MEASURE-PRESERVING DYNAMICAL SYSTEMS AND APPROXIMATION TECHNIQUES

JASON LIANG

Abstract. In this paper, we demonstrate how approximation structures called sufficient semirings can provide information about measure-preserving dynamical systems. We describe the basic properties of measure-preserving dynamical systems and illustrate their connections to sufficient semirings by investigating the dyadic adding machine.

Contents

1. Introduction 1
2. Invariant Transformations 2
3. Recurrence 3
4. Ergodicity 4
5. The Dyadic Adding Machine and Approximation Techniques 8
Acknowledgments 13
References 13

1. Introduction

Consider a measure space $(X, \mathcal{L}, \lambda)$, where X is a set, $\mathcal{L}$ is a σ-ring of measurable subsets of X, and λ is the measure. A **dynamical system** consists of the measure space and a **time rule** that describes how points in the space change over time. When the time rule is discrete, the dynamical system is given by $(X, \mathcal{L}, \lambda, T)$, where the time rule T is a transformation, or a mapping from a set to itself, on X. For a given natural number n, we can interpret the iterate $T^n(x)$ as the position at time n of an object that starts at x and the **orbit of x** $\{T^n(x)\}_{n=0}^{\infty}$ as the set of all points visited by the path beginning at x.

The first three sections of the paper-proper contain a discussion of the general properties of measure-preserving dynamical systems. In Section 2, we define invariant, or measure-preserving, transformations and prove a theorem relating invariant transformations to integrable functions. In Section 3, we define recurrence and prove Poincaré's Recurrence Theorem. In Section 4, we state some of the motivating questions of ergodic theory, demonstrate the connections between these questions, and answer them by proving and applying Birkhoff's Ergodic Theorem.

We then demonstrate, in Section 5, one method of analyzing dynamical systems for these properties by constructing and investigating a transformation called the dyadic adding machine. The dyadic adding machine is closely related to the set of dyadic intervals, or intervals whose endpoints are of the form $\frac{k}{2^n}$, on $[0, 1)$, and we define an approximation structure called a sufficient semiring using the set of dyadic intervals as an example. The main results of this section are Theorem 5.2, which allows us to extend the measure-preserving property on a sufficient semiring to the entire space, and Lemma 5.6, an approximation result that we can use to prove ergodicity.

The main prerequisite for this paper is an understanding of the definitions, constructions, and basic properties of the Lebesgue measure and the Lebesgue integral. Many results and definitions are stated and proved for general measure spaces, but readers should feel free to mentally substitute a Lebesgue space. Readers may also choose to begin with the construction of the dyadic adding machine on pp. 8-9 and to keep this transformation in mind while reading the first three sections.

Date: 29 AUGUST 2014; Revised 21 SEPTEMBER 2014.

2. Invariant Transformations

Given a map T from a measure space $(X, \mathcal{L}, \lambda)$ to a measure space $(Y, \mathcal{S}, \mu)$, two important questions are whether T preserves the measurability of sets, and, if so, whether it preserves the measure of sets. T is called **measurable** if $A \in \mathcal{S}$ implies $T^{-1}(A) \in \mathcal{L}$. If, in addition, T is invertible and T^{-1} is also measurable, then T is called **invertible-measurable**. Given that T is measurable, T is called **measure-preserving** if $A \in \mathcal{S}$ implies $\lambda(T^{-1}(A)) = \mu(A)$.

When T maps from a measure space to itself, it is called a **transformation**, and if T is also measure-preserving, then it is called **invariant**. A measure-preserving map is not called invariant unless it is also a transformation. This paper mainly concerns measure-preserving dynamical systems $(X, \mathcal{L}, \lambda, T)$ on a finite measure space. In other words, X is assumed to be of finite measure and T is assumed to be an invariant transformation.

The use of preimage, rather than image, in the definitions above preserves the σ-ring structure. This is demonstrated by the following proposition, which follows from the definitions of image and preimage.

Proposition 2.1. *Suppose f is a mapping from a set $\mathcal{X}$ to a set $\mathcal{Y}$. The following are true.*

(1) If $X_n \subset \mathcal{X}$, then $f^{-1}(\bigcup_{n=1}^{\infty} X_n) = \bigcup_{n=1}^{\infty} f^{-1}(X_n)$ and $f(\bigcup_{n=1}^{\infty} X_n) = \bigcup_{n=1}^{\infty} f(X_n)$.

(2) In addition, $f^{-1}(\bigcap_{n=1}^{\infty} X_n) = \bigcap_{n=1}^{\infty} f^{-1}(X_n)$ and $f(\bigcap_{n=1}^{\infty} X_n) \subset \bigcap_{n=1}^{\infty} f(X_n)$. For the latter case, there exist examples where inclusion is strict; however, if f is invertible, then the inclusion is always an equality.

(3) If $X_1, X_2 \subset \mathcal{X}$, then $f^{-1}(X_1 \setminus X_2) = f^{-1}(X_1) \setminus f^{-1}(X_2)$ and $f(X_1 \setminus X_2) \subset f(X_1) \setminus f(X_2)$. Once again, there exist examples where the latter inclusion is strict, but if f is invertible, then the inclusion is always an equality.

On the other hand, the next proposition shows that taking the image of a set under an invertible-measurable, invariant transformation also preserves its measure.

Proposition 2.2. *If T is an invertible-measurable, invariant transformation on X, then $A \in \mathcal{L}$ implies $\lambda(T(A)) = \lambda(A)$.*

Proof. Since T is invertible-measurable, $T(A)$ is a measurable set, and since T is measure-preserving

$$\lambda(T^{-1}(T(A))) = \lambda(T(A)).$$

By the invertibility of T, $\lambda(A) = \lambda(T(A))$. □

It is clear that measurable maps can be composed, when appropriate, to form new measurable maps. Many questions in dynamical systems pertain to the composition of an integrable function, or a measurable real function with a finite Lebesgue integral, f with an iterate of a transformation T. For physical applications, we can interpret $f(T^n(x))$ as the state of an object at x at time n. The next theorem allows us to characterize invariant transformations in terms of $f \circ T$. In the proof of this theorem and the rest of the paper, K_A denotes the indicator function for any set A.

