AN ABSTRACT OF THE DISSERTATION OF

David Josiah Wing for the degree of Doctor of Philosophy in Mathematics presented on March 31, 2011.

Title: Notions of Complexity in Substitution Dynamical Systems

Abstract approved: ______________________________

Robert M. Burton

There has been a lot of work done in recent decades in the field of symbolic dynamics. Much attention has been paid to the so-called "complexity" function, which gives a sense of the rate at which the number of words in the system grow. In this paper, we explore this and several notions of complexity of specific symbolic dynamical systems. In particular, we compute positive entropy and state some *k-balancedness* properties of a few specific (random) substitutions. We also view certain sequences as subsets of $\mathbf{Z}^2$, stating several properties and computing bounds on entropy in a specific example.

Notions of Complexity in Substitution Dynamical Systems

by

David Josiah Wing

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented March 31, 2011
Commencement June 2011

Doctor of Philosophy dissertation of David Josiah Wing presented on March 31, 2011

APPROVED:

Major Professor, representing Mathematics

Chair of the Department of Mathematics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

David Josiah Wing, Author

ACKNOWLEDGEMENTS

Academic

I first would like to thank my advisor, Bob Burton, for his encouragement, enthusiasm, and guidance throughout my research studies. I would also like to thank Dennis Garity for his generous patience and guidance. There are of course many other professors, instructors, students, and staff I have worked with here at Oregon State University and Portland State University who have contributed significantly to my education. Without them my advancement to this point would not have been possible. I am deeply grateful to them all.

Personal

I wish to sincerely thank my parents, Larry and Teressa Wing, and friends Gil Hartl, Jacques Rossignol, Dan Rockwell, and Jason Schmurr for their support and encouragement.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

NOTIONS OF COMPLEXITY IN SUBSTITUTION DYNAMICAL SYSTEMS

1 INTRODUCTION

The concept of a symbolic dynamical system (that is, a set of infinite strings of symbols under a certain continuous map, called the *left shift map*) has received much study in the last several decades. Originally, it was realized that a partition of a space could be coded under a transformation via the symbols that represent the partition elements, called *atoms*. Iterating the transformation moves points between atoms, thus yielding infinite strings of symbols. It became of great interest as to the connections (if any) of the space of these strings to the general space at hand. Attention then shifted from studying properties of a space to studying these spaces of infinite strings, of which it turned out there are a great many questions that can be asked.

In this paper, we mainly explore a certain subclass of symbolic dynamical systems called *substitution dynamical systems*. Substitutions provide a natural way to generate sequences and the substitutive map that generates them aids in their study. We will explore and further the results on several notions of complexity in these systems, called *entropy*, *k-balancedness*, and *recoverability*. In particular, we compute positive entropy (Section 4.2) and state some k-balancedness properties of a few specific (random) substitutions (Section 5.1). We also view certain sequences as subsets of $\mathbf{Z}^2$, stating several properties and computing bounds on entropy in a specific example (Chapter 6).

The paper is organized as follows. We first discuss the notion of a symbolic dynamical system in general and review some basic properties it can have. We then focus on substitutions in Chapter 3 and define the so called *FibMorse* substitution (which originated this study). After these general concepts, we discuss the idea of randomness in Chapter 4 and capture a way to quantify it (entropy). We attempt to give a helpful heuristical view, after which we apply these ideas to compute this quantity for several examples of substitution dynamical systems. We then study two different notions of how complex a symbolic dynamical system can be in terms of k-balancedness (Section 5.1), which tells us about how "uniform" a sequence is and recoverability (Section 5.2), which tells us how "well we know it". We then look briefly at how and if these notions are connected. Finally in the last two chapters, we look at how a geometric view can study complexity of sequences in general and to begin to see how balancedness looks in two dimensions. We thank J. Rossignol and D. Rockwell for many helpful conversations.

2 SYMBOLIC DYNAMICAL SYSTEMS

The study of spaces of infinite sequences of symbols, "symbolic dynamics", was originally introduced as a means of studying more general dynamical systems. The techniques involved eventually were applied to fields such as information theory, data compression, cryptology, coding, and number theory. In this chapter, we define a symbolic dynamical system and review some standard properties one may examine for such systems.

Definition 2.0.1 *A* ***dynamical system*** (X, T) *is a space* X *together with a transformation* $T : X \to X$. *If* X *is a topological space and* T *is continuous, then* (X, T) *is a* ***topological dynamical system.***

Definition 2.0.2 *Suppose* (X, Σ, μ) *is a measure space and* T *is* Σ*-measurable. Let* $\mathcal{P}$ *be a finite collection of disjoint measurable sets of* X *of positive measure that cover* X *up to a set of measure zero. Then* $(X, T) = (X, \Sigma, \mu, T)$ *is a* ***measure-theoretical dynamical system****.* $\mathcal{P} = \{P_1, \ldots P_n\}$ *is a* ***partition*** *of* X *if* X *is a disjoint union of the* P_i*. We then call* $(X, T) = (X, \Sigma, \mathcal{P}, \mu, T)$ *a* ***process****.*

Remark 2.0.1 *In ergodic theory,* T *is often* ***measure-preserving****, meaning that if* $A \in \Sigma$*, then* $T^{-1}(A) \in \Sigma$ *and* $\mu\left(T^{-1}(A)\right) = \mu(A)$.

Standing Assumption: All of our measure spaces (X, μ) will be probability spaces (that is, $\mu(X) = 1$).

Throughout, denote by $\mathbf{Z}$ and $\mathbf{Z}_+$ the set of integers and the set of non-negative integers, respectively. Let $\mathcal{A} = \{0, 1, \ldots, s-1\}$ be an **alphabet**. Elements of $\mathcal{A}$ are called **symbols** or **letters** and elements of $\mathcal{A}^k = \prod_{i=1}^{k} \mathcal{A}$ **words of length** k. Denote by

$\mathcal{A}^* = \cup_{k\geq 1}\mathcal{A}^k$ the set of all words on $\mathcal{A}$. For $i \in \mathcal{A}$ and $w \in \mathcal{A}^*$, let $|w|_i$ be the number of times i occurs in w. Let

$$\mathcal{A}^{\mathbf{Z}_+} = \prod_{j=0}^{\infty} \mathcal{A} = \{x = (x_0, x_1, x_2, \cdots) : x_j \in \mathcal{A} \text{ for all } j \geq 0\}$$

be the set of all infinite-tuples on $\mathcal{A}$. We will call elements of $\mathcal{A}^{\mathbf{Z}_+}$ **infinite strings** (or **strings**) and denote $x \in \mathcal{A}^{\mathbf{Z}_+}$ by $x_0x_1x_2\ldots$. A **subword** (or **word**) **of length** n of $x = x_0x_1x_2\ldots \in \mathcal{A}^{\mathbf{Z}_+}$ is a finite sequence of the form $w = x_ix_{i+1}\ldots x_{i+n-1}$. We also write $w = x_{[i,i+n-1]}$. In this case, we say that x **contains** w and denote this by $w \in x$. In this case we denote the length of w by $|w| = n$. Similarly if x is only finite. If $X \subseteq \mathcal{A}^{\mathbf{Z}_+}$, then w is a **word in** X if w is contained in some infinite string in X. The **empty word** ε is the word having no symbols. The **language** of X is the set of all words in X.

Put the discrete topology on $\mathcal{A}$ and the product topology on $\mathcal{A}^{\mathbf{Z}_+}$. We define a metric d_S on $\mathcal{A}^{\mathbf{Z}_+}$ given by

$$d_S(x,y) = \begin{cases} 2^{-N} & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases},$$

where $N = \min\{k \geq 0 : x_k \neq y_k\}$. We will call d_S the **sequence metric**.

Remark 2.0.2 *Often, the set of bi-infinite sequences on $\mathcal{A}$, defined by $\mathcal{A}^{\mathbf{Z}} = \prod_{j=-\infty}^{\infty} \mathcal{A} = \{x = \ldots x_{-2}x_{-1}x_0x_1x_2\cdots : x_j \in \mathcal{A} \text{ for all } j \in \mathbf{Z}\}$ is used instead of $\mathcal{A}^{\mathbf{Z}_+}$. Most all of the work here can be naturally extended to $\mathcal{A}^{\mathbf{Z}}$.*

Lemma 2.0.1 *d_S metrizes the product topology on $\mathcal{A}^{\mathbf{Z}_+}$.*

Proof. Let $x \in \mathcal{A}^{\mathbf{N}}$ and $B_{2^{-N}}(x) := \left\{y \in \mathcal{A}^{\mathbf{N}} : d_S(x,y) < 2^{-N}\right\}$ a d_S-open neighborhood of x. Let

$$U = \prod_{k=0}^{N-1} B_{\frac{1}{2}}(x) \times \prod_{k=N}^{\infty} \mathcal{A}.$$

Then U is open in the product topology and $x \in U \subseteq B_{2^{-N}}(x)$. Indeed, if $y \in U$ then $d(x_k, y_k) < \frac{1}{2}$ for all $0 \leq k \leq N-1$, where d is the discrete metric on $\mathcal{A}$. But then $x_k = y_k$ for all $0 \leq k \leq N-1$ so that $y \in B_{2^{-N}}(x)$.

Conversely, suppose that $U = \prod_{k=0}^{\infty} U_k$ is a neighborhood of x, open in the product topology so that $U_k \neq \mathcal{A}$ for at most finitely many k. Then $N = \max\{k : U_k \neq \mathcal{A}\}$ exists and is finite. We then have that $x \in B_{2^{-N}}(x) \subseteq U$ since $y \in B_{2^{-N}}(x)$ implies $x_k = y_k$ for all $0 \leq k \leq N-1$ and this means that $d(x_k, y_k) = 0$ for all $0 \leq k \leq N-1$ so that $y \in U$ (by the definition of N). ■

Definition 2.0.3 *Define a transformation $\sigma : \mathcal{A}^{\mathbf{Z}_+} \to \mathcal{A}^{\mathbf{Z}_+}$, called the (left)* ***shift*** *on $\mathcal{A}^{\mathbf{Z}_+}$, by*

$$(\sigma(x))_n = x_{n+1}.$$

If X is any closed, shift-invariant ($\sigma(X) \subseteq X$) subset of $\mathcal{A}^{\mathbf{Z}_+}$, then $(X, \sigma|_X)$ is called a ***symbolic dynamical system*** *with* ***alphabet*** *$\mathcal{A}$. If $X = \mathcal{A}^{\mathbf{Z}_+}$, then (X, σ) is called a* ***full shift***. *If (X, σ) is a symbolic dynamical system, then any word in X is called* ***admissible*** *and any word not in X is called* ***forbidden***.

It will follow from topological properties of "cylinder sets" (Definition 2.0.4) that any symbolic dynamical system has at most countably many forbidden words.

Let u be an element of $\mathcal{A}^{\mathbf{Z}_+}$ or $\mathcal{A}^{\mathbf{Z}}$. The orbit closure of u under the shift map $\overline{Orb_\sigma(u)}$ (with respect to the sequence metric d_S) is a closed shift-invariant subset of the full-shift, so is a symbolic dynamical system, called the **dynamical system arising from** u. A more general construction of a symbolic dynamical system is the following.

Lemma 2.0.2 *Let $\{w_n\}$ be a fixed sequence of words in $\mathcal{A}$ (that is, elements of $\mathcal{A}^*$) ordered by increasing length such that*

$$\limsup_{n\to\infty} |w_n| = \infty.$$

Let X be the set of all infinite sequences $x \in \mathcal{A}^{\mathbf{Z}_+}$ such that every finite subword of x is contained in infinitely many w_n's. Then $(X, \sigma|_X)$ is a symbolic dynamical system.

Proof. By hypothesis, the lengths of the w_n's increase without bound so that X is well-defined (that is, given ℓ, there is a word of length ℓ in some w_n). We show that X is closed. So let $z \in \mathcal{A}^{\mathbf{Z}_+}$ be a limit point of X. Then there is a sequence $\left\{z^{(m)}\right\}_{m=1}^{\infty}$ in X such that $z^{(m)} \to z$ as $m \to \infty$. So, for every $p \geq 0$, there is some M such that $\rho\left(z^{(m)}, z\right) < 2^{-p+1}$ whenever $m \geq M$ so that $z_k^{(m)} = z_k$ for all $0 \leq k \leq p-1$. Now, let w be a finite subword of z, say $w = z_{[a,a+|w|-1]}$. Then for every $m \geq \max\{M, a+|w|-1\}$, $z_k^{(m)} = z_k$ for every $0 \leq k \leq \max\{M, a+|w|-1\} - 1$. That is, w is contained in $z^{(m)}$ for every $m \geq \max\{M, a+|w|-1\}$. Since $z^{(m)} \in X$ for all m, we thusly have that w is contained in infinitely many of the w_n's. Hence, X is closed. Now shift invariance is clear: Any $x \in X$ has the property that any finite subword of x is contained in infinitely many of the w_n's. So, by definition of $\sigma|_X$, any finite subword of $\sigma|_X(x)$ is contained in infinitely many of the w_n's so that $\sigma|_X(x) \in X$. ■

We will often write $\sigma = (X, \sigma|_X)$ when unambiguous.

Definition 2.0.4 *Let (X, σ) be a symbolic dynamical system and $w = w_0 w_1 \cdots w_{n-1}$ an admissible word of length n. The set*

$$[w]_i = \{x \in X : x_i x_{i+1} \cdots x_{i+n-1} = w_0 w_1 \cdots w_{n-1}\}$$

*is called the **cylinder set** of w at position i. The sets $[a]_0$, where $a \in \mathcal{A}$, are called the **time-zero sets**.*

Useful basic topological properties are the following (see, for example, [13], Chapter 6, for details). Cylinder sets are useful because they form a basis for the topology of a symbolic dynamical system, which gives us that the shift-map σ is continuous. We also have that (X, σ) is compact (which follows from the Tychonoff Theorem since $\mathcal{A}$ is finite). Since a symbolic dynamical system X is closed, its complement is open and so is a countable union of cylinder sets since these form a basis for the topology. Thus, X has at most countably many forbidden words.

It is a desirable property of a space to want it to be "indecomposable" in some sense, or to be able to break it up in to "indecomposable pieces". For example, if (X, T, μ) is a measure-theoretic dynamical system, we might want the transformation T to move points all throughout the space X and not miss any part of it. This leads to the following ideas. A set E in a dynamical system (X, T) is T-**invariant** if $T(E) = E$. If μ is a measure on X, then E is called μ-**invariant** if $\mu\left(T^{-1}E\right) = \mu(E)$. A (measure-theoretic) dynamical system (X, μ, T) is **ergodic** if the only T-invariant sets are metrically trivial. That is, if E is μ-invariant, then $\mu(E) \in \{0, 1\}$ (recall that we are assuming (X) to be a probability space). In this case, we also say that the transformation T is ergodic. A dynamical system (or transformation) is **uniquely ergodic** if it has only one ergodic measure. A dynamical system (X, T) is **minimal** if the only closed T-invariant subsets are $\varnothing$ and X.

We note that if (X, μ, T) is uniquely ergodic, then μ is an ergodic measure. Indeed, suppose not. Then there is a T-invariant set E_0 with $0 < \mu(E_0) < 1$ and so we can define measures μ_1 and μ_2 on X by

$$\mu_1(B) := \frac{\mu(E_0 \cap B)}{\mu(E_0)}$$

and

$$\mu_2(B) := \frac{\mu((X \backslash E_0) \cap B)}{1 - \mu(E_0)}.$$

Then, since $T^{-1}(E_0) = E_0$, μ_1 and μ_2 are both T-invariant probability measures on X, at least one of which is distinct from μ. This contradicts the unicity of μ.

Lemma 2.0.3 *A dynamical system (X, T) is minimal if and only if every orbit is dense.*

Proof. *Suppose (X, T) is minimal and $x \in X$. $\overline{Orb(x)}$ is T-invariant by the continuity of T since $y \in \overline{Orb(x)} \implies y = \lim_{k \to \infty} T^{n_k}(x)$ for some sequence $\{n_k\}_{k=1}^{\infty}$ so that $T(y) \in \overline{Orb(x)}$. Therefore, $\overline{Orb(x)} = X$ by minimality. Conversely, let $E \subseteq X$ be* *a closed and nonempty T-invariant set and let $x \in X$. Since every orbit is dense, given $e \in E$, there is a sequence $\{n_k\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} T^{n_k}(e) = x$. Then, by invariance and continuity of T, $\{T^{n_k}(e)\}$ is a sequence in E converging to x. Since E is closed, this means that $x \in E$. Therefore, $X = E$.* $\blacksquare$

Example 2.0.1 *The map $x \mapsto 2x$ (mod 1) on the unit interval is ergodic, by definition. If α is irrational, then the map $x \mapsto x + \alpha$ (mod 1) on the unit interval is minimal (orbits are dense since α is irrational) and so is uniquely ergodic. The unique ergodic measure is Lebesgue measure (see, for example, [25], Theorem 6.20).*

Other basic properties orbits of points can have are as follows. A point x in a dynamical system (X, ρ, T) is **periodic** if there is an N such that $T^N(x) = x$ and **eventually periodic** if there is an N such that $T^{N+k}(x) = T^N(x)$ for all $k \geq 0$. Otherwise, x is called **aperiodic**. x is **almost periodic** if, given $\varepsilon > 0$, there is some $N = N(\varepsilon)$ such that the set $A = \{n \geq 1 : \rho(T^N(x), x) < \varepsilon\}$ has gaps of size at most N (that is, $n, m \in A \implies |n - m| \leq N$). In this case, we say that x returns arbitrarily close to its initial position with **bounded gaps**. If the gaps are not necessarily bounded, that is, if there is an increasing sequence $\{n_k\}_{k=0}^{\infty}$ such that $T^{n_k}(x) = x$, then x is **recurrent**.

Thus, if x is almost periodic, then it comes arbitrarily close to its initial position under iterations of T. Note that for a sequence u in $\mathcal{A}^{\mathbf{Z}_+}$ or $\mathcal{A}^{\mathbf{Z}}$, we have that u is almost periodic if every word in u occurs in u with bounded gaps: Given $B \in u$, there is a sequence $\{n_k\}_{k=1}^{\infty}$ such that $\sigma^{n_k}(u)_{[0,|B|-1]} = B$ for all $k \geq 1$. Note that almost periodicity is stronger than minimality in that the gaps in a minimal sequence may not be bounded. However, see Theorem 3.0.1, below.

Definition 2.0.5 *A function f on a dynamical system (X, T) is T-**invariant** if $f(Tx) = f(x)$ for almost all $x \in X$.*

Lemma 2.0.4 *$X = (X, \mu, T)$ is ergodic if and only if any T-invariant (complex-valued) function on X is constant almost everywhere.*

Proof. Suppose any T-invariant function is constant and let E be a T-invariant set, so that $T^{-1}(E) = E$. If χ_E is the characteristic function on E, then

$$\begin{aligned} \chi_E(Tx) &= \begin{cases} 0 & \text{if} \quad Tx \notin E \\ 1 & \text{if} \quad Tx \in E \end{cases} \\ &= \begin{cases} 0 & \text{if} \quad x \notin T^{-1}(E) = E \\ 1 & \text{if} \quad x \in T^{-1}(E) = E \end{cases} \\ &= \chi_E(x). \end{aligned}$$

Therefore, χ_E is T-invariant so that χ_E is constant on X. So either $\chi_E(x) = 0$ for all $x \in X$ (in which case $\mu E = 0$) or $\chi_E(x) = 1$ for all $x \in X$ (in which case $\mu E = \mu X = 1$). This shows that T is ergodic.

Conversely, suppose that T is ergodic and f invariant. Let $E_y = \{x \in X : f(x) > y\}$. Since f is invariant, so is E_y for every y and therefore has measure 0 or 1 by ergodicity. Now, if f is not constant almost everywhere, there is some E_y such that $0 < E_y < 1$, a contradiction. ■

Lemma 2.0.5 *If (X, μ, T) is minimal, then it is ergodic.*

Proof. Let $f : X \longrightarrow \mathbf{C}$ be a continuous T-invariant function. Then by invariance, f is constant on the orbit of any point in X. By minimality, every orbit is dense (Lemma [2.0.3]). Hence, by continuity of f, $f(x) = f(y)$ for every $x, y \in X$, which establishes ergodicity by Lemma 2.0.4 . ■

Other basic questions about dynamical systems concern the following definitions:

Definition 2.0.6 *A dynamical system* (X, μ, T) *is* ***weakly mixing*** *if, for all measurable* $A, B \subseteq X,$ *we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left|\mu\left(T^{-j}(A) \cap B\right) - \mu(A)\mu(B)\right| = 0.$$

Definition 2.0.7 *A dynamical system* (X, μ, T) *is* ***strongly mixing*** *if, for all measurable* $A, B \subseteq X,$ *we have*$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B)$.

Ergodicity and mixing tells us how "mixed up" a space is. We might instead be interested in how "complicated" a space is. A nice survey of the following so-called complexity function can be found in [8].

Definition 2.0.8 *The* ***complexity*** *of a symbolic dynamical system* (X, σ) *is the function* $p : \mathbf{Z}_+ \to \mathbf{Z}_+$ *where* $p(n)$ *is the number of distinct words of length* n *in* X.

The idea is that the more distinct words of length n there are, the more "complex" the system should be. Note that if $u_i u_{i+1} \dots u_{i+n-1}$ is a word of length n occurring in u, then $u_i u_{i+1} \dots u_{i+n-1} u_{i+n}$ is a word of length $n+1$ occurring in u. So p is a non-decreasing function.

Example 2.0.2 *Let* u *be the constant sequence* $\bar{0} = 00000\dots.$ *Then* $p(n) = 1$ *for all* n.

Example 2.0.3 *Let* u *be the concatenation of the positive integers, where each integer is written in base 2. That is,*

$$u = 0.1.10.11.100.101.110.111\dots.$$

u *is called the* ***Champernowne sequence****. Note that, for every* n, u *contains every word (on* $\{0, 1\}$*) of length* n. *So* $p(n) = 2^n$.

In terms of sequences, an element u of $\mathcal{A}^{\mathbf{Z}_+}$ is periodic if there is some d such that $u_{k+d} = u_k$ for all k and u is eventually periodic if there is some N and some d such that $u_{k+d} = u_k$ for all $k \geq N$. In either case, d is called the **period** of u.

The following is a basic fact about periodic sequences. We refer the reader to [14], chapter 1, for further details.

Proposition 2.0.1 *Let u be a sequence on a finite alphabet $\mathcal{A}$. If u is eventually periodic, then $p(n)$ is bounded and there is some n such that $p(n) \leq n$. Conversely, if there is some n such that $p(n) \leq n$, then u is eventually periodic.*

Proof. Suppose u is eventually periodic. Then there is some $N \geq 0$ and some period $d \geq 1$ such that $u_{k+d} = u_k$ for all $k \geq N$. We may assume that N and d are the smallest such. Then, for $k \geq N$, the words

$$\begin{gathered} u_k u_{k+1} \dots u_{k+d-1}, \\ u_{k+1} u_{k+2} \dots u_{k+d}, \\ \vdots \\ u_{k+d-1} u_{k+d} \dots u_{k+2d-1} \end{gathered}$$

are all distinct words of length d and any other word of length d occurring in $u_N u_{N+1} \cdots$ equals one of these d words. There can be at most $N-1$ words of length d in $u_0 u_1 \cdots u_{N+d-1}$. Therefore, $p(d) \leq N + d - 1$. Then, again by periodicity, $p(N + d - 1) \leq N + d - 1$. Conversely, suppose $p(n) \leq n$ for some n. Note that $p(1) \geq 2$ or else u is constant (and hence periodic). Then there is some k_0 such that $p(k_0 + 1) = p(k_0)$. For if not, then $p(k+1) > p(k)$ for all k (since p is increasing) so that $p(k+1) - p(k) \geq 1$ for all k. Then for any n,

$$\begin{aligned} p(n) &= p(1) + \sum_{k=1}^{n-1} p(k+1) - p(k) \\ &\geq 2 + (n-1) \\ &= n + 1, \end{aligned}$$

contradicting our hypothesis. Now, since there are finitely many words of length k_0 that occur in u, given i, there must exist some $j > i$ such that $u_i u_{i+1} \dots u_{i+k_0-1} = u_j u_{j+1} \dots u_{j+k_0-1}$. But for any word B of length k_0 occurring in u, Ba must also occur in u for some $a \in \mathcal{A}$. But since $p(k_0+1) = p(k_0)$, there is only one such word Ba. Thus, $u_{i+k_0} = u_{j+k_0}$. But then $u_{i+1}u_{i+2} \dots u_{i+k_0} = u_{j+1}u_{j+2} \dots u_{j+k_0}$ is a word of length k_0 and the same argument applies to show that $u_{i+k_0+1} = u_{j+k_0+1}$. Continuing in this way, we have that $u_{i+k} = u_{j+k}$ for every $k \geq 0$. Therefore, u is periodic with period $j - i$. ■

Define the **height** $h(w)$ of a word $w \in \{0,1\}^*$ to be the number of 1's occurring in w (that is, $h(w) = \sum_{k=0}^{n-1} w_k$) and the ratio $\frac{h(w)}{|w|}$ is called the **slope** of w. A sequence $u \in \{0,1\}^{\mathbf{Z}_+}$ is **balanced** if for any integer $n \geq 1$ and any two words w_1 and w_2 of length n, we have $|h(w_1) - h(w_2)| \leq 1$. This balancedness enables a bound on the complexity. We will see later in 4.3.2 that the aperiodic balanced sequences are the ones with minimal complexity.

