

Ergodic Theory Lecture Notes - Jan-Mar. 2011

MA 427

March 7, 2011

1 A very condensed history of Ergodic Theory

1.1 Boltzmann and gas particles

The motivation for ergodic theory comes from theoretical physics and statistical mechanics. Boltzmann (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 particle 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 $x \in X \subset \mathbb{R}^{6 \cdot 10^{20}}$. Let $Tx \in X$ be the new configuration of the gas particles if we wait for time $t = 1$.

The Boltzmann 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, \dots spends a proportion of time in any subset $B \subset X$ proportional to the size of the set, i.e.

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

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

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

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 committed 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 simultaneous work of von Neumann (1903-1959) and Birkhoff (1884-1944). We have to restrict to measure preserving transformations $T : (X, \mu) \rightarrow (X, \mu)$. 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. (μ)). The easier L^2 result was published earlier than the stronger pointwise result. There are various extensions of these results where \mathbb{Z} or \mathbb{R} can be extended 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 $0 \leq x \leq 1$ with decimal expansion $x = .a_1 a_2 a_3 \dots$ (where $a_1, a_2, a_3, \dots \in \{0, 1, \dots, 9\}$) 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 $T : [0, 1) \rightarrow [0, 1)$ given by $Tx = bx \pmod{1}$.*

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

1.3 Mixing and spectral properties

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

Example 1.4 (Heuristic picture). *One can think of an incompressible fluid stirred in a container, e.g., martini in a glass. The set B represents the gin and $X \setminus B$ 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 $m(X \setminus B) = \frac{2}{3}$, 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 theory of *entropy*, a concept adapted from information theory. This turned out to be one of the most important invariants for the classification of invariant measures μ , 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 isomorphism invariant (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 Approximation (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.,*

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

then it contains arithmetic 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 $T : X \rightarrow X$ be a measurable transformation, i.e., if $A \in \mathcal{B}$ then $T^{-1}A \in \mathcal{B}$.

Definition 2.1. *We say that μ is T -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 μ be the usual Lebesgue measure. Given $0 \leq \alpha < 1$ define $Tx = x + \alpha \pmod{1}$, 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 $T : [0, 1) \rightarrow [0, 1)$ given by $Tx = x + \alpha \pmod{1}$:

1. If α is irrational then Lebesgue measure μ is the only invariant probability measure
2. If α is rational then there are more invariant probability measures

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

Definition 2.4. If $T : X \rightarrow X$ 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 digit of 2^n being equal to $l \in \{1, \dots, 9\}$ 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 μ be the Lebesgue measure. Let $b \geq 2$ be a natural number and let $Tx = bx \pmod{1}$, i.e.,

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

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 $[c, d) \subset [0, 1)$ the preimage

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

is a union of b intervals of length $(d - c)/b$. 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 δ_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 let λ be the Lebesgue measure. Let $Tx = \frac{1}{x} \pmod{1}$.

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

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

The Gauss measure is preserved by the Gauss transformation and is absolutely continuous with respect to the Lebesgue measure λ .

We will return to this

Example 2.10 (Bernoulli measures). Let $\Sigma = \{1, 2, \dots, k\}^{\mathbb{Z}}$ and let $T : \Sigma \rightarrow \Sigma$ be the full shift on k symbols given by $T(x_n) = (x_{n+1})$. We can consider the (cylinder) sets

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

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

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

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

Example 2.11. 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 $dm = "d(\text{area on } X) \times d(\text{Haar measure on } S^1)"$. 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 2.12. A probability measure μ 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., μ is T -invariant.

Conversely, if μ 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. \square

Corollary 2.13. *If $T : X \rightarrow X$ is a continuous map on a compact metric space. A probability measure μ 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, μ 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 $\int g \circ T d\mu \neq \int g d\mu$. \square

2.2 The existence of invariant measures

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

Definition 2.14. *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., $\mu_n \rightarrow \mu$) precisely when $\int f d\mu_n \rightarrow \int f d\mu$ for every $f \in C(X)$.*

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

Theorem 2.15 (Alaoglu). *For any sequence $(\mu_n) \subset \mathcal{M}$ there exists a $\mu \in \mathcal{M}$ and subsequence (μ_{n_k}) converging to μ .*

Theorem 2.16 (Krylov–Bogolyubov). *There exists at least one T -invariant probability measure μ .*

