the Perron-Frobenious eigenvalue of both M' *and* M *are* 1.

Now the measures on cylinder set is defined as

$$\mu_{PF}(\Pi_w) = x'_{w_0} \cdot \prod_{j=0}^{n-2} M'_{w_j w_{j+1}}$$

Here M'_{ij} *is the element in the* i*-th row,* j*-th column of* M'.

It is easy to see that if a cylinder set is the disjoint union of other cylinder sets, then its μ_{PF} is the sum of the μ_{PF} of the other cylinder sets. Let $w \in A^k$,

$$\mu_{PF}(\Pi_w) = x'_{w_0} \cdot \prod_{j=0}^{k-2} M'_{w_j w_{j+1}}$$

$$\Pi_w = \cup_{w' \in A^{k+1}, w'_i = w_i \forall i < k} \Pi_{w'}$$

Then

$$\sum_{w' \in A^{k+1}, w'_i = w_i \forall i < k} \mu_{PF}(\Pi_{w'}) = \sum_{l \in \{l : M_{w_{k-1}, l} \neq 0\}} x'_{w_0} \cdot \prod_{j=0}^{k-2} M'_{w_j w_{j+1}} M'_{w_{k-1}, l}$$

$$= x'_{w_0} \cdot \prod_{j=0}^{k-2} M'_{w_j w_{j+1}} \sum_l M'_{w_{k-1},l} = \mu_{PF}(\Pi_w)$$

To show that μ_{PF} can be extended to a Radon measure on S_G, one can either use outer measure (define it as the infimum of the sum of μ_{PF} of countable unions of cylinder sets covering a given subset $A \subset S_G$), or use μ_{PF} to define "Riemann integral" for continuous functions on S_G then use Rietz representation Theorem.

Now we show that:

Proposition 3.7. *μ_{PF} is σ-invariant.*

Proof. We only need to show that it is σ-invariant when applied to cylinder sets. Let $w \in A^k$, then

$$\sigma^{-1}\Pi_w = \cup_{w' \in A^{k+1}, w'_{i+1} = w_i \forall i < k} \Pi_{w'}$$

So

$$\mu_{PF}(\sigma^{-1}\Pi_w) = \sum_{i \in A} x'_i M'_{iw_0} M'_{w_0 w_1} \dots M'_{w_{k-2} w_{k-1}}$$
$$= x'_{w_0} M'_{w_0 w_1} \dots M'_{w_{k-2} w_{k-1}} = \mu_{PF}(\Pi_w)$$

□

Example 3.8. *Let $\lambda = \frac{\sqrt{5}+1}{2}$. Consider map $T : [2-\lambda, 1]$, $T(x) = \lambda x$ when $x \le 1/\lambda$, $T(x) = 2 - \lambda x$ if $x > 1/\lambda$. Then $I_1 = [2-\lambda, 1/\lambda]$, $I_2 = [1/\lambda, 1]$ is a Markov decomposition, using it one gets semiconjugation from a one-sided geometric Markov chain to T, where $A = \{1, 2\}$,*

$$M = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

$$M' = \begin{bmatrix} 0 & 1 \\ 2-\lambda & \lambda - 1 \end{bmatrix}$$

$$x' = \begin{bmatrix} \frac{1}{2+\lambda} \\ \frac{1+\lambda}{2+\lambda} \end{bmatrix}$$

$$\mu_{PF}(\Pi_{122}) = \frac{1}{2+\lambda} \cdot 1 \cdot (\lambda - 1) = \frac{\lambda - 1}{2+\lambda}$$

$$\mu_{PF}(\sigma^{-1}(\Pi_{122})) = \mu_{PF}(\Pi_{2122}) = \frac{1+\lambda}{2+\lambda} \cdot (2-\lambda) \cdot 1 \cdot (\lambda - 1) = \frac{\lambda - 1}{2+\lambda}$$

10/14

Theorem 3.9. *σ is mixing under μ_{PF}.*

Proof. Firstly we show that if A and B are both cylinder sets,

$$\lim_{n\to\infty} \mu_{PF}(A\cap\sigma^{-n}(B)) = \mu_{PF}(A)\mu_{PF}(B)$$

Suppose $A = \Pi_w,\ B = \Pi_{w'},\ w \in \{1,\dots,m\}^{n_1},\ w' \in \{1,\dots,m\}^{n_2}$. Let $k > 0$, then the set

$$A\cap\sigma^{-n_1-k}(B) = \{w'' \in S : w''_i = w_i \forall 0 \le i < n_1, w''_{i+n_1+k} = w'_i \forall 0 \le i < n_2\}$$

Is the disjoint union of all cylinder sets of the form $\Pi_{wvw'}$, here v' is a word of length k, and wvw' means the concatenation of w, v and w'. Hence,

$$\mu_{PF}(A\cap\sigma^{-n_1-k}(B)) = \sum_{i_1,\dots,i_k} x'_{w_0} M'_{w_0w_1}\dots M'_{w_{n_1-2}w_{n_1-1}} M'_{w_{n_1-1},i_1}$$

$$M'_{i_1,i_2}\dots M'_{i_{k-1}i_k} M'_{i_k,w'_0} M'_{w'_0w'_1}\dots M_{w'_{n_2-2}w'_{n_2-1}}$$

$$= x'_{w_0} M'_{w_0w_1}\dots M'_{w_{n_1-2}w_{n_1-1}}((M')^{k+1})_{w_{n_1-1}w'_0} M'_{w'_0w'_1}\dots M_{w'_{n_2-2}w'_{n_2-1}}$$