It will be useful and interesting to talk about the slope of an infinite string.

Definition 2.0.9 *If the limit*

$$\mu[1] = \lim_{n\to\infty} \frac{h(u_0 u_1 \cdots u_{n-1})}{n}$$

*exists, then $\mu[1]$ is the **frequency of** 1 in u. We also call $\mu[1]$ the **slope** of u.*

Remark 2.0.3 *$h(u_0 u_1 \cdots u_{n-1}) \leq n$ for any n and the sequence $\{h(u_0 u_1 \cdots u_{n-1})\}_{n=1}^{\infty}$ of integers is non-decreasing.*

Note that the slopes of sequences do not always exist. For example, if $u \in \{0,1\}^{\mathbf{Z}_+}$ is such that

$$a_n := h\left(u_0 \dots u_{4^{2^{n-1}}-1}\right) = 3^{\frac{1}{2}((-1)^n+1)} \cdot 4^{2^{n-1}-1}$$

for $n \geq 1$ (such a u exists since a_n is increasing and $a_n < 4^{2^{n-1}}$ for all n so that the height is always less than the length), then

$$\frac{h\left(u_0 \ldots u_{4^{2^{n-1}}-1}\right)}{4^{2^{n-1}}} = 3^{\frac{1}{2}((-1)^n+1)} \cdot \frac{4^{2^{n-1}-1}}{4^{2^{n-1}}}$$

is the sequence $\frac{1}{4}, \frac{3}{4}, \frac{1}{4}, \frac{3}{4}, \ldots$. Thus, the limit $\lim_{n\to\infty} \frac{h(u_0 \ldots u_{n-1})}{n}$ does not exist.

Example 2.0.4 *It is known that the frequency of* 1 *in the Champernowne sequence is* $\frac{1}{2}$. *This shows that the slope* $\mu[1]$ *does not necessarily have anything to do with the complexity function* $p(n)$ *since the Champernowne sequence has complexity* $p(n) = 2^n$.

3 SUBSTITUTIONS

A lot of phenomena in nature are self-similar. That is, if we look close enough in a certain object, we can sometimes see copies of that object inside itself. Such is the case with fractals, for example. This phenomena also occurs in many crystalline structures. In studying such structures, it is useful to be able to model such self-similarities. It turns out that certain functions called "substitutions" allow us to model self-similar behavior study the object's properties using dynamical systems. We refer the reader to [21] for further properties and details. Define **word concatenation** of two words $w = w_1 w_2 \ldots w_n$ and $v = v_1 v_2 \ldots v_m$ as

$$wv = w_1 w_2 \ldots w_n v = v_1 v_2 \ldots v_m.$$

Let $s \geq 2$ and let $\mathcal{A} = \{0, \ldots, s-1\}$ be a finite alphabet on s symbols. Any function $\omega : \mathcal{A} \to \mathcal{A}^*$ is called a **substitution** on $\mathcal{A}$. If there is an $\ell \in \mathbf{N}$ such that $|\omega(a)| = \ell$ for all $a \in \mathcal{A}$, then ω is said to be of **constant length**. We extend ω to a map $\omega' : \mathcal{A}^* \to \mathcal{A}^*$ defined by

$$\omega'(w_0 w_1 \ldots w_{k-1}) := \omega(w_0)\,\omega(w_1) \ldots \omega(w_{k-1}),$$

which also extends to a map $\omega'' : \mathcal{A}^{\mathbf{Z}_+} \to \mathcal{A}^{\mathbf{Z}_+}$ defined by

$$\omega''(w_0 w_1 w_2 \ldots) := \omega(w_0)\,\omega(w_1)\,\omega(w_2) \ldots.$$

We shall abuse notation and write $\omega = \omega' = \omega''$. For every $n \geq 2$, we define $\omega^n(a) := \omega\left(\omega^{n-1}(a)\right)$, where $a \in \mathcal{A}$ and a map $\omega^n : \mathcal{A}^* \longrightarrow \mathcal{A}^*$ by

$$\omega^n(w_0 w_1 \ldots w_{k-1}) := \omega^n(w_0)\,\omega^n(w_1) \ldots \omega^n(w_{k-1}).$$

This again extends to a map $\omega^n : \mathcal{A}^{\mathbf{Z}_+} \to \mathcal{A}^{\mathbf{Z}_+}$ defined by

$$\omega^n(w_0 w_1 w_2 \ldots) := \omega^n(w_0)\,\omega^n(w_1)\,\omega^n(w_2) \ldots.$$

We define substitutions similarly on $\mathcal{A}^{\mathbf{Z}}$.

Remark 3.0.4 *Although the elements of the alphabet $\mathcal{A}$ formally consists of "symbols", we shall interpret these elements to be integers, when needed.*

Given $a \in \mathcal{A}$, if $\lim_{n\to\infty} |\omega^n(a)| = \infty$ for every $a \in \mathcal{A}$, there is an infinite sequence u on $\mathcal{A}$ such that $u_{[0,|\omega^n(a)|-1]} = \omega^n(a)$ for all $n \geq 1$. u is called a **sequence generated by** ω, and is denoted by $\omega^\infty(a)$. Note that $\omega^\infty(a)$ begins with $\omega^n(a)$ for every $n \in \mathbf{Z}_+$. The convention is to denote this a by 0. In this case, we also have $\omega^\infty(0) = \lim_{n\to\infty} \omega^n(0)$, where convergence is with respect to the metric d_S above. Any word in $\omega^\infty(a)$ is called ω-**admissible**. Note that a substitution may not give rise to an infinite sequence: For example, $\omega : \begin{cases} 0 \longrightarrow 1 \\ 1 \longrightarrow 0 \end{cases}$.

Standing Hypothesis: Unless otherwise stated, we shall henceforth assume that $\lim_{n\to\infty} |\omega^n(a)| = \infty$ for every $a \in \mathcal{A}$ and that such a symbol 0 exists.

Remark 3.0.5 *If ω is irreducible and there is some a such that $|\omega(a)| \geq 2$, then the sequence generated by ω is infinite and defined.*

Definition 3.0.10 *A substitution $\omega : \mathcal{A} \to \mathcal{A}^*$ is **irreducible** if, given $a, b \in \mathcal{A}$, there is some $N = N(a,b) \geq 0$ such that $\omega^N(a)$ contains b. ω is **primitive** if there is some $N \geq 0$ such that for every $a, b \in \mathcal{A}$, $\omega^N(a)$ contains b.*

Definition 3.0.11 *Let $\omega : \mathcal{A} \to \mathcal{A}^*$ be a substitution. Define the **incidence matrix** $M(\omega) = \{m_{ij}\}_{i,j=1}^{|\mathcal{A}|}$ of ω to be the $|\mathcal{A}| \times |\mathcal{A}|$ matrix such that m_{ij} is the number of times the symbol i occurs in $\omega(j)$.*

Remark 3.0.6 *Note that the ij-entry of M^n is the number of times the symbol i occurs in $\omega^n(j)$.*

Lemma 3.0.6 *ω is irreducible (primitive) if and only if its incidence matrix is irreducible (primitive).*

Proof. Let $M = [m]_{ij}$ be the incidence matrix of ω. Suppose ω is irreducible. Then for each $i, j \in \{0, \ldots s-1\}$, there is some N such that $\omega^N(j)$ contains i. This means by definition of M that $\left([m]_{ij}\right) > 0$. Therefore, M is irreducible. Conversely, if there is some N such that $\left([m]_{ij}\right)^N > 0$, then $\omega^N(j)$ contains i. Therefore, ω is irreducible. The same argument applies in the primitive case. ■

Example 3.0.5 *Let $\mathcal{A} = \{0, 1\}$ and*

$$\omega_1 : \begin{cases} 0 \longrightarrow 01 \\ 1 \longrightarrow 10 \end{cases}.$$

Then ω is primitive,

$$\omega_1^\infty(0) = 0110100110010110\ldots,$$

and

$$M(\omega_1) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

*ω_1 is called the **Morse substitution** and $\omega_1^\infty(0)$ the **Morse sequence**.*

Example 3.0.6 *Let $\mathcal{A} = \{0, 1\}$ and*

$$\omega_2 : \begin{cases} 0 \longrightarrow 01 \\ 1 \longrightarrow 0 \end{cases}.$$

Then ω is primitive,

$$\omega_2^\infty(0) = 0100101001001010010 1\ldots,$$

and

$$M(\omega_2) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

*ω_2 is called the **Fibonacci substitution** and $\omega_2^\infty(0)$ the **Fibonacci sequence**.*

Example 3.0.7 *Let* $\mathcal{A} = \{0, 1, 2\}$ *and*

$$\omega_3 : \begin{cases} 0 \longrightarrow 01 \\ 1 \longrightarrow 11 \\ 2 \longrightarrow 22 \end{cases}.$$

Then ω_3 *is not irreducible,*

$$\omega_3^{\infty}(0) = 011111111111\ldots,$$

and

$$M(\omega_3) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Example 3.0.8 *Let* $\mathcal{A} = \{0, 1\}$ *and*

$$\omega_4 : \begin{cases} 0 \longrightarrow 0010 \\ 1 \longrightarrow 1 \end{cases}.$$

Then ω_4 *is not primitive,*

$$\omega_4^{\infty}(0) = 0010.0010.1.0010.0010.0010.1.0010.1.0010.0010\ldots,$$

and

$$M(\omega_4) = \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix}.$$

ω_4 *is called the* ***Chacon substitution*** *and* $\omega_4^{\infty}(0)$ *the* ***Chacon sequence****.*

It is interesting to consider the possibility of a substitution mapping one symbol to two (or more) different words, according to some probability distribution. We shall call such things "random substitutions". Suppose a word w occurs in a set with probability p. Then we will denote this probability by $w(p)$.

Definition 3.0.12 *Let $\mathcal{A} = \{a_0, \ldots, a_{s-1}\}$. Define a discrete random map $\rho : \mathcal{A} \to \mathcal{A}^* = \cup_{n\geq 1}\mathcal{A}^n$ by*

$$\rho : \begin{cases} a_0 & \mapsto & \begin{cases} w_0^1\left(p_0^1\right) \\ \vdots \\ w_0^{i_0}\left(p_0^{i_0}\right) \end{cases} \\ a_1 & \mapsto & \begin{cases} w_1^1\left(p_1^1\right) \\ \vdots \\ w_1^{i_1}\left(p_1^{i_1}\right) \end{cases} \\ \vdots & & \vdots \\ a_{s-1} & \mapsto & \begin{cases} w_{s-1}^1\left(p_{s-1}^1\right) \\ \vdots \\ w_{s-1}^{i_{s-1}}\left(p_{s-1}^{i_{s-1}}\right) \end{cases} \end{cases},$$

where each $w_j^i \in \mathcal{A}^*$, $0 \leq p_j^i \leq 1$, and $\sum_{k=1}^{i_j} p_j^k = 1$ for each $j = 0, \ldots, s-1$. That is, the probability distribution of the random variable $\rho(a_j)$ is $\left(p_j^1, \ldots, p_j^{i_j}\right)$. We call such a random map ρ a ***random substitution on $\mathcal{A}$.***

For $a_j \in \mathcal{A}$, define the **image of a_j under** ρ by $Im_\rho\left(a_j\right) := \left\{w_j^i : 1 \leq i \leq i_j\right\}$ and the **image of** ρ by $\rho\left(\mathcal{A}\right) := \left\{Im_\rho\left(a_j\right) : 0 \leq j \leq s-1\right\}$. We wish to "generate" infinite strings via random substitutions as before. For a word $v = v_1 \ldots v_k \in \mathcal{A}^*$, define

$$\rho\left(v\right) := \rho\left(v_1\right)\rho\left(v_2\right)\cdots\rho\left(v_k\right)$$

and denote by ρ^n the composition of n independent copies of ρ (that is, for each $n \geq 1$, inductively define $\rho^{n+1}\left(v\right) = \rho\left(\rho^n\left(v\right)\right)$, where $\rho^1\left(v\right) := \rho\left(v\right)$). Define **set concatenation** ST for finite or countable sets $S = \{s_i\}$ and $T = \{t_j\}$ as

$$ST = \{s_i t_j : s_i \in S, t_j \in T\}$$

(where $s_i t_j$ is word concatenation). Then we may define

$$Im_\rho\left(v\right) = Im_\rho\left(v_1\right)Im_\rho\left(v_2\right)\cdots Im_\rho\left(v_k\right)$$

For $n \geq 2$, inductively define $\rho^n(\mathcal{A}) := \left\{ Im_{\rho^{n-1}}\left(w_j^i\right) : 0 \leq j \leq s-1, 1 \leq i \leq i_j \right\}$ to be the **image** of ρ^n. Now, order the words in $\bigcup_{n=0}^{\infty} \rho^n(\mathcal{A})$ by increasing length (and lexicographically for words of the same length). Here, we define $\rho^0(a) = a$. Then apply Lemma 2.0.2 to obtain a symbolic dynamical system $\mathcal{G}_\infty^\rho$. We call any element u of $\mathcal{G}_\infty^\rho$ a **sequence generated by** ρ. Any word in u is called ρ-**admissible**.

Specific examples of random substitutions are considered in [17], [18], and [12]. In this work, we study additional examples of random substitutions and various properties concerning them. Of considerable interest is a sort of "intertwining" of the Fibonacci and Morse substitutions (of which we will say more about in the sequel):

Example 3.0.9 *Let $0 < p < 1$ and $q = 1 - p$. Define the (p, q)-**Fibonacci-Morse Substitution** ("(p, q)-FibMorse substitution" or "FibMorse substitution") to be the random substitution $\zeta : \mathcal{A} \longrightarrow \mathcal{A}^*$ on the alphabet $\mathcal{A} = \{0, 1\}$ defined by*

$$\zeta : \begin{cases} 0 & \longmapsto & 01 \\ 1 & \longmapsto & \begin{cases} 0\,(p) \\ 10\,(q) \end{cases} \end{cases}.$$

We then have the sets

$$\begin{aligned} \zeta(\mathcal{A}) &= \{01, 0, 10\}, \quad \zeta^2(\mathcal{A}) = \{010, 0110, 01, 001, 1001\}, \\ \zeta^3(\mathcal{A}) &= \{01001, 011001, 010001, 0101001, 0110001, 01101001, \ldots\} \end{aligned}$$

*etc. That is, $\zeta^n(\mathcal{A})$ consists of "all possible" words that can arise from applying n independent copies of ζ to 0 or 1. The words obtained in $\zeta^n(\mathcal{A})$ arise from a finite sequence of "Fibonacci choices" or "Morse choices". We will call such a finite (or possible infinite) sequence a **driving sequence**: Let σ be the shift transformation on $X = \{0, 1\}^{\mathbf{Z}_+}$ and ζ_0, ζ_1 the substitutions on $Y = \{0, 1\}^* = \bigcup_{n \geq 1} \{0, 1\}^n$ given by*

$$\zeta_0 : \begin{cases} 0 \longmapsto 01 \\ 1 \longmapsto 0 \end{cases}$$

and

$$\zeta_1 : \begin{cases} 0 \longmapsto 01 \\ 1 \longmapsto 10 \end{cases} .$$

Let $\tau : X \times Y \to X \times Y$ be given by

$$\tau(x, w) = (\sigma(x), \zeta_{x_0}(w)).$$

Given a sequence $x \in X$, we obtain via τ a sequence $u \in X$ from the second coordinate of $\lim_{n\to\infty} \tau^n(x, 0)$. *Then x is the driving sequence that produces u.*

We have the following useful "set representation" of a random substitution on a finite alphabet. Let ρ be a random substitution on $\mathcal{A} = \{a_0, \ldots, a_{s-1}\}$. For each $i \in \{0, \ldots, s-1\}$, let $A_0^i = \{a_i\}$. Now, consider $j \in \{0, \ldots, s-1\}$ and $k \in \{1, \ldots, i_j\}$ (where i_j is as in the definition of ρ on page 18). Let $a = a_j \in \mathcal{A}$ and write $w = w_j^k$ as $w = w_0 w_1 w_2 \ldots w_{m-1}$ For each $n \geq 1$, inductively define

$$A_{n+1}^j = \bigcup_{\ell=1}^{i_{s-1}} A_n^{w_0} A_n^{w_1} A_n^{w_2} \cdots A_n^{w_{m-1}}.$$

So, for example, A_1^j "mimics" $Im_\rho(j)$. As above, order the elements of $\bigcup_{n\geq 0}^{\infty} A_n^0$ by increasing length (and then lexicographically) to obtain a symbolic dynamical system $(\mathcal{SR}_\infty^\rho, \sigma)$ by Lemma 2.0.2 (where $\sigma = \sigma|_{\mathcal{SR}_\infty^\rho}$ is the restriction of the shift map). We call $\mathcal{SR}_\infty^\rho$ the **set representation** of the random substitution ρ.

Example 3.0.10 *Let ζ by the FibMorse substitution and $A_0 = \{0\}, B_0 = \{1\}$. Inductively define $A_{n+1} = A_n B_n$ and $B_{n+1} = B_n A_n \cup A_n$. Note that $A_n^0 = A_n$ and $A_n^1 = B_n$. Then*

$$\begin{aligned} A_1 &= \{01\}, \quad B_1 = \{10, 0\} \\ A_2 &= \{0110, 010\}, \quad B_2 = \{1001, 001, 01\} \\ A_3 &= \{01101001, 0110001, 011001, 0101001, 010001, 01001\} \end{aligned}$$

etc.

Note that in this example, A_n consists of all words that "can arise" from $\zeta^n(0)$. For example, $\zeta^2(\mathcal{A}) = A_2 \cup B_2$. The next Proposition shows that this is true in general.

Proposition 3.0.2 *Let ρ be a random substitution and $\mathcal{G}_\infty^\rho$ the set of all infinite sequences generated by ρ. Then $\mathcal{G}_\infty^\rho = \mathcal{SR}_\infty^\rho$.*

Proof. It suffices to show that $A_n^j = Im_{\rho^n}(\mathcal{A})$ for all n, j. We proceed by induction. Indeed, the base case $n = 0$ is trivial by the definitions. Now suppose that for each j, there is some N such that $A_n^j = Im_{\rho^n}(\mathcal{A})$ for all $n \leq N$. If $v \in A_{N+1}^j$, then $v = v_N^0 v_N^1 \cdots v_N^{s-1}$ for some $v_N^i \in A_N^{w_i} = Im_{\rho^N}(w_i)$. This means that $v \in Im_{\rho^{N+1}}(\mathcal{A})$. The reverse inclusion is similar. ■

Remark 3.0.7 *The FibMorse system thus defined is equivalent to the dynamical system obtained from the Fibonacci-Morse substitution defined above (Definition 3.0.9).*

We will see in Corollary 4.4.1 that the dynamical system arising from irreducible substitutions are in a sense "deterministic". First, we ask about some more "basic ergodic" properties a dynamical system may have. For example, what does minimality say for substitutions? We first note the following Theorem about sequences.

Theorem 3.0.1 ([21], Theorem 4.12) *Let $u \in \mathcal{A}^{\mathbf{N}}$ and $X = \overline{Orb(u)}$. Then (X, σ) is minimal if and only if u is almost periodic.*

Proof. First, suppose that u is almost periodic but (X, σ) is not minimal. Then there is some $x \in X$ whose orbit under σ is not dense and so there is a word $u_0 \ldots u_j$ in u that is not in x. Since $x \in X = \overline{Orb(u)}$, there is a sequence $\{n_k\}_{k=1}^\infty$ such that $\lim_{k\to\infty} \sigma^{n_k}(u) = x$. Since u is almost periodic, $u_0 \ldots u_j$ occurs in u with bounded gaps so that $u_{n_k+s} \ldots u_{n_k+s+j} = u_0 \ldots u_j$ for some s and infinitely many k. Now, by continuity of σ, $\lim_{k\to\infty}\sigma^{n_k+s}(u)=\sigma^s(x)$. By definition of σ, this means that, for sufficiently large k, $u_{n_k+s}\dots u_{n_k+s+j}=x_s\dots x_{s+j}$. But then $u_0\dots u_j=x_s\dots x_{s+j}\in x$, a contradiction.

Conversely, suppose that (X,σ) is minimal. We will show that any word $u_0\dots u_j$ in u occurs in u with bounded gaps. Let V be neighborhood of u (so that any sequence in V agrees with u for a long time). Then $X=\cup_{n\geq 0}\sigma^{-n}(V)$ since $\sigma^{-n}(V)=\{y:\sigma^n(y)\in V\}$. By compactness (as noted after Definition 2.0.4), there is a finite set $\{n_k\}_{k=1}^N$ such that $X=\cup_{k=1}^N\sigma^{-n_k}(V)$. Let $M=\max\{n_k:1\leq k\leq N\}$. Then any sequence enters V after at most M iterations of σ. In particular, one of the sequences $\sigma^j(u),\sigma^{j+1}(u),\dots,\sigma^{j+M}(u)$ enters into V. Therefore, u is almost periodic. ■

Theorem 3.0.2 ([21], Theorem 5.2) *Let $u=\omega^\infty(a)$ be a sequence generated by a substitution ω such that $\omega(a)$ starts with a for some $a\in\mathcal{A}$ and let (X,σ) be the dynamical system arising from u. (X,σ) is minimal if and only if ω is irreducible.*

Proof. Suppose that (X,σ) is minimal. Then every word in u occurs in u with bounded gaps (Theorem 3.0.1). In particular, a occurs with bounded gaps in u. Let $b\in\mathcal{A}$. Since $u=\omega^n(u)$ for all $n\geq 0$, (recall the Standing Hypothesis on page 15) u contains $\omega^n(b)$ for every $n\geq 0$. Thus, since $\lim_{n\to\infty}|\omega^n(b)|=\infty$, there is some n_b such that $\omega^{n_b}(b)$ contains a.

Now suppose that for every $b\in\mathcal{A}$, there is some $n_b\geq 0$ such that $\omega^{n_b}(b)$ contains a. We will show that every word in u occurs in u with bounded gaps. For this, it is enough to show that a occurs with bounded gaps. Indeed, if a occurs in u with bounded gaps, then $\omega^n(a)$ occurs in u with bounded gaps for every n. Then any word B will occur in u with bounded gaps since $B\in\omega^n(a)$ for n large enough. For each $b\in\mathcal{A}$, let $M_b=\min\{n\geq 0:\omega^n(b)\text{ contains }a\}$ and let $N=\max_{b\in\mathcal{A}}\{M_b\}$. Note that, since $\omega^{n_b}(b)$ contains a, by induction $\omega^{n_b+k}(b)$ contains a for every $k\geq 1$. Then $\omega^N(b)$ contains a for every $b\in\mathcal{A}$. Since $u=\omega^N(u)=\omega^N(u_0u_1u_2\cdots)$ and $u_i\in\mathcal{A}$ for all i, u is therefore the concatenation of words of the form $\omega^N(b)$, each of which contains a. Hence, a occurs with bounded gaps, as was to be shown. ■

3.1 Ergodicity and Mixing of Substitutions

Do invariant probability measures exist for symbolic dynamical systems? If so, can they be constructed, or is their existence just theoretical? Krylov and Bogolioulov answers these questions in a general setting:

Theorem 3.1.1 *(Krylov-Bogolioulov) Let T be a continuous transformation of a compact metric space X. Then there is a T-invariant probability measure on X.*

Proof. Fix $x \in X$ and for each $n \in \mathbf{N}$, define a measure μ_n on X by

$$\mu_n(B) := \frac{1}{n} \sum_{k=0}^{n-1} \chi_B\left(T^k x\right),$$

where χ_B is the characteristic function on B. Then any weak-* limit point of the sequence $\{\mu_n\}_{n=1}^{\infty}$ is a T-invariant measure on X. ■

Alternatively, we can define the measure of a cylinder set $[w]$ to be the frequency with which w occurs. In fact, by Theorem 3.0.2, if the substitution is irreducible, then this is the only way to define the measure:

Theorem 3.1.2 ([21], Theorem 5.13) *If the dynamical system arising from a substitution ω is minimal, then it is uniquely ergodic.*

Ergodicity is, in a sense, the "weakest" notion of mixing in a system. On the other hand, it turns out that primitive substitutions are not strongly mixing. Here is the idea. Take, for example, the Fibonacci substitution sequence $u = \omega^{\infty}(0)$. In this case, you know that a 0 will *always* follow a 1. So the word $\omega^n(0)$ will always eventually occur after

$\omega^n(1)$ and, moreover, the distance between their first symbols, $s_n = |\omega^n(0)| - |\omega^n(1)|$, is known. But then shifting $\omega^n(0)$ to the left by s_n places in u matches up with $\omega^n(1)$. So if you looked at a cylinder set $[w]$ of a word $w \in \omega^n(1)$, you will see w in $\sigma^{-n}([w])$ at the "same place" so that these cylinder sets are the same. This means they're not "approximately independent" and so we can not have strongly mixing.