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

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

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

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

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

and the result follows. \square

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 sequence

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

is bounded (by $\|f_m\|_\infty$). Choose convergence subsequences $(n_i^{(1)})_{i=1}^\infty \supset (n_i^{(2)})_{i=1}^\infty \supset (n_i^{(3)})_{i=1}^\infty \supset \dots$ such that

$$\tau(f_k) := \lim_{i \rightarrow +\infty} \frac{1}{n_i^{(k)}} \sum_{l=0}^{n_i^{(k)}-1} f_k(T^l x).$$

Then for each $m \geq 1$, we can use the diagonal subsequence $(n_k^{(k)})_{k=1}^\infty$ to write

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

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

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

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. $L : C(X, \mathbb{R}) \rightarrow \mathbb{R}$ is linear;
3. $L(f \circ T) = L(f)$, $\forall f \in C(X, \mathbb{R})$, since

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

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 2.17 (Natural extensions). Assume that $T : X \rightarrow X$ 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 $[x_{-n}, \dots, x_0] = \{(x_n)_{-\infty}^0 \in \widehat{X} : x_i \in B_i, -n \leq i \leq 0\}$ where $B_i \in \mathcal{B}$ is the sigma algebra for X . We then define a map $\widehat{T} : \widehat{X} \rightarrow \widehat{X}$ by $\widehat{T}(x_n)_{-\infty}^0 = (Tx_n)_{-\infty}^0$. We can also uniquely extend the measure μ 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 2.18 (Poincaré Recurrence). *Let μ be a T -invariant probability measure 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. (μ) 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

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

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

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

Remark 2.19. This statement can (essentially) be found in the bible (Ecclesiastes 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.*

We return to an example mentioned in the introduction, which can be viewed as an application of Poincaré Recurrence.

Example 2.20 (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 2.21 (Poincaré-Furstenberg Multiple Recurrence). *Let $T_1, \dots, T_d : X \rightarrow X$ 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 $\mu(B \cap T_1^{-M} B \cap \dots \cap T_d^{-M} B) > 0$.*

The most famous application of this is the following.

Example 2.22 (Furstenberg-Szemerdi). Let $\mathcal{N} \subset \mathbb{Z}$ be a subset of positive density, i.e.,

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

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 $N, N+M, N+2M, \dots, N+dM \in \mathcal{N}$.

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

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

Let $T : \Sigma \rightarrow \Sigma$ 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 (μ_n) 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 Σ there is at least one weak-star limit point μ (i.e., there exists $n_k \rightarrow +\infty$ such that $\mu_{n_k} \rightarrow \mu$). Moreover, the same argument as in the existence of invariant measures shows that μ is T -invariant and, assuming we took the subsequence through the sequence corresponding to the limit supremum 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

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

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

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

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

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

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

2.5 Invariant measures and flows

Let $T^t : X \rightarrow X$ be a flow where $t \in \mathbb{R}$, i.e., $T^{s+t} = T^s \circ T^t$ and $T^0 = I$. (More generally, we can let G be a continuous group and let $t \in G$).

Definition 2.23. We say that a probability measure μ is T^t -invariant if $\mu(T^t B) = \mu(B)$, for all $B \in \mathcal{B}$, $t \in \mathbb{R}$.

Example 2.24. We can define a flow on $X = \mathbb{R}/\mathbb{Z}$ by $T^t x = x + t \pmod{1}$. This preserves the usual Lebesgue measure.

Example 2.25. Let $(\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d$. We can define a flow on $X = \mathbb{R}^d/\mathbb{Z}^d$ by $T^t(x_1, \dots, x_d) = (x_1 + t\alpha_1, \dots, x_d + t\alpha_d) \pmod{1}$. This preserves the usual Lebesgue measure.

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

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

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 2.26. 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 ν to assume that it is a probability measure.

Example 2.27 (Geodesic flow). We can define a flow $T^t : X \rightarrow X$ 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 2.28. The measure ν is T^t -invariant.

Proof. This follows from the fact that for any $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \Gamma \in G/\Gamma$ we have that

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

So $DT^t(g\Gamma) = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}$ which is independent of $g\Gamma$. 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. \square

Example 2.29 (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 2.30. ν is T^t -invariant.

Proof. This follows from the fact that we have

$$\begin{aligned} T^t g\Gamma &= h_t g\Gamma = h_t \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \Gamma = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \Gamma \\ &= \begin{pmatrix} \alpha + t\gamma & \beta + t\delta \\ \gamma & \delta \end{pmatrix} \Gamma. \end{aligned}$$

So $DT^t(g\Gamma) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ which is independent of $g\Gamma$. In particular, we see that

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

(since the shear in β is volume preserving) which implies that the flow preserves the measure. \square

Remark 2.31. 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 ν is T^t -invariant.

