Overlap-Free Morphisms

By

Christopher R. Tompkins

Bachelor of Science
Mathematics
Virginia Tech, 2007

Bachelor of Science
Engineering Science and Mechanics
Virginia Tech 2007

Submitted in Partial Fulfillment of the Requirements

for the Degree of Masters of Science in

Mathematics

College of Arts and Sciences

University of South Carolina

2009

Accepted by:

George F. McNulty, Major Professor

Stephen A. Fenner, Second Reader

James Buggy, Interim Dean of the Graduate School

Dedication

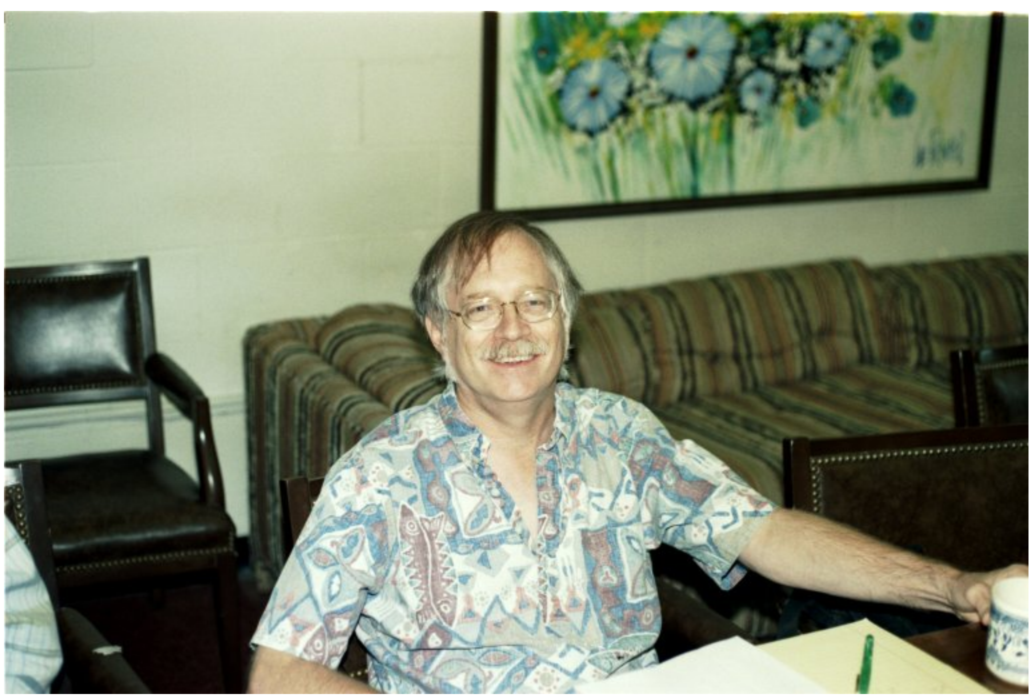
My thesis is dedicated to my dad, who only days before my defense suffered catastrophic spinal injury cycling. Be strong, you'll make it through.

Acknowledgments

I would like to begin by thanking my advisor Professor George McNulty for his guidance and patience with me, along with his encouragement of my continual pursuit of an area that I already had a good deal of interest in. I would also like to thank Dr. Stephen Fenner, my second reader.

Secondly, I would like to thank Dr. Griff Elder my undergraduate research advisor from Virginia Tech who helped me with my formalization of the ideas in the chapter on Latin Square Morphisms.

Finally, I would like to thank Dan Farkas who introduced me to the study of combinatorics on words.



Abstract

An alphabet Σ is a collection of letters. A word W is a possibly empty list of letters from an alphabet written horizontally, and we write Σ^* for the set of all words over the alphabet Σ. A morphism is a mapping $h : \Sigma^* \to \Delta^*$, where Σ and Δ are alphabets, such that for any words $X, Y \in \Sigma^*$, $h(XY) = h(X)h(Y)$. An overlap is the pattern $cXcXc$ where c is a letter and X is a word that is possibly empty. We say that a word is overlap-free provided there is no overlap in it. In this work we study morphisms h that map overlap-free words to overlap-free words. Our main result is the creation of a new condition sufficient to ensure that a morphism is overlap-free. As a result, we provide a new class of overlap-free morphisms.

Contents

List of Figures

Chapter 1

Introduction

1.1. History

The mathematical study of words dates back to the beginning of the 20^{th} century to the work of Axel Thue (Thue, 1906) and maybe even further back to the work of M. E. Prouhet (Prouhet, 1851). Thue's work appeared in a journal not usually read by mathematicians, so it attracted little notice for almost a generation.

Thue's work was primarily based on patterns in words and most notably that there is an infinite word on two letters that avoids cubes. We quickly note that XXX is a cube where X is some nonempty string of symbols, and a word W avoids cubes if there is no subword XXX in W. A word that avoids cubes is called cube-free. Note XX is a square and the definition of a square-free word follows in the same manner. Along with the idea of square-freeness and cube-freeness, the idea of overlap-freeness was studied by Thue. An overlap is the pattern $cXcXc$ where c is a letter and X is a word with possibly zero letters. The standard example of a word that is an overlap in its entirety is "alfalfa". An overlap-free word is a word in which no overlap occurs as a subword.

We know Thue's infinite cube-free word as the Thue-Morse infinite word

$$0110100110010110100101100110100 1\dots$$

(Thue, 1912). The sequential definition is given as

$$\begin{aligned} t_0 &= 0 \\ t_{2n} &= t_n \\ t_{2n+1} &= \bar{t_n}, \end{aligned}$$

where $\bar{0} = 1$ and $\bar{1} = 0$.

Despite the underlying morphic characteristic to the Prouhet words, he was not thinking of them with the idea of morphisms in mind. On the other hand, Thue was. Although we will more rigorously define a morphism later, a morphism h is a mapping that takes words as inputs which acts individually on letters and gives words as outputs. That is, if h is a morphism, $h(ab) = h(a)h(b)$ for two letters a, b. Every morphism is completely determined by the words it assigns as images of the individual letters of its domain alphabet. Thus, while the morphism itself is a function between two infinite sets, it will have a finite description, at least when the domain alphabet is finite.

For example, the morphism that generates the Thue-Morse infinite word is

$$\mu(t) = \begin{cases} 01, & \text{for } t = 0 \\ 10, & \text{for } t = 1. \end{cases}$$

We can compute the n^{th} Thue-Morse word in the following manner,

$$\begin{aligned} \mu(0) &= 01 \\ \mu^2(0) &= \mu(01) = \mu(0)\mu(1) = 0110 \\ \mu^3(0) &= \mu(0110) = 01101001 \\ &\vdots \\ \mu^n(0) &= \mu(\mu^{n-1}(0)) = 0110100110010110 1001\ldots. \end{aligned}$$

Because $\mu(0)$ begins with 0 we see that $\mu^2(0)$ must begin with $\mu(0)$ and, more generally, that $\mu^{n+1}(0)$ must begin with $\mu^n(0)$. In this sense, the sequence $0, \mu(0), \mu^2(0), \ldots$ converges to an infinitely long word. We express this as $\lim_{n\to\infty} \mu^n(0) = \mu^\omega(0)$, where $\mu^\omega(0)$ is the Thue-Morse infinite word. Further we note that there is an underlying topology to the situation, but we will not discuss that here. A morphism is called overlap-free if the image of every overlap-free word is itself overlap-free. Similarly, a morphism is said to be cube-free (respectively square-free) provided the image of every cube-free (square-free) word is itself cube-free (square-free). Thue (1912) noted that μ is overlap-free. Hence, the Thue-Morse infinite word is overlap free, and therefore alse cube-free.

From Thue's creation of the idea of the morphism, the study of morphisms was expanded at the 20$^{\text{th}}$ century progressed. Note that clearly the Thue-Morse morphism does not map square-free words to square-free words. However, (Thue, 1906; 1912) created square-free morphisms on the alphabet with 3 letters.

Using the definition, to determine whether a morphism is overlap-free (cube-free, square-free) might require the examination of the image of each overlap-free (cube-free, square-free) word. Since the set of overlap-free (cube-free, square-free) words is infinite, it would be desirable to be able the replace this infinite set with one of its finite subsets. Max Crochemore, Andrzej Ehrenfeucht, and Grzegorz Rozenberg in the early 80's produced results that characterized when finite sets of input words (called test-sets for square-freeness) that when tested with any morphism would display when the morphism was square-free (Crochemore, 1982), (Ehrenfeucht and Rozenberg, 1982).

Jean Berstel and Patrice Séébold in 1993 showed that the Thue-Morse morphism (along with it's complement and powers of it) is the only overlap-free morphism on the alphabet with two letters (Berstel and Séébold, 1993). They also found that $\{01101001\}$ is a test set for overlap-freeness in this context. In 2004 Gwenael Richomme and Francis Walzinski classified all of the overlap-free morphisms with a result considering test-sets for overlap-freeness (Richomme and Wlazinski, 2004). Further in 2007, we began exploring specific examples of morphisms that are overlap-free in hopes of producing a better characterization of overlap-free morphisms (Tompkins, 2007). But we have made significant progress towards such a characterization.

1.2. Thesis Outline

In Chapter 2 we recall the results of Jean Berstel and Patrice Séébold characterizing the Thue-Morse morphism as essentially the only overlap-free morphism (Berstel and Séébold, 1993). The results of the chapter are a simpler proof of the characterization of the overlap-free binary morphisms, as given in the Allouche and Shallit book (Allouche and Shallit, 2003).

In Chapter 3 we discuss the characterization that Richomme and Wlazinski proved about test-sets for overlap-free morphisms (Richomme and Wlazinski, 2004). Specifically we concentrate on the test-set result for uniform morphisms. We note that the Richomme and Wlazinski result in general is extremely interesting, but we only concern ourself with the uniform morphism result.

In Chapter 4 we discuss previous results that we have proved (Tompkins, 2007). We take the result about infinite overlap-free words and extend it to overlap-free morphisms fairly easily. This is the first portion of the thesis giving a specific description of overlap-free morphisms on the n-letter alphabet.

In Chapter 5 we consider the newest class of morphisms that we have proved to be overlap-free. We examine the Pooh morphism which is a generalization of the Leech square-free morphism (Leech, 1957). Not only do we generalize the Leech square-free morphism to a class overlap-free morphisms, but we also generalize the Leech square-free morphism to a considerable class of square-free morphisms.

1.3. Preliminaries and Definitions

We will use the standard definitions from the Lothaire book on combinatorics on words for our definitions with a few additions (Lothaire, 2002).

We begin by defining an alphabet Σ to be a finite set of symbols from which we will make words by concatenation (note, we will use capital Greek letters for alphabets). Further, we define a word W to be a list of symbols from any alphabet Σ written horizontally (we will use capital letters to denote words and lower case letters to denote letters). We will denote the word with no letters, that is the empty word, by ε.

1.3.1. Words. The length (or number of letters) for a word W will be written $|W|$. Note that we will use the same symbol to represent the size of a set or absolute value. The difference will be clear based on context. Notice that $|\varepsilon| = 0$. Further we will represent $|W|_a$ to represent the number of times the letter a occurs in W. Also we will use $|W|_{aba}$ to represent the number of times the word aba occurs in W. For example if $C = abaababa$, then we have $|C|_{aba} = 3$ along with $|C| = 8$.

A word U is a factor of a word V if there exist two (possibly empty) words S and T such that $V = TUS$. We will also say that U is a subword of V (or V contains U). If $T = \varepsilon$, then we call U the prefix of V. Similarly, if $S = \varepsilon$, then we call U the suffix of V.

Let W be a word, and let i, j be integers such that $0 \leq i - 1 \leq j \leq |W|$. We denote $W[i..j]$ to mean the subword of W that begins at the i^{th} letter of W and ends with the j^{th} letter of W. That is, that there exists two words S, T such that $W = SW[i..j]T$, where $|S| = i - 1$ and $|T| = |W| - j$. When $i = j$, we will denote $W[i..j]$ by $W[i]$ which represents the i^{th} letter of the word W.

Let $W = a_1 \ldots a_n$, with $a_i \in \Sigma$ and $W \neq \varepsilon$. We will define $\text{alph}(W)$ to be the set $\{a_i : 1 \leq i \leq |W|\}$.

For any set of words X, we define $\text{Fact}(X) = \{U : U \text{ is a subword for some } W \in X\}$.

For some alphabet Σ, Σ^* is the Kleene closure of our alphabet. That is, Σ^* is all of the possible words over the alphabet Σ. Notice that Σ^* is the free monoid over the set Σ.

1.3.2. Morphisms. A morphism h is a mapping from Σ^* into Δ^*, where Σ and Δ are alphabets, such that $h(WV) = h(W)h(V)$ for all words $W, V \in \Sigma^*$, and $h(\varepsilon) = \varepsilon$. Note that W and V could potentially be single letters. We note that if $X \subseteq \Sigma$ (X represents a set of words) for some alphabet Σ, $h(X)$ represents the set of words $\{h(W) : W \in X\}$. Further, we call h non-erasing if for all $a \in \Sigma$, where Σ is an alphabet, $h(a) \neq \varepsilon$.

Recall from earlier that the Thue-Morse morphism, μ defined as

$$\mu(t) = \begin{cases} 01, & \text{for } t = 0 \\ 10, & \text{for } t = 1, \end{cases}$$

is a morphism defined on the alphabet with two letters. For convenience, we will call the alphabet with n letters Σ_n. Infinite words are possible with such a morphism. We have displayed the n^{th} Thue-Morse word as being $\mu^n(0)$. We will use ω to represent the first infinite ordinal. So the Thue-Morse infinite word becomes

$$\mathbf{T} = \lim_{n\to\infty} \mu^n(0) = \mu^\omega(0),$$

as previously seen. Note that we will use bold capitol letters to represent infinite words, with $\mathbf{T}$ here representing the Thue-Morse infinite word.

When discussing $\Sigma_2 = \{0, 1\}$, the two letter alphabet we will use $\bar{0}$ to denote the complement of 0 (or 1 if we need that complement). That is $\bar{0} = 1$ and $\bar{1} = 0$. This will become necessary in Chapter 2.

For some morphism $h : \Sigma^* \to \Delta^*$, we will call h uniform if $|h(a)| = n$ for some integer n for all $a \in \Sigma$ (more exactly, in this case we will call h n-uniform). We will call a morphism $h : \Sigma^* \to \Delta^*$ square-free when $h(W)$ is square-free if and only if $X \in \Sigma^*$ is square-free. Similarly, we will call $h : \Sigma^* \to \Delta^*$ an overlap-free morphism when $h(W)$ is overlap-free if an only if $W \in \Sigma^*$ is overlap-free.

Chapter 2

The Berstel - Séébold Theorem

Jean Berstel and Patrice Séébold proved that a morphism h over the two-letter alphabet $\Sigma_2 = \{0, 1\}$ is overlap-free if and only if the word $h(01101001)$ is overlap-free (Berstel and Séébold, 1993). Further Berstel and Séébold proved that the only binary overlap-free morphism is the Thue-Morse morphism or it's complement or any power of the two.

2.1. Preliminaries

We know from earlier that the Thue-Morse morphism is

$$\mu(t) = \begin{cases} 01, & \text{for } t = 0 \\ 10, & \text{for } t = 1. \end{cases}$$

We will also use the notation introduced earlier that $\bar{0} = 1$ and $\bar{1} = 0$. Further, for $n \geq 0$ we will define the following

$$U_n = \mu^n(0), \qquad \text{and} \qquad V_n = \mu^n(1).$$

For example, we see that

$$\begin{aligned} U_0 &= 0 & V_0 &= 1 \\ U_1 &= 01 & V_1 &= 10 \\ U_2 &= 0110 & V_2 &= 1001 \\ U_3 &= 01101001 & V_3 &= 10010110. \end{aligned}$$

If we define the morphism E to be

$$E(0) = 1 \qquad \text{and} \qquad E(1) = 0,$$

then it becomes clear that $U_{n+1} = U_n V_n$, $V_{n+1} = V_n U_n$, $U_n = E(V_n)$, and $V_n = E(U_n)$. We can extend the morphism μ to infinite words as displayed earlier, and the two fixed points become

$$\mathbf{T} = \mu^{\omega}(0) = 01101001100101101001011001101001\ldots = \mu(\mathbf{T})$$
$$E(\mathbf{T}) = \mu^{\omega}(1) = 10010110011010010110100110010110\ldots = \mu(E(\mathbf{T})).$$

Definition 2.1. A morphic infinite word $\mathbf{X}$ with generating morphism $h : \Sigma^* \to \Delta^*$ is a word such that there is some integer $m \geq 0$ and letter $t \in \Sigma$ so that

$$\mathbf{X} = \lim_{n\to\infty} (h^{mn}(t)).$$

Clearly we see that $\mathbf{X} = h^m(\mathbf{X})$ and m must be bounded by $|\Sigma|$.

Later, Theorem 4.3 shows us that μ is an overlap-free morphism. So this further tells us that $\mathbf{T}$ is an overlap-free infinite word, which an original result of Axel Thue (Thue, 1912).

