# Ergodic Theory Lecture Notes - Jan-Mar. 2011

Lecture 1

May 1, 2011

## 1 A very condensed history of Ergodic Theory

### 1.1 Boltzmann and gas particles

The motivation for ergodic theory comes for theoretical physics and statistical mechanics. Bolzmann (1844-1906) proposed the so called *Ergodic Hypothesis* on the behaviour of particles (e.g., of gas molecules). Consider a box of unit size containing $10^{20}$ gas particles. The position of each particle will be given by three coordinates in space. The velocity of each particular will be given by a further three coordinates in space. Thus the configuration at any particular time is described by $3 \times 10^{20} + 3 \times 10^{20} = 6 \times 10^{20}$ coordinates, i.e., a point $x \in X \subset \mathbb{R}^{6.10^{20}}$. Let $Tx \in X$ be the new configuration of the gas particles it we wait for time $t = 1$.

The Botzmann ergodic hypothesis (1887) addresses the distribution of the evolution of the gas. The idea is that starting from a configuration $x$ the orbit $x, Tx, T^2x, \cdots$ spends a proportion of time in any subset $B \subset X$ proportional to the size of the set, i.e.

$$\lim_{n\to+\infty} \frac{1}{n} \text{Card}\{1 \le k \le n : T^k x \in B\} = \mu(B).$$

More generally, if we consider the continuous times then the position at time $t \in \mathbb{R}$ would be $T^t x$ and then we require that

$$\lim_{T\to+\infty} \frac{1}{T} \text{Card}\{0 \le t \le T : T^t x \in B\} = \mu(B).$$

Boltzmann also coined the term "Ergodic", based on the Greek words for "work" and "path". Tragically, Boltzmann's theories were not widely accepted and he commited suicide.

**Example 1.1** (Zermelo Paradox)**.** *Why doesn't all of the air in a lecture room move to one corner so that the audience suffocates?*

*Consider the air particles in a closed room. $X$ is the phase space (i.e., all possible positions of particles and their directions of motion). Let $m$ be the normalized volume. Let $T$ denote the motion of the particles in the phase space. If $A \subset X$ is the set of configurations where all of the particles are on one side of the room, then they will return again to this set $A$, i.e., a lot of people will suffocate.*

*However, $1cc$ of gas contains about $10^{20}$ molecules (and the times spent in this state is of order $10^{18}$ seconds, exceeding the age of the universe).*

## 1.2 Birkhoff and von Neumann

The rigorous formulation of these results came from almost simultaneously work of von Neumann (1903-1959) and Birkhoff (1884-1944). We have to restrict to measure preserving transformations $T : (X, \mu) \to (X, \mu)$. They showed that for $f \in L^\infty(X, \mu)$ we have $\frac{1}{n}\sum_{k=1}^{n} f(T^k x)$ converges both in $L^2$ and pointwise (for a.e.($\mu$)). The easier $L^2$ result was published earlier than the stronger pointwise result. There are various extensions of these results to more general groups ($\mathbb{R}^d, \mathbb{Z}^d$, Free groups, etc.) and other things than averages (e.g., subadditive sequences).

**Question 1.2** (Normal numbers)**.** *For a typical point $0 \leq x \leq$ with decimal expansion $x = \cdot a_1 a_2 a_3 \cdots$ (where $a_1, a_2, a_3, \cdots, \in \{0, 1, \cdots, 9\}$) what is the frequency that a fixed digit 7, say, occurs?*

The answer is that for almost all points (i.e., except on a set of zero Lebesgue measure) the frequency with which a certain digit appears in a base $b$ expansion is $\frac{1}{b}$.

**Example 1.3.** *Let $b \geq 2$ be a natural number, then consider the map $T : [0, 1) \to [0, 1)$ given by $Tx = bx \pmod 1$.*

*In particular, except for a set of measure zero in $x$ the Birkhoff Theorem shows that $\frac{1}{N}\sum_{k=0}^{N-1} \chi(T^k x) \to \frac{1}{b}$ where $\chi$ is the indicator function of $[\frac{i}{b}, \frac{i+1}{b})$. This proves almost all number are normal.*

## 1.3 Mixing and spectral properties

Another important ingredient in the area is the classification of ergodic maps. From the 1930s to the 1950s the main approach to the classification was via spectral results. One considers the operators $U_T : L^2(X, \mu) \to L^2(X, \mu)$ defined by $U_T f(x) = f(Tx)$ which are unitary operators on the Hilbert space. From these operators one can derive invariants from the spectral properties of the operators (e.g., the spectrum of the operator, or the mixing properties).

**Example 1.4** (Heuristic picture)**.** *One can think of an incompressible fluid stirred in a container, e.g., martini in a class. The set $B$ represents the gin and $X - B$ the vermouth. If $x \in B$ is the position of a molecule of vodka at time $t = 0$ seconds then $T^n x \in X$ can be its position at time $t = n$ seconds.*

*For the experimentalists: Usually $m(B) = \frac{1}{3}$ and $m(X - B) = \frac{2}{3}$ and you garnish with a twist of lemon.*

## 1.4 Entropy

In 1959 there was a major shift in the development of the subject with the introduction of the Entropy theory. This turned out to be one of the most important invariants $h_\mu(T)$ for the classification of invariant measures $\mu$, and was introduced by the russian mathematician Kolmogorov (1903-1987) and his student Sinai (1935- ) circa 1960. This was motivated by entropy in information theory.

Remarkably, for an important class of examples (Bernoulli measures ) Ornstein (1934- ) showed that the entropy is a complete isomorphisms (i.e., having the same entropy is equivalent to the measures being isomorphic.)

The entropy also turns out to be an important characteristic for transformations.

## 1.5 Applications

There have been quite remarkable applications of ergodic theory to number theory. These include proofs of results on arithmetic progressions (Szemerdi-Furstenberg theorem); quadratic forms (Margulis-Oppenheim); and Diophantine Appoximation (Littlewood-Einseidler-Katok-Lindenstrauss).

The classical example of an application of ergodic theory to number theorem is the following:

**Theorem 1.5** (Furstenberg-Szemerdi Theorem). *Assume that $S \subset \mathbb{Z}$ has positive density, i.e.,*

$$\liminf_{k\to\infty} \frac{Card(\{-k, \cdots, 0, \cdots, k-1, k\} \cap S)}{2k+1} > 0,$$

*then it contains arithmetrc projections of arbitrary length, i.e., for each $l \geq 1$ we can find $a, b \neq 0$ such that the arithmetic progression $\{a + bi : 0 \leq i \leq l-1\} \subset S$.*

**Example 1.6** (Trivial example). *If $S$ represents the odd numbers then the density is $\frac{1}{2}$. In this case, it is easy to see the conclusion with $a = 1$ and $b = 2$.*

Green and Tao (2004) extended this result to $S$ being the set of primes (which has zero density).

*Remark* 1.7. There is higher dimensional analogue of this theorem for $S \subset \mathbb{Z}^2$.

Another famous application of ergodic theory is the following.

**Theorem 1.8** (Oppenheim Conjecture, 1929). *Let $Q(n_1, n_2, n_3) = n_1^2 + n_2^2 - \sqrt{2}n_3^2$ (or any irrational indefinite quadratic form) then the countable set*

$$\{Q(n_1, n_2, n_3) : n_1, n_2, n_3 \in \mathbb{N}\} \subset \mathbb{R}$$

*is dense in $\mathbb{R}$.*

This was proved by Margulis in 1987 using the ergodic theory of flows in spaces of matrices.

# 2 Invariance and Recurrence

## 2.1 Invariance for transformations

Let $X$ be a set with associated sigma-algebra $\mathcal{B}$. Let $T : X \to X$ be a measurable transformation, i.e., if $A \in \mathcal{B}$ then $T^{-1}A \in \mathcal{B}$.

**Definition 2.1.** *We say that $(T, X, \mu)$ is* invariant *if for any $A \in \mathcal{B}$ we have that $\mu(T^{-1}A) = \mu(A)$*

**Example 2.2** (Rotation)**.** *Let $X = [0, 1)$. Let $\mathcal{B}$ be the Borel sigma algebra and let $\mu$ be the lebesgue measure. Given $0 \leq \alpha < 1$ and define $Tx = x + \alpha$ (mod 1), i.e.,*

$$T(x) = \begin{cases} x + \alpha & \text{if } x + \alpha < 1 \\ x + \alpha - 1 & \text{if } x + \alpha \geq 1 \end{cases}$$

*It is easy to see that Lebesgue measure is invariant. One only needs to observe that for any subinterval $[a, b) \subset [0, 1)$ the preimage $T^{-1}[a, b)$ is a subinterval (or union of two intervals) of the same length. It is easy to deduce from this that if $A$ is a disjoint union of intervals then $T^{-1}A$ is a disjoint union of intervals with the same Lebesgue measure. Finally, we can use the Kolmogorov extension principle to deduce that the measures $\mu(A)$ and $\mu(T^{-1}A)$ coincide for any $A \in \mathcal{B}$.*

**Lemma 2.3.** *For the rotation $T : [0, 1) \to [0, 1)$ given by $Tx = x + \alpha$ (mod 1):*

1. *If $\alpha$ is irrational then Lebesgue measure $\mu$ is the only invariant probability measure*
2. *If $\alpha$ is rational then there are more invariant probability measures*

*Proof.* Assume that $\alpha$ is irrational and $\mu$ is an invariant probability measure. For any $\alpha$ choose $n_1 > n_2$ such that the fractional parts $0 < \{n_1\alpha\}, \{n_2\alpha\} < 1$ satisfy $|\beta := \{n_1\alpha\} - \{n_2\alpha\}| < \epsilon$. In particular, $\mu$ is preserved by translation $T_\beta : [0, 1) \to [0, 1)$ by $\beta$, and thus $m\beta$, for any $m \in \mathbb{Z}$), i.e., a dense set of values. Finally (using the dominated convergence principle) we can show that $\mu$ is preserved by $T_\beta$ for any $0 < \beta < 1$, and so $\mu$ must be Lebesgue measure.

Assume that $\alpha = p/q$ is rational. Then $\mu = \frac{1}{q}\sum_{i=0}^{q-1} \delta_{i/q}$ is an other $T_\alpha$-invariant measure. □

**Definition 2.4.** *If $T : X \to X$ is a homeomorphism of a compact metric space then we say that it is uniquely ergodic when there is precisely one $T$-invariant probability measure.*

**Example 2.5** (Leading digits)**.** *The unique ergodicity can be used to show that the frequency of the leading digits of $2^n$ being equal to $l \in \{1, \cdots, 9\}$ is $\log(1 + \frac{1}{l})$.*

**Example 2.6** ($b$-transformation)**.** *Let $X = [0, 1)$. Let $\mathcal{B}$ be the Borel sigma algebra, and let $\mu$ be the lebesgue measure. Let $b \geq 2$ be a natural number and let $Tx = bx$ (mod 1), i.e.,*

$$T(x) = \begin{cases} bx & \text{if } 0 \leq x < \frac{1}{b} \\ bx & \text{if } \frac{1}{b} \leq x < \frac{2}{b} \\ & \vdots \\ bx & \text{if } \frac{b-1}{b} \leq x < 1 \end{cases}$$

*When $b = 2$ this is called the Doubling map.*

*It is easy to see that Lebesgue measure is invariant. One only needs to observe that for any subinterval* $[a, b) \subset [0, 1)$ *the preimage*

$$T^{-1}[c,d) = \left[\frac{c}{b}, \frac{d}{b}\right) \cup \left[\frac{c+1}{b}, \frac{d+1}{b}\right) \cup \cdots \cup \left[\frac{c+b-1}{b}, \frac{d+b-1}{b}\right)$$

*is a union of two intervals of half the length. It is easy to deduce from this that if* $A$ *is a disjoint union of intervals then* $T^{-1}A$ *is a disjoint union of intervals with the same Lebesgue measure. Finally, we can use the Kolmogorov extension principle to deduce that the measures* $\mu(A)$ *and* $\mu(T^{-1}A)$ *coincide for any* $A \in \mathcal{B}$.

*The Dirac measure* $\delta_0$ *is an example on another* $T$*-invariant measure.*

*Remark* 2.7. If $b$ is replaced by a real number $\beta > 1$ then there is an invariant measure which is absolutely continuous with respect to Lebesgue measure.

**Example 2.8** (Gauss transformation)**.** *Let* $X = [0, 1)$*. Let* $\mathcal{B}$ *be the Borel sigma algebra and l et* $\mu$ *be the lebesgue measure Let* $Tx = \frac{1}{x}$ *(mod 1).*

**Definition 2.9.** *We define the* Gauss measure $\mu$ *on* $[0, 1)$ *defined by*

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{dx}{1+x}, \text{ for } A \in \mathcal{B}.$$

**Lemma 2.10.** *The Gauss measure is preserved by the Gauss transformation*

(Proof later)

**Example 2.11** (Bernoulli measures)**.** *Let* $\Sigma = \{1, 2, \cdots, k\}^{\mathbb{Z}}$ *and let* $T : \Sigma \to \Sigma$ *be the full shift on two symbols given by* $T(x_n) = (x_{n+1})$*. We can consider the (cylinder) sets*

$$[i_{-l}, \cdots, i_0, \cdots, i_m] = \{x = (x_n) : x_j = i_j \text{ for } j = -l, \cdots, m\}$$

*where* $i_{-l}, \cdots, i_0, \cdots, i_m \in \{1, 2, \cdots, k\}$*. These form a sub-basis for a topology on* $\Sigma$ *and we let* $\mathcal{B}$ *be the associated sigma algebra.*

*Let* $(p_1, p_2, \cdots, p_k)$ *be a probability vector, i.e.,* $p_1 + \cdots + p_k = 1$ *(e.g.,* $p_1 = \cdots = p_k = \frac{1}{k}$*). We can define a Bernoulli measure* $\mu$ *on* $\Sigma$ *by letting* $\mu[i_{-l}, \cdots, i_0, \cdots, i_m] = p_{-i_l} \cdots p_{i_0} \cdots p_{i_m}$ *and then using the Kolmogorov extension theorem to extend this to* $\Sigma$*. This measure is easily seen to be invariant since*

$$\begin{aligned} \mu([i_{-l}, \cdots, i_0, \cdots, i_m]) &= p_{i_{-l}} \cdots p_{i_0} \cdots p_{i_k} \\ &= \mu(T^{-1}[i_{-l}, \cdots, i_0, \cdots, i_k]) \end{aligned}$$

**Example 2.12.** *Consider a polygon* $P$ *in the plane. Let* $X$ *denote the position (and velocity) of a particle which moves in a straight line in the interior the polygon, and bounces off the sides. Let* $T$ *denote the change after one second.* $T$ *preserves the natural measure* $m =$ *"area* $\times$ *direction". For generic polygons,* $m$ *is ergodic.*

*(N.B. If one puts obstacles inside the polygon this gives a "Sinai or dispersive billiard". If one makes the sides of the polygon curved, this gives a "Bunimovitch or Stadium type billiard". Both involve a lot of technical analysis).*

**Lemma 2.13.** *A probability measure $\mu$ is $T$-invariant iff $\int f \circ T d\mu = \int f d\mu$ for all $f \in L^1(X, \mu)$.*

*Proof.* Assume $\int f \circ T d\mu = \int f d\mu$ for all $f \in L^1(X, \mu)$. For any $A \in \mathcal{B}$ we can take $f = \chi_A$ to deduce that $\mu(A) = \mu(T^{-1}A)$, for $A \in \mathcal{B}$, i.e., $\mu$ is $T$-invariant.

Conversely, if $\mu$ is $T$-invariant then we can deduce that $\int f \circ T d\mu = \int f d\mu$ for simple functions and then extend the result to $f \in L^1(X, \mu)$ by approximation. □

**Corollary 2.14.** *If $T : X \to X$ is a continuous map on a compact metric space. A probability measure $\mu$ is $T$-invariant iff $\int g \circ T d\mu = \int g d\mu$ for all $g \in C(X)$.*

*Proof.* By the previous lemma, $\mu$ is $T$-invariant iff $\int f \circ T d\mu = \int f d\mu$ for all $f \in L^1(X, \mu)$ which implies $\int g \circ T d\mu = \int g d\mu$ for all $g \in C(X)$.

For the converse, assume there exists $f \in L^1(X, \mu)$ and $\epsilon > 0$ such that $|\int f \circ T d\mu - \int f d\mu| > \epsilon$. However, since we can find $g \in C(X)$ with $\|g - f\|_1 < \frac{\epsilon}{2}$, $\|g \circ T - f \circ T\|_1 < \frac{\epsilon}{2}$ then by the triangle inequality $\int g \circ T d\mu \neq \int g d\mu$. □

## 2.2 The existence of invariant measures

Let $X$ be a compact metric space and let $T : X \to X$ be a continuous map. Let $\mathcal{B}$ be the Borel sigma algebra (i.e., the smallest sigma algebra containing the topology on $X$).

**Definition 2.15.** *Let $\mathcal{M}$ be the space of probability measures on $X$. We can consider the weak-star topology on $\mathcal{M}$, corresponding to a sequence $(\mu_n) \in \mathcal{M}$ converging to $\mu \in \mathcal{M}$ (i.e., $\mu_n \to \mu$) precisely when $\int f d\mu_n \to \int f \mu$ for every $f \in C(X)$.*

The great thing about the weak-star topology is that $\mathcal{M}$ is compact:

**Theorem 2.16** (Alaoglu). *For any sequence $(\mu_n) \subset \mathcal{M}$ there exists a $\mu \in \mathcal{M}$ and subsequence $(\mu_{n_k})$ converging to $\mu$.*

**Theorem 2.17** (KrylovBogolyubov). *There exists at least one $T$-invariant probability measure $\mu$.*

*Proof.* Fix a point $x_1 \in X$. For each $n \geq 1$, we define a new probability measure $\mu_n$ by

$$\int f \mu_n = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$$

where $f \in C(X, \mathbb{R})$ is a continuous function. By Alaoglu's theorem we can find a $\mu$ and subsequence $(\mu_{n_k})$ converging to $\mu$.

It remains to show that $\mu$ is $T$-invariant. However, for any $f \in C(X)$ we see that

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

and the result follows. □

There is a direct proof that doesn't require weak star compactness.

*Proof.* Fix $x \in X$. Let $\{f_m\}_{m=0}^{\infty}$ be a countable dense subset of $C(X,\mathbb{R})$. For each $m \geq 1$ the set

$$\left\{ \frac{1}{n} \sum_{k=0}^{n-1} f_m(T^k x) \right\}_{n=0}^{\infty}$$

is bounded (by $\|f_m\|_\infty$). Choose convergence subsequences $(n_i^{(1)})_{i=1}^{\infty} \supset (n_i^{(2)})_{i=1}^{\infty} \supset \cdots \supset (n_i^{(m)})_{i=1}^{\infty} \supset \cdots$. Then for each $m \geq 1$, we can use the diagonal subsequence to write

$$L(f_m) := \lim_{k \to +\infty} \frac{1}{n_k^{(k)}} \sum_{l=0}^{n_k^{(k)}-1} f_m(T^l x).$$

This defines a linear function $L : C(X,\mathbb{R}) \to \mathbb{R}$. In particular, for $\epsilon > 0$ and $f \in C(X,\mathbb{R})$ choose $\|f - f_m\| < \epsilon$. Thus

$$\frac{1}{n_k^{(k)}} \sum_{l=0}^{n_k^{(k)}-1} f(T^l x) = \underbrace{\frac{1}{n_k^{(k)}} \sum_{l=0}^{n_k^{(k)}-1} f_m(T^l x)}_{\to L(f_k)} + \underbrace{\frac{1}{n_k^{(k)}} \sum_{l=0}^{n_k^{(k)}-1} f(T^l x) - f_k(T^l x)}_{|\cdot| < \epsilon}.$$

Since $\epsilon > 0$ is arbitrary we deduce that the Left Hand Side converges to $L(f)$, say.

Moreover, it is easy to see that

1. $f \geq 0 \implies L(f) \geq 0$ (positivity);
2. $L : C(X,\mathbb{R}) \to \mathbb{R}$ is linear;
3. $L(f \circ T) = L(f)$, $\forall f \in C(X,\mathbb{R})$, since

$$L(f \circ T) = \lim_{k \to +\infty} \frac{1}{n_k^{(k)}} \sum_{l=1}^{n_k^{(k)}} f(T^l x) = \lim_{k \to +\infty} \frac{1}{n_k^{(k)}} \sum_{l=0}^{n_k^{(k)}-1} f_m(T^l x) = L(f).$$

We can then define a $T$-invariant probability measure by $\int \phi d\mu := L(\phi)$ (by the Riesz Representation Theorem).

□

A useful device is often the following.

