Exploring Ergodic Theory

Alexander J. Clay

An honors thesis presented for the degree of
Bachelor of Arts

Department of Mathematics
University of Notre Dame
United States
April 4, 2022

Professor Liviu I. Nicolaescu, Advisor

Acknowledgements

I would first like to thank my advisor, Professor Liviu Nicolaescu, for his steadfast encouragement of my work and his faith in me as a person and as a mathematician. I appreciate greatly your advice and assistance with this project. It was an honor to collaborate with you and a privilege to take your classes. Thank you as well for helping me find the fun in math this year and encouraging me to apply to graduate school.

Additionally, I would like to recognize a few more faculty members who helped me navigate my journey through mathematics. First, I would like to thank Professor Michael Gekhtman for his kindness and generosity. Second, thank you to Professor Jeffrey Diller for his advice and helpfulness on my thesis and the graduate school application process. I would also like to thank Professor Eric Riedl and Professor David Galvin for encouraging me to stick with the Honors Math program early in my undergraduate career.

My friends and colleagues also deserve a huge thank you. Thank you to everyone from the Math Bunker, especially Nick and Alex. Thank you to Justin, Tarik, and the band family. Thank you to all my friends from back home.

Finally, I would like to thank my family for their constant support and encouragement throughout my academic career. Mom, thank you for always supporting me and encouraging me to do the things I love. Dad, thank you for giving me great advice and always keeping up with me while I'm at school. Ava, thank you for being the best sister I could ever ask for. To my grandparents, thank you for always believing in me. You are my inspirations.

In memory of Dan McCartney

EXPLORING ERGODIC THEORY

ALEX CLAY

Contents

Advisor: Liviu Nicolaescu. Started August 30, 2021. Last revised April 19, 2022.

Notation and convention

- We will denote by $|S|$ of $\#S$ the cardinality of a finite set S.
- We denote by $\mathbb{N}$ the set of natural numbers, $\mathbb{N} = \{1, 2, \dots\}$ and we set $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.
- Ffor any set X we denote by 2^X the collection of all the subsets of X
- For any set S, contained in some ambient space X, we denote by I_S the *indicator function* of S
$$I_S : X \to \{0,1\}, \;\; I_S(x) = \begin{cases} 1, & x \in S, \\ 0, & x \notin S. \end{cases}$$
- We will use the symbol $\mathbb{E}$ when referring to expectation and the symbol $\mathbb{P}$ when referring to probability.
- We set $\boldsymbol{i} := \sqrt{-1}$

Introduction

Mix cocoa powder into a dough. Walk along a circle with steps of equal sizes. These are all examples of ergodic transformations, and they describe systems that evolve in time. All the possible states of that particular system are collected in a space of states denoted by Ω. If, at the present, the system is in a state ω, then during one unit of time it will transition to a new state $T\omega$. The transitions are thus mathematically described by a map $T : \Omega \to \Omega$. Over n units of time, the transitions are described by the map $T^n := \underbrace{T \circ \cdots \circ T}_{n}$.

In many instances, one has a concept of measure associated to Ω, i.e., one can speak of the measure (think volume) of (certain) subsets of Ω. Formally, we have a collection $\mathscr{S}$ of subsets of Ω, the so-called measurable subsets, and a measure, i.e., a function $\mathbb{P} : \mathscr{S} \to [0, \infty)$. We will be interested exclusively in the case when Ω has finite measure or "volume". By choosing units appropriately, we can assume that Ω has measure 1. In this case, we say that $\mathbb{P}$ is a probability measure. Given a random point $\omega \in \Omega$, the probability that ω will land in a measurable subset $S \subset \Omega$ is $\mathbb{P}[S]$.

Ergodic theory is interested in the dynamics of measure preserving maps $T : \Omega \to \Omega$. Roughly speaking, the map T is measure preserving if T maps measurable sets to other measurable sets, but of equal measure. Physicists say that the evolution described by T is incompressible. For example, a C^1 diffeomeorphism such that the absolute value of the determinant of its Jacobian is equal to 1 is measure preserving.

Another important example of measure-preserving maps are the so-called shifts. We use Kolmogorov's Existence Theorem to construct an underlying, shift-invariant probability measure.

Rotations of the circle are other examples of measure preserving transformations.

The states of a system have measurable numerical characteristics, such as temperature and pressure. Mathematically, they are described by functions $f : \Omega \to \mathbb{R}$. Suppose we fix a certain measurable characteristic f. If the system starts in the state ω with characteristic $f(\omega)$, then after n epochs the characteristic is $f(T^n\omega)$. Thus the average value of this characteristic during the first n epochs of the evolution is

$$A_n[f](\omega) = \frac{1}{n+1}\big(\, f(\omega) + f(T\omega) + \cdots + f(T^n\omega)\,\big).$$

This averaging procedure yields new functions $A_n[f] : \Omega \to \mathbb{R}$ called the temporal averages.

One central question of ergodic theory concerns the long term behavior of the temporal averages, i.e., what happens to $A_n[f]$ as $n \to \infty$. The various ergodic theorems discussed in this thesis describe $\bar{f} = A_\infty[f]$, the limit of $A_n[f]$ as $n \to \infty$.

This description involves the so called invariant sets. More precisely, a subset $S \subset \Omega$ is called T-invariant if $T^{-1}(S) = S$. Explicitly, this means two things.

- If $\omega \in S$, then $T(\omega) \in S$ and, conversely
- If $T(\omega) \in S$, then $\omega \in S$.

We denote by $\mathscr{I}_T$ the collection of T-invariant subsets of Ω. This collection is a sigma-algebra. In probabilistic language, $\bar{f}$ is the conditional expectation of f given the sigma-algebra $\mathscr{I}_T$. Loosely speaking, this is the best approximation of f given the information contained in the collection $\mathscr{I}_T$.

The description of $\bar{f}$ simplifies considerably under the *ergodic hypothesis*. More precisely, we say that the map T is ergodic if any T-invariant subset has either measure 0, or measure 1, the same measure as the state space Ω. In other words, under the ergodic hypothesis, an invariant subset is either almost nothing, or almost everything, with nothing in between. In this case, the ergodic theorem states that

$$\lim_{n\to\infty} A_n[f] = \int_\Omega f(\omega)\,\mathbb{P}[d\omega].. \tag{E}$$

The term in the right-hand-side of (E) is the space average of f.

Let us explain why physicists appreciate the equality (E). From a physicist's point of view, the temporal averages are essentially non-computable experimentally since it involves the behavior of a system in perpetuity, and we do not have an infinite amount of time at our disposal. On the other hand, the left-hand-side, the space average, is something that can be determined experimentally or by direct computation. Thus, whenever (E) holds, we can say something about the distant future by performing experiments and computations in the present.

Ludwig Boltzmann was the first physicist to single out the ergodic hypothesis and give a heuristic explanantion of why it would imply (E). The rigorous proof appeared much later.

In this thesis, we discuss two landmark results from ergodic theory: the Mean Ergodic Theorem, proved by John Von Neumann in 1932, and the Birkhoff Ergodic Theorem, proved by George David Birkhoff in 1931. In fact, with the wisdom of hindsight, we first prove the Mean Ergodic theorem and, based on it, we prove Birkhoff's theorem using some simplifications that were discovered by various mathematicians in the decades that followed Birkhoff's proof.

Proving that a given preserving map satisfies the ergodic hypothesis is not a trivial matter, and each concrete example requires some ingenuity. We conclude this thesis by describing a few classical examples of ergodic maps.

1. Measure-Preserving Transformations

Let us recall a few measure-theoretic facts. For more details and proof we refer to [3]. Let $(\Omega, \mathscr{S})$ be a measurable space, i.e., $\mathscr{S}$ is a fixed sigma algebra of subsets of Ω. We denote by $\mathrm{Meas}(\Omega, \mathscr{S})$ the space of sigma-finite measures on $\mathscr{S}$ and by $\mathrm{Prob}(\Omega, \mathscr{S})$ the set of probability measures on $\mathscr{S}$. Any measurable map

$$T : (\Omega, \mathscr{S}) \to (\Omega, \mathscr{S})$$

induces a *pushforward* transformation

$$T_{\#} : \mathrm{Meas}(\Omega, \mathscr{S}) \to \mathrm{Meas}(\Omega, \mathscr{S}), \;\; \mathrm{Meas}(\Omega, \mathscr{S}) \ni \mu \mapsto T_{\#}\mu \in \mathrm{Meas}(\Omega, \mathscr{S}),$$
$$T_{\#}\mu[S] = \mu\big[T^{-1}(S)\big], \;\; \forall S \in \mathscr{S}, \;\; \mu \in \mathrm{Prob}(\Omega, \mathscr{S}).$$

Definition 1.1. Suppose that $(\Omega, \mathscr{S}, \mu)$ is a finite measured space and $T : \Omega \to \Omega$ is a map.

(i) The map T is said to be *measure preserving* if it is measurable and $\mu = T_{\#}\mu$. This means that

$$\mu\big[T^{-1}(S)\big] = \mu[S], \;\; \forall S \in \mathscr{S}. \tag{1.1}$$

(ii) The map T is called and *automorphism* of the measured space $(\Omega, \mathscr{S}, \mu)$ if it is bijective and both T and T^{-1} are measure preserving.

□

Note that if T is an automorphism, then for any $S \in \mathscr{S}$ the sets

$$S,\ T(S),\ T^{-1}(S)$$

are measurable and have the same measure. As explained in [3], for the equality (1.1) to hold, it suffices to check that

$$\mu\big[T^{-1}(S)\big] = \mu[S],\ \ \forall S \in \mathscr{P}, \tag{1.2}$$

where $\mathscr{P}$ is a π-system generating[1] $\mathscr{S}$.

Definition 1.2 (Invariance)**.** Let $T : (\Omega, \mathscr{S}, \mu) \to (\Omega, \mathscr{S}, \mu)$ be a measure preserving map.

(i) A measurable function $f : (\Omega, \mathscr{S}) \to \mathbb{R}$ is said to be (T-)**invariant** if $f \circ T = f$. A measurable set A to be **invariant** if its indicator function $\boldsymbol{I}_A$ is invariant.

(ii) A measurable function $f : (\Omega, \mathscr{S}) \to \mathbb{R}$ is said to be (T-)**quasi-invariant** if $f \circ T = f$ μ almost everywhere. A measurable set A to be **quasi-invariant** if its indicator function $\boldsymbol{I}_A$ is quasi-invariant.