2.2. Lemmas

For the proof of the theorem that any morphism on the two letter alphabet h is overlap-free if and only if $h(01101001)$ is an overlap-free word, we will use the following lemmas. Furthermore we will need this first lemma for establishing the relationship between the Thue-Morse morphsim μ and the arbitrary morphism on two letters h. We first notice two easy results:

Remark 2.2. Let $U \in \Sigma_2^*$. There exists some $V \in \Sigma_2^*$ such that $U = \mu(V)$ if and only if $U \in \{01, 10\}^*$.

REMARK 2.3. Let $X, Y \in \Sigma_2^*$. If $XY \in \{01, 10\}^*$ and $|X|$ is even then $Y \in \{01, 10\}^*$.

We note that the results in the Allouche and Shallit book on automatic sequences (Allouche and Shallit, 2003) gives a slightly better presentation of the Berstel and Séébold paper (Berstel and Séébold, 1993). We will proceed in the same manner as Allouche and Shallit. We begin with some of the lemmas needed for our result.

LEMMA 2.4. *Suppose $Y \in \Sigma_2^*$, and $a \in \Sigma_2$. If the word $a\bar{a}\bar{a}Y$ is overlap-free, then at least one of the following holds:*

(a) *Y begins with aa;*

(b) $|Y| \leq 3$*;*

(c) *Y begins with $a\bar{a}aa$.*

PROOF. If Y begins with aa or if $|Y| \leq 3$, then we are done. So we assume that $|Y| \geq 4$ and that Y does not begin with aa.

Y cannot begin with $\bar{a}$ because if it did, then we would have that $a\bar{a}\bar{a}Y = a\bar{a}\bar{a}\bar{a}Q$ for some Q. We assumed that $a\bar{a}\bar{a}Y$ has no overlap, so Y must begin with a. Also we assumed that Y does not begin with aa, so let $Y = a\bar{a}Z$ for some $Z \in \Sigma_2^*$ such that $|Z| \geq 2$.

We now want to consider what would happen if Z were to begin with $\bar{a}$. So we assume that Z begins with $\bar{a}$, then there is some W such that $|W| \geq 1$ such that

$$a\bar{a}\bar{a}Y = a\bar{a}\bar{a}a\bar{a}\bar{a}W.$$

If W begins with either a or $\bar{a}$ then we have a problem with $a\bar{a}\bar{a}$ begin overlap- free. Thus we must have that Z begins with an a as opposed to $\bar{a}$.

So we assume that $Z = aW$ for some W such that $|W| \geq 1$. Then, we see that

$$a\bar{a}\bar{a}Y = a\bar{a}\bar{a}a\bar{a}aW.$$

So the only possibility is that W begins with an a. Thus Y begins with $a\bar{a}aa$, and we are done. □

LEMMA 2.5. *If $Y, Y' \in \Sigma_2^*$, and if there exist $c, d \in \Sigma_2$ such that $U = c\mu(Y) = \mu(Y')d$, then $U = c(\bar{c}c)^{|Y|}$.*

PROOF. It is clear that $|Y| = |Y'|$. So we set

$$\begin{aligned} Y &= a_1 a_2 \ldots a_n \\ Y' &= b_1 b_2 \ldots b_n. \end{aligned}$$

Now we examine the equality $c\mu(Y) = \mu(Y')d$, which follows as

$$ca_1\overline{a_1}a_2\overline{a_2}\ldots a_n\overline{a_n} = b_1\overline{b_1}b_2\overline{b_2}\ldots b_n\overline{b_n}d.$$

Thus, it is clear that

$$\begin{aligned} c &= b_1 \\ a_1 &= \overline{b_1} = \bar{c} \\ b_2 &= \overline{a_1} = c \\ a_2 &= \overline{b_2} = \bar{c} \\ &\vdots \\ b_t &= \overline{a_{t-1}} = c \\ a_t &= \overline{b_1} = \bar{c} \\ \overline{a_t} &= d. \end{aligned}$$

Thus we must have that $c = d$ and further that $U = c(\bar{c}c)^{|Y|}$. □

LEMMA 2.6. *Let $Y, Z \in \Sigma_2^*$, and let $\mu(Y) = ZZ$. Then there must be some $X \in \Sigma_2^*$ such that $\mu(X) = Z$.*

PROOF. Begin by supposing that $\mu(Y) = ZZ$. If $|Z| \equiv 0 \pmod 2$ then we are done because $|\mu(Y)| \equiv 0 \pmod 4$, and so $|Y|$ is even. Then we can set $Z = \mu(W)$ where W is the first half of Y.

Now we want to show that it is impossible for $|Z|$ to be odd. So assume for a contradiction that $|Z|$ is odd. Let $U, V \in \Sigma_2^*$ and $a, b \in \Sigma_2$ be such that

$$Z = aU = Vb.$$

Because $|Z|$ is odd, it is clear that $|U|$ and $|V|$ are even. Now we have that

$$\mu(Y) = ZZ = VbaU.$$

Thus, we must have some $R, S \in \Sigma_2^*$ so that $U = \mu(R)$ and $V = \mu(S)$. Notice that

$$ZZ = \mu(S)ba\mu(R)$$

and $b = \bar{a}$. But we have that $Z = a\mu(R) = \mu(S)b$, so by Lemma 2.5 we must have that $Z = a(\bar{a}a)^{|R|}$. If this is true, then the last letter of Z must be a, but the last letter is $b = \bar{a}$. So we have our contradiction, and we are done □

LEMMA 2.7. *Let $W \in \Sigma_2^*$ and let $\mathbf{X} \in \Sigma_2^\omega$. Then*

(a) *W contains an overlap if and only if $\mu(W)$ contains an overlap.*

(b) *$\mathbf{X}$ contains an overlap if and only if $\mu(\mathbf{X})$ contains an overlap.*

This result is established by μ being Latin square morphism which is proven to be overlap-free in Theorem 4.3. We can use this lemma to prove the "factorization" result below.

LEMMA 2.8.

(a) *If $X \in \Sigma_2^*$ is overlap-free, then there exists U, V, Y with $U, V \in \{\varepsilon, 0, 1, 00, 11\}$ and $Y \in \Sigma_2^*$ an overlap-free word, such that $X = U\mu(Y)V$. Furthermore this* *factorization is unique if* $|X| \geq 7$ *and* U *(or* V*) is completely determined by the prefix (or suffix) of length* 7 *of* X *(Note: the bound* 7 *is the best possible).*

(b) *If* $\mathbf{X} \in \Sigma_2^\omega$ *is an infinite overlap-free word, then there exists a word* U *in* $\{\varepsilon, 0, 1, 00, 11\}$ *and an infinite overlap-free word* $\mathbf{Y} \in \Sigma_2^\omega$ *such that* $\mathbf{X} = U\mu(Y)$*. The prefix is completely determined by the prefix of* $\mathbf{X}$ *of length* 4*, except if* $\mathbf{X}$ *begins with* 0100 *or* 1011*, in which case the word* U *is completely determined by the prefix of* $\mathbf{X}$ *of length* 5*.*

PROOF.

(a) We begin by noticing that if our decomposition exists then we must have that Y is overlap-free by Lemma 2.7. We will begin by showing that the factorization exists, and we will proceed by induction on $|X|$.

The base case $|X| = k < 2$ is clear. So our inductive hypothesis becomes assume that we can apply our factorization for all X such that $|X| < k$, and we want to show it possible for $|X| = k$.

Thus we begin our induction step. Let X be overlap-free with $|X| \geq 3$. Further we want to assume that X begins with the letter $a \in \Sigma_2$, so let $X = aZ$, with $Z \in \Sigma_2^*$. Because of the induction hypothesis, we have that $Z = U\mu(Y)V$ for $U, V \in \{\varepsilon, 0, 1, 00, 11\}$. Now we approach all of the possibilities for U.

(i) If $U = \varepsilon$ or if $U = a$, then we must have that $X = a\mu(Y)V$ or $X = aa\mu(Y)V$ and we are done.

(ii) If $U = \bar{a}$ then we see that $X = a\bar{a}\mu(Y)V$. Thus, $X = \mu(aY)V$.

(iii) If $U = aa$, we have a problem because then $X = aaa\mu(Y)V$ which contains an overlap. So $U \neq aa$.

(iv) If $U = \bar{a}\bar{a}$ then we have that $X = a\bar{a}\bar{a}\mu(Y)V$, and we have three possibilities:

1. If $|Y| = 0$ then we have that $X = a\bar{a}\bar{a}V$. Thus, we must have that $V \in \{\varepsilon, a, aa\}$. So our possible configurations for X follow as

$$X = \mu(a)\bar{a},$$
$$X = \mu(a\bar{a}),$$
$$X = \mu(a\bar{a})a.$$

2. If $|Y| = 1$, then we must have $Y = a$. So we have

$$X = a\bar{a}\bar{a}a\bar{a}V.$$

The possibilities for V yield the following results

(a) $V = \varepsilon$ implies that $X = \mu(a\bar{a})\bar{a}$.

(b) $V = a$ implies that $X = \mu(a\bar{a}\bar{a})$.

(c) $V = \bar{a}$ implies that $X = \mu(a\bar{a})\bar{a}\bar{a}$.

(d) $V = aa$ implies that $X = \mu(a\bar{a}\bar{a})a$.

(e) $V = \bar{a}\bar{a}$ implies that the overlap $\bar{a}\bar{a}\bar{a}$ occurs in X, which is impossible.

3. If $|Y| \geq 2$, then we have that $|\mu(Y)| \geq 4$. Now Y cannot begin with aa because then we would have an overlap. So, by Lemma 2.4 $\mu(Y)$ must begin with $a\bar{a}aa$ which is impossible because $a\bar{a}aa$ cannot be the image of μ.

We will now show that the factorization of X is unique provided that $|X| \geq 7$. Let X be overlap- free with $|X| \geq 7$, and suppose that

$$X = U\mu(Y)V = U'\mu(Y')V', \tag{1}$$

with $Y, Y' \in \Sigma_2^*$ overlap-free and $U, U', V, V' \in \{\varepsilon, 0, 1, 00, 11\}$. Notice that $|X| \geq 7$ implies that $|Y|, |Y'| \geq 2$ because $|\mu(Y)|, |\mu(Y')| \geq 3$. Again we approach all of the possibilities for U.

(i) If $|U| = |U'|$, then we have $U = U'$. Thus, $\mu(Y)V = \mu(Y')V'$, and further we have $|V| \equiv |V'| \pmod 2$.

1. If $|V| = |V'|$ we are done because $V = V'$ and $\mu(Y) = \mu(Y')$ gives that $Y = Y'$.

2. If $|V| \neq |V'|$, then we have that $|V| = 2$ and $|V'| = 0$ (or vice versa). So without loss of generality assume that $\mu(Y)V = \mu(Y')$. In this case either $V = 00$ or $V = 11$, and V must then be the suffix of $\mu(Y')$. Thus, we must have that $V = \mu(0)$ or $V = \mu(1)$ which is impossible.

(ii) Set $|U| \neq |U'|$, and without loss of generality we can assume that $|U'| < |U|$. So, U' is a prefix of U. Now we can pick W such that $U = U'W$ with $W \neq \varepsilon$. By equating the letters in equation (1) we see that $W\mu(Y)V = \mu(Y')V'$. With W being a prefix of $\mu(Y')$ it is clear that $W \neq 00$ or $W \neq 11$. Thus, $W = b \in \Sigma_2$, and we have that $b\mu(Y)V = \mu(Y')V'$. This implies that $|V| \not\equiv |V'| \pmod 2$, and it is obvious that $|Y| \neq |Y'|$.

1. If $|V| < |V'|$ then $V' = ZV$ for some $Z \neq \varepsilon$. So $a\mu(Y) = \mu(Y')Z$. Notice that $|Z|$ must be odd so $Z = b \in \Sigma_2$. Now we have that $a\mu(Y) = \mu(Y')b$ which is not possible by Lemma 2.5 because $a\mu(Y) = a(a\bar{a})^{|Y|}$ and $|Y| \geq 2$ yielding an overlap.
2. If $|V'| < |V|$, then we can set $V = ZV'$ for some $Z \neq \varepsilon$. Thus, we have $a\mu(Y)Z = \mu(Y')$, and it is clear that $|Z|$ is odd so $Z = b \in \Sigma_2$. Notice that $a\mu(Y)b = \mu(Y')$ and further that $\mu(Y')$ must begin with a. So, we set $Y = aT$ with $T \neq \varepsilon$ because we know $|Y| \geq 2$. we see that $\mu(Y)b = \bar{a}\mu(T)$. This is not possible because of Lemma 2.5 along with $|Y| \geq 2$ as above.

In this factorization of X, U (or V) depends on the prefix (or suffix) of length 7 only. Table 2.1 shows all possibilities for prefixes (or suffixes). By inspection we see that U (or V) is uniquely determined assuming that the factorization exists. We also note that some words (for instance 1100100) cannot be extended to overlap-free words.

Also, we note that the bound 7 is best possible because if we consider the word $X = 001011$, then we could have $X = 00\mu(1)11$ or we could have $X = 0\mu(00)1$.

(b) We will now consider $\mathbf{X}$, the infinite word, and it's factorization. So let X_n be any prefix of $\mathbf{X}$ (with $|X_n| = n$). Then part (a) of our lemma gives that there is

Table 2.1. Prefixes of length 7

X	U	X	U	X	U	X	U
$0010011\cdots$	00	$0100110\cdots$	0	$1001011\cdots$	ε	$1011010\cdots$	1
$0010110\cdots$	0	$0101100\cdots$	ε	$1001100\cdots$	ε	$1100100\cdots$	1
$0011001\cdots$	0	$0101101\cdots$	ε	$1001101\cdots$	ε	$1100101\cdots$	1
$0011010\cdots$	0	$0110010\cdots$	ε	$1010010\cdots$	ε	$1100110\cdots$	1
$0011011\cdots$	0	$0110011\cdots$	ε	$1010011\cdots$	ε	$1101001\cdots$	1
$0100101\cdots$	0	$0110100\cdots$	ε	$1011001\cdots$	1	$1101100\cdots$	11
X	V	X	V	X	V	X	V
$\cdots 0010011$	1	$\cdots 0100110$	ε	$\cdots 1001011$	1	$\cdots 1011010$	ε
$\cdots 0010110$	ε	$\cdots 0101100$	0	$\cdots 1001100$	0	$\cdots 1100100$	00
$\cdots 0011001$	ε	$\cdots 0101101$	1	$\cdots 1001101$	1	$\cdots 1100101$	ε
$\cdots 0011010$	ε	$\cdots 0110010$	0	$\cdots 1010010$	0	$\cdots 1100110$	ε
$\cdots 0011011$	11	$\cdots 0110011$	1	$\cdots 1010011$	1	$\cdots 1101001$	ε
$\cdots 0100101$	ε	$\cdots 0110100$	0	$\cdots 1011001$	ε	$\cdots 1101100$	0

some $U_n, V_n \in \{\varepsilon, 0, 1, 00, 11\}$ and $Y_n \in \Sigma_2^*$ that is overlap-free such that

$$X_n = U_n\mu(Y_n)V_n.$$

Plus, we see that U_n does not depend on n for $n \geq 7$. So we set $U = U_7$. Then we have that

$$\mathbf{X} = \lim_{n\to\infty} X_n = \lim_{n\to\infty} U\mu(Y_n)V_n.$$

Now because we have $|\mu(Y_n)| \to \infty$ as $n \to \infty$, then we see that

$$\mathbf{X} = \lim_{n\to\infty} U\mu(Y_n).$$

Because $\lim \mu(Y_n)$ exists, we can set $\mathbf{Y} = \lim \mu(Y_n)$. Further we see that the sequence is overlap-free because it is the limit of overlap-free words. So we see that

$$\mathbf{X} = \mu(\mathbf{Y}).$$

Finally we set $X_n = U_n\mu(Y_n)V_n$. Because $V_n \in \{\varepsilon, 0, 1, 00, 11\}$, we must have that $|V_n| \leq 2$. So if we consider some $k \in \mathbb{N}$ large, then $\mu(Y_n)V_n$ is a prefix of $\mu(Y_k)$. Because $|\mu(Y_n)|$ is even, we see that V_n is the prefix for some image of μ. So because $\mu(T_n) = V_n$ for some word $T_n \in \Sigma_2^*$, it follows that $V_n \neq 00$ and $V_n \neq 11$.

