Infinite Sequences and Pattern Avoidance

by

Narad Rampersad

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2004

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The study of combinatorics on words dates back at least to the beginning of the 20^{th} century and the work of Axel Thue. Thue was the first to give an example of an infinite word over a three letter alphabet that contains no squares (identical adjacent blocks) xx. This result was eventually used to solve some longstanding open problems in algebra and has remarkable connections to other areas of mathematics and computer science as well. In this thesis we primarily study several variations of the problems studied by Thue in his work on repetitions in words, including some recent connections to other areas, such as graph theory.

In Chapter 1 we give a brief introduction to the subject of combinatorics on words. In Chapter 2 we use uniform morphisms to construct an infinite binary word that contains no cubes xxx and no squares yy with $|y| \geq 4$, thus giving a simpler construction than that of Dekking. We also use uniform morphisms to construct an infinite binary word avoiding all squares except 0^2, 1^2, and $(01)^2$, thus giving a simpler construction than that of Fraenkel and Simpson. We give some new enumeration results for these avoidance properties and solve an open problem of Prodinger and Urbanek regarding the perfect shuffle of infinite binary words that avoid arbitrarily large squares.

In Chapter 3 we examine ternary squarefree words in more detail, and in Chapter 4 we study words w satisfying the property that for any sufficiently long subword w' of w, w does not contain the reversal of w' as a subword. In Chapter 5 we discuss an application of the property of squarefreeness to colourings of graphs. In Chapter 6 we study strictly increasing sequences $(a(n))_{n \geq 0}$ of non-negative integers satisfying the equation $a(a(n)) = dn$. Finally, in Chapter 7 we give a brief conclusion and present some open problems.

Acknowledgements

I would like to thank Jeffrey Shallit for supervising the research presented in this thesis. I would also like to thank Ming Li and Alejandro López-Ortiz for serving as readers for this thesis, and Bryan Krawetz for many interesting discussions.

Contents

List of Tables

List of Figures

Chapter 1

Combinatorics on Words

This thesis focuses primarily on the subject of combinatorics on words. The study of combinatorics on words dates back at least to the beginning of the 20^{th} century and the work of Axel Thue [75, 76]. Unfortunately, Thue's work was published in a relatively obscure journal and therefore remained largely unknown for several decades. For this reason the study of combinatorics on words languished somewhat during the first half of the 20^{th} century (for some notable exceptions to this statement, see, for example, the work of Aršon [9] or Morse and Hedlund [55]). However, as we shall see later, there has been a tremendous increase in the amount of research being done in the field of combinatorics on words in recent decades.

Thue's work on combinatorics on words was largely concerned with repetitions in words, and his primary technique for studying such repetitions was the use of iterated morphisms (see Section 1.2 below). Chapters 2 and 3 of this thesis deal exclusively with questions regarding repetitions in words, and Chapter 2 in particular makes frequent use of iterated morphisms to construct words avoiding certain types of repetitions.

Chapter 4 deals not with repetitions in words but reversals. That is, we study words that do not contain the reversal of any of their sufficiently long subwords.

Some of the concepts of combinatorics on words, such as avoiding repetitions, have also been applied to other combinatorial structures, for instance, in the study of non-repetitive tilings of the plane or non-repetitive colourings of graphs. In Chapter 5 we examine some problems arising from the work of Alon *et al.* [7] regarding non-repetitive colourings of graphs in greater detail.

In Chapter 6 we move away from words over a finite alphabet and instead consider sequences over the non-negative integers. In particular, we examine strictly increasing sequences of non-negative integers satisfying a certain functional equation.

Finally, Chapter 7 concludes with a review of the material covered in this thesis, as well as some open problems.

In the remainder of this chapter we give an overview of some of the basic notions of combinatorics on words. For a more in-depth treatment of the subject of combinatorics on words, see one of the following: Berstel and Karhumäki [13], Choffrut and Karhumäki [23], or Lothaire [53].

1.1 Words

Let Σ be a finite, nonempty set called an *alphabet*; the elements of Σ are referred to as *symbols* or *letters*. We denote the set of all finite words over the alphabet Σ by Σ^*. We also write Σ^+ to denote the set $\Sigma^* - \{\epsilon\}$, where ϵ is the empty word. Let Σ_k denote the alphabet $\{0, 1, \ldots, k-1\}$. The length of a word w is denoted $|w|$. For $a \in \Sigma$ and $w \in \Sigma^*$, the number of occurrences of a in w is denoted $|w|_a$.

Let $\mathbb{N}$ denote the set $\{0, 1, 2, \ldots\}$. An *infinite word* is a map from $\mathbb{N}$ to Σ. The set of all infinite words over the alphabet Σ is denoted Σ^ω. We also write Σ^∞ to denote the set $\Sigma^* \cup \Sigma^\omega$.

A word w' is called a *subword*[1] of $w \in \Sigma^\infty$ if w can be written in the form $uw'v$ for some $u \in \Sigma^*$ and $v \in \Sigma^\infty$. If such a decomposition exists where $u = \epsilon$ (resp. $v = \epsilon$), then w' is called a *prefix* (resp. *suffix*) of w.

Throughout we will make implicit use of the following version of König's Infinity Lemma (see Allouche and Shallit [5, Section 1.9, Exercise 42]. For the original formulation of the lemma see König [49, 50]).

Theorem 1.1 (König's Infinity Lemma). *Let Σ be an alphabet. Let X be an infinite subset of Σ^*. There exists $\mathbf{w} \in \Sigma^\omega$ such that every prefix of $\mathbf{w}$ is a prefix of some $x \in X$.*

[1] Sometimes the term *factor* is also used.

A *square* is a nonempty word of the form xx, where $x \in \Sigma^*$. We say a word w is *squarefree* (or *avoids squares*) if no subword of w is a square. It is easy to see that every word of length at least four constructed from the symbols 0 and 1 contains a square; it is therefore impossible to avoid squares in infinite binary words. However, in 1906, Thue [75, Satz 5] proved that there exists an infinite squarefree word over an alphabet of size three. This result has been independently rediscovered several times (see, for example, Aršon [9] and Morse and Hedlund [55]).

Perhaps the most famous application of Thue's construction of an infinite squarefree ternary word is in the work done by Novikov and Adjan [57, 58, 59, 60] in solving the Burnside problem for groups [18] (see also Hall [41] and Adjan [1]).

By analogy with the definition of a square, a *cube* is a nonempty word of the form xxx, where $x \in \Sigma^*$. While it is impossible to avoid squares in infinite binary words, Thue [75, Satz 6] also proved that there exists an infinite cubefree word over a binary alphabet.

The concepts of *square* and *cube* can also be considered in a more general framework of pattern avoidance (see Bean, Ehrenfeucht, and McNulty [10], Zimin [78], Roth [68, 69], and Cassaigne [21]).

Another interesting generalization of *square* is the Abelian square. An *Abelian square* is a nonempty word xy, where y is a permutation of the symbols of x. This concept was motivated by a problem posed by Erdős [34]. For more on Abelian squares, see Evdokimov [36], Pleasants [62], Brown [17], Justin [44, 45], Dekking [30], Keränen [47], and Carpi [20].

In Chapter 3 we examine infinite squarefree ternary words in more detail, and in Chapter 5 we discuss an application of the property of squarefreeness to colourings of graphs.

1.2 Morphisms

A map $h : \Sigma^* \to \Delta^*$ is called a *morphism* if h satisfies $h(xy) = h(x)h(y)$ for all $x, y \in \Sigma^*$. A morphism may be specified by providing the *image words* $h(a)$ for all $a \in \Sigma$. For example, we may define a morphism $h : \Sigma_3^* \to \Sigma_3^*$ by

$$\begin{aligned} 0 &\to 01201 \\ 1 &\to 020121 \\ 2 &\to 0212021 \ . \end{aligned} \tag{1.1}$$

This definition is easily extended to infinite words.

A morphism $h : \Sigma^* \rightarrow \Sigma^*$ such that $h(a) = ax$ for some $a \in \Sigma$ and $x \in \Sigma^*$ is said to be *prolongable on* a; we may then repeatedly iterate h to obtain the *fixed point*

$$h^\omega(a) = axh(x)h^2(x)h^3(x)\cdots\ .$$[2]

The morphism h given by (1.1) is prolongable on 0, and so we have the fixed point

$$h^\omega(0) = 012010201210212021012010201 21\cdots\ .$$

From Thue [76, Satz 18] (see also Pleasants [62]) we have that $h^\omega(0)$ is squarefree.

A morphism is k-*uniform* if $|h(a)| = k$ for all $a \in \Sigma$; it is *uniform* if it is k-uniform for some k. For example, if the morphism $g : \Sigma_3^* \rightarrow \Sigma_3^*$ is defined by

$$\begin{aligned} 0 &\rightarrow 0121021201210 \\ 1 &\rightarrow 1202102012021 \\ 2 &\rightarrow 2010210120102\ , \end{aligned} \tag{1.2}$$

then g is 13-uniform. Moreover, Leech [52] showed that the infinite word $g^\omega(0)$ is squarefree. Note that the morphism h given by (1.1) is not uniform.

In Chapter 2 we use uniform morphisms to give simpler proofs of results due to Dekking [28] and Fraenkel and Simpson [38]. We also solve an open problem due to Prodinger and Urbanek [63].

1.3 Enumeration

We are often interested in counting the number of words of length n with a given avoidance property. For example, Brinkhuis [16] showed that there are exponentially many ternary squarefree words of length n; specifically, he showed that there are $\Omega(2^{n/24}) = \Omega(1.029^n)$ such words. Successive improvements to this bound have been found by Brandenburg [15], Ekhad and Zeilberger [32], and Grimm [39]. Currently, the best known lower bound is $\Omega(110^{n/42}) = \Omega(1.118^n)$ due to Sun [73]. Noonan and Zeilberger [56] have given an

[2] Such a word is sometimes also called a $D0L$ word (see Rozenberg and Salomaa [70]).

upper bound of $O(1.302^n)$ (Richard and Grimm [67] have subsequently given a very slight improvement).

In Chapter 2 we also give some new enumeration results for certain avoidance properties.

Chapter 2

Avoiding Large Squares

In this chapter[1] we use uniform morphisms to give simpler proofs of results due to Dekking [28] and Fraenkel and Simpson [38]. In addition, we give an enumeration of the words studied by Dekking and those studied by Fraenkel and Simpson. We also solve an open problem due to Prodinger and Urbanek [63].

2.1 Previous work

We have already noted in Chapter 1 that every binary word of length at least four contains a square. However, in 1974, Entringer, Jackson, and Schatz [33] proved the remarkable fact that there exists an infinite binary word containing no squares xx with $|x| \geq 3$. Furthermore, the bound 3 is best possible.

The construction of Entringer, Jackson, and Schatz is as follows. Let $\mathbf{w}$ be any infinite squarefree word over Σ_3. Let g be the morphism defined by

$$\begin{aligned} 0 &\rightarrow 1010 \\ 1 &\rightarrow 1100 \\ 2 &\rightarrow 0111 \ . \end{aligned}$$

[1]The contents of this chapter are taken largely verbatim from Rampersad, Shallit, and Wang [66]. For the most part, the formulation of the theorems and lemmata in this chapter are due to J. Shallit and M.-w. Wang; the contribution of the present author lies largely in the proofs.

Then $g(\mathbf{w})$ contains no squares xx with $|x| \geq 3$.

Entringer, Jackson, and Schatz also conjectured that any infinite cubefree binary word must contain arbitrarily large squares. Dekking [28] disproved this conjecture by showing that there exists an infinite binary word that contains no cubes xxx and no squares yy with $|y| \geq 4$. Furthermore, the bound 4 is best possible.

Dekking's construction is as follows. Let $\mathbf{w}$ be any infinite squarefree word over Σ_3. Let g be the morphism defined by

$$\begin{aligned} 0 &\rightarrow 00110101100101 \\ 1 &\rightarrow 0011011001011101001 \\ 2 &\rightarrow 0011011010010111001 \ . \end{aligned}$$

Then $g(\mathbf{w})$ contains no contains no cubes xxx and no squares yy with $|y| \geq 4$. Note that Dekking's morphism is not uniform. In this chapter we show how to obtain, using the image of a *uniform* morphism, an infinite binary word that is cubefree and avoids squares yy with $|y| \geq 4$. This gives a somewhat simpler construction than that given by Dekking.

Fraenkel and Simpson [38] strengthened the results of Entringer, Jackson, and Schatz by showing that there exists an infinite binary word avoiding all squares except 0^2, 1^2, and $(01)^2$. Their construction, however, was rather complicated, involving several steps and non-uniform morphisms. In this chapter we show how to obtain, again using a uniform morphism, an infinite binary word avoiding all squares except 0^2, 1^2, and $(01)^2$. This gives a somewhat simpler construction than that given by Fraenkel and Simpson.

We also consider the number of finite binary words satisfying the Dekking and Fraenkel–Simpson avoidance properties. We give exponential upper and lower bounds on this number in both cases.

Prodinger and Urbanek [63] also studied words avoiding large squares, in particular with reference to operations that preserve this property, such as the perfect shuffle $⧢$. Let $w = a_1 a_2 \cdots a_n$ and $x = b_1 b_2 \cdots b_n$ be words of length n. The *perfect shuffle* $w ⧢ x$ is defined to be the word $a_1 b_1 a_2 b_2 \cdots a_n b_n$ of length $2n$. The definition can easily be extended to infinite words. They stated the following open question: do there exist two infinite words avoiding arbitrarily large squares such that their perfect shuffle has arbitrarily large squares? We answer this question in the affirmative by giving two such words.

