Dynamical Systems and Ergodic Theory

Jiaqi Cui
East China Normal University

November 25, 2025

Abstract

This is my notes of the course "Dynamical Systems and Ergodic Theory" given by Manfred Einsiedler. See https://video.ethz.ch/lectures/d-math/2024/spring/401-2374-24L.html. I filled most of details omitted in the lectures.

Contents

0 Examples

Let X be a set and $T\colon X \to X$ be a map.

Definition 0.1. *fixed point, periodic point, period, orbit...*

Definition 0.2. *Assume X has a topology. The ω-limit of $x \in X$ is*

$$\omega^{\pm}(x) := \left\{ \lim_{k\to\infty} T^{n_k}x : n_k \nearrow \pm\infty \right\}.$$

We could also ask about the "distribution" of $x, Tx, T^2x, \cdots, T^n x$ inside X as $n \to \infty$.

More generally, a dynamical system can be defined as a group action.

Example 0.3. $X = \mathbb{R}$, $Tx = x + 1$. The ω-limits are empty set. Thus we will restrict to compact metric spaces.

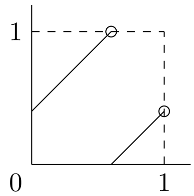
Example 0.4. $X = \mathbb{R} \cup \{\infty\}$ the one-point compactification of $\mathbb{R}$. $T(x) = \begin{cases} x+1, \ x \in \mathbb{R}; \\ \infty, \ x = \infty \end{cases}$. Then the ω-limits are all $\{\infty\}$.

Example 0.5. $X = \mathbb{R} \cup \{\pm\infty\}$. $T(x) = \begin{cases} x+1, \ x \in \mathbb{R}; \\ +\infty, \ x = +\infty; \\ -\infty, \ x = -\infty \end{cases}$. Then $\omega^+(x) = \begin{cases} +\infty, \ x \in \mathbb{R} \cup \{+\infty\}; \\ -\infty, \ x = -\infty \end{cases}$.

Example 0.6. North-South Dynamics

Example 0.7. $X = \mathbb{T} = \mathbb{R}/\mathbb{Z} \cong \mathbb{S}^1$ with the metric $d(x+\mathbb{Z}, y+\mathbb{Z}) = \min_{k \in \mathbb{Z}} |x - y + k|$. $R(x+\mathbb{Z}) := x+\alpha+\mathbb{Z}$ for a fixed $\alpha \in \mathbb{T}$. $R\colon \mathbb{T} \to \mathbb{T}$ is an isometry.

- If $\alpha = \frac{p}{q}$ is rational, then $R^q(x+\mathbb{Z}) = x + q\alpha + \mathbb{Z} = x + p + \mathbb{Z} = x + \mathbb{Z}$. Every point is periodic with period q.



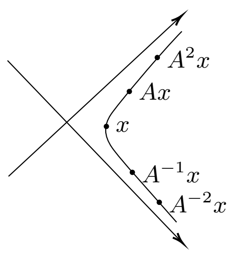
- If $\alpha \notin \mathbb{Q}/\mathbb{Z}$, then no point is periodic: say $R^n(x+\mathbb{Z}) = x + \mathbb{Z}$, then $n\alpha \in \mathbb{Z}$. Actually, all orbits are dense in this case.

Example 0.8. $X = \mathbb{T} = \mathbb{R}/\mathbb{Z}$. Fix $p \geqslant 2 \in \mathbb{N}$. $T(x) := px$. This map links to the base-p expansion of $x \in [0,1)$. Suppose $x = \sum_{k=1}^{\infty} \theta_k p^{-k}$ where $\theta_k\{0, \cdots, p-1\}$. Then $Tx = p\sum_{k=1}^{\infty} \theta_k p^{-k} + \mathbb{Z} = \sum_{k=1}^{\infty} \theta_{k+1} p^{-k} + \mathbb{Z}$.

Claim.

- There exist lots of periodic points– they are dense.
- There exist pre-periodic points that are not periodic, where x is pre-periodic if its orbit $|\mathcal{O}^+(x)| < \infty$.
- Many periods exist.
- The set of pre-periodic points is countable.
- There exist $x \in \mathbb{T}$ with $\omega^+(x) = \mathbb{T}$.
- There exist $x \in \mathbb{T}$ with $\omega^+(x)$ uncountable but not $\mathbb{T}$.
- There exist $x \in \mathbb{T}$ with $\omega^+(x)$ countable but not finite.

Example 0.9. $X = \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$, $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. This is called a hyperbolic toral automorphism, because the orbit of any $x \neq 0 \in X$ is on a hyperbola.



Example 0.10. $X = (0,1) - \mathbb{Q}$, $Tx = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor$. This relates to continued fraction expansion

$$x = \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}}$$

where $a_1, a_2, a_3, \cdots \in \mathbb{N}$. Note that

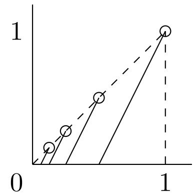
$$Tx = \cfrac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \cdots}}}.$$

Example 0.11 (Benford‘s law for powers of 2)**.** Given $j \in \{1, \cdots, 9\}$, the limits

$$d_j := \lim_{N\to\infty} \frac{1}{N} \sharp \{2^n : 1 \leqslant n \leqslant N,\ 2^n \text{ starts in digital expension with } j\}$$

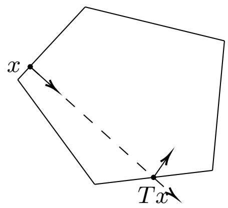
satisfy $d_1 > d_2 > \cdots > d_9 > 0$. In fact $d_1 = \log_{10} 2$.

Example 0.12. $X = [0,1]$, $T(x) = \begin{cases} 0, & x = 0, 1; \\ nx - 1, & x \in \left[\frac{1}{n}, \frac{1}{n-1}\right) \end{cases}$.



We claim that $\lim_{n\to\infty} T^n x = 0$, and if $x \in \mathbb{Q}$, there exists n with $T^n x = 0$ and if $x \notin \mathbb{Q}$, $T^n x > 0$ for all n. For $x = e$ this can be used to show that $e \notin \mathbb{Q}$.

Example 0.13 (Billiards)**.** X is the set of boundary points with a vector and T is the movement of a boundary point along its vector to the next boundary point with a reflected vector.



Given a triangle, are there always periodic points? When the triangle is acute, right or obtuse with rational angles, the answer is yes. For the other cases, people don‘t know.

Example 0.14 (Geodesic flow)**.** Given a nice manifold M and its unit tangent bundle. There exists a way of following any vector in the tangent. When M is a sphere, the orbits are great circles. When M is a torus, whether an orbit is closed depending on whether the initial vector is rational. When M is a orientable surface of genus greater than 2, the orbits will be more complicated and more sensitive to initial values.

1 Topological Dynamics

Assume X is a compact metric space and $T : X \to X$ is continuous or even a homeomorphism.

Definition 1.1. *A homeomorphism $T : X \to X$ is called (topological) transitive if there exists a point $x_0 \in X$ for which the orbit is dense, i.e. $\overline{\mathcal{O}(x_0)} = X$.*

Definition 1.2. *A comtinuous map $T : X \to X$ is called forward transitive if there exists $x_0 \in X$ with $\overline{\mathcal{O}^+(x_0)} = X$.*

Example 1.3. $T_p : \mathbb{T} \to \mathbb{T}$ for $p \geqslant 2$ an integer which maps x to px is forward transitive. We will construct x_0 using base-p-expansion. We first list all finite sequences in the symbols $0, 1, \cdots, p-1$, and consider the result as one sequence of digits $a_1, a_2, \cdots \in \{0, 1, \cdots, p-1\}$. Then we define $x_0 := \sum_{k=1}^{\infty} a_k p^{-k}$. For any $\left[\frac{j}{p^i}, \frac{j+1}{p^i}\right)$, we can find an l such that $T^l x_0$ is in this interval. Thus $\mathcal{O}^+(x_0)$ is dense in $\mathbb{T}$. For example, for $p = 2$, we write

$$0, 1, 00, 01, 10, 000, \cdots, 111, 0000, \cdots, 1111, \cdots$$

Then we write it as a number sequence

$$0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, \cdots, 1, 1, 1, 0, 0, 0, 0, \cdots, 1, 1, 1, 1, \cdots$$

and then

$$x_0 = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{8} + 0 \cdot \frac{1}{16} + 0 \cdot \frac{1}{32} + 1 \cdot \frac{1}{32} + 1 \cdot \frac{1}{64} + \cdots$$

When we apply T on x_0 for n times, the first n numbers of the number sequence will become 0. Then for any $\frac{j+1}{2^i}$, we can find an n such that the base-2-expansion of $\frac{j+1}{2^i}$ will at the start of the number sequence. This means $T^n x_0 \in \left[\frac{j}{2^i}, \frac{j+1}{2^i}\right)$.

Example 1.4. $R_\alpha \colon \mathbb{T} \to \mathbb{T}$ maps x to $x + \alpha$.

- If $\alpha \in \mathbb{Q}/\mathbb{Z}$, R_α only has periodic orbits and so is not transitive.
- If $\alpha \notin \mathbb{Q}/\mathbb{Z}$, R_α is topological transitive. See later.

Proposition 1.5. *Let $T \colon X \to X$ be a homeomorphism. The followings are equivalent:*

1. *T is topological transitive;*
2. *if $U \subset X$ is open and $TU = U$, then either $U = \varnothing$ or $\overline{U} = X$;*
3. *if $U, V \subset X$ are non-empty and open, then there exists $n \in \mathbb{Z}$ so that $T^n U \cap V \neq \varnothing$;*
4. *the set $\left\{x_0 \in X : \overline{\mathcal{O}(x_0)} = X\right\}$ is a dense G_δ-set.*

Definition 1.6. *A set G is called a G_δ-set if it is a countable intersection of open sets.*

Theorem 1.7 (Baire Category Theorem)**.** *Let X be a complete metric space. Let $O_n \subset X$ be a sequence of dense open sets. Then $\bigcap_{n=1}^{\infty} O_n$ is a dense G_δ-set.*

Proof. We only prove that $\bigcup_{n=1}^{\infty} O_n$ is dense. For any open set U, we want to find a point in $U \cap \bigcup_{n=1}^{\infty} O_n$. First, $U \cap O_1$ is non-empty and open because O_1 is open and dense. Then we can find a open ball $B_{\varepsilon_1}(x_1) \subset U \cap O_1$. Repeat this process. We find a open ball $B_{\varepsilon_2}(x_2) \subset B_{\varepsilon_1}(x_1) \cap O_2$... WLOG we can suppose $\varepsilon_n \leqslant \frac{1}{n}$. Thus we can claim that $\{x_n\}$ is a Cauchy sequence and $x := \lim_{n\to\infty} x_n \in U \cap \bigcap_{n=1}^{\infty} O_n$: by construction, $x_m \in B_{\varepsilon_{m-1}}(x_{m-1}) \subset \cdots \subset B_{\varepsilon_n}(x_n)$ and then $d(x_m, x_n) < \varepsilon_n \leqslant \frac{1}{n}$. Note that X is complete, the limit $x = \lim_{n\to\infty} x_n \in X$ exists. For all n, taking the limit of m, we obtain $x = \lim_{m\to\infty} x_m \in \overline{B_{\varepsilon_n}(x_n)} \subset O_n$. Moreover, $x \in \overline{B_{\varepsilon_1}(x_1)} \subset U$. □

Corollary 1.8. *A countable intersection of dense G_δ-sets is a dense G_δ-set.*

Proof of Proposition 1.5. **(1)⇒(2):** Let $x_0 \in X$ with $\mathcal{O}(x_0)$ dense in X. Because U is open, $\mathcal{O}(x_0) \cap U \neq \varnothing$. Then there exists $n \in \mathbb{Z}$ such that $T^n x_0 \in U$. Note that $TU = U$, we have $x_0 \in T^{-n} U = U = T^{-m} U$ for any $m \in \mathbb{Z}$. This shows $\mathcal{O}(x_0) \subset U$ and then U is dense in X.

(2)⇒(3): Define $\widetilde{U} = \bigcup_{n\in\mathbb{Z}} T^n U$. We note that $T\widetilde{U} = \widetilde{U}$ is non-empty and open. Then it is dense. Because V is open, $\widetilde{U} \cap V \neq \varnothing$ and then there exists $n \in \mathbb{Z}$ such that $T^n U \cap V \neq \varnothing$.

(3)⇒(4): For any $n \in \mathbb{N}$, $\bigcup_{x\in X} B_{\frac{1}{n}}(x)$ is a open cover. Because X is compact, there exists $k(n) \in \mathbb{N}$ such that $\bigcup_{i=1}^{k(n)} B_{\frac{1}{n}}(x_i)$ covers X. Denote $B_{\frac{1}{n}}(x_i)$, $i = 1, \cdots, k(n)$, $n \in \mathbb{N}$ by $O_1, O_2, \cdots$. To show density of a set, it suffices to show that the set intersects any of those open sets $O_1, O_2, \cdots$. For any $j \in \mathbb{N}$, we define $\widetilde{O_j} = \bigcup_{n\in\mathbb{Z}} T^n O_j$. By assumption, $\widetilde{O_j}$ intersects any other open set. This means $\widetilde{O_j}$ is dense. By Baire Category Theorem, $G := \bigcap_{j=1}^{\infty} \widetilde{O_j}$ is a dense G_δ-set and consists precisely of all points $x_o \in X$ with dense orbit. To see that, if x_0 has dense orbit, $\mathcal{O}(x_0)$ must intersect all open set $O_1, O_2, \cdots$. Then for any O_i, there is $n \in \mathbb{Z}$ such that $T^n x_0 \in O_i$. Thus $x_0 \in \widetilde{O_i}$ and then $x_0 \in G$. Conversely, for any $x_0 \in G$, $\mathcal{O}(x_0)$ intersects open balls with any small radius. Thus it is dense.

(4)⇒(1): Dense set can not be empty. □

Definition 1.9. *A homeomorphism $T \colon X \to X$ is called topological mixing if for any two non-empty open sets $U, V \subset X$, there exists N such that $T^n U \cap V \neq \varnothing$ for all $n \in \mathbb{Z}$ with $|n| > N$.*

Definition 1.10. *A homeomorphism $T \colon X \to X$ is called minimal if every orbit is dense.*

Proposition 1.11. *Let $T \colon X \to X$ be a homeomorphism of a compact metric space X. The followings are equivalent:*

1. *T is minimal;*
2. *if $E = TE \subset X$ is closed, then either $E = \varnothing$ or $E = X$;*
3. *if $U \subset X$ is open and non-empty, then $\bigcup_{n\in\mathbb{Z}} T^n U = X$.*

Proof. **(1)⇒(2):** Suppose $x \in E$. We have $X = \overline{\mathcal{O}(x)} \subset \overline{E} = E$.

(2)⇒(3): Denote $\widetilde{U} = \bigcup_{n\in\mathbb{Z}} T^n U$. $\widetilde{U}$ is open and $T\widetilde{U} = \widetilde{U}$. Hence $E := X - \widetilde{U}$ is closed and $TE = E$. Then we have $E = \varnothing$.

(3)⇒(1): Let $x_0 \in X$ and $U \neq \varnothing \subset X$ be open. Then there is $n \in \mathbb{Z}$ such that $x_0 \in T^n U$. This shows that $\mathcal{O}(x_0)$ intersects any non-empty open subset of X. Hence it is dense. □

Theorem 1.12. *Let X be a compact metric space and $T\colon X \to X$ be a homeomorphism. Then there exists a closed non-empty subset $Y \subset X$ so that $Y = TY$ and $T|_Y\colon Y \to Y$ is minimal.*

Proof. Define
$$\mathcal{E} := \{Y \subset X : Y \neq \varnothing, Y \text{ closed}, TY = Y\} \ni X.$$
We want to use Zorn's Lemma under $\subset$. We need to show that any chain in $\mathcal{E}$ has a lower bound in $\mathcal{E}$. Let $(Y_\alpha : \alpha \in \mathcal{I})$ be a chain. Define $Y = \bigcap_{\alpha \in \mathcal{I}} Y_\alpha$. Clearly, $TY = Y$ and Y is closed. Note that any intersection of finite Y_α is non-empty, by compactness of X, $Y \neq \varnothing$. These show that Y is the lower bound of chain $(Y_\alpha : \alpha \in \mathcal{I})$.

By Zorn's Lemma, there is a minimal element $Y \in \mathcal{E}$. Then $T|_Y\colon Y \to Y$ is minimal by Proposition 1.11. □

Corollary 1.13. *There exists $x_0 \in X$ so that $T^{n_k} x_0 \to x_0$ as $k \to \infty$ for some $n_k \nearrow \infty$.*

Proof. Let $Y \subset X$ be as in Theorem 1.12 and $x_0 \in Y$ be arbitrary. We have $\omega^+(x_0) \subset Y$ and $\omega^+(x_0)$ is T-invariant closed set. Note that $T|_Y\colon Y \to Y$ is minimal, by Proposition 1.11, $\omega^+(x_0) = \varnothing$ or Y. Note again that Y is a closed subset of the compact space X, it is also compact. Hence $\omega^+(x_0) \neq \varnothing$. This shows that $x_0 \in Y = \omega^+(x_0)$. □

Definition 1.14. *$x_0 \in X$ is called recurrent for $T\colon X \to X$ if $x_0 \in \omega_T^+(x_0)$.*

Example 1.15. Let $\alpha \notin \mathbb{Q}$ and $R\colon \mathbb{T} \to \mathbb{T}$ which maps x to $x + \alpha$. Then R is minimal: by 1.13, there is $x_0 \in \mathbb{T}$ which is recurrent. Let $\varepsilon > 0$. There is $n \in \mathbb{N}$ such that $\varepsilon > d(R^n x_0, x_0) = d(x_0 + n\alpha, x_0) = d(n\alpha, 0)$. Note that $n\alpha \neq 0$ in $\mathbb{T}$. For any $x \in \mathbb{T}$, $x, x \pm n\alpha, x \pm 2n\alpha, \cdots, x \pm \left\lfloor \frac{1}{d(n\alpha,0)} \right\rfloor n\alpha \in \mathcal{O}(x)$ is ε-dense in $\mathbb{T}$. This shows every orbit is dense.

Definition 1.16. *Let $T\colon X \to X$ be a homeomorphism. We say T is expansive if there exists (the expansive constant) $\delta > 0$ so that for any $x \neq y \in X$, there exists $n \in \mathbb{Z}$ so that $d(T^n x, T^n y) \geqslant \delta$.*

Example 1.17. $T_A\colon \mathbb{T}^2 \to \mathbb{T}^2$ given by matrix multiplication defined by $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ is expansive, topological transitive, mixing and not minimal.

Theorem 1.18 (Multiple Recurrence Theorem)**.** *Let X be a compact metric space and let $T_1, \cdots, T_d\colon X \to X$ be pairwise commutative homeomorphisms. Then there exists some $x_0 \in X$ and some $n_k \nearrow \infty$ as $k \to \infty$ so that $T_j^{n_k} x_0 \to x_0$ for $j = 1, \cdots, d$.*

Proof. We will use induction. For $d = 1$, Corollary 1.13 works. Now we assume the theorem holds for $d-1$. And we also assume X is d-minimal in the sense that $Y \subset X$ closed and $Y = T_1 Y = \cdots = T_d Y$ imply $Y = \varnothing$ or X. This can be done by reapplying the proof of Theorem 1.12.

Denote
$$S = T_1 \times \cdots \times T_d\colon X^d \to X^d,$$
$$\widehat{T_j} = T_j \times \cdots \times T_j\colon X^d \to X^d.$$
$S, \widehat{T_1}, \cdots, \widehat{T_d}$ are pairwise commutative. For $\underline{n} = (n_1, \cdots, n_d) \in \mathbb{Z}^d$, we define
$$T^{\underline{n}} = T_1^{n_1} \circ \cdots \circ T_d^{n_d}\colon X \to X,$$
$$\widehat{T}^{\underline{n}} = \widehat{T_1}^{n_1} \circ \cdots \circ \widehat{T_d}^{n_d}\colon X^d \to X^d.$$
Denote $\Delta(x) = (x, \cdots, x)$ the diagonal element in X^d and $\Delta_X = \{\Delta(x) : x \in X\}$. By commutativity, $\widehat{T}^{\underline{n}}(\Delta(x)) = \Delta(T^{\underline{n}} x)$. We need to prove that there exist some $x_0 \in X$ and $n_k \nearrow \infty$ such that $S^{n_k}(\Delta(x_0)) \to \Delta(x_0)$ as $k \to \infty$.

Claim (A)**.** $\Delta_X \cap \omega_S^+(\Delta_X) \neq \varnothing$, where for the subset $\Delta_X \subset X$, we define $\omega_S^+(\Delta_X) = \{\lim_{k\to\infty} S^{n_k}(\Delta(y_k)) : \Delta(y_k) \in \Delta_X, n_k \nearrow \infty\}$.

Proof of Claim (A). Denote
$$R_1 = T_1 T_d^{-1}, \cdots, R_{d-1} = T_{d-1} T_d^{-1}\colon X \to X.$$
By inductive hypothesis, there exists $x_0 \in X$ and $n_k \nearrow \infty$ so that $R_j^{n_k} x_0 \to x_0$ as $k \to \infty$ for all $j = 1, \cdots, d-1$. Define $y_k = T_d^{n_k} x_0$. For $j < d$, $T_j^{n_k} y_k = R_j^{n_k} x_0 \to x_0$; for $j = d$, $T_d^{n_k} y_k = x_0$, as $k \to \infty$. This means $\Delta(x_0) = \lim_{k\to\infty} S^{n_k}(\Delta(y_k)) \in \Delta_X \cap \omega_S^+(\Delta_X) \neq \varnothing$. □

Claim (B)**.** $\Delta_X \subset \omega_S^+(\Delta_X)$.

Proof of Claim (B). This proof needs the minimality assumption. Denote $Y = \{x \in X : \Delta(x) \in \omega_S^+(\Delta_X)\}$. Because $[x \mapsto \Delta(x)]$ is continuous and $\omega_S^+(\Delta_X)$ is closed by diagonal principle, Y is closed. By Claim (A), $Y \neq \varnothing$. By Proposition 1.11, we only need to prove that $T_j^{\pm 1} Y \subset Y$ for $j = 1, \cdots, d$, then $Y = X$ and the claim follows.

Let $x \in Y$. Then there exists $n_k \nearrow \infty$ and $y_k \in X$ such that $\Delta(x) = \lim_{k\to\infty} S^{n_k}(\Delta(y_k))$. Hence we have

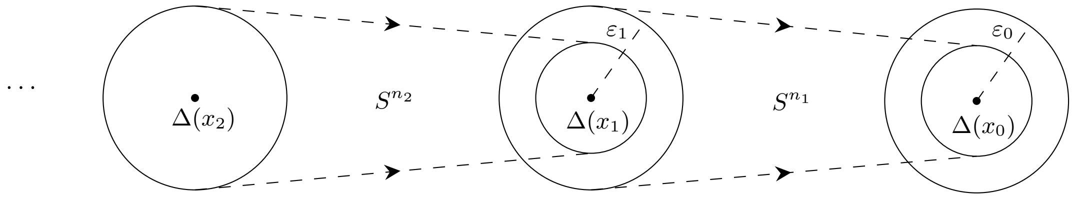
$$\Delta(T_j^{\pm 1}x) = \widehat{T_j}^{\pm 1}(\Delta(x)) = \lim_{k\to\infty} \widehat{T_j}^{\pm 1} S^{n_k}(\Delta(y_k)) = \lim_{k\to\infty} S^{n_k}\Delta(T_j^{\pm 1}y_k) \in \omega_S^+(\Delta_X).$$

This shows $T_j^{\pm 1}x \in Y$. $\square$

Claim (C)**.** For every $\varepsilon > 0$, there exists a point $x \in X$ and some $n \geqslant 1$ so that $d(S^n(\Delta(x)), \Delta(x)) < \varepsilon$.

Proof of Claim (C). Let $x_0 \in X$ be arbitrary and $\varepsilon_0 \in \left(0, \frac{\varepsilon}{2}\right)$. By Claim (B), there exists $x_1 \in X$, $n_1 \geqslant 1$ so that $d(S^{n_1}(\Delta(x_1)), \Delta(x_0)) < \varepsilon_0$. By continuity of S^{n_1} there exists $\varepsilon_1 \in (0, \varepsilon_0)$ so that if $y = (y_1, \cdots, y_d) \in B_{\varepsilon_1}(\Delta(x_1))$, then $S^{n_1}(y) \in B_{\varepsilon_0}(\Delta(x_0))$.

Continuing inductively we find new points $x_k \in X$ and $n_k \geqslant 1$ so that $d(S^{n_k}(\Delta(x_k)), \Delta(x_{k-1})) \leqslant \varepsilon_{k-1}$, where $\varepsilon_k \in (0, \varepsilon_{k-1})$ so that if $y = (y_1, \cdots, y_d) \in B_{\varepsilon_k}(\Delta(x_k))$, then $S^{n_k}(y) \in B_{\varepsilon_{k-1}}(\Delta(x_{k-1}))$.



By compactness of X, we can find $0 \leqslant k < l$ so that $d(\Delta(x_k), \Delta(x_l)) < \frac{\varepsilon}{2}$. Applying S^{n_l} to $\Delta(x_l)$, we obtain a point in $B_{\varepsilon_{l-1}}(\Delta(x_{l-1}))$. By construction, note that $S^{n_l}(\Delta(x_l)) \in B_{\varepsilon_{l-1}}(\Delta(x_{l-1}))$, we obtain $S^{n_{l-1}+n_l}(\Delta(x_l)) \in B_{\varepsilon_{l-2}}(\Delta(x_{l-2}))$. Continuing inductively, we obtain $S^{n_{k+1}+\cdots+n_l}(\Delta(x_l)) \in B_{\varepsilon_k}(\Delta(x_k))$. Then $n = n_{k+1} + \cdots + n_l$ and $x = x_l$ are as desired:

$$d(S^{n_{k+1}+\cdots,n_l}(\Delta(x_l)), \Delta(x_l)) \leqslant d(S^{n_{k+1}+\cdots+n_l}(\Delta(x_l)), \Delta(x_k)) + d(\Delta(x_k), \Delta(x_l)) < \varepsilon_k + \frac{\varepsilon}{2} < \varepsilon.$$

$\square$

We define

$$F(x) = \inf_{n\in\mathbb{N}} d(S^n(\Delta(x)), \Delta(x)).$$

Claim (D)**.** If there exists $x_0 \in X$ such that $F(x_0) = 0$, then the theorem also holds for d maps. Thus the proof will be completed.

Proof of Claim (D). If there exists n so that $d(S^n(\Delta(x)), \Delta(x)) = 0$, then for $n_k = nk$, we have $\Delta^{n_k}(\Delta(x_0)) = \Delta(x_0) \to \Delta(x_0)$. If not, there exists n_k with $S^{n_k}(\Delta(x_0)) \to \Delta(x_0)$. $\square$

To prove the existence of such $x_0 \in X$, we need a lemma:

Lemma 1.19. *Let X be a compact metric space. Let $f_n \colon X \to [0, \infty)$ be a sequence of continuous. We define $F(x) = \inf_{n\in\mathbb{N}} f_n(x)$ for any $x \in X$.*

1. *F is upper semi-continuous: for any $x \in X$ and $\varepsilon > 0$, there exists some $\delta > 0$ so that for any $y \in B_\delta(x)$, we have $F(y) < F(x) + \varepsilon$.*
2. *The sets*
$$\mathcal{J}_\varepsilon := \left\{ x \in X : \forall \eta > 0, \sup_{y,z\in B_\eta(x)} |F(y) - F(z)| > \varepsilon \right\}$$
are closed with empty interior for all $\varepsilon > 0$.
3. *F is continuous on a dense G_δ-set in X.*

Proof of Lemma 1.19. **(1):** By definition, there is some n so that $f_n(x) < F(x)+\frac{\varepsilon}{2}$. Because f_n is continuous, there exists $\delta > 0$ so that for any $y \in B_\delta(x)$, $f_n(y) < f_n(x) + \frac{\varepsilon}{2}$. Now we have

$$F(y) \leqslant f_n(y) < f_n(x) + \frac{\varepsilon}{2} < F(x) + \varepsilon.$$

(2): Let $\overline{x} \in \overline{\mathcal{J}_\varepsilon}$. Let $\eta > 0$. Then there exists $x \in \mathcal{J}_\varepsilon \cap B_{\frac{\eta}{2}}(\overline{x})$. By triangle inequality, $B_{\frac{\eta}{2}}(x) \subset B_\eta(\overline{x})$. Thus by definition,

$$\sup_{y,z\in B_\eta(\overline{x})} |F(y) - F(z)| \geqslant \sup_{y,z\in B_{\frac{\eta}{2}}(x)} |F(y) - F(z)| > \varepsilon.$$

This implies $\overline{x} \in \mathcal{J}_\varepsilon$ and then $\mathcal{J}_\varepsilon$ is closed.

Suppose the there exists some $x_0 \in \mathcal{J}_\varepsilon^\circ$. By (1), there is a $\delta_0 > 0$ so that $F(y) < F(x_0) + \frac{\varepsilon}{2}$ for all $y \in B_{\delta_0}(x_0)$. We choose δ_0 small enough such that $B_{\delta_0}(x_0) \subset \mathcal{J}_\varepsilon$. We claim that we can find a point $x_1 \in B_{\delta_0}(x_0)$ such that $F(x_1) \leqslant F(x_0) - \frac{\varepsilon}{2}$. If not, for any $y \in B_{\delta_0}(x_0)$, $F(y) > F(x_0) - \frac{\varepsilon}{2}$. Associating our choice of δ_0, $|F(y) - F(x_0)| < \frac{\varepsilon}{2}$. Then for any $y, z \in B_{\delta_0}(x_0)$, $|F(y) - F(z)| \leqslant |F(y) - F(x_0)| + |F(x_0 - F(z))| < \varepsilon$. This gives a contradiction of the definition of $\mathcal{J}_\varepsilon$. Now we can repeat this to find $B_{\delta_1}(x_1) \subset B_{\delta_0} \subset \mathcal{J}_\varepsilon$ and $x_2 \in B_{\delta_1}(x_1)$ such that $F(x_2) \leqslant F(x_1) - \frac{\varepsilon}{2} \leqslant F(x_0) - 2 \cdot \frac{\varepsilon}{2}$. Repeating this process, we get a sequence $x_n \in B_{\delta_0}(x_0)$ so that $F(x_n) \leqslant F(x_0) - n \cdot \frac{\varepsilon}{2}$. This contradicts to the fact that $F \geqslant 0$.

(3): By (2), $X - \mathcal{J}_{\frac{1}{n}}$ is open and dense. Thus by Baire Category Theorem, $G := \bigcap_{n\geqslant 0} \left(X - \mathcal{J}_{\frac{1}{n}}\right)$ is a dense G_δ-set. Fix $x \in G$. For any $\varepsilon > 0$, choose n such that $\frac{1}{n} < \varepsilon$. By construction, $x \notin \mathcal{J}_{\frac{1}{n}}$, i.e. there exists $\eta > 0$ so that $\sup_{y,z\in B_\eta(x)} |F(y) - F(z)| < \frac{1}{n} < \varepsilon$. In particular, $|F(y) - F(x)| < \varepsilon$ for all $y \in B_\eta(x)$. Hence F is continuous at x. □

Especially, we have

Claim (E)**.** There exists $x_0 \in X$ so that F is continuous at x_0.

Claim (F)**.** If F is continuous at x_0, then $F(x_0) = 0$.

Proof of Claim (F). Assume $F(x_0) > 0$. Then there is an open neighborhood U of x_0 and $\delta > 0$ such that $F(x) > \delta > 0$ in U. Note that $\widetilde{U} := \bigcup_{\underline{n}\in\mathbb{Z}^d} T^{-\underline{n}}U$ is non-empty and open, and $T_1\widetilde{U} = \cdots = T_d\widetilde{U} = \widetilde{U}$. By minimality and Proposition 1.11, considering the closed set $Y = X - \widetilde{U}$, we obtain $\widetilde{U} = X$. By compactness, there is a finite set $F \subset \mathbb{Z}^d$ such that $X = \bigcup_{\underline{n}\in F} T^{-\underline{n}}U$. By continuity of $\widehat{T}^{\underline{n}}$ for $\underline{n} \in F$ and compactness of X, $\widehat{T}^{\underline{n}}$ is uniform continuous on X. And because F is finite, there exists some $\varepsilon > 0$ so that for all $x, y \in X^d$, $d(x, y) < \varepsilon$, we have $d(\widehat{T}^{\underline{n}}x, \widehat{T}^{\underline{n}}y) < \delta$, for any $\underline{n} \in F$.

By Claim (C), for this $\varepsilon > 0$, we can find an $x_\varepsilon \in X$ with $F(x) < \varepsilon$. Especially, we can also find an $m \geqslant 1$ so that $d(S^m(\Delta(x_\varepsilon), \Delta(x_\varepsilon)) < \varepsilon$. Besides, we can find an $\underline{n} \in F$ so that $x_\varepsilon \in T^{-\underline{n}}U$. Now by continuity of $\widehat{T}^{\underline{n}}$ and commutativity of the maps, we have

$$\delta > d(\widehat{T}^{\underline{n}}(S^m(\Delta(x_\varepsilon))), \widehat{T}^{\underline{n}}(\Delta(x_\varepsilon))) = d(S^m(\Delta(T^{\underline{n}}(x_\varepsilon))), \Delta(T^{\underline{n}}(x_\varepsilon)))$$

for $T^{\underline{n}}(x_\varepsilon) \in U$. Recall that $F > \delta$ on U. We get a contradiction. □

□

Remark 1.20. If T_j are not commutative, it fails. For example, consider the North-South dynamics and "East-West" dynamics on $\mathbb{S}^1$.

As corollary, we have:

Theorem 1.21 (van der Werden)**.** *Let $\mathbb{Z} = B_1 \sqcup B_2 \sqcup \cdots \sqcup B_k$ be a finite partition. Then there exists $B = B_j$ such that it contains arbitrarily long arithmetic progressions: for arbitrary N, there exist some $a, d \in \mathbb{Z}$ such that*

$$a, a + d, a + 2d, \cdots, a + Nd \in B$$

Proof. To prove this theorem, we need to construct a related dynamical system.

Let $X_{full} = \{1, \cdots, k\}^{\mathbb{Z}}$. The full shift is defined as $\sigma\colon X_{full} \to X_{full}$, $(\sigma(x))_n = x_{n+1}$.

We can define a metric on X_{full}:

$$d(x, y) := \begin{cases} 0, & x = y; \\ \frac{1}{n+1}, & n = \min\{|k| : x_k \neq y_k\} \end{cases}.$$

Claim. X_{full} is compact under this metric.

Proof of Claim. Note that d induces the standard product topology. Then Tychonoff's Theorem gives the claim. $\square$

Now we turn to famous Furstenberg's correspondence. We define $z \in X_{full}$ by $z_n = j$ if $n \in B_j$ and then $X := \overline{\mathcal{O}_\sigma(z)} \subset X_{full}$. Consider the shift map $\sigma = \sigma|_X : X \to X$. Note that X is closed in a compact space, it's compact. Hence σ is a homeomorphism on a compact metric space X. Moreover, we define $T_1 = \sigma, \cdots, T_N = \sigma^N$.

By multiple recurrence theorem, there exists some $x \in X$ and some $n_l \nearrow \infty$ such that $T_1^{n_l}x \to x, \cdots, T_N^{n_l}x \to x$, i.e. $\sigma^{n_l}x \to x, \cdots, \sigma^{Nn_l}x \to x$ as $l \to \infty$. Then for $\varepsilon = 1$, we can find an n_l such that $d(\sigma^{n_l}x, x) < 1, \cdots, d(\sigma^{Nn_l}x, x) < 1$. Denote $d = n_l$. By definition, this happens if and only if $x_d = x_0, \cdots, x_{Nd} = x_0$.

