INTRODUCTION TO FURSTENBERG'S $\times 2 \times 3$ CONJECTURE

BEN CALL

ABSTRACT. In this paper, we introduce the rudiments of ergodic theory and entropy necessary to study Rudolph's partial solution to the $\times 2 \times 3$ problem posed by Furstenberg.

CONTENTS

1. INTRODUCTION

In 1967, Furstenberg conjectured that for transformations $Tx = px \mod 1$ and $Sx = qx \mod 1$ on S^1, where $p, q \neq 1$, that any invariant, ergodic, Borel probability measure is either atomic or the Lebesgue measure. Rudolph introduced a partial solution to this problem, proving that if this measure is not the Lebesgue measure and $\text{GCD}(p, q) = 1$, then both T and S have entropy zero with respect to the measure. In this paper, we seek to provide an overview of the path that Rudolph takes in his proof. We first give a brief overview of ergodic theory and symbolic dynamics before moving on to a discussion of entropy. We then discuss Rudolph's results.

We assume that the reader is already familiar with the basic concepts of Lebesgue measure theory.

2. ERGODIC THEORY

In this section, we will discuss the rudiments of ergodic theory, with a focus on the concepts of measure-preserving transformations and ergodicity. Conveniently, the transformations that Furstenberg uses in his conjecture provide useful examples, and so consequently, throughout this section, we will use the dynamical system $(S^1, \mathfrak{L}, \lambda, T)$ as an example throughout, where $Tx = 2x \mod 1$ is the doubling map

---

Date: August 25, 2013.

and S^1 is $[0,1]$ with endpoints identified.

We first provide some preliminary definitions.

Definition 2.1. A **measure space** is a triple $(X, \mathcal{S}, \mu)$ where X is a non-empty set, $\mathcal{S}$ is a σ-algebra on X, and μ is a measure on $\mathcal{S}$.

Example 2.2. We use $(\mathbb{R}, \mathfrak{L}, \lambda)$ to refer to the measure space of $\mathbb{R}$, with $\mathfrak{L}$ denoting the set the of Lebesgue measurable subsets of $\mathbb{R}$ and λ denoting Lebesgue measure.

Proving anything for measure spaces can prove difficult because of how large a σ-algebra can be, so we introduce the concept of semi-rings, which greatly simplifies this problem.

Definition 2.3. A **semi-ring** on a nonempty set X is a collection $\mathcal{R}$ of subsets of X such that

(1) $\mathcal{R}$ is nonempty
(2) $\mathcal{R}$ is closed under finite intersections
(3) If $A, B \in \mathcal{R}$, then $A \setminus B$ can be written as a finite union of disjoint elements of $\mathcal{R}$.

Definition 2.4. A semi-ring $\mathcal{C}$ is a **sufficient** semi-ring for a measure space $(X, \mathcal{S}, \mu)$ if for every $A \in \mathcal{S}$,

$$\mu(A) = \left\{ \sum_{j=1}^{\infty} \mu(I_j) : A \subset \bigcup_{j=1}^{\infty} I_j \text{ and } I_j \in \mathcal{C} \text{ for } j \geq 1 \right\}.$$

The reason that this construction is so useful is that in order to prove a property, such as ergodicity, on a measure space, one need only prove it on a sufficient semi-ring for that measure space. This fact and the following example will go without proof, as it would be digressive.

Example 2.5. The collection of dyadic intervals

$$\left\{ D_{n,k} = \left[\frac{k}{2^n}, \frac{k+1}{2^n} \right) \mid n \in \mathbb{Z}^+, 0 \leq k \leq 2^n - 1 \right\}$$

forms a sufficient semi-ring for the measure space $([0,1), \mathfrak{L}, \lambda)$.

Having introduced the concept of measure spaces and semi-rings, we move on to the notion of measure-preserving transformations.

Definition 2.6. A transformation $T : X \to X$ is **measure-preserving** if for all $A \in \mathcal{S}$, $\mu(A) = \mu(T^{-1}(A))$. In this case, we say that μ is an invariant measure for T.

Example 2.7. Lebesgue measure is an invariant measure for the doubling map $Tx = 2x \mod 1$.

Proof. Let $A \subset S^1$ be Lebesgue measurable. We may approximate A with a set of dyadic intervals $\{D_i\}$. Consider the preimage of any dyadic interval D_i. It consists of two disjoint dyadic intervals, each of measure $\frac{\lambda(D_i)}{2}$, and consequently, $\lambda(T^{-1}(D_i)) = \lambda(D_i)$. Thus, we see

$$\lambda(T^{-1}(A)) = \sum_{i=1}^{\infty} \lambda(T^{-1}(D_i)) = \sum_{i=1}^{\infty} \lambda(D_i) = \lambda(A)$$

and so, Lebesgue measure is an invariant measure for the doubling map. $\square$

We now introduce a notion of invariance that will be used in our definition of ergodicity.

Definition 2.8. A set $A \subset X$ is **strictly T-invariant** when $T^{-1}(A) = A$.