Observe that A is T-invariant iff $A = T^{-1}(A)$, i.e.,

(i) $A \subset T^{-1}(A) \longleftrightarrow \forall a \in A:\ \ T(a) \in A$

(ii) $T^{-1}(A) \subset A \longleftrightarrow \forall \omega,\ \ T(\omega) \in A \implies \omega \in A.$

Definition 1.3 (Orbit)**.** The orbit of a map $T : X \to X$ through a point $x \in X$ is the set

$$\boldsymbol{O}_x = \boldsymbol{O}_{T,x} = \big\{\, x, Tx, T^2x, \dots \,\big\}.$$

If T is bijective then we define the **orbit** of a point $x \in X$ to be the set of

$$\big\{T^n x, n \in \mathbb{Z}\,\big\}$$

□

The main property of an invariant set is that it contains any orbit that intersects it.

Theorem 1.4 (Poincaré Recurrence Theorem)**.** Suppose that T is an automorphism of the probability space $(\Omega, \mathscr{S}, \mu)$. Let $S \in \mathscr{S}$ such that $\mu[S] > 0$. Then, there exists a negligible set $N \subset S$ such that, for any $\omega \in S \setminus N$ the orbit of ω will go again through S infinitely many times in the future, i.e., the set $\boldsymbol{O}_\omega \cap S$ is infinite for any $\omega \in S \setminus N$.

[1] Aπ-system is a collection of subsets closed under intersection.

Proof. We follow the presentation in [6, Thm. 2.12]. For every $n \in \mathbb{N}$ we set

$$S_n := \bigcup_{i=n}^{\infty} T^{-i} S.$$

Clearly

$$S_0 \supset S_1 \supset \ldots \tag{1.3}$$

Furthermore, we have $T^{-n}S_0 = S_n$. Because T is measure-preserving, we have

$$\forall n \geq 0 \;\; \mu[S_0] = \mu[T^{-n}S_0] = \mu[S_n] = \mu[S_\infty], \;\; S_\infty = \bigcap_{n=0}^{\infty} S_n.$$

Since μ is finite and (1.3) holds, we have

$$\mu[S_0 \Delta S_\infty] = 0$$

for all n. Note that $S \subset S_0$ implies that $\mu[S \cap S_\infty] = \mu[A]$. Any point $\omega \in S \cap S_\infty$ has the claimed properties. ☐

In the next example, we describe a stronger form of recurrence.

Example 1.5 (Irrational Rotations on a Circle)**.** For any $\theta \in (0,1)$ we define $T_\theta : [0,1) \to [0,1)$,

$$T_\theta(x) = x + \theta \bmod 1. \tag{1.4}$$

Equivalently, if we identify $x \in [0,1)$ with the point $z = e^{2\pi \boldsymbol{i} x}$ on the unit circle in $\mathbb{C}$, $\{|z| = 1\}$, then we can think of T as acting on the unit circle and

$$T_\theta(z) = ze^{2\pi \boldsymbol{i}\theta}.$$

Thus T is a rotation of angle $2\pi\theta$, implying that T preserves the arclength measure on the circle. We have

$$T^k(z) = ze^{2k\pi \boldsymbol{i}\theta}.$$

We distinguish two cases. We want to investigate the orbits of this map.

1. The angle $2\pi\theta$ is rational. We write θ as $\frac{a}{b}$ with $\gcd(a,b) = 1$. Then, $T_\theta^b = \mathbb{1}$. We say that T is periodic. All orbits consist of b points and we can show that T is not ergodic.

2. The angle $2\pi\theta$ is irrational. We want to show that any orbit of T is dense on the circle of radius 1. More precisely, we will show that any orbit of T intersects any arc on the circle.

Observe that $T^n z = zT^n(1)$. Thus, it is only necessary to show that the orbit of 1 is dense. Fix an arc J of length $\varepsilon > 0$. Set $z_n = T^n(1) = e^{2n\pi \boldsymbol{i}}$. Then

$$m \neq n \implies z_m \neq z_n.$$

Denote by I_0 the arc of length $\varepsilon/2$ centered at z_0. We set $I_n = T^n(I_0)$ and we observe that I_n is the arc length $\varepsilon/2$ centered around x_n. Obviously, the arcs I_n are not pairwise disjoint, because the circle has finite length and each interval has equal measure $\varepsilon/2$. Thus, there exist m, n, $m < n$ such that $I_m \cap I_n \neq \emptyset$. If $k := n - m$, we observe that

$$I_m \cap I_m = T^m\big(\, I_0 \cap I_k \,\big)$$

so that $I_0 \cap I_k \neq \emptyset$. Thus, the distance between the centers of z_0 and z_k is $< \varepsilon/2$, i.e.

$$\operatorname{dist}(z_0, z_k) < \varepsilon.$$

The rotation $R = T^k$ is also measure preserving and $\operatorname{dist}(z_0, Rz_0) < \varepsilon$. Hence

$$\operatorname{dist}(z_{kj}, z_{k(j+1)}) = \operatorname{dist}(R^j z_0, R^{j+1} z_0) < \varepsilon/2, \;\; \forall j \geq 0.$$

Think of the points $R^j z_0$ as describing the walk along the circle with step sizes of equal lengths $< \varepsilon/2$. Such a walk cannot avoid a "ditch" J of width $\varepsilon > \varepsilon/2$. □

The next example is found in *differentiable dynamics*.

Example 1.6. Denote by $\boldsymbol{\lambda}$ the Lebesgue measure on $\mathbb{R}^n$ Suppose that $\Omega \subset \mathbb{R}^n$ is an open set and

$$F : \Omega \to \Omega$$

is a C^1-diffeomorphism onto. Let J_F be its Jacobian matrix. The change in variables formula shows that for any open subset $V \subset \Omega$ we have

$$\int_{F^{-1}(V)} \lambda[dx] = \int_V |\det J_{F^{-1}}(y)| \lambda[dy].$$

This shows that F preserves the Lebesgue measure iff $|\det J_{F^{-1}}(x)| = 1, \forall x \in U$. □

Example 1.7. [M. Kac] Consider the map $T : [0, 1) \to [0, 1)$,

$$T(x) = 2x - \lfloor 2x \rfloor = 2x \bmod 1.$$

We want to show that T preserves the Lebesgue measure $\boldsymbol{\lambda}$. Note that the collection of intervals $[0, a]$, $a \in [0, 1)$, is a π-system that generates the Borel algebra of $[0, 1)$ so, in view of (1.2), it suffices to show that

$$\boldsymbol{\lambda}\big[\, T^{-1}(I_a) \,\big] = \boldsymbol{\lambda}[I_a] = a, \;\; \forall a \in [0, 1).$$

This is immediate since

$$T^{-1}(I_a) = [0, a/2) \cup [1/2, 1/2 + a/2)$$

and obviously $\boldsymbol{\lambda}\big[\, T^{-1}(I_a) \,\big] = a$.

As in Example 1.5, we identify $x \in [0, 1)$ with the point $z = e^{2\pi \boldsymbol{i} x}$ on the unit circle $S^1 = \{|z| = 1\}$. The Lebesgue measure on $[0, 1)$ corresponds to the arc length on this circle normalized so that the total length is 1. We can view T as a map $S^1 \to S^1$ and, as such, it has the simple description $T(z) = z^2$. □

Example 1.8 (The tent map)**.** Consider the *tent map* $T : [0,1] \to [0,1]$, $T(x) = \min(2x, 2-2x)$. Equivalently, this is the unique continuous map such that $T(0) = T(1) = 0$, $T(1/2) = 1$, and it is linear on each of the intervals $[0,1/2]$ and $[1/2,1]$. Its graph looks like a tent with vertices $(0,0)$, $(1/2,1)$ and $(1,0)$.

This map preserves the Lebesgue measure. Indeed, if $I \subset [0,1]$ is a compact interval, then $T^{-1}(I)$ consists of two intervals $I_{\pm}$, symmetrically located with respect to the midpoint $1/2$ of $[0,1]$, and each having half the size of I. □

Example 1.9 (Shifts)**.** Let $\mathscr{A}$ be a finite set(alphabet). Suppose that $w : \mathscr{A} \to (0,1]$ is a function satisfying

$$\sum_{a\in\mathscr{A}} w(a) = 1.$$

It defines a probability P_w measure on $2^{\mathscr{A}}$,

$$\mathbb{P}_w[S] = \sum_{s\in S} w(s), \;\; \forall S \subset \mathscr{A}.$$

Let $\Omega = \mathscr{A}^{\mathbb{Z}}$. An element of Ω is a function $\omega : \mathbb{Z} \to \mathscr{A}$ or, equivalently, a doubly infinite sequence

$$\omega = (\dots, a_{-1}, a_0, a_1, \dots).$$

For every $S \subset \mathscr{A}$ and any $n \in \mathbb{Z}$ we set

$$C_{n,S} = \big\{ \, \omega \in \Omega; \;\; \omega(n) \in S \, \big\}.$$

The finite intersections of sets of the type $C_{n,S}$ are called *cylinders*. Note that a cylinder is a set of the type

$$C_{n_1,\dots,n_k|S_1,\dots,S_k} = \big\{ \, \omega \in \Omega; \;\; \omega(n_j) \in S_j, \;\; \forall j = 1,\dots,k \, \big\}, \tag{1.5}$$

where $n < \cdots < n_k$ are integers, and $S_1,\dots,S_k$ are subsets of $\mathscr{A}$. We denote by $\mathscr{C}$ the collection of cylinders. Note that $\mathscr{C}$ is π-system. We denote by $\mathscr{S}$ the σ-algebra generated by $\mathscr{C}$.

Kolmogorov's existence theorem [3, Sec. 36] shows that there exists a probability measure $\bar{\mathbb{P}}$ on $\mathscr{S}$ uniquely determined by the condition

$$\bar{\mathbb{P}}[C_{n_1,\dots,n_k|S_1,\dots,S_k}] = \mathbb{P}_w[S_1]\cdots\mathbb{P}_w[S_k], \;\; \forall C_{n_1,\dots,n_k|S_1,\dots,S_k} \in \mathscr{C}.$$

The *shift* is the map

$$T : \Omega \to \Omega, \;\; (T\omega)(n) = \omega(n+1).$$

Observe that

$$T_{\#}\bar{\mathbb{P}}[C_{n_1,\dots,n_k|S_1,\dots,S_k}] = \bar{\mathbb{P}}[C_{n_1+1,\dots,n_k+1|S_1,\dots,S_k}] = \bar{\mathbb{P}}[C_{n_1,\dots,n_k|S_1,\dots,S_k}].$$

Invoking (1.1), we deduce that $T_{\#}\bar{\mathbb{P}} = \bar{\mathbb{P}}$, so the shift is measure preserving. □

2. The Ergodic Theorems

Throughout this section, $(\Omega, \mathscr{S}, \mathbb{P})$ will denote a *probability* space. Suppose that $T : (\Omega, \mathscr{S}, \mathbb{P}) \to (\Omega, \mathscr{S}, \mathbb{P})$ is a measure preserving map.

