2015

Avoiding Doubled Words in Strings of Symbols

Michael Lane
University of South Carolina

Avoiding Doubled Words in Strings of Symbols

by

Michael Lane

Bachelor of Arts
Charleston Southern University 2009

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Mathematics

College of Arts and Sciences

University of South Carolina

2015

Accepted by:

George F. McNulty, Major Professor

Ognian Trifonov, Committee Member

Jerrold Griggs, Committee Member

Michael Filaseta, Committee Member

Stephen Fenner, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

Dedication

This dissertation is dedicated to Amy, Kyle, Skylar, and all my future descendants. This dissertation is the culmination of 11 years of college so that I may give you the life you always dreamed.

Acknowledgments

I would first like to thank my advisor, George McNulty, for his countless hours of mentoring and training in this discipline. If it were not for the many meetings, discussions, and pushes to keep working, this dissertation would not have been possible. In addition, his dedication to the algebra and logic seminar gave me a great breadth of additional knowledge in my discipline. I would also like to thank Ognian Trifonov, the graduate director during my last five years at the University of South Carolina, for his support of my needs as a graduate student, both in my understanding of the whole process to graduation and also in my financial needs. If it were not for him ensuring that I always had enough work to have a reasonable stipend, I would have had far more difficulty finishing my studies. I would also like to thank him and Michael Filaseta for building my knowledge in the areas of number theory. I would further like to thank Michael Filaseta for his work towards ensuring that I was able to teach the math education classes in which I was most interested. Next, I would like to thank Jerrold Griggs for teaching me discrete math and pushing me to learn it in depth rather than counting on my own problem solving abilities. With a dissertation classified in combinatorics, I cannot appreciate further my gaining the deeper understanding in this field. I would like to thank Steve Fenner for his enlightening connections of my field to that of computer science. His presentations in the algebra and logic seminar built my horizons to better understand how logic and computation fit together. Finally, I would like to thank my committee as a whole for their efforts to analyze my dissertation and provide valuable feedback as to how to make it the best it can be.

Next, I would like to thank my friends and family for their unwavering support through the whole process. In particular, I would like to thank all of my classmates in my first year of graduate school that kept me going and worked with me to ensure that I survived the qualifying exams in algebra and analysis. I would also like to particularly thank Travis Johnston for his support and the many discussions to build my knowledge of LaTeXand C++. Without him, many of the algorithms I have used would not have been possible. I would also like to particularly thank Andrew Dove, Danny Rorabaugh, and Heather Smith for being awesome friends and support through the whole process of being in graduate school. I hope to know all of you for a lifetime. I would also like to thank Anton Strizhov for his invaluable assistance in translating the Russian in Mel'nichuk's result.

Finally, I would like to thank my mother, father, and brother for their support and encouragement to keep moving throughout my college career and to never give up. And last but not least, I want to thank my amazing wife Amy Lane for her hours of listening to my excitement and to my concerns as I dug deep into my research. Her encouragement and pushing to keep working made this dissertation possible. I cannot thank her enough for all the support in my final stretch to graduate.

Abstract

A *word* on the n-letter alphabet is a finite length string of symbols formed from a set of n letters. A word is *doubled* if every letter that appears in the word appears at least twice. A word w *avoids* a word u if there is no non-erasing homomorphism h (a map that respects concatenation) such that $h(u)$ is a subword of w. Finally, a word w is n-avoidable if there is an infinite list of words on the n-letter alphabet that avoid w. In 1906, Thue showed that the simplest doubled word, namely xx, is 3-avoidable. In 1984, Dalalyan showed that each doubled word is 4-avoidable and that each doubled word on 6 or more letters is 3-avoidable. In 2013, Blanchet-Sadri and Woodhouse, building on the work of Bell and Goh that is similar to Dalalyan, strengthened the result by showing that all doubled words of length at least 12 are 3-avoidable. Cassaigne in his dissertation classified all the words on the 2-letter alphabet and most of the words on the 3-letter alphabet, and as a result, showed that each doubled word in which at most most 3 distinct letters appear is 3-avoidable. These results leave 7441 doubled words in which exactly 4 or 5 distinct letters appear to check for 3-avoidability. In this dissertation, we show that each doubled word in which at least one letter occurs 3 or more times is 3-avoidable. This leaves only the doubled words to check for 3-avoidability in which exactly 4 or 5 letters appear in the word and each letter appears exactly twice. In fact, we give a list of just 99 doubled words so that if each word on the list can be shown to be 3-avoidable, then we would know that each doubled word is 3-avoidable.

Table of Contents

Dedication iii

Acknowledgments iv

Abstract vi

List of Figures ix

Chapter 1 Introduction 1

Chapter 2 Avoiding Doubled Words Individually 5

2.1 Dalalyan's Results 5

2.2 An Improvement on Dalalyan's Results 10

2.3 Bell, Goh, Rampersad, Blanchet-Sadri, and Woodhouse's Results . . 21

2.4 A Second Proof of Theorem 2.2.1 25

2.5 Classification of All Binary and Ternary Words 27

2.6 The 3-Avoidability of Doubled Words 28

2.7 The 2-Avoidability of Tripled Words 30

Chapter 3 Avoiding Doubled Words Simultaneously 31

3.1 Previous Results 31

3.2 Mel'nichuk's Results 32

Chapter 4 Conclusion 65

Bibliography 66

List of Figures

Figure 1.1 Example of a Homomorphism 2

Figure 2.1 Graph for 3-Avoidability of Doubled Words 15

Figure 2.2 Graph for 2-Avoidability of Tripled Words 16

Figure 2.3 Graph for Doubled Words on an Alphabet of Size 2 18

Figure 2.4 Graph for Doubled Words on an Alphabet of Size 3, where $r_0 = 7$ 20

Figure 3.1 Example of the Method to Simultaneously Avoid All Doubled Words 35

Figure 3.2 Diagram for Case 1 in the Proof of Lemma 3.2.10 60

Figure 3.3 Diagram for Case 2 in the Proof of Lemma 3.2.10 61

Figure 3.4 Diagram for Case 3 in the Proof of Lemma 3.2.10 62

Chapter 1

Introduction

The study of the combinatorial properties of words traces back to 1906 when Axel Thue showed that there exists an infinite squarefree word. Define the *n-letter alphabet* to be the set $X = \{x_0, x_1, \ldots, x_{n-1}\}$. A *word* (or *pattern*) on the alphabet X is a finite length string of symbols formed from letters in X. We denote by X^* the set of all words on the alphabet X, including the empty word ϵ. We denote by X^+ the set of all non-empty words on the alphabet X. If every letter in a word w appears at least twice, we say that the word is *doubled*, and if every letter appears at least three times, we say that it is *tripled*. For a word w, let $|w|$ be the length of the word w, and let $\alpha(w)$ be the set of all letters occurring in the word w. If $w = uv$ for non-empty words u and v on the same alphabet as w, we call u an *initial segment* (or *prefix*) of w and v a *final segment* (or *suffix*) of w. A word w is said to be a *subword* of a word u if u can be written as awb for some (possibly empty words) a and b.

Where $\cdot$ references the operation of concatenation, $\langle X^*, \cdot, \epsilon \rangle$ is a free monoid generated by X, and the system $\langle X^+, \cdot \rangle$ is a free semigroup generated by X. Any homomorphism $h : \langle X^+, \cdot \rangle \to \langle Y^+, \cdot \rangle$ is uniquely determined by its restriction to X, and any map from X into Y^+ extends uniquely to such a homomorphism. We sometimes call these homomorphisms *non-erasing* to ensure that no letter can be mapped to the empty word. An *erasing homomorphism* removes the restriction that X maps into Y^+ and instead is allowed to map into Y^*. For example, suppose that we have a homomorphism $h : \{e, n, t\} \to \{i, m, p, s\}^+$ defined by $e \mapsto ssi$, $n \mapsto mi$, and $t \mapsto ppi$. Then $h(neet) = mississippi$. This is illustrated in Figure 1.1.

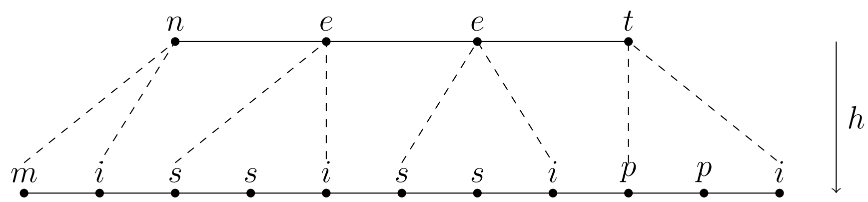

Figure 1.1 Example of a Homomorphism

The word u is said to *encounter* w if there is a homomorphism φ such that $\varphi(w)$ is a subword of u. For example, *mississippi* encounters xx using the homomorphism $x \mapsto s$ or the homomorphism $x \mapsto ssi$, among others. The word u is said to *avoid* w if it does not encounter w. That is, u avoids w if there is no homomorphism φ such that $\varphi(w)$ is a subword of u. In this case, the word u is sometimes called an *avoiding word* of w. A word w (or set of words S) is said to be n*-avoidable* if there is an infinite list of words on the n-letter alphabet that avoid w (or avoid every word in S). If such an n exists, then w is called *avoidable*, and if not, w is called *unavoidable*. Note that if a word is unavoidable, then for any n, there are only finitely many words on the n-letter alphabet that avoid w. Finally, the smallest n such that w is n-avoidable is called its *avoidability index*.

Bean, Ehrenfeucht, and McNulty (1979) proved a characterization of unavoidable words. They showed that a word is unavoidable if and only if it can be reduced to a word of length 1 through an iterative process that at each step deletes all occurrences of some letters in a set that satisfy specific constraints. In particular, they showed that each unavoidable word contains a letter that appears only once. Thus, each doubled and each tripled word is avoidable. Independently, Zimin (1982) gave the same characterization and also gave a second characterization involving a special group of unavoidable words.

The goal of this dissertation is to conclude how many letters are needed to avoid doubled and tripled words. We break our discussion into two parts: avoiding doubled and tripled individual words, and avoiding all doubled words on a particular given alphabet simultaneously. Bean, Ehrenfeucht, and McNulty (1979) obtain results in simultaneous avoidance of all words that give an exponential bound on the size of the avoiding alphabet in terms on the number of letters in the alphabet. Zimin (1982) obtains results that, upon extraction, yields a slightly better exponential bound. Baker, McNulty, and Taylor (1989) reduced the bound of the avoidance of all words to a linear bound in terms of the number of letters in the alphabet. The smallest known result in the avoidability of all words simultaneously is attributed to Mel'nichuk, and an exposition is given in (Lothaire 2002). This bound is $4\lfloor \frac{n}{2} \rfloor + 4$, where n is the size of the alphabet. Finally, Mel'nichuk (1985) gives the best linear bound currently known for simultaneous avoidance of doubled words: $3\lfloor \frac{n}{2} \rfloor + 3$, where n is the size of the alphabet.

Related to avoidability of the word w is counting the number of avoiding words of w of a certain length on the m-letter alphabet. If this number is never 0 for every length, then w is m-avoidable. We say that w exhibits *exponential growth in avoidability* if there is an exponential lower bound on the number of avoiding words of w of a certain length.

Dalalyan (1984) showed, among other results, that each doubled word exhibits exponential growth in avoidability on the 4-letter alphabet, thus showing that each doubled word is 4-avoidable. In recent years, Bell and Goh (2007) and Blanchet-Sadri and Woodhouse (2013) have taken on the problem of individually avoiding doubled words. Bell and Goh (2007), apparently independently of Dalalyan (1984), recreate Dalalyan's result on the 4-avoidability of doubled words, among other things, and Blanchet-Sadri and Woodhouse (2013) strengthen the arguments of Bell and Goh (2007), among other things. Each of these results is obtained by finding the exponential lower bound, which seems to exist for most doubled words. Even xx, shown to be 3-avoidable by Thue (1906), has an exponential lower bound, shown by Brandenburg (1983) and Brinkhuis (1983).

Cassaigne (1994) classified the avoidability index of all binary words (words on the 2-letter alphabet) and partially classifies the avoidability index of all ternary words (words on the 3-letter alphabet), and Ochem (2006) finishes this classification. This, coupled with the results of Dalalyan (1984), can be used to show that all tripled words are 2-avoidable.

In my dissertation, I improve the results of Dalalyan (1984) and Blanchet-Sadri and Woodhouse (2013). Prior to my results, there were 7441 doubled words in which exactly 4 and 5 distinct letters appear whose 3-avoidability was yet to be shown. As a consequence of my results, there are now only 99 doubled words in which exactly 4 or 5 distinct letters appear, each occurring exactly twice, not yet known to be 3-avoidable.

I also give an exposition of (Dalalyan 1984) and (Mel'nichuk 1985) due to the difficulty to obtain and translate their results. The results of Dalalyan (1984) were published in the Reports of the Academy of Sciences of the Armenian Soviet Socialist Republic in 1984, and this journal is not widely available. The work of Mel'nichuk (1985) also appeared in a collection of papers not in wide circulation in the west. Moreover, it is difficult to understand due to being very terse, so I give a more full exposition that fills in the details.

Chapter 2

Avoiding Doubled Words Individually

2.1 Dalalyan's Results

In 1984, A. G. Dalalyan published a result on the avoidability index of doubled and tripled words that strengthened previous results of Bean, Ehrenfeucht, and McNulty (1979). His approach was different from his predecessors in that he used a very combinatorial approach. His results are presented here.

Theorem 2.1.1. *Let w be a word on the n-letter alphabet with r_i being the number of times the letter x_i appears for all i, and let $r = \min(r_0, \ldots, r_{n-1})$. If there exists a λ that satisfies the constraints*

- $\lambda \leq m$
- $\lambda > \sqrt[r]{m}$
- $\left(\dfrac{m-\lambda}{\lambda}\right)(\lambda^r - m)^n \geq m^n,$

then there are at least λ^ℓ words of length ℓ on the m-letter alphabet that avoid w. In particular, w is m-avoidable.

Proof. Let λ satisfy the constraints above, and let $\gamma_m(\ell)$ be the number of words of length ℓ on the m-letter alphabet that avoid w. We desire to show that this function is positive for all values ℓ. To achieve this, we will show that λ satisfies $\gamma_m(\ell+1) \geq \lambda\gamma_m(\ell)$ for every natural number ℓ. This shows that $\gamma_m(\ell) \geq \lambda^\ell$ for all

ℓ. If $\lambda > 1$, this also gives us that γ grows exponentially with ℓ. We will induct on ℓ to prove our claim.

For the base step, we can easily see that $\gamma_m(0) = 1$ and $\gamma_m(1) = m$, so

$$\gamma_m(1) = m \geq \lambda = \lambda\gamma_m(0).$$

So, assume for the sake of induction that $\gamma_m(k+1) \geq \lambda\gamma_m(k)$ for all $k < \ell$, and we will show that $\gamma_m(\ell+1) \geq \lambda\gamma_m(\ell)$.

Let $\delta_m(\ell+1)$ be the number of words of length $\ell+1$ such that each word's initial segment of length ℓ avoids w but the whole word encounters w. From this, we see that $\gamma_m(\ell+1) = m\gamma_m(\ell) - \delta_m(\ell+1)$. We will work to achieve an upper bound on $\delta_m(\ell+1)$.

First, notice that $\delta_m(\ell+1)$ is bounded above by the number of words of the form $u\varphi(w)$ with length $\ell+1$, where u avoids w and φ is a nonerasing homomorphism. Given a choice of φ, let ℓ_0 be the length of the image of x_0, let ℓ_1 be the length of the image of x_1, and so on, finally letting ℓ_{n-1} be the length of the image of x_{n-1}. Then we see that the length of $\varphi(w)$ is $\sum_{i=0}^{n-1} r_i\ell_i$ (which is less than $\ell+1$ by construction), so the length of u is $\ell+1-\sum_{i=0}^{n-1} r_i\ell_i$. Finally, this gives us that the number of possible u's is $\gamma_m\left(\ell+1-\sum_{i=0}^{n-1} r_i\ell_i\right)$.

To count the number of possible homomorphisms, we can see that there are at most $m^{\ell_0}m^{\ell_1}\dots m^{\ell_{n-1}}$ possibilities of φ for a choice of $(\ell_0,\dots,\ell_{n-1})$. So, to find the total number of possible homomorphisms, we'll do a double sum, the outer of which is the sum of the ℓ_i's, and the inner of which sums over all choices of $(\ell_0,\dots,\ell_{n-1})$ with a specific sum. Putting this all together, we get the following:

$$\delta_m(\ell+1) \leq \sum_{j=n}^{\ell+1}\left(\sum_{\ell_0+\dots+\ell_{n-1}=j}\gamma_m\left(\ell+1-\sum_{i=0}^{n-1} r_i\ell_i\right)m^{\ell_0+\dots+\ell_{n-1}}\right).$$

Next, note that $\sum_{i=0}^{n-1} r_i\ell_i \geq \sum_{i=0}^{n-1} r\ell_i = rj$. Also, note that the number of ways to sum n positive integers to a total of j is given by $\binom{j-1}{n-1}$. This gives us:

$$\begin{aligned}\delta_m(\ell+1) &\leq \sum_{j=n}^{\ell+1} \gamma_m(\ell+1-rj) \sum_{\ell_0+\cdots+\ell_{n-1}=j} m^j \\ &= \sum_{j=n}^{\ell+1} \gamma_m(\ell+1-rj) \binom{j-1}{n-1} m^j.\end{aligned}$$

We now use our induction hypothesis, rewritten as $\lambda^{-s}\gamma_m(\ell) \geq \gamma_m(\ell-s)$. This gives us:

$$\begin{aligned}\delta_m(\ell+1) &\leq \sum_{j=n}^{\ell+1} \lambda^{1-rj}\gamma_m(l) \binom{j-1}{n-1} m^j \\ &= \sum_{j=n}^{\ell+1} \lambda\gamma_m(\ell) \binom{j-1}{n-1} \left(\frac{m}{\lambda^r}\right)^j \\ &= \lambda\gamma_m(\ell) \sum_{j=n}^{\ell+1} \binom{j-1}{n-1} \left(\frac{m}{\lambda^r}\right)^j \\ &\leq \lambda\gamma_m(\ell) \sum_{j=n}^{\infty} \binom{j-1}{n-1} \left(\frac{m}{\lambda^r}\right)^j.\end{aligned}$$

Now, we use the second constraint that $\lambda > \sqrt[r]{m}$. Then $\frac{m}{\lambda^r} < 1$, and we will use the geometric series. Note that as long as $|x| < 1$, this equation holds:

$$\left(\frac{1}{1-x}\right)^n = \left(\sum_{i=0}^{\infty} x^i\right)^n = \sum_{i=0}^{\infty} \binom{n-1+i}{n-1} x^i.$$

We can use this fact and rewrite our sum above by using $j = n + i$ as follows:

$$\begin{aligned}\delta_m(\ell+1) &\leq \lambda\gamma_m(\ell) \left(\frac{m}{\lambda^r}\right)^n \sum_{i=0}^{\infty} \binom{n-1+i}{n-1} \left(\frac{m}{\lambda^r}\right)^i \\ &= \lambda\gamma_m(\ell) \left(\frac{m}{\lambda^r}\right)^n \left(\frac{1}{1-\frac{m}{\lambda^r}}\right)^n \\ &= \lambda\gamma_m(\ell) \left(\frac{m}{\lambda^r - m}\right)^n.\end{aligned}$$

Finally, we put this inequality back into our original expression.

$$\begin{aligned}\gamma_m(\ell+1) &= m\gamma_m(\ell) - \delta_m(\ell+1) \\ &\geq m\gamma_m(\ell) - \lambda\gamma_m(\ell) \left(\frac{m}{\lambda^r - m}\right)^n \\ &= \gamma_m(\ell) \left(m - \lambda \left(\frac{m}{\lambda^r - m}\right)^n\right)\end{aligned}$$

So, in order to finish our induction, we must have that $m-\lambda\left(\frac{m}{\lambda^r-m}\right)^n \geq \lambda$. However, this is nothing more than a rewriting of our third constraint. So, we have shown that $\gamma_m(\ell+1) \geq \lambda\gamma_m(\ell)$, and hence there are infinitely many words that avoid w on the m-letter alphabet. Thus, w is m-avoidable. $\square$

With this theorem under our belt, we prove the main theorems of Dalalyan's paper, which are presented as corollaries to Theorem 2.1.1. The first 3 corollaries are the main results of Dalalyan (1984), and the others follow from his results.

