# Math 758 Introduction to Ergodic Theory and Dynamics

May 15, 2025

## Contents

# 1 Syllabus

Institution Name: University of Wisconson-Madison

Credits: 3

Requisites: Graduate/professional standing.

Course Designation: Grad 50% - Counts toward 50% graduate coursework requirement

Repeatable for Credit: No

Official course description: An introduction to ergodic theory and dynamics covering fundamental theorems of ergodic theory, classical examples of one and two dimensional dynamics as well as applications to study of group actions.

Instructor: Chenxi Wu

Email: cwu367@wisc.edu

Modes of Instruction: In person.

Lecture: 1-2:15 pm Tu Th VV B235

Exam Date: 5/4/25, 12:25-2:25pm

Office Hours: 10-11 am Wednesdays and Thursdays at VV 517 or by appointment.

Learning goal:

1. Analyze a range of mathematical proofs in discrete and continuous dynamical systems and recognize the underlying principles and techniques.
2. Write clear and well-reasoned mathematical arguments to describe the long time behavior of measure preserving transformations.
3. State the Birkhoff, von Neumann, and maximal ergodic theorems and reproduce their proofs.
4. Apply fundamental ergodic theorems from classical dynamics to low-dimensional examples such as the circle rotation, interval exchange transformations, and geodesic and horocycle flows on two dimensional manifolds.
5. Apply results in ergodic theory and dynamics to solve novel problems in geometric group theory and geometric topology.

Textbook: I will post detailed lecture notes. Some of the lectures might be based on the textbook by Einsiedler and Ward and the textbook by Anatole Katok.

How Credit Hours are met: This class meets for two, 75-minute class periods each week over the fall/spring semester and carries the expectation that students will work on course learning activities (reading, writing, problem sets, studying, etc) for about 3 hours out of the classroom for every class period. The syllabus includes more information about meeting times and expectations for student work.

Canvas Support: `https://kb.wisc.edu/luwmad/page.php?id=66546`

Grades: There will be homework assignments once every 2 weeks, to be submitted on Canvas, accounting for 30% of the grades, a final presentation of about 20 minutes on a topic related to dynamics that you are interested in, accounting for 30% of the grades, and a final exam accounting for 40% of the grade. A grade of 85 or above will get you an A and a grade of 70 or above will get you a letter grade of B or above. Missing HW assignments will be replaced by final grades, so don't worry if you need to miss a few week's lecture for personal reasons, just read the posted lecture notes afterwards and let me know if you have any questions!

This is a tentative list of topics we may cover in this semester:

- Ergodic Theorems
- Unique ergodicity
- Symbolic Dynamics
- Mixing and entropy
- Dynamics of group actions
- Dynamics on hyperbolic surfaces
- Dynamics of interval maps

Institutional Level Academic Policies: `https://teachlearn.wisc.edu/course-syllabi/course-credit-information-required-for-syllabi/`

# 2 Analysis Review

## 2.1 Topology and Metric

1. Let $X$ be a set. A **topology** is a subset $\mathcal{O} \subseteq P(X)$ such that:
   (a) $\emptyset \in \mathcal{O}$, $X \in \mathcal{O}$.
   (b) $A, B \in \mathcal{O} \implies A \cap B \in \mathcal{O}$.
   (c) $S \subseteq \mathcal{O} \implies \bigcup S \in \mathcal{O}$.
2. When $O = \{\emptyset, X\}$ we call it the **trivial topology**. When $O = P(X)$ we call it the **discrete topology**.
3. Let $\mathcal{B} \subseteq P(X)$, the smallest topology containing $\mathcal{B}$ is called the topology **generated by** $\mathcal{B}$, and $\mathcal{B}$ is called a **basis** of this topology.
4. Let $(X, \mathcal{O})$ be a topological space.
   (a) The elements of $\mathcal{O}$ are called **open sets**, their complements called **closed sets**. Let $x \in X$, an open set containing $x$ is called an **open neighborhood** of $X$, and a set containing an open neighborhood of $x$ as a subset is called a **neighborhood**.
   (b) Let $A \subseteq X$, the **closure** of $A$ is $\overline{A} = \bigcap\{V \text{ closed} : A \subseteq V\}$, and the **interior** of $A$ is $A^\circ = \bigcup\{U \text{ open} : U \subseteq A\}$. The **boundary** of $A$ is $\partial A = \overline{A} \backslash A^\circ$.
   (c) $A$ is called **dense** if $X = \overline{A}$. $X$ is called **separable** iff it has a countable dense subset.
   (d) Let $\{x_n\}$, $n \in \mathbb{N}$ be a sequence of elements in $X$. We say $y \in X$ is a **limit point** if for every neighborhood $N$ of $y$, $\{n \in \mathbb{N} : x_n \in N\}$ is infinite. We say that $y$ is a **limit** of the sequence if for any neighborhood $N$ of $y$, there is some $M > 0$ such that $n > M$ implies $x_n \in N$. When limit of a sequence exists we call it **convergent**.
   (e) We call $X$ **Hausdorff** if any two distinct points in $X$ have disjoint open neighborhoods. When $X$ is Hausdorff, the limit of a sequence, when exists, must be unique.
   (f) A subset of $O$ whose union equals $X$ is called an **open cover**. We say $X$ is **compact**, if any open cover has a finite subcover. We say $X$ is **locally compact**, if any point in $X$ has a compact neighborhood. We say $X$ is **sequentially compact**, if any sequence has a convergent subsequence.
   (g) If $X$ is compact, a nested sequence of non empty closed subsets has non empty intersection.
5. Let $(X, O)$ be a topological space, $A \subseteq X$, we define the **subspace topology** on $A$ as $\{A \cap U : U \in O\}$.

6. Let $(X_\alpha, O_\alpha)$, $\alpha \in \Omega$ be a non-empty family of topological spaces, we define the **product topology** on $\prod_{\alpha\in\Omega} X_\alpha$ as the topology generated by

$$\{\{p \in \prod_\alpha X_\alpha : \pi_{\alpha_i}(p) \in U_{\alpha_i}\} : \alpha_1, \ldots, \alpha_k \in \Omega, U_{\alpha_i} \in O_{\alpha_i},$$

$$\pi_\alpha \text{ are the projection to factor } \alpha\}$$

7. Let $(X, O)$ be a topological space, $\sim$ an equivalence relation on $X$, we define the **quotient topology** on $X/\sim$ as $\{A \subseteq X/\sim : \pi^{-1}(A) \in O\}$. Here $\pi : X \to X/\sim$ is the quotient map.

8. Let $(X, O)$, $(Y, O')$ be two topological spaces, $f : X \to Y$ is called **continuous** if for all $V \in O'$, $f^{-1}(V) \in O$. When $Y = \mathbb{R}$, $O'$ the topology generated by all open intervals, we call such a map a **(real valued) continuous function**. The set of continuous functions on $X$ form a vector space (you can show that via the observation that compositions of continuous functions are continuous), denoted as $C(X)$.

9. If $X$ is compact, $f : X \to Y$ is continuous, then $f(X)$ is compact. As a consequence, any continuous function on compact set has maximum and minimum.

10. Let $X$ be a set, $d$ a $\mathbb{R}_{\geq 0}$ function on $X \times X$ such that

    (a) $d(x, y) = 0$ iff $x = y$.

    (b) $d(x, y) = d(y, x)$.

    (c) $d(x, z) \leq d(x, y) + d(y, z)$.

    then we call $d$ a **metric** and $X$ a **metric space**. The topology on $X$ is then defined as the one generated by all the **open balls** $B(x, r) = \{y \in X : d(x, y) < r\}$.

11. Let $(X, d)$ be a metric space, $\Gamma < Isom(X)$ where $Isom(X)$ is the group of isometries on $X$, such that for every $x \in X$, there exists $\epsilon > 0$ such that $\{g \in \Gamma : d(x, g(x)) < \epsilon\}$ is finite. Define $\sim$ as $a \sim b$ iff there is $g \in \Gamma$, $b = g(a)$, then $X/\sim$ can be made into a metric space with **quotient metric** $d'([x], [y]) = \inf_{g\in\Gamma} d(x, g(y))$. As an example, let $X = \mathbb{R}$, $d(x, y) = |x-y|$, $\Gamma = \{x \mapsto x + n : n \in \mathbb{Z}\}$, then this quotient metric makes $\mathbb{R}/\mathbb{Z}$ into a metric space.

12. A metric space is compact iff it is sequentially compact.

13. Let $(X, d)$ be a metric space, we say $(X, d)$ is **complete**, if any Cauchy sequence converges. Completeness is NOT a topological property, for example, on $\mathbb{R}$, the metric $d(x, y) = |x - y|$ and $d'(x, y) = |e^x - e^y|$ defines the same topology, the former is complete while the latter isn't. By "adding" limits of Cauchy sequences to a metric space we can embed it into a larger complete metric space, called its **completion.**

14. (Baire Category Theorem) Let $X$ be a non empty complete metric space, then any countable intersection of dense open sets is non empty.

    - To prove it, use the fact that when $X$ is complete, a nested sequence of closed balls with radius $\to 0$ has non empty intersection.
    - As applications, $\mathbb{Q} \subseteq \mathbb{R}$ can not be the intersection of countably many open sets, a smooth function $f$ on $\mathbb{R}$ such that for every $x \in [0, 1]$, $f^{(n_x)}(x) = 0$ for some integer $n_x$, is a polynomial when restricted to $[0, 1]$.

15. When $X$ is compact, we define the metric on $C(X)$ as $d(f, g) = \max_{x\in X} |f(x)-g(x)|$. $C(X)$ is a complete metric space under this metric.

16. (Stone-Weierstrass Theorem) Let $X$ be a compact Hausdorff space. Let $L$ be a subspace of the vector space $C(X)$ such that:

    (a) $1 \in L$.

    (b) $f, g \in L \implies fg \in L$

    (c) For any $x, y \in X$, there is some $f \in L$ such that $f(x) \neq f(y)$.

    Then $L$ is dense in $C(X)$.

    - To prove it, first show that $max(x, y)$ can be approximated by polynomials, then, for any $f \in C(X)$, $a, b \in X$, let $g_{a,b} \in L$ be a function that is close to $f$ at points $a$ and $b$. Now pick finitely many $b_i$ such that $g_a = \min\{g_{a,b_i}\}$ satisfies $g_a < f + \epsilon$, pick finitely many $a_j$ such that $g = \max\{g_{a_j}\}$ satisfies $g > f - \epsilon$. The existence of $b_i$, $a_j$ is due to compactness of $X$.
    - If $X$ is a compact metric space, $X$ is separable, and by Stone-Weierstrass, $C(X)$ is separable as well.

17. Let $X$ be a separable metric space, $F \subseteq C(X)$ is called **uniformly bounded** if there is some $M$, such that for all $f \in F$, $x \in X$, $|f(x)| < M$. It is called **uniformly equicontinuous** if for any $\epsilon > 0$, there is $\delta > 0$, such that for any $f \in F$, any $x, y \in X$ such that $d(x, y) < \delta$, $|f(x) - f(y)| < \epsilon$.

    (Arzela-Ascoli Theorem) Let $X$ be a separable metric space, $\{f_n\} \in C(X)$ a sequence which is uniformly bounded and uniformly equicontinuous. Then there is a subsequence of $\{f_n\}$ converges uniformly on any compact subset $K \subseteq X$.

    - Proof: diagonal trick.
    - An application is Montel's Theorem in complex analysis: if $\mathcal{F}$ is a family of holomorphic functions on open set $\Omega \subseteq \mathbb{C}$ and is uniformly bounded, then any sequence in $\mathcal{F}$ has a subsequence convergent uniformly on any compact subset of $\Omega$.

## 2.2 Topological Vector Spaces

1. A vector space $V$ is called a **topological vector space** if it has a topology, where addition and scalar multiplication are both continuous.

2. Let $V$ be a vector space, a **norm** $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ is a function such that

   (a) $\|x\| = 0$ iff $x = 0$.

   (b) $\|cx\| = |c|\|x\|$.

   (c) $\|x + y\| \leq \|x\| + \|y\|$.

   A normed vector space has a metric $d(x, y) = \|x - y\|$, and it is called a **Banach Space** if this metric is complete.

3. (a) When $X$ is compact, $(C(X), f \mapsto \max_{x\in X} |f(x)|)$ is a Banach space.

   (b) Let $l^\infty$ be the vector space of bounded sequences, $\|(a_i)\|_\infty = \sup_i |a_i|$, then $(l^\infty, \|\|_\infty)$ is a Banach space.

   (c) Let $1 \leq p < \infty$, $l^p$ be the set of sequences $(a_i)$ such that $\sum_i |a_i|^p < \infty$, $\|(a_i)\|_p = \left(\sum_i \|a_i\|^p\right)^{1/p}$, then $(l^p, \|\cdot\|_p)$ is a Banach space.

4. A linear map between two Banach spaces $f : V \to W$ is **continuous** iff it is **bounded**, i.e. there is some $L$ such that $\|f(x)\| \leq L\|x\|$ for all $x \in V$.

   (a) (Open Mapping Theorem) $T : V \to W$ is a linear bounded map between Banach spaces and also surjective, then $T$ sends open sets to open sets. (Proof: if not true then $\{W \backslash \overline{\{T(v) : \|v\| \leq n\}}\}$ are nested dense open sets whose intersection is empty, contradicting with Baire Category Theorem).

   (b) (Closed Graph Theorem) $T : V \to W$ is a linear map between Banach spaces, then $T$ is continuous iff the graph of $T$ in $V \times W$ (under product topology) is closed. (A consequence of Open Mapping Theorem).

5. Let $V$ be a Banach space, the vector space of bounded linear functions on $V$ is called the **dual space**, denoted as $V^*$.

   (a) The **strong** topology on $V^*$ is the topology defined using the operator norm $\|f\| = \sup_{\|x\|\leq 1} |f(x)|$.

   (b) The **weak** topology on $V$ is the smallest (under containment) topology where $v \mapsto f(v)$ for all $f \in V^*$ are continuous.

   (c) The **weak-**$*$ topology on $V^*$ is the smallest topology where $f \mapsto f(v)$ for all $v \in V$ are continuous.

   (d) If $V$ is separable, the closed unit ball (under operator norm) under the weak-$*$ topology is metrizable (follows from definition), sequentially compact (by Arzela-Ascoli), and compact (by the previous two or by Tychonoff's theorem).

6. $H$ a real (complex) vector space, an **inner product** $(\cdot,\cdot): V \times V \to \mathbb{R}$ ($\mathbb{C}$) is a function satisfying:

   (a) $(a,b) = \overline{(b,a)}$, where $\bar{\cdot}$ is the complex conjugation.

   (b) $a \mapsto (a,b)$ is linear for all $b \in H$.

   (c) $(a,a) \geq 0$, $(a,a) = 0$ iff $a = 0$.

   A norm on $H$ can then be defined as $\|x\| = \sqrt{(x,x)}$. If $H$ is complete under this norm we call it a real (or complex) Hilbert space. $l^2$ is a Hilbert space.

7. If $H$ is a Hilbert space, $L \subseteq H$ a subspace, for every $z \in H$, $z$ can be uniquely written as $z = z_1 + z_2$, where $z_1 \in L$, $z_2 \in L^{\perp} = \{x \in H : (x,l) = 0 \text{ for all } l \in L\}$. $z_1$ is called the **orthogonal projection** of $z$ onto $L$. To prove this, let $z_1$ be the element in $L$ closest to $z$.

8. (Riesz Representation Theorem for Hilbert Space) If $H$ is a Hilbert space, $H^*$ the dual space with operator norm, there is a bijective isometry $i : H \to H^*$ such that $i(x) = (y \mapsto (y,x))$.

9. (Hahn-Banach Theorem) $V$ a real vector space, $p : V \to \mathbb{R}$ a function such that $p(rx) = rp(x)$ for $r \geq 0$, $p(x+y) \leq p(x) + p(y)$ (called **sublinear**), $V'$ a subspace, $f'$ a linear function on $V'$ such that $f' \leq p$, then $f'$ can be extended to a linear function $f$ on $V$ such that $f \leq p$. Proof is via Zorn's Lemma.

## 2.3 Measure Theory

1. Let $X$ be a set. A **$\sigma$-algebra** $\Sigma$ is a subset of $P(A)$ such that

   (a) $X \in \Sigma$

   (b) $A \in \Sigma$ then $X \backslash A \in \Sigma$.

   (c) Countable union of elements of $\Sigma$ is in $\Sigma$.

   The pair $(X, \Sigma)$ is called a measurable space. The smallest $\sigma$-algebra containing a subset $S \subseteq P(X)$ is called the $\sigma$-algebra generated by $S$.

2. If $X$ is a topological space, the smallest $\sigma$-algebra generated by the set of all open sets is called the **Borel** $\sigma$-algebra, its elements are called **Borel sets**.

3. Let $(X, \Sigma)$ and $(Y, \Sigma')$ be two measuable spaces, we call a map $f : X \to Y$ to be **measurable** if for any $A \in \Sigma'$, $f^{-1}(A) \in \Sigma$. When $Y = \mathbb{R}$ or $\mathbb{C}$ we set $\Sigma'$ as the Borel $\sigma$-algebra by default.

4. A **measure** on a measurable space $(X, \mu)$ is a function $\mu : \Sigma \to \mathbb{R} \cup \{\infty\}$ such that

   (a) For all $A \in \Sigma$, $\mu(A) \geq 0$.

(b) $\mu(\emptyset) = 0$.

(c) If $S \subseteq \Sigma$ is a countable subset whose elements are pairwise disjoint, then $\mu(\bigcup S) = \sum_{A\in S} \mu(A)$.