2.2 Cubefree words without arbitrarily large squares

In this section we construct an infinite cubefree binary word avoiding squares yy with $|y| \geq 4$. The techniques we use are also used in later sections, so in this section we spell them out in some detail.

Theorem 2.1. *There is a squarefree infinite word over Σ_4 with no occurrences of the subwords* 12*,* 13*,* 21*,* 32*,* 231*, or* 10302*.*

Proof. Let the morphism h be defined by

$$\begin{array}{rcl} 0 & \rightarrow & 0310201023 \\ 1 & \rightarrow & 0310230102 \\ 2 & \rightarrow & 0201031023 \\ 3 & \rightarrow & 0203010201 \end{array} .$$

Then we claim the fixed point $h^\omega(0)$ has the desired properties.

First, we claim that if $w \in \Sigma_4^*$ then $h(w)$ has no occurrences of 12, 13, 21, 32, 231, or 10302. For if any of these words occur as subwords of $h(w)$, they must occur within some $h(a)$ or straddling the boundary between $h(a)$ and $h(b)$, for some single letters a, b. They do not; this easy verification is left to the reader.

Next, we prove that if w is any squarefree word over Σ_4 having no occurrences of 12, 13, 21, or 32, then $h(w)$ is squarefree.

We argue by contradiction. Let $w = a_1 a_2 \cdots a_n$ be a squarefree string such that $h(w)$ contains a square, i.e., $h(w) = xyyz$ for some $x, z \in \Sigma_4^*$, $y \in \Sigma_4^+$. Without loss of generality, assume that w is a shortest such string, so that $0 \leq |x|, |z| < 10$.

Case 1: $|y| \leq 20$. In this case we can take $|w| \leq 5$. To verify that $h(w)$ is squarefree, it therefore suffices to check each of the 49 possible words $w \in \Sigma_4^5$ to ensure that $h(w)$ is squarefree in each case.

Case 2: $|y| > 20$. First, we establish the following result.

Lemma 2.2. *(a) Suppose $h(ab) = th(c)u$ for some letters $a, b, c \in \Sigma_4$ and strings $t, u \in \Sigma_4^*$. Then this inclusion is trivial (that is, $t = \epsilon$ or $u = \epsilon$) or u is not a prefix of $h(d)$ for any $d \in \Sigma_4$.*

(b) *Suppose there exist letters a, b, c and strings s, t, u, v such that $h(a) = st$, $h(b) = uv$, and $h(c) = sv$. Then either $a = c$ or $b = c$.*

Proof. (a) This can be verified with a short computation. In fact, the only a, b, c for which the equality $h(ab) = th(c)u$ holds nontrivially is $h(31) = th(2)u$, and in this case $t = 020301$, $u = 0102$, so u is not a prefix of any $h(d)$.

(b) This can also be verified with a short computation. If $|s| \geq 6$, then no two distinct letters have images under h that share a prefix of length 6. If $|s| \leq 5$, then $|t| \geq 5$, and no two distinct letters have images under h that share a suffix of length 5.

□

Once Lemma 2.2 is established, the rest of the argument is fairly standard. It can be found, for example, in [46], but for completeness we repeat it here.

For $i = 1, 2, \ldots, n$ define $A_i = h(a_i)$. Then if $h(w) = xyyz$, we can write

$$h(w) = A_1 A_2 \cdots A_n = A_1' A_1'' A_2 \cdots A_{j-1} A_j' A_j'' A_{j+1} \cdots A_{n-1} A_n' A_n''$$

where

$$\begin{aligned} A_1 &= A_1' A_1'' \\ A_j &= A_j' A_j'' \\ A_n &= A_n' A_n'' \\ x &= A_1' \\ y &= A_1'' A_2 \cdots A_{j-1} A_j' = A_j'' A_{j+1} \cdots A_{n-1} A_n' \\ z &= A_n'' \ , \end{aligned}$$

and $|A_1''|, |A_j''| > 0$. See Figure 2.1.

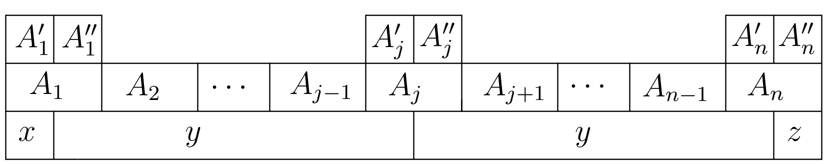

Figure 2.1: The string $xyyz$ within $h(w)$

If $|A''_1| > |A''_j|$, then $A_{j+1} = h(a_{j+1})$ is a subword of $A''_1 A_2$, hence a subword of $A_1 A_2 = h(a_1 a_2)$. Thus we can write $A_{j+2} = A'_{j+2} A''_{j+2}$ with

$$A''_1 A_2 = A''_j A_{j+1} A'_{j+2} \ .$$

See Figure 2.2.

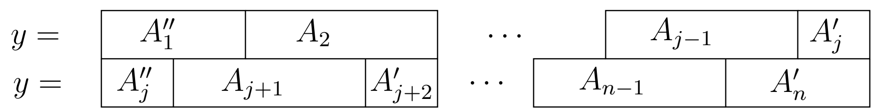

Figure 2.2: The case $|A''_1| > |A''_j|$

But then, by Lemma 2.2 (a), either $|A''_j| = 0$, or $|A''_1| = |A''_j|$, or A'_{j+2} is a not a prefix of any $h(d)$. All three conclusions are impossible.

If $|A''_1| < |A''_j|$, then $A_2 = h(a_2)$ is a subword of $A''_j A_{j+1}$, hence a subword of $A_j A_{j+1} = h(a_j a_{j+1})$. Thus we can write $A_3 = A'_3 A''_3$ with

$$A''_1 A_2 A'_3 = A''_j A_{j+1} \ .$$

See Figure 2.3.

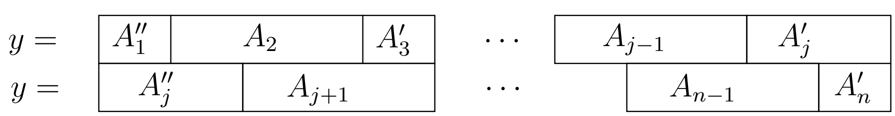

Figure 2.3: The case $|A''_1| < |A''_j|$

By Lemma 2.2 (a), either $|A''_1| = 0$ or $|A''_1| = |A''_j|$ or A'_3 is not a prefix of any $h(d)$. Again, all three conclusions are impossible.

Therefore $|A''_1| = |A''_j|$. Hence $A''_1 = A''_j$, $A_2 = A_{j+1}$, $\ldots$, $A_{j-1} = A_{n-1}$, and $A'_j = A'_n$. Since h is injective, we have $a_2 = a_{j+1}, \ldots, a_{j-1} = a_{n-1}$. It also follows that $|y|$ is divisible by 10 and $A_j = A'_j A''_j = A'_n A''_1$. But by Lemma 2.2 (b), either (1) $a_j = a_n$ or (2) $a_j = a_1$. In the first case, $a_2 \cdots a_{j-1} a_j = a_{j+1} \cdots a_{n-1} a_n$, so w contains the square $(a_2 \cdots a_{j-1} a_j)^2$, a contradiction. In the second case, $a_1 \cdots a_{j-1} = a_j a_{j+1} \cdots a_{n-1}$, so w contains the square $(a_1 \cdots a_{j-1})^2$, a contradiction.

It now follows that the infinite word

$$h^\omega(0) = 03102010230203010201031023010203102010230201031023\cdots$$

is squarefree and contains no occurrences of 12, 13, 21, 32, 231, or 10302. □

Theorem 2.3. *Let* $\mathbf{w}$ *be any infinite word satisfying the conditions of Theorem 2.1. Define a morphism* g *by*

$$\begin{aligned} 0 &\to 010011 \\ 1 &\to 010110 \\ 2 &\to 011001 \\ 3 &\to 011010 \ . \end{aligned}$$

Then $g(\mathbf{w})$ *is a cubefree word containing no squares* xx *with* $|x| \geq 4$.

Before we begin the proof, we remark that all the words 12, 13, 21, 32, 231, 10302 must indeed be avoided, because

$$\begin{array}{rl} g(12) & \text{contains the squares } (0110)^2,\ (1100)^2,\ (1001)^2 \\ g(13) & \text{contains the square } (0110)^2 \\ g(21) & \text{contains the cube } (01)^3 \\ g(32) & \text{contains the square } (1001)^2 \\ g(231) & \text{contains the square } (10010110)^2 \\ g(10302) & \text{contains the square } (100100110110)^2 \ . \end{array}$$

Proof. The proof parallels the proof of Theorem 2.1. Let $w = a_1 a_2 \cdots a_n$ be a squarefree string, with no occurrences of 12, 13, 21, 32, 231, or 10302. We first establish that if $g(w) = xyyz$ for some $x, z \in \Sigma_2^*$, $y \in \Sigma_2^+$, then $|y| \leq 3$. Without loss of generality, assume w is a shortest such string, so $0 \leq |x|, |z| < 6$.

Case 1: $|y| \leq 12$. In this case we can take $|w| \leq 5$. To verify that $g(w)$ contains no squares yy with $|y| \geq 4$, it suffices to check each of the 41 possible words $w \in \Sigma_4^5$.

Case 2: $|y| > 12$. First, we establish the analogue of Lemma 2.2.

Lemma 2.4. *(a) Suppose $g(ab) = tg(c)u$ for some letters $a, b, c \in \Sigma_4$ and strings $t, u \in \Sigma_2^*$. Then this inclusion is trivial (that is, $t = \epsilon$ or $u = \epsilon$) or u is not a prefix of $g(d)$ for any $d \in \Sigma_4$.*

(b) Suppose there exist letters a, b, c and strings s, t, u, v such that $g(a) = st$, $g(b) = uv$, and $g(c) = sv$. Then either $a = c$ or $b = c$, or $a = 2$, $b = 1$, $c = 3$, $s = 0110$, $t = 01$, $u = 0101$, $v = 10$.

Proof. (a) This can be verified with a short computation. The only a, b, c for which $g(ab) = tg(c)u$ holds nontrivially are

$$\begin{aligned} g(01) &= 010\ g(3)\ 110 \\ g(10) &= 01\ g(2)\ 0011 \\ g(23) &= 0110\ g(1)\ 10\ . \end{aligned}$$

But none of 110, 0011, 10 are prefixes of any $g(d)$.

(b) If $|s| \geq 5$ then no two distinct letters have images under g that share a prefix of length 5. If $|s| \leq 3$ then $|t| \geq 3$, and no two distinct letters have images under g that share a suffix of length 3. Hence $|s| = 4$, $|t| = 2$. But only $g(2)$ and $g(3)$ share a prefix of length 4, and only $g(1)$ and $g(3)$ share a suffix of length 2.

□

The rest of the proof is exactly parallel to the proof of Theorem 2.1, with the following exception. When we get to the final case, where $|y|$ is divisible by 6, we can use Lemma 2.4 to rule out every case except where $x = 0101$, $z = 01$, $a_1 = 1$, $a_j = 3$, and $a_n = 2$. Thus $w = 1\alpha 3\alpha 2$ for some string $\alpha \in \Sigma_4^*$. This special case is ruled out by the following lemma.

Lemma 2.5. *Suppose $\alpha \in \Sigma_4^*$, and let $w = 1\alpha 3\alpha 2$. Then either w contains a square, or w contains an occurrence of one of the subwords* 12, 13, 21, 32, 231, *or* 10302.

Proof. This can be verified by checking (a) all strings $w = 1\alpha 3\alpha 2$ with $|\alpha| \leq 5$, and (b) all strings $w = 1\alpha 3\alpha 2$ such that α is of the form $abc\alpha' def$, where $a, b, c, d, e, f \in \Sigma_4$ and $\alpha' \in \Sigma_4^*$. (Here α' may be treated as an indeterminate.) □

It now remains to show that if w is squarefree and contains no occurrence of 12, 13, 21, 32, 231, or 10302, then $g(w)$ is cubefree. If $g(w)$ contains a cube yyy, then it contains a square yy, and from what precedes we know $|y| \leq 3$. It therefore suffices to show that $g(w)$ contains no occurrence of 0^3, 1^3, $(01)^3$, $(10)^3$, $(001)^3$, $(010)^3$, $(011)^3$, $(100)^3$, $(101)^3$, $(110)^3$. The longest such string is of length 9, so it suffices to examine the 16 possibilities for $g(w)$ where $|w| = 3$. This is left to the reader.

The proof of Theorem 2.3 is now complete. □

Corollary 2.6. *If g and h are defined as above, then*

$$g(h^\omega(0)) = 0100110110100101100100110110010100110101100100110110\cdots$$

is cubefree, and avoids all squares xx with $|x| \geq 4$.

Next, based on the morphism h, we define the substitution (in the sense of Hopcroft and Ullman [42, p. 60]) $h' : \Sigma_4^* \to 2^{\Sigma_4^*}$ as follows:

$$\begin{aligned} 0 &\to \{h(0)\} \\ 1 &\to \{h(1), 0310230201\} \\ 2 &\to \{h(2)\} \\ 3 &\to \{h(3)\}\ . \end{aligned}$$

Thus, if $w \in \Sigma_4^*$, $h'(w)$ is a language of 2^r words over Σ_4, where $r = |w|_1$. Each of these words is of length $10|w|$.