Corollary 2.1.2. *Each doubled word is* 4*-avoidable.*

Proof. Let $m=4$, $r\geq 2$, and $\lambda=\sqrt{12}$. Then $\sqrt{12}\leq 4$ and $\sqrt{12}>\sqrt{4}\geq\sqrt[r]{4}$. For the third constraint, we know that when $n=2$, all doubled words contain a square and are hence 3-avoidable. When $n=3$, we see that

$$\left(\frac{4-\sqrt{12}}{\sqrt{12}}\right)\left(\sqrt{12}^2-4\right)^3 \approx 79.21 > 4^3 = 64.$$

For $n>3$, notice that as long as $\lambda^r-m\geq m$, the inequality still holds. To see this, let M_n denote the left side of the third constraint, and assume that $M_n\geq m^n$. Then

$$M_{n+1} = M_n\cdot(\lambda^r-m) \geq M_n\cdot m \geq m^n\cdot m = m^{n+1}.$$

So, in this case, we see that $\sqrt{12}^2-4=8\geq 4$, so the third constraint holds for all n. $\square$

Corollary 2.1.3. *Each doubled word on an alphabet of size at least* 6 *is* 3*-avoidable.*

Proof. Let $m=3$, let $r\geq 2$, let $\lambda=\sqrt{8}$, and let $n\geq 6$. Then $\sqrt{8}\leq 3$ and $\sqrt{8}>\sqrt{3}\geq\sqrt[r]{3}$. For the third constraint,

$$\left(\frac{3-\sqrt{8}}{\sqrt{8}}\right)\left(\sqrt{8}^2-3\right)^6 \approx 947.82 > 3^6 = 729.$$

For $n>6$, note that $\sqrt{8}^2-3=5\geq 3$. $\square$

Corollary 2.1.4. *Each tripled word on an alphabet of size at least* 4 *is* 2*-avoidable.*

Proof. Let $m = 2$, let $r \geq 3$, let $\lambda = \sqrt[3]{6}$, and let $n \geq 4$. Then $\sqrt[3]{6} \leq 2$ and $\sqrt[3]{6} > \sqrt[3]{2} \geq \sqrt[r]{2}$. For the third constraint,

$$\left(\frac{2 - \sqrt[3]{6}}{\sqrt[3]{6}}\right)\left(\sqrt[3]{6}^3 - 2\right)^4 \approx 25.76 > 2^4 = 16.$$

For $n > 4$, note that $\sqrt[3]{6}^3 - 2 = 4 \geq 2$. ◻

Corollary 2.1.5. *Each tripled word is* 3*-avoidable.*

Proof. Let $m = 3$, let $r \geq 3$, and let $\lambda = 2$. Then $2 \leq 3$ and $2 > \sqrt[3]{3} \geq \sqrt[r]{3}$. For the third constraint,

$$\left(\frac{3 - 2}{2}\right)\left(2^3 - 3\right) = 2.5 > 2^1 = 2.$$

For $n > 1$, note that $2^3 - 3 = 5 \geq 3$. ◻

Corollary 2.1.6. *Each word in which every letter appears at least $r \geq 4$ times is* 2*-avoidable.*

Proof. Let $m = 2$, let $r \geq 4$, and let $\lambda = \sqrt[4]{6}$. Then $\sqrt[4]{6} \leq 2$ and $\sqrt[4]{6} > \sqrt[4]{2} \geq \sqrt[r]{2}$. For the third constraint, we know that when $n = 1$, A encounters a cube and is hence 2-avoidable. For $n \geq 2$,

$$\left(\frac{2 - \sqrt[4]{6}}{\sqrt[4]{6}}\right)\left(\sqrt[4]{6}^r - 2\right)^2 \geq \left(\frac{2 - \sqrt[4]{6}}{\sqrt[4]{6}}\right)\left(\sqrt[4]{6}^4 - 2\right)^2 \approx 4.45 > 2^2 = 4.$$

For $n > 2$, note that $\sqrt[4]{6}^4 - 2 = 4 \geq 2$. ◻

Note: In all of the corollaries above we have that the specified words are not only m-avoidable but the growth function on their avoiding words is exponential.

Dalalyan's proof can be analyzed for various assumptions that, if removed, can tighten the result. One assumption that appears to be unnecessary is that a word in $\delta_m(\ell + 1)$ is bounded above by the number of words of the form $u\varphi(w)$ with length $\ell + 1$, where u avoids w and φ is a nonerasing homomorphism. It is possible that a word in $\delta_m(\ell+1)$ can be written in multiple ways, and hence this bound is an overcount. Another assumption is when r_i is replaced by r, and we see in the next section that removing this assumption yields stronger results.

2.2 An Improvement on Dalalyan's Results

We present here a number of stronger results derived by using the methods of Dalalyan (1984) without using the assumption that $r \leq r_i$.

Theorem 2.2.1. *Let w be a word on the n-letter alphabet with r_i being the number of times the letter x_i appears for all i, and let $r = \min(r_0, \ldots, r_{n-1})$. If there exists a λ that satisfies the constraints*

- $\lambda \leq m$
- $\lambda > \sqrt[r]{m}$
- $\left(\dfrac{m-\lambda}{\lambda}\right)\displaystyle\prod_{i=0}^{n-1}\left(\lambda^{r_i}-m\right) \geq m^n,$

then there are at least λ^ℓ words of length ℓ on the m-letter alphabet that avoid w. In particular, w is m-avoidable.

Proof. Let λ satisfy the constraints above, and let $\gamma_m(\ell)$ be defined as in Dalalyan's proof. We again desire to show that this function is positive for all values ℓ by showing that λ satisfies $\gamma_m(\ell+1) \geq \lambda\gamma_m(\ell)$ for every natural number ℓ. This will in turn show that $\gamma_m(\ell) \geq \lambda^\ell$ for all ℓ. We will again induct on ℓ to prove our claim.

For the base step, we can easily see that $\gamma_m(0) = 1$ and $\gamma_m(1) = m$, so

$$\gamma_m(1) = m \geq \lambda = \lambda\gamma_m(0).$$

So, assume for the sake of induction that $\gamma_m(k+1) \geq \lambda\gamma_m(k)$ for all $k < \ell$, and we will show that $\gamma_m(\ell+1) \geq \lambda\gamma_m(\ell)$.

Let $\delta_m(\ell+1)$ be defined as in Dalalyan's proof, and we will again use the fact that $\gamma_m(\ell+1) = m\gamma_m(\ell) - \delta_m(\ell+1)$ and work to achieve an upper bound on $\delta_m(\ell+1)$.

Again notice that $\delta_m(\ell+1)$ is bounded above by the number of words of the form $u\varphi(w)$ with length $\ell+1$, where u avoids w and φ is a nonerasing homomorphism. Given a choice of φ, again let ℓ_0 be the length of the image of x_0, let ℓ_1 be the length of the image of x_1, and so on, finally letting ℓ_{n-1} be the length of the image of x_{n-1}. Then the length of u is $\ell+1-\sum_{i=0}^{n-1} r_i\ell_i$ and the number of possible u's is $\gamma_m\left(\ell+1-\sum_{i=0}^{n-1} r_i\ell_i\right)$. Also, there are again at most $m^{\ell_0}m^{\ell_1}\dots m^{\ell_{n-1}}$ possibilities for a choice of $(\ell_0,\dots,\ell_{n-1})$. So, to find the total number of possible homomorphisms, we'll do a double sum, the outer of which is the sum of the ℓ_i's, and the inner of which sums over all choices of $(\ell_0,\dots,\ell_{n-1})$ with a specific sum. Putting this all together, we get the following:

$$\delta_m(\ell+1) \le \sum_{j=n}^{\ell+1}\left(\sum_{\ell_0+\cdots+\ell_{n-1}=j} \gamma_m\left(\ell+1-\sum_{i=0}^{n-1} r_i\ell_i\right) m^{\ell_0+\cdots+\ell_{n-1}}\right).$$

At this point, rather than use that fact that $r \le r_i$ for all the r_i's, we will go straight into the induction hypothesis. We again rewrite it as $\lambda^{-s}\gamma_m(\ell) \ge \gamma_m(\ell-s)$. This gives us that:

$$\begin{aligned}
\delta_m(\ell+1) &\le \sum_{j=n}^{\ell+1}\left(\sum_{\ell_0+\cdots+\ell_{n-1}=j} \lambda^{1-\sum_{i=0}^{n-1} r_i\ell_i}\gamma_m(\ell)\, m^{\ell_0+\cdots+\ell_{n-1}}\right)\\
&= \lambda\gamma_m(\ell)\cdot\sum_{j=n}^{\ell+1}\left(\sum_{\ell_0+\cdots+\ell_{n-1}=j} \frac{m^{\ell_0+\cdots+\ell_{n-1}}}{\lambda^{r_0\ell_0+\cdots+r_{n-1}\ell_{n-1}}}\right)\\
&= \lambda\gamma_m(\ell)\cdot\sum_{j=n}^{\ell+1}\left(\sum_{\ell_0+\cdots+\ell_{n-1}=j} \left(\frac{m}{\lambda^{r_0}}\right)^{\ell_0}\left(\frac{m}{\lambda^{r_1}}\right)^{\ell_1}\cdots\left(\frac{m}{\lambda^{r_{n-1}}}\right)^{\ell_{n-1}}\right)\\
&\le \lambda\gamma_m(\ell)\cdot\sum_{j=n}^{\infty}\left(\sum_{\ell_0+\cdots+\ell_{n-1}=j} \left(\frac{m}{\lambda^{r_0}}\right)^{\ell_0}\left(\frac{m}{\lambda^{r_1}}\right)^{\ell_1}\cdots\left(\frac{m}{\lambda^{r_{n-1}}}\right)^{\ell_{n-1}}\right)\\
&= \lambda\gamma_m(\ell)\cdot\sum_{\ell_0=1}^{\infty}\sum_{\ell_1=1}^{\infty}\cdots\sum_{\ell_{n-1}=1}^{\infty}\left(\frac{m}{\lambda^{r_0}}\right)^{\ell_0}\left(\frac{m}{\lambda^{r_1}}\right)^{\ell_1}\cdots\left(\frac{m}{\lambda^{r_{n-1}}}\right)^{\ell_{n-1}}\\
&= \lambda\gamma_m(\ell)\cdot\sum_{\ell_0=1}^{\infty}\left(\frac{m}{\lambda^{r_0}}\right)^{\ell_0}\sum_{\ell_1=1}^{\infty}\left(\frac{m}{\lambda^{r_1}}\right)^{\ell_1}\cdots\sum_{\ell_{n-1}=1}^{\infty}\left(\frac{m}{\lambda^{r_{n-1}}}\right)^{\ell_{n-1}}.
\end{aligned}$$

Now, we use the second constraint that $\lambda > \sqrt[r]{m}$. Then since $r \leq r_i$ for all $i < n$, we have that $\frac{m}{\lambda^{r_i}} \leq \frac{m}{t^r} < 1$ for all $i < n$. So, by geometric series, we get:

$$\begin{aligned}
\delta_m(\ell+1) &\leq \lambda\gamma_m(\ell) \cdot \left(\frac{m/\lambda^{r_0}}{1-m/\lambda^{r_0}}\right)\left(\frac{m/\lambda^{r_1}}{1-m/\lambda^{r_1}}\right)\cdots\left(\frac{m/\lambda^{r_{n-1}}}{1-m/\lambda^{r_{n-1}}}\right) \\
&= \lambda\gamma_m(\ell) \cdot \left(\frac{m}{\lambda^{r_0}-m}\right)\left(\frac{m}{\lambda^{r_1}-m}\right)\cdots\left(\frac{m}{\lambda^{r_{n-1}}-m}\right) \\
&= \lambda\gamma_m(\ell) \cdot m^n \prod_{i=0}^{n-1}\left(\frac{1}{\lambda^{r_i}-m}\right).
\end{aligned}$$

Finally, we put this inequality back into our original expression.

$$\begin{aligned}
\gamma_m(\ell+1) &= m\gamma_m(\ell) - \delta_m(\ell+1) \\
&\geq m\gamma_m(\ell) - \lambda\gamma_m(\ell) \cdot m^n \prod_{i=0}^{n-1}\left(\frac{1}{\lambda^{r_i}-m}\right) \\
&= \gamma_m(\ell)\left(m - \lambda \cdot m^n \prod_{i=0}^{n-1}\left(\frac{1}{\lambda^{r_i}-m}\right)\right)
\end{aligned}$$

So, in order to finish our induction, we must have that $m - \lambda \cdot m^n \prod_{i=0}^{n-1}\left(\frac{1}{\lambda^{r_i}-m}\right) \geq \lambda$. However, this is nothing more than a rewriting of our third constraint. So, we have shown that $\gamma_m(\ell+1) \geq \lambda\gamma_m(\ell)$, and hence $\gamma_m(\ell+1) \geq \lambda^n$. Thus, there are infinitely many words that avoid w on the m-letter alphabet, so w is m-avoidable. $\square$

Note in the third constraint that as long $\lambda > 1$, the value on the left will go up if any r_i is increased. So, to prove a few results, we define a partially ordered set $\langle S, \leq\rangle$, where

$$S = \{(r_0, r_1, \ldots, r_{n-1}) \mid r_i \text{ is the number of times } x_i \text{ appears in } w\}.$$

and $\leq$ is defined by

$$(r_0, r_1, \ldots, r_{n-1}) \leq (s_0, s_1 \ldots, s_{n-1}) \text{ if and only if } r_i \leq s_i \text{ for all } i < n.$$

It is clear that if a λ exists that satisfies the theorem for $(r_0, r_1, \ldots, r_{n-1})$ and

$$(r_0, r_1, \ldots, r_{n-1}) \leq (s_0, s_1 \ldots, s_{n-1}),$$

then this λ also satisfies the theorem for $(s_0, s_1 \ldots, s_{n-1})$.

Corollary 2.2.2. *Each doubled word on the n-letter alphabet in which some letter appears at least 3 times is 3-avoidable.*

Proof. Let $m = 3$, let $r \geq 2$, and let $\lambda = \sqrt{8}$. Then $\sqrt{8} \leq 3$ and $\sqrt{8} > \sqrt{3} \geq \sqrt[r]{3}$. We note that when $n = 2$, all doubled words contain a square and are hence 3-avoidable. When $n = 3$, consider the system $(r_0, r_1, r_2) = (3, 2, 2)$. Then

$$\left(\frac{3-\sqrt{8}}{\sqrt{8}}\right)\left(\sqrt{8}^3 - 3\right)\left(\sqrt{8}^2 - 3\right)^2 \approx 29.77 > 3^3 = 27,$$

and this value will not change if we consider a rearrangement of the values of r_0, r_1, and r_2. For $n > 3$ and the system $(r_0, r_1, \ldots, r_{n-1}) = (3, 2, \ldots, 2)$, notice that as long as $\lambda^r - m \geq m$, the inequality still holds. To see this, let M_n denote the left side of the third constraint, and assume that $M_n \geq m^n$. Then

$$M_{n+1} = M_n \cdot (\lambda^r - m) \geq M_n \cdot m =\geq m^n \cdot m = m^{n+1}.$$

So, in this case, we see that $\sqrt{8}^2 - 3 = 5 \geq 3$, so the third constraint holds for all n. □

In other words, each doubled word on the n-letter alphabet with length at least $2n + 1$ is 3-avoidable.

Corollary 2.2.3. *Each tripled word on the 3-letter alphabet in which some letter appears at least 4 times is 2-avoidable.*

Proof. Let $m = 2$, let $r \geq 3$, and let $\lambda = \sqrt[3]{6}$. Then $\sqrt[3]{6} \leq 2$ and $\sqrt[3]{6} > \sqrt[3]{2} \geq \sqrt[r]{2}$. When $n = 3$, consider the system $(r_0, r_1, r_2) = (4, 3, 3)$. Then

$$\left(\frac{2-\sqrt[3]{6}}{\sqrt[3]{6}}\right)\left(\sqrt[3]{6}^4 - 2\right)\left(\sqrt[3]{6}^3 - 2\right)^2 \approx 14.34 > 2^3 = 8,$$

and this value will not change if we rearrange the values of r_0, r_1, and r_2. □

In other words, each tripled word on the 3-letter alphabet with length at least 10 (think $3n + 1$) is 2-avoidable.

To get other results, the following lemma is helpful.

Lemma 2.2.4. *For $\lambda > 1$ and $i \geq j+2$, $(\lambda^i - m)(\lambda^j - m) < (\lambda^{i-1} - m)(\lambda^{j+1} - m)$.*

Proof. We start with the desired inequality and simplify it.

$$\left(\lambda^i - m\right)\left(\lambda^j - m\right) < \left(\lambda^{i-1} - m\right)\left(\lambda^{j+1} - m\right)$$

$$\lambda^{i+j} - m\lambda^i - m\lambda^j + m^2 < \lambda^{i+j} - m\lambda^{i-1} - m\lambda^{j+1} + m^2$$

$$-m\lambda^i - m\lambda^j < -m\lambda^{i-1} - m\lambda^{j+1}$$

$$\lambda^i + \lambda^j > \lambda^{i-1} + \lambda^{j+1}$$

$$\lambda^i - \lambda^{i-1} > \lambda^{j+1} - \lambda^j$$

$$\lambda^{i-1}(\lambda - 1) > \lambda^j (\lambda - 1)$$

This is clearly true since $\lambda > 1$ and $i - 1 > j$. $\square$

Corollary 2.2.5. *Each tripled word on the 2-letter alphabet in which some letter appears at least 5 times or in which at least 2 of its letters appear 4 times is 2-avoidable.*

Proof. Let $m = 2$, let $r \geq 3$, and let $\lambda = \sqrt[3]{6}$. Then $\sqrt[3]{6} \leq 2$ and $\sqrt[3]{6} > \sqrt[3]{2} \geq \sqrt[r]{2}$. When $n = 2$, consider the system $(r_0, r_1) = (5, 3)$ or $(r_0, r_1) = (4, 4)$. By Lemma 2.2.4, we need only check the third constraint for $(5, 3)$. Then

$$\left(\frac{2 - \sqrt[3]{6}}{\sqrt[3]{6}}\right)\left(\sqrt[3]{6}^5 - 2\right)\left(\sqrt[3]{6}^3 - 2\right) \approx 7.17 > 2^2 = 4,$$

and this value will not change if we rearrange the values of r_0 and r_1. $\square$

In other words, each tripled word on the 2-letter alphabet with length at least 8 (think $3n + 2$) is 2-avoidable.

These corollaries were discovered by analyzing pictures. We present a few graphs of the constraints to more clearly show what we are looking for in using Corollary 2.2.6. Figure 2.1 and Figure 2.2 show the left side of the third constraint of Corollary 2.2.6. In these figures, we desire for the curve to lie above m^n for some λ

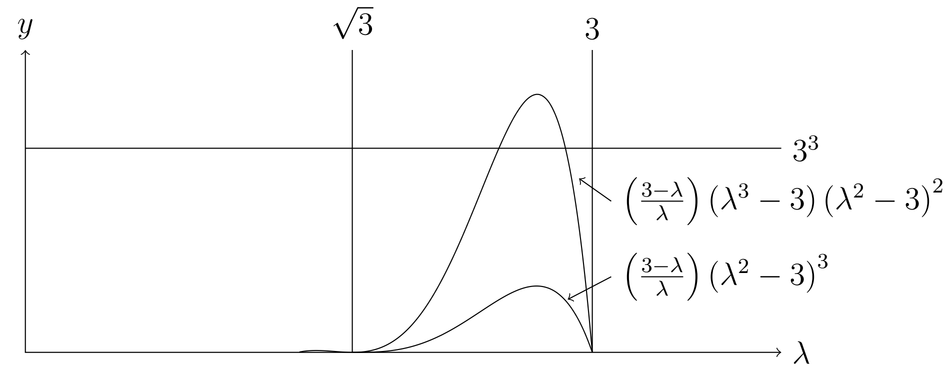

Figure 2.1 Graph for 3-Avoidability of Doubled Words

between the lines $\sqrt{m}$ and m. From these pictures, the λ chosen in each of the corollaries above were conveniently picked from where the curve lies above m^n between $\sqrt{m}$ and m.

Figure 2.1 shows the curves for when a doubled word on the 3-letter alphabet has length 6 and length 7. We see that this method fails when the length is exactly twice the number of letters, but works when adding a single letter. Figure 2.2 shows the curves for tripled words. The blue curves represent the tripled words on the 3-letter alphabet of lengths 9 and 10. We see that this method fails when the length is exactly three times the number of letters, but works when adding a single letter. The red curves represent the tripled words on the 2-letter alphabet of lengths 7 and 8. We see that this method fails when the length is one more than three times the number of letters, but works when adding two additional letters.

This result can also reprove Dalalyan's results. We use his same values for m, r, λ, and m, but we mention the systems here for reference. Note that the third constraint of Theorem 2.2.1 reduces to the third constraint of Theorem 2.1.1 when all the r_i's are equal.

- For Corollary 2.1.2, the system is $(2, 2, 2)$.

- For Corollary 2.1.3, the system is $(2, 2, 2, 2, 2, 2)$.