Thus we see that $V_n = \varepsilon, 0, 1$. Lastly, looking at Table 2.1 we see that the prefix of length 4 of a 7-letter word determines the the word U in all cases except when the word begins with 0010 or 1101, when we need to consider a prefix of length 5. □

LEMMA 2.9. *Let $W \in \Sigma_2^*$ be an overlap-free word with $|W| \geq 52$. Then W contains $\mu^3(0) = 01101001$ and $\mu^3(1) = 10010110$ as subwords.*

PROOF. Let $W \in \Sigma_2^*$ be overlap-free. Lemma 2.8 gives that

$$W = A\mu(W')B, \tag{2}$$

with $|W'| \leq 24$, and it is clear that W' is overlap free. Again by Lemma 2.8 and W' overlap-free we see that

$$W' = C\mu(W'')D, \tag{3}$$

where W'' is overlap-free and $|W''| \leq 10$. Lastly, we use Lemma 2.8 to write

$$W'' = G\mu(Y)H, \tag{4}$$

with Y overlap-free and $|Y| \geq 3$. By putting equations (2), (3), and (4) together we obtain,

$$W = A\mu(C)\mu^2(G)\mu^3(Y)\mu^2(H)\mu(D)B,$$

where $A, B, C, D, G, H \in \{\varepsilon, 0, 1, 00, 11\}$ and Y is overlap-free with $|Y| \geq 3$, as stipulated. Because Y is overlap-free, it must contain both a 0 and a 1. Thus, $\mu^3(Y)$ contains both $\mu^3(0)$ and $\mu^3(1)$ which means that by default W contains these subwords as well. □

2.3. Main Theorem of Berstel and Séébold and Corollaries

THEOREM 2.10. *Let μ be the morphism defined by $\mu(0) = 01$ and $\mu(1) = 10$. Let E be the morphism defined by $E(0) = 1$ and $E(1) = 0$. Let $h : \Sigma_2^* \to \Sigma_2^*$ be a nonerasing morphism. If the image by h of any overlap-free binary word of length 3 is* *overlap-free, then there exists an integer* $k \geq 0$ *such that either* $h = \mu^k$ *or* $h = E \circ \mu^k$.

PROOF. Let $h(0) = U$ and $h(1) = V$, where $|U|, |V| \geq 1$. Our proof will proceed by induction on $|U| + |V|$.

We will first notice some unavoidable facts about U and V. For instance, if U and V were to begin with the same letter, then we would have an overlap in the word $h(001)$. Further, if U and V were to end with the same letter, then $h(100)$ would contain an overlap. Clearly if h maps some 3-letter word to an overlap we have a problem, so we set

$$U = a \ldots = \ldots b$$

$$V = \bar{a} \ldots = \ldots \bar{b}.$$

We want to show now that neither U nor V can begin with 00 or 11. So assume that $U = aa \ldots$. Then, we see that

$$UU = \ldots baa \ldots$$

$$VU = \ldots \bar{b}aa \ldots.$$

This becomes a problem because either UU or VU contains an overlap and they are subwords of $h(001) = UUV$ and $h(100) = VUU$ respectively. Thus we see that neither U nor V can begin with a square.

Now we begin our induction. For the base case we consider $|U| + |V| = 2$. Clearly in this case $|U| = |V| = 1$ because otherwise our morphism is erasing. It is easy to see that now $h = \mu^0$ or $h = E \circ \mu^0$.

For the induction we assume that the theorem is true for all U and V with $|U| + |V| < j$, and we show that it is further true for $|U| + |V| = j$. So suppose that $|U| \geq 2$. For this to be true we must also have $|V| \geq 2$ because if $|V| = 1$, then we can set $V = b$ yielding $U = \bar{b}b \ldots b\bar{b}$. This formulation yields a contradiction because then $h(010)$ contains the overlap $b\bar{b}b\bar{b}b$. Similarly we can show that if $|V| \geq 2$, then we must also have $|U| \geq 2$. So we assume that $|U|, |V| \geq 2$.

Now, set

$$U = a\bar{a}\ldots = \ldots\bar{b}b$$

$$V = \bar{a}a\ldots = \ldots b\bar{b}.$$

We want to show that $U = \mu(W)$ for some $W \in \Sigma_2^*$, and we will proceed with cases here.

If $|U| = 2$, then $U = a\bar{a} = \bar{b}b = \mu(a)$.

If $|U| = 3$, then $\bar{a} = \bar{b}$ so $a = b$. Now we have that $U = a\bar{a}a$, and $V = \ldots a\bar{a}$. This yields a contradiction because $h(100)$ contains the overlap $a\bar{a}a\bar{a}a$ as a subword.

If $|U| = 4$, then $U = a\bar{a}\bar{b}b = \mu(a\bar{b})$.

If $|U| \geq 5$, then we have $U = a\bar{a}Z\bar{b}b$ for some $Z \in \Sigma_2^*$ with $|Z| \geq 1$. So we have

$$VUV = \ldots b\bar{b}\ a\bar{a}Z\bar{b}b\ \bar{a}a\ldots.$$

Now we must consider two cases $b = a$ and $b = \bar{a}$.

(i) If $b = a$, then we have $VUV = \ldots a\bar{a}\ a\bar{a}Z\bar{a}a\ \bar{a}a\ldots$. Noting that $|Z| \geq 1$ we see that Z cannot begin with a otherwise we would have an overlap. So we set $Z = \bar{a}T$ for some $T \in \Sigma_2^*$. Then, we have

$$VUV = \ldots a\bar{a}\ a\bar{a}\bar{a}T\bar{a}a\ \bar{a}a\ldots.$$

Clearly $T \neq \varepsilon$ because otherwise we would have $\bar{a}\bar{a}\bar{a}$ as a subword of U which is impossible. Further T cannot end in a so pick $X \in \Sigma_2^*$ so that $T = X\bar{a}$. Now, we have

$$VUV = \ldots a\bar{a}\ a\bar{a}\bar{a}X\bar{a}\bar{a}a\ \bar{a}a\ldots.$$

Notice $h(101) = VUV$ which must be overlap-free. Applying Lemma 2.8 gives us that $a\bar{a}\bar{a}X\bar{a}\bar{a}a = \mu(Y_1)$ for some $Y_1 \in \Sigma_2^*$, or more specifically $U = \mu(Y_1)$.

(ii) Assume that $b = \bar{a}$. Then $VUV = \ldots \bar{a}aa\bar{a}Za\bar{a}\bar{a}a \ldots$ for some $Z \in \Sigma_2^*$. Again we apply Lemma 2.8, and we see that'

$$\bar{a}aa\bar{a}Za\bar{a}\bar{a}a = \mu(Y_2),$$

for some $Y_2 \in \Sigma_2^*$. Notice that Y_2 begins and ends with $\bar{a}$. Set $Y_2 = \bar{a}Y_3\bar{a}$ for some $Y_3 \in \Sigma_2^*$. Then

$$U = a\bar{a}Za\bar{a} = \mu(Y_3).$$

Note that the same logic shows that V is the image of μ for some word in Σ_2^*.

Notice now that there must be two words $U', V' \in \Sigma_2^*$ so that $U = \mu(U')$ and $V = \mu(V')$. Consider the morphism h' such that $h'(0) = U'$ and $h'(1) = V'$, then clearly $h = \mu \circ h'$, and any image of any overlap-free word of length 3 under h' must also be overlap-free by Lemma 2.7. Finally, we see that $|U'| < |U|$ and $|V'| < |V|$, so the inductive hypothesis tells us that

$$h' = \mu^k \quad \text{or} \quad h' = E \circ \mu^k.$$

So the result is clear from here, and we are done. □

COROLLARY 2.11. *Using the same notation as in the theorem, let h be a morphism on the alphabet such that $\mu(01) \neq \varepsilon$. Then the following are equivalent.*

(a) *The morphism h is nonerasing and maps any overlap-free word of length 3 to an overlap- free word.*

(b) *There exists $k \geq 0$ such that $h = \mu^k$ or $h = E \circ \mu^k$.*

(c) *The morphism h maps any infinite overlap-free word to an infinite overlap-free word.*

(d) *There exists an infinite overlap-free word whose image under h is overlap-free.*

(e) *The morphism h maps the word 01101001 to an overlap-free word.*

PROOF. (a) $\Rightarrow$ (b) is proved in Theorem 2.10 above.

(b) $\Rightarrow$ (c) is easily seen from Lemma 2.7.

(c) $\Rightarrow$ (d) is clear from knowing that there must be an infinite overlap-free word, and both Theorem 4.3 and Lemma 2.7 give this result.

(d) $\Rightarrow$ (e) The hypothesis gives the existence of an infinite overlap-free word $\mathbf{X}$ whose image under h is overlap-free. Lemma 2.9 shows that clearly 01101001 is a subword of $\mathbf{X}$, and thus $h(01101001)$ is also overlap-free.

(e) $\Rightarrow$ (a) Suppose that $h(0) = \varepsilon$, then because $h(01) \neq \varepsilon$, we know that $h(1) \neq \varepsilon$. So it is clear that $h(01101001) = h(1)h(1)h(1)h(1)$ must contain an overlap, which is a contradiction. Similarly, $h(1) \neq \varepsilon$, so we can assume that h is nonerasing. But every overlap-free word of length 3 in Σ_2^* is a subword of 01101001, so we are done. $\square$

COROLLARY 2.12. *An infinite overlap-free binary word is a fixed point of a non-identity morphism if and only if it is equal to $\mathbf{T}$, the Thue-Morse word, or its complement $\overline{\mathbf{T}}$.*

PROOF. Let h be a morphism on Σ_2 such that h is not the identity, and let $\mathbf{X}$ be a fixed point of h that is overlap-free. From Corollary 2.11 it is clear that because h is mapping an infinite overlap-free word to another infinite overlap-free word that $h = \mu^k$ or $h = E \circ \mu^k$ for some $k \geq 0$. Because h has a fixed point we see that $h \neq E \circ \mu^k$, and further h is not the identity morphism so $k \geq 1$. Because $\mathbf{T}$, the Thue-Morse sequence, and its complement $\overline{\mathbf{T}}$ are the only fixed points of μ, then we are done. $\square$

Chapter 3

Test-Sets for the Overlap-Freeness of Morphisms

In 2004 Richomme and Wlazinski proved a result that classified all overlap-free morphisms using the notion of test-sets (Richomme and Wlazinski, 2004). This concept of a test-set was addressed in the last chapter. For example the test-set for an overlap-free morphism over the alphabet Σ_2 is the set of all of the binary words of length 3. That is, to tell if a morphism h over the alphabet Σ_2 is overlap-free, we take all $W \in \Sigma_2^*$ such that $|W| = 2$ and we check if $h(W)$ is overlap-free. If $h(W)$ is overlap-free, then we know that h is an overlap-free morphism. So we define what a test-set is in general.

Definition 3.1. Let Σ and Δ be alphabets. A set T ($\subseteq \Sigma^*$) is a test-set for overlap-freenes of morphisms from Σ^* to Δ^* if for each morphism $f : \Sigma^* \to \Delta^*$, f is overlap-free if and only if $f(W)$ is overlap-free for all words $W \in T$.

We first observe the theorem that Richomme and Wlazinski proved about overlap-free morphisms in general.

The Richomme and Wlazinski Test-Set Theorem. *Given alphabets Σ and Δ with $3 \leq |\Sigma| \leq |\Delta|$, there is no finite test-set for overlap-freeness of morphisms from Σ^* to Δ^*.*

This result seems reasonable because of a similar result was explored by Ehrenfeucht and Rozenberg concerning square-free morphisms (Ehrenfeucht and Rozenberg, 1982), which took Max Crochemore's paper (Crochemore, 1982) and added the notion of test sets. The Ehrenfeucht and Rozenberg result states that

The Ehrenfeucht and Rozenberg Test-Set Theorem. *Given alphabets* Σ *and* Δ *with* $|\Sigma| > 3$ *or* $|\Delta| > 2$*, then there is no finite test set for square-freeness of morphisms from* Σ^* *to* Δ^**.*

On the other hand a much more striking result is seen in (Richomme and Wlazinski, 2004) concerning uniform morphisms. Let Σ be an alphabet, we define the following sets

$$T_1(\Sigma) = \{xW_0x \big| x \in \Sigma, W_0 \in \Sigma^*, \forall a \in \Sigma, |xW_0|_a \leq 1\},$$

$$T_2(\Sigma) = \left\{ xW_1yW_2z \left| \begin{array}{l} x, y, z \in \Sigma, W_1, W_2 \in \Sigma^* \\ \forall a \in \Sigma, |W_1yW_2|_a \leq 1, \\ |W_1| = |W_2| \geq 1, |yW_2|_x = |W_1y|_z = 0 \end{array} \right. \right\}, \qquad (5)$$

$$T(\Sigma) = T_1(\Sigma) \cup T_2(\Sigma).$$

(Note that "$\forall a \in \Sigma, |xW_0|_a \leq 1$" means that the letters occurring in W_0 are all distinct and different from x). The assertion now is that $T(\Sigma)$ is a finite test set for any morphism $h : \Sigma^* \to \Delta^*$ such that $|h(a)|$ is the same for all $a \in \Sigma$.

3.1. Lemmas

We begin with the lemmas necessary to establish the Richomme and Wlazinski theorem.

Lemma 3.2. *Let* f *be a morphsism such that each image word of* f *begins and ends with a different letter. Let* W_1*,* W_2*,* W_3*,* V_0*,* V_2*,* V_3 *be words and* a *be a letter such that* $f(W_1) = V_0aV_1$*,* $f(W_2) = V_2aV_1$*, and* $f(W_3) = V_2aV_3$*. The word* $W_1W_2W_3$ *is not overlap-free.*

Proof. Let T be the longest common suffix of W_1 and W_2. Let P, T_0 be words such that $W_1 = PT, W_2 = T_0T$. Since none of the image words f end or begin with the same letter we have that $|f(T)| \geq |aV_1|$. Let P' be the word such that $f(T) = P'aV_1$.

We have that $V_0 f(P)P'$, and $V_2 = f(T_0)P'$. Thus, $f(T_0T) = f(W_2) = f(T_0)P'aV_1$ and $f(W_3) = f(T_0)P'aV_3$. Since none of the image words of f begin or end with the same letter there exists a non-empty word T_1 such that T_0T_1 is a common prefix of W_2 and W_3. Let T_2 be such that $T = T_1T_2$. The word $W_1W_2W_3$ begins with $PT_1T_2T_0T_1T_2T_0T_1$ and so it contains an overlap. □

LEMMA 3.3. *Given an alphabet Σ with $|\Sigma| \geq 3$, then the set $T(\Sigma)$ contains all overlap-free words over Σ of length at most 3 as factors.*

PROOF. It is clear that any overlap-free word of length 3 is a factor of *aab*, *bcc*, *aba*, or *abc* with a, b, c being three distinct letters. The words *aab* and *bcc* are factors of the word *aabcc* in $T_2(\Sigma)$. The word *aba* is a word in $T_1(\Sigma)$. □

LEMMA 3.4. *Let Σ and Δ be alphabets such that $3 \leq |\Sigma| \leq |\Delta|$. For any different letters $a, b \in \Sigma$, then word aba must be a factor of any test-set for overlap-freeness of morphisms from Σ^* to Δ^*.*

PROOF. Without loss of generality we can assume that $\Sigma \subseteq \Delta$. Let $a, b, c \in \Sigma$ be three distinct letters. Now define the morphism $f : \Sigma^* \to \Delta^*$ to be

$$f(a) = acbabca$$

$$f(b) = bacbcab$$

$$f(d) = dbbabbd \text{ for all letters } d \in \Sigma \setminus \{a, b\}.$$

To restate the lemma, we want to prove that this morphism verifies for any overlap-free word W over Σ that $f(W)$ contains an overlap if and only if *aba* is a factor of W.

We begin by observing that $f(aba) = (acba)b(cabac)b(cabac)b(abca)$.

Let $W \in \Sigma^*$ be overlap-free such that $f(W)$ is not overlap-free. Let a be a letter and P, V, S be words such that $f(W) = PaVaVaS$. We now must prove that

aba is a factor of W. Without loss of generality assume that $|Pa| \leq |f(W[1])|$ and $|aS| \leq f(W[|W|])$. Particularly we will show that $W = aba$.

Assume that $|W|_d \neq 0$ for some $d \in \Sigma \setminus \{a,b,c\}$. The hypothesis on $|Pa|$ and on $|aS|$, we see that $|aVaVa|_d \geq 1$ and so $|aV|_d \neq 0$. Now, let V_1, V_2 be words such that $aV = V_1 d V_2$. Clearly we have that $V_2V_1 \neq \varepsilon$ so $aVaVa \neq ddd$. So by the definition of f we see that $|V_2V_1| \geq 5$. So we either need that $|V_2| \geq 2$ or $|V_1| \geq 2$. If $|V_2| \geq 2$ and V_2 starts with bb, then there exists words W_1, W_2, W_3 such that $f(W_1) = pV_1$, $f(W_2) = dV_2V_1$, $f(W_3) = dV_2aS$, and $W = W_1W_2W_3$. If $|V_1| \neq \varepsilon$, a is the first letter of V_1 and we get a contradiction by Lemma 3.2 with W being overlap-free. If $V_1 = \varepsilon$, then $a = d$, W_2 begins with d, and because the image words of f begin and end with different letters, then W_3 starts with W_2d and again we have a contradiction with W begin overlap-free. If $|V_2| \geq 2$ and V_2 does not begin with bb, then there exist words W_1, W_2, W_3 such that $f(W_1) = PV_1d$, $f(W_2) = V_2V_1d$, $f(W_3) = V_2aS$, and $W = W_1W_2W_3$. Since the first letter of V_1d is a by Lemma 3.2, we have a contradiction with W being overlap-free. The case $|V_1| \geq 2$, considering the two last letters of V_1 leads to the same contradiction.

