Symbolic Dynamics and Substitutions: from a to b

By

Svetlana Mokhnach

Advisor: Dr. Lori Alvin

An Undergraduate Proseminar
In Partial Fulfillment of the Degree of
Bachelor of Science in Mathematical Sciences
The University of West Florida
December 2012

Approval Page

This Proseminar of Svetlana Mokhnach is approved:

Lori Alvin, Ph.D., Proseminar Advisor	Date
Jossy Uvah, Ph.D., Proseminar Committee Chair	Date

Accepted for the Department:

Jaromy Kuhl, Ph.D., Chair	Date

ABSTRACT

Symbolic Dynamics and Substitutions: from a to b

The area of symbolic dynamics is an active and fast-growing part of dynamical systems. We focus on the subarea of substitutive dynamical systems. We generate infinite binary strings over the alphabet $\{a, b\}$ using a variety of substitution mappings and explore patterns that arise. We will also discuss famous substitutions such as the Fibonacci, Thue-Morse, and Cantor substitutions.

Table of Contents

Title Page i

Approval Page ii

Abstract iii

Table of Contents iv

1 Introduction **1**

1.1 Statement of Problem 1

1.2 Relevance of Problem 1

1.3 Literature Review 2

2 Main Body **2**

2.1 Basic Notation 2

2.2 Topology 4

2.3 Substitutions 5

2.4 The Fractal-like Nature of Fixed Points 7

2.5 Famous Substitutions 7

2.5.1 Fibonacci substitution 7

2.5.2 Thue-Morse substitution 12

2.5.3 2^∞ Sequence 15

2.6 Comparing Fixed Points 17

3 Conclusion **19**

3.1 Summary 19

3.2 Suggestions For Further Study 19

References **21**

1 Introduction

1.1 Statement of Problem

The area of symbolic dynamics is an active and fast-growing part of dynamical systems. When working in a dynamical system, it is often convenient to think about the system combinatorically, reducing the behavior of the system under iteration to sequences of letters, each letter representing a different state under consideration. In this paper we focus on the subarea of substitutive dynamical systems. That is, we look at dynamical systems that are generated by mappings that take one letter to a finite word. We are interested in the infinite-length fixed points that are obtained from infinite iterations of the mapping applied to a letter.

In our exploration of substitutions, we looked at several famous substitutions, including the Fibonacci substitution, the Thue-Morse substitution, the 2^∞ sequence, and the Cantor substitution. We explore patterns that arise within the fixed points of these substitutions. We also explore some higher level topics such as the topology of the space of all possible infinite sequences under a fixed alphabet and the minimality of specific sequences.

1.2 Relevance of Problem

When describing the motion of the planets or the trajectory of a molecule over time, we can model the behavior by a dynamical system. Although the motion here occurs continuously, we can look at the state of the system at fixed intervals of time, thus discretizing the system. If we additionally represent certain states symbolically, we can form an infinite sequence that represents the behavior of the system at each tick of the clock. By analyzing the symbolic representation of the system, we can gain an understanding of the long-term behavior of the object under consideration, and analyze any patterns that arise.

Morse and Hedlund, forefathers of the field of symbolic dynamics, explained the significance of the area as follows [5].

The methods used in the study of recurrence and transitivity frequently combine classical differential analysis with a more abstract symbolic analysis. This involves a characterization of the ordinary dynamical trajectory by an unending sequence of symbols termed a symbolic trajectory such that the properties of recurrence and transitivity of the dynamical trajectory are reflected in analogous properties of its symbolic trajectory.

The dynamical properties exhibited by physical phenomena also occur in seemingly simple symbolic systems. By generating various symbolic systems and observing the associated dynamics, we will gain a better understanding of the many behaviors that can occur in the world around us. Thus, our goal is to better understand those symbolic systems generated by substitutions.

1.3 Literature Review

For a main understanding of symbolic dynamics and substitutions, we relied on two texts. The first text, *An Introduction to Symbolic Dynamics and Coding* by Lind and Marcus [4], provides a friendly introduction to the field of symbolic dynamics and discusses many ways in which symbolic systems can be generated. *Substitutions in Dynamics, Arithmetics, and Combinatorics* by Fogg [3] provides an in depth exploration of substitution systems, with specific examples including the Fibonacci, Thue-Morse, and Cantor substitutions. For a general book about dynamical systems, see *Topics from One-Dimensional Dynamics* [2]. We also utilized several papers about symbolic dynamics including [1, 5, 6].

2 Main Body

2.1 Basic Notation

In this section, we use the notation and terminology from [4].

Sequences are ordered lists made up of numbers, letters, or symbols which are elements of our finite *alphabet* set $\mathcal{A}$. Usually, we will use a binary alphabet where $\mathcal{A} = \{0, 1\}$ or $\mathcal{A} = \{a, b\}$, but alphabets can be of any size, including infinite. Sequences of length k, also called *words* of length k or k-blocks, are all sequences u such that $|u| = k$. We will denote the set of all finite words over $\mathcal{A}$ by $\mathcal{A}^*$. The *empty word*, or empty block, ϵ is used to refer to a sequence of length 0. The empty word allows concatenation of words to be well defined, such that for all words u, $\epsilon u = u\epsilon = u$, and for all words u and v, $|uv| = |vu| = |u| + |v|$. We will also use the notation $|u|_x$ to denote the number of times the letter x appears in the word u.