*Remark* 2.18 (Natural extensions). Assume that $T : X \to X$ is not necessarily invertible. We can associate a new space by defining

$$\widehat{X} = \{(x_n)_{-\infty}^{0} : Tx_n = x_{n+1} \text{ for } n \leq -1\}$$

and then the sigma algebra generated by sets $[x_{-n}, \cdots, x_0] = \{(x_n)_{-\infty}^{0} \in \widehat{X} : x_i \in B_i, -n \leq i \leq 0\}$ where $B_i \in \mathcal{B}$ is the sigma algebra for $X$. We then define a map $\widehat{T} : \widehat{X} \to \widehat{X}$ defined by $\widehat{T}(x_n)_{-\infty}^{0} = (Tx_n)_{-\infty}^{0}$. We can also uniquely extend the measure $\mu$ to $\widehat{\mu}$ on $\widehat{X}$. Finally, the map $\widehat{T}$ is invertible with $\widehat{T}(x_n)_{-\infty}^{0} = (x_{n-1})_{-\infty}^{0}$.

## 2.3 Poincaré Recurrence

The following simple result is extremely useful (and dates back to 1899).

**Theorem 2.19** (Poincaré Recurrence)**.** *Let $\mu$ be a $T$-invariant probability on $X$. For any set $B \in \mathcal{B}$ with $\mu(B) > 0$, almost every $x \in B$ returns to $B$. In fact, almost every point returns infinitely often, i.e., for a.e. $(\mu)$ we can choose an increasing subsequence $\{n_k\}_{k=1}^{\infty}$ of natural numbers such that $T^{n_k}(x) \in B$.*

*Proof.* The set of points $x \in B$ which return to $B$ infinitely often can be written as

$$B_\infty := B \cap \left( \cap_{N=0}^{\infty} \underbrace{\cup_{n=N}^{\infty} T^{-n} B}_{=:B_N} \right)$$

Since $T^{-1}B_N = B_{N+1}$ we have by $T$-invariance that $\mu(B_N) = \mu(B_{N+1})$. Moreover, since $B_0 \supset B_1 \supset B_2 \supset \cdots$ we see that $\mu(B_0) = \mu(\cap_{N=0}^{\infty} B_N)$.

Finally $\mu(B_\infty) = \mu(B \cap B_0) = \mu(B)$, where the last equality comes from $B \subset B_0 := \cup_{n=0}^{\infty} T^{-n} B$.

□

*Remark* 2.20. This is statement can (essentially) be found in the bible (Ecclestiastes 1.7): *All the rivers run into the sea. Yet the sea is not full. To the place from which the rivers come. There they return again.*

**Example 2.21** (Zermelo Paradox)**.** *Why doesn't all of the air in a lecture room move to one corner so that the audience suffocates?*

*Consider the air particles in a closed room. $X$ is the phase space (i.e., all possible positions of particles and their directions of motion). Let $m$ be the normalized volume. Let $T$ denote the motion of the particles in the phase space.*

*If $A \subset X$ is the set of configurations where all of the particles are on one side of the room, then they will return again to this set $A$, i.e., a lot of people will suffocate.*

*However, 1cc of gas contains about $10^{20}$ molecules (and the return times is of order $10^{18}$ seconds, exceeding the age of the universe).*

## 2.4 Multiple recurrence

A much more difficult result to prove is a $\mathbb{Z}^d$-version.

**Theorem 2.22** (Poincaré-Furstenberg Multiple Recurrence)**.** *Let $T_1, \cdots, T_d : X \to X$ be commuting transformations, i.e., $T_i \circ T_j = T_j \circ T_i$ for $1 \leq i, j \leq d$ . We can choose $M > 0$ such that $\mu(B \cap T_1^{-M} B \cap \cdots \cap T_d^{-M} B) > 0$.*

The most famous application of this is the following.

**Example 2.23** (Furstenberg-Szemerdi)**.** *Let $\mathcal{N} \subset \mathbb{Z}$ be a subset of positive density, i.e.,*

$$d := \limsup_{k \to +\infty} \frac{1}{2k+1} Card\{-k \leq n \leq k : n \in \mathcal{N}\}.$$

*Then for any $d > 0$ the set $\mathcal{N}$ contains an arithmetic progression of length $d$, i.e., there exists $N \in \mathbb{Z}$ and $M > 0$ such that the arithmetic progression $N, N+M, N+2M, \cdots, N+dM \in \mathcal{N}$.*

*We briefly show how to deduce this result from the Multiple Recurrence Theorem. Let $\Sigma = \{0,1\}^{\mathbb{Z}}$ be the space of all possible sequences of 0s and 1s. We can consider the point $x = (x_n)_{n=-\infty}^{\infty} \in \Sigma$ defined by*

$$x = (x_n)_{n=-\infty}^{\infty} \in \Sigma \text{ given by } x_n = 0 \text{ iff } n \in \mathcal{N}.$$

*Let $T : \Sigma \to \Sigma$ be the shift map defined by $T(x_n)_{n=-\infty}^{\infty} = (x_{n+1})_{n=-\infty}^{\infty}$.*

*As in the proof of the existence of invariant measures, define for each $n \geq 1$ a family of measures $(\mu_n)$ defined by*

$$\mu_n = \frac{1}{2n+1} \sum_{k=-n}^{n} \delta_{T^k x} \in \mathcal{M}.$$

*By compactness of the space of probability measures $\mathcal{M}$ on $\Sigma$ there is at least one weak-star limit point $\mu$ (i.e., there exists $n_k \to +\infty$ such that $\mu_{n_k} \to \mu$). Moreover, the same argument as in the existence of invariant measures shows that $\mu$ is $T$-invariant and, assuming we took the subsequence through the sequence corresponding to the limit supremeum $d$ we have that $B = \{y \in \Sigma : y_0 = 0\}$ satisfies $\mu(B) > 0$.*

*We can apply Poincaré-Furstenberg Multiple Recurrence to the transformations*

$$T_1 = T, T_2 = T^2, \cdots, T_d = T^d$$

*set $B = \{y \in \Sigma : y_0 = 0\}$ to deduce that there exists some $M > 0$ such that*

$$\mu(T^{-M} B \cap T^{-2M} \cap \cdots \cap T^{-dM} B) > 0$$

*Moreover, since $T^{-M} B \cap T^{-2M} \cap \cdots \cap T^{-dM} B$ is an open (and closed) set and $\mu_{n_k} \to \mu$ (in the weak-star topology) we have that for $k$ sufficiently large:*

$$\mu_{n_k}(T^{-M} B \cap T^{-2M} \cap \cdots \cap T^{-dM} B) > 0.$$

*In particular, there exists $0 \leq N \leq n_k - 1$ such that $T^l x \in T^{-M} B \cap T^{-2M} \cap \cdots \cap T^{-dM} B$. Equivalently, $N + m, N + 2M, \cdots, N + dM \in \mathcal{N}$ are an arithmetic progression.*

## 2.5 Invariant measures and flows

Let $T^t : X \to X$ be a flow, i.e., $T^{s+t} = T^s \circ T^t$ and $T^0 = I$.

**Definition 2.24.** *We say that a probability measure $\mu$ is $T^t$-invariant if $\mu(T^t B) = \mu(B)$, for all $B \in \mathcal{B}$, $t \in \mathbb{R}$.*

**Example 2.25.** *We can define a flow on $X = \mathbb{R}/\mathbb{Z}$ by $T^t x = x + t$ (mod 1). This preserves the usual Lebesgue measure.*

**Example 2.26.** *Let $\alpha_1, \cdots, \alpha_d \in \mathbb{R}$. We can define a flow on $X = \mathbb{R}^d/\mathbb{Z}^d$ by $T^t(x_1, \cdots, x_d) = (x_1 + t\alpha_1, \cdots, x_d + t\alpha_1)$ (mod 1). This preserves the usual Lebesgue measure.*

Let $G = SL(2,\mathbb{R})$ then we can let $\nu$ be the Haar measure, i.e., we can define a measure $\nu$ on $G$ by

$$d\nu \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \frac{d\delta d\beta d\gamma}{\delta}$$

i.e., $\nu(B) = \int_B \frac{1}{\delta} d\alpha d\beta d\gamma$ for any $B \in \mathcal{B}$. However, this measure isn't finite (i.e., $\nu(G) = \infty$).

Let $\Gamma = SL(2,\mathbb{Z}) < SL(2,\mathbb{R})$ be the subgroup with entries in $\mathbb{Z}$ and consider the quotient space $X = G/\Gamma$.

**Lemma 2.27.** *The volume of the quotient space is finite, i.e., $\nu(X) < +\infty$. In fact, one can show that $\nu(X) = \frac{2}{3}\pi^2$.*

We can therefore rescale the measure $\nu$ to assume that it is is a probability measure.

**Example 2.28** (Geodesic flow)**.** *We can define a flow $T^t : X \to X$ by $T^t g\Gamma = g_t g\Gamma$ where*

$$g_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}$$

**Lemma 2.29.** *The measure $\nu$ is $T^t$-invariant.*

*Proof.* This follows from the fact that

$$g_t \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \Gamma = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \Gamma = \begin{pmatrix} e^{t/2}\alpha & e^{t/2}\beta \\ e^{-t/2}\gamma & e^{-t/2}\delta \end{pmatrix} \Gamma.$$

In particular, we see that

$$DT^t \left( \frac{d\delta d\beta d\gamma}{\delta} \right) = \frac{(e^{-t/2}d\delta)(e^{t/2}d\beta)(e^{-t/2}d\gamma)}{e^{-t/2}\delta} = \frac{d\delta d\beta d\gamma}{\delta}$$

which implies that the flow preserves the measure. □

**Example 2.30** (Horocycle flow)**.** *Let $G = SL(2,\mathbb{R})$ and $\Gamma = SL(2,\mathbb{Z})$ and let $X = G/\Gamma$. We can define a flow on $X$ by $T^t g\Gamma = h_t g\Gamma$ where*

$$h_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

**Lemma 2.31.** *$\nu$ is $T^t$ :-invariant.*

*Proof.* This follows from the fact that

$$h_t \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \alpha + t\gamma & \beta + t\delta \\ \gamma & \delta \end{pmatrix}.$$

In particular, we see that

$$DT^t \left( \frac{d\delta d\beta d\gamma}{\delta} \right) = \frac{d\delta d\beta d\gamma}{\delta}$$

which implies that the flow preserves the measure. □

*Remark* 2.32. We can define a second version of the horocycle flow on $X = G/\Gamma$. We can define a flow on $X$ by $T^t g\Gamma = h_t^- g\Gamma$ where

$$h_t^- = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}.$$

The measure $\nu$ is $T^t$-invariant.

# 3 Ergodicity

## 3.1 Ergodicity for transformations

*Let $X$ be a space with associated sigma-algebra $\mathcal{B}$. Let $\mu$ be $T$-invariant probability measure.*

**Definition 3.1.** *We say that $(T, X, \mu)$ is* ergodic *if whenever $T^{-1}A = A \in \mathcal{B}$ then $\mu(A) = 0$ or $1$.*

*Remark* 3.2. As usual, we understand the equality of sets as being "up to a set of zero measure".

*The following characterization can be useful.*

**Lemma 3.3.** *We have the following.*

1. *$\mu$ is ergodic iff $f \circ T = f \in L^2(X, \mu)$ implies $f$ is a constant.*
2. *$\mu$ is ergodic iff $f \circ T \leq f$ implies $f$ is a constant.*
3. *$\mu$ is ergodic iff $f \circ T \geq f$ implies $f$ is a constant.*

*Proof.* We first observe that the function $f$ is non-constant iff we can choose $c \in \mathbb{R}$ such that $A = \{x : f(x) > c\}$ has $0 < \mu(A) < 1$.

For part 1, if $\mu$ is ergodic then $f \circ T = f$ implies that for any such $A$ we have $T^{-1}A = A$ and so by ergodicity $\mu(A) = 0$ or 1. In particular, $f$ is constant. On the other hand, if $\mu$ is not ergodic then we can choose $0 < \mu(B) < 1$ with $T^{-1}B = B$. Therefore, $\chi_B \circ T = \chi_B$ and we can choose the non-constant function $f = \chi_B$.

For part 2, if we know that "$f \circ T \leq f$ implies $f$ is a constant" then we have that "$f \circ T = f$ implies $f$ is a constant" as a special case and part 1 gives ergodicity. Conversely, assume that $\mu$ is ergodic. If $f \circ T \leq f$ then given any set of the form $A = \{x : f(x) < c\}$ we see that any $x \in T^{-1}A$ satisfies $f(Tx) \leq f(x) < c$ and thus $x \in A$. In particular, $T^{-1}A \subset A$. However, since $\mu$ is $T$-invariant this implies that $T^{-1}A = A$, and by ergodicity $\mu(A) = 0$ or 1. In particular, $f$ is constant.

The proof of the third part is similar to the second (Exercise). □

*Remark* 3.4 (Fourier Series). It is a useful fact that any (periodic) function $f \in L^2([0,1], \mu)$ can be uniquely written in the form

$$f(x) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x}, a_n \in \mathbb{C} \text{ with } \sum_{n \in \mathbb{Z}} |a_n|^2 = \|f\|_2^2 < +\infty.$$

**Example 3.5.** *Let $X = [0,1]$. Let $\mathcal{B}$ be the Borel sigma algebra, (i.e., the collection of sets which are the result of countable unions and intersections of subintervals). Let $\mu$ be the usual lebesgue measure.*

1. *Given $0 \leq \alpha < 1$ let $T(x) = x + \alpha$ (mod 1). This is ergodic if and only if $\alpha$ is irrational.*

   *(a) If $\alpha = \frac{p}{q}$ is rational then $\mu$ is not ergodic. For example, $f(x) = \sin(2\pi q x)$ is non-constant but satisfies $f \circ T = f$.*

*(b) If $\alpha$ is irrational then $\mu$ is ergodic. Any $f \in L^2([0,1],\mu)$ can be written uniquely as*

$$f(x) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n x}, a_n \in \mathbb{C}.$$

*Thus,*

$$f(Tx) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n (x+\alpha)},$$

*and if $f \circ T = f$ comparing coefficients gives $a_n = a_n e^{2\pi i n\alpha}$, for all $n \in \mathbb{Z}$. Thus $a_n = 0$ if $n \neq 0$, i.e., $f$ is constant.*

2. *Let $b \geq 2$. Consider $T(x) = bx$ (mod 1) defined by*

$$T(x) = \begin{cases} bx & \text{if } 0 < x < \frac{1}{b} \\ bx - 1 & \text{if } \frac{1}{b} \leq x < \frac{2}{b} \\ & \vdots \\ bx - (b-1) & \text{if } \frac{b-1}{b} \leq x < 1 \end{cases}$$

*The Lebesgue measure $\mu$ is ergodic. Any $f \in L^2([0,1],\mu)$ can be written uniquely as*

$$f(x) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n x}, a_n \in \mathbb{C}.$$

*where $\sum_{n\in\mathbb{Z}} |a_n|^2 = \|f\|_2^2 < +\infty$. Thus, if $f \circ T = f$ then*

$$\sum_{n\in\mathbb{Z}} a_n e^{2\pi i 2nx} = f(Tx) = f(x) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n x},$$

*and comparing coefficients gives $a_n = a_{2n}$. However, since $\sum_{n\in\mathbb{Z}} |a_n|^2 < +\infty$ we deduce that $a_n = 0$ if $n \neq 0$. In particular, $f$ is constant.*

3. *Let $T(x) = \frac{1}{x}$ (mod 1) then the Gauss measure $\mu$ is ergodic. (Proof later.)*

4. *A Bernoulli measure on a shift space is ergodic. (Exercise.)*

*Remark* 3.6. One can also show this without resorting to Fourier series Let $D_{k/2^n} = [\frac{k}{2^n}, \frac{k+1}{2^n}$ where $k \in \{0, 1, \cdots, 2^n - 1\}$ If $E \in \mathcal{B}$ then it is not hard to check that $\mu(T^{-n}E \cap D_{k/2^n}) = 2^{-n}\mu(E)$

## 3.2 Existence of Ergodic Measures

*Let $X$ be a compact metric space and let $T : X \to X$ be a continuous transformation. Let $\mathcal{B}$ be the Borel sigma algebra (i.e., the smallest sigma algebra containing the topology on $X$).*

*Let $\mathcal{M}$ be the space of probability measures on $X$. Let $\mathcal{M}_T$ be the space of $T$-invariant probability measures on $X$.*

**Lemma 3.7.** *There are two nice properties of this space:*

1. *$\mathcal{M}_T$ is compact; and*

2. *$\mathcal{M}_T$ is convex (i.e., If $0 \leq \lambda \leq 1$ and $\mu_0, \mu_1 \in \mathcal{M}_T$ then $\mu_\lambda = (1-\lambda)\mu_0 + \lambda\mu_1 \in \mathcal{M}_T$).*

*Proof.* 1. By Alaoglu's theorem, $\mathcal{M}$ is weak-star compact. Moreover, $\mathcal{M}_T = \cap_{f \in C(X)} \{\mu \in \mathcal{M} : \int (f \circ T - f) d\mu = 0\}$ is weak star closed, and thus compact.

2. For $0 \leq \lambda \leq 1$ we observe that for any $f \in C(X)$,

$$\int f \circ T d\mu_\lambda = (1-\lambda) \int f \circ T d\mu_0 + \lambda \int f \circ T d\mu_1 = (1-\lambda) \int f d\mu_0 + \lambda \int f d\mu_1 = \int f d\mu_\lambda.$$

In particular, $\mu_\lambda \in \mathcal{M}$. □

**Definition 3.8.** *We call $\mu \in \mathcal{M}_T$ extremal if whenever we can write $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$, where $\mu_1, \mu_2 \in \mathcal{M}_T$ and $0 \leq \lambda \leq 1$ then necessarily $\mu = \mu_1$ or $\mu = \mu_2$.*

*The next standard result related general $T$-invariant measures to ergodic ones.*

**Theorem 3.9** (Ergodic Decomposition)**.** *We have the following results:*

1. *A $T$-invariant probability measure $\mu$ is ergodic iff $\mu$ is extremal in $\mathcal{M}_T$;*

2. *Given any $T$-invariant probability measure $\mu$, there exists a measure $\Lambda_\mu$ on the extremal (ergodic) measures in $\mathcal{M}_T$ such that we can write*

$$\int f d\mu = \int_{Ext(\mathcal{M}_T)} \left( \int f d\nu \right) d\Lambda_\mu(\nu), \text{ for all } f \in C(X).$$

*Proof.* For part 1, assume that $\mu$ is not ergodic. Then we can find an invariant set $T^{-1}A = A$ with $0 < \mu(A), \mu(A^c) < 1$. We can then define

$$\mu_1(B) = \frac{\mu(A \cap B)}{\mu(A)} \text{ and } \mu_2(B) = \frac{\mu(A^c \cap B)}{\mu(A^c)}$$

for any $B \in \mathcal{B}$. Then with $\lambda = \mu(A)$ we can write

$$\mu = \lambda\mu_1 + (1-\lambda)\mu_2$$

showing that $\mu$ isn't extremal. Conversely, assume that $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$ with $\mu_1 \neq \mu_2$ and $0 < \lambda < 1$. In particular, $\mu \ll \mu_1, \mu_2$. Moreover, the Radon-Nikodyn derivatives $\frac{d\mu_1}{d\mu}, \frac{d\mu_1}{d\mu} \in L^1(X, \mu)$ are invariant, i.e., $\frac{d\mu_1}{d\mu} = \frac{d\mu_1}{d\mu} \circ T$ and $\frac{d\mu_2}{d\mu} = \frac{d\mu_2}{d\mu} \circ T$. If we assume for a contradiction that $\mu$ is ergodic, then we deduce that $\frac{d\mu_1}{d\mu}, \frac{d\mu_1}{d\mu} = 1$ and thus $\mu_1 = \mu_2 2 = \mu$.

The second part is an application of a classical result in functional analysis (the Choquet's theorem). □

*We can compare this with a more familiar finite dimensional version of Choquet's theorem.*

*Remark* 3.10. Finite dimensional polytope version Let $V$ be a convex polytope with vertices $v_1, \ldots, v_{n+1} \in \mathbb{R}^n$. Then for every point $x \in V$ can be written as a combination of Choquet's theorem vertices, i.e., $\exists 0 \leq \lambda_1, \ldots, \lambda_{n+1} \leq 1$ with $\sum_{i=1}^{n+1} \lambda_i = 1$ and $x = \sum_{i=1}^{n+1} \lambda_i v_i$.

