CLASSIC EXAMPLES IN DYNAMICAL SYSTEMS

SHANNA GOLDMAN

Abstract. This paper is intended to be an undergraduate-accessible introduction to the following classic examples of dynamical systems: rotations and expanding maps on S^1, shift maps on infinite sequences, quadratic maps, and the horseshoe map. We will investigate what happens over time when we iterate these transformations. We assume the reader knows point-set topology, and has taken an introductory measure theory class. The goal is to introduce different properties of dynamical systems and notions of equivalence, and use them to compare our examples.

First, we provide relevant background on measure theory and probability. Then we define rotations and expanding maps on S^1. We dedicate an entire section to defining the shift map and the space of infinite sequences. Then we look at examples of measure-preserving transformations and measure-theoretic isomorphism. We define ergodicity and mixing, which are ways of classifying how points get distributed in a dynamical system. We conclude the paper with a section on symbolic dynamics, which is the process of analyzing transformations by relating them to shift maps. Specifically, we will use symbolic dynamics to easily find periodic points of quadratic maps and the horseshoe map.

Contents

Date: August 31, 2024.

Introduction

A *dynamical system* is a transformation $T : X \to X$. We want to study what happens to X (which we call a *phase space*) when we apply T over and over again. For notation, we use T^n to mean $T \circ T \circ \cdots \circ T$. Think of each application of T as a step forward in time. For instance, if $x \in X$, then $T^3(x)$ represents where x is at time 3. Also, for a set $A \subseteq X$, we think of the pre-image $T^{-1}(A)$ as A one unit in the past.

When studying what happens to a dynamical system over time, there are some natural questions to ask. We may wonder what the *fixed points* are, the points $x \in X$ such that $T(x) = x$. Similarly, the *periodic points* are the points $x \in X$ such that there exists $n \in \mathbb{N}$ with $T^n(x) = x$. Another question is what are the *invariant sets*, the sets Λ such that $T^{-1}(\Lambda) = \Lambda$. For example, the set of all periodic points of T is an invariant set. Furthermore, if we put a probability measure on our dynamical system, then we can ask what probabilistically happens to the phase space. In this paper, we will investigate examples of dynamical systems that have interesting answers to each of these questions.

1. Preliminaries and Notation

1.1. **Measure Theory.**

This paper assumes the reader has had a formal introduction to measure theory. This section is meant to be a review of key definitions. The content in this section is from [8], [9], [12], and [17].

Notation 1.1. Let f and g be measurable transformations. We say f and g are *equal almost everywhere* if the set $\{x \mid f(x) \neq g(x)\}$ has measure 0. Throughout this paper, we will write $f = g$ a.e. to mean f and g are equal almost everywhere.

We also have a notion of "almost everywhere" for set equality. Define the *symmetric difference* of sets A and B (denoted $A\Delta B$) to be $(A\backslash B) \cup (B\backslash A)$. We say A and B are *equal mod* 0 if the measure of $A\Delta B$ is 0.

Definition 1.2. A *σ-algebra* on a set X is a collection $\mathscr{B}$ of subsets of X such that

1. $\varnothing \in \mathscr{B}$,
2. if $B \in \mathscr{B}$, then $B^c \in \mathscr{B}$, and
3. if $B_1, B_2, \cdots \in \mathscr{B}$, then $\bigcup_{n=1}^{\infty} B_n \in \mathscr{B}$.

We call $(X, \mathscr{B})$ a *measurable space*.

Example 1.3. For any set X, the power set 2^X is a σ-algebra.

Example 1.4. The set of all Lebesgue measurable sets in $\mathbb{R}$ is a σ-algebra.

Definition 1.5. Let $\mathcal{F}$ be a collection of subsets of X. Then $\sigma(\mathcal{F})$, the *σ-algebra generated by* $\mathcal{F}$, is the intersection of all σ-algebras that have $\mathcal{F}$ as a subset. Thus, $\sigma(\mathcal{F})$ is the smallest σ-algebra that has $\mathcal{F}$ as a subset.

Example 1.6. The Borel σ-algebra is the σ-algebra generated by the open sets of $\mathbb{R}$.

Definition 1.7. Let $\mathscr{B}$ be a σ-algebra on X. We say $\mu : \mathscr{B} \to [0, \infty]$ is a *measure* if

1. $\mu(\varnothing) = 0$, and

(2) for pairwise disjoint $B_1, B_2 \dots \in \mathscr{B}$, we have
$$\mu(\bigcup_{k=1}^{\infty} B_k) = \sum_{k=1}^{\infty} \mu(B_k).$$

We call $(X, \mathscr{B}, \mu)$ a *measure space.*

Example 1.8. The Lebesgue measure on $\mathbb{R}$ is a measure.

Example 1.9. On a σ-algebra $\mathscr{B}$, the *counting measure* $c : \mathscr{B} \to [0, 1]$ is defined as
$$c(B) := \begin{cases} |B| & \text{if } B \text{ is finite} \\ \infty & \text{if } B \text{ infinite.} \end{cases}$$

Example 1.10. Given a measurable transformation $f : (X, \mathscr{A}) \to (Y, \mathscr{B})$ with measure μ on $(X, \mathscr{A})$, we define the *pushforward measure* $f_*(\mu)$ on $(Y, \mathscr{B})$ by
$$f_*(\mu)(B) := \mu(f^{-1}(B)).$$

We are often in a position where we want to study a measure μ on a σ-algebra $\mathscr{B}$, but we don't understand what all the sets in $\mathscr{B}$ look like. However, we can usually find a family of well-understood sets $\mathcal{F}$ such that $\sigma(\mathcal{F}) = \mathscr{B}$. The following definitions and theorem will outline the circumstances in which we can apply what we know about μ on $\mathcal{F}$ to all of $\mathscr{B}$.

Definition 1.11. A *semi-algebra* on a set X is a collection $\mathcal{S}$ of subsets of X such that

(1) $\varnothing \in \mathcal{S}$,
(2) for all $S_1, S_2 \in \mathcal{S}$, $S_1 \cap S_2 \in \mathcal{S}$,
(3) for all $S \in \mathcal{S}$, S^c is a finite disjoint union of sets in $\mathcal{S}$.

Example 1.12. Let $\mathcal{I}$ denote the collection of all intervals (bounded or unbounded) in $\mathbb{R}$. Then $\mathcal{I} \cup \{\varnothing\}$ is a semi-algebra.

Definition 1.13. Let $\mathcal{S}$ be a semi-algebra on X. We say $\mu : \mathcal{S} \to [0, \infty]$ is a *pre-measure* if

(1) $\mu(\varnothing) = 0$,
(2) for pairwise disjoint $S_1, \dots, S_n \in \mathcal{S}$ such that $\bigcup_{k=1}^{n} S_k \in \mathcal{S}$, we have
$$\mu(\bigcup_{k=1}^{n} S_k) = \sum_{k=1}^{n} \mu(S_k).$$

The definition of a pre-measure is basically the same as the definition of a measure, except it is adjusted to make sense for semi-algebras rather than σ-algebras, because the pre-measure definition allows for the domain to not be closed under unions. A measure restricted to a semi-algebra is a pre-measure.

Example 1.14. The length function for intervals (where an interval from a to b has length $b - a$ and an unbounded interval has length ∞) is a pre-measure on $(\mathbb{R}, \mathcal{I} \cup \{\varnothing\})$.

Definition 1.15. Say μ is a pre-measure (or measure) on a semi-algebra (or σ-algebra) $\mathcal{S}$. Then μ is *σ-finite* if there exists $S_1, S_2 \dots \in \mathcal{S}$ such that $X = \bigcup_{k=1}^{\infty} S_k$ and $\mu(S_k) < \infty$.

Example 1.16. The counting measure in Example 1.9 on $([0,1], 2^{[0,1]})$ is not σ-finite.

Example 1.17. The length function for intervals in Example 1.14 is σ-finite.

Theorem 1.18. *(Carathéodory Extension Theorem). Let $\mathcal{S}$ be a semi-algebra, and μ be a σ-finite pre-measure on $(X, \mathcal{S})$. Then μ extends uniquely to a measure on $\sigma(\mathcal{S})$.*

Proof. See [12]. $\square$

Example 1.19. The length function for intervals in Example 1.14 extends to the Lebesgue measure on the Borel σ-algebra of $\mathbb{R}$.

1.2. **Probability Language.** Although we won't always use probability terms in this paper, understanding how certain definitions are interpreted in probability can help motivate and demystify concepts in dynamics.

Definition 1.20. A *probability space* is a triple $(X, \mathscr{B}, \mu)$, where X is a space, $\mathscr{B}$ is a σ-algebra of X, and μ is a measure on $(X, \mathscr{B})$ such that $\mu(X) = 1$. We call μ a *probability measure*.

Example 1.21. Because we identify S^1 with $\mathbb{R}/\mathbb{Z}$, we can consider S^1 with the normalized Lebesgue measure on $[0, 1)$. Call this the *circular Lebesgue measure*, denoted ℓ_c. Let $\mathscr{B}_c$ be the Borel σ-algebra on S^1. Then $(S^1, \mathscr{B}_c, \ell_c)$ is a probability space. Unless otherwise specified, we will consider S^1 with $(S^1, \mathscr{B}_c, \ell_c)$.

Example 1.22. Consider the measurable space $(\{0, \ldots, N-1\}, 2^{\{0,\ldots,N-1\}})$. Let $p = (p_0, \ldots, p_{N-1})$ be a *probability vector*, i.e.

$$p_0 + p_1 + \cdots + p_{N-1} = 1.$$

Then, $\mu_p : 2^{\{0,\ldots,N-1\}} \to [0, 1]$ defined by

$$\mu_p(B) = \sum_{x \in B} p_x$$

is a probability measure, and $(\{0, \ldots, N-1\}, 2^{\{0,\ldots,N-1\}}, \mu_p)$ is a probability space.