Infinite sequences are denoted $u = (u_n)_{n\in\mathbb{N}}$ in $\mathcal{A}^{\mathbb{N}}$, and bi-infinite sequences are denoted $u = (u_n)_{n\in\mathbb{Z}}$ in $\mathcal{A}^{\mathbb{Z}}$; note here $\mathbb{N} = \{0, 1, 2, 3, \dots\}$. Infinite sequences begin with an index of 0, such that

$$u = u_0u_1u_2u_3u_4u_5u_6u_7\dots,$$

while bi-infinite sequences are usually written with a decimal point to separate the u_i with $i \geq 0$ and u_i with $i < 0$, such that

$$u = \dots u_{-3}u_{-2}u_{-1}.u_0u_1u_2u_3\dots$$

These infinite sequences u can also be referred to as points in $\mathcal{A}^{\mathbb{N}}$ or $\mathcal{A}^{\mathbb{Z}}$. If $\mathcal{A}$ is a finite alphabet, then the full $\mathcal{A}$-shift is the collection of all bi-infinite sequences of symbols from $\mathcal{A}$.

The *shift map* s on the full shift $\mathcal{A}^{\mathbb{Z}}$ maps a point x to the point $y = s(x)$ whose i-th coordinate is $y_i = x_{i+1}$. Thus, the point

$$x = \dots x_{-3}x_{-2}x_{-1}.x_0x_1x_2x_3\dots$$

becomes

$$y = s(x) = \dots x_{-3}x_{-2}x_{-1}x_0.x_1x_2x_3\dots$$

The decimal point moves one to the right, effectively shifting the sequence one to the left. The inverse s^{-1} shifts the sequence one to the right. Thus, s is one-to-one and onto when defined on the full shift $\mathcal{A}^{\mathbb{Z}}$. The shift map s can also be defined on $\mathcal{A}^{\mathbb{N}}$. For a point $x \in \mathcal{A}^{\mathbb{N}}$, $s(x)$ will also be defined such that the i-th coordinate of $y = s(x)$ is $y_i = x_{i+1}$; however, since $y \in \mathcal{A}^{\mathbb{N}}$, we note $y_0 = x_1$ is the first coordinate of y. Thus, the point

$$x = x_0x_1x_2x_3x_4x_5x_6x_7\dots$$

becomes

$$y = s(x) = x_1x_2x_3x_4x_5x_6x_7\dots$$

Point u is *periodic* if there exists $t \in \mathbb{N}$ such that for all $n \in \mathbb{Z}$, $s^n(u) = s^{n+t}(u)$. The smallest possible such t is called the *period* of u. Points are *eventually periodic* if there exists $i \in \mathbb{Z}$ such that $u = (u_n)_{n\geq i}$ is periodic. The point u is *recurrent* if every word that appears in u appears infinitely often. Point u is *uniformly recurrent*, also called *almost periodic* or *minimal*, if every word occurring in u occurs in an infinite number of positions with bounded gaps between successive occurrences, i.e. for all $\epsilon > 0$, there exists N such that if $s^j(u) \in B(u, \epsilon)$, then $s^{j+k}(u) \in B(u, \epsilon)$ for some $k \leq N$. Note we will define an appropriate metric on our space of infinite sequences in Section 2.2.

2.2 Topology

For two points $u \neq v \in \mathcal{A}^{\mathbb{N}}$, the topology on $\mathcal{A}^{\mathbb{N}}$ is defined by the distance:

$$d(u, v) = d^{-\min\{n\in\mathbb{N}; u_n\neq v_n\}}$$

where $d = |\mathcal{A}|$. Hence, two points are near each other when their first few terms are equal. The space $\mathcal{A}^{\mathbb{N}}$ is a complete metric space. [3].

Theorem 1. *The set $\mathcal{A}^{\mathbb{N}}$ is a totally disconnected compact set without isolated points.*

Proof. Since $\mathcal{A}^{\mathbb{N}}$ is a complete metric space, to show that it is compact, we need to show that it is totally bounded. A totally bounded space can be covered in finitely many ϵ-balls for all ϵ. Fix $\epsilon > 0$. Let N be the smallest integer such that $\frac{1}{d^N} < \epsilon$. If x and y agree in the first $(N+1)$ positions, then $d(x, y) \leq \frac{1}{d^{N+1}} < \frac{1}{d^N} < \epsilon$. There are d^{N+1} possible initial blocks $x_0 x_1 \ldots x_N$, and thus d^{N+1} distinct words of the form $x_0 x_1 \ldots x_N (0)^\infty$. Choose the d^{N+1} ϵ-balls centered at each of these words. Hence, $\mathcal{A}^{\mathbb{N}}$ is compact.

Every point $x \in \mathcal{A}^{\mathbb{N}}$ is a limit point because we can construct a sequence of points (U_n) that limit to x, with each $U_n \neq x$. If $x = x_0 x_1 x_2 x_3 \ldots$, then let $U_n = x_0 x_1 x_2 x_3 \ldots x_n \ldots$, where each U_n matches x up to index n. Since x is arbitrary, and we found a sequence that converges to it, $\mathcal{A}^{\mathbb{N}}$ does not contain any isolated points and every point in $\mathcal{A}^{\mathbb{N}}$ is a limit point.