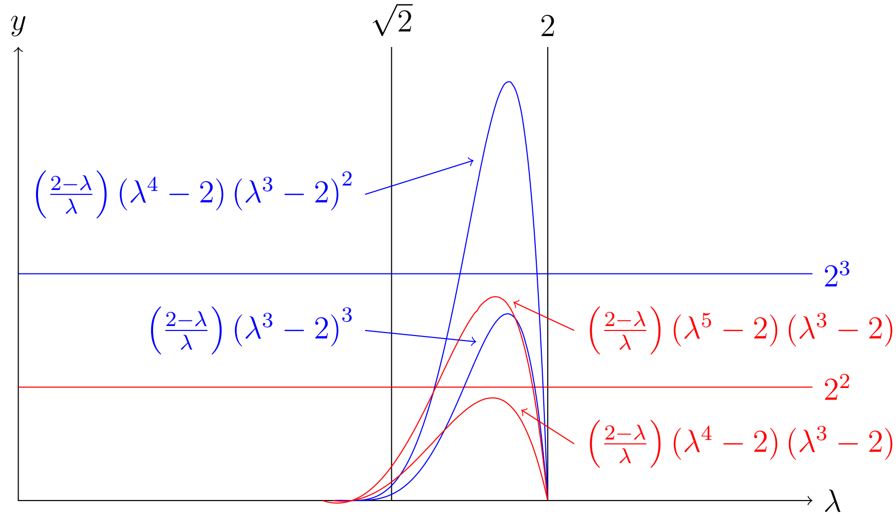

Figure 2.2 Graph for 2-Avoidability of Tripled Words

- For Corollary 2.1.4, the system is $(3, 3, 3, 3)$.
- For Corollary 2.1.5, the system is (3).
- For Corollary 2.1.6, the system is $(4, 4)$.

Note: In all of the corollaries above we have that the specified words are not only m-avoidable but the growth function on their avoiding words is exponential.

Remark 2.2.1. This leaves the following types of doubled words to check if they are 3-avoidable: words in which exactly 3 distinct letters appear with length 6, words in which exactly 4 distinct letters appear with length 8, and words in which exactly 5 distinct letters appear with length 10. To determine if these are 3-avoidable, we create a list of all the words of this type, then remove the words that contain squares and are hence 3-avoidable. We then remove words that are encountered by other words in this list. Next, we remove words encountered by other doubled words that are known to be 3-avoidable. Finally, we remove words that are a relettering of the reverse of another word. In doing so, we arrive at 2 words in which exactly 3 distinct letters appear with length 6, 11 words in which exactly 4 distinct letters appear with length 8, and 88 words in which exactly 5 disctint letters appear with length 10. The words *abacbc* and *abcacb* were shown by Cassaigne (1994) to be 3-avoidable, so we are left with 99 doubled words to check. A table of these words is given in Section 2.6.

Remark 2.2.2. This leaves the tripled words in which exactly 2 distinct letters appear with length 6 or 7 and the tripled words in which exactly 3 distinct letters appear with length 9 to check if they are 2-avoidable. To determine if these are 2-avoidable, we create a list of all words of this type, then remove the words that contain cubes and are hence 2-avoidable. We then remove the words that are encountered by other words in this list. Finally, we remove words that are a relettering of the reverse of another word. In doing so, we arrive at 4 words in which exactly 2 distinct letters appear of length 6, 1 word in which exactly 2 distinct letters appear of length 7, and 101 words in which exactly 3 distinct letters appear of length 9. Using are careful analysis of the tables provided by Cassaigne (1994) and Ochem (2006), we can show that all of these words are 2-avoidable. We more carefully describe this proof in Section 2.7.

Remark 2.2.3. This method cannot prove a theorem of the form "Each doubled word on at an alphabet of size at least n is 2-avoidable". To show this, let $m = 2$, let $r \geq 2$, and let λ be such that $\sqrt{2} < \lambda \leq 2$. Let the system be $(r_0) = (2)$. It is easy to check through calculus that $\left(\frac{2-\lambda}{\lambda}\right)(\lambda^2 - 2)$ maximizes at $0.16 < 2$ for any choice of λ in $\left(\sqrt{2}, 2\right]$. Further, note that $\lambda^2 - 2 \leq 2 = m$, so moving up to a higher n will maintain the $<$ inequality. Thus, the third constraint can never be satisfied in order to make a theorem of this form. This is illustrated in Figure 2.3, using an example of $n = 2$. Note that as n gets larger, the distance between the curve and the line above only grows proportionally.

However, we can find a basic result in the 2-avoidability of doubled words. First, we present a corollary of Theorem 2.2.1 that is useful to find this result.

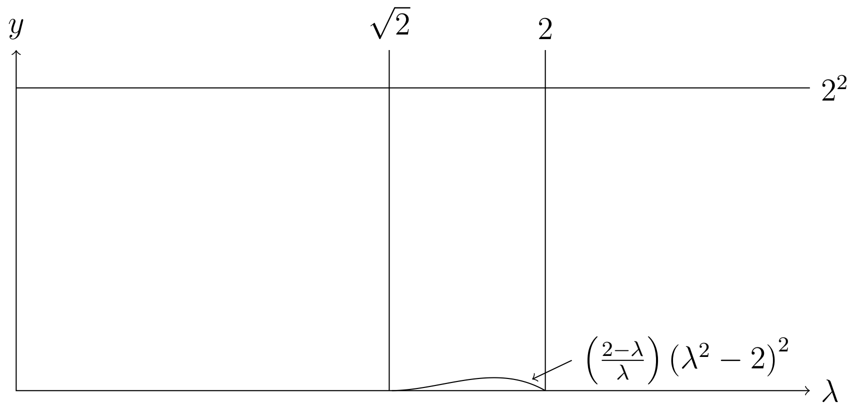

Figure 2.3 Graph for Doubled Words on an Alphabet of Size 2

Corollary 2.2.6. *Let w be a word on the n-letter alphabet with r_i being the number of times the letter x_i appears for all i, and let $r = \min(r_0, \ldots, r_{n-1})$. If there exists a λ that satisfies the constraints*

- $\lambda \leq m$
- $\lambda > \sqrt[r]{m}$
- $\left(\dfrac{m-\lambda}{\lambda}\right)\left(\lambda^{|w|-r(n-1)} - m\right)\left(\lambda^r - m\right)^{n-1} \geq m^n,$

then there are at least λ^ℓ words of length ℓ on the m-letter alphabet that avoid w. In particular, w is m-avoidable.

Proof. Using Lemma 2.2.4, we see that we can decrease each r_i for $i > 0$ to r and increase r_0 by the sum of the decreases. □

This corollary could also be proven from Theorem 2.2.1 by using modified form of Lemma 3 in (Blanchet-Sadri and Woodhouse 2013). It is worth noting, however, that Corollary 2.2.6 is not as strong as Theorem 2.2.1. For example, consider the word $xyxzxyxzxy$. Then $r_0 = 5$, $r_1 = 3$, and $r_2 = 2$. By Theorem 2.2.1, this word is 2-avoidable by applying $\lambda = 1.97$. However, using Corollary 2.2.6, there is no λ that satisfies $\left(\dfrac{2-\lambda}{\lambda}\right)\left(\lambda^6 - 2\right)\left(\lambda^2 - 2\right)^2 \geq 4$.

Corollary 2.2.7. *For all $n \geq 2$, there exists a length L dependent on n such that each doubled word on the n-letter alphabet with length at least L has an exponential lower bound on the number of their avoiding words on the 2-letter alphabet. In particular, these words are 2-avoidable.*

Proof. Let $m = 2$, let $r \geq 2$, and let λ be such that $\sqrt[r]{2} \leq \sqrt{2} < \lambda \leq 2$. Consider $r_0 = |w| - r\,(n-1)$. Then we want $\left(\frac{2-\lambda}{\lambda}\right)(\lambda^{r_0} - 2)\left(\lambda^2 - 2\right)^{n-1} > 2^n$. Every value in this expression is fixed with respect to $|w|$, so we are free to make r_0 as large as we like without changing other parameters. In particular, note that $\lambda^{r_0} - 2$ is positive since $\lambda > \sqrt{2}$ and $r_0 \geq 2$, so let r_0 be the smallest value such that the inequality holds. Then the desired length L is simply $r_0 + 2\,(n-1)$, and all words longer than L will still satisfy the third constraint. □

Remark 2.2.4. Corollary 2.2.7 lends itself to making a table of known lower bounds for L using these methods. To use Corollary 2.2.6, we simply need to find a large enough value of r_0 so that the third constraint of Corollary 2.2.6 holds. A sample of results based on Corollary 2.2.6 is given in Table 2.1, but these bounds are not likely to be tight. In addition, some shorter words can be found to be 2-avoidable by Theorem 2.2.1 (see the example of $xyxzxyxzxy$), but these words would be tedious to classify. To get each bound, we manipulate the functions for graphs like Figure 2.1 and increase r_0 one increment at a time until the curve lies above m^n. This is illustrated in Figure 2.4, using an example of $n = 3$. For the case when $n = 2$ and $n = 3$, we give a tight bound based on previous results.

Theorem 2.2.8. *Each doubled word on the 2-letter alphabet with length at least 6 is 2-avoidable, and this is the smallest bound possible.*

Proof. Roth (1992) proves that all words on the 2-letter alphabet with length at least 6 are 2-avoidable. Lothaire (2002) states that *aabab* is 2-unavoidable, which is easily verified by backtracking. (See Theorem 2.5.1) □

Table 2.1 Minimal L for a Doubled Word on the n-letter Alphabet to be 2-avoidable

n	Minimal r_0 to get λ	Upper bound for L
1	N/A	$\operatorname{ind}(xx) = 3$
2	6	8
3	7	11
4	7	13
5	8	16
6	8	18
7	8	20
8	8	22
9	8	24
10	8	26
11	9	29
12	9	31
13	9	33
14	9	35
15	9	37
16	9	39
17	9	41
18	9	43
19	9	45
20	9	47
21	9	49
22	9	51
23	10	54
24	10	56
25	10	58
26	10	60
27	10	62
28	10	64
29	10	66
30	10	68
31	10	70
32	10	72
33	10	74
34	10	76
35	10	78
36	10	80
37	10	82
38	10	84
39	10	86
40	10	88
41	10	90
42	10	92
43	10	94
44	10	96
45	10	98
46	11	101
47	11	103
48	11	105
49	11	107
50	11	109

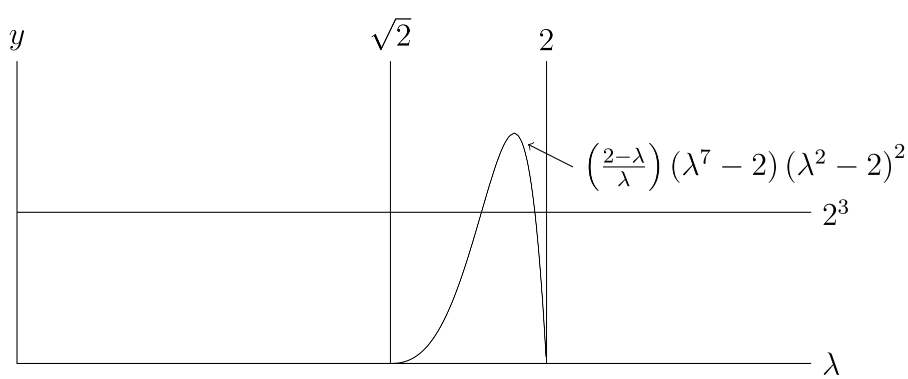

Figure 2.4 Graph for Doubled Words on an Alphabet of Size 3, where $r_0 = 7$

Theorem 2.2.9. *Each doubled word on the* 3*-letter alphabet with length at least* 7 *is* 2*-avoidable, and this is the smallest bound possible.*

Proof. It is easy to show that the words *aabccb*, *abaccb*, *abbacc*, *abbcac*, *abbcca*, *abcbca*, *abccab*, and *abccba* are all 2-unavoidable. We proceed using the classification of ternary words started by Cassaigne (1994) and finished by Ochem (2006). Between these two papers, we discover that all doubled words on 6 letters that are not listed (other than reletterings) have avoidability index 2. Thus, if any word of length 7 or more is to have avoidability index 3, it must contain as a prefix one of these listed words. However, appending the letters a, b, and c to each of these words creates 24 words that all have avoidability index 2. This proves the proposition. □

2.3 Bell, Goh, Rampersad, Blanchet-Sadri, and Woodhouse's Results

Bell and Goh (2007) obtain results that, independent of the work of Dalalyan (1984), include a proof that each doubled word has an exponential lower bound on the number of words on the 4-letter alphabet of length ℓ that avoid it. Their theorem can be stated in the following manner.

Theorem 2.3.1. *(Bell and Goh (2007), Theorem 1, Restated) Let w be a doubled word on the n-letter alphabet with $r_i \geq 2$ being the number of times the letter x_i appears for all i. If there exists a λ that satisfies the constraints*

- $\lambda \leq m$
- $\lambda > \sqrt[r]{m}$
- $\left(\dfrac{m-\lambda}{\lambda}\right)\left(\lambda^2 - m\right)^n \geq m^n,$

then there are at least λ^ℓ words of length ℓ on the m-letter alphabet that avoid w. In particular, w is m-avoidable.

This is found in a similar fashion to Dalalyan by using a power series argument to lower bound the number of words avoiding the doubled word.

Rampersad (2011) follows in a similar fashion to Bell and Goh (2007) in order to prove the following theorem.

Theorem 2.3.2. *(Rampersad (2011), Theorem 1) Let w be a word on the n-letter alphabet. Then:*

(1) *If w has length at least 2^n, then w is 4-avoidable.*

(2) *If w has length at least 3^n, then w is 3-avoidable.*

(3) *If w has length at least 4^n, then w is 2-avoidable.*

Blanchet-Sadri and Woodhouse (2013) obtain a refinement of the results of Rampersad (2011). Using a few technical lemmas, they achieve the following result.

Theorem 2.3.3. *(Blanchet-Sadri and Woodhouse (2013), Theorem 2) Let w be a word on the n-letter alphabet. Then:*

(1) *If w has length at least 2^n, then w is 3-avoidable.*

(2) *If w has length at least $3\left(2^{n-1}\right)$, then w is 2-avoidable.*

In addition, they cite the existence of a word of length $2^n - 1$ that is 3-unavoidable and a word of length $3\left(2^{n-1}\right) - 1$ that is 2-unavoidable, thus noting that these bounds are tight. Both of these theorems above also show an exponential lower bound on the number of words of length ℓ avoiding the word with the desired properties.

In each of these three papers, they use a Theorem due to Golod and Šafarevič (1964). In their paper, this theorem is stated in terms of ring theory. However, Rampersad (2011) states and proves it in a more combinatorial fashion. We present his formulation and proof here, with some modifications of the terminology.

Theorem 2.3.4. *(Rampersad (2011), Theorem 2) Let S be a set of words over an m-letter alphabet, each word of length at least 2. Suppose that for each $\ell \geq 2$, the set S contains at most c_i words of length i. If the power series expansion of*

$$G(x) = \left(1 - mx + \sum_{i \geq 2} c_i x^i\right)^{-1}$$

has nonnegative coefficients, then there are at least as many words of length ℓ over an m-letter alphabet that contain no word in S as a subword as the coefficient of x^ℓ in the power series expansion of $G(x)$.

Proof. For two power series $f(x) = \sum_{i \geq 0} a_i x^i$ and $g(x) = \sum_{i \geq 0} b_i x^i$, we write $f \geq g$ to mean that $a_i \geq b_i$ for all $i \geq 0$. Let $F(x) = \sum_{i \geq 0} a_i x^i$, where a_i is the number of words on the m-letter alphabet that contain no word in S as a subword and $G(x) = \sum_{i \geq 0} b_i x^i$ be the power series expansion of $G(x)$ as defined above. We will show that $F \geq G$, and since the coefficients of G are non-negative, this finishes the proof.

For $\ell \geq 1$, there are $m^\ell - a_\ell$ words w of length n on the m-letter alphabet that contain some word in S as a subword. Further, for any w of this form, observe that either $w = w'a$, where w' contains a word in S as a subword and a is a single letter, or $w = xy$, where x has length $\ell - j$ and contains no word in S as a subword and $y \in S$ is a word of length j. Then there are at most $\left(m^{\ell-1} a_{\ell-1}\right) m$ words of the first form and at most $\sum_{j=2}^{\ell} a_{n-j} c_j$ words of the second form. Thus,

$$m^\ell - a_\ell \leq \left(m^{\ell-1} - a_{\ell-1}\right) m + \sum_{j=2}^{\ell} a_{\ell-j} c_j.$$

Rearranging, we get

$$a_\ell - a_{\ell-1} m + \sum_{j=2}^{\ell} a_{\ell-j} c_j \geq 0.$$

Finally, consider the function

$$\begin{aligned} H(x) &= F(x)\left(1 - mx + \sum_{i\geq 2} c_i x^i\right) \\ &= \left(\sum_{i\geq 0} a_i x^i\right)\left(1 - mx + \sum_{i\geq 2} c_i x^i\right). \end{aligned}$$

We see that the coefficient of x^ℓ in $H(x)$ is $a_\ell - a_{\ell-1}m + \sum_{j=2}^{\ell} a_{\ell-j}c_j$, which was just shown to be ≥ 0. We also see that the coefficient of x^0 in $H(x)$ is 1. Thus, the inequality $H \geq 1$ holds, and in particular, $H - 1$ has non-negative coefficients. So, $F = HG = (H-1)G + G$, and since $H - 1$ and G both have non-negative coefficients, $(H-1)G$ has non-negative coefficients. Thus, $(H-1)G \geq 0$, so $F \geq G$, as desired. □

Bell and Goh (2007) and Rampersad (2011) assume that every letter appears at least r times, similar to that of Dalalyan (1984). Blanchet-Sadri and Woodhouse (2013), however, tightens the work of Bell and Goh (2007). Their theorem can be stated in the following manner.

Theorem 2.3.5. *(Blanchet-Sadri and Woodhouse (2013), Lemma 3, Restated) Let w be a doubled word on the n-letter alphabet with $r_i \geq 2$ being the number of times the letter x_i appears for all i. If there exists a λ that satisfies the constraints*

- $\lambda \leq m$
- $\lambda > \sqrt[r]{m}$
- $\left(\dfrac{m-\lambda}{\lambda}\right)\left(\lambda^{|w|-2(n-1)}\right)\left(\lambda^2 - m\right)^{n-1} \geq m^n,$

then there are at least λ^ℓ words of length ℓ on the m-letter alphabet that avoid w. In particular, w is m-avoidable.

Theorem 2.3.5, though only usable for doubled words in its current form, allows for the consideration of the length of the word rather than considering only the fact that each letter occurs at least twice. This theorem is very similar to Corollary 2.2.6, as the only difference is the use of r rather than 2. However, Blanchet-Sadri and Woodhouse (2013) only use their theorem to show the following results, one of which was also given by Dalalyan (1984).

- Each doubled word on an alphabet of size at least 6 is 3-avoidable.
- Each doubled word on an alphabet of size at least 2 with length at least 12 is 3-avoidable.

Theorem 2.3.5 is strong enough to prove Corollary 2.2.2, but Blanchet-Sadri and Woodhouse (2013) seem to only consider the fact that the constraint of the word being on an alphabet of size at least 6 can be replaced by the word having length at least 12.

2.4 A Second Proof of Theorem 2.2.1

We now proceed to give a second proof of Theorem 2.2.1 by modifying the arguments used by Bell and Goh (2007) and Blanchet-Sadri and Woodhouse (2013).

Theorem 2.2.1. Let w be a word on the n-letter alphabet with r_i being the number of times the letter x_i appears for all i, and let $r = \min(r_0, \ldots, r_{n-1})$. If there exists a λ that satisfies the constraints

- $\lambda \leq m$
- $\lambda > \sqrt[r]{m}$
- $\left(\dfrac{m-\lambda}{\lambda}\right)\displaystyle\prod_{i=0}^{n-1}(\lambda^{r_i} - m) \geq m^n,$

then there are at least λ^ℓ words of length ℓ on the m-letter alphabet that avoid w. In particular, w is m-avoidable.

Proof. We will use the following Lemma of Bell and Goh (2007).

Lemma 2.4.1. *Let $n \geq 1$ and let w be a word on the n-letter in which each letter x_i in w occurs $r_i \geq 1$ times for all i. Then for $\ell \geq 1$, the number of words of length ℓ on the m-letter alphabet that are homomorphic images of w is equal to the coefficient of x^ℓ in*

$$C(x) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} m^{i_1+\cdots+i_n} x^{r_1 i_1+\cdots+r_n i_n}.$$

Define S to be the set of all words on the m-letter alphabet that are homomorphic images of w. Then by Lemma 2.4.1, the number of words of length ℓ in S is equal to the coefficient of x^ℓ in

$$C(x) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} m^{i_1+\cdots+i_n} x^{r_1 i_1+\cdots+r_n i_n}.$$

We will use Theorem 2.3.4, so define

$$B(x) = \sum_{i \geq 0} b_i x^i = (1 - mx + C(x))^{-1}$$

We will show by induction that $b_i \geq \lambda b_{i-1}$ for all i. In doing so, we see that $b_i \geq \lambda^i$ for all i, and hence all the coefficients of B are non-negative. So, by Theorem 2.3.4, we get that there are at least λ^ℓ word of length ℓ that do not contain an element from S, and by the construction of S, these words also avoid w.