Lemma 2.7. *Let g, h, and h' be defined as above. Let $w = h^m(0)$ for some positive integer m. Then $g(h'(w))$ is a language of $2^{n/300}$ words over Σ_2, where $n = 60 \cdot 10^m$ is the length of each of these words. Furthermore, each of these words is cubefree and avoids all squares xx with $|x| \geq 4$.*

Proof. Note that there are exactly two 1's in every image word of h. Hence, $|w|_1 = \frac{1}{5}|w| = \frac{1}{5} \cdot 10^m$. We have then that $g(h'(w))$ consists of $2^{\frac{1}{5}\cdot 10^m}$ binary words. Since $n = 6 \cdot 10 \cdot 10^m$, we see that $g(h'(w))$ consists of $2^{n/300}$ words.

To see that the words in $g(h'(w))$ are cubefree and avoid all squares xx with $|x| \geq 4$, it suffices by Theorem 2.3 to show that the words in $h'(w)$ are squarefree and contain no occurrences of the subwords 12, 13, 21, 32, 231, or 10302. By the same reasoning as in Theorem 2.1, it is easy to verify that no word in $h'(w)$ contains an occurrence of 12, 13, 21, 32, 231, or 10302.

To show that the words in $h'(w)$ are squarefree, we will, as a notational convenience, prefer to consider h' to be a morphism defined as follows:

$$\begin{aligned} 0 &\rightarrow h(0) \\ 1 &\rightarrow h(1) \\ \hat{1} &\rightarrow 0310230201 \\ 2 &\rightarrow h(2) \\ 3 &\rightarrow h(3)\ . \end{aligned}$$

Here, 1 and $\hat{1}$ are considered to be the same alphabet symbol; the 'hat' simply serves to distinguish between which choice is made for the substitution. To show that $h'(w)$ is squarefree, it suffices to show that h' satisfies the conditions of Lemma 2.2. For Lemma 2.2 (a) we have $h'(22) = th'(\hat{1})u$, but we can rule this case out since w avoids the square 22. For Lemma 2.2 (b) we again have that no two distinct letters have images under h' that share a prefix of length 6 or a suffix of length 5 (since 1 and $\hat{1}$ are not considered to be distinct letters). Hence, h' satisfies the conditions of Lemma 2.2, and so $h'(w)$ is squarefree. □

Theorem 2.8. *Let G_n denote the number of cubefree binary words of length n that avoid all squares xx with $|x| \geq 4$. Then $G_n = \Omega(1.002^n)$ and $G_n = O(1.178^n)$.*

Proof. Noting that $2^{1/300} \doteq 1.002$, we see that the lower bound follows immediately from Lemma 2.7.

For the upper bound we reason as follows. The set of binary words of length n avoiding cubes and squares xx with $|x| \geq 4$ is a a subset of the set of binary words avoiding 000 and 111. The number G'_n of binary words avoiding 000 and 111 satisfies the linear recurrence $G'_n = G'_{n-1} + G'_{n-2}$ for $n \geq 3$. From well-known properties of linear recurrences, it follows that $G'_n = O(\alpha^n)$, where α is the largest zero of $x^2 - x - 1$, the characteristic polynomial of the recurrence. Here $\alpha < 1.618$, so $G'_n = O(1.618^n)$.

This argument can be extended by using a symbolic algebra package such as Maple. Noonan and Zeilberger [56] have written a Maple package, `DAVID_IAN`, that allows one to specify a list L of forbidden words, and computes the generating function enumerating words avoiding members of L. We used the package for a list L of 90 words of length ≤ 20:

$$000, 111, \ldots, 11011001001101100100$$

obtaining a characteristic polynomial of degree 44 with dominant root $\doteq 1.178$. □

The following table gives the number G_n of binary words of length n avoiding both cubes xxx and squares y with $|y| \geq 4$.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
G_n	1	2	4	6	10	16	24	36	52	72	90	116	142	178	220	264	332	414

2.3 Words avoiding all squares except 0^2, 1^2, and $(01)^2$

In this section we construct an infinite binary word avoiding all squares except 0^2, 1^2, and $(01)^2$.

Roughly speaking, verifying that the image of a morphism avoids arbitrarily large squares breaks up into two parts: checking a finite number of "small" squares, and checking an infinite number of "large" squares. The small squares can be checked by brute force, while for the large squares we need a version of Lemma 2.2. Referring to Lemma 2.2 (a), if $h(c)$ is a subword of $h(ab)$ for some letters a, b, c, we call this an "inclusion". Inclusions can be ruled out either by considering prefixes, as we did in Lemma 2.2 (a), or suffixes. Referring to Lemma 2.2 (b), if $h(a) = st$, $h(b) = uv$, and $h(c) = sv$, we call that an "interchange".

The basic idea of the proofs in this section parallels that of the previous section, so we just sketch the basic ideas, pointing out the properties of the inclusions and interchanges.

Consider the 24-uniform morphism h defined as follows:

$$\begin{aligned} 0 &\rightarrow 012321012340121012321234 \\ 1 &\rightarrow 012101234323401234321234 \\ 2 &\rightarrow 012101232123401232101234 \\ 3 &\rightarrow 012321234323401232101234 \\ 4 &\rightarrow 012321234012101234321234 \ . \end{aligned}$$

Theorem 2.9. *If $w \in \Sigma_5^*$ is squarefree and avoids the patterns* 02*,* 03*,* 04*,* 14*,* 20*,* 30*,* 41*, then $h(w)$ is squarefree and avoids the patterns* 02*,* 03*,* 04*,* 13*,* 14*,* 20*,* 24*,* 30*,* 31*,* 41*,* 42*,* 010*,* 434*.*

Proof. The only inclusion is $h(32) = 0123212343234\ h(0)\ 01232101234$, and 0123212343234 is not a suffix of the image of any letter.

There are no interchanges for this morphism. ☐

Now consider the 6-uniform morphism g defined by

$$\begin{aligned} 0 &\rightarrow 011100 \\ 1 &\rightarrow 101100 \\ 2 &\rightarrow 111000 \\ 3 &\rightarrow 110010 \\ 4 &\rightarrow 110001 \ . \end{aligned}$$

Theorem 2.10. *If w is squarefree and avoids the patterns* 02*,* 03*,* 04*,* 13*,* 14*,* 20*,* 24*,* 30*,* 31*,* 41*,* 42*,* 434010 *then the only squares in $g(w)$ are* 00*,* 11*,* 0101*.*

Proof. There are no examples of interchanges for g.

There are multiple examples of inclusions, but many of them can be ruled out by properties of w and g:

- $g(02) = 01110g(0)0$ but 02 cannot occur
- $g(24) = 1g(4)10001$ but 24 cannot occur

- $g(12) = 10110g(0)0$ but 10110 is not a suffix of any $g(a)$
- $g(32) = 11001g(0)0$ but 11001 is not a suffix of any $g(a)$
- $g(21) = 1g(4)01100$ but 01100 is not a prefix of any $g(a)$
- $g(23) = 1g(4)10010$ but 10010 is not a prefix of any $g(a)$.

Since $g(434010) = 1100(01110010110001)^2 1100$, we need a special argument to rule this out. There are four special cases that must be handled:

- $g(43) = 1100g(0)10$
- $g(34) = 1100g(1)01$
- $g(01) = 01g(3)1100$
- $g(10) = 10g(4)1100$.

In the first example, $g(43) = 1100g(0)10$, since 10 is only a prefix of $g(1)$, we can extend on the right to get $g(43)1100 = 1100g(01)$. But since 1100 is only a prefix of $g(3)$ or $g(4)$, this gives either the forbidden pattern 33 or the forbidden pattern 434.

In the second example, $g(34) = 1100g(1)01$, since 01 is only a prefix of $g(0)$, we can extend on the right to get $g(34)1100 = 1100g(10)$. But 1100 is a suffix of only $g(0)$ and $g(1)$, so on the right we get either the forbidden pattern 010 or the forbidden pattern 11.

The other two cases are handled similarly.

□

Ochem [61] has recently given a construction that uses a uniform morphism to generate an infinite binary word with the Fraenkel–Simpson property (*i.e.*, it contains no squares except 0^2, 1^2, and $(01)^2$).

As in the previous section, we now define the substitution $h' : \Sigma_5^* \to 2^{\Sigma_5^*}$ as follows:

$$\begin{aligned} 0 &\to \{h(0), 012101232123401234321234\} \\ 1 &\to \{h(1)\} \\ 2 &\to \{h(2)\} \\ 3 &\to \{h(3)\} \\ 4 &\to \{h(4)\}\ . \end{aligned}$$

Thus, if $w \in \Sigma_5^*$, $h'(w)$ is a language of 2^r words over Σ_5, where $r = |w|_0$. Each of these words is of length $24|w|$.

Lemma 2.11. *Let g, h, and h' be defined as above. Let $w = h^m(0)$ for some positive integer m. Then $g(h'(w))$ is a language of $2^{n/1152}$ words over Σ_2, where $n = 144 \cdot 24^m$ is the length of each of these words. Furthermore, these words avoid all squares except 0^2, 1^2, and $(01)^2$.*

Proof. The proof is analogous to that of Lemma 2.7. Note that there are at least three 0's in every image word of h. Hence, $|w|_0 \geq \frac{1}{8}|w| = \frac{1}{8} \cdot 24^m$. We have then that $g(h'(w))$ consists of at least $2^{\frac{1}{8} \cdot 24^m}$ binary words. Since $n = 6 \cdot 24 \cdot 24^m$, we see that $g(h'(w))$ consists of at least $2^{n/1152}$ words.

To see that the words in $g(h'(w))$ avoid all squares except 0^2, 1^2, and $(01)^2$ it suffices by Theorem 2.10 to show that the words in $h'(w)$ are squarefree and contain no occurrences of the subwords 02, 03, 04, 13, 14, 20, 24, 30, 31, 41, 42, or 434010. The reader may easily verify that the words in $h'(w)$ contain no occurrences of the subwords 02, 03, 04, 13, 14, 20, 24, 30, 31, 41, 42, or 434010.

To show that the words in $h'(w)$ are squarefree, we will, as before, consider h' to be a morphism defined as follows:

$$\begin{array}{rcl} 0 & \to & h(0) \\ \hat{0} & \to & 01210123212340123432 1234 \\ 1 & \to & h(1) \\ 2 & \to & h(2) \\ 3 & \to & h(3) \\ 4 & \to & h(4) \ . \end{array}$$

There are no inclusions for h' other than the one identified in the proof of Theorem 2.9. There are three interchanges: referring to Lemma 2.2 (b), we have that

$$(a, b, c) \in \{(2, 1, \hat{0}), (2, 4, \hat{0}), (\hat{0}, 3, 2)\}$$

satisfies $h'(a) = st$, $h'(b) = uv$, and $h'(c) = sv$. We may rule out the first two cases by showing that w avoids all subwords of the form $1\alpha0\alpha2$ and $4\alpha0\alpha2$, where $\alpha \in \Sigma_5^*$. Note that in the word w, any occurrence of 0 must be followed by a 1, since w avoids the patterns 02, 03, and 04. Let x be a subword of w of the form $1\alpha0\alpha2$ or $4\alpha0\alpha2$. Then x must begin with 11 or 41. This is a contradiction, as w avoids both 11 and 41.

We may rule out the third case by showing that w avoids all subwords of the form $3\alpha2\alpha0$, where $\alpha \in \Sigma_5^*$. Note that in the word w, any occurrence of 2 must be followed by either 1 or 3, since w avoids the patterns 20 and 24. Let x be a subword of w of the form $3\alpha2\alpha0$. Then x must begin with 31 or 33. This is a contradiction, as w avoids both 31 and 33. □

Theorem 2.12. *Let H_n denote the number of binary words of length n that avoid all squares except 0^2, 1^2, and $(01)^2$. Then $H_n = \Omega(1.0006^n)$ and $H_n = O(1.135^n)$.*

Proof. The proof is analogous to that of Theorem 2.8. Noting that $2^{1/1152} \doteq 1.0006$, we see that the lower bound follows immediately from Lemma 2.11.

For the upper bound, we again used the `DAVID_IAN` Maple package for a list of 65 words of length ≤ 20:

$$0000, 1010, \ldots, 1110001011100010$$

obtaining a characteristic polynomial of degree 58 with dominant root $\doteq 1.135$. □

The following table gives the number H_n of binary words of length n containing only the squares 0^2, 1^2, and $(01)^2$.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
H_n	1	2	4	8	13	22	31	46	58	78	99	124	144	176	198	234	262	300

2.4 The Prodinger–Urbanek problem

Prodinger and Urbanek [63] stated that they were unable to find an example of two infinite binary words avoiding large squares such that their perfect shuffle had arbitrarily large squares. In this section we give an example of such words.

Theorem 2.13. *There exist two infinite binary words* $\mathbf{x}$ *and* $\mathbf{y}$ *such that neither* $\mathbf{x}$ *nor* $\mathbf{y}$ *contain a square* ww *with* $|w| \geq 4$*, but* $\mathbf{x} ⧢ \mathbf{y}$ *contains arbitrarily large squares.*

Proof. Consider the morphism $f : \Sigma_2^* \to \Sigma_2^*$ defined as follows:

$$\begin{aligned} f(0) &= 001 \\ f(1) &= 110 \ . \end{aligned}$$

We will show that $f^\omega(0) = 001001110001001110110110001\cdots$ contains arbitrarily large squares and is the perfect shuffle of two words, each avoiding squares ww with $|w| \geq 4$.