Theorem 3.1.3 *The dynamical system (X, μ, σ) arising from a primitive substitution ω (so μ is an invariant probability measure) is not strongly mixing.*

Proof. We include the proof originally due to Dekking and Keane [7]. For simplicity, we assume ω is defined on the alphabet $\{0, 1\}$. Interchanging 0 and 1 if necessary, $\mu([00]) > 0$, where μ is a σ-invariant probability measure (whose existence is guaranteed by Krylov-Bogolioulov, Theorem 3.1.1). Let w be a word in u, $s_n = |\omega^n(0)|$ and $D_n = [w] \cap \sigma^{-s_n}([w])$. If X were strongly mixing, then $\lim_{n\to\infty} \mu(D_n) = \mu([w])^2$. However, we will show that there exists a constant $C > 0$ that is independent of w such that $\lim_{n\to\infty} \mu(D_n) \geq C\mu([w])$. This will show that X cannot be strongly mixing since $\mu([w])$ can be made arbitrarily small by choosing w arbitrarily long (by Theorem 3.0.2 and Theorem 3.1.2).

Now, $\mu(D_n)$ is the limit of the occurrence frequency of two w's at a distance s_n in the block $\omega^N(0)$, as N tends to infinity. That is, $D_n \in \omega^N(0)$ if $wBw \in \omega^N(0)$ for some word B in u of length $s_n - |w|$. Suppose that N is sufficiently bigger than n. Then

$$\begin{aligned} \left|\omega^N(0)\right|_{wBw} &\geq |\omega^n(00)|_{wBw} \cdot \left|\omega^N(0)\right|_{\omega^n(00)} \\ &\geq |\omega^n(0)|_w \cdot \left|\omega^{N-n}(0)\right|_{00}, \end{aligned}$$

since an occurrence of w in $\omega^n(0)$ implies an occurrence of wBw in $\omega^n(00)$, where B is *some* block. Therefore, we have that

$$\mu(D_n) \geq \lim_{N\to\infty} \frac{1}{s_N} |\omega^n(0)|_w \cdot \left|\omega^{N-n}(0)\right|_{00}.$$

Now, s_k grows asymptotically with $\rho\lambda_{PF}^k$, where $\rho > 0$ and $\lambda_{PF} > 1$ is the Perron-Frobenius eigenvalue of ω (see 3.2). Hence,

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{s_N} \left|\omega^{N-n}(0)\right|_{00} &= \lim_{N\to\infty} \frac{1}{s_N} \mu([00]) \cdot s_{N-n} \\ &= \mu([00]) \lim_{N\to\infty} \lambda_{PF}^{-n}. \end{aligned}$$

Thus,

$$\begin{aligned} \lim_{n\to\infty} \mu(D_n) &\geq \mu([00]) \liminf_{n\to\infty} \frac{|\omega^n(0)|_w}{s_n} \cdot \frac{s_n}{\lambda_{PF}^n} \\ &= C \cdot \mu([w]), \end{aligned}$$

where $C := \mu([00]) \cdot \rho > 0$ is independent of w. ■

Primitivity is not always required for mixing, however. For example, the dynamical system arising from the Chacon substitution

$$\omega_4 : \begin{cases} 0 \longrightarrow 0010 \\ 1 \longrightarrow 1 \end{cases}$$

turns out to be weakly mixing but not strongly mixing. See [15], Lemmas 5.5.1 and 5.5.4 for details.

3.2 Frequency of Symbols

It can be useful and interesting to know the frequencies with which the symbols in a substitution occur. It is a consequence of the Perron-Frobenius Theorem on primitive matrices that we can calculate this using the incidence matrix. We refer the reader to [21] or [13] for proofs and details.

Theorem 3.2.1 (Perron-Frobenius) *Let M be a non-zero primitive matrix with non-negative entries. Then M has a eigenvector $\mathbf{v}_{PF}$ with positive entries. There is a positive* *real $\lambda_{PF} > 0$ corresponding to $\mathbf{v}_{PF}$ that has both geometric and algebraic multiplicity one. Furthermore, if λ is any other eigenvalue for M, then $|\lambda| < \lambda_{PF}$.*

We call $\mathbf{v}_{PF}$ and λ_{PF} the **Perron-Frobenius eigenvector** and **Perron-Frobenius eigenvalue**, respectively. As a corollary to Perron-Frobenius, we have (see [21], Proposition 5.9)

Theorem 3.2.2 *Let ω be a primitive substitution and let $M = M(\omega)$ be its incidence matrix. Then for every $a, i \in \mathcal{A}$, the limit*

$$\lim_{n\to\infty} \frac{|\omega^n(a)|_i}{|\omega^n(a)|} = v_i$$

exists and is independent of a. Furthermore, $v_i > 0$ is precisely the ith coordinate of the Perron-Frobenius eigenvector $\mathbf{v}_{PF}$ of M, normalized to one (so that the sum of the entries of $\mathbf{v}_{PF}$ of M is 1).

We will also use the similar consequence (see [21]) of Perron-Frobenius that

$$\lim_{n\to\infty} \frac{|\omega^n(a)|}{\lambda_{PF}^n} = c_a > 0, \tag{3.2.1}$$

where c_a is a constant that involves the Perron-Frobenius eigenvector.

Example 3.2.1 *Let ω_1 be the Morse substitution as in Example 3.0.5. Then the incidence matrix of ω_1 has Perron-Frobenius eigenvector $\mathbf{v}_{PF} = (1,1)$ (corresponding to the eigenvalue $\lambda_{PF} = 2$, which is $\left(\frac{1}{2}, \frac{1}{2}\right)$ when normalized to 1. This means that both the symbols 0 and 1 occur with frequency $\frac{1}{2}$ in $\omega_1^\infty(0)$.*

Example 3.2.2 *Let ω_2 be the Fibonacci substitution as in Example 3.0.6. Let $\Phi = \frac{1+\sqrt{5}}{2}$ and $\varphi = \frac{1-\sqrt{5}}{2}$ be the eigenvalues of the incidence matrix of ω_2. Then $\mathbf{v}_{PF} = (\Phi, 1)$ corresponding to $\lambda_{PF} = \Phi$. $\mathbf{v}_{PF}$ normalized to 1 is the vector $\left(\Phi - 1, \varphi^2\right) \approx (0.62, 0.38)$. So the symbol 0 occurs about 62% and 1 occurs about 38% of the time in $\omega_2^\infty(0)$.*

Example 3.2.3 *Let ζ be the (p,q)-FibMorse substitution as in Example 3.0.9. The incidence matrix of ζ is*

$$M = M(\zeta) = \begin{bmatrix} 1 & 1 \\ 1 & q \end{bmatrix}.$$

M has Perron-Frobenius eigenvector

$$\mathbf{v}_{PF} = \left(\frac{1-q+\sqrt{q^2-2q+5}}{2}, 1 \right)$$

corresponding to the eigenvalue $\lambda_{PF} = \frac{1}{2}\left(1+q+\sqrt{q^2-2q+5}\right)$. Normalizing $\mathbf{v}_{PF}$ to 1 yields the vector

$$\mathbf{f}_{PF} = \left(\frac{-1-q+\sqrt{q^2-2q+5}}{2-2q}, \frac{3-q-\sqrt{-2q+q^2+5}}{2-2q} \right).$$

Now, when $q=0$, $\mathbf{f}_{PF} = \left(\Phi-1, \varphi^2\right)$ which agrees with the fact that in this case ζ is the Fibonacci substitution. On the other hand, when $q=1$,

$$\begin{aligned} \mathbf{f}_{PF} &= \left(\lim_{q\to 1^-} \frac{-1-q+\sqrt{q^2-2q+5}}{2-2q}, \lim_{q\to 1^-} \frac{3-q-\sqrt{-2q+q^2+5}}{2-2q} \right) \\ &= \left(\frac{1}{2}, \frac{1}{2}\right), \end{aligned}$$

as expected since, in this case, ζ is the Morse substitution.

Example 3.2.4 *Define the random substitution α by*

$$\alpha : \begin{cases} 0 & \longmapsto & \begin{cases} 10 & \frac{1}{2} \\ 02 & \frac{1}{2} \end{cases} \\ 1 & \longmapsto & 03 \\ 2 & \longmapsto & 04 \\ 3 & \longmapsto & 3 \\ 4 & \longmapsto & 4 \end{cases}$$

We have that

$$M = M(\alpha) = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

has Perron eigenvalue $\lambda_{PF} = \frac{1+\sqrt{5}}{2}$ *with normalized Perron eigenvector*

$$\mathbf{v}_{PF} = \begin{bmatrix} \frac{3-\sqrt{5}}{2} \\ \frac{\sqrt{5}-2}{2} \\ \frac{\sqrt{5}-2}{2} \\ \frac{3-\sqrt{5}}{4} \\ \frac{3-\sqrt{5}}{4} \end{bmatrix} \approx \begin{bmatrix} 0.38197 \\ 0.11803 \\ 0.11803 \\ 0.19098 \\ 0.19098 \end{bmatrix}.$$

3.3 Structure of Words

It can be beneficial to determine the structure of symbolic systems, such as knowing what admissible and forbidden words look like. Towards this end, the set representation of a random substitution can be useful.

Definition 3.3.1 *A forbidden word* w *of length* n *is* ***minimal*** *if it contains no forbidden word of length* $< n$*.*

Example 3.3.1 *Let* ζ *be the FibMorse substitution and* A_n *the sets that generate its set representation (as in Example 3.0.10). We show that neither* $w_1 = 111$ *nor* $w_2 = 11011$ *are contained in any element of any* A_n*.*

By definition of $A_k = A_{k-1}B_{k-1}$ *and* $B_k = (B_{k-1}A_{k-1}) \cup A_{k-1}$*, all words in* $\cup_n A_n$ *start the same (up to the* k*th Fibonacci number). Suppose* $w_1 \in A_n$*. Then, by definition* *of* A_k *and* B_k, *there is some* $k \geq 0$ *and some* $x \in A_k$ *and* $y \in B_k$ *such that* $w_1 = xy$ *where (A)* x *ends in* 1 *and* y *starts with* 11, *or (B)* x *ends in* 11 *and* y *starts with* 1. *The proofs for both cases and for* w_2 *are all similar. We provide a proof for case (A).*

Consider the case (A) where x *ends in* 1 *and* y *starts with* 11. *By definition of* B_k, y *starts with* 11 *means that either some element of some* B_m *starts with* 11 *or some element of some* A_m *starts with* 11. *By induction, since the lengths of elements of the A's and B's grow, we only need to look at* $m = 2$. *But no element of* B_2 *or of* A_2 *begins with* 11.

This result about the minimal forbidden words $w_1 = 111$ and $w_2 = 11011$ in terms of ζ translates to the following:

Corollary 3.3.1 *Any minimal forbidden word contains or extends to a word of either the form* $\zeta^k(111)$ *or of the form* $\zeta^k(11011)$.

For example, the minimal forbidden word 110101 extends to 1101010, which contains $\zeta(111)$.

The method of analyzing the set representation of a random substitution can also yield information about admissible words.

Example 3.3.2 *Let* w *be an admissible word in the FibMorse substitution. We show that if* $|w|$ *is sufficiently large, then* w *must contain one of the following 6 words:*

$$
\begin{aligned}
a &= 0110010, \\
b &= 001010, \\
c &= 1010010, \\
d &= 01100110, \\
e &= 0010110,
\end{aligned}
$$

or

$$f = 10100110.$$

Let w be admissible. Then w is contained in some element of $\bigcup_n A_n$, where $A_n = A_{n-1}B_{n-1}$ This means that there is an N such that w is contained in some element of A_N. Without loss of generality, we can assume N is the smallest such that w is not contained in an element of A_{N-1} (that is, w "breaks up" over A_{N-1} and B_{N-1}). Now, by inspection, we see that any element of any A_n ends in either 0110, 001*, or* 1010 *and any element of any B_n starts with either* 010 *or* 0110 *(for $n \geq 4$). Thus, since $A_N = A_{N-1}B_{N-1}$, either w must contain (at least) one of the 6 possible words* $\{a = 0110.010, b = 001.010, c = 1010.010, d = 0110.0110, e = 001.0110, f = 1010.0110\}$, *or some k-extension of w (that is, an extension by k-letters on the left and/or right) contains one of $\{a, b, c, d, e, f\}$. Therefore, if $|w|$ is sufficiently large (ie, large enough to "be" a k-extension), w must contain one of $\{a, b, c, d, e, f\}$.*

How do we know if $|w|$ is large enough? If we have a word w, we'd like to just check if it contains one of $\{a, b, c, d, e, f\}$ for admissability. If we do not find one of those contained in w, one of the words a-f might be contained in an extension of w. Since this is a finite list and the lengths of these words is small, it is not too difficult to get a lower bound on how long w needs to be to do this simple test. Any of the words a, b, c, d, e, or f will show up infinitely often, so some k-extension will contain it. Suppose w breaks up over A_{N-1} and B_{N-1} (that is, suppose $w \in uv$ where $u \in A_{N-1}$ and $v \in B_{N-1}$). Then, since the longest word in $\{a, b, c, d, e, f\}$ has 8 symbols, either a 7-extension on the right or a 7-extension on the left (or a combination totalling 7 symbols) contains w. Since we don't know if it needs to be extended on the right or left (or both), $|w| \geq 14$ is large enough to contain one of $\{a, b, c, d, e, f\}$.

We can generalize this reasoning. We took the length ℓ_3 of the longest string in A_3 and said that a $(\ell_3 - 1)$-extension on either the right or left (or a combination totalling $\ell_3 - 1$ symbols) must contain w. Then $|w| \geq 2(\ell_3 - 1)$ is long enough to contain one of

$\{a, b, c, d, e, f\}$. The length of the longest string in A_N is 2^N. So $|w| \geq 2\left(2^N - 1\right)$ is long enough to ensure that w contains an element of A_N. Thus, if w be an admissible word in the FibMorse substitution of length at least 6, then w must contain an element of A_N, where $N \leq \log_2\left(\frac{1}{2}|w| + 1\right)$.

Example 3.3.3 *Any word of length 14 must contain an element of* A_3 *and any word of length 30 must contain an element of* A_4*. However, the word* 100 *does not contain an element of* $A_1 = \{01\}$ *(*001 *is not of length at least 6, which is the smallest length of the words* a *-* f *above).*

4 RANDOMNESS

In many areas of mathematics and science, there are certain notions of "randomness". This is a way to quantify to what degree something is disordered. For example, if one were to flip a fair coin, then the probability of seeing a heads is $\frac{1}{2}$ and intuitively there is a lot of randomness involved in whether a heads or tails comes up. Much more so than if flipping a coin in which there is a $\frac{99}{100}$ chance of seeing a heads and a $\frac{1}{100}$ chance of seeing a tails. Indeed, with the latter biased coin, there is hardly any amount of randomness involved, as it is almost certain of seeing a heads. That is, there is almost no uncertainty in seeing a heads come up. We would like to apply and quantify the idea of randomness to substitutions. We have already introduced the Fibonacci substitution and the Morse substitution, and we will see that these (indeed, all primitive) substitutions are not random at all in the following sense: Suppose, for example, we flip a coin once every second forever, and the coin is biased so that when we write down the sequence of heads and tails, we get the Fibonacci sequence that is generated by the Fibonacci substitution (interpreting heads as 0 and tails as 1). Then a coin flip is not random at all in that we already know what is going to come up (assuming we know where we are in the Fibonacci sequence). In this sense, we will show later that any sequence arising from a primitive substitution is not random. On the other hand, consider the FibMorse Substitution

$$\zeta : \begin{cases} 0 & \longmapsto & 01 \\ 1 & \longmapsto & \begin{cases} 0 & \text{with probability } p \\ 10 & \text{with probability } q \end{cases} \end{cases},$$

defined above. Then it seems that there is then some randomness involved in generating an infinite string via ζ. This is the content of Theorem 4.2.2 below. But first, let's understand a little more about the precise nature of randomness.

4.1 Entropy

Entropy is an old concept in science, being initiated in 1803 by the French mathematician Lazare Carnot. The concept is generally interpreted as the amount of disorder or randomness in a system. In 1948, Claude Shannon [23] investigated this concept in information transmission (this theory was later made more precise in 1953 by Khinchine [10]). He wanted to somehow quantify the amount of information gained or uncertainty removed when bits of information was transmitted. From these ideas, he introduced the idea of entropy in information theory, a concept that was abstracted to dynamical systems in ergodic theory. In the next section, we discuss and define entropy (we refer the reader to [20] or [25] for further details). We then use a method to calculate it for several specific random substitutions (beyond previously known results).

4.2 Metric Entropy

There are many things in nature and science that are "random" or "unpredictable". We would like a way to measure this randomness. For example, flipping a fair coin is far more unpredictable than flipping a coin weighted so that heads come up 95% of the time and tails only 5% of the time. The fair coin seems more random than the weighted one and so a measure of "randomness" should be greater for the former than the latter. The coin can be modeled by a random variable Y taking values in $\{H, T\}$ with probability distribution $\left(\frac{1}{2}, \frac{1}{2}\right)$ (for the fair coin) or $(0.95, 0.5)$ (for the weighted coin).

We now view randomness as an amount of uncertainty or "information gained". Fix a point c in the unit interval $J = [0, 1]$ and divide J into two halves of equal length (call the left half H for Heads and the right half T for Tails). If someone tells you which half of J that c_0 is in, then you've gained one bit ($\log_2 2$) of information about the location of c_0. Now divide the half that c_0 is in into halves (so we've essentially divided J into 4 subintervals of equal lengths, call them HH, HT, TH, TT). If someone tells you which "new half" (i.e., quarter of J) that c_0 is in, you have gained two bits ($\log_2 4$) of information about the location of c_0. In general, divide J into 2^n subintervals of lengths $\frac{1}{2^n}$ each. If someone tells you which of the subintervals that c_0 is in, you have gained $n = \log_2 (2^n)$ bits of information about the location of c_0. Flipping a fair coin n times corresponds to a person telling you which subinterval c_0 is in. So, for example, a sequence of n random variables taking values in $\{H, T\}$, each with probability distribution $\left(\frac{1}{2}, \frac{1}{2}\right)$ gives you n bits of information.

We can generalize these ideas as follows. Let μ be a measure on J and let $\mathcal{P} = \{P_1, \ldots, P_N\}$ partition J into subintervals of lengths $\mu(P_i)$ $(i = 1, \ldots, N)$. If someone tells you that $c_0 \in P_i$, then you have gained $\log_2 \frac{1}{\mu(P_i))} = -\log_2 \mu(P_i)$ bits of information about the location of c_0. In general, for any partition $\mathcal{P} = \{P_1, \ldots, P_N\}$ (all of positive measure) of a probability space (X, μ), define the **information content** of the partition element P_i to be $H(P_i) = -\log \mu(P_i)$. If we think of the partition elements as events (for example, the outcomes of a random variable), then heuristically this definition makes sense because it matches what properties a "measure of information" should have: (a) It should be non-negative; (b) If the probability of an event P_i is one, then we should get no information from that event; (c) If two events occur independently, then the information gained from both evens should be the sum of the individual events; and (d) If we change the probability of an even occurring a little bit, then the amount of information in that even should only change by a little bit (that is, a measure of information should be "continuous").

For a parition $\mathcal{P} = \{P_1, \ldots, P_N\}$, how should we quantify the total amount of information in $\mathcal{P}$? A natural way would be to average the amount of information in each partition element P_i. Since each P_i may have different measures, we therefore would ask for a weighted average of the information content of all the partition elements. Thus, we define:

Definition 4.2.1 *The* ***entropy of a finite partition*** $\mathcal{P} = \{P_1, \ldots, P_N\}$ *of a measure space* (X, μ) *is defined to be*

$$H(\mathcal{P}) = -\sum_{i=1}^{N} \mu(P_i) \log_2 \mu(P_i),$$

where we define $0 \log 0 = 0$.

Remark 4.2.1 *The base of the logarithm is somewhat arbitrary. Sometimes the base is* 10 *or e. A base of* 2 *is natural because it can count the number of "Yes-No" decisions. It is because of this arbitrarity of base that we are not requiring* (X, μ) *to be a probability space. Entropy is usually just "normalized to 1", but can differ up to a constant (depending on the base). Henceforth we will often disregard the base.*

Remark 4.2.2 *Originally, entropy was denoted by the upper case Greek letter eta,* H*. But then* H *was used for other things in mathematical theory, and so the lower case Roman equivalent was used.*

So the entropy of a partition is the (weighted) average amount of information we get from an event in the partition. Note that by Jensen's Inequality ([20], p. 230), $H(\mathcal{P})$ is maximal when $\mu(P_i) = \frac{1}{N}$ for all i. For example, viewing the flip of a fair coin as the unit interval J divided into equal halves, the average amount of information in this coin flip is $\log_2 2 = 1$ bit whereas the average amount of information in flipping the unfair coin having probability distribution $(0.95, 0.05)$ is 0.2864 bits (implying that there is less information in a coin in which we are not surprised to see a heads most of the time). These ideas lead to:

Definition 4.2.2 *The* ***entropy of the random variable*** Y *is*

$$H(Y) = -\sum_{i=1}^{N} p_i \log_2 p_i,$$

where $(p_1, \ldots, p_N)$ *is the probability distribution of* Y.

Let us now take a different (though related) useful heuristic for entropy: Shannon's entropy as applied to information transmission (cf [20]). Imagine a source, such as a ticker-tape, which prints out a symbol from a finite set $\mathcal{A}$ of symbols, of cardinality s. Suppose that each symbol $a_i \in \mathcal{A}$ has a certain probability p_i (so that $\sum_i p_i = 1$) of being printed by the ticker-tape. Then there is a certain amount of "surprise", or information gained, from reading that symbol. So we would like to measure the amount of "information gained". If the event that a symbol is printed is independent of what symbols came before it, then the measurement function of the quantity we're looking for should be additive at the least. We would also want this function to be continuous since if the probability of a symbol decreases a little, then the surprise in seeing it should increase a little. In view of this, again the logarithm is the natural choice. That is, we might say that the **information content** of the event of seeing the symbol a_i is $-\log_2 p_i$. In this case, we might think to "average" the information content of each symbol. Thus, we define the average amount of information per symbol by the quantity

$$H = -\sum_{a_i \in \mathcal{A}} p_i \log_2 p_i.$$

Example 4.2.1 *Let $\{X_n\}_{n=0}^{\infty}$ be an i.i.d. (independent and identically distributed) and equiprobable stochastic process on $\{0,1\}$. Then*

$$\begin{aligned} H &= -\left(\Pr\left(X_n = 0\right)\log\Pr\left(X_n = 0\right) + \Pr\left(X_n = 1\right)\log\Pr\left(X_n = 1\right)\right) \\ &= \log 2. \end{aligned}$$

This particular process is called a **Bernoulli coin flip** because it models an experimenter flipping a fair coin each day forever (see Example 4.2.2).

Again we see that entropy H satisfies two other properties we would want an "information function" to have: If $p_i = 1$ and $p_j = 0$ for all $j \neq i$, then $H = 0$. That is, there is no information gained if we know that only one of the symbols will ever be printed. Also, it can be shown ([20], p. 230) by Jensen's Inequality that H is a maximum when $p_i = \frac{1}{s}$

for all i. Now, it may happen that the event that some symbol a_i is printed depends on some number of symbols printed before it. In this case, we are considering words of symbols of length n, each word occurring with a certain probability. Letting Ω_n be the collection of all words of length n, the quantity

$$-\frac{1}{n}\sum_{w\in\Omega_n}\Pr(w)\log_2 P(w) \tag{4.2.1}$$

is the average amount of information in reading a word w (where $\Pr(w)$ is the probability that the word w occurs). This leads to the following definition.