We now introduce useful notation that allows us to work with properties that are true almost everywhere.

Definition 2.9. A measure-preserving transformation T is **ergodic** if whenever A is strictly invariant and measurable, then either $\mu(A) = 0$ or $\mu(A^C) = 0$.

We will now demonstrate that the doubling map is ergodic.

Theorem 2.10. *The doubling map is ergodic.*

Proof. Let A be a measurable, strictly invariant set. By Example 2.5, we know that the set of dyadic intervals forms a sufficient semi-ring, so we may approximate A with them. Thus, there exists a collection of dyadic intervals $K = \bigcup_{i=1}^{\infty} D_i$ such that $\lambda(K \setminus A) = 0$. Additionally, because

$$\lambda(T^{-1}(K \setminus A)) = \lambda(T^{-1}(K) \setminus T^{-1}(A)) = \lambda(T^{-1}(K) \setminus A),$$

we see that for all $i \geq 1$, $\lambda(T^{-i}(K) \setminus A) = 0$. By subadditivity, we are given that

$$\lambda(\bigcup_{i=0}^{\infty} T^{-i}(K) \setminus A) \leq \sum_{i=1}^{\infty} T^{-i}(K) \setminus A = 0.$$

Thus, it suffices to show that $\lambda(\bigcup_{i=0}^{\infty} T^{-i}(K))$ is 0 or 1. Now let $D_{n,k} \in K$, and note that

$$\lambda(\bigcup_{i=0}^{\infty} T^{-i}(D_{n,k})) \leq \lambda(\bigcup_{i=0}^{\infty} T^{-i}(K)).$$

If each interval in K is the empty interval, then all intervals that approximate A are empty, then we are done, and $\lambda(A) = 0$. Thus, we choose $D_{n,k}$ to have some positive length. We can easily see that $\lambda(\bigcup_{i=0}^{\infty} T^{-i}(D_{n,k})) = 1$. Thus, $\lambda(A) = 1$, and we have shown that if A is an invariant set, then $\lambda(A) = 1$ or 0, and the proof is complete.

$\square$

Finally, we introduce a notion of isomorphism between two dynamical systems.

Definition 2.11. $(X', \mathcal{S}', \mu', T')$ is a **factor** of $(X, \mathcal{S}, \mu, T)$ if there exist measurable subsets $X_0 \subset X$ and $X_0{}' \subset X'$ of full measure such that $T(X_0) \subset X_0$ and $T'(X_0') \subset X_0'$ and there exists $\varphi : X_0 \to X_0'$ that is onto such that for all $A \in \mathcal{S}'(X_0')$:

(1) $\varphi^{-1}(A) \in \mathcal{S}(X_0)$
(2) $\mu(\varphi^{-1}(A)) = \mu'(A)$
(3) $\varphi(T(x)) = T'(\varphi(x)) \forall x \in X_0$

Definition 2.12. Let $(X, \mathcal{S}, \mu, T)$ and $(X', \mathcal{S}', \mu', T')$ be finite measure-preserving dynamical systems. They are **isomorphic** if the properties in 2.11 hold, and the factor map φ is a bijection.

We now move on to methods of constructing these isomorphisms.

3. Symbolic Dynamics

Another way to simplify the study of different dynamical systems is to encode them in terms of infinite strings of numbers. In this section, we will provide some of the basic tools used as well as demonstrate the usefulness of thinking about systems in this way.

Definition 3.1. For all $N \in \mathbb{Z}^+$, the set Σ_N^+ consists of all infinite sequences $a_1a_2a_3\dots$ of the symbols from $\{0, \cdots, N-1\}$.

We also need to define a σ-algebra, for which we need to define a metric d on Σ_N^+.

Definition 3.2. Letting x_i denote the ith symbol in the sequence $x \in \Sigma_N^+$, we define a metric d by

$$d(x,y) = \begin{cases} 2^{-\min\{i: x_i \neq y_i\}} & x \neq y \\ 0 & x = y \end{cases}.$$

Definition 3.3. Let $\mathcal{B}$ be the σ-algebra generated by the open sets in (Σ_N^+, d)

In order to create an isomorphic symbolic dynamical system, we need a measure as well. We first however must define a map from S^1 to Σ_N^+.

Definition 3.4. Almost every point in S^1 has a unique base N representation of the form $x = \sum\limits_{i=1}^{\infty} \frac{x_i}{N^i}$ where $0 \leq x_i \leq N-1$, and for all such x, we define $\varphi(x) = x_1x_2x_3\dots x_n\dots$ a map from S^1 to Σ_N^+.

Because each representation is unique, it is easy to see that φ is $1-1$, and we will not prove that φ is continuous here. However, we can now use it to define a measure on Σ_N^+ as follows.

Definition 3.5. For all measurable $A \subset \Sigma_N^+$, define $\nu(A) = \lambda(\varphi^{-1}(A))$.