Definition 1.23. Let $(\Omega, \mathscr{A}, \mu)$ be a probability space, let $\mathscr{B}$ be the Borel σ-algebra, and ℓ be the Lebesgue measure on $\mathbb{R}$.

(1) We call an element of a σ-algebra an *event*.
(2) We call a measurable function $Y : (\Omega, \mathscr{A}, \mu) \to (\mathbb{R}, \mathscr{B}, \ell)$ a *random variable*.
(3) If Y is a random variable, we call the pushforward measure $Y_*(\mu)$ the *distribution of Y*.

Example 1.24. When reading probability papers, you might see a statement that looks like this: $P(Y < 5) = \frac{1}{2}$. Let's unpack this notation: P is a probability measure, Y is a random variable (measurable function), and $Y < 5$ is shorthand for $\{a \mid Y(a) < 5\}$, which is an event (measurable set).

Notation 1.25. Although it is common in probability to use X to represent a random variable, in this paper we will use generally use X to refer to a phase space (the domain/codomain of a dynamical system). This notation matches [7] and [15].

Definition 1.26. Let $(X, \mathscr{B}, \mu)$ be a probability space. We say two events $A, B \in \mathscr{B}$ are *independent* if

$$\mu(A \cap B) = \mu(A)\mu(B).$$

What does independence mean intuitively? Let's use the Lebesgue measure on $[0,1]\times[0,1]$ as an example. Say A and B are independent Borel sets of $[0,1]\times[0,1]$. So

$$\frac{\ell(A\cap B)}{\ell(A)} = \ell(B).$$

Now, say that we pick a random point of $[0,1]\times[0,1]$. The probability that the point is in B is $\ell(B)$, because $\ell(B)$ can be thought of as the fraction of area of $[0,1]\times[0,1]$ that is taken up by B. If someone were to give us a hint and tell us that our point is in A, then we know the probability that the point is in B is $\frac{\ell(A\cap B)}{\ell(A)}$. This is because $\frac{\ell(A\cap B)}{\ell(A)}$ is the fraction of area of A that is taken up by B. Since $\frac{\ell(A\cap B)}{\ell(A)} = \ell(B)$, that means the "hint" actually doesn't change the probability of our point being in B. We can think of these events as not effecting each other.

Definition 1.27. The expectation of a random variable $f:(\Omega,\mathscr{A},\mu)\to(\mathbb{R},\mathscr{B},\ell)$ is

$$E(f) := \int f d\mu.$$

For information about the construction of the integral with respect to a measure, see [12].

1.3. **Transformations: Notation and Pocket Examples.** Here we will define the rational rotation, irrational rotation, and expanding map on S^1. We will continually return to these examples of dynamical systems throughout the paper.

Notation 1.28. For a transformation T, we use T^n to mean $\underbrace{T\circ T\circ\cdots\circ T}_{n\text{ times}}$.

Notation 1.29. To avoid clutter, we sometimes drop the parenthesis from a function input. For example, we write $T^{-1}A$ to mean $T^{-1}(A)$, and Tx to mean $T(x)$.

Now we introduce some of the "pocket examples" that we will study throughout the paper. In the following definitions, we identify S^1 with $\mathbb{R}/\mathbb{Z}$ so that points in S^1 can be described as points in $[0,1)$.

Definition 1.30. A *rotation on* S^1 is a transformation $R_\alpha : S^1\to S^1$ of the form

$$R_\alpha(x) = x+\alpha \bmod 1,$$

where $\alpha\in[0,1)$. See Figure 1. If α is irrational, we say R_α is an *irrational rotation*. If α is rational, we say R_α is an *rational rotation*. We will consider this transformation on the probability space $(S^1,\mathscr{B}_c,\ell_c)$ from Example 1.21.

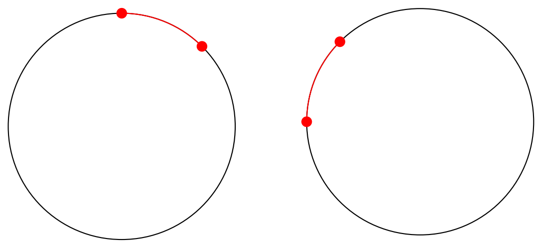
Even though the rational and irrational rotation may seem very similar, they have drastically different properties. For instance, every point of S^1 is periodic under a rational rotation, whereas no points of S^1 are periodic under an irrational rotation. In Section 4, we will show that any set that is invariant under an irrational rotation has measure 0 or 1, which is not the case for rational rotations.

Definition 1.31. An *expanding map on* S^1 is a transformation $E_k : S^1\to S^1$ of the form

$$E_k(x) = kx \bmod 1,$$

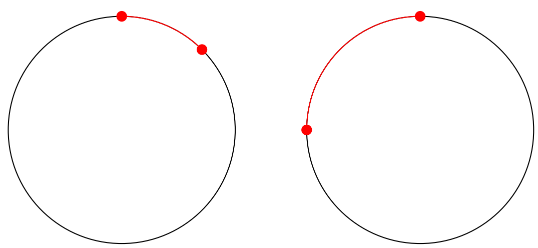
where $k\in\mathbb{Z}$ and |k|>1. See Figure 2. We will consider this transformation on the probability space $(S^1,\mathscr{B}_c,\ell_c)$ from Example 1.21.

In this paper, we will usually refer to E_2, since it is the easiest expanding map to visualize. In Section 5, we will show that this transformation is mixing, which means that the phase space gets "mixed up" over time.



(A) $[\frac{1}{8}, \frac{1}{4}]$ (B) $R_{\frac{1}{4}}[\frac{1}{8}, \frac{1}{4}]$

Figure 1. Image of $[\frac{1}{8}, \frac{1}{4}]$ under $R_{\frac{1}{4}}(x) = x + \frac{1}{4} \bmod 1$.



(A) $[\frac{1}{8}, \frac{1}{4}]$ (B) $E_2[\frac{1}{8}, \frac{1}{4}]$

Figure 2. Image of $[\frac{1}{8}, \frac{1}{4}]$ under $E_2(x) = 2x \bmod 1$.

2. The Shift Map

In this section, we define our final pocket example, the shift map. This map takes as input an infinite sequence of digits. It outputs a sequence with the same digits in the same order, but with the indexing shifted by 1. We will discuss what periodic points of the shift map look like. Also, we will take a moment to describe the space of infinite sequences where the shift map lives by putting a topology and a measure on it. The content of this section loosely follows [7].

Definition 2.1. For $N \in \mathbb{N}_{\geqslant 2}$, let

$$\Omega_N = \{(\dots \omega_{-1}, \omega_0, \omega_1, \dots) \mid \omega_i \in \{0, 1, \dots, N-1\}, i \in \mathbb{Z}\},$$

and

$$\Omega_N^R = \{(\omega_0, \omega_1, \dots) \mid \omega_i \in \{0, 1, \dots, N-1\}, i \in \mathbb{N}\}.$$

We say Ω_N is the space of two-sided sequences of N symbols, and Ω_N^R is the space of one-sided sequences of N symbols.

Example 2.2. For example, $(\dots 1, 0, 1, \dot{1}, 0, 1, 0 \dots)$ is an element of Ω_2. The dot indicates where the 0th coordinate is.

Definition 2.3. The *shift map* $\sigma_N : \Omega_N \to \Omega_N$ on N symbols is defined by

$$\sigma_N(\dots, x_{-1}, \dot{x_0}, x_1, x_2 \dots) := (\dots, x_0, \dot{x_1}, x_2, x_3 \dots)$$

where the dot indicates the 0th coordinate. So $\sigma_N(\omega) = \omega'$, where $\omega'_i = \omega_{i+1}$. The shift map $\sigma_N^R : \Omega_N^R \to \Omega_N^R$ on N symbols is defined by

$$\sigma_N^R(x_0, x_1, x_2 \dots) := (x_1, x_2 \dots).$$

Note that σ_N is invertible, and σ_N^R is not.

Example 2.4. For instance, $\sigma_2^R(1, 0, 1, 1, 0, \dots) = (0, 1, 1, 0, \dots)$.

Given a dynamical system $T : X \to X$, a *periodic point* of period n is a point $x \in X$ such that $T^n(x) = x$. Periodic points of period 1 are *fixed points.* One great thing about σ_N and σ_N^R is that it is very easy to find their periodic points. We start off with an example of finding all the periodic points of period 3 for σ_2^R. These are the sequences of 0s and 1s such that when we cut off the first three digits, the sequence is the same. For instance,

$${\sigma_N^R}^3(0, 1, 1, 0, 1, 1, 0, 1, 1 \dots) = (0, 1, 1, 0, 1, 1, 0, 1, 1 \dots).$$

The periodic points of period 3 are therefore all the sequences that repeat their first 3 coordinates forever. So there is one periodic point for each 3-digit string of 0s and 1s that we can form. Therefore, there are 2^3 periodic points. By the same reasoning, we get that there are N^n periodic points of period n for σ_N^R and σ_N.

Now we will define a topology on Ω_N and Ω_N^R.

Definition 2.5. A *cylinder set* of Ω_N is a set of the form

$$\prod_{i=-\infty}^{\infty} A_i$$

where $A_i \subseteq \{0, \dots, N-1\}$, and $A_i = \{0, \dots, N-1\}$ for all but finitely many i. A cylinder set of Ω_N^R is defined in the same way, except the product indexing starts at 0.

Example 2.6. The set

$$\dots \times \{0, 1\} \times \{0, 1\} \times \{0\} \times \{0, 1\} \times \{0, 1\} \times \dots$$

is a cylinder set of Ω_2.