In order to proceed, we will compute the coefficients of

$$\left(\sum_{i=0}^{\infty} b_i x^i\right) \left(1 - mx + \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} m^{i_1+\cdots+i_n} x^{r_i i_1+\cdots+r_n i_n}\right) = 1.$$

The coefficient of x^0 is b_0 on the left and 1 on the right, so $b_0 = 1$. Similarly, the coefficients of x give us $b_1 - b_0 m = 0$, so $b_1 = m$. So, when $n = 1$, we see that $b_1 = m > \lambda = \lambda b_0$. Now, assume that $b_i \geq \lambda b_{i-1}$ for all $i < \ell$, and we desire to show that $b_\ell \geq \lambda b_{\ell-1}$.

Computing the coefficient of x^i in both sides of the equation above, we get the equation

$$b_\ell - m b_{\ell-1} + \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} m^{i_1+\cdots+i_n} b_{\ell - r_1 i_1 - \cdots - r_n i_n} = 0,$$

where any values of $i_1, \ldots, i_n$ in which $\ell - r_1 i_1 - \cdots - r_n i_n < 0$ are not included in the sums. Hence,

$$b_\ell = \lambda b_{\ell-1} + (m-\lambda)\, b_{n-1} - \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} m^{i_1+\cdots+i_n} b_{\ell-r_1 i_1-\cdots-r_n i_n}.$$

Thus, in order to show that $b_\ell \geq \lambda b_{\ell-1}$, we must show that

$$(m-\lambda)\, b_{\ell-1} - \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} m^{i_1+\cdots+i_n} b_{\ell-r_1 i_1-\cdots-r_n i_n} \geq 0.$$

By the inductive hypothesis, we have that $b_{\ell-i} \leq \frac{b_{\ell-1}}{\lambda^{i-1}}$ for $1 \leq i \leq \ell$. Thus,

$$\begin{aligned}\sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} m^{i_1+\cdots+i_n} b_{\ell-r_1 i_1-\cdots-r_n i_n} &\leq \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} m^{i_1+\cdots+i_n} \frac{b_{\ell-1}}{\lambda^{r_1 i_1+\cdots+r_n i_n-1}} \\ &= \lambda b_{\ell-1} \sum_{i_1=1}^{\infty} \frac{m^{i_1}}{\lambda^{r_1 i_1}} \cdots \sum_{i_n=1}^{\infty} \frac{m^{i_n}}{\lambda^{r_n i_n}}.\end{aligned}$$

Since $\lambda > \sqrt{m} \geq \sqrt[r_i]{m}$, we get

$$\begin{aligned}\lambda b_{\ell-1} \sum_{i_1=1}^{\infty} \frac{m^{i_1}}{\lambda^{r_1 i_1}} \cdots \sum_{i_n=1}^{\infty} \frac{m^{i_n}}{\lambda^{r_n i_n}} &= \lambda b_{\ell-1} \left(\frac{m}{\lambda^{r_1}-m}\right) \cdots \left(\frac{m}{\lambda^{r_n}-m}\right) \\ &= \lambda b_{\ell-1} \prod_{i=1}^{n} \frac{m}{\lambda^{r_i}-m}.\end{aligned}$$

Finally, we desire for the following to be true: $m - \lambda \geq \lambda \prod_{i=1}^{n} \frac{m}{\lambda^{r_i}-m}$. However, this is simply a rewriting of our third constraint. So, we have that

$$\begin{aligned}&(m-\lambda)\, b_{\ell-1} - \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} m^{i_1+\cdots+i_n} b_{\ell-r_1 i_1-\cdots-r_n i_n} \\ &\geq \lambda \prod_{i=1}^{n} \frac{m}{\lambda^{r_i}-m} b_{\ell-1} - \lambda b_{\ell-1} \prod_{i=1}^{n} \frac{m}{\lambda^{r_i}-m} \\ &= 0.\end{aligned}$$

Thus, $b_\ell \geq \lambda b_{\ell-1}$, which completes the proof. □

2.5 Classification of All Binary and Ternary Words

The classification of all the binary words began with the work of Schmidt (1989), who proved that all the binary words with length at least 13 are 2-avoidable. This classification was continued by Roth (1992), who proved that all the binary words with length at least 6 are 2-avoidable. Finally, Cassaigne (1994) extended this work by including a full classification of all the binary words as well as a partial classification of all the ternary words. We include here his classification of the binary words, taken from (Lothaire 2002).

Theorem 2.5.1. *((Lothaire 2002), Theorem 3.3.3) Binary words fall in three categories:*

- *The* 7 *binary words ϵ, a, b, ab, ba, aba, and bab are unavoidable.*
- *The* 22 *binary words aa, bb, aab, abb, baa, bba, aaba, aabb, abaa, abab, abba, baab, baba, babb, bbaa, bbab, aabaa, aabab, ababb, babaa, bbaba, and bbabb have avoidability index* 3*.*
- *All other binary words, and in particular all binary words of length* 6 *or more, have avoidability index 2.*

Also included in (Cassaigne 1994), Appendix A, is a partial classification of all ternary words. Here, he states that what remains in the classification is the word *abcbabc* for 3-avoidability and 103 words of lengths 6 through 10 for 2-avoidability. Ochem (2006) completed this classification by using a new method to generate homomorphisms to show that a word is avoidable on the desired alphabet. It turns out that all of the above mentioned words are 2-avoidable, including *abcbabc*.

2.6 The 3-Avoidability of Doubled Words

Corollary 2.1.3, Corollary 2.2.2, and Cassaigne (1994) state that each doubled word is 3-avoidable except possibly those in which there are exactly 4 or 5 letters with each letter occurring exactly twice. We conjecture the following.

Conjecture 2.6.1. *Each doubled word is* 3*-avoidable.*

Table 2.2 Doubled words Left to Be Shown to Be 3-Avoidable

abacbdcd	abacdbdc	abacdcbd	abcadbdc
abcadcbd	abcadcdb	abcbdadc	abcdacbd
abcdadcb	abcdbadc	abcdbdac	
abacbdcede	abacbdeced	abacbdedce	abacdbcede
abacdbdece	abacdbeced	abacdbedce	abacdbedec
abacdcebde	abacdcebed	abacdcedbe	abacdebced
abacdebdce	abacdebedc	abacdecbed	abacdecebd
abacdedbce	abacdedbec	abacdedcbe	abcacdebed
abcacdedbe	abcadbecde	abcadbeced	abcadbedce
abcadbedec	abcadcebed	abcadcedbe	abcadcedeb
abcadebdce	abcadebecd	abcadebedc	abcadecbed
abcadecdbe	abcadecebd	abcadecedb	abcadedbec
abcadedcbe	abcadedceb	abcbadeced	abcbdaeced
abcbdaedec	abcbdeaced	abcbdeadce	abcbdeaecd
abcbdeaedc	abcbdecaed	abcbdecdae	abcbdecead
abcbdedace	abcbdedaec	abcbdedcae	abcdaebdce
abcdaebdec	abcdaebedc	abcdaecbed	abcdaecebd
abcdaecedb	abcdaedbec	abcdaedcbe	abcdaedceb
abcdaedecb	abcdbeaced	abcdbeadec	abcdbeaedc
abcdbecead	abcdbedaec	abcdbedeac	abcdceaebd
abcdceaedb	abcdeacbed	abcdeacebd	abcdeacedb
abcdeadbec	abcdeadcbe	abcdeadceb	abcdeaebdc
abcdeaecbd	abcdeaedcb	abcdebaced	abcdebaedc
abcdebdaec	abcdebeadc	abcdebecad	abcdebedac
abcdecaebd	abcdecaedb	abcdeceadb	abcdecebad

Table 2.2 is a list of 99 doubled words in which, if these words are shown to be 3-avoidable, the conjecture is proven. Each word seems to be 3-avoidable, evidenced by the fact that each can be shown via a computer program to be avoided by a word of length 200 on the 3-letter alphabet. Generally speaking, most 3-unavoidable words have each avoiding word on the 3-letter alphabet of length at most 30.

See Remark 2.2.1 for how these words were found.

2.7 The 2-Avoidability of Tripled Words

Dalalyan (1984) showed that each tripled word with at least 4 distinct letters is 2-avoidable (see Corollary 2.1.5). Schmidt (1989) and Roth (1992), working on arbitrary binary words, showed that all binary words of length 6 or more were 2-avoidable. This shows that each tripled word with exactly 2 distinct letters is 2-avoidable. Finally, Cassaigne (1994) and Ochem (2006) classified the avoidability index of all ternary words. While Cassaigne and Ochem were not specifically searching for avoidability indices of tripled words, examination of their results shows that each tripled word with exactly 3 distinct letters is 2-avoidable. Thus, we arrive at the following.

Theorem 2.7.1. *Each tripled word is* 2*-avoidable.*

Chapter 3

Avoiding Doubled Words Simultaneously

3.1 Previous Results

Bean, Ehrenfeucht, and McNulty (1979) and Zimin (1982) obtained several early results of simultaneous avoidance of doubled words. For the first result, define the *mesh* of a word w as a value k in which, whenever x is a letter and u is a word in which x does not occur, if $|u| > k$, then xux is not a subword of w.

Theorem 3.1.1. *(Bean, Ehrenfeucht, and McNulty (1979), Theorem 1.12) The set of all doubled words on a denumerable alphabet of mesh k is $(8k + 16)$-avoidable.*

This result is useful in that it is not dependent on the size of the alphabet of the doubled words but simply how far apart the letters can be. Another desirable result would be to find a bound on the avoidability index of all doubled words that is dependent on their length. For the set of all doubled words on an alphabet of size at most n, Bean, Ehrenfeucht, and McNulty (1979) gives a bound of $8 \cdot 2^n + 16$. The work of Zimin Zimin (1982) implicitly contains a bound of $6 \cdot 2^n + 14$ on the avoidability index of this set, though he did not prove any specific theorems about doubled words.

Baker, McNulty, and Taylor (1989) gave two bounds on the avoidability index of the set of all avoidable words on an alphabet of size at most n. These naturally extend to bounds on the set of all doubled words on an alphabet of size at most n.

Theorem 3.1.2. *(Baker, McNulty, and Taylor (1989), Theorem 1.2) The set of all avoidable words on an alphabet of size at most n is $(4(n+2)\lceil\log(n+2)\rceil)$-avoidable.*

Theorem 3.1.3. *(Baker, McNulty, and Taylor (1989), Theorem 1.3) The set of all avoidable words on an alphabet of size at most n is $(9n + 20)$-avoidable.*

In Lothaire (2002), Cassaigne states and proves a sharper bound on the avoidability index of the set of all avoidable words on an alphabet of size at most n, which he attributes to Mel'nichuk.

Theorem 3.1.4. *(Lothaire (2002), Theorem 3.3.4) The set of all avoidable words on an alphabet of size at most n is $\left(4\lfloor \frac{n}{2} \rfloor + 4\right)$-avoidable.*

Finally, Mel'nichuk (1985) establishes the smallest known linear bound on the size of the alphabet that avoids the set of all doubled words on an alphabet of size at most n, which we present in the next section.

3.2 Mel'nichuk's Results

Theorem 3.2.1. *The set of all doubled words on an alphabet of size at most n is $\left(3\lfloor \frac{n}{2} \rfloor + 3\right)$-avoidable.*

Note 1: The value $m = 3\lfloor \frac{n}{2} \rfloor + 3$ above cannot be replaced by a number less than $n + 1$. If all doubled words are simultaneously avoided by an infinite set S of words on the n-letter alphabet, then all doubled words are simultaneously avoided by an infinite word U on the n-letter alphabet. We construct this word by forming a full m-ary tree rooted with the empty word. At each node, we construct the next level by branching into m new nodes, each labeled with one of the m letters. We consider the word represented by each node as the concatenation of the letters going up the branch to reach that node. Next, we remove any nodes whose represented word is not in S or is not a prefix of a word in S. Now, we know that this tree is still infinite since S is infinite, and since each node branches into finitely many new nodes as we go up the tree, by König's Infinite Lemma, there must be some infinite branch in the tree. Let our infinite word U be the concatenation of the letters going up the branch.

So, since U avoids every doubled word on an alphabet of size at most n, one letter ξ must appear only once in U. If all letters appear twice, then U contains a doubled word on the n-letter alphabet. However, the suffix of U that does not contain ξ still avoids all the doubled words on an alphabet of size at most n, and this word only has $n-1$ letters. Similarly, in order for this to still avoid all the doubled words on an alphabet of size at most n, it must have a suffix that still avoids all the doubled words on an alphabet of size at most n and has only $n-2$ letters. Continuing this process, we arrive at an infinite word U' on only one letter, and this clearly encounters a doubled word on an alphabet of size at most n. Thus, only finitely many words in the list avoid all the doubled words on an alphabet of size at most n for each infinite branch in the tree, so the list S is not actually infinite.

Note 2: Theorem 3.2.1 is also true when $n = 1$. The set of all doubled words on the 1-letter alphabet is $\{xx, xxx, xxxx, \dots\}$. Thue (1906) showed xx is 3-avoidable, and since all others in this list encounter xx, the set of all doubled words on the 1-letter alphabet is 3-avoidable. Finally, we note that $m = 3\lfloor \frac{1}{2} \rfloor + 3 = 3$ when $n = 1$.

Proof. Fix a positive integer $n > 1$, and let $m = 3\lfloor \frac{n}{2} \rfloor + 3$. We consider the m-letter alphabet $X = \{x_0, x_1, \dots, x_{m-1}\}$. In order to prove this theorem, we desire to generate an infinite list of words $J_0,\ J_1,\ \dots$ on X^+ that avoid all doubled words on an alphabet of size at most n. We do so by forming a homomorphism $\Psi : X \to X^+$ such that $J_0 = x_0$ and $J_{i+1} = \Psi(J_i)$ for $i \geq 0$. We focus our proof on showing the following assertion, of which Theorem 3.2.1 easily follows.

Assertion 3.2.2. *Let w be a doubled word on an alphabet of size at most n. If $k > 1$ and an image of w is a subword of J_{k+1}, then J_k contains an image of some doubled word u satisfying $\alpha(u) \subseteq \alpha(w)$, where $\alpha(u) \neq \varnothing$.*

In order to use Assertion 3.2.2, suppose that w_0 is a doubled word on an alphabet of size at most n that is encountered by J_{k+1} for some k. Then there is a doubled word w_1 on an alphabet of size at most n that is encountered by J_k, then a doubled word w_2 on an alphabet of size at most n that is encountered J_{k-1}, etc., and finally, there is a doubled word on an alphabet of size at most n that is encountered by J_0. However, $J_0 = x_0$, which cannot encounter a doubled word. Thus, no doubled word is encountered by J_k for any k, so the set of all doubled words is $\left(3\lfloor \frac{n}{2} \rfloor + 3\right)$-avoidable.

Proof of Assertion 3.2.2. Let w be a doubled word on an alphabet of size at most n, let k be such that J_{k+1} encounters w, and let φ be such that $\varphi(w)$ is a subword of J_{k+1}. We seek to create a word u from w such that J_k encounters u. Figure 3.1 illustrates the setup of this problem. In the figure, φ can be any function, and the images making up $\varphi(w)$ may not line up with the natural breaks in the application of Ψ to J_k. Thus, we need to carefully define Ψ so that φ can be modified into an erasing homomorphism φ' that maps the letters of w into the natural breaks between images of Ψ. Note that in some cases, we may need to add an additional copy of a letter in w in order to complete this definition of φ'. In doing so, we let u be the word w with any letters removed that are erased by φ' and any letters added that are required to complete the definition of φ'. Then J_k will encounter u via φ', as desired.

We define Ψ in the following manner. Let $a_0, a_1, \ldots, a_{m-1} \in X^+$ with each a_i being of length m with each letter in X occurring exactly once and satisfying the properties that will be listed below. Before stating these properties, we must state a few definitions.

The word $x_{i_p}x_{i_q}$ is said to be *basic* if there exists $j \in \{0, \ldots, m-1\}$ such that $x_{i_p}x_{i_q}$ is a subword of a_j and i_p is in a position that is $\equiv 0 \pmod 3$. In this case, $x_{i_p}x_{i_q}$ is said to be *associated* with a_j. If $a_i = b_1b_2$ and $a_j = c_1c_2$ for $b_1, b_2, c_1, c_2 \in X^+$, then the word b_2c_1 is called *adjacent*.

We now assume that the a_i's satisfy the following properties. After we prove the assertion under this assumption, we will carefully define a_i for all $i < m$ and prove that they satisfy these properties.

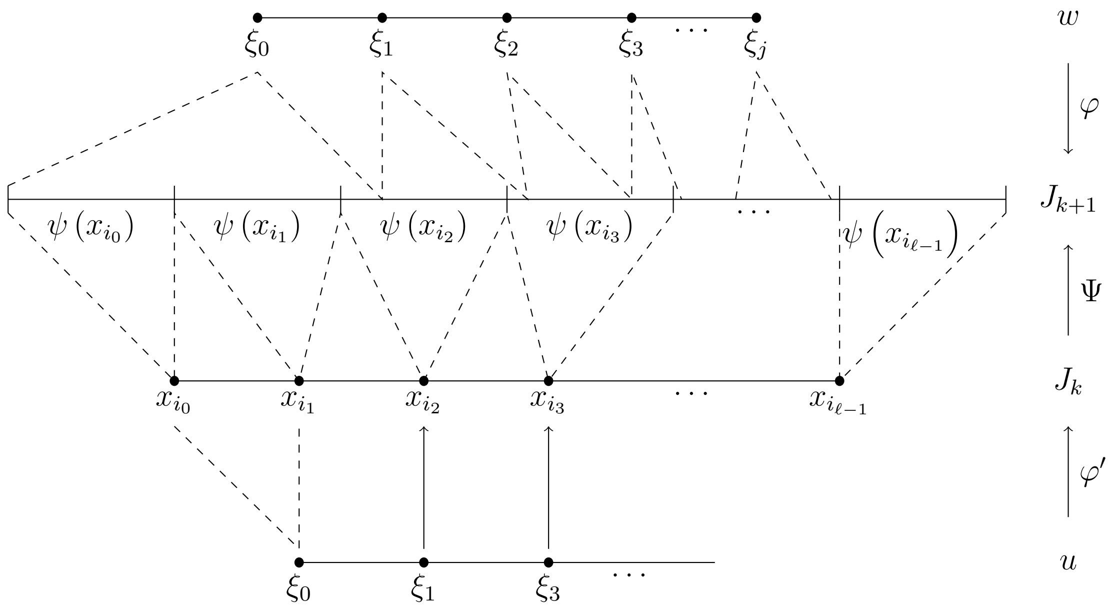

Figure 3.1 Example of the Method to Simultaneously Avoid All Doubled Words

(a) Each basic word is associated with a_i for only one i in the set $\{0, \dots, m-1\}$.

(b) The words a_i, a_j with $i \neq j$ do not contain identical subwords of length greater than 2.

(c) There are no adjacent words that are subwords of a_i for any $i \in \{0, \dots, m-1\}$.

(d) If $x_{i_p} x_{i_q}$ appears in a_i and a_j for $i \neq j$, then suppose $x_{i_p} x_{i_q}$ is a basic word associated with only one of a_i or a_j. If it is preceded by x_{i_r} where it is not associated, then x_{i_r} directly follows $x_{i_p} x_{i_q}$ in its associated word.

(e) If a subword of a_i has length $\equiv 0 \pmod 3$, begins in a position $\equiv 0 \pmod 3$, is composed of images of letters in w of length 1 or 2, and has more images of length 2 than length 1, then this subword contains a basic word.

(f) The word $a_i a_j$ with $i \neq j$ does not contain any image of any doubled word v as a subword.

Define $\Psi(x_i) = a_i$. We now seek to show that φ' and u can be created using this choice of Ψ.

Consider the set $W = \{a_0, \ldots, a_{m-1}\}$ as an alphabet, and consider the words over W, which we will call *chains*. Thus, a word a_i regarded in X^* has length m, but as a chain in W^*, it has length 1. Each chain over W^* naturally corresponds to a word in X^*. A word v in W^* that is a subword of a chain C is called a *subchain* of C, and if this v is an occurrence of a_i in C, we call it a *link*. We say that the chains C and D are *graphically equal* if they are graphically equal as words. Note that 2 links being graphically equal does not mean that they are the same link.