It is called a **probability measure** if $\mu(X) = 1$, called **finite** if $\mu(X) < \infty$, called **$\sigma$-finite** if $X$ is the union of countably subsets of finite measures.

5. $\delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$ is a measure, called the **Dirac measure**.

6. Let $(X, \Sigma, \mu)$ be a measurable space, $f : X \to Y$ a map. The **pushforward measure** $f_*(\mu)$ is defined on $\Sigma' = \{B \in P(Y) : f^{-1}(B) \in \Sigma\}$ as $(f_*(\mu))(B) = \mu(f^{-1}(B))$.

7. By **"almost every"** or **"generically"** we mean something is true other than a zero measure set.

8. A measurable space with measure $(X, \Sigma, \mu)$ is called **complete** if for any $\mu(A) = 0$, any $B \subseteq A$, we have $B \in \Sigma$. Any $(X, \Sigma, \mu)$ can be made complete by adding zero measure sets to $\Sigma$, and this process is called the **completion**. We often identify a measure with its completion implicitly.

9. On $\mathbb{R}^n$, define a Borel measure as $\mu(A)$ being the infimum of the sum of volumes of countably many boxes covering $A$. Its completion is called the **Lebesgue measure**.

10. Let $(X, \Sigma)$ be a measurable space, $\mu$ a measure on it, $f$ a non negative measurable function, define the **integral** of $f$ on $X$ as

$$\int_X f d\mu = \int_0^\infty \mu(f^{-1}((h, \infty)))dh$$

where the integral on the right hand side can be defined in the sense of Cauchy or Riemann. A real valued measurable function $g$ is called **integrable** if $\int_X |g|d\mu < \infty$, and its integral is defined as $\int_X g d\mu = \int_X \max(g, 0)d\mu - \int_X(-\min(g, 0))d\mu$.

11. This integration satisfies dominance convergence, monotone convergence, and Fubini's Theorem. For example, Dominance Convergence Theorem can be written as follows:

if $\{f_n\}$ are measurable on a measurable space $(X, \Sigma)$ with a measure $\mu$, $g$ is integrable, $|f_n| \le g$ almost everywhere, $f_n$ converges to $f$ when $n \to \infty$ almost everywhere, then $f$ is integrable, and $\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu$.

12. Let $(X, \Sigma)$ be a measurable space and $\mu$ a measure, the space $L^p_\mu(X)$ is defined as the set of functions $f$ on $X$ such that $|f|^p$ is integrable, quotient

by the relation of "equal almost everywhere". When $1 < p < \infty$, $L^p_\mu(X)$ is a Banach space with norm

$$\|f\|_p = \left(\int_X |f|^p d\mu\right)^{1/p}$$

$L^2_\mu(X)$ is a Hilbert space with inner product

$$(f,g) = \int_X f\overline{g} d\mu$$

Here completeness follows from dominance convergence above.

13. (Lebesgue Decomposition, Radon-Nikodyn Derivative) Let $\lambda$, $\mu$ be two measure on $(X, \Sigma)$, such that $\mu(X) < \infty$, $\lambda(X) < \infty$, then $\lambda$ can be uniquely written as the sum of two measures $\lambda = \lambda_a + \lambda_s$, called the **Lebesgue decomposition**, such that $\lambda_a = h\mu$ for some $h \in L^1_\mu(X)$ (called the **Radon-Nikodyn Derivative**), and there are $A$ and $B$ in $\Sigma$ such that $A \cup B = X$, $\mu(A) = \lambda_s(B) = 0$.
    (a) $\lambda_a$ is called **absolutely continuous** with respect to $\mu$, denoted as $\lambda_a << \mu$.
    (b) $\lambda_s$ is called **mutually singular** to $\mu$, denoted as $\lambda_s \perp \mu$.
    (c) For a proof, consider linear functional on $L^2_{\lambda+\mu}(X)$ defined as $f \mapsto \int_X f d\lambda$, then by Riesz representation of Hilbert space we have $g \in L^2_{\lambda+\mu}(X)$, such that $0 \leq g \leq 1$, and $\int_X f d\lambda = \int_X fg d\lambda + \int_X fg d\mu$. Now $\lambda_s = \lambda|_{\{x:g(x)=1\}}$, $\lambda_a = \lambda|_{\{x:g(x)<1\}}$, $h = \begin{cases} \sum_{n=1}^\infty g^n & g < 1 \\ 0 & g = 1 \end{cases}$ which is in $L^1_\mu$ by uniform convergence.
    (d) The theorem can be adapted to the case where $\lambda$ and $\mu$ are only $\sigma$-finite.

14. As an example of a pair of mutually singular measures, consider the injection from $[0,1]$ to $[0,1]$ defined by taking the base 2 expansion (disallowing infinitely many consecutive 1s), replace 1 with 2 and use it as the base 3 expansion. Then the pushforward Lebesgue measure is mutually singular with the Lebesgue measure, and is also **non atomic**, i.e. does not contain **atoms**, which are subsets with positive measure whose any measuable subset has either the same measure or measure 0.

15. If a $\sigma$-finite Borel measure $\mu$ satisfies
    (a) $\mu(K) < \infty$ when $K$ is compact
    (b) If $E$ is Borel then $\mu(E) = \inf_{U \text{ open }, E \subseteq U} \mu(U)$ (outer regular).
    (c) If $U$ is open then $\mu(U) = \sup_{K \text{ compact }, K \subseteq U} \mu(K)$ (inner regular).

then its completion is called a **Radon** measure.

16. The key property of Radon measure is the following **Riesz Representation Theorem**:

    If $X$ is locally compact and Hausdorff, $C_c(X)$ the space of real valued continuous functions with compact support, $L : C_c(X) \to \mathbb{R}$ a linear map that sends non negative functions to non negative numbers (called a **positive linear functional**), then there is a unique Radon measure $\mu$ on $X$ such that $\int_X f d\mu = L(f)$ for any $f \in C_c(X)$.

    For the proof, define $\mu$ via the regularity, and show that it is well defined.

17. An application of the **Riesz Representation Theorem** above is the following **Choquet's Theorem**:

    Let $V$ be a topological vector space, such that for any $x \in V$ there is some $l \in V^*$ such that $l(x) \neq 0$. Let $K$ be a metrizable compact subset. By **extremal point** we mean $x \in K$ such that there isn't $t \in (0,1)$, $y, z \in K$, such that $x = ty + (1-t)y$. Then for any $q \in K$, there is a Borel probability measure $\mu$ on the set of extremal points $E(K)$ of $K$ such that for any affine continuous function $f$ on $V$, $f(q) = \int_{E(K)} f d\mu$. (We can also write this as $q = \int_{E(K)} x d\mu(x)$.)

    The proof goes as follows:

    (a) Find a countable sequence of affine continuous functions $\{a_n\}$, such that $|a_n|_K| \leq 1$, and for any $x, y \in K$, $x \neq y$, there is some $n$ such that $a_n(x) \neq a_n(y)$. Let $a = \sum_n 2^{-n} a_n^2$. Then $a$ is strictly convex on $K$.

    (b) For any $g \in C(K)$, define the **envelope** $\overline{g}$ as

    $$\overline{g}(x) = \inf_{l \text{ affine continuous, } l|_K \geq g|_K} l(x)$$

    Apply Hahn-Banach to $V = C(K)$, $V' = \{l + ta : l \text{ affine continuous}, t \in \mathbb{R}\}$, $p(g) = \overline{g}(q)$, $f'(l + ta) = l(q) + t\overline{a}(q)$, we get $f \in (C(K))^*$.

    (c) Now apply Riesz Representation Theorem above to $f$ we get the desired $\mu'$ as a Borel measure on $K$. By calculating $\int_K a d\mu'$, we see that its support is a subset of $E(K)$. Restrict it to $E(K)$ we get the desired $\mu$.

18. (a) Fourier series for (complex valued) periodic functions on $\mathbb{R}$:
    $f \in L^2(\mathbb{R}/\mathbb{Z})$, then

    $$f(x) = \sum_{n \in \mathbb{Z}} \hat{f}(n) e^{2\pi n x \sqrt{-1}}$$

    (b) Fourier transform for (complex valued) $L^2$ functions on $\mathbb{R}$:

    $$f(x) = \int_{\mathbb{R}} \hat{f}(y) e^{2\pi x y \sqrt{-1}} dy$$

# 3 Ergodic Theory for Maps

- Dynamical system: The study of asymptotic behavior for iterated self maps, flows or infinite group actions.
- Ergodic Theory: Dynamics via measure theory.

For this section we consider the dynamics of a self map $T : X \to X$. We define $T^{n+1}(x) = T(T^n(x))$, and the sequence $\{x, T(x), T^2(x), \dots\}$ is called the **forward orbit** of $x \in X$.

## 3.1 Invariant Measures

**Definition 3.1.** Let $(X, \mathcal{M}, \mu)$ be a probability measure space. A measurable map $T : X \to X$ is called **measure preserving** if for any $A \in \mathcal{M}$, $\mu(T^{-1}(A)) = \mu(A)$. If $T$ is measure preserving we also call $\mu$ an **invariant probability measure** of $T$.

**Example 3.2.** $X = \mathbb{R}/\mathbb{Z}$. The Lebesgue measure is invariant under:

1. **Circle Rotation**: $x \mapsto x + \theta$.
2. **Circle Doubling**: $x \mapsto 2x$.

**Theorem 3.3.** Let $X$ be a compact metric space and $T$ a continuous map, then there is a Radon probability measure on $X$ invariant under $T$.

*Proof of Theorem 3.3.* Let $x \in X$, define $\mu_n$ as

$$\int_X f d\mu_n = \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(x))$$

for any $f \in C(X)$. $C(X)$ is separable by Stone-Weierstrass, hence by Arzela-Ascoli, there is a subsequence of $\{\mu_n\}$ which converges under weak-$*$ topology (here we identify the set of Radon measures with a subset of $(C(X))^*$ due to Riesz Representation). It is easy to see that this limit $L$ sends non negative functions to non negative numbers, and

$$L(1) = \lim_{j\to\infty}\int_X 1 d\mu_{n_j} = 1$$

hence by Riesz representation theorem there is a Radon probability measure $\mu$ such that $L(v) = \int_X c = v d\mu$ for all $v \in C(X)$.

Now we want to show that $\mu$ is invariant under $T$, i.e. for any $f \in C(X)$, $\int_X f d\mu = \int_X f \circ T d\mu$. For any $\epsilon > 0$, any $v \in C(X)$, there is some $M$, such that if $n_j > M$, then

$$\left|\int_X f d\mu - \int_X f d\mu_{n_j}\right| < \epsilon$$

$$\left|\int_X f\circ T d\mu - \int_X f\circ T d\mu_{n_j}\right| < \epsilon$$

Then

$$\left|\int_X f d\mu - \int_X f\circ T d\mu\right| \leq 2\epsilon + \left|\frac{1}{n_j}\sum_{i=1}^{n_j} f(T^i(x)) - \frac{1}{n_j}\sum_{i=0}^{n_j-1} f(T^i(x))\right|$$

$$\leq 2\epsilon + \frac{2}{n_j}\max_{x\in X}|f(x)|$$

Let $\epsilon \to 0$ and $j \to \infty$ we finished the proof. □

**Remark 3.4.** If $T$ preserves probability measure $\mu$, a measurable subset $B$ is called **invariant** under $T$ if $T^{-1}(B) = B$. Clearly $B$ is invariant iff $\chi_B = \chi_B \circ T$, where $\chi_B(x) = \begin{cases} 1 & x \in B \\ 0 & x \notin B \end{cases}$ is the characteristic function. If $T(B) \subseteq B$, then there is some $B'$ differs from $B$ only by a measure zero set and is invariant under $T$.

## 3.2 Ergodic Theorems

For this section, we always assume that $(M, \mathcal{M}, \mu)$ is a measurable space with a probability measure, and $T : X \to X$ is a measure preserving map.

### 3.2.1 Mean Ergodic Theorem

**Proposition 3.5.** Let $H$ be a Hilbert space (vector space with inner product $(\cdot,\cdot)$, complete under the induced metric), $U$ a norm preserving operator (linear map from $H$ to $H$, such that $(U(x), U(y)) = (x, y)$ for all $x, y \in H$), then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} U^j(x) = \pi_{H^U}(x)$$

where $\pi_{H^U}$ is the orthogonal projection to the closed subspace $H^U = \{x \in H : U(x) = x\}$, and the convergence is in the topology defined by the norm of $H$.

The original proof by von Neumann is via spectral theory of normal operators. The following elementary proof (which is basically a rewrite of the original spectral theory proof) is due to Riesz.

*Proof.* Let $V = \{x - U(x) : x \in H\}$. Then for any $y \in V^\perp$,

$$(y - U(y), y - U(y)) = (y, y - U(y)) - (U(y), y - U(y))$$

$$= (y, y) - (U(y), y) = (y - U(y), y) = 0$$

So $V^\perp = H^U$.

$$V^\perp = \{y \in H : (x, y) = (U(x), y) \text{ for all } x \in H\}$$

$$= \{y \in H : (U(x), U(y)) = (U(x), y) \text{ for all } x \in H\}$$
$$= \{y \in H : y = U(y)\} = H^U$$

For any $v = x - U(x) \in V$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} U^j(v) = \lim_{n\to\infty} \frac{1}{n}(x - U^n(x)) = 0 = \pi_{H^U}(v)$$

Because $\frac{1}{n}\sum_{j=0}^{n-1} U^j$ and $\pi_{H^U}$ are all bounded operators of norm no more than 1, the identity $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} U^j(v) = \pi_{H^U}(v)$ holds for $v \in \overline{V}$ as well. It is evident that this identity is also true on $H^U$, hence by linearity, it is true on $\overline{V} + H^U = H$. $\square$

**Theorem 3.6** (von Neumann's Mean Ergodic Theorem)**.** For any $f \in L^2_\mu(X)$, the sequence $\{\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j\}$ converges to some $T$-invariant function $f^* \in L^2_\mu(X)$ in the sense of $L^2_\mu(X)$.

*Proof.* Apply Proposition 3.5 by setting $H = L^2_\mu(X)$ and $U = U_T = (f \mapsto f \circ T)$. $\square$

**Remark 3.7.** By approximating $L^1_\mu(X)$ functions by $L^2_\mu(X)$ functions, one see that in the statement of Theorem 3.6 one can replace all the $L^2_\mu(X)$ with $L^1_\mu(X)$.

**Remark 3.8.** The proof of Proposition 3.5 implies that one can replace, in the statement of the proposition,

$$\frac{1}{n}\sum_{j=0}^{n-1} U^j(x)$$

with

$$\frac{1}{b_n - a_n}\sum_{j=a_n}^{b_n-1} U^j(x)$$

where $\{a_n\}$, $\{b_n\}$ are two sequences in $\mathbb{Z}$ such that $\lim_{n\to\infty} b_n - a_n = \infty$. Theorem 3.6 can be strengthened similarly.

**Corollary 3.9.** [Poincare Recurrence] For any $\mu$-measurable subset $E$ of $X$, $\mu$-almost every point $p$ in $E$ has that $\{n \in \mathbb{N} : T^n(p) \in E\}$ is an infinite set.

*Proof.* Suppose this is not true, then there is some $E' \subseteq E$ with positive measure, and some integer $N$, such that $T^n(E') \cap E = \emptyset$ for all $n > N$. Apply Theorem 3.6 to the characteristic function $\chi_{E'} \in L^2_\mu(X)$, we see that $(\chi_{E'})^* = \pi_{(L^2_\mu(X))^{U_T}}(\chi_{E'})$ vanishes on $E'$ hence is orthogonal to $\chi_{E'}$, which implies that $\chi_{E'}$ lies in the orthogonal complement of $(L^2_\mu(X))^{U_T}$. However $1 \in (L^2_\mu(X))^{U_T}$ and $(\chi_{E'}, 1) = \mu(E') > 0$, a contradiction. $\square$

### 3.2.2 Maximal and Pointwise Ergodic Theorem

**Proposition 3.10.** Let $f \in L^1_\mu(X)$, define $f_0 = 0$, $f_n = \sum_{j=0}^{n-1} f \circ T^j$, $F_N = \max\{f_n : 0 \le n \le N\}$. Let

$$B = \{x \in X : \sup\{F_N(x) : N \in \mathbb{N}\} = \infty\}$$

$$A_N = \{x \in X : F_N(x) > 0\}, A = \bigcup_N A_N$$

Then

1. $\int_B f d\mu \ge 0$.
2. (Maximal Inequaltiy) $\int_{A_N} f d\mu \ge 0$.
3. $\int_A f d\mu \ge 0$.

*Proof.* It is easy to see that $f_{n+1} = f_n \circ T + f$. Take maximum for $0 \le n \le N$ on both sides, we see that for every $x \in X$, either one of the two cases below is true:

i. $F_{N+1}(x) = F_N(x) = 0$, which implies that $f(x) \le -F_N(T(X)) \le 0$.

ii. $f(x) = F_{n+1}(x) - F_n(T(x))$.

As a consequence, $|F_{N+1} - F_N \circ T| \le |f|$.