Now we work with the facts that $x_0 = x_d = \cdots = x_{Nd}$ and $x \in X = \overline{\mathcal{O}_\sigma(z)}$. There exists some $a \in \mathbb{Z}$ such that $d(x, \sigma^a z) < \frac{1}{Nd+1}$. By definition, x and $\sigma^a z$ have the same symbols at coordinates $-Nd, \cdots, 0, \cdots, Nd$. Therefore,

$$\begin{tikzcd}
x_0 \arrow[r, equal] \arrow[d, equal] & x_d \arrow[r, equal] \arrow[d, equal] & \cdots \arrow[r, equal] \arrow[d, equal] & x_{Nd} \arrow[d, equal] \\
z_a & z_{a+d} & \cdots & z_{a+Nd}
\end{tikzcd}$$

This forces,

$$\begin{tikzcd}
x_0 \arrow[r, equal] \arrow[d, equal] & x_d \arrow[r, equal] \arrow[d, equal] & \cdots \arrow[r, equal] \arrow[d, equal] & x_{Nd} \arrow[d, equal] & \\
z_a \arrow[r, equal] & z_{a+d} \arrow[r, equal] & \cdots \arrow[r, equal] & z_{a+Nd} \arrow[r, equal] & j
\end{tikzcd}$$

for a $j \in \{1, \cdots, k\}$. Hence $a, a+d, \cdots, a+Nd \in B_j$, which are as desired.

$\square$

2 Symbolic Dynamics

Recall that the full shift on a finite alphabet $\mathcal{A} = \{1, \cdots, k\}$ is defined on $X_{full} = \mathcal{A}^{\mathbb{Z}}$ by $(\sigma(x))_n = x_{n+1}$. This defines a homeomorphism $\sigma: X_{full} \to X_{full}$ on a compact metric space. A shift space is a σ-invariant closed subset $X \subset X_{full}$ together with $\sigma = \sigma|_X : X \to X$.

Definition 2.1. *A cylinder set in X_{full} or X is defined by*

$$[w]_{m,n} = \{x \in X : w_m = x_m, \cdots, w_n = x_n\}$$

where $w \in X$ and $m \leqslant n$.

Proposition 2.2. *Cylinder sets are compact and open.*

Proof. For compactness, note that $[w]_{m,n} \cong \mathcal{A}^{((-\infty, m-1] \sqcup [n+1, +\infty)) \cap \mathbb{Z}}$ and Tychonoff's theorem works. For openness, note that if $m = -n$, then $[w]_{-n,n}$ is an open ball in X. Thus in general case, $[w]_{m,n} = \bigcup_{v \in [w]_{m,n}} [v]_{-N,N}$, where $N = \max\{|m|, |n|\}$, is open. $\square$

Example 2.3. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an oriented finite graph. We can define the vertex shift by

$$X_{\mathcal{G}} = \left\{x \in \mathcal{V}^{\mathbb{Z}} : \text{for any } n,\ x_n \text{ connects to } x_{n+1} \text{ by an edge}\right\}.$$

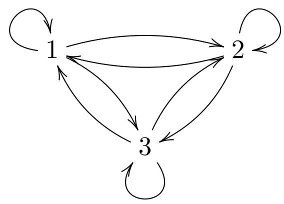
- Possibly $X_{\mathcal{G}} = \varnothing$, e.g.

$$\mathcal{G} = \begin{tikzcd} 1 \arrow[r] & 2 \arrow[r] & 3 \end{tikzcd}$$

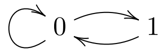
- Possibly $X_{\mathcal{G}}$ is finite, e.g.

$$\mathcal{G} = \begin{tikzcd} 1 \arrow[rr] & & 2 \arrow[dl] \\ & 3 \arrow[ul] & \end{tikzcd}$$

- Full shift is a vertex shift, .e.g.

$\mathcal{G} =$ 

- Golden mean shift:

$\mathcal{G} =$ 

The adjacency matrix $A = A_{\mathcal{G}}$ is defined by

$$A_{ij} = \begin{cases} 1, & i \text{ connectes to } i; \\ 0, & \text{otherwise} \end{cases}.$$

Lemma 2.4. *1. $(A^n)_{ij}$ is the number of paths from i to j. $\operatorname{Tr}(A^n)$ is the number of periodic points in $X_{\mathcal{G}}$ of period n or a divisor of n.*

2. If $\mathcal{G}$ is connected, i.e. $\forall i, j$, $\exists n \geqslant 1$ such that $(A^n)_{ij} > 0$, then $X_{\mathcal{G}}$ is topological transitive.

3. $\mathcal{G}$ is connected and aperiodic, i.e. there exists n with $(A^n)_{ij} > 0$ for all i, j, if and only if $X_{\mathcal{G}}$ is topological mixing.

Proof. **(1):** We use induction. For $n = 1$, the proof is by definition. Suppose the claim holds for n. Then

$$(A^{n+1})_{ij} = \sum_l A_{il}(A^n)_{lj}$$

where A_{il} is depending on whether l connects i and $(A^n)_{lj}$ is the number of paths from j to l. The claim follows.

Now $\operatorname{Tr}(A^n) = \sum_i (A^n)_{ii}$ is the number of closed paths with identified starting point, which equals to the number of periodic points with period n or a divisor of n.

(2): Suppose $U, V \subset X_{\mathcal{G}}$ be non-empty open subsets of the vertex shift. We wish to find some n such that $\sigma^{-n}(U) \cap V \neq \varnothing$. Then by Proposition 1.5, the claim follows.

We may assume $[w]_{-N,N} \subset U$ and $[v]_{-N,N} \subset V$. Denote $j = w_N \in \{1, \cdots, k\}$ and $i = v_{-N}$. Then by connectedness there exists some $l \geqslant 1$ such that $(A^l)_{ij} > 0$. This means it's possible to go from j to i in l steps. Denote $u = (u_0 = j, u_1, \cdots, u_{l-1}, u_l = i)$. Then we define

$$\begin{aligned} x = (\cdots, x_{-3N-l-1} = w_{-N-1}, x_{-3N-l} = w_{-N}, \cdots, x_{-N-l} = w_N = j = u_0, \\ x_{-N-l+1} = u_1, \cdots, x_{-N} = u_l = i = v_{-N}, \\ x_{-N+1} = v_{-N+1}, \cdots). \end{aligned}$$

In fact, we first go along w until $x_{-N-l} = w_N = u_0$, then we go along u, finally we go along v from $x_{-N} = u_l = v_{-N}$. It's easy to check that $x \in \sigma^{2N+l}([w]_{-N,N}) \cap [v]_{-N,N} \subset \sigma^{2N+l}(U) \cap V \neq \varnothing$. Thus $X_{\mathcal{G}}$ is topological transitive.

(3): (⇒): The proof is similar to (2).

Claim. For any $m \geqslant n$, $(A^m)_{ij} > 0$ for any i, j.

Proof of Claim. Suppose $(A^{m+1})_{ij} = 0$ and $(A^m)_{ij} > 0$ for any i, j. Then $(A^{m+1})_{ij} = \sum_l A_{il}(A^m)_{lj}$ shows that $A_{il} = 0$ for any l. Hence $(A^m)_{il} = 0$ for any m and i by induction. We get a contradiction. □

For any two open sets U, V, we may assume $[w]_{-N,N} \subset U$ and $[v]_{-N,N} \subset V$. For any $m \geqslant n$, we can go from w_N to v_{-N} in m steps along some path u. We first go along w until $x_{-N-l} = w_N = u_0$, then we go along u, finally we go along v from $x_{-N} = u_m = v_{-N}$. It's easy to check that $x \in \sigma^{2N+m}([w]_{-N,N}) \cap [v]_{-N,N} \subset \sigma^{2N+m}(U) \cap V$. It follows that for any $k > 2N + n$, $\sigma^k(U) \cap V \neq \varnothing$. By altering the roles of U, V, we can prove $\sigma^{-k}(U) \cap V = \sigma^{-k}(U \cap \sigma^k(V)) \neq \sigma^{-k}(\varnothing) = \varnothing$ for any $k > 2N + n$. Above all, for any $|k| > 2N + n$, $\sigma^k(U) \cap V \neq \varnothing$. Thus $X_{\mathcal{G}}$ is topological mixing.

(⇐): Note that $[(\cdots, i, i, i, \cdots)]_{0,0}$, $i \in \mathcal{V}$ are open balls. Then because $\mathcal{V}$ is finite, there is a universal constant N such that for any $|n| > N$, $\sigma^n([(\cdots, i, \cdots)]_{0,0}) \cap [(\cdots, j, \cdots)]_{0,0} \neq \varnothing$ for any $i, j \in \mathcal{V}$. By above construction, this means for any $|n| > N$, any $i, j \in \mathcal{V}$ can be connected by a path in n steps, i.e. $(A^n)_{ij} > 0$. Thus $\mathcal{G}$ is connected and aperiodic. □

Definition 2.5. *A shift of finite type (sft) is a (closed shift-invariant) subset $X \subset X_{full} = \mathcal{A}^{\mathbb{Z}}$ defined by a finite list of forbidden finite words. More precisely, there should exist N and a finite set $\mathcal{F} \subset \mathcal{A}^{\{1,\cdots,N\}}$ so that $X = \left\{x \in \mathcal{A}^{\mathbb{Z}} : \forall n, (x_{n+1}, \cdots, x_{n+N}) \notin \mathcal{F}\right\}$.*

Definition 2.6. *$X \subset \mathcal{A}^{\mathbb{Z}}$ is called a sofic system if there exists a shift of finite type $Y \subset \mathcal{B}^{\mathbb{Z}}$ and a continuous map $\Phi \colon Y \to X$ with $\Phi(Y) = X$ and $\Phi \circ \sigma_Y = \sigma_X \circ \Phi$.*

Lemma 2.7. *For any shift of finite type X, there exists a vertex shift $X_{\mathcal{G}}$ such that $X_{\mathcal{G}}$ is sofic to X.*

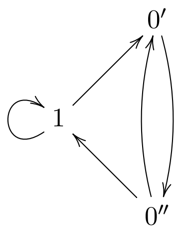
Proof. Suppose X is defined by $\mathcal{F} \subset \mathcal{A}^{\{1,\cdots,N+1\}}$. We define $\mathcal{V} = \mathcal{A}^{\{1,\cdots,N\}}$. We connect $v, v' \in \mathcal{V}$ if there exists $w \in \mathcal{A}^{\{1,\cdots,N+1\}}$ so that $w \notin \mathcal{F}$ and $w = (v, w_{N+1}) = (w_1, v')$. This defines a graph $\mathcal{G}$ and so also a vertex shift $X_{\mathcal{G}}$.

We define a map $\Phi \colon X \to X_{\mathcal{G}}$ by $x \mapsto (\Phi(x))_n = (x_n, \cdots, x_{n+N-1})$. By construction, x doesn't have forbidden words in $\mathcal{F}$. Thus $(\Phi(x))_n = (x_n, \cdots, x_{n+N-1})$ connects $(\Phi(x))_{n+1} = (x_{n+1}, \cdots, x_{n+N})$. This shows that Φ is well-defined. Moreover, $\Phi(X) = X_{\mathcal{G}}$. To see this, let $v \in X_{\mathcal{G}}$. By definition, there exists $x \in X_{full}$ so that $v_n = (x_n, \cdots, x_{n+N-1})$. It's easy to check that $x \in X$ by definition. Moreover, it's also easy to check that Φ is a homeomorphism and $\Phi \circ \sigma_X = \sigma_{X_{\mathcal{G}}} \circ \Phi$. $\square$

Example 2.8. The even shift $X_{even} \subset \{0,1\}^{\mathbb{Z}}$ is sofic but bot of finite type, where

$$X_{even} = \left\{x \in \{0,1\}^{\mathbb{Z}} : \text{any two 1's in the sequence are separated by an even number of 0's}\right\}.$$

Let Y be the vertex shift defined by



and Φ forget the primes over 0's. Hence we have defined $\Phi \colon Y \to X_{even}$.

Definition 2.9 (Complexity). *Let $X \subset X_{full} = \mathcal{A}^{\mathbb{Z}}$ be a shift. We define the complexity function*

$$p_X(n) := \left|\pi_{\{1,\cdots,n\}}(X)\right| := \textit{the number of words of length } n \textit{ appearing in any } x \in X.$$

Lemma 2.10. *If X is a shift of finite type, then $p_X(n)$ either grows polynomially (is constant in case X is transitive) or grows exponentially. If X is a topological mixing vertex shift, then either $|X| = 1$ or $p_X(n)$ grows exponentially.*

Proof. By Lemma 2.7, there exists a vertex shift $x_{\mathcal{G}}$ such that $X_{\mathcal{G}}$ is sofic to X. It's easy to check that p_X and $p_{X_{\mathcal{G}}}$ have the same growth type. Thus we can assume that $X = X_{\mathcal{G}}$ for simplicity. Moreover, we can remove all sinks and sources of $\mathcal{G}$ because $X = X_{\mathcal{G}}$ consists of bi-infinite paths. We also assume that any finite path in $\mathcal{G}$ can be extended to a bi-infinite path, i.e. a point in $X = X_{\mathcal{G}}$. Let A be the adjacent matrix of $\mathcal{G}$. By Lemma 2.4 (1),

$$p_X(n) = \sum_{i,j} (A^n)_{ij} \asymp \sum_{i,j} \left(TA^nT^{-1}\right)_{ij},$$

where T takes A to its Jordan normal form by conjugation. For a dimension k Jordan block $\lambda I + N$, we can calculate that $\sum_{i,j}((\lambda I + N)^n)_{ij} = \sum_{d=0}^{k-1}(k-d)\binom{n}{d}\lambda^{n-d}$. Thus if the eigenvalues of A are all 0 or 1, p_X grows polynomially, otherwise p_X grows exponentially.

Suppose $X = X_{\mathcal{G}}$ is topological mixing. By 2.4 (3), $\mathcal{G}$ is connected and aperiodic. Then there exists an n such that $(A^n)_{ij} > 0$ for any $i, j \in \mathcal{V}$. Unless A is the 1-all matrix, p_X grows exponentially. (The professor didn't give an explicit proof. But I think the Perron-Frobenius Theorem works.) $\square$

Theorem 2.11 (Morse-Hedlund). *Let X be a shift space. Then $|X| < \infty$ if and only if there exists some $n \in \mathbb{N}$ so that $p_X(n) \leqslant n$.*

Proof. **($\Rightarrow$):** We first claim that any $x \in X$ is periodic. If not, $\infty = |\mathcal{O}(x)| \leqslant |X|$. We get a contradiction. Suppose x has period N. For any $n > N$, words with length n must contain some copies of the period words and be added by some finite choices of symbols at the front or back. In fact, we have $p_x(n) = p_x(n+N) = p_x(n+2N) = \cdots$. Thus there exists C_x such that $p_x \leqslant C_x$. Because $|X| < \infty$, there exists an universal constant C such that $p_X \leqslant \sum_{x \in X} p_x \leqslant C$. Then the proof follows.

($\Leftarrow$): Let n be the minimal number such that $p_X(n) \leqslant n$. If $n = 1$, $p_X(1) = 1$ and so $|X| = 1$. The theorem follows. Suppose $n > 1$. We have

$$n - 1 < p_X(n-1) \leqslant p_X(n) \leqslant n.$$

This shows that $p_X(n-1) = p_X(n) = n$. Denote $\mathcal{L}_n = \pi_{\{1,\cdots,n\}}(X) = \{w_1, \cdots, w_n\}$. We define $L\colon \mathcal{L}_n \to \mathcal{L}_{n-1}$ by forgetting the last symbol and $R\colon \mathcal{L}_n \to \mathcal{L}_{n-1}$ by forgetting the first symbol. Note that any word in $\mathcal{L}_{n-1}$ is the last or first of a length n word. Thus L, R are surjective. Note that $|\mathcal{L}_n| = \mathcal{L}_{n-1} = n$, L, R are also injective.

This shows that for any $w \in \mathcal{L}_{n-1}$, there is only one way of adding a symbol to the right or left to obtain a word in $\mathcal{L}_n$. Explicitly, for $w = (w_1, \cdots, w_{n-1})$, there exists unique symbol w_n such that $(w_1, \cdots, w_n) \in \mathcal{L}_n$. Then considering $(w_2, \cdots, w_n)$, we get unique w_{n+1} such that $(w_2, \cdots, w_{n+1}) \in \mathcal{L}_n$. Iterating this process and doing the same things on the left, we get an unique $x \in X$ such that $(x_1, \cdots, x_{n-1}) = w = (w_1, \cdots, w_{n-1})$. This shows that $|X| = |\mathcal{L}_{n-1}| = n$. To see this, for any $x \in X$, $(x_1, \cdots, x_{n-1}) \in \mathcal{L}_{n-1}$. And by uniqueness, if $x, x' \in X$ are with $(x_1, \cdots, x_{n-1}) = (x'_1, \cdots, x'_{n-1})$, then $x = x'$. $\square$

Definition 2.12. *X is Sturmian if $p_X(n) = n + 1$ for all $n \in \mathbb{N}$.*

Example 2.13. $X_{\mathcal{G}}$ is Sturmian where

$$\mathcal{G} = \begin{tikzcd} 0 \arrow[loop left] \arrow[r] & 1 \arrow[loop right] \end{tikzcd}$$

Example 2.14. Let $\alpha \notin \mathbb{Q}$. We define $R_\alpha\colon \mathbb{T} \to \mathbb{T}$ by $x \mapsto x + \alpha$. Consider two intervals $J_1 = [0, 1-\alpha)$ and $J_2 = [1-\alpha, 1)$ as subsets of $\mathbb{T}$. We called a word $\underline{w} \in \{1,2\}^n$ allowed if $J_{\underline{w}} := J_{w_1} \cap R_\alpha^{-1}(J_{w_2}) \cap \cdots \cap R_\alpha^{-(n-1)}(J_{w_n}) \neq \varnothing$.

Claim. There are precisely $n+1$ allowed words of length n, the corresponding sets $J_{\underline{w}}$ are half-open intervals, the end points of these intervals are precisely $\{0, -\alpha, -2\alpha, \cdots, -n\alpha\}$.

Proof of Claim. For $n = 1$, the claim holds. We now assume that the claim holds for n. Let $\underline{w}$ be an allowed word with $J_{\underline{w}} = [a, b) \subset \mathbb{T}$. Note that

$$J_{\underline{w}} = \left\{x \in \mathbb{T} \text{ for which } \underline{w} \text{ describes the locations of } R_\alpha^j(x) \text{ for } j = 0, \cdots, n-1 \text{ with respect to } J_1 \text{ or } J_2\right\}.$$

The question of how the allowed word $\underline{w}$ extends corresponds to the question if $J_{\underline{w}} \supsetneqq J_{\underline{w}a}$ for $a \in \{1,2\}$.

- If $(n+1)\alpha \notin J_{\underline{w}}$, then $J_{\underline{w}} = J_{\underline{w}a}$ for some $a \in \{1,2\}$.
- If $(n+1)\alpha \in J_{\underline{w}}$, then $\underline{w}$ extends in two ways to allowed words.

$\square$

Hence the space of allowed words with the shift map σ is a Sturmian system.

3 Ergodic Theory

In the following, let $(X, \mathcal{B})$ be a measurable space, i.e. X is a set and $\mathcal{B}$ is a σ-algebra on X. Moreover, let μ be a probability measure on $(X, \mathcal{B})$.

Definition 3.1. *Let $(X, \mathcal{B}, \mu)$ be a probability space. Let $T\colon X \to X$ be measurable. Then T is called measure-preserving or μ is called T-invariant if $\mu(T^{-1}B) = \mu(B)$ for every $B \in \mathcal{B}$.*

Definition 3.2. *If $T\colon (X, \mathcal{B}) \to (Y, \mathcal{C})$ is measurable and μ is a measure on X, then the push forward measure $T_*\mu$ on Y is defined by*

$$(T_*\mu)(C) = u(T^{-1}C)$$

for all $C \in \mathcal{C}$.

Lemma 3.3. *If $f \geqslant 0$ is measurable on Y, then*

$$\int_Y f \,\mathrm{d}(T_*\mu) = \int_X f \circ T \,\mathrm{d}\mu.$$

Proof. Check the formula for simple positive functions. Then approach general non-negative functions by simple positive functions by monotone convergence theorem. $\square$

Theorem 3.4 (Poincaré Recurrence). *Let $T\colon (X,\mathcal{B},\mu) \to (X,\mathcal{B},\mu)$ be measure-preserving on a probability space. Let $E \subset X$ be measurable. Then for μ-a.e. $x \in E$, there exist infinitely many returns to E: there exists $n_k \nearrow \infty$ so that $T^{n_k}(x) \in E$.*

Proof. We define

$$A_m = \bigcup_{n \geqslant m} T^{-n}E = \{x \in X : x \text{ has a visit to } E \text{ at time } m \text{ or later}\}$$

and

$$A = \bigcap_{m=0}^{\infty} = \{x \in X : x \text{ has infinitely many visits to } E\}.$$

We note that $E \subset A_0$, $A_0 \supset A_1 \supset A_2 \supset \cdots \supset A$. And note that for any $m \geqslant 0$, $T^{-1}A_m = A_{m+1}$, we have $\mu(A_m) = \mu(X_{m+1})$. This implies that $\mu(A_m - A_{m+1}) = 0$. Then

$$\mu(A_0 - A) = \mu\left(\bigsqcup_{m=0}^{\infty}(A_m - A_{m+1})\right) \leqslant \sum_{m=0}^{\infty} \mu(A_m - A_{m+1}) = 0.$$

Associating with $E \subset A_0$, we have $E \overset{a.e.}{\subset} A$, which is as desired. □

Corollary 3.5. *Let X be a compact metric space, $\mathcal{B}$ be the Borel-σ-algebra on X, μ be a probability measure on $\mathcal{B}$, $T\colon X \to X$ be measure-preserving. Then μ-a.e. $x \in X$ is recurrent in the sense of topological dynamics: there exists $n_k \nearrow \infty$ so that $T^{n_k}(x) \to x$.*

Proof. As X is compact, it can be covered by finitely many balls of radius $\frac{1}{n}$ for every $n \in \mathbb{N}$. We apply Poincaré Recurrence to each of these countably many balls and denote the collection of the null sets of non-recurrence points by Y. Y is also a null set.

Let $x \in X - Y$. Then x belongs to one of the ball of radius 1. By recurrence, there exists n_1 so that $T^{n_1}(x)$ belongs to this ball. Hence we have $d(T^{n_1}(x), x) \leqslant 2 \cdot 1$. Repeating this process for balls of radius $\frac{1}{2}$, we get an $n_2 > n_1$ such that $d(T^{n_2}(x), x) \leqslant 2 \cdot \frac{1}{2}$. Repeating this process again and again, we get a sequence $n_k \nearrow \infty$ such that $T^{n_k}(x) \to x$, which is as desired. □

Example 3.6. $R_\alpha\colon \mathbb{T} \to \mathbb{T}$, $x \mapsto x + \alpha$ is measure-preserving under Lebesgue measure λ on $\mathbb{T} \cong [0,1)$. If $\alpha \in \mathbb{Q}$, other measures exists. If $\alpha \notin \mathbb{Q}$, λ is the only one. See below.

Example 3.7. $T_p\colon \mathbb{T} \to \mathbb{T}$, $x \mapsto px$ is measure-preserving for integer $p \geqslant 1$ under λ by some real analysis techniques.

Example 3.8. $T\colon \mathbb{T}^2 \to \mathbb{T}^2$ defined by $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ is measure-preserving with respect to Lebesgue measure λ on $[0,1)^2 \cong \mathbb{T}^2$, because the Jacobian $= |\det A| = 1$ and T is a bijection.

Example 3.9. $T\colon \mathbb{T}^2 \to \mathbb{T}^2$, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 + \alpha \\ x_2 \end{pmatrix}$ is measure-preserving.

Example 3.10. $T\colon \mathbb{T}^2 \to \mathbb{T}^2$, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix}$ is measure-preserving.

Definition 3.11. *Let $T\colon X \to X$ be measure-preserving. Then $U_T\colon L^2_\mu(X) \to L^2_\mu(X)$, $f \mapsto f \circ T$ is called the Koopman operator.*

Proposition 3.12. *U_T is linear, norm-preserving and preserves inner products in $L^2_\mu(X)$.*

Theorem 3.13 (Von Newmann mean ergodic theorem). *Let $(X,\mathcal{B},\mu)$ be a probability space and $T\colon X \to X$ be measure-preserving. Then for any $f \in L^2_\mu(X)$, we have*

$$A_n(f) := \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \xrightarrow{L^2_\mu} P_T(f)$$

where P_T is the orthogonal projection from $L^2_\mu(X)$ onto the subspace $\mathcal{I}_T = \{f \in L^2_\mu(X) : f \circ T = f\}$.

Proof. First. for $f \in \mathcal{I}_T$, the claim is obvious. Denote $B := \{h \circ T - h : h \in L^2_\mu(X)\}$. Note that, for any $h \in L^2_\mu(X)$, because T is measure-preserving, we have

$$\begin{aligned}\|A_n(h \circ T - h)\|_{L^2_\mu} &= \left\| \frac{1}{n} \sum_{k=0}^{n-1} \left(h \circ T^{k+1} - h \circ T^k\right) \right\|_{L^2_\mu} = \frac{1}{n} \|(h \circ T^n - h)\|_{L^2_\mu} \\ &\leqslant \frac{1}{n} \left(\|h \circ T^n\|_{L^2_\mu} + \|h\|_{L^2_\mu} \right) = \frac{2\|h\|_{L^2_\mu}}{n} \to 0,\end{aligned} \tag{1}$$

i.e. $A_n(h \circ T - h) \xrightarrow{L^2_\mu} 0$. This help us to guess:

Claim. $B^\perp = \mathcal{I}_T$.

Proof of Claim. For any $h \in L^2_\mu(X)$ and $f \in \mathcal{I}_T$, we have

$$\langle h \circ T - h, f \rangle = \langle h \circ T, f \circ T \rangle - \langle h, f \rangle = 0.$$

This shows $\mathcal{I}_T \subset B^\perp$.

Suppose $f \in B^\perp$ and $h \in L^2_\mu(X)$. Then

$$\langle h \circ T - h, f \rangle = 0 \Rightarrow \langle U_T(h), f \rangle = \langle h, f \rangle \Rightarrow \langle h, U_T^*(f) \rangle = \langle h, f \rangle.$$

This implies $U_T^* = f$. Hence we have

$$\begin{aligned}\|f - U_T(f)\|^2_{L^2_\mu} &= \langle f - U_T(f), f - U_T(f) \rangle \\ &= \langle f, f \rangle + \langle U_T(f), U_T(f) \rangle - \langle f, U_T(f) \rangle - \langle U_T(f), f \rangle \\ &= 2\|f\|^2 - \langle U_T^*(f), f \rangle - \langle f, U_T^*(f) \rangle = 0.\end{aligned}$$

This shows that $f \circ T = U_T(f) = f$ and then $f \in \mathcal{I}_T$. Hence $B^\perp \subset \mathcal{I}_T$. □

Note that $\mathcal{I}_T$ is the preimage of a continuous linear functional $[f \mapsto f \circ T - f]$ of $\{0\}$, it is closed. Hence in Hilbert spaces, we have $L^2_\mu(X) = \mathcal{I}_T \oplus \overline{B}$. For any $f \in L^2_\mu(X)$, we can find $f^* = P_T(f)$ and $g \in \overline{B}$ so that $f = f^* + g$. It follows that

$$\|A_n(f) - f^*\|_{L^2_\mu} = \|A_n(f^* + g) - f^*\|_{L^2_\mu} = \|A_n(f^*) - f^* + A_n(g)\|_{L^2_\mu} = \|A_n(g)\|_{L^2_\mu}.$$

For any $\varepsilon > 0$, we can find some function in B which is close to g, explicitly, we can find some $h \in L^2_\mu(X)$ so that $\|g - (h \circ T - h)\|_{L^2_\mu} < \varepsilon$. Now we have

$$\|A_n(g)\|_{L^2_\mu} = \|A_n(g) - A_n(h \circ T - h) + A_n(h \circ T - h)\|_{L^2_\mu} \leqslant \|A_n(g - (h \circ T - h))\|_{L^2_\mu} + \|A_n(h \circ T - h)\|_{L^2_\mu}.$$

For the first item,

$$\begin{aligned}\|A_n(g - (h \circ T - h))\|_{L^2_\mu} &= \left\| \frac{1}{n} \sum_{k=0}^{n-1} (h \circ T - h) \circ T^k \right\|_{L^2_\mu} \\ &\leqslant \frac{1}{n} \sum_{k=0}^{n-1} \left\|(g \circ T - h) \circ T^k\right\|_{L^2_\mu} = \frac{1}{n} \sum_{k=0}^{n-1} \|(g \circ T - h)\|_{L^2_\mu} < \varepsilon.\end{aligned}$$

For the second item, by equation (1), $\|A_n(h \circ T - h)\|_{L^2_\mu} < \varepsilon$ for n large enough. Above all, we have

$$\|A_n(f) - f^*\|_{L^2_\mu} = \|A_n(g)\|_{L^2_\mu} \leqslant \|A_n(g - (h \circ T - h))\|_{L^2_\mu} + \|A_n(h \circ T - h)\|_{L^2_\mu} < \varepsilon + \varepsilon = 2\varepsilon$$

for n large enough, i.e. $A_n(f) \xrightarrow{L^2_\mu} f^* = P_T(f)$, which is as desired. □

Proposition 3.14. *Let $(X, \mathcal{B}, \mu)$ be a probability space and $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. Then there exists unique linear operator $E(\cdot|\mathcal{A}) \colon L^1_\mu(X, \mathcal{B}) \to L^1_\mu(X, \mathcal{A})$ so that*

1. *$E(f|\mathcal{A})$ is $\mathcal{A}$-measurable and*
$$\int_A E(f|\mathcal{A}) \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu$$
for any $f \in L^1_\mu(X, \mathcal{B})$ and $A \in \mathcal{A}$,

2. *For any* $f \geqslant 0 \in L^1_\mu(\mathcal{B})$*,* $E(f|\mathcal{A}) \overset{a.s.}{\geqslant} 0$*.*

3. $\|E(f|\mathcal{A})\|_{L^1_\mu} \leqslant \|f\|_{L^1_\mu}$ *for any* $f \in L^1_\mu(X)$*.*

4. *For any* $g \in L^\infty_\mu(\mathcal{A})$ *and* $f \in L^2_\mu(\mathcal{B})$*,* $E(gf|\mathcal{A}) \overset{a.s.}{=} gE(f|\mathcal{A})$*.*

Definition 3.15. $E(\cdot|\mathcal{A})$ *is called the conditional expectation given* $\mathcal{A}$*.*

Example 3.16. Maybe $|\mathcal{A}| < \infty$. For example, $X = X_1 \sqcup \cdots \sqcup X_n$ has a finite partition. Then

$$E(f|\mathcal{A}) = \frac{1}{\mu(X_i)} \int_{X_i} f \, \mathrm{d}\mu$$

where $X_i \in \mathcal{A}$ is minimal with $x \in X_i$.

Example 3.17. Let $X = [0,1]^2$ and $\mathcal{A} = \mathcal{B}_{[0,1]} \times \{\varnothing, [0,1]\}$ and μ be the usual Lebesgue measure on $[0,1]^2$. Then

$$E(f|\mathcal{A})(x_1, x_2) = \int_0^1 f(x_1, y) \, \mathrm{d}y.$$

Example 3.18. Let $X = [0,1]^2$ and $\mathcal{A} = \mathcal{B}_{[0,1]} \times \{\varnothing, [0,1]\}$ and μ be the Lebesgue measure on the diagonal. Then

$$E(f|\mathcal{A})(x_1, x_2) = f(x_1, x_1).$$

In measurable situations one can show that $E(f|\mathcal{A})$ is an average of values of f over a part of the space that depends on X with respect to a probability measure on that part that depends on that part and on μ. (I can't understand this sentence, so I just copied it here.)

Proof of Proposition 3.14. We first prove the uniqueness in the sense of a.s.. Assume the existence. Suppose There exists another desired operator $E'(\cdot|\mathcal{A})$. For any $f \in L^1_\mu(\mathcal{B})$, denote $h_1 = E(f|\mathcal{A})$ and $h_2 = E'(f|\mathcal{A})$. By (1), we have $h_1, h_2 \in L^1_\mu(\mathcal{A})$ and

$$\int_A h_1 \, \mathrm{d}\mu = \int_A E(f|\mathcal{A}) \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu = \int_A E'(f|\mathcal{A}) \, \mathrm{d}\mu = \int_A h_2 \, \mathrm{d}\mu \tag{2}$$

for any $A \in \mathcal{A}$. We define $A = \{x \in X : h_2(x) > h_1(x)\} \in \mathcal{A}$ because it is the preimage of a $\mathcal{A}$-measurable function $h_2 - h_1$ of the measurable set $(0, \infty)$. We have

$$\int_A (h_2 - h_1) \, \mathrm{d}\mu \begin{cases} = 0, & \mu(A) = 0; \\ > 0, & \mu(A) > 0 \end{cases}.$$

By equation (2), we must have $\int_A (h_2 - h_1) \, \mathrm{d}\mu = 0$ and then $\mu(A) = 0$, i.e. $h_2 \overset{a.s.}{\leqslant} h_1$. Similarly, we have $h_1 \overset{a.s.}{\geqslant} h_2$. Hence we have $h_1 \overset{a.s.}{=} h_2$, i.e. $E(f|\mathcal{A}) \overset{a.s.}{=} E'(f|\mathcal{A})$ for any $f \in L^1_\mu(\mathcal{B})$. The uniqueness follows.

In fact, suppose there exists an operator which just satisfies property (1), then property (2) (3) (4) can be deduced. For (2), let $f \geqslant 0 \in L^1_\mu(\mathcal{B})$. Denote $A = \{x \in X : E(f|\mathcal{A}) < 0\} \in \mathcal{A}$. Then

$$0 \geqslant \int_A E(f|\mathcal{A}) \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu \geqslant 0.$$

Hence $\int_A E(f|\mathcal{A}) \, \mathrm{d}\mu = 0$. Note that $E(f|\mathcal{A}) < 0$ on A, we must have $\mu(A) = 0$ and $E(f|\mathcal{A}) \overset{a.s.}{\geqslant} 0$ follows. For (3), let $f \in L^1_\mu(\mathcal{B})$. Then

$$\begin{aligned} \|E(f|\mathcal{A})\|_{L^1_\mu} = \int_X |E(f|\mathcal{A})| \, \mathrm{d}\mu &= \int_{E(f|\mathcal{A})>0} E(f|\mathcal{A}) \, \mathrm{d}\mu + \int_{E(f|\mathcal{A})<0} (-E(f|\mathcal{A})) \, \mathrm{d}\mu \\ &= \int_{E(f|\mathcal{A})>0} f \, \mathrm{d}\mu + \int_{E(f|\mathcal{A})<0} (-f) \, \mathrm{d}\mu \\ &\leqslant \int_{E(f|\mathcal{A})>0} |f| \, \mathrm{d}\mu + \int_{E(f|\mathcal{A})<0} |f| \, \mathrm{d}\mu \leqslant \int_X |f| \, \mathrm{d}\mu = \|f\|_{L^1_\mu}. \end{aligned} \tag{3}$$

For (4), let $f \in L^1_\mu(\mathcal{B})$ and $g \in L^\infty_\mu(\mathcal{A})$. It's easy to check that (4) holds for characteristic function $g = \chi_{A_0}$ for any $A_0 \in \mathcal{A}$, because for any $A \in \mathcal{A}$,

$$\int_A gE(f|\mathcal{A}) \, \mathrm{d}\mu = \int_{A \cap A_0} E(f|\mathcal{A}) \, \mathrm{d}\mu = \int_{A \cap A_0} f \, \mathrm{d}\mu = \int_A gf \, \mathrm{d}\mu = \int_A E(gf|\mathcal{A}) \, \mathrm{d}\mu.$$

Hence $E(gf|\mathcal{A}) = gE(f|\mathcal{A})$ for these $g's$. Then use linear combinations to extend to simple bounded functions. Finally, using that any $g \in L^\infty_\mu(\mathcal{A})$ is a uniform limit and property (3), the general case follows.