Definition 2.1. A measurable function $h : (\Omega, S, \mathbb{P}) \to \mathbb{R}$ is called T-**quasi-invariant** if there exists $N \in \mathscr{S}$ such that $\mathbb{P}[S] = 0$ and

$$h(T\omega) = h(\omega), \ \ \forall \omega \in \Omega \setminus N.$$

In other words, $h \circ T = h$ almost everywhere.

A measurable set $A \in \mathscr{S}$ is called quasi-invariant if its indicator $\boldsymbol{I}_A$ is such, i.e. $\boldsymbol{I}_A = \boldsymbol{I}_A \circ T$ almost everywhere. We denote by $\mathscr{J} = \mathscr{J}_T$ the collection of quasi-invariant sets. □

The idea of invariance can be conveniently expressed in terms of the *Koopman operator*. We denote by $L^0(\Omega, \mathscr{S}, \mathbb{P})$ the vector space of $\mathscr{S}$-measurable functions $f : \Omega \to \mathbb{R}$ modulo almost everywhere equality.

Definition 2.2 (Koopman Operator)**.** Let $(\Omega, \mathscr{S}, \mathbb{P})$ be a probability space and let $T : \Omega \to \Omega$ be a measure-preserving map. For any measurable function $f : (\Omega, \mathscr{S}, \mathbb{P}) \to \mathbb{R}$, we denote by $\hat{T}$ the pullback of f by T, i.e., $\hat{T}f = f \circ T$. The induced *linear* map

$$\hat{T} : L^0(\Omega, \mathscr{S}, \mathbb{P}) \to L^0(\Omega, \mathscr{S}, \mathbb{P})$$

is called the **Koopman operator** determined by T. □

A measurable function h is quasi-invariant if and only if $\hat{T}h = h$. Thus,

$$h \ \text{ is quasi-invariant } \longleftrightarrow \ h \in \ker\bigl(1 - \hat{T}\bigr).$$

Proposition 2.3. The collection $\mathscr{J}_T$ is a sigma-subalgebra of $\mathscr{S}$. Moreover, a measurable function $h : (\Omega, \mathscr{S}, \mathbb{P}) \to \mathbb{R}$ is T-quasi-invariant if and only if it is $\mathscr{J}_T$-measurable, i.e.,

$$\bigl\{ h \leq r \bigr\} \in \mathscr{J}_T, \ \ \forall r \in \mathbb{R}.$$

Proof. Clearly, if f is quasi-invariant, then so are the sublevel sets $\{f \leq x\}$ $\forall x \in \mathbb{R}$, and thus f is $\mathscr{J}_T$-measurable.

Conversely, if f is $\mathscr{J}_T$-measurable, then so are $f_{\pm}$, and it suffices to show that if $f \geq 0$ is $\mathscr{J}$-measurable, then f is quasi-invariant. Clearly any $\mathscr{J}$-measurable elementary function is quasi-invariant. Since f is an increasing limit of $\mathscr{J}$-measurable elementary functions, it is therefore an increasing limit of quasi-invariant elementary functions and thus it is quasi-invariant. □

Now that we have established what it means for a function to be measurable on invariant sets, it is necessary to use this definition to formulate a notion of ergodicity. In chapter 1, we noted that a measure preserving map T is ergodic if and only if every T-invariant set has measure 0 or 1. Here, we will reformulate the definition in terms of the sigma-subalgebra $\mathscr{I}_T$. Then, we will regard the image of functions under $(\hat{T})$ as a vector space and give a definition of ergodicity in terms of vector spaces.

Definition 2.4. A measure preserving map $T : (\Omega, \mathscr{S}, \mathbb{P}) \to (\Omega, \mathscr{S}, \mathbb{P})$ is called *ergodic* if

$$\forall A \in \mathscr{I}_T : \ \ \mathbb{P}\big[\, A \,\big] \in \{0, 1\},$$

i.e., any quasi-invariant set has measure zero or one. □

Note that for any measure preserving map T the vector subspace $\ker\big(\, 1 - \hat{T} \,\big)$ has dimension ≥ 1 since the indicator function $\boldsymbol{I}_\Omega$ is obviously T-quasi-invariant.

Proposition 2.5. Let $p \in [1, \infty]$. The measure preserving map T is ergodic if and only if $\dim \ker\big(\, 1 - \hat{T} \,\big) \cap L^p(\Omega, \mathscr{S}, \mathbb{P}) = 1$ so that $\ker\big(\, 1 - \hat{T} \,\big) \cap L^p(\Omega, \mathscr{S}, \mathbb{P})$ consists only of the constant functions.

Proof. Suppose that T is ergodic and $h \in \ker\big(\, 1 - \hat{T} \,\big) \cap L^p$. We deduce from Proposition 2.3 that the sets $\{h \leq r\}$ are quasi-invariant so that

$$\mathbb{P}\big[\, \{h \leq r\} \,\big] \in \{0, 1\}, \ \ \forall r \in \mathbb{R}.$$

Since h is a.s. finite, and the function $r \mapsto \mathbb{P}\big[\, \{h \leq r\} \,\big]$ is right continuous, we deduce that there exists $r_0 \in \mathbb{R}$ such that

$$\mathbb{P}\big[\, \{h \leq r\} \,\big] = \begin{cases} 0, & r < r_0, \\ 1, & r \geq r_0. \end{cases}$$

Hence, h is a.s. constant.

Conversely, assume that $\ker\big(\, 1 - \hat{T} \,\big) \cap L^p$ consists only of the constant functions. If S is a quasi-invariant set, then $\boldsymbol{I}_S \in \ker\big(\, 1 - \hat{T} \,\big) \cap L^p$ so $\mathbb{P}\big[\, S \,\big] \in \{0, 1\}$. Therefore, T is ergodic. □

The Koopman operator associated to a measure preserving map T also has some nice analytic properties which will be helpful in proving the Mean Ergodic Theorem. First, we notice that the Koopman operator is an isometry of $L^p(\Omega, \mathscr{S}, \mathbb{P})$.

Proposition 2.6. Let $T : (\Omega, \mathscr{S}, \mathbb{P}) \to (\Omega, \mathscr{S}, \mathbb{P})$ be a measure preserving map and $p \in [1, \infty)$. Then

$$\|\hat{T} f\|_p = \|f\|_p, \ \ \forall f \in L^p(\Omega, \mathscr{S}, \mathbb{P}),$$

where $\| - \|_p$ denotes the norm in L^p.

Proof. The change in variables [3, Thm.16.11] implies that

$$\int_\Omega |\hat{T}f|^p d\mathbb{P} = \int_\Omega \hat{T}|f|^p d\mathbb{P} = \int_\Omega |f|^p dT_\# f\mathbb{P} = \int_\Omega |f|^p d\mathbb{P},$$

where at the last step we used the fact that T is measure preserving, i.e., $\mathbb{P} = T_\#\mathbb{P}$. □

A measure preserving map $T : (\Omega, \mathscr{S}, \mathbb{P}) \to (\Omega, \mathscr{S}, \mathbb{P})$ can be viewed as describing a discrete time evolution. If the system is at an initial state ω, then after one unit of time it evolves to a new state $T\omega$. This does not give us data; rather, it gives us a sort of photograph of the system at a certain time. Ideally, we would like to get some data out of this system, which is where our function will come in. A measurable function $f : (\Omega, \mathscr{S}, \mathbb{P}) \to \mathbb{R}$ should be viewed as describing a numerical (physical) characteristic of a state, e.g., it could be the heat released by an exothermic reaction or the pH of a solution. A chemist might measure the pH of a system at several time intervals during a titration. $f(T^n\omega)$ would be the pH at the nth time interval.

As the system evolves, this numerical characteristic of a state changes. Its average value over the first $(n-1)$ time intervals is

$$A_n f(\omega) = \frac{1}{n}\big(f(\omega) + f(T\omega) + \cdots f(T^{n-1}\omega) \big).$$

The temporal average A_n thus defines a linear operator

$$A_n : L^0(\Omega, \mathscr{S}, \mathbb{P}) \to L^0(\Omega, \mathscr{S}, \mathbb{P}),$$

that can be compactly described in terms of the Koopman operator

$$A_n = \frac{1}{n}\big(1 + \hat{T} + \cdots + \hat{T}^{n-1} \big).$$

We will now investigate these temporal averages and see what happens when we hit functions with them. Let's introduce and prove a few basic properties as a warm-up to the bigger theorems.

Proposition 2.7. Let $T : (\Omega, \mathscr{S}, \mathbb{P}) \to (\Omega, \mathscr{S}, \mathbb{P})$ be a measure preserving map and $p \in [1, \infty)$. Then, for any $n \in \mathbb{N}$ and any $f \in L^0(\Omega, \mathscr{S}, \mathbb{P})$,

$$f \geq 0 \ \text{ almost everywhere } \implies A_n f \geq 0 \ \text{ almost everywhere}, \tag{2.1a}$$

$$|A_n f| \leq A_n |f| \quad \text{almost everywhere}, \tag{2.1b}$$

$$\|A_n f\|_p \leq \|f\|_p, \tag{2.1c}$$

Proof. We have

$$A_n f = \frac{1}{n}\big(f + \hat{T}f + \cdots + \hat{T}^{n-1} f \big).$$

Since $f \geq 0$ a.e., we have that the right hand side is a sum of positive terms a.e., so $A_n f \geq 0$.

Note that

$$|A_n f| = \frac{1}{n}|f + \hat{T}f + \cdots + \hat{T}^{n-1}f|.$$

Using the triangle inequality, we obtain

$$\frac{1}{n}|f + \hat{T}f + \cdots + \hat{T}^{n-1}f| \leq \frac{1}{n}|f| + \hat{T}|f| + \cdots + \hat{T}^{n-1}|f| = A_n|f|.$$

For the third inequality, observe that

$$\|A_n f\|_p \leq \frac{1}{n}\big(\,\|f\|_p + \|\hat{T}f\|_p + \cdots + \|\hat{T}^{n-1}f\|_p\,\big) = \|f\|_p$$

since $\hat{T}$ is an isometry. □

Now that we have proven these facts about A_n, we have enough machinery to get to some more subtle results. The Mean Ergodic Theorem, proved in the L^2 case by John Von Neumann in 1931, describes the behavior of the temporal averages A_n as $n \to \infty$.