3 Ergodicity

3.1 Ergodicity for transformations

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

Definition 3.1. We say that (T, X, μ) 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 very useful throughout the course.

Lemma 3.3. *The following are equivalent characterisations of ergodicity.*

1. μ is ergodic iff $f \circ T = f \in L^2(X, \mu)$ implies f is a constant.
2. μ is ergodic iff $f \circ T \leq f$ implies f is a constant.
3. μ 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 $A = \{x : f(x) > c\}$ satisfies $0 < \mu(A) < 1$.

For part 1, if μ 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 μ 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 μ 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 μ 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 (and is left as an exercise). \square

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

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

(In fact, this formally applies to periodic functions on \mathbb{R} , requiring that $f(0) = f(1)$. However, since this single point has zero measure it is not an issue.)

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 μ be the usual Lebesgue measure.

1. Given $0 \leq \alpha < 1$ let $T(x) = x + \alpha \pmod{1}$. This is ergodic if and only if α is irrational.

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

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

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

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 $T(x) = bx \pmod{1}$ 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 μ 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

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

Thus, if $f \circ T = f$ then

$$\sum_{n \in \mathbb{Z}} a_n e^{2\pi i n b x} = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x},$$

and comparing coefficients gives $a_n = a_{nb}$. 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 $T(x) = \frac{1}{x} \pmod{1}$ then the Gauss measure μ is ergodic. (We will return to this example later.)
4. A Bernoulli measure on a shift space is ergodic. (We will return to this example later.)

Remark 3.6. One can often show ergodicity in such examples without resorting to Fourier series. An alternative approach in the case of the b -expansion, say, involves writing $D_{k/2^n} = [\frac{k}{2^n}, \frac{k+1}{2^n}) 2^n$ where $k \in \{0, 1, \dots, 2^n - 1\}$. If $E \in \mathcal{B}$ then it is not hard to check that $\mu(T^{-n}E \cap D_{k/2^n}) = 2^{-n}\mu(E)$ and then by an approximation of any invariant set $T^{-1}E = E$ by intervals of size b^{-n} one can show that $\mu(E) = 0$ or 1.

3.2 Existence of Ergodic Measures

Let X be a compact metric space and let $T : X \rightarrow X$ 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 useful properties of \mathcal{M}_T :*

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, $\mathcal{M}_T = \bigcap_{f \in C(X)} \{\mu \in \mathcal{M} : \int (f \circ T - f) d\mu = 0\}$ 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, $\mu_\lambda \in \mathcal{M}_T$. □

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 μ is ergodic iff μ is extremal in \mathcal{M}_T ;
2. Given any T -invariant probability measure μ , there exists a measure Λ_μ on the extremal (ergodic) measures in \mathcal{M}_T such that we can write

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

Proof. For part 1, assume that μ 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 μ 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, $\mu_1 \ll \mu$ and $\mu_2 \ll \mu$ since if $\mu(B) = 0$ then $\mu_1(B) = 0$ and $\mu_2(B) = 0$. Moreover, the Radon-Nikodym derivatives $\frac{d\mu_1}{d\mu}, \frac{d\mu_2}{d\mu} \in L^1(X, \mu)$ 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 μ is ergodic, then we deduce that $\frac{d\mu_1}{d\mu} = \frac{d\mu_2}{d\mu} = 1$ and thus $\mu_1 = \mu_2 = \mu$.

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

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

Theorem 3.11. Let $T : X \rightarrow X$ 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

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

by

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

Since $\mathcal{M}_i \ni \nu \rightarrow \int f_{i+1} d\nu$ 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) $\mathcal{M}_\infty := \bigcap_{n=0}^\infty \mathcal{M}_n \neq \emptyset$

We claim that \mathcal{M}_∞ 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

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

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

3.3 Recurrence and ergodic measures

We now come to a natural question: If almost all points return to a set infinitely often, then what is the average return time?

Let $B \in \mathcal{B}$ with $\mu(B) > 0$. We define a function $n_B : B \rightarrow \mathbb{R}$ 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. (μ) $x \in B$ returns to B (infinitely often) we can define the first return map $T_B : B \rightarrow B$ by

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

We can define a probability measure μ_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 $A \in \mathcal{B}$). We can easily check that T_B preserves μ_B .