We just need to prove the existence of $E(\cdot|\mathcal{A})$ which satisfies property (1). We first treat the case $f \in L^2_\mu(\mathcal{B})$. Let $P\colon L^2_\mu(\mathcal{B}) \to L^2_\mu(\mathcal{A})$ be the forget projection (measure theory and functional analysis give us this map). We claim that $P(f) \in L^1_\mu \mathcal{A}$ and satisfies property (1). To check this, let $A \in \mathcal{A}$. Then $\chi_A \in L^2_\mu(\mathcal{A})$ and hence

$$\langle f, \chi_A \rangle = \langle P(f), \chi_A \rangle \Rightarrow \int_A f \,\mathrm{d}\mu = \int_A P(f) \,\mathrm{d}\mu.$$

Using the same method in equation (3), we get this claim and we also have $\|P(f)\|_{L^1_\mu} \leqslant \|f\|_{L^1_\mu}$.

Suppose $f \in L^1_\mu(\mathcal{B})$. Then there exists a sequence $f_n \in L^2_\mu(\mathcal{B})$ with $f_n \xrightarrow{L^1} f$ (real analysis gives us this sequence). We obtain

$$\|P(f_m) - P(f_n)\|_{L^1_\mu} = \|P(f_m - f_n)\|_{L^1_\mu} \leqslant \|f_m - f_n\|_{L^1_\mu}.$$

This shows that $P(f_n) \in L^1_\mu(\mathcal{A})$ is a Cauchy sequence and so has a limit $h \in L^1_\mu(\mathcal{A})$. Then for any $A \in \mathcal{A}$, by some real analysis, we have

$$\int_A h \,\mathrm{d}\mu = \lim_{n\to\infty} \int_A P(f_n) \,\mathrm{d}\mu = \lim_{n\to\infty} \int_A f_n \,\mathrm{d}\mu = \int_A f \,\mathrm{d}\mu.$$

Therefore $E(f|\mathcal{A}) := h$ is as desired. □

Theorem 3.19 (Birkhoff pointwise ergodic theorem)**.** *Let $(X, \mathcal{B}, \mu)$ be a probability space and $T\colon X \to X$ be measure-preserving. Then for any $f \in L^1_\mu(\mathcal{B})$, we have*

$$A_n(f)(x) = \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) \to E(f|\mathcal{E}_T)(x)$$

for a.s. $x \in X$ pointwise, and also in L^1_μ, where $\mathcal{E}_T = \{B \in \mathcal{B} : T^{-1}B = B\}$.

Proof. For any function h on X, we define $S_0 h = 0$ and

$$S_n h(x) = h(x) + h \circ T(x) + \cdots + h \circ T^{n-1}(x) = h(x) + h(Tx) + \cdots + h(T^{n-1}x)$$

for all $n \in \mathbb{N}$. For any $f \in L^1_\mu(X)$ and $\varepsilon > 0$, we define

$$g = f - E(f|\mathcal{E}_T) - \varepsilon \in L^1_\mu(X).$$

We wish to show that

$$\limsup_{n\to\infty} A_n(g) \overset{a.s.}{\leqslant} 0. \tag{4}$$

We define $A = \{x \in X : \sup_{k\geqslant 1}(S_k g)(x) = \infty\}$ and just need the following claim:

Claim. $\mu(A) = 0$.

Proof of Claim. Note that

$$S_k g(Tx) = S_{k+1} g(x) - g(x)$$

for $k \geqslant 1$. Hence we have $T(x) \in A$ if and only if $x \in A$, because $S_k g(T(x)) \nearrow \infty$ if and only if $S_k(x) = S_{k-1}(x) + g(x) \nearrow \infty$. This implies $A \in \mathcal{E}_T$.

For $n \geqslant 1$, we define

$$M_n(x) = \max_{0\leqslant k\leqslant n} S_k g(x) = \max\{0, g(x), g(x) + g(Tx), \cdots, g(x) + \cdots + g(T^{n-1}x)\}.$$

Then we have

$$M_{n+1}(x) = \max\{0, g(x) + M_n(Tx)\}$$

and so

$$g(x) \leqslant M_{n+1}(x) - M_n(Tx) = \max\{-M_n(Tx), g(x)\} \leqslant \max\{0, g(x)\}. \tag{5}$$

For any $x \in A$,we have $S_k g(x) \to \infty$ and so $M_n(x) \to \infty$ as $n \to \infty$. Note that $M_{n+1}(x) - M_n(Tx) = \max\{-M_n(Tx), g(x)\}$ and $M_n(Tx) \geqslant 0$, we have

$$M_{n+1}(x) - M_n(Tx) \to g(x). \tag{6}$$

Combining equations (5)(6), with dominated convergence theorem, and by Proposition 3.14 (1), we obtain

$$\lim_{n\to\infty}\int_A (M_{n+1}-M_n\circ T)\,\mathrm{d}\mu = \int_A g\,\mathrm{d}\mu = \int_{A\in\mathcal{E}_T} (f-E(f|\mathcal{E}_T)-\varepsilon)\,\mathrm{d}\mu = -\varepsilon\mu(A). \tag{7}$$

Note again that $A\in\mathcal{E}_T$ and μ is T-invariant, we have

$$\int_A M_n\circ T\,\mathrm{d}\mu = \int_{TA} M_n\circ T\,\mathrm{d}\mu = \int_A M_n\,\mathrm{d}\left(T_*^{-1}\mu\right) = \int_A M_n\,\mathrm{d}\mu.$$

Hence by definition,

$$\int_A (M_{n+1}-M_n\circ T)\,\mathrm{d}\mu = \int_A (M_{n+1}-M_n)\,\mathrm{d}\mu \geqslant \int_A 0\,\mathrm{d}\mu = 0. \tag{8}$$

Combining equations (7)(8), we have

$$0 \leqslant \lim_{n\to\infty}\int_A (M_{n+1}-M_n\circ T)\,\mathrm{d}\mu = -\varepsilon\mu(A).$$

This implies $\mu(A)=0$, which is as desired. □

If $\limsup_{n\to\infty} A_n(g)(x) > 0$, then along a subsequence, $(S_n g)(x)\nearrow\infty$ and so $x\in A$. Hence $\mu(A)=0$ implies equation (4).

Claim. If h is $\mathcal{E}_T$-measurable, then $h\circ T = h$ and so $A_n(h)=h$.

Proof of Claim. Given an interval $U\subset\mathbb{R}$, $h^{-1}(I)$ is a measurable subset in $\mathcal{E}_T$, i.e. $h^{-1}(I)\in\mathcal{E}_T$ and so is T-invariant. Hence for any $x\in X$,

$$h(x)\in I \iff x\in h^{-1}(I) = T^{-1}\circ h^{-1}(I) \iff x\in h\circ T(I).$$

For any $x\in X$, by shrinking the interval I so that $h(x)\in I$, we have $h(x)=h\circ T(x)$. □

Associating with Proposition 3.14 (1), we have

$$A_n(g) = A_n(f) - A_n(E(f|\mathcal{E}_T)) - \varepsilon = A_n(f) - E(f|\mathcal{E}_T) - \varepsilon.$$

By equation (4), we have

$$\limsup_{n\to\infty} A_n(f) \overset{a.s.}{\leqslant} E(f|\mathcal{E}_T)+\varepsilon.$$

By the arbitrariness of ε, we have

$$\limsup_{n\to\infty} A_n(f) \overset{a.s.}{\leqslant} E(f|\mathcal{E}_T).$$

Hence for $-f$, we also have

$$-\limsup_{n\to\infty} A_n(f) = \limsup_{n\to\infty} A_n(-f) \overset{a.s.}{\leqslant} E(-f|\mathcal{E}_T) = -E(f|\mathcal{E}_T).$$

Above all, we have

$$\lim_{n\to\infty} A_n(f) \overset{a.s.}{=} E(f|\mathcal{E}_T),$$

which is as desired.

Now for L^1_μ convergence, we use the mean ergodic theorem (Theorem 3.13) for L^2_μ functions. Let $\varepsilon>0$ and $g\in L^2_\mu(X)$ with $\|f-g\|_{L^1_\mu}<\varepsilon$. This implies

$$\limsup_{n\to\infty}\|A_n(f)-E(f|\mathcal{E}_T)\|_{L^1_\mu} \leqslant \limsup_{n\to\infty}\left(\|A_n(f-g)\|_{L^1_\mu} + \|A_n(g)-P_T(g)\|_{L^1_\mu} + \|P_T(g)-E(f|\mathcal{E}_T)\|_{L^1_\mu}\right).$$

For the first term, because T is measure-preserving,

$$\begin{aligned}\left\|(f-g)\circ T^k\right\|_{L^1_\mu} &= \int_X \left|(f-g)\circ T^k\right|\,\mathrm{d}\mu = \int_{T^kX}\left|(f-g)\circ T^k\right|\,\mathrm{d}\mu \\ &= \int_X |f-g|\,\mathrm{d}\left(T_*^{-k}\mu\right) = \int_X |f-g|\,\mathrm{d}\mu = \|f-g\|_{L^1_\mu}\end{aligned}$$

for any k and hence

$$\begin{aligned}\limsup_{n\to\infty}\|A_n(f-g)\|_{L^1_\mu} &= \limsup_{n\to\infty}\frac{1}{n}\int_X\left|\sum_{k=0}^{n-1}(f-g)\circ T^k\right|\,\mathrm{d}\mu\\&\leqslant \limsup_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\int_X\left|(f-g)\circ T^k\right|\,\mathrm{d}\mu = \limsup_{n\to\infty}\|f-g\|_{L^1_\mu}=\varepsilon.\end{aligned}$$

For the second term, for any $h\in L^2_\mu(X)$, by Cauchy-Schwarz inequality,

$$\|h\|_{L^1_\mu}=\int_X 1\cdot|h|\,\mathrm{d}\mu=\int_X|h|\,\mathrm{d}\mu\leqslant\sqrt{\int_X 1\,\mathrm{d}\mu\cdot\int_X h^2\,\mathrm{d}\mu}=\|h\|_{L^2_\mu}.$$

Hence by mean ergodic theorem,

$$\limsup_{n\to\infty}\|A_n(g)-P_T(g)\|_{L^1_\mu}\leqslant\limsup_{n\to\infty}\|A_n(g)-P_T(g)\|_{L^2_\mu}=0.$$

For the third term, by definition, note that $P_T(g)=E(g|\mathcal{E}_T)$, we have

$$\limsup_{n\to infty}\|P_T(g)-E(f|\mathcal{E}_T)\|_{L^1_\mu}=\limsup_{n\to\infty}\|E(g-f|\mathcal{E}_T)\|_{L^1_\mu}\leqslant\limsup_{n\to\infty}\|g-f\|_{L^1_\mu}<\varepsilon$$

by Proposition (3.14) (3).

Above all, for any $\varepsilon>0$,

$$\begin{aligned}\limsup_{n\to\infty}\|A_n(f)-E(f|\mathcal{E}_T)\|_{L^1_\mu}&\leqslant\limsup_{n\to\infty}\Big(\|A_n(f-g)\|_{L^1_\mu}+\|A_n(g)-P_T(g)\|_{L^1_\mu}+\|P_T(g)-E(f|\mathcal{E}_T)\|_{L^1_\mu}\Big)\\&\leqslant\varepsilon+0+\varepsilon=2\varepsilon.\end{aligned}$$

Hence we have $A_n(f)\xrightarrow{L^1_\mu}E(f|\mathcal{E}_T)$, which is as desired. □

In the proof above, the following fact was used: $f\in L^2_\mu(X)$ is invariant under U_T i.e. $U_tf\overset{L^2_\mu}{=}f$ if and only if there is a representative of f that is $\mathcal{E}_T$-measurable.

Definition 3.20. *Let $\mathcal{A},\mathcal{C}\subset\mathcal{B}$ be two sub-σ-algebras, We write $\mathcal{A}\underset{\mu}{\subset}\mathcal{C}$ if for any $A\in\mathcal{A}$, there exists a $C\in\mathcal{C}$ with $\mu(A\Delta C)=0$. Moreover, $\mathcal{A}\underset{\mu}{=}\mathcal{C}$ if $\mathcal{A}\underset{\mu}{\subset}\mathcal{C}$ and $\mathcal{C}\underset{\mu}{\subset}\mathcal{A}$.*

Lemma 3.21. *Define $\mathcal{E}_{T,\mu}=\{B\in\mathcal{B}:\mu(T^{-1}B\Delta B)=0\}$. Then $\mathcal{E}_T\subset\mathcal{E}_{T,\mu}$ and $\mathcal{E}_T\underset{\mu}{=}\mathcal{E}_{T,\mu}$.*

Proof. By definition, $\mathcal{E}_T\subset\mathcal{E}_{T,\mu}$, and hence $\mathcal{E}_T\underset{\mu}{\subset}\mathcal{E}_{T,\mu}$. Conversely, for any $A\in\mathcal{E}_{T,\mu}$, we define $\widetilde{A}=\bigcap_{k=0}^{\infty}\bigcup_{j=k}^{\infty}T^{-j}A$. Up to null sets, we have $T^{-j}A$ is A and so $\bigcup_{j=k}^{\infty}$ is A and so $\widetilde{A}$ is A. Hence $\mu(\widetilde{A}\Delta A)=0$. Moreover

$$T^{-1}\widetilde{A}=\bigcap_{k=0}^{\infty}\bigcup_{j=k}^{\infty}T^{-j-1}A=\bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty}T^{-j}A=\bigcap_{k=0}^{\infty}\bigcup_{j=k}^{\infty}T^{-j}A=\widetilde{A}.$$

Above all, we find a $\widetilde{A}\in\mathcal{E}_T$ such that $\mu(\widetilde{A}\Delta A)=0$. Therefore $\mathcal{E}_{T,\mu}\underset{\mu}{\subset}\mathcal{E}_T$ and then $\mathcal{E}_T\underset{\mu}{=}\mathcal{E}_{T,\mu}$ □

Lemma 3.22. *If $f\in L^2_\mu(X)$ satisfies $f\circ T\overset{a.s.}{=}f$, then there exists a representative $\widetilde{f}$ of f in $L^2_\mu(X)$ so that $\widetilde{f}\circ T=\widetilde{f}$ everywhere.*

Definition 3.23. *Let $(X,\mathcal{B},\mu)$ be a probability space and $T\colon X\to X$ be measure-preserving. We say that (T,μ) is ergodic, if for any $B\in\mathcal{B}$ so that $\mu(T^{-1}B\Delta B)=0$ (equivalently for any T-invariant $B\in\mathcal{B}$), we have $\mu(B)\in\{0,1\}$.*

Lemma 3.24. *Let $(X,\mathcal{B},\mu,T)$ be ergodic on a probability space. Let $f\in L^1_\mu(X)$ such that $f\circ T\overset{a.s.}{=}f$. Then f is a constant a.s.*

Proof. For any interval I, $B\colon f^{-1}(I)\subset X$ is invariant, because $T^{-1}B=(f\circ T)^{-1}B=f^{-1}B=B$. Then $\mu(B)\in\{0,1\}$.

Note that $\mathbb{R}=\bigsqcup_{n\in\mathbb{N}}[n+n+1)$, there exists an interval $[n+n+1)$ such that $f^{-1}([n+n+1))=1$. Using dichotomy, we can find half a interval I' such that $f^{-1}(I')=1$. Repeating this process, by dichotomies, we get a point c such that $f^{-1}(\{c\})=1$, i.e. $f\overset{a.s.}{=}c$. □

Corollary 3.25. *Let $(X,\mathcal{B},\mu,T)$ be ergodic on a probability space. Let $f\in L^1_\mu(X)$. Then $A_n(f)\to\int_X f\,\mathrm{d}\mu$ a.s. and also in L^1_μ.*

Proof. By pointwise ergodic theorem (Theorem 3.19), for $f\in L^1_\mu(X)$, $\widetilde{f}:=E(f|\mathcal{E}_T)=\lim_{n\to\infty}A_n(f)$ is invariant. By Lemma 3.24, there exists a constant c such that $A_n(f)\to\widetilde{f}\overset{a.s.}{=}c$ a.s. and in L^1. The constant c must be $\int_X f\,\mathrm{d}\mu$. ☐

This corollary shows that in the setting of ergodic system, the "time average" equals to the "space average".

Lemma 3.26. *$(X,\mathcal{B},\mu,T)$ is ergodic if and only if $U_T\colon L^2_\mu(X)\to L^2_\mu(X)$ has only a one-dimensional eigenspace for eigenvalue* 1.

Proof. **($\Rightarrow$):** First constants are in the eigenspace for eigenvalue 1. Suppose f is in the eigenspace for eigenvalue 1, i.e. $f\overset{a.s.}{=}U_T(f)=f\circ T$. By Lemma 3.24, f is a constant a.s.. Above all, the eigenspace for eigenvale 1 is $\mathbb{R}$.

($\Leftarrow$): Suppose $B\in\mathcal{B}$ and $\mu(T^{-1}B\Delta B)=0$. It's easy to check that χ_B is an eigenfunction for eigenvalue 1. Note that the eigenspace for eigenvalue 1 is $\mathbb{R}$. Hence χ_B equals to a constant c a.s.. We have $\mu(B)=\chi_B^{-1}(\{1\})=\chi_B^{-1}([0.5,1.5])=c^{-1}([0.5,1.5])\in\{0,1\}$. Therefore $(X,\mathcal{B},\mu,T)$ is ergodic. ☐

Example 3.27. $T_p\colon\mathbb{T}\to\mathbb{T}$, $x\mapsto px$ for $p\geqslant 2$ is ergodic with respect to Lebesgue measure λ. To see this, let $f\in L^2(\mathbb{T})=L^2([0,1])$ satisfy $U_{T_p}f=f$. Denote $f=\sum_{n\in\mathbb{Z}}c_ne_n$ where $e_n(x)=\mathrm{e}^{2\pi\mathrm{i}nx}$ be the Fourier expansion. Therefore

$$U_{T_p}f(x)=f\circ T_p(x)=\sum_{n\in\mathbb{Z}}c_ne_n(px)=\sum_{n\in\mathbb{Z}}c_ne_{pn}(x).$$

Then $U_{T_p}f=f$ implies that

$$\sum_{n\in\mathbb{Z}}c_ne_n=\sum_{m\in\mathbb{Z}}c_me_{mp}.$$

By comparing coefficients, we have $c_n=0$ if $p\nmid n$. Let m be $\pm1,\pm2,\cdots$, we have that $\cdots=c_{\pm2p}=c_{\pm p}=c_{\pm1}=0$. Hence $c_n=0$ except when $n=0$, i.e. $f=c_0e_0=c_0$ is a constant. This shows that the eigenspace of U_{T_p} for eigenvalue 1 is $\mathbb{R}$. By Lemma 3.26, (T_p,λ) is ergodic.

Example 3.28. $R\alpha\colon\mathbb{T}\to\mathbb{T}$, $x\mapsto x+\alpha$ for $\alpha\notin\mathbb{Q}$ is ergodic with respect to the Lebesgue measure λ. To see this, let $f\in L^2(\mathbb{T})=L^2([0,1])$ satisfy $U_{T_p}f=f$. Denote $f=\sum_{n\in\mathbb{Z}}c_ne_n$ where $e_n(x)=\mathrm{e}^{2\pi\mathrm{i}nx}$ be the Fourier expansion.

$$U_{R_\alpha}f(x)=f\circ R_\alpha(x)=f(x+\alpha)=\sum_{n\in\mathbb{Z}}c_ne_n(x+\alpha)=\sum_{n\in\mathbb{Z}}c_n\mathrm{e}^{2\pi\mathrm{i}n\alpha}e_n(x).$$

Then $U_{R_\alpha}f=f$ implies that

$$\sum_{n\in\mathbb{Z}}c_n\mathrm{e}^{2\pi\mathrm{i}n\alpha}e_n(x)=\sum_{n\in\mathbb{Z}}c_ne_n(x).$$

Hence for any n, $c_n\mathrm{e}^{2\pi\mathrm{i}n\alpha}=c_n$. We must have $c_n=0$ for $n\neq0$, i.e. $f=c_0e_0=c_0$ is a constant. This shows that the eigenspace of U_{R_α} for eigenvalue 1 is $\mathbb{R}$. By Lemma 3.26, (R_α,λ) is ergodic.

Example 3.29 (Higher-dimensional Rotation)**.** Let $\underline{\alpha}=(\alpha_1,\cdots,\alpha_d)\in\mathbb{R}^d$. We define $R_{\underline{\alpha}}:=\mathbb{T}^d\to\mathbb{T}^d$, $\underline{x}\mapsto\underline{x}+\underline{\alpha}$. Let λ be the d-dimensional Lebesgue measure. It's easy to check that λ is invariant under $R_{\underline{\alpha}}$.

When $(\mathbb{R}_{\underline{\alpha}},\lambda)$ is ergodic?

Suppose $f=U_{R_{\underline{\alpha}}}=f\circ R_{\underline{\alpha}}$. Let $f=\sum_{\underline{n}\in\mathbb{Z}^d}c_{\underline{n}}e_{\underline{n}}$ where $e_{\underline{n}}(\underline{x})=\mathrm{e}^{2\pi\mathrm{i}\langle\underline{n},\underline{x}\rangle}$ be the Fourier expansion. Note that $e_{\underline{n}}(\underline{x}+\underline{\alpha})=\mathrm{e}^{2\pi\mathrm{i}(n_1\alpha_1+\cdots+n_d\alpha_d)}e_{\underline{n}}(\underline{x})$, $f=f\circ R_{\underline{\alpha}}$ shows that

$$\sum_{\underline{n}\in\mathbb{Z}^d}c_{\underline{n}}e_{\underline{n}}=\sum_{\underline{n}\in\mathbb{Z}^d}c_{\underline{n}}\mathrm{e}^{2\pi\mathrm{i}(n_1\alpha_1+\cdots n_d\alpha_d)}e_{\underline{n}}.$$

Comparing coefficients, we have $c_{\underline{n}}=c_{\underline{n}}\mathrm{e}^{2\pi\mathrm{i}(n_1\alpha_1+\cdots n_d\alpha_d)}$ for any $\underline{n}$. Hence f is constant if and only if $1,\alpha_1,\cdots,\alpha_d$ are linear independent over $\mathbb{Q}$. By Lemma 3.26, $(R_{\underline{\alpha}},\lambda)$ is ergodic if and only if $1,\alpha_1,\cdots,\alpha_d$ are linear independent over $\mathbb{Q}$.

Lemma 3.30. *Suppose $1,\alpha_1,\cdots,\alpha_d$ are linear independent. Let $f\in C(\mathbb{T}^d)$. Then $A_n(f)\to\int_{\mathbb{T}^d}f\,\mathrm{d}\lambda$ everywhere.*

Proof. We start with $f = e_{\underline{n}}$. Calculating directly, we have

$$\begin{aligned} A_n(e_{\underline{m}})(\underline{x}) &= \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{e}^{2\pi\mathrm{i}\langle \underline{m}, \underline{x}+j\underline{\alpha}\rangle} = \frac{1}{n}\mathrm{e}^{2\pi\mathrm{i}\langle \underline{m},\underline{x}\rangle}\sum_{j=0}^{n-1}\left(\mathrm{e}^{2\pi\mathrm{i}\langle \underline{m},\underline{\alpha}\rangle}\right)^j \\ &= \mathrm{e}^{2\pi\mathrm{i}\langle \underline{m},\underline{x}\rangle}\frac{1}{n}\frac{\left(\mathrm{e}^{2\pi\mathrm{i}\langle \underline{m},\underline{\alpha}\rangle}\right)^n - 1}{\mathrm{e}^{2\pi\mathrm{i}\langle \underline{m},\underline{\alpha}\rangle} - 1} \to 0 = \int_{\mathbb{T}^d} e_{\underline{n}}\,\mathrm{d}\lambda. \end{aligned}$$

Then the lemma holds for any $f \in C(\mathbb{T}^d)$ by Fourier series approximation. □

Lemma 3.31. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system on a probability space. Then (T, μ) is ergodic if and only if for any $A, B \in \mathcal{B}$, we have*

$$\frac{1}{N}\sum_{n=0}^{N-1}\mu\left(A \cap T^{-n}B\right) \to \mu(A)\mu(B).$$

Proof. **(⇐):** For any $B \in \mathcal{B}$ such that $B = T^{-1}B$, letting $A = B = T^{-1}B$, we have

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

Hence we have $\mu(B) \in \{0, 1\}$.

(⇒): Applying mean ergodic theorem (Theorem 3.13) for χ_B, we have

$$\frac{1}{N}\sum_{n=0}^{N-1}\chi_B \circ T^n \xrightarrow{L^2_\mu} P_T(\chi_B),$$

i.e.

$$\frac{1}{N}\sum_{n=0}^{N-1}\chi_{T^{-n}B} \xrightarrow{L^2_\mu} \mu(B)\chi_X.$$

Taking inner product with χ_A, we have

$$\frac{1}{N}\sum_{n=0}^{N-1}\langle \chi_A, \chi_{T^{-n}B}\rangle_{L^2_\mu} \to \langle \chi_A, \mu(B)\chi_X\rangle_{L^2_\mu},$$

where

$$\langle \chi_A, \chi_{T^{-n}B}\rangle_{L^2_\mu} = \int_X \chi_A\chi_{T^{-n}B}\,\mathrm{d}\mu = \mu\left(A \cap T^{-n}B\right)$$

and

$$\langle \chi_A, \mu(B)\chi_X\rangle_{L^2_\mu} = \int_X \mu(B)\chi_A\chi_X\,\mathrm{d}\mu = \mu(A)\mu(B).$$

The claim follows. □

Definition 3.32. *$(X, \mathcal{B}, \mu, T)$ is called (strongly) mixing if for any $A, B \in \mathcal{B}$, we have*

$$\mu\left(A \cap T^{-n}B\right) \to \mu(A)\mu(B).$$

Lemma 3.33. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system on a probability space. Then T is mixing if and only if for any two $f, g \in L^2_\mu(X)$, we have*

$$\langle f, U_T^n g\rangle_{L^2_\mu} \to \int_X f\,\mathrm{d}\mu \int_X \overline{g}\,\mathrm{d}\mu. \tag{9}$$

Proof. **(⇐):** Taking $f = \chi_A$ and $g = \chi_B$, by the same calculation as in the proof of Lemma 3.31, we get the claim.

(⇒): Also by the same calculations, equation (9) holds for characteristic functions. Then it also holds for simple functions. We can approximate any function by simple functions. By the continuity of $\langle \cdot, \cdot\rangle_{L^2_\mu}$, equation (9) holds for general functions. □

Example 3.34. Let $d \geqslant 2$ and $A \in \mathrm{SL}_d(\mathbb{Z})$. Define $T_A\colon \mathbb{T}^d \to \mathbb{T}^d$, $x \mapsto Ax$. Then $(\mathbb{T}^d, \mathcal{B}, \lambda, T_A)$ is mixing if and only if it's ergodic if and only if A has no root of unity as an eigenvalue. By Lemma 3.31 and Lemma 3.33, Fourier series gives us this claim.

Definition 3.35. *$(X, \mathcal{B}, \mu, T)$ is called weak mixing if for any $A, B \in \mathcal{B}$, we have*

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

Remark 3.36.

$$\text{strongly mixing} \Rightarrow \text{weak mixing} \Rightarrow \text{ergodic}$$

Definition 3.37. *A subset $J \subset \mathbb{N}$ has density*

$$d(J) := \lim_{N\to\infty}\frac{1}{N}\sharp\left(J \cap [0, N]\right)$$

if this limit exists.

Definition 3.38. *$f \neq 0 \in L^2_\mu(X)$ is an eigenfunction for T if $U_T f = \lambda f$.*

Definition 3.39. *T is said to have continuous spectrum if any eigenfunction is a constant. T has discrete spectrum if $L^2_\mu(X)$ is spanned by the eigenfunctions for T.*

Theorem 3.40. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system on a probability space. Then the followings are equivalent:*

1. *T is weak mixing.*
2. *$T \times T$ is ergodic with respect to $(\mu \times \mu)$.*
3. *$T \times T$ is weak mixing with respect to $(\mu \times \mu)$.*
4. *$T \times S$ is ergodic with respect to $\mu \times \nu$ where $(Y, \mathcal{C}, \nu, S)$ is any ergodic system.*
5. *T has continuous spectrum.*
6. *For any $A, B \in \mathcal{B}$, we have $\mu(A \cap T^{-n}B) \to \mu(A)\mu(B)$ outside a set of density zero.*

We first prove a lemma to characterize (1)⇔(6):

Lemma 3.41. *For a bounded sequence $a_n \geqslant 0$, the followings are equivalent:*

1. $\frac{1}{N}\sum_{n=0}^{N-1} a_n \to 0$.
2. *There exists a subset $J \subset \mathbb{N}$ of density zero so that $a_n \to 0$ as $n \in \mathbb{N} - J$ goes to ∞.*
3. $\frac{1}{N}\sum_{n=0}^{N-1} a_n^2 \to 0$.

Proof. **(1)⇒(2):** Define $J_k = \{n \in \mathbb{N} : a_n > \frac{1}{k}\}$. Note that $J_1 \subset J_2 \subset \cdots$, we will show that J_k has density zero for any k: For any $n \in J_k \cap [0, N-1)$, by definition, we get an $a_n > \frac{1}{k}$. Hence $\sharp\left(J_k \cap [0, N-1)\right) \leqslant \sum_{n \in J_k \cap [0,N-1)} k a_n \leqslant k\sum_{n=0}^{N-1} a_n$. Then we have

$$\lim_{n\to\infty}\frac{1}{N}\sharp\left(J_k \cap [0, N-1)\right) \leqslant \lim_{N\to\infty}\frac{k}{N}\sum_{n=0}^{N-1} a_n = k0 = 0.$$

Now for any k, we can find some l_k such that $\frac{1}{N}\sharp(J_k \cap [0, N-1)) < \frac{1}{k}$ for all $N \geqslant l_k$. WLOG, we can assume $l_k \nearrow \infty$. Define

$$J = \bigcup_{k=0}^{\infty}\left(J_k \cap [l_k, l_{k+1})\right).$$

Claim. J has density zero.

Proof of Claim. Note that for any N, suppose $N \in [l_k, l_{k+1})$, then

$$\frac{1}{N}\sharp\left(J \cap [0, N-1)\right) \leqslant \frac{1}{N}\sharp\left(J_k \cap [0, N-1)\right) < \frac{1}{k}.$$

Hence we have

$$\lim_{N\to\infty} \frac{1}{N}\sharp\left(J \cap [0, N-1)\right) = 0$$

which is as desired. ☐

Claim. $\lim_{\substack{n\to\infty \\ n\notin J}} a_n = 0.$

Proof of Claim. Note that for $n \notin J$, $n \in [l_k, l_{k+1})$, we have $n \notin J_k$. Then $a_n < \frac{1}{k}$. Hence we have $\lim_{\substack{n\to\infty \\ n\notin J}} a_n = 0.$ ☐

Hence J is as desired.

(2)⇒(1): Suppose $a_n \leqslant M$ for all n. For any $\varepsilon > 0$, there exists some l_1 so that $\frac{1}{N}\sharp(J \cap [0, N)) < \varepsilon$ for $N \geqslant l_1$, and some l_2 so that $a_n \leqslant \varepsilon$ for any $n \geqslant l_2$ and $n \notin J$. Hence for $N \geqslant l_1$, we have

$$\frac{1}{N}\sum_{n=0}^{N-1} a_n \leqslant \frac{1}{N}\left(\sum_{n\leqslant l_2} a_n + \sum_{\substack{n=0 \\ n\in J}}^{N-1} a_n + \sum_{\substack{n=l_2+1 \\ n\notin J}}^{N-1} a_n\right)$$

For the first term, we have $\sum_{n\leqslant l_2} a_n \leqslant Ml_2$. For the second term, we have

$$\sum_{\substack{n=0 \\ n\in J}}^{N-1} a_n \leqslant M\sum_{\substack{n=0 \\ n\in J}}^{N-1} 1 = M\sharp\left(J \cap [0, N)\right) \leqslant M\varepsilon N.$$

For the third term, we have

$$\sum_{\substack{n=l_2+1 \\ n\notin J}}^{N-1} a_n \leqslant \sum_{n=l_2+1}^{N-1} a_n \leqslant (N-1-l_2)\,\varepsilon \leqslant N\varepsilon.$$

Above all, we have

$$\frac{1}{N}\sum_{n=0}^{N-1} a_n \leqslant \frac{1}{N}\left(\sum_{n\leqslant l_2} a_n + \sum_{\substack{n=0 \\ n\in J}}^{N-1} a_n + \sum_{\substack{n=l_2+1 \\ n\notin J}}^{N-1} a_n\right) \leqslant \frac{1}{N}\left(Ml_2 + M\varepsilon N + N\varepsilon\right) \to (M+1)\varepsilon.$$

Therefore $\frac{1}{N}\sum_{n=0}^{N-1} a_n \to 0$.

(3)⇔(2): Using (1)⇔(2), we have $\frac{1}{N}\sum_{n=0}^{N-1} a_n^2 \to 0$ if and only if there exists a subset $J \subset \mathbb{N}$ of density zero so that $a_n^2 \to 0$ as $n \in \mathbb{N} - J$ goes to ∞, if and only if there exists a subset $J \subset \mathbb{N}$ of density zero so that $a_n \to 0$ as $n \in \mathbb{N} - J$ goes to ∞. ☐

Proof of Theorem 3.40. **(1)⇔(6):** It follows Lemma 3.41 using $a_n := |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)|$.

(6)⇒(3): For any $A_1, B_1, A_2, B_2 \in \mathcal{B}$, there exist J_1, J_2 of density zero so that $\mu(A_i \cap T^{-n}B_i) \to \mu(A_i)\mu(B_i)$ for $n \notin J_i \to \infty$, $i = 1, 2$. Denote $J = J_1 \cup J_2$. Note that

$$\begin{aligned}
d(J) = \lim_{N\to\infty} \frac{1}{N}\sharp\left(J \cap [0, N]\right) &= \lim_{N\to\infty} \frac{1}{N}\sharp\left((J_1 \cup J_2) \cap [0, N]\right) \\
&\leqslant \lim_{N\to\infty} \frac{1}{N}\left(\sharp\left(J_1 \cap [0, N]\right) + \sharp\left(J_2 \cap [0, N]\right)\right) \\
&= \lim_{N\to\infty} \frac{1}{N}\sharp\left(J_1 \cap [0, N]\right) + \lim_{N\to\infty} \frac{1}{N}\sharp\left(J_2 \cap [0, N]\right) = d(J_1) + d(J_2) = 0,
\end{aligned}$$

J is also of density zero. As $n \to \infty$ outside of J,

$$\begin{aligned}
(\mu \times \mu)\left((A_1 \times A_2) \cap (T \times T)^{-n}(B_1 \times B_2)\right) &= \mu\left(A_1 \cap T^{-n}B_1\right) \cdot \mu\left(A_2 \cap T^{-n}B_2\right) \\
&\to \mu(A_1)\,\mu(B_1) \cdot \mu(A_2)\,\mu(B_2) = \left(\mu(A_1)\,\mu(A_2)\right)\left(\mu(B_1)\,\mu(B_2)\right) \\
&= (\mu \times \mu)(A_1 \times A_2)\,(\mu \times \mu)(B_1 \times B_2).
\end{aligned}$$

This shows that the desired convergence for measurable rectangles. If A is a finite disjoint union of measurable rectangles and also B, then the statement for rectangles implies the statement for A, B. Finally we note that

$$\mathcal{C} := \{C \in \mathcal{B} \times \mathcal{B} : C \text{ can be approximated under } (\mu \times \mu) \text{ by finite disjoint unions of measurable rectangles}\}$$

is a σ-algebra. By measure theory, $\mathcal{C} = \mathcal{B} \times \mathcal{B}$. For any $A, B \in \mathcal{B} \times \mathcal{B}$, for any $\varepsilon > 0$, we can find finite unions of rectangles $\widetilde{A}, \widetilde{B} \in \mathcal{C}$ such that $(\mu \times \mu)(A \Delta \widetilde{A}) < \varepsilon$, $(\mu \times \mu)(B \Delta \widetilde{B}) < \varepsilon$. Then we have

$$\begin{aligned} &\limsup_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left|(\mu \times \mu)(A \cap (T \times T)^{-n} B) - (\mu \times \mu)(A)(\mu \times \mu)(B)\right| \\ \leqslant &\limsup_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left(\left|(\mu \times \mu)\left(\widetilde{A} \cap (T \times T)^{-n} \widetilde{B}\right) - (\mu \times \mu)\left(\widetilde{A}\right)(\mu \times \mu)\left(\widetilde{B}\right)\right| + 4\varepsilon\right) \leqslant 4\varepsilon. \end{aligned}$$

Hence for any $A, B \in \mathcal{B} \times \mathcal{B}$, we have $\frac{1}{N} \sum_{n=0}^{N-1} |(\mu \times \mu)(A \cap (T \times T)^{-n} B) - (\mu \times \mu)(A)(\mu \times \mu)(B)| \to 0$, i.e. $T \times T$ is weak mixing.