First, we define the morphisms $h : \Sigma_4^* \to \Sigma_4^*$, $g_1 : \Sigma_4^* \to \Sigma_2^*$, and $g_2 : \Sigma_4^* \to \Sigma_2^*$ defined as follows.

$$\begin{aligned} h(0) &= 012 & g_1(0) &= 001 & & & g_2(0) &= 010 \\ h(1) &= 302 & g_1(1) &= 101 & &\text{and} & g_2(1) &= 100 \\ h(2) &= 031 & g_1(2) &= 010 & & & g_2(2) &= 011 \\ h(3) &= 321 \ , & g_1(3) &= 110 \ , & & & g_2(3) &= 101 \ . \end{aligned}$$

We now show the following identity involving the morphisms h, g_1, and g_2.

Lemma 2.14. $f^\omega(0) = g_2(h^\omega(0)) ⧢ g_1(h^\omega(0))$.

Proof. We prove the following identities by induction on n:

$$f^{n+1}(00) = g_2(h^n(0)) ⧢ g_1(h^n(0)) \tag{2.1}$$

$$f^{n+1}(10) = g_2(h^n(1)) ⧢ g_1(h^n(1)) \tag{2.2}$$

$$f^{n+1}(01) = g_2(h^n(2)) ⧢ g_1(h^n(2)) \tag{2.3}$$

$$f^{n+1}(11) = g_2(h^n(3)) ⧢ g_1(h^n(3)) \ . \tag{2.4}$$

It is easy to verify that these equations hold for $n = 0$. We assume that they hold for $n = k$, where $k > 0$, and show that they hold for $n = k + 1$. We first consider $f^{k+2}(00)$, where we have

$$\begin{aligned} f^{k+2}(00) &= f^{k+1}(001001) \\ &= f^{k+1}(00)f^{k+1}(10)f^{k+1}(01) \\ &= \big(g_2(h^k(0)) ⧢ g_1(h^k(0))\big)\ \big(g_2(h^k(1)) ⧢ g_1(h^k(1))\big) \\ &\qquad \big(g_2(h^k(2)) ⧢ g_1(h^k(2))\big) \\ &= \big(g_2(h^k(0))g_2(h^k(1))g_2(h^k(2))\big) ⧢ \\ &\qquad \big(g_1(h^k(0))g_1(h^k(1))g_1(h^k(2))\big) \\ &= g_2(h^k(0)h^k(1)h^k(2)) ⧢ g_1(h^k(0)h^k(1)h^k(2)) \\ &= g_2(h^k(012)) ⧢ g_1(h^k(012)) \\ &= g_2(h^{k+1}(0)) ⧢ g_1(h^{k+1}(0)) \end{aligned}$$

as desired. The other cases of the induction for $f^{k+2}(10)$, $f^{k+2}(01)$, and $f^{k+2}(11)$ follow similarly. The result now follows from (2.1). □

We now prove

Lemma 2.15. *The infinite word $h^\omega(0)$ is squarefree.*

Proof. This follows immediately by the analogue of Lemma 2.2. An easy computation shows there are no inclusions or interchanges for h. □

We now define the set $\mathcal{A}$ as follows:

$$\mathcal{A} = \{010, 013, 021, 030, 032, 102, 121, 131, 202, 212, 231, 301, 303, 312, 320, 323\} \ .$$

Lemma 2.16. *(a) $h^\omega(0)$ contains no subwords x where $x \in \mathcal{A}$; and*

(b) $h^\omega(0)$ contains no subwords of the form $0\alpha 1\alpha 3$, $1\alpha 0\alpha 2$, $2\alpha 3\alpha 1$, or $3\alpha 2\alpha 0$, where $\alpha \in \Sigma_4^*$.

Proof. (a) This can be verified by inspection.

(b) We argue by contradiction. Let w be a shortest subword of $h^\omega(0)$ such that w is of the form $0\alpha 1\alpha 3$, $1\alpha 0\alpha 2$, $2\alpha 3\alpha 1$, or $3\alpha 2\alpha 0$. Suppose w is of the form $0\alpha 1\alpha 3$. Note that the only image words of h that contain the letter 1 are $h(0) = 012$, $h(2) = 031$, and $h(3) = 321$. Hence it must be the case that α is of the form $2\alpha'0$, $\alpha'03$, or $\alpha'32$ for some $\alpha' \in \Sigma_4^*$. We therefore have three cases.

Case 1: $w = 02\alpha'012\alpha'03$ for some $\alpha' \in \Sigma_4^*$. We have two subcases.

Case 1.i: $|w| \leq 12$. A short computation suffices to verify that, contrary to (a), all words w of the form $02\alpha'012\alpha'03$ with $|w| \leq 12$ contain either a square or a subword x where $x \in \mathcal{A}$.

Case 1.ii: $|w| > 12$. We first make the observation that any image word of h is uniquely specified by its first two letters and also by its last two letters. Thus, if $w = 02\alpha'012\alpha'03$, it must be the case that $3w1 = 302\alpha'012\alpha'031 = h(1)\alpha' h(0)\alpha' h(2)$ is also a subword of $h^\omega(0)$. Furthermore, since h has no inclusions, the infinite word $h^\omega(0)$ can be uniquely parsed into image words of h. Since we have that $h(1)\alpha' h(0)\alpha' h(2)$ is a subword of $h^\omega(0)$, this implies that $|\alpha'|$ is a multiple of 3 and that $h(1)\alpha' h(0)\alpha' h(2) = h(1\beta 0\beta 2)$ for some $\beta \in \Sigma_4^*$, $|\beta| < |\alpha'|$. So $1\beta 0\beta 2$ must also be a subword of $h^\omega(0)$. This contradicts the minimality of w.

Case 2: $w = 0\alpha'031\alpha'033$ for some $\alpha' \in \Sigma_4^*$. But then w contains the square 33, contrary to Lemma 2.15.

Case 3: $w = 0\alpha'321\alpha'323$ for some $\alpha' \in \Sigma_4^*$. But by (a) w cannot contain the subword 323.

The cases where w is of the form $1\alpha 0\alpha 2$, $2\alpha 3\alpha 1$, or $3\alpha 0\alpha 2$ follow similarly.

□

We now give the analogue of Lemma 2.2 for g_1 and g_2. Let g_i represent either g_1 or g_2. Then we have

Lemma 2.17. *(a) Suppose $g_i(ab) = tg_i(c)u$ for some letters $a, b, c \in \Sigma_4$ and words $t, u \in \Sigma_2^*$. Then at least one of the following holds:*

(i) this inclusion is trivial (that is, $t = \epsilon$ or $u = \epsilon$);

(ii) u is not a prefix of $g_i(d)$ for any $d \in \Sigma_4$;

(iii) t is not a suffix of $g_i(d)$ for any $d \in \Sigma_4$; or

(iv) for all $v, w \in \Sigma_2^*$ and all $e, e' \in \Sigma_4$, if $vg_i(ab)w = g_i(ece')$, then at least one of the following holds:

(A) this inclusion is trivial (that is, $v = \epsilon$ or $w = \epsilon$);

(B) w is not a prefix of $g_i(d)$ for any $d \in \Sigma_4$;

(C) v is not a suffix of $g_i(d)$ for any $d \in \Sigma_4$;

(D) either $e = c$ or $e' = c$;

(E) $ece' \in \mathcal{A}$;

(F) for all $x, y \in \Sigma_2$ and all $k \in \Sigma_4$, if $g_i(kab)x = yg_i(ece')$, then $k = a$; or

(G) for all $x, y \in \Sigma_2$ and all $k \in \Sigma_4$, if $xg_i(abk) = g_i(ece')y$, then $k = b$.

(b) Suppose there exist letters $a, b, c \in \Sigma_4$ and words $s, t, u, v \in \Sigma_2^*$ such that $g_i(a) = st$, $g_i(b) = uv$, $g_i(c) = sv$, and $b\alpha c\alpha a$ is a subword of $h^\omega(0)$ for some $\alpha \in \Sigma_4^*$. Then either $a = c$ or $b = c$.

Proof. (a) We give one example of each case and list the other non-trivial cases in a table below.

(i) Trivial.

(ii) $g_2(32) = 1g_2(0)11$, but 11 is not a prefix of $g_2(d)$ for any $d \in \Sigma_4$.

(iii) $g_1(02) = 00g_1(1)0$, but 00 is not a suffix of $g_1(d)$ for any $d \in \Sigma_4$.

(iv) (A) Trivial.

(B) $g_2(01) = 01g_2(0)0$ and $1g_2(01)11 = g_2(302)$, but 11 is not a prefix of $g_2(d)$ for any $d \in \Sigma_4$.

(C) $g_1(31) = 1g_1(1)01$ and $00g_1(31)0 = g_1(012)$, but 00 is not a suffix of $g_1(d)$ for any $d \in \Sigma_4$.

(D) $g_1(23) = 0g_1(1)10$ and $01g_1(23)1 = g_1(211)$, but $e' = c = 1$.

(E) $g_1(21) = 0g_1(1)01$ and $01g_1(21)0 = g_1(212)$, but $ece' = 212 \in \mathcal{A}$.

(F) $g_2(30) = 1g_2(0)10$, $01g_2(30)0 = g_2(201)$, and $g_2(330)0 = 1g_2(201)$, but $k = a = 3$.

(G) $g_1(12) = 10g_1(1)0$, $0g_1(12)01 = g_1(210)$, and $0g_1(122) = g_1(210)0$, but $k = b = 2$.

Table 2.1: Case analysis for Lemma 2.17

Case	$g_i(ab) = tg_i(c)u$	$vg_i(ab)w = g_i(ecj)$	$g_i(kab)x = yg_i(ecj)$ or $xg_i(abk) = g_i(ecj)y$
a.ii	$g_2(01) = 0g_2(3)\mathbf{00}$	-	-
	$g_2(02) = 0g_2(1)\mathbf{11}$	-	-
	$g_2(31) = 1g_2(2)\mathbf{00}$	-	-
	$g_2(32) = 1g_2(0)\mathbf{11}$	-	-
a.iii	$g_1(01) = \mathbf{00}g_1(3)1$	-	-
	$g_1(02) = \mathbf{00}g_1(1)0$	-	-
	$g_1(31) = \mathbf{11}g_1(2)1$	-	-
	$g_1(32) = \mathbf{11}g_1(0)0$	-	-
a.iv.B	$g_2(01) = 01g_2(0)0$	$1g_2(01)\mathbf{11} = g_2(302)$	-
	$g_2(32) = 10g_2(3)1$	$0g_2(32)\mathbf{00} = g_2(031)$	-
a.iv.C	$g_1(02) = 0g_1(2)10$	$\mathbf{11}g_1(02)1 = g_1(321)$	-
	$g_1(31) = 1g_1(1)01$	$\mathbf{00}g_1(31)0 = g_1(012)$	-
a.iv.D	$g_1(10) = 1g_1(2)01$	$00g_1(02)0 = g_1(0\mathbf{22})$	-
	$g_1(10) = 1g_1(2)01$	$10g_1(02)0 = g_1(1\mathbf{22})$	-
	$g_1(02) = 0g_1(2)10$	$01g_1(02)1 = g_1(\mathbf{22}1)$	-
	$g_1(23) = 0g_1(1)10$	$01g_1(23)1 = g_1(2\mathbf{11})$	-
	$g_1(23) = 0g_1(1)10$	$11g_1(23)1 = g_1(3\mathbf{11})$	-
	$g_1(31) = 1g_1(1)01$	$10g_1(31)0 = g_1(\mathbf{11}2)$	-

Table 2.1: Case analysis for Lemma 2.17 continued

Case	$g_i(ab) = tg_i(c)u$	$vg_i(ab)w = g_i(ecj)$	$g_i(kab)x = yg_i(ecj)$ or $xg_i(abk) = g_i(ecj)y$
	$g_2(01) = 1g_2(2)01$	$1g_2(01)10 = g_2(3\mathbf{00})$	-
	$g_2(13) = 1g_2(2)01$	$0g_2(13)00 = g_2(\mathbf{00}1)$	-
	$g_2(13) = 0g_2(2)10$	$0g_2(13)01 = g_2(\mathbf{00}3)$	-
	$g_2(20) = 0g_2(1)10$	$1g_2(20)10 = g_2(\mathbf{33}0)$	-
	$g_2(20) = 0g_2(1)10$	$1g_2(20)11 = g_2(\mathbf{33}2)$	-
	$g_2(32) = 10g_2(3)1$	$0g_2(32)01 = g_2(0\mathbf{33})$	-
a.iv.E	$g_1(12) = 10g_1(1)0$	$0g_1(12)10 = g_1(\mathbf{212})$	-
	$g_1(12) = 10g_1(1)0$	$1g_1(12)10 = g_1(\mathbf{312})$	-
	$g_1(12) = 1g_1(2)10$	$00g_1(12)1 = g_1(\mathbf{021})$	-
	$g_1(12) = 1g_1(2)10$	$10g_1(12)1 = g_1(\mathbf{121})$	-
	$g_1(21) = 0g_1(1)01$	$01g_1(21)0 = g_1(\mathbf{212})$	-
	$g_1(21) = 0g_1(1)01$	$11g_1(21)0 = g_1(\mathbf{312})$	-
	$g_1(21) = 01g_1(2)1$	$0g_1(21)01 = g_1(\mathbf{021})$	-
	$g_1(21) = 01g_1(2)1$	$1g_1(21)01 = g_1(\mathbf{121})$	-
	$g_2(03) = 01g_2(0)1$	$1g_2(03)00 = g_2(\mathbf{301})$	-
	$g_2(03) = 01g_2(0)1$	$1g_2(03)01 = g_2(\mathbf{303})$	-
	$g_2(03) = 0g_2(3)01$	$01g_2(03)0 = g_2(\mathbf{030})$	-
	$g_2(03) = 0g_2(3)01$	$01g_2(03)1 = g_2(\mathbf{032})$	-
	$g_2(30) = 1g_2(0)10$	$10g_2(30)0 = g_2(\mathbf{301})$	-
	$g_2(30) = 1g_2(0)10$	$10g_2(30)1 = g_2(\mathbf{303})$	-
	$g_2(30) = 10g_2(3)0$	$0g_2(30)10 = g_2(\mathbf{030})$	-
	$g_2(30) = 10g_2(3)0$	$0g_2(30)11 = g_2(\mathbf{032})$	-
a.iv.F	$g_2(03) = 0g_2(3)01$	$10g_2(03)0 = g_2(130)$	$g_2(\mathbf{00}3)0 = 0g_2(130)$
	$g_2(03) = 0g_2(3)01$	$10g_2(03)1 = g_2(132)$	$g_2(\mathbf{00}3)1 = 0g_2(132)$
	$g_2(30) = 1g_2(0)10$	$01g_2(30)0 = g_2(201)$	$g_2(\mathbf{33}0)0 = 1g_2(201)$
	$g_2(30) = 1g_2(0)10$	$01g_2(30)1 = g_2(203)$	$g_2(\mathbf{33}0)1 = 1g_2(203)$
a.iv.G	$g_1(12) = 10g_1(1)0$	$0g_1(12)01 = g_1(210)$	$0g_1(1\mathbf{22}) = g_1(210)0$