Definition 4.2.3 *The **entropy of the source** is defined to be*

$$h=\lim_{n\to\infty}-\frac{1}{n}\sum_{w\in\Omega_n}\Pr(w)\log_2\Pr(w)\,. \tag{4.2.2}$$

So h can be interpreted as the average amount of information gained in reading a symbol from a ticker-tape that prints forever. How does this translate in ergodic theory? Note that a ticker-tape that prints forever is printing out an infinite string x of elements of $\mathcal{A}$, that is, an element of $\mathcal{A}^{\mathbf{Z}_+}$ (or $\mathcal{A}^{\mathbf{Z}}$) and the printing of a symbol is just an application of the shift map σ on x. Now consider a partition $\mathcal{P}=\{P_1,P_2,\ldots,P_n\}$ of a measure-theoretical dynamical system (X,Σ,μ,T). Given $x_0\in X$, we may produce an element x of $\mathcal{A}^{\mathbf{Z}_+}$ by letting $x_i=P_j$ if $T^i x_0\in P_j$ for all $i\geq 0$. The $x\in\mathcal{A}^{\mathbf{Z}_+}$ obtained in this way is called the **coding** of x_0 by T (also known as the $(\mathcal{P},T)$-**name** of x_0). If we consider this coding as a "source", the average amount of information gained in seeing a symbol in x is just the quantity h in 4.2.2. To put this in terms of our coding, first define the **join** of two partitions $\mathcal{P}$ and $\mathcal{Q}$ to be $\mathcal{P}\vee\mathcal{Q}=\{P\cap Q : P\in\mathcal{P}, Q\in\mathcal{Q}\}$. Suppose a word $P_{j_0}P_{j_1}\cdots P_{j_{n-1}}$ of length n appears at the ith place in x. Then $T^i(x_0)\in P_{j_0}$, $T^{i+1}(x_0)\in P_{j_1}$, ..., $T^{i+n-1}(x_0)\in P_{j_{n-1}}$ since $T^{i+n}(x_0)\in P_{j_n}$ if and only if $T^i(x_0)\in T^{-n}(P_{j_n})$ for all n. So $T^i(x_0)\in P_{j_0}\cap T^{-1}(P_{j_1})\cap\cdots\cap T^{-n+1}\left(P_{j_{n-1}}\right)$. Thus, by the definition 4.2.1 of the entropy of a partition, the quantity 4.2.1 becomes

$$\frac{1}{n}H\left(\mathcal{P}\vee\mathcal{T}^{-1}(\mathcal{P})\vee\cdots\vee T^{-n+1}(\mathcal{P})\right),$$

where $T^{-n}(\mathcal{P})$ is the finite partition $\{T^{-n}(P) : P \in \mathcal{P}\}$. We may think of a partition $\mathcal{Q}$ of X has a set of possible outcomes (events) of an experiment and $\mathcal{P} \vee \mathcal{Q}$ as the possible outcomes of a joint experiment. In this case, $H(\mathcal{Q})$ is a measurement in the amount of surprise, or information gained, in performing the experiment and $H(\mathcal{P} \vee \mathcal{Q})$ that of the joint experiment. As noted above, $H(\mathcal{Q})$ is maximal when $\mu(Q) = \frac{1}{|\mathcal{Q}|}$ for each $Q \in \mathcal{Q}$. If μ is a probability measure, this means we are asking for each $Q \in \mathcal{Q}$ to occur with the same frequency, which heuristically means that we should have "maximal surprise".

We will show in Lemma 4.2.2 that the following limit exists for a measure-preserving transformation T.

Definition 4.2.4 *The* ***metric entropy*** $h_\mu(\mathcal{P}, T)$ ***of the partition*** $\mathcal{P}$ *of the dynamical system* $(X, \mathcal{P}, \mu, T)$ *is*

$$h_\mu(\mathcal{P}, T) = h(\mathcal{P}, T) := \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})\right).$$

This represents the average information gained in knowing into which element of $\mathcal{P}$ that T moves points of X. Equivalently, $h(\mathcal{P}, T)$ represents the average amount of information gained in reading an element of $\mathcal{P}$. If the measure of a certain partition element P is small, then we should gain more information (and more surprise) in knowing that $T(x) \in P$. However, if too many sets in $\mathcal{P}$ are "too big", then we may not gain much information at all in knowing where T moves points. We therefore would like to consider all possible partitions if we are to measure how much information is gained in knowing where T moves points. We thus arrive at the following definition:

Definition 4.2.5 *The* ***metric entropy*** $h_\mu(X, T)$ ***of the dynamical system*** $(X, \mathcal{P}, \mu, T)$ *is*

$$h_\mu(X, T) = h(X, T) := \sup_{\mathcal{P}} h_\mu(\mathcal{P}, T),$$

where the supremum is taken over all partitions $\mathcal{P}$ *of* X*. If* $h_\mu(X, \sigma) = 0$*, then we call the dynamical system* (X, T) ***deterministic****.*

We can interpret $h_\mu(X)$ as the average amount of information gained in knowing where T moves points of X. In order to show that the limit $h_\mu(X)$ exists, we need a property of the entropy of partitions and an analytical Lemma.

Proposition 4.2.1 *If $\mathcal{P}$ and $\mathcal{Q}$ are finite partitions of X, then $H(\mathcal{P}\vee\mathcal{Q}) \leq H(\mathcal{P}) + H(\mathcal{Q})$.*

Proof. We have

$$\begin{aligned}
-H(\mathcal{P}\vee\mathcal{Q}) &= \sum_{P\cap Q\in\mathcal{P}\vee\mathcal{Q}} \mu(P\cap Q)\log\mu(P\cap Q) \\
&= \sum_{P,Q} \mu(P\cap Q)\log\mu(P\cap Q) \\
&= \sum_{P,Q} \mu(P\cap Q)\log\left(\mu P\frac{\mu(P\cap Q)}{\mu P}\right) \\
&= \sum_{P,Q} \mu(P\cap Q)\log\mu P + \sum_{P,Q} \mu(P\cap Q)\log\frac{\mu(P\cap Q)}{\mu P}.
\end{aligned}$$

For each $P\in\mathcal{P}$, we have that $\sum_Q \mu(P\cap Q) = \mu(P)$ since the sets in $\mathcal{Q}$ are disjoint and cover X. So

$$-H(\mathcal{P}\vee\mathcal{Q}) = \sum_{P} \mu P\log\mu P + \sum_{P,Q} \mu(P\cap Q)\log\frac{\mu(P\cap Q)}{\mu P}.$$

Now, let

$$\varphi(x) = \begin{cases} 0 \text{ if } x=0 \\ x\log x \text{ if } x\neq 0 \end{cases}.$$

φ is continuous and concave on $[0,\infty)$. Then by Jensen's Inequality,

$$\varphi(ax+by) \leq a\varphi(x) + b\varphi(y)$$

for all $x, y \in [0, \infty)$ and $a, b \geq 0$ such that $a + b = 1$. Then, for each $Q \in \mathcal{Q}$,

$$\begin{aligned} \mu(Q)\log\mu(Q) &= \varphi(\mu Q) \\ &= \varphi\left(\sum_P \mu P \frac{\mu(P\cap Q)}{\mu P}\right) \\ &\leq \sum_P \mu P \cdot \varphi\left(\frac{\mu(P\cap Q)}{\mu P}\right) \\ &= \sum_P \mu(P\cap Q)\log\frac{\mu(P\cap Q)}{\mu P}. \end{aligned}$$

Thus, summing over all $Q \in \mathcal{Q}$, we have that

$$\sum_Q \mu(Q)\log\mu(Q) \leq \sum_{P,Q} \mu(P\cap Q)\log\frac{\mu(P\cap Q)}{\mu P}$$

so that

$$\begin{aligned} -H(\mathcal{P}\vee\mathcal{Q}) &\geq \sum_P \mu P\log\mu P + \sum_Q \mu(Q)\log\mu(Q) \\ &= -H(\mathcal{P}) - H(\mathcal{Q}) \end{aligned}$$

and thus

$$H(\mathcal{P}\vee\mathcal{Q}) \leq H(\mathcal{P}) + H(\mathcal{Q}),$$

as desired. ■

This makes sense if we think in terms of experiments: We should gain more information from performing experiment $\mathcal{P}$ and $\mathcal{Q}$ separately than if we perform $\mathcal{P}$ and $\mathcal{Q}$ as one experiment.

A sequence $\{a_n\}$ of real numbers is **subadditive** if $a_{n+m} \leq a_n + a_m$ for all n, m.

Lemma 4.2.1 *If $\{a_n\}$ is a subadditive sequence, then $\lim_{n\to\infty} \frac{a_n}{n}$ exists and equals $L := \inf_n \frac{a_n}{n}$.*

Proof. Fix m and let $n \geq m$. Write $n = km + \ell$ for some $k \geq 0$ and $0 \leq \ell < m$. Then

$$a_n \leq k a_m + a_\ell$$

by subadditivity. As $n \to \infty$, $k \to \infty$ so that $\frac{n}{k} = m + \frac{\ell}{k} \to m$ since m is fixed and $0 \le \ell < m$. Thus,

$$\limsup_{n\to\infty} \frac{a_n}{n} \le \limsup_{n\to\infty} \frac{k}{n} a_m + \limsup_{n\to\infty} \frac{a_\ell}{n} = \frac{a_m}{m}.$$

Since m was arbitrary, we then have that

$$\begin{aligned} \liminf_{n\to\infty} \frac{a_n}{n} &\le \limsup_{n\to\infty} \frac{a_n}{n} \\ &\le \inf_{n\ge 1} \frac{a_n}{n} \\ &\le \liminf_{n\to\infty} \frac{a_n}{n} \end{aligned}$$

since $\liminf_{n\to\infty} \frac{a_n}{n} = \lim_{n\to\infty} \left(\inf_{k\ge n} \frac{a_k}{k}\right)$. Therefore,

$$\lim_{n\to\infty} \frac{a_n}{n} = \liminf_{n\to\infty} \frac{a_n}{n} = L.$$

■

Remark 4.2.3 *A similar proof shows that* $\lim_{n\to\infty} \frac{a_n}{n} = \limsup_{n\to\infty} \frac{a_n}{n}$ *if* $\{a_n\}$ *is **superadditive** (that is,* $a_{n+m} \ge a_n + a_m$ *for all* n, m*).*

Proposition 4.2.2 *If* T *is measure-preserving, then the limit* $h_\mu(X,T)$ *exists.*

Proof. We show that the sequence $\{a_n\}_{n=1}^{\infty}$ defined by

$$a_n = H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})\right)$$

is subadditive. Indeed, we have

$$\begin{aligned} a_{n+m} &= H\left(\bigvee_{i=0}^{n+m-1} T^{-i}(\mathcal{P})\right) \\ &\le H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})\right) + H\left(\bigvee_{i=n}^{n+m-1} T^{-i}(\mathcal{P})\right) && (4.2.3) \\ &= H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})\right) + H\left(\bigvee_{i=0}^{m-1} T^{-i}(\mathcal{P})\right) && (4.2.4) \\ &= a_n + a_m. \end{aligned}$$

Equation 4.2.3 is true because of Proposition 4.2.1 and equation 4.2.4 is true because T is measure-preserving. Therefore, $h_\mu(T)$ exists by Lemma 4.2.1. ■

In general, computing entropy can be very difficult since one must consider all possible partitions of the dynamical system. In certain cases, however, we only need one "nice" partition. This would be a partition which captures all the information at once.

Definition 4.2.6 *A partition $\mathcal{P}$ of a dynamical system (X, Σ, T) is called a* ***generator*** *of (X, Σ, T) if $\bigvee_{i=-\infty}^{\infty} T^i(\mathcal{P}) = \Sigma$.*

In this case, we have the following useful result (see, for example, [20] for a proof).

Theorem 4.2.1 (Kolmogorov-Sinai) *If $\mathcal{P}$ is a generator of (X, Σ, μ, T), then*

$$h_\mu(X, T) = h_\mu(\mathcal{P}, T).$$

One of the great uses of entropy is that it is invariant under isomorphism (see [25], Theorem 4.11).

Proposition 4.2.3 *Let $(X, \mathcal{P}, \mu, T)$ and $(Y, \mathcal{Q}, \nu, S)$ be isomorphic measure-theoretic dynamical systems. Then $h_\mu(T) = h_\nu(S)$.*

Example 4.2.2 *Let $\mathcal{A} = \{a_0, a_1, \ldots, a_{s-1}\}$ be a finite alphabet and assign to each a_i a "weight" p_i, such that $p_i \geq 0$ for all i and $\sum_i p_i = 1$ (we think of each a_i occurring with probability p_i). Let $X = \mathcal{A}^{\mathbf{Z}_+}$, let T be the left-shift on X, and let μ be product measure (so that $\mu[a_{i_0}a_{i_1}a_{i_2}\cdots a_{i_k}] = p_{i_0}p_{i_1}p_{i_2}\cdots p_{i_k}$). The dynamical system (X, μ, T) is called a* ***Bernoulli shift****, denoted $\mathcal{B}(p_0, p_1, \ldots p_{s-1})$. We think of $\mathcal{B}\left(\frac{1}{2}, \frac{1}{2}\right)$ as modeling an experimenter flipping a fair coin every day for eternity. The Bernoulli shifts $\mathcal{B}\left(\frac{1}{2}, \frac{1}{2}\right)$ and $\mathcal{B}\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ are not isomorphic since $h\left(\mathcal{B}\left(\frac{1}{2}, \frac{1}{2}\right)\right) = \log 2$ and $h\left(\mathcal{B}\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)\right) = \log 3$.*

Entropy is not in general a *complete* invariant. That is, two dynamical systems with the same entropy may not be isomorphic. For example, let X_m be a space with exactly m points and let $T_m : X_m \to X_m$ be a permutation of these points. Then T_m and T_n are not isomorphic for all $m \neq n$. However, if we assign each singleton in X_m the same measure, then the entropy of (X_m, T_m) is zero for all m. In 1970 [19], Donald Ornstein showed that entropy is a complete invariant for the class of Bernoulli shifts (see also [24] for a detailed exposition or [20] for a sketch of the theory). It is interesting to note that it was unknown for a long time whether $\mathcal{B}_1 = \mathcal{B}\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ and $\mathcal{B}_2 = \mathcal{B}\left(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$, both having entropy $2\log 2$, were isomorphic. In 1959, Melshalkin [16] constructed an isomorphism between $\mathcal{B}_1$ and $\mathcal{B}_2$. However, it is this "Ornstein Isomorphism Theorem" that answered the general case affirmatively.

A method of computation of entropy

What do the definitions involved in entropy say in the case that (X, T) is a symbolic dynamical system? Let $\mathcal{C}_n$ denote the set of all length-n cylinder sets at 0. That is, $\mathcal{C}_n$ is the set of all cylinder sets $[w] := \{x \in X : x_0 x_1 \dots x_{n-1} = w\}$ where $|w| = n$. $\mathcal{C}_0$ is a generating partition. So by Kolmogorov-Sinai,

$$h_\mu(\mathcal{P}, T) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{C}_0)\right).$$

Now, $\mathcal{P} \vee T^{-1}(\mathcal{P}) \vee \cdots \vee T^{-n+1}(\mathcal{P}) = \mathcal{C}_n$ and so

$$H(\mathcal{C}_n) = -\sum_{C \in \mathcal{C}_n} \mu C \log \mu C$$

is the weighted average amount of information content we get in a word of length n. Dividing this by n then gives us the average amount of information content per symbol in a word of length n. Therefore,

$$h_\mu(\mathcal{P}, T) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{C}_n) \tag{4.2.5}$$

is the average amount of information per symbol in an infinite string in X.

In [12] and [17], the entropy of a randomized variant of the Fibonacci substitution (which is similar to the FibMorse substitution defined above) is considered and bounds given. This example is generalized in [18]. We now further this work to compute exactly the entropy for several examples of random substitutions and compute a positive lower bound for FibMorse. The method that accomplishes this is to use equation 4.2.5 and note that $H(\mathcal{C}_n)$ is the probability distribution of a certain random variable.

Example 4.2.3 *For fixed p and $q = 1 - p$, define the **Bridge Substitution** $\beta_{p,q}$ as*

$$\beta : \begin{cases} 0 \longmapsto \begin{cases} 010 & (p) \\ 020 & (q) \end{cases} \\ 1 \longmapsto \begin{cases} 101 & (p) \\ 121 & (q) \end{cases} \\ 2 \longmapsto \begin{cases} 202 & (p) \\ 212 & (q) \end{cases} \end{cases}.$$

We wish to compute the entropy h_β of $\beta = \beta_{p,q}$ when $p = q = \frac{1}{2}$. We therefore want to know the average amount of information gained in reading a symbol in $\beta^\infty(0)$. Since $|\beta^n(i)| = 3^n$, from equation 4.2.5 we have

$$h_\beta = \lim_{n\to\infty} \frac{1}{3^n} H(\mathcal{C}_{3^n}),$$

where $H(\mathcal{C}_{3^n})$ is the weighted average of the information content in $\beta^n(0)$. Note that $H(\mathcal{C}_{3^n})$ is the entropy of the probability distribution of $\beta^n(0)$, which we now calculate. Let B_n be the set of all possible words arising from $\beta^n(0)$. By induction,

$$|B_n| = \prod_{k=1}^{n} 2^{3^{k-1}}.$$

Since $p = q = \frac{1}{2}$, each $\beta^n(i)$ is uniformly distributed and so each element of B_n occurs with the same probability. The probability distribution of $\beta^n(0)$ is then $\left(\frac{1}{|B_n|}, \dots, \frac{1}{|B_n|}\right)$

so that

$$\begin{aligned}
H\left(\mathcal{C}_{3^n}\right) &= -\left|B_n\right|\left(\frac{1}{\left|B_n\right|}\log\frac{1}{\left|B_n\right|}\right) \\
&= \log\left|B_n\right| \\
&= \sum_{k=1}^{n}\log 2^{3^{k-1}} \\
&= (\log 2)\sum_{k=1}^{n}3^{k-1}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
h_\beta &= \lim_{n\to\infty}\frac{1}{3^n}H\left(\mathcal{C}_{3^n}\right) \\
&= (\log 2)\lim_{n\to\infty}\frac{1}{3^n}\sum_{k=1}^{n}3^{k-1} \\
&= (\log 2)\lim_{n\to\infty}\sum_{k=1}^{n}\frac{1}{3^k} \\
&= (\log 2)\sum_{k=1}^{\infty}\frac{1}{3^k} \\
&= \frac{1}{2}\log 2
\end{aligned}$$

by geometric series. Heuristically, this says that on average, every other symbol in an infinite string generated by β carries one bit of information (or that every symbol carries $\frac{1}{2}$ bits of information).

Note that we can obtain the formula for $|B_n|$ via the Set Representation $\mathcal{SR}_\infty^\beta$ of β. Let $a_n = |A_n|$, $b_n = |B_n|$, and $c_n = |C_n|$, where A_n, B_n, and C_n are the sets A_n^0, A_n^1, and A_n^2 in the definition of the set representation $\mathcal{SR}_\infty^\beta$ of β). So $A_0 = \{0\}$, $B_0 = \{1\}$, $C_0 = \{2\}$, and

$$\begin{aligned}
A_{n+1} &= A_nB_nA_n \cup A_nC_nA_n, \\
B_{n+1} &= B_nA_nB_n \cup B_nC_nB_n, \\
C_{n+1} &= C_nA_nC_n \cup C_nB_nC_n.
\end{aligned}$$

Then

$$\begin{cases} a_{n+1} = a_n^2\left(b_n + c_n\right) \\ b_{n+1} = b_n^2\left(a_n + c_n\right) \\ c_{n+1} = c_n^2\left(a_n + b_n\right) \end{cases},$$

where $a_0 = b_0 = c_0 = 1$. Noting that $a_n = b_n = c_n$, we obtain

$$a_{n+1} = 2a_n^3,$$

which has the closed form

$$\begin{aligned} a_n &= \prod_{k=1}^{n} 2^{3^{k-1}} \\ &= \sqrt{2}^{3^n-1}. \end{aligned}$$

Then we may compute the entropy h_β as above:

$$\begin{aligned} h_\beta &= \lim_{n\to\infty} \frac{1}{n} H\left(\mathcal{C}_n\right) \\ &= \lim_{n\to\infty} \frac{1}{3^n} H\left(\mathcal{C}_{3^n}\right) \\ &= -\lim_{n\to\infty} \frac{1}{3^n} \sum_{k=1}^{a_n} \frac{1}{a_n} \log \frac{1}{a_n} \\ &= -\lim_{n\to\infty} \frac{1}{3^n} \log \frac{1}{a_n} \\ &= \lim_{n\to\infty} \frac{1}{3^n} \log \sqrt{2}^{3^n-1} \\ &= \left(\frac{1}{2}\log 2\right) \lim_{n\to\infty} \frac{3^n-1}{3^n} \\ &= \frac{1}{2}\log 2. \end{aligned}$$

We can apply the ideas in this previous example to other random substitutions. We only needed to know two things: The lengths of elements of $\beta^n(0)$ and the probability distribution of $\beta^n(0)$.

Example 4.2.4 *Let ϕ be the random substitution*

$$\varphi : \begin{cases} 0 & \longmapsto & 0011 \\ 1 & \longmapsto & \begin{cases} 1001 & p \\ 0101 & q \end{cases} \end{cases}.$$

We will show that $h_\phi = \frac{1}{6}\log 2$. The set representation of φ is as follows: $A_0 = \{0\}$, $B_0 = \{1\}$, $A_{n+1} = A_n A_n B_n B_n$, and $B_{n+1} = B_n A_n A_n B_n \cup A_n B_n A_n B_n$. So if $a_n = |A_n|$ and $b_n = |B_n|$ for $n \geq 1$, then $a_{n+1} = a_n^2 b_n^2$ and $b_{n+1} = 2a_n^2 b_n^2$. Then

$$a_{n+1} = 4a_n^4$$

where $a_1 = 1$. Then it can be shown by induction that

$$a_n = 2^{\frac{2}{3}\left(4^{n-1}-1\right)}$$

for $n \geq 2$. So the probability distribution of $\varphi^n(0)$ is $\left(\frac{1}{a_n}, \ldots, \frac{1}{a_n}\right)$ (where this vector has a_n elements). Thus, the entropy of the probability distribution of $\varphi^n(0)$ is

$$\begin{aligned} H\left(\mathcal{C}_{4^n}\right) &= -\sum_{i=1}^{a_n} \frac{1}{a_n} \log \frac{1}{a_n} \\ &= \log a_n \\ &= \frac{2}{3}\left(4^{n-1}-1\right)\log 2. \end{aligned}$$

Note that every element of A_n has length 4^n for $n \geq 1$. Therefore,

$$\begin{aligned} h_\varphi &= \lim_{n\to\infty} \frac{1}{4^n} H\left(\mathcal{C}_{4^n}\right) \\ &= \left(\frac{2}{3}\log 2\right) \lim_{n\to\infty} \frac{1}{4^n}\left(4^{n-1}-1\right) \\ &= \left(\frac{2}{3}\log 2\right)\left(\frac{1}{4}\right) \\ &= \frac{1}{6}\log 2. \end{aligned}$$

Note that we can view this heuristically either as saying that on average every symbol in an infinite string generated by ϕ carries $\frac{1}{6}$ bits of information or that we gain information only about 16.7 % of the time.

Example 4.2.5 *Recall from Example 3.2.4 the 5-state substitution α defined by*

$$\alpha : \begin{cases} 0 & \longmapsto & \begin{cases} 10 & \frac{1}{2} \\ 02 & \frac{1}{2} \end{cases} \\ 1 & \longmapsto & 03 \\ 2 & \longmapsto & 04 \\ 3 & \longmapsto & 3 \\ 4 & \longmapsto & 4 \end{cases}$$

where we computed the frequency of 0 *to be We compute the entropy h_α. Let $\mathcal{SR}_\infty^\alpha$ be the set representation of α so that $A_0 = \{0\}$, $B_0 = \{1\}$, $C_0 = \{2\}$, $D_0 = \{3\}$, $E_0 = \{4\}$, and*

$$\begin{aligned} A_{n+1} &= B_n A_n \cup A_n C_n \\ B_{n+1} &= A_n D_n \\ C_{n+1} &= A_n E_n \\ D_{n+1} &= D_n \\ E_{n+1} &= E_n \end{aligned}$$

(As an aside, we will encounter this Example again when we talk about "recoverability"). Note that for each n, every word in A_n has the same length ℓ_n. Let $a_n = |A_n|$. Then since $|D_n| = |E_n| = 1$, by the definition of A_n above we see that $a_{n+1} = 2a_n a_{n-1}$ with $a_0 = 1$ and $a_1 = 2$. From this is follows (see SLOAN's online encyclopedia of integer sequences) that

$$2a_n = 2^{f_{n+2}}$$

where $\{f_n\}_{n=0}^\infty = 0, 1, 1, 2, 3, \ldots$ is the Fibonacci sequence (easily proved by induction). Note that since $f_{n+2} = f_{n+1} + f_n$, $lim_{n\to\infty} \frac{f_n}{f_{n+1}} = \Phi - 1$, where $\Phi = \frac{1+\sqrt{5}}{2} \approx 1.618$ is the

Golden Mean. We see from the definition of A_n that $\ell_n = f_{n+3} - 1$. Then we have that

$$\begin{aligned} h_\alpha &= \lim_{n\to\infty} \frac{1}{\ell_n} H\left(\mathcal{C}_{\ell_n}\right) \\ &= \lim_{n\to\infty} \frac{\log a_n}{\ell_n} \\ &= \lim_{n\to\infty} \frac{\log\left(\frac{1}{2}\cdot 2^{f_{n+2}}\right)}{f_{n+3}-1} \\ &= \lim_{n\to\infty} \frac{f_{n+2}\log 2}{f_{n+3}} \\ &= (\Phi - 1)\log 2. \end{aligned}$$

Example 4.2.6 *Define the substitution ρ_1 by*

$$\rho_1 : \begin{cases} 0 & \longmapsto \begin{cases} 01 & \textit{with probability } p \\ 02 & \textit{with probability } q \end{cases} \\ 1 & \longmapsto \qquad\qquad 00 \\ 2 & \longmapsto \qquad\qquad 00 \end{cases}.$$

We compute the entropy h_{ρ_1} for $p = q = \frac{1}{2}$. Let $\mathcal{SR}_\infty^{\rho_1}$ be the set representation of ρ_1 so that $A_0 = \{0\}$, $B_0 = \{1\}$, $C_0 = \{2\}$, and

$$\begin{aligned} A_{n+1} &= A_n B_n \cup A_n C_n \\ B_{n+1} &= A_n A_n \\ C_{n+1} &= A_n A_n \end{aligned}$$

So

$$A_{n+1} = A_n A_{n-1} A_{n-1}.$$

Let $a_n = |A_n|$. We must be careful in using the set representation of ρ_1 to calculate a_n since iterates of ρ_1 may produce duplicate words. For example$\rho_1(01)$ and $\rho_1(02)$ produce the same words 0100 *and* 0200 *(and no more). We therefore argue combinatorially to calculate a_n. Note that each* 0 *in a word in A_n produces two distinct words in A_{n+1}. Also, each word in A_n has the same number of zeros z_n and produce the same words in A_{n+1}. Therefore, $a_{n+1} = 2^{z_n}$. The incidence matrix of ρ_1 is*

$$M = \begin{bmatrix} 1 & 2 & 2 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix}$$

and so, by diagonalization,

$$M^n = \begin{bmatrix} \frac{1}{3}\left(2^{n+1} + (-1)^n\right) & \frac{2}{3}\left(2^n - (-1)^n\right) & \frac{2}{3}\left(2^n - (-1)^n\right) \\ \frac{1}{6}\left(2^n - (-1)^n\right) & \frac{1}{3}\left(2^{n-1} + (-1)^n\right) & \frac{1}{3}\left(2^{n-1} + (-1)^n\right) \\ \frac{1}{6}\left(2^n - (-1)^n\right) & \frac{1}{3}\left(2^{n-1} + (-1)^n\right) & \frac{1}{3}\left(2^{n-1} + (-1)^n\right) \end{bmatrix}.$$

Thus, as in Remark 3.0.6,

$$z_n = \frac{2^{n+1} + (-1)^n}{3}.$$

Since each element of A_n has length 2^n, we have that

$$\begin{aligned} h_{\rho_1} &= \lim_{n\to\infty} \frac{1}{2^n} H\left(\mathcal{C}_{2^n}\right) \\ &= \lim_{n\to\infty} \frac{\log a_n}{2^n} \\ &= \left(\frac{1}{3}\log 2\right) \lim_{n\to\infty} \frac{2^{n+1} + (-1)^n}{2^n} \\ &= \frac{2}{3}\log 2. \end{aligned}$$

Heuristically, this says that $\frac{2}{3}$ of the symbols in an infinite string u generated by ρ_1 contains a binary choice (ie, information). Note that the normalized Perron-Frobenius eigenvector (see 3.2.2) of M is $\left(\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)$ (corresponding to the eigenvalue $\lambda = 2$, the length of ρ_1), signifying that $\frac{2}{3}$ of the symbols in u is a 0, the only symbol that yields a choice.