1. One can check that $B$ is $T$ invariant, and on $B$ the sequence of functions $F_{N+1} - F_N \circ T$ decreases as $N$ increases, and at every point it will eventually stablized at $f$. Apply dominated convergence we get

$$\int_B f d\mu = \lim_{N\to\infty} \int_B (F_{N+1} - F_N \circ T) d\mu$$

$$\int_B (F_{N+1} - F_N \circ T) d\mu = \int_B F_{N+1} d\mu - \int_B F_N d\mu \ge 0$$

2. It is clear that the elements of $A_N$ must all be in Case ii above. Hence

$$\int_{A_N} f d\mu = \int_{A_N} F_{N+1} d\mu - \int_{A_N} F_N \circ T d\mu \ge \int_{A_N} F_N d\mu - \int_{A_N} F_N \circ T d\mu$$

$$\ge \int_X F_N d\mu - \int_X F_N \circ T d\mu = 0$$

3. This follows immediately from Part 2 above and dominated convergence.

$\square$

**Theorem 3.11** (Maximal Ergodic Theorem)**.** For any $f \in L^1_\mu(X)$, any $a \in \mathbb{R}$, let

$$E_a = \left\{ x \in X : \sup_{n \geq 1} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) > a \right\}$$

Then $a\mu(E_a) \leq \int_{E_a} f d\mu$.

*Proof.* Apply Part 3 of Proposition 3.10 to the function $f - a$. $\square$

**Theorem 3.12** (Birkhoff Ergodic Theorem)**.** For any $f \in L^1_\mu(X)$, there is some $f^* \in L^1_\mu(X)$ which is $T$-invariant, such that for almost every $x \in X$

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

*Proof.* Let $f^*$ be constructed as in Remark 3.7. Suppose there is some $\epsilon > 0$ such that the set $P_\epsilon$ where

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j > f^* + \epsilon$$

has positive measure, then apply Part 1 of Proposition 3.10 to the function $f - f^* - \epsilon/2$, we get that

$$0 < \frac{\epsilon}{2}\mu(P_\epsilon) \leq \frac{\epsilon}{2}\mu(B) \leq \int_B (f - f^*) d\mu = \int_B \left( \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j - f^* \right) d\mu = 0$$

which is an contradiction. Hence we must have

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

The condition for $\liminf$ can be shown similarly. $\square$

### 3.2.3 Ergodicity

**Definition 3.13.** $\mu$ is called an **ergodic measure** of $T$ if the only $T$ invariant measurable function on $X$ are almost everywhere constant. Or, equivalently, any measurable $T$ invariant subset of $X$ has either zero measure or full measure.

**Example 3.14.**

- The Lebesgue measure is ergodic for circle rotation where $\theta$ is irrational, and also for the circle doubling map. This can be shown by Fourier series.
- For rotation where $\theta$ is rational, the Lebesgue measure is not ergodic. For example, the distance function to the finite set $\{n\theta : n \in \mathbb{Z}\}$ is $T$ invariant and not almost everywhere constant.

**Remark 3.15.** If $\mu$ is ergodic, Theorem 3.12 implies that for any measurable subset $A$ of $X$, for almost every $x \in X$,

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

**Example 3.16.** Remark 3.15 and Example 3.14 implies that for almost every real number, there are asymptotically the same number of 0 and 1 in its binary representation.

**Proposition 3.17.** Let $(X, \mathcal{M})$ be a measurable space and $T : X \to X$ a measurable map, $\mu$, $\mu'$ are two distinct ergodic measures of $T$, then $\mu$ and $\mu'$ are mutually singular.

*Proof.* Let $B$ be a measurable set such that $\mu(B) \neq \mu'(B)$. Then by Theorem 3.12, there are $N$, $N'$, $\mu(N) = \mu'(N') = 0$, such that for any $x \in X \backslash N$,

$$\lim_{n\to} \frac{1}{n} \sum_{i=0}^{n-1} \chi_B(T^i(x)) = \mu(B)$$

and for any $x \in X \backslash N'$

$$\lim_{n\to} \frac{1}{n} \sum_{i=0}^{n-1} \chi_B(T^i(x)) = \mu'(B)$$

Hence $N \cup N' = X$. □

**Remark 3.18.**

1. A sqeuence $\{x_n\}$ is said to be **equidistributed** with respect to a Radon measure $\mu$, if $\frac{1}{n}\sum_{i=0}^{n-1} \delta_{x_i}$ converges (in weak-$*$ sense in $(C_c(X))^*$) to $\mu$.
2. We say $x$ is **generic** with respect to $T$ and $\mu$, if the forward orbit $\{T^i(x)\}$ of $x$ equidistributes with respect to $\mu$.
3. When $X$ is a compact metric space, $C(X)$ is separable. Hence as long as $\mu$ is ergodic, by Theorem 3.12, $\mu$-almost every point in $X$ is generic with respect to $T$ and $\mu$.

## 3.3 Ergodic decomposition, Unique Ergodicity

For this section we always assume $X$ is a compact metric space and $T : X \to X$ is a continuous map. It is easy to see that the set of invariant Radon probability measures $\mathcal{D}$ is a compact metrizable convex subset of $(C(X))^*$.

**Lemma 3.19.** Under the notations above, $\mu$ is an extremal point of $\mathcal{D}$ iff it is ergodic under $T$.

*Proof.* 1. If $\mu$ is not ergodic, there is some invariant subset $A$ such that $\mu(A) \in (0,1)$, hence

$$\mu = \mu(A) \cdot \frac{1}{\mu(A)}\mu|_A + \mu(X\backslash A) \cdot \frac{1}{\mu(X\backslash A)}\mu|_{X\backslash A}$$

where $\mu|_A(B) = \mu(A \cap B)$.

2. If $\mu = t\mu' + (1-t)\mu''$, $\mu$, $\mu'$ and $\mu''$ all in $D$, $t \in (0,1)$ and $\mu \neq \mu'$, then let $f \in C(X)$ satisfies $\int_X f d\mu \neq \int_X f d\mu'$. Suppose further that $\mu$ is ergodic, then by Theorem 3.12, for $\mu$-a.e. $x \in X$,

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

Because a set $N$ have $\mu(N) = 0$ implies that $\mu'(N) = 0$, the above limit holds for $\mu'$ a.e. $x \in X$ as well. Integrate both sides by $\int_X \cdot d\mu'$ and use dominated convergence we get

$$\int_X f d\mu' = \int_X f d\mu$$

a contradiction.

□

Hence, by Choquet's Theorem we have

**Theorem 3.20** (Ergodic decomposition)**.** Let $\mu$ be an invariant probability Radon measure on $X$, then there is a Radon probability measure $v$ on the set $\mathcal{E}$ of ergodic Radon probability measures on $X$ (with weak-$*$ topology) such that $\mu = \int_{\mathcal{E}} x dv(x)$. This is called the **ergodic decomposition** of $\mu$.

□

For Choquet's Theorem, the measure on extremal sets is not necessarily unique, but because distinct ergodic measures are mutually singular, the $v$ above is unique.

### 3.3.1 Properties of unique ergodicity

**Definition 3.21.** $T$ is called **uniquely ergodic** if there is only a single Radon invariant probability measure on $X$.

In other words, the ergodic Radon probability measure on $X$ is unique.

**Theorem 3.22.** Let $T : X \to X$ be continuous and $\mu_0$ a Radon invariant probability measure on $X$. The followings are equivalent:

1. $T$ is uniquely ergodic.
2. For any $f \in C(X)$, $\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i$ converges uniformly to a constant function $\int_X f d\mu_0$.

3. There is a dense subset $Y$ of $C(X)$, such that for any $f \in Y$, $\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i$ converges pointwise to a constant function $C_f$.

4. There is a dense subset $Y$ of $C(X)$, such that for any $f \in Y$, let $f_n = \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i$,

$$\lim_{n\to\infty}\left(\sup_{x\in X} f_n(x) - \inf_{x\in X} f_n(x)\right) = 0$$

*Proof.*

- 1 $\implies$ 2: Suppose 2 is not true, then there is $\epsilon > 0$, for any natural number $j$, there is some $n_j > j$, $x_j \in X$ such that $|\frac{1}{n_j}\sum_{i=0}^{n_j-1} f \circ T^i(x_j) - \int_X f d\mu_0| > \epsilon$. Now consider the sequence $\{\frac{1}{n_j}\sum_{i=0}^{n_j-1}\delta_{T^i(x_j)}\}$. It has a convergent subsequence under weak-$*$ topology, and the weak-$*$ limit of this subsequence is an invariant Radon probability measure (see the proof of Theorem 3.3). Because the integrations of $f$ under them are different, it is also distinct from $\mu_0$, which contradicts with 1.

- 2 $\implies$ 3: This is obvious.

- 3 $\implies$ 1: Suppose $\mu$ is a Radon probability measure invariant under $T$. Then by dominant convergence, for any $f \in Y$,

$$\int_X f d\mu = \lim_{n\to\infty}\int_X \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i d\mu = C_f$$

$$\int_X f d\mu_0 = \lim_{n\to\infty}\int_X \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i d\mu_0 = C_f$$

  So $\mu = \mu_0$.

- 2 $\implies$ 4: This is obvious.

- 4 $\implies$ 1: Suppose $\mu$ is a Radon probability measure invariant under $T$. Then for any $f \in Y$,

$$\left|\int_X f d\mu - \int_Z f d\mu_0\right| = \left|\int_X f_n d\mu - \int_X f_n d\mu_0\right| \le |\sup f_n - \inf f_n|$$

  Now let $n \to \infty$ we see that $\mu = \mu_0$.

□

**Example 3.23.** By calculation on functions of the form $e^{2\pi nx}$, we see that irrational circle rotation is uniquely ergodic.

**Example 3.24.** The circle doubling map is not uniquely ergodic. For example, $\delta_0$ is an invariant measure different from the Lebesgue measure.

**Example 3.25.** Let $X$ be a compact metric space, $T : X \to X$ an isometry such that there is some $x_0 \in X$ whose forward orbit $\{T^n(x_0)\}$ is dense in $X$. We now show that $T$ is uniquely ergodic. By compactness, for any $f \in C(X)$, any $\epsilon > 0$, there is $\delta > 0$ such that $x, y \in X$, $d(x, y) \leq \delta$ then $|f(x) - f(y)| < \epsilon$. Pick $m$ such that any $y \in X$ is no more than $\delta$ distance away from some $T^{j_y}(x_0)$ where $0 \leq j_y \leq m - 1$. So for any $x \in X$,

$$\left|\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) - \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x_0))\right| = \frac{1}{n}\sum_{i=0}^{n-1} \left|f(T^i(x)) - f(T^{i+j_x}(x_0))\right|$$

$$+ \left|\frac{1}{n}\sum_{i=0}^{j_x-1} f(T^i(x_0))\right| + \left|\frac{1}{n}\sum_{i=n}^{n+j_x-1} f(T^i(x_0))\right| \leq \epsilon + \frac{2m}{n}\sup_X |f_n|$$

So as $n \to \infty$, $|\sup f_n - \inf f_n| = 0$.

As a consequence, the rotation $x \mapsto x + a$ on $\mathbb{R}^d/\mathbb{Z}^d$ is uniquely ergodic if the orbit of 0 is dense. And when this map is uniquely ergodic, for any ball $B$ in $\mathbb{R}^d/\mathbb{Z}^d$ (under the Euclidean metric), any $x \in X = \mathbb{R}^d/\mathbb{Z}^d$,

$$\lim_{n\to\infty} \frac{1}{n}|\{k \in \{0, \dots, n-1\} : T^k \in B\}|$$

equals the Euclidean volume of $B$.

**Remark 3.26.** Theorem 3.22 implies that

1. If $T$ is uniquely ergodic, every point in $X$ is generic with respect to $T$ and the invariant Radon probability measure $\mu$.

2. If there is a Radon invariant probability measure $\mu$ such that every point in $X$ is generic with respect to $T$ and $\mu$, then $T$ is uniquely ergodic.

### 3.3.2 Furstenberg's Skew Product Theorem

By a **topological group** we mean a group $G$ which is also a topological space, where group multiplication and taking inverses are both continuous. A Radon probability measure $\mu$ is called **invariant** on a topological group, or a **Haar measure**, if it is invariant under multiplication by any element from the left or from the right.

**Theorem 3.27** (Furstenberg's Skew Product)**.** Let $G$ be a compact metrizable topological group, $\mu$ a Haar measure on $G$. Let $X$ be a compact metric space, $T : X \to X$ a continuous map which is uniquely ergodic, and the unique invariant Radon probability measure on $X$ is $m$. Let $c : X \to G$ be continuous. Consider the map

$$S : X \times G \to X \times G, (x, g) \mapsto (T(x), c(x)g)$$

Then clearly $m \times \mu$ is invariant under $S$. If furthermore $S$ is ergodic under $m \times \mu$, then $S$ is uniquely ergodic.

*Proof.* The fact that $m \times \mu$ is invariant can be shown by Fubini or by looking at products of open sets.

Let $R \subseteq X \times G$ be the subset consisting of points that are generic with respect to $S$ and $m \times \mu$. If $(x, g) \in R$, then for any $h \in G$, $(x, gh) \in R$ as well, because for any $f \in C(X \times G)$,

$$\frac{1}{n}\sum_{j=0}^{n-1} f(S^j(x, gh)) = \frac{1}{n}\sum_{j=0}^{n-1} f_h(S^j(x, g))$$

where $f_h(x, g) = f(x, gh)$. Hence $R$ is of the form $Y \times G$ where $Y \subseteq X$. By Remark 3.18, $m(Y) = 1$.

Let $v$ be another $S$-invariant measure on $X \times G$, the projection of $v$ onto $X$ is invariant under $T$ hence equals $m$, which means $v(R) = 1$ as well. Use Remark 3.18 again we see that there must be some element of $R$ which is generic with respect to $S$ and $v$, hence $v = m \times \mu$. □

As an application, we have:

**Corollary 3.28.** Let $f$ be a real coefficient polynomial with leading coefficient irrational and degree $d \geq 1$. The sequence $f(n) - \lfloor f(n) \rfloor$ equidistributes on $[0, 1]$ with respect to the Lebesgue measure.

*Proof.* Consider the map

$$S_k : \mathbb{R}^k/\mathbb{Z}^k \to \mathbb{R}^k/\mathbb{Z}^k$$

$$(x_1, \ldots x_k) \mapsto (x_1 + a, x_2 + x_1, \ldots, x_k + x_{k-1})$$

Where $a$ is irrational. By Fourier series, they are all ergodic with respect to the Lebesgue measure. By Example 3.23, $S_1$ is uniquely ergodic, and by Theorem 3.27, if $S_j$ uniquely ergodic then so is $S_{j+1}$. Hence $S_n$ is uniquely ergodic for all $n$, and by Remark 3.26, every point in $\mathbb{R}^n/\mathbb{Z}^n$ is generic with respect to $S_k$ and the Lebesgue measure. The corollary follows. □

**Example 3.29.** For example, to show that $\{bn^2\}$ equidistributes on $\mathbb{R}/\mathbb{Z}$ to the Lebesgue measure, we show, via Theorem 3.27, that $(x, y) \mapsto (x + 2b, x + y)$ is uniquely ergodic, and consider the orbit of $(b, 0)$.

## 3.4 Mixing

The circle doubling map is not uniquely ergodic, but it satisfies another property called **mixing**:

**Definition 3.30.** Let $\mu$ be an invariant probability measure under $T$.

1. We say $T$ is **weakly mixing** if for any measurable sets $A$ and $B$,

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

2. We say $T$ is **(strongly) mixing**, if

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

### 3.4.1 $L^2$-characterization

We can characterize ergodicity and mixing with $L^2_\mu(X)$ as follows:

**Theorem 3.31.** Let $\mu$ be an invariant probability measure under $T : X|rightarrowX$. Then

1. $T$ is mixing iff for any $f, g \in L^2_\mu(X)$,

$$\lim_{n\to\infty} \left| \int_X (f\circ T^n) g d\mu - \int_X f d\mu \int_X g d\mu \right| = 0$$

2. $T$ is **weakly mixing** iff for any $f, g \in L^2_\mu(X)$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \left| \int_X (f\circ T^j) g d\mu - \int_X f d\mu \int_X g d\mu \right| = 0$$

3. $T$ is ergodic iff for any $f, g \in L^2_\mu(X)$,

$$\lim_{n\to\infty} \left| \int_X \frac{1}{n}\sum_{j=0}^{n-1} (f\circ T^j) g d\mu - \int_X f d\mu \int_X g d\mu \right| = 0$$

*Proof.* For Part 1 and Part 2, the "if" direction can be shown by letting $f = \chi_A$, $g = \chi_B$. For Part 3, the "if" direction can be shown by letting $f = g = \chi_A$ for any invariant measurable subset $A$.

Now we show the "only if" direction. For Part 3, this is due to mean ergodic theorem. For Part 1 and Part 2, one see that the left hand side of the equation before taking limit, when seen as a real valued function $\Phi$ with parameters $f$ and $g$, is uniformly Lipschitz hence equicontinuous on the subset where the $L^2_\mu(X)$ norms of $f$ and $g$ are both bounded, and have

$$\Phi(cf, g) = \Phi(f, cg) = |c|\Phi(f, g) \text{ for all } c \in \mathbb{R}$$

$$\Phi(f + f', g) \leq \Phi(f, g) + \Phi(f', g)$$

$$\Phi(f, g + g') \leq \Phi(f, g) + \Phi(f, g')$$

$$\Phi(f, g) \geq 0$$

Hence, it is easy to verify that if $\mathcal{B}$ is a subset of $L^2_\mu(X)$, where the set of finite linear combinations of elements of $\mathcal{B}$ is dense in $L^2_\mu(X)$, then to verify these limits, we only need to verify them for a pair of elements of $\mathcal{B}$. Now let $\mathcal{B}$ be the set of characteristic functions and we finished the proof. $\square$

**Remark 3.32.** The proof of Theorem 3.31 above implies that to check mixing, weakly mixing or ergodic we only need to check that for a subset $\mathcal{B} \subseteq L^2_\mu(X)$ which spans a dense subset of $L^2_\mu(X)$.