So we see that $W \in \{a,b,c\}^*$.

Now, by the definition of f for any letters $x, y, z \in \{a,b,c\}$, $f(x)$ cannot be an internal factor of $f(yz)$ (that is $f(yz) = \delta_1 f(x) \delta_2$ with $\delta_1 \neq \varepsilon$, $\delta_2 \neq \varepsilon$).

Assume that there exists an integer i with $1 \leq i \leq |W|$, and two words V_1, V_2 such that $f(W[1..i]) = PaV_1$, $V = V_1 f(W[i+1]) V_2$, and $f(W[i+2..|W|]) = V_2 aVaS$. In this case, we have $f(W[i+2..|W|]) = V_2aV_1f(W[i+1])V_2aS$. From the last paragraph the must be some integer j so that $i+2 \leq j < |W|$, such that $f(W[i+2..j]) = V_2aV_1$, $f(W[j+1]) = f(W[i+1])$, $f(W[j+2..|W|]) = V_2aS$. Let $W_1, W_2, W_3, \mathcal{V}_2$ be the words such that $W_1 = W[1..i]$, $W_2 = W[i+1..j]$, $W_3 = W[j+1..|W|]$, $\mathcal{V}_2 = f(W[i+1])V_2$. We have $f(W_1) = PaV_1$, $f(W_2) = \mathcal{V}_2aV_1$, $f(W_3) = \mathcal{V}_2aS$. By Lemma 3.2, $W = W_1W_2W_3$ is not overlap-free which is a contradiction.

A similar contradiction arises if there is some integer i and words V_1, V_2 such that $f(W[1..i]) = PaVaV_2$, $V = V_1 f(W[i+1])V_2$, and $f(W[i+2..|W|) = V_2 aS$.

From $|Pa| \leq |f(W[1])|$ and $|aS| \leq |f(W[|W|])$, it follows that $|W| \leq 3$ (otherwise one of the two previous contradictory situations holds). An exhaustive verification shows that we must have $W = aba$. $\square$

LEMMA 3.5. *Let $p \geq 1$ be an integer, Σ be an alphabet containing at least $2p+2$ letters, and Δ an alphabet with $|\Sigma| \leq |\Delta|$. Given any letters $c_1, \ldots, c_p, d_1, \ldots d_p, \beta, \gamma \in \Sigma$ which are distinct, and any letters $x \in \{c_1, \ldots, c_p, \gamma\}$, and $y \in \{d_1, \ldots, d_p\}$, then word $xc_1 \cdots c_p \beta d_1 \cdots d_p y$ must be a factor of any test-set for overlap-freeness of morphisms from Σ^* to Δ^**

PROOF. Similarly to Lemma 3.4 because $|\Sigma| \leq |\Delta|$ we can assume without loss of generality that $\Sigma \subseteq \Delta$.

Let $p, \Sigma, c_1, \ldots, c_p, d_1, \ldots, d_p, \beta, \gamma, x, y$ be as the hypothesis of the lemma describes. Now, we want to consider the morphism $f : \Sigma^* \to \Delta^*$ as defined by

$$\begin{aligned}
f(c_i) &= c_i\ d_{i-1}\ d_i\ \beta\ c_i\ \beta\ \beta\ c_i \quad \text{for all } 1 \leq i \leq p, \\
f(d_i) &= d_i\ \beta\ c_i\ \beta\ \beta\ c_i\ c_{i+1}\ d_i \quad \text{for all } 1 \leq i \leq p, \\
f(\gamma) &= \gamma\ \beta\ d_1\ \beta c_1\ \beta\ \beta\ \gamma, \\
f(\beta) &= \beta\ d_p\ y\ \beta\ \beta\ x\ c_1\ \beta, \\
f(a) &= a\ d_1\ d_1\ c_1\ d_1\ c_1\ c_1\ a \quad \text{for all } a \in \Sigma \setminus \{c_1, \ldots, c_p, \beta, d_1, \ldots, d_p, \gamma\}.
\end{aligned} \tag{6}$$

Now to restate the lemma, we need to prove that this morphism verifies for any overlap-free word $W \in \Sigma^*$ that $f(W)$ contains an overlap if and only if W contains $xc_1 \cdots c_p \beta d_1 \cdots d_p y$ as a factor.

Observe first that $\beta\beta x f(c_1 \cdots c_p)\beta d_p y \beta = \beta\beta x c_1 \beta f(d_1 \cdots d_p) y \beta$. Thus we see that $f(xc_1 \cdots c_p \beta d_1 \cdots d_p y)$ contains an overlap.

Conversely, let W be an overlap-free word over Σ such that $f(W)$ contains an overlap and let $n = |W|$. There exists some words P_1, V, S_n and a letter α such that $f(W) = P_1\alpha V\alpha V\alpha S_n$. We must now prove that $xc_1 \cdots c_p\beta d_1 \cdots d_p y$ is a factor of W. Without loss of generality, we can assume that $|P_1\alpha| \leq |f(W[1])|$ and $|\alpha S_n| \leq |f(W[n])|$. In this particular case we show that $W = xc_1 \cdots c_p\beta d_1 \cdots d_p y$.

Claim. $\text{alph}(W) \subseteq \{c_1, \ldots, c_p, d_1, \ldots, d_p, \beta, \gamma\}$.

Proof of the Claim. If the claim is not true, then there must be some letter $a \in \Sigma \setminus \{c_1, \ldots, c_p, d_1, \ldots, d_p, \beta, \gamma\}$ that occurs in αV. Let V_1, V_2 be two words such that $\alpha V = V_1 a V_2$. Notice that $f(W) = P_1 V_1 a V_2 V_1 a V_2 \alpha S_n$, and also observe that aaa cannot occur in $f(W)$. So we see that $V_1V_2 \neq \varepsilon$, and consequently we see that $|V_2V_1| \geq 6$. This implies $|V_1| \geq 2$ or $|V_2| \geq 2$.

If $|V_1| \geq 2$ and V_1 ends with c_1c_1, then there exists three words W_1, W_2, W_3 such that $f(W_1) = P_1V_1a$, $f(W_2) = V_2V_1a$, $f(W_3) = V_2\alpha S_n$, and $W = W_1W_2W_3$. By Lemma 3.2, we get a contradiction with W being overlap-free.

If $|V_1| \geq 2$ and V_1 does not end with c_1c_1, then there exist three words W_1, W_2, W_3 such that $f(W_1) = P_1V_1$, $f(W_2) = aV_2V_1$, $f(W_3) = aV_2\alpha S_n$, and $W = W_1W_2W_3$. In the case when $V_1 = \varepsilon$, then $\alpha = a$. By the definition of f given in equation (6), W_2 starts with a and W_2a is a prefix of W_3. So W is not overlap- free, which is a contradiction. If $V_1 \neq \varepsilon$, we get the same contradiction by Lemma 3.2.

In the case $|V_2| \geq 2$, looking at the beginning of V_2, we arrive at the same contradiction. This finishes our claim. $\square$

Note that for all letters $a, b \in \Sigma$, $f(ab)$ must be overlap-free. It follows that $|P_1\alpha V\alpha| > |f(W[1])|$, and $|\alpha V\alpha S_n| > |f(W[n])|$. So, there exists an integer i with $1 < i < n$ and two words P_i and S_i such that $f(W[i]) = P_i\alpha S_i$ and

$$V = S_1 f(W[2..i-1]) P_i = S_i f(W[i+1..n-1]) P_n. \tag{7}$$

If $|S_1| = |S_i|$, then $S_1 = S_i$. Taking $W_1 = W[1]$, $V_0 = P_1$, $V_1 = S_1$, $W_2 = W[2..i-1]$, $V_2 = f(W[2..i1])P_i = f(W[i+1..n-1])P_n$, $W_3 = W[i+1..n]$, and $V_3 = S_n$, Lemma 3.2 shows that W contains an overlap, which is a contradiction.

Thus we know that $|S_1| \neq |S_i|$ and, since f is uniform, this implies that $|P_i| \neq |P_n|$. This also implies that for any k and l with $1 \leq k \leq i-1 < l \leq n-1$, that $|S_1 f(W[2..k])| \neq |S_i f(W[i+1..l])|$.

We begin by considering the case when $n = 3$, that is $i = 2$, $W[2..i-1] = \varepsilon$ and $W[i+1..n-1] = \varepsilon$. If $|S_1| > |S_i|$, then equation (7) gives that there is a non-empty word Z such that $S_1 = S_i Z$, $P_n = ZP_i$. This is impossible since there are no values of P_i, α, S_i, or Z such that $P_i \alpha S_i \in f(\Sigma)$, $Z \neq \varepsilon$, $ZP_i\alpha$ a prefix of $f(\Sigma)$, and $\alpha S_i Z$ a suffix of $f(\Sigma)$. If $|S_1| < |S_i|$, let Z be the non-empty word such that $S_i = S_1 Z$, and $P_i = ZP_n$. Since Z is both a prefix and a suffix of $f(W[i])$, the word Z must be a letter. This is again impossible since there are no values of Z, P_n, S_1 such that $ZP_n\alpha S_1 Z \in f(\Sigma)$, $P_n\alpha$ a prefix of $f(\Sigma)$, and αS_1 a suffix of $f(\Sigma)$.

From now on , $n > 3$, and so $W[2..i-1] \neq \varepsilon$ or $W[i+1..n-1] \neq \varepsilon$.

Assume first that $|S_1| < |S_i|$. In this case $W[2..i-1] \neq \varepsilon$ (in particular $i \geq 3$). Indeed if $W[i+1..n-1] \neq \varepsilon$, from equation (7) we get that $|f(W[2..i-1)| > |fW[i+1..n-1])| \geq 8$. Further because $|P_i| \leq 7$, we see that $W[2..i-1] \neq \varepsilon$.

From equation (7), there must be two non-empty words $\mathcal{P}_2$ and $\mathcal{S}_2$ such that $S_i = S_1\mathcal{P}_2$, $f(W[2]) = \mathcal{P}_2\mathcal{S}_2$, and

$$\mathcal{S}_2 f(W[3..i-1])P_i = f(W[i+1..n-1])P_n. \tag{8}$$

Observe that $\mathcal{P}_2\mathcal{S}_2 \in f(\Sigma)$, $\mathcal{P}_2$ is a non-empty suffix of $f(\Sigma)$, and $\mathcal{S}_2$ is a non-empty prefix of $f(\Sigma)$. By the definition of f in equation (6), there exists an integer j with $1 < j < p$ such that $(\mathcal{P}_2, \mathcal{S}_2, W[2]) = (d_j\beta c_j\beta\beta c_j,\ c_{j+1}d_j,\ d_j)$ or $(\mathcal{P}_2, \mathcal{S}_2, W[2]) = (c_j d_{j-1},\ d_j\beta c_j\beta\beta c_j,\ c_j)$. We now consider both of these cases.

Case $(\mathcal{P}_2, \mathcal{S}_2, W[2]) = (d_j\beta c_j\beta\beta c_j,\ c_{j+1}d_j,\ d_j)$. The last letter of S_i is c_j the last letter of $\mathcal{P}_2$. , and the first letter of $f(W[i+1])$ is c_{j+1} which is the first letter of

$\mathcal{S}_2$ (note that this is even true if $i+1=n$). By the definition of f we can see that $W[i]=c_j$ and $W[i+1]=c_{j+1}$. Let $d_{p+1}=y$ since $P_i\alpha$ is a prefix of $f(W[i])$, since $P_n\alpha$ is a prefix of $f(W[n])$, and since $f(W[i+1]$ starts with $c_jd_jd_{j+1}$. From equation (8) we deduce that $f(W[3])$ starts with d_{j+1} (even if $i=3$ or $i+1=n$). Consequently $W[3]=d_{j+1}$.

If $i>3$, then $n>i+1$. We cannot have $j+1=p$. Indeed otherwise $\mathcal{S}_2f(W[3])$ begins with $\mathcal{S}_2y\beta\beta$ which is not a prefix of $f(W[i])$, a contradiction. Consequently, there exist two words $\mathcal{P}_3, \mathcal{S}_3$ such that $(\mathcal{P}_3, \mathcal{S}_3, W[3]) = (d_{j+i}\beta c_{j+1}\beta\beta c_{j+1},\ c_{j+2}d_{j+1},\ d_{j+1})$.

By induction, we can state that $W[2..i]=d_j\cdots d_{j+i-2}$, $W[i..2i-2]=c_j\cdots c_{j+i-2}$ with $j+i\leq p+1$ (and $2i-2\leq n$). Since we cannot have $d_{j+i-2}=c_j$, this case is impossible.

Case $(\mathcal{P}_2, \mathcal{S}_2, W[2]) = (c_jd_{j-1},\ d_j\beta c_j\beta\beta c_j,\ c_j)$. Similarly to the previous case, we can state that $W[2..i]=c_j\cdots c_{j+i-2}$, $W[i..2i-2]=d_{j-1}\cdots d_{j+i-3}$ with $j+i-2\leq p+1$ (and $21-2\leq n$. The only possibility to have $c_{j+i-2}=d_{j-1}$ is that $j=1$ and $j+i-2=p+1$. More simply we have $j=1$ and $i=p+2$. In this condition $W[2..p+1]=c_1\cdots c_p$, $W[p+2]=\beta$ and $W[p+3..2p+2]=d_1\cdots d_p$. The last letter of $f(W[1])$ is $W[1]$ by the definition of f. It is also the last letter of αS_1. From $\alpha S_i=\alpha S_1\mathcal{P}_2$, $\mathcal{P}_2=c_1\beta$, $W[i]=\beta$, we deduce that $W[1]=x$. From equation (7) and $\alpha S_i=\alpha S_1\mathcal{P}_2=\alpha S_1c_1\beta$, we get that $n=2p+3$ and $P_i=\beta d_pP_n\alpha$. By the definition of f we must have $\alpha=\beta$, $S_1=x$, and $P_n=y$ which implies that $W[n]=y$. Thus, $W=xc_1\cdots c_p\beta d_1\cdots d_py$.

To end this proof we must consider what happens when $|S_1|>|S_i|$. This situation can be treated as the previous one. First, we can see that $W[i+1..n-1]\neq\varepsilon$. Consequently there exist words $\mathcal{P}_{i+1}$ and $\mathcal{S}_{i+1}$ such that $S_1=S_i\mathcal{P}_{i+1}$, $f(W[i+1])=\mathcal{P}_{i+1}\mathcal{S}_{i+1}$, and

$$\mathcal{S}_{i+1}f(W[i+2..n-1])P_n = f(W[2..i-1])P_i.$$

Thus, there is an integer j between 1 and p such that $(\mathcal{P}_{i+1}, \mathcal{S}_{i+1}, W[i+1]) = (d_j\beta c_j\beta\beta c_j,\ c_{j+1}d_j,\ d_j)$ or $(\mathcal{P}_{i+1}, \mathcal{S}_{i+1}, W[i+1]) = (c_jd_{j-1},\ d_j\beta c_j\beta\beta c_j,\ c_j)$. Similarly to the cases for $(\mathcal{P}_2, \mathcal{S}_2, W[2])$, we can show these two cases are impossible. $\square$

LEMMA 3.6. *Let $p \geq 1$ be an integer, let Σ be an alphabet containing at least $2p+1$ letters, and let Δ be an alphabet such that $|\Sigma| \leq |\Delta|$. Given any distinct letters $c_1, \ldots, c_p, d_1, \ldots, d_p, \beta \in \Sigma$, given $x \in \{c_1, \ldots, c_p\}$, and $y \in \{d_1, \ldots, d_p\}$ the word $xc_1 \cdots c_p\beta d_1 \cdots d_p y$ must be a factor of any test-set for overlap-freeness of morphisms from Σ^* to Δ^*.*

The proof of Lemma 3.6 is similar to that of Lemma 3.5 considering the same morphism but excluding the definition of $f(\gamma)$.

LEMMA 3.7. *Let $p \geq 1$ be an integer, let Σ be an alphabet containing at least $2p+1$ letters, and let Δ be an alphabet with $|\Sigma| \leq |\Delta|$. Given any distinct letters $c_1, \ldots, c_p, d_1, \ldots, d_p, \gamma \in \Sigma$, the word $\gamma c_1 \cdots c_p d_1 \cdots d_p \gamma$ must be a factor of any test-set for overlap-freeness of morphisms from Σ^* to Δ^*.*

PROOF. Without loss of generality we assume that $\Sigma \subseteq \Delta$. Suppose that we have $p, \Sigma, c_1, \ldots, c_p, d_1, \ldots, d_p, \gamma$ as in the hypothesis of Lemma 3.7. In what follows $c_0 = d_{p+1} = \gamma$. We consider the morphism $f : \Sigma^* \to \Delta^*$ defined as

$$f(c_i) = c_i\ d_i\ d_{i+1}\ d_{i+1}\ d_i\ d_{i+1}\ c_i \quad \text{for all } 1 \leq i \leq p,$$

$$f(d_i) = d_i\ c_{i-1}\ c_i\ d_i\ d_{i+1}\ d_{i+1}\ d_i \quad \text{for all } 1 \leq i \leq p,$$

$$f(\gamma) = \gamma\ c_p\ d_1\ c_p\ d_1\ d_1\ \gamma,$$

$$f(a) = a\ c_1\ c_1\ d_1\ c_1\ c_1\ a \quad \text{for all } a \in \Sigma \setminus \{c_1, \ldots, c_p, d_1, \ldots, d_p, \gamma\}.$$

We can observe that $d_1\gamma f(c_1 \cdots c_p)d_1 = f(d_1 \cdots d_p)\gamma c_p d_1$.