The topology we use on Ω_N and Ω_N^R is the one generated by cylinder sets. This topology is the same as the product topology on $\{0, \dots, N-1\}$ with the discrete topology. Also, this topology is metrizable with the distance function

$$d_\lambda(\omega, \omega') = \sum_{-\infty}^{\infty} \frac{|\omega_n - \omega'_n|}{\lambda^{|n|}}.$$

With this metric, points that share more middle coordinates in common are closer together. For more information about this metric, see [7].

Proposition 2.7. *The transformations σ_N and σ_N^R are continuous.*

Proof. We will show σ_N is continuous. Because cylinder sets of Ω_N form a basis for the topology on Ω_N, we only need to check that the pre-image of any cylinder set is open. Let $A = \prod_{i=-\infty}^{\infty} A_i$ be a cylinder set of Ω_N, meaning $A_i = \{0, \dots, N-1\}$ for all but finitely many indices. Then,

$$\sigma_N^{-1}(A) = \prod_{i=-\infty}^{\infty} A_{i-1}.$$

Note $A_{i-1} = \{0, \dots, N-1\}$ for all but finitely many indices. So $\sigma_N^{-1}(A)$ is a cylinder set and is therefore open. Thus σ_N is continuous. The proof for continuity of σ_N^R follows by the same argument. $\square$

Now we will discuss the topological structure of Ω_N and Ω_N^R.

Definition 2.8. A *Cantor space* is a space that is homeomorphic to the middle thirds Cantor set, i.e. a space that is metrizable, compact, totally disconnected, and perfect (closed and has no isolated points).

Proposition 2.9. *The spaces Ω_N and Ω_N^R are Cantor spaces.*

Proof. We already know that Ω_N is metrizable. We will show that Ω_N is compact, perfect, and totally disconnected.

(1) By Tychonoff's theorem, Ω_N is compact because it is the product of compact sets with the product topology.
(2) We will show that Ω_N is perfect. Let $x \in \Omega_N$, and say $\prod_{i=-\infty}^{\infty} A_i$ contains x. There exists $i_0 \in \mathbb{Z}$ such that $A_{i_0} = \{0, \dots, N-1\}$. Let y be equal to x, except let y_{i_0} have a different value than x_{i_0}. Therefore, $x \neq y$ and $y \in \prod_{i=-\infty}^{\infty} A_i$. So every point in Ω_N is a limit point. Since Ω_N is the universal space, it is closed. Therefore, Ω_N is perfect.
(3) Finally, we will show that Ω_N is totally disconnected. Let $x, y \in \Omega_N$ such that $x \neq y$. There exists n such that $x_n \neq y_n$. So $x \in \prod_{i=-\infty}^{\infty} B_i$, where $B_i = \{0, \dots, N-1\}$ for all $i \neq n$, and $B_n = \{x_n\}$. Note $\prod_{i=-\infty}^{\infty} B_i$ is open. Also, $y \in (\prod_{i=-\infty}^{\infty} B_i)^c$, which is also open. Therefore, x and y cannot be in the same connected component. Thus Ω_N is totally disconnected.

The same argument holds for Ω_N^R. $\square$

We end this section by constructing the product measure on $(\Omega_N, \mathscr{B}_\infty)$, where $\mathscr{B}_\infty$ is the σ-algebra generated by the set of cylinder sets of Ω_N.

Proposition 2.10. *Let $\mathcal{C}$ denote the set of cylinder sets of Ω_N as well as $\varnothing$ and Ω_N. Then $\mathcal{C}$ is a semi-algebra.*

Proof. First, $\mathcal{C}$ contains $\varnothing$. Also, the intersection of two cylinder sets is a cylinder set. Now we show that the compliment of a set $C \in \mathcal{C}$ is a disjoint finite union of elements of $\mathcal{C}$. If $C = \varnothing$ or Ω_N, then the compliment is a single element of $\mathcal{C}$. If $C \neq \varnothing$ and $C \neq \Omega_N$, then C is a cylinder set, and we can write C as

$$C = \dots X \times X \times C_1 \times \dots \times C_n \times X \times X \times \dots,$$

Where $X = \{0, \dots, N-1\}$. We can express C^c as a disjoint finite union of cylinder sets of the form

$$\dots X \times X \times C_1' \times \dots \times C_n' \times X \times X \times \dots$$

where C_i' is either C_i or C_i^c. For example, the compliment of

$$\dots \times X \times C_1 \times C_2 \times X \times \dots$$

can be expressed as the union of

$$\cdots \times X \times C_1^c \times C_2 \times X \times \dots,$$
$$\cdots \times X \times C_1 \times C_2^c \times X \times \dots,$$

and

$$\cdots \times X \times C_1^c \times C_2^c \times X \times \dots$$

□

Now we will construct the product measure on $(\Omega_N, \mathscr{B}_\infty)$. To do so, we will construct a pre-measure on $(\Omega_N, \mathcal{C})$, and then use the Carathéodory Extension Theorem (see Theorem 1.18) to obtain a measure on $(\Omega_N, \mathscr{B}_\infty)$. Let p be the probability vector $(\frac{1}{N}, \dots, \frac{1}{N})$, and $\mu := \mu_p$ be the probability measure on $(\{0, \dots, N-1\}, 2^{\{0,\dots,N-1\}})$ described in Example 1.22. Define the pre-measure $\nu_\infty : \mathcal{C} \to [0, 1]$ by

$$\nu_\infty(\prod_{i=-\infty}^{\infty} A_i) = \prod_{i=-\infty}^{\infty} \mu(A_i).$$

Since $A_i = \{0, \dots, N-1\}$ for all but finitely many i, that means $\mu(A_i) = 1$ for all but finitely many i. Therefore, $\prod_{i=-\infty}^{\infty} \mu(A_i)$ is always a finite product, and ν_∞ is σ-finite. By the Carathéodory Extension Theorem, ν_∞ extends uniquely to a measure μ_∞ on $\sigma(\mathcal{C}) = \mathscr{B}_\infty$. Also, by construction, $(\Omega_N, \mathscr{B}_\infty, \mu_\infty)$ is a probability space. The construction for the product measure μ_∞^R on $(\Omega_N^R, \mathscr{B}_\infty^R)$, where $\mathscr{B}_\infty^R$ is the σ-algebra generated by the set of cylinder sets of Ω_N^R, follows the same steps.

3. Measure Preserving Transformations

When studying dynamical systems, our guiding question is, "what happens to the phase space over time?" By putting a probability measure on the phase space, we can talk about what is probable to happen to points or sets in our dynamical system. When studying dynamical systems on probability spaces, we want to look at transformations that preserve the measure-theoretic structure of the space (just like how we study continuous functions on topological spaces, and linear transformation on vector spaces). As such, we narrow our discussion to measure-preserving transformations. In this section, we show that our pocket examples are measure-preserving, and introduce a notion of what it means for two dynamical systems to be "equivalent" in the measure-theoretic sense. The definitions and theorems in this section follow [15].

Definition 3.1. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two probability spaces. Then we say a transformation $T : X \to Y$ is

(1) *measure-preserving* if it is measurable and $\mu(T^{-1}(B)) = \nu(B)$ for all $B \in \mathscr{B}$, and
(2) an *invertible measure-preserving transformation* if it is a measure-preserving bijection with a measure-preserving inverse.

Example 3.2. The identity function on any probability space is always an invertible measure-preserving transformation.

Example 3.3. The map $f(x) = x^2$ on $[0, 1]$ with the Lebesgue measure and Borel σ-algebra is *not* measure-preserving. This is because $f^{-1}[0, \frac{1}{4}] = [0, \frac{1}{2}]$, and $[0, \frac{1}{4}]$ does not have the same Lebesgue measure as $[0, \frac{1}{2}]$.

Before we discuss more examples of maps that are measure-preserving, we will prove a theorem that provides a method for showing that a map is measure-preserving when we don't have a good idea of what sets in the σ-algebra look like.

Theorem 3.4. *Let $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ be two probability spaces, and $T : X \to Y$ be a measurable transformation. Say $\mathcal{S}$ be a semi-algebra such that $\sigma(\mathcal{S}) = \mathscr{B}$. Then T is measure-preserving if and only if*

$$\mu(T^{-1}S) = \nu(S) \quad \text{for all } S \in \mathcal{S}. \tag{3.5}$$

Proof. We will follow the proof from [14]. First, we will show that (3.5) implies that T is measure-preserving. Assume that (3.5) holds. Then, the pushforward measure of μ, $T_*(\mu)$ (see Example 1.10) is equivalent to ν on $\mathcal{S}$. Since μ is a probability measure, so is $T_*(\mu)$, and thus both $T_*(\mu)$ and ν are σ-finite. By Carathéodory Extension Theorem (1.18), $T_*(\mu)$ must also be equivalent to ν on $\mathscr{B}$. Therefore, for all $B \in \mathscr{B}$,

$$\nu(B) = T_*(\mu)(B) = \mu(T^{-1}B),$$

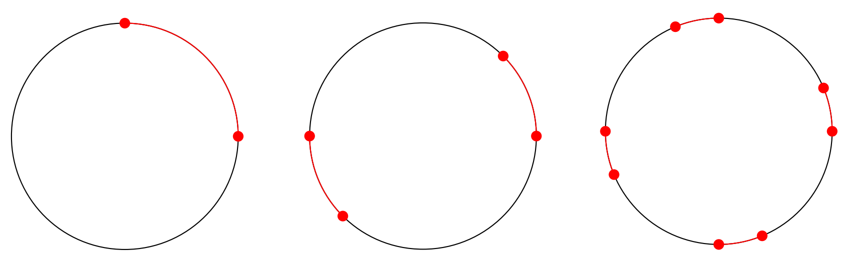
and thus T is measure-preserving. The reverse implication is immediate. ☐

Example 3.6. Consider a rotation R_α on $(S^1, \mathscr{B}_c, \ell_c)$ (refer to Definition 1.30). The Borel sets of S^1 can be generated by the semi-algebra of arcs (open, closed, or neither). By Theorem 3.4, since R_α preserves arc-length, R_α is measure preserving.