Example 4.2.7 *Define the substitution ρ_2 by*

$$\rho_2 : \begin{cases} 0 & \longmapsto \begin{cases} 01 & \text{with probability } p \\ 02 & \text{with probability } q \end{cases} \\ 1 & \longmapsto \quad 0 \\ 2 & \longmapsto \quad 0 \end{cases}.$$

As usual, we will assume $p = q = \frac{1}{2}$. We would expect the entropy h_{ρ_2} to be greater than h_{ρ_1} in the previous example since "less space is used up" by $\rho_2(1)$ and $\rho_2(2)$. This time, *we have that*

$$\begin{aligned} A_{n+1} &= A_n B_n \cup A_n C_n \\ &= A_n A_{n-1} \end{aligned}$$

so that $a_0 = 1$ *and* $a_{n+1} = a_n a_{n-1}$. *Let* $f_n = f_{n-1} + f_{n-2}$ *be the Fibonacci sequence (with* $f_0 = 0,\ f_1 = 1$*). We see from the definition of* A_{n+1} *that the length of any word in* A_n *is* f_{n+2}. *Now*

$$\log a_{n+1} = \log a_n + \log a_{n-1}$$

so that $\{\log a_n\}_{n\geq 0}$ *is the Fibonacci sequence. Thus,* $\log a_n = f_n$ *so that*

$$a_n = 2^{f_n}.$$

We can alternatively reason combinatorially as in the previous example: Since every 0 *in a word in* A_n *produces two words in* A_{n+1} *and all words in* A_n *produce the same words in* A_{n+1}, $a_{n+1} = 2^{f_n}$. *Thus,*

$$\begin{aligned} h_{\rho_2} &= \lim_{n\to\infty} \frac{\log a_n}{f_{n+2}} \\ &= \lim_{n\to\infty} \frac{f_n \log 2}{f_{n+2}} \\ &= \frac{1}{\Phi^2} \log 2 \\ &\approx 0.382 \log 2, \end{aligned}$$

where $\Phi = \frac{1+\sqrt{5}}{2} \approx 1.618$ *is the Golden Mean. Note that* $h_{\rho_2} > h_{\rho_1}$, *as expected.*

Example 4.2.8 ((p,q)-"Recoverable" constant-length FibMorse) *Define the substitution* ζ *by*

$$\zeta : \begin{cases} 0 \longmapsto & 01 \\ 1 \longmapsto & \begin{cases} 10 & \textit{with probability } p \\ 20 & \textit{with probability } q \end{cases} \\ 2 \longmapsto & 22 \end{cases}.$$

Suppose that $p = q = \frac{1}{2}$. *Let* $\mathcal{SR}_\infty^\zeta$ *be the set representation of* ζ, *so that* $A_0 = \{0\}$, $B_0 = \{1\}$, $C_0 = \{2\}$, *and*

$$\begin{aligned} A_{n+1} &= A_n B_n \\ B_{n+1} &= C_n A_n \cup B_n A_n \\ C_{n+1} &= C_n C_n \end{aligned}$$

Let $a_n = |A_n|$, $b_n = |B_n|$, *and* $c_n = |C_n|$. *Then,* $a_0 = b_0 = c_0 = 1$ *and*

$$\begin{aligned} a_{n+1} &= a_n b_n \\ b_{n+1} &= a_n b_n + a_n c_n \\ c_{n+1} &= 1 \end{aligned}$$

for $n \geq 0$. *This gives us that*

$$\begin{aligned} a_{n+1} &= a_n \left(a_{n-1} b_{n-1} + a_{n-1} c_{n-1}\right) \\ &= a_n \left(a_n + a_{n-1}\right). \end{aligned}$$

Therefore, we have the recurrence

$$a_{n+2} = a_{n+1} \left(a_{n+1} + a_n\right) \tag{4.2.6}$$

for all $n \geq 2$ *and* $a_0 = a_1 = 1$. *We apply techniques from [2] to find an asymptotic solution to this recursion and thus get a formula for and approximate the topological entropy* h_ζ. *Let*

$$y_n = \log a_n$$

(where the $\log$ *is taken to the base 2) and*

$$L_n = \log\left(1 + \frac{a_n}{a_{n+1}}\right)$$

for $n \geq 0$. *Note that by the recurrence (4.2.6),*

$$\frac{a_n}{a_{n+1}} = \frac{1}{a_n + a_{n-1}}$$

so that $\lim_{n\to\infty} L_n = 0$.

Using our recurrence (4.2.6), we have that

$$\begin{aligned} 2y_n + L_{n-1} &= \log a_n^2 + \log\left(1 + \frac{a_n a_{n-1}}{a_n^2}\right) \\ &= \log\left[a_n^2\left(1 + \frac{a_{n+1} - a_n^2}{a_n^2}\right)\right] \\ &= \log a_{n+1} \\ &= y_{n+1} \end{aligned}$$

so that

$$y_{n+1} = 2y_n + L_{n-1} \tag{4.2.7}$$

for all $n \geq 1$ *and* $y_0 = \log a_0 = 0$. *Note that the recurrence (4.2.7) has the solution*

$$\begin{aligned} y_n &= 2^{n-1}\left(\frac{L_0}{2^1} + \frac{L_1}{2^2} + \frac{L_2}{2^3} + \cdots + \frac{L_{n-2}}{2^{n-1}}\right) \\ &= \sum_{j=1}^{n-1} 2^{n-j-1} L_{j-1}. \end{aligned}$$

for $n \geq 2$. *Indeed,*

$$2\left(\sum_{j=1}^{n-1} 2^{n-j-1} L_{j-1}\right) + L_{n-1} = \left(\sum_{j=1}^{n-1} 2^{n-j} L_{j-1}\right) + L_{n-1}$$

whereas

$$\begin{aligned} \sum_{j=1}^{(n+1)-1} 2^{(n+1)-j-1} L_{j-1} &= \sum_{j=1}^{n} 2^{n-j} L_{j-1} \\ &= \left(\sum_{j=1}^{n-1} 2^{n-j} L_{j-1}\right) + 2^{n-n} L_{n-1} \\ &= \left(\sum_{j=1}^{n-1} 2^{n-j} L_{j-1}\right) + L_{n-1}. \end{aligned}$$

Thus, $2y_n + L_n = y_{n+1}$. *We note in passing that this solution did not depend on* L_n *and we thus have a solution to any equation of the form (4.2.7) (see [22], P. 26). Since each* *word in A_n has length $\ell_n = 2^n$, we therefore have that*

$$
\begin{aligned}
h_\zeta &= \lim_{n\to\infty} \frac{\log a_n}{\ell_n} \\
&= \lim_{n\to\infty} \frac{y_n}{2^n} \\
&= \lim_{n\to\infty} \frac{1}{2^n} \sum_{j=1}^{n-1} 2^{n-j-1} L_{j-1} \\
&= \sum_{j=1}^{\infty} 2^{-j-1} L_{j-1},
\end{aligned}
$$

which is positive. We compute the first few partial sums $S_n = \sum_{j=1}^{n} 2^{-j-1} L_{j-1}$ of this series:

n	0	1	2	3	4	5	6	7	8
S_n	–	0.25	0.323 12	0.349 06	0.354 37	0.354 78	0.354 79	0.354 79	0.354 79

We then speculate that the entropy is (with log *base* 2*)*

$$h_\zeta \approx 0.3548 \log 2.$$

Theorem 4.2.2 *The entropy $h_\mu(X(\zeta), \sigma)$ of the dynamical system arising from the $\left(\frac{1}{2}, \frac{1}{2}\right)$-FibMorse substitution ζ is positive.*

Proof. The set representation of ζ is $A_0 = \{0\}$, $B_0 = \{1\}$, $A_{n+1} = A_n B_n$, and $B_{n+1} = B_n A_n \cup A_n$. Let $a_n = |A_n|$ and $b_n = |B_n|$ for $n \geq 1$. Then $a_{n+1} = a_n b_n$ and $b_{n+1} = a_n b_n + a_n$ so that

$$
\begin{aligned}
a_{n+1} &= a_n \left(a_{n-1} b_{n-1} + a_{n-1}\right) \\
&= a_n \left(a_n + a_{n-1}\right).
\end{aligned}
$$

This is the same recursive formula as in the Constant-Length FibMorse substitution above (Example 4.2.8), where we obtained that

$$\lim_{n\to\infty} \frac{1}{2^n} H\left(\mathcal{C}_{2^n}\right) = \sum_{j=1}^{\infty} \frac{1}{2^{j+1}} \log\left(1 + \frac{a_{j-1}}{a_j}\right).$$

Notice that the partial sums of this series form an increasing sequence since $\log\left(1+\frac{a_{j-1}}{a_j}\right) > 0$, so this series is positive. Now, if $x \in A_n$, then $f_n \le |x| \le 2^n$, where f_n is the nth Fibonacci number. So $h_\zeta \ge \lim_{n\to\infty} \frac{1}{2^n} H\left(\mathcal{C}_{2^n}\right) > 0$. ■

4.3 Topological Entropy

In [1], the authors generalize the notion of entropy to topological dynamical systems (X, T), where X is compact Hausdorff and T is a homeomorphism (see also [20] for an exposition). Notice that if the partition $\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})$ of X contains many small sets, then the quantity $H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})\right)$ is large. We cannot necessarily measure the "size" of partition elements in an arbitrary topological space, but we can count how many there are. Thus, instead of partitions, consider an open cover $\mathcal{U}$ of X. Since X is compact, there is a subcover of $\mathcal{U}$ of minimal cardinality. Denote this cardinality by $N(\mathcal{U})$ and let

$$H(\mathcal{U}) = \log_2 N(\mathcal{U}).$$

Defining the **join** of open covers $\mathcal{U}$ and $\mathcal{V}$ to be $\mathcal{U} \vee \mathcal{V} = \{U \cap V : U \in \mathcal{U}, V \in \mathcal{V}\}$, we may let

$$h(\mathcal{U}, T) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{U})\right).$$

Proposition 4.3.1 *The limit $h(\mathcal{U}, T)$ exists.*

Proof. *Since $\mathcal{U} \vee \mathcal{V} = \{U \cap V : U \in \mathcal{U}, V \in \mathcal{V}\}$, $N(\mathcal{U} \vee \mathcal{V}) \le N(\mathcal{U}) N(\mathcal{V})$. So $H(\mathcal{U} \vee \mathcal{V}) \le H(\mathcal{U}) + H(\mathcal{V})$. Hence, Lemma 4.2.1 applies.* ■

Definition 4.3.1 *Let X be compact Hausdorff and T be a homeomorphism. The* ***topological entropy*** *of the topological dynamical system (X, T) is*

$$h_{top}(X, T) = \sup_{\mathcal{U}} h(\mathcal{U}, T),$$

where the supremum is taken over all open covers $\mathcal{U}$ of X. If $h_{top}(X,T) = 0$, then we call the dynamical system (X,T) ***deterministic****.*

Just as generating partitions are useful for computing metric entropy, "refining sequences" of open covers are useful for computing topological entropy. We say that an open cover $\mathcal{U}$ **refines** an open cover $\mathcal{V}$ if every $U \in \mathcal{U}$ is contained in some $V \in \mathcal{V}$. In this case, we write $\mathcal{U} \geq \mathcal{V}$. We also say that $\mathcal{U}$ is a **refinement** of $\mathcal{V}$ and that $\mathcal{V}$ is **refined by** $\mathcal{U}$.

Definition 4.3.2 *Let (X,T) be a topological dynamical system. Suppose $\{\mathcal{U}_n\}_n$ is a set of open covers of X such that $\mathcal{U}_n \leq \mathcal{U}_{n+1}$ for all n. Then $\{\mathcal{U}_n\}_n$ is a* ***refining sequence*** *of X.*

For example, the cylinder sets in a symbolic dynamical system form a refining sequence.

Proposition 4.3.2 *If $\mathcal{V} \leq \mathcal{U}$, then $H(\mathcal{V}) \leq H(\mathcal{U})$ so that $h(\mathcal{V},T) \leq h(\mathcal{U},T)$.*

Proposition 4.3.3 *Suppose that $\{\mathcal{U}_n\}_n$ is a refining sequence of a compact space X such that any finite open cover is refined by some $\mathcal{U}_n$. Then*

$$h_{top}(X,T) = \lim_{n\to\infty} h(\mathcal{U}_n,T).$$

Proof. Since the entropy of a partition increases with refinement (Proposition 4.3.2), $h_{top}(X,T)$ is the limit of the net $\{h(\mathcal{C}_\alpha,T)\}_{\alpha\in J}$, where the index set J is the collection of all open covers of X. Since X is compact, each $\mathcal{C}_\alpha$ has a finite subcover $\mathcal{V}_\alpha$ which is refined by some $\mathcal{U}_n$ (by hypothesis). So $\mathcal{U}_n$ refines $\mathcal{C}_\alpha$. Therefore, since $\{\mathcal{U}_n\}_n$ is a refining sequence, $h_{top}(X,T) = \lim_{n\to\infty} h(\mathcal{U}_n,T)$. ■

Example 4.3.1 *Let* $(X,\sigma) \subseteq \{0,1\}^{\mathbf{Z}_+}$ *be a symbolic dynamical system. Since the more words of length* n *there are, the more complicated the symbolic system is, it would make sense that the topological entropy of* (X,σ) *should increase with the number of words. This is indeed the case, as we now show. Let* $\mathcal{U}_n$ *be the collection of cylinder sets of the form*

$$[w]_0 = \{x \in X : x_0 x_1 \cdots x_{n-1} = w_0 w_1 \cdots w_{n-1}\},$$

where $w_i \in \{0,1\}$. *That is,* $\mathcal{U}_n = \{[w]_0 : |w| = n\}$ *(so, for example,* $\mathcal{U}_0$ *consists of the time-zero sets). Then* $\{\mathcal{U}_n\}_n$ *is a refining sequence of* X *and*

$$\mathcal{U}_n = \bigvee_{i=0}^{n-1} \sigma^{-i}(\mathcal{U}_0).$$

We now show that $h(\mathcal{U}_n,\sigma) = h(\mathcal{U}_0,\sigma)$ *for all* n. *Indeed,*

$$\begin{aligned}
h(\mathcal{U}_0,\sigma) &\leq h(\mathcal{U}_n,\sigma)\, Prop.4.3.2 && (4.3.1)\\
&= \lim_{k\to\infty} \frac{1}{k} H\left(\bigvee_{i=0}^{k-1} \sigma^{-i}(\mathcal{U}_n)\right)\\
&= \lim_{k\to\infty} \frac{1}{k} H\left(\bigvee_{i=0}^{n+k-2} \sigma^{-i}(\mathcal{U}_0)\right)\\
&\leq \lim_{k\to\infty} \frac{1}{k} H\left(\bigvee_{i=0}^{k-1} \sigma^{-i}(\mathcal{U}_0)\right) Prop.4.3.2 && (4.3.2)\\
&= h(\mathcal{U}_0,\sigma).
\end{aligned}$$

Therefore,

$$\begin{aligned}
h_{top}(X,\sigma) &= \lim_{n\to\infty} h(\mathcal{U}_n,\sigma)\, Prop.4.3.3 && (4.3.3)\\
&= \lim_{n\to\infty} h(\mathcal{U}_0,\sigma)\\
&= h(\mathcal{U}_0,\sigma)\\
&= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} \sigma^{-i}(\mathcal{U}_0)\right)\\
&= \lim_{n\to\infty} \frac{1}{n} H(\mathcal{U}_n)\\
&= \lim_{n\to\infty} \frac{\log_2 N(\mathcal{U}_n)}{n}.
\end{aligned}$$

Now, $N(\mathcal{U}_n)$ is the (least) number of open sets needed to cover $\mathcal{U}_n$. But $\mathcal{U}_n$ consists of the length-n cylinders, which form a basis for the topology of X (as noted after Definition 2.0.4). So $N(\mathcal{U}_n)$ is the number of admissible words of length n in X.

Recalling the definition 2.0.8 of the complexity function $p(n)$, Example 4.3.1 motivates the following definition.

Definition 4.3.3 *The **topological entropy** of a symbolic dynamical system (X, σ) is given by*

$$h_{top}(X, \sigma) = \lim_{n \longrightarrow \infty} \frac{\log_{|\mathcal{A}|} p(n)}{n}.$$

One may wonder if there is a connection between metric and topological entropy. There is a general "Variational Principle" that says that, for a dynamical system (X, T), $h_{top}(X, T) = \sup h_\mu$, where the supremum is taken over all T-invariant probability measures on X, so that $h_{top} \geq h_\mu$ for all μ. This inequality is easily seen to be true in the case of a symbolic dynamical system (a general proof can be found in [20]).

Theorem 4.3.1 *Let (X, σ) be a symbolic dynamical system on the alphabet $\mathcal{A} = \{0, \ldots, s-1\}$, with σ the shift on X. Then $h_\mu(X, \sigma) \leq h_{top}(X, \sigma)$ for all μ.*

Proof. As noted above, the quantity

$$H(\mathcal{Q}) := -\sum_{Q \in \mathcal{Q}} \mu Q \log \mu Q$$

is maximal when each $Q \in \mathcal{Q}$ occurs with the same frequency. So if μ is a probability measure on X, then h_μ is a maximum when each every word w of length n occurs with the same frequency $\frac{1}{p(n)}$ for all n. Let μ be the σ-invariant probability measure on X given by $\mu[w] = \frac{1}{p(n)}$ for all length n cylinder sets $[w]$. Then μ is a measure of maximal entropy. Since the time-zero sets $\mathcal{P} = \{[0]_0, [1]_0, \ldots [s-1]_0\}$ generate and $\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})$ is the set of all length-n cylinder sets $\mathcal{C}_n$, by the Kolmogorov-Sinai Theorem 4.2.1 we have

$$\begin{aligned} h_\mu &= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})\right) \\ &= \lim_{n\to\infty} \frac{1}{n}\left[-\sum_{C\in\mathcal{C}_n} \mu(C)\log\mu(C)\right] \\ &= \lim_{n\to\infty} \frac{1}{n} p(n)\left[\frac{1}{p(n)}\log p(n)\right] \\ &= h_{top}. \end{aligned}$$

■

We note that the entropy calculations in Section 4.2 reduced to calculating h_{top}. This is because of the assumption that $p = q = \frac{1}{2}$ in all the random substitution examples, which gave a uniform distribution to a the random variable in question. A uniform distribution is essentially how the entropy was maximized in the preceding Theorem 4.3.1.

Sturmian Sequences

It is well known that the Fibonacci sequence that arises from the Fibonacci substitution has complexity function $p(n) = n + 1$ for all n (see, for example, [4]). We now digress to mention certain interesting aperiodic sequences that exhibit a minimal complexity. In [9], Hedlund and Morse initiated a study of these sequences, called *Sturmian sequences*. These sequences have been greatly studied and many properties and equivalent representations are now known (see [3] and [4] for recent surveys). The most basic of these involves the complexity function $p(n)$. We saw in Proposition 2.0.1 that any aperiodic sequence satisfies $p(n) \geq n + 1$ for all n. The aperiodic sequences with minimal complexity are called Sturmian.

Definition 4.3.4 *A* ***Sturmian sequence*** *u is a sequence such that $p_u(n) = n + 1$ for all n.*

This was not the original definition Hedlund and Morse gave in [9], which was more of a “balancing” nature. The (equivalent) definition we give here was given 30 years later by Coven and Hedlund in [5]. A few basic properties of Sturmian sequences are:

Proposition 4.3.4 *Let u be a Sturmian sequence on an alphabet $\mathcal{A}$. Then*

(a) $\mathcal{A} = \{0,1\}$*;*

(b) any admissible word is recurrent (occurs an infinite number of times) in u; and

(c) either 00 *or* 11 *is forbidden, but not both.*

Proof. For (a), simply note that $p(1) = 2$.
For (b), suppose there is a word w that occurs a finite number of times in u. Then there is an N such that w occurs in $u_0u_1 \ldots u_N$ but does not occur in $v = u_{N+1}u_{N+2} \ldots$. Now, the set Ω_v of all admissible words of v is contained in the set Ω_u of all admissible words of u and $w \notin \Omega_v$ so that $p_v(n) \leq n$ (since $p_u(n) = n+1$ for all n). But this implies that v, hence u, is eventually periodic, by Proposition 2.0.1. Therefore, every word in u occurs an infinite number of times.
For (c), we know that exactly one of 00, 01, 10, or 11 is forbidden since $p(2) = 3$. By (a), both 0 and 1 occur in u and so by (b), they both occur an infinite number of times in u. So 01 and 10 are both admissible. ■

Proposition 4.3.5 ([4], Prop. 2.1.2) *If $u \in \{0,1\}^{\mathbf{Z}_+}$ is balanced, then $p(n) \leq n+1$.*

Proof. Since u is balanced, 00 and 11 cannot both be in u. So the claim is true for $n = 2$. By way of contradiction, suppose there is some $n_0 \geq 3$ such that $p(n_0) \geq n_0 + 2$ and assume n_0 is the smallest such. Then $p(n_0 - 1) \leq n_0$. Since there are at least $n_0 + 2$ words of length n_0 and at most n_0 words of length $n_0 - 1$, there are at least two distinct words w, w' of length $n_0 - 1$ that can be extended on the left in more than one way. So $0w$, $1w$, $0w'$, and $1w'$ are all distinct words of length n_0 in u. Let K be the smallest such that $x = w_0w_1 \ldots w_K = w'_0w'_1 \ldots w'_K$ but $w_{K+1} \neq w'_{K+1}$, say $w_{K+1} = 0$ and $w'_{K+1} = 1$. Then $0x0$ and $1x1$ are both words in u, contradicting that u is balanced. ■

Since aperiodic sequences satisfy $p(n) \geq n+1$ (see 2.0.1), this leads to an interesting characterization of Sturmian sequences is the following (see, for example, [4] for a proof):

Theorem 4.3.2 *A sequence u on $\{0, 1\}$ is Sturmian if and only if it is aperiodic and balanced.*

So Sturmian sequences are exactly the aperiodic sequences with minimal complexity. For example, if (X, σ) is the symbolic dynamical system arising from any Sturmian sequence, then $h_{top}(X, \sigma) = 0$.