(3)⇒(1): For any $A, B \in \mathcal{B}$, we have

$$\begin{aligned} &\frac{1}{N} \sum_{n=0}^{N-1} \left|\mu\left(A \cap T^{-n} B\right) - \mu(A)\mu(B)\right| \\ = &\frac{1}{N} \sum_{n=0}^{N-1} \left|(\mu \times \mu)((A \times X) \cap (T \times T)^{-n}(B \times X)) - (\mu \times \mu)(A \times X)(\mu \times \mu)(B \times X)\right| \to 0. \end{aligned}$$

Therefore T is weak mixing.

(1)⇒(4): Let $A_1, B_1 \in \mathcal{B}$ and $A_2, B_2 \in \mathcal{C}$. Divide $\mu \times \nu\left((A_1 \times A_2) \cap (T \times S^{-n}\left(B_1 \times B_2\right)\right)$ into two parts:

$$\begin{aligned} &\mu \times \nu\left((A_1 \times A_2) \cap (T \times S^{n}\left(B_1 \times B_2\right)\right) = \mu\left(A_1 \cap T^{-n} B_1\right) \cdot \nu\left(A_2 \cap T^{-n} B_2\right) \\ = &\left(\mu\left(A_1 \cap T^{-n} B_1\right) - \mu\left(A_1\right)\mu\left(B_1\right)\right)\nu\left(A_2 \cap S^{-n} B_2\right) + \mu\left(A_1\right)\mu\left(B_1\right)\nu\left(A_2 \cap S^{-n} B_2\right). \end{aligned}$$

For the first term, because T is weak mixing, as $n \to \infty$,

$$\begin{aligned} &\frac{1}{N} \sum_{n=0}^{N-1} \left(\mu\left(A_1 \cap T^{-n} B_1\right) - \mu\left(A_1\right)\mu\left(B_1\right)\right)\nu\left(A_2 \cap S^{-n} B_2\right) \\ \leqslant &\frac{1}{N} \sum_{n=0}^{N-1} \left|\mu\left(A_1 \cap T^{-n} B_1\right) - \mu\left(A_1\right)\mu\left(B_1\right)\right| \cdot 1 \to 0. \end{aligned}$$

For the second term, because S is ergodic, as $N \to \infty$,

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} \mu\left(A_1\right)\mu\left(B_1\right)\nu\left(A_2 \cap S^{-n} B_2\right) = &\mu\left(A_1\right)\mu\left(B_1\right) \frac{1}{N} \sum_{n=0}^{N-1} \nu\left(A_2 \cap S^{-n} B_2\right) \\ \to &\mu\left(A_1\right)\mu\left(B_1\right)\nu\left(A_2\right)\nu\left(B_2\right) = \mu \times \nu\left(A_1 \times A_2\right)\mu \times \nu\left(B_1 \times B_2\right). \end{aligned}$$

Above all, as $N \to \infty$,

$$\begin{aligned} &\frac{1}{N} \sum_{n=0}^{N-1} (\mu \times \mu)\left((A_1 \times A_2) \cap (T \times S^{n}\left(B_1 \times B_2\right)\right) = \\ = &\frac{1}{N} \sum_{n=0}^{N-1} \left(\mu\left(A_1 \cap T^{-n} B_1\right) - \mu\left(A_1\right)\mu\left(B_1\right)\right)\nu\left(A_2 \cap S^{-n} B_2\right) + \frac{1}{N} \sum_{n=0}^{N-1} \mu\left(A_1\right)\mu\left(B_1\right)\nu\left(A_2 \cap S^{-n} B_2\right) \\ \to &\mu \times \nu\left(A_1 \times A_2\right)\mu \times \nu\left(B_1 \times B_2\right). \end{aligned}$$

Therefore our statement holds for measurable rectangles. Using finite disjoint unions of rectangles and approximation argument, we again obtain

$$\frac{1}{N} \sum_{n=0}^{N-1} \mu \times \nu\left(A \cap (T \times S)^{-n} B\right) \to \mu \times \nu(A)\mu \times \nu(B)$$

as $N \to \infty$, for any measurable $A, B \in \mathcal{B} \times \mathcal{C}$. Hence $T \times S$ is ergodic.

(4)⇒(2): Note that $(\{x_0\}, 2^{\{x_0\}}, \lambda, \mathrm{Id})$ is ergodic, setting $(Y, \mathcal{C}, \nu, S) = (\{x_0\}, 2^{\{x_0\}}, \lambda, \mathrm{Id})$, we have $T \times \mathrm{Id}$ is ergodic and so is T. Then we can set $(Y, \mathcal{C}, \nu, S) = (X, \mathcal{B}, \mu, T)$ and get (2).

(2)⇒(6): By Lemma 3.41 (2)⇔(3), it suffices to show that $\frac{1}{N}\sum_{n=0}^{N-1}(\mu(A \cap T^{-n}B) - \mu(A)\mu(B))^2 \to 0$ as $N \to \infty$. Expanding it, we have

$$\begin{aligned}&\frac{1}{N}\sum_{n=0}^{N-1}\left(\mu\left(A \cap T^{-n}B\right) - \mu(A)\mu(B)\right)^2 \\ =&\frac{1}{N}\sum_{n=0}^{N-1}\mu\left(A \cap T^{-n}B\right)^2 - 2\mu(A)\mu(B)\frac{1}{N}\sum_{n=0}^{N-1}\mu\left(A \cap T^{-n}B\right) + (\mu(A)\mu(B))^2.\end{aligned}$$

For the first term,

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1}\mu\left(A \cap T^{-n}B\right)^2 &= \frac{1}{N}\sum_{n=0}^{N-1}(\mu \times \mu)\left((A \times A) \cap (T \times T)^{-n}(B \times B)\right) \\ &\to (\mu \times \mu)(A \times A)(\mu \times \mu)(B \times B) = \mu(A)^2\mu(B)^2.\end{aligned}$$

For the second term, we need a claim:

Claim. If $T \times T$ is ergodic for $(\mu \times \mu)$, then T is ergodic for μ.

Proof of Claim. Suppose $B \in \mathcal{B}$ satisfies $\mu(T^{-1}B\Delta B) = 0$. Then $(\mu \times \mu)((T \times T)^{-1}(B \times B)\Delta(B \times B)) = \mu(T^{-1}B\Delta B)^2 = 0$. Hence $(\mu \times \mu)(B \times B) = \mu(B)^2 \in \{0, 1\}$. We have $\mu(B) \in \{0, 1\}$. ☐

By Lemma 3.31,

$$\frac{1}{N}\sum_{n=0}^{N-1}\mu\left(A \cap T^{-n}B\right) \to \mu(A)\mu(B).$$

Therefore, as $N \to \infty$,

$$\begin{aligned}&\frac{1}{N}\sum_{n=0}^{N-1}\left(\mu\left(A \cap T^{-n}B\right) - \mu(A)\mu(B)\right)^2 \\ =&\frac{1}{N}\sum_{n=0}^{N-1}\mu\left(A \cap T^{-n}B\right)^2 - 2\mu(A)\mu(B)\frac{1}{N}\sum_{n=0}^{N-1}\mu\left(A \cap T^{-n}B\right) + (\mu(A)\mu(B))^2 \\ \to&\mu(A)^2\mu(B)^2 - 2\mu(A)\mu(B)\mu(A)\mu(B) + \mu(A)^2\mu(B)^2 = 0.\end{aligned}$$

(2)⇒(5): Suppose $f \in L^2_\mu(X)$ is an eigenfunction for U_T, i.e. $f \circ T \overset{a.s.}{=} \lambda f$. Define $g \in L^2_{\mu\times\mu}(X \times X)$ by $g(x, y) = f(x)\overline{f(y)}$. Then

$$\left(U_{T\times T}g\right)(x, y) = g \circ (T \times T)(x, y) = f \circ T(x)\overline{f \circ T(y)} = \lambda f(x)\overline{\lambda f(y)} = \lambda\overline{\lambda}f(x)\overline{f(y)} = 1 \cdot g(x, y).$$

This shows g is an eigenfunction of $U_{T\times T}$. Because $T \times T$ is ergodic, by Lemma 3.26, g is a.s. a constant. Hence f is a.s. a constant. By definition, T has continuous spectrum.

(5)⇒(2): We need some facts about the spectral theorem for compact selfadjoint operators, which are listed in Appendix A.

Let $B \subset X \times X$ be a $T \times T$-invariant measurable subset. We suppose $(\mu \times \mu)(B) \notin \{0, 1\}$. Then χ_B is not a.s. a constant and so isn't $k_0(x, y) := \chi_B(x, y) - (\mu \times \mu)(B)$. Define

$$k_1(x, y) = \frac{1}{2}\left(k_0(x, y) + \overline{k_0(y, x)}\right), \quad k_2(x, y) = \frac{1}{2\mathrm{i}}\left(k_0(x, y) - \overline{k_0(y, x)}\right).$$

We have $k_i(x, y) = \overline{k_i(y, x)}$ for $i = 1, 2$ and $k_0(x, y) = k_1(x, y) + \mathrm{i}k_2(x, y)$ for any $x, y \in X$. Hence k_1 or k_2 is not a.s. a constant. Denote it by k.

Because $\int_{X\times X} k_0 \,\mathrm{d}(\mu \times \mu) = \int_{X\times X}\chi_B \,\mathrm{d}(\mu \times \mu) - \mu(B) = 0$.

$$\int_{X\times X} k \,\mathrm{d}(\mu \times \mu) = \begin{cases}\frac{1}{2}\int_{X\times X}\left(k_0(x, y) + \overline{k_0(y, x)}\right)\mathrm{d}(\mu \times \mu) \\ \frac{1}{2\mathrm{i}}\int_{X\times X}\left(k_0(x, y) - \overline{k_0(y, x)}\right)\mathrm{d}(\mu \times \mu)\end{cases} = 0 \tag{10}$$

By definition, $k_0 \circ (T \times T) = k_0$. Hence we have

$$k \circ (T \times T) = k. \tag{11}$$

By Example A.3,

$$(Af)(x) = \int_X k(x,y) f(y) \, \mathrm{d}\mu(y)$$

is a selfadjoint compact operator of $L^2_\mu(X)$.

Claim. $A \neq 0$.

Proof of Claim. Suppose $A = 0$. Then for any $C_1, C_2 \subset X$,

$$0 = \langle A\chi_{C_1}, \chi_{C_2} \rangle = \int_{X \times X} k(x,y) \chi_{C_1}(y) \overline{\chi_{C_2}(x)} \, \mathrm{d}(\mu \times \mu) = \int_{C_1 \times C_2} k(x,y) \, \mathrm{d}(\mu \times \mu).$$

This forces $k \overset{a.s.}{=} 0$, which contradicts to our construction of k. □

By Theorem A.4, A has an eigenvalue $\lambda \neq 0$ with a finite-dimensional eigenspace V_λ.

Claim. V_λ is invariant under U_T.

Proof of Claim. Suppose $f \in V_\lambda$. By equation (11), we have

$$\begin{aligned} A(U_T f)(x) = \int_X (U_T f)(y) k(x,y) \, \mathrm{d}\mu(y) &= \int_X f(Ty) k(Tx, Ty) \, \mathrm{d}\mu(Ty) \\ &= \int_X f(z) k(Tx, z) \, \mathrm{d}\mu(z) = (Af)(Tx) = \lambda f(Tx) = \lambda (U_T f)(x). \end{aligned}$$

This shows $U_T f \in V_\lambda$. □

By linear algebra, $U_T|_{V_\lambda} : V_\lambda \to V_\lambda$ has an eigenvector $f \in V_\lambda - \{0\}$. If f is a.s. a constant c, then by equation (10) we have

$$\lambda c = \lambda \int_X c \, \mathrm{d}\mu = \int_X Ac \, \mathrm{d}\mu = \int_{X \times X} k(x,y) \cdot c \, \mathrm{d}(\mu \times \mu) = 0$$

which is a contradiction. Hence U_T has a non-a.s.-constant eigenvetor, i.e. T has a non-a.s.-constant eigenfunction. Therefore we have T doesn't have continuous spectrum which contradicts to (5). □

Corollary 3.42. *If $T_1, \cdots, T_k$ are weak mixing, then $T_1 \times \cdots \times T_k$ is weak mixing.*

Proof. We just need prove for $k = 2$. By Theorem 3.40 (1)⇒(4), for any ergodic S, $T_2 \times S$ is ergodic. Using this theorem again, we have $T_1 \times T_2 \times S$ is ergodic. Hence for any ergodic S, $(T_1 \times T_2) \times S$ is ergodic. By Theorem 3.40 (4)⇒(1), $T_1 \times T_2$ is weak mixing. □

Corollary 3.43. *If T is weak mixing, then T^n is weak mixing for any $n \in N$.*

Proof. This can be proved directly by Theorem 3.40 (1)⇔(5) or (1)⇔(6). □

Remark 3.44. These two corollaries are false if $T_1, \cdots, T_k$ or T are just ergodic. Hence ergodic property doesn't behave well under direct products but weak mixing property does.

4 Interplay: Topological Dynamics & Ergodic Theory

Theorem 4.1 (Existence)**.** *Let X be a compact metric space and $T : X \to X$ be continuous. Then there exists at least one ergodic invariant probability measure.*

To prove this theorem, we need some lemmas from functional analysis,. See Appendix B.

We prove the theorem by the following lemmas.

Lemma 4.2. *For any sequence of measures ν_n on X, we define*

$$\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} T_*^j \nu_n \in \mathcal{M}(X).$$

Then any limit of μ_n on a subsequence in the weak topology is T-invariant.*

Proof. Suppose $n_k \nearrow \infty$ so that $\mu_{n_k} \to \mu \in \mathcal{M}(X)$. Let $f \in C(X)$. Then

$$\begin{aligned}\int_X (f - f\circ T)\,\mathrm{d}\mu &= \lim_{k\to\infty}\int_X (f - f\circ T)\,\mathrm{d}\mu_{n_k} = \lim_{k\to\infty}\frac{1}{n_k}\sum_{j=0}^{n_k-1}\int_X (f - f\circ T)\,\mathrm{d}\left(T_*^j\nu_{n_k}\right)\\ &= \lim_{k\to\infty}\frac{1}{n_k}\sum_{j=0}^{n_k-1}\int_X (f - f\circ T)\circ T^j\,\mathrm{d}\nu_{n_k} = \lim_{k\to\infty}\sum_{j=0}^{n_k-1}\int_X \left(f\circ T^j - f\circ T^{j+1}\right)\mathrm{d}\nu_{n_k}\\ &= \lim_{k\to\infty}\frac{1}{n_k}\int_X \left(f - f\circ T^{n_k}\right)\mathrm{d}\nu_{n_k} \leqslant \lim_{k\to\infty}\frac{1}{n_k}\cdot 2\cdot \|f\|_\infty = 0.\end{aligned}$$

Hence

$$\int_X f\,\mathrm{d}\mu = \int_X f\circ T\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\left(T_*\mu\right),$$

i.e. μ and $T_*\mu$ give all $f \in C(X)$ same integral. This shows $\mu = T_*\mu$, i.e. μ is T-invariant. $\square$

Lemma 4.3. *The set $\mathcal{M}(X)^T := \{\mu : \mu$ is a T-invariant probability measure on $X\}$ is a convex weak*-compact subset of $\mathcal{M}(X)$.*

Proof. If μ_0, μ_1 are T-invariant probability measure and $s \in [0,1]$, then it's easy to check that $\mu := (1-s)\mu_0 + s\mu_1$ is also a T-invariant probability measure. Hence $\mathcal{M}(X)$ is convex.

By Theorem B.3, $\mathcal{M}(X)$ is compact in weak* topology. Hence we only need to prove that $\mathcal{M}(X)^T \subset \mathcal{M}(X)$ is closed. For any $f \in C(X)$, both

$$\mu \in \mathcal{M}(X) \mapsto \int_X f\,\mathrm{d}\mu \quad \text{and} \quad \mu \in \mathcal{M}(X) \mapsto \int_X f\circ T\,\mathrm{d}\mu$$

are continuous. Hence $\left\{\mu \in \mathcal{M}(X) : \int_X f\,\mathrm{d}\mu = \int_X f\circ T\,\mathrm{d}\mu\right\}$ is closed. Note that a probability measure μ is T-invariant if and only if for any $f \in C(X)$,

$$\int_X f\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\left(T_*\mu\right) = \int_X f\circ T\,\mathrm{d}\mu.$$

Therefore

$$\mathcal{M}(X)^T = \bigcap_{f\in C(X)}\left\{\mu \in \mathcal{M}(X) : \int_X f\,\mathrm{d}\mu = \int_X f\circ T\,\mathrm{d}\mu\right\}$$

is an intersection of close sets, so it's closed. $\square$

Lemma 4.4. *μ is ergodic if and only if $\mu \in \mathcal{M}(X)^T$ is an extreme point, i.e. if $\mu = (1-s)\mu_0 + s\mu_1$ for some $\mu_0, \mu_1 \in \mathcal{M}(X)^T$ and $s \in (0,1)$, then $\mu = \mu_0 = \mu_1$.*

Proof. **($\Leftarrow$):** Suppose μ is not ergodic. Then there exists a T-invariant subset $B_0 \subset X$ with $\mu(B_0) \in (0,1)$. We define

$$\mu_0 = \frac{1}{\mu\left(B_0\right)}\mu|_{B_0} \quad \text{and} \quad \mu_1 = \frac{1}{1-\mu\left(B_0\right)}\mu|_{X-B_0}.$$

Then μ_0, μ_1 are both T-invariant: For any $B \subset X$,

$$\begin{aligned}\mu_0\left(T^{-1}B\right) &= \frac{1}{\mu\left(B_0\right)}\mu\left(T^{-1}B \cap B_0\right)\\ &= \frac{1}{\mu\left(B_0\right)}\mu\left(T^{-1}B \cap T^{-1}B_0\right) = \frac{1}{\mu\left(B_0\right)}\mu\left(T^{-1}\left(B\cap B_0\right)\right)\\ &= \frac{1}{\mu\left(B_0\right)}\mu\left(B\cap B_0\right) = \mu_0(B)\end{aligned}$$

and the same process works for μ_1. Note that $\mu_0, \mu_1 \neq \mu$ and

$$\mu = \mu\left(B_0\right)\mu_0 + \left(1 - \mu\left(B_0\right)\right)\mu_1$$

we get a contradiction.

($\Rightarrow$): Suppose $\mu = (1-s)\mu_0 + s\mu_1$ for some $s \in (0,1)$ and $\mu_0, \mu_1 \in \mathcal{M}(X)^T$. We wish to show that $\mu = \mu_0 = \mu_1$.

Note that we have $\mu_0 \ll \mu$, i.e. μ_0 is absolutely continuous, i.e. if $B \subset X$ has $\mu(B) = 0$, then we have $\mu_0(B) = 0$. Ny the Radon-Nikodym derivative theorem, there exists

$$f = \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu} \geqslant 0$$

so that

$$\mu_0(B) = \int_B f \,\mathrm{d}\mu$$

for any $B \subset X$. We define

$$B_0 = \{x : f(x) < 1\}$$

and note that $\mu(B_0) < 1$ because $1 = \mu_0(X) = \int_X f \,\mathrm{d}\mu$.

We calculate

$$\int_{B_0} f \,\mathrm{d}\mu = \mu_0 (B_0) = (T_* \mu_0)(B_0) = \mu_0 \left(T^{-1} B_0\right) = \int_{T^{-1}B_0} f \,\mathrm{d}\mu.$$

Then we abstract the intersection part of the integral domains:

$$\int_{B_0 - T^{-1}B_0} f \,\mathrm{d}\mu = \int_{T^{-1}B_0 - B_0} f \,\mathrm{d}\mu.$$

Note that for the left term, $f < 1$ on $B_0 - T^{-1}B_0$; for the right term $f \geqslant 1$ on $T^{-1}B_0 - B_0$. Toge ther with the fact that $\mu(B_0) = (T_*\mu)(B_0) = \mu(T^{-1}B_0)$ implies $\mu(B_0 - T^{-1}B_0) = \mu(T^{-1}B_0 - B_0)$, we must have

$$\mu \left(B_0 \Delta T^{-1} B_0\right) = \mu \left(\left(B_0 - T^{-1}B_0\right) \sqcup \left(T^{-1}B_0 - B_0\right)\right) = 0 + 0 = 0.$$

Because μ is ergodic, $\mu(B_0) \in \{0, 1\}$ and hence $\mu(B_0) = 0$. Note that μ, μ_0 are both probability measures, we have $f \overset{a.s.}{1}$. Hence $\mu = \mu_0$ and this also implies $\mu = \mu_1$. Therefore μ is extreme. □

Theorem 4.5 (Krein-Milman)**.** *Let $K \subset V$ be a convex compact subset of a locally convex vector space V. Then K is the close of all finite convex combinations of extreme points in K.*

Proof of Theorem 4.1. Let $K = \mathcal{M}(X)^T$ inside $V < C(X)^*$. By Lemma 4.2 and Lemma 4.3, $K \neq \varnothing$ is convex and compact in weak* topology. By Theorem 4.5, K has extreme points. By Lemma 4.4, this shows that the ergodic T-invariant probability measure exists. □

Theorem 4.6 (Ergodic Decomposition)**.** *Let X be compact and $T\colon X \to X$ be continuous. Let μ be a T-invariant probability measure on X. Then there exists a probability space $(\Omega, \mathcal{B}_\Omega, \rho)$ and a measurable function $\omega \in \Omega \mapsto \mu_\omega \in \mathcal{M}(X)^T$ so that μ_ω is ergodic a.s. and $\mu = \int_\Omega \mu_\omega \,\mathrm{d}\rho(\omega)$.*

This theorem can be proved by

- Choquet's theorem from functional analysis.
- Conditional measures. And this proof can give a bit more information.

Example 4.7. Let $X = \mathbb{T}^2$ and $R_{(\alpha,0)}\colon \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 + \alpha \\ x_2 \end{pmatrix}$ for an irrational α. In this case, the ergodic probability measures are precisely of the form $\lambda_{\mathbb{T}} \times \delta_{x_2}$ for $x_2 \in \mathbb{T}$. A measure of the form $\lambda_{\mathbb{T}} \times \mu$ for any μ would also be invariant. The ergodic decomposition in this case looks loke

$$\lambda_{\mathbb{T}} \times \mu = \int_{\mathbb{T}} \left(\lambda_{\mathbb{T}} \times \delta_{x_2}\right) \mathrm{d}\mu \left(x_2\right).$$

Example 4.8. $\mathcal{M}(X)^T$ can be crazy:

- Let $T = T_p$ on $\mathbb{T}$. The set of ergodic measures are dense in the space of all T-invariant probability measures on $\mathbb{T}$, i.e. the convex weak* compact set $\mathcal{M}(\mathbb{T})^{T_p}$. To see that, for any $\mu \in \mathcal{M}(\mathbb{T})^{T_p}$, by Theorem 4.5, we can approximate μ arbitrarily well in the weak* topology by a finite convex combination of ergodic measures: For any finite list $f_1, \cdots, f_k \in C(\mathbb{T})$ and $\varepsilon > 0$, there exist ergodic T_p-invariant probability measures $\mu_1, \cdots, \mu_l$ and $c_1, \cdots, c_l > 0$ with $c_1 + \cdots + c_l = 1$ so that

$$\left| \int_{\mathbb{T}} f_i \,\mathrm{d}\mu - \int_{\mathbb{T}} f_i \,\mathrm{d}\left(\sum_{j=1}^{l} c_j \mu_j\right)) \right| < \varepsilon \tag{12}$$

for all $i = 1, \cdots, k$. Applying Theorem 3.19 and Proposition 3.14 (1), for any $1 \leqslant i \leqslant k$, $1 \leqslant j \leqslant l$, for μ_j-a.e. $x \in \mathbb{T}$, we have

$$\frac{1}{n} \sum_{a=0}^{n} -1 f_i \left(T_p^a x\right) \to E\left(f_i | \mathbb{T}\right) \mathrm{d}\mu_j = \int_{\mathbb{T}} f_i \,\mathrm{d}\mu_j. \tag{13}$$

Let $x_j \in \mathbb{T}$ be a point satisfying equation (13) for all $i = 1, \cdots, k$. Let N be large and ;et $a_{j,1}, a_{j,2}, \cdots \in \{0, 1, \cdots, p-1\}$ be the base p-expansion of x_j. We define x to be the ratianl number whose base p-expansion equals

$$\overline{a_{1,1} a_{1,2} \cdots a_{1,\lceil c_1 N \rceil}, a_{2,1} \cdots a_{2,\lceil c_2 N \rceil} \cdots a_{l,1} \cdots a_{l,\lceil c_l N \rceil}}.$$

Then x is periodic with period close to N. We define $\widetilde{\mu}$ as the normalized counting measure on the orbit. Note that $\widetilde{\mu}$ is ergodic. As f_i is continuous, most points in the orbit of x up to $T_p^{\lceil c_1 N \rceil} x$ are very close to x_1 respect to its orbit points. We obtain that

$$\frac{1}{\lceil c_1 N \rceil} \sum_{a=0}^{\lceil c_1 N \rceil - 1} \left| f_i \left(T_p^a x\right) - f_i \left(T_p^a x_1\right) \right| < \varepsilon$$

for large N. Applying the same argument for $j = 2, \cdots, l$, we obtain in total an estimate of the form

$$\left| \int_{\mathbb{T}} f_i \,\mathrm{d}\mu - \int_{\mathbb{T}} f_i \,\mathrm{d}\widetilde{\mu_N} \right| < 7\varepsilon$$

for $i = 1, \cdots, k$.

- This is a non-dynamical example. Let

$$K = \{f \colon [0,1] \to \mathbb{R} : f(0) = 0,\ f \text{ is 1-Lipschitz}\} \subset C([0,1]).$$

Then K is convex, bounded and compact. Hence it has extreme points. And the set of extreme points are dense. To see that, note that for any function in $C([0,1])$ and its ε-neighborhood, we can find a polyline function in this neighborhood so that the slope of every segment is ± 1. These polyline functions are the extreme points. They are dense in $C([0,1])$ by finite linear combinations.

Definition 4.9. *Let X be compact, $T \colon X \to X$ be continuous and $\mu \in \mathcal{M}(X)^T$. Then $x \in X$ is called generic (for (T, μ)) if*

$$\frac{1}{n} \sum_{j=0}^{n-1} f\left(T^j x\right) \to \int_X f \,\mathrm{d}\mu$$

as $n \to \infty$ for all $f \in C(X)$.

Proposition 4.10. *Let X be compact, $T \colon X \to X$ be continuous and $\mu \in \mathcal{M}(X)^T$. Let μ be in addition ergodic. Then μ-a.e. x is generic.*

Proof. Let $D \subset C(X)$ be dense with respect to $\|\cdot\|_\infty$ and countable. By Theorem 3.19 applied to all $f \in D$, we have

$$\frac{1}{n} \sum_{j=0}^{n-1} f\left(T^j x\right) \to \int_X f \,\mathrm{d}\mu$$

for μ-a.e. x and all $f \in D$.

Claim. Those points are generic.

Proof of Claim. Let $f_0 \in C(X)$ arbitrarily. Let x be any point as above. For any $\varepsilon > 0$ there exists $f \in D$ with $\|f - f_0\|_\infty < \varepsilon$. Then

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f_0\left(T^j x\right) &\leqslant \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left(f\left(T^j x\right) + \varepsilon\right) \leqslant \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(T^j x\right) + \varepsilon \\ &= \int_X f \,\mathrm{d}\mu + \varepsilon \leqslant \int_X \left(f_0 + \varepsilon\right) \mathrm{d}\mu + \varepsilon = \int_X f_0 \,\mathrm{d}\mu + 2\varepsilon.\end{aligned}$$

This shows that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f_0\left(T^j x\right) = \int_X f_0 \,\mathrm{d}\mu.$$

Above all, those points are generic. ☐

□

Definition 4.11. *(X,T) is called uniquely ergodic if there exists only one T-invariant probability measure on X.*

Theorem 4.12. *The followings are equivalent:*

1. *(X,T) is uniquely ergodic.*
2. *There exists only one T-invariant ergodic probability measure.*
3. *For any $f \in C(X)$, there exists a constant C_f so that*

$$\frac{1}{n}\sum_{j=0}^{n-1} f\left(T^j x\right) \to C_f$$

for every $x \in X$.

4. *Same as in (3) with uniform convergence.*

Proof. **(1)⇒(2)** is by definition.

(2)⇒(1) follows Theorem 4.5.

(1)⇒(3): Let $\mu \in \mathcal{M}(X)^T$ be the unique T-invariant probability measure on X. Let $f \in C(X)$ and $x \in X$ arbitrarily. Then any weak* limit of $\frac{1}{n}\sum_{j=0}^{n-1} T_*^j\delta_x$ is T-invariant by Lemma 4.2, where δ_x is the Dirac measure on x. However, uniqueness means that μ is the only limit of any converging subsequence. This implies that

$$\frac{1}{n}\sum_{j=0}^{n-1} T_*^j\delta_x \to \mu$$

because if it doesn't, then there exists a neighborhood N of μ in $\mathcal{M}(X)^T$ and a subsequence staying outside N. As $\mathcal{M}(X)^T$ is compact, this subsequence would have a subsequence converging to μ. This gives a contradiction. Hence

$$\frac{1}{n}\sum_{j=0}^{n-1} f\left(T^j x\right) = \frac{1}{n}\sum_{j=0}^{n-1}\int_X f \,\mathrm{d}\left(T_*^j\delta_x\right) = \int_X f\,\mathrm{d}\left(\frac{1}{n}\sum_{j=0}^{n-1} T_*^j\delta_x\right) \to \int_X f\,\mathrm{d}\mu =: C_f.$$

Therefore, $C_f = \int_X f\,\mathrm{d}\mu$ is as desired.

(3)⇒(2): By Theorem 4.1, we only need to prove the uniqueness. Suppose μ_1, μ_2 are both T-invariant and ergodic. By Proposition 4.10, we can suppose x_i be generic for μ_i for $i = 1, 2$. Then by definition, for any $f \in C(X)$,

$$\frac{1}{n}\sum_{k=0}^{n-1} f\left(T^k x_i\right) \to \int_X f\,\mathrm{d}\mu_i = C_f.$$

Hence

$$\int_X f\,\mathrm{d}\mu_1 = C_f = \int_X f\,\mathrm{d}\mu_2$$

for any $f \in C(X)$. This shows $\mu_1 = \mu_2$.

(4)⇒(3) is obvious.

(1)⇒(4): By (1)⇒(3), we have

$$\frac{1}{n}\sum_{j=0}^{n-1} f\left(T^j x\right) \to C_f = \int_X f\,\mathrm{d}\mu$$

for any $x \in X$, where μ is the unique T-invariant probability measure. Suppose it is not uniform convergent. Then there exists an $\varepsilon >$ so that for any N, we can find some $x_N \in X$ and $n_N > N$ so that

$$\left|\frac{1}{n_N}\sum_{k=0}^{n_N-1} f\left(T^k x_N\right) - \int_X f\,\mathrm{d}\mu\right| > \varepsilon. \tag{14}$$

As the proof of (1)⇒(3), consider the sequence of N: $\left\{ \frac{1}{n_N} \sum_{j=0}^{n_N-1} T_*^j \delta_{x_N} \right\}_{N=0}^{\infty}$. Note that $\mathcal{M}(X)^T$ is compact (Lemma 4.3), it has some weak* convergent subsequence. Its any weak* limit is T-invariant (Lemma 4.2) and will be μ. But equation (14) shows that

$$\left| \int_X f \,\mathrm{d} \left(\frac{1}{n_N} \sum_{j=0}^{n_N-1} T_*^j \delta_{x_N} \right) - \int_X f \,\mathrm{d}\mu \right| = \left| \frac{1}{n_N} \sum_{k=0}^{n_N-1} f\left(T^k x_N\right) - \int_X f \,\mathrm{d}\mu \right| > \varepsilon$$

for any N. Hence any subsequence of $\left\{ \frac{1}{n_N} \sum_{j=0}^{n_N-1} T_*^j \delta_{x_N} \right\}_{N=0}^{\infty}$ can not convergence to μ. We get a contradiction.

□

Theorem 4.13 (Weyl)**.** *For $\alpha \in \mathbb{R} - \mathbb{Q}$ and $f \in C(\mathbb{T})$, we have*

$$\frac{1}{n} \sum_{j=0}^{n-1} f\left(j^2 \alpha\right) \to \int_{\mathbb{T}} f \,\mathrm{d}x.$$

This theorem can be deduced by

Theorem 4.14 (Furstenberg)**.** *For $\alpha \in \mathbb{R} - \mathbb{Q}$, we define $T \colon \mathbb{T}^2 \to \mathbb{T}^2$, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 + 2x_2 + \alpha \\ x_2 + \alpha \end{pmatrix}$. Then T is uniquely ergodic with $\lambda_{\mathbb{T}^2}$ being the only T-invariant probability measure on $\mathbb{T}^2$.*

Proof of Theorem4.13 using Theorem 4.14. Consider the orbit of $x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ under T. We have $T^j x = \begin{pmatrix} j^2\alpha \\ j\alpha \end{pmatrix}$. Hence for any $f \in C(\mathbb{T})$, let $F(x, y) := f(x) \in C(\mathbb{T}^2)$. Then

$$\frac{1}{n} \sum_{j=0}^{n-1} F\left(T^j x\right) \to \int_{\mathbb{T}^2} F \,\mathrm{d}x\mathrm{d}y.$$

For the left side,

$$F\left(T^j x\right) = F\left(j^2\alpha, j\alpha\right) = f\left(j^2\alpha\right).$$

For the right side,

$$\int_{\mathbb{T}^2} F \,\mathrm{d}x\mathrm{d}y = \int_0^1 \int_0^1 f \,\mathrm{d}x 1 \,\mathrm{d}y = \int_{\mathbb{T}} f \,\mathrm{d}x.$$

Therefore

$$\frac{1}{n} \sum_{j=0}^{n-1} f\left(j^2 \alpha\right) \to \int_{\mathbb{T}} f \,\mathrm{d}x.$$

□

Proof of Theorem 4.14. **Step 1: $\lambda_{\mathbb{T}^2}$ is ergodic.**

We use Fourier series and Lemma 3.26. Suppose $f \in L^2(\mathbb{T}^2, \lambda_{\mathbb{T}^2})$ is an eigenfunction of U_T for eigenvalue 1, i.e. $f \circ T = f$. Suppose f has Fourier expansion

$$f = \sum_{\underline{n} \in \mathbb{Z}^2} c_{\underline{n}} \mathrm{e}^{2\pi\mathrm{i}\underline{n}\underline{x}}.$$

Then

$$f \circ T = \sum_{\underline{n} \in \mathbb{Z}^2} c_{\underline{n}} \mathrm{e}^{2\pi\mathrm{i}\langle \underline{n}, T\underline{x} \rangle} = \sum_{\underline{n} \in \mathbb{Z}^2} c_{\underline{n}} \mathrm{e}^{2\pi\mathrm{i}(n_1 x_2 + (n_2 + 2n_1)x_2 + (n_1 + n_2)\alpha)} =: \sum_{\underline{m} \in \mathbb{Z}^2} c_{\underline{m}} \mathrm{e}^{2\pi\mathrm{i}\underline{m}\underline{x}}.$$

We have

$$c_{\underline{n}} \mathrm{e}^{2\pi\mathrm{i}(n_1+n_2)\alpha} = c_{(n_1, n_2 + 2n_1)} \tag{15}$$

for all $\underline{n} \in \mathbb{Z}$.