A common transformation on Σ_N^+ is the shift, which is defined as follows:

Definition 3.6. The **shift** σ maps $x_1x_2x_3x_4\dots$ to $x_2x_3x_4\dots$.

In other words, it sets $x_i = x_{i+1}$ for all $i \geq 1$.

Symbolic dynamics allows us to establish the relation between certain dynamical systems like the following:

Example 3.7. $(S^1, \mathfrak{L}, \lambda, T)$ is a factor of $(\Sigma_2^+, \mathcal{B}, \nu, \sigma)$, where T is the doubling map.

Proof. We use the map φ defined in 3.2 as the factor map. We already know that φ is onto almost everywhere from our definition. We know from our definition of ν that given $A \in \mathcal{B}$, $\varphi^{-1}(A) \in \mathfrak{L}$. Furthermore, we see from that same definition $\lambda(\varphi^{-1}(A)) = \nu(A)$, so all we must show is that $\varphi(T(x)) = \sigma(\varphi(x))$ for all x with unique binary representations. Suppose that $x = \sum\limits_{i=1}^{\infty} \frac{a_i}{2^i}$. Then

$$T(x) = a_1 + \sum_{i=1}^{\infty} \frac{a_{i+1}}{2^i} \mod 1 = \sum_{i=1}^{\infty} \frac{a_{i+1}}{2^i},$$

so $\varphi(T(x)) = a_2a_3a_4 \dots a_n \dots$. We see

$$\sigma(\varphi(x)) = \sigma(a_1a_2a_3 \dots a_n \dots) = a_2a_3a_4 \dots a_n \dots$$

and consequently, $\varphi(T(x)) = \sigma(\varphi(x))$, and our proof is complete. $\square$

Symbolic dynamics serve as a powerful tool for us to use when studying dynamical systems because they allow us to easily represent the actions that we take in an easy to conceptualize manner. Rudolph uses this in [3] as a method of simplifying the proofs about the doubling map into proofs about shifts.

4. Entropy

We now introduce the concept of measure-theoretic entropy, which formalizes a concept of uncertainty and information in dynamical systems. It also serves as an invariant under isomorphism.

Throughout this section we work in a measure space $(X, \mathcal{S}, \mu)$ where $\mu(X) = 1$.

Definition 4.1. A **partition** of X is a finite collection of essentially disjoint measurable sets C_i such that $\bigcup_i C_i = X \mod \mu$.

Definition 4.2. Let ξ and ξ' be partitions of X. Then ξ' is a refinement of ξ, or $\xi \leq \xi'$ if for all $C \in \xi'$, there exists $D \in \xi$ such that $C \subset D$.

Definition 4.3. The **common refinement** of partitions ξ and η is denoted by $\xi \vee \eta$, where

$$\xi \vee \eta = \{C \cap D \mid C \in \xi, D \in \eta\}$$

Definition 4.4. The entropy of a finite partition $\xi = \{C_1, C_2, \cdots, C_n\}$ is denoted

$$H(\xi) = -\sum_{i=1}^{n} \mu(C_i) \log \mu(C_i)$$

where we assign $0 \log 0 = 0$

In other words, the entropy of a partition is the average information of the elements of the partition, where the lower the probability of an element of a partition, or event, the higher the information gained.

We now consider the notion of entropy of a given partition with respect to another partition

Definition 4.5. If $\xi = \{C_i\}$ and $\eta = \{D_j\}$, then the **conditional entropy of ξ with respect to η** is

$$H(\xi \mid \eta) = -\sum_j \mu(D_j) \left(\sum_i \mu(C_i \mid D_j) \log \mu(C_i \mid D_j) \right)$$

where

$$\mu(C \mid D) = \frac{\mu(C \cap D)}{\mu(D)}$$

Here we provide some basic facts without proof that will be used in later proofs.

Proposition 4.6. *Let ξ, η, and ζ be finite partitions, and T be a measure-preserving transformation. Then:*

(1) If $\eta \leq \zeta$, then $H(\xi \mid \eta) \geq H(\xi \mid \zeta)$.

(2) $H(\xi \vee \eta) = H(\xi) + H(\eta \mid \xi)$.
(3) $H(T^{-1}(\xi)) = H(\xi)$.

Having established the notion of entropy of partitions, we extend the notion to transformations on these partitions. Before we do that though, we need to establish some notation.

Definition 4.7. Let T be a measure-preserving transformation, and let $\xi = \{C_\alpha\}$ be a partition. Then for all integers k, $T^{-k}(\xi) = \{T^{-k}(C_\alpha) \mid C_\alpha \in \xi\}$.

Finally, we need a notion of applying the transformation to the partition multiple times. Thus, for a partition ξ and a transformation T, we use the notation

$$\xi^n = \xi \vee T^{-1}(\xi) \vee \cdots \vee T^{-n+1}(\xi).$$

We are now ready to define a measure of entropy of a transformation given a particular partition.