The end of the proof of Lemma 3.7 can be done similarly to the proof of Lemma 3.5, proving that this morphism verifies for any overlap-free word $W \in \Sigma^*$ that $f(W)$ contains an overlap if and only if W contains $\gamma c_1 \cdots c_p d_1 \cdots d_p \gamma$ as a factor. $\square$

3.2. Main Theorem of Richomme and Wlazinski

Theorem 3.8. *Given alphabets Σ and Δ with $3 \leq |\Sigma| \leq |\Delta|$, the set $T(\Sigma)$ is a test-set for overlap-freeness of uniform morphisms from Σ^* to Δ^*.*

Proof. Let f be a L-uniform morphism for an integer $L \geq 0$. All words in $T(\Sigma)$ are overlap-free. Thus if f is overlap free, then $f(T(\Sigma))$ is overlap-free. (Note the definition of $T(\Sigma)$ is in equation (5)).

Now assume that f is not overlap-free. In particular assume $f \neq \varepsilon$, that is, $L \neq 0$. Because f is uniform it is clearly non-erasing. We now want to show that $f(T(\Sigma))$ is not overlap-free. If two of the image words in f begin or end with the same letter, then there exist two letters x and y such that $f(xxy)$ or $f(xyy)$ contains an overlap. The set $f(T(\Sigma))$ is not overlap-free because Lemma 3.3 tells us that xxy and xyy are in $\mathrm{Fact}(T(\Sigma))$.

From now on we will assume that all of the image words in f begin with different distinct letters and end with different distinct letters.

Let W be one of the shortest overlap-free words over Σ such that $f(W)$ is not overlap-free. Let a be a letter, and let P, V, S be words such that $f(W) = PaVaVaS$. By our hypotheses on W, we have $|f(W[1])| \geq |Pa|$, and setting $n = |W|$, $|f(W[n]) \geq |aS|$.

By Lemma 3.3, we can assume that $n \geq 4$.

Let i be the smallest integer such that $|PaVa| \leq f(W[1..i])|$. We have $|f(W[i+1..n])| \leq |VaS|$, and so $|f(W[i+1..n-1])| \leq |V| < |f([1..i])|$. Since f is uniform, and since $n \geq 4$, $i \neq 1$. By definition of i, we also have $|PaVa| > |f(W[1..i-1])|$, and so $|VaS| < |f(W[i..n])|$. It follows that $|f(W[2..i-1])| < |aV| < |f(W[i..n])|$. Thus, since $n \geq 4$, $i \neq n$.

Let S_i, P_i, S_i, P_n be the words such that $f(W[1]) = PaS_1$, $f(W[i]) = P_iaS_i$, $f(W[n]) = P_naS$, and

$$V = S_1 f(W[2..i-1])P_i = S_i f(W[i+1..n-1])P_n. \tag{9}$$

Since f is L-uniform, and since both $|P_i| < L$ and $|P_n| < L$, we have $S_1 = S_i$ if and only if $P_i = P_n$. In this case, by Lemma 3.2, with $W_1 = W[1]$, $W_2 = W[2..i]$, $W_3 = W[i+1..n]$, $V_0 = P$, $V_1 = S_1$, $V_2 = f(W[2..i-1])P_i = f(W[i+1..n-1])P_n$, and $V_3 = S$, we get that W is not overlap-free, a contradiction.

Thus $|S_1| \neq |S_i|$ and $|P_i| \neq |P_n|$. Let j, k be two different integers such that $1 \leq j < k \leq n$ and $W[j] = W[k]$. We will prove:

Claim. One of the three following assertions is verified:

(1) $j = 1$, $1 < k < i$, and $|S_1| < |S_i|$,

(2) $i < j < n$, $k = n$, and $|P_n| < |P_i|$,

(3) $j = 1$ and $k = n$.

Proof. By contradiction, we successively exclude the other cases.

Case 1. We cannot have $1 < j < k < i$, or $1 = j < k < i$ with $|S_1| > |S_i|$.

Assume for a contradiction that $1 < j < k < i$, or $1 = j < k < i$ with $|S_1| < |S_i|$.

Let $\ell \geq i$ be the greatest integer so that $aS_i f(W[i{+}1..\ell])$ is a prefix of $aS_1 f(W[2..j])$ (note that if $j = 1$, then $\ell = i$). Since f is uniform, $\ell + k - j$ is the greatest integer such that $aS_i f(W[i+1..\ell+k-j])$ is a prefix of $aS_1 f(W[2..k])$. Let Y be the word such that

$$aS_1 f(W[2..j]) = aS_i f(W[i+1..\ell])Y. \tag{10}$$

Since $0 < ||S_1| - |S_i|| < L$, the word Y is non-empty, and $|Y| < |f(W[j])|$. Since f is uniform, and since $W[j] = W[k]$, we also have

$$aS_1 f(W[2..k]) = aS_i f(W[i+1..\ell+k-j])Y. \tag{11}$$

From the above, we have the following deductions.

(i) $W[l] = W[\ell + k - j]$. From the two previous equations, we get that the last letter of $f(W[l])$ is also the last letter of $f(W[\ell + k - j])$. In fact, it is the $(L - |Y|)^{\text{th}}$

letter of $f(W[j])$. Because the image words of f do not begin or end with the same letters, we see that $W[\ell] = W[\ell + k - j]$.

(ii) $W[j-1] = W[k-1]$. Note that we must have $j > 1$. Now let X be the word such that $f(W[j]) = XY$. From equation (10), X is a suffix of $f(W[\ell])$ (even if $\ell = i$). Since $|X| = L - |Y| < L$, there exists a letter b such that X is a suffix of $f(W[\ell])$. From equations (10) and (11), this letter is also the last letter of $f(W[j-1])$ (more precisely of aS_1 if $j = 2$) and of $f(W[k-1])$. Because the image words of f do not begin or end with the same letter, we see that $W[j-1] = W[k-1]$.

(iii) $W[\ell{+}1] = W[\ell{+}k{-}j{+}1]$. From equations (9), (10), and (11), we see that Y is the first letter of $f(W[\ell{+}1])$ and of $f(W[\ell{+}k{-}j{+}1])$. Thus, $W[\ell{+}1] = W[\ell{+}k{-}j{+}1]$. (Note that in the case with $k = i$ and $|P_n| < |P_i|$, we can still define $\ell(\ell{+}k{-}j = n{-}1)$ and the previous techniques can be applied to prove that $W[\ell + 1] = W[n]$).

(iv) $W[j{+}1] = W[k{+}1]$. Let γ be the letter such that γY is a prefix of $f(W[l{+}1])$. It is also the first letter of $f(W[i])$ and of $f(W[k+1])$. Thus, $W[j+1] = W[k+1]$.

By induction, we get $W[j'] = W[j' + k - j]$, for all j' such that, $1 \leq j' < j' + k - j \leq i$, and for all j' such that, $i \leq j' < j' + k - j \leq n$. Recall that P is a prefix of $f(W[1]) = f(W[1 + k - j])$, and P_i is a prefix of $f(W[i]) = f(W[i + k - j])$. We now have that $|S_1 P| = |S_i P_i| = L - 1$. Thus from equation (9), because f is uniform, we get $S_1 f(W[2..k-j])P = S_i f(W[i+1..i+k-j-1]P_i$. It follows that

$$aS_1 f(W[k-j+2..i-1])P + ia = aS_i f(W[i+k-j+1..n-1])P_n a. \qquad (12)$$

Let $X = W[1]W[k{-}j{+}2..i{-}1]W[i]W[i{+}k{-}j{+}1..n]$. From equation (12), $f(X)$ is not overlap-free. We have also proved that $W[1] = W[1{+}k{-}j]$ and $W[i{+}k{-}j{+}1..n] = W[i + 1..n - k + j]$. Thus $X = W[1 + k - j..n - k + j]$ is a factor of W, and so it is overlap-free. Finally, since $k - j \geq 1$, $|X| < |W|$. This contradicts the assumption that W is one of the shortest overlap-free words such that $f(W)$ not overlap-free.

Similarly (as it is the same logic), we can prove case 2 below.

Case 2. We cannot have $i < j < k < n$, or $i < j < k = n$ with $|P_n| > |P_i|$.

Case 3. We cannot have $1 \leq j \leq k = 1$.

Otherwise let b be the last letter of aS_i. Since $W[i] = W[k] = W[j]$, it is the last letter of $aS_1 f(W[2..j])$. We first consider the case $|aS_1 f(W[2..j])| > |aS_i|$. Let M be the word such that $aS_1 f(W[2..j]) = aS_i Mb$. From equation (9), the word Mb is a prefix of $f(W[i+1..n])$. Let $m \leq n$ be the least integer such that Mb is a prefix of $f(W[i+1..m])$. If $|aS_1| < |aS_1|$, let $X = W[1..j]W[i+1..m]$. If $|aS_1| < |aS_i|$, let $X = W[2..j]W[i+1..m]$. Since $j < i$, we have $|X| < |W|$. Moreover, the word $f(X)$ contains the overlap $b\mathcal{V}b\mathcal{V}b$. From Case 1, each letter occurring in $W[2..j]$, or in $W[1..j]$ only occurs once as long as $|aS_1| > |aS_i|$. From Case 2 (not that since $j < i$, by the definition of m, in this case $m = n$, we have $|P_n| > |P_i|$), each letter occurring in $W[i+1..m]$ occurs only once. Thus, X is overlap-free. This contradicts the hypothesis "W is one of the shortest overlap-free words such that $f(W)$ is not overlap-free."

The same contradiction arises when $j = 1$ and $|aS_1| < |aS_i|$, since in this case we can show as previously that $f(W[1]W[2])$ is not overlap-free.

Similarly we can prove the three following cases.

Case 4. We cannot have $i = j < k \leq n$ (in the proof, take b as the first letter of $P_i a$).

Moreover with the same techniques as in Cases 3 and 4, we can prove the following cases.

Case 5. We cannot have $1 < j < i$ and $k = n$ (in the proof, take b as the first letter of $f(W[i]))$.

Case 6. We cannot have $j = 1$ and $i < k < n$ (in the proof, take b as the last letter of $f(W[k]))$.

Case 7. We cannot have $1 < j < i < k < n$.

Since $|S_1| \neq |S_i|$ and since f is uniform we have $|S_1 f(W[2..j])| \neq |S_f(W[i+1..k])|$.

We consider the case $|S_1 f(W[2..j])| < |S_f(W[i+1..k])|$ (as the logic in the case with $|S_1 f(W[2..j])| > |S_f(W[i+1..k])|$ is virtually identical). Let b be the first letter of $f(W[j]) = f(W[k])$. From equation (9), there exists a word M such that $S_i f(W[i+1..k-1]) = S_1 f(W[2..j-1])bM$. The word bM is a suffix of $f(W[i..k-1]$, and the word bMb is a prefix of the word $f(W[j..i])$. From Case 3, each letter occurring in $W[j..i]$ occurs only once. From Case 4, each letter occurring in $W[i..k-1]$ occurs only once. Thus the word $X = W[i..k-1]W[j..i]$ is overlap- free. Moreover $|X| < |W|$ and $f(W)$ contains the overlap $bMbMb$. This once again contradicts the "shortest" hypothesis on W. This concludes the argument for the claim.

We now end the proof of Theorem 3.8.

If $|S_1| > |S_i|$ and $|P_n| > |P_i|$, or if $|S_1| < |S_i|$ and $|P_n| < |P_i|$, since f is uniform, from equation (9) we deduce that $|W[2..i-1]| = |W[i+1..n-1]|$. Let $W_1 = W[2..i-1]$, $W_2 = W[i+1..n-1]$, $x = W[1]$, $y = W[i]$, $z = W[n]$. Since $n \geq 4$, $|W_1| = |W_2| \geq 1$. By the Claim, $\forall a \in \Sigma\ |W_1 y W_2|_a \leq 1$, $|W_1|_z = 0 = |W_2|_x$, $y \neq x$ and $y \neq z$. Thus we have $W = xW_1 y W_2 z \in T_2(\Sigma)$.

If $|S_1| > |S_i|$ and $|P_n| > |P_i|$, we have $|W[2..i-1]| = |W[i+1..n-1]| + 1$. Let $W_1 = W[2..i-1]$, $W_2 = W[i+1..n-1]$, $x = W[1]$, $y = W[i]$, and $z = W[n]$. From the Claim we have that $|W_1|_z = 0 = |W_2|_x$, $\forall a \in \Sigma\ |W_1 y W_2|_a \leq 1$, $y \neq x$ and $y \neq z$. Moreover, $|W_2|_z = 0$ from assertion 2 in the Claim. If $x = z$, then $W \in T_1(\Sigma)$. Otherwise $x \neq z$ and we have $Wy \in T_2(\Sigma)$.

The case $|S_1| > |S_i|$ and $|P_n| < P_i|$ can be treated the same way and we are done. □

Corollary 3.9. *Given alphabets Σ and Δ with $3 \leq |\Sigma| \leq |\Delta|$, a set T of non-overlap-free words over Σ is a test-set for overlap-freeness of uniform morphisms from Σ^* to Δ^* if and only if $T(\Sigma) \subseteq \mathrm{Fact}(T)$.*

Proof. Let Σ and Δ be alphabets such that $3 \leq |\Sigma| \leq |\Delta|$. Let T be a set of overlap- free words over Σ.

If $T(\Sigma) \subseteq \text{Fact}(T)$, then by Theorem 3.8, the set T is a test-set for overlap-freeness of uniform morphisms from Σ^* to Δ^*.

Conversely assume that T is a test-set for overlap-freeness of uniform morphisms from Σ^* to Δ^*. We must prove that $T_1(\Sigma) \subseteq \text{Fact}(T)$ and $T_2(\Sigma) \subseteq \text{Fact}(T)$.

Let $xW_0x \in T_1(\Sigma)$ with $x \in \Sigma$ and $W_0 \in \Sigma^*$. Since for $a \in \Sigma$, $|xW_0|_a \leq 1$, $|\Sigma| \geq |xW_0$. If $|W_0$ is even, let $p = |W_0|/2$. We have that $|\Sigma| \geq 2p + 1$. There exist $2p$ distinct letters $c_1, \ldots, c_p, d_1, \ldots, d_p$ such that $W = c_1 \cdots c_p d_1 \cdots d_p$. Taking $x = \gamma$, Lemma 3.7 implies that xW_0x belongs to $\text{Fact}(T)$. If $|W_0|$ is odd, we obtain the same conclusion using Lemma 3.5 if $|W_0| \geq 3$, and if $|W_0| = 1$, we use Lemma 3.4.

Let $xW_1yW_2z \in T_2(\Sigma)$ with $x, y, z \in \Sigma$ and $W_1, W_2 \in \Sigma^*$. let $p = |W_1| = |W_2| \geq 1$. Since for $a \in \Sigma$, $|W_1yW_2| \leq 1$, there exist $2p$ distinct letters $c_1, \ldots, c_p, d_1, \ldots, d_p$ so that $W_1 = c_1 \cdots c_p$ and $W_2 = d_1 \cdots d_p$ where $y \notin \{c_1, \ldots, c_p, d_1, \ldots, d_p\}$. Note that from $|yW_2|_x = 0 = |W_1y|_z$, $x \notin \{d_,1 \ldots, d_p, y\}$, and $z \notin \{c_1, \ldots, c_p, y\}$. If $x \in \{c_1, \ldots, c_p\}$ and $z \in \{d_1, \ldots, d_p\}$, then by Lemma 3.6, $xc_1 \cdots c_p y d_1 \cdots d_p z \in \text{Fact}(T)$. If $x \in \{c_1, \ldots, c_p\}$ and $z \notin \{d_1, \ldots, d_p\}$, then $|\Sigma| \geq 2p + 2$. By taking $\gamma = z$, we get Lemma 3.5 showing that $xc_1 \cdots c_p y d_1 \cdots d_p z \in \text{Fact}(T)$. Similarly the same conclusion holds if $x \notin \{c_1, \ldots, c_p\}$ and $z \in \{d_1, \ldots, d_p\}$. Finally, we consider the case $x \notin \{c_1, \ldots, c_p\}$ and $z \notin \{d_1, \ldots, d_p\}$. If $x = z$, then $xW_1yW_2z \in T_1(\Sigma) \subseteq \text{Fact}(T)$. If $x \neq z$, then $xW_1yW_2zx \in T_1(\Sigma) \subseteq \text{Fact}(T)$. Thus we are done. $\square$

Chapter 4

The Latin Square Morphism

4.1. Latin Squares

Now that we have a classification for overlap-free morphisms, we want to explore some specific morphisms that are overlap-free. So we discuss the concept of a Latin square. In (Tompkins, 2007), an infinite overlap-free word was introduced using the concept of Latin squares. We now want to focus on the morphisms created from that paper.