To prove that $\mathcal{A}^{\mathbb{N}}$ is totally disconnected, we need to show that for all points $x, y \in \mathcal{A}^{\mathbb{N}}$, there exist separated sets A and B with $x \in A$, $y \in B$, and $\mathcal{A}^{\mathbb{N}} = A \cup B$. Two sets $A, B \subseteq \mathcal{A}^{\mathbb{N}}$ are separated if $\overline{A} \cap B$ and $A \cap \overline{B}$ are both empty, where $\overline{S}$ denotes the closure of the set S. Let $x, y \in \mathcal{A}^{\mathbb{N}}$ be two distinct arbitrary points. Let u_n be the first coordinate of x that differs from y. Then $d(x, y) = d^{-n}$. Let $A = \{a \in \mathcal{A}^{\mathbb{N}} | d(a, x) < d^{-n}\} = \{a \in \mathcal{A}^{\mathbb{N}} | d(a, x) \leq d^{-(n+1)}\}$ and $B = \{a \in \mathcal{A}^{\mathbb{N}} | d(a, x) \geq d^{-n}\}$. Then $x \in A$ since $d(x, x) = 0$, and $y \in B$ since $d^{-n} \geq d^{-n}$. We know that $\mathcal{A}^{\mathbb{N}} = A \cup B$ because all points in $\mathcal{A}^{\mathbb{N}}$ are either less than d^{-n} distance from x or greater than or equal to d^{-n} distance from x. Both A and B are closed because they contain all of their limit points. Hence, A and B are separated, and we can conclude that $\mathcal{A}^{\mathbb{N}}$ is a totally disconnected. $\square$

After briefly exploring the set of all infinite sequences $\mathcal{A}^{\mathbb{N}}$, let us now focus on the subarea of substitutive symbolic dynamics and look at which points in $\mathcal{A}^{\mathbb{N}}$ are generated by substitutions.

2.3 Substitutions

In this section, we use the notation and terminology from [3].

Substitutions are mappings σ from the alphabet $\mathcal{A}$ into the set $\mathcal{A}^* - \{\epsilon\}$. The notation for a substitution can be written as both $\sigma(u_0)$ or $u_0 \to \sigma(u_0)$ for some $u_0 \in \mathcal{A}$. Substitutions can be of constant length k if $|\sigma(u)| = k$ for all $u \in \mathcal{A}$ or can be of varying lengths. Substitutions extend to words by concatenation such that for letters $u_0, u_1 \in \mathcal{A}$, $\sigma(u_0 u_1) = \sigma(u_0)\sigma(u_1)$ and $\sigma(\epsilon) = \epsilon$. If $\sigma(u) = u$ for a point $u \in \mathcal{A}^{\mathbb{N}}$, then u is a *fixed point* for σ. A substitution will have a fixed point if there exists $a \in \mathcal{A}$ such that $\sigma(a)$ begins with a. If there are multiple such $a \in \mathcal{A}$, then there will be multiple fixed points for a given substitution.

There are two main ways to construct fixed points of substitutions. The clearest way, and the way that easily leads to being able to prove patterns about the substitutions by induction, is to think of iterations of substitutions upon a starting letter $u_0 \in \mathcal{A}$. Thus, $\sigma(u_0)$ is the first iteration, $\sigma^2(u_0)$ is the second iteration where σ is performed on each letter of $\sigma(u_0)$ using the concatenation property, etc., with the fixed point being

$$\lim_{n\to\infty} \sigma^n(u_0).$$

The other way to construct a fixed point is to begin with the letter u_0 such that $\sigma(u_0)$ begins with u_0. If $|\sigma(u_0)| = n$, then we write out $\sigma(u_0)$ as $u_0 u_1 \dots u_n$ and then append $\sigma(u_1)$ on the end, followed by $\sigma(u_2)$, up to $\sigma(u_n)$. Then, instead of starting at the beginning as in with the previous method, we continue on by appending $\sigma(\sigma(u_1))$ one letter at a time. This method generates the same fixed point as the other method if started with the same letter u_0 and is faster in computation time.

Every substitution can be also be associated with an *incidence matrix*. If σ is a substitution defined over an alphabet $\mathcal{A} = \{a_1, \dots, a_d\}$ of cardinality d, then the incidence matrix is the $d \times d$ matrix M_σ with $|\sigma(a_j)|_{a_i}$ in the entry of index (i, j), i.e. the matrix counts the number of occurrences of a_i in $\sigma(a_j)$. For every $n \in \mathbb{N}$, $|\sigma^n(a_j)|_{a_i}$ is equal to the coefficient of index (i, j) in the matrix M_σ^n.

2.4 The Fractal-like Nature of Fixed Points

The concatenation property of substitutions mentioned above, i.e. for all letters $u_0, u_1 \in \mathcal{A}$, $\sigma(u_0u_1) = \sigma(u_0)\sigma(u_1)$, leads to fractal-like behavior when we consider fixed points $u \in \mathcal{A}^{\mathbb{N}}$ of constant length substitutions.

Consider the substitution $\sigma(a) = aba$ and $\sigma(b) = bbb$. The fixed point u that begins with a for this substitution is

$$u = ababbbababbbbbbbbbababbbaba\ldots.$$

If we group letters together, assigning $A = aba$ and $B = bbb$, then our fixed point is

$$u = ABABBBABABBBBBBBBBABABBBABA\ldots.$$

The two sequences contain all the same letters in the same order. In fact, this holds for any power of k for a substitution of constant length k. For example, we could also assign $A = ababbbaba$ and $B = bbbbbbbbb$ to obtain $u = ABABBBABABBBBBBBBBABABBBABA\ldots$ as well.

2.5 Famous Substitutions

Let us examine a few famous substitutions.

2.5.1 Fibonacci substitution

The Fibonacci substitution is generated by the substitution $\sigma(a) = ab$ and $\sigma(b) = a$. The first few iterations when beginning with a are

$$a \to ab \to aba \to abaab \to abaababa \to \ldots$$

This substitution has only one fixed point

$$u = abaababaabaababaababaabaababaabaababaababaabaababaababa\ldots$$

The Fibonacci substitution has several interesting properties that inspired its namesake.