Definition 4.8. Let T be a measure-preserving transformation and let ξ be a partition. Then the **metric entropy of T relative to ξ** is

$$h(T, \xi) = \lim_{n\to\infty} \frac{1}{n} H(\xi^n).$$

We will now prove the following proposition that we will use to simplify our calculations, as well as provide an intuitive explanation of measure-theoretic entropy.

Proposition 4.9. $h(T, \xi) = \lim_{n\to\infty} H(\xi \mid T^{-1}(\xi^n))$

Proof. From 4.6 we have that $H(\xi \mid \eta) \geq H(\xi \mid \zeta)$ if $\eta \leq \zeta$. When combined with the fact that $T^{-1}(\xi^n) \leq T^{-1}(\xi^{n+1})$, we see that $H(\xi \mid T^{-1}(\xi^n))$ is non-increasing with n. We apply 4.6.2 and 4.6.3

$$\begin{aligned}
H(\xi^n) &= H(T^{-1}(\xi^{n-1}) \vee \xi) \\
&= H(\xi^{n-1}) + H(\xi \mid T^{-1}(\xi^{n-1})) \\
&= H(\xi^{n-2}) + H(\xi \mid T^{-1}(\xi^{n-2})) + H(\xi \mid T^{-1}(\xi^{n-1})) \\
&\vdots \\
&= H(\xi) + \sum_{i=1}^{n-1} H(\xi \mid T^{-1}(\xi^i)).
\end{aligned}$$

When we divide this by n and take the limit, we obtain the result

$$h(T, \xi) = \lim_{n\to\infty} \frac{1}{n} H(\xi^n) = \lim_{n\to\infty} H(\xi \mid T^{-1}(\xi^n)).$$

□

With this proposition, we see that measure-theoretic entropy can be thought of as the information obtained by a transformation of the state conditioned on the knowledge of all past states.

Given this notion of entropy of a transformation relative to a partition, we wish to establish a notion of entropy independent of an individual partition, so that it depends only on the transformation itself.

Definition 4.10. Let T be a measure preserving transformation. Then the **measure-theoretic entropy** of T is defined as

$$h(T) = \sup\{h(T, \xi) \mid \xi \text{ is a finite partition}\}$$

We now provide some definitions which will allow us to more easily calculate entropy.

Definition 4.11. A sequence of partitions (ξ_n) is **generating** if for every measurable set $A \subset X$, there exists $n \in \mathbb{N}$ such that A can be approximated up to a set of measure 0 by a union of elements of $\bigvee_{i=1}^{n} \xi$.

To obtain an equivalent definition, we introduce a function to quantify the difference between finite partitions.

Definition 4.12. Let ξ and η be finite partitions with m elements, (we may add null sets to whichever has fewer sets). Then

$$d(\xi, \eta) = \min_{\sigma \in S_m} \sum_{i=1}^{m} \mu(C_i \Delta D_{\sigma(i)}).$$

We see that an equivalent definition is

Definition 4.13. A sequence of partitions (ξ_n) is **generating** if for every finite partition ζ and for all $\delta > 0$, there exists $n \in \mathbb{N}$ such that for all $m \geq n$, there exists a partition $\zeta_m \leq \bigvee_{i=1}^{m} \xi_i$ such that $d(\zeta_m, \zeta) < \delta$.

Definition 4.14. A **generator** for T is a partition ξ such that the sequence (ξ_n) is generating.

In order to prove the following theorem, we now present some facts without proof.

Proposition 4.15. *If ξ and η are finite partitions and T is a measure-preserving transformation, then:*

(1) $h(T, \xi) \leq h(T, \eta) + H(\xi \mid \eta)$.
(2) $h(T, \xi \vee \eta) \leq h(T, \xi) + h(T, \eta)$.
(3) $h(T, \xi) = h(T, \bigvee_{i=0}^{n} T^{-1}(\xi))$.

Lemma 4.16. *Let η be a finite partition, and let ξ_n be a sequence of finite partitions such that $d(\xi_n, \eta) \to 0$. Then there exist finite partitions $\zeta_n \leq \xi_n$ such that $H(\eta \mid \zeta_n) \to 0$.*

Theorem 4.17. *Let ξ be a generator for T. Then $h(T) = h(T, \xi)$.*

Proof. Let η be an arbitrary finite partition. Then because ξ is a generator, there exist partitions $\zeta_n \leq \bigvee_{i=0}^{n} T^{-i}(\xi)$ such that $d(\zeta_n, \eta) \to 0$. For any $\delta > 0$, there exists $n \in \mathbb{N}$ and a partition ξ_n such that $\xi_n \leq \zeta_n \leq \bigvee_{i=0}^{n} T^{-i}(\xi)$ and $H(\xi_n \mid \eta) < \delta$ by the previous lemma. Combined with the facts from the previous proposition, we see

$$h(T, \eta) \leq h(T, \xi_n) + H(\eta \mid \xi_n) \leq h(T, \bigvee_{i=0}^{n} T^{-i}(\xi)) + \delta = h(T, \xi) + \delta.$$

Because δ was arbitrary, we obtain the desired result. $\square$

We will now provide an example calculation of entropy of the doubling map.