**Remark 3.33.** It is evident that mixing implies weakly mixing, which implies ergodicity.

**Example 3.34.**

1. The irrational rotation is ergodic but not weakly mixing. To show this, consider $f = g = \sin(2\pi x)$, then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1}\left|\int_0^1 f\circ T^j g dx - \int_0^1 f dx \int_0^1 g dx\right| = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\frac{|\cos(ja)|}{2}$$

$$= \int_0^1 \frac{|\cos(x)|}{2} dx \neq 0$$

The last equal sign is because 0 is generic with respect to irrational rotation and the Lebesgue measure, see Remark 3.26.

2. The circle doubling is mixing which can be shown by Fourier series.

3. Consider $X = \mathbb{R}^2/\mathbb{Z}^2$ with the Lebesgue measure, then $T : (x, y) \mapsto (y, x + y)$ is also mixing, which can be shown in Fourier series.

By Remark 3.32, to check the second and third example above, we only need to check them for a pair of functions in $\{\sin(2\pi inx), \cos(2\pi inx)\}$ and $\{\sin(2\pi i(mx+ny)), \cos(2\pi i(mx + ny))\}$, respectively.

**Proposition 3.35.** Let $\mu$ be an invariant probability measure under $T : X \to X$. Then $T$ is weakly mixing iff $T \times T : X \times X \to X \times X$ is weakly mixing with respect to $\mu \times \mu$. Here $(T \times T)(a, b) = (T(a), T(b))$.

*Proof.* To show the "if" part, consider sets of the form $A \times X$ and $B \times X$. To show the "only if" part, recall that by Remark 3.32, $T \times T$ is weakly mixing iff for any measurable subsets $A, A', B, B'$,

$$0 = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\left|\int_X (\chi_{A\times B}\circ(T\times T)^j)\chi_{A'\times B'}d\mu - \int_X \chi_{A\times B}d\mu\int_X \chi_{A'\times B'}d\mu\right|$$

Now

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\left|\int_X (\chi_{A\times B}\circ(T\times T)^j)\chi_{A'\times B'}d\mu - \int_X \chi_{A\times B}d\mu\int_X \chi_{A'\times B'}d\mu\right|$$

$$= \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\left|\mu(T^{-j}(A)\cap A')\mu(T^{-j}(B)\cap B') - \mu(A)\mu(B)\mu(A')\mu(B')\right|$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \big| (\mu(T^{-j}(A) \cap A') - \mu(A)\mu(A'))\mu(T^{-j}(B) \cap B') \\
&\qquad + \mu(A)\mu(A')(\mu(T^{-j}(B) \cap B') - \mu(B)\mu(B'))\big| \\
&\le \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left|\mu(T^{-j}(A) \cap A') - \mu(A)\mu(A')\right| \mu(T^{-j}(B) \cap B') \\
&+ \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(A)\mu(A') \left|\mu(T^{-j}(B) \cap B') - \mu(B)\mu(B')\right| = 0
\end{aligned}
$$

□

**Theorem 3.36.** Let $\mu$ be an invariant probability measure under $T : X \to X$. Then the followings are equivalent:

1. $T$ is weakly mixing.
2. For any $S : Y \to Y$, $\mu'$ an ergodic probability measure on $Y$, the map $T \times S : X \times Y \to X \times Y$ is ergodic with respect to $\mu \times \mu'$.
3. $T \times T : X \times X \to X \times X$ is ergodic with respect to $\mu \times \mu$.

*Proof.* $2 \implies 3$ is obvious: suppose 2 is true, let $Y$ be a single point and $S$ be the identity map, we know that $\mu$ is an ergodic measure. Now apply 2 for $(Y, S, \mu') = (X, T, \mu)$.

To show $1 \implies 2$, by remark 3.32 and a calculation similar to the proof of Proposition 3.35, we only need to show that for any measurable $A, A' \subseteq X$, $B, B' \subseteq Y$,

$$
0 = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap A')\mu'(S^{-j}(B) \cap B') - \mu(A)\mu(A')\mu'(B)\mu'(B')
$$

However,

$$
\begin{aligned}
&\frac{1}{n} \sum_{j=0}^{n-1} \big(\mu(T^{-j}(A) \cap A')\mu'(S^{-j}(B) \cap B') - \mu(A)\mu(A')\mu'(B)\mu'(B')\big) \\
&\qquad = \frac{1}{n} \sum_{j=0}^{n-1} \big((\mu(T^{-j}(A) \cap A') - \mu(A)\mu(A'))\mu'(S^{-j}(B) \cap B')\big) \\
&\qquad\quad + \frac{\mu(A)\mu(A')}{n} \sum_{j=0}^{n-1} \big(\mu'(S^{-j}(B) \cap B') - \mu'(B)\mu'(B')\big)
\end{aligned}
$$

and as $n \to \infty$, the first part goes to 0 due to $T$ being weakly mixing, while the second goes to 0 due to $S$ being ergodic.

Now we show $3 \implies 1$. Firstly by considering subsets of the form $A \times X$, $B \times X$, we see that $T$ must be ergodic. Now for any measurable $A$, $B$,

$$\left|\mu(T^{-j}(A) \cap B) - \mu(A)\mu(B)\right| = \sqrt{(\mu(T^{-j}(A) \cap B) - \mu(A)\mu(B))^2}$$

$$= \sqrt{((\mu(T^{-j}(A) \cap B))^2 - (\mu(A))^2(\mu(B))^2) + 2\mu(A)\mu(B)\left(\mu(A)\mu(B) - \mu(T^{-j}(A) \cap B)\right)}$$

So by Cauchy-Schwartz,

$$\left(\frac{1}{n}\sum_{j=0}^{n-1}\left|\mu(T^{-j}(A) \cap B) - \mu(A)\mu(B)\right|\right)^2$$

$$\leq \frac{1}{n}\sum_{j=0}^{n-1}\left((\mu(T^{-j}(A) \cap B))^2 - (\mu(A))^2(\mu(B))^2\right)$$

$$+\frac{2\mu(A)\mu(B)}{n}\sum_{j=0}^{n-1}\left(\mu(A)\mu(B) - \mu(T^{-j}(A) \cap B)\right)$$

When $n \to \infty$, the first sum converges to 0 because $T \times T$ is ergodic, while the second converges to 0 because $T$ itself is ergodic. □

### 3.4.2 Markov Shifts

The mixing of circle rotation can be shown without the use of Fourier series, and this idea can be extended to a large family of maps, the **Markov shifts**.

**Definition 3.37.** Let

$$\Gamma = \{1, \ldots, n\}^{\mathbb{N}} = \{(a_0, a_1, \ldots) : a_i \in \{1, \ldots, n\}\}$$

$$\Gamma' = \{1, \ldots, n\}^{\mathbb{Z}} = \{(\ldots, a_{-1}, a_0, a_1, \ldots) : a_i \in \{1, \ldots, n\}\}$$

Both are given a topology as the product of discrete topologies. Let $\sigma : \Gamma \to \Gamma$ be defined as $\sigma((a_i)) = (b_i)$, $b_i = a_{i+1}$, and $\sigma' : \Gamma' \to \Gamma'$ is defined analogously. $\sigma$ and $\sigma'$ are called **shift maps**. Then the dynamical system $(\Gamma, \sigma)$ is called a **one-sided full shift** of $n$ letters, and $(\Gamma', \sigma')$ is called a **two-sided full shift** of $n$ letters.

**Remark 3.38.** The topologies of $\Gamma$ and $\Gamma'$ can both be metrized with metric $d((a_i), (b_i)) = e^{-t}$ where $t = \max\{a \in \mathbb{N} : a_j = b_j \text{ for all } |j| < a\}$. It is clear that under this metric they are both compact, and are homeomorphic to the Cantor set. Open balls of radius $e^{-m}$, which are of the form $\{(b_n) : b_i = a_i \text{ for all } |i| \leq m\}$, form a basis of this topology, and are called **cylinder sets**.

For now we will focus on the case of the one sided shifts. The case for two sided shifts are analogous.

**Example 3.39.** Let $\mu_0$ be a probability measure on $\{1, 2, \dots, n\}$ with the $\sigma$ algebra being the power set. Then we can define a Radon probability measure $\mu$ on $\Gamma$ as the product measure

$$\mu(\{(b_n) \in \Gamma : b_i = a_i \text{ for all } 0 \leq i \leq m\}) = \prod_{i=0}^{m} \mu_0(\{a_i\})$$

Then $\mu$ is an invariant measure under $\sigma$, and by setting the $\mathcal{B}$ in Remark 3.32 as the characteristic functions of the cylinder sets, we see that $\sigma : \Gamma \to \Gamma$ is mixing with respect to this measure.

**Definition 3.40.** A $\sigma$-invariant subset of $\Gamma$ or $\Gamma'$ as in Definition 3.37 is called a **subshift**. If $A = [a_{ij}]$ is a $n \times n$ matrix with all entries being 0 or 1, a subshift is called a **Markov shift with transition matrix** $A$, if it equals

$$\{(b_n) : a_{b_i b_{i+1}} = 1 \text{ for all } i\}$$

**Example 3.41.** Let $\Sigma \subseteq \Gamma$ be a subshift with transition matrix $A$. Let $A' = [r_{ij}]$ be a non negative matrix, where $r_{ij} \neq 0$ implies $a_{ij} \neq 0$, and $\sum_j r_{ij} = 1$ for all $i$. Then let $[y_1, \dots, y_n]$ be a non-negative left eigenvector of $A'$ with eigenvalue 1, such that $\sum_i y_i = 1$. Define $\mu$ on $\Sigma$ as

$$\mu(\{(b_n) \in \Sigma : b_i = a_i \text{ for all } 0 \leq i \leq m\}) = y_{a_0} \prod_{i=0}^{m-1} r_{a_i a_{i+1}}$$

Then $\mu$ is $\sigma$-invariant because

$$\mu(\sigma^{-1}(\{(b_n) \in \Sigma : b_i = a_i \text{ for all } 0 \leq i \leq m\}))$$

$$= \sum_{j \in \{1, \dots, n\}, a_{ja_0} = 1} \mu(\{(b_n) \in \Sigma : b_0 = j, b_{i+1} = a_i \text{ for all } 0 \leq i \leq m\})$$

$$= \sum_j y_j r_{ja_0} \prod_{i=0}^{m-1} r_{a_i a_{i+1}} = y_{a_0} \prod_{i=0}^{m-1} r_{a_i a_{i+1}}$$

To show mixing, we need the following:

**Theorem 3.42** (Perron-Frobenius)**.** Let $A$ be a $n \times n$ matrix with non negative entries, and there is a natural number $m > 0$ such that $A^m$ has all entries being positive. Then the characteristic polynomial $\det(A - tI)$ has a simple largest positive root $\lambda$, such that all other roots (in $\mathbb{C}$) $z$ satisfies $|z| < \lambda$. And $\lambda$ is also the unique eigenvalue with an eigenvector with all positive entries.

**Remark 3.43.** $\lambda$ is called the **Perron-Frobenius eigenvalue** and the corresponding eigenvector (up to scalar multiplication) the **Perron-Frobenius eigenvector**.

*Proof.* Let $\Delta$ be the simplex $\{[x_i]^T \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i, \sum_i x_i = 1\}$, and let $P([x_i]^T) = \frac{1}{\sum_i x_i}[x_i]$ be the projection from the set of non negative vectors to $\Delta$. Let $T_A : \Delta \to \Delta$ be $T_A(x) = P(Ax)$. Then by assumption, $T_A^n(\Delta)$ is a sequence of nested subsimplices whose diameter decreases exponentially, which means that there is a unique fixed point $x$ of $T_A$. $x$ is a positive eigenvector which is unique up to scalar multiplication, and the corresponding eigenvalue is positive which we denote as $\lambda$.

Now suppose $x'$ is another eigenvector corresponding to another real eigenvalue, or a real vector in the sum of eigenspaces of some eigenvalue and its complex conjugate, then by considering the orbit of $P(x + tx')$ under $T_A$, for some $t << 1$, we see that the other eigenvalues have to have norm strictly smaller than $\lambda$. Lastly, suppose the Jordan block of $\lambda$ has size larger than $1 \times 1$, then, by linear algebra, there is some positive vector $x''$ close to $x$ such that $(\frac{1}{\lambda}A)^n(x'')$ goes to infinity as $n \to \infty$. This is not possible, because the fact that $(\frac{1}{\lambda}A)^n$ is non negative, and $(\frac{1}{\lambda}A)^n x = x$, provided a bound on the norm of $(\frac{1}{\lambda}A)^n$ independent from $n$. $\square$

**Corollary 3.44.** The shift map in Example 3.41 is mixing as long as there is some $m > 0$ such that $(A')^m$ have all entries being positive.

*Proof.* Similar to Example 3.39, we only need to consider cylinder sets. Let $U = \{(b_i) : b_j = c_j \text{ for all } j \leq d\}$, $V = \{(b_i) : b_j = c'_j \text{ for all } j \leq d'\}$, $n >> d'$, then

$$\mu(T^{-n}(U) \cap V) = \sum_{c'_{d'+1},\ldots,c'_{n-1}, r_{c'_j,c'_{j+1}}>0, r_{c'_{n-1},c_0}>0} \left( y_{c'_0} \prod_{k=0}^{n-2} r_{c'_k c'_{k+1}} r_{c'_n c_0} \prod_{k=0}^{d-1} r_{c_k c_{k+1}} \right)$$

$$= \mu(V)\mu(U) \left( \frac{e^T_{c'_{d'}} A'^{n-d'-1} e_{c_0}}{y_{c_0}} \right)$$

where $e_j$ is the $j$-th standard basis vector, i.e. the column vector with the $j$-th entry being 1 and all others being 0. By Theorem 3.42 we have

$$\lim_{n\to\infty} e^T_{c'_{d'}} A'^{n-d'-1} e_{c_0} = y_{c_0}$$

$\square$

**Definition 3.45.** Let $f : X \to X$ and $f' : X' \to X'$ be two maps. We say $h : X \to X'$ is a **conjugation** between the two, if $h \circ f = f' \circ h$.

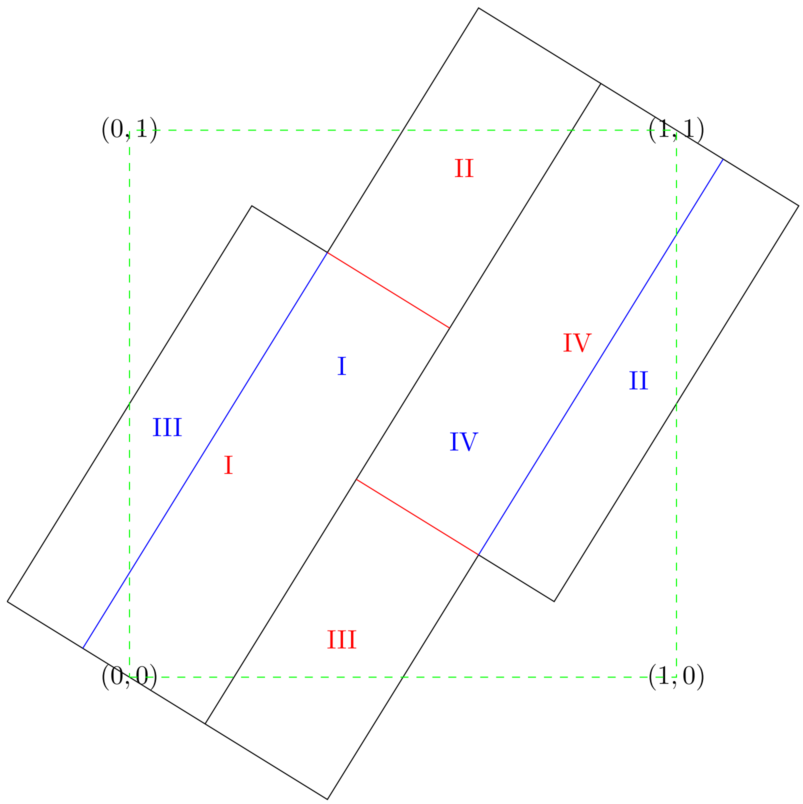
**Proposition 3.46.** Let $f : X \to X$ be a measurable map with invariant probability measure $\mu$, $f' : X' \to X'$ another measurable map, with a measurable conjugation $h : X \to X'$. Then $h_*(\mu)$ is an invariant probability measure on $X'$, and if $\mu$ is mixing, weakly mixing or ergodic then so is $\mu'$.

*Proof.* Let $A$ be a measurable set on $X'$, then $(h_*(\mu))(A) = \mu(h^{-1}(A)) = \mu(f^{-1}h^{-1}(A)) = \mu(h^{-1}f'^{-1}(A)) = (f_*(\mu))(f'^{-1}(A))$. The argument for mixing, weakly mixing and ergodic are similar. $\square$

**Example 3.47.** Let $X$ be the full shift of two letters $\{1,2\}$, $X' = \mathbb{R}/\mathbb{Z}$, $f = \sigma$, $h((a_i)) = \sum_{i=0}^{\infty}(a_i - 1)2^{-i-1}$, then pushforward of the $\mu$ in Example, where 3.39 for $\mu_0$ being uniform measure, is just the Lebesgue measure on $\mathbb{R}/\mathbb{Z}$, and $f'$ is the circle doubling map $x \mapsto 2x$. Hence Example 3.39 and Proposition 3.46 imply that the circle doubling is mixing with respect to the Lebesgue measure.