Proposition 2. *The n-th iteration of the fixed point beginning with a has length F_n, with F_{n-1} a's and F_{n-2} b's, where F_n is the n-th Fibonacci number, defined as $F_0 = 1$, $F_1 = 2$, and $F_n = F_{n-1} + F_{n-2}$ for all $n \geq 2$.*

Proof. We can prove this by mathematical induction. When $n = 2$, $|\sigma^2(a)| = |aba| = 3$, $|\sigma^2(a)|_a = |aba|_a = 2$, and $|\sigma^2(a)|_b = |aba|_b = 1$. Since, $F_0 = 1$, $F_1 = 2$, and $F_2 = 3$, the base case step is confirmed. Suppose the k-th iteration of the fixed point beginning with a has length F_{k+1}, with F_k a's and F_{k-1} b's. Consider the process of generating the next iteration. All of the a's in the k-th iteration, of which there are F_k, become ab. All of the b's from the k-th iteration, of which there are F_{k-1}, become a's. Hence, the $(k+1)$-th iteration has $F_{k-1} + F_k = F_{k+1}$ a's and F_k b's, and $F_{k+1} + F_k = F_{k+2}$ length total. Therefore, the n-th iteration of the fixed point beginning with a has length F_{n+1}, with F_n a's and F_{n-1} b's, where F_n is the n-th Fibonacci number. □

As a direct consequence of this Proposition and the way Fibonacci numbers are defined, we now have that for all $n \in \mathbb{N}$,

$$|\sigma^{n+2}(a)| = |\sigma^{n+1}(a)| + |\sigma^n(a)|$$

$$|\sigma^{n+2}(a)|_a = |\sigma^{n+1}(a)|_a + |\sigma^n(a)|_a$$

$$|\sigma^{n+2}(a)|_b = |\sigma^{n+1}(a)|_b + |\sigma^n(a)|_b.$$

Proposition 3. *Successive iterations end in different letters, i.e. the last letter of $\sigma^n(a)$ is the opposite of the last letter of $\sigma^{n+1}(a)$.*

Proof. Assume the last letter of $\sigma^n(a)$ was a. Since $\sigma(a) = ab$, the last letter of $\sigma^{n+1}(a)$ will now be b. Now assume the last letter of $\sigma^n(a)$ was b. Since $\sigma(b) = a$, the last letter of $\sigma^{n+1}(a)$ will now be a. Since those are the only two cases, we have shown that successive iterations will end in different letters. □

Theorem 4. *For all $k \in \mathbb{N}$, $u_{F_k-1} = 1 - u_{F_{k+1}-1}$ where $u \in \mathcal{A}^{\mathbb{N}}$ is the fixed point of the Fibonacci substitution.*

Proof. Since our alphabet $\mathcal{A}$ is binary and can be equivalently defined by $\{a, b\}$ and $\{0, 1\}$ with $a \sim 0$ and $b \sim 1$, the statement of the theorem translates as the letter in position $F_k - 1$ is opposite of the letter in position $F_{k+1} - 1$ in the fixed point u of the Fibonacci substitution. By Proposition 2, the n-th iteration of the fixed point beginning with a has length F_{n+1}, so the letter $u_{F_{n+1}-1}$ occurs in the last position of the n-th iteration and the letter u_{F_n-1} occurs in the last position of the $(n-1)$-th iteration. By Proposition 3, successive iterations end in different letters. Hence, u_{F_n-1} and $u_{F_{n+1}-1}$ are opposites, i.e. for all $k \in \mathbb{N}$, $u_{F_k-1} = 1 - u_{F_{k+1}-1}$ where $u \in \mathcal{A}^{\mathbb{N}}$ is the fixed point of the Fibonacci substitution. □

The *Zeckendorff numeration system* represents every nonnegative integer n in a unique way as

$$n = \sum_{i \geq 0} n_i \cdot F_i$$

with $n_i \in \{0, 1\}$ and $n_i \cdot n_{i+1} = 0$ for any $i \geq 0$ [3]. That is, in this numeration system, no two consecutive terms both equal 1. This should be intuitive as the sum of two consecutive Fibonacci numbers is again a Fibonacci number, and we want to write each non-negative integer in what is often referred to as the "greedy algorithm."

Proposition 5. *Every $n \in \mathbb{N}$ can be represented as $n = \sum_{i \geq 0} n_i \cdot F_i$ with $n_i \in \{0, 1\}$ and $n_i \cdot n_{i+1} = 0$ for any $i \geq 0$.*

Proof. (Adapted from [3].) We can prove by mathematical induction that such a decomposition exists for all $n \in \mathbb{N}$. When $n = 0$, $0 = 0 \cdot F_0 = 0 \cdot 1$. When $n = 1$, $1 = 1 \cdot F_0 = 1 \cdot 1$. When $n = 2$,

$2 = 1 \cdot F_1 + 0 \cdot F_0 = 1 \cdot 2 + 0 \cdot 1$. Hence, the base case step is confirmed. Suppose that this property is true for all $n < F_k$ such that $k \geq 2$, i.e. every $n < F_k$ can be represented as $\sum_{i=0}^{k-1} n_i \cdot F_i$. Let $F_k \leq n < F_{k+1}$. Since the Fibonacci sequence is defined to have the property $F_{k+1} = F_k + F_{k-1}$, we now have that $n < F_k + F_{k-1}$, or $n - F_k < F_{k-1}$. By the inductive hypothesis, we now know that $n - F_k = \sum_{i=0}^{k-2} n_i \cdot F_i$. Hence, $n = F_k + \sum_{i=0}^{k-2} n_i \cdot F_i$, and the property is true by induction for all k. □

We now show another result about the Zeckendorff numeration system that will help us prove its uniqueness for all $n \in \mathbb{N}$.