Example 4.18. We can see that $\xi = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$ is a generator for the doubling map as the dyadic intervals form a semi-ring for S^1. Thus, using the notation $D_{n,k} = [\frac{k}{2^n}, \frac{k+1}{2^n})$,

$$\begin{aligned} h(T) &= h(T, \xi) \\ &= \lim_{n\to\infty} \frac{1}{n} H(\xi^n) \\ &= \lim_{n\to\infty} \frac{-1}{n} \sum_{i=0}^{2^n-1} \mu(D_{n,i}) \log(\mu(D_{n,i})) \\ &= \lim_{n\to\infty} \frac{-1}{n} \sum_{i=0}^{2^n-1} \frac{1}{2^n} \log(2^{-n}) \\ &= \lim_{n\to\infty} \frac{-1}{n} \log(2^{-n}) \\ &= \log(2) \end{aligned}$$

and our proof is complete.

In fact, it can be shown that for any map $Tx = px$ mapped onto S^1 that $h(T) = \log(p)$. This notion of entropy is central to Rudolph's proof.

5. Overview of the paper

We now have the necessary background to give an overview of the direction that Rudolph takes in his proof in [3]. This can be broken down into three main portions. First, he constructs the symbolic descriptions of the transformations in question. He then relates the entropies of the two transformations. Finally, he proves that if the measure is not Lebesgue measure, then the transformations have entropy 0. Recall that throughout this section, we refer to the transformations $T = px \mod 1$ and $S = qx \mod 1$ on S^1, with $p, q \neq 1$.

5.1. **The symbolic representation.**

In order to construct a symbolic representation, we identify points in S^1 with arrays of symbols. To this end, we partition S^1 into pq intervals

$$I_j = \left[\frac{j}{pq}, \frac{j+1}{pq}\right], 0 \leq j \leq pq - 1$$

which is the common refinement of the partition according to T and the partition according to S following the same methods used in the doubling map. From here, we define transition matrices, $M_T = [a_{i,j}]$ and $M_S = [b_{i,j}]$, where $a_{i,j} = 1$ if and only if $I_j \subset T(I_i)$ and $b_{i,j} = 1$ if and only if $I_j \subset S(I_i)$.

We consider Σ_{pq}^+, which throughout this section we will call Σ, and introduce a useful definition.

Definition 5.1. A word of elements of Σ, $\mathbf{i} = [i_0, i_1, \cdots]$ is M_T**-allowed** if $a_{i_k, i_{k+1}} = 1$ for all $i_k \in \mathbf{i}$. The definition is similar for M_S-allowed words.

Using this idea, we associate each infinite word with a point in S^1 in the following manner.

Definition 5.2. For each infinite M_T-allowed word $\mathbf{i}$, let $x_i \in S^1$ be defined by

$$x_i = \bigcap_{j=1}^{\infty} T^{-j}(I_{i_j})$$

and for each infinite M_S-allowed word, there the construction is similar.

In order to construct an isomorphism, we wish to make the mapping $i \mapsto x_i$ $1-1$ almost everywhere, so we ignore the set of points on which either of these constructions are not $1-1$, which is

$$V = \left\{ x \in S^1 \mid x = \frac{t}{p^n q^m}, t, n, m \in \mathbb{Z} \right\}$$

and use $V' = S^1 \setminus V$.

Using these as building blocks, we construct the symbolic representation as follows.

Definition 5.3. An **array** y is an infinite collection of symbols, each of which is assigned a unique pair of indices i and j, such that $y(i,j)$ is the element in the ith row and the jth column.

We define Y to be the set of all arrays in which all rows are M_T-allowed and all columns are M_S-allowed. As demonstrated in our earlier example, on each array, T and S can be related to a left shift T_0 and a down shift S_0 respectively.

We are finally ready to construct the relationship between S^1 and Y. explicitly. In order to do this, we will show that Y consists of exactly the points in V'. First, for every $x \in V'$, there is a unique point $y_x \in Y$, by setting

$$y_x(n,m) = j \text{ if } T^n S^m(x) \in I_j \ \forall n,m \in \mathbb{N}$$

where $y_x(n,m)$ is the symbol in the nth position to the right and the mth position from the bottom. We will demonstrate that there does not exist $y \in Y$ that does not come from a point in S^1.

Lemma 5.4. *For any $a, b \in \{0, 1, \ldots, pq-1\}$ and $i \in \mathbb{N}$, there exist $y \in Y$ such that $y(i,i) = a$ and $y(i+1,i+1) = b$, but all such arrays agree on $y(i+1,i)$ and $y(i,i+1)$.*

Theorem 5.5. *The conjugation*

$$\varphi(y) = \bigcap_{i=j} T^{-j} S^{-j}(I_{y(i,j)})$$

is almost $1-1$ between (S_0, T_0, Y) and (S, T, S^1).