For any $p \in [1, \infty)$ we set

$$Q_T^p := \ker\big(\,1 - \hat{T}\,\big) \cap L^p(\Omega, \mathscr{S}, \mathbb{P}) = L^p(\Omega, \mathscr{I}_T, \mathbb{P}).$$

For simplicity we will write $Q_T := Q_T^2$. Observe that $Q_T = L^2(\Omega, \mathscr{I}_T, \mathbb{P})$ is a closed subspace of the Hilbert space $L^2(\Omega, \mathscr{S}, \mathbb{P})$. We denote by P_T the orthogonal projection onto Q_T. Later, we will provide a concrete description of P_T in L^2 and find something analogous to it in L^1.

Theorem 2.8 (Mean Ergodic Theorem)**.** Suppose that $(\Omega, \mathscr{S}, \mathbb{P})$ is a probability space and $T : \Omega \to \Omega$ be measure-preserving. Then, for all $f \in L^2(\Omega, \mathscr{S}, \mathbb{P})$, the temporal averages $A_n f$ converge in L^2 to the orthogonal projection of f onto Q_T, i.e.

$$\frac{1}{n}(1 + \hat{T} + \hat{T}^2 + \hat{T}^3 + ... + \hat{T}^{n-1})f \to P_T f.$$

In particular, if T is ergodic, then $A_n f$ converges to the expected value of f over Ω.

Proof. Let $\mathbb{X}_2$ be the set of functions $f \in L^2(\Omega, \mathscr{S}, \mathbb{P})$ such that $A_n f$ converges in L^2 to some function $A_\infty f$. Note that $\mathbb{X}_2$ is a vector space. We want to show two things:

i.) $X_2 = L^2(\Omega, \mathscr{S}, \mathbb{P})$, and

ii.) $A_\infty = P_T$.

1. We have $Q_T \subset \mathbb{X}_2$ and $A_\infty f = f$ for all $f \in Q_T$.

This is obvious since $\hat{T}f = f$ for invariant f.

2. For all $f \in \mathbb{X}_2$, $\hat{T}f$ is in $\mathbb{X}_2$ and $A_\infty f \in Q_T$.

Observe that $\hat{T}$ and A_n are commutative since A_n is a linear combination of $\hat{T}^i$. $\hat{T}$ is also continuous, so we can take limits and obtain $A_\infty \hat{T} = \hat{T} A_\infty$. This shows $\hat{T} f \in \mathbb{X}_2$. We also have

$$nA_n \hat{T} f = (n+1)A_{n+1} f - f.$$

Dividing by n, we get

$$A_n \hat{T} f = \frac{n+1}{n} A_{n+1} f - \frac{1}{n} f.$$

Using the assumption that $f \in \mathbb{X}_2$, we can let $n \to \infty$, and get

$$A_\infty \hat{T} f = A_\infty f.$$

Commutativity yields $\hat{T} A_\infty f = A_\infty \hat{T} f = A_\infty f$, which means $A_\infty f$ is quasi-invariant under T.

3. $\hat{T} f - f \in \mathbb{X}_2$ for all $f \in L^2(\Omega, \mathscr{S}, \mathbb{P})$.

We observe that

$$A_n(\hat{T} f - f) = \frac{1}{n}(\hat{T}^n f - f).$$

We have

$$\|A_n(\hat{T} f - f)\|_2 = \|\frac{1}{n}(\hat{T}^n f - f)\|_2 \leq \frac{1}{n}(\|\hat{T} f\|_2 + \|f\|_2)$$

using substitution and the triangle inequality. Furthermore, $\hat{T}$ is unitary because it is surjective, bounded, and preserves the integral of f over Ω. (We showed this in the properties of $\hat{T}$). This means $\|\hat{T}\| = 1$. Thus,

$$\frac{1}{n}(\|\hat{T} f\|_2 + \|f\|_2) = \frac{2}{n}\|f\|_2 \to 0.$$

Since $A_n(\hat{T} f - f)$ converges, we conclude that $\hat{T} f - f \in \mathbb{X}_2$.

4. $\hat{T}^k f - f \in Q_T^\perp$ for all $k \in \mathbb{N}$ and all $f \in L^2(\Omega, \mathscr{S}, \mathbb{P})$.

It suffices to show that, for any $g \in Q_T^\perp$, the inner product of $\hat{T}^k f - f$ and g is 0. We will start with the case where $k = 1$. For any $g \in Q_T^\perp$, we have

$$(\hat{T} f - f, g) = (\hat{T} f, g) - (f, g) = (\hat{T} f, \hat{T} g) - (f, g) = 0,$$

using the fact that since $\hat{T}$ is unitary, $(\hat{T} f, \hat{T} g) = (f, g)$. This completes the $k = 1$ case. To extend to the general case, we use more properties of $\hat{T}$. We have

$$\hat{T}^k f - f = \sum_{j=1}^{k} (\hat{T}^j f - f) = \sum_{j=1}^{k} (\hat{T}(\hat{T}^{j-1} f) - \hat{T}^{j-1} f).$$

Since f was arbitrary, let $f_j = \hat{T}^{j-1} f$. Using the truth of the $k = 1$ case, we conclude that $\hat{T} f_j - f_j \in Q_T^\perp$.

5. For all $f \in \mathbb{X}_2$, $A_\infty f = P_T f$.

We deduce that

$$f - A_n f = \frac{1}{n}\sum_{k=1}^{n-1}(f - \hat{T}^k f) \in Q_T^{\perp}.$$

Using the result from step **4** and letting $n \to \infty$, we have that $f - A_\infty f \in Q_T^{\perp}$ and thus, using the definition of an orthogonal projection, $A_\infty f = P_T f$.

6. $\mathbb{X}_2$ is a closed subset of L^2.

To show this, we need to show that any convergent sequence in $\mathbb{X}_2$ converges to a function in $\mathbb{X}_2$. Let $(f_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{X}_2$ that converges to f in L^2. We will show that $(A_n f)$ is Cauchy, which will therefore show that $f \in \mathbb{X}_2$ because its temporal averages converge to a function in L^2. Let $\varepsilon > 0$ be arbitrary.

Note that

$$\|A_n f - A_m f\|_2 = \|A_n f - A_n f_k + A_n f_k - A_m f_k + A_m f_k - A_m f\|_2.$$

Applying the triangle inequality, we have

$$\|A_n f - A_m f\|_2 \leq \|A_n f - A_n f_k\|_2 + \|A_n f_k - A_m f_k\|_2 + \|A_m f_k - A_m f\|_2. \quad (2.2)$$

Since $\hat{T}$ is unitary, the opreators A_n are contractions, i.e.,

$$\|A_n g\|_2 \leq \|g\|_2, \ \ \forall g \in L^2.$$

Hence we can eliminate the A_n and the A_m from the first and last norms on the right-had-side of (2.2). We get

$$\|A_n f - A_n f_k\|_2 + \|A_n f_k - A_m f_k\|_2 + \|A_m f_k - A_m f\|_2 \leq \|f - f_k\|_2 + \|A_n f_k - A_m f_k\|_2 + \|f - f_k\|_2.$$

We have control over the quantity $\|f - f_k\|_2$ by assumption. Set $\|f - f_k\|_2 < \frac{\varepsilon}{3}$. Since $f_k \in \mathbb{X}_2$, $(A_n f_k)$ is convergent. Hence, $(A_n f_k)$ is Cauchy, so we can set $\|A_n f_k - A_m f_k\|_2 < \frac{\varepsilon}{3}$ as well. Putting it all together, we get

$$\|A_n f - A_m f\|_2 < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

7. $\mathbb{X}_2 = L^2(\Omega, \mathscr{S}, \mathbb{P})$.

This is the most difficult step. We need to use some facts about Hilbert spaces since L^2 is a Hilbert space.

From step **1**, we know that $Q_T \subset X_2$. Also, Range$(\hat{T} - 1) \subset Q_T^{\perp} \cap \mathbb{X}_2$. We use the following result from functional analysis.

If $S : H \to H$ is a bounded linear operator on a Hilbert space, then $closure(Range(S)) = (\ker S^*)^{\perp}$.

$\hat{T}$ is bounded and unitary. This means $\hat{T}^* = \hat{T}^{-1}$. Let $S = \hat{T} - 1$ as in the result. We have

$$S^* = \hat{T}^* - 1 = \hat{T}^{-1} - 1,$$

again using the fact that $\hat{T}$ is unitary. Plugging into our result, we get

$$closure(Range(\hat{T}-1)) = (\ker(\hat{T}^{-1}-1))^{\perp} = Q_T^{\perp}.$$

From step **6**, we have that $\mathbb{X}_2$ is closed and thus $Q_T^{\perp} \subset \mathbb{X}_2$. $Q_T^{\perp} \cup Q_T = L^2(\Omega, \mathscr{S}, \mathbb{P})$. Since both $Q_T^{\perp} \subset \mathbb{X}_2$ and $Q_T \subset \mathbb{X}_2$, we conclude that $\mathbb{X}_2 = L^2(\Omega, \mathscr{S}, \mathbb{P})$. □

Before we state and prove our next result, we need to digress and present a more convenient description of the projection P_T.

Let $f \in L^2(\Omega, \mathscr{S}, \mathbb{P})$ and set $\bar{f} = P_T f \in Q_T := L^2(\Omega, \mathscr{I}_T, \mathbb{P})$. This means that for any $g \in L^2(\Omega, \mathscr{I}_T, \mathbb{P})$ we have

$$\int_{\Omega} fg d\mathbb{P} = \int_{\Omega} \bar{f} g d\mathbb{P}.$$

In particular, if we choose g to be of the form $g = \boldsymbol{I}_S$ we deduce

$$\int_{\Omega} f \boldsymbol{I}_S d\mathbb{P} = \int_{\Omega} \bar{f} \boldsymbol{I}_S d\mathbb{P}, \;\; \forall S \in \mathscr{I}_T. \tag{2.3}$$

The orthogonal projection onto $L^2(\Omega, \mathscr{I}_T, \mathbb{P})$ is well defined only for functions in L^2. We want to show that we can make sense of such a projection even when we work in L^1. More precisely, we have the following result.