So

$$\frac{\mu_{PF}(A\cap\sigma^{-n_1-k}(B))}{\mu_{PF}(A)\mu_{PF}(B)} = \frac{((M')^{k+1})_{w_{n_1-1}w'_0}}{x'_{w'_0}} = \frac{((M'^T)^{k+1})_{w'_0w_{n_1-1}}}{x'_{w'_0}}$$

$$= 1 + \frac{((M'^T)^{k+1})_{w'_0w_{n_1-1}} - x'_{w'_0}}{x'_{w'_0}}$$

Let $e_{w_{n_1-1}}$ be the w_{n_1-1}-th basis vector in $\mathbb{R}^m$, i.e. the vector where the w_{n_1-1}-th entry is 1 and all other entries are zero, then $((M'^T)^{k+1})_{w'_0w_{n_1-1}} - x'_{w'_0}$ is the w'_0-th entry of

$$(M'^T)^{k+1}e_{w_{n_1-1}} - x' = (M'^T)^{k+1}(e_{w_{n_1-1}} - x')$$

Let $L = \{x : \sum_i x_i = 0\}$ be a $m-1$ dimensional subspace of $\mathbb{R}^m$, it is easy to check that L is invariant under M'^T, and the Perron-Frobenious eigenvector x' is not in L. Hence, by Perron Frobenious, all the eigenvalues of $M'^T|_L$ has to be smaller than 1, in other words for any vector $v \in L$, $(M'^T)^k v \to 0$ when $k\to\infty$. Because $e_{w_{n_1-1}} - x' \in L$, we have $(M'^T)^{k+1}e_{w_{n_1-1}} - x' \to 0$ when $k\to\infty$. This proved the condition for mixing when A and B are cylinder sets.

Now we show that the mixing condition is true for all measurable sets A and B. To show that we use the following lemma:

Lemma 3.10. *(Katok, prop. 4.2.10) To test for mixing, one only need to test for a family F of measurable sets that is called* sufficient, *i.e. for any measurable set B, there is a finite collection of disjoint sets F_i, such that $\mu(B\setminus\cup_i F_i) < \epsilon$, $\mu(\cup_i F_i\setminus B) < \epsilon$.*

Proof. Let A and B be two measurable sets. Let $\{F_i\}$, $\{G_i\}$ be two finite collections of disjoint sets in F such that

$$\mu(A\backslash \cup_i F_i) < \epsilon$$

$$\mu(\cup_i F_i \backslash A) < \epsilon$$

$$\mu(B\backslash \cup_i G_i) < \epsilon$$

$$\mu(\cup_i G_i \backslash B) < \epsilon$$

Let $A' = A\backslash \cup_i \{F_i\}$, $A'' = \cup_i\{F_i\}\backslash A$, $B' = B\backslash \cup_i \{G_i\}$, $B'' = \cup_i\{G_i\}\backslash B$, $A_0 = \cup_i F_i$, $B_0 = \cup_j G_j$. Then by assumption,

$$\mu(A_0 \cap T^{-n}(B_0)) \to \mu(A_0)\mu(B_0)$$

which is less than $2\epsilon + \epsilon^2$ from $\mu(A)\mu(B)$.

$$A \cap T^{-n}(B) = (A_0\backslash A'' \cup A') \cap (T^{-n}(B_0\backslash B'' \cup B')$$

So

$$(A_0 \cap T^{-n}(B_0))\backslash(A'' \cup T^{-n}(B'')) \subset A \cap T^{-n}(B) \subset (A_0 \cap T^{-n}(B_0)) \cup A' \cup T^{-n}(B')$$

Hence

$$|\mu(A \cap T^{-n}(B)) - \mu(A_0 \cap T^{-n}(B_0))| \leq 2\epsilon$$

Now let $\epsilon \to 0$ we proved the lemma. □

The verification that cylinder sets are sufficient is a standard measure theory argument, and will be left as an exercise. □

To define Perron-Frobenious invariant measures for bi-infinite shifts, use semiconjugation $h : A^{\mathbb{Z}} \to A^{\mathbb{N}}$, $(a_i) \mapsto (b_i)$, $b_i = a_i$ to define μ_{PF} on $h^{-1}(\Pi_w)$, then use shift invariance to extend it to all sylinder sets. More concretely, for any cylinder set in $\{1, \ldots, m\}^{\mathbb{Z}}$ of the form:

$$\Pi_w = \{w' \in S : w'_i = w_i \forall k \leq i \leq k'\}$$

Where $k < k' \in \mathbb{Z}$, $w \in \{1, \ldots, m\}^{\{k,k+1,\ldots,k'\}}$.

We have

$$\mu_{PF}(\Pi_w) = x'_{w_k} \prod_{j=k}^{k'-1} M'_{w_j,w_{j+1}}$$

One can verify by computation that μ_{PF} is additive, defines a Radon measure, invariant under σ. The proof that μ_{PF} in the two sided shift case is mixing is the same as the proof in the one sided shift case.

Example 3.11. *Under the semiconjugacy between $\{0,1\}^{\mathbb{N}}, \sigma$ and the cycle doubling map, it is easy to see that the pushforward of μ_{PF} in the shift map is the Lebesgue measure on $S^1 = \mathbb{R}/\mathbb{Z}$. Hence the cycle doubling map is mixing, hence ergodic, under the Lebesgue measure.*

Example 3.12. *Let* $\lambda = (\sqrt{5}+1)/2$, $X = [2-\lambda, 1]$, $T(x) = \lambda x$ *when* $x \leq 1/\lambda$ *and* $T(x) = 2 - \lambda x$ *when* $x > 1/\lambda$. *Use Markov decomposition* $I_1 = [2-\lambda, 1/\lambda]$, $I_2 = [1/\lambda, 1]$ *to build a semiconjugacy between it and subshift of finite type* $S \subset \{1,2\}^{\mathbb{N}}$, *here a sequence in* S *if* 1 *is followed by* 2 *and while* 2 *can be followed by either* 1 *or* 2. *Now by calculation, one can show that* μ_{PF} *on* S *induces the following measure on* X*:*

$$\mu_1 = h dx, \text{ where } h(x) = \begin{cases} \frac{\lambda}{3-\lambda} & x \leq 1/\lambda \\ \frac{\lambda+1}{3-\lambda} & x > 1/\lambda \end{cases}$$

Therefore, for a.e. $x \in X$, *the forward orbit of* x *will be dense in* X, *and the amount of time it spent in* I_1 *and* I_2 *will have ratio* $1 : \lambda^2$.