**Example 3.48.** The map $(x, y) \mapsto (y, x + y)$ on $\mathbb{R}^2/\mathbb{Z}^2$ can be conjugated to a two sided Markov shift.

Consider the following decomposition of a fundamental domain of $\mathbb{R}^2/\mathbb{Z}^2$:




Now $T$ is sending the red blocks continuously and affinely to the blue blocks, and

$$A' = \begin{bmatrix} (\sqrt{5}-1)/2 & (3-\sqrt{5})/2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & (3-\sqrt{5})/2 & (\sqrt{5}-1)/2 \end{bmatrix}$$

Then, other than a measure zero set, there is a measure preserving bijection from $X$ to the two sided Markov shift $\Sigma$ which gives a conjugation between $T$ and $\sigma$, sending every $x \in X$ to $(a_n) \in \Sigma$ such that $T^n(x)$ lies in the $a_n$-th red block.

**Remark 3.49.** Decompositions like in the previous 2 examples that can be used to establish conjugations with one or two sided Markov shifts are called Markov decompositions.

## 3.5 Entropy

Proposition 3.46 can be used to show the non existence of measure preserving conjugations. Another invariant that can be used for this purpose is the concept of **entropy**.

### 3.5.1 Measure Theoretic Entropy

For this subsection we always let $T : X \to X$ be measurable and $\mu$ an invariant probability measure.

**Definition 3.50.**

1. By a **partition** of $X$ we mean a finite set of disjoint measurable subsets whose union is $X$.
2. (Boltzmann) The **entropy** of a partition $\mathbb{P}$ is defined as
$$h(\mathcal{P}) = -\sum_{U \in P} \mu(U) \log(\mu(U))$$
Here we set $0 \log(0) = 0$.
3. Let $\mathcal{P}$ and $\mathcal{P}'$ be two partitions, let partition $\mathcal{P} \vee \mathcal{P}'$ be defined as
$$\mathcal{P} \vee \mathcal{P}' = \{A \cap B : A \in \mathcal{P}, B \in \mathcal{P}'\}$$
4. Let $\mathcal{P}$ be a partition, $T : X \to X$ a measurable map, we define a partition
$$T^{-1}(\mathcal{P}) = \{T^{-1}(U) : U \in \mathcal{P}\}$$
5. The entropy of a map $T$ with respect to invariant measure $\mu$ and partition $\mathcal{P}$ (with finite entropy) is defined as
$$h(T, \mu, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} h\left(\bigvee_{j=0}^{n-1} T^{-j}(\mathcal{P})\right)$$

**Remark 3.51.** It is easy to show, via calculus, that $h(\mathcal{P} \vee \mathcal{P}') \leq h(\mathcal{P}) + h(\mathcal{P}')$, and the maximum is archived when for any $A \in \mathcal{P}$, $A' \in \mathcal{P}'$, $\mu(A \cap A') = \mu(A)\mu(A')$. So $H : n \mapsto h(\bigvee_{j=0}^{n-1} T^{-j}(\mathcal{P}))$ satisfies $H(m+n) \leq H(m) + H(n)$. Because for any $n > 0$, any $m \in \mathbb{N}$, $m = kn + r$ where $k$ is a natural number and $0 \leq r < n$, we have
$$\frac{H(m)}{m} = \frac{H(kn+r)}{kn+r} \leq \frac{kH(n)}{kn+r} + \frac{H(r)}{kn+r}$$

Let $m \to \infty$ we get

$$\limsup_{m\to\infty} \frac{H(m)}{m} \leq \frac{H(n)}{n}$$

Let $n \to \infty$ we get

$$\limsup_{m\to\infty} \frac{H(m)}{m} \leq \liminf_{n\to\infty} \frac{H(n)}{n}$$

Hence, $h(T, \mu, \mathcal{P}) = \lim_{n\to\infty} \frac{H(n)}{n}$ always exists.

**Remark 3.52.** Intuitively, one can think of entropy as measuring "how much information you gained via knowing $x \in X$ belongs to a fixed element of $\mathcal{P}$". Hence the inequaltiy in Remark 3.51 can be thought of as "knowing membership in two partitions gets at most the sum of the information one gets from knowing membership in each".

**Example 3.53.** Consider one sided full shifts (Example 3.39) and Markov shifts (Example 3.41), let the partition $\mathcal{P}$ be the set of cylinder sets of the form $\{(a_n) : a_0 = k\}$, then for full shift,

$$h(\sigma, \mu, \mathcal{P}) = -\sum_i \mu_0(\{i\}) \log(\mu_0(\{i\}))$$

and for Markov shifts,

$$h(\sigma, \mu, \mathcal{P}) = -\sum_i y_i \sum_j r_{ij} \log(r_{ij})$$

**Definition 3.54.** The **measure theoretic entropy** of $T$ with respect to invariant probability measure, is defined as

$$h(T, \mu) = \sup_{\mathcal{P}} h(T, \mu, \mathcal{P})$$

**Remark 3.55.** It is easy to see that if $T : X \to X$ is conjugate to $T' : X' \to X'$ via $h : X \to X'$, then $h(T, \mu) \geq h(T', h_*(\mu))$.

Measure theoretic entropy can be calculated via Kolmogorov-Sinai Theorem below:

**Theorem 3.56** (Kolmogorov-Sinai)**.** We say a partiton $\mathcal{P}$ is a **generator**, if either $\bigcup_n \bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})$, or, when $T$ is invertible, $\bigcup_n \bigvee_{i=-n}^{n} T^i(\mathcal{P})$, generates a $\sigma$ algebra whose completion is the same as the completion of the $\sigma$ algebra of measurable subsets on $X$. If $\mathcal{P}$ is a generator, then $h(T, \mu, \mathcal{P}) = h(T, \mu)$.

*Proof.* It is easy to see, by definition, that

$$h(T, \mu, \mathcal{P}) = h(T, \mu, \bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P}))$$

and, if $T$ is invertible,

$$h(T,\mu,\mathcal{P}) = h(T,\mu,\bigvee_{i=-n}^{n} T^i(\mathcal{P}))$$

Furthermore, if $\mathcal{Q}$ is a partition that is finer than $\mathcal{P}$ (i.e. every element in $\mathcal{Q}$ is a subset of some element of $\mathcal{P}$), then $h(T,\mu,\mathcal{P}) \leq h(T,\mu,\mathcal{Q})$.

Now let $\mathcal{P}'$ be another partition of $X$. By assumption, there is a partition $\mathcal{P}''$, coarser than some $\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})$ or some $\bigvee_{i=-n}^{n} T^i(\mathcal{P})$, with the same cardinality as $\mathcal{P}'$, such that the set difference between each element of $\mathcal{P}''$ and $\mathcal{P}'$ have measures that can be made arbitrarily small, which means $h(\mathcal{P}' \vee \mathcal{P}'') - h(\mathcal{P}'')$, which we call **relative entropy** and denote as $h(\mathcal{P}'|\mathcal{P}'')$, can be made arbitrarily small. Furthermore, by calculation, if $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{Q}_1$, $\mathcal{Q}_2$ are all partitions, then, by a calculation very similar to Remark 3.51,

$$h(\mathcal{Q}_1 \vee \mathcal{Q}_2|\mathcal{P}_1 \vee \mathcal{P}_2) \leq h(\mathcal{Q}_1|\mathcal{P}_1 \vee \mathcal{P}_2) + h(\mathcal{Q}_2|\mathcal{P}_1 \vee \mathcal{P}_2) \leq h(\mathcal{Q}_1|\mathcal{P}_1) + h(\mathcal{Q}_2|\mathcal{P}_2)$$

Hence,

$$h(T,\mu,\mathcal{P}') \leq h(T,\mu,\mathcal{P}' \vee \mathcal{P}'') = \lim_{n\to\infty} \frac{1}{n} h(\bigvee_{j=0}^{n-1} T^{-j}(\mathcal{P}' \vee \mathcal{P}''))$$

which can be made arbitrarily close to

$$\lim_{n\to\infty} \frac{1}{n} h(\bigvee_{j=0}^{n-1} T^{-j}(\mathcal{P}'')) = h(T,\mu,\mathcal{P}'') \leq h(T,\mu,\mathcal{P})$$

□

**Example 3.57.**

1. The partitions in Example 3.53 are all generators.
2. The entropy of irrational cycle rotation with the Lebesgue measure is 0. The entropy of cycle doubling with the Lebesgue measure is $\log(2)$.

**Proposition 3.58.** $h(T^k,\mu) = kh(T,\mu)$

*Proof.* By considering $\bigvee_{j=0}^{k-1} T^{-j}(\mathcal{P})$, we see that $h(T^k,\mu) \geq kh(T,\mu)$. The other direction follows from the fact that $\bigvee_{j=0}^{n-1} T^{-kj}(\mathcal{P})$ is coarser than $\bigvee_{j=0}^{nk-1} T^{-j}(\mathcal{P})$. □

### 3.5.2 Topological Entropy, Variational Principle

For this subsection, we let $X$ be a compact metric space and $T : X \to X$ a continuous map.

**Theorem 3.59.** The following three numbers are equal:

$$h_1(T) = \sup_{\mu \text{ Radon invariant probability measure}} h(T,\mu)$$

$$h_2(T) = \sup_{\mathcal{C} \text{ finite open cover of } X} \limsup_{n\to\infty} \frac{1}{n} \log MCov\left(\left\{\bigcap_{j=0}^{n-1} T^{-j}(U_j) : U_j \in \mathcal{C}\right\}\right)$$

where $MCov(\mathcal{C})$ is the smallest number of elements of $\mathcal{C}$ whose union is $X$.

$$h_3(T) = \lim_{\epsilon\to 0}\left(\limsup_{n\to\infty} \frac{1}{n}\log N(n,\epsilon)\right)$$

where $N(n,\epsilon)$ is the largest number of points in $X$ such that for any two distinct points $x$, $x'$ among them,

$$\max_{j=0,\dots,n-1} d(T^j(x), T^j(x')) \geq \epsilon$$

(called $(n,\epsilon)$ **separability**).

**Remark 3.60.** By the same argument as in Remark 3.51, in the definition of $h_2(T)$ above, the $\limsup_{n\to\infty}$ can be replaced by $\lim_{n\to\infty}$. The $\lim_{\epsilon\to 0}$ is well defined because $N(n,\epsilon)$ is non decreasing with respect to $\epsilon$.

**Remark 3.61.** Proposition 3.58 implies that $h_1(T^k) = kh_1(T)$. An analogous argument shows that $h_2(T^k) = kh_2(T)$.

*Proof.* 1. $h_1 \leq h_2$: For any measurable partition $\mathcal{P} = \{B_1, \dots, B_m\}$, let $V_j \subseteq B_j$ be compact and has measure very close to $B_j$, let $\mathcal{P}' = \{V_j, V_0 = X\backslash \bigcup_j V_j\}$, then $h(\mathcal{P}'|\mathcal{P})$ can be made arbitrarily small, hence by the proof of Theorem 3.56, $h(T,\mu,\mathcal{P})$ can be arbitrarily approximated by $h(T,\mu,\mathcal{P}')$. We observe that $h(\bigvee_{j=0}^{n-1} T^{-j}(\mathcal{P}'))$ is bounded by the number of non empty elements of $\bigvee_{j=0}^{n-1} T^{-j}(\mathcal{P}')$, which is bounded by

$$2^n MCov\left(\left\{\bigcap_{j=0}^{n-1} T^{-j}(V_0 \cup V_{i_j})\right\}\right)$$

hence $h_1(T) \leq \log 2 + h_2(T)$. Now apply Remark 3.61 one gets the conclusion.

2. $h_2 \leq h_1$: Given any finite open cover $\mathcal{C}$, let $\epsilon > 0$ be small enough such that for any $x, x' \in X$, if $d(x,x') \leq \epsilon$ then they both lie in a single element of $\mathcal{C}$. Now any set of $(n,\epsilon)$ separable points gives rise to a finite subset of $\{\bigcap_{j=0}^{n-1} T^{-j}(U_j) : U_j \in \mathcal{C}\}$ of the same or fewer elements which is an open cover of $X$. The inequality follows.

3. $h_3 \leq h_1$: Pick $\epsilon$ small enough, let $v_n$ be the average of the Dirac $\delta$ function on $N(n,\epsilon)$ points that form a $(n,\epsilon)$ separable set, let $\mu_n = \sum_{j=0}^{n-1} T_*^j(\mu_n)$. Let $\mu$ be a weak-$*$ limit of $\mu_n$, then by construction it is an Radon invariant probability measure.

   Let $\mathcal{P}$ be a finite partition whose pieces are all of diameter smaller than $\epsilon$, and all boundaries have zero $\mu$ measure. Let $h(\mathcal{P}, v)$ be the entropy of partition $\mathcal{P}$ under probability measure $v$, then by remark 3.51 and the two observations below:

   (a) $h(T^{-1}(\mathcal{P}), \mu) = h(\mathcal{P}, T_*(\mu))$.
   (b) $t_i \geq 0, \sum_i t_i = 1$, then $h(\mathcal{P}, \sum_i t_i\mu_i) \geq \sum_i t_i h(\mathcal{P}, \mu_i)$.

   we get that if $p << n$,

$$\frac{1}{n}h\left(\bigvee_{j=0}^{n-1} T^{-j}(\mathcal{P}), v_n\right) \leq \frac{1}{p}h\left(\bigvee_{j=0}^{p} T^{-j}(\mathcal{P}), \mu_n\right) + O(p/n)$$

   And the inequality follows.

□

**Definition 3.62.** $h_1(T) = h_2(T) = h_3(T)$ is called the **topological entropy**, denoted as $h_{top}(T)$.

**Example 3.63.** By $h_3$ above, the topological entropy of circle rotation is 0 and of circle doubling is $\log(2)$.

**Example 3.64.** The topological entropy of a one sided full shift with $m$ letters is $\log(m)$.

**Example 3.65.** The topological entropy of a one sided Markov shift $\Sigma_A$, where $A$ satisfies Theorem 3.42, is log of the spectral radius of $A$. An invariant Radon probability measure $\mu$ on $\Sigma_A$ such that $h_{top}(T) = h(T,\mu)$ (called **measure of maximal entropy**) can be written down as follows: let $x = (x_1, \dots, x_m)$ be the non negative eigenvalue corresponding to leading eigenvector $\lambda > 0$, normalized such that $\sum_i x_i = 1$. Then let $\mu$ be constructed as in example 3.41, with $y_i$ defined via $\sum_i \frac{a_{ij}y_ix_j}{x_i} = \lambda y_j$, $r_{ij} = \frac{a_{ij}x_j}{\lambda x_i}$.

### 3.5.3 Dimension Theory, Thermodynamics Formalism

Thermodynamics Formalism (TF) is a way to obtain invariant measures with "good" geometric properties. We will illustrate the idea with an example:

**Example 3.66.** Let $X = \Sigma_A, T = \sigma$ be the one-sided Markov shift, such that $A$ satisfies Theorem 3.42. The measure of maximal entropy in Example 3.65 can also be interpreted as follows:

1. Consider linear map $\mathcal{L} : C(X) \to C(X)$ defined as $f \mapsto (x \mapsto \sum_{y \in T^{-1}(x)} f(y))$, $\mathcal{L}^*$ its dual on $(C(X))^*$.

2. The spectral radius of both $\mathcal{L}$ and $\mathcal{L}^*$ are both $\lambda$ which is the spectral radius of $A$.

3. The leading eigenvector of $\mathcal{L}$ is the locally constant function $h$ sending $(a_0, a_1, \dots) \in X$ to $z_{a_0}$ such that $\lambda z_j = \sum_i a_{ij} z_i$, while the leading eigenvector of $\mathcal{L}^*$ is $\mu/h$, so their product is exactly the measure of maximal entropy $\mu$ as in Example 3.65.

Now let $\psi$ a Hölder continuous function called the **potential**, $\varphi = e^{\psi}$.

1. The **Ruelle-Perron-Frobenius operator** on $C(X)$ is defined as $(\mathcal{L}_\varphi(f))(x) = \sum_{y \in T^{-1}(x)} \varphi(y) f(y)$. $\mathcal{L}_\varphi^*$ is an operator on Radon measures.

2. The *Ruelle-Perron-Frobenius theorem* says that both $\mathcal{L}_\varphi$ and $\mathcal{L}_\varphi^*$ have a unique leading positive eigenvalue and eigenvector, and the leading eigenvalues are the same, denoted as $\lambda > 0$, the eigenvectors are denoted as $h$ and $\mu$ respectively.

3. $h\mu$ is an invariant measure, because

$$\int_X f(T(x))h(x)d\mu(x) = \int_X f(T(x))(\varphi(x)h(x))((\varphi(x))^{-1}d\mu(x))$$

$$= \int_X f(y)\lambda h(y)d(\lambda^{-1}\mu)(y) = \int_X f(y)h(y)d\mu(y)$$

called the **Gibbs measure**. $\log(\lambda)$ is called the **topological pressure**.

Some applications:

1. When $\psi = 0$, or $\varphi = 1$, the Gibbs measure is the measure of maximal entropy and the topological pressure equals the topological entropy.

2. Let $i$ be an injection from $X$ to $\mathbb{R}$, such that disjoint cylinder sets are sent to subsets of disjoint intervals, and $T = \sigma$ is conjugated to some smooth map $S$ via $i$. Let $\varphi_\beta(x) = |S'(i(x))|^{-\beta}$. Then the $\beta$ that makes the topological pressure equals 0 is the **Hausdorff dimension** of $i(X)$.