Proposition 2.9. For any function $f \in L^1(\Omega, \mathscr{S}, \mathbb{P})$, there exists a function $\hat{f} \in L^1(\Omega, \mathscr{I}_T, \mathbb{P})$ satisfying

$$\int_{\Omega} f \boldsymbol{I}_S d\mathbb{P} = \int_{\Omega} \hat{f} \boldsymbol{I}_S d\mathbb{P}, \;\; \forall S \in \mathscr{I}_T. \tag{2.4}$$

Moreover, if $\tilde{f}$ is another function in $L^1(\Omega, \mathscr{I}_T, \mathbb{P})$ satisfying (2.4), then $\tilde{f} = \hat{f}$ almost everywhere.

Proof. Define

$$\mu_f : \mathscr{I}_T \to [0, \infty), \;\; \mu_f\big[\, S \,\big] = \int_S f d\mathbb{P}.$$

Note that if $\mathbb{P}[S] = 0$, then $\mu_f\big[\, S \,\big] = 0$, i.e., μ_f is absolutely continuous relative to the restriction of $\mathbb{P}$ to the sigma-subalgebra $\mathscr{I}_T$. The Radon-Nicodym theorem implies that there exists an integrable $\mathscr{I}_T$-measurable function $\hat{f}$ such that

$$\mu_f\big[\, S \,\big] = \int_S \hat{f} d\mathbb{P}, \;\; \forall S \in \mathscr{I}_T.$$

The last equality is exactly (2.4). The uniqueness of $\hat{f}$ up to almost everywhere equality is also a consequence of the Radon-Nicodym theorem. □

Definition 2.10. For any $f \in L^1(\Omega, \mathscr{S}, \mathbb{P})$ we will denote by $\mathbb{E}\big[\, f | \mathscr{J}_T \,\big]$ a function in $L^1(\Omega, \mathscr{J}_T, \mathbb{P})$ satisfying (2.4). We will refer to $\mathbb{E}\big[\, f \mid \mathscr{J}_T \,\big]$ as the *conditional expectation of f given $\mathscr{J}_T$.*

Corollary 2.11. Suppose that $(\Omega, \mathscr{S}, \mathbb{P})$ is a probability space and $T : \Omega \to \Omega$ is a measure preserving map. Then, $\forall f \in L^1(\Omega, \mathscr{S}, \mathbb{P})$ the temporal averages $A_n f$ converge in L^1 to $\mathbb{E}\big[\, f \mid \mathscr{J}_T \,\big]$.

Proof. Let $\mathbb{X}_1$ and $\mathbb{X}_2$ be collections of functions $f \in L^1(\Omega, \mathscr{S}, \mathbb{P})$ and $g \in L^2(\Omega, \mathscr{S}, \mathbb{P})$, respectively, such that $A_n f$ converges in L^1 to some function $A_\infty f$ and $A_n g$ converges in L^2 to some function $A_\infty g$. Since $L^2(\Omega) \subset L^1(\Omega)$, we deduce that the tempral averateges that converge in L^2 also converge in L^1, i.e., $\mathbb{X}_2 \subset \mathbb{X}_1$.

We want to prove that $\mathbb{X}_1$ is closed in L^1. Let $(f_k)_{k \in \mathbb{N}}$ be a sequence in $\mathbb{X}_1$ that converges to f in L^1. We want to show that $A_n f$ converges in L^1. Fix $\varepsilon > 0$. Using the triangle inequality and the fact that $A_n : L^1 \to L^1$ is a contraction, we get

$$
\begin{aligned}
\|A_n f_k - A_m f\| &\le |\|A_n f - A_n f_k\|_1 + \|A_n f_k - A_m f_k\|_1 + \|A_m f_k - A_m f\|_1 \\
&\le \|f - f_k\|_1 + \|A_n f_k - A_m f_k\|_1 + \|f - f_k\|_1.
\end{aligned}
$$

Since $f_k \to f$ in L^1, there exists k such that $\|f - f_k\|_1 < \frac{\varepsilon}{3}$. Since $f_k \in \mathbb{X}_1$, $(A_n f_k)$ is convergent in L^1, which means that $(A_n f_k)$ is Cauchy, so we also have power over $\|A_n f_k - A_m f_k\|_1$. Choose $N = N(\varepsilon, k) > 0$ such that Set this norm $< \frac{\varepsilon}{3}$ as well. We obtain

$$
\|A_n f - A_m f\|_1 < \varepsilon, \ \ \forall m, n \ge N.
$$

This proves that $(A_n f)$ is Cauchy, so $\mathbb{X}_1$ is closed in L^1.

Since $\mathbb{X}_2 = L^2 \subset \mathbb{X}_1$ by Theorem 2.7, and L^2 is dense in L^1, we conclude that $\mathbb{X}_1 = L^1$.

Next, we need to describe the limits of these averages. By Proposition 2.5, we have that for any $f \in L^1(\Omega, \mathscr{S}, \mathbb{P})$ and any $J \in \mathscr{J}_T$,

$$
\int_\Omega f \boldsymbol{I}_J d\mathbb{P} = \int_\Omega (A_n f) \boldsymbol{I}_J d\mathbb{P}.
$$

We have

$$
\begin{aligned}
\left| \int_\Omega (A_n f) \boldsymbol{I}_J d\mathbb{P} - \int_\Omega (A_\infty f) \boldsymbol{I}_J d\mathbb{P} \right| &= \left| \int_\Omega \Big((A_n f) - A_\infty f \Big) \boldsymbol{I}_J d\mathbb{P} \right| \\
\le \int_\Omega \Big| (A_n f) - A_\infty f \Big| \boldsymbol{I}_J d\mathbb{P} &\le \int_\Omega \Big| (A_n f) - A_\infty f \Big| d\mathbb{P}
\end{aligned}
$$

We let $n \to \infty$, and we obtain

$$
\int_\Omega f \boldsymbol{I}_J d\mathbb{P} = \int_\Omega (A_\infty f) \boldsymbol{I}_J d\mathbb{P}.
$$

By the definition of conditional expectation, $A_\infty f = \mathbb{E}\big[\, f | \mathscr{I}_T \,\big]$.

$\square$

The Birkhoff Ergodic Theorem to be discussed next states that the temporal averages converge almost everywhere case. This is a more difficult result that does not follow from the L^2 and L^1 convergence cases presented in Theorem 2.8 and Corollary 2.11, respectively.

Theorem 2.12 (Birkhoff Ergodic Theorem)**.** Suppose that $(\Omega, \mathscr{S}, \mathbb{P})$ is a probability space and $T : \Omega \to \Omega$ is measure-preserving. Then, for all $f \in L^1(\Omega, \mathscr{S}, \mathbb{P})$, the temporal averages $A_n f$ converge almost surely to $\mathbb{E}\big[\, f \,|\, \mathscr{I} \,\big]$.

Proof. In a similar way as in the L^2 proof, let $\mathbb{X}_0$ be the collection of functions $f \in L^1(\Omega, \mathscr{S}, \mathbb{P})$ such that $A_n f$ converges almost surely to a function $A_\infty f$. We already know that $A_n f$ converges L^1 to $\mathbb{E}\big[\, f \,|\, \mathscr{I} \,\big]$, so the a.e. limit of $A_n f$ must be the the same because a subsequence of the L^1 convergent sequence $A_n f$ converges a.e. to this limit.

Hence, we need to show that $\mathbb{X}_0 = L^1(\Omega, \mathscr{S}, \mathbb{P})$. Our proof will involve three steps.

1. Assuming that $\mathbb{X}_0$ is closed in $L^1(\Omega, \mathscr{S}, \mathbb{P})$, we will show that $\mathbb{X}_0 = L^1(\Omega, \mathscr{S}, \mathbb{P})$.

2. We will state and prove the **Maximal Ergodic Lemma**.

3. Using the **Maximal Ergodic Lemma**, we will show that $\mathbb{X}_0$ is, in fact, closed in $L^1(\Omega, \mathscr{S}, \mathbb{P})$. Then, step 1 will imply that $\mathbb{X}_0 = L^1(\Omega, \mathscr{S}, \mathbb{P})$.

Step 1. We first prove that for any $f \in L^1$, we have $\hat{T} f - f \in \mathbb{X}_0$. For any $f \in L^1$, we have

$$A_n(\hat{T} f - f) = \frac{1}{n}(\hat{T}^n f - f).$$

Using similar reasoning as in the L^∞ case, we have

$$\|A_n(\hat{T} f - f)\|_\infty \leq \frac{2}{n}\|f\|_\infty.$$

Thus, $A_n(\hat{T} f - f) \to 0$ in L^∞. We conclude that $\hat{T} f - f \in \mathbb{X}_0$ if $f \in L^\infty$. But that is not our goal. We need to show that this happens when $f \in L^1$.

Let $f \in L^1$. Write $f = f_+ - f_-$ and note that $f_\pm \in L^1$. We have

$$(\hat{T} f)_\pm = (\hat{T} f)_+ - (\hat{T} f)_- = \hat{T} f_+ - \hat{T} f_- = \hat{T} f_\pm.$$

This means that it suffices to prove only that if $f \in L^1$ and $f \geq 0$, then $\hat{T} f - f \in \mathbb{X}_0$. Working with positive functions in L^1 is nice because we can always find a sequence of functions that converges increasingly to a positive function. This will allow us to apply the Monotone Convergence Theorem.

Construct a sequence of elementary functions f_n that converges increasingly to f. Then, Tf_n are also elementary and $Tf_n \nearrow Tf$. The Monotone Convergence theorem implies

$$\hat{T}f_n - f_n \to \hat{T}f - f.$$

in L^1. The functions f_n are bounded because they are elementary, so $\hat{T}f_n - f_n \in \mathbb{X}_0$. Assuming that $\mathbb{X}_0$ is closed in L^1 we deduce that $\hat{T}f - f \in \mathbb{X}_0$.

We have thus proved the range of $\hat{T} - 1 : L^1 \to L^1$ is contained in $\mathbb{X}_0$. Hence, its L^2-closure, which is contained in its L^1-closure, is also contained in $\mathbb{X}_0$. We have already shown in the L^2 proof that $Q_T^2 \subset \mathbb{X}_0$ and the equality

$$L^2 = \mathbb{X}_2 = Q_T^2 + (Q_T^2)^\perp = Q_T^2 + cl_{L^2}(range(\hat{T} - 1)) \subset \mathbb{X}_0.$$

The space L^2 is dense in L^1, and $\mathbb{X}_0$ is closed in L^1, so $\mathbb{X}_0 = L^1$. This completes the part of the proof that assumes that $\mathbb{X}_0$ is closed and uses this to show that $\mathbb{X}_0 = L^1$.