Example 3.7. Consider the expanding map $E_2(x) = 2x \bmod 1$ on $(S^1, \mathscr{B}_c, \ell_c)$. Just as in Example 3.6, the fact that

$$\ell_c(E_2^{-1}A) = \ell_c(A) \quad \text{for any arc } A \tag{3.8}$$

ensures that E_2 is measure preserving. See Figure 3 for an example demonstrating how (3.8) holds. Also, any expanding map E_k on S^1 (not just E_2) is measure-preserving.



(A) $[0, \frac{1}{4}]$ (B) $E_2^{-1}[0, \frac{1}{4}]$ (C) $E_2^{-2}[0, \frac{1}{4}]$

FIGURE 3. Example preimages under $E_2(x) = 2x \bmod 1$.

Remark 3.9. For a Borel set B of S^1, it is not generally the case that $\ell_c(E_k(B)) = \ell_c(B)$. For example, $\ell_c([0, \frac{1}{4}]) = \frac{1}{4}$, but $\ell_c(E_2[0, \frac{1}{4}]) = \frac{1}{2}$.

Example 3.10. The map σ_N on $(\Omega_N, \mathscr{B}_\infty, \mu_\infty)$ is measure-preserving. Let $A = \prod_{i=-\infty}^{\infty} A_i$ be a cylinder set. So

$$\mu_\infty(A) = \prod_{i=-\infty}^{\infty} \mu(A_i).$$

Also,

$$\mu_\infty(\sigma_N^{-1}A) = \mu_\infty(\prod_{i=-\infty}^{\infty} A_{i-1}) = \prod_{i=-\infty}^{\infty} \mu(A_i).$$

Therefore, $\mu_\infty(A) = \mu_\infty(\sigma_N^{-1}A)$. Since the set of cylinder sets of Ω_N union $\{\varnothing, \Omega_N\}$ is a semi-algebra that generates $\mathscr{B}_\infty$, by Theorem 3.4, σ_N is measure preserving.

Example 3.11. The map σ_N^R on $(\Omega_N^R, \mathscr{B}_\infty^R, \mu_\infty^R)$ is measure-preserving. Let $A = A_1 \times A_2 \times \dots$ be a cylinder set. So

$$\mu_\infty^R(A) = \mu(A_1) \cdot \mu(A_2) \cdot \dots.$$

Also,

$$\sigma_N^{R^{-1}} A = \{0, \dots, N-1\} \times A_1 \times A_2 \times \dots.$$

So

$$\mu_\infty^R(\sigma_N^{R^{-1}} A) = \mu(\{0, \dots, N-1\}) \cdot \mu(A_1) \cdot \mu(A_2) \dots.$$

Because $\mu(\{0, \dots, N-1\}) = 1$, then

$$\mu_\infty^R(A) = \mu_\infty^R(\sigma_N^{R^{-1}} A).$$

Since the set of cylinder sets of Ω_N^R is the semi-algebra that generates $\mathscr{B}_\infty^R$, by Theorem 3.4, σ_N^R is measure preserving.

We now provide a notion of what it means for two measure-preserving transformations to be equivalent in the measure-theoretic sense.

Definition 3.12. Let $T : (X, \mathscr{A}, \mu) \to (X, \mathscr{A}, \mu)$ and $S : (Y, \mathscr{B}, \nu) \to (Y, \mathscr{B}, \nu)$ be measure-preserving transformations. We say T and S are *measure-theoretically isomorphic* is there exists $X' \in \mathscr{A}$ and $Y' \in \mathscr{B}$ and $R : X' \to Y'$ such that

(1) $\mu(X') = 1$ and $\nu(Y') = 1$,
(2) $T(X') \subseteq X'$ and $S(Y') \subseteq Y'$,
(3) R is bijective,
(4) R is an invertible measure-preserving transformation, and
(5) $S \circ R = R \circ T_{|X'}$.

This notion of measure-theoretic equivalence is useful, because measure-theoretically isomorphic transformations have the same measure-theoretic properties. This means that given two measure-theoretically isomorphic transformations, we only need to study one in order to learn about the other. We can also conclude that two transformations are not measure-theoretically isomorphic if they do not have the same measure-theoretic properties.

Example 3.13. The map σ_2^R on $(\Omega_2^R, \mathscr{B}_\infty^R, \mu_\infty^R)$ is measure-theoretically isomorphic to the map $E_2(x) = 2x \bmod 1$ on $(S^1, \mathscr{B}_c, \ell_c)$. The following is a sketch of a proof. Let X' be $[0, 1) - \Gamma$, where Γ is the set of points in $[0, 1)$ that have multiple representations in binary. Since Γ is countable, it has measure 0. Given a point $x \in X'$ with binary representation $0.x_0x_1x_2\dots$, define

$$R(x) := (x_0, x_1, \dots),$$

and let $Y' = R(X')$. To show that R is measure-preserving, use Theorem 3.4 with the semi-algebra of cylinder sets. Note, this proof can be generalized to show that any expanding map E_N is measure-theoretically isomorphic to σ_N^R.

This result is really useful. We can now visualize the behavior of one-sided shifts by looking at expanding maps. Also, we can analyze expanding maps by studying one-sided shifts, which have a much simpler definition. In the next couple of sections, we will define ergodicity and mixing. These are properties of measure-preserving transformations that describe how points get distributed throughout the phase space. They are also invariants of measure-theoretic isomorphism. Therefore, we can show that expanding maps are ergodic and mixing by showing that one-sided shifts are ergodic and mixing. Also, we can show that two maps are not measure-theoretically isomorphic simply by showing that one is mixing or ergodic, and the other isn't (rather than somehow checking that no isomorphisms exist).

4. Ergodicity

In this section, we define what an ergodic transformation is, and show which of our pocket examples are ergodic. We state Birkhoff's Ergodic Theorem and discuss what it means intuitively. We also prove that ergodicity is an invariant of measure-theoretic isomorphism. The content of this section is from [15], [8], [11], [4], and [13].

Definition 4.1. Let T be a measure-preserving transformation on the probability space $(X, \mathscr{B}, \mu)$. We say T is *ergodic* if for all $A \in \mathscr{B}$ such that $T^{-1}(A) = A$, $\mu(A) = 0$ or 1.

Example 4.2. Consider a probability space $(X, \mathscr{B}, \mu)$ where $\mu : \mathscr{B} \to \{0, 1\}$. Any transformation on $(X, \mathscr{B}, \mu)$ is ergodic.

Before we talk about other examples and non-examples of ergodic transformations, we will introduce some equivalent definitions that we can use in proofs.

Theorem 4.3. *Let T be a measure-preserving transformation on the probability space $(X, \mathscr{B}, \mu)$. The following are equivalent:*

(1) T is ergodic
(2) For all $A \in \mathscr{B}$ such that $\mu((T^{-1}A)\Delta A) = 0$, $\mu(A) = 0$ or 1.
(3) If $f : X \to \mathbb{C}$ is measurable and $f \circ T = f$ a.e., then f is constant a.e.

Proof. In this paper, we use the fact that (3) $\Rightarrow$ (1) much more than the other implications. As such, we will prove (3) $\Rightarrow$ (1), and omit the rest (see [11] for the full proof). Assume that if $f : X \to \mathbb{C}$ is a measurable function and $f \circ T = f$ a.e., then f is constant a.e. We will now show T is ergodic. Say $A \in \mathscr{B}$ such that $T^{-1}A = A$. Let 1_A be the characteristic function of A, which is measurable. Also, since $T^{-1}A = A$,

$$1_A \circ T = 1_A.$$

By assumption, 1_A must be constant a.e. If 1_A is 1 a.e., then $\mu(A) = 1$. If 1_A is 0 a.e., then $\mu(A) = 0$. Because 1_A only takes on values 0 or 1, that means $\mu(A) = 0$ or 1. Thus, T is ergodic. □

Proposition 4.4. *In (3), we can replace $f \circ T = f$ a.e. with $f \circ T = f$, and instead of measurable f we can consider $f \in L^2(\mu)$.*

Proof. See [15]. □

Example 4.5. The rational rotation on $(S^1, \mathscr{B}_c, \ell_c)$ is not ergodic. To prove this, it is helpful to represent the unit circle as

$$S^1 = \{z \in \mathbb{C} \mid |z| = 1\}.$$

With this representation of S^1, we can write a rational rotation as $R(x) = ax$, where a is a root of unity. Since a is a root of unity, there exists k such that $a^k = 1$. Let $f : S^1 \to \mathbb{C}$ be defined by $f(x) = x^k$. We know f is measurable. Also,

$$(f \circ R)(x) = f(ax) = (ax)^k = a^k x^k = x^k = f(x)$$

for all $x \in S^1$. However, f is not constant a.e. Therefore, R is not ergodic.

Example 4.6. The irrational rotation on $(S^1, \mathscr{B}_c, \ell_c)$ is ergodic. Again, represent the unit circle as

$$S^1 = \{z \in \mathbb{C} \mid |z| = 1\}.$$

An irrational rotation is a map of the form $R(x) = ax$, where a is not a root of unity. For this proof, we will use a theorem about Fourier series called *Carlson's Theorem*: if $f : S^1 \to \mathbb{C}$ is in $L_2(\mu)$, then

$$f(x) = \sum_{n=-\infty}^{\infty} c_n x^n \quad \text{a.e.}$$

Also, the coefficients c_n are unique. For more information about this theorem and Fourier series in general, see [13]. Now, let $R(x) = ax$ be an irrational rotation, so a is not a root of unity. Assume $f \in L^2(\mu)$ such that $f \circ R = f$. Let $\sum_{-\infty}^{\infty} c_n x^n$ be the Fourier series for f. Then, for all $x \in S^1$,

$$\begin{aligned} f(x) &= (f \circ R)(x) \\ &= f(ax) \\ &= \sum_{n=-\infty}^{\infty} c_n (ax)^n \\ &= \sum_{n=-\infty}^{\infty} (c_n a^n) x^n \quad \text{a.e.} \end{aligned}$$

So $\sum_{n-\infty}^{\infty} (c_n a^n) x^n$ is also a Fourier series for f. Since coefficients are unique, $c_n a^n = c_n$. So $c_n a^n - c_n = 0$ and thus $c_n(a^n - 1) = 0$. Because a is not a root of unity, $a^n - 1 \neq 0$, and thus $c_n = 0$ for all $n \neq 0$. Therefore,

$$f(x) = c_0 x^0 = c_0 \quad \text{a.e.}$$

Since f is constant *a.e.*, that means R is ergodic.