Two interesting geometrical interpretations of Sturmian sequences in the they arise from irrational rotations and from irrational slopes of lines it 2-space (see, for example [4] and [14]):

Proposition 4.3.6 (Hedlund and Morse [9]) *A sequence u is Sturmian if and only if u is the coding of an irrational rotation of the circle $S^1 = \mathbf{R}/\mathbf{Z}$.*

Proposition 4.3.7 *A sequence u is Sturmian if and only if u is the "cutting" sequence of a line $y = \alpha x + b$ with irrational slope α.*

There are many other interesting (and sometimes surprising) properties of Sturmian sequences and the reader is encouraged to read the aforementioned references.

4.4 Topological Entropy of Substitutions

If an infinite string is generated by a (non-random) substitution, then heuristically it would seem that this string is deterministic in the sense that we know (theoretically) exactly what symbol will occur and where. That is, there is no information gained in reading a symbol in the string. Thus, the entropy of a substitution dynamical system should be zero. This is a Corollary of the following Proposition.

Proposition 4.4.1 *Let ω be an irreducible substitution on a finite alphabet $\mathcal{A} = \{0, \ldots, s-1\}$. Then there is a constant C such that $p(n) \leq Cn$ for all n.*

We first provide a proof for the constant length case to illustrate the ideas. So let ω be a substitution of constant length ℓ. Let u be a string generated by ω. That is, u is a fixed point of ω^{∞}. If w is an admissible word of length ℓ^k, then $w \in u = \omega^m(u)$ for every $m \geq k$. Then, since ω is irreducible, either w is contained in some $\omega^m(a)$ or $\omega^m(ab)$, where $a, b \in \mathcal{A}$. Since there are $p(2)$ words ab and at most ℓ^k words of length ℓ^k, we therefore have that

$$\begin{aligned} p\left(\ell^k\right) &\leq p(2)\,\ell^k \\ &\leq s^2\ell^k. \end{aligned} \tag{4.4.1}$$

where $s = |\mathcal{A}|$. Now suppose that $n \geq 1$. Then there is some n such that

$$\ell^{k-1} \leq n \leq \ell^k.$$

Then

$$\begin{aligned} p(n) &\leq p\left(\ell^k\right) \\ &\leq s^2\ell^k \\ &\leq s^2\ell n \end{aligned}$$

so that $p(n) \leq Cn$ where $C = s^2\ell$ (note we could also let $C = p(2)\,\ell$).

For the non-constant length case, the Perron-Frobenius Theorem 3.2 essentially provides an “average length” to the substitution so as to allow it to be treated as a constant-length substitution. We now follow the general proof in [21] (Proposition 5.19) to prove Proposition 4.4.1.

Proof. Since the sequence $\left\{\min_{a\in\mathcal{A}}\left|\zeta^j(a)\right|\right\}_{j=1}^{\infty}$ is nondecreasing, for every $n \geq 1$ we can find a $j_n \geq 1$ such that

$$\min_{a\in\mathcal{A}}\left|\zeta^{j_n-1}(a)\right| \leq n \leq \min_{a\in\mathcal{A}}\left|\zeta^{j_n}(a)\right|.$$

So every word w of length n is contained either in $\zeta^{j_n}(a)$ or $\zeta^{j_n}(ab)$ for some $a, b \in \mathcal{A}$. Now by the limit 3.2.1, $\lim_{j\to\infty}\frac{\left|\zeta^j(a)\right|}{\lambda_{PF}^j} = c_a > 0$ and so for j large enough, there are constants $\alpha_1, \alpha_2 > 0$ such that

$$\alpha_1\lambda_{PF}^j \leq \min_{a\in\mathcal{A}}\left|\zeta^j(a)\right| \leq \max_{a\in\mathcal{A}}\left|\zeta^j(a)\right| \leq \alpha_2\lambda_{PF}^j.$$

Now, fix n and j large enough so as to satisfy these last two inequalities. Then $\lambda^{j-1} \leq \frac{n}{\alpha_1}$. There are at most s^2 words of length 2 in $u = \zeta^{\infty}(0)$ and at most $\max_{a\in\mathcal{A}}\left|\zeta^j(a)\right|$ words of length n (that starts in some $\zeta^j(a)$ and is therefore contained in some $\zeta^j(ab)$). Therefore,

$$\begin{aligned} p(n) &\leq s^2 \cdot \max_{a\in\mathcal{A}}\left|\zeta^j(a)\right| \\ &\leq s^2 \cdot \alpha_2\lambda_{PF}^j \\ &\leq s^2 \cdot \alpha_2\lambda_{PF}\frac{n}{\alpha_1}. \end{aligned}$$

So let $C = \frac{s^2\alpha_2\lambda_{PF}}{\alpha_1}$. ■

Corollary 4.4.1 *The dynamical system arising from an irreducible substitution on a finite alphabet is deterministic.*

Proof. By the Proposition, we have that

$$\begin{aligned} h_{top}(X,\sigma) &= \lim_{n\longrightarrow\infty}\frac{\log_{|\mathcal{A}|}p(n)}{n} \\ &\leq \lim_{n\longrightarrow\infty}\frac{\log_{|\mathcal{A}|}Cn}{n} \\ &= \lim_{n\longrightarrow\infty}\left(\frac{\log_{|\mathcal{A}|}C}{n} + \frac{\log_{|\mathcal{A}|}n}{n}\right) \\ &= 0 \end{aligned}$$

by L'Hôpital's Rule. ■

Example 4.4.1 *Recall the "Recoverable" constant length FibMorse substitution in Example 4.2.8. In that example, we approximated the entropy to be* $h_\zeta \approx 0.3548 \log 2$*. In doing so, we have also approximated values of the complexity function at* 2^n*. For large* n*,*

$$h_\zeta \approx \frac{\log a_n}{2^n}$$

so that

$$\begin{aligned} p\left(2^n\right) &= a_n \\ &= 2^{h_\zeta \cdot 2^n} \\ &\approx 1.2788^{2^n}. \end{aligned}$$

We note that although in the original Example 4.2.8 we could have approximated h_ζ *from the start via the sequence* $\frac{\log_2 a_n}{2^n}$*, the calculations above give formulas for* h_ζ *and* $p\left(2^n\right)$*.*

Recall from examples in section 4.2 that the entropy of random substitutions can easily be positive, meaning that the complexity function is exponential. Empirically, we find that the complexity function for the FibMorse substitution is

$$p\left(n\right) = 2, 4, 7, 12, 20, 33, 52, 82, 128, 197, 302, \ldots,$$

which does not seem linear (it is interesting to note that the differences are the Fibonacci numbers for the first 5 differences). One might wonder why the above Proposition does not apply to random substitutions. The problem arises in the fact that for a random substitution ζ, there may be more than $\max_\alpha \left|\zeta^k\left(\alpha\right)\right|$ words of length n. That is, we cannot control the upper bound on $p\left(n\right)$. For example, if ζ is the FibMorse substitution, then we can take, say, $n = 4$ and $k = 3$ in the proof of Proposition 4.4.1. But $p\left(4\right) = 12 > \max_\alpha \left|\zeta^3\left(\alpha\right)\right|$.

5 ENTROPY, BALANCEDNESS, AND RECOVERABILITY

5.1 Balancedness

In this chapter, we explore another way to view how complicated a space of sequences is. This notion extends the previously known concept of balancedness.

In Theorem 4.3.2, we saw that an aperiodic balanced sequence was Sturmian, hence has entropy zero and is deterministic. We generalize the concept of balancedness to obtain dynamical systems with positive entropy. Define the **height** $h(w)$ of a word $w \in \{0,1\}^*$ to be the number of 1's that occur in w and define the **height distance** between two words x and y of the same length to be $\delta(x,y) = |h(x) - h(y)|$.

Definition 5.1.1 *Let $k \geq 1$ be an integer. A sequence $u \in \{0,1\}^{\mathbf{Z}_+}$ is k-**balanced** if for any integer $n \geq 1$ and any two words x and y of length n, we have $\delta(x,y) \leq k$. If $k = 1$, then u is also called **balanced**. A random substitution ρ is k-**balanced** is any sequence arising from ρ is k-balanced. If a sequence or random substitution is not k-balanced for any k, then it is **unbalanced**.*

In the literature, "k-balancedness" refers to balancedness for sequences on alphabets having k symbols. At the time of this paper, the author knows of no notion of k-balancedness as presented here.

Example 5.1.1 *The Fibonacci sequence $u = 0100101001\cdots$ is balanced (see [4]).*

Example 5.1.2 *Let ρ be the random substitution*

$$\rho : \begin{cases} 0 \longmapsto & 0011 \\ 1 \longmapsto & \begin{cases} 1001 & p \\ 0101 & q \end{cases} \end{cases} .$$

We show that ρ is 2-balanced. First note that we may write any ρ-admissible w in the form

$$w = a_1 a_2 a_3 \rho(x) b_1 b_2 b_3$$

for some ρ-admissible word x (possibly empty) and $a_i, b_i \in \{\varepsilon, 0, 1\}$ (where ε is the empty word). Indeed, suppose $w = a_1 a_2 a_3 a' \rho(x) b_1 b_2 b_3$ with each $a_i \neq \varepsilon$. Then $a_1 a_2 a_3 a' = \rho(j)$ for some $j \in \{0, 1\}$ so that $w = \rho(x') b_1 b_2 b_3$ with $x' = jx$. Similarly if $w = a_1 a_2 a_3 \rho(x) b_1 b_2 b_3 b'$. Since $\rho(0)$ and $\rho(1)$ both end in 1, $a_3 = 1$. Also, both $a_1 a_2 a_3$ and $b_1 b_2 b_3$ end and start a substituted symbol, respectively. Because of this, $h(a_1 a_2) \in \{0, 1\}$ and $h(b_1 b_2 b_3) \in \{0, 1\}$. Finally, note that $h(\rho(j)) = 2$ for $j \in \{0, 1\}$ so that $h(x) = 2|x|$. Thus, if $v = c_1 c_2 c_3 \rho(y) d_1 d_2 d_3$ is ρ-admissible where $|y| = |x|$, then

$$\begin{aligned} \delta(w, v) &= |h(a_1 a_2 a_3 \rho(x) b_1 b_2 b_3) - h(c_1 c_2 c_3 \rho(y) d_1 d_2 d_3)| \\ &\leq |h(a_1 a_2) - h(c_1 c_2)| + |h(x) - h(y)| + |h(b_1 b_2 b_3) - h(d_1 d_2 d_3)| \\ &= 2. \end{aligned}$$

This example is a specific case of the more general:

Proposition 5.1.1 *Let ρ be a random substitution*

$$\rho : \begin{cases} 0 & \longmapsto \quad z \\ 1 & \longmapsto \begin{cases} s_1 & p \\ s_2 & q \end{cases} \end{cases}$$

of constant length ℓ. Suppose that $g = h(z) = h(s_1) = h(s_2)$. Then ρ is $2(\ell - g)$-balanced (this may not be a minimal upper bound).

Proof. *Arguing as in Example 5.1.2, any ρ-admissible word w can be written*

$$w = a_1 \ldots a_{\ell-1} \rho(x) b_1 \ldots b_{\ell-1}$$

where $a_i, b_i \in \{\varepsilon, 0, 1\}$ and x is a (possibly empty) ρ-admissible word. Since $a_1 \ldots a_{\ell-1}$ and $b_1 \ldots b_{\ell-1}$ end and start a substituted symbol, respectively, $h(a_1 \ldots a_{\ell-1}), h(b_1 \ldots b_{\ell-1}) \in$

$\{g, g-1\}$. *If* $v = c_1 \dots c_\ell \rho(y) d_1 \dots d_\ell$ *is* ρ*-admissible with* $|v| = |w|$, *then* $|y| = |x|$ *and* $h(\rho(x)) = h(\rho(y)) = g|x|$. *Since* $g = h(z) = h(s_1) = h(s_2)$,

$$\begin{aligned} \delta(w, v) &\leq |h(a_1 \dots a_{\ell-1}) - h(c_1 \dots c_{\ell-1})| + |h(\rho(x)) - h(\rho(y))| + |h(b_1 \dots b_{\ell-1}) - h(d_1 \dots d_{\ell-1})| \\ &\leq 2(\ell - g). \end{aligned}$$

■

Example 5.1.3 *Let* τ *be the substitution on* $\{0, 1\}$ *given by*

$$\tau : \begin{cases} 0 & \longmapsto & 0100 \\ 1 & \longmapsto & 11 \end{cases}$$

τ *has entropy zero by Corollary 4.4.1. We show that* τ *is unbalanced by showing that* δ *grows without bound. The incidence matrix of* τ *is*

$$M = M(\tau) = \begin{bmatrix} 3 & 0 \\ 1 & 2 \end{bmatrix}$$

and so by diagonalization,

$$M^n = \begin{bmatrix} 3^n & 0 \\ 3^n - 2^n & 2^n \end{bmatrix}.$$

So $h(\tau^n(0)) = 3^n - 2^n$ *and* $h(\tau^n(1)) = 2^n$. *Thus,*

$$\delta\left(\tau^{(n)}(0), \tau^n(1)\right) = 3^n - 2^{n+1},$$

which tends to infinity with n. *So given* N, *there are words* $w = \tau^{(n)}(0)$ *and* $v = \tau^n(1)$ *so that* $\delta(w, v) > N$.

Not surprisingly, randomness can also allow for unbalancedness, as in the following Example.

Example 5.1.4 *Let* ζ *be the random substitution*

$$\zeta : \begin{cases} 0 & \longmapsto & 01 \\ 1 & \longmapsto & \begin{cases} 10\,(p) \\ 00\,(q) \end{cases} \end{cases}$$

We show that ζ is unbalanced. The idea is that we can find admissible words v_τ and v_θ so that $\delta(v_\tau, v_\theta)$ grows without bound with the lengths of v_τ and v_θ. Let w be ζ-admissible of length n and denote by $|w|_0$ the number of 0's in w. Define the substitutions τ and θ by

$$\tau : \begin{cases} 0 \longmapsto 01 \\ 1 \longmapsto 10 \end{cases}$$

and

$$\theta : \begin{cases} 0 \longmapsto 01 \\ 1 \longmapsto 00 \end{cases} .$$

Let $v_\tau = \tau(w)$ and $v_\theta = \theta(w)$. Then $|v_\tau| = |v_\tau|$, $h(v_\tau) = |w|_0 + |w|_1 = n$, and $h(v_\theta) = |w|_0$. So $\delta(v_\tau, v_\theta) = n - |w|_0 = |w|_1$. Since 0 always produces a 1 in both τ and θ, $|w|_1 \to \infty$ as $n \to \infty$. So given k, there is some n such that $\delta(v_\tau, v_\theta) > k$. This shows that ζ is unbalanced.

The ideas in the previous example yield a condition for unbalancedness:

Theorem 5.1.1 *Let τ and θ be irreducible substitutions on $\{0,1\}$ such that $\tau(0) = \theta(0)$ and $|\tau(1)| = |\theta(1)|$. Let ρ be a random substitution on $\{0,1\}$ such that*

$$\rho(1) = \begin{cases} \tau(1) \ \textit{with probability } p \\ \theta(1) \ \textit{with probability } 1-p \end{cases} ,$$

where $0 < p < 1$. If $h(\tau(1)) > h(\theta(1))$, then ρ is unbalanced.

Proof. ρ is the random substitution

$$\rho : \begin{cases} 0 \longmapsto \tau(0) = \theta(0) \\ 1 \longmapsto \begin{cases} \tau(1) \\ \theta(1) \end{cases} \end{cases} .$$

Let w be ρ-admissible, $v_\tau = \tau(w)$ and $v_\theta = \theta(w)$. Note that v_τ and v_θ are ρ-admissible since $0 < p < 1$. Since $|\tau(1)| = |\theta(1)|$, $|v_\tau| = |v_\theta|$. We will show that $\delta(v_\tau, v_\theta)$ grows without bound as $|w| \to \infty$. Let $|x|_i$ denote the number of i's in the word x (for $i = 0, 1$). We have that

$$\begin{aligned} h(v_\tau) &= h(\tau(w)) \\ &= |w|_0 \cdot |\tau(0)|_1 + |w|_1 \cdot |\tau(1)|_1 \end{aligned}$$

and

$$\begin{aligned} h(v_\theta) &= h(\theta(w)) \\ &= |w|_0 \cdot |\theta(0)|_1 + |w|_1 \cdot |\theta(1)|_1 . \end{aligned}$$

So, since $\tau(0) = \theta(0)$,

$$\begin{aligned} \delta(v_\tau, v_\theta) &= |w|_0 \cdot (|\tau(0)|_1 - |\theta(0)|_1) + |w|_1 \cdot (|\tau(1)|_1 - |\theta(1)|_1) \\ &= |w|_1 \cdot (|\tau(1)|_1 - |\theta(1)|_1) . \end{aligned}$$

Since $h(\tau(1)) > h(\theta(1))$, $h(\tau(1)) - h(\theta(1)) \geq 1$ so that $\delta(v_\tau, v_\theta) \geq |w|_1$. Since τ and θ are irreducible and $0 < p < 1$, given k there is some N_k such that $|w|_1 > k$. Thus, $\delta(v_\tau, v_\theta)$ grows without bound, as was to be shown. ■

We would like to apply ideas in Example 5.1.4 and Theorem 5.1.1 to non-constant length substitutions. A problem that arises is in choosing words v_τ and v_θ of the *same* length. We consider the case of FibMorse:

Example 5.1.5 *Let ζ be the (p, q)-FibMorse substitution*

$$\zeta : \begin{cases} 0 \longmapsto & 01 \\ 1 \longmapsto & \begin{cases} 0 & \textit{with probability } p \\ 10 & \textit{with probability } q \end{cases} \end{cases}$$

(where $q = 1 - p$ and $0 < p < 1$). We will show that ζ is unbalanced. Let u be any sequence arising from ζ, and let

$$\zeta_f : \begin{cases} 0 \longrightarrow 01 \\ 1 \longrightarrow 0 \end{cases}$$

and

$$\zeta_m : \begin{cases} 0 \longrightarrow 01 \\ 1 \longrightarrow 10 \end{cases}$$

be the Fibonacci and Morse substitutions, respectively. The idea is that we can find words v_f *and* v_m *from* u *so that* $\delta(v_f, v_m)$ *grows without bound. Denote by* $|w|_0$ *the number of* 0*'s in* w*. Let* w *be* ζ*-admissible (so* $w \in u$*) and let* w' *be any* ζ*-admissible word of length* $|w'| = |w| - \frac{1}{2}h(w)$*. Let* $v_f = \zeta_f(w)$ *and* $v_m = \zeta_m(w')$*. Note that* v_f *and* v_m *are* ζ*-admissible since* $0 < p < 1$*. Since every symbol in* w' *produces two symbols via* ζ_m*, we have that*

$$\begin{aligned} |v_m| &= 2|w'| \\ &= 2|w| - h(w) \\ &= |w| + |w|_0 \end{aligned}$$

and similarly

$$\begin{aligned} |v_f| &= 2|w|_0 + h(w) \\ &= |w|_0 + |w| \end{aligned}$$

so that $|v_m| = |v_f|$*. Now, since only the symbol* 0 *produces the symbol* 1 *via* ζ_f*,* $h(v_f) = |w|_0$*. Similarly,* $h(v_m) = |w'|_0 + h(w') = |w'|$*. We then have*

$$\begin{aligned} \delta(v_m, v_f) &= \left| |w'| - |w|_0 \right| \\ &= \left| |w| - \frac{1}{2}h(w) - |w|_0 \right| \\ &= \left| h(w) - \frac{1}{2}h(w) \right| \\ &= \frac{1}{2}h(w). \end{aligned}$$

Since w *was arbitrary, we can pick* w *so that* $\delta(v_m, v_f)$ *tends to infinity. This shows that for any given integer* k*, we can find two* ζ*-admissible words* v_f *and* v_m *whose height distance is greater than* k*. Therefore,* ζ *is unbalanced.*

Having a few examples at our disposal, we now present some properties of k-balanced sequences.

Proposition 5.1.2 *If the sequence $u \in \{0,1\}^{\mathbf{N}}$ is k_0-balanced, then the limit $\mu[1]$ exists.*

Proof. Let

$$a_n = \frac{1}{p(n)} \sum_{w \in \Omega_n} h(w)$$

be the "average height" of a word of length n. Enumerate $\Omega_n = \left\{w_1, w_2, \ldots, w_{p(n)}\right\}$. Since u us k_0-balanced, for any word w of length n,

$$\begin{aligned}
|h(w) - a_n| &= \left| \frac{1}{p(n)} \left(p(n) h(w) - \sum_{j=1}^{p(n)} h(w_j) \right) \right| \\
&= \frac{1}{p(n)} \left| p(n) h(w) \sum_{j=1}^{p(n)} \frac{1}{p(n)} - \sum_{j=1}^{p(n)} h(w_j) \right| \\
&\leq \frac{1}{p(n)} \sum_{j=1}^{p(n)} |h(w) - h(w_j)| \\
&\leq k_0.
\end{aligned}$$

We now show that $\lim_{n\to\infty} \frac{a_n}{n}$ exists. Fix m and write $n = dm + r$ for some $d \geq 0$ and $0 \leq r < m$. Write any word of length n as the concatenation of d words of length m and a word of length $r < d$. Then since $h(w) \leq a_m + k_0$ for any word w of length m, $a_n \leq (d+1)(a_m + k_0)$. Now, as $n \to \infty$, $d \to \infty$ so that $\frac{d+1}{dm+r} \to \frac{1}{m}$. So

$$\begin{aligned}
\limsup_{n\to\infty} \frac{a_n}{n} &\leq \limsup_{n\to\infty} \left(\frac{(d+1) a_m}{n} + \frac{k_0 (d+1)}{n} \right) \\
&= \limsup_{n\to\infty} \frac{(d+1) a_m}{dm + r} \\
&= \frac{a_m}{m}.
\end{aligned}$$

Since m was arbitrary, we have

$$\begin{aligned}
\liminf_{n\to\infty} \frac{a_n}{n} &\leq \limsup_{n\to\infty} \frac{a_n}{n}. \\
&\leq \inf_m \frac{a_m}{m} \\
&\leq \liminf_{n\to\infty} \frac{a_n}{n}
\end{aligned}$$

Therefore, $\liminf_{n\to\infty} \frac{a_n}{n} = \limsup_{n\to\infty} \frac{a_n}{n}$ so that $\lim_{n\to\infty} \frac{a_n}{n}$ exists. Now, since $k_0 - a_n \leq h(u_0 \ldots u_{n-1}) \leq k_0 + a_n$ for all n,

$$\lim_{n\to\infty} \frac{a_n}{n} \leq \lim_{n\to\infty} \frac{h(u_0 \ldots u_{n-1})}{n} \leq \lim_{n\to\infty} \frac{a_n}{n}$$

so that the slope $\mu[1] = \lim_{n\to\infty} \frac{h(u_0 \ldots u_{n-1})}{n}$ exists. ■

Note that this shows that unbalancedness does not necessarily imply the nonexistence of slope. For example, the FibMorse substitution is unbalanced and has positive slope (Examples 5.1.5 and 3.2.3) and the random substitution in 5.1.4 was seen to be unbalanced but has slope $\frac{2}{5}$ for $p = q = \frac{1}{2}$ (by Theorem 3.2.2).

Corollary 5.1.1 *If u is k_0-balanced, then for all n there is some a_n such that $h(w) \leq a_n \pm k_0$ for all admissible words w of length n.*

As a partial converse to this Corollary, we have that if $h(w) \leq a_n \pm k_0$ for all admissible w in u of length n, then the triangle inequality immediately shows that u is $2k$-balanced, where $k \leq k_0$.

Corollary 5.1.2 *If u is k_0-balanced and eventually periodic, then the slope $\mu[1]$ of u is rational.*

Proof. Since u is k_0-balanced, $\mu[1]$ exists. Since u is eventually periodic,

$$u = xy^\infty = \lim_{n\to\infty} xy^n$$

for some admissible words x, y. Then

$$\begin{aligned}
\mu[1] &= \lim_{n\to\infty} \frac{h(xy^n)}{|xy^n|} \\
&= \lim_{n\to\infty} \frac{h(x) + nh(y)}{|x| + n|y|} \\
&= \frac{h(y)}{|y|} \in \mathbf{Q}.
\end{aligned}$$

■

It can be useful to know that we don't have to "start with zero".

Proposition 5.1.3 *Suppose $u \in \{0,1\}^{\mathbf{Z}_+}$ is a k-balanced sequence with slope α. Then for any positive integer M, we have that*

$$\alpha = \lim_{n\to\infty} \frac{h\left(u_M u_{M+1} \dots u_{M+n-1}\right)}{n}.$$

Proof. Let $\varepsilon > 0$. Then there are N_1 and N_2 such that

$$\left|\frac{h\left(u_0 \dots u_{n-1}\right)}{n} - \alpha\right| < \frac{\varepsilon}{2}$$

whenever $n \geq N_1$ and $\frac{k}{n} < \frac{\varepsilon}{2}$ whenever $n \geq N_2$. Let $N = \max\left\{N_1, N_2\right\}$. Then for any $n \geq N$,

$$\begin{aligned}
\left|\frac{h\left(u_M \dots u_{M+n-1}\right)}{n} - \alpha\right| &\leq \left|\frac{h\left(u_M \dots u_{M+n-1}\right)}{n} - \frac{h\left(u_0 \dots u_{n-1}\right)}{n}\right| + \left|\frac{h\left(u_0 \dots u_{n-1}\right)}{n} - \alpha\right| \\
&\leq \frac{k}{n} + \left|\frac{h\left(u_0 \dots u_{n-1}\right)}{n} - \alpha\right| \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon.
\end{aligned}$$

■

Let $M_j = \max\left\{n : j^n \text{ is admissible}\right\}$ for $j = 1, 2$.