For $n_1 \neq 0$, we use the fact that $\left|c_{\underline{n}}\right| = \left|c_{(n_1, n_2+2n_1)}\right|$ and square-summability of the coefficients to obtain $c_{\underline{n}} = 0$ for any $\underline{n} \in \mathbb{Z}^2$ with $n_1 \neq 0$. If $n_1 = 0$, equation 15 becomes

$$c_{(0,n_2)} \mathrm{e}^{2\pi\mathrm{i}n_2\alpha} = c_{(0,n_2)} \neq 1$$

unless $n_2 = 0$. This shows that $c_{(0,n_2)} = 0$ for $n_2 \neq 0$. Above all, f is a.s. a constant, i.e. U_T has only one 1-dimensional eigenspace for eigenvalue 1. Hence by Lemma 3.26, $\lambda_{\mathbb{T}^2}$ is ergodic.

Step 2: $(\mathbb{T}, R_\alpha, \lambda_{\mathbb{T}})$ **is uniquely ergodic.**

We can use Fourier series again.

For $x \in \mathbb{T}$ and $k \neq 0$, we have

$$\frac{1}{n}\sum_{j=0}^{n-1} \mathrm{e}^{2\pi \mathrm{i} k R_\alpha^j(x)} = \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{e}^{2\pi \mathrm{i} k(x+j\alpha)} = \frac{\mathrm{e}^{2\pi \mathrm{i} kx}}{n}\sum_{j=0}^{n-1}\left(\mathrm{e}^{2\pi \mathrm{i} k\alpha}\right)^j = \frac{\mathrm{e}^{2\pi \mathrm{i} kx}}{n}\frac{\left(\mathrm{e}^{2\pi \mathrm{i} k\alpha}\right)^{n+1}-1}{\mathrm{e}^{2\pi \mathrm{i} k\alpha}-1} \to 0 = \int_{\mathbb{T}} \mathrm{e}^{2\pi \mathrm{i} kx}\,\mathrm{d}\lambda_{\mathbb{T}}$$

as $n \to \infty$. Hence by taking linear combinations, we have

$$\frac{1}{n}\sum_{j=0}^{n-1} f\left(R_\alpha^j x\right) \to \int_{\mathbb{T}} f\,\mathrm{d}\lambda_{\mathbb{T}}$$

for any triangle-polynomial f. Then by uniform approximation of $f \in C(\mathbb{T})$, this formula also holds for constant functions. This is equivalent to say that

$$\int_{\mathbb{T}} f\,\mathrm{d}\left(\frac{1}{n}\sum_{j=0}^{n-1}\left(R_\alpha^j\right)_* \delta_x\right) = \frac{1}{n}\sum_{j=0}^{n-1} f\left(R_\alpha^j x\right) \to \int_{\mathbb{T}} f\,\mathrm{d}\lambda_{\mathbb{T}}$$

for any constant function f. Hence we have

$$\frac{1}{n}\sum_{j=0}^{n-1}\left(R_\alpha^j\right)_* \delta_x \to \lambda_{\mathbb{T}}.$$

Then for any $f \in C(\mathbb{T})$,

$$\frac{1}{n}\sum_{j=0}^{n-1} f\left(R_\alpha^j x\right) = \int_{\mathbb{T}} f\,\mathrm{d}\left(\frac{1}{n}\sum_{j=0}^{n-1}\left(R_\alpha^j\right)_* \delta_x\right) \to \int_{\mathbb{T}} f\,\mathrm{d}\lambda_{\mathbb{T}} =: C_f$$

for any $x \in \mathbb{T}$. By Theorem 4.12, $(\mathbb{T}, R_\alpha, \lambda_{\mathbb{T}})$ is unique ergodic .

Step 3: $\frac{1}{n}\sum_{j=0}^{n-1} T_*^j\left(\lambda_{\mathbb{T}} \times \delta_{x_2}\right) \to \lambda_{\mathbb{T}^2}$ **in weak* topology.**

By definition. for any $f \in C(\mathbb{T}^2)$,

$$\begin{aligned}\int_{\mathbb{T}^2} f\,\mathrm{d}\left(\frac{1}{n}\sum_{j=0}^{n-1} T_*^j\left(\lambda_{\mathbb{T}} \times \delta_{x_2}\right)\right) &= \frac{1}{n}\sum_{j=0}^{n-1}\int_{\mathbb{T}^2} f\left(y_1, y_2\right)\,\mathrm{d}\left(T_*^j\left(\lambda_{\mathbb{T}} \times \delta_{x_2}\right)\right) \\ &= \frac{1}{n}\sum_{j=0}^{n-1}\int_{\mathbb{T}^2} f \circ T^j\left(y_1, y_2\right)\,\mathrm{d}\left(\lambda_{\mathbb{T}} \times \delta_{x_2}\right)\left(y_1, y_2\right)\end{aligned}$$

For any j,

$$f \circ T^j\left(y_1, y_2\right) = f\left(y_1 + a y_2 + b\alpha, y_2 + j\alpha\right)$$

for some integers a, b. But after integral,

$$\begin{aligned}\int_{\mathbb{T}^2} f \circ T^j\left(y_1, y_2\right)\,\mathrm{d}\left(\lambda_{\mathbb{T}} \times \delta_{x_2}\right)\left(y_1, y_2\right) &= \int_{\mathbb{T}^2} f\left(y_1 + a y_2 + b\alpha, y_2 + j\alpha\right)\,\mathrm{d}\left(\lambda_{\mathbb{T}} \times \delta_{x_2}\right)\left(y_1, y_2\right) \\ &= \int_{\mathbb{T}} f\left(y_1 + a x_2 + b\alpha, x_2 + j\alpha\right)\,\mathrm{d}\lambda_{\mathbb{T}}\left(y_1\right) \\ &= \int_{\mathbb{T}} f\left(y_1, x_2 + j\alpha\right)\,\mathrm{d}\lambda_{\mathbb{T}}\left(y_1\right) =: F\left(x_2 + j\alpha\right) \in C(\mathbb{T})\end{aligned}$$

because the Lebesgue measure $\lambda_{\mathbb{T}}$ is invariant under the shift $y_1 \mapsto y_1 + a x_2 + b\alpha$. Then by step 2 and Fubini Theorem,

$$\frac{1}{n}\sum_{j=0}^{n-1} F\left(x_2 + j\alpha\right) = \frac{1}{n}\sum_{j=0}^{n-1} F\left(R_\alpha^j x_2\right) \to \int_{\mathbb{T}} F\,\mathrm{d}\lambda_{\mathbb{T}} = \int_{\mathbb{T}^2} f\,\mathrm{d}\lambda_{\mathbb{T}^2}.$$

Above all, for any $f \in C(\mathbb{T}^2)$, as $n \to \infty$,

$$\int_{\mathbb{T}^2} f \,\mathrm{d}\left(\frac{1}{n}\sum_{j=0}^{n-1} T_*^j\left(\lambda_{\mathbb{T}} \times \delta_{x_2}\right)\right) = \frac{1}{n}\sum_{j=0}^{n-1}\int_{\mathbb{T}^2} f \circ T^j\left(y_1, y_2\right) \mathrm{d}\left(\lambda_{\mathbb{T}} \times \delta_{x_2}\right)\left(y_1, y_2\right)$$
$$= \frac{1}{n}\sum_{j=0}^{n-1} F\left(x_2 + j\alpha\right) \to \int_{\mathbb{T}^2} f \,\mathrm{d}\lambda_{\mathbb{T}^2}.$$

Hence we have $\frac{1}{n}\sum_{j=0}^{n-1} T_*^j\left(\lambda_{\mathbb{T}} \times \delta_{x_2}\right) \to \lambda_{\mathbb{T}^2}$ in weak* topology, which is as desired.

Step 4: Let $0 < \eta < 1$ and define $\nu_\eta = \frac{1}{\eta}\lambda_{[x_1, x_1+\eta]} \times \delta_{x_2}$. Then any weak* limit μ_η of $\left\{\frac{1}{n}\sum_{j=0}^{n-1} T_*^j \nu_\eta\right\}_{n=0}^{\infty}$ is T-invariant and $\mu_\eta = \lambda_{\mathbb{T}^2}$.

Also define
$$\widetilde{\nu_\eta} = \frac{1}{1-\eta}\lambda_{\mathbb{T}-[x_1, x_1+\eta]} \times \delta_{x_2}.$$

Note that
$$\eta\nu_\eta + (1-\eta)\widetilde{\nu_\eta} = \lambda_{\mathbb{T}} \times \delta_{x_2}.$$

By Lemma 4.2, any weak* limit μ_η is T-invariant. Suppose μ_η is the limit of the subsequence along n_k. Then we assume $\widetilde{\mu_\eta}$ is a weak* limit of $\left\{\frac{1}{n_k}\sum_{j=0}^{n_k-1} T_*^j \widetilde{\nu_\eta}\right\}_{k=0}^{\infty}$. By Lemma 4.2, $\widetilde{\mu_\eta}$ is also T-invariant.

Claim. $\eta\mu_\eta + (1-\eta)\widetilde{\mu_\eta} = \lambda_{\mathbb{T}^2}$.

Proof of Claim. Let $f \in C(\mathbb{T}^2)$ arbitrarily. Calculating directly, we have

$$\int_{\mathbb{T}^2} f \,\mathrm{d}\left(\eta\mu_\eta + (1-\eta)\widetilde{\mu_\eta}\right) = \eta\int_{\mathbb{T}^2} f \,\mathrm{d}\mu_\eta + (1-\eta)\int_{\mathbb{T}^2} f \,\mathrm{d}\widetilde{\mu_\eta}$$
$$= \lim_{k\to\infty}\int_{\mathbb{T}^2} f \,\mathrm{d}\left(\frac{1}{n_k}\sum_{j=0}^{n_k-1} T_*^j\left(\nu_\eta + \widetilde{\nu_\eta}\right)\right)$$
$$= \lim_{k\to\infty}\int_{\mathbb{T}^2} f \,\mathrm{d}\left(\frac{1}{n_k}\sum_{j=0}^{n_k-1} T_*^j\left(\lambda_{\mathbb{T}} \times \delta_{x_2}\right)\right) \overset{step3}{=} \int_{\mathbb{T}^2} f \,\mathrm{d}\lambda_{\mathbb{T}^2}.$$

Therefore we have $\eta\mu_\eta + (1-\eta)\widetilde{\mu_\eta} = \lambda_{\mathbb{T}^2}$. □

Then by step 1, $\lambda_{\mathbb{T}^2}$ is ergodic. Because $\mu_\eta, \widetilde{\mu_\eta}$ are T-invariant, by Lemma 4.4, we have $\mu_\eta = \lambda_{\mathbb{T}^2}$. Moreover, by compactness,
$$\frac{1}{n}\sum_{j=0}^{n-1} T_*^j \nu_\eta \to \lambda_{\mathbb{T}^2}$$
as $n \to \infty$ without the need to consider a subsequence.

Step 5: Use uniform continuity of $f \in C(\mathbb{T}^2)$ to prove the theorem.

Let $f \in C(\mathbb{T}^2)$ arbitrarily. For any $\varepsilon > 0$, we can choose $\eta \in (0,1)$ such that for any $d(x_1, x_2) < \eta$, $|f(x_1) - f(x_2)| < \varepsilon$. We will use Theorem 4.12 (3)⇒(1). For any $(x_1, x_2) \in \mathbb{T}^2$, We calculate

$$\left|\frac{1}{n}\sum_{j=0}^{n-1} f\left(T^j\left(x_1, x_2\right)\right) - \int_{\mathbb{T}^2} f \,\mathrm{d}\lambda_{\mathbb{T}^2}\right| \leqslant \left|\frac{1}{n}\sum_{j=0}^{n-1} f\left(T^j\left(x_1, x_2\right)\right) - \int_{\mathbb{T}^2} f \,\mathrm{d}\left(\frac{1}{n}\sum_{j=0}^{n-1} T_*^j\left(\frac{1}{\eta}\lambda_{[x_1, x_1+\eta]} \times \delta_{x_2}\right)\right)\right|$$
$$+ \left|\int_{\mathbb{T}^2} f \,\mathrm{d}\left(\frac{1}{n}\sum_{j=0}^{n-1} T_*^j\left(\frac{1}{\eta}\lambda_{[x_1, x_1+\eta]} \times \delta_{x_2}\right)\right) - \int_{\mathbb{T}^2} f \,\mathrm{d}\lambda_{\mathbb{T}^2}\right|$$

For the second term, it goes to 0 as $n \to \infty$ by step 4. Now we calculate the first term.

$$\left|\frac{1}{n}\sum_{j=0}^{n-1} f\left(T^j\left(x_1, x_2\right)\right) - \int_{\mathbb{T}^2} f \,\mathrm{d}\left(\frac{1}{n}\sum_{j=0}^{n-1} T_*^j\left(\frac{1}{\eta}\lambda_{[x_1, x_1+\eta]} \times \delta_{x_2}\right)\right)\right|$$
$$= \left|\frac{1}{n}\sum_{j=0}^{n-1}\left(f\left(T^j\left(x_1, x_2\right)\right) - \frac{1}{\eta}\int_{x_1}^{x_1+\eta} f \circ T^j\left(y_1, x_2\right) \mathrm{d}\lambda\left(y_1\right)\right)\right|$$
$$= \left|\frac{1}{n}\sum_{j=0}^{n-1}\frac{1}{\eta}\int_{x_1}^{x_1+\eta}\left(f\left(T^j\left(x_1, x_2\right)\right) - f \circ T^j\left(y_1, x_2\right)\right) \mathrm{d}\lambda\left(y_1\right)\right| \leqslant \left|\frac{1}{n}\sum_{j=0}^{n-1}\frac{1}{\eta}\int_{x_1}^{x_1+\eta} \varepsilon \,\mathrm{d}\lambda\left(y_1\right)\right| = \varepsilon.$$

Above all, we have

$$\lim_{n\to\infty}\left|\frac{1}{n}\sum_{j=0}^{n-1} f\left(T^j\left(x_1,x_2\right)\right)-\int_{\mathbb{T}^2} f\,\mathrm{d}\lambda_{\mathbb{T}^2}\right|\leqslant\varepsilon$$

i.e.

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f\left(T^j\left(x_1,x_2\right)\right)=\int_{\mathbb{T}^2} f\,\mathrm{d}\lambda_{\mathbb{T}^2}=:C_f$$

for any $(x_1,x_2)\in\mathbb{T}^2$. By Theorem 4.12 (3)$\Rightarrow$(1), $(\mathbb{T}^2,T)$ is unique ergodic with $\lambda_{\mathbb{T}^2}$ being the only T-invariant probability measure.

□

5 Introduction to Hyperbolic Surfaces

I omit some details in this section because I have learnt some hyperbolic geometry before.

5.1 Hyperbolic plane

Definition 5.1. *The hyperbolic plane is defined by*

$$\mathbb{H}:=\{x+\mathrm{i}y:y>0\}\subset\mathbb{C}$$

with a metric d defined by

$$d(p,q):=\inf\left\{\ell(\phi):\phi\colon[a,b]\to\mathbb{H},\phi(a)=p,\phi(b)=q\right\},$$

where we use the inner product at $z=x+\mathrm{i}y$

$$\langle V,W\rangle_z:=\frac{1}{y^2}\langle V,W\rangle_{\mathbb{C}}$$

to calculate the length $\ell(\phi)$ of any curve ϕ.

Remark 5.2. We note that d induces the same topology on $\mathbb{H}$.

Remark 5.3. Direct calculation shows that

$$d\left(y_0\mathrm{i},y_1\mathrm{i}\right)=\ln\frac{y_1}{y_0}.$$

5.2 Isometries

Proposition 5.4. *For any $\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\mathrm{SL}_2(\mathbb{R})$, the map $z\mapsto\frac{az+b}{cz+d}$ defined on $\mathbb{H}$ is an isometry.*

Remark 5.5. $-\mathrm{Id}_2$ acts trivially on $\mathbb{H}$. Actually, $\mathrm{PSL}_2(\mathbb{R}):=\mathrm{SL}_2(\mathbb{R})/\pm\mathrm{Id}_2$ acts faithfully on $\mathbb{H}$.

Remark 5.6. This action can be extended to an isometric action $D.$ on $T\mathbb{H}$: For any $g=\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\mathrm{SL}_2(\mathbb{R})$, we define

$$D_g(z,v):=\left(g(z),g'(z)v\right)=\left(gz,\frac{v}{(cz+d)^2}\right).$$

Proposition 5.7. *1.* $\mathrm{PSL}_2(\mathbb{R})$ *acts transitively on* $\mathbb{H}$.

2. $\mathrm{Stab}_{\mathrm{PSL}_2(\mathbb{R})}(\mathrm{i})=\mathrm{SO}_2(\mathbb{R})$.

3. $\mathbb{H}=\mathrm{PSL}_2(\mathbb{R})/\mathrm{PSO}_2(\mathbb{R})\cong\mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$.

Proof. **(1):** By direct calculations, i can be acted by $\begin{pmatrix}\sqrt{y}&0\\0&\frac{1}{\sqrt{y}}\end{pmatrix}$ and $\begin{pmatrix}1&x\\0&1\end{pmatrix}$ to any $z=x+\mathrm{i}y$.

(2): This can be proved by solving the equation

$$\mathrm{i}=\begin{pmatrix}a&b\\c&d\end{pmatrix}\mathrm{i}=\frac{a\mathrm{i}+b}{c\mathrm{i}+d}$$

directly.

(3) is given by (1) and (2).

□

Note that $D_g \curvearrowright T\mathbb{H}$ can not be transitive because it preserves $|v|$ and then any orbit only has tangent vectors of the same length. We consider the unit tangent bundle $T^1\mathbb{H} = \{(z,v) : |v| = 1\}$.

Proposition 5.8. $\mathrm{SL}_2(\mathbb{R})$ *acts on* $T^1\mathbb{H}$ *transitively without a stabilizer.*

Proof. By Proposition 5.7 (1)(2), we only need to show that $\mathrm{SO}_2(\mathbb{R})$ acts transitively on $T^1_{\mathrm{i}}\mathbb{H} = \{(\mathrm{i}, v) : |v| = 1\}$.

Denote $h_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \in \mathrm{SO}_2(\mathbb{R})$. We have

$$D_{h_\theta}(\mathrm{i},\mathrm{i}) = \left(\mathrm{i}, \mathrm{e}^{\pi\mathrm{i}\theta}\mathrm{i}\right).$$

Hence any $\left(\mathrm{i}, \mathrm{e}^{\pi\mathrm{i}\theta}\mathrm{i}\right)$ can be obtain by action of D_{h_θ} on (i,i). Therefore the action $\mathrm{SO}_2(\mathbb{R}) \curvearrowright T^1_{\mathrm{i}}\mathbb{H}$ is transitive. Moreover, $(\mathrm{i},\mathrm{i}) = \left(\mathrm{i}, \mathrm{e}^{\pi\mathrm{i}\theta}\mathrm{i}\right)$ if and only if $\theta = k\pi$, if and only if $h_\theta = \pm\mathrm{Id}_2$. Hence there isn't any non-trivial stabilizer. ☐

Remark 5.9. By Proposition 5.8, we have $\mathrm{PSL}_2(\mathbb{R}) \cong T^1\mathbb{H}$ given by $g \mapsto D_g(\mathrm{i},\mathrm{i})$. Hence the action $D_g\colon \mathrm{PSL}_2(\mathbb{R}) \curvearrowright T^1\mathbb{H}$ induces an action $\mathrm{PSL}_2(\mathbb{R}) \curvearrowright \mathrm{PSL}_2(\mathbb{R})$ by left-multiplications.

5.3 Understand all geodesics

After normalization, we assume that all geodesics are with speed 1.

For any $z_1, z_2 \in \mathbb{H}$, there exists unique element $g \in \mathrm{PSL}_2(\mathbb{R})$ so that it maps the vertical segment of length $d(z_1,z_2)$ start at i to the geodesic segment from z_1 to z_2. This can be done as follows: The Möbius translations act sharply 3-transitively on $P^1(\mathbb{C})$: For any $z_i \neq w_i$, $i = 1,2,3$, there exists unique $g \in \mathrm{GL}_2(\mathbb{R})$ so that $g(z_1,z_2,z_3) = (w_1,w_2,w_3)$. This is because the map $z \mapsto \frac{(z-z_1)(z-z_2)}{(z-z_3)(z_2-z_1)}$ maps $(0,\mathrm{i},\infty)$ to (z_1,z_2,z_3). Denote the intersections of x-axis and the geodesic (a semi circle vertical to x-axis) through z_1, z_2 by x, x'. Then we find unique isometry g maps $(0,\mathrm{i},\infty)$ to (x, z_1, x'). Note that g is an isometry, it keeps geodesics. g maps $y > 0$-axis to the geodesic through z_1, z_2. g also keeps length, hence g is as desired.

Definition 5.10. *The geodesic flow on* $T^1\mathbb{H}$ *is defined by*

$$g_t(z,v) = \textit{following the geodesic through } (z,v) \textit{ for time } t.$$

Remark 5.11. $g_t(z,v) = D_g \cdot \mathrm{e}^t(\mathrm{i},\mathrm{i})$, where g is the unique element with $D_g \cdot (\mathrm{i},\mathrm{i}) = (z,v)$. Note that

$$g_t(z,v) = D_g \cdot g_t(\mathrm{i},\mathrm{i}) = D_g \cdot D_{a_t^{-1}}(\mathrm{i},\mathrm{i}) = D_{g\cdot a_t^{-1}} \cdot (\mathrm{i},\mathrm{i})$$

where $a_t = \begin{pmatrix} \mathrm{e}^{-\frac{t}{2}} & \\ & \mathrm{e}^{\frac{t}{2}} \end{pmatrix}$. We have a corresponding

$$\text{right multiplication by } a_t^{-1} \longleftrightarrow \text{geodesic flow.}$$

6 Geodesic Flow

For a sphere $\mathbb{S}^2$ with constant curvature $+1$, geodesics on $\mathbb{S}^2$ are great circles. Hence its geodesics are all periodic. For a torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ with constant curvature 0, its geodesics depend on their slope. Hence they are either periodic or with 2-dimensional closure. No geodeisc is dense in $T^1(\mathbb{T}^2)$. But on $\mathbb{H}^2$, all geodesics are same: any geodesic can be mapped to any another geodesic by some isometry of $\mathbb{H}^2$.

Goal. *We wish to define quotients of* $\mathbb{H}^2$ *– maybe compact – so that the geodesic flow on it is more interesting that on* $\mathbb{H}^2$ *itself.*

We recall and conclude the last section with

Theorem 6.1.

1. $\mathrm{PSL}_2(\mathbb{R})$ *acts on* $\mathbb{H}^2$ *transitively preserving hyperbolic distance.*
2. *By induced action* D_g *of* $g \in \mathrm{PSL}_2(\mathbb{R})$, $\mathrm{PSL}_2(\mathbb{R})$ *acts on* $T^1\mathbb{H}^2$ *simply transitive.*
3. *Using* (i,i)*, we have* $\mathrm{PSL}_2(\mathbb{R}) \cong T^1\mathbb{H}^2$ *by* $g \mapsto D_g(\mathrm{i},\mathrm{i}) = \left(g\cdot\mathrm{i}, \frac{1}{(cz+d)^2}\mathrm{i}\right)$.

4. *The geodesic orbit for* (i,i) *is* $(\mathrm{e}^t\mathrm{i},\mathrm{e}^t\mathrm{i}) \in T^1\mathbb{H}^2$ *for* $t \in \mathbb{R}$. *For* $D_g(\mathrm{i},\mathrm{i})$, *it geodesic orbit is* $D_g\left(\mathrm{e}^t\mathrm{i},\mathrm{e}^t\mathrm{i}\right)$. *Denote* $a_t = \begin{pmatrix} \mathrm{e}^{-\frac{t}{2}} & \\ & \mathrm{e}^{\frac{t}{2}} \end{pmatrix}$.
$$D_g\left(\mathrm{e}^t\mathrm{i},\mathrm{e}^t\mathrm{i}\right) = D_g D_{a_t^{-1}}(\mathrm{i},\mathrm{i}).$$
Within $\mathrm{PSL}_2(\mathbb{R})$, *the geodesic throught* g *looks like* ga_t^{-1} *for* $t \in \mathbb{R}$. *Note that the Möbius transformations correspond to left multiplication. Especially, if* $t = 1$, *the right multiplication by* a_1^{-1} *is called the time-one-map.*

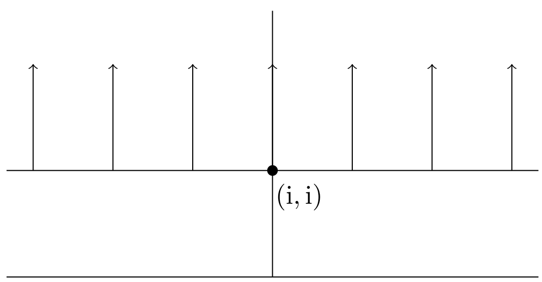
5. *Right-multiplication by* $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ *on* $g \in \mathrm{PSL}_2(\mathbb{R})$ *corresponds to rotating tangent vectors with angle* $\pm 2\theta$ *in* $T^1\mathbb{H}^2$ *without changing the base point.*
$$K := \left\{ k_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\} = \mathrm{Stab}_{\mathrm{PSL}_2(\mathbb{R})}(\mathrm{i}).$$
Note that the derivative of k_θ *is multiplication by* $\frac{1}{(\mathrm{i}\cos\theta - \sin\theta)^2} = \cos 2\theta + \mathrm{i}\sin 2\theta = \mathrm{e}^{2\theta\mathrm{i}}$. *As this happens at* (i,i), *the transitive* $\mathrm{PSL}_2(\mathbb{R})$*-action propagates this property to all of* $\mathrm{PSL}_2(\mathbb{R}) \cong T^1\mathbb{H}^2$*:*
$$D_{gk_\theta}(\mathrm{i},\mathrm{i}) = D_g\left(\mathrm{i},\mathrm{e}^{2\theta}ii\right) = \left(g\cdot\mathrm{i}, \frac{\mathrm{e}^{2\theta\mathrm{i}}}{(cz+d)^2}\mathrm{i}\right).$$

Definition 6.2. *The (stable) horocycle flow is defined/corresponds to right multiplication by upper unipotent matrices*
$$U^- = \left\{ \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} : s \in \mathbb{R} \right\}.$$

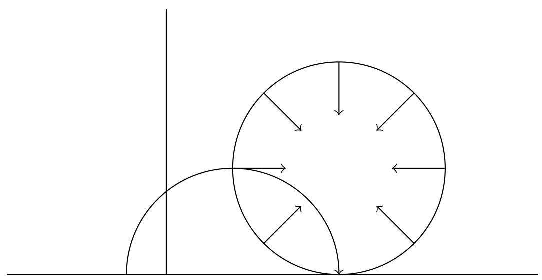
Starting at (i,i), we obtain

- In $\mathrm{PSL}_2(\mathbb{R})$, $\mathrm{Id} \mapsto \mathrm{Id}\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$.
- In $T^1\mathbb{H}^2$, $(\mathrm{i},\mathrm{i}) \mapsto D_{\left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)}(\mathrm{i},\mathrm{i}) = (\mathrm{i}+s,\mathrm{i})$.

One can check that horocycle orbits of vectors pointing north are horizontal lines. Hence the stable horospherical orbit consists of all vectors for which the forward geodesic orbit is asymptotic to the given starting geodesic path.



As before, we now argue that this holds not only at (i,i) but everywhere. In general, the horocycles look like horizontal lines or circles tangent to x-axis. Any geodesic towards the intersection of a horocycle and the x-axis is perpendicular to the horocycle.



To achieve the goal of obtaining a compact (or finite volume) surface, we need a “big” discrete subgroup $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ so that $M = \Gamma\backslash\mathbb{H}^2$.

For $M = \mathbb{T}^2$, we have $\mathbb{R}^2$ up to the isometries defined by $\underline{n}\colon v \in \mathbb{R}^2 \mapsto v + \underline{n} \in \mathbb{R}^2$ where $\underline{n} \in \mathbb{Z}^2$.
Formally,

$$M = \Gamma\backslash\mathbb{H}^2 = \left\{\Gamma \cdot z : z \in \mathbb{H}^2\right\}$$

and

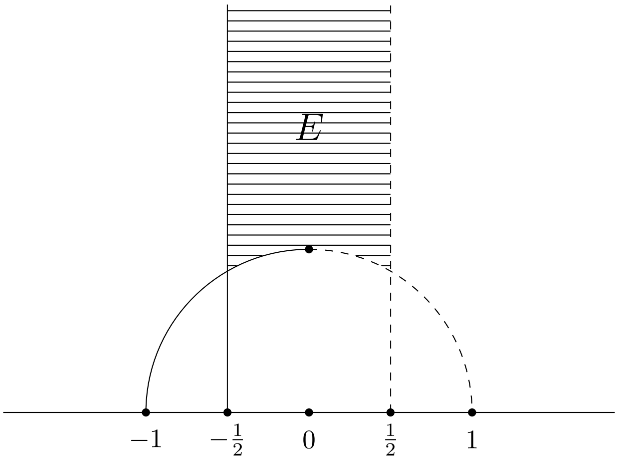
$$X = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R}) = \left\{\Gamma g : g \in \mathrm{PSL}_2(\mathbb{R})\right\}.$$

We will think of X as the unit tangent bundle of M.

Definition 6.3. *Let $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ be a subgroup. $E \subset \mathbb{H}^2$ is called a fundamental domain of Γ (or of $\Gamma\backslash\mathbb{H}^2$) if $\sharp(\Gamma \cdot z \cap E) = 1$ for all $z \in \mathbb{H}^2$.*

Theorem 6.4. *Let $\Gamma = \mathrm{PSL}_2(\mathbb{Z}) < \mathrm{PSL}_2(\mathbb{R})$. Then a fundamental domain of Γ is given by the triangle*

$$E := \left\{z = x + \mathrm{i}y \in \mathbb{H}^2 : -\frac{1}{2} \leqslant x < \frac{1}{2}, |z| < 1\right\} \cup \left\{z = x + \mathrm{i}y \in \mathbb{H}^2 : |z| = 1, -\frac{1}{2} \leqslant x < 0\right\}.$$



Proof. $g := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} : z \mapsto z + 1 \in \mathrm{PSL}_2(\mathbb{Z})$ maps $\mathbb{H}^2$ into the column $\left\{z = x + \mathrm{i}y \in \mathbb{H}^2 : -\frac{1}{2} \leqslant x \leqslant \frac{1}{2}\right\}$. $h := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} : z \mapsto -\frac{1}{z} \in \mathrm{PSL}_2(\mathbb{Z})$ maps inner of the the semicircle $\left\{z \in \mathbb{H}^2 : |z| = 1\right\}$ into the outer of the semicircle. On the semicircle, h maps the arc in the first quadrant to the are in the second quadrant. Hence the fundamental domain is contained in E. In fact, let $z \in \mathbb{H}$ be arbitrary. We wish to find $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ so that $\gamma \cdot z \in E$. Recall that

$$\mathrm{Im}\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z\right) = \frac{\mathrm{Im}(z)}{|cz + d|^2}. \tag{16}$$

As $\mathrm{Im}(z) > 0$ and $(c, d) \in \mathbb{Z}^2 - \{(0, 0)\}$, the imaginary part of $\gamma \cdot z$ can not be arbitrarily big for any $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$. Suppose the maximum of $\mathrm{Im}(\gamma \cdot z)$ is achieved at $\gamma_0 \cdot z$. We apply $g^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$ to $\gamma_0 \cdot z$ for a suitable n so that $\mathrm{Re}\left(g^n\gamma_0 \cdot z\right) \in \left[-\frac{1}{2}, \frac{1}{2}\right)$. Denote $\gamma_1 = g^n\gamma_0$.

Claim. $|\gamma_1 \cdot z| \geqslant 1$, i.e. $\gamma_1 \cdot z \in E$.

Proof of Claim. If not, say $|\gamma_1 \cdot z| < 1$, then

$$\mathrm{Im}\left(h\gamma_1 \cdot z\right) > \mathrm{Im}\left(\gamma_1 \cdot z\right) = \mathrm{Im}\left(\gamma_0 \cdot z\right).$$

This contradicts to our choice of γ_0. □

Now we prove that the orbit of any $z \in \mathbb{H}^2$ under the action of $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ in E is a single point. WLOG, we suppose $z \in E$. Suppose $\gamma \cdot z \in E$. We wish to prove that $\gamma \cdot z = z$. WLOG, we assume $\mathrm{Im}(\gamma \cdot z) \geqslant \mathrm{Im}(z)$. If not, we just consider γ^{-1} maps $\gamma \cdot z \in E$ to $z \in E$ and get $\gamma \cdot z = \gamma^{-1}(\gamma \cdot z) = z$. Denote $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. By formular (16), our assumption tells us that $|cz + d| \leqslant 1$. Because $z \in E$, this forces $c = 0, \pm 1$.

- If $c = 0$, then $d = \pm 1$. Hence we have $\gamma = \pm\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = g^{\pm b}$ for some integer b. $\gamma \cdot z \in E$ shows that $b = 0$ and then $\gamma = \mathrm{Id}$.

- If $c = 1$, then $|z + d| \leqslant 1$.
  - If $d \neq 0$, then z is the third root of unity $w = \mathrm{e}^{\frac{2\pi}{3}\mathrm{i}}$ and $d = 1$. Suppose $\gamma = \begin{pmatrix} a & a-1 \\ 1 & 1 \end{pmatrix}$. $\gamma \cdot z \in E$ gives us two equations $|\gamma \cdot z| \geqslant 1$ and $-\frac{1}{2} \leqslant \mathrm{Re}(\gamma \cdot z) < \frac{1}{2}$. After direct calculations, we have
  $$\begin{cases} 0 \leqslant a < 1, \\ a(a-1) \geqslant 0. \end{cases}$$
  We must have $a = 0$. Now we have $\gamma \cdot z = z$.
  - If $d = 0$, we have $\gamma = \begin{pmatrix} a & -1 \\ 1 & 0 \end{pmatrix}$ and $|z| \leqslant 1$. Because $z \in E$, $|z| = 1$ and $z = \mathrm{e}^{\theta \mathrm{i}}$ for some $\theta \in \left[\frac{\pi}{2}, \frac{2\pi}{3}\right]$. $|\gamma \cdot z| \geqslant 1$ and $-\frac{1}{2} \leqslant \mathrm{Re}(\gamma \cdot z) < \frac{1}{2}$ give us
  $$\begin{cases} -\frac{1}{2} \leqslant a - \cos\theta < \frac{1}{2}, \\ a(a - 2\cos\theta) \geqslant 0. \end{cases}$$
  One solution is $a = 0$. Now $\gamma \cdot z = -\frac{1}{z} \notin E$ except for $z = \mathrm{i}$. We get a contradiction. If $z = \mathrm{i}$, $\gamma \cdot z = \mathrm{i} = z$. Another solution is $a = -1, \theta = \frac{2\pi}{3}$. Now $\gamma \cdot z = -\frac{1}{2} + \frac{\sqrt{3}}{2}\mathrm{i} = z$.
- If $c = -1$, then $|-z + d| \leqslant 1$.
  - If $d \neq 0$, then z is the third root of unity $w = \mathrm{e}^{\frac{2\pi}{3}\mathrm{i}}$ and $d = -1$. Suppose $\gamma = \begin{pmatrix} a & a+1 \\ -1 & -1 \end{pmatrix}$. $\gamma \cdot z \in E$ gives us two equations $|\gamma \cdot z| \geqslant 1$ and $-\frac{1}{2} \leqslant \mathrm{Re}(\gamma \cdot z) < \frac{1}{2}$. After direct calculations, we have
  $$\begin{cases} -1 < a \leqslant 0, \\ a(a+1) \geqslant 0. \end{cases}$$
  We must have $a = 0$. Now we have $\gamma \cdot z = z$.
  - If $d = 0$, we have $\gamma = \begin{pmatrix} a & 1 \\ -1 & 0 \end{pmatrix}$ and $|-z| \leqslant 1$. Because $z \in E$, $|z| = 1$ and $z = \mathrm{e}^{\theta \mathrm{i}}$ for some $\theta \in \left[\frac{\pi}{2}, \frac{2\pi}{3}\right]$. $|\gamma \cdot z| \geqslant 1$ and $-\frac{1}{2} \leqslant \mathrm{Re}(\gamma \cdot z) < \frac{1}{2}$ give us
  $$\begin{cases} -\frac{1}{2} \leqslant -a - \cos\theta < \frac{1}{2}, \\ a(a + 2\cos\theta) \geqslant 0. \end{cases}$$
  One solution is $a = 0$. Now $\gamma \cdot z = -\frac{1}{z} \notin E$ except for $z = \mathrm{i}$. We get a contradiction. If $z = \mathrm{i}$, $\gamma \cdot z = \mathrm{i} = z$. Another solution is $a = 1, \theta = \frac{2\pi}{3}$. Now $\gamma \cdot z = -\frac{1}{2} + \frac{\sqrt{3}}{2}\mathrm{i} = z$.