Suppose we have an occurrence of the word v of X^+ in the chain $C = a_{i_0} \ldots a_{i_{p-1}}$, where $p > 1$ and v is not contained in a single link. Then the word C can be written as $C = a_{i_0} \ldots a_{i_r} c_1 v c_2 a_{i_m} \ldots a_{i_{p-1}}$, where $c_1, c_2 \in X^*$, where $|c_1|$ and $|c_2| < m$, where c_1 is an initial segment of $a_{i_{r+1}}$, and where c_2 is a final segment of $a_{i_{m-1}}$. Let $\mathcal{C}(v)$ be the smallest subchain $a_{i_{r+1}} \ldots a_{i_{m-1}}$ of the chain C that contains an occurrence of v. Let $\tau_1(v)$ and $\tau_2(v)$ denote the words of X^+ such that $c_1\tau_1(v)$ is the first link of the chain $\mathcal{C}(v)$ and $\tau_2(v)\, c_2$ is the last link of this chain. In this manner, $\tau_1(v)$ is an initial segment of v and $\tau_2(v)$ is a final segment of v. Note: If v starts a link, we do not define $\tau_1(v)$, and if v ends a link, we do not define $\tau_2(v)$. Finally, if v is contained in a single link, we do not define $\tau_1(v)$ and $\tau_2(v)$.

Let $\mathcal{C}$ be the smallest subchain of J_k containing $\varphi(w)$. By Property (f), the chain $\mathcal{C}$ contains more than two links. For each letter ξ of $\alpha(w)$, we consider each of its occurrences, where $\xi^{(p)}$ denotes the p^{th} occurrence of ξ. Let $\mathcal{C}_p(\xi)$ denote the smallest subchain of $\mathcal{C}$ that contains $\varphi\left(\xi^{(p)}\right)$.

Suppose $\mathcal{C}_p(\xi)$ has length at least 2 links for some $\xi \in \alpha(w)$. Let T_1 denote the set of links a_i of $\mathcal{C}$ such that a_i is the first link in $\mathcal{C}_p(\xi)$, and define $\tau_R(a_i)$ to be $\tau_1\left(\varphi\left(\xi^{(p)}\right)\right)$. Let T_2 denote the set of links a_i of $\mathcal{C}$ such that a_i is the last link in $\mathcal{C}_p(\xi)$, and define $\tau_L(a_i)$ to be $\tau_2\left(\varphi\left(\xi^{(p)}\right)\right)$.

Note: Property (c) implies that for any two occurrences, say the p^{th} and r^{th}, of any letter ξ of $\alpha(w)$, the following holds:

(1) Chains $\mathcal{C}_p(\xi)$ and $\mathcal{C}_r(\xi)$ are the same length. If this were not the case, then one chain could be at most one link longer by shifting the image of ξ. Suppose without loss of generality that $\mathcal{C}_r(\xi)$ has an additional link to the right. Then the occurrence $\varphi\left(\xi^{(p)}\right)$ must be shifted right when considering the occurrence $\varphi\left(\xi^{(r)}\right)$. Consider the word u that consists of the last letter in the next to last link of $\mathcal{C}_r(\xi)$ and the first letter of its last link. Then this word appears in the last link of $\mathcal{C}_p(\xi)$, contradicting Property (c).

(2) If the chain $\mathcal{C}_p(\xi)$ contains more than one link, then $\tau_1\left(\varphi\left(\xi^{(p)}\right)\right) = \tau_1\left(\varphi\left(\xi^{(r)}\right)\right)$ and $\tau_2\left(\varphi\left(\xi^{(p)}\right)\right) = \tau_2\left(\varphi\left(\xi^{(r)}\right)\right)$. If this were not the case, then the image of ξ would be shifted like in the previous claim, and this would again contradict Property (c).

For the p^{th} occurrence of the letter ξ in the word w define $\mathrm{Cl}_p(\xi)$, the closure of the p^{th} occurrence of ξ, a subchain of $\mathcal{C}$, as follows.

(1) In the case when $\mathcal{C}_p(\xi) = a_j$,

$$\mathrm{Cl}_p(\xi) = \begin{cases} a_j & \text{if } |\varphi(\xi)| \geq 3 \text{ or } \varphi(\xi) \text{ is a basic word} \\ \epsilon & \text{otherwise.} \end{cases}$$

(2) In the case when $\mathcal{C}_p(\xi) = a_{j_0} \dots a_{j_{q-1}}$,

$$\mathrm{Cl}_p(\xi) = \begin{cases} a_{j_0} \dots a_{j_{q-1}} & \text{if } |\tau_R(a_{j_0})| \geq 3 \text{ and } \left|\tau_L\left(a_{j_{q-1}}\right)\right| \geq 2 \\ a_{j_1} \dots a_{j_{q-1}} & \text{if } |\tau_R(a_{j_0})| < 3 \text{ and } \left|\tau_L\left(a_{j_{q-1}}\right)\right| \geq 2 \\ a_{j_0} \dots a_{j_{q-2}} & \text{if } |\tau_R(a_{j_0})| \geq 3 \text{ and } \left|\tau_L\left(a_{j_{q-1}}\right)\right| < 2 \\ a_{j_1} \dots a_{j_{q-2}} & \text{otherwise.} \end{cases}$$

Lemma 3.2.3. *For all $\xi \in \alpha(w)$, its p^{th} and r^{th} occurrences have the property that $Cl_p(\xi) = Cl_r(\xi)$.*

Proof. By the note above, $\mathcal{C}_p(\xi)$ and $\mathcal{C}_r(\xi)$ are the same length, say $q > 1$ links. Suppose that $\mathcal{C}_p(\xi) = a_{i_0} \dots a_{i_{q-1}}$ and $\mathcal{C}_r(\xi) = a_{j_0} \dots a_{j_{q-1}}$. By the note above, $\tau_R(a_{i_0}) = \tau_R(a_{j_0})$ and $\tau_L\left(a_{i_{q-1}}\right) = \tau_L\left(a_{j_{q-1}}\right)$. Thus, $\mathrm{Cl}_p(\xi)$ and $\mathrm{Cl}_r(\xi)$ are defined by the same case in its definition. In particular, if $q > 2$, since $\varphi(\xi)$ is the same regardless of which encounter, we see that $a_{i_1} = a_{j_1}, \dots, a_{i_{q-2}} = a_{j_{q-2}}$. Considering each case in the definition, $a_{i_0} = a_{j_0}$ if $|\tau_R(a_{i_0})| \geq 3$ by Property (b), and $a_{i_{q-1}} = a_{j_{q-1}}$ if $\left|\tau_L\left(a_{j_{q-1}}\right)\right| \geq 2$ by Property (a). These parts are omitted in the definition of $\mathrm{Cl}_p(\xi)$ whenever the appropriate sizes of τ_R or τ_L are not reached, so $\mathrm{Cl}_p(\xi) = \mathrm{Cl}_r(\xi)$. If $q = 2$, the same arguments work, even if the center is empty.

If $\mathcal{C}_p(\xi) = \mathrm{Cl}_p(\xi) = a_i$ and $\mathcal{C}_r(\xi) = a_j$, then $\varphi(\xi)$ is either a basic word or has length at least 3. If $\varphi(\xi)$ has length at least 3, then $a_i = a_j$ by Property (b). Thus, $\mathrm{Cl}_p(\xi) = \mathrm{Cl}_r(\xi)$. If $\varphi(\xi)$ is a basic word associated with a_i, then suppose that $\varphi(\xi)$ is not associated with a_j. Then by Property (d), we have that the letter x_{i_s} that comes after $\varphi\left(\xi^{(p)}\right)$ must come before $\varphi\left(\xi^{(r)}\right)$. Thus, $\varphi\left(\xi^{(p)}\right)$ is preceded by x_{i_s} and followed by x_{i_s}, and since it is basic and x_{i_s} cannot be repeated in any one a_i, this means that $x_{i_s}\varphi(\xi)$ is an adjacent word. Finally, this word must appear in a_j since $\varphi\left(\xi^{(r)}\right)$ is not basic, contradicting Property (c). Thus, $\varphi(\xi)$ is associated with a_j, so by Property (a), $a_i = a_j$.

Let $\mathcal{C}_p(\xi) = a_i$ and $\mathrm{Cl}_p(\xi) = \epsilon$. If $\mathcal{C}_r(\xi) = a_j$, then $\varphi(\xi)$ has length at most 2 and is not a basic word associated with x_i. If $\varphi(\xi)$ is a basic word associated with x_j, we arrive at the same contradiction as before. Thus, $\mathrm{Cl}_r(\xi) = \epsilon$ as well. $\square$

With this lemma in place, we will now drop the subscripts and simply refer to $\mathrm{Cl}(\xi)$ as the closure of ξ. Note also by this lemma that if an image of a letter ξ is basic in some a_i, then it is basic in every a_i in which it appears.

The word a in X^+ is called *composite* if a satisfies the following:

(1) The word a is a subword of a_i for some $i \in \{0, \ldots, m-1\}$, and

(2) The word $a = \varphi(\xi_{i_0}) \ldots \varphi\left(\xi_{i_{r-1}}\right)$, and for each $j \in \{0, \ldots, r-1\}$, we have that $\left|\varphi\left(\xi_{i_j}\right)\right| \leq 2$ and $\varphi\left(\xi_{i_j}\right)$ is not basic if $\left|\varphi\left(\xi_{i_j}\right)\right| = 2$. Note that being not basic in one location requires that it not be basic in any location by the note above.

Lemma 3.2.4. *The length of any composite word is less than* $3\lfloor \frac{n}{2} \rfloor$.

Proof. We begin by showing by induction on k that J_k avoids xx. For the base cases, we have $k = 0$ and $k = 1$. Note that $J_0 = x_0$ and clearly avoids xx. Also, $J_1 = a_0$, and a_0 avoids xx because it has distinct letters. Now, assume that J_k avoids xx, and suppose that UU appears as a subword of J_{k+1}. Let v be a minimal length subword of J_k such that the chain $\Psi(v)$ contains UU. By Property (f), $\Psi(v)$ contains at least 3 links, and hence $|v| \geq 3$. Thus, since v is minimal, UU sits in at least 3 links. Observe that $|U| \geq m$ since each link avoids xx. Thus, some consecutive two of the links of $\Psi(v)$ contains the same subword of U that has at least 3 characters, so by Property (b), these two links are graphically equal. Hence, v encounters xx, contradicting the inductive hypothesis. So, J_k avoids xx for all k.

Now, let a be a composite subword of a_i with $a = \varphi(\xi_{i_0}) \ldots \varphi\left(\xi_{i_{r-1}}\right)$. We denote the set of letters $\left\{\xi_{i_0}, \ldots, \xi_{i_{r-1}}\right\}$ that make up a by M, the number of letters ξ of M for which $|\varphi(\xi)| = 1$ by n_1, and the number of letters of M for which $|\varphi(\xi)| = 2$ by n_2. Suppose for the sake of contradiction that $|a| \geq 3\lfloor \frac{n}{2} \rfloor$. Then the following inequalities must be true:

(1) $n_1 + n_2 \leq n$

(2) $n_1 + 2n_2 \geq 3\lfloor \frac{n}{2} \rfloor$

(3) $n_2 \leq \lfloor \frac{n}{2} \rfloor + 1$.

For (1), note that $r = n_1 + n_2$. Since a is composed of images of letters in w, we know that $r \leq n$. For (2), we know by definition that $|a| = n_1 + 2n_2$, so (2) holds since we're supposing $|a| \geq 3\lfloor \frac{n}{2} \rfloor$.

For (3), this uses Property (e), but a doesn't necessarily start in a position of a_i that is $\equiv 0 \pmod 3$. Suppose n_2 is at least 3 more than n_1. If a starts in a position of a_i that is $\equiv 0 \pmod 3$, then the proof is done by Property (e). If a starts in a position $\equiv 2 \pmod 3$, then starting with an image of length 1 brings us back to Property (e). If it starts with an image of length 2, then having another image of length 2 takes us back to a position of $\equiv 2 \pmod 3$, and we still have more images of length 2 than images of length 1 and can finish with Property (e). If it starts with an image of length 2 and then an image of length 1, we are done by an easy induction argument. If a starts in a position $\equiv 1 \pmod 3$, then starting with an image of length 2 yields Property (e), and starting with an image of length 1 brings us back to the case of a in a position $\equiv 2 \pmod 3$ with even more images of length 2 than images of length 1.

So, with an additional 3 images of length 2 than images of length 1, this forces a basic word to appear. If $n_2 = \lfloor \frac{n}{2} \rfloor + 2$, then $n_1 \leq \lceil \frac{n}{2} \rceil - 2$ by (1), and having 4 more 2's than 1's will result in one of the images being a basic word. Thus, $n_2 \leq \lfloor \frac{n}{2} \rfloor + 1$.

Suppose that $n_2 \leq \lfloor \frac{n}{2} \rfloor - 1$, and consider $n_2 = \lfloor \frac{n}{2} \rfloor - 1 - g$ for $g \geq 0$. If n is even, then by (1),

$$n_1 \leq n - n_2 = n - \frac{n}{2} + 1 + g = \frac{n}{2} + 1 + g.$$

Similarly, by (2),

$$n_1 \geq \frac{3n}{2} - 2\left(\frac{n}{2} - 1 - g\right) = \frac{n}{2} + 2 + 2g.$$

However, these inequalities contradict each other. If n is odd, then by (1),

$$n_1 \leq n - \lfloor \frac{n}{2} \rfloor + 1 + g = \lfloor \frac{n}{2} \rfloor + 2 + g,$$

and by (2),

$$n_1 \geq 3\lfloor\frac{n}{2}\rfloor - 2\left(\lfloor\frac{n}{2}\rfloor - 1 - g\right) = \lfloor\frac{n}{2}\rfloor + 2 + 2g.$$

If $g > 0$, these inequalities contradict each other. If $g = 0$, then $n_1 = \lfloor\frac{n}{2}\rfloor + 2$, so

$$n_1 + n_2 = \lfloor\frac{n}{2}\rfloor + 2 + \lfloor\frac{n}{2}\rfloor - 1 = 2\lfloor\frac{n}{2}\rfloor + 1 = n.$$

Thus, $r = n$, so by Property (f),

$$n_1 + 2n_2 > m = 3\lfloor\frac{n}{2}\rfloor + 3.$$

However,

$$n_1 + 2n_2 = \lfloor\frac{n}{2}\rfloor + 2 + 2\left(\lfloor\frac{n}{2}\rfloor - 1\right) = 3\lfloor\frac{n}{2}\rfloor,$$

contradicting the previous sentence. So, $n_2 > \lfloor\frac{n}{2}\rfloor - 1$.

Suppose $n_2 = \lfloor\frac{n}{2}\rfloor$. Then $n_1 \leq \lceil\frac{n}{2}\rceil$ by (1) and $n_1 \geq \lfloor\frac{n}{2}\rfloor$ by (2). We wish to show that $n_1 = \lfloor\frac{n}{2}\rfloor$. If n is even, then this is true. If n is odd, then suppose that $n_1 = \lceil\frac{n}{2}\rceil$. Then $r = n$ again, and by (2),

$$n_1 + 2n_2 = \lceil\frac{n}{2}\rceil + 2\lfloor\frac{n}{2}\rfloor = 3\lfloor\frac{n}{2}\rfloor + 1.$$

Since this is not at least m, this contradicts Property (f). Thus, $n_1 = \lfloor\frac{n}{2}\rfloor$, so if $r = n$, then the proof is the same as above. So, suppose that there is one remaining letter $\xi \in \alpha(w)$ such that $\xi \notin M$. If $|\varphi(\xi)| \leq 2$, then the proof is the same as above. So, suppose that $|\varphi(\xi)| \geq 3$, and let $\mathcal{C}_1(\xi) = a_{j_0} \dots a_{j_{q-1}}$. Note that the letters named in n_1 and n_2 have images whose lengths total to $m - 3$.

If $q = 1$, then since all other letters of w total $m - 3$, two consecutive links contain $\varphi(\xi)$. By Property (b), this says that two consecutive are graphically equal, contradicting the fact that J_{k+1} avoids xx. If $q > 2$, then by Properties (b) and (c), all the links, except maybe the first and last, of $\mathcal{C}_1(\xi)$ and $\mathcal{C}_2(\xi)$ must be graphically equal. In addition, this gives that $\tau_R(a_{j_0})$ and $\tau_L\left(a_{j_{q-1}}\right)$ are the same for both $\varphi\left(\xi^{(1)}\right)$ and $\varphi\left(\xi^{(2)}\right)$, if they exist. If you were to shift the image of ξ, this would force an adjacent word to appear in a link, contradicting Property (c). Consider the link in which a appears, and without loss of generality, assume that a comes from the first occurrences of its associated letters in w and a appears in the link x. We consider the letters that appear in w after the associated letters in w. If the first letter is not ξ, then a is at the end of the link x in order to avoid having the same letter appear twice in x. If ξ appears at the beginning of w, then the last links of $\mathcal{C}_1(\xi)$ and $\mathcal{C}_2(\xi)$ are graphically equal by Property (b). We have that $\varphi(\xi)$ must begin in the link after x since all the other letters total to $m-3$. We also have that the first links of $\mathcal{C}_1(\xi)$ and $\mathcal{C}_2(\xi)$ are graphically equal by Property (b). This gives an encounter of xx, so we will instead assume that ξ does not appear at the beginning of w. However, we still have that the first links of $\mathcal{C}_1(\xi)$ and $\mathcal{C}_2(\xi)$ are graphically equal by Property (b). If the last links are graphically equal again, we arrive at a contradiction, so assume that $\varphi(\xi)$ extends into its last link with length 1. Then $\varphi(\xi)$ must begin again in the same link where it ended, giving the contradiction once more.

So, we have shown that the letter ξ must appear in w directly after the letters associated with a. If $|\tau_R(x)| = 3$, then the first links of $\mathcal{C}_1(\xi)$ and $\mathcal{C}_2(\xi)$ are again graphically equal. In order to avoid the contradiction once again, we need the last links to not be graphically equal, so we again assume that $\varphi(\xi)$ extends into its last link with length 1. If $\varphi(\xi)$ begins again in this link, we are finished, so ξ must be at the beginning of w. However, a starts the link, stopping $\varphi(\xi)$ from extending 1 into that link. If $\tau_R(x) = 1$, then the last links of $\mathcal{C}_1(\xi)$ and $\mathcal{C}_2(\xi)$ are graphically equal. If $\varphi(\xi)$ starts in the same link, then we again get an encounter of xx. So, ξ must be the first letter in w, but this is enough to again get an encounter of xx by not including the beginning link of the first occurrence of ξ. Finally, if $|\tau_R(x)| = 2$, then $\varphi(\xi)$ cannot extend past its last link more than one to still avoid xx. We note that if $\varphi\left(\xi_{i_j}\right)$ has length 2 and appears in another link, then it cannot be basic. Thus, by Property (d), any two links containing the same image of length 2 must be graphically equal in this setup. Thus, since $m-3$ characters must be filled in the link after x before ξ can appear again, there must be an image of length 2. Hence, x and the link after it are graphically equal, giving another contradiction. So, $n_2 \neq \lfloor \frac{n}{2} \rfloor$, meaning that $n_2 = \lfloor \frac{n}{2} \rfloor + 1$.

With this in place, we get $n_1 \leq \lceil \frac{n}{2} \rceil - 1$ by (1) and $n_1 \geq \lfloor \frac{n}{2} \rfloor - 2$ by (2). If $n_1 = \lceil \frac{n}{2} \rceil - 1$, then $r = n$ again and $n_1 + 2n_2 = \lceil \frac{n}{2} \rceil - 1 + 2\lfloor \frac{n}{2} \rfloor + 2 \leq 3\lfloor \frac{n}{2} \rfloor + 2 < m$, contradicting Property (f). If $n_1 = \lfloor \frac{n}{2} \rfloor - 2$, then there are at least 3 more images of length 2 than length 1, contradicting the fact that a is composite by the earlier argument. If $n_1 = \lfloor \frac{n}{2} \rfloor - 1$ and n is even, we arrive at $r = n$ and the same contradiction of Property (f). If $n_1 = \lfloor \frac{n}{2} \rfloor - 1$ and n is odd, then $n_1 + n_2 = n - 1$, meaning that there is one remaining letter $\xi \in \alpha(w)$ such that $\xi \notin M$. If $|\varphi(\xi)| \leq 1$, then the proof is the same as above since $n_1 + 2n_2 = 3\lfloor \frac{n}{2} \rfloor + 1$. So, suppose that $|\varphi(\xi)| \geq 2$, and let $\mathcal{C}_1(\xi) = a_{j_0} \ldots a_{j_{q-1}}$. Note that the letters named in n_1 and n_2 have images whose lengths total to $m - 2$, and the remainder of the proof uses the same cases as above. $\square$

Recall that T_1 consists of the a_i's that show up as a first link of some occurrence of $\varphi(\xi)$ for some $\xi \in \alpha(w)$ and T_2 consists of the a_i's that show up as a final link of some other occurrence of $\varphi(\xi)$ for some $\xi \in \alpha(w)$ (not necessarily the same ξ). Let $T = T_1 \cap T_2$.