Proposition 6. *For $n_i \in \{0, 1\}$ and $n_i \cdot n_{i+1} = 0$ for any $i \geq 0$, $\sum_{i=0}^{k} n_i \cdot F_i < F_{k+1}$.*

Proof. (Adapted from [3].) We can prove this by mathematical induction. When $k = 0$, both $0 \cdot F_0 = 0 \cdot 1 = 0$ and $1 \cdot F_0 = 1 \cdot 1 = 1$ are less than $F_1 = 2$. Suppose that this property is true for all $k < m$. If $n_m = 0$, then $\sum_{i=0}^{m} n_i \cdot F_i = \sum_{i=0}^{m-1} n_i \cdot F_i$. This is smaller than F_m by our hypothesis. If $n_m = 1$, then $n_{m-1} = 0$ to satisfy the restraint $n_i \cdot n_{i+1} = 0$ for any $i \geq 0$. Then $\sum_{i=0}^{m} n_i \cdot F_i = F_m + 0 + \sum_{i=0}^{m-2} n_i \cdot F_i$. But by our hypothesis, $\sum_{i=0}^{m-2} n_i \cdot F_i < F_{m-1}$. Hence, $\sum_{i=0}^{m} n_i \cdot F_i < F_m + F_{m-1} = F_{m+1}$. Therefore, $\sum_{i=0}^{k} n_i \cdot F_i < F_{k+1}$ for all $k \in \mathbb{N}$. □

By combining our results from Propositions 5 and 6, we are now ready to prove that the numeration system represents nonnegative integers in a unique way.

Theorem 7. *Every nonnegative integer $n \in \mathbb{N}$ can be represented in a unique way as $n = \sum_{i \geq 0} n_i \cdot F_i$ with $n_i \in \{0, 1\}$ and $n_i \cdot n_{i+1} = 0$ for any $i \geq 0$.*

Proof. (Adapted from [3].) Suppose that there are two representations $\sum_{i \geq 0} n_i \cdot F_i = \sum_{i \geq 0} n_i' \cdot F_i$ with $n_i, n_i' \in \{0, 1\}$ and $n_i \cdot n_{i+1} = n_i' \cdot n_{i+1}' = 0$ for any $i \geq 0$. Let k be the highest index where the two expansions disagree, i.e. $n_k \neq n_k'$ and $n_i = n_i'$ for all $i > k$, since we can simply subtract $\sum_{i>k} n_i \cdot F_i = \sum_{i>k} n_i' \cdot F_i$ from both sides. Without loss of generality, suppose $n_k = 1$ and $n_k' = 0$. Then, $\sum_{i=0}^{k} n_i \cdot F_i \geq F_k$, and $\sum_{i=0}^{k} n_i' \cdot F_i = \sum_{i=0}^{k-1} n_i' \cdot F_i < F_k$ by Proposition 6. This is a contradiction and implies that $\sum_{i\geq 0} n_i \cdot F_i \neq \sum_{i\geq 0} n_i' \cdot F_i$. Hence, the expansions are unique for all $n \in N$. □

Let $\mathbb{F}_a$ be the set of integers n such that the n-th letter of the Fibonacci fixed point is a and $\mathbb{F}_b$ be the set of integers n such that the n-th letter of the Fibonacci fixed point is b. Hence,

$$\mathbb{F}_a = \{0, 2, 3, 5, 7, 8, 10, 11, 13, 15, \ldots\}, \text{ and}$$

$$\mathbb{F}_b = \{1, 3, 6, 9, 12, 14, 17, 19, 22, 25, \ldots\}.$$

If $n = \sum_{i=0}^{k} n_i \cdot F_i$ with $n_k = 1$, $n_i \in \{0, 1\}$ and $n_i \cdot n_{i+1} = 0$ for any $i \leq k$, we define $\text{Fib}(n) = n_k n_{k-1} \ldots n_0$ as the Fibonacci expansion of the positive integer n. We can write out the Fibonacci expansion of each of the above numbers and notice a pattern.

$$\mathbb{F}_a = \{0, 10, 100, 1000, 1010, 10000, 10010, 10100, 100000, 100010, \ldots\}$$

$$\mathbb{F}_b = \{1, 101, 1001, 10001, 10101, 100001, 100101, 101001, 1000001, 1000101, \ldots\}$$

Theorem 8. *All of the $n \in \mathbb{F}_a$ have $n_0 = 0$, and $n \in \mathbb{F}_b$ have $n_0 = 1$ in their Fibonacci expansion.*

Proof. From Theorem 4, we know that for all $k \in \mathbb{N}$, $u_{F_k - 1} = 1 - u_{F_{k-1} - 1}$ where u is the Fibonacci fixed point. Hence, $u_{F_k - 1} = u_{F_{k-2} - 1}$. Therefore, if $n = \sum_{i=0}^{k} n_i \cdot F_i$ with $n_k = 1$, $n_i \in \{0, 1\}$ and $n_i \cdot n_{i+1} = 0$ for any $i \leq k$, $n \in \mathbb{F}_a$ implies that $\sum_{i=0}^{k-2} n_i \cdot F_i = n - F_k \in \mathbb{F}_a$ as well, and similarly for when $n \in \mathbb{F}_b$. Since $0 \in \mathbb{F}_a$ and $1 \in \mathbb{F}_b$, then $n \in \mathbb{F}_a$ if and only if $n_0 = 0$ in $\text{Fib}(n)$, and $n \in \mathbb{F}_b$ if and only if $n_0 = 1$ in $\text{Fib}(n)$. □

We now study another famous substitution.