Now, we must prove that $\mathbb{X}_0$ is closed in L^1. In order to do this, we will need the following lemma, which is **step 2** of the proof.

Proposition 2.13 (Maximal Ergodic Lemma)**.** Given $f \in L^1(\Omega, \mathscr{S}, \mathbb{P})$, set

$$M(f)(\omega) := \sup_{n \geq 1} A_n f(\omega).$$

For all $\lambda > 0$ and all $f \in L^1(\Omega, \mathscr{S}, \mathbb{P})$, we have

$$\lambda \mathbb{P}[\{M(|f|) > \lambda\}] \leq \|f\|_1.$$

Proof. It suffices to show that

$$\int_{\{M(f)>0\}} f d\mathbb{P} \geq 0$$

for all $f \in L^1(\Omega, \mathscr{S}, \mathbb{P})$. This is enough because, if we let $f = g - \lambda$ with $\lambda > 0$, then we can integrate over the set $\{M(f) > \lambda\}$ and get

$$\|g\|_1 \geq \int_{\{M(f)>\lambda\}} g d\mathbb{P} \geq \lambda \mathbb{P}\big[\, \{M(g) > \lambda\}\,\big],$$

as desired.

Set

$$Q := \{M(f) > 0\} \subset \Omega,$$

$$S_n(f) := \sum_{j=0}^{n-1} \hat{T}^j f,$$

$M_n(f) := \max_{1\leq k\leq n} S_k(f)$, $F_n := M_n(f)$, $F_n^+ := \max(F_n, 0)$, and $Q_n = \{F_n \geq 0\}$. Since $F_n \nearrow M(f)$ as $n \to \infty$, we get that $Q_n \nearrow Q$. Thus, we want to show that

$$\int_{Q_n} M(f) \geq 0$$

for all n. Observe that the map $f \to \hat{T}f$ is monotone, so $f_0 \leq f_1 \Rightarrow \hat{T}f_0 \leq \hat{T}f_1$. For $1 \leq k \leq m$, we have

$$S_{k-1}(f) \leq F_{m-1} \leq F_{m-1}^+$$

and

$$S_k(f) = f + \hat{T}S_{k-1}(f) \leq f + \hat{T}F_{m-1} \leq f + \hat{T}F_{m-1}^+.$$

Thus, $\forall m \in \mathbb{N}$,

$$F_{m-1} \leq F_m \leq f + \hat{T}F_{m-1}^+.$$

Rewriting, we obtain, $\forall n \in \mathbb{N}$,

$$f \geq F_n - \hat{T}F_n^+.$$

Integrating over Q_n, we get

$$\int_{Q_n} f \geq \int_{Q_n} F_n - \int_{Q_n} \hat{T}F_n^+.$$

Note that $\hat{T}F_n^+ \geq 0$ on Ω, and $F_n = F_n^+$ on Q_n with $F_n^+ = 0$ on $\Omega \backslash Q_n$. Also note that $\int_\Omega \hat{T}F_n^+ = \int_\Omega F_n^+$ because T is measure-preserving. Hence,

$$\int_{Q_n} F_n - \int_{Q_n} \hat{T}F_n^+ \geq \int_{Q_n} F_n^+ - \int_\Omega \hat{T}F_n^+ = \int_\Omega F_n^+ - \int_\Omega F_n^+ = 0.$$

Let $f = g - \lambda$ with $\lambda > 0$. We have

$$\|g\|_1 \geq \int_{\{M(f)>\lambda\}} g d\mathbb{P} \geq \lambda \mathbb{P}\big[\,\{M(g) > \lambda\}\,\big],$$

as desired. □

Now, we may return to the proof of the Birkhoff Ergodic Theorem. We will show that $\mathbb{X}_0$ is closed in L^1 using the Maximal Ergodic Lemma. This is **step 3**.

We have to prove that if (f_k) is a sequence in $\mathbb{X}_0$ that converges in L^1 to a function $f \in L^1$, then $f \in \mathbb{X}_0$, i.e., $A_n f$ converges a.e.. We will show that the sequence $(A_n f)$ is a.e. Cauchy. Recall the measure-theoretic definition of a Cauchy sequence. We want to show that, for any $\epsilon > 0$,

$$\mathbb{P}\Big[\bigcup_N \bigcap_{m,n>N} \{|A_n(f) - A_m| < \epsilon\}\Big] = 1.$$

Set

$$X_N(f, \epsilon) = \bigcap_{m,n>N} \{|A_n(f) - A_m| < \epsilon\},$$

and observe that $X_N(f,\epsilon) \subset X_{N'}(f,\epsilon)$ if $N < N'$. Passing this to the definition of limits, we get that

$$\mathbb{P}\big[\bigcup_N \bigcap_{m,n>N} \{|A_n(f) - A_m| < \epsilon\}\big] = 1 \iff \lim_{N\to\infty} \mathbb{P}\big[\, X_N(f,\epsilon)\,\big] = 1.$$

Fix $\epsilon > 0$. Using similar logic as in the L^2 proof and repeatedly applying the triangle inequality, we deduce that

$$\begin{aligned} |A_n(f) - A_m(f)| &\leq |A_n(f) - A_n(f_k)| + |A_n(f_k) - A_m(f_k)| + |A_m(f_k) - A_m(f)| \\ &\leq |A_n(|f - f_k|)| + |A_n(f_k) - A_m(f_k)| + |A_m(|f - f_k|)| \\ &\leq 2M(|f_k - f|) + |A_n(f_k) - A_m(f_k)|. \end{aligned}$$

Thus,

$$\big(\{2M(|f_k - f|) < \frac{\epsilon}{2}\} \cap X_N(f_k, \frac{\epsilon}{2})\big) \subset X_N(f,\epsilon)$$

for all N, k. Let $N \to \infty$ and take the measure of both sides. We have

$$\lim_{N\to\infty} \mathbb{P}\big[\, X_N(f,\epsilon)\,\big] \geq \lim_{N\to\infty} \mathbb{P}\big[\,\{2M(|f_k - f|) < \frac{\epsilon}{2}\} \cap X_N(f_k, \frac{\epsilon}{2})\,\big].$$

The inclusion-exclusion principle implies that

$$\begin{aligned} \mathbb{P}\big[\,\{2M(|f_k - f|) < \frac{\epsilon}{2}\} \cap X_N(f_k, \frac{\epsilon}{2})\,\big] &= \mathbb{P}\big[\,\{2M(|f_k - f|) < \frac{\epsilon}{2}\}\,\big] + \mathbb{P}\big[\, X_N(f_k, \frac{\epsilon}{2})\,\big] \\ &\quad - \mathbb{P}\big[\,\{2M(|f_k - f|) < \frac{\epsilon}{2}\} \cup X_N(f_k, \frac{\epsilon}{2})\,\big]. \end{aligned}$$

Since $f_k \in \mathbb{X}_0$, $\big(A_n(f_k)\big)_{n\geq 1}$ is almost surely a Cauchy sequence for any k. Thus,

$$\lim_{N\to\infty} \mathbb{P}\big[\,\{2M(|f_k - f|) < \frac{\epsilon}{2}\} \cup X_N(f_k, \frac{\epsilon}{2})\,\big] = \lim_{N\to\infty} \mathbb{P}\big[\, X_N(f_k, \frac{\epsilon}{2})\,\big] = 1.$$

Eliminating these terms, we find that

$$\lim_{N\to\infty} \mathbb{P}\big[\,\{2M(|f_k - f|) < \frac{\epsilon}{2}\} \cap X_N(f_k, \frac{\epsilon}{2})\,\big] = \lim_{N\to\infty} \mathbb{P}\big[\,\{2M(|f_k - f|)\}\,\big].$$

Finally, we invoke the maximal ergodic lemma. Let $\lambda = \frac{\epsilon}{4}$.

$$\mathbb{P}[\{2M(|f_k - f|) \geq \frac{\epsilon}{2}\}] \leq \frac{4}{\epsilon}\|f - f_k\|_1.$$

so that

$$\begin{aligned} \mathbb{P}\big[\,\{2M(|f_k - f|) < \frac{\epsilon}{2}\}\,\big] &= 1 - \mathbb{P}\big[\,\{2M(|f_k - f|) \geq \frac{\epsilon}{2}\}\,\big] \\ &\geq 1 - \leq \frac{4}{\epsilon}\|f - f_k\|_1, \quad \forall k \end{aligned}$$

Let $k \to \infty$. Since $f_k \to f$, we conclude that

$$\lim_{N\to\infty} \mathbb{P}\big[\, X_N(f,\epsilon)\,\big] = 1.$$

Hence $(A_n f)$ is a.e. Cauchy and thus converges a.e. to $A_\infty f$ so $f \in \mathbb{X}_0$. Thus, $\mathbb{X}_0$ is closed in L^1.

Step 1 implies that $\mathbb{X}_0 = L^1(\Omega, \mathscr{S}, \mathbb{P})$. This completes the proof. □

Corollary 2.14. Suppose that T is an ergodic map on the probability space $(\Omega, \mathscr{S}, \mathbb{P})$. Then for every measurable set $S \in \mathscr{S}$, there exists a negligible subset $\mathscr{N} \subset \Omega$ such that for any $\omega \in \Omega \setminus S$ we have

$$\lim_{n\to\infty} \frac{1}{n} \#\big\{ \, k; \;\; T^k(\omega) \in S, \;\; 0 \le k < n \, \big\} = \mathbb{P}\big[\, S \,\big]. \tag{2.5}$$

Proof. Let $f = \boldsymbol{I}_S$. Observe that for any $\omega \in \Omega$, we have

$$A_{n-1}[f](\omega) \frac{1}{n} \#\big\{ \, k; \;\; T^k(\omega) \in S, \;\; 0 \le k < n \, \big\}$$

and

$$\int_\Omega f d\mathbb{P} = \mathbb{P}\big[\, S \,\big].$$

Hence, the equality (2.5) is a special case of Theorem 2.12. □

Think of $T^n\omega$ as the location at time n of a particle initially located at ω, and regard S as a small hole in Ω and N as an infinitesimally tiny part of Ω. The quantity

$$\frac{1}{n} \#\big\{ \, k; \;\; T^k(\omega) \in S, \;\; 0 \le k < n \, \big\}$$

is the fraction of the first n epochs the particle spends in the hole. It is usually referred to as the frequency. The Birkhoff Ergodic Theorem tells as that in the long run and for most starting points, the fraction of the time the particle spends in the hole approaches the fraction of the volume of Ω occupied by S.