## 3.6 Some other important topics in ergodic theory

- Subadditative and multiplicative ergodic theorems
- Lyapunov exponents
- Factors and joinings.
- Quantative version of ergodicity and mixing
- ...

# 4 Further topics

## 4.1 Interval Exchange Transformation

**Definition 4.1.** Let $T : X \to X$ be a map, $Y \subseteq X$, the **first return map** is a map $T'$ from $Y$ to itself such that $T'(y) = T^n(y)$ where $n$ is the smallest natural number such that $T^n(y) \in Y$. Let $\mu$ be an invariant probability measure on $X$, then Poincare recurrence (Corollary 3.9) shows that $T'$ is $\mu$-a.e. well defined.

**Definition 4.2.** An **interval exchange transformation** (IET) is defined with the following data: $x = (x_1, \dots, x_n) \in \mathbb{R}^n_{>0}$, such that $\sum_i x_i = 1$; $\sigma \in S_n$. $T_{x,\sigma} : I = [0,1] \to I = [0,1]$ is carried out with the following steps:

1. Divide $I$ into $n$ subintervals $I_1, \dots, I_n$, such that the interval $I_j$ has length $x_j$.
2. Permute the subintervals according to $\sigma$, put them together into $I$ again.

An IET of $n$ intervals is also called an $n$-IET.

**Remark 4.3.**

1. The interval exchange transformation is well defined on all but countable many points on $I$.
2. The Lebesgue measure is an invariant probability measure.
3. When $n = 2$ and $\sigma = (1, 2)$, this reduces to cycle rotation.
4. Let $S$ be a surface formed by gluing parallel sides of a polygon together via translation, then the first return map of the straight line flow on any embedded interval is an IET. Such surfaces corresponds to holomorphic 1-forms on closed Riemann surfaces, hence the dynamics of IETs is deeply connected to the study of Teichmuller dynamics.

The key tool for the study of IET is the following observation, which one can prove by calculation:

**Proposition 4.4.** Let $T$ be an IET of $m$ intervals defined using $x \in \mathbb{R}^m$, $\sigma \in S_m$. Then:

1. The first return map on $[0, 1 - \min\{x_m, x_{\sigma^{-1}(m)}\}]$ is an IET of no more than $m$ intervals.
2. The first return map of any subinterval $I' \subseteq I$ is an IET of no more than $m + 1$ intervals.

□

**Remark 4.5.** The procedure of taking first return map on $[0, 1-\min\{x_m, x_{\sigma^{-1}(m)}\}]$ is called the **Rauzy-Veech induction**. For 2-IETs, Rauzy-Veech induction is related to continued fraction.

**Theorem 4.6** (Katok, 1980)**.** IETs can never be strongly mixing under the Lebesgue measure.

*Proof.* Suppose $T : I \to I$ is an $m$-IET. Let $\Delta$ be a subinterval of $I = [0, 1]$, then the first return map of $T$ on $\Delta$ is an IET of no more than $m + 1$ sub intervals. Denote these subintervals as $I_j$, then $I$ has a partition by $I_j$ and their images under $T$, denoted as $\mathcal{P}_\Delta$.

For each $I_j$, consider first return map on it, one gets another IET, and $I_j$ is then decomposed into no more than $m + 1$ subintervals $I_{jk}$. Let $r_{jk} > 0$ be the smallest positive integer such that $f^{r_{jk}}(I_{jk}) \subseteq I_{jk}$, then, for any element $B$ of the partition $\mathcal{P}_\Delta$, we have that up to a measure 0 set, $B \subseteq \bigcup_{j,k} T^{-r_{jk}}(B)$. Hence for any $A$ consisting of unions of elements of $\mathcal{P}$, there is some $r_{jk}$ such that the Lebesgue measure of $A \cap T^{-r_{jk}}(A)$ is no less than $\frac{1}{(m+1)^2}$ times the measure of $A$.

Now let $A \subseteq I$ be a measurable set with a very small measure, $N > 0$ be any natural number, then either one can pick $\Delta$ very small, such that one can approximate $A$ by unions of elements of $\mathcal{P}_\Delta$, and $r_{jk} > N$ for all $j, k$, or, when such $\Delta$ is impossible, the set of periodic points would have positive measure, hence the Lebesgue measure would not be an ergodic measure. □

Gutkin and Katok (1987) showed some IETs are weakly mixing. Avila and Forni (2007) proved that almost every IET is either a rotation or weakly mixing, by showing that the only measurable eigenfunction of eigenvalue 1 is constant, and then do Rauzy-Veech induction.

## 4.2 Group Actions

**Definition 4.7.** Let $G$ be a locally compact topological group.

1. A Radon measure on $G$ $m$ is called a (left) Haar measure, if for any measurable $A \subseteq G$, $g \in G$, $m(A) = m(gA)$.

2. $G$ is called **amenable**, if for any compact subset $K$, any $\epsilon > 0$, there is some measurable subset $F$ with compact closure, such that

$$m(KF \backslash F) + m(F \backslash KF) < \epsilon m(F)$$

Here $AB = \{ab : a \in A, b \in B\}$.

**Example 4.8.**

1. Let $G$ be a discrete group, then the counting measure is a Haar measure.

2. Let $G$ be $\mathbb{R}$ with the Euclidean topology, then the Lebesgue measure is a Haar measure.

3. $\mathbb{Z}$ as a discrete group is amenable. One can further show that a discretr group is amenable iff all of its finitely generated subgroups are amenable, hence any abelian group as a discrete group is amenable.

4. Free group of 2 generators as a discrete group is not amenable.

**Definition 4.9.** A sequence of measurable subset $\{F_n\}$ is called a **Følner sequence**, if for any compact subset $K$, any $\epsilon > 0$, there is some $N$, such that if $n > N$, then

$$m(KF_n \backslash F_n) + m(F_n \backslash KF_n) < \epsilon m(F_n)$$

**Theorem 4.10.** Let $X$ be a compact metric space with a continuous $G$ action for some amenable group $G$, $\mu$ a Radon measure on $X$ invariant under group action, $\{F_n\}$ a Følner sequence, then in $L^2(X, \mu)$ (or $L^1(X, \mu)$),

$$\lim_{n\to\infty} \frac{1}{\mu(F_n)} \int_{F_n} f(gx)dm(g)$$

exists and is $G$-invariant.

*Proof.* The proof is analogous to the proof of Theorem 3.11. □

**Remark 4.11.** When $G = \mathbb{R}$ with Euclidean topology, we call a continuous $G$ action a **flow**. Theorem 4.10 implies that flows have mean ergodic theorem and Poincare recurrence.

**Theorem 4.12.** Let $G$ be an amenable locally compact group. There is a linear map $A$ from bounded continuous functions to real numbers, such that

1. $A(1) = 1$
2. If $f(x) \geq 0$ for all $x \in G$ then $A(f) \geq 0$
3. If $f'(gx) = f(x)$ for all $x \in G$ and some $g \in G$, then $A(f) = A(f')$.

*Proof.* Let $B(X)$ be the space of bounded continuous functions on $G$, with norm $\|f\| = \sup_{g\in G} |f(g)|$. Let $H$ be a subspace spanned by elements of the form $f(x) - f(gx)$ and 1. Define real valued linear map $A'$ on $H$ such that $A'(f(\cdot) - f(g\cdot)) = 0$.

Now if we can show that the operator norm of $A'$ is 1 (if $\|f\| \leq 1$ then $|A'(f)| \leq 1$), then by Hahn-Banach we can extend it to $B$ which satisfies all the requirements. To show this, we only need to show that there can not be $\epsilon > 0$, $f_j \in B$, $g_j \in G$, such that $\sum_{j=1}^n (f_j(x) - f_j(g_j x)) > \epsilon$. Now pick $\epsilon' > 0$, let $F$ be a subset of $G$ with finite measure, such that $m(F\backslash g_i F) + m(g_i F \backslash F) < \epsilon' m(F)$, where $m$ is the Haar measure. The existence of such $F$ follows from the definition of amenability. Then

$$\epsilon \leq \frac{1}{m(F)} \int_F \sum_j (f_j(x) - f_j(g_j x))dm(x) \leq \epsilon' \sum_j 2\|f_j\|$$

Now make $\epsilon'$ small enough we get a contradiction. □

**Theorem 4.13.** If $G$ is a locally compact amenable group acting continuously on a compact metric space $X$, then there is a Radon invariant probability measure on $X$.

*Proof.* Let $A$ be as in Theorem 4.12, pick $x \in X$, then for any $f \in C(X)$, define $\int_X f d\mu = A(g \mapsto f(gx))$. □

**Remark 4.14.** One can write down a pair of homeomorphisms from $\mathbb{R}/\mathbb{Z}$ to itself whose set of invariant Radon probability measures are disjoint. Hence, by Theorem 4.13, the free group of two generators $F_2$ is non amenable.

**Remark 4.15.** The staring point of the concept of amenability, and ergodic theory on groups, is the Banach-Ruziewicz problem, raised initially by Stanislaw Ruziewicz: if a function on subsets of the sphere is invariant under isometry and finitely additive, does the function have to be a multiple of Lebesgue measure.

Banach provided the answer for the first case, the case of dimension 0, in the negative. To show that, first observe that $\mathbb{R}/\mathbb{Z}$ is abelian, hence as a discrete group it is amenable. This discrete group acts on $\mathbb{R}/\mathbb{Z}$ by addition. Pick $V$ as the intersection of countably many dense open subsets of $\mathbb{R}/\mathbb{Z}$ (under the Euclidean topology). Let $B$ be the space of bounded functions on $\mathbb{R}/\mathbb{Z}$, with the sup norm. Let $H$ be the subspace generated by $f(\cdot) - f(g\cdot)$, 1 and $\chi_V$, and define $A' : H \to \mathbb{R}$ as a linear map such that $A'(f(\cdot) - f(g\cdot)) = 0$, $A'(1) = A'(\chi_V) = 1$, then by the same argument as in the proof of Theorem 4.12, and the Baire Category Theorem, $A'$ has norm 1 hence can be extended to a function on $B$ with norm 1. Now $Y \mapsto A(\chi_Y)$ is the desired function.

**Remark 4.16.** In the remark above, by unique ergodicity of rotation, for any continuous function $f$ on $X$, $A(f)$ equals the integral of $f$ with respect to the Lebesgue measure. However $\chi_Y$ is not continuous.

**Theorem 4.17** (Lindenstrauss, 2001)**.** The convergence in Theorem 4.10 is a.e. pointwise if $f \in L^1(X, \mu)$ and $\{F_n\}$ satisfies the condition that there is $C > 0$, such that for all $n \in \mathbb{N}$, $m(\bigcup_{k<n} F_k^{-1} F_n) \leq Cm(F_n)$.

The idea of the proof is as follows:

- The Birkhoff ergodic theorem can be deduced from maximal ergodic theorem.
- Maximal ergodic theorem can be reduced, via integration, to maximal ergodic theorem on a $G$ orbit.
- Maximal ergodic theorem on single $G$ orbit can be reduced to Vitali covering lemma when $G = \mathbb{Z}$ or $\mathbb{R}$.
- For more general amenable groups like in Lindenstrauss's theorem, consider "probabilistic covering lemma" as a replacement.

## 4.3 Hyperbolic Plane

Let $H$ be the upper half plane $\{z = x + yi \in \mathbb{C} : y > 0\}$ with the metric $\frac{dx^2+dy^2}{y^2}$. The group $SL(2, \mathbb{R})$ acts on $H$ as isometries by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} z = \frac{az+b}{cz+d}$$

It induces an action on the unit tangent bundle $T_1H$, and via it one can identify $T_1H$ with the group $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\{\pm I\}$.

There are two commuting left actions of $PSL(2,\mathbb{R})$ on $PSL(2,\mathbb{R})$, as $(g,x) \mapsto gx$ and $(g,x) \mapsto xg^{-1}$. We can define a Riemannian metric on $PSL(2,\mathbb{R})$ which is invariant under one action, then the other action preserves the volume. For example, the Riemannian metric on $H$ itself induces a metric which is invariant under the first action, and the corresponding volume form is invariant under the second action.

Now let $\Gamma$ be a discrete subgroup, the set of right $\Gamma$ cosets $\Gamma\backslash PSL(2,\mathbb{R})$ is an orbifold with Riemannian metric induced by the metric described above. We further assume that it has finite volume (such subgroups are called **lattices**). Now a group homomorphism to $PSL(2,\mathbb{R})$ composed with the second action $(g,x) \mapsto xg^{-1}$ gives a group action with a finite invariant measure. This is one of the first example in the field of **homogenuous dynamics**.

**Remark 4.18.** An example that will give finite volume is $\Gamma = PSL(2,\mathbb{Z})$. Now elements of $\Gamma\backslash PSL(2,\mathbb{R})$ would be related to the space of flat tori of area 1 with a direction, and geodesic flow has connection to the dynamics of Rauzy-Veech induction, see Remark 4.5.

**Definition 4.19.** The **geodesic flow** $g_t$ is the $\mathbb{R}$ action defined via the group homomorphism

$$t \mapsto \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$$

The **Horocycle flow** $h_t$ is the $\mathbb{R}$ action defined via the group homomorphism

$$t \mapsto \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

### 4.3.1 Hopf's argument

A fundamental result in the study of this dynamical system is that the geodesic flow is ergodic, which we will show via the **Hopf's argument** below:

**Theorem 4.20.** If $PSL(2,\mathbb{R})$ acts transitively and continuously to some metric space $X$ that has finite invariant probability measure $\mu$, and the induced (right) action on $L^2(X,\mu)$ is continuous, then the map $x \mapsto g_t x$ is ergodic for any $t > 0$.

*Proof.* Let $f \in L^2(X,\mu)$ be invariant under $U : f \mapsto f \circ g_t$. Then, for any $s$, as $n \to \infty$,

$$g_{-nt} \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} g_{nt} \to I$$

$$g_{nt} \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix} g_{-nt} \to I$$

Let $V_s : f(z) \mapsto f\left(\begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} z\right)$, $W_s : f(z) \mapsto f\left(\begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix} z\right)$, then

$$\lim_{n\to\infty} U^{-n} V_s U^n f = f$$

hence

$$\|f\|^2 = (f,f) = \lim_{n\to\infty}(U^{-n}V_sU^nf,f) = \lim_{n\to\infty}(V_sU^nf,U^nf) = \lim_{n\to\infty}(V_sf,f)$$

Because $\|V_sf\| = \|f\|$, $V_sf = f$. Similarly $W_sf = f$. However the group $PSL(2,\mathbb{R})$ is generated by elements of the form

$$\begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix}$$

Hence $f$ is a.e. constant on $X$, i.e. the map $x \mapsto g_tx$ is ergodic. $\square$

**Remark 4.21.**

$$h_{nt}\begin{bmatrix} 1 & 0 \\ \epsilon & 1 \end{bmatrix}g_{mt} = \begin{bmatrix} 1+n\epsilon t & mt+nt+mnt^2\epsilon \\ \epsilon & 1+mt\epsilon \end{bmatrix}$$

Now set $\epsilon$ small and pick $m, n$ appropriately such that $1+n\epsilon t = 2$, $1+m\epsilon t = 1/2$, one see that any $h_t$ invariant function must be invariant under $\begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix}$, hence by Theorem 4.20, it is also a constant. This shows that $x \mapsto h_tx$ is ergodic as well.

### 4.3.2 Mixing and unique ergodicity

**Theorem 4.22.** Let $\Gamma$ be a lattice, $\{a_n\}$ a sequence in $PSL(2,\mathbb{R})$. Let $v \in L^2(\Gamma\backslash PSL(2,\mathbb{R}))$, if $v_0$ is a weak-* limit of $v(a_nx)$, $\lim_{n\to\infty} a_nga_n^{-1} = 1$, then $v_0(x) = v_0(gx)$.

*Proof.* For any $L^2$ function $w$,

$$\begin{aligned}(v_0(g\cdot),w) &= \lim_{n\to\infty}(v(a_ng\cdot),w) = \lim_{n\to\infty}(v(a_nga_n^{-1}\cdot),w(a_n^{-1}\cdot)) \\ &= \lim_{n\to\infty}(v,w(a_n^{-1}\cdot)) = (v_0,w)\end{aligned}$$

$\square$

**Remark 4.23.** By Theorem 4.22 and Remark 4.21, both the geodesic flow and the horocycle flow are mixing.

A consequence of the mixing of geodesic flow is the following:

**Theorem 4.24.** When $\Gamma\backslash PSL(2,\mathbb{R}$ is compact, the horocycle flow is uniquely ergodic.

**Remark 4.25.** This is the first non trivial case of **Ratner's theorem** in homogenuous dynamics.

# 5 Final Review

1. Existence of invariant Radon probability measure.
2. Mean Ergodic theorem
3. Poincare recurrence
4. Maximal identity, Maximal Ergodic Theorem
5. Birkhoff Ergodic Theorem
6. Unique ergodicity and uniform convergence
7. Mixing and weakly mixing
8. Markov shifts
9. Measure theoretic entropy
10. Topological entropy

**Example 5.1.**

1. Apply the mean, maximal and pointwise ergodic theorems to circle doubling map $x \mapsto 2x$ on $\mathbb{R}/\mathbb{Z}$ and $\chi_{[0,1/2]}$, we have:
   - Let $f_n(x)$ be the number of 0s in the first $n$ binary expansions of $x$. Then in $L^2$ sense $\frac{f_n}{n} \to 1/2$.
   - For any $\alpha \in (0,1)$, let $E_\alpha$ be the set consisting of number $x$ in $(0,1)$ where there is some $n \in \mathbb{N}$ such that the first $n$ digits in the binary expansion of $x$ has more than $n\alpha$ zeros. Then $\alpha$ times the Lebesgue measure of $E_\alpha$ is no more than $1/2$.
   - For almost every $x \in (0,1)$, $\lim_{n\to\infty} \frac{f_n(x)}{n} = 1/2$.
2. Unique ergodicity of $x \mapsto x + \log 2/\log 10$ in $\mathbb{R}/\mathbb{Z}$ can be used to prove that
$$\lim_{n\to\infty} \frac{\text{number of } k \in \{0,\dots,n-1\} \text{ such that } 2^k \text{ has first digit } 1}{n} = \frac{\log(2)}{\log(10)}$$
3. Consider the map $T : [0,1] \to [0,1]$ defined as $T(x) = \begin{cases} 3x & x < 1/3 \\ 2-3x & 1/3 \le x < 2/3 \\ 3x-2 & x \ge 2/3 \end{cases}$.