2.5.2 Thue-Morse substitution

The Thue-Morse substitution is generated by the substitution $\sigma(a) = ab$ and $\sigma(b) = ba$. The first few iterations when beginning with a are

$$a \to ab \to abba \to abbabaab \to abbabaabbaababba \to \dots$$

There are two fixed points

$$x = abbabaabbaababbabaababbaabbabaabbaababbaabbabaababbabaabbaababba \dots$$

and

$$y = baababbaabbabaababbabaabbaababbaabbabaabbaababbabaababbaabbabaab \dots$$

Theorem 9. *For all $n \in \mathbb{N}$, $u_{2n+1} = 1 - u_n$ where $u \in \mathcal{A}^{\mathbb{N}}$ is a fixed point of the Thue-Morse substitution.*

Proof. Since our alphabet $\mathcal{A}$ is binary and can be equivalently defined by $\{a, b\}$ or $\{0, 1\}$ with $a \sim 0$ and $b \sim 1$, the statement of the theorem translates as the letter in position $2n+1$ is opposite of the letter in position n in a fixed point u of the Thue-Morse substitution. This Theorem effectively states that each coordinate of the initial block of length 2^n is opposite the equivalent coordinate in the second block of length 2^n. That is, with each successive iteration in the construction of the fixed point, the first half of the word is exactly opposite the second half. We can prove this by mathematical induction, by inducting on k, the number of iterations when generating the fixed point. Without loss of generality, we can assume our fixed point begins with a, as $a \to ab$ and $b \to ba$.

When $k = 1$, $\sigma(a) = ab$. Since the letter in position u_0 is opposite the letter in position $u_{2 \cdot 0+1} = u_1$, or the first half of the word is opposite the second half of the word, the base case step is confirmed. Suppose that this property is true for all $n < 2^k - 1$, such that in the k-th iteration of generating the Thue-Morse fixed point that begins with a, which has length 2^k or highest letter index of $2^k - 1$, $u_{2n+1} = 1 - u_n$ or the first 2^{k-1} letters are opposite the second 2^{k-1} letters. Thus, the fixed point in the k-th iteration looks like

$$u_0 u_1 \dots u_{2^{k-1}-2} u_{2^{k-1}-1} (1-u_0)(1-u_1)\dots(1-u_{2^{k-1}-2})(1-u_{2^{k-1}-1}).$$

In the $(k+1)$-th iteration, the word is $2^{k+1} = 2 \cdot 2^k$ letters long. The first 2^k letters are identical to the previous iteration, since we are generating a fixed point. If we consider the second method of generating fixed points, appending the substitution of each letter onto the end of the starting letter, the first 2^k letters were generated by the first 2^{k-1} letters and the second 2^k letters were generated by the second block of length 2^{k-1}. Since, by our inductive hypothesis, we have that the first 2^{k-1} letters are opposite the second 2^{k-1} letters, and the substitution is generated by the mapping such that $\sigma(a)$ is opposite $\sigma(b)$, the word generated by the first 2^{k-1} letters will be opposite the word generated by the second 2^{k-1} letters. Thus, in the $(k+1)$-th iteration, the first 2^k letters are opposite the second 2^k letters. This is equivalent to considering the indices of the letters and getting that for all $n \in \mathbb{N}$, $u_{2n+1} = 1 - u_n$ where $u \in \mathcal{A}^{\mathbb{N}}$ is a fixed point of the Thue-Morse substitution. □

Let $\mathbb{B}_a$ be the set of integers n such that the letter of n-th index in the Thue-Morse fixed point that begins with a is a and $\mathbb{B}_b$ be the set of integers n such that the letter of n-th index in the Thue-Morse fixed point that begins with a is b. Hence,

$$\mathbb{B}_a = \{0, 3, 5, 6, 9, 10, 12, 15, 17, 18, 20, \dots\}, \text{ and}$$

$$\mathbb{B}_b = \{1, 2, 4, 7, 8, 11, 13, 14, 16, 19, \dots\}.$$

If we write each of the above numbers in binary, we notice a pattern.

$$\mathbb{B}_a = \{0, 11, 101, 110, 1001, 1010, 1100, 1111, 10001, 10010, 10100, \dots\}$$

$$\mathbb{B}_b = \{1, 10, 100, 111, 1000, 1011, 1101, 1110, 10000, 10011, \dots\}$$

Theorem 10. *Each $n \in \mathbb{B}_a$ has an even number of 1's in its binary expansion, while each $n \in \mathbb{B}_b$ has an odd number of 1's in its binary expansion.*

Proof. We can prove this by mathematical induction, by inducting on k, the number of iterations when generating the fixed point. Without loss of generality, we can assume our fixed point begins with a.

When $k = 1$, $\sigma(a) = ab$. The a is in index 0, which has an even number of 1's in its binary expansion, and the b is in index 1, which has an odd number of 1's in its binary expansion. Hence, the base case step is confirmed. Suppose that this property is true for all $n < 2^k - 1$, such that in the k-th iteration of generating the Thue-Morse fixed point that begins with a, all of the a's appear in indices that when represented in binary expansion have an even number of 1's and all of the b's appear in indices that when represented in binary expansion have an odd number of 1's.