Above all, $\sharp\left(\mathrm{PSL}_2(\mathbb{Z}) \cdot z \cap E\right) = 1$ for any $z \in \mathbb{H}^2$. Hence E is a fundamental domain of $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$.

□

Definition 6.5. *The quotient in Theorem 6.4 is called the modular surface.*

Proposition 6.6. *The hyperbolic area A defined by the volume form $\mathrm{d}A = \frac{\mathrm{d}x\mathrm{d}y}{y^2}$ is preserved by Möbius transformations.*

Proof. Let $f \in C_c(\mathbb{H}^2)$ and $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{R})$. We calculate
$$\int_{\mathbb{H}^2} f \circ g(z)\,\mathrm{d}A = \int_{\mathbb{H}^2} f \circ g(z)\frac{1}{y^2}\,\mathrm{d}x\mathrm{d}y = \int_{\mathbb{H}^2} f \circ g(z)\frac{1}{(\mathrm{Im}(z))^2}\,\mathrm{d}x\mathrm{d}y.$$
By equation (16),
$$\left(\mathrm{Im}(gz)\right)^2 = \frac{(\mathrm{Im}z)^2}{(cz+d)^4}.$$
Hence we have
$$\int_{\mathbb{H}^2} f \circ g\,\mathrm{d}A = \int_{\mathbb{H}^2} f \circ g(z)\frac{1}{(\mathrm{Im}(z))^2}\,\mathrm{d}x\mathrm{d}y = \int_{\mathbb{H}^2} f \circ g(z)\frac{1}{(\mathrm{Im}(gz))^2}\frac{1}{|cz+d|^4}\,\mathrm{d}x\mathrm{d}y.$$

Note that $\frac{1}{|cz+d|^4}$ is the Jacobian of $z' = gz$: Let $z' = x' + \mathrm{i}y' = gz$. Then direct calculations show that

$$f\left(z'\right)\,\mathrm{d}x'\mathrm{d}y' = f \circ g(z) \frac{1}{|cz+d|^4}\,\mathrm{d}x\mathrm{d}y.$$

These calculations can be done by doing $\mathbb{R}$-derivatives for x, y in $\mathrm{Mat}_{2\times 2}(\mathbb{R})$ and taking determinant. Then we have

$$\int_{\mathbb{H}^2} f \circ g\,\mathrm{d}A = \int_{\mathbb{H}^2} f \circ g(z) \frac{1}{(\mathrm{Im}(gz))^2} \frac{1}{|cz+d|^4}\,\mathrm{d}x\mathrm{d}y = \int_{\mathbb{H}^2} f\left(z'\right) \frac{1}{\left(\mathrm{Im}\left(z'\right)\right)^2}\,\mathrm{d}x'\mathrm{d}y' = \int_{\mathbb{H}^2} f\,\mathrm{d}A.$$

Hence for any $f \in C_c(\mathbb{H}^2)$,

$$\int_{\mathbb{H}^2} f\,\mathrm{d}\left(g^* A\right) = \int_{\mathbb{H}^2} f \circ g\,\mathrm{d}A = \int_{\mathbb{H}^2} f\,\mathrm{d}A.$$

Therefore $\mathrm{d}A$ is preserved by any Möbius transformation $g \in \mathrm{PSL}_2(\mathbb{R})$. □

Now we generalize this proposition to the unit tangent bundle of $\mathbb{H}^2$.

Definition 6.7. *Recall that $T^1\mathbb{H}^2 \cong \mathbb{H}^2 \times \mathbb{S}^1$. We define the volume m on $T^1\mathbb{H}^2$ by*

$$\mathrm{d}m = \mathrm{d}A\mathrm{d}\theta$$

where θ is the coordinate in $\mathbb{S}^1$.

Proposition 6.8. *m is preserved by the induced action D_g of any $g \in \mathrm{PSL}_2(\mathbb{R})$.*

Proof. The proof is similar to Proposition 6.6.

Let $f \in C_c(T^1\mathbb{H}^2)$ and $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{R})$. We calculate

$$\begin{aligned}
\int_{T^1\mathbb{H}^2} f \circ D_g(z,v)\,\mathrm{d}m &= \int_{T^1\mathbb{H}^2} f \circ D_g\left(z, \mathrm{e}^{\mathrm{i}\theta}\right)\,\mathrm{d}A\mathrm{d}\theta \\
&= \int_{T^1\mathbb{H}^2} f\left(gz, g'(z)\mathrm{e}^{\mathrm{i}\theta}\right)\,\mathrm{d}A\mathrm{d}\theta = \int_{T^1\mathbb{H}^2} f\left(gz, \mathrm{e}^{\mathrm{i}\left(\theta + \mathrm{Angle}\left(g'(z)\right)\right)}\right) \frac{1}{(\mathrm{Im}(z))^2}\,\mathrm{d}x\mathrm{d}y\mathrm{d}\theta \\
&= \int_{T^1\mathbb{H}^2} f\left(gz, \mathrm{e}^{\mathrm{i}\left(\theta + \mathrm{Angle}\left(g'(z)\right)\right)}\right) \frac{1}{(\mathrm{Im}(gz))^2} \frac{1}{|cz+d|^4}\,\mathrm{d}x\mathrm{d}y\mathrm{d}\theta.
\end{aligned}$$

Now we calculate the Jacobian of $(z', v') = (gz, D_g v)$. The coordinates representations are

$$\left(x', y', \theta'\right) = \left(x'(x,y), y'(x,y), \theta + \mathrm{Angle}\left(g'(z)\right)\right).$$

Hence the marix is

$$\begin{pmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} & 0 \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} & 0 \\ * & * & 1 \end{pmatrix}.$$

Then the Jacobian is $J_g \cdot 1 = \frac{1}{|cz+d|^4}$.

Therefore now

$$\begin{aligned}
\int_{T^1\mathbb{H}^2} f \circ D_g(z,v)\,\mathrm{d}m &= \int_{T^1\mathbb{H}^2} f\left(gz, \mathrm{e}^{\mathrm{i}\left(\theta + \mathrm{Angle}\left(g'(z)\right)\right)}\right) \frac{1}{(\mathrm{Im}(gz))^2} \frac{1}{|cz+d|^4}\,\mathrm{d}x\mathrm{d}y\mathrm{d}\theta \\
&= \int_{T^1\mathbb{H}^2} f\left(z', v'\right) \frac{1}{\left(\mathrm{Im}\left(z'\right)\right)^2}\,\mathrm{d}x'\mathrm{d}y'\mathrm{d}\theta' \\
&= \int_{T^1\mathbb{H}^2} f\left(z', v'\right)\,\mathrm{d}A'\mathrm{d}\theta' = \int_{T^1\mathbb{H}^2} f\left(z', v'\right)\,\mathrm{d}m' = \int_{T^1\mathbb{H}^2} f\left(z, v\right)\,\mathrm{d}m.
\end{aligned}$$

Hence for any $f \in C_c(T^1\mathbb{H}^2)$,

$$\int_{T^1\mathbb{H}^2} f\,\mathrm{d}\left(D_g^* m\right) = \int_{T^1\mathbb{H}^2} f \circ D_g\,\mathrm{d}m = \int_{T^1\mathbb{H}^2} f\,\mathrm{d}m.$$

Therefore $\mathrm{d}m$ is preserved by any induced Möbius transformation D_g for any $g \in \mathrm{PSL}_2(\mathbb{R})$.

□

Definition 6.9. $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ *is called a uniform lattice if* $\Gamma \backslash \mathrm{PSL}_2(\mathbb{R})$ *is compact (in the quotient topology). We also say* Γ *is co-compact.*

Definition 6.10. $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ *is a lattice if*

- *there exists a fundamental domain in* $\mathbb{H}^2$ *with finite volume,*
- *or there exists a fundamental domain in* $\mathrm{PSL}_2(\mathbb{R})$ *with finite volume.*

Proposition 6.11. $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ *is a (non-uniform) lattice in* $\mathrm{PSL}_2(\mathbb{R})$*.*

Proof. By Theorem 6.4, E is a fundamental domain of Γ. Then we calculate directly

$$\mathrm{Area}(E) = \int_E \mathrm{d}A = \int_E \frac{1}{y^2}\,\mathrm{d}x\mathrm{d}y \leqslant \int_{\frac{\sqrt{3}}{2}}^{\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{y^2}\,\mathrm{d}x\mathrm{d}y < \infty.$$

□

We conclude the above by the following remarks:

Remark 6.12. Recall that $\mathrm{PSL}_2(\mathbb{R}) \cong T^1\mathbb{H}^2$ by $g \mapsto D_g(\mathrm{i},\mathrm{i})$.

- When we take Möbius transformation D_g on $T^1\mathbb{H}^2$, the corresponding operation on $\mathrm{PSL}_2(\mathbb{R})$ is left-multiplication by g.
- When we go along a geodesic flow on $T^1\mathbb{H}^2$, the corresponding operation on $\mathrm{PSL}_2(\mathbb{R})$ is right-multiplication by a diagonal matrix.
- When we rotate tangent vector without moving base point on $T^1\mathbb{H}^2$, the corresponding operation on $\mathrm{PSL}_2(\mathbb{R})$ is right-multiplication by a matrix in $\mathrm{PSO}_2(\mathbb{R})$.
- When go along a horocycle flow on $T^1\mathbb{H}^2$, the corresponding operation on $\mathrm{PSL}_2(\mathbb{R})$ is right-multiplication by a matrix in $\mathrm{U}^- = \left\{\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} : s \in \mathbb{R}\right\}$.

Back to our goal, we have

Remark 6.13. If $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ is a lattice, then there is a probability measure m_X on $X := \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ that is invariant under the action of $\mathrm{PSL}_2(\mathbb{R})$.

Proposition 6.14. *Let* G *be a "nice" topological group and* $\Gamma < G$ *be a lattice. Then the left Haar measure is also right invariant (If so,* G *is called unimodular). Moreover, there exists a finite* G*-invariant measure on* $X = \Gamma\backslash G$ *which is also* G*-ergodic.*

Before we prove this proposition, we talk about something about Haar measure.
We assume G is a σ-compact locally compact metrizable topological group.

Definition 6.15. *A measure* $m = m_G$ *is a left Haar measure on* G *if*

- $m(gB) = m(B)$ *for any* $g \in G$ *and* $B \subset G$*.*
- $m(K) < \infty$ *for compact* $K \subset G$*.*
- $m(O) > 0$ *for a nonempty open* $O \subset G$*.*

Theorem 6.16. *The left Haar measure always exists.*

Lemma 6.17 (Positive Overlap)**.** *Let* $B_1, B_2 \subset G$ *be measurable with* $m(B_1) > 0$ *and* $m(B_2) > 0$*. Then there exist* $g \in G$ *with* $m(gB_1 \cap B_2) > 0$*.*

Proof. Note that

$$m\left(gB_1 \cap B_2\right) = \int_G \chi_{gB_1}(h)\chi_{B_2}(h)\,\mathrm{d}m(h) = \int_G \chi_{hB_1^{-1}}(g)\chi_{B_2}(h)\,\mathrm{d}m(h).$$

By Fubini theorem, we have

$$\begin{aligned}\int_G m\left(gB_1 \cap B_2\right)\,\mathrm{d}m(g) &= \int_G \chi_{B_2(h)} \int_G \chi_{hB_1^{-1}}(g)\,\mathrm{d}m(g)\,\mathrm{d}m(h) \\ &= m\left(hB_1^{-1}\right)\int_G \chi_{B_2}(h)\,\mathrm{d}m(h) = m\left(B_1^{-1}\right)m\left(B_2\right).\end{aligned} \tag{17}$$

Using this formula, we claim

Claim. If $m(B_1) > 0$, then $m(B_1^{-1}) > 0$.

Proof of Claim. Suppose $m\left(B_1^{-1}\right) = 0$. Let $B_2 = G$. Equation (17) becomes

$$\int_G (gB_1)\ \mathrm{d}m(g) = m\left(B_1^{-1}\right) m\left(B_2\right) = 0.$$

But for $g = \mathrm{Id}$, $m(gB_1) = m(B_1) > 0$. Hence we have $\int_G m(gB_1)\,\mathrm{d}m(g) > 0$. This is a contradiction. □

Now if for any $g \in G$, $m(gB_1 \cap B_2) = 0$, equation (17) shows that

$$0 = \int_G m\left(gB_1 \cap B_2\right)\ \mathrm{d}m(g) = m\left(B_1^{-1}\right) m\left(B_2\right) > 0.$$

We get a contradiction. □

Lemma 6.18 ("Ergodicity")**.** *The left action of G on G is ergodic in the sense that*

- *If $B \subset G$ satisfies $m(gB\Delta B) = 0$ for any $g \in G$, then $m(B) = 0$ or $m(G - B) = 0$.*
- *If $f : G \to \mathbb{R}$ is meaasurable and for any $g \in G$, f and $[h \in G \mapsto f(gh)]$ agree m-a.e., then f is a constant m-a.e..*

Proof. For the first version, we set $B_1 = B$ and $B_2 = G - B$. If $m(B_1) > 0$ and $m(B_2) > 0$, by Lemma 6.17, there exists $g \in G$ so that $m(gB_1 \cap B_2) > 0$, i.e. $m(gB \cap (G-B)) > 0$. Then $m(gB\Delta B) = m(gB \cap (G-B)) > 0$. We get a contradiction.

For the second version, let $I_1, I_2 \subset \mathbb{R}$ be two disjoint intervals and define $B_j = f^{-1}(I_j)$ for $j = 1, 2$. We have $m(B_1 \cap B_2) = m(\varnothing) = 0$. If $m(B_1) > 0$ and $m(B_2) > 0$, by Lemma 6.17, there exists $g \in G$ so that $m(gB_1 \cap B_2) > 0$. Note that f and $f(g^{-1}\cdot)$ agree m-a.e., we have $gB_1 = (f(g^{-1}\cdot))^{-1}(I_1) \overset{a.s.}{=} f^{-1}(I_1) = B_1$. Hence

$$0 = m\left(B_1 \cap B_2\right) = m\left(gB_1 \cap B_2\right) > 0.$$

This is a contradiction. Therefore $m(B_1) = 0$ or $m(B_2) = 0$, i.e. the value domain of f doesn't contain I_1 or I_2 m-a.s.. By covering $\mathbb{R}$ with countable many intervals and subdividing the intervals, the value domain of f is contained in an arbitrarily small interval, i.e. f is a constant m-a.s.. □

Lemma 6.19. *The left Haar measure is unique up to scalars.*

Proof. Suppose m_1, m_2 are two left Haar measure. Then $m_1 + m_2$ is another Haar measure so that m_1, m_2 are absolutely continuous with respect to m. Let $f_1 = \frac{\mathrm{d}m_1}{\mathrm{d}m}$ be the Radon-Nikodym derivative.

Claim. f_1 is G-invariant in the sense that $[h \mapsto f_1(h)]$ and $[h \mapsto f_1(gh)]$ agree m-a.s. for any $g \in G$.

Proof of Claim. Let $B \subset G$ measurable arbitrarily. Then

$$\int_B f_1(gh)\,\mathrm{d}m(h) = \int_G \chi_B(h) f_1(gh)\,\mathrm{d}m(h).$$

Take variable alternation $\widetilde{h} := gh$. Because m is a left Haar measure, $\mathrm{d}m(\widetilde{h}) = \mathrm{d}m(h)$. Hence we have

$$\begin{aligned}\int_B f_1(gh)\,\mathrm{d}m(h) &= \int_G \chi_B(h) f_1(gh)\,\mathrm{d}m(h) = \int_G \chi_B\left(g^{-1}\widetilde{h}\right) f_1\left(\widetilde{h}\right)\,\mathrm{d}m\left(\widetilde{h}\right) \\ &= \int_G \chi_{g^{-1}B}\left(\widetilde{h}\right) f_1\left(\widetilde{h}\right)\,\mathrm{d}m\left(\widetilde{h}\right) = \int_{g^{-1}B} \left(\widetilde{h}\right) f_1\left(\widetilde{h}\right)\,\mathrm{d}m\left(\widetilde{h}\right) \\ &= m_1\left(g^{-1}B\right) = m_1(B) = \int_B f_1(h)\,\mathrm{d}m(h).\end{aligned}$$

Hence f_1 and $f_1(g\cdot)$ agree m-a.s.. □

By the second version of Lemma 6.18, f_1 is a constant m-a.s.. Therefore m is a scalar of m_1 and then m_1, m_2, m are all multiples of each other. □

Corollary 6.20. *Given $g \in G$, we define*

$$m_g(B) := m(Bg)$$

for any measurable $B \subset G$. m_g is also a left Haar measure. Hence there exists $\Delta(g) > 0$ so that

$$m(Bg) = m_g(B) = \Delta(g)m(B)$$

for all measurable $B \subset G$. Then G is unimodular, i.e. m is also a right Haar measure, if and only if $\Delta(g) = 1$ for any $g \in G$.

Now we turn to prove Proposition 6.14.

Proof of Proposition 6.14. We begin with a claim.

Claim. Let $B_1, B_2 \subset G$ with $\bigsqcup_{\gamma\in\Gamma} \gamma B_1 = \bigsqcup_{\gamma\in\Gamma} \gamma B_2$. Then $m(B_1) = m(B_2)$.

Proof of Claim. Note that

$$B_1 = B_1 \cap \bigsqcup_{\gamma\in\Gamma} \gamma B_2 = \bigsqcup_{\gamma\in\Gamma} B_1 \cap \gamma B_2.$$

Hence we have

$$m(B_1) = \sum_{\gamma\in\Gamma} m(B_1 \cap \gamma B_2) = \sum_{\gamma\in\Gamma} m\left(\gamma^{-1}B_1 \cap B_2\right).$$

Note that the sum is over all $\gamma \in \Gamma$,

$$m(B_1) = \sum_{\gamma\in\Gamma} m\left(\gamma^{-1}B_1 \cap B_2\right) = \sum_{\gamma\in\Gamma} m(\gamma B_1 \cap B_2) = m\left(\bigsqcup_{\gamma\in\Gamma} (\gamma B_1 \cap B_2)\right) = m(B_2).$$

□

This claim shows that for two fundamental domain F_1, F_2, $m(F_1) = m(F_2)$, because by definition, $\bigsqcup_{\gamma\in\Gamma} \gamma F_1 = G = \bigsqcup_{\gamma\in\Gamma} \gamma F_2$.

Let $F \subset G$ be a fundamental domain with $m(F) < \infty$. For any $g \in G$, Fg is another fundamental domain. Hence we have

$$m(F) = m(Fg) = \Delta(g)m(F).$$

Then $\Delta(g) = 1$ for any $g \in G$. By Corollary 6.20, G is unimodular.

We define a measure m_X on $X = \Gamma\backslash G$ by

$$m_X(B) = m(\{g \in F : \Gamma g \in B\}) \leqslant m(F)$$

for any measurable $B \subset X$. For any $g_0 \in G$,

$$m_X(Bg_0) = m(\{g \in F : \Gamma g \in Bg_0\}) = m\left(\{g \in F : \Gamma g g_0^{-1} \in B\}\right).$$

By substitution $hg_0 = g$, we have

$$m_X(Bg_0) = m\left(\{g \in F : \Gamma g g_0^{-1} \in B\}\right) = m\left(\{h \in Fg_0^{-1} : \Gamma h \in B\}\, g_0^{-1}\right) = m\left(\{h \in Fg_0^{-1} : \Gamma h \in B\}\right).$$

Note again that Fg_0^{-1} is also a fundamental domain, $m(Fg_0^{-1}) = m(F)$. Then

$$m_X(Bg_0) = m\left(\{h \in Fg_0^{-1} : \Gamma h \in B\}\right) = m(\{h \in F : \Gamma h \in B\}) = m_X(B).$$

Therefore m_X is a finite G-invariant measure on X.

Finally, we prove that m_X is G-ergodic. Let $B \subset X$ be such that $m_X(B\Delta Bg_0) = 0$ for all $g_0 \in G$. We define

$$B_G = \{g \in G : \Gamma g \in B\}.$$

Note that $B_G = \gamma B_G$ for any $\gamma \in \Gamma$. Let $g_0 \in G$ arbitrarily. For any $\gamma \in \Gamma$, by the definition of m_X, because γF is also a fundamental domain,

$$m((B_G \Delta B_G g_0) \cap \gamma F) = m_X(B_G \Delta B_G g_0) = 0.$$

Because F is a fundamental domain, we must have $m(B_G \Delta B_G g_0) = 0$ for any $g_0 \in G$. By Lemma 6.18, $m(B_G) = 0$ or $m(G - B_g) = 0$. Then $m_X(B) \leqslant m(B_G) = 0$ or $m_X(X - B) \leqslant m(G - B_G) = 0$. Therefore m_X is also G-ergodic. □

Proposition 6.21 (Unitary Representation). *Let $\Gamma < G$ be a lattice and $X := \Gamma\backslash G$. For any $g \in G$, we define $\pi_g \colon L^2(X) \to L^2(X)$ by*

$$(\pi_g f)(x) = f(xg).$$

Then

$$\begin{aligned} \pi \colon G &\to \mathrm{GL}\left(L^2(X)\right) \\ g &\mapsto \pi_g \end{aligned}$$

is a unitary representation, i.e.

$$\pi_e = \mathrm{Id}, \quad \pi_{g_1}\pi_{g_2} = \pi_{g_1 g_2}, \quad \pi_g \text{ is unitary}$$

for any $g_1, g_2, g \in G$. Moreover, for any $f \in L^2(X)$,

$$g \in G \mapsto \pi_g f \in L^2(X)$$

is continuous.

Proof. Unitarity follows from the discussion of Koopman operator (Proposition 3.12) where T is the right multiplication by g.

Let $g_1, g_2 \in G$, $x \in X$ and $f \in L^2(X)$. We have

$$\left(\pi_{g_1}\left(\pi_{g_2} f\right)\right)(x) = \left(\pi_{g_2} f\right)(x g_1) = f\left(x g_1 g_2\right) = \left(\pi_{g_1 g_2} f\right)(x).$$

Then we prove continuity. For $f \in C_c(X)$ and $g_n \to g$ in G, the convergence of $\pi_{g_n} f \to \pi_g f$ holds pointwisely by continuity of f. Because f has compact support, $|f|$ can achieve the maximum, say $|f(x_0)|$. Using dominated convergence theorem, because $|\pi_{g_n} f - \pi_g f| \leqslant 2\,|f(x_0)|$ for any $x \in X$ and $n \geqslant 0$, we have

$$\|\pi_{g_n} f - \pi_g f\|_{L^2(X)}^2 = \int_X |\pi_{g_n} f - \pi_g f| \, \mathrm{d}m_X \to \int_X 0 \, \mathrm{d}m_X = 0.$$

For any general $f \in L^2(X)$, for any $\varepsilon > 0$, by density of $C_c(X)$ in $L^2(X)$, there exists $f_0 \in C_c(X)$ with $\|f - f_0\|_{L^2(X)} < \varepsilon$. Suppose $g_n \to g$. Then because unitary operators preserve norms, as $n \to \infty$,

$$\begin{aligned} \|\pi_{g_n} f - \pi_g f\|_{L^2(X)} &\leqslant \|\pi_{g_n} f - \pi_{g_n} f_0\|_{L^2(X)} + \|\pi_{g_n} f_0 - \pi_g f_0\|_{L^2(X)} + \|\pi_g f - \pi_g f_0\|_{L^2(X)} \\ &= \|f - f_0\|_{L^2(X)} + \|\pi_{g_n} f_0 - \pi_g f_0\|_{L^2(X)} + \|f_0 - f\|_{L^2(X)} < 3\varepsilon. \end{aligned}$$

Hence $\pi_{g_n} f \to \pi_g f$ in $L^2(X)$. ☐

This Proposition can be used to prove ergodicity or mixing property of many system as follows. We remark that although the following theorems are proved for $\mathrm{PSL}_2(\mathbb{R})$ for simplicity, they also hold for any "nice" group G.

Theorem 6.22 (Ergodicity of geodesic flow). *Let $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ be a lattice, Then the geodesic flow, or even the time-one-map of the geodesic flow, acts ergodically on $X := \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ with respect to m_X.*

Proof. Recall that the time-one-map is given by $T \colon x \in X \mapsto xa^{-1}$ where $a = \begin{pmatrix} \mathrm{e}^{-\frac{1}{2}} & 0 \\ 0 & \mathrm{e}^{\frac{1}{2}} \end{pmatrix}$. (See Theorem 6.1.) We use Lemma 3.26. Suppose $f \in L^2(X)$ with $U_T f = f$, i.e. $\pi_{a^{-1}} f = f$. We will show that f is a constant a.s.. Note that we also have $\pi_a f = \pi_a \pi_{a^{-1}} f = f$.

Denote $u = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ for any $s \in \mathbb{R}$. We have

$$a^n u a^{-n} = \begin{pmatrix} \mathrm{e}^{-\frac{n}{2}} & 0 \\ 0 & \mathrm{e}^{\frac{n}{2}} \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathrm{e}^{\frac{n}{2}} & 0 \\ 0 & \mathrm{e}^{-\frac{n}{2}} \end{pmatrix} = \begin{pmatrix} 1 & \mathrm{e}^{-n} s \\ 0 & 1 \end{pmatrix} \to \mathrm{Id}$$

as $n \to \infty$. Then by Proposition 6.21, π_{a^n} is unitary, it preserves norms. We have

$$\begin{aligned} \|\pi_u f - f\| &= \|\pi_{a^n}\left(\pi_u f - f\right)\| = \|\pi_{a^n}\pi_u f - \pi_{a^n} f\| \\ &\overset{\pi_{a^{-n}} f = f = \pi_{a^n} f}{=} \|\pi_{a^n}\pi_u\pi_{a^{-n}} f - f\| = \|\pi_{a^n u a^{-n}} f - f\| \to \|\mathrm{Id} f - f\| = 0. \end{aligned}$$

Therefore $\pi_u f = f$ a.s., i.e. f is invariant under $\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ for any $s \in \mathbb{R}$. By similar calculations, f is also invariant under $\begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}$ for any $s \in \mathbb{R}$. Note that $\left\{ \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} : s \in \mathbb{R} \right\}$ generates $\mathrm{PSL}_2(\mathbb{R})$. It follows that $f \in L^2(X)$ is invariant under all of $\mathrm{PSL}_2(\mathbb{R})$. That is to say, for any $g \in \mathrm{PSL}_2(\mathbb{R})$, f and $h \mapsto f(hg)$ agree a.s.. Then by Lemma 6.18, f is a constant a.s.. (Lemma 6.18 is for left multiplications. But note that the left Haar measure is also right Haar measure, this lemma also holds for right multiplications.) We obtain that the time-one-map is ergodic. The proof for ergodicity of geodesic flow is similar, by substituting 1 by all t. ☐

Theorem 6.23 (Ergodicity of horocycle flow)**.** *Let $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ be a lattice and $X := \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$. Then the horocycle flow acts ergodically with respect to m_X.*

Proof. Recall that the horocycle flow is defined by right multiplication by upper unipotent matrices

$$U^- = \left\{ \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} : s \in \mathbb{R} \right\}.$$

See Definition 6.2. We also use Lemma 3.26. Suppose $f \in L^2(X)$ with $\pi_{\left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)} f = f$ for all $s \in \mathbb{R}$. We will show that f is a constant a.s..

Let $\varepsilon > 0$. Denote $s = \frac{1}{\varepsilon}$ and $s' = -\frac{s}{2} = -\frac{1}{2\varepsilon}$ for simplicity. We have

$$\begin{aligned} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \varepsilon & 1 \end{pmatrix} \begin{pmatrix} 1 & s' \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} 1 + s\varepsilon & s \\ \varepsilon & 1 \end{pmatrix} \begin{pmatrix} 1 & s' \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 + s\varepsilon & s'(1 + s\varepsilon) + s \\ \varepsilon & 1 + s'\varepsilon \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ \varepsilon & \frac{1}{2} \end{pmatrix} \to \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} =: a \end{aligned}$$

as $\varepsilon \to \infty$. Because the unitary representation is continuous (Proposition 6.21), we calculate

$$\|\pi_a f - f\| = \lim_{\varepsilon \to 0} \left\| \pi_{\left(\begin{smallmatrix} 2 & 0 \\ \varepsilon & \frac{1}{2} \end{smallmatrix}\right)} f - f \right\| = \lim_{\varepsilon \to 0} \left\| \pi_{\left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)} \pi_{\left(\begin{smallmatrix} 1 & 0 \\ \varepsilon & 1 \end{smallmatrix}\right)} \pi_{\left(\begin{smallmatrix} 1 & s' \\ 0 & 1 \end{smallmatrix}\right)} f - f \right\|.$$

By our assumption, $\pi_{\left(\begin{smallmatrix} 1 & s' \\ 0 & 1 \end{smallmatrix}\right)} f = f = \pi_{\left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)} f$. By Proposition 6.21, $\pi_{\left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)}$ is unitary, it preserves norm. Proposition 6.21 also gives us continuity. Hence we have

$$\begin{aligned} \|\pi_a f - f\| &= \lim_{\varepsilon \to 0} \left\| \pi_{\left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)} \pi_{\left(\begin{smallmatrix} 1 & 0 \\ \varepsilon & 1 \end{smallmatrix}\right)} \pi_{\left(\begin{smallmatrix} 1 & s' \\ 0 & 1 \end{smallmatrix}\right)} f - f \right\| = \lim_{\varepsilon \to 0} \left\| \pi_{\left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)} \pi_{\left(\begin{smallmatrix} 1 & 0 \\ \varepsilon & 1 \end{smallmatrix}\right)} f - \pi_{\left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)} f \right\| \\ &= \lim_{\varepsilon \to 0} \left\| \pi_{\left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)} \left(\pi_{\left(\begin{smallmatrix} 1 & 0 \\ \varepsilon & 1 \end{smallmatrix}\right)} f - f \right) \right\| = \lim_{\varepsilon \to 0} \left\| \pi_{\left(\begin{smallmatrix} 1 & 0 \\ \varepsilon & 1 \end{smallmatrix}\right)} f - f \right\| = \|\mathrm{Id} f - f\| = 0. \end{aligned}$$

This means $\pi_a f = f$ a.s., i.e. f is invariant under a. Note that a is an element of the geodesic flow. By Theorem 6.22 and Lemma 3.26, f is a constant a.s.. We obtain that the horocycle flow is ergodic. ☐

Theorem 6.24 (Mixing, Howe-Moore)**.** *Let $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ be a lattice and $X := \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$. Then the geodesic flow is mixing with respect to m_X.*

Proof. By Lemma 3.33, we wish to show that for any $f_1, f_2 \in L^2(X)$,

$$\langle U_T^n f_1, f_2 \rangle = \langle \pi_{a^n} f_1, f_2 \rangle = \langle \pi_{a_n} f_1, f_2 \rangle \to \int_X f_1 \,\mathrm{d}m_X \int_X \overline{f_2} \,\mathrm{d}m_X$$

as $n \to \infty$, where $a_n = \begin{pmatrix} \mathrm{e}^{-\frac{n}{2}} & 0 \\ 0 & \mathrm{e}^{\frac{n}{2}} \end{pmatrix}$ are elements of geodesic flow. We denote $a_n = \begin{pmatrix} \lambda_n & 0 \\ 0 & \lambda_n^{-1} \end{pmatrix}$ with $\lambda_n \to \infty$ for simplicity.

By Banach-Alaoglu theorem, the sequence $\pi_{a_n} f_1$ has a weak* converging subsequence $\pi_{a_{n_k}} f_1 \rightharpoonup \widetilde{f} \in L^2(X)$.

Claim. $\widetilde{f}$ is invariant under $u = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ for any $s \in \mathbb{R}$.