Theorem 2.3. *Suppose T is a transformation of a finite measure space $(X, \mathcal{L}, \lambda)$. Then T is an invariant transformation iff for any integrable function f on X,*

$$\int_X f \circ T = \int_X f.$$

Proof. First, assume that the equality of integrals holds for any integrable function f. Let A be a measurable set. Then $K_A(T(x)) = 1$ iff $T(x) \in A$, so $K_A \circ T = K_{T^{-1}(A)}$. Therefore

$$\lambda(T^{-1}(A)) = \int_X K_A \circ T = \int_X K_A = \lambda(A),$$

so that T is measure-preserving.

Now assume that T is a measure-preserving function. We use the standard strategy of proving the theorem first for indicator functions then generalizing to integrable functions. If A is a measurable set, then

$$\int_X K_A \circ T = \lambda(T^{-1}(A)) = \lambda(A) = \int_X K_A.$$

This shows that the theorem holds for indicator functions, and by the linearity of the integral, the theorem is also true for simple functions. Now suppose that f is a non-negative, integrable function. It is well known that there exists a sequence $\{s_n\}$ of bounded simple functions that converges pointwise to f and that $\lim_{n\to\infty} \int_X s_n = \int_X \lim_{n\to\infty} s_n = \int_X f$. Then

$$\int_X f \circ T = \lim_{n\to\infty} \int_X s_n \circ T = \lim_{n\to\infty} \int_X s_n = \int_X f.$$

Finally, rewrite any integrable function f as $f = f^+ - f^-$, where $f^+ := \max(f, 0)$ and $f^- := \max(-f, 0)$, and apply the argument for the non-negative case to complete the proof of the theorem. □

3. Recurrence

Suppose T is a measurable transformation on a measure space $(X, \mathcal{L}, \lambda)$. One question is whether, given a set A and an element $x \in A$, there exists a natural number n such that $T^n(x) \in A$. In other words, does the path starting at x ever return to A? Is this true of almost every x in A? T is called **recurrent** if for every set A of positive measure and almost every $x \in A$, there exists a natural number n such that $T^n(x) \in A$.

The main theorem, called Poincaré's Recurrence Theorem, of this section shows that invariant transformations of a finite measure space are always recurrent. In the proof of this theorem, we use the following equivalent definition of recurrence. T is called **conservative** if for every set A of positive measure, there exists a natural number n such that $\lambda(T^{-n}(A) \cap A) > 0$. That is, there is a subset B of A such that B is of positive measure and every object that starts in B lies in A at some specific time n.

Proposition 3.1. *Suppose T is a measurable transformation on a measure space X. T is recurrent iff it is conservative.*

Proof. Suppose that T is not conservative. Then there exists a set A of positive measure such that for any $n \in \mathbb{N}$, the set of x such that $T^n(x) \in A$ has measure zero, and this means T is not recurrent.

Now suppose that T is not recurrent. Then there exist sets A and B of positive measure such that $B \subset A$ and for all $x \in B$, $T^n(x) \notin B \subset A$ for any n. Since $x \in (T^{-n}(B) \cap B)$ means that $T^n(x) \in B$, $\lambda(T^{-n}(B) \cap B) = 0$ for all n. This means that T is not conservative. □

If the sets $T^{-n}(A)$ are of equal positive measure, it seems that if X is of finite measure, then eventually one of the preimages must overlap with A. The proof of Poincaré's Recurrence Theorem is a formalization of this idea.

Theorem 3.2 (Poincaré's Recurrence Theorem)**.** *Suppose T is an invariant transformation on a finite measure space $(X, \mathcal{L}, \lambda)$. Then T is recurrent.*

Proof. Assume, in contradiction to the theorem, that T is not recurrent. Then there exists a set A of positive measure such that $\lambda(T^{-n}(A) \cap A) = 0$ for all natural numbers n. Now suppose that j and k are natural numbers. Then $\lambda(T^{-(j+k)}(A) \cap T^{-j}(A)) = \lambda(T^{-j}(T^{-k}(A) \cap A)) = 0$. The last equality holds because T is measure-preserving. If $j < k$, then $\lambda(T^{-k}(A) \cap T^{-j}(A)) = 0$. This means that

$$\lambda(\bigcup_{n=1}^{\infty} T^{-n}(A)) = \sum_{n=1}^{\infty} \lambda(T^{-n}(A)) = \sum_{n=1}^{\infty} \lambda(A) = \infty,$$

and this contradicts the fact that X is of finite measure. □

To reach the contradiction, we needed X to be of finite measure. The transformation T on the Lebesgue space $\mathbb{R}$ defined by $T(x) := x + 3$ provides an example of an invariant transformation that is not recurrent.

One last question is whether the orbit of an element under a recurrent transformation has an infinite number of returns. The next proposition answers this question in the affirmative.

Proposition 3.3. *If T is recurrent, then for every set A of positive measure and almost every x in A, there exists a sequence $\{n_{x_i}\}$ such that $T^{n_{x_i}}(x) \in A$.*

Proof. Let N_1 be the set of all $x \in A$ such that there does not exist a natural number n_{x_1} such that $T^{n_{x_1}}(x) \in A$. Since T is recurrent, N_1 has measure zero. Now let N_2 be the set of all $x \in A \setminus N_1$ such that there does not exist a natural number n_{x_2} such that $T^{n_{x_2}}(x) \in A \setminus N_1$, and define N_3, N_4, etc. analogously.

Since T is recurrent, each N_i has measure zero. For all x in $A \setminus \bigcup_{i=1}^{\infty} N_i$, there exists a sequence $\{n_{x_i}\}$ such that $T^{n_{x_i}}(x) \in A$, and we also have $\lambda(\bigcup_{i=1}^{\infty} N_i) = 0$. $\square$

4. Ergodicity

Among the most prominent questions in dynamical systems is the one that asks whether the time average of a function is equal to its space average. Today, this question is typically posed in the following fashion. Suppose $(X, \mathcal{L}, \lambda, T)$ is a measure-preserving dynamical system on a finite measure space, and let f be an arbitrary integrable function. The long-term time average of f over a particular orbit is given by $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$ while the space average of f is $\frac{1}{\lambda(X)} \int_X f$. Since λ can be scaled so that $\lambda(X) = 1$, we assume that X is a probability space. The question is whether the equation

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \int_X f$$

is true at least for almost every $x \in X$.