   Then
   - $T$ is 3-Lipschitz, and for any $\epsilon$, $[0,1]$ can be covered by $O(3^n/\epsilon)$ balls of diameter $\frac{\epsilon}{3^n \cdot 2}$, each containing no more than one element in any $(n,\epsilon)$ separated set. Hence $h_{top}(T) \le \log(3)$.

- Consider the Lebesgue measure and the partition

$$\mathcal{P} = \{[0, 1/3), [1/3, 2/3), [2/3, 1]\}$$

we see that $h(T, \mu_{Lebesgue}) \geq \log(3)$.

- Because $h_{top}(T) \geq h(T, \mu_{Lebesgue})$, $h_{top}(T) = h(T, \mu_{Lebesgue}) = \log(3)$

Practice Problems

- Consider the flat torus $\mathbb{R}^2/\mathbb{Z}^2$ as a quotient group, show that any group homomorphism to itself, of the form $(x, y) \mapsto (ax + by, cx + dy)$, which is ergodic with respect to the Lebesgue measure, must also be mixing with respect to the Lebesgue measure.

- Let $\Sigma$ be a subshift of $\{0, 1\}^{\mathbb{Z}}$ consisting of sequences with no more than 2 consecutive 0s. Find a Radon invariant probability measure on $\Sigma$ and calculate the measure theoretic entropy.

- If $X$ is a measurable space with probability measure $\mu$, $T$ a measure preserving map. Let $f$ be a non negative measurable function on $X$ such that $\int_X f d\mu = \infty$. What can you say about the convergence of $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$?

- Show that the set of ergodic measures on one sided full shift of 2 symbols is dense in the set of all invariant measures under the weak $*$ topology.

- Let $X$ be a compact metric space and $f : X \to X$ a continuous map. What is the relationship between the topological entropy of $f$ and of $g : X \times X \to X \times X$ defined as $g(a, b) = (b, f(b))$?

Answer

- By Fourier series, the Lebesgue measure is ergodic iff $|ad - bc| = 1$ and the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has two real and irrational eigenvalues. Mixing follows from the $L^2$ characterization.

- $\Sigma$ is conjugate via homeomorphism to a Markov subshift of $\{0, 0', 1\}^{\mathbb{Z}}$, where the 0s followed by 0 are sent to 0 and the 0s followed by 1 are sent to $0'$. Now one can continue the calculation as in the lectures.

- It converges almost everywhere to either infinity or a measurable function taking value on a measurable subset of $X$.

- Two invariant probability measures are close under weak $*$ topology iff they are close when evaluating at all the $2^N$ cylinder sets where $N$ consecutive digits are fixed. Now do a conjugation via homeomorphism from the original full shift to a Markov subshift of $2^N$ symbols, where each digit is sent to the $N$ digits starting at itself. Pick $A'$ appropriately such that it is

Perron-Frobenious and the eigenvector is close to the prescribed measures on the cylinder sets, and define the invariant measure as in Example 3.41. Such measures are mixing hence ergodic.

- The topological entropies are the same. By Poincare recurrence, any invariant probability measure for $g$ is supported in the compact subset $\{(b, f(b))\}$, hence the topological entropy of $g$ equals the topological entropy of $g|_{\{(b,f(b))\}}$, which is conjugate to $f$ via a homeomorphism.

# A Homeworks

## A.1 HW1

1. Let $(X, \Sigma)$ be a measurable space with a complete probability measure $\mu$, $T : X \to T$ a measurable map such that $T_*(\mu) = \mu$.
   (a) Show that $\mu(T(X)) = 1$.
   (b) If $B \in \Sigma$ satisfies $T(B) \subseteq B$, then there is some $B' \in \Sigma$, such that $T^{-1}(B') = B'$, and two sets $Z, Z'$ with measure zero such that $B \cup Z = B' \cup Z'$.
   (c) Let $A \in \Sigma$ be a set, such that for any $a \in A$, there is some $n_a > 0$, such that if $n > n_a$ then $T^n(a) \notin A$. Show that $\mu(A) = 0$.
2. Let $X$ be a compact metric space and $T : X \to X$ a continuous map.
   (a) Show that if $T(x) = x$ then $\delta_x$ is an ergodic measure of $T$.
   (b) Write down an $X$ and a $T$, so that $T$ has three Radon invariant ergodic measures.
3. Let $X$ and $Y$ be two mesurable spaces, $S : X \to X$, $T : Y \to Y$ measurable self maps. Let $\mu$ be a probability measure on $X$ ergodic under $S$, $\mu'$ a probability measure on $Y$ ergodic under $T$. Let $S \times T$ be a map from $X \times Y$ to itself defined as $(S \times T)(x, y) = (S(x), T(y))$.
   (a) Show that $\mu \times \mu'$ is a probability measure on $X \times Y$ invariant under $S \times T$.
   (b) Find an example of $X$, $Y$, $S$, $T$, $\mu$ and $\mu'$, such that $\mu \times \mu'$ is ergodic under $S \times T$.
   (c) Find an example of $X$, $Y$, $S$, $T$, $\mu$ and $\mu'$, such that $\mu \times \mu'$ is not an ergodic measure under $S \times T$.

Answer

1. (a) $\mu(T(X)) = \mu(T^{-1}(T(X)) = \mu(X) = 1$
   (b) Because $T(B) \subseteq B$, $B \subseteq T^{-1}(B)$. Because $\mu(B) = \mu(T^{-1}(B))$, we have $\mu(T^{-1}(B) \backslash B) = 0$. Now let $B' = \bigcup_i T^{-i}(B)$, then $T^{-1}(B') = B'$, $Z' = \emptyset$, $Z = \cup_i T^{-i}(T^{-1}(B) \backslash B)$ has measure 0.
   (c) This follows immediately from Poincare recurrence.
2. (a) $T_*(\delta_x) = \delta_{T(x)} = \delta_x$, hence it is an invariant measure. Any invariant subset, by construction, either contains $x$ or doesn't contain $x$, hence has measure of either 1 or 0.
   (b) Let $X$ be the interval $[-1, 1]$, $T : x \mapsto x^3$. By 1(c) above, for any $\epsilon > 0$, any invariant probability measure $\mu$, $\mu([-1+\epsilon, -\epsilon] \cup [\epsilon, 1-\epsilon]) = 0$, hence the support of $\mu$ is contained in $\{0, \pm 1\}$. The three radon ergodic probability measures are $\delta_0$, $\delta_{\pm 1}$.

3. (a) This follows from, for example, the Fubini's theorem.
   (b) Let $X = Y = \mathbb{R}/\mathbb{Z}$, with the invaruant measure being the Lebesgue measure, $S : x \mapsto x + a$, $T : x \mapsto x + b$, such that $1, a, b$ are $\mathbb{Q}$-linearly independent. Then for any invariant measurable subset $A$, in $L^2(X \times Y)$ we have $\chi_A = \sum_{m,n} c_{m,n} e^{2\pi i(mx+ny)}$. Invariance of $A$ implies that $c_{m,n} = c_{m,n} e^{2\pi i(ma+nb)}$, hence $c_{m,n} = 0$ unless $m = n = 0$, which implies that $A$ has either zero or full measure.
   (c) $X$, $Y$, $S$ and $T$ same as above but $a = b$ and is irrational, then the subset consisting of $(x, y)$ where the distance between them is no more than $1/3$ is an inavriant subset of measure $2/3$.

## A.2 HW2

1. Let $X = \mathbb{R}/\mathbb{Z}$, $T : X \to X$ defined as $T(x) = x + a$ where $a \in \mathbb{R}\backslash\mathbb{Q}$.
   (a) Show that for any open subinterval $I = (a, b) \subseteq X$, any $x \in X$, $\lim_{n\to\infty} \frac{1}{n} \left|\{j \in \{0, 1, \ldots, n-1\} : T^j(x) \in I\}\right|$ equals $b-a$, where $|A|$ is the cardinality of the set $A$.
   (b) Show that there is an open subset $U \subseteq X$, $x \in X$, such that $\lim_{n\to\infty} \frac{1}{n} \left|\{j \in \{0, 1, \ldots, n-1\} : T^j(x) \in U\}\right|$ does not equal the Lebesgue measure of $U$.
2. Let $X = \mathbb{R}/\mathbb{Z}$, $\mu$ be the Lebesgue measure, $T : x \mapsto 2x$. Find an $a \in X$ such that $\{a, T(a), T^2(a), \ldots\}$ is dense in $X$ but does not equidistributes to $\mu$.
3. Let $X$ be a compact metric space, $\mathcal{M}$ the set of Radon probability measure on $X$ with the weak $*$ topology (Here we identify $\mathcal{M}$ with a subset of $(C(X))^*$ via Riesz Representation Theorem). Then by Stone-Weierstrass, $\mathcal{M}$ with the weak $*$ topology is homeomorphic to a compact metric space.
   (a) Let $T : X \to X$ be a continuous map. Show that $\mu \mapsto T_*(\mu)$ is a map from $\mathcal{M}$ to itself that is continuous under the weak $*$ topology.
   (b) Let $X = \mathbb{R}/\mathbb{Z}$, $T(x) = x + a$, $a \notin \mathbb{Q}$. Find a non atomic measure on $\mathcal{M}$ which is an ergodic measure of the map $\mu \mapsto T_*(\mu)$.
   (c) Let $X = \mathbb{R}/\mathbb{Z}$, find a $T$ such that $\mu \mapsto T_*(\mu)$ is uniquely ergodic, or prove that such a $T$ does not exist.

Answer:

1. (a) Let $A = (a, b)$, $d(\cdot, \cdot)$ the Euclidean distance function, $\epsilon > 0$ be very small, define $f, g \in C(X)$ as

$$f(x) = \begin{cases} 1 & x \in A \\ 1 - \frac{d(x,A)}{\epsilon} & x \notin A, d(x, A) \le \epsilon \\ 0 & d(x, A) > \epsilon \end{cases}$$

$$g(x) = \begin{cases} 0 & x \notin A \\ \frac{d(x,A)}{\epsilon} & x \in A, d(x, X\backslash A) \leq \epsilon \\ 1 & d(x, X\backslash A) > \epsilon \end{cases}$$

Then by construction,

$$\frac{1}{n}\sum_{i=0}^{n-1} g(T^i(x)) \leq \frac{1}{n}\left|\{j \in \{0, 1, \ldots, n-1\} : T^j(x) \in I\}\right| \leq \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$$

Because $T$ is uniquely ergodic, for any $x$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} g(T^i(x)) = b - a - \epsilon$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) = b - a + \epsilon$$

Now let $\epsilon \to 0$ we finish the proof.

(b) Let $U = \bigcup_{i=0}^{\infty}(T^i(x) - 10^{-1-i}, T^i(x) + 10^{-1-i})$, then the limit equals 1 and the measure of $U$ is no more than $\frac{2}{9}$.

2. Let $a = \sum_{n=4}^{\infty}\sum_{j=0}^{2^n-1} j2^{-(10+2^{n+1}+j)!-n}$. By construction, $\{T^{2^{(10+2^{n+1}+j)!}}(a)\}$ is dense in $X$, and in the binary expansion there are way more 0 than 1s asymptotically so it can not equidistribute on $X$ (see 1(a) above).

3. (a) Let $U \subseteq \mathcal{M}$ be open under weak-$*$ topology, we need to show $(T_*)^{-1}(U)$ is open as well. Let $\mu \in U$, then there is a neighborhood $\mu \in U' \subseteq U$ defined as

$$U' = \left\{\mu \in \mathbb{M} : a_i < \int_X f_i d\mu < b_i, f_1, \ldots, f_n \in C(X)\right\}$$

So $(T_*)^{-1}(\mu) \in (T_*)^{-1}(U') \subseteq (T_*)^{-1}(U)$ and

$$(T_*)^{-1}(U') = \left\{\mu \in \mathbb{M} : a_i < \int_X (f_i \circ T) d\mu < b_i, f_1, \ldots, f_n \in C(X)\right\}$$

which is open.

(b) Consider continuous map $F : x \mapsto \delta_x$. It is easy to see that it gives a conjugation from $T$ to $T_*$. The unique ergodic measure on $X$ pushed forward into $\mathcal{M}$ via $F$ is a non atomic ergodic measure of $T_*$.

(c) If $T$ has multiple ergodic measures, like $\mu$, $\mu'$, then $\delta_\mu$ and $\delta_{\mu'}$ are distinct ergodic measures of $T_*$, so $T_*$ won't be uniquely ergodic.

If $T$ is uniquely ergodic, let $\mu$ be the unique ergodic Radon measure, then $F_*(\mu)$ and $\delta_\mu$ are both ergodic measures of $T_*$. If we want $T_*$ to be uniquely ergodic, they must be identical. So we can, for example, let $T$ be the constant map sending everything to 0, then $T_*$ sends $\mathcal{M}$ to $\delta_0$. Now if $\mu$ is a Radon probability measure invariant under $T_*$ then

$$\mu(U) = \mu(T^{-1}(U)) = \begin{cases} \mu(\mathcal{M}) = 1 & \delta_0 \in U \\ \mu(\emptyset) = 0 & \delta_0 \notin U \end{cases} = \delta_{\delta_0}$$

## A.3 HW3

1. Let $\Sigma = \{(a_n), n \in \mathbb{Z} : a_j \in \{1, 2\}, a_j = 2 \implies a_{j+1} = 1\}$ be a two sided Markov subshift.

   (a) Write down a Radon probability measure $\mu$ invariant under the shift map $\sigma$, such that $\sigma$ is mixing with respect to this measure.

   (b) Calculate the measure theoretic entropy of $\sigma$ with respect to this measure $\mu$.

2. Prove that if $P$ and $P'$ are two finite partitions of $X$ with probability measure $\mu$, then $h(P \vee P') \leq h(P) + h(P')$.

3. Let $X$ be a measurable space with a probability measure $\mu$, $T : X \to X$ a measure preserving map. Show that if for any measurable subset $A$, $\lim_{n\to\infty} \mu(T^{-n}(A) \cap A) = (\mu(A))^2$, then $T$ is mixing.

4. Write down an ergodic Radon probability measure for a two sided full shift of 3 letters, such that the shift map is not weakly mixing with respect to this measure, but has positive measure theoretic entropy.

5. Let $X$ be a compact metric space, $V_1, \dots, V_m$ be $n$ compact subsets of $X$ whose union is $X$, and $T : X \to X$ a continuous map. Let $A$ be a $m \times m$ matrix whose entries are either 0 or 1. Suppose further that if the $i, j$-th entry of $A$ is 1, then $V_j \subseteq T(V_i)$, and if $x, x'$ lies in the same $V_i$, we have $d(T(x), T(x')) > Ld(x, x')$ for some $L > 1$, then there is a continuous conjugation $h$ from one sided Markov subshift

   $$\Sigma_A = \{(a_n) \in \{1, \dots, m\}^{\mathbb{N}} : \text{ the } a_j, a_{j+1}\text{-th entry of } A \text{ is } 1\}$$

   to $X$, such that $T \circ h = h \circ \sigma$ ($\sigma((a_n)) = (b_n)$ then $b_j = a_{j+1}$), and $T^k(h((a_n))) \in V_{a_k}$.

**Remark:** As a further exercise, can you write down a sufficient condition for the existence of a continuous conjugation from a two sided Markov shift to a homoemorphism on a compact metric space?

Answer:

1. (a) $\mu(\{(a_n) : a_j = b_j, \dots, a_k = b_k\}) = y_{b_j} \prod_{l=k}^{j-1} r_{b_l b_{l+1}}$, where $y_1 = 2/3$, $y_2 = 1/3$, $r_{11} = r_{12} = 1/2$, $r_{21} = 1$, $r_{22} = 0$.

   (b) The entropy is $h = 2/3 \log(2)$.

2. Let the measures of the elements of $P$ and $P'$ be $\mu(U_i) = s_i$ and $\mu(V_j) = t_j$ respectively, $\mu(U_i \cap V_j) = r_{ij}$. Then

$$h(P \vee P') - h(P) = -\sum_{ij} r_{ij} \log(r_{ij}/s_i) = \sum_i s_i \left( -\sum_j \frac{r_{ij}}{s_i} \log\left(\frac{r_{ij}}{s_i}\right)\right)$$

By concavity of the function $-\sum_j x_j \log(x_j)$, the above is bounded from above by $h(P')$.