**Theorem 3.11.** *Let $T : X \to X$ be a continuous map on a compact metric space. There exists at least one ergodic measure.*

*Proof.* Let $\{f_n\}_{n=0}^{\infty}$ be a dense subset of $C(X, \mathbb{R})$. Let $\mathcal{M} = \mathcal{M}_0$ be the $T$-invariant probability measures and define inductively

$$\mathcal{M}_0 \supset \mathcal{M}_1 \supset \mathcal{M}_2 \supset \cdots$$

by

$$\mathcal{M}_{i+1} = \{\mu \in \mathcal{M}_i : \int f_{i+1} d\mu = \max_{\nu \in \mathcal{M}_i} \int f_{i+1} d\nu\}.$$

Since $\mathcal{M}_i \ni \nu \to \int f_{i+1} d\nu$ is weak star continuous we can deduce that $\mathcal{M}_{i+1}$ is compact, non-empty and convex. Thus by compactness (by Alaoglu's theorem) $\mathcal{M}_\infty := \cap_{n=0}^{\infty} \mathcal{M}_n \neq \emptyset$

We claim that $\mathcal{M}_\infty$ consists only of extreme points. Let $\mu \in \mathcal{M}_\infty$ and assume

$$\mu = \lambda\mu_1 + (1 - \lambda)\mu_2 \text{ for } \mu_1, \mu_2 \in \mathcal{M}_\infty, 0 < \lambda < 1.$$

Then $\int f d\mu = \lambda \int f d\mu_1 + (1 - \lambda) \int f d\mu_2$, for all $f \in C(X.\mathbb{R})$. Thus, inductively we see that

$$\int f_i d\mu_1 = \int f_i d\mu_2 = \int f_i d\mu \text{ for } i \geq 1.$$

By density of $\{f_n\}_{n=0}^{\infty} \subset C(X, \mathbb{R})$ we have that

$$\int f d\mu_1 = \int f d\mu_2, \forall f \in C(X, \mathbb{R}).$$

Thus $\mu_1 = \mu_2$, i.e., $\mu$ is extremal.

□

## 3.3 Recurrence and ergodic measures

*If almost all points return infinitely often, then what is the average return time?*

*Let $B \in \mathcal{B}$ with $\mu(B) > 0$. We define a function $n_B : B \to \mathbb{R}$ by*

$$n_B(x) = \inf\{n \geq 1 : T^n x \in B\}.$$

*The proof uses one of the more useful constructions in ergodic theory: The induced system. Since by the Poincaré recurrence theorem a.e. ($\mu$) $x \in B$ returns to $B$ (infinitely often) we can define the first return map $T_B : B \to B$ by*

$$T_B(x) = T^{n_B(x)} \in B.$$

*We can define a probability measure $\mu_B$ on $B$ by $\mu_B(A) = \frac{\mu(A \cap B)}{\mu(B)}$ (with respect to the obvious sigma algebra on $B$ consisting of sets $A \cap B$, where $A \in \mathcal{A}$). We can easily check that $T_B$ preserves $\mu_B$.*

**Theorem 3.12** (Kac)**.** *Assume that $\mu$ is ergodic, then $\int_B n_B(x) d\mu_B(x) = \frac{1}{\mu(B)}$*

*Proof.* Furthermore, by ergodicity we can partition the space $X$ by

$$X = \cup_{n=1}^{\infty} \cup_{k=0}^{n-1} \{T^k x : x \in B, n_B(x) = n\}$$

and therefore write

$$\begin{aligned} 1 = \mu(X) &= \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} \mu(\{T^k x : x \in B, n_B(x) = n\}) \\ &= \sum_{n=1}^{\infty} n\mu(\{x : x \in B, n_B(x) = n\}) = \int n_B d\mu \\ &= \mu(B) \int n_B d\mu_B \end{aligned}$$

□

**Example 3.13.** *Let $T : \Sigma \to \Sigma$ be the full shift on two symbols and $\mu$ the Bernoulli measure with $(p_1, p_2) = (\frac{1}{2}, \frac{1}{2})$. Let us first take $B = [0] = \{x \in \Sigma : x_0 = 0\}$. We see that $x \in B$ with $n_B(x) = n$ must be of the form:*

$$x = (\cdots, 0, \underbrace{1, 1, \cdots, 1}_{\times n}, 0, \cdots)$$

*with a 0 in the zeroth and nth place and 1s in between. In particular, we see that*

$$\mu(\{x \in B : n_B(x) = 0\}) = \frac{1}{2^{n+2}}.$$

*As in the proof of Kac's theorem, we see that the average return time is*

$$\int_B n_B(x) d\mu_B(x) = \sum_{n=1}^{\infty} \frac{n}{2^{n+2}} = 2 = \frac{1}{\mu(B)}.$$

*More generally, we can consider $B_N = [\underbrace{0, \cdots, 0}_{\times N}]$. In this case we see that $x \in B_N$ with $n_{B_N}(x) = n$ must be of the form:*

$$x = (\cdots, \underbrace{0 \cdots, 0}_{\times N}, \underbrace{1, 1, \cdots, 1}_{\times n}, \underbrace{0 \cdots, 0}_{\times N}, \cdots)$$

*with an $N$-blocks of 0s starting in the zeroth and nth places and 1s in between. In particular, we see that*

$$\mu(\{x \in B_N : n_{B_N}(x) = 0\}) = \frac{1}{2^{n+2N}}.$$

*As in the proof of Kac's theorem, we see that the average return time is*

$$\int_{B_N} n_{B_N}(x) d\mu_{B_N}(x) = \sum_{n=1}^{\infty} \frac{n}{2^{n+2N}} = 2^{2N} = \frac{1}{\mu(B_N)}.$$

*We can also consider what happens as $N \to +\infty$. In this case we can see there is an exponential distribution of return times, i.e., for any $\alpha > 0$,*

$$\mu_{B_N}\{x \in B_N : n_{B_N}(x) \geq \alpha 2^N\} = \left(1 - \frac{1}{2^N}\right)^{\alpha 2^N} \to \exp\left(\frac{1}{\alpha}\right)$$

### 3.4 Moore ergodicity theorem

*This is a more general result about group actions.*

**Theorem 3.14.** *Let $H < SL(2,\mathbb{R})$ be a closed subgroup and consider the action $H \times X \to X$ defined by $(h, g\Gamma) \mapsto hg\Gamma$. This action is ergodic if and only if $H$ is not compact.*

*In particular we see that ergodicity of the geodesic and horocycle flows follows from this result.*

## 4 Mean Ergodic Theorems

*We can consider the von- Neumann Ergodic Theorem.*

**Theorem 4.1** (von Neumann Mean Ergodic Theorem)**.** *Let $\mu$ be a $T$-invariant probability measure and let $f_1, f_2 \in L^1(X,\mu)$. Then*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1}\int f_1(T^k x) f_2(x) d\mu(x) = \int f_1 d\mu \int f_2 d\mu$$

*Proof.* Let us write $\|f\|_2 = \left(\int |f|^2 d\mu\right)^{1/2}$. Let

$$\mathcal{G} = \left\{ f \in L^2(X,\mathcal{B},m) : \frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k \text{ concerges w.r.t. } \|\cdot\|_2 \right\}.$$

We want to show that $\mathcal{G} = L^2(X,\mathcal{B},\mu)$. We begin by showing it is closed.

**Lemma 4.2.** $\mathcal{G}$ *is closed.*

*Proof.* Let $(f_k)_{k=1}^\infty \subset \mathcal{G}$ and assume that $f_k \to f$. We want to show that $f \in \mathcal{G}$ too. Given $\epsilon > 0$, choose $k$ sufficiently large that $\|f - f_k\| < \epsilon$. Then for $n, m \geq 1$:

$$\begin{aligned}\|\frac{1}{m}\sum_{k=0}^{m-1} f\circ T^k - \frac{1}{m}\sum_{k=0}^{m-1} f\circ T^k\| \leq & \underbrace{\|\frac{1}{m}\sum_{k=0}^{m-1} f(f-f_k)\circ T^k\|}_{\leq\|f-fk\|\leq\epsilon} \\ & + \underbrace{\|\frac{1}{m}\sum_{k=0}^{m-1} f\circ T^k - \frac{1}{m}\sum_{k=0}^{m-1} f\circ T^k\|}_{\to 0 \text{ as } n,m\to+\infty} \\ & + \underbrace{\|\frac{1}{n}\sum_{k=0}^{n-1} f(f-f_k)\circ T^k\|}_{\leq\|f-f_k\|\leq\epsilon}\end{aligned}$$

In particular, since $\epsilon > 0$ can be chosen arbitrarily small,

$$\left\{\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k\right\}$$

is Cauchy. Since $L^2(X,\mu)$ is complete, the sequence converges (i.e., $f \in \mathcal{G}$). Thus $\mathcal{G}$ is closed. □

Two special types of functions can be found in $\mathcal{G}$:

1. If $f$ satisfies $f \circ T = f$ then

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i = \frac{1}{n}\sum_{i=0}^{n-1} f = f \to f \text{ as } n \to +\infty.$$

2. If $f = g - g \circ T$ for some $g \in L^2(X,\mu)$ then

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \to 0 \text{ as } n \to +\infty.$$

(since $\|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i\| \leq \|\frac{1}{n}(g - gT^n)\| \leq 2\|g\|/n \to 0$ as $n \to +\infty$).

Let $\langle f_1, f_2\rangle = \int f_1 f_2 d\mu$ for $f_1, f_2 \in L^2(X,\mathcal{B},\mu)$. To show that $\mathcal{G} = L^2(X,\mu)$, we can assume for a contradiction that there is $h \in L^2(X,\mu)$ with $h \neq 0$ and $\langle h, f\rangle = 0$, for all $f \in \mathcal{G}$ (since $\mathcal{G}$ is closed).

Observe that

$$\|h - h \circ T\|^2 = \|h\|^2 - \langle h, h \circ T\rangle - \langle h \circ T, h\rangle + \|h \circ T\|^2$$

and

1. $\|h \circ T\|^2 = \int h^2 \circ T d\mu == \int h^2 d\mu = \|h\|^2$
2. $\langle h, h \circ T\rangle = \int h.h \circ T d\mu = \langle h \circ T, h\rangle$
3. $\langle h, \underbrace{h - h \circ T}_{\in \mathcal{G}}\rangle$ so that $\|h\|^2 = \langle h, h \circ T\rangle$

In particular, $\|h - h \circ T\| = 0$, i.e., $h = h \circ T \in \mathcal{G}$ giving a contradiction.

Finally, to deduce the von Neumann ergodic theorem if $\frac{1}{N}\sum_{k=0}^{N-1} f_1(T^k x) \to \overline{f}_1 \in L^2(X,\mu)$ as $N \to +\infty$ then

$$\frac{1}{N}\sum_{k=0}^{N-1}\int f_1(T^k x) f_2(x) d\mu(x) \to \int \overline{f}_1 d\mu \int f_2 d\mu$$

and $\int \overline{f}_1 d\mu = \int f d\mu$. □

Finally, there are a wealth of examples that come from Hamiltonian flows, where motion is described by classical laws of motion. In this case a natural invariant measure is the Liouville measure (which is absolutely continuous with respect to Lebesgue measure).

**Question 4.3** (Harder Question)**.** *For which Hamiltonian (or even geodesic flows) is the (normalized) Liouville measure ergodic?*

# 5 Pointwise Ergodic theorems

## 5.1 Birkhoff Ergodic Theorem

Let $T : X \to X$ be a transformation that preserves a probability measure $\mu$.

The original pointwise ergodic theorem is due to Birkhoff

**Theorem 5.1** (Birkhoff, 1931 (Ergodic version))**.** *Assume that $T : X \to X$ is ergodic. Let $f \in L^1(X, \mu)$ and let $f^n(x) := \sum_{k=0}^{n-1} f(T^k x)$. Then $\frac{1}{n} f^n(x) \to \int f d\mu$, a.e. $(\mu)$.*

We will prove the theorem under the simplifying assumption that $f \in L^\infty(X, \mu)$, and return to the general case later.

*Proof.* Let us define $\underline{f}(x) := \liminf_{n\to+\infty} \frac{1}{n} f^n(x)$, for a.e. $(\mu)$ $x \in X$. In particular, since $\underline{f}(x) = \underline{f}(Tx)$ then by $\underline{f}$ is constant.

- Fix $\epsilon > 0$ and define $n(x) := \inf\{n \geq 1 : f^n(x) \leq n(\underline{f} + \epsilon)\}$.
- Fix $M > 0$ and define $A = \{x : n(x) \geq M\}$.

*Claim*: For $n \geq 1$, $f^n(x) \leq n(\underline{f} + \epsilon) + \sum_{i=0}^{n-1} \chi_A(T^i x)\|f\|_\infty$.

*Proof of Claim* We can cover the set $\{1, 2, \cdots, n-1\}$ by sets of the form

1. $\{k : T^k x \in A\}$;
2. $\{l, l+1, \cdots, l + n(T^l x) - 1\}$; or
3. $\{n - M, \cdots, n-1\}$.

This completes the proof of the claim. □

Thus

$$\frac{f^n(x)}{n} \leq (\underline{f} + \epsilon) + \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i x)\|f\|_\infty + \frac{M}{n}\|f\|_\infty$$

and integrating gives

$$\int f d\mu \leq n(\underline{f} + \epsilon) + \mu(A)\|f\|_\infty + \frac{M}{n}\|f\|_\infty$$

First let $n \to +\infty$ and then $M \to +\infty$, which implies that $\mu(A) \to 0$. Then let $\epsilon \to 0$.

Therefore $\int f d\mu \leq \underline{f}$. Replacing $f$ by $-f$ gives $\overline{f} \leq \int f d\mu$.

Therefore we can conclude that

$$\overline{f} = \underline{f} = \int f d\mu = \int f d\mu$$

which completes the proof of the Birkhoff ergodic theorem. □

We next relax the assumption that $\mu$ is ergodic. In this case we cannot assume that $\overline{f}$ and $\underline{f}$ are constant. However, it suffices to show that $\int \overline{f} d\mu \leq \int f d\mu \leq \int \underline{f} d\mu$ since then

$$\int \underbrace{(\underline{f} - \overline{f})}_{\leq 0} d\mu \geq 0$$

which implies that $\overline{f} = \underline{f}$ a.e. $(\mu)$.

If we don't assume ergodicity then integrating

$$\frac{f^n(x)}{n} \leq (\underline{f} + \epsilon) + \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i x) \|f\|_\infty + \frac{M}{n} \|f\|_\infty$$

gives us

$$\int f d\mu \leq (\int \underline{f} d\mu + \epsilon) + \mu(A) \|f\|_\infty + \frac{M}{n} \|f\|_\infty$$

First let $n \to +\infty$ and then $M \to +\infty$, which implies that $\mu(A) \to 0$. Then let $\epsilon \to 0$.

Therefore $\int f d\mu \leq \int \underline{f} d\mu$. Replacing $f$ by $-f$ gives $\int \overline{f} d\mu \leq \int f d\mu$.

As we observed above, this suffices to prover the theorem for invariant measures and $f \in L^\infty$.

We leave it as an exercise to do the general case $f \in L^1$.

## 5.2 A second proof of the Birkhoff Ergodic Theorem

Let $T : X \to X$ be preserve the probability measure $m$ and let $\phi : X \to \mathbb{R}$ be in $L^1(X)$.
**Ergodic case** Given $\epsilon > 0$, let us define $f(x) = \phi(x) - \int \phi(x) dm(x) - \epsilon$. Then $\int f(x) dm(x) = -\epsilon$. For each $n \geq 1$, denote the sums $f_n(x) := \sum_{k=0}^{n-1} f(T^k x)$. For each $N \geq 1$ we can denote the maximum of the first $N$ terms to be $F_N(x) = \max\{f_n(x) : 1 \leq n \leq N\}$. Observe that $F_{N+1}(x) \geq F_N(x)$, since we take the maximum over more terms;

Moreover, since $f_{n+1}(x) = f(x) + f_n(Tx)$ we see

$$F_{N+1}(x) = f(x) \text{ if } F_N(Tx) \leq 0 f(x) + F_N(Tx) \text{ otherwise}$$

$\int F_N(Tx) dm(x) = \int F_N(x) dm(x)$ since the measure $m$ is $T$ invariant.

Let $A = \{x : F_N(x) \to \infty \text{ as } N \to +\infty\}$. We can see that

$$\begin{aligned} 0 &\leq \lim_{N \to +\infty} \int_A [F_{N+1}(x) - F_N(x)] \, dm(x) \\ &= \lim_{N \to +\infty} \int_A [F_{N+1}(x) - F_N(Tx)] \, dm(x) = \int_A f(x) dm(x). \end{aligned} \tag{1}$$

(by dominated convergence).

Clearly $T^{-1}A = A$, and so by ergodicity either $m(A) = 0$ or $m(A) = 1$. However, if $m(A) = 1$ then the last term in (1) becomes $\int f dm < -\epsilon$, giving a contradiction. Therefore, $m(A) = 0$. In particular, almost every point $x \in X$ actually lies in $X - A$ and so be definition

$$\limsup_{n \to +\infty} \frac{f_n(x)}{n} \leq \limsup_{n \to +\infty} \frac{F_n(x)}{n} \leq 0$$

In particular,

$$\limsup_{n\to+\infty}\frac{f_n(x)}{n}=\limsup_{n\to+\infty}\frac{\phi_n(x)}{n}-\int\phi dm-\epsilon\le 0$$

Since $\epsilon>0$ can be chosen arbitrarily small we deduce that

$$\limsup_{n\to+\infty}\frac{\phi_n(x)}{n}\le\int\phi dm \tag{2}$$

If we replace $\phi$ by $-\phi$ this inequality becomes

$$\liminf_{n\to+\infty}\frac{\phi_n(x)}{n}\ge\int\phi dm \tag{3}$$

Comparing (2) and (3) gives the result.

**Invariant measure case** Let $\epsilon>0$ and then write $f:=\phi-E(\phi|I)-\epsilon$, then $E(\phi|I)<-\epsilon$.

Following the previous argument as far as the inequality (1), we can then write $\int_A f(x)dm(x)=\int_A E(f|I)(x)dm(x)$. Since we are assuming that $E(\epsilon|I)(x)<0$ then we can deduce that $m(A)=0$. In particular, almost every point $x$ lies in $X-A$ and then

$$\limsup_{N\to+\infty}\frac{f_n(x)}{n}\le\limsup_{N\to+\infty}\frac{F_n(x)}{n}\le 0$$

In particular,

$$\limsup_{n\to+\infty}\frac{f_n(x)}{n}=\limsup_{n\to+\infty}\frac{\phi_n(x)}{n}-E(\phi|I)-\epsilon\le 0$$

Since $\epsilon>0$ can be chosen arbitrarily small we deduce that

$$\limsup_{n\to+\infty}\frac{\phi_n(x)}{n}\le E(\phi|I)(x) \tag{4}$$

If we replace $\phi$ by $-\phi$ this inequality becomes

$$\liminf_{n\to+\infty}\frac{\phi_n(x)}{n}\ge E(\phi|I)(x) \tag{5}$$

Comparing (4) and (5) gives the result.

□

**Example 5.2** (Normal numbers)**.** *Let*

$$x=\sum_{n=1}^{\infty}\frac{b_n(x)}{b^n}$$

*be the b-expansion of* $0\le x\le 1$ *where* $b\ge 2$ *and* $b_n(x)\in\{0,1,2,\cdots,b-1\}$. *For a.e., (m)* $x\in[0,1]$ *this expansion is unique (where m is Lebesgue measure). Let*

$$N_b=\left\{0\le x\le 1:\ \lim_{N\to+\infty}\frac{Card\{1\le n\le N:\ b_n(x)=i\}}{N}=\frac{1}{b}\ for\ i=0,1,\cdots,b-1\right\}$$

*then we can see that* $m(N_b) = 1$. *More precisely,* $Tx = bx \pmod 1$ *then writing*

$$\frac{Card\{1 \leq n \leq N : b_n(x) = i\}}{N} = \frac{1}{N}\sum_{n=1}^{N} \chi_{[i/b,(i+1)/b]}(T^n x) \to \int \chi_{[i/b,(i+1)/b]}(x)dx = \frac{1}{b}.$$

*for a.e., (m)* $x \in [0, 1]$.

*The set of normal numbers* $\cap_{b=2}^{\infty} N_b$ *to all bases is again a set of full measure (since the intersection of a countable union of sets of full measure again has full measure). However, there are very few explicit examples of normal numbers know.*

*Remark* 5.3 (Central Limit Theorem). In fact, there is a stronger result, i.e., $\exists \sigma > 0$, $\forall y \in \mathbb{R}$

$$\lim_{N\to+\infty} m\left(\left\{0 \leq x \leq 1 : \frac{\text{Card}\{1 \leq n \leq N : b_n(x) = i\} - N/b}{\sqrt{N}}\right\}\right) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{y} e^{-t^2/2\sigma^2} dt$$

**Example 5.4** (Leading Digits of $2^n$)**.** *Consider the sequence* $\{2^n\} = \{2, 4, 8, 16, 32, 64, 128, 256, \cdots\}$. *The sequence of leading digits is* $\{x_n\} = \{2, 4, 8, 1, 3, 6, 1, 2, \cdots\}$.

**Lemma 5.5.** *The frequency with which* $x_n = l \in \{1, 2, \cdots, 9\}$ *is* $\log_{10}(1 + \frac{1}{l})$*, i.e.,*

$$\lim_{N\to\infty} \frac{Card\{1 \leq n \leq N : x_n = l\}}{N} = \log_{10}\left(1 + \frac{1}{l}\right)$$

# 6 Subadditive Ergodic Theorem

More generally, we can consider a subadditive sequence of functions $f_n(x) \in L^1(X, \mu)$, i.e., $f_{n+m}(x) \leq f_n(x) + f_m(T^n x)$, for $n, m \geq 1$, where $T : X \to X$ preserves a probability measure $\mu$.