The following specific case of this question for indicator functions asks whether, on average, T distributes elements proportionally throughout the space. Suppose X is a set of locations such that each location contains exactly one object. If A and B are measurable sets, is the proportion of objects that start in B that are in A at time n equal to, on average, the proportion of the space that A takes up? In other words, is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \lambda(T^{-i}(A) \cap B) = \frac{\lambda(A)}{\lambda(X)} \lambda(B) = \lambda(A)\lambda(B)?$$

While these questions are of a physical nature, we will show that they are directly related to the algebraic question of reducibility. It is with this property that the section begins.

It is natural to ask whether T can be decomposed into smaller transformations. Given an invariant transformation T on a probability space $(X, \mathcal{L}, \lambda)$, a subset A of X is called **T-invariant** if $x \in A$ iff $T(x) \in A$. If there exists a T-invariant set A of positive but not full measure, then T can be decomposed into two transformations on two disjoint sets of positive measure. Namely, T restricted to A and T restricted to A^c is one such decomposition. T is called **ergodic** if the only sets that are T-invariant are those that have zero measure or whose complements have zero measure. Since sets which differ by a set of zero measure are typically identified, the ergodic transformations are those that are irreducible in a non-trivial sense.

In the next theorem, we present several equivalent characterizations of ergodicity that follow immediately from this definition. The first is a relaxation of the T-invariant condition, while the last two provide more geometric interpretations of ergodicity. In both the statement of this theorem and the rest of this paper, $\triangle$ denotes the symmetric difference operation.

Theorem 4.1. *For an invariant transformation T on a probability space $(X, \mathcal{L}, \lambda)$, the following statements are equivalent.*

(1) T is ergodic.
(2) If A is measurable and $\lambda(A \triangle T^{-1}(A)) = 0$, then either A or A^c has zero measure.
(3) If A is a set of positive measure, then for almost every $x \in X$, there exists a natural number n such that $T^n(x) \in A$. In other words, every area of positive measure is reached by the orbit of almost every point in the space.
(4) If A and B are both sets of positive measure, then there exists a natural number n such that $\lambda(T^{-n}(A) \cap B) > 0$. That is, there is a region of positive area within B such that every object that starts in that area lies in A at some particular time.

Proof. Assume that T is ergodic. We will show that any set A satisfying $\lambda(A \triangle T^{-1}(A)) = 0$ can be approximated up to a difference of zero measure by a T-invariant set A'. Suppose A is such a set, and let

$A^{'} := \bigcap_{N=0}^{\infty} \bigcup_{n=N+1}^{\infty} T^{-n}(A)$. Then $T^{-1}(A^{'}) = \bigcap_{N=-1}^{\infty} \bigcup_{n=N+1}^{\infty} T^{-n}(A) = A^{'}$, so $A^{'}$ is T-invariant. Since

$$(\bigcap_{n=N+1}^{\infty} A \setminus T^{-n}(A)) \subset (\bigcap_{n=N+2}^{\infty} A \setminus T^{-n}(A))$$

for all N,

$$\lambda(A \setminus A') = \lambda(\bigcup_{N=0}^{\infty} \bigcap_{n=N+1}^{\infty} A \setminus T^{-n}(A)) = \lim_{N\to\infty} \lambda(\bigcap_{n=N+1}^{\infty} A \setminus T^{-n}(A)) = 0.$$

We can apply a similar argument to show that $\lambda(A' \setminus A) = 0$, so $\lambda(A \triangle A') = 0$. Since T is ergodic, either $A^{'}$ or $(A^{'})^c$ has zero measure, and thus the same is true of A. This means that (1) implies (2).

Now assume that (2) is true, and suppose that A is a set of positive measure. Define $A^* := \bigcup_{n=1}^{\infty} T^{-n}(A)$. Then $T^{-1}(A^*) = \bigcup_{n=2}^{\infty} T^{-n}(A) \subset A^*$, and $\lambda(T^{-1}(A^*)) = \lambda(A^*)$ since T is an invariant transformation. Thus $\lambda(A^* \triangle T^{-1}(A^*)) = 0$, and since T is ergodic, $\lambda(A^*) = 1$. Therefore (2) implies (3).

Next assume that (3) is true, and suppose that A and B are sets of positive measure. Since $\bigcup_{n=1}^{\infty} T^{-n}(A) \cap B$ is the set of all $x \in B$ such that the orbit of x intersects A, (3) implies $\lambda(\bigcup_{n=1}^{\infty} T^{-n}(A) \cap B) = \lambda(B) > 0$. Therefore there must exist an m such that $\lambda(T^{-m}(A) \cap B) > 0$, so (3) implies (4).

Lastly, assume that (4) is true, and suppose that A is a T-invariant set of positive measure. If, in contradiction to the theorem, A does not have full measure, then A^c also has positive measure and $\lambda(T^{-m}(A) \cap A^c) > 0$ for some m. At the same time, $\lambda(T^{-n}(A) \cap A^c) = 0$ for all n since A is T-invariant. This leads to a contradiction, so T must be ergodic.

□

Corollary 4.2. *If T is an invertible-measurable, invariant transformation and T is ergodic, then T^{-1} is also ergodic.*

Proof. Suppose A and B are both sets of positive measure. By Theorem 4.1, there exists a natural number n such that $\lambda(T^{-n}(A) \cap B) > 0$. Since, by Proposition 2.2, T^{-1} is measure-preserving,

$$\lambda(A \cap T^n(B)) = \lambda(T^n(T^{-n}(A) \cap B)) = \lambda(T^{-n}(A) \cap B) > 0.$$

□

We are now in a position to answer the opening question of this section. The next theorem says that when T is an ergodic transformation, the time average of any integrable function over almost every orbit is equal to its space average.

Theorem 4.3 (Birkhoff's Ergodic Theorem)**.** *Suppose T is an ergodic transformation on a probability space $(X, \mathcal{L}, \lambda)$. If T is ergodic, then for any integrable function f on X,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \int_X f$$

for almost every x.

Before beginning the proof, we introduce some notation to keep track of important expressions. The notation and fundamental argument of our proof closely follow those in Section 5.1 of Silva's book [3, pp. 176-187]. A more general form of the theorem and a proof of that general form are also found there.