To put it differently, almost any orbit of T spends equal time in regions of equal size, i.e., it is uniformly distributed, or equidistributed.

Let us observe that, conversely, if a measure preserving map T satisfies the above equidistribution property, then it has to be ergodic. Indeed, suppose that S is a T-invariant set. Let $\mathscr{N}$ be a negligible set as in (2.5). Then for any $\omega \in (\Omega \setminus S) \setminus \mathscr{N}$ the orbit $\boldsymbol{O}_{T,\omega}$ does not intersect S since S is invariant. In this case the left-hand-side of (2.5) is equal to zero so $\mathbb{P}\big[\, S \,\big] = 0$. Thus if S is invariant and its complement $\Omega \setminus S$ is not negligible, then S must be so. Hence T is ergodic.

3. Examples of ergodic maps

The strength or range of a theory is measured by the amount of situations it applies to. To give an idea about the range of applicability of ergodic theory, we discuss a few examples of ergodic maps.

Example 3.1 (Ergodic rotations)**.** Fix $\phi \in (0, 2\pi)$ and let R_ϕ be the counterclockwise rotation in the plane of angle ϕ about the origin. This will induce a transformation of the unit circle $S^1 := \{z \in \mathbb{C}; |z| = 1\}$ which is measure-preserving with respect to the canonical probability measure on S^1,

$$\mu[d\theta] = \frac{1}{2\pi} d\theta,$$

where $d\theta$ is the change in angle for a given rotation. We want to show that if $2\pi\phi$ is irrational, then R_ϕ is ergodic.

The associated Koopman operator is

$$\hat{R}_\phi : L^2(S^1, \mu) \to L^2(S^1, \mu), \hat{R}_\phi f(\theta) = f(\theta + \phi).$$

Above $L^2(S^1)$ denotes the space of complex valued functions

$$f : [0, 2\pi] \to \mathbb{C} \text{ such that } \int_0^{2\pi} |f(\theta)|^2 d\theta < \infty.$$

For $n \in \mathbb{Z}$, we define $e_n \in L^2(S^1)$, $e_n(\theta) = e^{\boldsymbol{ni\theta}}$.

The collection $\{e_n\}_{n\in\mathbb{Z}}$ is a complete orthonormal system of the Hilbert space $L^2(S^1)$. Note that

$$\hat{R}_\phi e_n(\theta) = e^{\boldsymbol{in(\theta+\phi)}} = e^{\boldsymbol{in\phi}} e_n(\theta).$$

Hence e_n is an eigenfunction of $\hat{R}_\phi$ corresponding to the eigenvalue $e^{\boldsymbol{in\phi}}$. Thus, the eigenspace corresponding to the eigenvalue 1 of $\hat{R}_\phi$ is

$$\ker\big(1 - \hat{R}_\phi\big) = \operatorname{span}\Big\{e_n, \frac{n\phi}{2\pi} \in \mathbb{Z}\Big\}.$$

Hence, if $\frac{\phi}{2\pi} \in \mathbb{R} \setminus \mathbb{Q}$, the space $\ker(1 - \hat{R}_\phi)$ is one-dimensional, and thus R_ϕ is ergodic according to Proposition 2.5 . □

From Corollary 2.14, we deduce the following celebrated result of H. Weyl [7].

Theorem 3.2 (Weyl's Equidistribution Theorem)**.** The transformation $\hat{R}_\phi$ is ergodic if and only if $\frac{\phi}{2\pi}$ is irrational. Moreover, for any open arc $A \subset S^1$ from angles θ_0 to $\theta_1 \in S^1$ and almost every $\theta \in S^1$, we have

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbf{I}_A(\theta + k\phi) \to \frac{\theta_1 - \theta_0}{2\pi}$$

almost surely, where $\mathbf{I}_A$ is the indicator function over A. □

Mixing is closely related to ergodicity.

Definition 3.3. Let $T : (\Omega, \mathscr{S}, \mathbb{P}) \to (\Omega, \mathscr{S}, \mathbb{P})$ be a measure preserving self-map of a probability space. We say that T is *mixing* if, for any $A, B \in \mathscr{S}$

$$\lim_{n\to\infty} \mathbb{P}\big[\,(T^{-n}A) \cap B\,\big] = \mathbb{P}\big[\,A\,\big]\mathbb{P}\big[\,B\,\big]. \tag{3.1}$$

□

Let us observe that (3.1) can be rewritten as

$$\lim_{n\to\infty}\int_\Omega (\hat{T}^n \boldsymbol{I}_A)\boldsymbol{I}_B d\mathbb{P} = \left(\int_\Omega \boldsymbol{I}_A d\mathbb{P}\right)\left(\int_\Omega \boldsymbol{I}_B d\mathbb{P}\right). \tag{3.2}$$

Proposition 3.4. A mixing map $T : (\Omega, \mathscr{S}, \mathbb{P}) \to (\Omega, \mathscr{S}, \mathbb{P})$ is ergodic.

Proof. Suppose that A is a quasi-invariant set. Then $B := A^c = \Omega \setminus A$ is also quasi-invariant. Hence $\hat{T}^n \boldsymbol{I}_A = \boldsymbol{I}_A$, $\forall n \in \mathbb{N}$ so that

$$\int_\Omega (\hat{T}^n \boldsymbol{I}_A)\boldsymbol{I}_{A^c} d\mathbb{P} = \int_\Omega \boldsymbol{I}_A \boldsymbol{I}_{A^c} d\mathbb{P} = 0, \ \ \forall n \in \mathbb{N}.$$

We deduce from (3.2) that $\mathbb{P}\big[\,A\,\big]\mathbb{P}\big[\,A^c\,\big] = 0$, i.e.,

$$\mathbb{P}\big[\,A\,\big]\big(\,1 - \mathbb{P}\big[\,A\,\big]\,\big) = 0$$

so that $\mathbb{A}\big[\,A\,\big] = 0$ or $\mathbb{P}\big[\,A\,\big] = 1$. This proves that T is ergodic.

□

Let us investigate the condition (3.2). We say that the pair

$$(f, g) \in L^2(\Omega, \mathscr{S}, \mathbb{P}) \times L^2(\Omega, \mathscr{S}, \mathbb{P})$$

satisfies the mixing condition if

$$\lim_{n\to\infty}\int_\Omega (\hat{T}^n f) g d\mathbb{P} = \left(\int_\Omega f\, d\mathbb{P}\right)\left(\int_\Omega g\, d\mathbb{P}\right). \tag{3.3}$$

Thus, T is mixing if and only if, for any $A, B \in \mathscr{S}$, the pair $(\boldsymbol{I}_A, \boldsymbol{I}_B)$ satisfies the mixing condition.

For $f \in L^2(\Omega, \mathscr{S}, \mathbb{P})$ we denote by $\mathscr{R}_f$ the collection of functions $g \in L^2(\Omega, \mathscr{S}, \mathbb{P})$ such that (f, g) satisfies the mixing condition. Similarly, for $g \in L^2(\Omega, \mathscr{S}, \mathbb{P})$ we denote by $\mathscr{L}_g$ the collection of functions $f \in L^2(\Omega, \mathscr{S}, \mathbb{P})$ such that (f, g) satisfies the mixing condition.

Lemma 3.5. For any $f, g \in L^2(\Omega, \mathscr{S}, \mathbb{P})$ the collections $\mathscr{L}_g$ and $\mathscr{R}_f$ are *closed* vector subspaces of $L^2(\Omega, \mathscr{S}, \mathbb{P})$.

Proof. First, we will show that $\mathscr{L}_g$ and $\mathscr{R}_f$ are vector subspaces of $L^2(\Omega, \mathscr{S}, \mathbb{P})$. For any $f_1, f_2 \in \mathscr{L}_g$, we have

$$\int_\Omega \big(\,\hat{T}^n(f_1 + f_2)\,\big) g\, d\mathbb{P} = \int_\Omega \big(\,\hat{T}^n(f_1)\,\big) g\, d\mathbb{P} + \int_\Omega \big(\,\hat{T}^n(f_2)\,\big) g\, d\mathbb{P}.$$

so that

$$\lim_{n\to\infty}\int_\Omega \big(\,\hat{T}^n(f_1 + f_2)\,\big) g\, d\mathbb{P} = \lim_{n\to\infty}\int_\Omega \big(\,\hat{T}^n(f_1)\,\big) g\, d\mathbb{P} + \lim_{n\to\infty}\int_\Omega \big(\,\hat{T}^n(f_2)\,\big) g\, d\mathbb{P},$$

because $\big(\,f_1, g\,\big)$ and $\big(\,f_2, g\,\big)$ satisfy the mixing condition. Therefore, $\mathscr{L}_g$ is a vector subspace. A similar argument proves that $\mathscr{R}_f$ is a vector subspace.

Let (f_k) be a converging sequence of functions in $\mathscr{L}_g$. Set

$$f_\infty = \lim_{k\to\infty} f_k.$$

For any $k \in \mathbb{N}$, we have

$$\lim_{n\to\infty} \int_\Omega \big(\hat{T}^n f_k\big) g \, d\mathbb{P} = \int_\Omega f_k d\mathbb{P} \int_\Omega g \, d\mathbb{P}.$$

Note that

$$\begin{aligned} \left| \int_\Omega \big(\hat{T}^n f_k\big) g \, d\mathbb{P} - \int_\Omega \big(\hat{T}^n f_\infty\big) g \, d\mathbb{P} \right| &= \left| \int_\Omega \big(\hat{T}^n (f_k - f_\infty)\big) g \, d\mathbb{P} \right| \\ &\le \|\hat{T}^n (f_k - f_\infty)\|_{L^2} \|g\|_{L^2} = \|f_k - f_\infty\|_{L^2} \|g\|_{L^2}, \end{aligned} \tag{3.4}$$

since $\hat{T}$ is an isometry.