**Example 6.1.** *Let us fix* $f \in L^1(X, \mu)$ *and then define* $f_n(x) := \sum_{k=0}^{n-1} f(T^k x)$. *We can then see that for a.e. (*$\mu$*)* $x \in X$*:*

$$f_{n+m}(x) = f_n(x) + f_m(T^n x).$$

**Example 6.2.** *Let* $M : X \to GL(d, \mathbb{R})$ *be a function taking values in the family (invertible)* $d \times d$ *matrices. We can define a norm on* $GL(d, \mathbb{R})$ *by*

$$\|A\| = \sup\{\|Ax\|_2 : x \in \mathbb{R}^2 \text{ such that } \|x\|_2 \leq 1\}.$$

*We assume that* $\int \log_+ \|M(x)\| d\mu(x) < +\infty$ *(where* $\log+(\xi) = \max\{\xi, 0\}$*). Let* $f_n(x) = \log \|M(x) \cdot M(Tx) \cdots M(T^{n-1}x)\|$ *then the subadditivity follows from the standard norm inequality* $\|A_1 A_2\| \leq \|A_1\|\|A_2\|$.

**Theorem 6.3** (Kingman, 1967)**.** *Assume that* $\mu$ *is ergodic. Let* $f_n \in L^1(X, \mu)$ *be subadditive. Then the following limit exists*

$$\lim_{n\to+\infty} \frac{1}{n} f_n(x) = l := \inf\left\{\int f_n d\mu : n \geq 1\right\},$$

*a.e. (*$\mu$*)* $x \in X$.

*Proof.* Let us define $\underline{f}(x) := \liminf_{n\to+\infty} \frac{1}{n} f_n(x) \le \overline{f}(x) := \limsup_{n\to+\infty} \frac{1}{n} f_1(T^i x) = 0$, for a.e. $(\mu)$ $x \in X$. Since

$$\underbrace{\liminf_{n\to+\infty} \frac{1}{n+1} f_{n+1}(x)}_{=\underline{f}(x)} \le \underbrace{\liminf_{n\to+\infty} \frac{1}{n+1} f_1(x)}_{=0} + \underbrace{\liminf_{n\to+\infty} \frac{1}{n+1} f_n(Tx)}_{=\underline{f}(Tx)}$$

we have that $\underline{f}(x) \le \underline{f}(Tx)$ and thus ergodicity implies that $\underline{f}$ is constant. Similarly, we see that $\overline{f}(x) \le \overline{f}(Tx)$ and thus ergodicity implies that $\overline{f}$ is constant as well.

Fix $N \ge 1$ and choose $1 \le i \le N$. We can then write any $n \ge i$ as $n = i + mN + r$ for unique choices $0 \le r \le N-1$ and $m \ge 0$. By subadditivity

$$\begin{aligned} f_n(x) &\le f_i(x) + f_{nm}(T^i x) + f_r(T^{i+nm} x) \\ &\le f_i(x) + \sum_{j=0}^{m-1} f_N(T^{jN+i} x) + f_r(T^{(m-1)N+i} x). \end{aligned}$$

Summing this inequality over $1 \le i \le N$ gives:

$$N f_n(x) \le \sum_{i=1}^{N} f_i(x) + \underbrace{\sum_{i=1}^{N} \sum_{j=0}^{m-1} f_N(T^{jN+i} x)}_{=\sum_{k=0}^{mN-1} f_N(T^k x)} + \sum_{i=1}^{N} f_{\underbrace{n-i-mN}_{=r}}(T^{nN+i} x).$$

Dividing by $nN$ and letting $n \to +\infty$ gives

$$\underbrace{\limsup_{n\to+\infty} \frac{1}{n} f_n(x)}_{=:\overline{f}(x)} \le 2 \underbrace{\limsup_{n\to+\infty} \frac{1}{n} \left( \max_{1\le i\le N} \|f_i\|_\infty \right)}_{=0} + \underbrace{\limsup_{n\to+\infty} \frac{1}{nN} f_N^{nN}(x)}_{=\frac{1}{N}\int f_N d\mu}$$

using the Birkhoff theorem for $f_N$, for any $N \ge 1$. In particular, we see that

$$\overline{f}(x) \le l =: \inf_N \left\{ \frac{1}{N} \int f_N d\mu \right\}.$$

Next we can turn our attention to $\underline{f}$. Moreover, we may as well assume $l > -\infty$.

Fix $\epsilon > 0$. Let us define $n : X \to \mathbb{N}$ by $n(x) = \inf\{n \ge 1 : f_n(x) \le n(\underline{f} + \epsilon)\}$.

Fix $M > 0$. Let use define the set of points $A = \{x : n(x) \ge M\}$.

*Claim*: For $n \ge 1$, $f_n(x) \le n(\underline{f} + \epsilon) + \sum_{i=0}^{n-1} \chi_A(T^i x)\|f_1\|_\infty + M\|f_1\|_\infty$ for a.e. $(\mu)$ $x \in X$.

*Proof of Claim.* This is similar to the claim in the first proof of the Birkhoff ergodic theorem. We can cover the set $\{1, 2, \cdots, n-1\}$ by sets of the form

1. $\{k : T^k x \in A\}$;
2. $\{l, l+1, \cdots, l + n(T^l x) - 1\}$; or
3. $\{n - M, \cdots, n-1\}$.

This completes the proof of the claim. □

Thus

$$\frac{f_n(x)}{n} \le (\underline{f} + \epsilon) + \frac{1}{n}\sum_{i=0}^{n-1} \chi_A(T^i x)\|f_1\|_\infty + \frac{M}{n}\|f_1\|_\infty.$$

Integrating this inequality gives that

$$\frac{\int f_n d\mu}{n} \le (\underline{f} + \epsilon) + \mu(A)\|f_1\|_\infty + \frac{M}{n}\|f_1\|_\infty.$$

and therefore

$$l = \inf_{n\ge 1}\left\{\frac{\int f_n d\mu}{n}\right\} \le (\underline{f} + \epsilon) + \mu(A)\|f_1\|_\infty + \frac{M}{n}\|f_1\|_\infty.$$

for any $n \ge 1$.

First we can let $n \to +\infty$. Next we let $M \to \infty$, which in turn implies that $\mu(A) \to 0$. Finally, we let $\epsilon \to 0$. This therefore gives that $l \le \underline{f}$ a.e. $(\mu)$ $x \in X$.

Therefore $\overline{f}(x) = \underline{f}(x) = l$, i.e., $l = \lim_{n\to+\infty} \frac{f_n(x)}{n}$. □

# 7 Ergodicity of flows and the Hopf method

Let $X$ be a space with associated sigma-algebra $\mathcal{B}$.

**Definition 7.1.** *We say that $\mu$ is $T$-*invariant *if for any $A \in \mathcal{B}$ we have that $\mu(T^{-t}A) = \mu(A)$ for all $t \in \mathbb{R}$.*

**Definition 7.2.** *We say that $(T^t, X, \mu)$ is* ergodic *if whenever $T^{-t}A = A \in \mathcal{B}$ for all $t \in \mathbb{R}$ then $\mu(A) = 0$ or 1.*

**Theorem 7.3** (Birkhoff - Ergodic Theory case)**.** *Let $\mu$ be an ergodic measure. Let $f \in L^1(X,\mu)$ then $\lim_{t\to+\infty} \frac{1}{t}\int_0^t f(T^u x)du = \int f d\mu$, a.e. $(\mu)$.*

We also need to consider the case that the invariant measure is not necessarily ergodic. This requires a little more notation. Let $\mathcal{I} \subset \mathcal{B}$ be the sub-sigma-algebra of $T^t$-invariant sets, i.e.,

$$\mathcal{I} = \{B \in \mathcal{B} : T^{-t}B = B \text{ (up to a set of zero measure) } \forall t \in \mathbb{R}\}.$$

Given $f \in L^1(X,\mathcal{B},\mu)$ there is a unique $g \in L^1(X,\mathcal{I},\mu)$ such that

1. $g$ is $\mathcal{I}$-measurable (i.e., for any Borel measurable set $B \subset \mathbb{R}$, $g^{-1}B \in \mathcal{I}$ );
2. $\int_B f d\mu = \int_B g d\mu$, $\forall B \in \mathcal{I}$.

We then denote $g = E(f|\mathcal{I})$, the conditional expectation of $f$ with respect to $\mathcal{I}$.

**Theorem 7.4** (Birkhoff - Invariant measure case)**.** *Let $f \in L^1(X,\mu)$ then $\lim_{t\to+\infty} \frac{1}{t}\int_0^t f(T^u x)du = E(f|\mathcal{B})$, a.e. $(\mu)$.*

We want to consider two important examples.

**Example 7.5** (Geodesic flows)**.** *We want to show the following for the geodesic flow on* $X = SL(2,\mathbb{R})/SL(2,Z)$.

**Theorem 7.6.** *The geodesic flow* $g_t : X \to X$ *defined by*

$$g_t g SL(2,\mathbb{Z}) = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} g SL(2,\mathbb{Z})$$

*is ergodic.*

The proof uses the so called Hopf method. We define flows $h_t : X \to X$ and $h_t^- : X \to X$ defined by

$$h_s g SL(2,\mathbb{Z}) = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} g SL(2,\mathbb{Z}) \text{ and } h_s^- g SL(2,\mathbb{Z}) = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} g SL(2,\mathbb{Z}).$$

We begin with the crucial connection between $g_t$ and $h_t$, and $g_t$ and $h_t^-$.

**Lemma 7.7.** *We can write* $g_t h_s g_{-t} = h_{se^t}$ *and* $g_t h_s^{--} g_{-t} = h_{se^{-t}}^-$

*Proof.* This follows by matrix multiplication

$$g_t h_s g_{-t} = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-t/2} & 0 \\ 0 & e^{t/2} \end{pmatrix} = \begin{pmatrix} 1 & se^t \\ 0 & 1 \end{pmatrix} = h_{se^t}$$

and

$$g_t h_s^- g_{-t} = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} \begin{pmatrix} e^{-t/2} & 0 \\ 0 & e^{t/2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ se^{-t} & 1 \end{pmatrix} = h_{se^{-t}}^-$$

□

We can apply the invariant measure version of the ergodic theorem for flows to deduce that for a.e. $(\mu)$ $x \in X$ we have that for any $f \in L^1(X,\mu)$.

$$E(f|\mathcal{I})(x) = \lim_{T\to+\infty} \frac{1}{T} \int_0^T f(g_t x) dt$$

and

$$E(f|\mathcal{I})(x) = \lim_{T\to+\infty} \frac{1}{T} \int_0^T f(g_{-t} x) dt.$$

**Lemma 7.8.** *If* $E(f|\mathcal{I})(x)$ *is constant if a.e.* $(\mu)$ *and all* $f \in C(X)$ *then the flow* $g_t$ *is ergodic.*

*Proof.* Recall that to show $\mu$ is ergodic it suffices to show that for any $f = f \circ T \in L^2(X,\mu)$ we have that $f$ is constant. In terms of the operator $E(\cdot|\mathcal{I})(x)$ this is equivalent to saying that the image of $L^2(X,\mu)$ are the constant functions. However, since $E(\cdot|\mathcal{I}) : L^2(X,\mu) \to L^2(X,\mu)$ contracts the norm, and $C(X) \subset L^2(X,\mu)$ is dense, this equivalent to the same statement for continuous functions. □

Clearly, if $y = g_s(x)$ we have that

$$\begin{aligned} E(f|\mathcal{I})(x) - E(f|\mathcal{I})(y) &= E(f|\mathcal{I})(x) - E(f|\mathcal{I})(g_s x) \\ &= \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_t x)dt - \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_{s+t}x)dt \\ &= -\lim_{T\to+\infty} \frac{1}{T}\int_{-s}^0 f(g_t x)dt + \lim_{T\to+\infty} \frac{1}{T}\int_T^{T-s} f(g_t x)dt = 0. \end{aligned}$$

The important point is that if $y = h_s x$ then

$$\begin{aligned} E(f|\mathcal{I})(x) - E(f|\mathcal{I})(y) &= E(f|\mathcal{I})(x) - E(f|\mathcal{I})(h_s x) \\ &= \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_{-t} x)dt - \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_{-t}h_s x)dt \\ &= \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_{-t} x)dt - \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(h_{se^{-t}}g_{-t} x)dt \end{aligned}$$

However, given $\epsilon > 0$ for $T_0 > 0$ sufficiently large we have that $\|f(x) - f(h_{se^{-t}x})\|_\infty < \epsilon/2$ for $t > T_0$. In particular, we can bound

$$\left|\frac{1}{T}\int_0^T f(g_t x)dt - \frac{1}{T}\int_0^T f(h_{se^{-t}}g_t x)dt\right| \leq \frac{2\|f\|_\infty T_0}{T} + \frac{(T-T_0)\epsilon}{2T}.$$

and deduce that for $y = h_s x$ we have that $|E(f|\mathcal{I})(x) - E(f|\mathcal{I})(y)| < \epsilon$.

Similarly, if $y = h_s^- x$ then

$$\begin{aligned} E(f|\mathcal{I})(x) - E(f|\mathcal{I})(y) &= E(f|\mathcal{I})(x) - E(f|\mathcal{I})(h_s^- x) \\ &= \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_t x)dt - \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_t h_s^- x)dt \\ &= \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_t x)dt - \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(h_{se^t}^- g_t x)dt \end{aligned}$$

However, given $\epsilon > 0$ for $T_0 > 0$ sufficiently large we have that $\|f(\cdot) - f(h_{se^t\cdot}^-)\|_\infty < \epsilon/2$ for $t > T_0$. In particular, we can bound

$$\left|\frac{1}{T}\int_0^T f(g_t x)dt - \frac{1}{T}\int_0^T f(h_{se^t}^- g_t x)dt\right| \leq \frac{2\|f\|_\infty T_0}{T} + \frac{(T-T_0)\epsilon}{2T}.$$

and deduce that for $y = h_s^- x$ we have that $|E(f|\mathcal{I})(x) - E(f|\mathcal{I})(y)| < \epsilon$.

We next need the following result.

**Lemma 7.9.** *We can write any element* $\gamma = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in G = SL(2,\mathbb{R})$ *with* $\delta \neq 0$ *in the form* $\gamma = h_{s_1} g_t h_{s_2}^-$.

*Proof.* It suffices to observe that every matrix in $SL(2,\mathbb{R})$ with $\delta \neq 0$ can be written in the form

$$\begin{pmatrix} 1 & s_1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ s_2 & 1 \end{pmatrix} = \begin{pmatrix} e^{t/2} + s_2 s_1 e^{-t/2} & s_1 e^{-t/2} \\ s_2 e^{-t/2} & e^{-t/2} \end{pmatrix}$$

since we can choose $t$ such that $e^{-t/2} = \delta$. □

Thus given almost any two points $x, x' \in G/\Gamma$ we can choose $\gamma \in G$ such that $x' = \gamma x$. In particular, we can write $\gamma x = h_{s_1} g_t h^-_{s_2} x$ and define

$$y_1 = h^-_{s_2} x, y_2 = g_t y_1, x' = h_{s_1} y_2.$$

Thus we can show that the averages $E(f|\mathcal{I})(x) = E(f|\mathcal{I})(x')$.

**Example 7.10** (Horocycle flows). *We define probability measures $\nu_t \in \mathcal{M}$, for $t > 0$, by*

$$\int f d\nu_t = \frac{1}{t}\int_0^t f(h_s x) ds \text{ for } f \in C^0(X).$$

*Let us denote*

$$M_t f(x) = \int_0^1 f(g_{-\log t} h_s x) ds \text{ for } t > 0.$$

*We first observe that*

$$\begin{aligned} \frac{1}{t}\int_0^t f(h_s x) ds &= \int_0^1 f(h_{st} x) ds \text{ (by a change of variable)} \\ &= \int_0^1 f(g_{-\log t} h_s g_{-\log t} x) ds \text{ (since } h_{st} = g_{-\log t} h_s g_{-\log t}\text{ )} \\ &= M_t f(g_{-\log t} x) \text{ (by definition of } M_t \text{ above)} \end{aligned} \tag{1}$$

We begin with the following estimate.

**Lemma 7.11.** *The family $\{M_t f(x) : t > 0\}$ is equicontinuous, i.e., for all $\epsilon > 0$ there exists $\delta > 0$ such that if $d(x, x') < \delta$ then $|M_t f(x) - M_t f(x')| < \epsilon$ for all $t > 0$.*

*Proof.* Let us define

$$D_x(\epsilon) := \{g_t h^-_s(x) : |t|, |s| \le \epsilon\} \text{ and } V_x(\epsilon) := h_{[0,1]} D_x(\epsilon) (= \cup_{0 \le t \le 1} h_t D_x(\epsilon)).$$

Then by uniform continuity of $f$ we have

$$\left| M_t f(x) - \frac{1}{\mathrm{Vol}(V_x(\epsilon))} \int_{V_x(\epsilon)} f(g_{-\log t} \cdot) d\mu \right| \to 0 \text{ as } \epsilon \to 0.$$

Moreover, for $x, x' \in X$ sufficiently close we can estimate

$$\begin{aligned} &\left| \frac{1}{\mathrm{Vol}(V_{x'}(\epsilon))} \int_{V_{x'}(\epsilon)} f(g_{-\log t} \cdot) d\lambda - \frac{1}{\mathrm{Vol}(V_x(\epsilon))} \int_{V_x(\epsilon)} f(g_{-\log t} \cdot) d\lambda \right| \\ =& \left| \int f(g_{-\log t} \cdot) \left( \frac{\chi_{V_{x'}(\epsilon)}}{\mathrm{Vol}(V_{x'}(\epsilon))} - \frac{\chi_{V_x(\epsilon)}}{\mathrm{Vol}(V_x(\epsilon))} \right) d\lambda \right| \\ \le& \underbrace{\|f \circ g_{-\log t}\|_2}_{\|f\|_2} \| \frac{1}{\mathrm{Vol}(V_{x'}(\epsilon))} \chi_{V_{x'}(\epsilon))} - \frac{1}{\mathrm{Vol}(V_x(\epsilon))} \chi_{V_x(\epsilon))} \|_2 \end{aligned}$$

by the Cauchy-Schwartz theorem (and $g_t$-invariance of $\mu$). Moreover, it is easy to see the term on the Right Hand Side tends to zero as $d(x, x') \to 0$. □

If $X$ is compact then we can apply the Arzela-Ascoli theorem there exists an accumulation point $\overline{f} \in C(X)$ of $\{M_t f(x) : t > 0\}$, i.e. there exists a subsequence such that $M_{t_k} f \to \overline{f}$ as $t_k \to +\infty$. If $X$ is not compact then we still know that this converges on compact sets (and in $L^2$).

**Lemma 7.12.** *$\overline{f}$ is constant on $h_s$-orbits.*

*Proof.* We have by (1) that

$$\begin{aligned} \|\frac{1}{t_k}\int_0^{t_k} f(h_s \cdot)ds - \overline{f} \circ g_{-\log t_k}\|_2 &= \|(M_{t_k} f) \circ g_{-\log t_k} - \overline{f} \circ g_{-\log t_k}\|_2 \\ &= \|M_{t_k} f - \overline{f}\|_2 \to 0 \end{aligned} \tag{2}$$

as $t_k \to 0$, by $g_t$-invariance of the probability measure $\mu$. Moreover, by the von Neumann mean ergodic theorem (for invariant probability measures)

$$\frac{1}{t_k}\int_0^{t_k} f(h_s \cdot)ds \to \widehat{f} = E(f|\mathcal{I}) \in L^2(X) \tag{3}$$

as $k \to +\infty$ in the $L^2$ norm. Moreover, this limit is naturally $h_s$-invariant, i.e., $\widehat{f} = \widehat{f} \circ h_s$ for all $s \in \mathbb{R}$. Thus,

1. We have $\|\overline{f} - \widehat{f} \circ g_{\log t_k}\|_2 = \|\overline{f} \circ g_{-\log t_k} - \widehat{f}\|_2 \to 0$ by the invariance of the measure and (2) and (3).
2. We can bound

$$\begin{aligned} &\|\overline{f} \circ h_s - \widehat{f} \circ g_{\log t_k}\|_2 \\ &= \|\overline{f} - \widehat{f} \circ g_{\log t_k} \circ h_{-s}\|_2 \text{ (by } h_s\text{-invariance of } \mu) \\ &= \|\overline{f} - \widehat{f} \circ h_{-st_k} \circ g_{\log t_k}\|_2 \text{ (since } g_{\log t_k} \circ h_{-s} = h_{-st_k} \circ g_{\log t_k}) \\ &= \|\overline{f} - \widehat{f} \circ g_{\log t_k}\|_2 \to 0 \text{ (since } \widehat{f} \circ h_{-st_k} = \widehat{f} \text{ by } h_s\text{-invariance of } \widehat{f}) \end{aligned}$$

Comparing 1 and 2 shows $\overline{f} = \overline{f} \circ h_s$, as required. □

Finally, we need the following simple topological fact (whose proof we omit).