Graphical equality on the set T is an equivalence relation. We let $[x]$ denote the equivalence class in T over graphical equality that contains the link x. Note that two links $x, y \in [x]$ may not necessarily have the same values for $\tau_L(x)$ and $\tau_R(x)$ due to the fact that two graphically equal links are not the same link. Graphical equality of two links simply means that the two links are the same when considered as words. Let S_1 be a subset of T consisting of links that are a closure for some occurrence of some letter ξ in w. In other words, if $x \in S_1$, since $x = a_i$ for some i, we know that x either (1) contains either $\varphi(\xi)$ as a basic word or with length at least 3 or (2) $\varphi(\xi)$ extends past x by no more than 2 on the left and 1 on the right. Assume that $[S_1] = \bigcup_{x \in S_1} [x]$. Next, let $T' = T \setminus [S_1]$. We now define by the following relation on elements of T': For links $x, y \in T'$, we say that $x \sim y$ if $\tau_L(x) = \tau_L(y)$ or $\tau_R(x) = \tau_R(y)$. Note that this relation is reflexive and symmetric but not generally transitive.

Let S_2 and S_3 be subsets of T' such that $S_2 = \{x \mid x \in T' \text{ and } |\tau_L(x)| = 1\}$, and $S_3 = \{x \mid x \in T' \text{ and } |\tau_R(x)| \leq 2\}$. Note that if $x \in S_2 \cap S_3$, then $x = \tau_L(x)\, a \tau_R(x)$, where $|a| \geq 3\lfloor \frac{n}{2} \rfloor$. Since $x \notin S_1$, this says that $\varphi\left(\xi^{(p)}\right)$ appears in x with $\left|\varphi\left(\xi^{(p)}\right)\right| \leq 2$ and $\varphi\left(\xi^{(p)}\right)$ not basic for any ξ and any p. Thus, a is composite, contradicting Lemma 3.2.4, and hence $S_2 \cap S_3 = \varnothing$. We now define sets P_i recursively in order to finish our definition of $\operatorname{Ind}(x)$.

STEP 0: Let $P_0 = S_2 \cup S_3$.

By definition, each link of $T' \setminus P_0$, since it is neither an element of S_2 nor S_3, can be written as $x = b_1 b_2 b_3$, where $|b_1| \geq 2$, $|b_3| \geq 3$, $b_1 = \tau_L(x)$, $b_3 = \tau_R(x)$, and b_2 is either composite or empty. Note that if $x \in T' \setminus P_0$, $y \in T'$, and $x \sim y$, we have that x and y are graphically equal by Properties (a) and (b). However, this does not imply that $\tau_L(x) = \tau_L(y)$ and $\tau_R(x) = \tau_R(y)$.

STEP $i+1$: Denote

$$P_{i+1} = \{x \mid x \in T' \setminus (P_0 \cup \cdots \cup P_i) \text{ and there exists } y \in P_i \text{ such that } x \sim y\}.$$

The set T is finite, so for some i we have $P_i = \varnothing$. For this i, denote $P = P_0 \cup \cdots \cup P_i$. Next, we describe a relation, $\operatorname{Ind}(x)$, from the set T into the set $\{1, 2, 3\}$ as follows:

$$\operatorname{Ind}(x) = \begin{cases} 1 & \text{if } x \in [S_1] \\ 2 & \text{if } x \in S_2 \text{ or } x \in T' \setminus P \\ 3 & \text{if } x \in S_3 \\ \operatorname{Ind}(y) & \text{if } x \in P_{i+1} \text{ and } y \in P_i \text{ with } x \sim y. \end{cases}$$

We now determine a few properties about this relation in order to show that it is actually a function.

Suppose that $a_i = b_1 b_2 b_3 b_4$, where $b_1, b_4 \in X^*$ and $b_2, b_3 \in X^+$. Then the words b_2 and b_3 are said to be *adjacent*. If $a_j = c_1 c_2 c_3$, where $c_1, c_2, c_3 \in X^*$, and either $|c_1| = 1$ or $|c_3| \leq 2$, then the word c_2 is called the *middle*.

Lemma 3.2.5. *If $Ind(x) = 2$, then x has a middle subword adjacent to $\tau_R(x)$ that is composite or empty. If $Ind(x) = 3$, then x has a middle subword adjacent to $\tau_L(x)$ that is composite or empty. Further, for all $x \in P_{i+1}$, if $x \sim y$ and $x \sim z$ for $y, z \in P_i$, then $Ind(y) = Ind(z)$.*

Note that the subword being composite here does not mean that the word satisfies the composite properties within the link x but in some other link. Also, this means that if the middle subword is empty, then $x \in P_0$.

Proof. We will use induction on the formulation of P_i to prove the first claim. By the construction of P_0, if $x \in P_0$, then $|\tau_R(x)| \leq 2$ or $|\tau_L(x)| = 1$ and $x \notin [S_1]$. Thus, x is not a closure of any image of any letter, so $x = \tau_L(x)\, m \tau_R(x)$, where m is composite or empty. Then m satisfies the lemma since $|\tau_R(x)| = 1$ or $|\tau_L(x)| \leq 2$.

Suppose, for steps up to i the lemma holds. Let $x \in P_{i+1}$, and suppose that $x \sim y$ for $y \in P_i$ such that $\text{Ind}(y) = 2$. Then $\tau_L(x) = \tau_L(y)$ or $\tau_R(x) = \tau_R(y)$. Since $x \in T' \backslash P_0$, we know that $|\tau_R(x)| > 2$ and $|\tau_L(x)| > 1$. By the induction hypothesis, y has a middle subword a' adjacent to $\tau_R(y)$ that is either composite or empty. Note again that if a' is composite, it may not satisfy the composite properties in y, but by the definition of composite, it must satisfy these properties in some link.

If $\tau_R(x) = \tau_R(y)$, then x is graphically equal to y by Property (b). Hence, a' is a middle subword adjacent to $\tau_R(x)$ that is either composite or empty. Let $\tau_L(x) = \tau_L(y)$. Then x is again graphically equal to y by Properties (a) or (b). If $|\tau_R(x)| \geq |\tau_R(y)|$, note that $x = \tau_L(x)\, m \tau_R(x)$, where m is composite or empty.

Then m is a subword of a', so consider $a' = m'mm''$, where m' is a suffix of $\tau_L(x)$ and m'' is a possibly empty prefix of $\tau_R(x)$. Then $m'm$ satisfies the lemma. If $|\tau_R(x)| < |\tau_R(y)|$, then denote by a'' the subword of x such that $\tau_R(y) = a''\tau_R(x)$. Then a'' is composite since $x \notin [S_1]$, and hence, $a'a''$ satisfies the lemma.

Now, suppose that $x \sim z$ for $z \in P_i$ such that $\operatorname{Ind}(z) = 3$. Then by the induction hypothesis, z has a middle subword adjacent to $\tau_L(z)$ that is either composite or empty, and using similar reasoning as that above, x has a middle subword adjacent to $\tau_L(x)$ that is either composite or empty, finishing the second part of the lemma.

Finally, we wish to show the second part of the lemma by contradiction. Let $x \in P_{i+1}$, and let $x \sim y$ and $x \sim z$ with $y, z \in P_i$. Suppose that $\operatorname{Ind}(y) = 2$, and $\operatorname{Ind}(z) = 3$. Let m_1 be the middle composite subword of x adjacent to $\tau_L(x)$ and m_2 be the middle composite subword of x adjacent to $\tau_R(x)$. Note that m_1 and m_2 are not empty since $x \notin P_0$. Consider the word $m_1'mm_2'$, where $m_1 = m_1'm$ and $m_2 = mm_2'$ for m possibly empty. If m is empty, then m_1m_2 is composite and has length at least $3\lfloor \frac{n}{2} \rfloor$, contradicting Lemma 3.2.4. If m is nonempty, then $m_1'mm_2'$ has length at least $3\lfloor \frac{n}{2} \rfloor$, and in order for m_1 and m_2 to be composite, we must have at most 2 additional images of length 2 than images of length 1. In order to have this, we must use at least $n-2$ letters to make m_1 and m_2, so there can be no images that go between m_1' and m or m and m_2'. If such an image appeared, then we would only have one image to make $\tau_R(x)$ and $\tau_L(x)$. By Property (c), any image that crosses a boundary between links must always appear in the same place. With only one image, all links have the same τ_R and τ_L, so in particular, J_k would encounter xx. Thus, $m_1'mm_2'$ is composite and has length at least $3\lfloor \frac{n}{2} \rfloor$, contradicting Lemma 3.2.4. Hence, $\operatorname{Ind}(y) = \operatorname{Ind}(z)$. □

One particular consequence of this lemma is that $\operatorname{Ind}(x)$ is actually a function. It is clear that $\operatorname{Ind}(x)$ is well-defined for $x \in T' \backslash P$ or $x \in [S_1]$. It is also clear since $S_2 \cap S_3 = \varnothing$ that if $x \in P_0$, then $\operatorname{Ind}(x)$ is well-defined. Then if we assume $\operatorname{Ind}(y)$

is well-defined for all $y \in P_i$, then if $x \in P_{i+1}$, we know that any $z \in P_i$ such that $x \sim z$ will have the same index. Thus, $\mathrm{Ind}(x)$ is well-defined for $x \in P_{i+1}$, and hence $\mathrm{Ind}(x)$ is well-defined for all $x \in P$ by induction.

We now have enough properties built to create u. We will consider the closures of each letter in $\alpha(w)$, parse them to remove overlaps without losing any links in $\mathcal{C}$ (except maybe the first and last), and then associate these parsed closures with one of the letters of $\alpha(w)$. First, we need the following claim:

Claim 3.2.6. *Every link x of $\mathcal{C}$, except maybe the first and the last link, occur in closures of $\varphi(\xi)$ for some occurrence of some $\xi \in \alpha(w)$.*

Proof. Consider some link x of $\mathcal{C}$ that is not the first or last link, and suppose that x does not occur in the closure of any occurrence of any letter. Let $\xi \in \alpha(a)$, and consider the r^{th} occurrence of ξ in w. If $\mathcal{C}_r(\xi) = x$ and $\mathrm{Cl}_r(\xi) = \epsilon$, then $|\varphi(\xi)| \leq 2$ and $\varphi(\xi)$ is not a basic word. If x is part of $\mathcal{C}_r(\xi)$ but $\mathrm{Cl}_r(\xi)$ does not include it, then x must lie on the end of $\mathcal{C}_r(\xi)$ and either $|\tau_R(x)| \leq 2$ or $|\tau_L(x)| \leq 1$. With this in mind, any part of an image of any ξ that appears in x must have length at most 2, not be a basic word, and the first image or part of the image of a letter that appears in x must have length 1. Thus, x is composite and has length $3\lfloor \frac{n}{2} \rfloor + 3$, contradicting Lemma 3.2.4. $\square$

Let the chain that contains all the centers of the occurrences of ξ in w for all $\xi \in \alpha(w)$ be denoted by $\mathrm{Cl}(w)$. Then $\mathrm{Cl}(w)$ is identical to $\mathcal{C}$ except possibly missing the first or last link. Since $\mathcal{C}$ contains 3 links by Property (f), $\mathrm{Cl}(w)$ is not empty. Define

$$M_0(\mathcal{C}) = \{y \mid y \text{ is a subchain of } \mathcal{C} \text{ such that } \mathrm{Cl}(\xi) = y \text{ for some } \xi \in \alpha(w)\}.$$

Note that any two subchains y_i and y_j of $M_0(\mathcal{C})$ with $i \neq j$ can have at most one common link. If it had two common links, then the first link would be filled with part of the image of its respective ξ but also would have to contain some part of the image of the respective ξ of the other link. Also, note that every link of $\mathrm{Cl}\,(w)$ is a link of some chain in $M_0\,(\mathcal{C})$ by the definition of $\mathrm{Cl}\,(w)$ and the fact that every link except maybe the first and last links of $\mathcal{C}$ are links in $\mathrm{Cl}\,(w)$. Finally, note that if a link x is common for two subchains y_i and y_j of $M_0\,(\mathcal{C})$ with $i \neq j$, then the link $x \in T$. Let $M\,(\mathcal{C})$ be the set of all subchains of $\mathcal{C}$, including the chain of length 0 (denoted E). Construct mappings $\psi_0 : M_0\,(\mathcal{C}) \to M\,(\mathcal{C})$ as follows.

(1) Suppose that $y = a_i$. Let

$$\psi_0\,(y) = \begin{cases} a_i & \text{if } a_i \in [S_1] \text{ or } a_i \notin T_1 \cup T_2 \\ E & \text{if } a_i \in T_1 \backslash T_2 \text{ or } a_i \in T_2 \backslash T_1. \end{cases}$$

Note: If $a_i \notin [S_1]$, then $a_i \notin T$.

(2) Suppose $y = a_{i_0} \ldots a_{i_{q-1}}$ where $q > 1$. Let

$$\psi_0\,(y) = \begin{cases} a_{i_1} \ldots a_{i_{q-2}} & \text{if } \mathrm{Ind}\,(a_{i_0}) \in \{1,2\} \text{ and } \mathrm{Ind}\left(a_{i_{q-1}}\right) \in \{1,3\} \\ a_{i_1} \ldots a_{i_{q-1}} & \text{if } \mathrm{Ind}\,(a_{i_0}) \in \{1,2\}\,, \text{ and } a_{i_{q-1}} \notin T \text{ or } \mathrm{Ind}\left(a_{i_{q-1}}\right) = 2 \\ a_{i_0} \ldots a_{i_{q-2}} & \text{if } a_{i_0} \notin T \text{ or } \mathrm{Ind}\,(a_{i_0}) = 3, \text{ and } \mathrm{Ind}\left(a_{i_{q-1}}\right) \in \{1,3\} \\ a_{i_0} \ldots a_{i_{q-1}} & \text{otherwise.} \end{cases}$$

Claim 3.2.7. *Two chains in $\psi_0\,(M_0\,(\mathcal{C}))$ have no common links, and every link in $Cl\,(w)$ appears in $\psi_0\,(M_0\,(\mathcal{C}))$.*

Proof. To show that there are no common links, let $y, z \in M_0\,(\mathcal{C})$, and let $y' = \psi_0\,(y)$ and $z' = \psi_0\,(z)$. Suppose that y' and z' share a common link x, and without loss of generality suppose that x is the last link of y' and the first link of z'. By above, note that x must be unique since y and z can share at most one common link. Let $y = a_{i_0} \ldots a_{i_{p-1}}$ and $z = a_{j_0} \ldots a_{j_{q-1}}$ with $p, q > 1$. Then $x = a_{i_{p-1}} = a_{j_0}$. Then since $x = a_{i_{p-1}}$, we have that $x \notin T$ or $\mathrm{Ind}\,(x) = 2$. Since $x = a_{j_0}$, we know that $x \notin T$

or $\text{Ind}(x) = 3$. Thus, since $\text{Ind}(x)$ maps to a single value, $x \notin T$, but by the above note, since x is common to y and z, we know that $x \in T$. If $y = a_{i_0} \dots a_{i_{p-1}}$ for $p > 1$ and $z = a_j$, then $x = a_{i_{p-1}} = z$. So, $x \notin T$ or $\text{Ind}(x) = 2$ and $x \in [S_1]$ or $x \notin T_1 \cup T_2$. If $x \notin T$, then $x \notin T_1 \cup T_2$. Thus, the closure of every letter whose image makes up x has length at most 1 link, so x is not the last link of y. So, $\text{Ind}(x) = 2$, which means that $x \notin [S_1]$ and $x \in T$. Thus, $x \neq z$. Similarly, x cannot be common to y' and z' if $y = a_i$ and $z = a_{j_0} \dots a_{j_{q-1}}$ with $q > 1$. If $y = a_i$ and $z = a_j$, then $y = x = z$ and thus y and z don't share a common link. Hence, different subchains of $\psi_0(M_0(\mathcal{C}))$ have no common links.

Next, suppose that some link x of $\text{Cl}(w)$ does not belong to any subchain of $\psi_0(M_0(\mathcal{C}))$. Then by the definition of ψ_0, since x is a link of some chain in $M_0(\mathcal{C})$, this says that x is on the beginning of the chain and $\text{Ind}(x) \in \{1, 2\}$, x is on the end of the chain and $\text{Ind}(x) \in \{1, 3\}$, or x is a chain of $M_0(\mathcal{C})$ and $x \in (T_1 \backslash T_2) \cup (T_2 \backslash T_1)$. If $\text{Ind}(x) = 1$, then $x \in [S_1]$, meaning that it is in $M_0(\mathcal{C})$ and hence in $\psi_0(M_0(\mathcal{C}))$. If x is on the beginning of a chain $y \in M_0(\mathcal{C})$ and $\text{Ind}(x) = 2$, then by Lemma 3.2.5, x has a middle subword m adjacent to $\tau_R(x)$ that is either composite or empty. By the definition of y being a closure, $|\tau_R(x)| \geq 3$, so $x = b_1 m \tau_R(x)$, where $|b_1| = 1$. If m is empty, then $|\tau_L(x)| = 1$ and hence $x \in S_3$ and $\text{Ind}(x) = 3$, a contradiction. Thus, m is nonempty, and $|\tau_1(x)| \geq 2$. So, x is on the end of the closure of the variable in which its τ_R is defined, and since it has index 2, it is in the φ_0 of that closure. If x is on the end of the chain and $\text{Ind}(x) = 3$, the result holds similarly. If $x \in M_0(\mathcal{C})$ and $x \in (T_1 \backslash T_2) \cup (T_2 \backslash T_1)$, then either the first image of some ξ starts x or the last image of some ξ ends x, but not both. Then, since $x \notin T$, we know that x will be included in one of the chains from the middle two cases in the definition of φ_0. $\square$

We now have a means by which to parse $\text{Cl}(w)$ into non-overlapping subchains that are linked to closures of letters in $\alpha(w)$. Our final step is to determine how to associate these chains with letters in $\alpha(w)$.

Fix a homomorphism $\varphi_0 : [\psi_0 (M_0 (\mathcal{C}))] \to \psi_0 (M_0 (\mathcal{C}))$, satisfying if $y = \varphi_0 ([x])$, then $y \in [x]$. For each $y \in \varphi_0 ([\psi_0 (M_0 (\mathcal{C}))])$ select a random letter ξ of $\alpha (w)$ such that $\psi_0 (\mathrm{Cl} (\xi)) = y$. We denote ξ by $f_0 (y)$. For each $y \in \psi_0 (M_0 (\mathcal{C}))$ we set $\psi_1 (y) = f_0 (\varphi_0 ([y]))$. In other words, φ_0 is a choice function that takes an equivalence class $[x]$ for some $x \in [\psi_0 (M_0 (\mathcal{C}))]$ and outputs an element $y \in \psi_0 (M_0 (\mathcal{C}))$ that is graphically equivalent to x, and f_0 is a choice function that randomly chooses a letter ξ such that y is the closure of ξ after the reduction by ψ_0. So, ψ_1 takes an element $y \in \psi_0 (M_0 (\mathcal{C}))$ and associates every element in $[y]$ with a letter ξ such that $\varphi_0 (\mathrm{Cl} (\xi))$ is graphically equivalent to y. More concisely, ψ_1 maps all graphically equivalent outputs of $\psi_0 (M_0 (\mathcal{C}))$ to the same letter ξ, which is what we need to finish the proof.

Claim 3.2.8. *ψ_1 an invertible function, and the output of its inverse is a word over W. (Recall that $W = \{a_0, a_1, \ldots, a_{m-1}\}$)*

Proof. Let $\xi \in \alpha (w)$, and suppose that $\psi_1 (x) = \xi$ and $\psi_1 (y) = \xi$. Then x is graphically equivalent to $\varphi_0 (\mathrm{Cl} (\xi))$ and y is graphically equivalent to $\varphi_0 (\mathrm{Cl} (\xi))$, so x is graphically equivalent to y. Thus, if a ξ is given, we can define $\psi_1^{-1} (\xi)$ to be the string of letters making up the subchain x, and this is clearly a word over W. $\square$

Since the chains of $\psi_0 (M_0 (\mathcal{C}))$ are disjoint, we can order them in the same order as in $\mathrm{Cl} (w)$. Finally, using the same ordering, concatenate the letters of $\psi_1 (\psi_0 (M_0 (\mathcal{C})))$ to form u. We note that by Property (f), there must be at least 3 links in $\mathcal{C}$, and hence $\mathrm{Cl} (w)$ has at least one link. Thus, u is not empty since each link in $\mathrm{Cl} (w)$ appears in $\psi_0 (M_0 (\mathcal{C}))$.