Example 4.7. The map σ_N on $(\Omega_N, \mathscr{B}_\infty, \mu_\infty)$ is ergodic. We will follow the proof from [4]. Note that for all $B, C \in \mathscr{B}_\infty$,

$$|\mu_\infty(B) - \mu_\infty(C)| \leqslant \mu_\infty(B \Delta C).$$

See [18] for a proof. Say $A \in \mathscr{B}_\infty$ such that $\sigma_N^{-1} A = A$. We will prove that $\mu_\infty(A) = 0$ or 1 by showing that $\mu_\infty(A) = \mu_\infty(A)^2$. Let $\epsilon > 0$. Since $\mathscr{B}_\infty$ is generated by the semi-algebra of cylinder sets, there exists a finite union of cylinder sets A_0 such that

$$\mu_\infty(A \Delta A_0) < \frac{\epsilon}{4}.$$

Therefore,

$$|\mu_\infty(A) - \mu_\infty(A_0)| < \frac{\epsilon}{4}. \tag{4.8}$$

Because A_0 is a finite union of cylinder sets, its measure only depends on finitely many coordinates (see the end of Section 2 to review the definition of the product measure). So there exists $n \in \mathbb{N}$ such that the measure of $\sigma_N^{-n} A_0$ depends on entirely different coordinates than A_0 does. Therefore,

$$\mu_\infty(A_0 \cap \sigma_N^{-n} A_0) = \mu_\infty(A_0)\mu_\infty(\sigma_N^{-n} A_0).$$

Because σ_N is measure-preserving, $\mu_\infty(\sigma_N^{-n} A_0) = \mu_\infty(A_0)$ and thus

$$\mu_\infty(A_0 \cap \sigma_N^{-n} A_0) = \mu_\infty(A_0)^2. \tag{4.9}$$

Also, since $\sigma_N^{-1} A = A$,

$$\begin{aligned} \mu_\infty(A \Delta \sigma_N^{-1} A_0) &= \mu_\infty(\sigma_N^{-1} A \Delta \sigma_N^{-1} A_0) \\ &= \mu_\infty(\sigma_N^{-1}(A \Delta A_0)) \\ &= \mu_\infty(A \Delta A_0) \\ &< \frac{\epsilon}{4}. \end{aligned} \tag{4.10}$$

Note that

$$A \Delta (A_0 \cap \sigma_N^{-1} A_0) \subseteq (A \Delta A_0) \cup (A \Delta \sigma_N^{-1} A_0).$$

So by (4.9) and (4.10),

$$\begin{aligned} |\mu_\infty(A) - \mu_\infty(A_0 \cap \sigma_N^{-1} A_0)| &\leqslant \mu_\infty(A_0 \cap \sigma_N^{-1} A_0) \\ &\leqslant \mu_\infty((A \Delta A_0) \cup (A \Delta \sigma_N^{-1} A_0)) \\ &\leqslant \mu_\infty(A \Delta A_0) + \mu_\infty(A \Delta \sigma_N^{-1} A_0) \\ &< \frac{\epsilon}{4} + \frac{\epsilon}{4} \\ &= \frac{\epsilon}{2}. \end{aligned} \tag{4.11}$$

So

$$\begin{aligned} |\mu_\infty(A) - \mu_\infty(A)^2| &\leqslant |\mu_\infty(A) - \mu_\infty(A_0 \cap \sigma_N^{-1} A_0)| \\ &\qquad + |\mu_\infty(A_0 \cap \sigma_N^{-1} A_0) - \mu_\infty(A)^2| \\ &< \frac{\epsilon}{2} + |\mu_\infty(A_0 \cap \sigma_N^{-1} A_0) - \mu_\infty(A)^2| && \text{by (4.11)} \\ &= \frac{\epsilon}{2} + |\mu_\infty(A_0)^2 - \mu_\infty(A)^2| && \text{by (4.9)} \\ &= \frac{\epsilon}{2} + |\mu_\infty(A_0)(\mu_\infty(A_0) - \mu_\infty(A)) \\ &\qquad + \mu_\infty(A)(\mu_\infty(A_0) - \mu_\infty(A))| \\ &\leqslant \frac{\epsilon}{2} + \mu_\infty(A_0)|\mu_\infty(A_0) - \mu_\infty(A)| \\ &\qquad + \mu_\infty(A)|\mu_\infty(A_0) - \mu_\infty(A)| \\ &\leqslant \frac{\epsilon}{2} + |\mu_\infty(A_0) - \mu_\infty(A)| + |\mu_\infty(A_0) - \mu_\infty(A)| \qquad (4.12) \\ &< \epsilon && \text{by (4.8)}. \end{aligned}$$

Note (4.12) is because μ_∞ is a probability measure, and thus $\mu_\infty(A) \leqslant 1$ and $\mu_\infty(A_0) \leqslant 1$. Because ϵ is arbitrary, we have shown that $\mu_\infty(A) = \mu_\infty(A)^2$, and thus $\mu_\infty(A) = 0$ or 1. Therefore, σ_N is ergodic.

Example 4.13. Note that σ_N^R on $(\Omega_N^R, \mathscr{B}_\infty^R, \mu_\infty^R)$ is ergodic by the same argument.

Now that we have seen some examples, we can talk about what ergodic intuitively means. We will do so by discussing Birkhoff's Ergodic Theorem.

Theorem 4.14. *(Birkhoff's Ergodic Theorem) Let T be an ergodic transformation on $(X, \mathscr{B}, \mu)$. Then for any $f \in L^1(\mu)$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j = \int_X f d\mu \quad \text{a.e. and in } L^1(\mu). \tag{4.15}$$

Proof. See [8]

□

Let's unpack this theorem. Think of $\int_X f d\mu$ as the average value of f on X. For a point x, think of $(f \circ T^j)(x)$ as sampling the value of f at the point that x is at time j. So $\frac{1}{n}\sum_{j=0}^{n-1}(f \circ T^j)(x)$ is the average of n sampled values of f, where each sample is taken at a point that x visits. The theorem states that if we pick an $x \in X$ at random, then with probability 1,

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

So over time, x samples values of f by traveling all over X in an evenly distributed manner, such that the average of the sampled values approaches the average value of f on X. If this is not the case, and x spends a disproportionate amount of time visiting and sampling from a subset A of X, then the average would be closer to $\int_A f d\mu$ than $\int_X f d\mu$. Simply put, Birkhoff's Ergodic Theorem can be interpreted as "the time average is equal to the space average."

Note that Birkhoff's Ergodic Theorem does not hold if T is not ergodic. Say T is not ergodic. There exists $A \in \mathscr{B}$ such that $T^{-1}A = A$ and $0 < \mu(A) < 1$. Consider 1_A, the characteristic function of A. Then, since $T^{-1}A = A$, we know

$$1_A = 1_A \circ T = 1_A \circ T^2 = \dots$$

Therefore,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} 1_A \circ T^j = 1_A,$$

and Birkhoff's Ergodic Theorem doesn't hold [8]. Using this intuition, we can think of ergodicity as meaning that almost all points have a forward orbit that is distributed evenly throughout X.

Now we will show that ergodicity is an invariant of measure-theoretic isomorphism. We will use this fact to prove that expanding maps on $(S^1, \mathscr{B}_c, \ell_c)$ are ergodic, and to show that the rational rotation and irrational rotation are not measure-theoretically isomorphic.

Theorem 4.16. *Let $T : (X, \mathscr{A}, \mu) \to (X, \mathscr{A}, \mu)$ and $S : (Y, \mathscr{B}, \nu) \to (Y, \mathscr{B}, \nu)$ be measure-theoretically isomorphic. Then T is ergodic if and only if S is ergodic.*

Proof. We will prove an equivalent statement, that T is not ergodic if and only if S is not ergodic. Since T and S are measure-theoretically isomorphic, there exists $X' \in \mathscr{A}$, $Y' \in \mathscr{B}$, and $R : X' \to Y'$ such that all the conditions listed in Definition 3.12 are met. Assume that S is not ergodic. So there exists $B \in \mathscr{B}$ such that $0 < \nu(B) < 1$ and $S^{-1}B = B$. Let $B' = Y' \cap B$. Because $\nu(Y') = 1$, we know $\nu(B') = \nu(B)$. One can check that $S^{-1}_{|Y'}B' = B'$. We know $R^{-1}B' \in \mathscr{A}$. Also, because $S \circ R = R \circ T_{|X'}$,

$$T^{-1}(R^{-1}B) = R^{-1}(S^{-1}_{|Y'}B') = R^{-1}B'.$$

Therefore, $R^{-1}B'$ is T-invariant. Also, since R is measure-preserving

$$\mu(R^{-1}B') = \nu(B') = \nu(B).$$

Since $0 < \nu(B) < 1$, then $0 < \mu(R^{-1}B') < 1$. Therefore, T is not ergodic. Because measure-theoretic isomorphism is reflexive, the reverse implication follows by duality. $\square$

Example 4.17. By Theorem 4.16, the irrational rotation and rational rotation on $(S^1, \mathscr{B}_c, \ell_c)$ are not measure-theoretically isomorphic, because the irrational rotation is ergodic, and the rational rotation is not.