**Lemma 7.13.** *There exists a dense $h_s$ orbit in $X$.*

In particular, since $\overline{f}$ is a continuous function and it is constant on $h_s$-orbits this implies that $\overline{f}$ is a constant. Since we had in the case that $X$ is compact convergence of the averages to a constant along every horocycle orbit, this is enough to deduce the following.

**Theorem 7.14.** *If $X$ is compact then the horocycle flow $h_t : X \to X$ is (uniquely) ergodic.*

More generally, even if $X$ isn't compact the $L^2$ convergence allows us to obtain.

**Theorem 7.15.** *The horocycle flow $h_t : X \to X$ is ergodic.*

# 8 Continued fractions

Let $0 \leq x \leq 1$ be an irrational number. There is a unique "Continued Fraction Expansion" of the form

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}} \text{ where } a_1, a_2, \ldots \in \mathbb{N}.$$

Let us consider $T : [0,1) \to [0,1)$ by

$$Tx = \begin{cases} \{1/x\}(= 1/x - [1/x]) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Then

$$x = \frac{1}{\frac{1}{x}} = \frac{1}{[1/x] + \{1/x\}} = \frac{1}{a_1 + Tx}$$

and

$$Tx = \frac{1}{(Tx)^{-1}} = \frac{1}{\frac{1}{x}} = \frac{1}{[1/Tx] + \{1/Tx\}} = \frac{1}{a_2 + T^2x}.$$

Thus

$$x = \frac{1}{a_1 + \left(\frac{1}{a_2 + T^2x}\right)} = \frac{a_2 + T^2x}{(a_1a_2 + 1) + a_1(T^2x)}$$

In particular, proceeding inductively, for $n \geq 2$

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots \cfrac{1}{a_n + T^n x}}}} \text{ where } a_n = [1/T^{n-1}x] \in \mathbb{N}.$$

We can rearrange this as

$$x = \frac{p_n + p_{n-1}(T^n x)}{q_n + q_{n-1}(T^n x)} \text{ for } n \geq 1.$$

We recall some basic facts.

**Lemma 8.1.** *1. We define $p_n, q_n$ inductively by:*

$$p_0 = 0, q_0 = 1, p_1 = 1, q_1 = a_1$$

$$p_n = a_n p_{n-1} + p_{n-2} \text{ and } q_n = a_n q_{n-1} + q_{n-2}$$

*2.* $p_{n-1}q_n - p_n q_{n-1} = (-1)^n$

*Proof.* 1. Substituting $T^n(x) = \frac{1}{a_{n+1} + T^{n+1}x}$ we can write

$$\frac{p_n + p_{n-1}(T^n x)}{q_n + q_{n-1}(T^n x)} = \frac{p_n + p_{n-1}\left(\frac{1}{a_{n+1} + T^{n+1}x}\right)}{q_n + q_{n-1}\left(\frac{1}{a_{n+1} + T^{n+1}x}\right)}$$

$$= \frac{\underbrace{(a_{n+1}p_n + p_{n-1})}_{=p_{n+1}} + p_n(T^{n+1}x)}{\underbrace{(a_{n+1}q_n + q_{n-1})}_{=q_{n+1}} + q_n(T^{n+1}x)}$$

by induction. It also holds with $n = 1$.

2. We can write

$$p_n q_{n+1} - p_{n+1} q_n = p_n(a_{n+1}q_n + q_{n-1}) - (a_{n+1}p_n + p_{n-1})q_n = p_n q_{n-1} - p_{n-1} q_n = -(-1)^{n-1} = (-1)^n$$

by induction. It also holds with $n = 1$.

□

**Corollary 8.2.** *$q_n \geq 2^{(n-1)/2}$ and $p_n \geq 2^{(n-2)/2}$*

*Proof.* The proof is by induction. By the above

$$q_n \geq q_{n-1} + q_{n-2} \geq 2^{(n-2)/2} + 2^{(n-3)/2} \geq 2 \cdot 2^{(n-3)/2} = 2^{(n-1)/2}.$$

by induction. It also holds for $n = 1$. Similarly for $p_n$. □

We now want to study the ergodic properties of $T$.

**Lemma 8.3** (Gauss)**.** *The probability measure $\mu$ defined by*

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{dx}{1+x} \text{ for any } A \in \mathcal{B}$$

*is $T$-invariant.*

*Proof.* It suffices to show that for any interval $(0, \alpha)$ we have that $\mu((0,\alpha)) = \mu(T^{-1}(0,\alpha))$. In particular,

$$\begin{aligned}
\mu(T^{-1}(0,\alpha)) &= \sum_{k=1}^{\infty} \mu\left(\left(\frac{1}{k+\alpha}, \frac{1}{k}\right)\right) = \frac{1}{\log 2} \sum_{k=1}^{\infty} \int_{\frac{1}{k+\alpha}}^{\frac{1}{k}} \frac{dx}{1+x} \\
&= \frac{1}{\log 2} \sum_{k=1}^{\infty} \left(\log\left(1 + \frac{1}{k}\right) - \log\left(1 + \frac{1}{k+\alpha}\right)\right) \\
&= \frac{1}{\log 2} \sum_{k=1}^{\infty} \left(\log\left(1 + \frac{\alpha}{k}\right) - \log\left(1 + \frac{\alpha}{k+1}\right)\right) \\
&= \frac{1}{\log 2} \sum_{k=1}^{\infty} \int_{\frac{\alpha}{k+1}}^{\frac{\alpha}{k}} \frac{dx}{1+x} \\
&= \frac{1}{\log 2} \int_0^{\alpha} \frac{dx}{1+x} = \mu\left((0,\alpha)\right).
\end{aligned}$$

□

**Theorem 8.4.** *The measure $\mu$ is ergodic.*

*Proof.* Consider the intervals

$$I_n(x) := \begin{cases} \left[\frac{p_n}{q_n}, \frac{p_n + p_{n-1}}{q_n + q_{n-1}}\right] & \text{if } n \text{ is even} \\ \left[\frac{p_n + p_{n-1}}{q_n + q_{n-1}}, \frac{p_n}{q_n}\right] & \text{if } n \text{ is odd} \end{cases}$$

then the length is $l(I_n(x)) = \frac{1}{q_n(q_n+q_{n-1})}$ (by Lemma (ii)), where $l$ denotes Lebesgue measure. We can define $\phi : [0,1] \to I_n$ by

$$\phi(t) = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots \frac{1}{a_n+t}}}} = \frac{p_{n-2} + (a_n+t)p_{n-1}}{q_{n-2} + (a_n+t)q_{n-1}}.$$

Note that $\phi$ is a bijection. Assume that $E = T^{-1}E \subset [0,1]$ then we can write

$$l(E \cap I_n(x)) = \int_0^1 \chi_E(\phi(t))|\phi'(t)|dt$$

and $|\phi'(t)| = \frac{1}{(q_{n-2}(a_n+t)q_{n-1})^2} \in \left[\frac{1}{(q_n+q_{n-1})^2}, \frac{1}{q_n^2}\right]$ by (i) and (ii).

In particular,

$$\frac{l(E \cap I_n)}{l(I_n)} \geq \frac{1}{4} l(E) \qquad (*)$$

since $l(E \cap I_n) \geq \frac{l(E)}{(q_n+q_{n-1})^2} \geq \frac{l(E)}{(q_n+q_{n-1})^2} \underbrace{(l(I_n)q_n)^2}_{\leq 1} \geq l(E)l(I_n)/4$.

Given $\epsilon > 0$ we can cover $E^c$ by intervals $\cup_n I_n$ of the above form such that

$$\sum_n l(I_n) \geq l(E^c) \geq \sum_n l(I_n) - \epsilon.$$

Then

$$\begin{aligned}
l(E)l(E^c) &\leq l(E)\sum_n l(I_n) \leq 8\sum_n l(E \cap I_n) \qquad \text{(by (*))} \\
&= 8\left(\sum_n l(E \cap I_n) + l(\underbrace{E^c \cap (\cup_n I_n)}_{=E^c}) - l(E^c)\right) \\
&\leq 8\left(\sum_n l(E \cap I_n) + \sum_n l(E^c \cap I_n) - l(E^c)\right) \\
&= \left(\sum_n l(I_n) - l(E^c)\right) \leq 8\epsilon.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, either $l(E) = 0$ or $l(E^c) = 0$, i.e., $\mu(E) = 0$ or $\mu(E^c) = 0$.

□

By applying the Birkhoff ergodic theorem we have the following.

**Theorem 8.5.** *For a.e. ($\mu$) $x \in (0,1)$,*

1. *the limit of the geometric averages satisfies*

$$\lim_{n\to+\infty}(a_1 \cdots a_n)^{1/n} = \prod_{k=1}^{\infty}\left(1 + \frac{1}{k^2+2k}\right)^{\frac{\log k}{\log 2}} ; \textit{and}$$

*2. the growth rate of the $q_n$ is*

$$\lim_{n\to+\infty}\frac{\log q_n}{n}=\frac{\pi^2}{12\log 2}$$

*Proof.* 1. Let $f(x)=\log[1/y]$. By the Birkhoff ergodic theorem:

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1}f(T^kx) &= \frac{1}{n}\sum_{k=1}^{n}f(T^{k-1}x)\\ &= \frac{1}{n}\sum_{k=1}^{n}\log\left[\frac{1}{T^{k-1}(x)}\right]\\ &= \frac{1}{n}\sum_{k=1}^{n}\log a_k\\ &\to \frac{1}{\log 2}\int_{\frac{1}{k+1}}^{\frac{1}{k}}\frac{\log k}{1+x}dx\\ &= \sum_{k=1}^{\infty}\frac{\log k}{\log 2}\log\left(1+\frac{1}{k^2+2k}\right)\end{aligned}$$

where $f\in L^1(X,m)$ we have that $\int f d\mu \leq \frac{1}{\log 2}\sum_{k=1}^{\infty}\frac{\log k}{k(k+1)}$. Taking exponentials gives the result.

2. Since we can write

$$\begin{aligned}x=\frac{p_n(x)}{q_n(x)} &= [a_1,\cdots,a_n]\\ &= \frac{1}{1+[a_2,\cdots,a_n]}\\ &= \frac{1}{1+\frac{p_{n-1}(Tx)}{q_{n-1}(Tx)}}\\ &= \frac{q_{n-1}(Tx)}{p_{n-1}(Tx)+q_{n-1}(Tx)}\end{aligned}$$

we deduce that $p_n(x)=q_{n-1}(Tx)$, by comparing numerators. In particular,

$$\frac{p_n(x)}{q_n(x)}\frac{p_{n-1}(Tx)}{q_{n-1}(Tx)}\cdots\frac{p_1(T^{n-1}x)}{q_1(T^{n-1}x)}=\frac{1}{q_n(x)}$$

and so

$$\frac{1}{n}\log q_n(x)=-\frac{1}{n}\sum_{k=0}^{n-1}\log\left(\frac{p_{n-k}}{q_{n-k}}(T^kx)\right).$$

If $f(x)=\log x$ then

$$\frac{1}{n}\log q_n(x)=-\frac{1}{n}\sum_{k=0}^{n-1}f(T^kx)+\underbrace{\frac{1}{n}\left(\sum_{k=0}^{n-1}\log(T^kx)-\log\left[\frac{p_{n-k}}{q_{n-k}}(T^kx)\right]\right)}_{(*)}.$$

By the Birkhoff ergodic theorem and the expansion $\log(1+x) = \sum_{k=0}^{\infty}(-1)^k \frac{(-1)^k}{k+1}$

$$\begin{aligned}\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) &= \frac{1}{\log 2}\int_0^1 \frac{\log x}{1+x}dx \\ &= -\frac{1}{\log 2}\int_0^1 \frac{\log(1+x)}{x}dx \text{ (Integrating by parts)} \\ &= -\frac{1}{\log 2}\sum_{k=0}^{\infty}(-1)^k \int_0^1 \frac{x^k}{k+1}dx \text{ (by expanding the logarithm)} \\ &= -\frac{1}{\log 2}\sum_{k=0}^{\infty}\frac{(-1)^k}{(k+1)^2} = -\frac{\pi^2}{12\log 2}.\end{aligned}$$

It still remains to show that the contribution from (*) is trivial:

$$\sum_{k=0}^{n-1}\left|\log\left(\frac{T^k x}{\left[\frac{p_{n-k}}{q_{n-k}}(T^k x\right]}\right)\right| \le \sum_{k=0}^{n-1}\left|\frac{T^k x)}{\left[\frac{p_{n-k}}{q_{n-k}}(T^k x)\right]} - 1\right|$$

since $|\log y| \le |y-1|$. Moreover,

$$\left|\frac{x}{p_k/q_k} - 1\right| = \frac{q_k}{p_k}\left|x - \frac{p_k}{q_k}\right| \le \frac{1}{p_k q_{k+1}} \le \frac{1}{2^{k-1}}$$

giving a bound

$$\sum_{k=0}^{n-1}\left|\frac{T^k x}{\left[\frac{p_{n-k}}{q_{n-k}}(T^k x)\right]} - 1\right| \le \sum_{k=0}^{n-1}\frac{1}{2^{n-k-1}} \le 1.$$

□

# 9 Markov Operators

Let $T_1, \cdots, T_n : X \to X$ be measure preserving transformations with respect to a probability measure $\mu$ (i.e., for all $B \in \mathcal{B}$ we have that $\mu(T_i^{-1}B) = \mu(B)$). Let $(p_1, \cdots, p_n)$ be a probability vector (with $p_1 + \cdots + p_n = 1$). We define a linear operator, $T : L^1(X,\mu) \to L^1(X,\mu)$ by

$$Tf(x) = \sum_{i=1}^{n} p_i f(T_i x), \forall f \in L^1(X,\mu), a.e.(\mu) x \in X.$$

**Definition 9.1.** *We call* $T : L^1(X,\mu) \to L^1(X,\mu)$ *a* Dunford-Schwartz operator *if*

$$\|Tf\|_1 \le \|f\|_1 \text{ and } \|Tf\|_\infty \le \|f\|_\infty$$

*In addition, we say*

1. $T$ *is* positive *if* $f_1(x) \le f_2(x)$ *implies that* $Tf_1(x) \le Tf_2(x)$ *a.e.* $(\mu)$

*2. $T$ is* Markov *if $T1 = 1$ where 1 is a constant function.*

*Finally, we say that $T$ is* ergodic *if $T\chi_A = \chi_A$ where $A \in \mathcal{A}$ implies that $\mu(A) = 0$ or 1.*

**Definition 9.2.** *We say that the family $\{T_1, \cdots, T_n\}$ is ergodic whenever $A \in \mathcal{B}$ satisfies $T_i^{-1}A = A$, $i = 1, \cdots, n$ then $\mu(A) = 0$ or $\mu(A) = 1$.*

If $T\chi_A(x) = \chi_A(x)$ then by convexity $T_i^{-1}A = A$ for $i = 1, \cdots, n$ and thus $\mu(A) = 0$ or 1, i.e., the operator $T$ is ergodic.

**Theorem 9.3** (Hopf-Dunford-Schwartz)**.** *If a Dunford-Schwartz operator $T : L^1(X, \mu) \to L^1(X, \mu)$ is ergodic then*

$$\frac{1}{n}\sum_{k=0}^{n-1} T^k\phi(x) \to \int \phi d\mu, \quad a.e.(\mu)x \in X$$

*and $\phi \in L^1(X, \mu)$.*

*Remark* 9.4. In particular, this subsumes the Birkhoff Ergodic Theorem by letting $n = 1$ and $p = (1)$.

The proof is slightly reminsicent of that for the von Neumann ergodic theorem. We begin with the following simple lemma (whose proof we omit).

**Lemma 9.5.** *Let $Fix(T) = \{g \in L^\infty(X, \mu) : Tg = g\}$ then $B = Fix(T) \oplus (I - T)L^1(X, \mu) \subset L^1(X, \mu)$ is dense (in the $L^1$-norm).*

We use the following notation $S_n\phi(x) = \sum_{k=0}^{n-1} T^k\phi(x)$, $n \geq 1$.

**Lemma 9.6.** *It $\phi \in B$ then $\frac{1}{n}S_n\phi$ converges in $\|\cdot\|_\infty$.*

*Proof.* Let $f = g+(I-T)h$ with $g \in \mathrm{Fix}(T)$ and $h \in L^\infty(X, \mu)$. Then $\frac{1}{n}S_n\phi = g+\frac{1}{n}(h-T^nh)$ and

$$\|\frac{1}{n}S_n\phi - g\|_\infty \leq \frac{\|h\|_\infty + \|T^nh\|_\infty}{n}$$

as $n \to +\infty$. □

Moreover, since $T$ is ergodic we see that $\mathrm{Fix}(T)$ consists of constant functions (i.e., $\phi \in B$ implies $\frac{1}{n}S_n\phi \to \int \phi d\mu$).

**Aim:** We want to extend the convergence to the $L^1$-closure of $B$ Then the ergodic theorem follows by Lemma **??**).

**Definition 9.7.** *We associate a maximal "operator" for $f \in L^1(X, \mu)$ by*

$$Mf(x) = \sup_n \left|\frac{1}{n}S_nf(x)\right|$$

**Lemma 9.8.** *1. $Mf \geq 0$ for $f \in L^1(X, \mu)$*

*2. $M(\alpha f) = |\alpha|M(f)$ for $f \in L^1(X, \mu)$, $\alpha \in \mathbb{C}$*

*3.* $M(f+g) \leq M(f) + M(g)$ *for* $f, g \in L^1(X,\mu)$

These follow from the definition of $M$.

**Definition 9.9.** *We say that* $(\frac{1}{n}S_n f)$ *satisfies a maximal inequality if*

$$\mu\{x : Mf(x) > \lambda\} \leq \frac{\|f\|_1}{\lambda} \tag{M}$$

*for all* $\lambda > 0$ *and* $f \in L^1(X,\mu)$.

To achieve our aim we use the following.

**Lemma 9.10** (Banach's Principle)**.** *Assuming (M) holds we have that*

$$C := \left\{ f \in L^1(X,\mu) : \left(\frac{S_n f}{n}\right)_{n=1}^{\infty} \text{ converges a.e.}(\mu) \right\} \subset L^1(X,\mu)$$

*is a* $\|\cdot\|_1$*-closed subspace.*

*Proof.* To see that $C$ is closed: Let $f \in \overline{C} \subset L^1(X,\mu)$ and $\epsilon > 0$. Choose $g \in C$ with $\|g-f\|_1 < \epsilon$. By the triangle inequality: For a.e. $(\mu)$ $x \in X$:

$$\begin{aligned}\left|\frac{1}{k}S_k f(x) - \frac{1}{l}S_l f(x)\right| &\leq \underbrace{\left|\frac{1}{k}S_k f(x) - \frac{1}{k}S_k g(x)\right|}_{\leq M(f-g)(x)} \\ &+ \underbrace{\left|\frac{1}{k}S_k g(x) - \frac{1}{l}S_l g(x)\right|}_{\to 0 \text{ as } k,l\to+\infty} + \underbrace{\left|\frac{1}{l}S_l g(x) - \frac{1}{l}S_l f(x)\right|}_{\leq M(f-g)(x)}\end{aligned}$$

Thus

$$h(x) = \limsup_{k,l\to+\infty} \left|\frac{1}{k}S_k f(x) - \frac{1}{l}S_l f(x)\right| \leq 2M(f-g)(x).$$

Fix $\lambda > 0$ then

$$\begin{aligned}\mu\{x : h(x) > 2\lambda\} &\leq \mu\{x : M(f-g)(x) > 2\lambda\} \\ &\leq \frac{\|f-g\|_1}{\lambda} \qquad \text{(by (M))} \\ &\leq \frac{\epsilon}{\lambda}\end{aligned}$$

Since $\epsilon > 0$ was arbitrary $\mu\{x : h(x) > \lambda\} = 0$. Since $\lambda > 0$ was arbitrary $\mu\{x : h(x) > 0\} = 0$, i.e., $h = 0$ a.e. $(\mu)$. In particular, $f \in C$ (i.e., we deduce $C$ is closed). □

Recall that the Theorem follows from Lemma **??**, Lemma **??** and Lemma **??**.
It remains to show (M) holds. The following is the key.

**Definition 9.11.** *Given* $f \in L^1(X,\mu)$ *we define* $M_n f(x) = \max_{1\leq k\leq n}\{S_k f(x)\}$, $n \geq 1$ *for a.e.* $(\mu)$ $x \in X$.