Proof. The previous lemma proves that the symbols $y(0,0), y(1,1), \ldots$ determine y completely. Furthermore, we also know that they determine a unique

$$x = \bigcap_i T^{-i} S^{-i}(I_{y(i,i)})$$

for all $x \in V'$. Thus, we see that $y(i,j) \in \Sigma$ is such that $T^i S^j(x) \in I_{y(i,j)}$ because the array is in Y and agrees with y on the diagonal. $\square$

From here, we then construct $\hat{Y}$, which is the set of doubly infinite arrays made up of doubly infinite sequences of elements of $\{0, 1 \ldots, pq-1\}$ in which all rows are M_T-allowed and all columns are M_S-allowed. We then define for $\hat{y} \in \hat{Y}$, $\hat{\varphi}(\hat{y})$ to be the restriction of $\hat{y}$ to Y, or in other words

$$\hat{\varphi}(\hat{y}) = \{y \in Y \mid y(a,b) = \hat{y}(a,b)\ \forall a, b \in \mathbb{N}\}.$$

We also define

$$\varphi(\hat{y}) = \varphi(\hat{\varphi}(\hat{y}))$$

to be its image in S^1, where φ is defined in Theorem 5.5. All of the previous results extend to arrays in $\hat{Y}$. We must use $\hat{Y}$ because we will use the notion of inverse it allows in Lemma 5.4 to construct probability functions. With the construction of the symbolic array complete, we now proceed to a discussion of entropy of invariant measures.

5.2. **Entropy and Invariant Measures.**

In this section, we consider all T and S invariant Borel probability measures on S^1, and we use μ to represent one of these measures. The ultimate goal of this section is to quantify a relationship between the entropies of T_0 and S_0 with respect to ergodic, invariant measures.

We begin by demonstrating that for any invariant measure μ, the relation constructed in the previous section can be extended to an isomorphism.

Lemma 5.6. *If $x \in V$ and $x \neq 0$, then $\mu(x) = 0$.*

Proof. Assume for contradiction that $x \in V$, $x \neq 0$, and $\mu(x) > 0$. Then because $x = \frac{t}{p^n q^m}$ for some t, n, and m, we know that there exist n, m such that $T^n S^m(x) = 0$. Thus, $\mu(0) > 0$, because μ is an invariant measure. But, we also see

$$\mu(0) = \mu(S^{-m}T^{-n}(0)) \geq \mu(\{0, x\}) \geq \mu(0) + \mu(x)$$

which implies that $0 \geq \mu(x)$, a contradiction. □

With this, we are now able to provide a unique T_0 and S_0 invariant measure on $\hat{Y}$ given μ.

Theorem 5.7. *If $\mu(0) = 0$, then μ lifts to a unique T_0 and S_0 invariant Borel probability measure on $\hat{Y}$.*

Proof. Because $\mu(V) = 0$ from the previous lemma, we can see that μ lifts to Y. Because $\hat{Y}$ is the limit of applying the inverse of $T_0 S_0$ to Y repeatedly, and μ is T_0 and S_0 invariant, we can see that μ lifts to a unique measure on $\hat{Y}$. □

At this point, we now stop dealing with measures on S^1, and begin considering T_0 and S_0 invariant measures.

We partition $\hat{Y}$ according to the symbol $\hat{y}(0,0)$, meaning that $\hat{y}_1, \hat{y}_2 \in \hat{Y}$ belong to the same element of the partition if and only if $\hat{y}_1(0,0) = \hat{y}_2(0,0)$, and we call this partition P. We omit the proof of the following lemma, but it is easily seen to be true because any measurable set may be approximated with small enough sets, which can be obtained after applying the shift transformation a suitable number of times.

Lemma 5.8. *P is a generator for $\hat{Y}$ under S_0 and T_0.*

Thus, we know that for any ergodic, T_0 and S_0 invariant measure μ, $h_\mu(T_0) = h_\mu(T_0, P)$ and $h_\mu(S_0) = h_\mu(S_0, P)$ by 4.17.

We are now ready to prove the major theorem of this section. However, we use a notion of entropy with respect to σ-algebras. We omit a discussion of these from our paper in order to keep it brief, and only note that many of the same propositions with regards to partitions hold here as well. However, we do provide the following equivalence, relating metric entropy of a transformation T_0 with respect to a partition ξ and a σ-algebra to the metric entropy of a partition with respect to a σ-algebra.

Definition 5.9. $h(T_0, \xi \mid \mathscr{A}) = \lim_{n\to\infty} \frac{1}{n} H(\xi_0^n \mid \mathscr{A})$.