Table 2.1: Case analysis for Lemma 2.17 continued

Case	$g_i(ab) = tg_i(c)u$	$vg_i(ab)w = g_i(ecj)$	$g_i(kab)x = yg_i(ecj)$ or $xg_i(abk) = g_i(ecj)y$
	$g_1(12) = 10g_1(1)0$	$1g_1(12)01 = g_1(310)$	$1g_1(1\mathbf{22}) = g_1(310)0$
	$g_1(21) = 01g_1(2)1$	$0g_1(21)10 = g_1(023)$	$0g_1(2\mathbf{11}) = g_1(023)1$
	$g_1(21) = 01g_1(2)1$	$1g_1(21)10 = g_1(123)$	$1g_1(2\mathbf{11}) = g_1(123)1$

(b) The only a, b, c that satisfy $g_1(a) = st$, $g_1(b) = uv$, and $g_1(c) = sv$ such that $a \neq c$ and $b \neq c$ are $(a, b, c) \in \{(0, 3, 2), (1, 2, 3), (2, 1, 0), (3, 0, 1)\}$. But by Lemma 2.16, the infinite word $h^\omega(0)$ contains no subwords of the form $3\alpha2\alpha0$, $2\alpha3\alpha1$, $1\alpha0\alpha2$, or $0\alpha1\alpha3$. This contradicts the assumption that $b\alpha c\alpha a$ is a subword of $h^\omega(0)$ for some $\alpha \in \Sigma_4^*$. The same result holds true for g_2.

□

Lemma 2.18. *Neither $g_1(h^\omega(0))$ nor $g_2(h^\omega(0))$ contain squares yy with $|y| \geq 4$.*

Proof. As in the case of Lemma 2.2, this follows from Lemma 2.17 with the following modification. Recall from the proof of Theorem 2.1 that the main idea of the argument involved exhaustively checking a finite number of squares yy, where $|y| \leq l$ for some constant l (for Theorem 2.1, $l = 20$), and then applying Lemma 2.2 to check an infinite number of squares yy, where $|y| > l$. To determine the value of l, recall that in order to apply Lemma 2.2 (a), it must have been the case that y contained a subword $h(c)$ such that $h(ab) = th(c)u$. Since the h of Theorem 2.1 was 10-uniform, in order to ensure that y contained at least one such subword $h(c)$, it sufficed to consider those y where $|y| > 2 \cdot 10 = 20$. However, for Lemma 2.18 we sometimes require something stronger, for example, that y contains a subword $g_i(kab)$ such that $g_i(kab)x = zg_i(ece')$. Since g_i is 3-uniform, in order to ensure that y contains such a subword $g_i(kab)$, it suffices to consider those y where $|y| > 4 \cdot 3 = 12$. Hence, the proof breaks into two cases: the case where $|y| \leq 12$ and the case where $|y| > 12$. The remainder of the argument exactly parallels that of Theorem 2.1. □

We can now complete the proof of Theorem 2.13. Let

$$\mathbf{x} := g_2(h^\omega(0)) = 010100011101010011\cdots$$

and

$$\mathbf{y} := g_1(h^\omega(0)) = 001101010110001010\cdots .$$

Then by Lemma 2.14 we have $\mathbf{x} \sqcup\!\sqcup\, \mathbf{y} = f^\omega(0)$. But $f^\omega(0) = f^\omega(001)$ and so $f^\omega(0)$ begins with $f^n(0)f^n(0)$ for all $n \geq 0$. Hence $f^\omega(0)$ begins with an arbitrarily large square.

On the other hand, by Lemma 2.18, we have that $\mathbf{x}$ and $\mathbf{y}$ avoid all squares ww with $|w| \geq 4$. □

2.5 Open problems

Entringer, Jackson, and Schatz [33] also showed that, in contrast with the case of ordinary squares, all infinite binary words contain arbitrarily large Abelian squares (recall the definition of *Abelian square* given in Chapter 1). Specifically, they proved the following theorem.

Theorem 2.19 (Entringer, Jackson, and Schatz). *Let $k > 0$ be a fixed integer constant and let w be a binary word containing no Abelian squares xy, where $|x| = |y| \geq k$. Then $|w| < k^2 + 6k$.*

Mäkelä [54] (see also Keränen [48]) has posed the following open problems.

Problem 2.20. *Does there exist an infinite ternary word containing no Abelian squares xy, where $|x| = |y| \geq 2$?*

Problem 2.21. *Does there exist an infinite binary word containing no Abelian cubes xyz, where $|x| = |y| = |z| \geq 2$?*

Computer searches have yielded words of length at least 3160 satisfying the conditions of Problem 2.20. Computer searches performed by Mäkelä have yielded words of length at least 250 satisfying the conditions of Problem 2.21.

Such computer searches can be performed in the following way. Let $P \subseteq \Sigma^*$ be a set of subwords to be avoided. We create a tree T as follows. The root of the tree is labeled by the empty word, ϵ. For any node labeled w in T, if $w \notin P$, then we create child nodes labeled wa for each $a \in \Sigma$. The leaves of T are thus the nodes labeled w, where w contains a subword $x \in P$. An example of such a tree T, where P is the set of squares over Σ_2, is given in Figure 2.4. Note that the tree T given in Figure 2.4 is symmetric under permutation of the alphabet symbols 0 and 1. We can therefore label the root by 0 (or, equivalently, 1) without losing any information. The resulting tree T' is given in Figure 2.5.

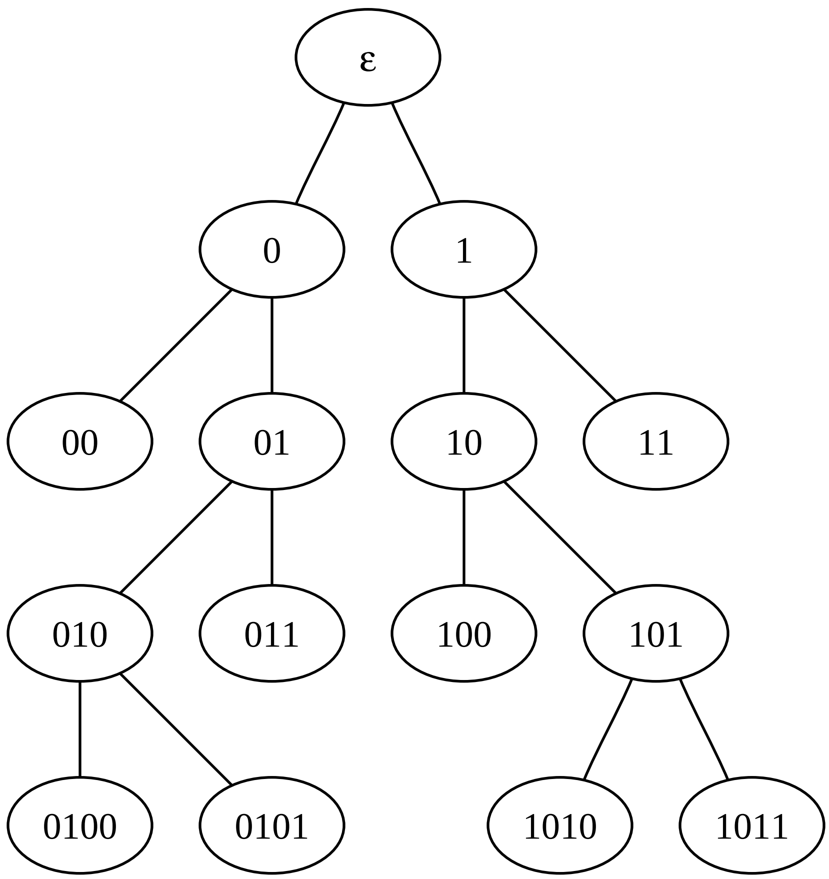

Figure 2.4: Tree T for $P = \{xx \mid x \in \Sigma_2^+\}$

If T is finite, then it has a height h; moreover, all words $w \in \Sigma^*$ with $|w| \geq h$ contain

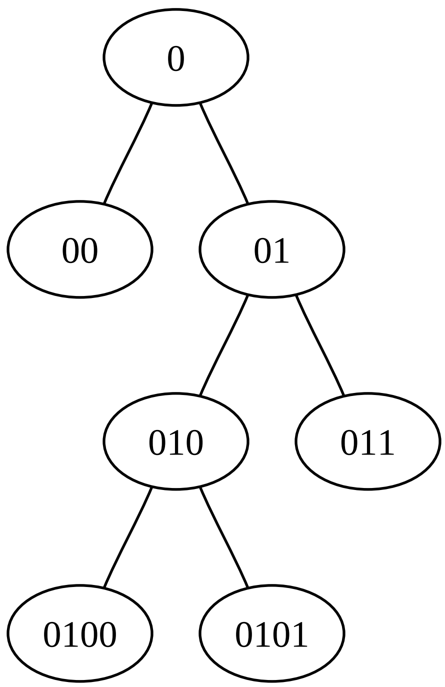

Figure 2.5: Tree T' (with root labeled 0) for $P = \{xx \mid x \in \Sigma_2^+\}$

some subword $x \in P$. We may verify that T is finite by a depth-first traversal of T.[2]

If, on the other hand, T is infinite, then there exists an infinite word over Σ that avoids all $x \in P$ (see Theorem 1.1). A depth-first search of such an infinite T will eventually visit non-leaf nodes with longer and longer labels. The order in which these nodes are visited depends on how the children of each node have been ordered. If the children of each node in T have been ordered from left to right based on the lexicographical ordering of their labels (as in Figure 2.5), then for any n we can perform a depth-first traversal of T to find the lexicographically least word of length n that avoids all $x \in P$. We call such a traversal a *lexicographical depth-first search*.

A lexicographical depth-first search is often very fast when there are many infinite words that avoid all $x \in P$; however, if there are very few such words, a lexicographical depth-first search may instead spend a great deal of time backtracking through T, and so may run rather slowly. Sometimes it is possible to improve the performance of the search by changing the order in which we create child nodes in T. For instance, suppose at each level i of the tree T we choose a random permutation σ_i of the alphabet symbols in Σ and then order the children of each node at level i from left to right based on the permutation σ_i. We call a depth-first traversal of a tree ordered in such a way a *random depth-first search*.

A random depth-first search often succeeds in finding a word of length n that avoids all $x \in P$ much faster than a lexicographical depth-first search. Note, however, that the word found by a random depth-first search will now no longer necessarily be the lexicographically least such word. We used a random depth-first search to determine that, as mentioned previously, there are words of length at least 3160 satisfying the conditions of Problem 2.20.

[2] Actually, any type of tree traversal will suffice; however, we will be exclusively concerned with depth-first search here.