10/19

3.2 Measure-theoretic entropy

(Section 4.3, Katok)

Definition 3.13.
- *Let* μ *be a probability measure on* X. *A* **(finite) partition** $P = \{P_i\}$ *is a finite set of measurable subsets of non-zero measure such that* $\mu(X \backslash (\cup_i P_i)) = 0$, $\sum_i \mu(P_i) = 1$.
- *If* T *is a measure preserving map, define the pullback of a partition* P *as*

$$T^{-1}(P) = \{T^{-1}(P_i) : P_i \in P\}$$

- *The* **entropy** *of a partition is*

$$h(P) = -\sum_{i \in I} \mu(P_i) \log \mu(P_i)$$

- *Let* P *and* Q *be two partitions, we define the* **joint partition**

$$P \vee Q = \{P_i \cap Q_j : \mu(P_i \cap Q_j) = 0\}$$

- *Let* T *be a measure preserving map on* X. *we define*

$$h_T(\mu, P) = \lim \sup_{n \to \infty} \frac{h(\vee_{i=0}^n T^{-i}(P))}{n}$$

- *The* **measure theoretic entropy** *of* T *is defined as*

$$h_T(\mu) = \sup_P h_T(\mu, P)$$

Remark 3.14. *This concept of entropy is identical to the one in statistical physics or information theory. It quantifies how "chaotic" a map is after many iterations.*

Example 3.15. *Let S be a one sided or two sided subshifts of finite type, where the corresponding graph is strongly connected with lengths of loops having gcd 1. Let P consists of 1-cylinder sets, then $h_\sigma(\mu, P) = \log(\lambda)$ where λ is the Perron-Frobenious eigenvalue. To show this for one sided subshifts, note that for a k-cylinder set Π_w, there is some uniform $C > 0$ independent from w such that*

$$\mu_{PF}(\Pi_w) \in [\lambda^k/C, C\lambda^k]$$

Hence

$$h(\vee_{i=0}^n T^{-i}(P)) \in [(n+1)\log(\lambda) - \log(C), (n+1)\log(\lambda) + \log(C)]$$

The argument for two sided subshifts are similar.

The key result for calculating measure theoretic entropy is the Kolmogorov-Sinai theorem:

Definition 3.16. *A finite partition P by Borel set is called a* **generator** *if any σ-algebra containing either all elements in $\cup_n(\vee_{i=0}^n T^i(P))$, or, when T is invertible, all elements in $\cup_n(\vee_{i=-n}^n T^i(P))$, is the Borel σ-algebra.*

Theorem 3.17. *(Kolmogorov-Sinai, Katok Cor. 4.3.14) If P is a generator, then $h_T(\mu) = h_T(\mu, P)$.*

The key idea of the proof is the concept of relative entropy:

Definition 3.18. *Let P and Q be two partitions, we define*

$$h(P|Q) = -\sum_{P_i \in P, Q_j \in Q, \mu(P_i \cap Q_j) \neq 0} \mu(P_i \cap Q_j)(\log(\mu(P_i \cap Q_j)) - \log(Q_j))$$

Proof. (Sketch)

1. Step 1: use definition to prove that $h_T(\mu, P) = h_T(\mu, \vee_{i=0}^n T^i(P))$, and when T is invertible, $h_T(\mu, P) = h_T(\mu, \vee_{i=-n}^n T^i(P))$
2. Step 2: Prove by calculation that $h(P \vee Q) \geq h(P)$, $h(P|Q \vee R) \leq h(P|Q)$. Hence $h(P \vee Q|R) = h(P \vee Q \vee R) - h(R) = h(P|Q \vee R) + h(Q|R) \leq h(P|R) + h(Q|R)$.
3. Step 3: Use Step 2, prove that $h_T(\mu, Q) \leq h_T(\mu, P) + h(Q|P)$.
4. Step 4: Show that if P is a generator, then for any other partition Q, any $\epsilon > 0$, there is some partition P' whose elements are unions of elements in some $\vee_{i=0}^n T^i(P)$ or $\vee_{i=-n}^n T^i(P)$, and $h(Q|P') < \epsilon$.

Now the theorem follows from Step 1, 3 and 4. □

10/21

Remark 3.19. *The inequalities in Step 2 of the proof of Kolmogorov-Sinai show that in definition of $h_T(\mu, P)$, the* $\limsup$ *can be replaced with* $\lim$.

Remark 3.20. *From definition it is easy to see that:*

- *Entropy is invariant under measure preserving conjugacy (or semiconjugacy which has a.e. inverse).*
- *Entropy is non-increasing under semiconjugacy.*

Example 3.21. *If $X = \mathbb{R}/\mathbb{Z}$ and T is irrational rotation $x \mapsto x + a$, $a \notin \mathbb{Q}$, then $P = \{[0, 1/2], [1/2, 1]\}$ is a generator and by induction, $\vee_{i=0}^{n} T^{-i}(P)$ has at most $2(n+1)$ elements, hence has entropy no more than $\log(2n+1)$. Hence the entropy of T is 0.*

3.3 Topological entropy

(Katok Section 3.1, 4.5)

Let X be a compact metric space, T a continuous map on X.