Clearly, $\alpha (u) \subseteq \alpha (w)$ since the range of ψ_1 is $\alpha (w)$. If $\xi \in \alpha (u)$, then it is an output of ψ_1 at least as many times as the letter appears in w due to the use of consistently choosing the same letter when graphical equality occurred. Thus, u is doubled. Finally, for $\xi \in \alpha (u)$, define $\varphi' : \alpha (u) \to X^+$ such that $\varphi' (\xi) = \Psi^{-1} \left(\psi_1^{-1} (\xi)\right)$, where

Ψ^{-1} makes sense since $\psi_1^{-1}(\xi) \in W^+$. Then $\Psi(\varphi'(u)) = \mathrm{Cl}(w)$ since there are no gaps and overlaps in $\psi_0(M_0(\mathcal{C}))$, so $\varphi'(u)$ is encountered by J_k. This completes the proof of Assertion 3.2.2, leaving only to formally define $a_0, a_1, \ldots, a_{m-1}$ and show that they satisfy Properties (a)-(f). $\square$

To define, $a_0, a_1, \ldots, a_{m-1}$, consider in the symmetric group S_m, the permutations of $\{0, \ldots, m-1\}$. Define the following permutations.

$$g_0 = \text{identity permutation}$$

$$g_1 = (1,2)(4,5)\ldots(m-2, m-1)$$

$$g_2 = (0,1)(3,4)\ldots(m-3, m-2)$$

$$f = (0,3,6,\ldots,m-3)$$

Define m permutations $\sigma_0, \ldots, \sigma_{m-1}$ by $\sigma_{3i+j} = f^i g_j$ for $0 \le i \le \frac{m}{3} - 1$, $0 \le j \le 2$, and let $a_1, \ldots, a_m$ be defined by $a_i = x_{\sigma_i(1)} \ldots x_{\sigma_i(m)}$ for all i.

Lemma 3.2.9. *The words $a_0, a_1, \ldots, a_{m-1}$ satisfy the following properties:*

(a) Each basic word is associated with a_i for only one i in the set $\{0, \ldots, m-1\}$.

(b) The words a_i, a_j with $i \neq j$ do not contain identical subwords of length greater than 2.

(c) There are no adjacent words that are subwords of a_i for any $i \in \{0, \ldots, m-1\}$.

(d) If $x_{i_p}x_{i_q}$ appears in a_i and a_j for $i \neq j$, then suppose $x_{i_p}x_{i_q}$ is a basic word associated with only one of a_i or a_j. If it is preceded by x_{i_r} where it is not associated, then x_{i_r} directly follows $x_{i_p}x_{i_q}$ in its associated word.

(e) If a subword of a_i has length $\equiv 0 \pmod 3$, begins in a position $\equiv 0 \pmod 3$, is composed of images of letters in w of length 1 or 2, and has more images of length 2 than length 1, then this subword contains a basic word.

(f) *The word $a_i a_j$ with $i \neq j$ does not contain any image of any doubled word v as a subword.*

Proof for Property (a). Suppose that a basic word $x_{i_p} x_{i_q}$ is associated with a_c and a_d with $c, d \in \{0, 1, \ldots, m-1\}$. Let $\sigma_c = f^{c_1} g_{c_2}$ and $\sigma_d = f^{d_1} g_{d_2}$ be the generating permutations for a_c and a_d, respectively, and let p and p' be positive integers such that $\sigma_c(p) = \sigma_d(p') = i_p$ and $\sigma_c(p+1) = \sigma_d(p'+1) = i_q$. Note that p and $p+1$ (resp. p' and $p'+1$) are the positions of x_{i_p} and x_{i_q} in a_c (resp. a_d), and also note that $p, p' \equiv 0 \pmod 3$. We now break into cases depending the value of i_p modulo 3.

Case: $i_p \equiv 0 \pmod 3$

Neither permutation uses g_2. Then, since $p \equiv 0 \pmod 3$, we know that

$$i_p = f^{c_1}(g_{c_2}(p)) = f^{c_1}(p).$$

Similarly,

$$i_p = f^{d_1}(g_{d_2}(p')) = f^{d_1}(p').$$

Thus, $f^{c_1}(p) = f^{d_1}(p')$. If the permutations use different g_i's, suppose without loss of generality that σ_c uses g_0 and σ_d uses g_1. Then

$$i_q = f^{c_1}(p+1) = p+1$$

and

$$i_q = f^{d_1}(g_1(p'+1)) = f^{d_1}(p'+2) = p'+2.$$

Thus, $p = p'+1$, but this is a contradiction because $p, p' \equiv 0 \pmod 3$. If both permutations use g_0, then $i_q = p+1$ again and $i_q = f^{d_1}(p'+1) = p'+1$. If both permutations use g_1, then

$$i_q = f^{c_1}(g_1(p+1)) = f^{c_1}(p+2) = p+2$$

and

$$i_q = f^{d_1}(g_1(p'+1)) = f^{d_1}(p'+2) = p'+2.$$

In either case, this says $p = p'$ and thus $f^{c_1}(p) = f^{d_1}(p)$. Hence, $c_1 = d_1$ since we know $p \equiv 0 \pmod 3$, so $\sigma_c = \sigma_d$ and $a_c = a_d$. Intuitively, the idea is that x_{i_q} appears in the same positions, and x_{i_p} can only precede x_{i_q} if the permutations are the same.

Case: $i_p \equiv 1 \pmod 3$

Both permutations use g_2. Then

$$i_p = f^{c_1}(g_2(p)) = f^{c_1}(p+1) = p+1.$$

Since $i_p \equiv 1 \pmod 3$,

$$i_p = g_2(p) = p+1.$$

Similarly,

$$i_p = f^{d_1}(g_2(p')) = p'+1,$$

so $p = p'$. Next,

$$i_q = f^{c_1}(g_2(p+1)) = f^{c_1}(p),$$

and similarly,

$$i_q = f^{d_1}(g_2(p'+1)) = f^{d_1}(p').$$

So, $f^{c_1}(p) = f^{d_1}(p)$ since $p = p'$, so by the same reasoning as above, $a_c = a_d$. Intuitively, the idea is that x_{i_p} appears in the same positions, and x_{i_q} can only follow x_{i_p} if the permutations are the same.

Case: $i_p \equiv 2 \pmod 3$

It is not possible for any σ to take p to the desired i_p since there is no σ that takes a value $\equiv 0 \pmod 3$ to a value $\equiv 2 \pmod 3$.

Therefore, every basic word is associated with a unique a_i. $\square$

Proof of (b). Suppose that the words a_c and a_d contain a common subword $x_{i_p}x_{i_q}x_{i_r}$ of length 3. Let σ_c, σ_d, p and p' be defined as in the proof of Property (a), and note that p, $p+1$, and $p+2$ (resp. p', $p'+1$, and $p'+2$) are the positions of x_{i_p}, x_{i_q}, and x_{i_r} in a_c (resp. a_d). By Property (a), if $x_{i_p}x_{i_q}x_{i_r}$ contains a basic subword, then

$a_c = a_d$. So, suppose $p, p' \equiv 1 \pmod{3}$. We now break into cases depending on the value of i_p modulo 3.

Case: $i_p \equiv 0 \pmod{3}$

Both permutations use g_2. So,

$$i_q = f^{c_1}\left(g_2\left(p+1\right)\right) = f^{c_1}\left(p+1\right) = p+1,$$

and similarly, $i_q = p' + 1$, so $p = p'$. Next,

$$i_p = f^{c_1}\left(g_2\left(p\right)\right) = f^{c_1}\left(p-1\right),$$

and similarly, $i_p = f^{d_1}\left(p' - 1\right)$. Since $p - 1, p' - 1 \equiv 0 \pmod{3}$, we again have that $c_1 = d_1$ and $a_c = a_d$.

Case: $i_p \equiv 1 \pmod{3}$

Both permutations use g_0. So, $i_p = f^{c_1}\left(p\right) = p$ and $i_p = f^{d_1}\left(p'\right) = p'$. So, $p = p'$. Next, $i_r = f^{c_1}\left(p+2\right)$ and $i_r = f^{d_1}\left(p'+2\right)$. So, since $p+2, p'+2 \equiv 0 \pmod{3}$, we know that $i_r \equiv 0 \pmod{3}$. Thus, since $p = p'$, $c_1 = d_1$, and hence $a_c = a_d$ by the same argument as Property (a).

Case: $i_p \equiv 2 \pmod{3}$

Both permutations use g_1. So,

$$i_p = f^{c_1}\left(g_1\left(p\right)\right) = f^{c_1}\left(p+1\right) = p+1,$$

and similarly, $i_p = p' + 1$, giving $p = p'$. Next,

$$i_r = f^{c_1}\left(g_1\left(p+2\right)\right) = f^{c_1}\left(p+2\right),$$

and similarly, $i_r = f^{d_1}\left(p'+2\right)$. Since $p+2, p'+2 \equiv 0 \pmod{3}$, we again have that $c_1 = d_1$ and $a_c = a_d$.

Intuitively, in each case, one of the letters are fixed in the same position in both words, and in order for the other letters to fall in the right place, the permutations must be the same. Therefore, any two a_i and a_j with $i \neq j$ have no common subwords of length greater than 2. $\square$

Proof of (c). Suppose that v is an adjacent word contained in the word a_j. Then by the definition of v being an adjacent word, v can be broken down into v_1v_2, where v_1 is a final segment of a_c and v_2 is an initial segment of a_d for some $c, d \in \{0, 1, \ldots, m-1\}$. Note that $c \neq j$ and $d \neq j$ since this would cause some letter to appear twice in the same word, but it does not contradict the definition of adjacent word for $c = d$. Then $|v_1| \leq 2$ or we contradict Property (b), and $|v_2| = 1$ or we contradict Property (a). Now, we'll consider the location of where this adjacent word could occur by considering the possible generators of a_c.

Case: a_c is generated by $\sigma_c = f^{c_1}g_0$

Since $m-1$ and $m-2$ are not $\equiv 0 \pmod 3$, the last two characters of a_c are $x_{m-2}x_{m-1}$. In order to retain that $x_{m-2}x_{m-1}$ appears in a_j, we must have that σ_j sends ℓ to $m-1$ for some $\ell \neq m-1$ (if $\ell = m-1$, then v_1 ends a_j and hence v doesn't appear in a_j) and sends $\ell - 1$ to $m-2$. If $\ell \equiv 1 \pmod 3$, then $\ell = m-2$ and σ_j uses g_1. In this case, $\ell - 2$ is not sent to $m-2$, however. If $\ell \equiv 0 \pmod 3$, then there is no permutation that will take ℓ to $m-1$. If $\ell \equiv 2 \pmod 3$, then σ_j uses either g_0 or g_2. If σ_j uses g_2, then $\ell = m-1$, which is already a contradiction since $\ell \neq m-1$ in order to ensure v appears in a_j.

Case: a_c is generated by $\sigma_c = f^{c_1}g_1$

Since $m-1$ and $m-2$ are not $\equiv 0 \pmod 3$, the last two characters of a_c are $x_{m-1}x_{m-2}$. In order to retain that $x_{m-1}x_{m-2}$ appears in a_j, we must have that σ_j sends ℓ to $m-2$ for some ℓ ($\ell \neq m-1$ as in the previous case) and sends $\ell - 1$ to $m-1$. If $\ell \equiv 1 \pmod 3$, then $\ell = m-2$ and σ_j uses g_0. In this case, $\ell - 1$ is not sent to $m-1$. If $\ell \equiv 0 \pmod 3$, then $\ell = m-3$ and σ_j uses g_2. In this case, $\ell - 1$ is not sent to $m-1$. If $\ell \equiv 2 \pmod 3$, then $\ell = m-1$.

Case: a_c is generated by $\sigma_c = f^{c_1} g_2$

Since $m-1$ is not $\equiv 0 \pmod 3$, the last two characters of a_c are $x_{i_{m-2}} x_{m-1}$, where $i_{m-2} \equiv 0 \pmod 3$. In order to retain that $x_{i_{m-2}} x_{m-1}$ appears in a_j, we must have that σ_j sends ℓ to $m-1$ for some ℓ (again $\ell \neq m-1$) and sends $\ell - 1$ to i_{m-2}. If $\ell \equiv 2 \pmod 3$, then $\ell = m-1$. If $\ell \equiv 0 \pmod 3$, then there is no permutation that will take ℓ to $m-1$. If $\ell \equiv 1 \pmod 3$, then $\ell = m-2$ and σ_j uses g_1. In this case, $\ell - 1$ can be sent to i_{m-2}, so we now consider the last letter of a_j. Since σ_j uses g_1, we see that the last letter of a_j is x_{m-2}. This says that x_{m-2} is the first letter of a_d. However, there is no permutation that will send 1 to $m-2$, a contradiction.

Thus, the adjacent word $v = v_1 v_2$ cannot appear in any a_j. $\square$

Proof of (d). Let a_c and a_d with $c \neq d$ contain $x_{i_p} x_{i_q}$. Let σ_c, σ_d, p and p' be defined as in (a), and note that p and $p+1$ (resp. p' and $p'+1$) are the positions of x_{i_p} and x_{i_q} in a_c (resp. a_d).

Suppose that $p \not\equiv 0 \pmod 3$, and we will show that if $p' \equiv 0 \pmod 3$ and $x_{i_p} x_{i_q}$ is preceded by x_{i_r} in a_c, then $x_{i_p} x_{i_q}$ is followed by x_{i_r} in a_d. First, we consider that $p \equiv 1 \pmod 3$ and break into cases with the generator of a_c.

Case: a_c is generated by $\sigma_c = f^{c_1} g_0$

Then $i_p = 1 \pmod 3$ and $i_q \equiv 2 \pmod 3$. If a_d uses g_0, then $i_p = f^{d_1}(g_0(p'))$, meaning that $p' \equiv 1 \pmod 3$. So, $x_{i_p} x_{i_q}$ is not a basic word associated with a_d. If a_d uses g_1, then $i_p = f^{d_1}(g_1(p'))$, meaning that $p' \equiv 2 \pmod 3$. So, $x_{i_p} x_{i_q}$ is not a basic word associated with a_d. If a_d uses g_2, then $i_p = f^{d_1}(g_2(p'))$. Since $i_p \equiv 1 \pmod 3$, this says that $p' \equiv 0 \pmod 3$, and hence $p'+1 \equiv 1 \pmod 3$. So, $i_q = f^{d_1}(g_2(p'+1)) = p' \equiv 0 \pmod 3$, contradicting $i_q \equiv 2 \pmod 3$.

Case: a_c is generated by $\sigma_c = f^{c_1} g_1$

Then $i_p = 2 \pmod 3$ and $i_q \equiv 1 \pmod 3$. If a_d uses g_0, then $i_p = f^{d_1}(g_0(p'))$, meaning that $p' \equiv 2 \pmod 3$. So, $x_{i_p} x_{i_q}$ is not a basic word associated with a_d.

If a_d uses g_1, then $i_p = f^{d_1}(g_1(p'))$, meaning that $p' \equiv 1 \pmod 3$. So, $x_{i_p}x_{i_q}$ is not a basic word associated with a_d. If a_d uses g_2, then $i_p = f^{d_1}(g_2(p'))$, meaning that $p' \equiv 2 \pmod 3$. So, $x_{i_p}x_{i_q}$ is not a basic word associated with a_d.

Case: a_c is generated by $\sigma_c = f^{c_1}g_2$

Then $i_p \equiv 0 \pmod 3$ and $i_q \equiv 2 \pmod 3$. If a_d uses g_0, then $i_p = f^{d_1}(g_0(p'))$. Since $i_p \equiv 0 \pmod 3$, this says that $p' \equiv 0 \pmod 3$ and hence $p' + 1 \equiv 1 \pmod 3$. So, $i_q = f^{d_1}(g_0(p'+1)) \equiv 1 \pmod 3$, contradicting $i_q \equiv 2 \pmod 3$. If a_d uses g_1, then $i_p = f^{d_1}(g_1(p'))$, meaning that $p' \equiv 0 \pmod 3$. Thus, $x_{i_p}x_{i_q}$ is a basic word associated with a_d. Further,

$$f^{c_1}(g_2(p+1)) = f^{d_1}(g_1(p'+1))$$
$$p + 1 = p' + 2.$$

Also,

$$f^{c_1}(g_2(p)) = f^{d_1}(g_1(p'))$$
$$f^{c_1}(p-1) = f^{d_1}(p') = f^{d_1}(p-1).$$

This gives us that $c_1 = d_1$. Finally, consider the letter x_{i_r} that appears directly before $x_{i_p}x_{i_q}$ in a_c. Then $p-1$ is the position of x_{i_r} in a_c, and we wish to show that $p'+2$ is the position of x_{i_r} in a_d.

$$f^{d_1}(g_1(p'+2)) = f^{c_1}(g_1(p+1)) = p = f^{c_1}(g_2(p-1)) = i_r,$$

as desired. If a_d uses g_2, then $i_p = f^{d_1}(g_2(p'))$, meaning that $p' \equiv 1 \pmod 3$. So, $x_{i_p}x_{i_q}$ is not a basic word associated with a_d.

Now, suppose that $p \equiv 2 \pmod 3$, and we will proceed similarly.

Case: a_c is generated by $\sigma_c = f^{c_1}g_0$

Then $i_p \equiv 2 \pmod 3$ and $i_q \equiv 0 \pmod 3$. If a_d uses g_0, then $i_p = f^{d_1}(g_0(p'))$, meaning that $p' \equiv 2 \pmod 3$, so $x_{i_p}x_{i_q}$ is not a basic word associated with a_d. If a_d

uses g_1, then $i_p = f^{d_1}(g_1(p'))$. Since $i_p \equiv 2 \pmod 3$, this says that $p' \equiv 1 \pmod 3$, so $x_{i_p}x_{i_q}$ is not a basic word associated with a_d. If a_d uses g_2, then $i_p = f^{d_1}(g_2(p'))$. Since $i_p \equiv 2 \pmod 3$, this says that $p' \equiv 2 \pmod 3$, so $x_{i_p}x_{i_q}$ is not a basic word associated with a_d.

Case: a_c is generated by $\sigma_c = f^{c_1}g_1$

Then $i_p \equiv 1 \pmod 3$ and $i_q \equiv 0 \pmod 3$. If a_d uses g_0, then $i_p = f^{d_1}(g_0(p'))$, meaning that $p' \equiv 2 \pmod 3$. So, $x_{i_p}x_{i_q}$ is not a basic word associated with a_d. If a_d uses g_1, then $i_p = f^{d_1}(g_1(p'))$, meaning that $p' \equiv 1 \pmod 3$. If a_d uses g_2, then $i_p = f^{d_1}(g_2(p'))$, meaning that $p' \equiv 0 \pmod 3$. Thus, $x_{i_p}x_{i_q}$ is a basic word of a_d. Further,

$$f^{c_1}(g_1(p)) = f^{d_1}(g_2(p'))$$
$$p - 1 = p' + 1.$$

Also,

$$f^{c_1}(g_1(p+1)) = f^{d_1}(g_2(p'+1))$$
$$f^{c_1}(p+1) = f^{d_1}(g_2(p-1)) = f^{d_1}(p-2).$$

This gives us that $c_1 = d_1 - 1$, possibly modulo n. Finally, consider the letter x_{i_r} that appears directly before $x_{i_p}x_{i_q}$ in a_c. Then $p-1$ is the position of x_{i_r} in a_c, and we wish to show that $p'+2$ is the position of x_{i_r} in a_d.

$$f^{d_1}(g_2(p'+2)) = f^{c_1+1}(g_2(p)) = p = f^{c_1}(g_1(p-1)) = i_r,$$

as desired.

Case: a_c is generated by $\sigma_c = f^{c_1}g_2$

Then $i_p \equiv 2 \pmod 3$ and $i_q \equiv 1 \pmod 3$. If a_d uses g_0, then $i_p = f^{d_1}(g_0(p'))$, meaning that $p' \equiv 2 \pmod 3$. So, $x_{i_p}x_{i_q}$ is not a basic word associated with a_d.