Example 4.18. Recall from Example 3.13 that the expanding map E_N on $(S^1, \mathscr{B}_c, \ell_c)$ is measure-theoretically isomorphic to σ_N^R on $(\Omega_N^R, \mathscr{B}_\infty^R, \mu_\infty^R)$. By Theorem 4.16, since σ_N^R is ergodic, so is E_N. To see a direct proof that E_2 is ergodic, see [13].

5. Mixing

Ergodicity is an interesting invariant, but it doesn't indicate at all whether a transformation "mixes up" the phase space (we saw that the irrational rotation, which is an isometry, is ergodic). In this section, we talk about mixing, which is an invariant that indicates whether a phase space gets scrambled up over time. The definitions and theorems in this section follow [15].

Definition 5.1. Let $(X, \mathscr{B}, \mu)$ be a probability space, and $T : X \to X$ be a measure-preserving transformation. We say T is *(strong) mixing* if for all $A, B \in \mathscr{B}$,

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

Notice that this definition looks similar to the definition of independence (see Definition 1.26). Think of $T^{-n}A$ as being what A looked like n units in the past. So a transformation being mixing means that for all $A, B \in \mathscr{B}$, B and the "infinite past" of A are independent.

Example 5.2. Consider the probability space $(X, \mathscr{B}, \mu)$ where $\mu : \mathscr{B} \to \{0, 1\}$. The identity function on $\{0, 1\}$ is mixing. This is because

$$\lim_{n\to\infty} \mu(\mathrm{id}^{-n}A \cap B) = \mu(A \cap B).$$

If $\mu(A)$ or $\mu(B)$ is 0, then $\mu(A \cap B) = 0 = \mu(A)\mu(B)$. If both $\mu(A) = \mu(B) = 1$, then A and B are equal mod 0 to the whole space, and $\mu(A \cap B) = 1 = \mu(A)\mu(B)$.

This example goes against our intuition of what "mixing" should mean, but is included to highlight the importance of the measure that we consider on the phase space.

Theorem 5.3. *If a measure-preserving transformation is mixing, then it is ergodic.*

Proof. Let T be a mixing transformation on $(X, \mathscr{B}, \mu)$. Say $A \in \mathscr{B}$ such that $T^{-1}A = A$. Since T is mixing,

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

Also, since $T^{-1}A = A$, then $T^{-n}A = A$ for all $n \in \mathbb{N}$. So

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

Therefore, $\mu(A)\mu(A) = \mu(A)$. Thus $\mu(A) = 0$ or 1, and T is ergodic. □

Example 5.4. Because the rational rotation on $(S^1, \mathscr{B}_c, \ell_c)$ is not ergodic, it is not mixing.

We showed that mixing implies ergodic. However, it is *not* the case that ergodic implies mixing. The following two examples are transformations that are ergodic but not mixing.

Example 5.5. Consider the probability space $(\{0,1\}, 2^{\{0,1\}}, \mu_p)$ where $p = (0,1)$ (see Example 1.22). Then $T : \{0,1\} \to \{0,1\}$ defined by $T(0) = 1$ and $T(1) = 0$ is ergodic, but not mixing. We know T is ergodic because the only invariant set is $\{0,1\}$. However, it is not mixing because for $A = \{0\}$ and $B = \{1\}$, the limit

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

does not exist.

Example 5.6. The irrational rotation R on $(S^1, \mathscr{B}_c, \ell_c)$ is not mixing. We will provide a sketch of proof. Consider two small intervals $A, B \in \mathscr{B}_c$. We can use the fact that the irrational rotation is ergodic to show that $R^{-n}A$ will be disjoint from B for infinitely many values of n (intuitively, this is because a point in A will visit all over S^1 in an evenly distributed way). Therefore, the limit does not exist or is 0, which is not equal to $\ell_c(A)\ell_c(B)$.

By the last two examples, ergodic does not imply mixing. Now we will show more examples of mixing transformations.

Example 5.7. The map σ_N on $(\Omega_N, \mathscr{B}_\infty, \mu_\infty)$ is mixing. To prove this, we only need to check that for two cylinder sets A and B,

$$\lim_{n\to\infty} \mu_\infty(T^{-n}A \cap B) = \mu_\infty(A)\mu_\infty(B).$$

This is because $\mathscr{B}_\infty$ is the σ-algebra generated by cylinder sets (in general, we need only check sets in a semi-algebra to check mixing; see [15]). So, let $A = \prod_{n=-\infty}^{\infty} A_i$ and $B = \prod_{n=-\infty}^{\infty} B_i$ be cylinder sets of Ω_N. For notation, let $X = \{0, \dots, N-1\}$. There exist $a_1, a_2 \in \mathbb{Z}$ such that when $i \notin [a_1, a_2]$, then $A_i = X$. Similarly, there exist $b_1, b_2 \in \mathbb{Z}$ such that when $i \notin [a_1, a_2]$, then $B_i = X$. So there exists n_0 such that when $n \geqslant n_0$, $T^{-n}A \cap B$ is of the form

$$\cdots \times X \times B_{b_1} \times B_{b_1+1} \times \cdots \times B_{b_2} \times X \times \cdots \times X \times A_{a_1} \times A_{a_1+1} \times \cdots \times A_{a_2} \times X \times \dots$$

Therefore, when $n \geqslant n_0$,

$$\begin{aligned}\mu_\infty(T^{-n}A \cap B) &= \mu(B_{b_1})\mu(B_{b_1+1}) \dots \mu(B_{b_2})\mu(A_{a_1})\mu(A_{a_1+1}) \dots \mu(A_{a_2}) \\ &= \mu_\infty(B)\mu_\infty(A).\end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \mu_\infty(T^{-n}A \cap B) = \mu_\infty(A)\mu_\infty(B),$$

and T is mixing.

Example 5.8. Note that σ_N^R on $(\Omega_N^R, \mathscr{B}_\infty^R, \mu_\infty^R)$ is mixing by the same argument.

Theorem 5.9. *Say $T : (X, \mathscr{A}, \mu) \to (X, \mathscr{A}, \mu)$ and $S : (Y, \mathscr{B}, \nu) \to (Y, \mathscr{B}, \nu)$ are measure-theoretically isomorphic. Then T is mixing if and only if S is mixing.*

Proof. See chapter 2 of [15]. □

Example 5.10. Recall from Example 3.13 that the expanding map E_N on $(S^1, \mathscr{B}_c, \ell_c)$ is measure-theoretically isomorphic to σ_N^R on $(\Omega_N^R, \mathscr{B}_\infty^R, \mu_\infty^R)$. By Theorem 5.9, since σ_N^R is mixing, so is E_N.

So we have learned that rotations of the circle are not mixing, but expanding maps on the circle are. This matches our expectation, because rotations are isometries, and so all the points stay in the same place relative to each other. On the other hand, the points of an expanding map on S^1 can move away from each other or towards each other depending on their location, and the phase space gets scrambled up over time.

6. Symbolic Dynamics

Earlier, we showed that $E_2(x) = 2x \bmod 1$ is measure-theoretically isomorphic to the shift map σ_2^R. Then, when we proved that σ_2^R is ergodic and mixing, we instantly got that E_2 is ergodic and mixing. This process of analyzing transformations by using shift maps is an entire field of dynamics, called symbolic dynamics. In this section, we define topological conjugacy, which is a topological notion of equivalent transformations. Then we use symbolic dynamics to identify the periodic points of two classic examples: the quadratic map and the horseshoe map. The content in this section follows [7], [5], [2], and [16].

Definition 6.1. Two continuous maps $f : X \to X$ and $g : Y \to Y$ are *topologically conjugate* if there exists a homeomorphism $h : X \to Y$ with

$$f = h^{-1} \circ g \circ h.$$

We call h a *topological conjugacy.*

Recall from linear algebra that two matrices A and B represent the same linear function (with respect to different bases) if and only if there exists an invertible matrix P such that $A = PBP^{-1}$. A change of basis is an example of a topological conjugacy. In fact, we can think of topological conjugacy as a nonlinear change of basis.

Theorem 6.2. *Say $h : X \to Y$ is a topological conjugacy between $f : X \to X$ and $g : Y \to Y$. If $x \in X$ is a periodic point of f with period n, then $h(x)$ is a periodic point of g with period n. Similarly, if $y \in Y$ is a periodic point of g, then $h^{-1}(x)$ is a periodic point of f.*

Proof. Say $x \in X$ such that $f^n(x) = x$. Since h is a topological conjugacy, $f = h^{-1} \circ g \circ h$. Therefore

$$\begin{aligned} x &= f^n(x) \\ &= \underbrace{(h^{-1} \circ g \circ h) \circ (h^{-1} \circ g \circ h) \circ \cdots \circ (h^{-1} \circ g \circ h)}_{n \text{ times}}(x) \\ &= (h^{-1} \circ g^n \circ h)(x). \end{aligned}$$

So $x = h^{-1}(g^n(h(x)))$, and thus $h(x) = g^n(h(x))$. We have thus proven that $h(x)$ is a periodic point of g with period n. The proof of the second statement follows by the same argument, as $h^{-1} : Y \to X$ is also a topological conjugacy. □

Topological conjugacy also preserves topological properties of transformations, like topological transitivity and topological mixing, which you can read more about in chapter 1 of [7].

Example 6.3. An irrational rotation and a rational rotation on S^1 are not topologically conjugate, since every point in S^1 is periodic under a rational rotation, whereas no points in S^1 are periodic under an irrational rotation.