**Lemma 9.12** (Hopf)**.** *We have*

$$\int_{\{x\ :\ M_n f(x) \geq 0\}} f d\mu \geq 0$$

*Proof.* By definition, for $k = 2, 3, \cdots, n$:

$$S_{k-1} f(x) \leq M_n f(x) \leq \underbrace{\max\{M_n f(x), 0\}}_{=:(M_n f)_+(x)} \tag{$*$}$$

In particular, for $k = 2, \cdots, n$,

$$\begin{aligned} S_k f(x) &= f(x) + T S_{k-1} f(x) \\ &\leq f(x) + T M_n f(x) \text{ (Since } T \text{ is positive and (*))} \\ &\qquad \text{and} \\ S_1 f(x) = f(x) &\leq f(x) + \underbrace{T M_n f(x)}_{\geq 0} \text{ (Since } (M_n f)_+ \geq 0 \text{ and } T \text{ is positive)} \end{aligned}$$

and thus

$$M_n f(x) := \max_{1 \leq k \leq n} S_k f(x) \leq f(x) + T(M_n f)_+(x) \tag{$**$}$$

Therefore,

$$\begin{aligned} \int (M_n f)_+ d\mu &= \int_{\{x\ :\ M_n f \geq 0\}} M_n f d\mu \text{ (by defintion of } (M_n f)_+) \\ &\leq \int_{\{x\ :\ M_n f \geq 0\}} f d\mu + \underbrace{\int_{\{x\ :\ M_n f \geq 0\}} T M_n f d\mu}_{\leq \int (M_n f)_+ d\mu} \text{ (since } T \text{ is positive and contracts)} \end{aligned}$$

Cancelling the first and last terms gives $\int_{\{x\ :\ M_n f \geq 0\}} f d\mu \geq 0$, as required. □

Finally, this leads to the following

**Corollary 9.13.** *Let* $A_n f(x) = \max_{1 \leq k \leq n}\{\frac{S_k f(x)}{k}\}$, $n \geq 1$, $f \in L^1(X, \mu)$ *the for any* $\lambda > 0$

$$\mu\{x : A_n f(x) > \lambda\} \leq \frac{\|f\|_1}{\lambda}$$

(i.e, a "finite dimensional" version of (M))

*Proof.* We can apply the Hopf Lemma (Lemma **??**) to $g = \overline{f} - \lambda 1$ to get

$$\int_{\{x\ :\ M_n g > 0\}} (\overline{f} - \lambda 1) d\mu \geq 0 \tag{$*$}$$

since $\{x : M_n f = 0\} \subset \{x : g(x) \leq 0\}$ by definition of $M_n g$ (and in the set integrated over $\geq$ is replaced by $>$).

Let $1 \leq k \leq n$. Let $0 \leq j \leq k-1$ then

$$T^j g = T^j f - \lambda T^j 1 = T^j f - \lambda 1 \text{ } (T1 = 1, \text{ Markov Property})$$

Averaging over $0 \leq j \leq k-1$ we have:

$$\frac{1}{k} S_k g = \frac{1}{k} S_k f - \lambda 1.$$

Taking the supremum over $1 \leq k \leq n$ gives

$$A_n g = A_n f - \lambda \tag{**}$$

and thus

$$\{x : A_n f(x) > \lambda\} \subset \{x : A_n g(x) > 0\} = \{x : M_n f(x) > \lambda\} \tag{***}$$

by (**) and comparing the definitions of $A_n g$ and $M_n g$. Thus

$$\begin{aligned} 0 &\leq \int_{\{x \,:\, M_n g(x) > 0\}} (f - \lambda 1) d\mu \text{ by (*)} \\ &= \int_{\{x \,:\, A_n g(x) > 0\}} (f - \lambda 1) d\mu + \underbrace{\int_{\{x \,:\, M_n g(x) > 0, A_n f(x) \leq \lambda\}} (f - \lambda 1) d\mu}_{\leq 0 \text{ since } f \leq A_n f} \text{ using (***)} \end{aligned}$$

Thus

$$\mu\{x : A_n f(x) > \lambda\} \leq \frac{1}{\lambda} \int_{\{x \,:\, A_n f(x) > \lambda\}} f d\mu \leq \frac{\|f\|_1}{\lambda}$$

as required. □

We can now prove that (M) holds. For $f \in L^1(X, \mu)$ we have that

$$\begin{aligned} \mu\{x : \max_{1 \leq k \leq n} \left\| \frac{1}{k} S_k f(x) \right\| > \lambda\} &\leq \mu\{x : \max_{1 \leq k \leq n} \left\| \frac{1}{k} S_k f(x) \right\| > \lambda\} \text{ (since } T \text{ positive)} \\ &\leq \frac{\|f\|_1}{\lambda} \end{aligned}$$

Letting $n \to +\infty$ then

$$\max_{1 \leq k \leq n} \left\{ \frac{1}{k} S_k f(x) \right\} \nearrow \underbrace{\sup_{k \geq 1} \left\{ \frac{1}{k} S_k f(x) \right\}}_{=: Mf(x)}$$

Thus $\mu\{x : Mf(x) > \lambda\} \leq \frac{\|f\|_1}{\lambda}$ (i.e., (M) holds as required).

## 9.1 Application to Ergodic Theorems Free groups

There are also some interesting variations of the ergodic theorems to other groups. Recently, there was a nice proof of an ergodic theorem for the free groups. More precisely, assume that $T_1, \cdots, T_d : X \to X$ are each (invertible) measure preserving transformations that generate a free group, that is there is no relationship between the generators, i.e., some combination $T_{i_1}^{n_1} \cdots T_{i_k}^{n_k}$ with $i_1, \ldots, i_k \in \{1, \ldots, d\}$ can only be the identity map when $n_1 = \ldots = n_k = 0$. Unless your choice of transformations is unlucky, this is what one would expect.

**Definition 9.14.** *We say that the action is ergodic if the only sets $B \in B$ satisfying $T_i^{-1}B = B$, for every $1 \le i \le d$ have either $\mu(B) = 0$ or $\mu(X - B) = 0$.*

For $f \in L^1(X)$ and $x \in X$ we can write

$$\sigma_n f(x) = \frac{\sum_{|n_1|+\cdots+|n_k|=n} f(T_{i_1}^{n_1} \circ \cdots \circ T_{i_k}^{n_k} x)}{\#\{(n_1, \cdots, n_k) : |n_1| + \cdots + |n_k| = n\}},$$

for the average over points the "shell" corresponding to acting on $x$ a total of $n$ times by different combinations of $T_1, \ldots, T_d$. It is now natural to consider the averages of these terms. In the case that $(X, B, \mu)$ is a probability space, we have the following ergodic theorem.

**Theorem 9.15** (Nevo-Stein, Bufetov)**.** *Let $m$ be an ergodic measure then, for any $f \in L^1(X)$, we have*

$$\frac{1}{N}\sum_{n=0}^{N-1} \sigma_n f(x) \to \int f(x)dm(x), \text{ as } n \to +\infty, \text{ a.e.}$$

We can observe that $\#\{(n_1, \cdots, n_k) : |n_1| + \cdots + |n_k| = n\} = 2k(2k-1)^{n-1}$ (The $L^2$ version of this was proved much earlier by Guivarc'h). There is also an application of Rota's theorem

*Proof.* Consider the operator $\widehat{T} : (L^1(X, \mu))^k \to (L^1(X, \mu))^k$ defined by

$$\widehat{T}(f_1, \cdots, f_{2k}(x))_j = \frac{1}{(2k-1)} \sum_{T_i^{-1} \neq T_j} f_i(T_j x).$$

□

# 10 Lyapunov exponents and Oseledets' Theorem

Let $F : X \to SL(2, \mathbb{R})$ be a measurable tranformation.

**Theorem 10.1** (Oseledets)**.** *Assume that $\log^+ \|F(x)\| \in L^1(X, \mu)$. For almost all $x \in X$ there exist*

1. *measurable functions $\lambda_1, \lambda_2 : X \to \mathbb{R}$ $\lambda_1(x) \ge \lambda_2(x)$*
2. *a measurable splitting $\mathbb{R}^2 = E_1(x) \oplus E_2(x)$*

*such that*

$$\lim_{n\to+\infty} \frac{1}{n} \log \|F(x)^n v_i\| = \lambda_i(x)$$

*for all non-zero $v_i \in E_i(x)$.*

## 10.1 Automorphisms of the disk

Let Mod($\mathbb{D}$) be the set of all maps $f : \mathbb{D} \to \mathbb{D}$ on $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ of the form

$$\phi_{a,b}(z) = \frac{az + b}{bz - \overline{a}}$$

where $|a|^2 - |b|^2 = 1$.

**Definition 10.2.** *We define the Poincaré metric on $\mathbb{D}$ by $d(z_1, z_2) = 2\tanh^{-1} \frac{|z_1 - z_2|}{|1 - z_1 \overline{z}_2|}$. Equivalently, we can write this as*

$$ds^2 = \frac{dx^2 + dy^2}{(1 - (x^2 + y^2))^2}.$$

**Lemma 10.3.** *The metric is invariant under maps $\phi \in Mod(\mathbb{D})$, i.e., $d(z_1, z_2) = d(\phi z_1, \phi z_2)$.*

By a simple calculation (which we omit):

**Lemma 10.4.** *For $f \in Mod(\mathbb{D})$ and $\xi \in K = \{\xi \in \mathbb{C} : |\xi| = 1\}$ we have that*

$$|f'(\xi)| = \frac{1 - |f(0)|^2}{|\xi - f^{-1}(0)|^2}.$$

We can relate matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{R})$ to maps $\phi \in$ Mod($\mathbb{D}$) by $\phi_A(z) = h \circ f_A \circ h^{-1}$ where

$$f_A(z) = \frac{az + b}{cz + d} \text{ and } h(z) = \frac{i - z}{i + z}.$$

**Lemma 10.5.** *We have $f_A : \mathbb{D} \to \mathbb{D}$ is a biijection.*

Let $A(0), A(1) \in SL(2, \mathbb{R})$. We can consider $x = (x_n) \in \{0, 1\}^{\mathbb{Z}^+}$ and for $n \geq 1$ we can write

$$f_{A^n(x)}(\xi) = (f_{A(x_0)} \circ f_{A(x_1)} \circ \cdots \circ f_{A(x_{n-1})})(\xi)$$

We can apply the subadditive ergodic theorem to $\phi_n(x) = d(f_{A^n(x)}0, 0)$, the displacement of $0 \in \mathbb{D}$. This uses the invariance of the Poincaré metric and then by the triangle inequality to write

$$\phi_{n+m}(x) \leq \phi_m(x) + \phi_n(\sigma^m x), \quad n, m \geq 1$$

In particular, there exists $\lambda \in \mathbb{R}$ such that for a.e. $(\mu)$ $x \in X$

$$\lim_n \frac{1}{n} \phi_n(x) = 2\lambda$$

We can concentrate on the case $\lambda > 0$. In particular, from the definition of the Poincaré metric

$$\lim_n \frac{1}{n} \log(1 - |f^{-1}_{A^n(x)}(0)|) = -2\lambda,$$

i.e., the Euclidean distance of $f^{-1}_{A^n(x)}(0)$ from the unit circle $K$ tends to zero. Moreover, we have the following:

**Lemma 10.6.** *For a.e. $(\mu)$ $x \in X$*

$$\lim_n \frac{1}{n} \log |f^{-1}_{A^{n+1}(x)}(0) - f^{-1}_{A^n(x)}(0)| \leq -2\lambda.$$

In particular, this shows that $|f^{-1}_{A^{n+1}(x)}(0)|$ is a Cauchy sequence in the Euclidean metric for a.e. $(\mu)$ $x \in X$. Let $w^s(x) \in K$ be the limit point on the unit circle in the Euclidean metric which can also be seen as a limit of the conformal action on $K$.

If $A \in SL(2, \mathbb{R})$ has a maximal eigenvalue $\lambda$ and eigenvector $v = v_A$ then let $s(A) \in K$ be the corresponding point on $K$.

**Lemma 10.7.** $s(A) = f_A^{-1}(0)/|f_A^{-1}(0)|$*and* $\lambda = |f'_A(s(A))|^{-1/2}$.

We denote by $u(A) \in K$ the direction corresponding to the eigenvector of the smaller eigenvalue.

We can write

$$w^s(x) = \lim_{n \to +\infty} s(f_{A^n(x)})$$

In particular,

$$|(f_A^n)'(z)| = (1 + |w_n|)\frac{1 - |w_n|}{|1 - w_n|^2}$$

and thus

$$\lim_{n \to +\infty} \frac{1}{n} \log |(f_A^n)'(z)| = -2 - 2 \lim_{n \to +\infty} \frac{1}{n} \log |z - w_n|$$

For all $z \neq w^s(x)$ this gives that the limit is $-2\lambda$.

# 11 Spectral Theory

Consider the Hilbert space

$$L^2(X, \mu) = \{f : X \to \mathbb{R} : \underbrace{\int |f|^2 d\mu}_{=:\|f\|_2^2} < +\infty\}.$$

We can associate to a measure preserving transformation $T : X \to X$ a linear operator $U_T : L^2(X, \mu) \to L^2(X, \mu)$ defined by

$$U_T f(x) = f(Tx).$$

The following lemma is easy to prove:

**Lemma 11.1.** *The operator $U_T$ is an isometry (i.e., $\|U_T f\|_2 = \|f\|_2$ for all $f \in L^2(X, \mu)$).*

**Definition 11.2.** *We say $\lambda \in \mathbb{C}$ is an* eigenvalue *if there exists a non-zero $f \in L^2(X, \mu)$ such that $U_T f = \lambda f$.*

In particular, 1 is an eigenvalue corresponding to the eigenfunctions consisting of constant functions.

**Definition 11.3.** *We say that $T$ has* continuous spectrum *if and only if 1 is the only eigenvalue of $T$ and $T : X \to X$ is ergodic.*

*Proof.* We know that $U_T f = f \circ T = f$ if and only if $T$ is ergodic. □

## 11.1 Spectral theorem

Let us assume for convenience that $T : X \to X$ is invertible.

One of the main results on the spectrum of the operator $U_T$ is the following.

**Theorem 11.4** (Spectral Theorem). *Given $f \in L^2(X,\mu)$ there exists a probability measure $\mu_f$ on $K = \{z \in \mathbb{C} : |z| = 1\}$ such that*

$$\int U_T^n f.f d\mu = \int z^n d\mu_f(z), \forall n \in \mathbb{Z}.$$

If $\int f d\mu = 0$ and $T$ is ergodic (i.e., $U_T$ has continuous spectrum) then $\mu_f(\{z\}) = 0$ for all $z \in K$.

## 11.2 Ergodic Theorems for subsequences

We can reprove the mean ergodic theorem using the spectral theorem.

*Proof.* Assume that $\int f d\mu = 0$ then $\mu_f(\{1\}) = 0$. We can write

$$\begin{aligned}\frac{1}{n}\sum_{i=0}^{n-1}\int f \circ T^i.f d\mu &= \frac{1}{n}\sum_{i=0}^{n-1}\int U_T^i f f d\mu \\ &= \frac{1}{n}\sum_{i=0}^{n-1}\int z^i d\mu_f(z) \\ &= \frac{1}{n}\int \left(\frac{1-z^n}{1-z}\right) d\mu_f(z)\end{aligned}$$

We see that for $z \neq 1$ we have $\lim_{n\to+\infty}\left|\frac{1}{n}\frac{1-z^n}{1-z}\right| = 0$.

By the dominated convergence theorem we have that

$$\frac{1}{n}\sum_{i=0}^{n-1}\int f \circ T^i.f d\mu \to \mu_f(\{1\}) = 0.$$

□

We can consider the corresponding problem for ergodic theorems for the subsequence $\{n^2\}$. In particular, we can consider

$$\frac{1}{n}\sum_{i=0}^{n-1}\int f \circ T^{i^2}.f d\mu \to \mu_f(\{1\}).$$

We can write this as

$$\int_K \left(\frac{1}{n}\sum_{i=0}^{n-1} z^{i^2}\right) d\mu_f(z)$$

By uniform distribution of the sequence $\{\alpha n^2\}$ we have the following.

**Lemma 11.5** (Weyl)**.** *For $m \in \mathbb{Z} \setminus \{0\}$ and irrational $\alpha$ we have*

$$\frac{1}{n}\sum_{k=0}^{n-1} e^{2\pi i \alpha k^2} \to 0 \text{ as } n \to +\infty.$$

In particular, we can use the spectral theorem and the dominated convergence theorem to write

$$\frac{1}{n}\sum_{k} i = 0^{n-1} \int f \circ T^{k^2}.f d\mu = \frac{1}{n}\sum_{k=0}^{n-1} \int U_T^{k^2} f f d\mu$$
$$= \frac{1}{n}\sum_{k=0}^{n-1} \int z^{k^2} d\mu_f(z)$$

By the dominated convergence theorem we have that

$$\frac{1}{n}\sum_{k=0}^{n-1} \int f \circ T^{k^2}.f d\mu \to \mu_f(\{1\}) = 0.$$

## 11.3 Weak mixing

We begin with the following equivalent definition of ergodicity.

**Theorem 11.6.** *Let $T : X \to X$ preserve a probability measure $\mu$. $T$ is ergodic if and only if for all $A, B \in \mathcal{B}$ we have that*

$$\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) \to \mu(A)\mu(B), \quad \text{as } n \to +\infty \tag{$*$}$$

*Proof.* Assume that $\mu$ is ergodic then by the Birkhoff ergodic theorem (with $f = \chi_A$):

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_A(T^i x) \to \mu(A) \text{ a.e. } (\mu)\ x \in X$$

Thus

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_A(T^i x)\chi_B(x) \to \mu(A)\chi_B(x) \text{ a.e. } (\mu)\ x \in X$$

and integrating

$$\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) \to \mu(A)\mu(B), \quad \text{as } n \to +\infty$$

(by the dominated convergence theorem) as required.

Conversely, assume (*) holds and $T^{-1}A = A$. Let $B = A$ then by (*)

$$\mu(A) = \sum_{i=0}^{n-1} m(A \cap A) \to m(A)^2$$

and thus $\mu(A) = 0$ or 1.

□

Given a transformation $T : X \to X$, ergodicity of an invariant measure $m$ is only the first in a sequence of increasingly stronger possible properties.

**Definition 11.7.** *We say that a measure $m$ is weak mixing if for any $A, B \in \mathcal{B}$ we have that*

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

*This is different to ergodicity because of the $|\cdot|$ used on each term.*

The following is immediate

**Lemma 11.8.** *If $T$ is weak-mixing then it is ergodic.*

We also have the following equivalent formulation.

**Lemma 11.9.** *$T$ is weak mixing if and only if for all $f_1, f_2 \in L^2(X)$ we have that*

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n=0}^{N-1} |\int f_1 \circ T^n f_2 dm - \int f_1 dm \int f_2 dm| = 0.$$

*Proof.* If we assume that for any $f_1, f_2 \in L^2(X)$ we have that

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n=0}^{N-1} |\int f_1 \circ T^n f_2 dm - \int f_1 dm \int f_2 dm| = 0.$$

then we can take $f_1 = \chi_A$ and $f_2 = \chi_B$ to deduce that

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

For the reverse implication we can assume that

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

and then approximate $\|f_i - g_i\| < \epsilon$ for $i = 1, 2$ where $g_1, g_2 \in L^2(X, \mu)$ are simple functions. □

**Example 11.10** (Ergodic, but not weak mixing)**.** *Let $T : [0, 1) \to [0, 1)$ be an irrational rotation then $T$ is ergodic with respect to Lebesgue measure $\mu$. Let $A = B = [0, \frac{1}{2}]$ . Then by considering the function $f \in L^2[0, 1]$ defined by $f(x) = |x - \frac{1}{2}|$ we see that*

$$\mu(T^{-i}A \cap A) = f(T^{-i}0)$$

*In fact (by uniform distribution)*

$$\frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap A) - \mu(A)^2| = \frac{1}{n} \sum_{i=0}^{n-1} |f(T^{-i}0) - \frac{1}{4}| \to \frac{19}{32} \neq 0$$

There is an interesting equivalent formulation:

**Lemma 11.11.** *$T$ is weak mixing if and only if given $f_1, f_2 \in L^2(X)$ there exists a subsequence $J \subset \mathbb{Z}$ of zero density (i.e., $\lim_{n\to+\infty} \frac{1}{n} Card \{0 \le i \le n-1 : i \in J\} = 0$) of the natural numbers such that $\lim_{J \not\ni n \to +\infty} \int f_1 \circ T^{n_k} f_2 dm = \int f_1 dm \int f_2 dm$.*

This is an easy consequence of the following lemma (where $a_n = \int f_1 \circ T^n f_2 dm - \int f_1 dm \int f_2$).

**Lemma 11.12.** *Let $\{a_n\}$ be a bounded sequence of real numbers then the following are equivalent*

1. $\lim_{n\to+\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0$;
2. $\exists J \subset \mathbb{Z}^+$ *of density zero, i.e.,* $\lim_{n\to+\infty} \frac{1}{n} Card \{0 \le i \le n-1 : i \in J\} = 0$ *such that* $\lim_{n\to+\infty} a_n = 0$ *provided* $n \notin J$.

(The proof is an exercise, or can be found in Walters' book).

**Lemma 11.13.** *Let $T : X \to X$ preserve a $T$-invariant probability measure. The following are equivalent:*

1. *$T$ is weak mixing;*
2. *$T \times T$ is ergodic; and*
3. *$T \times T$ is weak mixing.*

(The proof is an exercise, or can be found in Walters' book).