Proposition 5.1.4 *If $u \in \{0,1\}^{\mathbf{Z}_+}$ is not eventually constant and is k_0-balanced, then M_0 exists and is finite. Furthermore, $\min\left\{M_0, M_1\right\} \leq k_0$ and M_1 exists and is finite if $M_0 \geq k_0 + 1$.*

Proof. Suppose M_0 is not finite and take n large enough so that there is a word w of length n with $h\left(w\right) = k_0 + 1$. Then 0^n is admissible so that

$$\left|h\left(w\right) - h\left(0^n\right)\right| = k_0 + 1,$$

contradicting k_0-balancedness. So $M_0 < \infty$. This means that also $M_1 < \infty$ if $M_0 \geq k_0+1$.

If $M_1 \geq M_0$, then 1^{M_0} is admissible so that $M_0 = \left|h\left(1^{M_0}\right) - h\left(0^{M_0}\right)\right| \leq k$. Similarly, if $M_0 \leq M_1$ then $M_1 \leq k$. So $\min\{M_0, M_1\} \leq k$. ∎

Let B_k be the set of all k-balanced sequences that are not eventually periodic. Let m be the minimum number of consecutive 1's in any element of B_1. Then 01^m0 is admissible. By balanced, 1^{m+2} is therefore forbidden. Hence, the number of consecutive $1's$ in B_1 is either m or $m+1$. If $m \geq 2$, then 00 is forbidden so that 0 is isolated. Similarly, if $m = 1$, then there is a smallest m' such that $10^{m'}1$ is admissible. Then $0^{m'+2}$ is forbidden. We have thus shown that:

Lemma 5.1.1 *Either* 0 *or* 1 *is isolated for any element of* B_1.

Since $\left|h\left(0^{k+1}\right) - h\left(1^{k+1}\right)\right| = k+1 > k$, we can generalize this a bit to:

Lemma 5.1.2 *If u is k-balanced (with k the minimal such) and not eventually constant, then either 0^{k+1} or 1^{k+1} is forbidden.*

Let $u \in \{0,1\}^{\mathbf{Z}_+}$ be balanced (and aperiodic) and let M_0 be the maximal length of a string of 0's in u. Let w be a finite word in u. If $|w| \leq M_0$, then by maximality of M_0 there is a word $z_0 = 0^{|w|}$ that occurs in u with $h(z_0) = 0$ and $|z_0| = |w|$. Then by balanced,

$$h(w) = h(w) - h(z_0) \leq 1.$$

Now suppose that $|w| > M_0$. Then we can write w as the concatenation of $N = \left\lfloor \frac{|w|}{M_0} \right\rfloor$ words in u of length M_0 and a "left-over" piece of length $< M_0$. That is,

$$w = w_1 w_2 w_3 \ldots w_N w_L$$

where $|w_j| = M_0$ for all $j \leq N$ and $|w_L| < M_0$. Then $h(w) = \sum_{j=1}^{N} h(w_j) + h(w_L) \leq N + 1 = \left\lfloor \frac{|w|}{M_0} \right\rfloor + 1$. So the heights of words in a balanced infinite string are bounded by a constant (depending on the length of the word).

Proposition 5.1.5 *If u has slope $\frac{1}{q}$, then 0^{q-1} is admissible (i.e. $0^{q-1} \in u$).*

Proof. If 0^{q-1} is forbidden, then the minimum number of 1's in any word w of length $(q-1)\,c$ is c. That is, $h(w) \geq c$. Since u has slope $\frac{1}{q}$, for $\varepsilon = \frac{1}{q^2}$, there is some N so that

$$\left|\frac{h(u_0 \ldots u_{n-1})}{n} - \frac{1}{q}\right| < \frac{1}{q^2}$$

whenever $n \geq N$. Then, in particular,

$$\frac{h\left(u_0 \ldots u_{(q-1)N-1}\right)}{(q-1)\,N} < \frac{1}{q} + \frac{1}{q^2}.$$

But $h\left(u_0 \ldots u_{(q-1)N-1}\right) \geq N$ so that

$$\frac{1}{q-1} = \frac{N}{(q-1)\,N} \leq \frac{h\left(u_0 \ldots u_{(q-1)N-1}\right)}{(q-1)\,N} < \frac{1}{q} + \frac{1}{q^2},$$

a contradiction since $q^2 - 1 < q^2$ implies

$$\begin{aligned} \frac{1}{q^2-1} &> \frac{1}{q^2} \\ \frac{1}{q-1} &> \frac{q+1}{q^2} \\ &= \frac{1}{q} + \frac{1}{q^2}. \end{aligned}$$

■

For example, if u is 2-balanced and has slope $\frac{1}{4}$, then 111 is forbidden by the preceding Proposition and Lemma 5.1.2. If u is aperiodic, then 010 is admissible since otherwise, $u = w\,(011)^\infty$ for some finite word w, contradicting aperiodicity.

The following Proposition illustrates the useful tool involved in analyzing k-balanced sequences. This is the tool of "breaking up" words into smaller words and using balancedness to approximate properties of words. As before, let M_0 be the length of the "longest string" of zeros.

Proposition 5.1.6 *Let $M_0 = \max\{n : 0^n \in u\}$. Suppose u is k-balanced and has slope $\alpha > 0$. Then M_0 is finite. Furthermore, $M_0 \leq \left\lfloor \frac{k}{\alpha} \right\rfloor$.*

Proof. Suppose that $0^{\lfloor \frac{k}{\alpha} \rfloor + 1}$ is admissible. Then, since u is k-balanced, $h(w) \leq k$ for any word w of length $\lfloor \frac{k}{\alpha} \rfloor + 1$. We may then "break up" any word into a concatenation of words of lengths less than or equal to $\lfloor \frac{k}{\alpha} \rfloor + 1$ so that, for any n,

$$h(u_0 u_1 \ldots u_{n-1}) \leq k \left(\frac{n}{\lfloor \frac{k}{\alpha} \rfloor + 1} + 1 \right).$$

Then

$$\begin{aligned} \lim_{n\to\infty} \frac{h(u_0 \ldots u_{n-1})}{n} &\leq \lim_{n\to\infty} \frac{k}{n} \left(\frac{n}{\lfloor \frac{k}{\alpha} \rfloor + 1} + 1 \right) \\ &= \frac{k}{\lfloor \frac{k}{\alpha} \rfloor + 1} \\ &< \alpha \end{aligned}$$

since $\frac{k}{\alpha} < \lfloor \frac{k}{\alpha} \rfloor + 1$. But this contradicts that u has slope α. Therefore, $0^{\lfloor \frac{k}{\alpha} \rfloor + 1}$ is forbidden so that $M_0 < \lfloor \frac{k}{\alpha} \rfloor$ is finite. ■

Corollary 5.1.3 *If u is k-balanced and has slope $\frac{1}{q}$, then $q - 1 \leq M_0 \leq kq$ so that $h(w) \leq k$ for any word w of length $q - 1$.*

Example 5.1.6 *The Fibonacci sequence (which arises from the Fibonacci substitution) is 1-balanced with slope $\frac{3-\sqrt{5}}{2} \approx 0.382$ and $M_0 = 2$.*

Corollary 5.1.4 *Suppose u is k-balanced and has slope $\alpha > 0$. If 1^{M_0} is admissible, then $M_0 \leq k$.*

Proof. By the Proposition, M_0 is finite. Since u is k-balanced,

$$M_0 = \left| h\left(1^{M_0}\right) - h\left(0^{M_0}\right) \right| \leq k.$$

■

Let M_1 be maximal such that 1^{M_1} is admissible. Then we can show similarly that M_1 is finite (especially if $\alpha \leq \frac{1}{2}$).

Proposition 5.1.7 *Let M_0 and M_1 be minimal such that 0^{M_0} and 1^{M_1} are admissible, respectively. If u is k-balanced with slope α then*

$$|h(u_0 \ldots u_{n-1}) - h(u_0 \ldots u_{m-1})| \leq M_1$$

for all n, m and

$$h(u_0 \ldots u_{n-1}) = h(u_0 \ldots u_{m-1}) \Longrightarrow |n - m| \leq M_0.$$

Proof. M_0 and M_1 are finite by the preceding Proposition. Suppose $n > m$. Then

$$h(u_0 \ldots u_{n-1}) - h(u_0 \ldots u_{m-1}) = h(u_m \ldots u_{n-1}) \leq M_1$$

and similarly if $n < m$. If $n > m$, then $h(u_0 \ldots u_{n-1}) = h(u_0 \ldots u_{m-1})$ implies that $h(u_m \ldots u_{n-1}) = 0$ so that $(n-1) - m + 1 = n - m \leq M_0$. ■

Remark 5.1.1 *Note that if $M_1 \leq M_0$, then 0^{M_1} is admissible so that $M_1 = h\left(1^{M_1}\right) - h\left(0^{M_1}\right) \leq k$. Similarly, if $M_0 \leq M_1$, then $M_0 \leq k$. So either $M_0 \leq k$ or $M_1 \leq k$.*

Theorem 5.1.2 *Let u be k-balanced with slope α and let M_0 and M_1 be maximal such that 0^{M_0} and 1^{M_1} are admissible. Then*

$$\max\left\{\frac{1}{M_0+1}, 1 - \frac{k}{M_1}\right\} \leq \alpha \leq \min\left\{\frac{M_1}{M_1+1}, \frac{k}{M_0}\right\}.$$

Proof. *Since 0^{M_0} is admissible and u is k-balanced, $h(w) = \left|h(w) - h\left(0^{M_0}\right)\right| \leq k$ for any word w of length M_0 so that $h(w) \leq kn$ for any word of length $M_0 n$. Then*

$$\begin{aligned} \alpha &= \lim_{n\to\infty} \frac{h(u_0 \ldots u_{M_0 n-1})}{M_0 n} \\ &\leq \lim_{n\to\infty} \frac{kn}{M_0} \\ &= \frac{k}{M_0}. \end{aligned}$$

Since M_0 is maximal, 0^{M_0+1} is forbidden so that $h(w) \geq 1$ for any word of length M_0+1. Therefore

$$\begin{aligned} \alpha &= \lim_{n\to\infty} \frac{h\left(u_0 \dots u_{(M_0+1)n-1}\right)}{(M_0+1)\,n} \\ &\geq \frac{n}{(M_0+1)\,n} \\ &= \frac{1}{M_0+1}. \end{aligned}$$

Similarly, $h(w) \leq M_1$ for all words w of length M_1+1 so that $\alpha \leq \frac{M_1}{M_1+1}$. Since 1^{M_1} is admissible, $|M_1 - h(w)| \leq k$ for all words w of length M_1 so that $h(w) \geq M_1 - k$. Then

$$\begin{aligned} \alpha &= \lim_{n\to\infty} \frac{h\left(u_0 \dots u_{M_1 n-1}\right)}{M_1 n} \\ &\geq \lim_{n\to\infty} \frac{(M_1-k)\,n}{M_1 n} \\ &= 1 - \frac{k}{M_1}. \end{aligned}$$

■

Example 5.1.7 *If u is k-balanced, 0^{2k} admissible, and 0^{2k+1} forbidden, then $\alpha \leq \frac{1}{2}$.*

Since either $M_0 \leq k$ or $M_1 \leq k$ (by the above remark), we might actually have weaker bounds. For example, if u is 2-balanced, then 0^5 is forbidden (by the above Proposition). If $M_0 = 4$, then $M_1 \leq 2$. Then $h(w) \leq 2$ for all words w of length 3 and $h(w) \geq 1$ for all words w of length 5. Then $\frac{1}{5} \leq \alpha \leq \frac{2}{3}$.

Corollary 5.1.5 *If $k \geq 2$, then*

$$\max\left\{\frac{n}{M_0+1}, n - \frac{kn}{M_1}\right\} \leq h(u_0 \dots u_{n-1}) \leq \min\left\{\frac{nM_1}{M_1+1}, \frac{nk}{M_0}\right\}.$$

for all n.

Proof. $h(u_0 \dots u_n) \leq h(u_0 \dots u_m)$ for all $m \geq n$. ■

Proposition 5.1.8 *If u has slope $\frac{1}{q}$, then 0^{q-1} is admissible (i.e. $0^{q-1} \in u$).*

Proof. If 0^{q-1} is forbidden, then the minimum number of 1's in any word w of length $(q-1)c$ is c. That is, $h(w) \geq c$. Since u has slope $\frac{1}{q}$, for $\varepsilon = \frac{1}{q^2}$, there is some N so that

$$\left| \frac{h(u_0 \dots u_{n-1})}{n} - \frac{1}{q} \right| < \frac{1}{q^2}$$

whenever $n \geq N$. Then, in particular,

$$\frac{h\left(u_0 \dots u_{(q-1)N-1}\right)}{(q-1)N} < \frac{1}{q} + \frac{1}{q^2}.$$

But $h\left(u_0 \dots u_{(q-1)N-1}\right) \geq N$ so that

$$\frac{1}{q-1} = \frac{N}{(q-1)N} \leq \frac{h\left(u_0 \dots u_{(q-1)N-1}\right)}{(q-1)N} < \frac{1}{q} + \frac{1}{q^2},$$

a contradiction since $q^2 - 1 < q^2$ implies

$$\begin{aligned} \frac{1}{q^2-1} &> \frac{1}{q^2} \\ \frac{1}{q-1} &> \frac{q+1}{q^2} \\ &= \frac{1}{q} + \frac{1}{q^2}. \end{aligned}$$

■

Corollary 5.1.6 *Suppose $q \geq k+2$. If u is k-balanced and has slope $\frac{1}{q}$, then 1^{q-1} is forbidden.*

Proof. If 1^{q-1} were admissible, then $\left|h\left(1^{q-1}\right) - h\left(0^{q-1}\right)\right| \leq k$. But then $q \leq k+1$, a contradiction. ■

5.2 Recoverability

The idea of *recoverability* is not a new one (for example, see [21], Definition 5.6). It is, however, interesting to explore the idea in regards to random substitutions.

Definition 5.2.1 *Let ζ be a (possibly random) substitution on $\mathcal{A} = \{0, 1, \ldots, s-1\}$. An infinite string $u \in \mathcal{A}^{\mathbf{Z}_+}$ is ζ-**recoverable** if for every $i \leq j$, $u_i, \ldots, u_j = \zeta(w_i)$ for some unique ζ-admissible w_i and $w_0 w_1 w_2 \ldots \in \mathcal{A}^{\mathbf{Z}_+}$ is ζ-admissible. u is **weakly** ζ-**recoverable** if for every $i \leq j$ there is some $n_i \geq 1$ such that $u_i, \ldots, u_j = \zeta^{n_i}(w_i)$ for some unique ζ-admissible w_i. ζ is (**weakly**) **recoverable** if every sequence u generated by ζ is (weakly) ζ-recoverable.*

Example 5.2.1 *Recall the 2-balanced random substitution ρ from Example 5.1.2:*

$$\rho : \begin{cases} 0 & \longmapsto \quad 0011 \\ 1 & \longmapsto \begin{cases} 1001 & p \\ 0101 & q \end{cases} \end{cases}$$

We will show that ρ is recoverable. Let u be an infinite string arising from ρ. The word 111 is admissible and 01 is the unique word such that $111 = \rho(01)$. So

$$\begin{aligned} u &= \cdots 111 \cdots \\ &= \cdots \rho(01) \cdots \\ &= \cdots u_n u_{n+1} u_{n+2} 00111001 u_m u_{m+1} \cdots . \end{aligned}$$

Furthermore, $u_n u_{n+1} u_{n+2}$ and $u_m u_{m+1}$ end and start a substituted symbol, respectively and these substituted words is uniquely determined by $u_n u_{n+1} u_{n+2}$ and $u_m u_{m+1}$. This unicity then also completely determines the rest of u. Therefore, ρ is recoverable.

We may "mimic" the FibMorse substitution so as to be recoverable. Recall Example 3.2.4:

Example 5.2.2 *We have that*

$$\alpha : \begin{cases} 0 & \longmapsto & \begin{cases} 10 & \frac{1}{2} \\ 02 & \frac{1}{2} \end{cases} \\ 1 & \longmapsto & 03 \\ 2 & \longmapsto & 04 \\ 3 & \longmapsto & 3 \\ 4 & \longmapsto & 4 \end{cases}$$

Let u be an infinite string generated by α. By the definition of α, the word $a3$ is forbidden for $a = 1, 2, 3, 4$. Any 1 or 2 in u is uniquely the image of 0. The word 103 is uniquely the image of 03 and so any other occurrence of 03 in u is uniquely the image of 1. A similar analysis of words of the form $a4$ yield unicity of images. We conclude that α is recoverable.

5.3 Recoverability and Entropy

Is there a relationship between a deterministic (entropy zero) process and a recoverable substitution? It seems that being recoverable determines exactly what a sequence generated by a substitution looks like. However, this is not the case. For example, it is known that the Fibonacci substitution ω is recoverable (see for instance [14], Chapter 1) but the substitution

$$\omega' : \begin{cases} 0 \mapsto 010 \\ 1 \mapsto 101 \end{cases}$$

is not recoverable as it generates the periodic sequence $u = 0101010\ldots$. Both these substitutions are deterministic by Corollary 4.4.1. Then what is the real difference between being deterministic and recoverable? Recall that heuristically, to have entropy 0 means that there is no "surprise" or information gained in reading a symbol in an infinite string. However, to be recoverable just means that preimages of words are unique.

5.4 Entropy, Balancedness, and Recoverability - the Possibilities

Below is a table of the possibilities that can occur. Entropy zero is because of Corollary 4.4.1.

Example	Entropy	k-Balanced	Recoverable
$\begin{cases} 0 \longmapsto 01 \\ 1 \longmapsto 0 \end{cases}$	0	$k=1$ Ex. 5.1.1	yes
$\begin{cases} 0 \longmapsto 0100 \\ 1 \longmapsto 11 \end{cases}$	0	$k=\infty$ Ex. 5.1.3	yes
$\begin{cases} 0 \longmapsto 0011 \\ 1 \longmapsto \begin{cases} 0101 & \frac{1}{2} \\ 1001 & \frac{1}{2} \end{cases} \end{cases}$	$\frac{1}{6}\log 2$ Ex. 4.2.4	$k=2$ Ex. 5.1.2	yes Ex. 5.2.1
$\begin{cases} 0 \longmapsto 01 \\ 1 \longmapsto \begin{cases} 0 & \frac{1}{2} \\ 10 & \frac{1}{2} \end{cases} \end{cases}$	positive Thm. 4.2.2	$k=\infty$ Ex. 5.1.5	no
$\begin{cases} 0 \longmapsto 010 \\ 1 \longmapsto 101 \end{cases}$	0	$k=1$	no
$(\frac{1}{2}, \frac{1}{2})$-Bridge substitution	$\frac{1}{2}\log 2$ Ex. 4.2.3	NA	no
$\begin{cases} 0 \longmapsto 01 \\ 1 \longmapsto \begin{cases} 01 & \frac{1}{2} \\ 10 & \frac{1}{2} \end{cases} \end{cases}$	positive	$k=2$	no

6 A GEOMETRIC PICTURE OF SEQUENCES

We will view binary sequences as subsets of $\mathbf{Z}^2$ as follows. Given $u \in \{0,1\}^{\mathbf{Z}_+}$, define the sequence $\{(n_k, m_k)\}_{k=0}^{\infty} \subseteq \mathbf{Z}_+^2$ by letting $(n_0, m_0) = (0,0)$ and $(n_k, m_k) = (k, h(u_0 u_1 \ldots u_{k-1}))$ for $k \geq 1$ (where $h(w)$ is the number of 1's in w). Note that $\{(n_k, m_k)\}_{k=0}^{\infty}$ is nondecreasing and has slope at most 1. Conversely, given any nondecreasing sequence $\{(n_k, m_k)\}_{k=0}^{\infty}$ in $\mathbf{Z}_+^2$ with slope at most 1, define a sequence $u \in \{0,1\}^{\mathbf{Z}_+}$ by letting $u_k = m_{k+1} - m_k$. In this way, we have a bijection $\varphi : \{0,1\}^{\mathbf{Z}_+} \longrightarrow \mathbf{Z}_{1/8}^2 := \left\{(n,m) \in \mathbf{Z}^2 : 0 \leq m \leq n\right\}$ and to say that u **is in** a subset S of $\mathbf{Z}_{1/8}^2$ is to consider the sequence $\varphi(u)$, and we write $u \in S$. We can use this geometric description to calculate the complexity and analyze the structure of certain spaces of sequences. This is therefore a new prospect that would be interesting to compare and combine with previous theory.

A **line-approximating path** (LA-path) in $\mathbf{Z}^2$ is a sequence $\{(x_n, y_n)\}_n \subseteq \mathbf{Z}^2$ such that $(x_n, y_n) \in \{(x_{n-1}+1, y_{n-1}), (x_{n-1}+1, y_{n-1}+1)\}$ for all $n \in \mathbf{Z}$. Let LA be the set of all line-approximating paths. Call an LA-path **finite** if it is a finite subsequence of an LA-path. Let $k \geq 2$ be an integer, $0 < \alpha \leq \frac{1}{2}$, $M_0 = \left\lfloor \frac{k}{\alpha} \right\rfloor$, and $M_1 = \left\lfloor \frac{1}{\alpha} \right\rfloor$. Define the **lower envelope** $e_{k,\alpha}^{\ell} : \mathbf{Z}_+^2 \to \mathbf{Z}_+^2$ and **upper envelope** $e_{k,\alpha}^{u} : \mathbf{Z}_+^2 \to \mathbf{Z}_+^2$ functions as follows (see Figure 6.1):

$$e_{k,\alpha}^{\ell}(n) = \begin{cases} 0 \text{ if } n \leq M_0 \\ \lceil \alpha n \rceil - k \text{ if } n \geq M_0 + 1 \end{cases}$$

and

$$e_{k,\alpha}^{u}(n) = \begin{cases} n \text{ if } n \leq M_1 \\ \lfloor \alpha n \rfloor + k \text{ if } n \geq M_1 + 1 \end{cases}.$$

Note that the upper and lower envelopes are the best lattice approximations below the line $y = \alpha x + k$ and above the line $y = \alpha x - k$, respectively. Let $M_0 = max\{n : 0^n \text{ is admissible}\}$. If $y = \alpha x$, then the lower envelope "hugs" the line $y = \alpha x - k$, which has x-intercept $x = \frac{k}{\alpha}$. So it makes sense that we have $M_0 \leq \frac{k}{\alpha}$, as in Proposition 5.1.6.

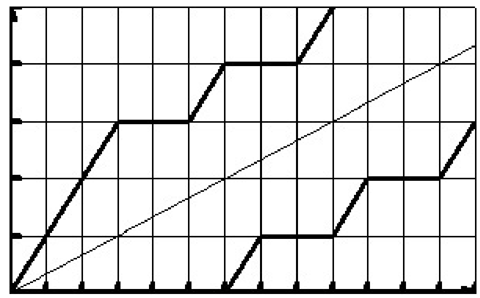

FIGURE 6.1: $e^{\ell}_{k,\alpha}$ and $e^{u}_{k,\alpha}$ for $k = 2$, $\alpha = \frac{1}{3}$

Theorem 6.0.1 *Let $M_0 = \left\lfloor \frac{k}{\alpha} \right\rfloor$, $e^{\ell}_{k,\alpha}$ and $e^{u}_{k,\alpha}$ be the lower and upper envelope functions as defined above. Let $E_{k,\alpha} = \left\{(n,m) \in \mathbf{Z}^2 : n \geq 0, e^{\ell}_{k,\alpha}(n) \leq m \leq e^{u}_{k,\alpha}(n)\right\}$. If $u \in E_{k,\alpha}$, then u has slope α. Conversely, any k-balanced $u \in \{0,1\}^{\mathbf{Z}_+}$ with slope α is in $E_{k,\alpha}$ (that is, $\varphi(u) \in E_{k,\alpha}$, where φ is the bijection described above).*

Proof. If $u \in E_{k,\alpha}$, then

$$\alpha n - k \leq \lceil \alpha n \rceil - k \leq h(u_0 \ldots u_{n-1}) \leq \lfloor \alpha n \rfloor + k \leq \alpha n + k$$

so that by dividing by n and taking limits gives

$$\alpha \leq \lim_{n \to \infty} \frac{h(u_0 \ldots u_{n-1})}{n} \leq \alpha.$$

Therefore, u has slope α.