In the $(k+1)$-th iteration, the highest index is $2^{k+1} - 1$. Consider the binary expansions of the numbers 0 through $2^k - 1$ and the numbers 2^k through $2^{k+1} - 1$. If they are all written in $(k+1)$ characters long, with enough leading 0's to fill in the powers of 2 that are higher than the number itself, then the binary expansions of the numbers 0 through $2^k - 1$ all begin with a 0 since they are all less than 2^k, and the binary expansions of the numbers 2^k through $2^{k+1} - 1$ all begin with a 1. By Theorem 9, we have that the letters in indices 0 through $2^k - 1$ are opposite the letters in indices 2^k through $2^{k+1} - 1$. By our inductive hypothesis, all of the a's appear in indices between 0 and $2^k - 1$ that when represented in binary expansion have an even number of 1's and all of the b's appear in those indices that when represented in binary expansion have an odd number of 1's. Combining these two facts with the comment on binary expansions, we get that an a in the index $n \in \{0, \ldots, 2^k - 1\}$ is opposite a b in the index $2^k + n \in \{2^k, \ldots, 2^{k+1} - 1\}$ and, since n had an even number of 1's in its binary expansion, $2^k + n$ will have an odd number of 1's in its binary expansion. Conversely, a b in the index $n \in \{0, \ldots, 2^k - 1\}$ is opposite an a in the index $2^k + n \in \{2^k, \ldots, 2^{k+1} - 1\}$ and, since n had an odd number of 1's in its binary expansion, $2^k + n$ will have an even number of 1's in its binary expansion. Hence, each $n \in \mathbb{B}_a$

has an even number of 1's in its binary expansion, while each $n \in \mathbb{B}_b$ has an odd number of 1's in its binary expansion. □

If we take a Thue-Morse fixed point and generate a new point by writing 0 when consecutive letters of the fixed point are equal and 1 when consecutive letters of the fixed point are not equal, i.e.

$$u_n = \begin{cases} 0 & \text{if } x_n = x_{n+1} \\ 1 & \text{if } x_n \neq x_{n+1} \end{cases}$$

we generate the sequence $101110101011101110111010101110101\ldots$. This is the famous 2^∞ sequence that we shall discuss more in depth in the next section.

2.5.3 2^∞ Sequence

The 2^∞ sequence is defined using $\mathcal{A} = \{0, 1\}$. We will introduce the notation that whenever

$$A = a_0 a_1 \ldots a_{2^n-2} a_{2^n-1},$$

$$A' = a_0 a_1 \ldots a_{2^n-2}(1 - a_{2^n-1}).$$

Hence, the words A and A' are identical everywhere except the last letter. The 2^∞ sequence is then generated recursively by

$$\lim_{n\to\infty} A_n,$$

where $A_0 = 1$ and $A_n = A_{n-1}A'_{n-1}$, where $A_{n-1}A'_{n-1}$ denotes appending the word A'_{n-1} on the end of the word A_{n-1} at each iteration [2].

The substitution $\sigma(0) = 11$, $\sigma(1) = 10$ generates the 2^∞ sequence as well. The first few iterations when beginning with 1 are

$$1 \to 10 \to 1011 \to 10111010 \to 1011101010111011 \to \ldots$$

There is only one fixed point

$$u = 1011101010111011101110101011101010111010101110111011101010111011\ldots$$

Theorem 11. *The fixed point of the substitution $\sigma(0) = 11$, $\sigma(1) = 10$ is the 2^∞ sequence.*

Proof. We can prove this by mathematical induction, by inducting on k, the number of iterations when generating the fixed point.

Notice that $A_0 = 1$, $A_0' = 0$. Thus, $\sigma(A_0) = \sigma(1) = 10 = A_0 A_0' = A_1$ and $\sigma(A_0') = \sigma(0) = 11 = A_0 A_0 = A_1'$. Suppose that this property is true for k, i.e. $A_k = \sigma(A_{k-1})$ and $A_k' = \sigma(A_{k-1}')$. Then, we have $\sigma(A_k) = \sigma(A_{k-1}A_{k-1}') = \sigma(A_{k-1})\sigma(A_{k-1}') = A_k A_k' = A_{k+1}$.

Since $A_k = \sigma(A_{k-1})$ for all $k \in \mathbb{N}$ and $A_0 = 1$, we can write that $A_k = \sigma^k(1)$. Hence,

$$\lim_{n\to\infty} A_n = \lim_{n\to\infty} \sigma^n(1),$$

and the 2^∞ sequence is the fixed point of the substitution $\sigma(0) = 11$, $\sigma(1) = 10$. $\square$

We can also construct the 2^∞ sequence in a third, seemingly unrelated, way that helps us see that the 2^∞ sequence is minimal. We begin with an infinite number of blank spaces and generate the 2^∞ sequence by placing a 1 in every other blank space:

$$1\_1\_1\_1\_1\_1\_1\_1\_1\_1\_1\_1\_1\_1\_1\_1\_1\_1\ldots$$

We then place a 0 in every other blank space:

$$101\_101\_101\_101\_101\_101\_101\_101\_1\ldots$$

We then place a 1 in every other blank space:

$$1011101\_1011101\_1011101\_1011101\_1\ldots$$

Continuing to alternate in this way, placing 0 in every other blank space and then placing 1 in every other blank space ad infinitum, we get the 2^∞ sequence:

$$101110101011101110111010101110101\ldots$$

A point u is a *Toeplitz sequence* if for all $k \in \mathbb{N}$, there exists a $p_k > 1$ such that $u_k = u_{k+n\cdot p_k}$ for all $n \in \mathbb{N}$ [1]. By construction, the 2^∞ sequence is Toeplitz and non-periodic. For a more detailed study on Toeplitz sequences, see [6]. In [1], it is shown that all non-periodic Toeplitz sequences are uniformly recurrent, and thus minimal. Hence, the 2^∞ sequence is minimal. We also note that the Thue-Morse and Fibonacci fixed points are both minimal [3].