Notation 4.4. Define the following:

(1) for a given natural number n, let $f_n(x) := \sum_{i=0}^{n-1} f(T^i(x))$,
(2) let $f^*(x) := \limsup_{n\to\infty} \frac{1}{n} f_n(x)$,
(3) let $f_*(x) := \liminf_{n\to\infty} \frac{1}{n} f_n(x)$,

(4) given a real number r, let $C_r := \{x : f_*(x) < r\}$, and
(5) given a real number r and a natural number p, let $E_p^r := \{x : f_n(x) \geq r \text{ for all } 1 \leq n \leq p\}$.

The outline of the proof is as follows. If we can show that $f_*(x) \geq \int_X f$ is true for almost every x, then $f^*(x) \leq \int_X f$ by application to $-f$, and this proves the theorem. Assuming the first inequality does not hold for almost every x means that there is some rational $r < \int_X f$ such that $C_r = \{x : f_*(x) < r\}$ has positive measure. We will show that $\lambda(C_r \triangle T^{-1}(C_r)) = 0$, so $\lambda(\bigcap_{p=1}^{\infty} E_p^r) \leq \lambda(X) - \lambda(C_r) = 1 - 1 = 0$. Our main goal in the next series of lemmas is to prove an inequality relating E_p^r to $\int_X f$ which shows that this cannot be the case.

Lemma 4.5. *The following statements are true.*
(1) For any $x \in X$, $f^*(T(x)) = f^*(x)$ and $f_*(T(x)) = f_*(x)$.
(2) For any real r, $\lambda(C_r \triangle T^{-1}(C_r)) = 0$.

Proof. For any natural number n,

$$\frac{1}{n} f_n(T(x)) = \frac{1}{n} \sum_{i=0}^{n-1} f(T(T^i(x)))$$

$$= \frac{1}{n} \sum_{i=1}^{n} f(T^i(x)) = \frac{1}{n}[f_{n+1} - f(x)]$$

$$= \frac{n+1}{n} \frac{1}{n+1} f_{n+1} - \frac{f(x)}{n}.$$

Since $\lim_{n\to\infty} \frac{n+1}{n} = 1$, we can take the appropriate limits using the last form of the equation to get

$$f^*(T(x)) = f^*(x)$$

and

$$f_*(T(x)) = f_*(x).$$

From the last equation, we get that if $x \in C_r$, then $T(x) \in C_r$. This means that $C_r \subset T^{-1}(C_r)$. Since T is measure-preserving, $\lambda(C_r \triangle T^{-1}(C_r)) = 0$. $\square$

Lemma 4.6. *Suppose g is a real function on a set X, p is a natural number, and $\tau : X \to \{1, 2, ..., p\}$ is a function such that for every x in X, $g_{\tau(x)}(x) = \sum_{i=0}^{\tau(x)-1} g(T^i(x)) \leq 0$. Then for any $n \geq p$,*

$$g_n(x) \leq \sum_{i=n-p}^{n-1} |g(T^i(x))|$$

for all x.

Proof. Let x and $n \geq p$ be given. Since $g_{\tau(x)}(x) \leq 0$,

$$g_n(x) = \sum_{i=0}^{n-1} g(T^i(x)) \leq \sum_{i=0}^{n-1} g(T^i(x)) - \sum_{i=0}^{\tau(x)-1} g(T^i(x)) = \sum_{\tau(x)}^{n-1} g(T^i(x)) \leq \sum_{\tau(x)}^{n-1} |g(T^i(x))|.$$

We can raise the lower bound on the last sum by removing the next $\tau(\tau(x))$ numbers, and this is permissible provided there are more than $\tau(\tau(x))$ terms remaining. We can keep raising the lower bound until there are not enough terms remaining, and since $\tau(x) < p$, there are at most p terms remaining when we have to stop. Therefore $g_n(x) \leq \sum_{i=n-p}^{n-1} |g(T^i(x))|$. $\square$

Lemma 4.7. *Suppose f is an integrable function and p is a natural number. The following statements are are true.*

(1) For any $n \geq p$,

$$f_n(x) \leq \sum_{i=0}^{n-1} f(T^i(x))K_{E_p^0}(x) + \sum_{i=n-p}^{n-1} |f(T^i(x))|.$$

(2) For any real r,

$$\int_X f \leq \int_{E_p^r} f + r(1 - \lambda(E_p^r)).$$

Proof. Define the functions $h(x) := f(x)K_{E_p^0}(x)$ and $g(x) := f(x) - h(x)$. Since $x \in E_p^0$ implies $f(x) \geq 0$, $h_n(x) \geq 0$ for all n. Suppose we could find a function τ corresponding to g and p and satisfying the hypothesis of Lemma 4.6. Then for any $n \geq p$,

$$\begin{aligned} g_n(x) &= \sum_{i=0}^{n-1} f(T^i(x)) - \sum_{i=0}^{n-1} f(T^i(x))K_{E_p^0}(T^i x) \\ &\leq \sum_{i=n-p}^{n-1} |f(T^i(x)) - f(T^i(x))K_{E_p^0}(T^i x)| \\ &\leq \sum_{i=n-p}^{n-1} |f(T^i(x))|. \end{aligned}$$

Define a function $\tau : X \to \{1, 2, ...p\}$ by the rule $\tau(x) = 1$ for x in E_p^0 and $\tau(x) = \min\{k \mid f_k(x) < 0\}$ for all other x. Then for x in E_p^0, $g_{\tau(x)}(x) = g(x) = f(x) - f(x)K_{E_p^0}(x) = 0$ and for x not in E_p^0, $g_{\tau(x)}(x) = f_{\tau(x)}(x) - h_{\tau(x)}(x) \leq 0 - h_{\tau(x)}(x) \leq 0$. Thus τ satisfies the hypothesis of Lemma 4.6, and this proves the first part of Lemma 4.7.