6.1. **Quadratic Maps.** In this subsection, we investigate the quadratic map $f_\lambda : \mathbb{R} \to \mathbb{R}$ defined by

$$f_\lambda(x) = \lambda x(1 - x),$$

where $\lambda > 2+\sqrt{5}$. The content in this subsection loosely follows [7]. For a discussion about the quadratic map for other values of λ, see [2].

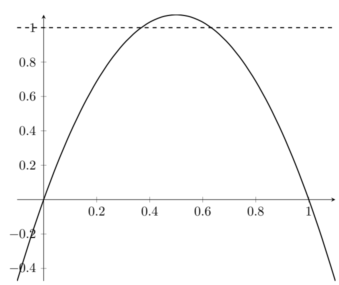


FIGURE 4. $f_\lambda(x)$ when $\lambda = 4.3$

Our goal for this section is to identify the periodic points of f_λ. A naïve approach would be to set $f_\lambda^n(x) = x$ and solve. However, with each one-step increase in n, the degree of the polynomial doubles. Once we reach degree 5 or above, there is no closed form for calculating roots, and we are basically lost. We know what the number of roots is, but we don't know how many of them are imaginary or are double roots. Instead, we will show that f_λ (when restricted to a set of interest) is topologically conjugate to σ_N^R. Then we can apply everything we know about the periodic points of σ_N^R to f_λ.

We know that if a point is periodic, then its orbit is bounded. Our first step in identifying the periodic points of f_λ is to rule out the points with unbounded orbits.

Proposition 6.4. *Let $x \in (-\infty, 0) \cup (1, \infty)$. Then $f_\lambda^n(x) \to -\infty$.*

Proof. Say $x < 0$. Then $1 - x > 1$. Multiply both sides by λx (and flip the inequality) to get $\lambda x(1 - x) < \lambda x$. Since $\lambda > 1$, then $\lambda x < x$. Therefore,

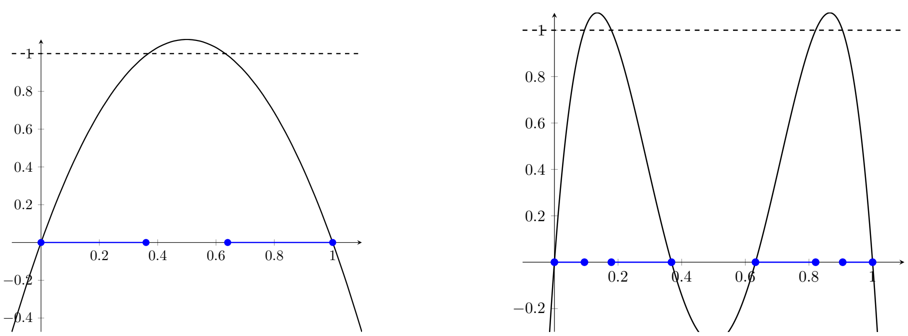
$$\lambda x(1 - x) < x.$$

So $\{f_\lambda^n(x)\}$ is a decreasing sequence. Say for a contradiction that $f_\lambda^n(x)$ converges to some p. Then $f_\lambda^{n+1}(x) \to f_\lambda(p) < p$. However, this is a contradiction, because $f_\lambda^{n+1}(x)$ must converge to the same point as $f_\lambda^n(x)$. Therefore, $f_\lambda^n(x) \to -\infty$. Now assume $x > 1$. By what we just showed, since $f_\lambda^n(x) < 0$, then $f_\lambda^{n+1}(x) \to -\infty$. Therefore, $f_\lambda^n(x) \to -\infty$. □

If at any time a point gets mapped out of $[0, 1]$, it will go to negative infinity. So we know that the points with bounded orbits are the points x such that $f^n(x) \in [0, 1]$, or equivalently $x \in f^{-n}[0, 1]$, for all $n \in \mathbb{N}$. We can express the set of all such points as

$$\Lambda = \bigcap_{n=0}^{\infty} f_\lambda^{-n}[0, 1].$$

Figure 5 shows $f^{-1}[0, 1]$ and $f^{-2}[0, 1]$. With each intersection, a "middle third" is removed, and the length of the intervals decreases exponentially. The set Λ is therefore a Cantor space.



(A) $f_\lambda^{-1}[0, 1]$ (B) $f_\lambda^{-2}[0, 1]$

FIGURE 5

Note that λ is an invariant set, i.e. $f_\lambda^{-1}\Lambda = \Lambda$. Now we can continue our search for periodic points by just looking at f_λ restricted to Λ. Because Λ and Ω_2^R are both Cantor spaces, it is reasonable that there would be a topological conjugacy h between $f_{\lambda|_\Lambda}$ and σ_2^R. We will now construct $h : \Lambda \to \Omega_2^R$. Let

$$\Delta_0 = \left[0, \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{\lambda}}\right] \quad \text{and} \quad \Delta_1 = \left[\frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{\lambda}}, 1\right].$$

See Figure 6. We define h by $h(x) = \omega$, where $\omega_n = 0$ if $f_\lambda^n(x) \in \Delta_0$, and $\omega_n = 1$ if $f_\lambda^n(x) \in \Delta_1$. We call h the *itinerary sequence* of x, because it describes the "itinerary" of where x will travel over time. This function h is well-defined because Δ_0 and Δ_1 are disjoint.

Now we will show h is a bijection. The set Δ_0 is the preimage of the set of sequences that have a 0 in the 0th coordinate, and the set Δ_1 is the preimage of

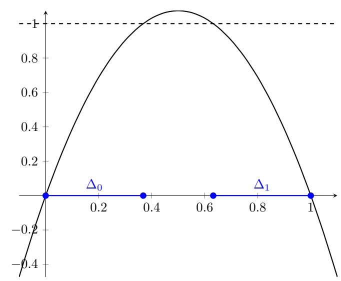


FIGURE 6. Δ_0 and Δ_1

the set of sequences that have a 1 in the 0th coordinate. Let $\Delta_{i_0 i_1}$ be the preimage of all sequences that have i_0 in the 0th coordinate and i_1 in the 1st coordinate.

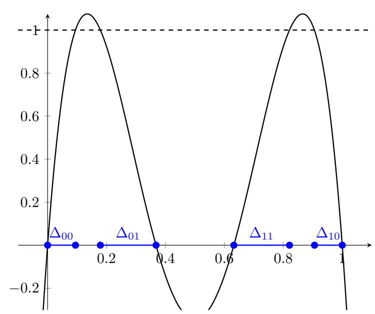


FIGURE 7. $\Delta_{00}, \Delta_{01}, \Delta_{11}$, and Δ_{10}

With each coordinate that is specified, the preimage corresponds to an exponentially smaller segment. Because these segments become a Cantor space, specifying all coordinates corresponds to exactly one point. Therefore, the pre-image of any point in Ω_2^R is exactly one point in Λ. Thus, h is a bijection. Also, $\varphi_\lambda = h^{-1} \circ \sigma_2^R \circ h$. Let $x \in \Lambda$, and $f_\lambda^n(x) \in \Delta_{i_n}$, where $\{i_n\}$ is a sequence in $\{0, 1\}$. So $f_\lambda(x)$ is a point such that $f_\lambda^n(f(x)) = f_\lambda^{n+1}(x) \in \Delta_{i_{n+1}}$. On the other hand, $h(x) = (i_0, i_1, i_2 \ldots)$, so $\sigma_2^R(h(x)) = (i_1, i_2 \ldots)$, and thus $(h^{-1} \circ \sigma_2^R \circ h)(x)$ is the point in Λ such that $f_\lambda^n((h^{-1} \circ \sigma_2^R \circ h)(x)) \in \Delta_{i_{n+1}}$. Therefore, $\varphi_\lambda = h^{-1} \circ \sigma_2^R \circ h$. Finally, h is a homeomorphism. We will not prove this, because the proof of Theorem 6.2 doesn't use the fact that the conjugacy is a homeomorphism. See [7] for a proof. Therefore, h is a topological conjugacy. So if x is a periodic point of σ_2^R, we know $h^{-1}(x)$ is a periodic point of f_λ.

Therefore, we can apply all of our knowledge about the periodic points of σ_2^N to f_λ. For instance, we know that f_λ has 2^7 periodic points of period 7.

6.2. **The Horseshoe Map.** We can think of the horseshoe map as being the 2-dimensional analogue of the quadratic map. In this subsection, we will briefly summarize how the same process of symbolic dynamics outlined in Section 6.1 can be executed for the horseshoe, but with σ_N instead of σ_N^R. For more details on this process, see [16] and [5].

The horseshoe map T is a map on the unit square $Q = [0, 1] \times [0, 1]$ that first stretches it vertically, then folds it into a horseshoe. See Figure 8.

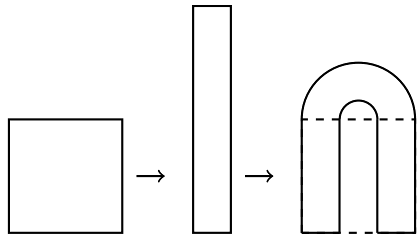
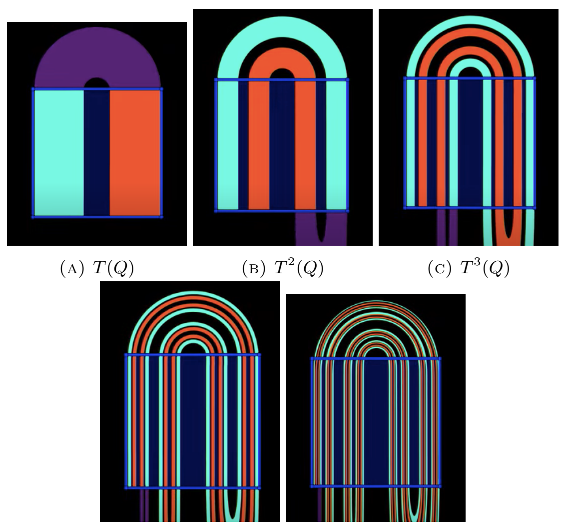


FIGURE 8. The Horseshoe Map

Just like how we ignored the points that leave $[0,1]$ in the quadratic map, in this map we ignore the points that leave the unit square. For each $n \in \mathbb{N}$, $T^n(Q) \cap Q$ is a bunch of vertical rectangles (see Figure 9, or go to [10] to see an excellent animation).