Definition 3.22. *Let $N(\epsilon, n)$ be the smallest cardinality of finite subsets $A \subset X$ such that for any $x \in X$, there is some $a \in A$ such that $d(x, a) < \epsilon$, $\dots$, $d(T^{n-1}(a), T^{n-1}(x)) < \epsilon$. Then the topological entropy is*

$$h_{top}(T) = \lim_{\to 0} \lim \sup_{n \to \infty} \frac{\log(N(\epsilon, n))}{n}$$

An alternative definition is the following:

Definition 3.23. *Let $C = \{U_i\}$ be a finite open cover of X. Let $N(C, n)$ be the smallest number of sets of the form $\cap_{i=0}^{n-1}(T^{-i}(U_{j_i}))$ that covers the whole X. Then*

$$h_{top}(T) = \sup_{C} \lim \sup_{n \to \infty} \frac{\log(N(C, n))}{n}$$

Proposition 3.24. *The two definitions are equivalent.*

Proof. To show that the first definition is no bigger than the second, let ϵ_0 be such that $\lim \sup_{n \to \infty} \frac{\log(N(\epsilon_0, n))}{n} > h - \epsilon'$, pick $C = \{U_i\}$ such that the diameter of U_i is less than ϵ_0, and let $\epsilon' \to 0$.

To show the other direction, let C_0 be a cover such that $\lim \sup_{n \to \infty} \frac{\log(N(C_0, n))}{n} > h - \epsilon'$, ϵ be small enough that any two points with distance less that ϵ lies in the same element of C_0, and let $\epsilon' \to 0$.

□

Hence, the definition of topological entropy depends on the topology and not the metric.

10/26

Example 3.25. • *If T is isometry then the topological entropy is 0.*

- *For subshifts calculation of topological entropy reduces to the growth of cylinder sets. Let G be a strongly connected graph whose lengths of loops has gcd 1. In S_G where the 0-1 matrix for directed graph G is M, the number of k-cylinder sets is $[1, \dots, 1]M^{k-1}[1, \dots, 1]^T$ which grows like λ^k. Hence if it is a subshift of finite type the topological entropy should be* $\log(\lambda)$.

Theorem 3.26. *(Goodwyn, Goodman 1971)(Katok Thm. 4.5.3) Topological entropy is the sup of measure theoretic entropy for all probability measures.*

Proof. To show that measured theoretic entropy is bounded from above by topological entropy, given any partition $P = \{P_i\}$, pick $B_i \subset P_i$ to be closed subsets that has almost the same measure, $B_0 = X \setminus \cup_i B_i$, then $B = \{B_0, B_i\}$ is another partition "close" to P, hence $h_T(\mu, B) \approx h_T(\mu, P)$, the former is bounded above by $h_{top} + \log(2)$ due to convexity of $-\log$. Now

$$h_T(\mu, B) \leq \frac{1}{k} h_{T^k}(\mu, \vee_{i=0}^{k-1} T^{-i}(B)) \leq \frac{h_{top}(T^k) + \log(2)}{k}$$
$$\leq \frac{k h_{top}(T) + \log(2)}{k} = h_{top}(T) + \frac{\log(2)}{k}$$

Now let $k \to \infty$ this is proved.

To show that the supremum of measured theoretic entropy is the topological entropy, given any ϵ, take weak limit of the average of the δ measures supported on the points that realizes $N(\epsilon, n)$. □

10/28

Remark 3.27. *Topological entropy is non increasing under continuous semi-conjugacy. Hence, if a dynamical system admits a Markov decomposition, so that it is "almost conjugate" to a subshift of finite type with strongly connected transition graph (i.e. the place where it is not one to one has zero Perron-Frobenious measure), then the topological entropy must be no larger than the* log *of Perron-Frobenious eigenvalue λ_{PF}. The Perron-Frobenious measure pushes forward into an invariant measure with measure-theoretic entropy* $\log(\lambda_{PF})$, *hence the topological entropy must be* $\log(\lambda_{PF})$.

In particular, this shows e.g. the golden ratio tent map has topological entropy $\log(\frac{\sqrt{5}+1}{2})$, *and the baker's map has topological entropy* $\log(2)$.

3.4 Application: Bowen-Series coding

Let $\mathbb{H}^2$ be the hyperbolic plane (unit disc with metric $g = 4(dx^2 + dy^2)/(1 - x^2 - y^2)$, or Gaussian curvature -1). Let M be any compact oriented surface with genus at least 2, then there is a $\pi_1(M)$ action on $\mathbb{H}^2$ by isometry whose quotient is M. Pick an action where there is a right-angled (or all angles being $\frac{\pi}{n}$) fundamental domain. Then the fundamental domain D_0 and all its images under $\pi(M) = \Gamma$ action tiles $\mathbb{H}^2$ in a way that the sides form infinite geodesic rays in $\mathbb{H}^2$ called "walls". Let A be a generating set of $\pi_1(M)$ consisting of those that sends D_0 to the other fundamental domains separated from it by a wall.