2.6 Comparing Fixed Points

In the Thue-Morse substitution, the two fixed points are different from each other at every letter. We sought to explore the question of when do two fixed points of the same substitution have similarities and differences in general.

Consider the Cantor substitution, generated by the substitution $\sigma(a) = aba$ and $\sigma(b) = bbb$. There are two fixed points, $x = ababbbababbbbbbbbbababbbaba\ldots$ and $y = (b)^\infty$. The indices where the two words differ are the indices where the letter a appears in the fixed point x. The indices where the letter a appears is the set $\{0, 2, 6, 8, 18, 20, 24, 26, 54, 56, 60, 62, 72, 74, 78, 80, \ldots\}$. These numbers are all the numbers whose base 3 expansion contains no 1's.

Let us look at a different substitution $\sigma(a) = aab$ and $\sigma(b) = bbb$. There are two fixed points, $x = aabaabbbbaabaabbbbbbbbbbbb\ldots$ and $y = (b)^\infty$. The indices where the two words differ are once again the indices where the letter a appears in the fixed point x. The indices where the letter a appears is the set $\{0, 1, 3, 4, 9, 10, 12, 13, 27, 28, 30, 31, 36, 37, 39, 40, \ldots\}$. These numbers are all the numbers whose base 3 expansion contains no 2's.

There appears to be a pattern. When the substitution mapping for the two letters was the same only in index 1, the differences between the fixed points appeared in indices that contained no 1's in their base 3 expansion, and when the substitution mapping for the two letters was the same only in index 2, the differences between the fixed points appeared in indices that contained no 2's in their base 3 expansion. Note that we begin at index 0. This pattern extends to longer substitutions.

When a substitution is of fixed length k, with $\sigma(a) = aA$ and $\sigma(b) = bB$ where $A, B \in \mathcal{A}^*$ are two distinct words of length $(k-1)$ that are equal in the non-empty set of indices I, then the indices where the two fixed points will differ are all the numbers in base k expansion that contain no digits that are in I. For a substitution on a binary alphabet to have two fixed points, the 0-th index must be different. Thus, it is impossible to obtain a set of indices where the two fixed points differ that contains no 0's in their base k expansion. While we offer no proof of this conjecture here, we believe that the fractal-like nature of substitutions leads to this result.

If we now look at where the two fixed points are similar, we also find an interesting pattern. In the Cantor substitution, if we let $B_0 = b, B_1 = (b)^3, B_2 = (b)^9, \ldots, B_n = (b)^{3^n}, \ldots$, then the similarities appear in the pattern

$$B_0B_1B_0B_2B_0B_1B_0B_3B_0B_1B_0B_2B_0B_1B_0B_4B_0B_1B_0B_2B_0B_1B_0B_3B_0B_1B_0B_2B_0B_1\ldots,$$

where we have removed all the coordinates that disagree between the two fixed points. This looks like the Toeplitz version of the 2^∞ sequence construction except with an infinite alphabet. This Toeplitz-like behavior of similarities between fixed points appears to work in general as well.

3 Conclusion

3.1 Summary

Substitutions can be used to generate dynamical systems. We focused on various properties and patterns that emerge from the fixed points of several substitutions. In order to do this, we needed to gain an understanding of the way in which the fixed point of a substitution was created. We observed that there is a fractal-like property that arises within the construction of a fixed point and discussed some basic topological properties of sequences.

We then focused on several examples of substitutions and proved several associated properties. In particular, in the Fibonacci substitution section, we related the substitution to the Zeckendorff numeration system. Similarly, in the Thue-Morse substitution section, we demonstrated a connection between the fixed point and a binary numeration system. In studying the Thue-Morse sequence, we derived the 2^∞ sequence and discussed the substitution that generates the 2^∞ sequence. Finally, in the case where a substitution yields two fixed points, we compare the two fixed points to find a pattern similar to the Toeplitz property observed in the 2^∞ sequence.

3.2 Suggestions For Further Study

We have made several conjectures about patterns coming from comparing the fixed points for a given substitution that we did not have time to thoroughly justify. In the future, it would be interesting to further explore these conjectures and provide proofs.

We can also begin to apply substitutions to other areas of symbolic dynamics. In the area of one-dimensional dynamical systems, one of the simplest ways to exhibit complex behavior is to study various unimodal maps restricted to the unit interval. Each unimodal map has a specific sequence of 0s and 1s associated with it, called the *kneading sequence*, and this sequence completely characterizes the dynamical behavior of the system. The 2^∞ sequence discussed in

Chapter 2 is an example of a kneading sequence; it belongs to the unique quadratic map with periodic points of period 2^n for all n and no other periodic points.

This raises the question of which substitutions are relevant or interesting in other areas of dynamical systems, such as in the study of unimodal maps. That is, which substitutions have fixed points that are the kneading sequences for unimodal maps? Further, how do properties within substitutions, such as minimality, affect the behavior of the associated unimodal map?

Also, there is currently a lot of interest in the patterns that arise in numeration systems for various substitutions. In the future, we could investigate and generalize some of the properties and patterns that appear from numeration systems.

References

[1] L. Alvin, *Toeplitz kneading sequences and adding machines*, preprint (2012).

[2] K.M. Brucks and H. Bruin, *Topics from one-dimensional dynamics*, Cambridge University Press (2004).

[3] N. P. Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, Springer-Verlag: New York (2002).

[4] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, (1995).

[5] M. Morse and G.A. Hedlund, *Symbolic dynamics,* Amer. J. Math **60** (1938), 815–866.

[6] S. Williams, *Toeplitz minimal flows which are not uniquely ergodic*, Zeitschrift fuer Wahrscheinlichkeitstheorie verw. Geb. **67** (1984), 95–107.