(A) $T(Q)$ (B) $T^2(Q)$ (C) $T^3(Q)$

(D) $T^4(Q)$ (E) $T^5(Q)$

FIGURE 9. Images of Q under T from [10]

The set $\bigcap_{n=0}^{\infty} T^n(Q)$ looks like $C \times [0,1]$, where C is the middle thirds Cantor set. If $x \in \bigcap_{n=0}^{\infty} T^n(Q)$, then $T^{-n}x \in Q$ for all $n \in \mathbb{N}$. So $\bigcap_{n=0}^{\infty} T^n(Q)$ is the set of points that will stay in Q under all backwards iterates of T. On the other hand, for each $n \in \mathbb{N}$, $T^{-n}(Q) \cap Q$ looks like a bunch of horizontal rectangles. Also, $\bigcap_{n=0}^{\infty} T^{-n}(Q)$ looks like $[0,1] \times C$, and is the set of all points that remain in Q under all forwards iterates of T. The set of all points that stay in Q throughout forward and backwards time is the intersection of these two sets,

$$\Lambda = \left(\bigcap_{n=0}^{\infty} T^n(Q) \right) \cap \left(\bigcap_{n=0}^{\infty} T^{-n}(Q) \right) = \bigcap_{n=-\infty}^{\infty} T^n(Q).$$

Note $\bigcap_{n=-\infty}^{\infty} T^n(Q)$ looks like $C \times C$, which is a Cantor space! Compare Figure 10 with Figure 11.

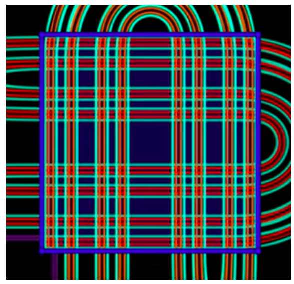


FIGURE 10. $T^5(Q)$ and $T^{-5}(Q)$, from [10].

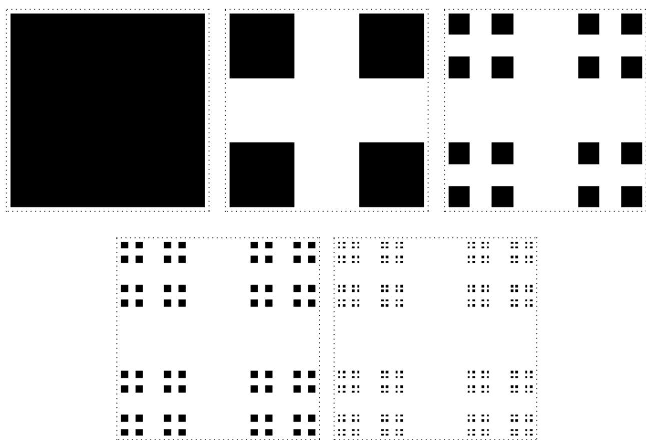


FIGURE 11. Construction of $C \times C$, from [3].

Now we will define the topological conjugacy $h : \Lambda \to \Omega_2$ between $T_{|\Lambda}$ and σ_2. Let V_0 be the left vertical rectangle of $T(Q) \cap Q$, and V_1 be the right vertical rectangle of $T(Q) \cap Q$. Let $H_0 = T^{-1}V_0$ and $H_1 = T^{-1}V_1$.

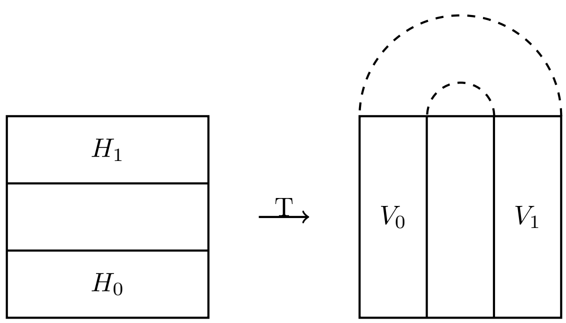


FIGURE 12. The sets H_0, H_1, V_0, and V_1.

Define h to be the itinerary map using H_0 and H_1. So $h(x) = \omega$, where $\omega_n = 0$ if $T^n(x) \in H_0$ and $\omega_n = 1$ if $T^n(x) \in H_1$. For a proof that h is a topological conjugacy, refer to [16]. Now we know that the periodic points of T look like the periodic points of σ_2.

7. Conclusion

We learned that rotations, expanding maps, and shift maps are measure-preserving. We defined measure-theoretic isomorphism, and showed that an expanding map E_N is measure-theoretically isomorphic to the shift map σ_N^R on N symbols. Then we talked about ergodicity and mixing, which are invariants of measure-theoretic isomorphism that describe how points get distributed throughout the phase space. If a transformation is mixing, then it is ergodic. The rational rotation is not ergodic, and therefore not mixing. The irrational rotation is ergodic, but not mixing. Thus, ergodic does not imply mixing, and the irrational rotation is not measure-theoretically isomorphic to the rational rotation. Shift maps are mixing, and since expanding maps are measure-theoretically isomorphic to shift maps, expanding maps are also mixing. Lastly, we talked about how to use symbolic dynamics to find periodic points of quadratic maps and the horseshoe map. Because shift maps are well-understood, we can immediately learn a lot about a map simply by showing that it is topologically conjugate to a shift map.

Acknowledgments

I want to give a big thank you to my mentor, Meg Doucette. She introduced me to the world of dynamical systems, guided me through the readings that culminated in this paper, and edited my outlines and drafts. Over the course of the summer, she has successfully convinced me of the philosophy that math learning should be grounded in examples. I am also deeply inspired by her dedication to teaching a new generation of mathematicians. I would like to thank my advisor at Reed College, David Meyer, for encouraging me to explore the world of analysis and measure theory. Lastly, I want to thank my mom, Debby Heicklen. She helped me understand the intuition behind ergodicity and mixing, and provided helpful advice regarding the structure of this paper.

References

[1] Bruin, Henk. “Lecture 9.” Ergodic Theory I . University of Vienna, 2020. https://www.mat.univie.ac.at/~bruin/ET1_lect9.pdf.

[2] Devaney, Robert. *An Introduction to Chaotic Dynamical Systems*. 1989. Reprint, CRC Press, 2018.

[3] Dickau, Robert. “Cantor Dust - Robert Dickau.” www.robertdickau.com, May 2014. https://www.robertdickau.com/cantor.html.

[4] Donald, Antonio. “Shifts and Ergodic Theory.” University of Chicago, September 2021. https://math.uchicago.edu/~may/REU2021/REUPapers/Donald.pdf.

[5] Dr. Shane Ross. “Horseshoe Map - Essence of Chaos, Symbolic Dynamics, and the Shift Map.” YouTube, May 10, 2021. https://www.youtube.com/watch?v=4OL3IVp25WQ&list=PLKD1Nmyjcd6QDYmPTQayw9USdka5RWD3g&index=2.

[6] Iakov Grigorevich Sinai. *Topics in Ergodic Theory (PMS-44), Volume 44*. Princeton University Press, 2017.

[7] Katok, A B, and Boris Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Cambridge ; New York, Ny, Usa: Cambridge University Press, 1995.

[8] Lalley, Steve. Statistics 381: Measure-Theoretic Probability 1 Course Notes. University of Chicago, 2019. https://galton.uchicago.edu/~lalley/Courses/381/index.html.

[9] Penrose, Mathew. "The Caratheodory Extension Theorem." Measure Theory and Integration (MA40042) Semester I Course Notes. University of Bath, 2023. https://people.bath.ac.uk/masmdp/measdir.bho/notesp17to21.pdf.

[10] Prof Ghrist Math. "AppDynSys : Smale Horseshoe : Iteration." YouTube, November 21, 2019. https://www.youtube.com/watch?v=SrJm6bkLuPs&list=PLKD1Nmyjcd6QDYmPTQayw9USdka5RWD3g&index=7.

[11] Proof Wiki. "Equivalence of Defintions of Ergodic Measure-Preserving Transformation," n.d. https://proofwiki.org/wiki/Equivalence_of_Defintions_of_Ergodic_Measure-Preserving_Transformation.

[12] Stein, Elias M, and Rami Shakarchi. *Real Analysis : Measure Theory, Integration, and Hilbert Spaces*. Princeton, N.J. ; Oxford: Princeton University Press, 2005.

[13] Ulcigrai, Corinna. "Dynamical Systems and Ergodic Theory." Course Notes for Lectures 16-30. University of Bristol, 2016. https://people.maths.bris.ac.uk/~ip13935/dyn/CorinnaIII.pdf.

[14] Ulcigrai, Corinna. "Measure Preserving Transformations," 2018. https://indico.ictp.it/event/8325/session/1/contribution/9/material/0/0.pdf.

[15] Walters, Peter. *An Introduction to Ergodic Theory*. 1975. Reprint, Springer Science & Business Media, 2000.

[16] Wen, Lan. *Differentiable Dynamical Systems : An Introduction to Structural Stability and Hyperbolicity*. Providence, Rhode Island: American Mathematical Society, 2016.

[17] Wikipedia Contributors. "Pre-Measure." Wikipedia. Wikimedia Foundation, June 28, 2022. https://en.wikipedia.org/wiki/Pre-measure.

[18] Wikipedia Contributors. "Symmetric Difference." Wikipedia. Wikimedia Foundation, May 10, 2024. https://en.wikipedia.org/wiki/Symmetric_difference#Symmetric_difference_on_measure_spaces.