Now suppose that u is k-balanced with slope α. We first show by way of contradiction that any such u satisfies $h(u_0 \ldots u_{n-1}) \geq e^{\ell}_{k,\alpha}(n)$ for all n. Indeed, suppose there is some n_0 such that $h(u_0 \ldots u_{n_0-1}) < e^{\ell}_{k,\alpha}(n_0)$ and that n_0 is the smallest such. Then $n_0 > M_0$ and so $h(u_0 \ldots u_{n_0-1}) < \alpha n_0 - k$ since $h(u_0 \ldots u_{n_0-1})$ is an integer. Since u is k-balanced, this means that $|h(w) - h(u_0 \ldots u_{n_0-1})| \leq k$ for any word w of length n_0 so that $h(w) < \alpha n_0$. Thus,

$$\begin{aligned}\alpha &= \lim_{m\to\infty} \frac{h(u_0 \ldots u_{mn_0-1})}{mn_0} \\ &< \lim_{m\to\infty} \frac{m(\alpha n_0)}{mn_0} \\ &= \alpha,\end{aligned}$$

a contradiction. Thus, $h(u_0 \ldots u_{n-1}) \geq e^{\ell}_{k,\alpha}(n)$ for all n.

We now show similarly that $h(u_0 \ldots u_{n-1}) \leq e^{u}_{k,\alpha}(n)$ for all n. Suppose there is some n_0 such that $h(u_0 \ldots u_{n_0-1}) > e^{u}_{k,\alpha}(n_0)$ and n_0 is the smallest such. Then $n_0 > M_1$ and so $h(u_0 \ldots u_{n_0-1}) > \alpha n_0 + k$ since $h(u_0 \ldots u_{n_0-1})$ is an integer. Since u is k-balanced, $|h(u_0 \ldots u_{n_0-1}) - h(w)| \leq k$ for any word w of length n_0 so that $h(w) > \alpha n_0$. Thus,

$$\begin{aligned}\alpha &= \lim_{m\to\infty} \frac{h(u_0 \ldots u_{mn_0-1})}{mn_0} \\ &> \lim_{m\to\infty} \frac{m(\alpha n_0)}{mn_0} \\ &= \alpha,\end{aligned}$$

a contradiction. Thus, $h(u_0 \ldots u_{n-1}) \leq e^{u}_{k,\alpha}(n)$ for all n. ■

So we have a nice geometric description of a set $\mathcal{S}_{k,\alpha}$ *containing* k-balanced binary sequences with slope α. Note that this is not an exact description as there are sequences in $\mathcal{S}_{k,\alpha}$ that are not necessarily k-balanced. For example, $u = 0^1 1^1 0^2 1^2 0^3 1^3 \cdots = 010011000111\cdots$ has slope $\frac{1}{2}$, hence is in $E_{k,1/2}$ but is not k-balanced for any k. Another example of slope $\frac{1}{2}$ but not 2-balanced is $u = (110110110110110001000100)^{\infty}$ (period $=$ 24). Note that $\left|\frac{1}{2}n - h(u_0 \cdots u_{n-1})\right| = 3 > 2$ when $n = 14$. u is not 2-balanced since $|h(u_0 \cdots u_5) - h(u_{14} \cdots u_{19})| = 3$.

This geometric description of $\mathcal{S}_{k,\alpha}$ can help us detect properties of sequences. For example:

Corollary 6.0.1 *If u, v are k-balanced with slope α, then for all n,*

$$|h(u_0 \dots u_{n-1}) - h(v_0 \dots v_{n-1})| \leq \begin{cases} 2k & \text{if } \alpha n \in \mathbf{Z} \\ 2k-1 & \text{if } \alpha n \notin \mathbf{Z} \end{cases}.$$

Thus,

$$|h(u_j \dots u_{n-1}) - h(v_j \dots v_{n-1})| \leq \begin{cases} 4k & \text{if } \alpha n \in \mathbf{Z} \\ 4k-2 & \text{if } \alpha n \notin \mathbf{Z} \end{cases}$$

for all j and n.

Proof. $u, v \in E_{k,\alpha}$ and $h(u_j \dots u_{n-1}) = h(u_0 \dots u_{n-1}) - h(u_0 \dots u_{j-1})$. ■

Corollary 6.0.2 *Let u be k-balanced with slope α. Then $h(u_0 \dots u_{n-1})$ can take at most $2k+1$ values if $\alpha n \in \mathbf{Z}$ and at most $2k$ values otherwise.*

Proof. By Theorem 6.0.1, $u \in E_{k,\alpha}$. The least number of 1's occurs if $\varphi(u) = e^{\ell}$ so that $\lceil \alpha n \rceil - k \leq h(u_0 \dots u_{n-1})$. The greatest number of 1's occurs when $\varphi(u) = e^{u}$ so that $h(u_0 \dots u_{n-1}) \leq \lfloor \alpha n \rfloor + k$. Then

$$\begin{aligned} (\lfloor \alpha n \rfloor + k) - (\lceil \alpha n \rceil - k) + 1 &= \lfloor \alpha n \rfloor - \lceil \alpha n \rceil + 2k + 1 \\ &= \begin{cases} 2k+1 & \text{if } \alpha n \in \mathbf{Z} \\ 2k & \text{if } \alpha n \notin \mathbf{Z} \end{cases}. \end{aligned}$$

■

6.1 Complexity via Geometry

Let $P_\alpha|_{(a,b)\to(c,d)}$ denote the set of finite sub-paths of elements of $E_{k,\alpha}$ that start at (a, b) and end at (c, d). Let S_{PC} be the infinite matrix whose ijth entry is 1 if $i = j$ or if $i = j + 1$ and 0 otherwise ("S" is for "Step"). Let F_{PC} be the infinite matrix whose ijth entry is 1 if $i = j$ or if $i = j - 1$ and 0 otherwise ("F" is for "Flat"). So

$$S_{PC} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

and

$$F_{PC} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 1 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

By a **standard** $n \times m$ **submatrix** of an infinite matrix M, we mean an $n \times m$ matrix A such that $A_{ij} = M_{ij}$ for all $0 \leq i \leq n$ and $0 \leq j \leq m$. Let e^u and e^ℓ be the upper and lower envelope functions as above:

$$e^\ell_{k,\alpha}(n) = \begin{cases} 0 \text{ if } n \leq M_0 \\ \lceil \alpha n \rceil - k \text{ if } n \geq M_0 + 1 \end{cases}$$

and

$$e^u_{k,\alpha}(n) = \begin{cases} n \text{ if } n \leq M_1 \\ \lfloor \alpha n \rfloor + k \text{ if } n \geq M_1 + 1 \end{cases}.$$

Note that

$$e^u_{k,\alpha}(n) - e^\ell_{k,\alpha}(n) = \begin{cases} 2k & \text{if} \quad \alpha n \in \mathbf{Z} \\ 2k-1 & \text{if} \quad \alpha n \notin \mathbf{Z} \end{cases}.$$

Let $S_{PC,k}(n)$ be the standard $\left(e^u_{k,\alpha}(n) - e^\ell_{k,\alpha}(n)\right) \times \left(e^u_{k,\alpha}(n+1) - e^\ell_{k,\alpha}(n+1)\right)$ submatrix of S_{PC} and let $F_{PC,k}(n)$ be the standard $\left(e^u_{k,\alpha}(n) - e^\ell_{k,\alpha}(n)\right) \times \left(e^u_{k,\alpha}(n+1) - e^\ell_{k,\alpha}(n+1)\right)$ submatrix of F_{PC}. We call both $S_{PC,k}(n)$ and $F_{PC,k}(n)$ the **path-connecting matrices at** n. What does a path-connecting matrix $M = [m_{ij}]$ represent? For each $n \in \mathbf{Z}_+$, denote by $v_{n,i}$ the point

$$v_{n,i} = \left(n, e^\ell_{k,\alpha}(n) + i - 1\right)$$

for $i = 1, \ldots, e^u_{k,\alpha}(n)$. Note that $m_{ij} = 1$ if

$$e^\ell_{k,\alpha}(n+1) - e^\ell_{k,\alpha}(n) + i - j = \left(e^\ell_{k,\alpha}(n+1) + i - 1\right) - \left(e^\ell_{k,\alpha}(n) + j - 1\right) \in \{0,1\}$$

and $m_{ij} = 0$ otherwise. So m_{ij} is the number of finite LA-paths of length 1 from $v_{n,i}$ to $v_{n+1,j}$. If $M = [m_{ij}]$ and $M' = \left[m'_{ij}\right]$ are path-connecting matrices at n and $n+1$, respectively, then the ij entry of MM' is

$$\sum_{k=1}^{e^u_{k,\alpha}(n+1) - e^\ell_{k,\alpha}(n+1)} m_{ik} m'_{kj},$$

the number of finite LA-paths of length 2 from $v_{n,i}$ to v_{n+2}, j. By induction, the entries of the product of N path-connecting matrices M_n at n give the number of finite LA-paths of length N that start at $v_{n,e^\ell_{k,\alpha}(n)}$. We have therefore established the following Theorem.

Theorem 6.1.1 *Let $\alpha \in (0,1)$ and $P = \prod_{n=1}^{N} P_n$ where*

$$P_n = \begin{cases} S_{PC,k}(n) & \text{if} \quad e^\ell_{PC,k}(n+1) - e^\ell_{PC,k}(n) = 1 \\ F_{PC,k}(n) & \text{if} \quad e^\ell_{PC,k}(n+1) - e^\ell_{PC,k}(n) = 0 \end{cases}$$

(so $P_n = S_{PC,k}(n)$ if the lower envelope "takes a step" and $P_n = F_{PC,k}(n)$ if the lower envelope "stays flat"). Then the sum of the entries in P is the total number of finite LA-paths of length N in $E_{k,\alpha}$ that start at $(0,0)$.

Corollary 6.1.1 *Let $\mathcal{B}_{k,\alpha}$ be the space of k-balanced sequences having slope α. Then*

$$p(n) \leq \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} P^n \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

where P is as in the Theorem. Thus,

$$h_{top}(\mathcal{B}_{k,\alpha}) \leq \lim_{n \to \infty} \frac{\log p(n)}{n}.$$

If $\alpha = \frac{p}{q} > 0$ *is in lowest terms, then*

$$p(n) \geq \begin{bmatrix} 1 & 1 \end{bmatrix} \left(\prod_{j=1}^{p} \begin{bmatrix} C_j - C_{j-1} & C_j - C_{j-1} - 1 \\ 1 & 1 \end{bmatrix} \right) \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

where $C_0 = 1$ *and* $C_j = \min\left\{n : e_{\min}^{\ell}(n) = j\right\}$ *for* $j = 1, \ldots, p$.

Proof. We have the upper bound since $\mathcal{B}_{k,\alpha} \subseteq E_{k,\alpha}$ (under the bijection φ). For the lower bound, redefine the upper and lower envelopes $e_{\min}^{u}$ and $e_{\min}^{\ell}$ by

$$e_{\min}^{u}(n) = \lceil \alpha n \rceil$$

and

$$e_{\min}^{\ell}(n) = \lfloor \alpha n \rfloor .$$

Let $E_{\min} = LA \cap \left\{(n,m) \in \mathbf{Z}^2 : n \geq 0, e_{\min}^{\ell}(n) \leq m \leq e_{\min}^{u}(n)\right\}$ (where LA is the set of all LA-paths). Then $E_{\min} \subseteq \mathcal{B}_{k,\alpha}$. Let $C_0 = 1$ and $C_j = \min\left\{n : e_{\min}^{\ell}(n) = j\right\}$ for $j = 1, \ldots, p$. Then there are $C_j - C_{j-1} - 1$ integers in the interval $[C_{j-1}, C_j)$ so that the sum of the entries of

$$\left[F_{PC,k}(n)\right]^{C_j - C_{j-1} - 1} S_{PC,k}(n) = \begin{bmatrix} C_j - C_{j-1} & C_j - C_{j-1} - 1 \\ 1 & 1 \end{bmatrix}$$

is the number of finite LA-paths in $E_{k,p/q}$ that start at a point having x-coordinate C_{j-1} and end at a point having x-coordinate C_j. This gives the result. ■

Example 6.1.1 *We will show that the space* $\mathcal{B}_{2,1/q}$ *of 2-balanced binary sequences with slope* $\frac{1}{q}$ $(q \geq 2)$ *satisfies*

$$\frac{1}{q} \log q \leq h_{top}\left(\mathcal{B}_{2,1/q}\right) \leq \frac{1}{q} \log \left(\frac{q + 2 + \sqrt{(q+2)^2 - 12}}{2} \right).$$

By Theorem 6.0.1, any 2*-balanced sequence with slope* $\frac{1}{q}$ *is in* $E_{2,1/q}$*. The path-connecting matrix at* $cq+1$ *is* $\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$, *at* $c(q+1)$ *is* $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$, *and at* $cq+j$ *is* $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ *for*

$2 \leq j \leq q-1$ *(for* $q \geq 3$*). Let*

$$P = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{q-2} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

for $q \geq 3$ *and*

$$P = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

for $q = 2$*. There are*

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} P \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

finite LA-paths of length q *between* $cq+1$ *and* $c(q+1)+1$ *for any integer* $c \geq 2$ *and hence*

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} P^n \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

paths of length qn *between* $cq+1$ *and* $c(q+n)+1$*. Now,*

$$P = \begin{bmatrix} 1 & q-1 & q-2 \\ 1 & q & q-1 \\ 0 & 1 & 1 \end{bmatrix}$$

and by diagonalization of P *we have that*

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} P^n \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3r}\left(\left(\frac{1}{2}q + \frac{1}{2}r + 1\right)^n (8q + 4r - 8) + \left(\frac{1}{2}q - \frac{1}{2}r + 1\right)^n (4r - 8q + 8)\right),$$

where $r = \sqrt{(q+2)^2 - 12}$. *Therefore,*

$$\begin{aligned} h_{top}\left(\mathcal{B}_{2,1/q}\right) &= \lim_{n\to\infty} \frac{\log p\,(qn)}{qn} \\ &\le \frac{1}{q} \lim_{n\to\infty} \frac{1}{n} \log \left[\left(\frac{1}{2}q + \frac{1}{2}r + 1 \right)^n \left((8q + 4r - 8) + \frac{\left(\frac{1}{2}q - \frac{1}{2}r + 1\right)^n}{\left(\frac{1}{2}q + \frac{1}{2}r + 1\right)^n} (4r - 8q + 8) \right) \right] \\ &= \frac{1}{q} \lim_{n\to\infty} \frac{1}{n} \log \left(\frac{1}{2}q + \frac{1}{2}r + 1 \right)^n \\ &= \frac{1}{q} \left(\frac{1}{2}q + \frac{1}{2}r + 1 \right). \end{aligned}$$

As for the lower bound, define the lower and upper envelopes $e^{\ell}_{\min}$ *and* $e^{u}_{\min}$ *by*

$$e^{\ell}_{\min}(n) = \left\lfloor \frac{n}{q} \right\rfloor$$

and

$$e^{u}_{\min}(n) = \left\lceil \frac{n}{q} \right\rceil.$$

Let $E_{\min} = LA \cap \left\{(n,m) \in \mathbf{Z}^2 : n \ge 0, e^{\ell}_{\min}(n) \le m \le e^{u}_{\min}(n)\right\}$ *(where* LA *is the set of all LA-paths). Then* $E_{\min} \subseteq \mathcal{B}_{2,1/q}$. *There are* q *paths of length* q *in* $E_{\min}$ *between* cq *and* $c\,(q+1)$ *for any integer* $c \ge 2$ *and hence* q^n *paths of length* qn *between* cq *and* $c\,(q+n)$. *Thus,*

$$\begin{aligned} h_{top}\left(\mathcal{B}_{2,1/q}\right) &= \lim_{n\to\infty} \frac{\log p\,(qn)}{qn} \\ &\ge \lim_{n\to\infty} \frac{\log q^n}{qn} \\ &= \frac{1}{q} \log q. \end{aligned}$$

Corollary 6.1.2 *The upper bound on* $h_{top}\left(\mathcal{B}_{2,1/q}\right)$ *is maximal when* $q = 2$.

Proof. The maximum of the function

$$f(x) = \frac{1}{x} \log \left(\frac{x + 2 + \sqrt{(q+2)^2 - 12}}{2} \right).$$

occurs at $x = 2$. This proves the Corollary since $f(q)$ is the upper bound in the Theorem.

■

7 A BEGINNING TO SYMBOLICS IN $\mathbf{Z}^2$

We would like to generalize the results in Section 5.1 to higher dimensions. This turns out to be a more difficult subject because adding dimensions significantly add to the complexity (both in terms of balancedness and the complexity function $p(n)$) of the symbolic system. This is a subject for further study, but we would like to suggest a speculative notion of balancedness for $\mathbf{Z}^2$.

7.1 Balancedness in $\mathbf{Z}^2$

For a word $w \in \{0,1\}^n$, let the **height** $h(w) = \sum_{i=0}^{n-1} w_i$ of w be the number of 1's in w (as before). For a finite rectangle $R = \{r_{ij}\} \in \{0,1\}^{n \times m}$, let the **box-height** $h_{box}(R) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} r_{ij}$ of R be the number of 1's in the rectangle R.

Definition 7.1.1 *An array $\{u_{ij}\} \in \{0,1\}^{\mathbf{Z}_+^2}$ is* ***balanced*** *if the difference of the box-heights of any two rectangles of the same size have sum less than or equal to* 1. *That is, if R, S are $n \times m$ submatrices of $\{u_{ij}\}$, then $|h_{box}(R) - h_{box}(S)| \leq 1$.*

Proposition 7.1.1 *Let $u \in \{0,1\}^{\mathbf{Z}_+}$ be a balanced sequence and let σ be the left-shift map. Then the infinite array $A = \{a_{ij}\}$ whose ith row is $\sigma^i(u)$ is balanced.*

Proof. Note that the jth column of A is $\sigma^j(u) = u_j u_{j+1} u_{j+2} \cdots$. Let $R = \{a_{ij}\}_{\substack{i=n_1 \ldots m_1 \\ j=n_2 \ldots m_2}}$ and $R' = \{a_{ij}\}_{\substack{i=n_1' \ldots m_1' \\ j=n_2' \ldots m_2'}}$ be $n \times m$ submatrices of A. Since $\sigma^i(u)$ is balanced for all $i \geq 0$, we have that

$$\begin{aligned}
\left|h_{box}\left(R\right) - h_{box}\left(R'\right)\right| &= \left|\sum_{i=n_1}^{m_1} \sum_{j=n_2}^{m_2} a_{ij} - \sum_{i=n_{1'}}^{m'_1} \sum_{j=n'_2}^{m'_2} a_{ij}\right| \\
&= \left|\sum_{i=n_1}^{m_1} h\left(\sigma^j\left(u\right)\right) - \sum_{i=n'_1}^{m'_1} h\left(\sigma^j\left(u\right)\right)\right| \\
&\leq 1
\end{aligned}$$

since u is balanced. ■

From here, one can ask all the questions as in the one-dimensional case, and even more. For example, the theory of "Hankel" matrices (see, for example, [11]) might be very useful for this study.

8 CONCLUSION AND FUTURE DIRECTIONS

There are currently many unanswered questions about symbolic dynamical systems in general. Most noteworthy is the exploration of the complexity function $p(n)$ (a nice survey can be found in [8]). In this paper, we have explored methods of computing the topological entropy of certain sequences, which can give information about $p(n)$. We have also explored a general balancedness of sequences and recoverability of substitutions. These various notions of complexity in a symbolic dynamical system all help to understand how complex and interesting a space is. We have looked at sequences as paths in the plane, which give a nice geometric description of what a geometry of these symbolic dynamical systems may look like.

Recall from Theorem 4.2.2 that we could only obtain the result that the FibMorse substitution had positive entropy. The difficulty in computing the exact entropy lies both in the fact that it is not a constant-length substitution and in solving a non-linear recurrence relation. A general study in this regard of non-constant length substitutions could prove beneficial, as well as understanding the separate study of non-linear recurrence relations.

We would actually like to fully generalize this concept. For example, in the FibMorse substitution, there is a deterministic part ($0 \to 01$, no uncertainty) and a non-deterministic part (1 maps to two different words with nonzero probabilities). It's possible, for example, that the dynamical system arising from the FibMorse substitution is isomorphic to some other dynamical system. This would mean that the other dynamical system is, in some sense, the "sum" of deterministic (ie, zero entropy) and non-deterministic (ie, positive entropy) factors. Could it be that this positive-entropy part is a Bernoulli shift as in Example 4.2.2? Since Bernoulli shifts have positive entropy, this would yield an entropic characterization of dynamical systems (we note that these are not new questions, but we would like to study them from the specific viewpoint of random substitutions). In a similar vein, there seems to be a mathematical connection between the frequency of symbols and entropy (as illustrated in Examples 3.2.3 and 5.1.4, by Theorem 3.2.2). In another, but similar, direction, is entropy a *complete* invariant among random substitutions? This could be a very interesting study indeed!

Balancedness is a concept that can analytically lead to results about the structure of words, as the results in Section 5.1. Does balancedness imply a more general condition on the structure of words? Or does k-balancedness for $k \geq 2$ imply positive entropy/ That is, does this ensure exponential word growth? Is there a general connection between balancedness and the complexity function $p(n)$? We only gave a very brief introduction to what balancedness means in 2-dimensions (Section 7.1). A general study of balancedness in $\mathbf{Z}^d$, for $d \geq 2$ could be very interesting and enlightening. Also, we would like to explore the entropy of spaces of lattice paths defined by balancedness and slope, as in Section 6.1, in more detail.

There are many interesting questions and concepts involved with these notions of complexity and the different ways to view them. It is the author's hope that answers to these questions could lead to a better understanding and description of symbolic dynamical systems.

BIBLIOGRAPHY

1. Adler, Roy L., Konheim, A. G., and McAndrew, M. H., “Topological Entropy”, Trans. Amer. Math. Soc., **114** (1965), 309-319.

2. A.V. Aho and Sloane, N. J. A., “Some Doubly Exponential Sequences”, Fibonacci Quearterly, **11** (1970), 429-437.

3. Berstel, Jean, “Recent Results On Extensions of Sturmian Words”, International Journal of Algebra and Computation, **12** (2002), No. 1-2, 371-385.

4. Berstel and Seebold, “Sturmian Words” (chapter 2), Algebraic Combinatorics on Words. Lothaire, M. (2002). Cambridge UK: Cambridge University Press. ISBN 0521812208.

5. Coven, Ethan M. and Hedlund, Gustav. A., “Sequences with Minimal Block Growth”, Math. Systems Theory, **7** (1973), 138-153.

6. “Regularity and Irregularity of Sequences Generated by Automata”, Sém. Th. Nombres Bordeaux 1979-1980, 901-910.

7. Dekking, F. M. and Keane, M., “Mixing Properties of Substitutions”, Z. Wahrscheinlichkeitstheorie verw. Gebiete, **42** (1978), 23-33.

8. Ferenczi, Sébastien, “Complexity of sequences and dynamical systems”, Discrete Mathematics, **206** (1999), Issues 1-3, 145-154.

9. Morse, Marston and Hedlund, Gustav. A,“’Symbolic Dynamics II: Sturmian Trajectories”, Amer. J. Math., **62** (1940) No. 1, 1-42.

10. Khinchine, A. I., “Mathematical Foundations of Information Theory”, 1957

11. Gantmacher, F. R., The Theory of Matrices, Vol. 1, 1959, AMS Chelsea Publishing.

12. C. Godrèche, J. M. Luck, “Quasiperiodicity and randomness in tilings of the plane”, Journal of Statistical Physics, Volume 55, Issue 12, 128.

13. Lind, Doug and Marcus, Brian, An Introduction to Symbolic Dynamics and Coding, (1995), Cambridge University Press.

14. “Topics in Symbolic Dynamics and Applications”, London Mathematical Society Lecture Note Series #279.

15. “Substitutions in Dynamics, Arithmetics, and Combinatorics”, Lecture Notes In Mathematic s #1794.

16. L. D. Melshalkin, “A Case of Isomorphism of Bernoulli Schemes”, Dokl. Akad. Nauk. SSSR, **128** (1959), 31-44.

17. J. Nilsson, “On the Entropy of Random Fibonacci Words”, arXiv:1001.3513.

18. J. Nilsson, “On the Entropy of Generalised Random Fibonacci Words”, arXiv:1103.4777.

19. Ornstein, Donald S., “Bernoulli Shifts With the Same Entropy Are Isomorphic”, Adv. in Math., **4**, 337-352.

20. Petersen, Karl, Ergodic Theory, (1983), Cambridge University Press.

21. Queffélec, Martine, Substitution Dynamical Systems - Spectral Anaysis, (1987), Lecture Notes in Mathematics #1294, Springer-Verlag.

22. Riordan, John, “An Introduction to Combinatorial Analysis”, John Wiley, N. Y., 1958.

23. Shannon, C. E., “A Mathematical Theory of Communication”, Bell System Tech. J., **27** (1948), 379-423, 623-656.

24. Shields, Paul, The Theory of Bernoulli Shifts, (1973), The University of Chicago Press.

25. Walters, Peter, An Introduction to Ergodic Theory, (1982), Springer-Verlag.

INDEX