Definition 4.1. A Latin square is an $n \times n$ table filled with n different symbols such that each symbol appears only once in each column and only once in each row.

As with words, we will use the natural numbers $\{0, 1, \ldots, n-1\}$ for our symbols to fill each Latin square. Furthermore, we will call a Latin square reduced if the first column is $[0\ 1\ \cdots\ n-1]^\top$. Notice that the only reduced Latin square on two letters is the Latin square

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

which represents two things. It first represents the multiplication table (i.e. Cayley table) for $\mathbb{Z}/2\mathbb{Z}$, and also the rows represent the image words of the Thue-Morse morphism.

It is clear that any finite group can be represented as a latin square. For instance the Latin square representing the Klien-4 group follows as

$$\begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 3 & 2 \\ 2 & 3 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix}$$

Not only can we represent the multiplication tables for groups with latin squares, but we can even represent the multiplication tables for quasigroups using latin squares. We note that quasigroups (groups without association) are algebraic structures with a closed multiplication that is not necessarily associative. For example the following latin square represents the multiplication of an algebraic structure that is not a group.

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & 5 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 & 5 & 0 \\ 3 & 4 & 0 & 5 & 1 & 2 \\ 4 & 5 & 3 & 0 & 2 & 1 \\ 5 & 2 & 1 & 4 & 0 & 3 \end{bmatrix}$$

4.2. The Latin Square Morphism and Tilings

4.2.1. Latin Squares. Using latin squares we wish to produce an overlap-free morphism. So we consider the following definition.

Definition 4.2. Let $\mathcal{L}$ be a reduced $n \times n$ latin square, and let L_i be the i-th row of $\mathcal{L}$. We call the morphism given by

$$\ell(i) = L_i,$$

the latin square morphism, where $i \in \Sigma_n$.

We begin by noticing that the Thue-Morse morphism is a latin-square morphism

$$\begin{array}{ccc} & \mu & \\ 0 & \mapsto & 01 \\ 1 & \mapsto & 10. \end{array}$$

Further we could use the latin square for the cyclic group $\mathbb{Z}/3\mathbb{Z}$,

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

The corresponding morphism for this latin square appears as

$$\ell(t) = \begin{cases} 012, & \text{for } t = 0 \\ 120, & \text{for } t = 1 \\ 201, & \text{for } t = 2. \end{cases}$$

Two examples of the latin square morphism are the Frid symmetric morphism (Frid, 2001) which essentially produce the Prouhet words (Séébold, 2002).

4.2.2. Tilings. For the ease of discussing images of morphisms, we will discuss the notion of tiling by the image words of the morphism. So considering the morphism from the latin square for $\mathbb{Z}/3\mathbb{Z}$ and the word $W = 012120$, we see that

$$\ell(W) = |012|120|201|120|201|012|,$$

which is tiled by the image words 012, 120, and 201.

For nonuniform morphisms, the notion of tiling offers more challenges due to irregularly placed edges of tiles. Thus the following argument for the overlap-freeness of the latin square morphism would seem to be made much more difficult if applied to nonuniform morphisms.

4.3. Main Theorem for Latin Square Morphisms

Theorem 4.3. *Let ℓ be a latin square morphism. A word $W \in \Sigma_n^*$ is overlap-free if and only if $\ell(W)$ is overlap free. That is, ℓ is an overlap-free morphism.*

Proof. We will prove both sections of the argument by contrapositive. Further we begin with the $\Leftarrow$ portion of the argument due to it being easier.

4.3.1. The $\Leftarrow$ portion of the argument. Assume that the word W has an overlap which we can write as

$$W = AcXcXcB,$$

where A is the initial segment of W, c is a single letter, X is a possibly empty word, and B is the final segment of W. Now we notice that

$$\ell(W) = \ell(AcXcXcB) = \ell(A)\ell(c)\ell(X)\ell(c)\ell(X)\ell(c)\ell(B).$$

Now we set $\ell(c) = dC$ where d is a single letter and C is the final segment of c's image word under ℓ. Then we see that $\ell(W)$ contains the overlap $dC\ell(X)dC\ell(X)d$.

4.3.2. The $\Rightarrow$ portion of the argument. Conversely we assume that for some word $W \in \Sigma_n^*$, $\ell(W)$ contains an overlap. We want to show that W must also contain an overlap. Write $\ell(W) = AcXcXcB$, where c is a single term, X is a finite string with $|cX| \geq n$, A is a finite string, and B is the finite tail of our word $\ell(W)$. Notice that $|cX| \geq n$ (bound by the length of the tiles) because each tile is a permutation of $0, 1, \ldots, n-1$, and we cannot have two of the three copies of c contained in one tile. Our subscripts place this overlap in our sequence. For $i \in \{0, 1, 2\}$, let j_i denote the subscript of the $(i+1)^{\text{th}}$ c. Thus,

$$\begin{aligned} A &= t_0 \cdots t_{j_0-1} \\ c &= t_{j_0} = t_{j_1} = t_{j_2} \\ X &= t_{j_0+1} \cdots t_{j_1-1} = t_{j_1+1} \cdots t_{j_2-1} \\ B &= t_{j_2+1} t_{j_2+2} t_{j_2+3} \cdots t_{j_2+k}, \end{aligned} \tag{13}$$

Our argument proceeds as follows: there are two cases $|cX| \not\equiv 0 \pmod{n}$ and $|cX| \equiv 0 \pmod{n}$. In the first case we use the fact that we have a tiling of $\ell(W)$ by the rows of a Latin square, to show that the overlap $cXcXc$ is not possible. In the second case, when $|cX| \equiv 0 \pmod{n}$, we will see that W must contain an overlap proving the contrapositive.

4.3.3. Case 1: $|cX| \not\equiv 0 \pmod{n}$

For each $i \in \{0,1,2\}$, let $r_i \in \{0,1,\ldots,n-1\}$ such that $r_i \equiv j_i \pmod{n}$. In other words t_{j_i} is the ${r_i}^{\text{th}}$ term in its tile in $\ell(W)$. Also, we will refer to the tile containing t_{j_i} as T_{m_i}. It is now possible to write the length of cX as $|cX| \equiv r_1 - r_0 \equiv r_2 - r_1 \pmod{n}$. So,

$$r_2 \equiv 2r_1 - r_0 \pmod{n}. \tag{14}$$

4.3.3.1. *Six Cases.* Since $r_1 - r_0 \equiv |cX| \not\equiv 0 \pmod{n}$ there are two main cases that we will first consider: $r_0 < r_1$ and $r_1 < r_0$. However, for the explicit details of our conclusions we will consider all six possibilities depending on the value of r_2,

$$r_2 = 2r_1 - r_0 \longleftrightarrow \begin{cases} r_0 < r_1 < r_2 \\ \quad\text{or} \\ r_2 < r_1 < r_0 \end{cases}$$

$$r_2 = 2r_1 - r_0 - n \longleftrightarrow \begin{cases} r_0 \le r_2 < r_1 \\ \quad\text{or} \\ r_2 < r_0 < r_1 \end{cases}$$

$$r_2 = 2r_1 - r_0 + n \longleftrightarrow \begin{cases} r_1 < r_0 \le r_2 \\ \quad\text{or} \\ r_1 < r_2 < r_0 \end{cases}$$

The equalities on the left arise out of equation (14), and the fact that the integer $2r_1 - r_0$ satisfies, $-n \le 2r_1 - r_0 \le 2n$. This means that r_2 is the element in the set $\{2r_1 - r_0 + n, 2r_1 - r_0, 2r_1 - r_0 - n\}$ that lies in the interval $0 < r_2 \le n$. Notice that $r_2 = 2r_1 - r_0$ in both cases when $r_0 < r_1$ and $r_1 < r_0$.

4.3.3.2. *G and the beginning of each cX.* When $r_0 < r_1$, pick $G \subset \Sigma$ to be the set of the last $r_1 - r_0$ letters in T_{m_0} such that G has no specific order and $G \neq \emptyset$. Of course, the remainder of the letters in T_{m_0} are in $\overline{G}$, the complement of G. Notice that this puts $c = t_{j_0} \in \overline{G}$. By equating the letters in T_{m_0} with the corresponding letters in $t_{j_1}\mathbf{x}t_{j_2}$, we find that the last $n - r_1 + 1$ letters of T_{m_1} (starting with $c = t_{j_1}$) are in $\overline{G}$. Also, we find that the first $r_1 - r_0$ letters of T_{m_1+1} are G.

When $r_1 < r_0$, pick $G \subset \Sigma$ to be set of the the last $r_0 - r_1$ letters in T_{m_1} such that G has no specific order and $G \neq \emptyset$. Obviously, the remainder of letters in T_{m_1} must be those that make up $\overline{G}$ again placing $c = t_{j_1} \in \overline{G}$. By equating the letters in T_{m_1} with the corresponding letters in $t_{j_0}Xt_{j_1}$ we find that the last $n - r_0 + 1$ letters of T_{m_0} (starting with $c = t_{j_0}$) are in $\overline{G}$. Also, we find that the first $r_0 - r_1$ letters of T_{m_0+1} are G.

We have discussed the appearance of G and its complement $\overline{G}$ in the beginning of each cX. So, we set forth to describe G and $\overline{G}$ at the end of each cXS.

4.3.3.3. *Following G through the overlap.* It is a basic observation that because each tile is a permutation of the elements in Σ, each tile can be partitioned into G and its complement $\overline{G}$. It is fundamental to our argument that because of the equality $t_{j_0}Xt_{j_1} = cXc = t_{j_1}Xt_{j_2}$, the elements of G form a contiguous collection of elements in each tile involved in our overlap excluding T_{m_i}, either the beginning or the ending of each tile. The idea involved in following G through the overlap is quite simple, we illustrate it in one particular case $r_0 < r_1 < r_2$.

We have explicitly described the location of G at the beginning of each cX. We will now use our example $r_0 < r_1 < r_2$ to show to the reader how the tiling of our sequence can be used to find the location of G at the end of each cX. In doing so, we will refer to Figure 4.1.

In Figure 4.1, we have displaced the overlap from our sequence (represented by the continuous solid horizontal line). We have also split our overlap in half leaving T_{m_1} intact for equality purposes. We have placed $t_{j_0}Xt_{j_1}$ over $t_{j_1}Xt_{j_2}$ with t_{j_0} directly over t_{j_1} and t_{j_1} directly over t_{j_2} so that we can see equality of letters simply by looking straight up or straight down (displayed by vertical arrows). The set of letters G is represented by a black shaded region on our sequence line, and the set of terms $\overline{G}$ is represented by a garnet shaded portion of the sequence line. Also, notice that we have drawn in the edges of the tiles with smaller vertical black lines.

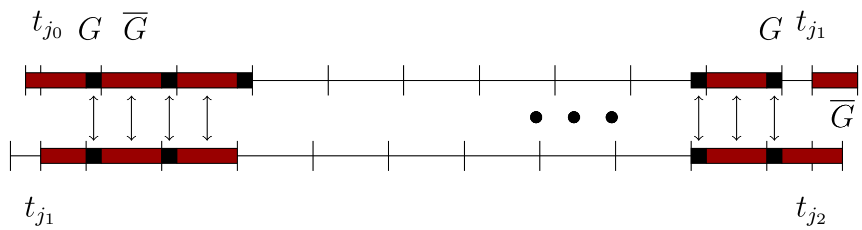


FIGURE 4.1. The situation when $r_0 < r_1 < r_2$.

Now notice that by using the tiles we can equate letters in $t_{j_0}Xt_{j_1}$ with $t_{j_1}Xt_{j_2}$ all the way through the overlap. Since we know that G occurs in the first $r_1 - r_0$ letters of T_{m_1+1}, then $\overline{G}$ is the last $n - (r_1 - r_0)$ letters of T_{m_1+1}. This causes $\overline{G}$ to be the first $n - (r_1 - r_0)$ letters of T_{m_0+1}, and thus G appears in the last $r_1 - r_0$ letters of T_{m_0+1}. Thus we can conclude that G occurs in the last $r_1 - r_0$ letters of all the tiles in $t_{j_0}Xt_{j_1}$ except for T_{m_1}. We can also conclude that G occurs in the first $r_1 - r_0$ letters of all the tiles in $t_{j_1}Xt_{j_2}$ up through T_{m_2-1}. We can approach every case by the same process.

4.3.3.4. *G and how each cX ends.* Now, we will explain the conclusions for the six possible cases that we defined earlier, leaving the actual drawing to the reader.

Case $r_0 < r_1 < r_2$ (as seen in Figure 4.1). After we follow G through the overlap, we find that G occurs in the first $r_1 - r_0$ letters of T_{m_2}. Recall $r_2 = 2r_1 - r_0$. So we have that the next $r_2 - (r_1 - r_0) = r_1$ letters of T_{m_2} are not in G. Notice that the size of G, $r_1 - r_0$, added to r_1 make up all of r_2. This places the boundary between T_{m_1-1} and T_{m_1} exactly in line with the end of G in T_{m_2} and the beginning of $\overline{G}$. We then equate the first letters in T_{m_2} with those in T_{m_1} to find that G occurs nowhere in T_{m_1}. So now, we have described T_{m_1} fully. Earlier we defined G such that $\overline{G}$ occurred from t_{j_1} to the end of the tile, and we have just shown that the first r_1 letters of T_{m_1} (which includes t_{j_1} must be in $\overline{G}$. So G does not appear in anywhere in T_{m_1}, and since $G \neq \emptyset$, we must have a contradiction.

Cases $r_0 \leq r_2 < r_1$ and $r_2 < r_0 < r_1$. After we follow G through the overlap, we find that G occurs in the first $r_1 - r_0$ letters of T_{m_2-1}. So, $\overline{G}$ occurs in the final

$n-(r_1-r_0)$ letters of T_{m_2-1} causing the first $n-(r_1-r_0)$ letters of T_{m_1} to be $\overline{G}$. Notice that $r_1 = [n-(r_1-r_0)] + r_2$. So the boundary between $\overline{G}$ and G in T_{m_1} coincides with the boundary between T_{m_2-1} and T_{m_2}. This means that $t_{j_1} \in G$, but we assumed that $c \notin G$ earlier which is a contradiction.

Case $r_2 < r_1 < r_0$. After we follow G through the overlap, we find that G occurs in the last $r_0 - r_1$ letters of T_{m_2-1}. This causes G to occur in the first $r_0 - r_1$ letters of T_{m_1} by equality of $t_{j_0}Xt_{j_1}$ and $t_{j_1}Xt_{j_2}$. To describe the remaining letters of T_{m_1} up to and including t_{j_1} consider $r_1 - (r_0 - r_1) = r_2$. So $\overline{G}$ occurs in the next r_2 letters after G. Thus we have that G is repeated twice in T_{m_1} so we have our contradiction.

Cases $r_1 < r_0 \leq r_2$ and $r_1 < r_2 < r_0$. After we follow G through the overlap we find that G occurs in the first $r_0 - r_1$ letters of T_{m_1-1}. This causes $\overline{G}$ to occur in the final $n-(r_0-r_1)$ letters of T_{m_1-1} and thus the first $n-(r_0-r_1)$ letters of T_{m_2}. Since $r_1 = r_2 - [n-(r_0-r_1)]$, we see that the left boundary of T_{m_1} coincides with the right boundary of these first $n-(r_0-r_1)$ letters of T_{m_2}. In particular, this means that the last $r_0 - r_1$ letters of T_{m_2}, which include c, are in G. But, this contradicts the fact that $c \notin G$.

4.3.4. Case 2: $|cX| \equiv 0 \pmod{n}$

We begin by considering some $\pi \in S_n$ the symmetric group on n terms. Note that we can apply π to any string by requiring π to act on each individual term, so $\pi(t_1t_2\dots t_s) = \pi(t_1)\pi(t_2)\dots\pi(t_s)$. Thus π can be treated as a morphism. Moreover, $\pi : \Sigma^* \to \Sigma^*$ is an invertible map because $\pi \in S_n$. Thus $w \in \Sigma^*$ contains an overlap if and only if $\pi(w) \in \Sigma^*$ contains an overlap.