**Theorem 11.14.** *Let $T : X \to X$ be an invertible measure preserving transformation. Then $T$ is weak-mixing if and only if $T$ has continuous spectrum.*

*Proof.* Assume that $T$ is weak mixing. Let $U_T f = \lambda f$ and assume $\lambda \neq 1$. Since $\int f d\mu = \int U_T f d\mu = \lambda \int f d\mu$ we deduce that $\int f d\mu = 0$. Thus, by the weak mixing property

$$\frac{1}{n}\sum_{i=0}^{n-1} \int f \circ T^i f d\mu = \frac{1}{n}\sum_{i=0}^{n-1} \int U_T^i f f d\mu = \frac{1}{n}\sum_{i=0}^{n-1} \lambda^i \int |f|^2 d\mu \to 0.$$

Moreover, since $\lambda \neq 1$ this gives that $\int |f|^2 d\mu = 0$, i.e., $f$ isn't an eigenfunction.

Assume that $\int f d\mu = 0$. Let $K$ denote the unit circle. By the spectral theorem

$$\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n-1}|\int f\circ T^i f d\mu|^2 &= \frac{1}{n}\sum_{i=0}^{n-1}|\int U_T^i f.f d\mu|^2 \\
&= \frac{1}{n}\sum_{i=0}^{n-1}|\int_K \lambda^i d\mu_f(\lambda)|^2 \\
&= \frac{1}{n}\sum_{i=0}^{n-1}\int_K \lambda^i d\mu_f(\lambda)\int_K \lambda^{-i} d\mu_f(\lambda) \\
&= \frac{1}{n}\sum_{i=0}^{n-1}\int_K\int_K (\lambda\overline{\tau})^i d(\mu_f\times\mu_f)(\lambda,\tau) \\
&= \frac{1}{n}\int_K\int_K\left(\frac{1-(\lambda\overline{\tau})^n}{1-(\lambda\overline{\tau})}\right) d(\mu_f\times\mu_f)(\lambda,\tau)
\end{aligned}$$

By bounded convergence (and since $(\mu_f\times\mu_f)(\{(x,x) : x\in K\}) = 0$) we deduce that the final expression tends to zero. By a previous result, $T$ is weak-mixing. □

Every weak-mixing measure is ergodic, but there are examples of ergodic measures which are not weak-mixing:

**Example 11.15.** *We saw that the irrational rotation (by $\beta$) was ergodic. However, it is not weak mixing. For example, if we choose $f_1 = f_2 = \chi_J$, for some small interval $J$ then we can never hope to find a subsequence satisfying (1) and (2) above.*

## 11.4 Strong mixing

**Definition 11.16.** *We say that a measure $m$ is strong mixing if for any $A, B\in\mathcal{B}$ we have that*

$$\lim_{N\to+\infty} m(T^{-N}A\cap B) = m(A)m(B)$$

**Lemma 11.17.** *$T : X\to X$ is strong mixing if and only if for any $f_1, f_2\in L^2(X)$ we have that*

$$\lim_{N\to+\infty}\int f_1\circ T^n f_2 dm = \int f_1 dm\int f_2 dm.$$

The proof is similar to that for weak mixing transformations.

The following is immediate

**Lemma 11.18.** *Every strong mixing transformation is weak mixing*

There are examples of weak mixing measures which are not strong mixing (e.g., certain types of interval exchange maps).

*Remark* 11.19 (Open Problem). One can generalize the notion of strong mixing to 3-mixing by considering bounded functions $f_1, f_2, f_3 : X\to\mathbb{R}$ and requiring that

$$\lim_{n_1,n_2,n_2-n_1\to\infty}\int f_1\circ T^{n_1} f_2\circ T^{n_2} f_3 dm = \int f_1 dm\int f_2 dm\int f_3 dm.$$

Although three mixing implies strong mixing, it is unknown if the converse it true. (In the analogous case of $\mathbb{Z}^2$ actions, Ledrappier showed this was not the case).

*Remark* 11.20 (Decay of correlations). The natural heuristic interpretation of strong mixing is that it measures how quickly the system approaches an equilibrium ( ... whatever that means). In some cases one would like to show that for nice enough $f_1, f_2$ the "correlation function" $\rho(n) = \int f_1 \circ T^n f_2 dm - \int f_1 dm \int f_2 dm$ tends to zero quickly (e.g., exponentially fast). Generally speaking, one expects that provided $\sigma^2 := \sum_{n=-\infty}^{\infty} \rho(n) < +\infty$ then the improvement of the Birkhoff ergodic theorem to the Central Limit Theorem holds. (Ideas of Gordin).

# 12 Bernoulli Automorphisms and K-transformations

Bernoulli transformations (the invertible case) Let us begin with an example before we give the definition!

**Definition 12.1** (Bernoulli shift). *Let $\Sigma = \prod_{n=-\infty}^{\infty} \{1, \ldots, k\}$ denote the space of all sequences $(x_n)_{n=-\infty}^{\infty}$ where $x_n \in \{1, \ldots, k\}$. There is a natural Tychonoff product topology and let $\mathcal{B}$ be the Borel sigma algebra. We can define the measure of an algebra of sets*

$$[i_r, \ldots, i_s] := \{(x_n)_{n=-\infty}^{\infty} : x_j = i_j, \text{ for } r \leq j \leq s\}$$

*by $m([i_r, \ldots, i_s]) = k^{-(s-r)}$. This extends uniquely to a measure $\mu$ on $B$ by the Kolmogorov extension theorem. The invertible transformation $T : \Sigma \to \Sigma$ defined by shifting a sequence one place to the left is called a Bernoulli shift*

A Bernoulli transformation $T : X \to X$ with respect to a measure $\mu$ is one that looks like a Bernoulli shift "up to isomorphism". More precisely, we call a map $\phi : X \to \Sigma$ an isomorphism if it is measurable; it has a measurable inverse $\phi^{-1} : \Sigma \to X$; the map $\phi$ preserves the measure (i.e., $\mu(\phi^{-1}(B)) = m(B)$, for all $B \in \mathcal{B}$), and $\phi$ commutes with the two transformations (i.e., $\phi \circ T = \sigma \circ \phi$). If we can find such an isomorphism, then we say that $T$ and $\sigma$ are isomorphic.

Finally, we call $T : X \to X$ Bernoulli if we can find a Bernoulli shift $\sigma : \Sigma \to \Sigma$ and an isomorphism between them.

It is easy to show (first with step functions, and then by approximation) that the Bernoulli shift $\sigma : \Sigma \to \Sigma$ is strong-mixing. It is then easy to use the isomorphism to show that Bernoulli transformations are also strong mixing.

**Example 12.2.** *Let $\phi_t : SM \to SM$ be the geodesic flow on a compact manifold of negative curvature. Let $B$ be the Borel sigma algebra and let $m$ be the (normalized) Liouville measure. If we fix $t = 1$ then the discrete transformation $T = \phi_{t=1}$ is Bernoulli. (This was shown by Ratner, for more general hyperbolic flows).*

## 12.1 K-automorphisms (the non-invertible case)

Since $T$ is measurable, we have that $T^{-1}\mathcal{B} \subset \mathcal{B}$. Similarly, we deduce that $\ldots \subset T^{-n}\mathcal{B} \subset \ldots \subset T^{-1}\mathcal{B} \subset \mathcal{B}$. The intersection $\cap_{k=0}^{\infty} T^{-k}\mathcal{B}$ consists of sets $B \in \mathcal{B}$ so that for any $n \geq 0$ we can find $B_n \in \mathcal{B}$ with $B = T^n B_n$.

We say that $T$ is a K-automorphism if $\cap_{k=0}^{\infty} T^{-k}B$ is the trivial sigma algebra, i.e., contains only $X$ and the empty set $\emptyset$.

K-automorphisms are also necessarily strong mixing.

**Example 12.3.** *Let $I = [0, 1)$ be the unit interval, let $B$ be the Borel sigma algebra and let $m$ be the usual Lebesgue measure. We let $T : I \to I$ be the "doubling map"*

$$T(x) = \begin{cases} 2x & \text{if } 0 \leq x < \frac{1}{2} \\ 2x - 1 & \text{if } \frac{1}{2} \leq x < 1 \end{cases}$$

*This is an example of a K-automorphism.*

# 13 Entropy

## 13.1 Entropy of partitions

Let $X$ be a set and let $\mathcal{B}$ be a sigma algebra.

**Definition 13.1.** *A partition $\alpha = \{A_1, \cdots, A_n\}$ is a (finite) collection of measurable sets $A_i \in \mathcal{B}$ such that*

1. *$X = A_1 \cup \cdots \cup A_n$; and*
2. *$A_i \cap A_j \neq \emptyset$ for $i \neq j$.*

We can define the entropy $H_\mu(\alpha)$ of a finite partition $\alpha$ of a probability space $(X, \mu)$.

**Definition 13.2.** *Let $\alpha = \{A_1, \cdots, A_k\}$ be a partition of $X$. We define the* entropy *by*

$$H_\mu(\alpha) = -\sum_{i=1}^{k} \mu(A_i) \log \mu(A_i).$$

**Definition 13.3.** *Let $\alpha = \{A_1, \cdots, A_k\}$ be a partition of $X$ and let $\beta = \{B_1, \cdots, B_l\}$ be a partition of $X$. We define the conditional entropy by*

$$H_\mu(\beta|\alpha) = -\sum_{i=1}^{k} \sum_{j=1}^{l} \mu(A_i \cap B_j) \log \left( \frac{\mu(A_i \cap B_j)}{\mu(A_i)} \right).$$

In particular, if we take $\alpha = \{X\}$ to be the trivial partition then we see that $H_\mu(\beta) = H_\mu(\beta|\{X\})$.

**Definition 13.4.** *Assume that $\alpha = \{A_i\}$ and $\beta = \{B_j\}$ are two partitions then we define $\alpha \vee \beta = \{A_i \cap B_j\}$.*

The following estimates are standard

**Proposition 13.5.** *We have the following.*

1. $H_\mu(\alpha \vee \beta) = H_\mu(\beta|\alpha) + H_\mu(\alpha)$.
2. $H_\mu(\beta|\alpha) \leq H_\mu(\beta)$.
3. $H_\mu(\alpha) = H_\mu(T^{-1}\alpha)$.

*Proof.* 1. We can then write

$$\begin{aligned} H_\mu(\alpha \vee \beta) &= -\sum_{i,j} \mu(A_i \cap B_j) \log \mu(A_i \cap B_j) \\ &= -\sum_{i,j} \mu(A_i \cap B_j) \log\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right) - \sum_{i,j} \mu(A_i \cap B_j) \log \mu(A_i) \\ &= H_\mu(\beta|\alpha) + H_\mu(\alpha) \end{aligned}$$

2. Since the function $f : (0,1) \to \mathbb{R}$ defined by $f(t) = -t \log t$ is concave we can write

$$\begin{aligned} H_\mu(\beta|\alpha) &= -\sum_{i=1}^{k} \sum_{j=1}^{l} \mu(A_i \cap B_j) \log\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right) \\ &= -\sum_{j=1}^{l} \left( \sum_{i=1}^{k} \mu(A_i) \underbrace{\frac{\mu(A_i \cap B_j)}{\mu(A_i)} \log\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right)}_{f\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right)} \right) \\ &\leq \sum_{j=1}^{l} f\left(\sum_{i=1}^{k} \mu(A_i) \frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right) \\ &= \sum_{j=1}^{l} f(\mu(B_j)) = H(\beta) \end{aligned}$$

3. Since $\mu(T^{-1}A_i) = \mu(A_i)$ we see that

$$H_\mu(T^{-1}\alpha) = -\sum_{i=1}^{k} \mu(T^{-1}A_i) \log \mu(T^{-1}A_i) = -\sum_{i=1}^{k} \mu(A_i) \log \mu(A_i). = H_\mu(\alpha).$$

□

We have the following corollary.

**Corollary 13.6.** $H_\mu(\alpha \vee \beta) \leq H_\mu(\alpha) + H_\mu(\beta)$

*Proof.* Combining inequalities 1 and 2 in the above proposition we have that

$$H_\mu(\alpha \vee \beta) = H_\mu(\beta|\alpha) + H_\mu(\alpha) \leq H_\mu(\beta) + H_\mu(\alpha)$$

□

We can define refinements by induction.

**Definition 13.7.** *We define* $\alpha^{(n)} := \vee_{i=0}^{(n-1)} T^{-i}\alpha$.

We then have the following.

**Lemma 13.8.** *We have that* $H_\mu(\alpha^{(n+m)}) \leq H_\mu(\alpha^{(n)}) + H_\mu(\alpha^{(m)})$.

*Proof.* We can write

$$\alpha^{(n+m)} = \bigvee_{k=0}^{n+m-1} T^{-k}\alpha = \bigvee_{k=0}^{n-1} T^{-k}\alpha \bigvee T^{-n}\left(\vee_{k=0}^{m-1} T^{-k}\alpha\right)$$

from which the result follows from the corollary and part 3 of the proposition. □

We say that a sequence $(a_n)_{n=1}^{\infty}$ is subadditive. if $a_{n+m} \leq a_n + a_m$ for $n, m \geq 1$.

**Lemma 13.9** (Subadditive Sequence Lemma)**.** *If* $(a_n)_{n=1}^{\infty}$ *is a subaddittive sequence then*

$$\lim_{n\to+\infty} \frac{a_n}{n} = l := \inf_{n\geq 1}\left\{\frac{a_n}{n}\right\}$$

*Proof.* Given $\epsilon > 0$, we choose $N > 0$ such that $\frac{a_N}{N} < l + \epsilon$. Then for any $n$ we can write $n = lN + r$ where $l \geq 0$ and $0 \leq r \leq N-1$. Then by subadditivity $a_n \leq l a_N + a_r$. We can then write

$$\frac{a_n}{n} \leq l\frac{a_N}{n} + \frac{a_r}{n} \leq \underbrace{\frac{a_N}{N + r/l}}_{\to \frac{a_N}{N} \text{ as } l\to+\infty} + \underbrace{\frac{\max_{0\leq r\leq N-1} a_r}{n}}_{\to 0 \text{ as } n\to+\infty} \leq l + 2\epsilon$$

for $l, n$ sufficiently large. Since $\epsilon > 0$ is arbitrary the result follows. □

We see from the above that $a_n = H_\mu(\alpha^{(n)})$ is subadditive. Applying the Subadditive Sequence Lemma gives Lemma 13.8.

**Definition 13.10.** *We define the* entropy *of the partition* $\alpha$ *with respect to the transformation* $T$ *and the measure* $\mu$ *by the limit*

$$h_\mu(T, \alpha) := \lim_{n\to+\infty} \frac{H_\mu(\alpha^{(n)})}{n},$$

*which exists by the last two lemmas.*

## 13.2 Examples

**Example 13.11** (Doubling map)**.** *Let* $T : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ *be the doubling map defined by* $Tx = 2x$ *(mod 1). Consider the partition* $\alpha = \left\{\left[0, \frac{1}{2}\right), \left[\frac{1}{2}, 1\right)\right\}$. *Let* $\mu$ *denote Lebesgue measure.*

*For each* $n$ *we see that* $\alpha^{(n)}$ *consists of intervals of the form* $\left[\frac{i}{2^{n+1}}, \frac{i+1}{2^{n+1}}\right)$ *for* $i = 0, \cdots, 2^{n+1} - 1$ *(each of which has Lebesgue measure* $\frac{1}{2^{n+2}}$*). Thus we have that*

$$H_\mu(\alpha^{(n)}) = -2^{n+1} \times \frac{1}{2^{n+1}} \log\left(\frac{1}{2^{n+1}}\right) = (n+1)\log 2.$$

*Thus*

$$h_\mu(T, \alpha) = \lim_{n\to+\infty} \frac{1}{n} H_\mu(\alpha^{(n)}) = \lim_{n\to+\infty} \frac{(n+1)}{n} \log 2 = \log 2.$$

## 13.3 Entropy for transformations

We define the entropy of $T : X \to X$ and a $T$-invariant probability measure $\mu$ as follows:

**Definition 13.12.** *We define the* entropy *of $T : X \to X$ with respect to $\mu$ by*

$$h_\mu(T) = \sup\{h_\mu(T, \alpha) : \alpha \text{ is a finite partition}\}$$

*Remark* 13.13. The entropy can be infinite. Consider, for example, the case of the continued fraction transformation. We can take the "Bernoulli measures" on the countably many intervals.

Let $T_1 : X_1 \to X_1$ preserve a probability measure $\mu_1$ (for a sigma algebra $\mathcal{A}_1$). Let $T_2 : X_2 \to X_2$ preserve a probability measure $\mu_2$ (for a sigma algebra $\mathcal{A}_2$).

We recall the following definition.

**Definition 13.14.** *We define an isomorphism $\pi : X_1 \to X_2$ to be a bijection such that $\pi, \pi^{-1}$ are measurable (i.e., $A \in \mathcal{B}_1$ iff $\pi(A) \in \mathcal{B}_2$); $T_2 \circ \pi = \pi \circ T_1$ and $\pi_* \mu_1 = \mu_2$.*

The main result is the following.

**Theorem 13.15** (Kolmogorov-Sinai)**.** *The entropy is an isomorphism invariant.*

*Proof.* If $\alpha_1 = \{A_1, \cdots, A_n\}$ is any partition for $X_1$ then $\pi(\alpha_1) = \{\pi(A_1), \cdots, \pi(A_n)\}$ is a partition for $X_2$. From the definitions $h_{\mu_1}(T_1, \alpha_1) = h_{\mu_2}(T_2, \pi(\alpha_1))$. In particular, we see that $h_{\mu_1}(T_1) \leq h_{\mu_1}(T_2)$. Conversely, if $\alpha_2 = \{A_1, \cdots, A_n\}$ is any partition for $X_2$ then $\pi^{-1}(\alpha_2) = \{\pi^{-1}(A_1), \cdots, \pi^{-1}(A_n)\}$ is a partition for $X_1$. From the definitions $h_{\mu_1}(T_1, \pi^{-1}\alpha_2) = h_{\mu_2}(T_2, \alpha_2)$. In particular, we see that $h_{\mu_1}(T_1) \geq h_{\mu_2}(T_1)$. Thus $h_{\mu_1}(T_1) = h_{\mu_2}(T_2)$, as required. ☐

## 13.4 Entropy from generators

In practice, it is enough to use special partitions called generators.

**Definition 13.16.** *We define a* generator *to be a finite partitions $\beta$ such that $\mathcal{B} = \vee_{k=0}^{\infty} T^{-k}\beta$. This means that for any $A \in \mathcal{B}$ and $\epsilon > 0$ there exists $N > 0$ and $B \in \vee_{k=0}^{N} T^{-k}\beta$ such that $\mu(A \Delta B) < \epsilon$ (where we use the notation $A \Delta B = (A \setminus B) \cup (B \setminus A)$).*

The main result is the following:

**Theorem 13.17** (Sinai)**.** *If $\alpha$ is a generator then $h_\mu(T) = h_\mu(T, \alpha)$.*

Let $\beta$ be a finite partition and let $\alpha$ be a generator then it suffices to show that $h_\mu(T, \beta) \leq h_\mu(T, \alpha)$. This is based on a few observations on the entropy of partitions.