Chapter 3

Squarefree Words over Three Symbols

Recall from Chapter 1 that Thue gave an example of a infinite squarefree ternary word. There are certain words that are avoidable in infinite squarefree ternary words and others that are unavoidable; *e.g.*, the word 101 is avoidable, whereas the word 012 is not. In this chapter we characterize all words that can be avoided in infinite squarefree ternary words.[1]

3.1 Characterization of the avoidable words

Theorem 3.1. *Let w be any infinite squarefree word over Σ_3. Then w contains at least one occurrence of each of the following words:* 012*,* 021*,* 102*,* 120*,* 201*,* 210*.*

Proof. This can be verified by an exhaustive computer search. It suffices to check all 34422 squarefree words of length 30 over Σ_3. ☐

Theorem 3.2. *Let a, b, and c be distinct letters of Σ_3. Then there exists an infinite squarefree word over Σ_3 that contains no occurrences of each of the words $abca$ and $acba$.*

Proof. It is easy to see that $(abca, acba) \in \{(0120, 0210), (1021, 1201), (2012, 2102)\}$. Hence it suffices to show that there exists an infinite squarefree word over Σ_3 that avoids 0120

[1] The author recently discovered that Theorems 3.1 and 3.2 were previously known to Thue [76].

and 0210, as we may simply rename a, b, and c to get the desired avoidance. Consider the morphism h defined as follows:

$$\begin{aligned} 0 &\rightarrow 12 \\ 1 &\rightarrow 102 \\ 2 &\rightarrow 0 \ . \end{aligned}$$

Then the fixed point $h^\omega(0)$ is squarefree (see Istrail [43], also Dekking [29]) and avoids 101 and 202. The only way to obtain 0120 from the morphism h is $h(202) = 0120$, but $h^\omega(0)$ avoids 202. Similarly, the only way to obtain 0210 is as a subword of $h(11) = 102102$, but $h^\omega(0)$ avoids the square 11. The result now follows. □

Theorem 3.3. *Let x be any word over Σ_3 such that $|x| \geq 4$. Then there exists an infinite squarefree word over Σ_3 that contains no occurrences of x.*

Proof. It suffices to prove the theorem for $|x| = 4$. Consider the set $\mathcal{A}$ of all squarefree words of length 4 over Σ_3. We have $\mathcal{A} = \mathcal{A}' \cup \mathcal{A}''$, where

$$\mathcal{A}' = \{0102, 0121, 0201, 0212, 1012, 1020, 1202, 1210, 2010, 2021, 2101, 2120\}$$

and

$$\mathcal{A}'' = \{0120, 0210, 1021, 1201, 2012, 2102\} \ .$$

Note that all words in $\mathcal{A}'$ contain a subword of the form aba, where a and b are distinct letters of Σ_3. It is well known that for any such subword aba, there exists an infinite squarefree word over Σ_3 that avoids aba (see Thue [75, Satz 5]). Hence, for any word $x \in \mathcal{A}'$, there exists an infinite squarefree word over Σ_3 that avoids x.

Now consider the set $\mathcal{A}''$. Note that all words in $\mathcal{A}''$ are of the form $abca$, where a, b, and c are distinct letters of Σ_3. Theorem 3.2 implies that for any such word $abca$, there exists an infinite squarefree word over Σ_3 that avoids $abca$. Hence, for any word $x \in \mathcal{A}$, there exists an infinite squarefree word over Σ_3 that avoids x. The result now follows. □

Chapter 4

Words Avoiding Reversed Subwords

In this chapter[1] we examine a question of Szilard regarding words w satisfying the property that for any subword w' of w, w does not contain the reversal of w' as a subword.

4.1 Definitions and notation

If $w \in \Sigma^*$ is written $w = w_1 w_2 \cdots w_n$, where each $w_i \in \Sigma$, then the *reversal* of w, denoted w^R, is the word $w_n w_{n-1} \cdots w_1$.

If $y \in \Sigma^*$ is a nonempty word, then the word $yyy\cdots$ is written as y^ω. If an infinite word $\mathbf{w}$ can be written in the form y^ω for some $y \in \Sigma^*$, then $\mathbf{w}$ is said to be *periodic*. If $\mathbf{w}$ can be written in the form $y'y^\omega$ for some $y, y' \in \Sigma^*$, then $\mathbf{w}$ is said to be *ultimately periodic*.

4.2 Avoiding reversed subwords

Szilard [74] asked the following question:

[1] The contents of this chapter are taken largely verbatim from Rampersad and Shallit [65]. For the most part, the formulation of the theorems in this chapter are due to J. Shallit; the contribution of the present author lies largely in the proofs.

> Does there exist an infinite word $\mathbf{w}$ such that if x is a subword of $\mathbf{w}$, then x^R is not a subword of $\mathbf{w}$?

Clearly there must be some restriction on the length of x: if $|x| = 1$, then all nonempty words fail to have the desired property. For $|x| \geq 2$, however, we have the following result.

Theorem 4.1. *There exists an infinite word* $\mathbf{w}$ *over* Σ_3 *such that if* x *is a subword of* $\mathbf{w}$ *and* $|x| \geq 2$*, then* x^R *is not a subword of* $\mathbf{w}$*. Furthermore,* $\mathbf{w}$ *is unique up to permutation of the alphabet symbols.*

Proof. Note that if $|x| \geq 3$ and both x and x^R are subwords of $\mathbf{w}$, then there is a prefix x' of x such that $|x'| = 2$ and $(x')^R$ is a suffix of x^R. Hence it suffices to show the theorem for $|x| = 2$. We show that the infinite word

$$\mathbf{w} = (012)^\omega = 012012012012\cdots$$

has the desired property. To see this, consider the set $\mathcal{A}$ consisting of all subwords of $\mathbf{w}$ of length two. We have $\mathcal{A} = \{01, 12, 20\}$. Noting that if $x \in \mathcal{A}$, then $x^R \notin \mathcal{A}$, we conclude that if x is a subword of $\mathbf{w}$ and $|x| \geq 2$, then x^R is not a subword of $\mathbf{w}$.

To see that $\mathbf{w}$ is unique up to permutation of the alphabet symbols, consider another word $\mathbf{w}'$ satisfying the conditions of the theorem, and suppose that $\mathbf{w}'$ begins with 01. Then 01 must be followed by 2, 12 must be followed by 0, and 20 must be followed by 1. Hence,

$$\mathbf{w}' = (012)^\omega = 012012012012\cdots = \mathbf{w} \ .$$

□

Note that the solution given in the proof of Theorem 4.1 is periodic. In the following theorem, we give a nonperiodic solution to this problem for $|x| \geq 3$.

Theorem 4.2. *There exists an infinite nonperiodic word* $\mathbf{w}$ *over* Σ_3 *such that if* x *is a subword of* $\mathbf{w}$ *and* $|x| \geq 3$*, then* x^R *is not a subword of* $\mathbf{w}$*.*

Proof. By reasoning similar to that given in the proof of Theorem 4.1, it suffices to show the theorem for $|x| = 3$. Let $\mathbf{w}'$ be an infinite nonperiodic word over Σ_2. For example, if

$\mathbf{w}' = 11010010001\cdots$, then $\mathbf{w}'$ is nonperiodic. Define the morphism $h : \Sigma_2^\omega \to \Sigma_3^\omega$ by

$$\begin{aligned} 0 &\to 0012 \\ 1 &\to 0112 \ . \end{aligned}$$

Then $\mathbf{w} = h(\mathbf{w}')$ has the desired property. Consider the set $\mathcal{A}$ consisting of all subwords of $\mathbf{w}$ of length three. We have

$$\mathcal{A} = \{001, 011, 012, 112, 120, 200, 201\} \ .$$

Noting that if $x \in \mathcal{A}$, then $x^R \notin \mathcal{A}$, we conclude that if x is a subword of $\mathbf{w}$ and $|x| \geq 3$, then x^R is not a subword of $\mathbf{w}$.

To see that $\mathbf{w}$ is not periodic, suppose the contrary; *i.e.*, suppose that $\mathbf{w} = y^\omega$ for some $y \in \Sigma_3^*$. Clearly, $|y| > 4$. Suppose then that y begins with $h(0)$. Noting that the only way to obtain 00 from $h(ab)$, where $a, b \in \Sigma_2$, is as a prefix of $h(0)$, we see that $y = h(y')$ for some $y' \in \Sigma_2^*$. Hence, $\mathbf{w} = (h\,(y'))^\omega = h\,((y')^\omega)$, and so $\mathbf{w}' = (y')^\omega$ is periodic, contrary to our choice of $\mathbf{w}'$. □

Over a two letter alphabet we have the following negative result.

Theorem 4.3. *Let $k \leq 4$ and let w be a word over Σ_2 such that if x is a subword of w and $|x| \geq k$, then x^R is not a subword of w. Then $|w| \leq 8$.*

Proof. As mentioned previously, if $k = 1$ the result holds trivially. If $k = 2$, note that all binary words of length at least three must contain one of the following words: 00, 11, 010, or 101. Similarly, if $k = 3$, note that all binary words of length at least five must contain one of the following words: 000, 010, 101, 111, 0110, or 1001; and if $k = 4$, note that all binary words of length at least nine must contain one of the following words: 0000, 0110, 1001, 1111, 00100, 01010, 01110, 10001, 10101, or 11011. Hence, $|w| \leq 8$, as required. □

For $|x| \geq 5$, however, we find that there *are* infinite words with the desired property.

Theorem 4.4. *There exists an infinite word $\mathbf{w}$ over Σ_2 such that if x is a subword of $\mathbf{w}$ and $|x| \geq 5$, then x^R is not a subword of $\mathbf{w}$.*

Proof. By reasoning similar to that given in the proof of Theorem 4.1, it suffices to show the theorem for $|x| = 5$. We show that the infinite word

$$\mathbf{w} = (001011)^\omega = 001011001011001011\cdots$$

has the desired property. To see this, consider the set $\mathcal{A}$ consisting of all subwords of $\mathbf{w}$ of length five. We have

$$\mathcal{A} = \{00101, 01011, 01100, 10010, 10110, 11001\} \ .$$

Noting that if $x \in \mathcal{A}$, then $x^R \notin \mathcal{A}$, we conclude that if x is a subword of $\mathbf{w}$ and $|x| \geq 5$, then x^R is not a subword of $\mathbf{w}$. □

Let z be the word 001011. We denote the *complement* of z by $\bar{z}$, *i.e.*, the word obtained by substituting 0 for 1 and 1 for 0 in z. Let $\mathcal{B}$ be the set defined as follows:

$$\mathcal{B} = \{x \,|\, x \text{ is a cyclic shift of } z \text{ or } \bar{z}\} \ .$$

We have the following characterization of the words satisfying the conditions of Theorem 4.4.

Theorem 4.5. *Let* $\mathbf{w}$ *be an infinite word over* Σ_2 *such that if* x *is a subword of* $\mathbf{w}$ *and* $|x| \geq 5$*, then* x^R *is not a subword of* $\mathbf{w}$*. Then* $\mathbf{w}$ *is ultimately periodic. Specifically,* $\mathbf{w}$ *is of the form* $y'y^\omega$*, where* $y' \in \{\epsilon, 0, 1, 00, 11\}$ *and* $y \in \mathcal{B}$*.*

Proof. By reasoning similar to that given in the proof of Theorem 4.1, it suffices to show the theorem for $|x| = 5$. We call a word $w \in \Sigma_2^*$ *valid* if w satisfies the property that if x is a subword of w and $|x| = 5$, then x^R is not a subword of w. We have the following two facts, which may be verified computationally.

1. All valid words of length 9 are of the form $y'yy''$, where $y' \in \{\epsilon, 0, 1, 00, 11\}$, $y \in \mathcal{B}$, and $y'' \in \Sigma_2^*$.

2. Let w be a valid word of the form yy'', where $y \in \mathcal{B}$ and $y'' \in \Sigma_2^*$. Then if $|w| = 15$, y is a prefix of y''.

We will prove by induction on n that for all $n \geq 1$, $y'y^n$ is a prefix of $\mathbf{w}$, where $y' \in \{\epsilon, 0, 1, 00, 11\}$ and $y \in \mathcal{B}$.

If $n = 1$, then by applying the first fact to the prefix of $\mathbf{w}$ of length 9, we have that $y'y$ is a prefix of $\mathbf{w}$, as required.

Assume then that $y'y^n$ is a prefix of $\mathbf{w}$. We can thus write $\mathbf{w} = y'y^{n-1}y\mathbf{w}'$, for some $\mathbf{w}' \in \Sigma_2^\omega$. By applying the second fact to the prefix of $y\mathbf{w}'$ of length 15, we have that y is a prefix of $\mathbf{w}'$. Hence $\mathbf{w} = y'y^{n-1}yy\mathbf{w}'' = y'y^{n+1}\mathbf{w}''$, for some $\mathbf{w}'' \in \Sigma_2^\omega$, as required.

We therefore conclude that if $\mathbf{w}$ satisfies the conditions of the theorem, then $\mathbf{w}$ is of the form $y'y^\omega$, where $y' \in \{\epsilon, 0, 1, 00, 11\}$ and $y \in \mathcal{B}$. □

Next we give a nonperiodic solution to this problem for $|x| \geq 6$.

Theorem 4.6. *There exists an infinite nonperiodic word* $\mathbf{w}$ *over* Σ_2 *such that if* x *is a subword of* $\mathbf{w}$ *and* $|x| \geq 6$*, then* x^R *is not a subword of* $\mathbf{w}$*.*

Proof. By reasoning similar to that given in the proof of Theorem 4.1, it suffices to show the theorem for $|x| = 6$. Let $\mathbf{w}'$ be an infinite nonperiodic word over Σ_2. Define the morphism $h : \Sigma_2^\omega \to \Sigma_2^\omega$ by

$$\begin{aligned} 0 &\to 0001011 \\ 1 &\to 0010111 \ . \end{aligned}$$

We show that the infinite word $\mathbf{w} = h(\mathbf{w}')$ has the desired property. To see this, consider the set $\mathcal{A}$ consisting of all subwords of $\mathbf{w}$ of length six. We have

$$\begin{aligned} \mathcal{A} = \{&000101, 001011, 010110, 010111, 011000, 011001, 011100, \\ &100010, 100101, 101100, 101110, 110001, 110010, 111000, 111001\} \ . \end{aligned}$$

Noting that if $x \in \mathcal{A}$, then $x^R \notin \mathcal{A}$, we conclude that if x is a subword of $\mathbf{w}$ and $|x| \geq 6$, then x^R is not a subword of $\mathbf{w}$.