Definition 3.28. *(Bowen-Series map and Bowen-Series coding Define a map T on $\mathbb{H}^2 \cup \partial\mathbb{H}^2$ as follows: the $2g$ walls touching the closure of D_0 separates $\mathbb{H}^2 \cup \partial\mathbb{H}^2$ into finitely many regions. If a region is separated from D_0 by a set of walls that contain sides of D_0, pick one in the set, apply g_i^{-1} where g_i is the group element that sends D_0 across that wall. The map sends D_0 to itself by identity.*

Remark 3.29. *When D_0 is the regular $4g$-gon where the angles are all $\pi/2g$, the number of regions would be $4g(4g-2)+1$.*

11/2

- When restricted to $\partial\mathbb{H}^2$, the set of allowable itineraries under the decomposition P via the walls touching D_0 is a subshift of finite type. The corresponding directed graph is strongly connected with lengths of loops having gcd 1.
- The construction can be generalized to surface with punctures and even orbifolds with punctures. In particular, it can be seen as a generalization to the continued fraction map.
- The closed geodesics on M correspond to periodic itineraries of this subshift of finite type. Hence there is a counting result on the number of closed geodesics that can be represented by an element of $\pi_1(M)$ within certain word length.

An application of B-S coding is the CLT for geodesic length (c.f. I. Gekhtman, S. J. Taylor, G. Tiozzo, A Central Limit Theorem for random closed geodesics: proof of the Chas-Li-Maskit conjecture):

Key steps:

- Counting closed geodesics (except for finitely many exceptions) is same as counting finite closed orbits of the shift map from B-S coding, which is same as integrating using B-S coding.
- Via "Theormodynamics formalism" one can get a family of invariant measures including the P-F measure (or "measure of maximal entropy"), called Gibbs measures.
- M. Ratner proved in 1973 (The Central Limit Theorem for geodesic flows on $n-$dimensional manifolds of negative curvature, Israel. J. of Math, 1973) that:

  Theorem 3.30. *Let $f : S_G \to \mathbb{R}$ be Hölder continuous and there is no u such that $f = u - u \circ T + \int f dv$, where v is Gibbs, then under v the sum of $f(T^i x)$ satisfies CLT.*

- For a finite word w, let g_w be the corresponding group element, and define $f(w) = d(z, g_w(z)) - d(z, g_{\sigma(w)}(z))$. Extend it to S_G, apply the CLT above.

Same argument works for intersection numbers between a random geodesic and a given closed geodesic.

4 Hyperbolic Dynamics and Structural Stability

4.1 Hyperbolic fixed points

Definition 4.1. *f is a smooth diffeomorphism on manifold M. A fixed point $p \in M$ is called* **hyperbolic** *if the eigenvalues of $df|_p$ has no eigenvalue on the unit circle.*

For flows, by considering first return map on a section, one can similarly define hyperbolic closed orbits.

4.1.1 Stable and unstable manifolds

Theorem 4.2. *(Hadamard-Perron) (Katok, Thm. 6.2.8) If p is a hyperbolic fixed point, there are two immersed submanifolds of M, passing through p via the subspace corresponding to eigenvalues with norm greater than 1 and smaller than 1 respectively, called the* **unstable** *and* **stable** *submanifolds.*

Proof. (Sketch)
Step 1: We only need to show the existence of these two manifolds in a small neigborhood, then use f and f^{-1} to expand them. WLOG we can assume $M = \mathbb{R}^n$, $p = 0$, the unstable and stable subspaces V^u and V^s are spanned by the first k and next $n-k$ coordinate vectors respectively. And we need to show that there is some submanifold W_0^u, W_0^s such that their tangent space at 0 are V^u and V^s respectively, and $W_0^u \subset f(W_0^u)$, $f(W_0^s) \subset W_0^s$. Now $W^u = \cup_{i=0}^{\infty} f^i(W_0^u)$, $W^s = \cup_{i=0}^{\infty} f^{-i}(W_0^s)$.

Let's assume $n = 2$, $k = 1$ for simplicity. The proof for the general case is similar. And let's focus on W_0^u, as the argument for W_0^s is identical. Let π_i be the orthogonal projection to the x_i direction.

Step 2: Pick T such that on square $S_T = \{(x_1, x_2) : |x_1| < T, |x_2| < T\}$, $f(x) = Ax + h(x)$, where $h'(0) = 0$, $|h'(x)| \leq \epsilon$, $|h(x)| \leq \epsilon$. As T decreases ϵ can be made arbitrarily small. Now if ϵ is small enough, the cone C defined by $|x_2| \leq |x_1|$ has $f(C \cap S_T) \subset C$, and $\partial S_T \cap C$ are sent to outside S_T. The reason is that

$$|\pi_1(f(x_1, x_2))| = |\lambda_1 x_1 + \pi_1(h(x_1, x_2))| \geq (|\lambda_1| - \sqrt{2}\epsilon)|x_1|$$

$$|\pi_2(f(x_1, x_2))| = |\lambda_1 x_1 + \pi_1(h(x_1, x_2))| \leq (|\lambda_2| + \sqrt{2}\epsilon)|x_2|$$

So as long as $\sqrt{2}\epsilon < \min\{|\lambda_1| - 1, 1 - |\lambda_2|\}$ this would be true.

Step 3: Now let N consisting of 1−Lipschitz functions from $[-T, T]$ to itself such that the graph is in C, then N is a complete metric space under the metric

$$d(g_1, g_2) = \sup_{x \in [-T,T], x \neq 0} \frac{|g_1(x) - g_2(x)|}{|x|}$$

Now define a map Ψ sending N to itself, such that the map f sends the graph of g to $\Psi(g)$. Similar argument as in Step 2 shows that if g is in N then the image of the graph of g under f is the graph of another element in N. Now start with the constant function $g_0 = 0$, let $g_{i+1} = \Psi g_i$, then the graph of the limit of g_i would be the candidate of W_0^u. The only thing left to proof is that the limit is also smooth.