Theorem 5.10. *For any ergodic measure μ and any T_0 and S_0 invariant algebra A,*

$$h_\mu(T_0, A) = \frac{\log(p)}{\log(q)} h_\mu(S_0, A).$$

Proof. We are given that $h(T_0, P) = h(T_0, A) + h(T_0, P \mid A)$ and that $h(S_0, P) = h(S_0, A) + h(S_0, P \mid A)$. Thus, if we can show that

$$h(T_0, P \mid A) = \frac{\log p}{\log q} h(S_0, P \mid A)$$

then we see

$$\begin{aligned} h(T_0, A) &= h(T_0, P) - h(T_0, P \mid A) \\ &= h(T_0) - \frac{\log(p)}{\log(q)} h(S_0, P \mid A) \\ &= h(T_0) - \frac{\log(p)}{\log(q)} (h(S_0) - h(S_0, A)) \\ &= \log(p) - \frac{\log(p)}{\log(q)} (\log(q) - h(S_0, A)) \\ &= \frac{\log(p)}{\log(q)} h(S_0, A) \end{aligned}$$

and we see that this is all we need to show.

Choose n_i, m_i tending to infinity such that

$$\left| n_i - m_i \frac{\log(p)}{\log(q)} m \right| < \frac{\log(1.1)}{\log(p)}.$$

Thus, we see that

$$0.9 < \frac{p^{n_i}}{q^{m_i}} < 1.1 \text{ and } \lim_{i\to\infty} \frac{m_i}{n_i} = \frac{\log(p)}{\log(q)}.$$

We know that for $f \in \bigvee_{i=0}^n$, $\varphi^{-1}(f) \subset S^1$ is an interval of length $q^{-1}p^{-n-1}$, and for any $g \in \bigvee_{i=0}^m S_0^{-i}(P)$, $\varphi^{-1}(g) \subset S^1$ is an interval of length $p^{-1}q^{-m-1}$. We also know that

$$\frac{q^{-1}p^{-n_i-1}}{p^{-1}q^{-m_i-1}} = \frac{p^{-n_i}}{q^{-m_i}} = \frac{q^{m_i}}{p^{n_i}} < \frac{10}{9}.$$

Thus, any element of $\bigvee_{i=0}^{n_i} T_0^{-i}(P)$ is contained in the union of at most three elements of $\bigvee_{i=0}^{m_i} S_0^{-i}(P)$. Through some calculations, we show that

$$\left| h\left(\bigvee_{i=0}^{n_i} T_0^{-i}(P) \mid A\right) - h\left(\bigvee_{i=0}^{m_i} S_0^{-i}(P) \vee A\right)\right| < 2\log(3)$$

which means that

$$\left| \frac{1}{n_i} h\left(\bigvee_{i=0}^{n_i} T_0^{-i}(P) \mid A\right) - \left(\frac{m_i}{n_i}\right)\frac{1}{m_i} h\left(\bigvee_{i=0}^{m_i} S_0^{-i}(P) \vee A\right)\right| < 2\log(3)$$

converges to 0 with i. However, this limit is also

$$\left| h(T_0, P \mid A) - \frac{\log(p)}{\log(q)} h(S_0, P \mid A)\right|$$

and our proof is complete. □

With this shown, we move on to the final section of the paper, in which we prove the main result.

5.3. **Completion of the result.**

In this section we show that if μ, the T_0 and S_0 invariant measure from the previous section, is not Lebesgue measure, then it is possible to construct an algebra $\mathscr{H}$ such that $h(T_0, \mathscr{H}) = h(T_0)$ and that $h(S_0, \mathscr{H}) = 0$, which will demonstrate that $h(S_0) = h(T_0) = 0$, and the proof will be complete. Throughout this section, we will work with an ergodic, T_0 and S_0 invariant measure μ. We will construct this algebra by constructing collections of distributions and taking the minimal σ-algebra for which they are measurable.

Recall that $\hat{Y}$ is the set of doubly infinite arrays of elements from $\{0, 1, \dots, pq-1\}$ in which rows are M_T-allowed and columns are M_S-allowed, and that $\varphi(\hat{Y})$ is the image of the first quadrant of $\hat{Y}$ in S^1. Let $x = \varphi(\hat{y})$. Then we know there are p points of the form $x_1 + \frac{i}{p} \mod 1$ for $0 \le i \le p-1$ such that $T(\frac{x_1+i}{p}) = x$. Thus, there are p^n points of the form $x_1 + \frac{i}{p^n} \mod 1$ for $0 \le i \le p^n - 1$ such that $T^n(\frac{x_1+i}{p^n}) = x$. For simplicity, we write $\varphi(T_0^{-n}(\hat{y})) = x_1(\hat{y})$. We write the μ conditional expectation that the M_T-allowed name $(i_0, i_1, \dots)$ of $\varphi(\hat{y})$ will extend to $(i_{-n}, i_{-n+1}, \dots, i_0, i_1 \dots)$, the M_T-allowed name of $x_1(\hat{y}) + \frac{t}{p^n}$ as

$$\delta(\hat{y}, t, n) = E\left(T^n\left(\varphi^{-1}\left(x_1(\hat{y}) + \frac{t}{p^n}\right)\right)\middle|\varphi^{-1}(x_1(\hat{y})\right).$$

Using this function, we get a distribution on the points $\left\{0, \frac{1}{p^n}, \cdots, \frac{p^n-1}{p^n}\right\}$ which is given by $\delta(\hat{y}, n)\left(\frac{t}{p^n}\right) = \delta(\hat{y}, t, n)$.