Proof of Claim. For any $f_2 \in L^2(X)$, we calculate

$$\left\langle \pi_u \widetilde{f}, f_2 \right\rangle = \left\langle \widetilde{f}, \pi_u^* f_2 \right\rangle = \lim_{k\to\infty} \left\langle \pi_{a_{n_k}} f_1, \pi_u^* f_2 \right\rangle = \lim_{k\to\infty} \left\langle \pi_{u a_{n_k}} f_1, f_2 \right\rangle = \lim_{k\to\infty} \left\langle \pi_{a_{n_k}} \pi_{a_{n_k}^{-1} u a_{n_k}} f_1, f_2 \right\rangle.$$

Note that

$$a_n^{-1} u a_n = \begin{pmatrix} \lambda_n^{-1} & 0 \\ 0 & \lambda_n \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_n & 0 \\ 0 & \lambda_n^{-1} \end{pmatrix} = \begin{pmatrix} 1 & \lambda_n^{-2} s \\ 0 & 1 \end{pmatrix} \to \mathrm{Id}$$

as $n \to \infty$. Then by continuity, because π is unitary by Proposition 6.21, we have

$$\begin{aligned} \lim_{k\to\infty} \left| \left\langle \pi_{a_{n_k}} \left(\pi_{a_{n_k}^{-1} u a_{n_k}} f_1 - f_1 \right), f_2 \right\rangle \right| &\leqslant \lim_{k\to\infty} \left\| \pi_{a_{n_k}} \left(\pi_{a_{n_k}^{-1} u a_{n_k}} f_1 - f_1 \right) \right\| \cdot \|f_2\| \\ &= \lim_{k\to\infty} \left\| \pi_{a_{n_k}^{-1} u a_{n_k}} f_1 - f_1 \right\| \cdot \|f_2\| = 0. \end{aligned}$$

Hence we have

$$\left\langle \pi_u \widetilde{f}, f_2 \right\rangle = \lim_{k\to\infty} \left\langle \pi_{a_{n_k}} \pi_{a_{n_k}^{-1} u a_{n_k}} f_1, f_2 \right\rangle = \lim_{k\to\infty} \left\langle \pi_{a_{n_k}} f_1, f_2 \right\rangle = \left\langle \widetilde{f}, f_2 \right\rangle.$$

We must have $\pi_u \widetilde{f} = \widetilde{f}$ a.s.. □

Then by Theorem 6.23 and Lemma 3.26, $\widetilde{f}$ = constant a.s.. Moreover, WLOG, assuming $m_X(X) = 1$, we have

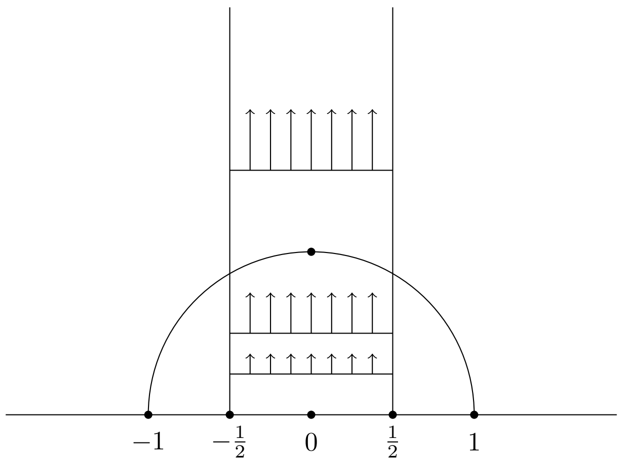
$$\text{constant} = \left\langle \widetilde{f}, \chi_X \right\rangle = \lim_{k\to\infty} \left\langle \pi_{a_{n_k}} f_1, \chi_X \right\rangle = \lim_{k\to\infty} \left\langle f_1, \pi_{a_{n_k}}^* \chi_X \right\rangle = \lim_{k\to\infty} \langle f_1, \chi_X \rangle = \int_X f_1 \, \mathrm{d}m_X.$$

This means $\widetilde{f} = \int_X f_1 \, \mathrm{d}m_X \cdot \chi_X$. By compactness and the trick where the limit is independent of the subsequence (I think this is by functional analysis), we see that $\pi_{a_n} f_1 \rightharpoonup \widetilde{f} = \int_X f_1 \, \mathrm{d}m_X \cdot \chi_X$. Then we have

$$\begin{aligned} \langle U_T^n f_1, f_2 \rangle = \langle \pi_{a_n} f_1, f_2 \rangle &= \int_X (\pi_{a_n} f_1) \, \overline{f_2} \, \mathrm{d}m_X \\ &\to \int_X \left(\int_X f_1 \, \mathrm{d}m_X \right) \chi_X \cdot \overline{f_2} \, \mathrm{d}m_X = \int_X f_1 \, \mathrm{d}m_X \int_X \overline{f_2} \, \mathrm{d}m_X \end{aligned}$$

as $n \to \infty$, which is as desired. □

Theorem 6.25 (Sarnak)**.** *Let $X = \mathrm{PSL}_2(\mathbb{Z}) \backslash \mathrm{PSL}_2(\mathbb{R})$. Consider the following periodic orbits (horizontal segment).*



Then the low-lying periodic orbits equidistribute. The periodic orbits measures for the horocycle flow converge in weak topology to the normalized Haar measure on X.*

Theorem 6.26 (Furstenberg)**.** *If $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ is a uniform lattice, then the horocycle flow is uniquely ergodic.*

Theorem 6.27 (Dani)**.** *Let $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ is a non-uniform lattice and deonte $X := \Gamma \backslash \mathrm{PSL}_2(\mathbb{R})$. Then any horocycle invariant and ergodic probability measure is either m_X or a periodic orbit measure.*

7 Entropy

We want to give a measure-preserving system $(X, \mathcal{B}, \mu, T)$ a numerical invariant so that it is isomorphism invariant: (X, μ, T) and (Y, ν, S) are called isomorphic if there exists an invertible and measure-preserving map $\phi\colon X \to Y$ with $\phi_* \mu = \nu$, and $\phi \circ T = S \circ \phi$, i.e. the following diagram is commutative.

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "\phi"'] & X \arrow[d, "\phi"] \\ Y \arrow[r, "S"'] & Y \end{tikzcd}$$

If we don't need ϕ to be invertible, we say Y is a factor of X. Then if (X, μ, T) and (T, ν, S) are isomorphic, their "numerical invariants" are same.

Example 7.1. Let $X = \{0, 1\}^{\mathbb{Z}}$ and T be the shift under uniform measure $\mu_2 = \left(\frac{1}{2}, \frac{1}{2}\right)$. Let $Y = \{0, 1, 2\}^{\mathbb{Z}}$ and S be the shift under uniform measure $\mu_3 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Then they are isomorphic. In fact, they are spectrally isomorphic: There exists a linear map $W\colon L^2\mu_2(X) \to L^2_{\mu_3}(Y)$ with $\langle f, g\rangle_{\mu_2} = \langle Wf, Wg\rangle_{\mu_3}$ for any $f, g \in L^2_{\mu_2}(X)$, and the following diagram is commutative.

$$\begin{tikzcd} L^2_{\mu_2}(X) \arrow[r, "W"] \arrow[d, "U_T"'] & L^2_{\mu_3}(Y) \arrow[d, "U_S"] \\ L^2_{\mu_2}(X) \arrow[r, "W"'] & L^2_{\mu_3}(Y) \end{tikzcd}$$

Example 7.2. Let $X = \{0, 1, 2, 3\}^{\mathbb{Z}}$ and T be the shift under uniform measure $\mu_4 = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$. Let $Y = \{0, 1, 2, 3, 4\}^{\mathbb{Z}}$ and S be the shift under the measure $\mu = \left(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$. Then they are also isomorphic.

Let (X, μ, T) be a measure-preserving system. (We omit the symbol of σ-algebra $\mathcal{B}$.) Then its entropy $h_\mu(X, T)$ will be

$$h_\mu(X, T) = \text{"rate in which we lose information overtime"}.$$

Then high entropy means chaos while zero entropy means order.

7.1 Entropy of partition

Let $\mathcal{P}$ be a partition of X. The entropy $H(\mathcal{P})$ will be

$$H(\mathcal{P}) = \text{uncertainty in } \mathcal{P} = \text{amount of information gained by performing the "partition experiment"}.$$

Formally, we first fix some notations.

Definition 7.3. *Let $(X, \mathcal{B}, \mu)$ be a measure space.*

- $\zeta = (A_1, \cdots, A_n)$ *or* $\zeta = (A_1, A_2, \cdots)$ *be a partition of* X
- *For any* $x \in X$, $[x]_\zeta$ *is the partition element contains* x.
- $\sigma(\zeta)$ *is the* σ*-algebra generated by* ζ.
- *For another partition* $\xi = (B_1, B_2, \cdots)$, $\zeta \vee \xi$ *is the joint of* ζ *and* ξ, *whose partition elements are look like* $A_i \cap B_j$.
- $\sigma(\zeta) \vee \sigma(\xi) := \bigcap_{\sigma(\zeta), \sigma(\xi) \subset \mathcal{B}} \mathcal{B}$. *We have* $\sigma(\zeta) \vee \sigma(\xi) = \sigma(\zeta \vee \xi)$. *For countablely many* ζ_n, *we also define*
$$\bigvee_{n=1}^{\infty} \zeta_n := \bigcap_{\sigma(\zeta_n) \subset \mathcal{B}, \forall n} \mathcal{B}.$$
- *For a map* $T\colon X \to X$, $T^{-1}(\zeta) := \left\{T^{-1}A_1, T^{-1}A_2, \cdots\right\}$.

Definition 7.4 (Entropy of a partition)**.** *Let (X,μ) be a measure space and ζ be a partition of X. The entropy of ζ is defined by*

$$H_\mu(\zeta) = H\left(\mu\left(A_1\right), \mu\left(A_2\right), \cdots\right) \coloneqq \sum_{i=1}^{\infty} -\mu\left(A_i\right) \cdot \log\left(\mu\left(A_i\right)\right) \in [0,\infty]$$

where we define $0\log 0 = 0$.

Let $\xi = (B_1, B_2, \cdots)$ be another partition of X. The entropy of ζ given ξ is defined by

$$\begin{aligned} H_\mu(\zeta|\xi) &\coloneqq \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mu\left(A_j \cap B_i\right) \cdot \log\left(\frac{\mu\left(A_j \cap B_i\right)}{\mu\left(B_i\right)}\right) \\ &= \sum_{i=1}^{\infty} \mu\left(B_i\right) \cdot \sum_{j=1}^{\infty} \frac{\mu\left(A_j \cap B_i\right)}{\mu\left(B_i\right)} \cdot \log\left(\frac{\mu\left(A_j \cap B_i\right)}{\mu\left(B_i\right)}\right) \\ &= \sum_{i=1}^{\infty} \mu\left(B_i\right) \cdot H_{\mu|_{B_i}}\left(\zeta|_{B_i}\right) \end{aligned}$$

where $\zeta|_{B_i} \coloneqq \{A_1 \cap B_i, A_2 \cap B_2, \cdots\}$ and $\mu|_{B_1}(A) \coloneqq \frac{1}{\mu(B_i)}\mu(A \cap B_i)$.

Example 7.5. Let $X = \{0, 1, 2, \cdots, 127\}$ under uniform measure μ. Then

$$H_\mu(\{1\}, \{2\}, \cdots, \{127\}) = \sum_{i=0}^{127} -\frac{1}{128}\log\left(\frac{1}{128}\right) = 7.$$

This can be explained by a game as below: Let someone choose a number in X randomly (i.e. under uniform measure). You should ask he/she some questions to obtain the number. The worth way is to ask "Is it 73?" and so on. A better way is to ask "Is it $\leqslant 63$?", "If so, is it $\leqslant 31$?" and so on. By asking by dichotomy, we need at most 7 times to get the right number.

Note that our definition of H also validates for number sequence. Denote

$$\Delta_k = \left\{(p_1, \cdots, p_k) : \sum_{i=1}^{k} p_i = 1\right\}$$

and

$$\Delta = \bigcup_{k\in\mathbb{N}} \Delta_k.$$

Then H will be a function

$$\begin{aligned} H \colon \Delta &\to [0,\infty] \\ (p_1, \cdots, p_k) &\mapsto \sum_{i=1}^{k} -p_i \log p_i. \end{aligned}$$

Proposition 7.6. *H is the unique function on Δ so that*

1. *$H \geqslant 0$. $H(p_1, \cdots, p_k) = 0$ if and only if there exists $1 \leqslant i_0 \leqslant k$ with $p_{i_0} = 1$.*
2. *$H(p_1, \cdots, p_k) = H(p_1, \cdots, p_k, 0)$.*
3. *$\max_{\underline{x}\in\Delta_k} H(\underline{x}) = \log k$ with equality if and only if $\underline{x} = \left(\frac{1}{k}, \frac{1}{k}, \cdots, \frac{1}{k}\right)$.*
4. *$H_\mu(\zeta \vee \xi) = H_\mu(\xi) + H(\zeta|\xi)$.*

Proof. We will not prove this whole proposition. We only prove that the entropy H has properties (1)-(4). **(1)(2)(4)** are by direct calculation.

Before prove (3), we need some facts about convex functions. We omit their proof. Recall that a function $\psi \colon I \to \mathbb{R}$ is called convex if

$$\psi\left(\sum_{i=1}^{n} t_i x_i\right) \leqslant \sum_{i=1}^{n} t_i \psi\left(x_i\right)$$

and equality holds if and only if $x_1 = \cdots = x_n$ and $t_1 = \cdots = t_n = \frac{1}{n}$. For a convex function ψ, we have the Jenson inequality: For any $s < t < u$,

$$\frac{\psi(t)-\psi(s)}{t-s} \leqslant \frac{\psi(u)-\psi(s)}{u-s}.$$

For a continuous convex function $\psi\colon I \to \mathbb{R}$ and a measurable function $f\colon X \to I$ on a probability space $(X, \mathcal{B}, \mu)$, we have

$$\psi\left(\int_X f(x)\,\mathrm{d}\mu(x)\right) \leqslant \int_X \psi(f(x))\,\mathrm{d}\mu(x).$$

(3): Note that $\phi(x) := x\log x$ is convex. Then

$$\frac{1}{k}\log\left(\frac{1}{k}\right) = \phi\left(\frac{1}{k}\right) = \phi\left(\sum_{i=1}^{k}\frac{1}{k}p_i\right) \leqslant \sum_{i=1}^{k}\frac{1}{k}\phi\left(p_i\right) = \frac{1}{k}\sum_{i=1}^{k}p_i\log p_i.$$

Hence for any, $\underline{x} = (p_1, \cdots, p_k) in \Delta_k$, we have

$$H\left(\underline{x}\right) = H\left(p_1, \cdots, p_k\right) = -\sum_{i=1}^{k} p_i \log p_i \leqslant -\log\left(\frac{1}{k}\right) = \log k.$$

Therefore

$$\max_{\underline{x}\in\Delta_k} H(\underline{x}) = \log k$$

with equality if and only if $\underline{x} = \left(\frac{1}{k}, \frac{1}{k}, \cdots, \frac{1}{k}\right)$.

□

Definition 7.7. *The information function of a partition ζ of $(X, \mathcal{B}, \mu)$ is defined by*

$$I_\mu(\zeta)(x) := -\log\left(\mu\left([x]_\zeta\right)\right) \geqslant 0.$$

For another partition ξ, the conditional information function given ξ is defined by

$$I_\mu(\zeta|\xi)(x) := -\log\left(\frac{\mu\left([x]_\zeta \cap [x]_\xi\right)}{\mu\left([x]_\xi\right)}\right).$$

Proposition 7.8 (Additivity and Monotonicity)**.** *Let ζ, ξ, η be partitions of $(X, \mathcal{B}, \mu)$. Then*

1.

$$H_\mu(\zeta) = \int_X I_\mu(\zeta)\,\mathrm{d}\mu \quad \text{and} \quad H_\mu(\xi|\eta) = \int_X I_\mu(\xi|\eta)\,\mathrm{d}\mu.$$

2.

$$I_\mu(\xi\vee\eta|\zeta) = I_\mu(\eta|\zeta) + I_\mu(\xi|\eta\vee\zeta) \quad \text{and} \quad H_\mu(\xi\vee\eta|\zeta) = H_\mu(\eta|\zeta) + H_\mu(\xi|\eta\vee\zeta).$$

3. Monotonicity for the measured partition:

$$I_\mu(\eta|\zeta) \leqslant I_\mu(\eta\vee\xi|\zeta) \quad \text{and} \quad H_\mu(\eta|\zeta) \leqslant H_\mu(\eta\vee\xi|\zeta).$$

4. Monotonicity for the given partition and entropy:

$$H_\mu(\xi|\eta\vee\zeta) \leqslant H_\mu(\xi|\zeta).$$

5. Subadditivity:

$$H_\mu(\xi\vee\eta|\zeta) \leqslant H_\mu(\xi|\zeta) + H_\mu(\eta|\zeta).$$

Proof. Let $\zeta = (A_1, A_2, \cdots), \xi = (B_1, B_2, \cdots), \eta = (C_1, C_2, \cdots)$.

(1): By definition,

$$\begin{aligned}\int_X I_\mu(\zeta)\,\mathrm{d}\mu &= \sum_{i=1}^{\infty}\int_{A_i} I_\mu(\zeta)\,\mathrm{d}\mu = \sum_{i=1}^{\infty}\int_{A_i} -\log\left(\mu\left([x]_\zeta\right)\right)\,\mathrm{d}\mu(x)\\ &= \sum_{i=1}^{\infty}\int_{A_i} -\log\left(\mu\left(A_i\right)\right)\,\mathrm{d}\mu = \sum_{i=1}^{\infty} -\mu\left(A_i\right)\cdot\log\left(\mu\left(A_i\right)\right) = H_\mu(\zeta),\end{aligned}$$

$$\int_X I_\mu(\xi|\eta)\,\mathrm{d}\mu = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\int_{B_j\cap C_i} I_\mu(\xi|\eta)\,\mathrm{d}\mu = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\int_{B_j\cap C_i} -\log\left(\frac{\mu\left(B_j\cap C_i\right)}{\mu\left(C_i\right)}\right)\mathrm{d}\mu$$
$$= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} -\mu\left(B_j\cap C_i\right)\cdot\log\left(\frac{\mu\left(B_j\cap C_i\right)}{\mu\left(C_i\right)}\right)\mathrm{d}\mu = H_\mu(\xi|\eta).$$

(2): For any $x\in X$, we denote $P=[x]_\zeta, Q=[x]_\xi, R=[x]_\eta$ for simplicity. Then $P\cap Q=[x]_{\zeta\vee\xi}, Q\cap R=[x]_{\xi\vee\eta}, R\cap P=[x]_{\eta\vee\zeta}$ by definition. Therefore

$$I_\mu(\eta|\zeta)(x)+I_\mu(\xi|\eta\vee\zeta)(x) = -\log\left(\frac{\mu(R\cap P)}{\mu(P)}\right)-\log\left(\frac{\mu(Q\cap(R\cap P))}{\mu(R\cap P)}\right)$$
$$= -\log\left(\frac{\mu(R\cap P)}{\mu(P)}\cdot\frac{\mu(Q\cap R\cap P)}{\mu(R\cap P)}\right)$$
$$= -\log\left(\frac{\mu(Q\cap R)\cap P)}{\mu(P)}\right) = I_\mu(\xi\vee\eta|\zeta)(x).$$

And by (1),

$$H_\mu(\xi\vee\eta|\zeta) = \int_X I_\mu(\xi\vee\eta|\zeta)\,\mathrm{d}\mu = \int_X I_\mu(\eta|\zeta)\,\mathrm{d}\mu + \int_X I_\mu(\xi|\eta\vee\zeta)\,\mathrm{d}\mu = H_\mu(\eta|\zeta)+I_\mu(\xi|\eta\vee\zeta).$$

(3): For any $x\in X$,

$$I_\mu(\eta|\zeta)(x) = -\log\left(\frac{\mu(P\cap R)}{\mu(P)}\right) \leqslant -\log\left(\frac{\mu(P\cap(R\cap Q))}{\mu(P)}\right) = I_\mu(\eta\vee\xi|\zeta)(x)$$

and then by (1),

$$H_\mu(\eta|\zeta) = \int_X I_\mu(\eta|\zeta)\,\mathrm{d}\mu \leqslant \int_X I_\mu(\eta\vee\xi|\zeta) = H_\mu(\eta\vee\xi|\zeta).$$

(4): We first prove for the case $\zeta=\{X\}$. Let $\phi(x)=x\log x$ be convex. By Jenson inequality,

$$H_\mu(\xi|\eta) = -\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\mu\left(B_i\cap C_j\right)\cdot\log\left(\frac{\mu\left(B_i\cap C_j\right)}{\mu\left(C_j\right)}\right) = -\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\mu\left(C_j\right)\cdot\phi\left(\frac{\mu\left(B_i\cap C_j\right)}{\mu\left(C_j\right)}\right)$$
$$\leqslant -\sum_{i=1}^{\infty}\phi\left(\sum_{j=1}^{\infty}\mu\left(C_j\right)\cdot\frac{\mu\left(B_i\cap C_j\right)}{\mu\left(C_j\right)}\right) = -\sum_{i=1}^{\infty}\phi\left(\sum_{j=1}^{\infty}\mu\left(B_i\cap C_j\right)\right)$$
$$= -\sum_{i=1}^{\infty}\phi\left(B_i\right) = -\sum_{i=1}^{\infty}\mu\left(B_i\right)\cdot\log\left(\mu\left(B_i\right)\right) = H_\mu(\xi).$$

Then we have

$$H_\mu(\xi|\eta\vee\zeta) = H_\mu(\xi|\eta) = H_\mu(\xi) = H_\mu(\xi|\zeta).$$

For the general case, we first change the measure to $\mu_i := \frac{\mu|_{A_i}}{\mu(A_i)}$. Then by previous case,

$$H_{\mu_i}(\xi|\eta) \leqslant H_{\mu_i}(\xi),$$

where

$$H_{\mu_i}(\xi|\eta) = -\sum_{j,k=1}^{\infty}\mu_i\left(B_j\cap C_k\right)\cdot\log\left(\frac{\mu_i\left(B_j\cap C_k\right)}{\mu_i\left(C_k\right)}\right)$$
$$= -\sum_{j,k=1}^{\infty}\frac{\mu\left(A_i\cap B_j\cap C_k\right)}{\mu\left(A_i\right)}\cdot\log\left(\frac{\mu\left(A_i\cap B_j\cap C_k\right)}{\mu\left(A_i\cap C_k\right)}\right)$$

and

$$H_{\mu_i}(\xi) = -\sum_{j=1}^{\infty}\mu_i\left(B_j\right)\cdot\log\left(\mu_i\left(B_j\right)\right)$$
$$= -\sum_{j=1}^{\infty}\frac{\mu\left(A_i\cap B_j\right)}{\mu\left(A_i\right)}\cdot\log\left(\frac{\mu\left(A_i\cap B_j\right)}{\mu\left(A_i\right)}\right).$$

Hence we have

$$-\sum_{j,k=1}^{\infty} \mu\left(A_i \cap B_j \cap C_k\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_j \cap C_k\right)}{\mu\left(A_i \cap C_k\right)}\right) \leqslant -\sum_{j=1}^{\infty} \mu\left(A_i \cap B_j\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_j\right)}{\mu\left(A_i\right)}\right).$$

Summing over i, we have

$$-\sum_{i,j,k=1}^{\infty} \mu\left(A_i \cap B_j \cap C_k\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_j \cap C_k\right)}{\mu\left(A_i \cap C_k\right)}\right) \leqslant -\sum_{i,j=1}^{\infty} \mu\left(A_i \cap B_j\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_j\right)}{\mu\left(A_i\right)}\right),$$

i.e.

$$\begin{aligned}
H_\mu(\xi|\eta \vee \zeta) &= -\sum_{i=1}^{\infty}\sum_{j,k=1}^{\infty} \mu\left(B_i \cap (A_j \cap C_k)\right) \cdot \log\left(\frac{\mu\left(B_i \cap (A_j \cap C_k)\right)}{\mu\left(A_j \cap C_k\right)}\right) \\
&= -\sum_{i,j,k=1}^{\infty} \mu\left(A_i \cap B_j \cap C_k\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_j \cap C_k\right)}{\mu\left(A_i \cap C_k\right)}\right) \\
&\leqslant -\sum_{i,j=1}^{\infty} \mu\left(A_i \cap B_j\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_j\right)}{\mu\left(A_i\right)}\right) = H_\mu(\xi|\zeta).
\end{aligned}$$

(5): By (2)(4),

$$H_\mu(\xi \vee \eta|\zeta) = H_\mu(\eta|\zeta) + H_\mu(\xi|\eta \vee \zeta) \leqslant H_\mu(\eta|\zeta) + H_\mu(\xi|\zeta) = H_\mu(\xi|\zeta) + H_\mu(\eta|\zeta).$$

□

7.2 Entropy of maps

From now on we assume $T\colon X \to X$ is measure-preserving on a probability space $(X, \mathcal{B}, \mu)$.

Let η be a finite (or finite entropy) partition of X. We can think of η as an experiment.

For indices $k \leqslant l$, we define

$$\eta_k^l := \bigvee_{j=k}^{l} T^{-j}\eta$$

which corresponds to asking which partition elements $T^k x, \cdots, T^l x$ belong to.

Lemma 7.9. *Let ξ, η be two finite partitions of X. Then*

$$H_\mu(\xi|\eta) = H_\mu\left(T^{-1}\xi|T^{-1}\eta\right) \quad \text{and} \quad I_\mu(\xi|\eta) \circ T = I_\mu\left(T^{-1}\xi|T^{-1}\eta\right).$$

Proof. For any $x \in X$, we denote $A = [Tx]_\xi$ and $B = [Tx]_\eta$ for simplicity. Note that $x \in T^{-1}A = [x]_{T^{-1}\xi}$ and $x \in T^{-1}B = [x]_{T^{-1}\eta}$. Because T is measure-preserving, we obtain

$$\begin{aligned}
I_\mu(\xi|\eta) \circ T(x) &= -\log\left(\frac{\mu(A \cap B)}{\mu(B)}\right) \\
&= -\log\left(\frac{\mu\left(T^{-1}(A \cap B)\right)}{\mu\left(T^{-1}B\right)}\right) = -\log\left(\frac{\mu\left(T^{-1}A \cap T^{-1}B\right)}{\mu\left(T^{-1}B\right)}\right) = I_\mu\left(T^{-1}\xi|T^{-1}\eta\right)(x).
\end{aligned}$$

And then by Proposition 7.8 (1), because T is measure-preserving, we have

$$\begin{aligned}
H_\mu(\xi|\eta) &= \int_X I_\mu(\xi|\eta)\,\mathrm{d}\mu = \int_X I_\mu(\xi|\eta) \circ T\,\mathrm{d}\left(T_*^{-1}\mu\right) \\
&= \int_X I_\mu(\xi|\eta) \circ T\,\mathrm{d}\mu = \int_X I_\mu\left(T^{-1}\xi|T^{-1}\eta\right)\,\mathrm{d}\mu = H_\mu\left(T^{-1}\xi|T^{-1}\eta\right).
\end{aligned}$$

□

Definition 7.10. *Let ξ be a finite partition of a measure-preserving probability system $(X, \mathcal{B}, \mu, T)$. Then*

$$h_\mu(T, \xi) := \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\xi_0^{n-1}\right).$$

Moreover,

$$h_\mu(T) := \sup_{\xi \text{ is finite}} h_\mu(T, \xi).$$

Goal. *Our next goal is: Prove the limit exists. Find some good properties of $h_\mu(T,\xi)$. Show that $h_\mu(T)$ is an invariant for measure-preserving transformations.*

Lemma 7.11. *Let $a_n \geqslant 0$ be a subadditive sequence, i.e.*

$$a_{m+n} \leqslant a_m + a_n$$

for any $m, n \in \mathbb{N}$. Then

$$\lim_{n\to\infty} \frac{1}{n} a_n = \inf_{n\in\mathbb{N}} \frac{1}{n} a_n.$$

Proof. First we have

$$\liminf_{n\to\infty} \frac{1}{n} a_n = \geqslant \inf_{n\in\mathbb{N}} \frac{1}{n} a_n.$$

On the other hand, for any $\varepsilon > 0$, there exists $m \in \mathbb{N}$ such that

$$\frac{1}{m} a_m < \inf_{n\in\mathbb{N}} \frac{1}{n} a_n + \varepsilon.$$

For any $n \in \mathbb{N}$, suppose $n = qm + r$ be the division with remainder. We have

$$a_n = a_{qm+r} \leqslant q a_m + a_r$$

and then

$$\frac{1}{n} a_n \leqslant \frac{q}{n} a_m + \frac{1}{n} a_r = \left(1 - \frac{r}{n}\right) \frac{1}{m} a_m + \frac{1}{n} a_r.$$

Note that $0 \leqslant r \leqslant m$, $a_r \in \{a_0, \cdots, a_m\}$ has an upper bound. Then as $n \to \infty$,

$$\limsup_{n\to\infty} \frac{1}{n} a_n \leqslant \limsup_{n\to\infty} \left(\left(1 - \frac{r}{n}\right) \frac{1}{m} a_m + \frac{1}{n} a_r \right) = \frac{1}{m} a_m < \inf_{n\in\mathbb{N}} \frac{1}{n} a_n + \varepsilon.$$

Because ε is arbitrary,

$$\limsup_{n\to\infty} \frac{1}{n} a_n \leqslant \inf_{n\in\mathbb{N}} \frac{1}{n} a_n.$$

Above all, we have

$$\inf_{n\in\mathbb{N}} \frac{1}{n} a_n \leqslant \liminf_{n\to\infty} \frac{1}{n} a_n \leqslant \limsup_{n\to\infty} \frac{1}{n} a_n \leqslant \inf_{n\in\mathbb{N}} \frac{1}{n} a_n$$

and then

$$\lim_{n\to\infty} \frac{1}{n} a_n = \inf_{n\in\mathbb{N}} \frac{1}{n} a_n$$

which is as desired. □

Lemma 7.12. *$a_n := H_\mu(\xi_0^{n-1})$ is subadditive. Then by Lemma 7.11, its limit*

$$h_\mu(T,\xi) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_0^{n-1}) = \inf_{n\in\mathbb{N}} \frac{1}{n} H_\mu(\xi_0^{n-1})$$

exists.

Proof. For any $m, n \in \mathbb{N}$, by Proposition 7.8 (5),

$$a_{m+n} = H_\mu\left(\xi_0^{m+n-1}\right) = H_\mu\left(\xi_0^{m-1} \vee \xi_m^{n+m-1}\right) \leqslant H_\mu\left(\xi_0^{m-1}\right) + H_\mu\left(\xi_m^{m+n-1}\right).$$

Note that

$$\xi_m^{m+n-1} = \bigvee_{i=m}^{n+m-1} T^{-j}\xi = T^{-m}\left(\bigvee_{i=0}^{n-1} T^{-j}\xi\right) = T^{-j}\left(\xi_0^{n-1}\right).$$

By Lemma 7.9,

$$H_\mu\left(\xi_m^{m+n-1}\right) = H_\mu\left(T^{-m}\left(\xi_0^{n-1}\right)\right) = H_\mu\left(\xi_0^{n-1}\right).$$

Then we have

$$a_{m+n} \leqslant H_\mu\left(\xi_0^{m-1}\right) + H_\mu\left(\xi_m^{m+n-1}\right) = H_\mu\left(\xi_0^{m-1}\right) + H_\mu\left(\xi_0^{n-1}\right) = a_m + a_n$$

which is as desired. □

Proposition 7.13 (Future formula)**.** *Let ξ be a finite partition. Then*

$$h_\mu(T,\xi) = \lim_{n\to\infty} H_\mu\left(\xi|\xi_1^n\right).$$

Proof. Denote $b_{n-1} = H_\mu\left(\xi|\xi_1^n\right) = H_\mu\left(\xi|T^{-1}\xi \vee \cdots \vee T^{-(n-1)}\xi\right)$. By Proposition 7.8 (2) and Lemma 7.9,

$$\begin{aligned} H_\mu\left(\xi_0^{n-1}\right) &= H_\mu\left(\xi \vee T^{-1}\xi \vee \cdots \vee T^{-(n-1)}\xi\right) \\ &= H_\mu\left(T^{-1}\xi \vee \cdots \vee T^{-(n-1)}\xi\right) + H_\mu\left(\xi|T^{-1}\xi \vee \cdots \vee T^{-(n-1)}\xi\right) \\ &= H_\mu\left(T^{-1}\left(\xi \vee \cdots \vee T^{-(n-2)}\xi\right)\right) + b_{n-1} \\ &= H_\mu\left(\xi \vee \cdots \vee T^{-(n-2)}\xi\right) + b_{n-1} = H_\mu\left(\xi_0^{n-2}\right) + b_{n-1}. \end{aligned}$$

We repeat this process for the first term and get

$$H_\mu\left(\xi_0^{n-2}\right) = H_\mu\left(\xi_0^{n-3}\right) + b_{n-2}.$$

Repeating this process inductively, we have

$$H_\mu\left(\xi_0^{n-1}\right) = b_0 + b_1 + \cdots + b_{n-1} = \sum_{j=0}^{n-1} b_j.$$

Recall that by Lemma 3.41, if $b_n \to b$ in $\mathbb{R}$, then $\frac{1}{n}\sum_{j=0}^{n-1} b_j \to b$. Then by Lemma 7.12,

$$h_\mu(T,\xi) = \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\xi_0^{n-1}\right) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} b_j = \lim_{n\to\infty} b_n = \lim_{n\to\infty} H_\mu\left(\xi|\xi_1^n\right).$$

□

Proposition 7.14. *Let ξ, η be finite partitions. Then*

1. *$h_\mu(T,\xi) \leqslant H_\mu(\xi)$.*
2. *$h_\mu(T,\xi \vee \eta) \leqslant h_\mu(T,\xi) + h_\mu(T,\eta)$.*
3. *Continuity bound: $h_\mu(T,\eta) \leqslant h_\mu(T,\xi) + H_\mu(\eta|\xi)$.*

Proof. **(1):** By Lemma 7.12,

$$h_\mu(T,\xi) = \inf_{n\in\mathbb{N}} \frac{1}{n} H_\mu\left(\xi_0^{n-1}\right) \leqslant \frac{1}{1} H_\mu\left(\xi_0^{1-1}\right) = H_\mu(\xi).$$

(2): Note that

$$(\xi \vee \eta)_0^{n-1} = \xi_0^{n-1} \vee \eta_0^{n-1}.$$

By Proposition 7.8 (5), we calculate directly:

$$\begin{aligned} h_\mu(T,\xi \vee \eta) &= \lim_{n\to\infty} \frac{1}{n} H_\mu\left((\xi \vee \eta)_0^{n-1}\right) = \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\xi_0^{n-1} \vee \eta_0^{n-1}\right) \\ &\leqslant \lim_{n\to\infty} \frac{1}{n}\left(H_\mu\left(\xi_0^{n-1}\right) + H_\mu\left(\eta_0^{n-1}\right)\right) \\ &= \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\xi_0^{n-1}\right) + \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\eta_0^{n-1}\right) = h_\mu(T,\xi) + h_\mu(T,\xi). \end{aligned}$$

(3): By Proposition 7.8 (3)(2),

$$h_\mu(T,\eta) = \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\eta_0^{n-1}\right) \leqslant \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\xi_0^{n-1} \vee \eta_0^{n-1}\right) = \lim_{n\to\infty} \frac{1}{n}\left(H_\mu\left(\xi_0^{n-1}\right) + H_\mu\left(\eta_0^{n-1}|\xi_0^{n-1}\right)\right).$$

For the first term,

$$\lim_{n\to\infty} \frac{1}{n} H_\mu\left(\xi_0^{n-1}\right) = h_\mu(T,\xi).$$

For the second term, by Proposition 7.8 (5),

$$\begin{aligned} H_\mu\left(\eta_0^{n-1}|\xi_0^{n-1}\right) &= H_\mu\left(\eta \vee T^{-1}\eta \vee \cdots \vee T^{-(n-1)}\eta|\xi_0^{n-1}\right) \\ &\leqslant H_\mu\left(\eta|\xi_0^{n-1}\right) + H_\mu\left(T^{-1}\eta|\xi_0^{n-1}\right) + \cdots + H_\mu\left(T^{-(n-1)}\eta|\xi_0^{n-1}\right) = \sum_{j=0}^{n-1} H_\mu\left(T^{-j}\eta|\xi_0^{n-1}\right). \end{aligned}$$

And by Proposition 7.8 (4), for any $0 \leqslant j \leqslant n-1$,

$$H_\mu\left(T^{-j}\eta|\xi_0^{n-1}\right) = H_\mu\left(T^{-j}\eta|\xi \vee T^{-1}\xi \vee \cdots \vee T^{-(n-1)}\xi\right) \leqslant H_\mu\left(T^{-j}\eta|T^{-j}\xi\right).$$

Hence by Lemma 7.9, the second term becomes

$$H_\mu\left(\eta_0^{n-1}|\xi_0^{n-1}\right) \leqslant \sum_{j=0}^{n-1} H_\mu\left(T^{-j}\eta|\xi_0^{n-1}\right) \leqslant \sum_{j=0}^{n-1} H_\mu\left(T^{-j}\eta|T^{-j}\xi\right) = \sum_{j=0}^{n-1} H_\mu\left(\eta|\xi\right).$$

Above all,

$$\begin{aligned} h_\mu(T,\eta) &= \lim_{n\to\infty}\frac{1}{n}\left(H_\mu\left(\xi_0^{n-1}\right) + H_\mu\left(\eta_0^{n-1}|\xi_0^{n-1}\right)\right) = \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\xi_0^{n-1}\right) + \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\eta_0^{n-1}|\xi_0^{n-1}\right) \\ &\leqslant h_\mu(T,\xi) + \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} H_\mu\left(\eta|\xi\right) = h_\mu(T,\xi) + H_\mu\left(\eta|\xi\right). \end{aligned}$$

□

Proposition 7.15 (Iterates)**.** *1. $h_\mu(T,\xi) = h_\mu\left(T,\xi_0^k\right)$ for any $k \geqslant 1$.*

2. If T is invertible, then $h_\mu(T,\xi) = h_\mu\left(T^{-1},\xi\right)$.

3. $h_\mu\left(T^k\right) = k \cdot h_\mu(T)$ for any $k \geqslant 1$.

4. $h_\mu(T) = h_\mu\left(T^{-1}\right)$ if T is invertible.