To see that $\mathbf{w}$ is not periodic, suppose the contrary; *i.e.*, suppose that $\mathbf{w} = y^\omega$ for some $y \in \Sigma_2^*$. Clearly, $|y| > 7$. Suppose then that y begins with $h(0)$. Noting that the only way to obtain 000 from $h(ab)$, where $a, b \in \Sigma_2$, is as a prefix of $h(0)$, we see that $y = h(y')$ for some $y' \in \Sigma_2^*$. Hence, $\mathbf{w} = (h(y'))^\omega = h((y')^\omega)$, and so $\mathbf{w}' = (y')^\omega$ is periodic, contrary to our choice of $\mathbf{w}'$. □

Finally we consider words avoiding squares as well as reversed subwords. Over a four letter alphabet we have the following negative result, which may be verified computationally.

Theorem 4.7. *Let w be a squarefree word over Σ_4 such that if x is a subword of w and $|x| \geq 2$, then x^R is not a subword of w. Then $|w| \leq 20$.*

In contrast with the result of Theorem 4.7, Alon *et al.* [7] have noted that over a four letter alphabet there exists an infinite squarefree word that avoids palindromes x where $|x| \geq 2$ (see Theorem 5.5). (A *palindrome* is a word x such that $x = x^R$.) However, over a five letter alphabet there are infinite words with an even stronger avoidance property.

Theorem 4.8. *There exists an infinite squarefree word $\mathbf{w}$ over Σ_5 such that if x is a subword of $\mathbf{w}$ and $|x| \geq 2$, then x^R is not a subword of $\mathbf{w}$.*

Proof. By reasoning similar to that given in the proof of Theorem 4.1, it suffices to show the theorem for $|x| = 2$. Let $\mathbf{w}'$ be an infinite squarefree word over Σ_3. Define the morphism $h : \Sigma_3^\omega \to \Sigma_5^\omega$ by

$$\begin{array}{ccc} 0 & \to & 012 \\ 1 & \to & 013 \\ 2 & \to & 014 \ . \end{array}$$

We show that the infinite word $\mathbf{w} = h(\mathbf{w}')$ has the desired property.

First we note that to verify that $\mathbf{w}$ is squarefree, it suffices by a theorem of Thue [76] (see also Bean, Ehrenfeucht, and McNulty [10], Berstel [12], and Crochemore [25]) to verify that $h(w)$ is squarefree for all 12 squarefree words $w \in \Sigma_3^*$ such that $|w| = 3$. This is left to the reader.

To see that if x is a subword of $\mathbf{w}$ and $|x| = 2$, then x^R is not a subword of $\mathbf{w}$, consider the set $\mathcal{A}$ consisting of all subwords of $\mathbf{w}$ of length 2. We have

$$\mathcal{A} = \{01, 12, 13, 14, 20, 30, 40\} \ .$$

Noting that if $x \in \mathcal{A}$, then $x^R \notin \mathcal{A}$, we conclude that if x is a subword of $\mathbf{w}$ and $|x| \geq 2$, then x^R is not a subword of $\mathbf{w}$. □

Chapter 5

Non-repetitive Colouring of Graphs

In the chapter we consider a generalization of squarefreeness from words to graph colourings. We also examine some open problems posed by Alon *et al.* [7] and Grytczuk [40].

5.1 Graph theoretic preliminaries

In this section we review some basic terminology from graph theory. There are numerous texts on graph theory (see, for example, Bollobás [14] or Diestel [31]).

A graph G is a pair (V, E) of sets. The set V is called the *vertex set* of G, and the set E is called the *edge set* of G. The edge set E consists of unordered pairs of vertices $\{u, v\}$, where $u, v \in V$. A subgraph G' of a graph G is a pair (V', E'), where $V' \subseteq V$, $E' \subseteq E$, and if $v \in V$ is contained in some $e \in E'$, then $v \in V'$.

A *walk* in G is a sequence of vertices $v_1, \ldots, v_k$ such that $\{v_i, v_{i+1}\} \in E$ for $1 \leq i \leq k-1$. A *path* is a walk where all the vertices are distinct. A *cycle* is a walk where all the vertices are distinct except $v_1 = v_k$. A graph G is *connected* if there is a path between any pair of vertices in G. A *connected component* of a graph G is a maximal connected subgraph of G.

A *forest* is a graph with no cycles. A *tree* is a connected graph with no cycles.

A *planar* graph is a graph that can be drawn in the plane such that no two edges intersect except at their endpoints. Such a drawing we call a *planar embedding*. The edges and vertices of an embedded planar graph divide the plane into connected regions; such regions are called *faces*. The face with infinite area is called the *exterior* face (also the *infinite* face or *outer* face).

An *outerplanar* graph is a planar graph that can be drawn such that each vertex lies on the boundary of the exterior face. For outerplanar graphs see Chartrand and Harary [22].

The *dual* graph G^* of an embedded planar graph G is the graph whose vertex set is the set of faces of G and whose edge set consists of all unordered pairs of faces of G that share an edge. For more on dual graphs see Whitney [77].

This notion of a dual graph was studied earlier in a slightly different form by Cayley and Tait in the late 19$^{\text{th}}$ century (see Coxeter [24]). The version of the dual graph used by Cayley and Tait did not consider the vertex corresponding to the exterior face of a planar graph G to be part of its dual. We call this form of the dual graph the *weak dual* graph, and we denote it (following Fleischner, Geller, and Harary [37]) by G^w.

We will need the following characterization of outerplanar graphs due to Fleischner, Geller, and Harary [37].

Theorem 5.1 (Fleischner, Geller, and Harary). *A graph G is outerplanar if and only if it has a weak dual G^w where G^w is a forest.*

A vertex (resp. edge) colouring of a graph G is called *non-repetitive* if for any path P in G, the sequence of vertex (resp. edge) colours along P is squarefree.

If G is an embedded planar graph, then a colouring of the faces of G is called *non-repetitive* if for any sequence of distinct faces such that each consecutive pair of faces shares an edge, the sequence of corresponding colours is squarefree.

5.2 Previous work and open questions

Alon *et al.* [7] proved the following result regarding non-repetitive colourings of graphs.

Theorem 5.2 (Alon *et al.*). *There exists an absolute constant c such that the edges of any graph G with maximum degree at most Δ can be non-repetitively coloured using at most $c\Delta^2$ colours.*

The proof of this theorem given by Alon *et al.* is probabilistic and uses the Lovász Local Lemma. (The Lovász Local Lemma was given by Erdős and Lovász [35] and generalized by Spencer [72]; for more details regarding the Lovász Local Lemma and the probabilistic method in general, see Alon and Spencer [8].) For some interesting applications of the Lovász Local Lemma to problems of combinatorics on words see Alon and Spencer [8, Section 5.8, Exercise 2], Beck [11], Grytczuk [40], and Currie [26].

Grytczuk [40] asked the following question.

Problem 5.3. *Is there a natural number k such that the faces of any planar graph can be non-repetitively coloured using at most k colours?*

We answer this question in the affirmative for all outerplanar graphs. Alon *et al.* [7] asked a similar question.

Problem 5.4. *Is there a natural number k such that the vertices of any planar graph can be non-repetitively coloured using at most k colours?*

Here we give lower bounds for k for both outerplanar and planar graphs.

For similar problems regarding colourings of integer lattices see Carpi [19] and Currie and Simpson [27].

Alon *et al.* [7] mention that the vertices of any tree can be non-repetitively coloured using at most four colours. We will make use of this observation below; we therefore state it as the following theorem.

Theorem 5.5 (Alon *et al.*). *For any tree T, the vertices of T can be non-repetitively coloured using at most 4 colours.*

Proof. We first show the existence of arbitrarily large squarefree words over Σ_4 that are also palindrome-free. We say a word w is *palindrome-free* if no subword w' of w satisfies $w' = (w')^R$ (see Chapter 4 for palindromes and the notation w^R). Let $\tilde{w}$ be any squarefree word over Σ_3. We form the word $w = w_1w_2\dots$ by inserting the symbol 4 between consecutive blocks of length two; *e.g.*, if $\tilde{w} = 10212012$, then $w = 104214204124$. Such a word w is both squarefree and palindrome-free.

We now use w to give a non-repetitive colouring of T. Choose an arbitrary vertex v to be the root of T. Colour each vertex at distance i from v with colour w_{i+1}. Then since w is squarefree and palindrome-free, all paths in T are coloured non-repetitively. ☐

5.3 Main results

Theorem 5.6. *If G is an outerplanar graph, then the faces of G can be non-repetitively coloured using at most five colours.*

Proof. We assume that we have an outerplanar embedding of G. It suffices to show that the vertices of the dual graph G^* can be coloured non-repetitively using at most five colours. Consider the weak dual graph G^w formed by deleting the vertex of G^* corresponding to the exterior face of G. By Theorem 5.1 we have that G^w is a forest. Applying Theorem 5.5, we have that the vertices of the weak dual graph G^w can be non-repetitively coloured with at most four colours. Finally, add back the vertex initially deleted from G^* and colour it using a fifth colour. The resulting 5-colouring of the dual graph G^* induces a non-repetitive colouring of the faces of the outerplanar map G. □

Theorem 5.7. *There exists an outerplanar graph G such that the vertices of G cannot be non-repetitively coloured using fewer than five colours.*

Proof. We will construct such a graph G. We begin with the graph P_4, *i.e.*, the graph consisting of a simple path on four vertices. Since there are no non-repetitive binary sequences of length four, we require at least three colours to non-repetitively colour P_4. Next we add a vertex v, connecting it with an edge to each of the vertices of P_4, thus forming the so-called *fan graph* F_4. Let us call the vertex v the *rivet* of the fan. Since v is connected to vertices of three different colours, it is evident that v must be coloured a fourth colour. The graph G then consists of *five* disjoint copies of F_4, with an additional vertex r connected to the rivet of each fan (see Figure 5.1). Clearly G is an outerplanar graph. If we assume that we only have four colours with which to work, then by the pigeonhole principle, two rivets, say v and v', must be coloured the same colour, say x. The vertex r cannot be coloured x, so r must be coloured with one of the three remaining colours, say y. However, the subgraph P_4 connected to the rivet v contains vertices coloured with three distinct colours different from x. Hence, we can always find a vertex w such that the path $wvrv'$ has colouring $yxyx$. This is clearly a repetition, and so we see that we need at least five colours to non-repetitively colour G. □

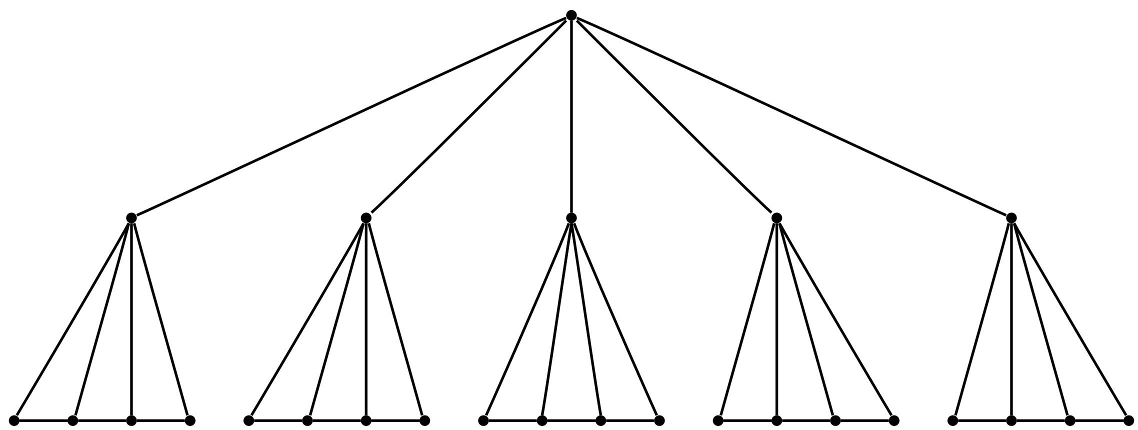

Figure 5.1: Graph G from Theorem 5.7

Theorem 5.8. *There exists a planar graph G such that the vertices of G cannot be non-repetitively coloured using fewer than seven colours.*

Proof. The construction of G is readily apparent from Figure 5.2. Let us label the two vertices of G with degree eight r and s. Let us call each of the connected components of the subgraph formed by deleting r and s from G a *diamond*. By reasoning similar to that used in the proof of Theorem 5.7, we may conclude that each diamond requires at least five colours for a non-repetitive colouring. Assume that we have a non-repetitive 6-colouring of G. Now consider the seven vertices of G with degree seven. By the pigeonhole principle, two of these vertices must have the same colour. Let us call these two vertices v and v', and let us assume that they are each coloured x. Let us call each of the two diamonds containing v and v' D and D' respectively. Suppose that D and D' are each coloured using exactly five colours, but the five colours used are not the same for each diamond. In this case, between D and D' all six colours are used, and so for all choices y, $y \neq x$, for the colour of r we can always find a vertex w in one of D or D' such that the path $wvrv'$ has colouring $yxyx$. Hence, it must be the case that D and D' are each coloured using exactly the same colours. If D and D' are each coloured using all six colours, then again we can always find a vertex w in one of D or D' such that the path $wvrv'$ has colouring $yxyx$. It is therefore the case that D and D' are each coloured using exactly the same five colours. Thus we may colour r using the colour that does not appear in D or D'; any other choice y, $y \neq x$, will allow us to find a vertex w in one of D or D' such that the path $wvrv'$ has colouring $yxyx$. However, by this same argument we find that s must be coloured the same colour as r. Since r and s are adjacent, this is a contradiction, and so we have that G cannot be non-repetitively coloured using fewer than seven colours. □

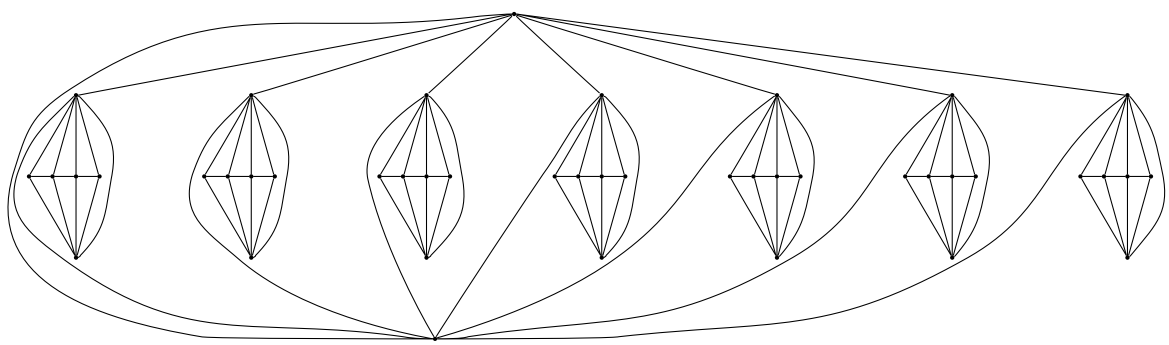

Figure 5.2: Graph G from Theorem 5.8

A very recent result of Albertson *et al.* [2] implies that the lower bound of Theorem 5.7 can be improved from five colours to six colours and that the lower bound of Theorem 5.8 can be improved from seven colours to ten colours. Furthermore, Kündgen and Pelsmajer [51] recently gave an upper bound for the number of colours needed to non-repetitively colour the vertices of an outerplanar graph by showing that the vertices of every outerplanar graph can be non-repetitively coloured using at most 12 colours.