To prove the second part of the lemma, we first apply (1) to the function $(f - r)$. Then for any $n \geq p$,

$$(f - r)_n(x) \leq \sum_{i=1}^{n} (f - r)(T^i(x))K_{E_p^r}(x) + \sum_{i=n-p}^{n-1} |(f - r)(T^i(x))|.$$

Since T is measure-preserving, we can integrate both sides then divide by n to get

$$\int_X f \leq \int_{E_p^r} f + r(1 - \lambda(E_p^r)) + \frac{p \int_X |f - r|}{n},$$

and since this holds for arbitrary n, the lemma is true. □

Readers who are struggling with the last inequality in the proof above may want to recall that we assumed that $\lambda(X) = 1$.

Proof of Theorem 4.3 Assume, in contradiction to the theorem, that the inequality $\int_X f \leq f_*(x)$ fails on a set A of positive measure. Let $P := \{q \in \mathbb{Q} < \int_X f\}$. Since $A = \bigcup_{q \in P} C_q$, there exists a rational $r < \int_X f$ such that $\lambda(C_r) > 0$. We apply Lemma 4.5 in combination with the ergodicity of T to get that $\lambda(C_r) = 1$, so $\lambda(\bigcap_{p=1}^{\infty} E_p^r) = \lim_{p\to\infty} \lambda(E_p^r) = 0$ as discussed earlier. Now we apply (2) from Lemma 4.7 to get $\int_X f \leq r$, and this contradicts the choice of r. □

We can use Birkhoff's Ergodic Theorem to show that the transformations that distribute elements proportionally in the sense described at the beginning of this section are exactly those that are ergodic.

Corollary 4.8. *Suppose T is an invariant transformation on a probability space $(X, \mathcal{L}, \lambda)$. Then T is ergodic iff*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \lambda(T^{-i}(A) \cap B) = \lambda(A)\lambda(B)$$

whenever A and B are measurable.

Proof. First, assume that T is ergodic. By Birkhoff's Ergodic Theorem, for each natural number N there is a set Y_N of full measure such that for all $y \in Y_N$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}K_A(T^{n+i}(y))K_B(T^i(y)) = \lambda(T^{-n}(A)\cap B).$$

Since each Y_N has full measure, we can fix $x \in X$ such that x is in Y_N for every N. Then

$$\lim_{m\to\infty}\frac{1}{m}\sum_{k=0}^{m-1}\lambda(T^{-k}(A)\cap B) = \lim_{m\to\infty}\frac{1}{m}\sum_{k=0}^{m-1}\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}K_A(T^{k+i}(x))K_B(T^i(x)).$$

We apply Birkhoff's Ergodic Theorem to convert the inner limit into an integral, and the equation becomes

$$\lim_{m\to\infty}\frac{1}{m}\sum_{k=0}^{m-1}\lambda(T^{-k}(A)\cap B) = \lim_{m\to\infty}\frac{1}{m}\sum_{k=0}^{m-1}\int_X K_A(T^k(x))K_B(x).$$

Since the integrand has an absolute value of at most 1, it satisfies the hypothesis of Lebesgue's Dominated Convergence Theorem. This allows us to move the limit inside the integral to yield the equation

$$\lim_{m\to\infty}\frac{1}{m}\sum_{k=0}^{m-1}\lambda(T^{-k}(A)\cap B) = \int_B \lim_{m\to\infty}\frac{1}{m}\sum_{k=0}^{m-1}K_A(T^k(x)) = \lambda(A)\lambda(B).$$

Next, assume that the equation in the corollary holds for all measurable sets A and B. If A is T-invariant, then $\lambda(T^{-n}(A)\cap A^c) = \lambda(A\cap A^c) = 0$ for all n. This means that $\lambda(A)\lambda(A^c) = 0$, so either A or A^c has measure zero. $\square$

5. The Dyadic Adding Machine and Approximation Techniques

There are several methods for proving the ergodicity of an invariant transformation. Here we introduce a method based on approximation structures, while Walken's lecture notes [4, p. 44-50] provide an example of an alternate technique that uses Fourier series. We begin by defining a transformation known as the dyadic adding machine, and we will develop the approximation techniques mentioned by using this transformation as an example. The construction of the dyadic adding machine uses the idea of replication, which is helpful for generating examples and counterexamples in ergodic theory. Friedman's article [1] and the sections of Silva's book [3, pp. 109-115 and 218-226] that deal with the Hajian-Kakutani transformation and Chacón's transformation demonstrate this latter point.

Suppose $I_1 = [a, a+b)$ and $I_2 = [(a+c), (a+c)+b)$ are two disjoint intervals of equal length. If the map $T : I_1 \to I_2$ is said to be represented by Figure 1, then T is given by the translation $T(x) = x + c$.

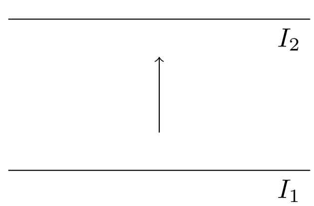

Figure 1. A representation of T

Similarly, if the interval I_3 is of the same length and type (left-closed, right-open) as I_1 and I_2 and disjoint from both, then the map $T^{'} : I_1 \cup I_2 \to I_2 \cup I_3$ represented by Figure 2 is defined by translating I_1 to I_2 and I_2 to I_3, and maps involving any finite number of disjoint intervals of equal length and identical type may be represented and defined analogously.

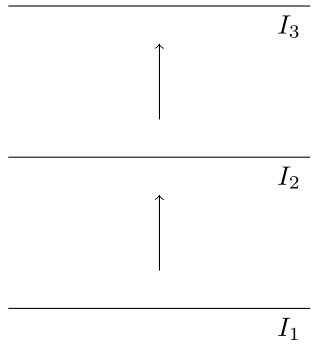

FIGURE 2. A representation of $T^{'}$

A map represented in the form above is called a **tower**, its representation is called a **stack**, and an interval in the stack is called a **level**. For example, the image in Figure 1 is a stack, the map $T(x) = x+c$ is the tower it represents, and I_1 and I_2 are the only levels of the stack. We construct the dyadic adding machine by inductively defining for each integer $n > 0$ a tower T_n and using these towers to define an infinitely piecewise function.

Let I be the *interval* $[0, 1)$. To define the *stack* S_1, divide I in half into $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1)$ and place the latter interval above the former. Let T_1 be the tower represented by S_1. To define S_2, divide S_1 in half vertically to form two stacks and place the right stack above the left stack. Let T_2 be the tower represented by S_2. As before, the intervals are left-closed and right-open. This process is illustrated in Figure 3.