Theorem 3.12 (Kac). *Assume that μ is ergodic, then $\int_B n_B(x) d\mu_B(x) = \frac{1}{\mu(B)}$ for any $B \in \mathcal{B}$ such that $\mu(B) > 0$.*

Proof. In order to give a simple proof, we assume that T is invertible.

By ergodicity we can partition the space X by

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

since the expression on the Right Hand Side is T -invariant and has positive measure (since $\mu(B) > 0$) and using ergodicity. We 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 $T : \Sigma \rightarrow \Sigma$ be the full shift on two symbols and μ 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:

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

with a 0 in the zeroth and n th 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 $B^{(N)} = \underbrace{[0, \dots, 0]}_{\times N}$. In this case we can partition B_N into sets

$$B_n = \left\{ x \in B^{(N)} : x = (\dots, \underbrace{0 \dots, 0}_{\times N}, x_N, \dots, x_{N+n-1}, \underbrace{0 \dots, 0}_{\times N}, \dots) \right\}$$

with two N -blocks of 0s starting in the zeroth and $(N + n)$ th places and no such blocks in between. As in the proof of Kac's Theorem, we have by ergodicity

$$\begin{aligned} 1 &= \mu(\Sigma) = \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} \mu(T^{-k} B_n) \\ &= \sum_{n=1}^{\infty} n \mu(B_n) = \int n_{B^{(N)}} d\mu \\ &= \mu(B^{(N)}) \int n_{B^{(N)}} d\mu_{B^{(N)}}. \end{aligned}$$

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 $H \times X \rightarrow X$ 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 μ be a T -invariant probability measure and let $f_1, f_2 \in L^1(X, \mu)$. Then*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\mathcal{G} = L^2(X, \mathcal{B}, \mu)$, 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. $\langle h, h \circ T \rangle = \int h \cdot h \circ T d\mu = \langle h \circ T, h \rangle$
3. $\langle h, \underbrace{h - h \circ T}_{\in \mathcal{G}} \rangle = 0$ 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 $\frac{1}{N} \sum_{k=0}^{N-1} f_1(T^k x) \rightarrow \bar{f}_1 \in L^2(X, \mu)$ as $N \rightarrow +\infty$ then

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

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

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 $T : X \rightarrow X$ be a transformation that preserves a probability measure μ .

The original pointwise ergodic theorem is due to Birkhoff

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

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 $\underline{f}(x) := \liminf_{n \rightarrow +\infty} \frac{1}{n} f^n(x)$, for a.e. $(\mu) x \in X$. In particular, since $\underline{f}(x) = \underline{f}(Tx)$ then by ergodicity \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 $n \geq 1$, $f^n(x) \leq n(\underline{f} + \epsilon) + \sum_{i=0}^{n-1} \chi_A(T^i x) \|f\|_\infty + M \|f\|_\infty$.

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

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

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

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

since by T -invariance of μ we have

$$\int \left(\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \right) d\mu = \int f d\mu$$

and

$$\int \left(\frac{1}{n} \sum_{k=0}^{n-1} \chi_A \circ T^k \right) d\mu = \int \chi_A d\mu = \mu(A).$$

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

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

Therefore we can conclude that

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

which completes the proof of the Birkhoff ergodic theorem. □

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

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

which implies that $\bar{f} = f$ a.e. (μ).

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

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

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

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

As we observed above, this suffices to prove 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 $T : X \rightarrow X$ be preserve the probability measure m and let $\phi : X \rightarrow \mathbb{R}$ 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

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

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

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

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

(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 by definition

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

In particular,

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

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

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

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

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

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 $X \setminus A$ and then

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

In particular,

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

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

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

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

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

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 $0 \leq x \leq 1$ where $b \geq 2$ and $b_n(x) \in \{0, 1, 2, \dots, b-1\}$. For a.e., (m) $x \in [0, 1]$ this expansion is unique (where m is Lebesgue measure). Let

$$N_b = \left\{ 0 \leq x \leq 1 : \lim_{N \rightarrow +\infty} \frac{\text{Card}\{1 \leq n \leq N : b_n(x) = i\}}{N} = \frac{1}{b} \text{ for } i = 0, 1, \dots, b-1 \right\}$$

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

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

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., $\exists \sigma > 0, \forall y \in \mathbb{R}$

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

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

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

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