**Definition 13.18.** *Given two partitions we say that $\alpha_1 < \alpha_2$ if the elements of $\alpha_1$ are unions of elements of $\alpha_2$ (and in particular $Card(\alpha_1) \leq Card(\alpha_2)$).*

**Lemma 13.19.** *If $\alpha_1 < \alpha_2$ then $H_\mu(\alpha_1) \leq H_\mu(\alpha_2)$.*

*Proof.* If we assume that $\alpha_1 < \alpha_2$ then we see that $\alpha_1 \vee \alpha_2 = \alpha_2$. Thus

$$H_\mu(\alpha_2) = H_\mu(\alpha_1 \vee \alpha_2) = H_\mu(\alpha_1) + \underbrace{H(\alpha_2|\alpha_1)}_{\geq 0}$$

and the result follows. □

**Lemma 13.20.** *We have the following:*

1. *Given partitions $\alpha, \beta, \gamma$ with $\alpha < \beta$ then $H(\alpha|\gamma) \leq H(\beta|\gamma)$*
2. *Given partitions $\alpha, \beta, \gamma$ with $\alpha < \beta$ then $H(\gamma|\alpha) \geq H(\gamma|\beta)$*
3. *Given partitions $\alpha$ and $\beta$ we have that $H(\alpha|\beta) = H(T^{-1}\alpha|T^{-1}\beta)$.*

*Proof.* For the first part we can write

$$H_\mu(\beta|\gamma) = H_\mu(\alpha \vee \beta|\gamma) = H_\mu(\alpha|\gamma) + \underbrace{H_\mu(\beta|\gamma \vee \alpha)}_{\geq 0} \geq H_\mu(\alpha|\gamma).$$

For the second part we again use that $f(t) = -t \log t$ is concave. Thus for each $A_i$ and $C_j$ we can use convexity and $\sum_k \frac{\mu(A_i \cap B_k)}{\mu(A_i)} = 1$ to write that

$$f\left(\sum_k \frac{\mu(A_i \cap B_k)}{\mu(A_i)} \frac{\mu(C_i \cap B_k)}{\mu(B_k)}\right) \geq \sum_k \frac{\mu(A_i \cap B_k)}{\mu(A_i)} f\left(\frac{\mu(C_j \cap B_k)}{\mu(B_k)}\right)$$

Since $\alpha < \beta$ we can write $A_i \in \alpha$ as unions of elements of $\beta$

$$\sum_k \frac{\mu(A_i \cap B_k)}{\mu(A_i)} \frac{\mu(C_j \cap B_k)}{\mu(B_k)} = \frac{\mu(A_i \cap C_j)}{\mu(A_i)}$$

and thus the previous line gives

$$-\left(\frac{\mu(A_i \cap C_j)}{\mu(A_i)}\right) \log\left(\frac{\mu(A_i \cap C_j)}{\mu(A_i)}\right) \geq -\sum_k \frac{\mu(A_i \cap B_k)}{\mu(A_i)} \left(\frac{\mu(C_j \cap B_k)}{\mu(B_k)}\right) \log\left(\frac{\mu(C_j \cap B_k)}{\mu(B_k)}\right).$$

Multiplying each side by $\mu(A_i)$ and summing gives

$$\begin{aligned}
\underbrace{-\sum_{ij} \mu(A_i \cap C_j) \log\left(\frac{\mu(A_i \cap C_j)}{\mu(A_i)}\right)}_{H_\mu(\gamma|\alpha)} &\geq -\sum_{i,j} \sum_k \mu(A_i \cap B_k) \left(\frac{\mu(C_j \cap B_k)}{\mu(B_k)}\right) \log\left(\frac{\mu(C_j \cap B_k)}{\mu(B_k)}\right) \\
&= -\sum_j \sum_k \mu(B_k) \left(\frac{\mu(C_j \cap B_k)}{\mu(B_k)}\right) \log\left(\frac{\mu(C_j \cap B_k)}{\mu(B_k)}\right) \\
&= \underbrace{-\sum_j \sum_k \mu(C_j \cap B_k) \log\left(\frac{\mu(C_j \cap B_k)}{\mu(B_k)}\right)}_{H_\mu(\gamma|\beta)}
\end{aligned}$$

For the last part

$$\begin{aligned} H_\mu(T^{-1}\alpha|T^{-1}\beta) &= -\sum_{i,j} \mu(T^{-1}A_i \cap T^{-1}B_j) \log\left(\frac{\mu(T^{-1}A_i \cap T^{-1}B_j)}{\mu(T^{-1}B_i)}\right) \\ &- \sum_{i,j} \mu(A_i \cap B_j) \log\left(\frac{\mu(A_i \cap B_j)}{\mu(B_i)}\right) \\ &= H_\mu(\alpha|\beta) \end{aligned}$$

by $T$-invariance of $\mu$. □

The proof of the theorem comes from the following:

*Proof.* (of theorem)

1. If $\alpha^{(n)} := \vee_{i=0}^{n-1} T^{-i}\alpha$, $\beta^{(n)} := \vee_{i=0}^{n-1} T^{-i}\beta$ then

$$\begin{aligned} H(\alpha^{(n)}|\beta^{(n)}) &\le \sum_{i=0}^{n-1} H(T^{-i}\alpha|\beta^{(n)}) \text{ (by Lemma 13.20.1)} \\ &\le \sum_{i=0}^{n-1} H(T^{-i}\alpha|T^{-i}\beta) \text{ (by Lemma 13.20.2)} \\ &\le nH(\alpha|\beta) \text{ (by Lemma 13.20.3)} \end{aligned}$$

Therefore, we have that

$$\frac{1}{n}H(\alpha^{(n)}) \le \frac{1}{n}H(\beta^{(n)}) + H(\alpha|\beta).$$

Moreover, if we replace $\beta$ by $\beta^{(N)} = \vee_{i=0}^{N-1} T^{-i}\beta$ we have that

$$\frac{1}{n}H(\alpha^{(n)}) \le \frac{1}{n}H(\beta^{(n+N)}) + H(\alpha|\beta^{(N)}).$$

2. Finally, $H(\alpha|\beta^{(N)}) \to 0$ as $N \to \infty$. [Exercise]

Putting all of this together we now see that

$$\begin{aligned} h_\mu(T, \alpha) &= \lim_{n\to+\infty} \frac{1}{n} H_\mu\left(\alpha^{(n)}\right) \\ &\le \lim_{n\to+\infty} \frac{1}{n} H_\mu\left(\beta^{(n+N)}\right) + H(\alpha|\beta^{(N)}) \\ &= \underbrace{h_\mu(T, \beta^{(N)})}_{h_\mu(T,\beta)} + \underbrace{H(\alpha|\beta^{(N)})}_{\to 0 \text{ as } N\to+\infty} \\ &= h_\mu(T, \beta) \end{aligned}$$

which proves the theorem. □

**Example 13.21** (Doubling map)**.** *We saw before that for* $\alpha = \{[0, \frac{1}{2}), [\frac{1}{2}, 1]\}$ *we have that* $h_\mu(T, \alpha) = \log 2$*. However, we also see that* $\alpha^{(n)} = \vee_{i=0}^{n-1} T^{-i}\alpha$ *consists of intervals of the form* $[\frac{i}{2^n}, \frac{i+1}{2^n}]$*. We therefore see that* $\alpha$ *is a generator. This* $h_\mu(T) = h_\mu(T, \alpha) = \log 2$*.*

## 13.5 Examples

**Example 13.22** (Rotations)**.** *Let $\rho \in \mathbb{R}$. Let $T : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be the rotation defined by $Tx = x + \rho \pmod 1$. Consider the partition $\alpha = \left\{\left[0, \frac{1}{2}\right), \left[\frac{1}{2}, 1\right)\right\}$. Let $\mu$ denote Lebesgue measure.*

*If $\rho = \frac{p}{q}$ is rational then for all $n$ sufficiently large we have that $\alpha^{(n)} = \{[\frac{i}{q}, \frac{i+1}{q})\}$ and thus*

$$h_\mu(T) = \lim_{n\to+\infty} \frac{1}{n} H_\mu(\alpha^{(n)}) = 0$$

*If $\rho$ is irrational then since $T^{-1}\alpha$ generates $\mathcal{B}$ we see that $H\left(\alpha | \vee_{i=1}^{n-1} T^{-i}\alpha\right) \to 0$ as $n \to +\infty$ and then since we can write*

$$\frac{1}{n} H\left(\vee_{i=0}^{n-1} T^{-i}\alpha\right) = \frac{1}{n} H(\alpha) + \frac{1}{n} H\left(\alpha | \vee_{i=1}^{n-1} T^{-i}\alpha\right)$$

*we see that*

$$h_\mu(T, \alpha) = \lim_{n\to+\infty} \frac{1}{n} H\left(\vee_{i=0}^{n-1} T^{-i}\alpha\right) = 0.$$

*In fact, the same is true for any finite partition and so we deduce that $h_\mu(T) = 0$.*

*Remark* 13.23. In fact if $T : X \to X$ is any transformation preserving $\mu$ then developing the argument

$$h_\mu(T, \alpha) = \underbrace{\lim_{n\to+\infty} \frac{1}{n} H\left(\alpha | \vee_{i=1}^{n-1} T^{-i}\alpha\right)}_{=:H\left(\alpha | \vee_{i=1}^{\infty} T^{-i}\alpha\right)}$$

where we know that $H\left(\alpha | \vee_{i=1}^{\infty} T^{-i}\alpha\right) := \lim_{n\to+\infty} \frac{1}{n} H\left(\alpha | \vee_{i=1}^{n-1} T^{-i}\alpha\right)$ exists by the monotonicity

$$\frac{1}{n+1} H\left(\alpha | \vee_{i=1}^{n} T^{-i}\alpha\right) \le \frac{1}{n+1} H\left(\alpha | \vee_{i=1}^{n-1} T^{-i}\alpha\right) \le \frac{1}{n} H\left(\alpha | \vee_{i=1}^{n-1} T^{-i}\alpha\right)$$

*Remark* 13.24. It is an exercise to show that $h_\mu(T^k, \alpha) = k h_\mu(T, \alpha)$

**Example 13.25** (Entropy of Bernoulli shifts)**.** *Let $\sigma : \Sigma \to \Sigma$ be a full shift on $\Sigma = \{1, \cdots, k\}^{\mathbb{N}}$. Let $p = (p_1, \cdots, p_k)$ be a probability vector (i.e., $0 < p_i < 1$ and $p_1 + \cdots + p_k = 1$). We can let $\alpha = \{[i]_0 : i = 1, \cdots, k\}$ be the partition into $k$ cylinders of length 1. We then see that*

$$\alpha^{(n)} = \vee_{i=0}^{n-1} T^{-i}\alpha = \{[i_0, \cdots, i_{n-1}] : i_0, \cdots, i_{n-1} \in \{1, \cdots, k\}\}$$

*corresponds to the partition into cylinders of length $n$. We can then write*

$$\begin{aligned} H_\mu(\alpha^{(n)}) &= -\sum_{i_0,\cdots,i_{n-1}} \mu([i_0, \cdots, i_{n-1}]) \log \mu([i_0, \cdots, i_{n-1}]) \\ &= -\sum_{i_0,\cdots,i_{n-1}} \prod_{r=0}^{n-1} p_{i_r} \log\left(\prod_{r=0}^{n-1} p_{i_r}\right) \\ &= -n \sum_{i=1}^{k} p_i \log p_i \end{aligned}$$

*In particular, we can write*

$$h_\mu(\alpha) = \lim_{n\to+\infty} \frac{1}{n} H\left(\vee_{i=0}^{n-1} T^{-i}\alpha\right) = -\sum_{i=1}^{k} p_i \log p_i.$$

**Example 13.26** (Entropy of Markov shifts)**.** *Let $\sigma : \Sigma \to \Sigma$ be a subshift associated to a matrix $A$ with entries $0$ or $1$ then*

$$\Sigma = \{(x_n) \in \{1, \cdots, k\}^{\mathbb{Z}^+} : A(x_n, x_{n+1}) = 1, n \geq 0\}$$

*Let $P$ be a stochastic matrix (with $P(i,j) = 0$ iff $A(i,j) = 0$) We can let $\alpha = \{[i]_0 : i = 1, \cdots, k\}$ be the partition into $k$ cylinders of length $1$. We then see that*

$$\alpha^{(n)} = \vee_{i=0}^{n-1} T^{-i}\alpha = \{[i_0, \cdots, i_{n-1}] : i_0, \cdots, i_{n-1} \in \{1, \cdots, k\} \text{ and } A(i_j, i_{j+1}) = 1\}$$

*corresponds to the partition into cylinders of length $n$. Let $p = pP$ be the left eigenvector for $P$.*

*We can then write*

$$\begin{aligned} H_\mu(\alpha^{(n)}) &= -\sum_{i_0,\cdots,i_{n-1}} \mu([i_0, \cdots, i_{n-1}]) \log \mu([i_0, \cdots, i_{n-1}]) \\ &= -\sum_{i_0,\cdots,i_{n-1}} p_{i_0} \prod_{j=0}^{n-2} P(i_j, i_{j+1}) \log \left( p_{i_0} \prod_{j=0}^{n-2} P(i_j, i_{j+1}) \right) \\ &= -\sum_{i=1}^{k} \sum_{j=1}^{k} p_i \log P(i,j) \end{aligned}$$

**Example 13.27.** *The $\beta$-translation on the interval with respect to the Lebesgue measure $m$ has entropy $0$.*

**Example 13.28.** *If $T : X \to X$ has a fixed point $T(x_0) = x_0$ and $m = \delta_{x_0}$ is a Dirac measure, then $h(m) = 0$.*

## 13.6 Shannon-McMillan-Breiman Theorem

The entropy of a transformation and a measure has an interesting dynamical interpretation. Let $T : X \to X$ preserve an ergodic probability measure $\mu$.

For a.e. $(\mu)$ $x \in X$ we denote by $A_n(x) \in \alpha^{(n)}$ the element of the refinement containing $x$.

**Theorem 13.29** (Shannon-McMillan-Breiman Theorem)**.** *For a.e. $(\mu)$ $x \in X$ we have that*

$$\lim_{n\to+\infty} \frac{1}{n} \log \mu(A_n(x)) = -h_\mu(T, \mu).$$

*Remark* 13.30. If we only assume that $\mu$ is $T$-invariant then the limit still exists.

Let us denote $f_n(x) = \log\left(\frac{A_n(x)}{A_{n-1}(Tx)}\right)$. We can then write

$$\begin{aligned} &-\log\mu(A_n(x)) \\ &= -\log\left(\frac{A_n(x)A_{n-1}(Tx)\cdots A_1(T^{n-1}x)}{A_{n-1}(Tx)A_{n-2}(T^2x)\cdots A_1(T^{n-1}x)}\right) = -\sum_{j=0}^{n-2} f_{n-j}(T^jx) + \log\mu(A_1(T^{n-1}x)). \end{aligned}$$

We claim that $f(x) := \lim_{n\to+\infty} f_n(x)$ exists and lies in $L^1(X,\mu)$. (This requires the Increasing Martingale Theorem, so we omit the proof). We can the write

$$-\frac{1}{n}\log\mu(A_n(x)) = \underbrace{-\frac{1}{n}\sum_{j=0}^{n-1} f(T^jx)}_{\to\int f d\mu} - \frac{1}{n}\sum_{j=0}^{n-1}\left(f(T^jx) - f_{n-j}(T^jx)\right) + \frac{1}{n}\log\mu(A_1(T^{n-1}x))$$

Let us write $f^*(x) = \sup_{j\geq 1}\{f_j(x)\}$.

**Claim.** We claim that $\int |f^*| d\mu < +\infty$.

Let us define for any $m \geq 1$,

$$e_m(x) = \sup_{j\geq m} |f_j(x) - f(x)| \in L^1(X,\mu).$$

Thus

$$\left|\frac{1}{n}\sum_{j=0}^{n-1}\left(f(T^jx) - f_{n-j}(T^jx)\right)\right| \leq \underbrace{\frac{1}{n}\sum_{j=0}^{m-1}\left|f(T^jx) - f_{n-j}(T^jx)\right|}_{\to 0 \text{ as } n\to+\infty} + \underbrace{\frac{1}{n}\sum_{j=m}^{n-1}\left|e_m(T^jx)\right|}_{\to\int e_m d\mu \text{ as } n\to+\infty}.$$

Since $f_n(x) \to f(x)$ for a.e. $(\mu)$ we have by dominated convergence $\lim_{n\to+\infty}\int e_m d\mu = 0$. In particular, we see that

$$\limsup_{n\to+\infty}\left|\frac{1}{n}\sum_{j=0}^{n-1}\left(f(T^jx) - f_{n-j}(T^jx)\right)\right| \leq \int e_m d\mu \to 0 \text{ as } m \to +\infty.$$

This completes the proof (modulo the proof of the claim).

*Proof of Claim.* We want to show that $f^*(x)$ is integrable. Given a partition $\alpha = \{A_1, \cdots, A_n\}$ we can associate to each $1 \leq i \leq n$ a function:

$$f_n^{(i)}(x) = -\log\left(\frac{\mu(A_i \cap A^{(n-1)}(Tx))}{\mu(A^{(n-1)}(Tx))}\right) \in \vee_{k=1}^{n-1} T^{-k}\alpha.$$

Fix $\lambda > 0$. Let

$$B_i^{(n)} = \left\{x : n = \min\{m : f_m^{(i)}(x) > \lambda\}\right\} \in \vee_{k=1}^{n-1} T^{-k}\alpha.$$

Then

$$\mu(B_i^{(n)} \cap A_i) = \int_{B_i^n} \chi_{A_i} d\mu = \int_{B_i^{(n)}} \left( \frac{\mu(A_i \cap A^{(n-1)}(Tx))}{\mu(A^{(n-1)}(Tx))} \right) d\mu(x) \le e^{-\lambda} \mu(B_i^{(n)})$$

We can then write

$$\{x \in A_i : f^*(x) > \lambda\} = \cup_{n=0}^{\infty} B_i^{(n)} \cap A_i$$

In particular,

$$\mu(\{x \in A_i : f^*(x) > \lambda\}) = \sum_n \mu(B_i^{(n)} \cap A_i) \le e^{-\lambda} \sum_n \mu(B_i^{(n)}) \le e^{-\lambda}$$

Thus $\int_{A_i} f^* d\mu = \int_0^{\infty} \mu(\{x : f^*(x) > \lambda\}) d\lambda \le \int_0^{\infty} e^{-\lambda} d\lambda < +\infty$.

**Example 13.31.** *Let $\sigma : \Sigma \to \Sigma$ be a full shift with a Bernoulli probability measure $\mu$ associated to a probability vector $(p_1, \cdots, p_n)$. The measure of a cylinder is $\mu([i_0, \cdots, i_{n-1}]) = p_{i_0} p_{i_1} \cdots p_{i_{n-1}}$ and*

$$\frac{1}{n} \log \mu([i_0, \cdots, i_{n-1}]) = \frac{1}{n} (\log p_{i_0} + \log p_{i_1} + \cdots + \log p_{i_{n-1}}) \to \int_{\Sigma} \log p_{x_0} d\mu(x) = \sum_{i=1}^{k} p_i \log p_i$$

*for a.e. ($\mu$) $x$ by the Birkhoff Ergodic Theorem.*

## 13.7 Ornstein-Weiss Theorem

We can also consider problems related to return times.

Let $T : X \to X$ preserve an ergodic probability measure $\mu$. We let

$$R_n(x) = \min\{k \ge 1 : A_n(x) = A_n(T^k x)\}$$

**Theorem 13.32** (Ornstein-Weiss Theorem)**.** *For a.e. ($\mu$) $x \in X$ we have that*

$$\lim_{n \to +\infty} \frac{1}{n} \log R_n(x) = h_\mu(T, \mu).$$

**Example 13.33.** *Let $\sigma : \Sigma \to \Sigma$ be a full shift with a Bernoulli probability measure $\mu$ associated to a probability vector $(p_1, \cdots, p_n)$. The measure of a cylinder is $\mu([i_0, \cdots, i_{n-1}]) = p_{i_0} p_{i_1} \cdots p_{i_{n-1}}$*

*Then for a.e. ($\mu$) $x \in X$ we have that*

$$\lim_{n \to +\infty} \frac{1}{n} \log R_n(x) = -\sum_{i=1}^{k} p_i \log p_i$$

*Proof.* Let $\overline{R}(x) = \limsup_{n \to +\infty} \frac{1}{n} \log R_n(x)$.

Let $\Gamma(N) = \{x : -\log \mu([x]_n) \in (h_\mu(T) - \epsilon, h_\mu(T) + \epsilon), n \ge N\}$ Let $> 0$. Then for $n \ge N$:

$$\sum_{n \ge N} \mu\{x \in \Gamma(N) : R_n(x) \ge e^{-n(h_\mu(T)+\epsilon)}\} \le \sum_{n \ge N} \int_{\Gamma(N)} R_n(x) d\mu(x)$$

$$\le e^{-n(h_\mu(T)+\epsilon)} \sum_{|C|=n} \int_{C \cap \Gamma(N) \tau_C d\mu} \le e^{-\epsilon n}$$

□

## 13.8 Entropy and isomorphism

One of the major problems in ergodic theory before 1960 was the classification of ergodic transformations. The natural approach to classifying transformations is to say they are equivalent if they are isomorphic, since isomorphic systems share many of the same properties.

**Question 13.34.** *Given transformations $T_1 : X_1 \to X_1$ (with ergodic measure $m_1$) and $T_2 : X_2 \to X_2$ (with ergodic measure $m_2$) when are the isomorphic?*

The main break through in this area came from work of Kolmogorov and Sinai. They associated a very useful isomorphism invariant called *entropy* (i.e., any two isomorphic transformations necessarily share the same entropy). The entropy is a numerical invariant and, in particular, one can establish that two transformations are not isomorphic if they have different entropies.

The main property of the entropy is the following:

**Theorem 13.35.** *The entropy is an isomorphism invariant, i.e., if $T_1 : X_1 \to X_1$ (with ergodic measure $m_1$) and $T_2 : X_2 \to X_2$ (with ergodic measure $m_2$) then they are isomorphic.*

Of course, it is quite possible that two non-isomorphic transformations can have the same entropy, but not be isomorphic. However, there is a subclass of transformations we have already mentioned where the same entropy does imply they are isomorphic.

**Theorem 13.36** (Ornstein's Theorem)**.** *Entropy is a complete invariant for Bernoulli transformations.*

This remarkable result was proved by Ornstein in 1968, and his methods introduced a wealth of powerful tools, including the $\overline{d}$-metric on measures (a stronger metric than the weak star metric) and joinings.