We omit the proof of the following lemma.

Lemma 5.11. *If $\delta(\hat{y_1}, n)$ and $\delta(\hat{y_2}, n)$ agree on the positive horizontal axis, then they differ by a translation by $\varphi(T_0^{-n}(\hat{y_2})) - \varphi(T_0^{-n}(\hat{y_1}))$.*

We now provide a definition of **symmetric** for a point, which will be central to the rest of this paper.

Definition 5.12. A point $\hat{y}$ is **symmetric** if there are points $\hat{y_1} \neq \hat{y_2}$ such that $\varphi(\hat{y_1}) = \varphi(\hat{y_2}) = \varphi(\hat{y})$ and for all $n > 0$ and $m \geq 0$,

$$\delta(T_0^m(\hat{y_1}), n) = \delta(T_0^m(\hat{y_2}), n).$$

The set of symmetric points is both T_0 and S_0 invariant if $\mathrm{GCD}(p, q) = 1$.

We now wish show that if almost every point is symmetric, then μ is Lebesgue measure. In order to do this, we need the following lemma.

Lemma 5.13. *If $\hat{y}$ is a symmetric point, then $\delta(\hat{y}, n)$ converges weakly to Lebesgue measure on S^1.*

Proof. Let $\hat{y_1} \neq \hat{y_2}$ be the points from the definition and $-i_0$ be the first index such that $\hat{y_1}(-i_0, 0) \neq \hat{y_2}(-i_0, 0)$. We know that $\delta(\hat{y}, n)$ is invariant under a shift by $a_n = \varphi(T_0^{-n}(\hat{y_2})) - \varphi(T_0^{-n}(\hat{y_1}))$. We can show that a_n is a fraction which in least terms has a denominator greater than or equal to 2^{-n-i_0+1}. Using this, we can show that the group of shifts preserving $\delta(\hat{y}, n)$ is at least of order 2^{n-i_0+1}, and consequently, its minimal element $d_n \leq \frac{1}{2^{n-i_0+1}}$, which for any continuous f on S^1 with $f(0) = f(1)$, forces

$$\lim_{n\to\infty} \int f \, d(\delta(\hat{y}, n)) = \int f \, d\lambda$$

and the proof is complete. □

This leads us to one of the major results of this section.

Theorem 5.14. *If almost every $\hat{y} \in \hat{Y}$ is symmetric, then μ is Lebesgue measure.*

From here, we need to show that if it is not the case that almost every $\hat{y}$ is symmetric, then T and S have zero entropy.

Theorem 5.15. *If $GCD(p, q) = 1$, μ is T and S invariant, and ergodic, but μ is not Lebesgue measure, then $h(T, P) = h(S, P) = 0$.*

As the material necessary to prove this is out of the scope of this paper, we will summarize the arguments used. We first construct a sigma algebra $\mathscr{H}$ which is the minimal one for which every $\delta(\hat{y}, n)$ is measurable. In other words, every sigma algebra for which each $\delta(\hat{y}, n)$ is measurable contains $\mathscr{H}$. We also construct a sequence of sigma algebras $\mathscr{H}_n$ which converges to $\mathscr{H}$. It is at this point where we make use of the fact that $\gcd(p, q) = 1$, as everything previous to this holds without this fact. Using this, we have that for all $A \in \mathscr{H}_n$ there exists some $m_n \in \mathbb{N}$ such that S^{m_n} is the identity on $\mathscr{H}_n$, and thus, S is periodic on $\mathscr{H}_n$. Using this, we can show that each ergodic component of the dynamical system is a finite rotation. Thus, all ergodic components of the dynamical system $(S, \mathscr{H}, \hat{\mu})$ have rational pure point spectrum, which roughly means that we are dealing with atomic measures, and consequently, $h(S, \mathscr{H}) = 0$. We then show

$$h(S, \mathscr{H}) = h(S, P) = h(T, P) = 0$$

and thus, obtain our result.

Acknowledgements. I would like to thank my mentor, Rachel Vishnepolsky for all of her help in the writing and revising of this paper, as well as for introducing me to such an interesting branch of mathematics, as well as Peter May for organizing this REU.

References

[1] Brin, Michael and Garrett Stuck. *Introduction to Dynamical Systems.* Cambridge University Press 2002.

[2] Furstenberg, Hillel. Disjointness in ergodic theory, minimal sets, and a problem in diophantine approximation. *Math. Sys. Theory 1* (1967), 1-49.

[3] Rudolph, Daniel J. 2 and 3 invariant measures and entropy. *Ergodic Theory Dynam. Systems*, 10(2):395-406, 1990.

[4] Silva, Cesar E. *Invitation to Ergodic Theory.* American Mathematical Society. 2008.