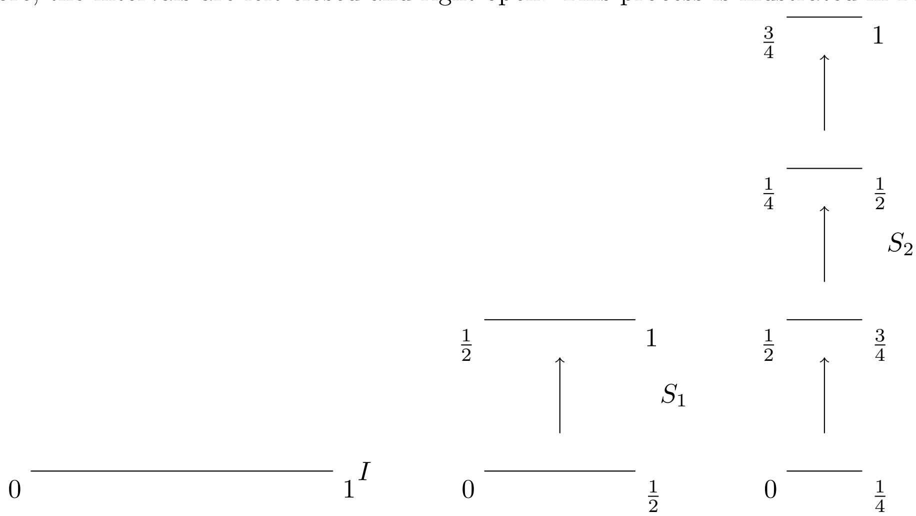

FIGURE 3. Constructing the dyadic adding machine

If S_n has been constructed for some $n \in \mathbb{N}$ and T_n is the tower corresponding to S_n, then S_{n+1} is the stack formed by dividing S_n in half and stacking the right stack above the left stack in the manner described above, and T_{n+1} is defined to be the tower corresponding to S_{n+1}. Readers may find it helpful to explicitly construct a few more stacks.

It is apparent from this process that these towers are closely related to the **dyadic numbers**, or numbers of the form $\frac{k}{2^n}$ where both k and n are integers. For the rest of this paper, let $\mathcal{D}$ refer to the set of left-closed, right-open **dyadic intervals**, or intervals whose endpoints are dyadic numbers, in $[0, 1)$. For a given n, let $\mathcal{D}^n$ refer to the subset of $\mathcal{D}$ consisting of intervals whose endpoints are of the form $\frac{k}{2^n}$. Several key facts about the towers T_n follow immediately from the construction above, and we summarize these in the next proposition.

Proposition 5.1. *Suppose $n \in \mathbb{N}$, and let S_n and T_n be as above. The following statements are true.*

(1) Each level of S_n is in $\mathcal{D}^n$.

(2) *T_n maps from* $\bigcup_{D\in\mathcal{D}^n} D \setminus [\frac{2^n-1}{2^n}, 1)$ *onto* $\bigcup_{D\in\mathcal{D}^n} D \setminus [0, \frac{1}{2^n})$.
(3) *For any natural number k less than n, $T_n = T_k$ on the domain of the latter.*
(4) *Let m be the Lebesgue measure. If $D \in \mathcal{D}^n \setminus [0, \frac{1}{2^n})$, then $m(T_n^{-1}(D)) = m(D)$.*
(5) *If $D \in \mathcal{D}^n \setminus [\frac{2^n-1}{2^n}, 1)$, then $m(T_n(D)) = m(D)$.*

The **dyadic adding machine** is the transformation $\mathcal{K}$ on $[0,1)$ defined by the following piecewise function.

$$\mathcal{K}(x) := \begin{cases} T_1(x) & : x \in [0, \frac{1}{2}) \\ T_2(x) & : x \in [\frac{1}{2}, \frac{3}{4}) \\ T_3(x) & : x \in [\frac{3}{4}, \frac{7}{8}) \\ ... \end{cases}$$

We see from this definition that if $n \in \mathbb{N}$, then $\mathcal{K}$ agrees with T_n on the domain of the latter. Take care to note that the intervals in $\mathcal{D}^n$ do not appear in S_n in order. Re-examine Figure 3 to see this.

It is easy to see from the piecewise definition of $\mathcal{K}$ that $\mathcal{K}$ is an invariant transformation of $[0,1)$. We will re-prove this by introducing some approximation techniques. This provides greater generality and at the same time gives us tools to prove the ergodicity of $\mathcal{K}$. We can see from Proposition 5.1 that $\mathcal{K}$ preserves the measure of dyadic intervals in $[0,1)$, and most of the next two pages are aimed at understanding and proving the following theorem, which allows us to extend the measure-preserving property of $\mathcal{K}$ on $\mathcal{D}$ to all of $[0,1)$.

Theorem 5.2. *Suppose that $(X, \mathcal{L}, \lambda)$ is a finite measure space, $\mathcal{S}$ is a sufficient semiring on X, and T is a transformation on X such that $\lambda(T^{-1}(S)) = \lambda(S)$ for any $S \in \mathcal{S}$. Then T is an invariant transformation on X.*

Note that T is not even assumed to be measurable. In our proof of this theorem, we show that the measurability of T can be derived form the fact that T is measure-preserving on $\mathcal{S}$.

The set $\mathcal{D}$ has the following structure. If $D_1, D_2 \in \mathcal{D}$, then

(1) $D_1 \cap D_2 \in \mathcal{D}$, and
(2) there exists a finite number of disjoint elements $C_1, .., C_n$ in $\mathcal{D}$ such that $D_1 \setminus D_2 = \bigcup_{i=1}^{n} C_i$.

Any non-empty collection of sets satisfying the above properties is called a **semiring**. Note that any semiring contains $\emptyset$. If the elements of a semiring $\mathcal{S}$ are subsets of some set X, then $\mathcal{S}$ is called a **semiring on** X. Suppose that $(X, \mathcal{L}, \lambda)$ is a measure space and $\mathcal{S}$ is a semiring consisting of subsets of X of finite measure. $\mathcal{S}$ is called a **sufficient semiring on X** if $A \in \mathcal{L}$ implies that

$$\lambda(A) = \inf_{A \subset \bigcup_{n=1}^{\infty} S_n} \sum_{n=1}^{\infty} \lambda(S_n),$$

where $S_n \in \mathcal{S}$. These definitions do not appear to be widely used and follow those in Silva's book [3, pp. 39-41].