Step 4: Firstly we show that the limit g_* is C^1. For every $x \in (-T, T)$, consider the set of tangent directions $T_x = \{k : \exists x_i \to x, \lim_{i\to\infty} \frac{|g_*(x_i)-g_*(x)|}{|x_i-x|} = k\}$. Then $T_x \subset [-1, 1]$, and if f sends $(x', g_*(x'))$ to $(x, g_*(x))$, df induces a fractional linear map f_* sending $T_{x'}$ to T_x. Pick ϵ small enough one see that this map is contracting, hence T_x must all have diameter 0, which implies that g_* is differentiable. On the other hand, these fractional linear maps induces a map from continuous functions on $[-T, T]$ with values in $[-1, 1]$ to continuous functions Φ on $[-T, T]$ with values in $[-1, 1]$, by $(\Phi(g))(x) = f_*(g(x'))$, and one can show that if ϵ is sufficiently small, Φ will be contracting under the uniform norm, which means that g'_* being the fixed point of Φ must be continuous.

Step 5: Now one can use chain rule to increase the smoothness to C^∞. □

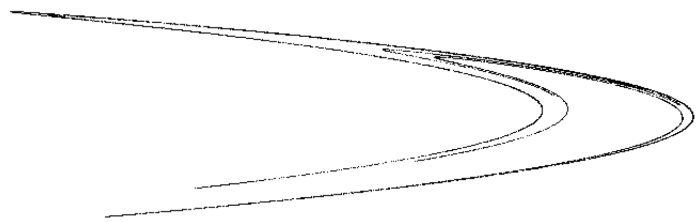
11/9

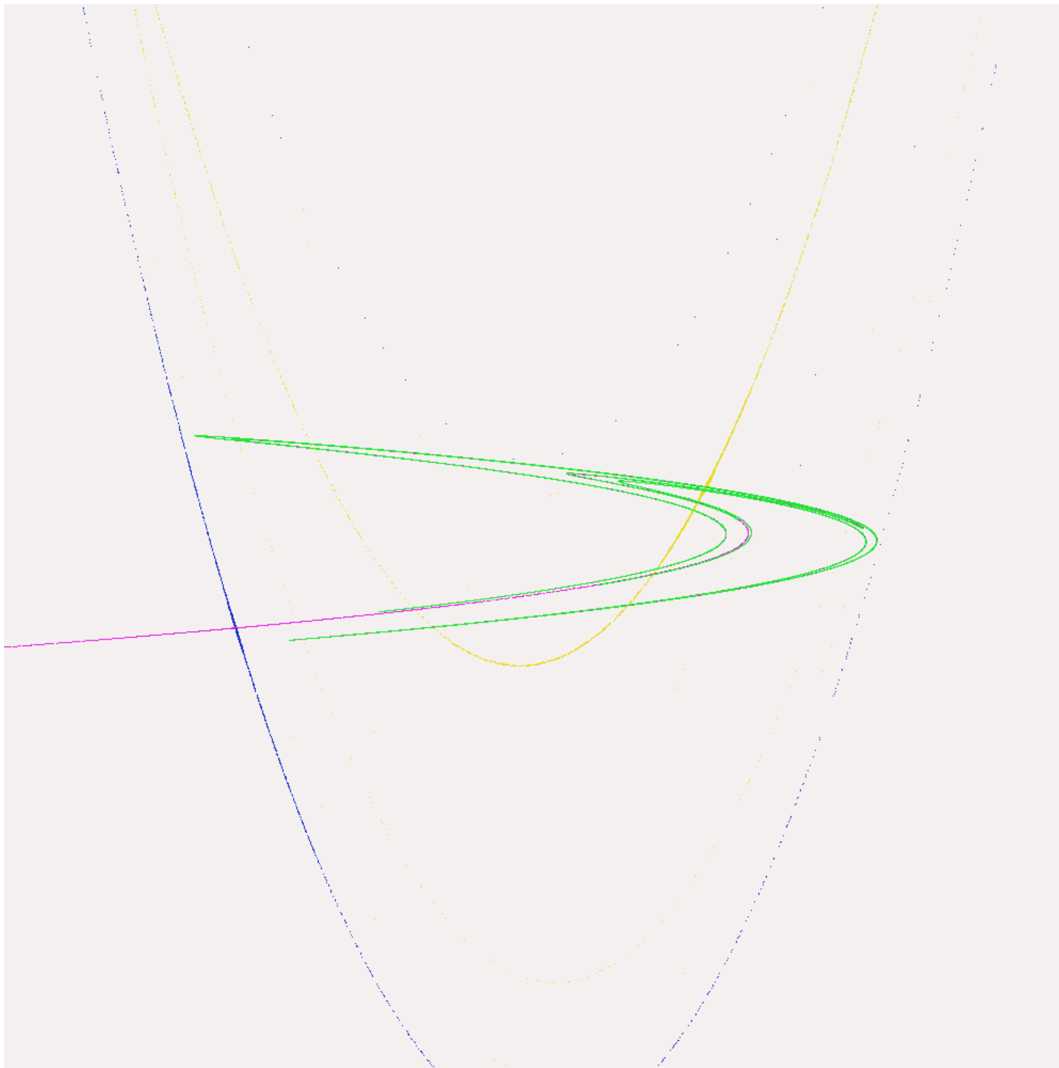
Applications for stable/unstable manifolds:

- They provide a characterization of the map at a local neighborhood of a fixed point.
- When they intersect once transversely, they intersects infinitely many times, and the map has chaotic behavior on the closure of the intersection sets.