If a_d uses g_1, then $i_p = f^{d_1}(g_1(p'))$, meaning that $p' \equiv 1 \pmod 3$. So, $x_{i_p}x_{i_q}$ is not a basic word associated with a_d. If a_d uses g_2, then $i_p = f^{d_1}(g_2(p'))$, meaning that $p' \equiv 2 \pmod 3$. So, $x_{i_p}x_{i_q}$ is not a basic word associated with a_d. □

Proof of (e). This problem is equivalent to writing a sum of 1's and 2's, with more 2's than 1's, that adds up to multiple of 3, and we need to show that there is a location in the sum where 2 is added and the sum before is a multiple of 3. We will prove this by induction on the size of the sum as a multiple 3. For a sum of 3, it is impossible to write this. For a sum of 6, the only form possible is $2+2+2$, and the first 2 satisfies the conclusion. So, suppose that the conclusion holds when summing to $b-3$ and that we are summing to b with more 2's than 1's. If we start with 2, we are finished. If we start with $1+2$, we are finished by the inductive hypothesis. If we start with $1+1+1$, we are again finished by the induction hypothesis. If we start with $1+1+2+1+1$ or $1+1+2+2$, we are again finished by the inductive hypothesis. Finally, if we start with $1+1+2+1+2$, then the $1+2$ does not contribute to the problem. So, we can remove it, leaving us again in a smaller case. Hence, we are again finished by the inductive hypothesis. □

Before proving that that the a_i's satisfy Property (f), we state and prove a lemma.

Lemma 3.2.10. *If a is an initial segment of a_c, b is a final segment of a_d, and $\alpha(b) = \alpha(a)$, then $|a| = |b| = |a_c|$.*

Proof. Let a be an initial segment of a_c and b be a final segment of a_d such that $\alpha(a) = \alpha(b)$. Then since a_c and a_d are composed of distinct letters, $|a| = |b|$. Let σ_c and σ_d be defined as in (a), and consider the last letter of b, say $x_{i_{m-1}}$. Then $i_{m-1} = f^{c_1}g_{c_2}(m-1)$. Note that for all possibilities of c_2, the position i_{m-1} is either $m-1$ or $m-2$. Now, we know that $x_{i_{m-1}} \in \alpha(a)$ by assumption, so let m' be the position of $x_{i_{m-1}}$ in a_c. Then $i_{m-1} = f^{d_1}g_{d_2}(m')$. If $i_{m-1} = m-1$, then m' is either $m-1$ or $m-2$. If $i_{m-1} = m-2$, then m' is either $m-1$, $m-2$, or $m-3$. If

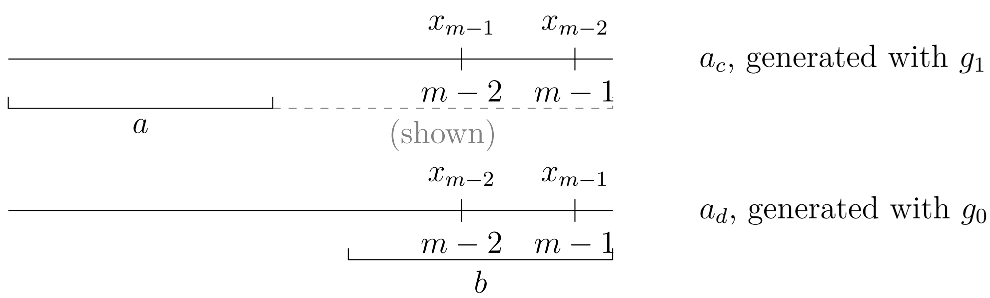

Figure 3.2 Diagram for Case 1 in the Proof of Lemma 3.2.10

$m' = m-1$, then $x_{i_{m-1}}$ is the last letter of a_c, and hence $|a| = |a_c|$. This leaves us to handle 3 cases: (1) $i_{m-1} = m-1$ and $m' = m-2$, (2) $i_{m-1} = m-2$ and $m' = m-2$, and (3) $i_{m-1} = m-2$ and $m' = m-3$. Before we go into the cases, note that $|b| = |a| \geq m-2 > 1$ since $n > 1$, so we can always assume that the next to last letter of a_d is in b. (Note: In case 3, if $n = 1$, x_2 can appear in the third letter of a_d and the first letter of a_c, contradicting this conclusion.)

Case 1: $i_{m-1} = m-1$ **and** $m' = m-2$

Recall that $m-1$ is the position of $x_{i_{m-1}}$ in a_c and m' is the position of $x_{i_{m-1}}$ in a_d. Since $i_{m-1} = m-1$, it must be that σ_d maps $m-1$ to $m-1$ and hence uses g_0 or g_2. Since $m' = m-2$, it must be that σ_c maps $m-2$ to $m-1$ and hence uses g_1. We now consider the next to last letter of b, which is $x_{i_{m-2}}$. So,

$$i_{m-2} = f^{d_1} g_{d_2} (m-2) = m-2.$$

Then x_{m-2} appears in a, so we desire to determine its position in a_c, say m''. Then

$$m-2 = f^{c_1} (g_1 (m'')) = g_1 (m'') = m'' - 1$$

since $m-2 \equiv 1 \pmod 3$. So, $m'' = m-1$, meaning that x_{m-2} appears in b and is the last letter of a_c. Thus, $|a| = |a_c|$, which is illustrated in Figure 3.2.

Case 2: $i_{m-1} = m-2$ **and** $m' = m-2$

Since $i_{m-1} = m-2$, it must be that σ_d maps $m-1$ to $m-2$ and hence uses g_1. Since $m' = m-2$, it must be that σ_c maps $m-2$ to $m-2$ and hence uses g_0. We

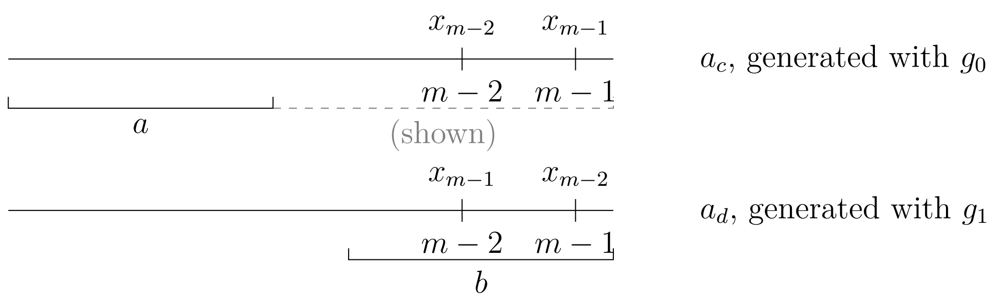

Figure 3.3 Diagram for Case 2 in the Proof of Lemma 3.2.10

again consider the next to last letter of b. Then

$$i_{m-2} = f^{d_1}\left(g_1\left(m-2\right)\right) = m-1.$$

Let m'' again be the position of x_{m-1} in a. Then

$$m-1 = f^{c_1}\left(g_0\left(m''\right)\right) = m'',$$

so $m'' = m-1$ again. Thus, $|a| = |a_c|$, which is illustrated in Figure 3.3.

Case 3: $i_{m-1} = m-2$ **and** $m' = m-3$

Since $i_{m-1} = m-2$, it must be that σ_d maps $m-1$ to $m-2$ and hence uses g_1. Since $m' = m-3$, it must be that σ_c maps $m-3$ to $m-2$ and hence uses g_2. Then

$$i_{m-2} = f^{d_1}\left(g_1\left(m-2\right)\right) = m-1,$$

and letting m'' be the position of x_{m-1} in a, we have that

$$m-1 = f^{c_1}\left(g_2\left(m''\right)\right) = g_2\left(m''\right) = m''$$

since $m-1 \equiv 2 \pmod 3$. Thus, $m'' = m-1$ and $|a| = |a_c|$ again, which is illustrated in Figure 3.4.

Therefore, either the last or next to last letter of b is the last letter of a_c and is in a, so $|a| = |b| = |a_c|$. $\square$

Proof of (f). Let c be an instance of the doubled word w, where $|\alpha\left(w\right)| \leq n$, and let φ be the mapping such that $\varphi\left(w\right) = c$. Suppose that c is a subword of $a_i a_j$, where

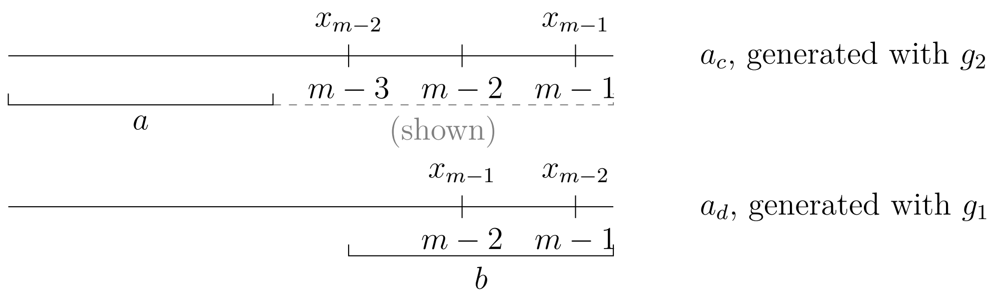

Figure 3.4 Diagram for Case 3 in the Proof of Lemma 3.2.10

$i \neq j$. Then since the words a_i and a_j do not contain two occurrences of the same letter, $c = c_1 c_2$, where c_1 is a final segment of a_i, c_2 is an initial segment of a_j, and $\alpha(c_1) = \alpha(c_2)$. Then by Lemma 3.2.10, $|c_1| = |c_2| = |a_i|$, so $c_1 = a_i$ and $c_2 = a_j$. So, since w is doubled, we have that $w = w_1 w_2$, where $c_1 = \varphi(w_1)$, $c_2 = \varphi(w_2)$, and hence w_2 is a reordering of the letters of w_1. Now, since $|\alpha(w)| \leq n$, we consider the images of the letters in w. No image can have size greater than 2, as this will create identical subwords in a_i and a_j, contradicting Property (b) since $i \neq j$.

Assume that $|\alpha(w)| = n$. Then by the pigeonhole principle, since a_i has m letters and w_1 has n letters, we must have $m - n$ images of letters of a_i of length 2. Now,

$$m - n = 3\lfloor\frac{n}{2}\rfloor + 3 - n \geq 3\frac{n-1}{2} + 3 - n = \frac{n+3}{2} > \lceil\frac{n}{2}\rceil + 1.$$

Thus, there are more images of letters of w of length 2 in a_i than there are images of length 1, even if $|\alpha(w)| \leq n$. So, by Property (e), there will be a letter $\xi \in \alpha(w)$ such that $\varphi(\xi)$ is a basic word, say $\varphi(\xi) = x_{i_p} x_{i_q}$. Because $\alpha(w_1) = \alpha(w_2)$, $\varphi(\xi)$ appears in a_j. So, consider the permutations σ_i and σ_j that generate a_i and a_j, respectively, and let p, p' be such that $\sigma_i(p) = \sigma_j(p') = i_p$ and $\sigma_i(p+1) = \sigma_j(p'+1) = i_q$. By definition of $x_{i_p} x_{i_q}$ being basic, we have that $p \equiv 0 \pmod 3$.

Case: σ_i uses g_0

We have that $p + 1 = i_q \equiv 1 \pmod 3$ and $i_p \equiv 0 \pmod 3$. If σ_j uses g_1, then $p' + 1 \equiv 2 \pmod 3$. Hence, $p' \equiv 1 \pmod 3$, so $i_p \equiv 2 \pmod 3$, contradicting the fact that $i_p \equiv 0 \pmod 3$. If σ_j uses g_2, then since $i_q \equiv 1 \pmod 3$, we know that $p' + 1 \equiv 0 \pmod 3$. Hence $p' \equiv 2 \pmod 3$, so $i_p \equiv 2 \pmod 3$, again a contradiction. If σ_j uses g_0, then $p' + 1 \equiv 1 \pmod 3$. Hence, $p' \equiv 0 \pmod 3$, so $x_{i_p} x_{i_q}$ is associated with a_j as a basic word. But, by Property (a), this says that $i = j$, contradicting the hypothesis.

Case: σ_i uses g_1

We have that $p+1 = i_q - 1$, so $i_q \equiv 2 \pmod 3$. Also, $i_p \equiv 0 \pmod 3$. If σ_j uses g_0, then since $i_q \equiv 2 \pmod 3$, we know that $p' + 1 \equiv 2 \pmod 3$. Hence $p' \equiv 1 \pmod 3$, so $i_p \equiv 1 \pmod 3$, contradicting $i_p \equiv 0 \pmod 3$. If σ_j uses g_1, then we know that $p' + 1 \equiv 1 \pmod 3$. Hence $p' \equiv 0 \pmod 3$, so $x_{i_p} x_{i_q}$ is associated with a_j as a basic word, again contradicting $i \neq j$ by appealing to Property (a). If σ_j uses g_2, then by Property (e), we see that if $x_{i_p} x_{i_q} x_{i_r}$ appears in a_i, then $x_{i_r} x_{i_p} x_{i_q}$ appears in a_j, and this holds for every basic word $x_{i_p} x_{i_q}$ associated with a_i. Consider another basic word associated with a_i, say $x_{i_c} x_{i_d}$, which is followed by x_{i_e}, and suppose that it is split between the images of two letters β and γ of w_1. That is $\varphi(\beta)$ ends in x_{i_c} and $\varphi(\gamma)$ begins in x_{i_d}. If $|\varphi(\beta)| = 2$, then the letter preceding x_{i_c} is x_{i_e} in a_j but is not x_{i_e} in a_i and a_j. Similarly, if $|\varphi(\gamma)| = 2$, then since the letter succeeding x_{i_d} is different in a_i and a_j, we have a contradiction. So, we must have that $|\varphi(\beta)| = |\varphi(\gamma)| = 1$ for any basic word that is not the image of some letter of w_1. With this in mind, the images of every letter of w_1 that is in a position $\equiv 2 \pmod 3$ must have size 1 in order to preserve the breakdown that every basic word is either the complete image of one or two letters. Thus, there are at least as many images of length 1 as there are images of length 2, meaning that a_i has length at most $2\lfloor \frac{n}{2} \rfloor + \left(n - \lfloor \frac{n}{2} \rfloor\right)$. But, this is the same as

$$2\left\lfloor \frac{n}{2} \right\rfloor + \left\lceil \frac{n}{2} \right\rceil \leq 2\left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{n}{2} \right\rfloor + 1 = 3\left\lfloor \frac{n}{2} \right\rfloor + 1 < m,$$

contradicting the length of a_i.

Case: σ_i uses g_2

We have that $p+1 = i_q + 1$, so $i_q \equiv 0 \pmod 3$. Also, $i_p \equiv 1 \pmod 3$. If σ_j uses g_0, then since $i_q \equiv 0 \pmod 3$, we know that $p' + 1 \equiv 0 \pmod 3$. Hence, we have that $p' \equiv 2 \pmod 3$, so $i_p \equiv 2 \pmod 3$, contradicting $i_p \equiv 1 \pmod 3$. If σ_j uses g_2, then since $i_q \equiv 0 \pmod 3$, we know that $p' + 1 \equiv 1 \pmod 3$. Hence $p' \equiv 0 \pmod 3$, so $x_{i_p} x_{i_q}$ is again associating with a_j as a basic word, a contradiction. If σ_j uses g_1, then this proof is the same as when σ_i uses g_1 and σ_j uses g_2 by instead assuming $\varphi(\xi)$ is basic in σ_j.

In essence, if a letter is mapped to a basic word, in order for this mapping to still hold in a_j, the permutations must be similar enough to either contradict Property (a) or contradict the construction of a_i. Thus, c is not a subword of $a_i a_j$ with $i \neq j$. $\square$

This completes the proof of the Theorem. $\square$

It is worth noting that the fact that w is doubled is only used in the proof of Property (f). However, Property (f) is needed to ensure that the word u is non-empty, among other things.

Chapter 4

Conclusion

The power series methods from Chapter 2 seem to not be applicable to the 99 doubled words remaining to check for 3-avoidability. These may not even have exponential lower bounds on the number of avoiding words. Ochem's approach to classifying all ternary words may be useful to finish the check for the 3-avoidability of every doubled word, but his approach generally does not imply any lower bound. Some new method would likely be needed to show an exponential lower bound, if one exists. It could be another power series argument that doesn't use the geometric series. It could be some variation on the methods of Brandenburg (1983) and Brinkhuis (1983). It could be a further squeeze on the inequalities used, but it is not obvious how to do so.

In the work of Mel'nichuk, it remains to consider how this argument could be squeezed. As noted at the end of Chapter 3, the only place in the proof that requires that the word be doubled is in Property (f). How much does the proof of Property (f) depend on the word being doubled? Could these methods be applied to find a bound on the avoidability index of all tripled words simultaneously? Mel'nichuk's result on the avoidability index of all avoidable words simultaneously is rather close to the bound on doubled words, but the methods are rather different. For even alphabets, we get a bound of $2\,(n+2)$ on avoidable words and a bound of $\frac{3}{2}\,(n+2)$ on doubled words. For odd alphabets, we get a bound of $2\,(n+1)$ on avoidable words and a bound of $\frac{3}{2}\,(n+1)$ on doubled words. It remains to show whether these bounds could be squeezed even further.

Bibliography

Baker, Kirby A., George F. McNulty, and Walter Taylor (1989). "Growth problems for avoidable words". In: *Theoret. Comput. Sci.* 69.3, pp. 319–345. ISSN: 0304-3975. DOI: 10.1016/0304-3975(89)90071-6. URL: http://dx.doi.org/10.1016/0304-3975(89)90071-6.

Bean, Dwight R., Andrzej Ehrenfeucht, and George F. McNulty (1979). "Avoidable patterns in strings of symbols". In: *Pacific J. Math.* 85.2, pp. 261–294. ISSN: 0030-8730. URL: http://projecteuclid.org/euclid.pjm/1102783913.

Bell, Jason P. and Teow Lim Goh (2007). "Exponential lower bounds for the number of words of uniform length avoiding a pattern". In: *Inform. and Comput.* 205.9, pp. 1295–1306. ISSN: 0890-5401. DOI: 10.1016/j.ic.2007.02.004. URL: http://dx.doi.org/10.1016/j.ic.2007.02.004.

Blanchet-Sadri, Francine and Brent Woodhouse (2013). "Strict bounds for pattern avoidance". In: *Developments in language theory*. Vol. 7907. Lecture Notes in Comput. Sci. Springer, Heidelberg, pp. 106–117. DOI: 10.1007/978-3-642-38771-5_11. URL: http://dx.doi.org/10.1007/978-3-642-38771-5_11.

Brandenburg, F.-J. (Mar. 1983). "Uniformly growing k-th power-free homorphisms". In: *Theoretical Computer Science* 23.1, pp. 69–82. ISSN: 0304-3975 (print), 1879-2294 (electronic).

Brinkhuis, Jan (1983). "Non-Repetetive Sequences on Three Symbols". In: *The Quarterly Journal of Mathematics* 34.2, pp. 145–149. DOI: 10.1093/qmath/34.2.145. eprint: http://qjmath.oxfordjournals.org/content/34/2/145.full.pdf+html. URL: http://qjmath.oxfordjournals.org/content/34/2/145.short.

Cassaigne, Julian (1994). "Motifs évitables et régularité dans les mots". PhD thesis. Université Paris VI.

Dalalyan, A. G. (1984). "Word eliminability". In: *Akad. Nauk Armyan. SSR Dokl.* 78.4, pp. 156–158. ISSN: 0321-1339.

Golod, E. S. and I. R. Šafarevič (1964). "On the class field tower". In: *Izv. Akad. Nauk SSSR Ser. Mat.* 28, pp. 261–272. ISSN: 0373-2436.

Lothaire, M. (2002). *Algebraic combinatorics on words*. Encyclopedia of mathematics and its applications. New York: Cambridge university press. ISBN: 0-521-81220-8. URL: `http://opac.inria.fr/record=b1098502`.

Mel'nichuk, I. L. (1985). "Existence of infinite finitely generated free semigroups in certain varieties of semigroups". In: *Algebraic systems with one action and relation*. Leningrad. Gos. Ped. Inst., Leningrad, pp. 74–83.

Ochem, Pascal (2006). "A generator of morphisms for infinite words". In: *ITA* 40.3, pp. 427–441. DOI: `10.1051/ita:2006020`. URL: `http://dx.doi.org/10.1051/ita:2006020`.

Rampersad, Narad (2011). "Further applications of a power series method for pattern avoidance". In: *Electron. J. Combin.* 18.1, Paper 134, 8. ISSN: 1077-8926.

Roth, Peter (1992). "Every binary pattern of length six is avoidable on the two-letter alphabet". English. In: *Acta Informatica* 29.1, pp. 95–107. ISSN: 0001-5903. DOI: `10.1007/BF01178567`. URL: `http://dx.doi.org/10.1007/BF01178567`.

Schmidt, Ursula (Jan. 1989). "Avoidable patterns on two letters". In: *Theoretical Computer Science* 63.1, pp. 1–17. ISSN: 0304-3975 (print), 1879-2294 (electronic).

Thue, A. (1906). *Über unendliche Zeichenreihen*. Skrifter udgivne af Videnskabsselskabet i Christiania. URL: `https://books.google.com/books?id=-gwpGwAACAAJ`.

Zimin, A. I. (1982). "Blocking sets of terms". In: *Mat. Sb. (N.S.)* 119(161).3, pp. 363–375, 447. ISSN: 0368-8666.