Define the function $d_{(a,n)} : \mathbb{N} \to \mathbb{N}$ by $d_{(a,n)}(m) = (m-1)n + a$. Now if we let $M = (t_s)$ be a sequence, then define the sequence given by the function $D_{(a,n)}(M)$ to be the subsequence $(t_{d_{(a,n)}(s)})$ of M. So for $i \in \{1, 2, \dots, n\}$ arbitrary we have that

$$D_{(i,n)}(t_1t_2\dots t_s) = t_it_{i+n}t_{i+2n}\dots.$$

Define $\pi_i : \Sigma \to \Sigma$ with $\pi_i \in S_n$, such that if $\mathcal{L}_{t_1} = \{t_1, t_2, \ldots, t_i, \ldots, t_n\}$, $\pi_i(t_1) = t_i$. Recall that $\mathcal{L}_t$ refers to the t^{th} row of our Latin square $\mathcal{L}$. So we have that π_i maps each term in the first column of our Latin square, to the i^{th} entry of its corresponding row. Now, we want to show that $\pi_i(U) = D_{(i,n)}(\ell(U))$ for all $U \in \Sigma^*$. So, let $U = t_0 \ldots t_{m-1}$, and take

$$\begin{aligned} D_{(i,n)}(\ell(U)) &= D_{(i,n)}(\ell(t_0)\ell(t_1)\ell(t_2)\cdots t_{m-1}) \\ &= \pi_i(t_0)\pi_i(t_1)\pi_i(t_2)\cdots\pi_i(t_{m-1}) \\ &= \pi_i(\ell(U)). \end{aligned}$$

Since $\pi_i \in S_n$ is invertible we can conclude that $D_{(i,n)}(\ell(W))$ contains an overlap if and only if $\ell(W)$ contains an overlap.

Since $|cX| \equiv 0 \pmod{n}$ pick $i \equiv j_0 \equiv j_1 \equiv j_2 \pmod{n}$, by applying $D_{(i,n)}$ to (14) we obtain

$$D_{(i,n)}(\ell(W)) = A_i\ t_{j_0}\ X_i\ t_{j_1}\ X_i\ t_{j_2}\ B_i$$

where

$$\begin{aligned} A_i &= D_{(i,n)}(A) = t_i t_{i+n} t_{i+2n} \ldots, \\ X_i &= D_{(i,n)}(X) = t_{j_0+n} t_{j_1+2n} \ldots t_{j_0+(m-1)n} \\ &= t_{j_1+n} t_{j_1+2n} \ldots t_{j_1+(m-1)n}, \\ B_i &= D_{(i,n)}(B) = t_{j_2+n} t_{j_2+2n} t_{j_2+3n} \ldots. \end{aligned}$$

Observe that $D_{(i,n)}(\ell(W))$ contains an overlap. Thus we see that W must also contain an overlap because W is a permutation of $D_{(i,n)}(\ell(W))$. So we are done. $\square$

4.4. Prouhet Words and the Frid Symmetric Morphism

We begin by noting that both the Prouhet words and the Frid symmetric morphism are instances of the Latin square morphism.

4.4.1. Prouhet Words. As seen in Séébold's paper on overlap-free morphisms (Séébold, 2002) and originally in (Prouhet, 1851), we begin by taking a circle and putting around it the letters $0, 1, \ldots, n-1$ in this order To obtain the Prouhet word

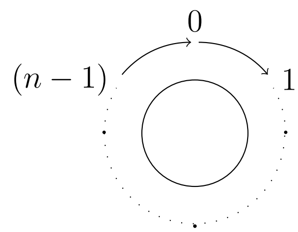


FIGURE 4.2. The Prouhet word setup for n letters.

$\mathbf{P}_n$, start from the letter 0 and turn all around the circle, always in the same direction, in writing the letters as and when encountered. At each round (when n letters have been written) skip one letter; all n rounds skip one more letter; all n^2 rounds skip one more letter; and so on infinitely many times.

For example the Prouhet word on four letters $\mathbf{P}_4$ is produced by taking the following circle

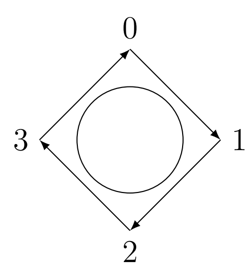


FIGURE 4.3. The Prouhet word setup for 4 letters.

During the first round of the circle we write 0123 and then skip 0. Next round we have skipped zero so we start with 1 and write 1230. So thus far the word is

$$0123\ 1230.$$

For the third round we skip 1 and begin with 2 writing 2301, and the fourth round we skip 2 and begin with 3 writing 3012. So now the beginning of the word is

$$0123\ 1230\ 2301\ 3012.$$

After proceeding with this infinitely we see that the beginning of $\mathbf{P}_4$ is

$$\mathbf{P}_4 = 0123\ 1230\ 2301\ 3012\ 1230\ 2301\ 3012\ 0123\ 2301\ 3012\ldots$$

It is interesting to note that the Frid symmetric morphism $\phi_n(t)$ for $n = 4$, produces this exact same word after countably infinite iterations. Further we notice that $\phi_n(t)$ is a Latin square morphism.

4.4.2. The Frid Symmetric Morphism.

Definition 4.4. Let $i \in \Sigma_n$, the Frid symmetric morphism is defined by

$$\phi_n(i) = \overline{i+0}\,\overline{i+1}\ldots\overline{i+(n-1)},$$

where $\bar{i}$ is the residue modulo n (Frid, 2001).

Notice that $\phi_2(i)$ is the standard definition of the Thue-Morse morphism, and further consider $\phi_4(i)$, which is given by

$$\phi_4(i) = \begin{cases} 0123, & \text{for } i = 0 \\ 1230, & \text{for } i = 1 \\ 2301, & \text{for } i = 2 \\ 3012, & \text{for } i = 3. \end{cases}$$

It is clear that

$$\phi_4^{\omega}(0) = \mathbf{P}_4.$$

Further it is clear that the Frid symmetric morphism is a Latin square morphism. Thus all of the Prouhet words can be derived by infinite iteration of Latin square morphisms. Namely those Latin square morphisms using the addition table for $\mathbb{Z}/n\mathbb{Z}$.

Chapter 5

The Pooh Morphism

In a vain attempt to classify all of the overlap-free morphisms using the latin square morphism (Tompkins, 2007), we stumbled across the Leech square-free morphism in (Allouche and Shallit, 2003). The following is the Leech square-free morphism

$$
\begin{aligned}
& \quad h \\
0 &\mapsto 0121021201210 \\
1 &\mapsto 1202102012021 \\
2 &\mapsto 2010210120102,
\end{aligned}
$$

which originally appeared in (Leech, 1957). Noticing that this morphism was overlap-free put a hole in our attempt to classify all of the overlap-free morphisms using Latin square morphisms. But on the other hand, we now could potentially find another class of overlap-free morphisms that could be explained in a better manner than with test-sets as in (Richomme and Wlazinski, 2004).

Using the test-set result in Chapter 3, we found the following overlap-free morphisms on four letters

$$
\begin{aligned}
& \quad f \\
0 &\mapsto 0123123010321321 0 \\
1 &\mapsto 12302301210320321 \\
2 &\mapsto 23013012321031032 \\
3 &\mapsto 30120123032102103,
\end{aligned}
$$

and

$$\begin{array}{rcl} & g & \\ 0 & \mapsto & 012301221211203210 \\ 1 & \mapsto & 123013003033010321 \\ 2 & \mapsto & 230120123310221032 \\ 3 & \mapsto & 301230110100132103. \end{array}$$

The morphism g raised a considerable number of questions as to why it was overlap-free. It seemed to avoid a considerable number of the techniques used in the proof For the Latin square morphisms. So the natural question was: what does the morphism g have in common with the Leech square-free morphism that causes its overlap-freeness.

5.1. The Pooh Overlap-Free Morphism

The overlap-free morphisms displayed are tied together with the following definition.

Definition 5.1. Let $h : \Sigma^* \to \Delta^*$ be an n-uniform morphism. We call h a Pooh overlap-free morphism if it satisfies the following properties:

(i) $h(W)$ is overlap-free for all overlap-free words $W \in \Sigma^*$ with $|W| = 3$.

(ii) For $a, b \in \Sigma$, and for all $V \in \Sigma^*$ such that $|V| \leq \lfloor n/2 \rfloor$,

$$h(a) = SV \quad \text{and} \quad h(b) = VU$$

if and only if S is not a suffix of any image word of h and U is not a prefix of any image word of h.

We now prove a lemma that captures the combinatorial properties in the first portion of Definition 5.1.

Lemma 5.2. *Let Σ be an alphabet with more than one letter. Let $h : \Sigma^* \to \Delta^*$ be a morphism such that $h(W)$ is overlap-free for all overlap-free $W \in \Sigma^*$ with $|W| = 3$. We then have the following properties:*

(i) *$h(a)$ is overlap-free for all $a \in \Sigma$.*

(ii) *$h(a)h(b)$ is overlap-free for all $a, b \in \Sigma$.*

(iii) *$h(a)$ and $h(b)$ do not begin or end with the same letter, whenever $a, b \in \Sigma$ and $a \neq b$.*

Proof. (i) Let us first state that the result does not apply when $|\Sigma| = 1$ because there are no overlap-free word of length three for this alphabet. For $|\Sigma| > 1$ this result is clear because if we assume for a contradiction that $h(a)$ contained an overlap for any $a \in \Sigma$, then $h(bab)$, with $b \neq a$, would contain an overlap which contradicts our assumption.

(ii) Similar to (i), if we assume that $h(a)h(b) = h(ab)$ contained an overlap, then $h(aba)$ would contain an overlap. Again this contradicts our assumption. We also must show that $h(aa)$ does not contain an overlap. Assume for a contradiction that it does, and we quickly obtain our contradiction by observing that then $h(aab)$ must contain an overlap.

(iii) Assume that for some $a, b \in \Sigma$, $h(a)$ and $h(b)$ begin with the same letter. Then, $h(aab)$ would contain an overlap, which contradicts our assumption. The argument for $h(b)$ and $h(b)$ ending with different letters is similar. $\square$

Theorem 5.3. *All Pooh overlap-free morphisms are overlap-free.*

Proof. We begin by assuming that h is a Pooh morphism with $|h(a)| = n$ for all $a \in \Sigma$. We must show that for all $W \in \Sigma^*$, W is overlap-free if and only if $h(W)$ is overlap-free. We will begin with the easy direction first.

5.1.1. The $\Leftarrow$ direction. Assume that $W = AcXcXcB$, so that we can argue by contrapositive that $h(W)$ must also contain an overlap. Notice that

$$h(W) = h(A)h(c)h(X)h(c)h(X)h(c)h(B).$$

Set $h(c) = dY$ where $d \in \Sigma$ and $Y \in \Sigma^*$, then $h(W) = h(A)dYh(X)dYh(x)dYh(B)$. So then $h(W)$ contains the overlap $dYh(X)dYh(X)d$, and we are done with the first portion of our argument.

5.1.2. The $\Rightarrow$ direction. Conversely we will argue by contrapositive. We will assume that $h(W)$ contains an overlap and show that W must also contain an overlap. So assume that for some $W \in \Sigma^*$ we have

$$h(W) = Ac_{j_0}Xc_{j_1}Xc_{j_2}B,$$

where $c = c_{j_0} = c_{j_1} = c_{j_2}$. We use the 0, 1 and 2 to denote which c we will refer to. Further, the index j_i will refer to which letter in the word $h(W)$ we are referring to, noting that we are indexing beginning with 0.

We will proceed with two separate arguments. The first argument will be that it is not possible to write $h(W)$ with $|cX| \not\equiv 0 \pmod{n}$. The second argument will be that W must contain an overlap if $|cX| \equiv 0 \pmod{n}$.

5.1.2.1. *The* $|cX| \not\equiv 0 \pmod{n}$ *case.* Notice that we must have the overlap in $h(W)$ contained in $h(Z)$ where $|Z| > 3$ is some subword of W. Otherwise we would be breaking hypothesis (i) in the definition of Pooh morphisms.

We begin by setting

$$r_i \equiv j_i \pmod{n},$$

where $i \in \{0, 1, 2\}$ and $r_i \in \{0, 1, \ldots, n-1\}$. We will argue first based on the number of tiles that the overlap occurs in, and then by cases. When the overlap occurs over four tiles (noting that occurring over three tiles contradicts the hypothesis), we will observe four cases. The cases are

$$r_0 \leq r_2 < r_1,$$
$$r_2 < r_0 < r_1,$$
$$r_1 < r_0 \leq r_2,$$
$$r_1 < r_2 < r_0.$$

We note that the cases $r_0 < r_1 < r_2$ and $r_2 < r_1 < r_0$ force the overlap to occur in a number other than four tiles. When the overlap occurs in more than four tiles we will more simply consider the two cases $r_0 < r_1$ and $r_1 < r_0$. Finally, we note the following relationship between r_0, r_1, and r_2.

$$r_2 \equiv 2r_1 - r_0 \pmod{n}. \tag{15}$$

Consider the notion of the tiling of a line segment. We will use this notion of tiling in application to working with $h(W)$. The tiles we speak of are the image words of h. Note that all the image words must be of the same length n, this is crucial to our argument. For ease we will use T_{s_i} with $i \in \{0, 1, 2\}$ to denote the tile containing c_{j_i}. Note that s_i is the number of the tile if we numbered them starting with the first tile as T_0.

The overlap is contained in 4 tiles. Let us consider the case where there is some subword of W, say Z, with $|Z| = 4$ and the overlap in $h(W)$ is contained in $h(Z)$. As in the argument for a Latin square morphism to be overlap-free we will consider the word $h(Z)$ to be a line. We will draw in small vertical lines to signify the edges of the tiles, and we will draw taller labeled vertical lines to signify the c's in the overlap.

In Figure 5.1, we have taken $c_{j_0}Xc_{j_1}Xc_{j_2}$ and written it twice aligning $c_{j_0}Xc_{j_1}$ in the upper line with $c_{j_1}Xc_{j_2}$ in the lower line for the purpose of equating the terms through the overlap. Figure 5.1 displays the case when $r_2 < r_0 < r_1$. We remark here that the case when $r_2 = r_0 < r_1$ proceeds in the same manner.

Let V to be the final $r_1 - r_0$ letters in the tile T_{s_0}, as we have drawn in Figure 5.1. Similarly we choose U to be the first $r_1 - r_2$ letters in T_{s_1}. Now equation (15) gives

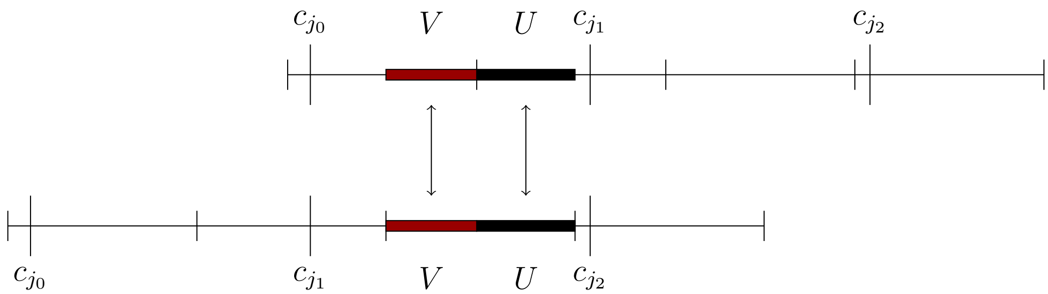


FIGURE 5.1. The short overlap with $r_2 < r_0 < r_1$

that in the $r_2 < r_0 < r_1$ situation we have that $n - (r_1 - r_0) = r_1 - r_2$. Clearly then we must have $|U| = r_1 - r_2 \leq \lfloor n/2 \rfloor$ or $|V| = r_1 - r_0 \leq \lfloor n/2 \rfloor$. In the case when $|V| \leq \lfloor n/2 \rfloor$ we cannot equate U with any prefix of a tile which leads to a contradiction. In the other case when $|U| \leq \lfloor n/2 \rfloor$ we cannot equate V with any suffix of a tile which leads to a contradiction. So this case is not possible.

We now consider the case where $r_0 < r_2 < r_1$ as shown in Figure 5.2.

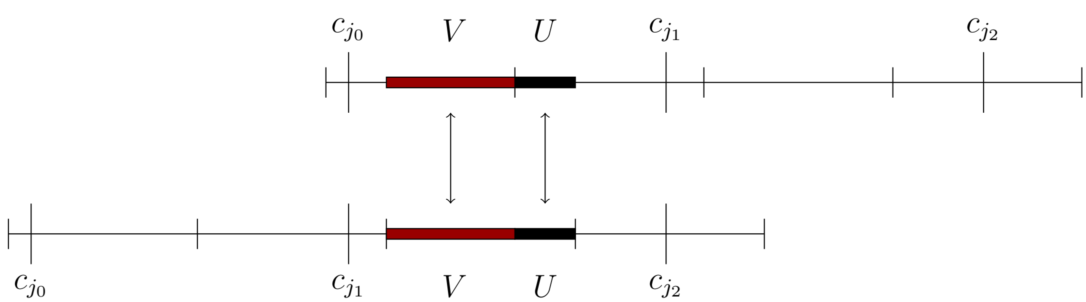


FIGURE 5.2. The short overlap with $r_0 < r_2 < r_1$

In this case we choose V to be the final $r_1 - r_0$ letters in T_{s_0}, and we also pick U to be the first $r_1 - r_2$ letters in T_{s_1} as drawn in Figure 5.2. Again we notice that $n - (r_1 - r_0) = r_1 - r_2$ so either $V \leq \lfloor n/2 \rfloor$ or $|U| \leq \lfloor n/2 \rfloor$, either of which is impossible. So we cannot have this case occurring either.