Let m denote the Lebesgue measure. We show that $\mathcal{D}$ is a sufficient semiring on $[0,1)$ by demonstrating that any open interval in $[0,1)$ can be approximated arbitrarily closely by a finite union of dyadic intervals. Suppose that $I = (a,b)$ is an interval contained in $[0,1)$ and that $\epsilon > 0$ is given. Choose $n \in \mathbb{N}$ so that $\frac{1}{2^n} < \frac{\epsilon}{2}$. Since the numbers of the form $\frac{k}{2^n}$ where $k = 0, 1, 2, ..., 2^n - 1$ partition $[0,1)$ into intervals of length $\frac{1}{2^n}$, there exists a non-negative integer m such that $\frac{m}{2^n}$ is to the left of a (or equal to a if $a = 0$) and $a - \frac{m}{2^n} \leq \frac{1}{2^n}$. Similarly, there exists a non-negative integer l such that $\frac{l}{2^n}$ is to the right of b (or equal to b if $b = 1$) and $\frac{l}{2^n} - b \leq \frac{1}{2^n}$. Let $\mathcal{D}_I := D^n \cap [\frac{m}{2^n}, \frac{l}{2^n})$. Then

$$I \subset \bigcup_{D \in \mathcal{D}_I} D$$

and

$$m(I) \leq \bigcup_{D \in \mathcal{D}_I} m(D) < m(I) + \epsilon.$$

Suppose that A is a measurable set. Then there exists a collection of open intervals $\{I_n\}$ in $[0,1)$ such that

$$m(A) \leq \sum_{n=1}^{\infty} m(I_n) < m(A) + \epsilon.$$

Since ϵ is arbitrary, for each I_n, there exists $\mathcal{D}_n := \mathcal{D}_{I_n}$ such that

$$m(I_n) \leq \bigcup_{D \in \mathcal{D}_n} m(D) < m(I_n) + (m(A) + \epsilon - \sum_{m=1}^{\infty} m(I_m))\frac{1}{2^n}.$$

Then

$$\sum_{n=1}^{\infty} \sum_{D \in \mathcal{D}_n} m(D) < m(A) + \epsilon.$$

Now let

$$m_*(A) := \inf_{A \subset \bigcup\limits_{n=1}^{\infty} D_n} \sum_{n=1}^{\infty} m(D_n),$$

where $D_n \in \mathcal{D}$. The preceding inequality shows that $m_*(A) < m(A) + \epsilon$, and since ϵ is arbitrary,

$$m_*(A) \leq m(A).$$

Since, by definition, $m(A) \leq m_*(A)$, this proves that $\mathcal{D}$ is a sufficient semiring on $[0,1)$.

Carathéodory's condition states that given an outer measure λ^*, a set A is measurable iff for every set S, $\lambda^*(S) = \lambda^*(S \cap A) + \lambda^*(S \setminus A)$. Therefore any null set, or set of outer measure zero, is measurable, and we will prove Theorem 5.2 by writing A as the difference of a null set N and a set H_A constructed using members of the sufficient semiring. In the next lemma, we construct H_A and highlight its essential properties.

Lemma 5.3. *Suppose that $(X, \mathcal{L}, \lambda)$ is a finite measure space and $\mathcal{S}$ is a sufficient semiring on X. If A is measurable, then there exists a set H_A that satisfies the following:*

(1) $H_A = \bigcap\limits_{n=1}^{\infty} H_n$, where each H_n is a countable union of elements of $\mathcal{S}$,
(2) $H_1 \supset H_2 \supset H_3 ... \supset H_n ... \supset H_A \supset A$, and
(3) $\lambda(H_A \setminus A) = 0$.

Proof. By definition, given $\epsilon > 0$ there exists a set

$$H^{\epsilon} = \bigcup_{n=1}^{\infty} S_n^{\epsilon}$$

where $S_i^{\epsilon} \in \mathcal{S}$ such that $A \subset H^{\epsilon}$ and $\lambda(H^{\epsilon} \setminus A) < \epsilon$. For any natural number n, define

$$H_n := H^1 \cap H^{\frac{1}{2}} \cap ... H^{\frac{1}{n}} = \bigcap_{m=1}^{\infty} \bigcup_{i=1}^{n} S_m^{\frac{1}{i}}.$$

Since $\mathcal{S}$ is closed under finite intersections, H_n is a countable union of elements of $\mathcal{S}$. Now let H_A be defined as in (1). Then (2) holds by the definition of H_n, and for any natural number n,

$$\lambda(H \setminus A) \leq \lambda(H_n \setminus A) \leq \lambda(H^{\frac{1}{n}} \setminus A) < \frac{1}{n}.$$

Thus H_A also satisfies (3). □

Proof of Theorem 5.2 Suppose A is a measurable set, and let H_A and $H_n = \bigcup\limits_{m=1}^{\infty} S_{n,m}$ be as in the proof of Lemma 5.3. Since taking preimages preserves, in the sense of Proposition 2.1, unions and intersections, the sets H_n and the set H_A are measurable. Since T^{-1} preserves the measure of elements of $\mathcal{S}$,

$$\lambda(T^{-1}(H_A)) = \lambda(\bigcap_{n=1}^{\infty} T^{-1}(H_n)) = \lambda(\bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} T^{-1}(S_{n,m})) = \lambda(\bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} (S_{n,m})) = \lambda(H_A).$$

Define the set $N := H_A \setminus A$; N is a null set by (3) of Lemma 5.3. We can apply the argument in the preceding paragraph to N to show that both H_N and $T^{-1}(H_N)$ are null sets, so

$$0 \leq \lambda^*(T^{-1}(N)) \leq \lambda^*(T^{-1}(H_N)) = 0.$$

$T^{-1}(N)$ is a null set and is measurable by Carathéodory's condition. Since

$$T^{-1}(A) = T^{-1}(H_A) \setminus T^{-1}(N),$$

$T^{-1}(A)$ is measurable and $\lambda(T^{-1}(A)) = \lambda(A)$ for any measurable set A. $\square$

Since $\mathcal{K}$ is an invariant transformation, we can ask whether $\mathcal{K}$ is ergodic. To prove that $\mathcal{K}$ is ergodic, we need to derive some further results about sufficient semirings. At this point, readers are encouraged to draw a $\mathcal{K}$-invariant set on some of the stacks S_n. This will provide some intuition for the discussion that follows.