Example 4.3. *For anosov maps, the stable, unstable manifolds and their intersections are dense on M.*

Example 4.4. *Henon map:* $(x, y) \mapsto (1 - 1.4x^2 + y, 0.3x)$





4.1.2 Smale's Horseshoe and symbolic dynamics (6.5 in Katok)

More concretely, we have:

Theorem 4.5. *(Lambda Lemma) If p is a hyperbolic fixed point, $i : B \to M$ a small coordinate chart near p such that $i(0) = p$, $B = \{(x_u, x_s) : \|x_u\| < T, \|x_s\| < T'\}$, and $\{i(x_u, 0)\}$ and $\{i(0, x_s)\}$ are neighborhoods of p in W^u and W^s respectively. Suppose there is a smooth manifold V intersecting with $\{i(0, x_s)\}$ transversely and under i^{-1} is the graph of a C^1 Lipschitz function, then*

- *When n is sufficiently large, $i^{-1}(f^n(V) \cap U)$ is the graph of a C^1 Lipschitz function.*

- *As $n \to \infty$, these functions C^1 converges to 0. In order words, $f^n(V)$* **locally C^1 converges to W^u**.

The proof is similar to Hadamard-Perron.

A consequence is that if W^u and W^s has transverse intersection (there is a transverse homoclinic point) then there is a neighborhood of p which is a trivial disc bundle over a neighborhood of p in W^u, which we identify with $B = U^u \times D$, where U^u is a neighborhood of p in W^u. After applying f^N for $N >> 1$, B will intersect itself at $B_i = U^u \times D_i$ where D_i are two smaller discs. Now consider all elements in B whose forward and backward orbits under f^N are in B_0 or B_1, this is a Cantor set homeomorphic to $\{0,1\}^{\mathbb{Z}}$ where f^N action is a shift map. This f^N is called **Smale's Horseshoe**, and f^N restricted to the Cantor set has topological entropy $\log 2$, hence:

Theorem 4.6. *Existence of transverse homoclinic point implies non-zero topological entropy.*

11/11

In the argument above we made use of certain concepts about differentiable manifolds which we will review below:

Remark 4.7. *On a differentiable manifold M,*

- *A* **coordinate chart** *is a bijection from an open subset of $\mathbb{R}^n$ to M, such that the coordinate change maps between them are smooth.*
- *An* **immersion** *is a smooth map from another manifold of dimension m to M such that the derivative is always of rank m.*
- *An* **embedding** *is an immersion which is also the homormorphism to its image. Given any point in an immersed submanifold there is a neighbor of that submanifold where the inclusion map restricted to it is an embedding.*
- *For embedded submanifolds one can use any Riemannian metric to construct* **tubular neighborhood**.

Remark 4.8. *For solutions of smooth ODEs, there are two ways to turn a flow into a map: time-1 maps or first return maps. For time-1 map the fixed points are constant solutions or periodic solutions with period $1/n$. In the former case, the stable and unstable manifolds are unions of flow lines so can't have transverse intersection. For first return maps though transverse intersection is possible. Hence for smooth ODE to get horseshoe like chaotic behavior the dimension must be at least 3.*

4.1.3 Structural stability (6.3 in Katok)

Another key property for hyperbolic fixed points is the **structural stability**:

Theorem 4.9. *(Hartman-Grobman) Near a hyperbolic fixed point, smooth maps are locally continuously conjegate to its linearization.*

Proof. One approach is similar to the Example in 1.2.2.

Alternatively, WLOG assume $M = \mathbb{R}^n$, $f(x) = Ax + g(x)$ where $A = diag(A_1, A_2)$, A_1 has size $k \times k$, A_2 has size $(n-k) \times (n-k)$, $\|A_1^{-1}\| < 1$, $\|A_2\| < 1$, and $g(0) = g'(0) = 0$, $g \in C^\infty$. We shall show that at a neighborhood of 0 there is some continuous function v such that $(A+g)\circ(id+v) = (id+v)\circ A$.

Let $*_u$ and $*_s$ be the first k and next $n-k$ components respectively, and define $v_0 = 0$, let

$$(A+g)_u(id + [v_{n+1,u}, v_{n,s}]^T) = A_1 \circ \pi_u + v_{n,u} \circ A$$

$$(A+g)_s(id + v_n) = A_2 \circ \pi_s + v_{n+1,s} \circ A$$

Here π_u and π_s are the orthogonal projection the first k and next $n-k$ coordinates respectively. The assumptions on the norms of A_1^{-1} and A_2 implies that the map from v_n to v_{n+1} is uniform contracting in a small neighborhood of 0, hence there is a unique fixed point which is the desired v. Carry out the same argument one gets v' such that $A \circ (id + v') = (id + v') \circ (A + g)$, and one can further show that $id + v'$ is the inverse of $id + v$, hence both are homeomorphisms. □

11/16

4.2 Hyperbolic sets (6.4 and 18 in Katok)

Definition 4.10. *We call f-invariant set Λ to be a hyperbolic set, if the tangent bundle at Λ can be decomposed into 2 invariant subbundles, where there is a Riemannian metric on it such that $|df^{-1}| < \lambda < 1$ on one and $|df| < \lambda < 1$ on the other.*

These splitting can be obtained via nested subcones, similar to proof of Hadamard-Perron, hence is stable. Hadamard-Perron can be applied to hyperbolic sets to create stable and unstable submanifolds:

$$W^s(x) = \{y : \lim_{n\to\infty} d(f^n(y), f^n(x)) = 0\}, W^u(x) = \{y : \lim_{n\to\infty} d(f^{-n}(y), f^{-n}(x)) = 0\}$$

Theorem 4.11. *If Λ is hyperbolic, then there is an open neighborhood U of Λ where any g sufficiently close to f in C^1, $\Lambda^g = \cap_{n\in\mathbb{Z}} g^n(\overline{V})$ is hyperbolic.*

The proof is by definition.

As a consequence, if $\Lambda = M$ (called **anosov**), then any map C^1 close to it is also anosov.