On the other hand, we know that $f_k \to f_\infty$ in L^2. Therefore, for any $\varepsilon > 0$, there exists $M_1 \in \mathbb{N}$ such that $\forall k > M_1$, we have

$$\forall k > M_1, \;\; \|f_k - f_\infty\|_{L^2} \|g\|_{L^2} < \frac{\varepsilon}{3}. \tag{3.5}$$

Hence

$$\left| \int_\Omega \big(\hat{T}^n f_k\big) g \, d\mathbb{P} - \int_\Omega \big(\hat{T}^n f_\infty\big) g \, d\mathbb{P} \right| < \frac{\varepsilon}{3}. \tag{3.6}$$

Observe that for $k > M_1$

$$\begin{aligned} \left| \int_\Omega f_k \, d\mathbb{P} \int_\Omega g \, d\mathbb{P} - \int_\Omega f_\infty g \, d\mathbb{P} \right| &= \left| \int_\Omega \big(f_\infty - f_k\big) \, d\mathbb{P} \int_\Omega g \, d\mathbb{P} \right| \\ &\le \|f_\infty - f_k\|_{L^2} \|g\|_{L^2} \overset{(3.5)}{<} \frac{\varepsilon}{3}. \end{aligned} \tag{3.7}$$

Fix $k_0 > M_1$. Since (f_{k_0}, g) satisfies the mixing condition, we have that there exists $N \in \mathbb{N}$, depending on k_0 and ε, such that $\forall n > N$,

$$\left| \int_\Omega \big(\hat{T}^n f_{k_0}\big) g \, d\mathbb{P} - \int_\Omega f_{k_0} \, d\mathbb{P} \int_\Omega g \, d\mathbb{P} \right| < \frac{\varepsilon}{3}. \tag{3.8}$$

We have, using the triangle inequality,

$$\begin{aligned} \left| \int_\Omega \big(\hat{T}^n f_\infty\big) g \, d\mathbb{P} - \int_\Omega f_\infty \, d\mathbb{P} \int_\Omega g \, d\mathbb{P} \right| &\le \left| \int_\Omega \big(\hat{T}^n f_\infty\big) g \, d\mathbb{P} - \int_\Omega \big(\hat{T}^n f_{k_0}\big) g \, d\mathbb{P} \right| \\ &\quad + \left| \int_\Omega \big(\hat{T}^n f_{k_0}\big) g \, d\mathbb{P} - \int_\Omega f_{k_0} d\mathbb{P} \int_\Omega g \, d\mathbb{P} \right| \\ &\quad + \left| \int_\Omega f_{k_0} d\mathbb{P} \int_\Omega g \, d\mathbb{P} - \int_\Omega f_\infty g \, d\mathbb{P} \right|. \end{aligned} \tag{3.9}$$

By equations (3.6), (3.7), and (3.8), $\forall n > N$,

$$\left| \int_\Omega \big(\hat{T}^n f_\infty\big) g \, d\mathbb{P} - \int_\Omega f_\infty \, d\mathbb{P} \int_\Omega g d\mathbb{P} \right| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,$$

and thus,

$$\lim_{n\to\infty}\int_\Omega (\hat{T}^n f_\infty) g\, d\mathbb{P} = \left(\int_\Omega f_\infty\, d\mathbb{P}\right)\left(\int_\Omega g\, d\mathbb{P}\right)$$

showing that $f_\infty \in \mathscr{L}_g$. Since $\lim_{n\to\infty}(f_k) \in \mathscr{L}_g$, $\mathscr{L}_g$ is a closed vector subspace of L^2. A similar argument shows that $\mathscr{R}_f$ is a closed vector subspace.

□

Proposition 3.6. Let $T : (\Omega, \mathscr{S}, \mathbb{P}) \to (\Omega, \mathscr{S}, \mathbb{P})$ be a measure preserving self-map of the probability space $(\Omega, \mathscr{S}, \mathbb{P})$. Then the following are equivalent.

(i) The map T is mixing (Definition 3.3).

(ii) For any $f, g \in L^2(\Omega, \mathscr{S}, \mathbb{P})$, the pair (f, g) satisfies the mixing condition (3.3).

(iii) There exists a collection $\mathscr{C} \subset L^2(\Omega, \mathscr{S}, \mathbb{P})$ with the following properties.

(a) Any pair $(f, g) \in \mathscr{C} \times \mathscr{C}$ satisfies (3.3).

(b) The vector space spanned by $\mathscr{C}$ is dense in $L^2(\Omega, \mathscr{S}, \mathbb{P})$.

Proof. (i) ⇒(ii) The mixing condition in (3.3) states that the pair $(\boldsymbol{I}_A, \boldsymbol{I}_B)$ satisfies the mixing condition for any $A, B \in \mathscr{S}$. We deduce from Lemma 3.5 that for any f, g in the L^2-closure of $\mathrm{Elem}(\Omega, \mathscr{S})$,

$$\lim_{n\to\infty}\int_\Omega \big(fT^n\big) g\, d\mathbb{P} = \int_\Omega f\, d\mathbb{P}\int_\Omega g\, d\mathbb{P}.$$

Since $\mathrm{Elem}(\Omega, \mathscr{S})$ is dense in $L^2(\Omega, \mathscr{S}, \mathbb{P})$, the above equality holds for all $f \in L^2(\Omega, \mathscr{S}, \mathbb{P})$.

(ii) ⇒(i) The indicator function of a measurable set is L^2. By (ii), $\forall A, B \in \mathscr{S}$, the pair $(\boldsymbol{I}_A, \boldsymbol{I}_B)$ satisfies the mixing condition, so T is mixing.

(ii) ⇒ (iii) Given (ii), we take $\mathscr{C} = L^2(\Omega, \mathscr{S}, \mathbb{P})$.

(iii) ⇒ (ii) Let $f \in \mathscr{C}$. From part (a) of (iii), we have that $g \in \mathscr{R}_f, \forall g \in \mathscr{C}$. Lemma 3.5 implies that $\mathrm{closure}(\mathrm{span}\mathscr{C}) \subset \mathscr{R}_f$ since $\mathscr{R}_f$ is a closed subspace. We also have that $\mathrm{closure}(\mathrm{span}\mathscr{C}) = L^2$ because $\mathrm{span}(\mathscr{C})$ is dense in L^2. Hence $\mathscr{R}_f = L^2$.

Hence, for any $f \in \mathscr{C}$ and any $g \in L^2$, the pair (f, g) satisfies the mixing condition, i.e. $\forall g \in L^2$, $\mathscr{C} \subset \mathscr{L}_g$. $\mathscr{L}_g$ is a closed subspace of L^2, so $\mathrm{closure}(\mathrm{span}\mathscr{C}) \subset \mathscr{L}_g$. We deduce that $\mathscr{L}_g = L^2$, $\forall g \in L^2$. □

Example 3.7. Let us revisit the system described in Example 1.7. We previously showed that the map $T : [0, 1) \to [0, 1)$,

$$T(x) = 2x - \lfloor 2x \rfloor = 2x \text{ mod } 1.$$

is measure-preserving with respect to the Lebesgue measure $\boldsymbol{\lambda}$. In fact, T is mixing. Let $S^1 = \{|z| = 1\}$. Note that we can view T as a map $S^1 \to S_1$, with $T(z) = z^2$, and $z = e^{\boldsymbol{i}\theta}$, $\theta = 2\pi x$. In this case, the Lebesgue measure $\boldsymbol{\lambda}$ corresponds to the canonical probability measure on S^1,

$$\mu[d\theta] = \frac{1}{2\pi} d\theta.$$

For $n \in \mathbb{Z}$, let $e_n \in L^2(S^2, u)$, $e_n(\theta) = e^{n\boldsymbol{i}\theta}$. Set

$$\mathscr{C} := \left\{ e_n; \ \ n \in \mathbb{Z} \right\} \subset L^2(S^1).$$

The Weierstrass approximation theorem implies that $\mathscr{C}$ spans a dense subspace of $L^2(S^1)$. We want to show that $\mathscr{C}$ also satisfies condition (iii.a) in Proposition 3.6.

If we write $z = e^{\boldsymbol{i}\theta}$, then $e_j(z) = z^j$, $\hat{T}e^j(z) = z^{2j} = e_{2j}(z)$, and

$$\hat{T^n} e_j = e_{2^n j}, \ \ \forall j \in \mathbb{Z}.$$

Observe that $e_j e_k = e_{j+k}$, $\forall j, k \in \mathbb{Z}$. We set

$$I_m := \int_{S^1} e_m(\theta) \mu\big[\, d\theta \,\big], \ \ m \in \mathbb{Z}.$$

Thus, the pair (e_j, e_k) satisfies the mixing condition iff

$$\int_{S^1} (\hat{T} e_j) e_k d\mu = I_j \cdot I_k \iff \lim_{n\to\infty} I_{2^n j+k} = I_j \cdot I_k. \tag{3.10}$$

Note that

$$I_0 = \int_{S^1} 1 d\mu = 1$$

since μ is a probability measure. For $m \neq 0$ we have

$$I_m = \frac{1}{2\pi} \int_{S^1} e^{m\boldsymbol{i}\theta} d\theta = \frac{1}{2\pi} \left(\frac{e^{m\boldsymbol{i}\theta}}{m\boldsymbol{i}} \Big|_{\theta=0}^{\theta=2\pi} \right) = 0.$$

This proves that the collection $\mathscr{C}$ also satisfies the condition (iii.a), and we deduce from the implication (iii) $\Rightarrow$ (i) in Proposition 3.6 that T is mixing.□

Remark 3.8. The irrational rotation in Example 3.1 is ergodic but not mixing. A simple computation shows that the pair (e_1, e_{-1}) *does not satisfy the mixing condition.* □

4. Further Reading

The Ergodic Theorem marked the birth of a new area of mathematics. In the nearly one hundred years from its discovery, ergodic theory has matured into a field that has found applications in diverse areas, such as dynamical systems, geometry, probability, information theory, and number theory. During the last two decades, research in ergodic theory earned some of the most prestigious mathematical awards, including the Fields Medal (E. Lindenstrauss 2010, P. Avila 2014) and the Abel Prize (Y. Sinai 2014, G. Margulis 2020).

To the reader curious about advances in theory after the discovery of the Ergodic Theorem, we recommend the classical monographs [1, 4] as good entry points to further study.

References

[1] V. I. Arnold, A. Avez: *Ergodic Problems of Classical Mechanics*, Addison Wesley, 1968.
[2] P. Billingsley: *Ergodic Theory and Information*, John Wiley & Sons, 1965.
[3] P. Billingsley: *Probability and Measure*, 3rd Edition, John Wiley & Sons, 1995.
[4] I. P. Cornfeld, S. V. Fomin, Ya. G. Sinai: *Ergodic Theory*, Springer Verlag, 1982.
[5] P. Halmos: *Lectures on Ergodic Theory*, Chelsea Publishing, 1956.
[6] J. Hawkins: *Ergodic Dynamics. From Basic Theory to Applications*, Springer Verlag, 2021.
[7] H. Weyl: *Über die Gleichverteilung von Zahlen mod Eins*, Math. Ann, **77**(1916), 313-352.

Department of Mathematics, University of Notre Dame, Notre Dame, IN 46556-4618.

Email address: aclay2@nd.edu