Proposition 5.4. *The dyadic adding machine $\mathcal{K}$ is an ergodic transformation on the Lebesgue space $[0, 1)$.*

If A is a $\mathcal{K}$-invariant set of positive measure and D and $D^{'}$ are dyadic intervals in $[0, 1)$, then

$$m(A) = \frac{m(A \cap [0,1))}{m([0,1))} = \frac{m(A \cap D)}{m(D)} = \frac{m(A \cap D')}{m(D')}.$$

In other words, the measure of A is equal to the proportion of any dyadic interval that A occupies. If, for any given $0 < \delta < 1$, we can find a dyadic interval D^δ that satisfies the inequality

$$m(D^\delta) \geq m(A \cap D^\delta) > (1 - \delta)m(D^\delta),$$

then $m(A) = 1$. In the the next two lemmas, we show that D^δ exists for any δ.

Lemma 5.5. *Suppose that $(X, \mathcal{L}, \lambda)$ is a finite measure space and $\mathcal{S}$ is a sufficient semiring on X. For any measurable set A and any $\epsilon > 0$, there exists a set G^A_ϵ that is a finite union of elements of $\mathcal{S}$ such that $\lambda(A \triangle G^A_\epsilon) < \epsilon$.*

Proof. Let $\delta := \frac{\epsilon}{2}$, and let $H^\delta = \bigcup_{i=1}^{\infty} S_i$ be as in the proof of Lemma 5.3. Since $\bigcup_{i=1}^{N} S_i \subset \bigcup_{i=1}^{M} S_i$ for $M > N$, $\lim_{n\to\infty} \lambda(\bigcup_{i=1}^{n} S_i) = \lambda(\bigcup_{i=1}^{\infty} S_i)$. Thus there exists an n such that $\lambda(H^\delta \setminus \bigcup_{i=1}^{n} S_i) < \epsilon$. Let $G^A_\epsilon := \bigcap_{i=1}^{n} S_i$. Then $\lambda(A \triangle G^A_\epsilon) = \lambda(A \triangle G^A_\epsilon) + \lambda(G^A_\epsilon \triangle A) \leq \lambda(H^\delta \triangle G^A_\epsilon) + \lambda(H^\delta \triangle A) < \epsilon$. $\square$

Lemma 5.6. *Suppose that $(X, \mathcal{L}, \lambda)$ is a finite measure space and $\mathcal{S}$ is a sufficient semiring on X. For any set A of positive measure and any $0 < \delta < 1$, there exists a set $S^\delta \in \mathcal{S}$ such that $\lambda(A \cap S^\delta) > (1 - \delta)\lambda(S^\delta)$.*

Proof. Choose ϵ so that $0 < \epsilon < \frac{\delta}{2-\delta}$, and define $\beta := \epsilon\lambda(A)$. Let $G := G^A_\beta = \bigcap_{i=1}^{N} S_i$ be as in the statement of Lemma 5.5. We apply λ to set inclusion $G = [(G \cap A) \cup (G \setminus A)] \subset [A \cup (A \setminus G) \cup (G \setminus A)]$ to get

$$\lambda(G \setminus A) \leq \lambda(A \cup [A \triangle G]) < (1 + \epsilon)\lambda(A).$$

Similarly, we apply λ to the set inclusion $A = [(A \cap G) \cup (A \setminus G)] \subset [(A \cap G) \cup (A \triangle G)]$ to get

$$\lambda(A) < \lambda(A \cap G) + \epsilon\lambda(A),$$

or

$$\lambda(A \cap G) > (1 - \epsilon)\lambda(A).$$

Suppose, in contradiction to the lemma, that there is no set $S^\delta \in \mathcal{S}$ that satisfies the stated inequality. In particular, none of the sets S_i satisfy the inequality. Then $\lambda(A \cap G) \leq (1 - \delta)\lambda(G) < (1 - \delta)(1 + \epsilon)\lambda(A)$, so

$$(1 - \delta)(1 + \epsilon) > (1 - \epsilon).$$

On the other hand, $(1 + \epsilon) < \frac{2}{2-\delta}$ and $(1 - \epsilon) > \frac{2-2\delta}{2-\delta} = \frac{2(1-\delta)}{2-\delta}$ by the choice of ϵ, so

$$\frac{(1 - \epsilon)}{(1 + \epsilon)} > (1 - \delta).$$

This is a contradiction, so the lemma is true. $\square$

Proof of Proposition 5.4 Suppose A is a $\mathcal{K}$-invariant set of positive measure, and let $0 < \delta < 1$ be given. Since $\mathcal{D}$ is a sufficient semiring on the Lebesgue space $[0, 1)$, there exists a dyadic interval D^δ such that

$$\frac{m(A \cap D^\delta)}{m(D^\delta)} > (1 - \delta).$$

By the remarks above,

$$1 \geq m(A) > (1 - \delta).$$

Since δ is arbitrary, $m(A) = 1$. $\square$

Silva's book [3, pp. 100-101 and 107-108] contains some more examples of using sufficient semirings to prove ergodicity. The latter group of pages cited presents a proof that $\mathcal{K}$ is ergodic that is different from the one in this paper.

Acknowledgments. I would like to thank my mentor Gong Chen for answering my many questions and reading my drafts.

References

[1] N.A. Friedman, *Replication and Stacking in Ergodic Theory*, Amer. Math. Monthly **99** (1992), no.1, 31–41.
[2] P.R. Halmos, *Measurable Transformations*, Bull. Amer. Math. Soc. **55** (1949), no. 11, 1015–1034.
[3] C.E. Silva, *Invitation to Ergodic Theory*, Stud. Math. Libr., Vol. 42, Amer. Math. Soc., Providence, RI, 2007.
[4] C. Walken, *MATH41112/61112 Ergodic Theory*, Course notes, retrieved July 14, 2014 from http://`www.maths.manchester.ac.uk/~cwalkden/ergodic-theory/ergodic_theory.pdf`