If $\Lambda = \Lambda^f$ we call Λ **locally maximal**.

The key property of hyperbolic set is **shadowing lemma**:

Definition 4.12. *A ϵ-pseudoorbit is a sequence $\{x_i\}$ such that $d(x_{i+1}, f(x_i)) < \epsilon$.*

Theorem 4.13. *Given a hyperbolic set Λ, there is a neighborhood U such that for any $\delta > 0$, there is ϵ such that any ϵ-pseudo orbit can be δ-approximated by a true orbit.*

Proof. (Sketched) The argument is similar to Hartman-Grobman in a sense. Suppose there is just expanding subbundle and no contracting subbundle, x_i is an ϵ-pseudo orbit, then $y_i = f(x_{i-1})$ is a nearby $\lambda\epsilon$-pseudo orbit. Continue the process we can make them convergent, local maximality implies that limit is in Λ. Similarly one can do it for the case where there is just contracting subbundle. In the general case, what needs to be done is to separate x_i into "coordinates" in the contracting and expanding direction and carry out the two processes separately. □

Applications:

- Compact locally maximal hyperbolic sets admits Markov decomposition. (Theorem 18.7.3). The basic idea is to find a finite set of points in Λ such that their ϵ-ball cover Λ, and make directed graph with these points as vertices, an edge from p to q if $d(f(p), q) < \epsilon$. Then paths on this graph correspond to ϵ pseudo orbits that can be approximated by true orbits.
- Hyperbolic sets are structurally stable. The idea is that perturbation turn orbits into pseudo-orbits, which can then be approximated.

Remark 4.14. *If the ω-limit set is hyperbolic and compact, and periodic orbits are dense in it (which implies locally maximal) then we call it a* **Axiom A** *map. Axiom A dynamical systems can be studied using geometric Markov chain and are also structurally stable.*

11/18

Some examples of hyperbolic flows (whose first return map is hyperbolic)

Example 4.15.

- *Lorenz system.*
- *Geodesic flow on hyperbolic manifolds (manifold with constant negative sectional curvature). This is actually Anosov, in the sense that tangent space has invariant subbundles $TM = E_0 + E_1 + E_2$ where E_0 is flow direction, and the flow is uniform expanding on E_1 and uniform contracting on E_2. To see the decomposition, consider the universal cover of M which is a n-dimensional hyperbolic space. Now geodesics on M are lifted to geodesics on this hyperbolic space $\mathbb{H}^n$, and as time goes to $\pm\infty$ the geodesic "converges" to two points at the boundary. Here E_0 means parallel translation in the flow line direction, E_1 means moving to nearby geodesics but keeping the limit at $t \to -\infty$, E_2 means moving to nearby geodesics but keeping the limit at $t \to +\infty$.*

5 Dynamics on Homogeneous Spaces

We will discuss the simplest case: dynamics on $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$.

- $SL(2,\mathbb{R})$ acts on upper half plane via fractional linear transformation from the left. This action preserves the hyperbolic metric there.
- The induced action on the unit tangent bundle is transitive, where the stablizer is $\pm I$. Hence the unit tangent bundle is identified with $SL(2,\mathbb{R})/\pm I$.
- Now one can write down $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$ as the unit tangent bundle of a hyperbolic orbifold.

Definition 5.1.
- *The* **geodesic flow** g_t *is defined as right multiplication by* $diag(e^t, e^{-t})$. *It is Anosov via Example 4.15.*
- *The* **horocycle flow** h_t *is defined as right multiplication by* $\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$.
- **Renormalization:** $g_{-s}h_t g_s = h_{e^{-2s}t}$

11/23

There is an invariant measure for both flows which is the unique $SL(2,\mathbb{R})$-invariant probability measure on $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$ called the **Harr measure**.

Proposition 5.2. *Under identification between $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$ and the unit tangent bundle of an orbifold, this Harr measure becomes locally product of hyperbolic and Euclidean measure.*

(Einsiedler-Ward Section 11.3)

Theorem 5.3. *Let H be a hilbert space, $SL(2,\mathbb{R})$ acts on H continuously via unitary maps. If $v \in H$ is fixed by a subgroup L then it is also fixed by any other element h such that for every δ, the δ ball of h, $B(h,\delta)$, has non empty intersection with $LB(e,\delta)L$.*

Proof. WLOG assume $\|v\| = 1$. Let $p(h) = (h(v), v)$ be a function on $SL(2,\mathbb{R})$, then p is continuous, and $p(l_1 h l_2) = p(h)$ for any $l_1, l_2 \in L$, hence $1 = p(e) = p(h)$ by assumption on the δ-balls which implies $h(v) = v$ due to Cauchy-Schwartz. □

Theorem 5.4. *The geodesic and horocycle flows are both ergodic under the Harr measure.*

Proof. Let L be diagonal and h be unipotent (due to renormalization) then one shows that functions invariant under geodesic flow are also invariant under unipotent flow, hence invariant under the whole $SL(2,\mathbb{R})$ action and be the Harr measure, hence the geodesic flow is ergodic. Do it the other way, one get that the horocycle flow is ergodic. □

The geodesic flow can further be shown to be mixing, which result in classification of horocycle invariant measures.

The ideas above can be generalized to general Lie groups, which was used by Margulis to prove that if Q is indefinite non degerate quadratic form with 3 or more variables, not a multiple of rational form, then $Q(\mathbb{Z}^n)$ is dense in $\mathbb{R}$.

Review:

- Ergodicity, unique ergodicity, entropy
- Geometric Markov chains
- Reducing dynamical systems to geometric Markov chain via hyperbolicity
- Homogenuous dynamics via renormalization.