We now consider the cases with $r_1 < r_2 \leq r_0$ and $r_1 < r_0 < r_2$. Figure 5.3 gives the situation when $r_1 < r_2 < r_0$ (note that the case when $r_1 < r_2 = r_0$ is similar, and the same applies to the arguments above). Notice that in both of these cases we have that $n - (r_2 - r_1) = r_0 - r_1$.

For the case depicted in Figure 5.3, $r_1 < r_2 < r_0$, we assume that V is the final $r_0 - r_1$ letters in T_{s_1}, and we also assume that U is the first $r_2 - r_1$ letters in T_{s_2}. Now either $|U| \leq \lfloor n/2 \rfloor$ or $|V| \leq \lfloor n/2 \rfloor$. In either case we have a contradiction.

Now we consider the case where $r_1 < r_0 < r_2$, which is displayed in Figure 5.4.

In the case displayed here in Figure 5.4, we again assume that V occurs in the final $r_0 - r_1$ letters of T_{s_1}, and we also assume that U occurs in the first $r_2 - r_1$ letters of T_{s_2}. We then have that either $|V| \leq \lfloor n/2 \rfloor$ or that $|U| \leq \lfloor n/2 \rfloor$. Either case is a contradiction. So we cannot have our overlap occurring in four tiles. Thus, we consider the case when the overlap occurs in more than four tiles.

We also not that if we are in the case when the overlap occurs in five tiles, the same arguments hold.

The overlap is contained in more than four tiles. We will look at the cases with $r_0 < r_1$ and $r_1 < r_0$, and we will only look at the beginning of the overlap. So we consider Figure 5.5 for the case when $r_0 < r_1$.

We will consider the case with $r_1 - r_0 \geq \lfloor n/2 \rfloor$, as we will cover the logic behind the argument for $r_1 - r_0 \leq \lfloor n/2 \rfloor$ in Figure 5.6.

Let V be the final $r_1 - r_0$ letters in T_{s_0}, then equating yields V as the beginning $r_0 - r_1$ letters of T_{s_1+1}. Since $|V| \geq \lfloor n/2 \rfloor$ we can equate the suffix of T_{s_1+1}, call it U (which is labeled with a dotted line in Figure 5.5), with T_{s_0+1}. So we have $T_{s_1+1} = VU$. Similarly we can set $S \in \Delta^*$ such that $T_{s_0+1} = US$. Notice now that

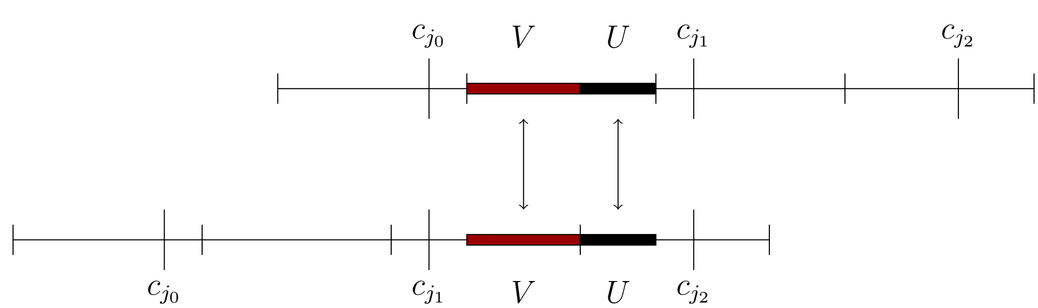


FIGURE 5.3. The short overlap with $r_1 < r_2 < r_0$

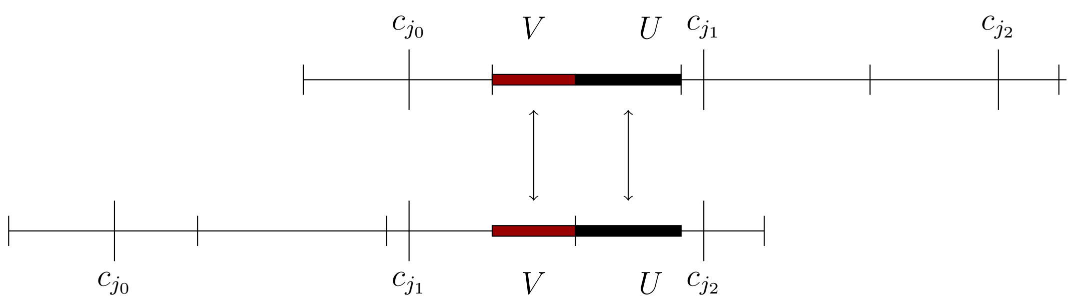


FIGURE 5.4. The short overlap with $r_1 < r_0 < r_2$

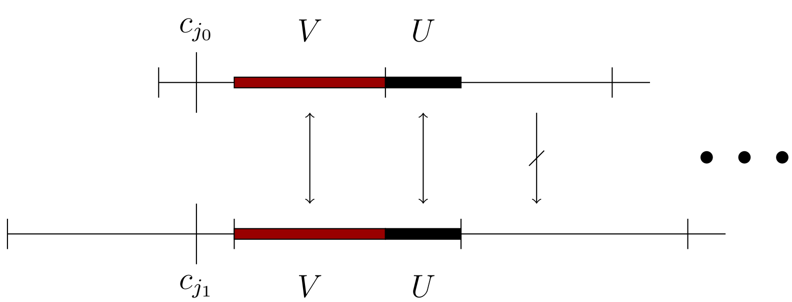


FIGURE 5.5. The long overlap with $r_0 < r_1$ and $r_1 - r_0 \geq \lfloor n/2 \rfloor$

$|U| = n - (r_1 - r_0) \leq \lfloor n/2 \rfloor$. Thus S cannot begin any image word of h so the overlap is impossible.

A note for the case when $r_1 - r_0 \leq \lfloor n/2 \rfloor$. In this case $|V| \leq \lfloor n/2 \rfloor$ and we would not be able to equate U.

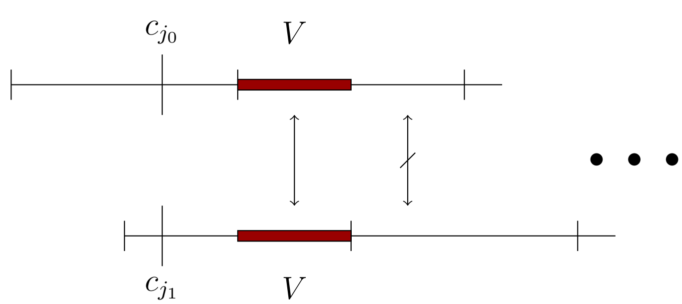


FIGURE 5.6. The long overlap with $r_1 < r_0$ and $r_0 - r_1 \leq \lfloor n/2 \rfloor$

Figure 5.6 gives the case when $r_1 < r_0$ with $r_0 - r_1 \leq \lfloor n/2 \rfloor$. In a similar manner to the case where $r_0 < r_1$ we pick V to be the suffix of $r_0 - r_1$ letters in T_{s_1}. Now we can possibly find an image word $T_{s_0+1} = VU$ for some $U \in \Delta^*$. But because

$|V| = r_0 - r_1 \leq \lfloor n/2 \rfloor$, U cannot be the prefix of any image word, so this formulation of the overlap is impossible.

So we see that in order for $h(W)$ to contain an overlap, it must be one so that $|cX| \equiv 0 \pmod{n}$.

5.1.2.2. *The* $|cX| \equiv 0 \pmod{n}$ *case.* From Lemma 5.2 we know that the beginning letters and ending letter for each image word in h must be distinct. Further we know that the suffix of T_{s_1} must be identical to the suffix of T_{s_0} as $r_0 = r_1$. This implies that $T_{s_0} = T_{s_1}$. Similarly $T_{s_1} = T_{s_2}$.

Pick d to be the letter such that $h(d) = T_{s_0} = T_{s_1} = T_{s_2} = T$. Further because

$$h(W) = Ac_{j_0}Xc_{j_1}Xc_{j_2}XB,$$

we can find subwords C, D, Y of W so that

$$h(W) = h(CdYdYdD) = AcXcXcB.$$

Now we must have that $W = CdYdYdD$ which contains an overlap. Thus we are done.

□

5.2. The Pooh Square-Free Morphism

Similarly to the pooh overlap-free morphism we can define the pooh square-free morphism in the following manner.

Definition 5.4. Let $h : \Sigma^* \to \Delta^*$ be an n-uniform morphism. We call h a Pooh square-free morphism if it satisfies the following properties:

(i) $h(W)$ is square-free for all square-free words $W \in \Sigma^*$ with $|W| = 3$

(ii) $h(a)$ and $h(b)$ do not begin or end with the same letter for all $a, b \in \Sigma$ with $a \neq b$.

(iii) For $a, b \in \Sigma$, and for all $V \in \Sigma^*$ such that $|V| \leq \lfloor n/2 \rfloor$,

$$h(a) = SV \quad \text{and} \quad h(b) = VU$$

if and only if S is not a suffix of any image word of h and U is not a prefix of any image word of h.

Because we cannot consider words like aab to put into h, we must add property (ii) in 5.4 so that we can use a similar preimage argument in the final portion of the argument. Thus the Pooh square-free theorem follows.

THEOREM 5.5. *All Pooh square-free morphisms are square-free.*

PROOF. Assume that h is a Pooh square-free morphism such that $|h(a)| = n$ for all $a \in \Sigma$. We must show that for some $W \in \Sigma^*$, W is square-free if and only if $h(W)$ is square-free. We will begin with the easy direction.

5.2.1. The $\Leftarrow$ direction. We will proceed by contrapositive. So assume that $W = AXXB$, where $X \in \Sigma^+$ and $A, B \in \Sigma^*$. Write

$$h(W) = h(AXXB) = h(A)h(X)h(X)h(B),$$

which contains the square $h(X)h(X)$. So we are done with this direction.

5.2.2. The $\Rightarrow$ direction. Again we proceed by arguing the contrapositive. So we assume that

$$h(W) = Ac_{j_0}Xd_{i_0}c_{j_1}Xd_{i_1}B, \tag{16}$$

where $c = c_{j_0} = c_{j_1} \in \Sigma$, $d = d_{i_0} = d_{i_1} \in \Sigma$ and $A, X, B \in \Sigma^*$. Note that we are using c_{j_0} and c_{j_1} so that we can mark the beginning of the square, and similarly for the d's and the end of the square.

There are two cases to consider here $|cXd| \not\equiv 0 \pmod{n}$ and $|cXd| \equiv 0 \pmod{n}$. We show that it is impossible for $|cXd| \not\equiv 0 \pmod{n}$ in an analogous manner as in Theorem 5.3, as seen in section 5.1.2.1.

So assume that $|cXd| \equiv 0 \pmod{n}$. From the definition of the pooh square-free morphism we know that each image word for h must begin and end with distinct letters. Further we know the suffix of the tile containing c_{j_0} must be identical to the suffix of the tile containing c_{j_1}. Thus they are the same image word, call it $h(z)$ for some $z \in \Sigma$. So because

$$h(W) = AcXcXB,$$

we can find subwords C, D, Y of W such that

$$h(W) = h(CzYzYD) = AcXcXB.$$

Thus we have that $W = AzXzXB$ which contains a square, and we are done. □

5.3. Examples of the Pooh Morphism

Theorem 5.3 gives us that any Pooh overlap-free morphism is overlap-free, but it remains to be checked if in fact the Leech square-free morphism lies in the class of Pooh overlap-free morphisms. Again we see that the Leech square-free morphism is

$$\begin{array}{ccl} & h & \\ 0 & \mapsto & 0121021201210 \\ 1 & \mapsto & 1202102012021 \\ 2 & \mapsto & 2010210120102. \end{array}$$

Notice that each row is a permutation of the first row, so we need only check the first image word to see if it satisfies the properties of Definition 5.1.

It is easy to check that $h(W)$ is overlap-free for all words $W \in \Sigma_3^*$ of length 3 which is part (i) of the definition. Part (ii) of the definition is to check the suffix/prefix condition on the image words. So we check

$$\begin{array}{l} 01210212\ 0121\ 0 \\ \qquad\qquad\quad 0121\ 0\ 21201210 \\ \qquad\qquad\qquad\quad 0\ 121021201210 \end{array}$$

to see if we can equate the end of the second or third line with any other image word. What we have done here is set the first line $h(0) = UV$ where $V = 01210$ in the first case and $V = 0$ in the second. In either case we see that $|V| \leq \lfloor 13/2 \rfloor = 6$. Further the suffix on the second line begins 212 which cannot begin any image word of h. The suffix on the third line begins 121 which similarly cannot begin any image word. We also notice that these are the only possibilities for finding a V that is the suffix of $h(0)$ so that V is also the prefix for any image word of h.

Another cited example at the beginning of the chapter is the morphism

$$\begin{array}{ccc} & f & \\ 0 & \mapsto & 012312301032132 10 \\ 1 & \mapsto & 1230230121032032 1 \\ 2 & \mapsto & 2301301232103103 2 \\ 3 & \mapsto & 3012012303210210 3. \end{array}$$

Again we notice that we only need consider $f(0)$ for our analysis. Again we can easily check to see that f is overlap-free on the words of length three over Σ_4. Now the only way that we can equate the prefix of an image word to the suffix of $f(0)$ by equating the 0 at the end of $f(0)$ to the 0 at the beginning of $f(0)$. And the remainder of the second $f(0)$ begins 1231, which does not start of any other image words of f. So f is a Pooh overlap-free morphism.

We note that the final example in the chapter the morphism g from earlier can also (and in the same fashion) be easily checked to see if it satisfies the properties of Definition 5.1. Thus g is yet another example of an overlap-free morphism.

Bibliography

Allouche, Jean-Paul and Jeffrey Shallit. 1999. *The ubiquitous Prouhet-Thue-Morse sequence*, Sequences and their applications (Singapore, 1998), Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London, pp. 1–16. MR1843077 (2002e:11025)

______. 2003. *Automatic sequences*, Cambridge University Press, Cambridge. Theory, applications, generalizations. MR1997038 (2004k:11028)

Berstel, J. and P. Séébold. 1993. *A characterization of overlap-free morphisms*, Discrete Appl. Math. **46**, no. 3, 275–281 (English, with English and French summaries). MR1243728 (95d:20097)

Crochemore, Max. 1982. *Sharp characterizations of squarefree morphisms*, Theoret. Comput. Sci. **18**, no. 2, 221–226. MR652471 (84a:20070)

Ehrenfeucht, A. and G. Rozenberg. 1982. *On finite sets testing square free property fo all homomorphisms between two given alphabets*, Department of Computer Science, University of Colorado at Boulder, Technical Report No. 230.

Frid, Anna E. 2001. *Overlap-free symmetric* D0L *words*, Discrete Math. Theor. Comput. Sci. **4**, no. 2, 357–362 (electronic). MR1877528 (2002j:68078)

Leech, J. 1957. *A problem on strings of beads*, Math. Gazette **41**, 277-278.

Lothaire, M. 2002. *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge. MR1905123 (2003i:68115)

Prouhet, M. E. 1851. *Mémoire sur quelques relations entre les puissances des nombres*, C. R. Acad. Sci. Paris **33**, 225.

Richomme, G. and F. Wlazinski. 2004. *Overlap-free morphisms and finite test-sets*, Discrete Appl. Math. **143**, no. 1-3, 92–109. MR2087872 (2005f:68122)

Séébold, Patrice. 1985. *Sequences generated by infinitely iterated morphisms*, Discrete Appl. Math. **11**, no. 3, 255–264. MR792893 (86m:20070)

———. 2002. *About some overlap-free morphisms on a n-letter alphabet*, J. Autom. Lang. Comb. **7**, no. 4, 579–597. Descriptional complexity of automata, grammars and related structures (Vienna, 2001). MR1990460 (2004f:68125)

Thue, Axel. 1906. *Über unendliche Zeichenreihen*, Christiania Vidensk. Selsk. Skr. **7**, 3–22. Jbuch 37, 66, Reprinted in (Thue, 1977).

———. 1912. *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*, Norske vid. Selsk. Skr. Mat. Nat. Kl. **1**, 1–67. Reprinted in (Thue, 1977).

———. 1977. *Selected mathematical papers*, Universitetsforlaget, Oslo. With an introduction by Carl Ludwig Siegel and a biography by Viggo Brun; Edited by Trygve Nagell, Atle Selberg, Sigmund Selberg, and Knut Thalberg. MR0460050 (57 #46)

Tompkins, C. Robinson. 2007. *Latin square Thue-Morse sequences are overlap-free*, Discrete Math. Theor. Comput. Sci. **9**, no. 1, 239–246 (electronic). MR2350266 (2008h:68099)