3. Let $f \in L^2(X)$, we only need to show that

$$\lim_{n\to\infty} \int_X f(\chi_A \circ T^n) d\mu = \mu(A) \int_X f d\mu$$

Consider $H = span\{1, \chi_A \circ T^n\}$. For any element $g = c + \sum_j a_j(\chi_A \circ T^j) \in H$,

$$\lim_{n\to\infty} \int_X g(\chi_A \circ T^n) d\mu = c\mu(A) + \lim_{n\to\infty} \sum_j a_j \mu(T^{-j}(A) \cap T^{-n}(A))$$

$$= \left( c + \sum_j a_j \mu(A) \right) \mu(A) = \mu(A) \int_X g d\mu$$

By continuity of inner product,

$$\lim_{n\to\infty} \int_X f(\chi_A \circ T^n) d\mu = \mu(A) \int_X f d\mu$$

is true for any $f \in \overline{H}$. It is evident that for any $f \in H^\perp$,

$$\lim_{n\to\infty} \int_X f(\chi_A \circ T^n) d\mu = 0 = \mu(A) \int_X f d\mu$$

Hence the identity is proved for all $f$ due to orthogonal projection.

4. $\mu(\{(a_n) : a_j = b_j, \ldots, a_k = b_k\}) = y_{b_j} \prod_{l=k}^{j-1} r_{b_l b_{l+1}}$, where $y_1 = y_2 = 1/4$, $y_3 = 1/2$,

$$[r_{ij}] = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

By considering subsets $A = B = \{(a_n) : a_n = 3 \text{ iff } n \text{ is even}\}$ we see that the shift map $\sigma$ is not weakly mixing. However, $\sigma^2$ is strongly mixing on $A$, hence any measurable invariant subset intersecting with $A$ must have either zero measure or full measure ($\mu(A) = 1/2$), which implies that the invariant subset itself must have a measure of 0 or 1. By calculation it is easy to see that the topological entropy is $\log(2)/2 > 0$.

5. $h$ is defined as follows:

$$h(a_n) = \bigcap_n (T|_{V_{a_0}})^{-1}(T|_{V_{a_1}})^{-1}\dots(T|_{V_{a_{n-1}}})^{-1}V_{a_n}$$

Now one can verify that $h$ is well defined (it is the intersection of a nested sequence of compact sets with diameter goes to 0), continuous, and satisfies all the other assumptions.

## A.4 HW4

1. Let $X$ be a compact metric space and $T : X \to X$ is a continuous map. Suppose $A \subseteq X$ is a closed subset such that $T^{-1}(A) = A$. What is the relationship between the topological entropy of $T$ and of $T|_A$? Justify your answer.

2. Write down a continuous map from $\mathbb{R}/\mathbb{Z}$ to itself, such that the Lebesgue measure is invariant unfer this map, and the measure theoretic entropy with respect to the Lebesgue measure is infinity.

3. Let $X$ be a compact metric space and $T : X \to X$ a continuous map. Let $D$ be the set of invariant Radon probability measures, with weak $*$ topology. For any $\mu \in D$, let $h(T, \mu)$ be the measure theoretic entropy. Is the function $\mu \to h(T, \mu)$ a continuous function on $D$? Prove it or write down a counter example.

Answer:

1. The topological entropy of $T$ is no less than the topological entropy of $T|_A$. This is because any open cover of $X$ must always induce an open cover of $A$.

2. From definition it is easy to see that if $X$ can be decomposed into disjoint unions of $T$-invariant subsets $A_i$ with positive measures, then $h(T, \mu) = \sum_i \mu(A_i)h(T|_{A_i}, \mu|_{A_i})$. Here for any measurable $B \subseteq A_i$, $\mu|_{A_i}(B) = \mu(B)/\mu(A_i)$. Hence we can define the map $T$ as:

$$T(x) = \begin{cases} \frac{1}{n+1} + 3^{2^n}\left(x - \frac{1}{n+1} - \frac{2j}{3^{2^n}n(n+1)}\right) & \frac{1}{n+1} + \frac{2j}{3^{2^n}n(n+1)} < x \le \frac{1}{n+1} + \frac{2j+1}{3^{2^n}n(n+1)} \\ \frac{1}{n+1} + 3^{2^n}\left(\frac{1}{n+1} + \frac{2k+2}{3^{2^n}n(n+1)} - x\right) & \frac{1}{n+1} + \frac{2k+1}{3^{2^n}n(n+1)} < x \le \frac{1}{n+1} + \frac{2k+2}{3^{2^n}n(n+1)} \end{cases}$$

Here $n$ goes through all positive integers, $j$ integer such that $0 \le j \le (3^{2^n} - 1)/2$, $k$ integer such that $0 \le k \le (3^{2^n} - 3)/2$.

3. It is not continuous. As an example, consider one sided full shift of 2 symbols. Let $\mu$ be the measure of maximal entropy, then $h(T, \mu) = \log(2)$. Let $\mu_n$ be the average of the $\delta$-measure on $x \in X$ such that $T^n(x) = x$, then it is easy to see that $\mu_n$ converges to $\mu$ under weak-$*$ topology, yet $h(T, \mu_n) = 0$ for all $n$.

## A.5 HW5

1. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function such that $f(x+n) = f(x) + 2n$ for all integer $n$. Let $T : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be defined as $T([x]) = [f(x)]$, where $[a]$ is the equivalence class represented by $a$ in the quotient space. Show that the topological entropy of $T$ is no less than $\log(2)$.

2. Let $X$ be a compact metric space and $T : X \to X$ a continuous map. Let $T' : X \times X \to X \times X$ be defined as $T'(a,b) = (T(a), T(b))$. What is the relationship between the topological entropy of $T$ and the topological entropy of $T'$? Here the metric on product space $X \times X$ is chosen as $d_{X\times X}((a,b),(a',b')) = \max\{d_X(a,a'), d_X(b,b')\}$.

3. Let $X$ be a compact metric space with a continuous map $H : X \times \mathbb{R}_{\geq 0} \to X$, such that $H(H(x,s),t) = H(x, s+t)$.
   (a) Show that there is a Radon probability measure $\mu$ on $X$ such that for any measurable set $A$, any $t \in \mathbb{R}_{\geq 0}$, $\mu(\{a \in X : H(a,t) \in A\}) = \mu(A)$.
   (b) Find $X$ and $H$ such that this Radon invariant probability measure is unique.

4. (This is how we get continued fraction from the Rauzy-Veech induction of a 2-IET) Let $\alpha$ be an irrational number between 0 and 1, consider sequence $a_n$ and $b_n$ defined as follows: $a_0 = \alpha$, $b_n = \begin{cases} 0 & a_n < 1/2 \\ 1 & a_n > 1/2 \end{cases}$, $a_{n+1} = \begin{cases} a_n/(1-a_n) & b_n = 0 \\ (1-a_n)/a_n & b_n = 1 \end{cases}$. Can you write down the continued fraction of $\alpha$ from the sequence $\{b_n\}$?

Answer:

1. By compactness, there is some $\epsilon > 0$, such that for any $x \in \mathbb{R}/\mathbb{Z}$, there are $y, y' \in T^{-1}(x)$ such that the distance between them is more than $\epsilon$. Now for any natural number $n$, let $x_{1,1}, x_{1,2} \in T^{-1}(0)$ with distance larger than $\epsilon$, $x_{2,1}, x_{2,2} \in T^{-1}(x_{1,1})$, $x_{2,3}, x_{2,4} \in T^{-1}(x_{1,2})$, such that the distance between $x_{2,1}$ and $x_{2,2}$, as well as $x_{2,3}$ and $x_{2,4}$, are both larger than $\epsilon$. Continue the process to $x_{n-1,i}$, we see that they form a $(n,\epsilon)-$separated set with $2^{n-1}$ elements, which implies that the topological entropy is no less than $\log(2)$.

2. $h(T') = 2h(T)$. For any $\epsilon > 0$, the product of an $(n,\epsilon)$-separated set in $X$ with itself is an $(n,\epsilon)$-separated set in $X \times X$, hence $h(T') \geq 2h(T)$. Now consider the product of a maximal $(n,\epsilon/3)$-separated set in $X$ with itself, the open balls centered at these points, with radius $\epsilon/2$, cover the whole $X \times X$, and each such ball contains at most one element in an $(n,\epsilon)$-separated set in $X \times X$, hence $h(T') \leq 2h(T)$.

3. (a) Pick $x \in X$, consider $\mu_n$ defined as follows: for any $f \in C(X)$,

$$\int_X f d\mu_n = \frac{1}{n}\int_0^n f(H(x,t))dt$$

and let $\mu^*$ be the weak-* limit of a subsequence. Now for any $n$, any $f \in C(X)$, any $t \geq 0$,

$$\left|\int_X f(H(a,t))d\mu_n(a) - \int_X f(a)d\mu_n(a)\right|$$

$$= \left|\frac{1}{n}\left(\int_t^{n+t} f(H(x,s))ds - \int_0^n f(H(x,s))ds\right)\right| \leq \frac{2t\max(|f|)}{n}$$

By construction of $\mu^*$, one can always find arbitrarily large $n$ such that $\mu_n$ is close to $\mu^*$ under weak-$*$ topology, and the above inequality shows that $\mu^*$ is $H(\cdot,t)$-invariant for all $t$.

(b) Let $X = \mathbb{R}/\mathbb{Z}$ and $H(x,t) = x + t$.

4. Suppose $\alpha = [0; a_1, a_2, a_3, \dots]$, then the sequence $\{b_n\}$ would be $a_1 - 1$ zeros followed by 1 followed by $a_2 - 1$ zeros followed by 1 followed by $a_3 - 1$ zeros followed by 1 ....

# B Final Practice Problems

1. Let $X$ be a compact closed subset of $\mathbb{R}^2$, with Euclidean metric, $T : X \to X$ such that $d(T(x), T(b)) \leq 2d(x,y)$ for any $x, y \in X$. Show that the topological entropy of $T$ is no more than $\log(4)$.

2. Let $X$ be a compact metric space, $T : X \to X$ a continuous map. If $T \circ T = id_X$ and $T$ is uniquely ergodic, then $T$ contains no more than two points.

3. Let $X = [0,1]$ be the closed interval, $T : X \to X$ defined as $T(x) = \begin{cases} 2x & x \leq 1/2 \\ 2-2x & x > 1/2 \end{cases}$.

   (a) Show that the topological entropy of $T$ is $\log(2)$.

   (b) Show that the Lebesgue measure is an invariant Radon probability measure, and the measure theoretic entropy of $T$ under the Lebesgue measure is $\log(2)$.

   (c) Write down a Radon invariant probability measure on $X$ which is ergodic but not mixing.

4. Let $A$ be a compact metric space, $X = \mathbb{A}^{\mathbb{N}}$, $T$ sends $(a_0, a_1, \dots) \in X$ to $(a_1, a_2, \dots)$.

(a) Show that $T$ is continuous.

(b) Show that if $A$ has more than one points then $T$ is not uniquely ergodic.

(c) Show that if $A$ has infinitely many points then $T$ has topological entropy $\infty$.

Answer:

1. Given any $\epsilon > 0$, cover $X$ with discs of diameter $\epsilon/2^n$, then the number of such discs are bounded from above by $O(4^n/\epsilon^2)$. Each disc contains at most one element of a $(n, \epsilon)$ separated set, hence the topological entropy is no more than $\log(4)$.

2. For any $x \in X$, $\frac{\delta_x + \delta_{T(x)}}{2}$ is an ergodic measure. Hence if there is a single ergodic measure then $X$ must have 2 points.

3. You can use a method similar to problem 1 to show that the topological entropy is no more than $\log(2)$. By making use of Markov decomposition $\{[0, 1/2], [1/2, 1]\}$ we can find a conjugacy from full shifts of two letters to this map (see also HW3 problem 5), and the measure of maximal entropy is sent to the Lebesgue measure by this conjugacy, hence the topological entropy is no less than $\log(2)$. To find an ergodic but not mixing measure, find a point with finite forward orbit and consider the average of the $\delta$ measures along this orbit.

4. (a) This is by the definition of product topology.

   (b) If $A$ has two distinct points $a$ and $b$, then $\delta_{(a,a,a,\dots)}$ and $\delta_{(b,b,b,\dots)}$ are both ergodic Radon measures.

   (c) For any $n$ there is a closed $T$-invariant subset of $X$ $\{a_1, \dots, a_n\}^{\mathbb{N}}$, where $a_1, \dots a_n$ are distinct elements of $A$. This subset is $T$ invariant and the topological entropy of $T$ on it is $\log(n)$, hence the topological entropy of $T$ on $X$ is no smaller than $\log(n)$ for all $n$, which means it has to be infinity.

# C Final Exam

1. Let $X = \{0, 1, 2\}$, with the discrete topology. Let continuous map $T : X \to X$ be defined as $T(0) = T(2) = 1$, $T(1) = 0$. Let $\mu = \frac{1}{2}(\delta_0 + \delta_1)$ be a probability measure on $X$.

   (a) Show that $\mu$ is invariant under $T$, and it is also an ergodic measure. (15 points)

   (b) Is $T$ uniquely ergodic? Why or why not? (10 points)

   (c) Is $T$ weakly mixing with respect to $\mu$? Why or why not? (10 points)

2. Let $Y = \mathbb{R}/\mathbb{Z}$ with the Euclidean metric, and $S : Y \to Y$ be defined as $S(x) = x + a$, where $a \notin \mathbb{Q}$. Let $X$ be a compact metric space, whose elements are all the non empty compact subsets of $Y$, and distance function is the "Hausdorff metric" defined as

$$d_H(A, B) = \max\left\{\sup_{a\in A}\left(\inf_{b\in B} d(a, b)\right), \sup_{b\in B}\left(\inf_{a\in A} d(a, b)\right)\right\}$$

You do not need to write down a proof that $(X, d_H)$ is compact as a metric space. Define $T : X \to X$ as $T(A) = \{S(a) : a \in A\}$. Let $\mu$ be a Radon probability measure on $X$ defined as $\mu(U) = \mu_Y(\{y \in Y : \{y\} \in U\})$, where $\mu_Y$ is the Lebesgue measure.

   (a) Show that $\mu$ is an ergodic measure of $T$. (15 points)

   (b) Write down the measure theoretic entropy of $T$ with respect to $\mu$. Hint: The measure theoretic entropy of $h(S, \mu_Y)$ is bounded from above by the topological entropy of $S$, hence equals 0. (15 points)

   (c) Show that $T$ is not uniquely ergodic by writing down a different Radon ergodic probability measure for $T$. (5 points)

   (d) Does $X$ have a Radon invariant probability measure $\mu'$ such that the measure theoretic entropy $h(T, \mu')$ is positive? Why or why not? (5 points)

3. Let $A = \{z \in \mathbb{C} : |z| = 1\}$, with the Euclidean metric on $\mathbb{C}$. Let $\Sigma \subseteq A^{\mathbb{Z}}$ be defined as follows:

$$(\dots, a_{-n}, \dots, a_{-1}, a_0, a_1, \dots, a_n, \dots) \in \Sigma \text{ iff } a_i^2 = a_{i+1}^3 \text{ for all } i \in \mathbb{Z}$$

Let $\sigma$ be the shift map, i.e. $\sigma(z_i) = (w_i)$ then $w_i = z_{i+1}$.

   (a) Show that $\Sigma$ is a closed subset of $A^{\mathbb{Z}}$ under the product topology, and $\sigma(\Sigma) = \sigma^{-1}(\Sigma) = \Sigma$ (15 points)

   (b) Calculate the topological entropy of $\sigma|_\Sigma : \Sigma \to \Sigma$. (10 points)

Answer

1. (a) There are only 8 possible subsets of $X$ and you can check them one by one.

   (b) Yes. Any probability measure is of the form $a\delta_0 + b\delta_1 + c\delta_2$, and being invariant under $T$ implies that $a = b$ and $c = 0$.

   (c) No. Let $A = B = \{0\}$.

2. (a) The map $h : Y \to X$, $h(y) = \{y\}$, is a conjugation from $S$ to $T$ that sends $\mu_Y$ to $\mu$, hence $\mu$ is ergodic.

   (b) It is 0, because $h(T, \mu) = h(S, \mu_Y) \le h(S) = 0$.

(c) No, because the topological entropy of any isometry on compact metric space is 0, which one can see easily from the $n, \epsilon$-separable set definition of topological entropy.

3. (a) $\sigma(\Sigma) = \sigma^{-1}(\Sigma) = \Sigma$ follows immediately from definition. $a_i^2 = a_{i+1}^3$ involves only finitely many $a_i$s hence are all closed conditions, and a set defined by a set of closed conditions is closed.

   (b) Let the metric on $\Sigma$ be $d((a_i), (b_i)) = \sum_{i \in \mathbb{Z}} 2^{-|i|}|a_i - b_i|$. Let $\epsilon < 1/10$. For any $n$, pick $3^{n-1}$ elements of the form $(z_i)$ in $\Sigma$ such that $z_0 = 1$, and $z_1, \ldots, z_{n-1}$ goes through all $3^{n-1}$ possibilities, then these elements form a $n, \epsilon$-separated set. Hence $h(\sigma) \leq \log(3)$. Now given any $\epsilon$, pick $M$ points in $\Sigma$ such that the $\epsilon/2$-balls centered at these points cover $\Sigma$. Pick $N'$ such that $2^{-N'+2} < \epsilon/2$. Now for every $n$, for each of the $M$ points $(z_i)$, pick $3^{n+N'}$ points of the form $(w_i)$, such that $w_i = z_i$ when $i \leq 0$ and $w_1, \ldots w_{n+N'}$ goes through all possibilities. Then for any $a = (a_i) \in \Sigma$, $a, \sigma(a), \ldots, \sigma^{n-1}(a)$ must all be in the $\epsilon/2$-neighborhood of one of the $M \cdot 3^{n+N'}$ points, hence $h(\sigma|_\Sigma) \leq \log(3)$.