Proof. **(1):** For any $k \geqslant 1$, we have

$$h_\mu\left(T,\xi_0^k\right) = \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\left(\xi_0^k\right)_0^{n-1}\right).$$

Note that

$$\left(\xi_0^k\right)_0^{n-1} = \left(\bigvee_{i=0}^{k} T^{-i}\xi\right)_0^{n-1} = \bigvee_{j=0}^{n-1} T^{-j}\left(\bigvee_{i=0}^{k} T^{-i}\xi\right) = \bigvee_{i=0}^{k+n-1} T^{-i}\xi = \xi_0^{k+n-1}.$$

Hence we have

$$\begin{aligned} h_\mu\left(T,\xi_0^k\right) &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\left(\xi_0^k\right)_0^{n-1}\right) = \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\xi_0^{k+n-1}\right) \\ &= \lim_{n\to\infty}\frac{k+n}{n}\frac{1}{k+n}H_\mu\left(\xi_0^{k+n-1}\right) = 1\cdot\lim_{n\to\infty}\frac{1}{k+n}H_\mu\left(\xi_0^{k+n-1}\right) = h_\mu(T,\xi). \end{aligned}$$

(2): If T is invertible, by Lemma 7.9,

$$\begin{aligned} h_\mu\left(T^{-1},\xi\right) &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\bigvee_{j=0}^{n-1}\left(T^{-1}\right)^{-j}\xi\right) = \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\xi_{-(n-1)}^0\right) \\ &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(T^{n-1}\xi_0^{n-1}\right) = \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\xi_0^{n-1}\right) = h_\mu(T,\xi). \end{aligned}$$

(3): Recall that $h_\mu(T) = \sup_{\xi \text{ is finite}} h_\mu(T,\xi)$. Let $k \geqslant 1$. For any finite partition ξ,

$$\begin{aligned} h_\mu\left(T^k,\xi_0^{k-1}\right) &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\bigvee_{j=0}^{n-1}\left(T^k\right)^{-j}\left(\xi_0^{k-1}\right)\right) = \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\bigvee_{j=0}^{n-1}T^{-kj}\left(\bigvee_{i=0}^{k-1}T^{-i}\xi\right)\right) \\ &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\bigvee_{j=0}^{kn-1}T^{-j}\xi\right) \\ &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\xi_0^{kn-1}\right) = k\cdot\lim_{n\to\infty}\frac{1}{kn}H_\mu\left(\xi_0^{kn-1}\right) = k\cdot h_\mu(T,\xi). \end{aligned} \tag{18}$$

Then by Proposition 7.8 (3),

$$\begin{aligned}h_\mu\left(T^k,\xi\right) &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\bigvee_{j=0}^{n-1}\left(T^k\right)^{-j}\xi\right) \leqslant \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\bigvee_{j=0}^{n-1}\left(T^k\right)^{-j}\left(\bigvee_{i=0}^{k-1}T^{-i}\xi\right)\right)\\ &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\bigvee_{j=0}^{n-1}\left(T^k\right)^{-j}\left(\xi_0^{k-1}\right)\right) = h_\mu\left(T^k,\xi_0^{k-1}\right) = k\cdot h_\mu(T,\xi) \leqslant k\cdot h_\mu(T).\end{aligned}$$

Hence we have

$$h_\mu\left(T^k\right) \leqslant k\cdot h_\mu(T).$$

Conversely, by equation (18) again,

$$h_\mu(T,\xi) = \frac{1}{k}h_\mu\left(T^k,\xi_0^{k-1}\right) \leqslant \frac{1}{k}h_\mu\left(T^k\right).$$

Hence we have

$$h_\mu(T) \leqslant \frac{1}{k}h_\mu\left(T^k\right).$$

Above all, we conclude

$$h_\mu\left(T^k\right) = k\cdot h_\mu\left(T\right).$$

(4): If T is invertible, by (2),

$$h_\mu(T) = \sup_{\xi\text{ is finite}} h_\mu(T,\xi) = \sup_{\xi\text{ is finite}} h_\mu\left(T^{-1},\xi\right) = h_\mu\left(T^{-1}\right).$$

□

Proposition 7.16. *Entropy is an invariant for measure-preserving isomorphisms: Suppose $(X,\mathcal{B},\mu,T)$ and $(Y,\mathcal{C},\nu,S)$ are two measure-preserving systems on probability spaces. Suppose also there exists a measure-preserving invertible map $\phi\colon X\to Y$ with $\phi\circ T = S\circ\phi$. Then $h_\mu(T) = h_\nu(S)$.*

Proof. For any finite partition $\xi = (B_1,\cdots,B_n)$ of Y, $\phi^{-1}\xi = \left(\phi^{-1}A_1,\cdots,\phi^{-1}A_n,A\right)$ is a finite partition of X, where $A := X - \bigcup_{i=1}^n \phi^{-1}A_i$ is a zero measure set because ϕ is measure-preserving. By definition, we have

$$\begin{aligned}H_\mu\left(\phi^{-1}\xi\right) &= -\mu(A) - \sum_{i=1}^n \mu\left(\phi^{-1}A_i\right)\log\left(\mu\left(\phi^{-1}A_i\right)\right)\\ &= 0 - \sum_{i=1}^n \nu\left(A_i\right)\log\left(\nu\left(A_i\right)\right) = -\sum_{i=1}^n \nu\left(A_i\right)\log\left(\nu\left(A_i\right)\right) = H_\nu(\xi).\end{aligned}$$

Then

$$\begin{aligned}h_\mu\left(T,\phi^{-1}\xi\right) &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\phi^{-1}\xi\vee T^{-1}\left(\phi^{-1}\xi\right)\vee\cdots\vee T^{-(n-1)}\left(\phi^{-1}\xi\right)\right)\\ &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\phi^{-1}\xi\vee\phi^{-1}\left(S^{-1}\xi\right)\vee\cdots\vee\phi^{-(n-1)}\left(S^{-1}\xi\right)\right)\\ &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\phi^{-1}\left(\xi\vee S^{-1}\xi\vee\cdots\vee S^{-1}\xi\right)\right)\\ &= \lim_{n\to\infty}\frac{1}{n}H_\nu\left(\xi\vee S^{-1}\xi\vee\cdots\vee S^{-1}\xi\right) = h_\nu(S,\xi).\end{aligned}$$

By taking limit, for any finite partition ξ of Y,

$$h_\nu(S,\xi) = h_\mu\left(T,\phi^{-1}\xi\right) \leqslant \sup_{\eta\text{ is finite}} h_\mu(T,\eta) = h_\mu(T).$$

Taking limit again, this means

$$h_\nu(S) = \sup_{\xi\text{ is finite}} h_\nu(S,\xi) \leqslant \sup_{\xi\text{ is finite}} h_\mu(T) = h_\mu(T).$$

Note that

$$T\circ\phi^{-1} = \phi^{-1}\circ\phi\circ T\circ\phi^{-1} = \phi^{-1}\circ S\circ\phi\circ\phi^{-1} = \phi^{-1}\circ S.$$

We can do the same process with exchanging the two systems and with ϕ^{-1}, and obtain

$$h_\mu(T) \leqslant h_\nu(S).$$

Above all, we have

$$h_\mu(T) = h_\nu(S)$$

which is as desired.

□

Example 7.17.
- $T_p \curvearrowright \mathbb{T}$ is isomorphic to the shift $\sigma \curvearrowright \{0, \cdots, p-1\}^{\mathbb{N}}$.
- $T_{p^2} \curvearrowright \mathbb{T}$ is isomorphic to $T_p \curvearrowright \mathbb{T}^2$, because $T_{p^2} \curvearrowright \mathbb{T}$ is isomorphic to the shift on p^2-many symbols, and $T_p \curvearrowright \mathbb{T}^2$ is isomorphic to shift on two coordinates where each coordinate is isomorphic to the shift on p-many symbols.

Definition 7.18. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system on a probability space. Let ξ be a finite partition of X. If*

$$\bigvee_{j=0}^{\infty} T^{-j}\xi = \sigma\left(\bigcup_{n=1}^{\infty} \xi_0^n\right) = \mathcal{B},$$

ξ is called a one-side generator. If T is invertible and

$$\bigvee_{j=-\infty}^{\infty} T^{-j}\xi = \sigma\left(\bigcup_{n=-\infty}^{\infty} \xi_0^n\right) = \mathcal{B},$$

ξ is called a generator.

Theorem 7.19 (Kolmogoroff-Sinai)**.** *If ξ is a one-side generator or a generator, then*

$$h_\mu(T) = h_\mu(T, \xi).$$

Before we prove Kolmogoroff-Sinai Theorem, we first need a lemma.

Lemma 7.20. *Suppose ξ is a one-side generator and η is any finite partition. Then*

$$H_\mu\left(\eta | \xi_0^k\right) \to 0$$

as $k \to \infty$.

Proof. We remark that $\mathcal{B}$ is the smallest sub-σ-algebra (of itself) that contains $T^{-j}\xi$ for $j = 0, 1, 2, \cdots$.

Claim. Let $\mathcal{C}$ be the set of $B \in \mathcal{B}$ so that: for any $\varepsilon > 0$, there exists $k \in \mathbb{N}$ and an element of the algebra generated by $\xi, T^{-1}\xi, \cdots, T^{-k}\xi$ so that $\mu(B \Delta A) < \varepsilon$. Then $\mathcal{C}$ is a σ-algebra.

Proof of Claim. **Closed under countable union:** Suppose $B_1, B_2, \cdots \in \mathcal{C}$. We will show that $\bigcup_{j=1}^{\infty} B_j \in \mathcal{C}$. Let $\varepsilon > 0$ arbitrarily. For B_1, B_2, there exists one $k = \max\{k_1, k_2\}$ and $A_1, A_2 \in \sigma\left(\xi, T^{-1}\xi, \cdots, T^{-k}\xi\right)$ so that $\mu(B_1 \Delta A_1) < \frac{\varepsilon}{2}$ and $\mu(B_2 \Delta A_2) < \frac{\varepsilon}{2}$. This implies

$$\mu\left((B_1 \cup B_2)\, \Delta\, (A_1 \cup A_2)\right) \leqslant \mu\left((B_1 \Delta A_1) \cup (B_2 \Delta A_2)\right) \leqslant \mu\left(B_1 \Delta A_1\right) + \mu\left(B_2 \Delta A_2\right) < \varepsilon.$$

Note that $A_1 \cup A_2 \in \sigma\left(\xi, T^{-1}\xi, \cdots, T^{-k}\xi\right)$. Hence $B_1 \cup B_2 \in \mathcal{C}$. By induction, we obtain

$$\bigcup_{j=1}^{n} B_j \in \mathcal{C}$$

for any $n \geqslant 1$.

Let $\varepsilon > 0$ arbitrarily again. We can find $n \geqslant 1$ so that

$$\mu\left(\left(\bigcup_{j=1}^{\infty} B_j\right) - \left(\bigcup_{j=1}^{n} B_j\right)\right) < \frac{\varepsilon}{2}.$$

Note that $\bigcup_{j=1}^{n} B_j \in \mathcal{C}$. There exists $k \geqslant 1$ and $A \in \sigma\left(\xi, T^{-1}\xi, \cdots, T^{-k}\xi\right)$ so that

$$\mu\left(\left(\bigcup_{j=1}^{n} B_j\right) \Delta A\right) < \frac{\varepsilon}{2}.$$

Note that

$$\begin{aligned}
\left(\bigcup_{j=1}^{\infty} B_j\right) \Delta A &= \left(\left(\bigcup_{j=1}^{\infty} B_j\right) - A\right) \cup \left(A - \left(\bigcup_{j=1}^{\infty} B_j\right)\right) \\
&= \left(\left(\bigcup_{j=1}^{\infty} B_j\right) - \left(\bigcup_{j=1}^{n} B_j\right)\right) \cup \left(\left(\bigcup_{j=1}^{n} B_j\right) - A\right) \cup \left(A - \left(\bigcup_{j=1}^{\infty} B_j\right)\right) \\
&\subset \left(\left(\bigcup_{j=1}^{\infty} B_j\right) - \left(\bigcup_{j=1}^{n} B_j\right)\right) \cup \left(\left(\bigcup_{j=1}^{n} B_j\right) - A\right) \cup \left(A - \left(\bigcup_{j=1}^{n} B_j\right)\right) \\
&= \left(\left(\bigcup_{j=1}^{\infty} B_j\right) - \left(\bigcup_{j=1}^{n} B_j\right)\right) \cup \left(\left(\bigcup_{j=1}^{n} B_j\right) \Delta A\right).
\end{aligned}$$

Hence we have

$$\mu\left(\left(\bigcup_{j=1}^{\infty} B_j\right) \Delta A\right) \leqslant \mu\left(\left(\bigcup_{j=1}^{\infty} B_j\right) - \left(\bigcup_{j=1}^{n} B_j\right)\right) + \mu\left(\left(\bigcup_{j=1}^{n} B_j\right) \Delta A\right) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This means $\bigcup_{j=1}^{\infty} B_j \in \mathcal{C}$.

Closed under complement: Suppose $B \in \mathcal{C}$, we will show that $X - B \in \mathcal{C}$. For nay $\varepsilon > 0$, there exists $k \geqslant 1$ and $A \in \sigma\left(\xi, T^{-1}\xi, \cdots, T^{-k}\xi\right)$ so that $\mu(B\Delta A) < \varepsilon$. Note that $X - A \in \sigma\left(\xi, T^{-1}\xi, \cdots, T^{-k}\xi\right)$ and

$$(X - B)\Delta(X - A) = ((X - B) - (X - A)) \cup ((X - A) - (X - B)) = (A - B) \cup (B - A) = B\Delta A.$$

We have

$$\mu((X - B)\Delta(X - A)) = \mu(B\Delta A) < \varepsilon.$$

Hence $X - B \in \mathcal{C}$.

Above all, $\mathcal{C}$ is a σ-algebra. $\square$

Note that $T^{-j}\xi \in \mathcal{C}$ for any $j = 0, 1, 2, \cdots$. (Just choose $k = j$ and $A = T^{-j}\xi$.) The claim implies that $\mathcal{C} = \mathcal{B}$ because $\mathcal{B}$ is the smallest one.

For any finite partition $\eta = (B_1, \cdots, B_m)$ of X, we apply the claim for each B_j and get an approximation finite partition $\zeta = (A_1, \cdots, A_m)$ so that $A_j \in \mathcal{B} = \mathcal{C}$ as follows: For any $\delta > 0$,

- Find $A_1 \in \mathcal{C}$ so that $\mu(A_1\Delta B_1) < \frac{\delta}{2^m}$.
- Find $A_2' \in \mathcal{C}$ so that $\mu(A_2'\Delta B_2) < \frac{\delta}{2^m}$. Let $A_2 = A_2' - A_1$. Then

  $$A_2\Delta B_2 = (A_2 - B_2) \cup (B_2 - A_2) \subset \left(A_2' - B_2\right) \cup \left(B_2 - A_2'\right) \cup (A_1 \cap B_2) = \left(A_2'\Delta B_2\right) \cup (A_1 \cap B_2).$$

  And note that $B_1 \cap B_2 = \varnothing$,

  $$A_1 \cap B_2 \subset A_1 - B_1 \subset A_1\Delta B_1.$$

  We have

  $$A_2\Delta B_2 \subset \left(A_2'\Delta B_2\right) \cup (A_1 \cap B_2) \subset \left(A_2'\Delta B_2\right) \cup (A_1\Delta B_1)$$

  and then

  $$\mu(A_2\Delta B_2) \leqslant \mu\left(\left(A_2'\Delta B_2\right) \cup (A_1\Delta B_1)\right) \leqslant \mu\left(A_2'\Delta B_2\right) + \mu(A_1\Delta B_1) < \frac{\delta}{2^m} + \frac{\delta}{2^m} = \frac{\delta}{2^{m-1}}.$$

- Iterates this process: Find $A_j' \in \mathcal{C}$ so that $\mu(A_j'\Delta B_j) < \frac{\delta}{2^{m-j+2}}$ and let $A_j = A_j' - \bigsqcup_{i=1}^{j-1} A_i$. We have $\mu(A_j\Delta B_j) < \frac{\delta}{2^{m-j+1}}$.

As conclusion, for any finite partition $\eta = (B_1, \cdots, B_m)$ of X, for any $\delta > 0$, we can find an approximation finite partition $\zeta = (A_1, \cdots, A_m)$ with $A_j \in \mathcal{C} = \mathcal{B}$ and $\mu(A_j \Delta B_j) < \delta$. Moreover, by constructions, we have a universal $K \geqslant 1$ so that for any $k \geqslant K$, $A_j \in \sigma(\xi, T^{-1}\xi, \cdots, T^{-k}\xi) = \sigma(\xi_0^k)$. Note that ξ_0^k is a subdivision of ζ. By Proposition 7.8 (4), we obtain

$$\begin{aligned} H_\mu\left(\eta|\xi_0^k\right) \leqslant H_\mu\left(\eta|\zeta\right) &= -\sum_{i,j=1}^m \mu\left(A_i \cap B_j\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_j\right)}{\mu\left(A_i\right)}\right) \\ &= -\sum_{i=1}^m \mu\left(A_i \cap B_i\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_i\right)}{\mu\left(A_i\right)}\right) - \sum_{\substack{i,j=1 \\ i \neq j}}^m \mu\left(A_i \cap B_j\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_j\right)}{\mu\left(A_i\right)}\right). \end{aligned}$$

For the first term, note that

$$A_i = \left(A_i \cap B_i\right) \sqcup \left(A_i - B_i\right).$$

Then

$$\mu\left(A_i\right) = \mu\left(A_i \cap B_i\right) + \mu\left(A_i - B_i\right) \leqslant \mu\left(A_i \cap B_i\right) + \mu\left(A_i \Delta B_i\right) < \mu\left(A_i \cap B_i\right) + \delta.$$

Hence

$$\begin{aligned} -\mu\left(A_i \cap B_i\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_i\right)}{\mu\left(A_i\right)}\right) &= \mu\left(A_i \cap B_i\right) \cdot \log\left(\frac{\mu\left(A_i\right)}{\mu\left(A_i \cap B_i\right)}\right) \\ &\leqslant 1 \cdot \log\left(\frac{\mu\left(A_i \cap B_i\right) + \delta}{\mu\left(A_i \cap B_i\right)}\right) = \log\left(1 + \frac{\delta}{\mu\left(A_i \cap B_i\right)}\right). \end{aligned}$$

Besides,

$$B_i \subset \left(A_i \cap B_i\right) \sqcup \left(A_i \Delta B_i\right).$$

Then

$$\mu\left(B_i\right) \leqslant \mu\left(A_i \cap B_i\right) + \mu\left(A_i \Delta B_i\right) < \mu\left(A_i \cap B_i\right) + \delta,$$

i.e.

$$\mu\left(A_i \cap B_i\right) > \mu\left(B_i\right) - \delta.$$

We conclude

$$-\sum_{i=1}^m \mu\left(A_i \cap B_i\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_i\right)}{\mu\left(A_i\right)}\right) \leqslant \sum_{i=1}^m \log\left(1 + \frac{\delta}{\mu\left(A_i \cap B_i\right)}\right) < \sum_{i=1}^m \log\left(1 + \frac{\delta}{\mu\left(B_i\right) - \delta}\right).$$

For the second term, for $i \neq j$, note that $B_i \cap B_j = \varnothing$,

$$A_i \cap B_j \subset A_i - B_i.$$

Hence

$$\mu\left(A_i \cap B_j\right) \leqslant \mu\left(A_i - B_i\right) \leqslant \mu\left(A_i \Delta B_i\right) < \delta.$$

Note that the function $-x \log x$ is increasing near 0. For δ small enough,

$$-\mu\left(A_i \cap B_j\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_j\right)}{\mu\left(A_i\right)}\right) \leqslant -\delta \cdot \log\left(\frac{\delta}{\mu\left(A_i\right)}\right).$$

Note that

$$A_i = B_i \sqcup \left(A_i - B_i\right).$$

We have

$$\mu\left(A_i\right) = \mu\left(B_i\right) + \mu\left(A_i - B_i\right) \leqslant \mu\left(B_i\right) + \mu\left(A_i \Delta B_i\right) < \mu\left(B_i\right) + \delta.$$

Hence

$$-\mu\left(A_i \cap B_j\right) \cdot \log\left(\frac{\mu\left(A_i \cap B_j\right)}{\mu\left(A_i\right)}\right) \leqslant -\delta \cdot \log\left(\frac{\delta}{\mu\left(A_i\right)}\right) < -\delta \cdot \log\left(\frac{\delta}{\mu\left(B_i\right) + \delta}\right).$$

Above all, we conclude

$$
\begin{aligned}
H_\mu\left(\eta|\xi_0^k\right) &\leqslant -\sum_{i=1}^{m}\mu\left(A_i\cap B_i\right)\cdot\log\left(\frac{\mu\left(A_i\cap B_i\right)}{\mu\left(A_i\right)}\right)-\sum_{\substack{i,j=1\\ i\neq j}}^{m}\mu\left(A_i\cap B_j\right)\cdot\log\left(\frac{\mu\left(A_i\cap B_j\right)}{\mu\left(A_i\right)}\right)\\
&<\sum_{i=1}^{m}\log\left(1+\frac{\delta}{\mu\left(B_i\right)-\delta}\right)-\sum_{\substack{i,j=1\\ i\neq j}}^{m}\delta\cdot\log\left(\frac{\delta}{\mu\left(B_i\right)+\delta}\right).
\end{aligned}
$$

Note that when $\delta\to 0$ the right side goes to 0.

This implies, for a fixed finite partition $\eta=(B_1,\cdots,B_m)$ of X, for any $\varepsilon>0$, we can find $\delta>0$ so that

$$
\sum_{i=1}^{m}\log\left(1+\frac{\delta}{\mu\left(B_i\right)-\delta}\right)-\sum_{\substack{i,j=1\\ i\neq j}}^{m}\delta\cdot\log\left(\frac{\delta}{\mu\left(B_i\right)+\delta}\right)<\varepsilon.
$$

Then we find an approximation partition $\zeta=(A_1,\cdots,A_m)$ so that $\mu(A_i\Delta B_i)<\delta$, and some K so that for any $k\geqslant K$, $A_j\in\sigma(\xi,T^{-1}\xi,\cdots,T^{-k}\xi)=\sigma(\xi_0^k)$. Hence for any $k\geqslant K$,

$$
H_\mu\left(\eta|\xi_0^k\right)<\sum_{i=1}^{m}\log\left(1+\frac{\delta}{\mu\left(B_i\right)-\delta}\right)-\sum_{\substack{i,j=1\\ i\neq j}}^{m}\delta\cdot\log\left(\frac{\delta}{\mu\left(B_i\right)+\delta}\right)<\varepsilon.
$$

By definition, that's to say,

$$
H_\mu\left(\eta|\xi_0^k\right)\to 0
$$

as $k\to\infty$.

□

Proof of Theorem 7.19. Let η be any finite partition of X. By Lemma 7.20, for any $\varepsilon>0$, we can find k so that

$$
H_\mu\left(\eta|\xi_0^k\right)<\varepsilon.
$$

Then by Proposition 7.14 (3) and Proposition 7.15 (1),

$$
h_\mu(T,\eta)\leqslant h_\mu\left(T,\xi_0^k\right)+H_\mu\left(\eta|\xi_0^k\right)=h_\mu(T,\xi)+H_\mu\left(\eta|\xi_0^k\right)<h_\mu(T,\xi)+\varepsilon.
$$

Because ε is arbitrary, we have

$$
h_\mu(T,\eta)\leqslant h_\mu(T,\xi).
$$

Because finite partition η is arbitrary, we have

$$
h_\mu(T)=\sup_{\eta\text{ is finite}}h_\mu(T,\eta)\leqslant\sup_{\eta\text{ is finite}}h_\mu(T,\xi)=h_\mu(T,\xi).
$$

Conversely, by definition,

$$
h_\mu(T,\xi)\leqslant\sup_{\eta\text{ is finite}}h_\mu(T,\eta)=h_\mu(T).
$$

Above all, we have

$$
h_\mu(T)=h_\mu(T,\xi)
$$

which is as desired. □

Corollary 7.21. *If T is invertible and there exists a one-side generator ξ, then $h_\mu(T)=0$.*

Proof. By Kolmogoroff-Sinai Theorem (Theorem 7.19), future formula (Proposition 7.13), T-invariance of entropy (Lemma 7.9) and Lemma 7.20,

$$
h_\mu(T)=h_\mu(T,\xi)=\lim_{n\to\infty}H_\mu\left(\xi|\xi_1^{n-1}\right)=\lim_{n\to\infty}H_\mu\left(T\xi|T\xi_1^{n-1}\right)=\lim_{n\to\infty}H_\mu\left(T\xi|\xi_0^{n-2}\right)=0.
$$

□

Example 7.22. One can use Corollary 7.21 to prove that $h_\lambda(R_\alpha)=0$ where $R_\alpha\colon x\in\mathbb{T}\mapsto x+\alpha\in\mathbb{T}$ and λ is the Lebesgue measure on $\mathbb{T}$.

Example 7.23. Consider $T_p\colon x \in \mathbb{T} \mapsto px \in \mathbb{T}$. Let

$$\xi = \left\{\left[0, \frac{1}{p}\right), \left[\frac{1}{p}, \frac{2}{p}\right), \cdots, \left[1-\frac{1}{p}, 1\right)\right\}$$

be a finite partition of $\mathbb{T}$. Then ξ_0^{n-1} is the partition defined by looking at the first n digits of elements of $[0,1)$ in the base-p-expansion, i.e.

$$\xi_0^{n-1} = \left\{\left[0, \frac{1}{p^n}\right), \left[\frac{1}{p^n}, \frac{2}{p^n}\right), \cdots, \left[1-\frac{1}{p^n}, 1\right)\right\}.$$

One can check that any open set in $\mathbb{T}$ is contained in $\sigma\left(\bigcup_{n=1}^{\infty} \xi_0^{n-1}\right)$. We see that ξ is a generator. Hence for Lebesgue measure λ on $\mathbb{T}$,

$$h_\lambda(T_p) = h_\lambda(T_p, \xi) = \lim_{n\to\infty} \frac{1}{n} H_\lambda\left(\xi_0^{n-1}\right) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{p^n-1} \left(-\frac{1}{p^n} \log\left(\frac{1}{p^n}\right)\right) = \lim_{n\to\infty} \frac{1}{n} \log(p^n) = \log p.$$

Besides, because $x\log x$ is convex, by Lemma 7.12, for any T_p-invariant measure μ on $\mathbb{T}$,

$$\begin{aligned} h_\mu(T_p) = h_\mu(T_p, \xi) &= \inf_{n\geqslant 1} \frac{1}{n} H_\mu\left(\xi_0^{n-1}\right) \\ &\leqslant \inf_{n\geqslant 1} \frac{1}{n} \sum_{i=0}^{p^n-1} \left(-\frac{1}{p^n} \log\left(\frac{1}{p^n}\right)\right) = \lim_{n\to\infty} \frac{1}{n} \log(p^n) = \log p = h_\lambda(T_p). \end{aligned}$$

This implies the Lebesgue measure λ is the "biggest" T_p-invariant measure in the sense of entropy function.

The following theorem is useful in information theory.

Theorem 7.24 (Shannon-McMillan-Breiman)**.** *Suppose $(X, \mathcal{B}, \mu, T)$ is a ergodic system on a probability space. Let ξ be a finite partition of X. Then*

$$\frac{1}{n} I_\mu\left(\xi_0^{n-1}\right) \to h_\mu(T, \xi)$$

a.e. and in L^1 as $n \to \infty$, or

$$\mu\left(\text{partition elements of } \xi_0^{n-1} \text{ containing } X\right) = \mathrm{e}^{-(h_\mu(T,\xi)+\mathrm{o}(1))n}$$

a.s..

Sketch of the Proof. Denote $f_n = I_\mu(\xi|\xi_1^n)$. Then by Proposition 7.8 (2),

$$\begin{aligned} I_\mu\left(\xi_0^{n-1}\right) &= I_\mu\left(\xi \vee T^{-1}\xi \vee \cdots \vee T^{-(n-1)}\xi\right) \\ &= I_\mu\left(\xi | T^{-1}\xi \vee \cdots \vee T^{-(n-1)}\xi\right) + I_\mu\left(T^{-1}\xi \vee \cdots \vee T^{-(n-1)}\xi\right) \\ &= f_{n-1} + I_\mu\left(T^{-1}\xi_0^{n-1}\right) = f_{n-1} + I_\mu\left(\xi_0^{n-1}\right) \circ T. \end{aligned}$$

Repeating this process, we have

$$\begin{aligned} I_\mu\left(\xi_0^{n-1}\right) &= f_{n-1} + I_\mu\left(\xi_0^{n-1}\right) \circ T \\ &= f_{n-1} + \left(f_{n-2} + I_\mu\left(\xi_0^{n-2}\right) \circ T\right) \circ T \\ &= \cdots \\ &= f_{n-1} + f_{n-2} \circ T + f_{n-3} \circ T^2 + \cdots + f_0 \circ T^{n-1}. \end{aligned}$$

Note that

$$f_n = I_\mu\left(\xi | \xi_1^{n-1}\right) = -\log\left(\mathbb{E}\left(\chi_{[x]_\xi} | \sigma\left(\xi_1^n\right)\right)\right)$$

where $\mathbb{E}(\cdot|\sigma(\xi_1^n))$ is the conditional expectation. Denote

$$f = I_\mu\left(\xi|\xi_1^\infty\right) = -\log\left(\mathbb{E}\left(\chi_{[x]_\xi} | \sigma\left(\xi_1^\infty\right)\right)\right).$$

Then

$$\frac{1}{n} I_\mu\left(\xi_0^{n-1}\right)=\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k+\frac{1}{n} \sum_{k=0}^{n-1}\left(f_{n-1-k}-f\right) \circ T^k.$$

For the first term, by Theorem 3.13,

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \rightarrow \int_X f \,\mathrm{d} \mu=H_\mu\left(\xi | \xi_1^{\infty}\right)=h_\mu(T, \xi).$$

For the second term, fix $m \geqslant 1$ and denote $F_m=\sup_{l \geqslant m}\left|f_l-f\right|$. Suppose the sequence f_n is dominated by $I^* \in L^1(X)$. Then

$$\frac{1}{n} \sum_{k=0}^{n-1}\left(f_{n-1-k}-f\right) \circ T^k=\frac{1}{n} \sum_{k=0}^{n-m-1}\left(f_{n-1-k}-f\right) \circ T^k+\frac{1}{n} \sum_{k=n-m}^{n-1}\left(f_{n-1-k}-f\right) \circ T^k$$

For the first term, note that when $0 \leqslant k \leqslant n-m+1$, $n-1-k \geqslant m$. We have

$$\frac{1}{n} \sum_{k=0}^{n-m-1}\left(f_{n-1-k}-f\right) \circ T^k \leqslant \frac{1}{n} \sum_{k=0}^{n-m-1} F_m \circ T^k.$$

Note that $F_m \rightarrow 0$ and $\frac{1}{n} \sum_{k=n-m}^{n-1} 2 I^* \circ T^k$ by Theorem 3.13. Hence we have

$$\begin{aligned}
\frac{1}{n} \sum_{k=0}^{n-1}\left(f_{n-1-k}-f\right) \circ T^k & =\frac{1}{n} \sum_{k=0}^{n-m-1}\left(f_{n-1-k}-f\right) \circ T^k+\frac{1}{n} \sum_{k=n-m}^{n-1}\left(f_{n-1-k}-f\right) \circ T^k \\
& \leqslant \frac{1}{n} \sum_{k=0}^{n-m-1} F_m \circ T^k+\frac{1}{n} \sum_{k=n-m}^{n-1} 2 I^* \circ T^k \rightarrow 0.
\end{aligned}$$

We conclude

$$\frac{1}{n} I_\mu\left(\xi_0^{n-1}\right)=\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k+\frac{1}{n} \sum_{k=0}^{n-1}\left(f_{n-1-k}-f\right) \circ T^k \rightarrow h_\mu(T, \xi)+0=h_\mu(T, \xi)$$

as $n \rightarrow \infty$, which is as desired. □

Roughly speaking, Shannon-McMillan-Breiman Theorem tells us the follwoing thing. Suppose we want to use some algorithm to encode the partition elements ξ_0^{n-1} into 0-1-sequences of finite length. Then

$$\liminf_{n \rightarrow \infty} \frac{1}{n}(\text{length of 0-1-code of the partition element containing } x) \geqslant \frac{1}{\log 2} h_\mu(T, \xi)$$

a.s..

Example 7.25. Under the same condition in Example 7.23, by Shannon-McMillan-Breiman Theorem (Theorem 7.24), one can show that any T_p-invariant ergodic probability measure μ is exact dimensional: as $\delta \rightarrow 0$, the limit

$$\frac{\log \left(\mu\left(B_\delta(x)\right)\right)}{\log \delta} \rightarrow d \geqslant 0$$

exists.

Taking $\delta=\frac{1}{p^n}$, $B_\delta(x)$ is nearly the partition elements of ξ_0^{n-1} containing x. Hence in fact, we have

$$d=\frac{h_\mu\left(T_p\right)}{\log p}.$$

A Spectral Theorem for Compact Selfadjoint Operators

Definition A.1. *Let V, W be two Banach spaces and $T\colon V \rightarrow W$ be linear. We say that T is a compact operator if*

$$T\left(B_1^V\right) \subset W$$

is compact.

Definition A.2. *Let V be a Hilbert space and $A\colon V \to V$ be linear. We say that A is selfadjoint if*

$$\langle Av_1, v_2\rangle = \langle v_1, Av_2\rangle$$

for any $v_1, v_2 \in V$.

Example A.3. Let $(X, \mathcal{B}, \mu)$ be a probability space. Let $k \in L^2_{\mu\times\mu}(X \times X)$. We define $A\colon L^2_\mu(X) \to L^2_\mu(X)$ by

$$f \mapsto (Af)(x) := \int_X k(x,y)f(y)\,\mathrm{d}\mu(y).$$

Then A is a compact operator. These are called Hilbert-Schmidt integral operators. We claim that $k(x,y) \overset{a.s.}{=} \overline{k(y,x)}$ implies A being sefladjoint:

$$\begin{aligned}\langle Af_1, f_2\rangle = \int_X Af_1(x)\overline{f_2(X)}\,\mathrm{d}\mu(x) &= \int_X\int_X k(x,y)f_1(y)\overline{f_2(x)\,\mathrm{d}(\mu\times\mu)(x,y)} \\ &= \int_X\int_X f_1(y)\overline{k(y,x)f_2(x)}\,\mathrm{d}(\mu\times\mu)(y,x) = \langle f_1, Af_2\rangle.\end{aligned}$$

Theorem A.4 (Spectral Theorem for Compact Selfadjoint Operators)**.** *Let V be a Hilbert space and let $A\colon V \to V$ be a compact selfadjoint operator. Then A is completely diagonalizable: there exists an orthonormal basis of V into eigenvectors. Moreover, for any eigenvalue $\lambda \neq 0$, the eigenspace $\{v \in V : Av = \lambda v\}$ is finite-dimensional.*

B Some Remarks about Probability Measures

Theorem B.1. *Let X be a compact metric space and let $\mathcal{M}(X)$ be the space of probability measures on X. Then $\mathcal{M}(X)$ is itself a compact metric space when equipped with the weak* topology: $\mu_n \overset{w^*}{\rightharpoonup} \mu$ as $n \to \infty$ in $\mathcal{M}(X)$ if for any $f \in C(X)$, $\int_X f\,\mathrm{d}\mu_n \to \int_X f\,\mathrm{d}\mu$ as $n \to \infty$.*

The proof combines the following results from functionnal analysis:

Theorem B.2. *Riesz representation for measures: Denote $C_{\mathbb{R}}(X)^*$ the set of signed measures on X. Any positive functional on $C_{\mathbb{R}}(X)$ is represented by a uniquely defined measure.*

Theorem B.3. *Banach-Alaoglu or Tychonoff: For a Banach space V, the dual V^* can be equipped with the weak* topology and the unit ball $\overline{B_1^{V^*}}$ is compact.*

Theorem B.4. *If V is separable, then the weak* topology restrict to $\overline{B_1^{V^*}}$ is metrizable. This applies to $C(X)$ if X is a compact metric space.*