Chapter 6

Self-referential Integer Sequences

In previous chapters we have been entirely concerned with words (*i.e.*, sequences) over a finite alphabet. In this chapter[1] we consider sequences over an infinite alphabet, namely the non-negative integers. Specifically, we study strictly increasing sequences $(a(n))_{n\geq 0}$ of non-negative integers satisfying the equation $a(a(n)) = dn$.

6.1 Preliminaries

Consider the functional equation

$$a(a(n)) = dn \ , \tag{6.1}$$

where $(a(n))_{n\geq 0}$ is a strictly increasing sequence of non-negative integers and d is a positive constant. We are interested in possible solutions to (6.1). For what values of d does (6.1) have a solution? For $d = 2$ there is no solution to (6.1) for all $n \geq 0$ (to see this, consider possible values for $a(1)$). However, C. L. Mallows gave a sequence (A007378 in Sloane's *On-line Encyclopedia of Integer Sequences* [71]) that does satisfy (6.1) for $n \geq 2$. Moreover, such a solution is unique. For $d = 3$ there is a unique solution to (6.1) for all $n \geq 0$ (see Propp [64], also A003605 in Sloane [71]). On the other hand, for $d \geq 4$ we show in Section 6.2 that there are uncountably many solutions to (6.1) for all $n \geq 0$.

[1]The contents of this chapter are taken largely verbatim from Allouche, Rampersad, and Shallit [3].

6.2 Number of sequences satisfying $a(a(n)) = 4n$

Theorem 6.1. *Let $b_1, \ldots, b_n$ be positive integers. For all choices*

$$\begin{aligned} b_1 &\in \{5, 6\} \\ b_2 &\in \{29, 30\} \\ &\vdots \\ b_i &\in \{2 \cdot 4^i - 3, 2 \cdot 4^i - 2\} \ , \end{aligned}$$

where $i = 1, 2, \ldots$, there exists a strictly increasing sequence $\mathbf{a} = (a(n))_{n \geq 0}$ of non-negative integers such that $a(a(n)) = 4n$ and $a(4^i - 1) = b_i$.

Proof. We first define a *partial sequence.*

Definition 6.2. *Let k be a positive integer and let $s : \{0, 1, 2, \ldots, 4^k - 1\} \to \mathbf{N}$. Then s is called a* partial sequence of length 4^k *if*

1. *s is strictly increasing;*
2. *$s(s(n)) = 4n$ for all $n < 2 \cdot 4^{k-1}$;*
3. *for $i = 1, 2, \ldots, k$, we have $s(4^i - 1) \in \{2 \cdot 4^i - 3, 2 \cdot 4^i - 2\}$.*

We will show that for any $k \geq 0$ and any partial sequence a of length 4^k, we can always extend the domain of a to $\{0, 1, 2, \ldots, 4^{k+1} - 1\}$ such that, for both choices of $a(4^{k+1} - 1) \in \{2 \cdot 4^k - 3, 2 \cdot 4^k - 2\}$, a is a partial sequence of length 4^{k+1}.

Certainly this is true for $k = 0$ since the partial sequence of length 1

$$a(0) = 0$$

can be extended to the partial sequence of length 4 given by

$$a(0) = 0, \quad a(1) = 2, \quad a(2) = 4, \quad a(3) = 5$$

or the partial sequence of length 4 given by

$$a(0) = 0, \quad a(1) = 2, \quad a(2) = 4, \quad a(3) = 6 \ .$$

Now assume that for some $k \geq 0$, we have a partial sequence a of length 4^k. We will show that the domain of a can be extended such that a becomes a partial sequence of length 4^{k+1}.

We first note that a must have

$$a(0) = 0, \quad a(1) = 2, \quad a(2) = 4$$

since any other choices would violate either the rule $a(a(n)) = 4n$ or the requirement that the sequence be strictly increasing. By induction then, for $i = 0, 1, \ldots, k$, we have $a(2 \cdot 4^i) = 4^{i+1}$.

Note that for each choice for the value of $a(4^{k-i} - 1)$, where $i = 0, \ldots, k-1$, the requirement that $a(a(n)) = 4n$ fixes exactly two values of the sequence in the range $4^k \leq n < 4^{k+1}$. Specifically, an easy induction shows that these values are:

$$a(2 \cdot 4^k - 4^i \cdot 3) = 4^{k+1} - 4^{i+1} \tag{6.2}$$

$$a(4^{k+1} - 4^{i+1}) = 2 \cdot 4^{k+1} - 4^{i+1} \cdot 3 \tag{6.3}$$

if $a(4^{k-i} - 1) = 2 \cdot 4^{k-i} - 3$, and

$$a(2 \cdot 4^k - 4^i \cdot 2) = 4^{k+1} - 4^{i+1} \tag{6.4}$$

$$a(4^{k+1} - 4^{i+1}) = 2 \cdot 4^{k+1} - 4^{i+1} \cdot 2 \tag{6.5}$$

if $a(4^{k-i} - 1) = 2 \cdot 4^{k-i} - 2$.

Our first task is to show that each of the terms defined by (6.2)–(6.5) is distinct. Specifically, we need to show that for all distinct integers i and j, such that $0 \leq i, j < k$, the following inequalities hold:

$$2 \cdot 4^k - 4^i \cdot 3 \neq 2 \cdot 4^k - 4^j \cdot 2 \tag{6.6}$$

and for all integers i and j such that $0 \leq i, j < k$,

$$2 \cdot 4^k - 4^i \cdot 3 \neq 4^{k+1} - 4^{j+1} \tag{6.7}$$

and

$$2 \cdot 4^k - 4^i \cdot 2 \neq 4^{k+1} - 4^{j+1} \quad . \tag{6.8}$$

Equation (6.6) is clearly true, since $4^i \cdot 3 \neq 4^j \cdot 2$ for all i, j. Equations (6.7) and (6.8) are also true, since $2 \cdot 4^k - 4^i \cdot 3 < 2 \cdot 4^k$, $2 \cdot 4^k - 4^i \cdot 3 < 2 \cdot 4^k$, and $4^{k+1} - 4^{i+1} > 2 \cdot 4^k$ for all integers i such that $0 \leq i < k$.

We have already seen that the rule $a(a(n)) = 4n$ determines the values of the terms in the sequence for certain integers n such that $4^k \leq n < 4^{k+1}$. Our next task is to show that for those terms that have not been fixed by this rule, we are able to choose values that preserve the required properties of the sequence.

We give a procedure to assign values to $a(n)$, with $4^k \leq n < 4^{k+1}$, such that at the end of the procedure a is a partial sequence of length 4^{k+1}.

EXTEND(a)

1. Either define $a(4^{k+1} - 1) = 2 \cdot 4^{k+1} - 3$ or define $a(4^{k+1} - 1) = 2 \cdot 4^{k+1} - 2$.
2. Repeat (3)–(5) until for all $h = 0, 1, \ldots, 4^{k+1} - 1$ the value of $a(h)$ is defined.
3. For $h = 0$ to $2 \cdot 4^{k-1} - 1$ repeat the following:
   If $a(h)$ is defined and $a(a(h))$ is not defined, set $a(a(h)) = 4h$.
4. Let l be the smallest integer such that there exists an integer i, with $2 \leq i \leq 4$, satisfying the following:
   (a) $a(l)$ and $a(l + i)$ are defined, and
   (b) $a(l + 1), a(l + 2), \ldots, a(l + i - 1)$ are not defined.
5. Let j be the integer such that $a(j) = l$ and $a(j + 1) = l + i$. Choose values for $a(l + 1), a(l + 2), \ldots, a(l + i - 1)$ from $\{4j + 1, 4j + 2, 4j + 3\}$ so that $a(l + 1), a(l + 2), \ldots, a(l + i - 1)$ is a strictly increasing sequence.

To see that steps 4 and 5 are correct, note that since for all $n \leq l$ the values of $a(n)$ have already been defined, and since $a(l)$ and $a(l+i)$ have already been determined by the rule $a(a(n)) = 4n$ (steps 3 and 4 guarantee this), it must be the case that there exists a number j such that $a(j) = l$ and $a(j+1) = l+i$. Hence, we see that $a(a(j)) = a(l) = 4j$ and $a(a(j+1)) = a(l+i) = 4j+4$. Additionally, since $a(l+i) - a(l) = 4$, we also see that i is at most 4 (otherwise, a would not be strictly increasing). Therefore, if $i = 4$ we must choose values for $a(l+1), a(l+2), a(l+3)$ in the following way:

$$a(l+1) = 4j+1, \quad a(l+2) = 4j+2, \quad a(l+3) = 4j+3 \ .$$

Similarly, if $i = 3$ we can choose values for $a(l+1)$ and $a(l+2)$ in three different ways among $\{4j+1, 4j+2, 4j+3\}$, and if $i = 2$ we can choose $a(l+1)$ to be one of $\{4j+1, 4j+2, 4j+3\}$. In this way we ensure that a is strictly increasing.

The last thing that remains to be shown is that choosing values in this way does not affect our freedom to choose the value of $a(4^{k+1}-1)$ in step 1. To do this, we first note that the only terms of a that can possibly influence the choice of $a(4^{k+1}-1)$ are $a(2 \cdot 4^k - 3)$, $a(2 \cdot 4^k - 2)$, and $a(2 \cdot 4^k - 1)$, as will be shown below. (Note that we have already shown that $a(2 \cdot 4^k)$ must have the value 4^{k+1}).

Since $a(4^k - 1) \in \{2 \cdot 4^k - 3, 2 \cdot 4^k - 2\}$, we must have either $a(2 \cdot 4^k - 3) = 4^{k+1} - 4$ or $a(2 \cdot 4^k - 2) = 4^{k+1} - 4$. This implies that the value of $a(4^{k+1} - 4)$ is always either $2 \cdot 4^{k+1} - 3 \cdot 4$ or $2 \cdot 4^{k+1} - 2 \cdot 4$. Therefore, the only terms of a that can affect the choice of $a(4^{k+1} - 1)$ are those that affect the values of $a(4^{k+1} - 4)$, $a(4^{k+1} - 3)$, $a(4^{k+1} - 2)$, and $a(4^{k+1} - 1)$. However, since $n = 2 \cdot 4^k - 3$ is the smallest possible value of n such that $a(n) = 4^{k+1} - 4$, and since $a(2 \cdot 4^k) = 4^{k+1}$, we see that the only terms of a that can affect the values of $a(4^{k+1} - 4)$, $a(4^{k+1} - 3)$, $a(4^{k+1} - 2)$, and $a(4^{k+1} - 1)$ are $a(2 \cdot 4^k - 3)$, $a(2 \cdot 4^k - 2)$, and $a(2 \cdot 4^k - 1)$. We will now see how to ensure that suitable values are assigned to $a(2 \cdot 4^k - 3)$, $a(2 \cdot 4^k - 2)$, and $a(2 \cdot 4^k - 1)$.

Suppose that we had chosen $a(4^k - 1) = 2 \cdot 4^k - 3$. Then, since $a(2 \cdot 4^k) = 4^{k+1}$, and since we must be left free to choose the value of $a(4^{k+1} - 1)$, we must have the following:

$$\begin{aligned} a(2 \cdot 4^k - 3) &= 4^{k+1} - 4 \\ a(2 \cdot 4^k - 2) &= 4^{k+1} - 3 \\ a(2 \cdot 4^k - 1) &= 4^{k+1} - 2 \ . \end{aligned}$$

The iterative rule $a(a(n)) = 4n$ yields in turn:

$$\begin{aligned} a(4^{k+1}-4) &= 2\cdot 4^{k+1} - 3\cdot 4 \\ a(4^{k+1}-3) &= 2\cdot 4^{k+1} - 2\cdot 4 \\ a(4^{k+1}-2) &= 2\cdot 4^{k+1} - 1\cdot 4 \ . \end{aligned}$$

leaving us free to choose the value of $a(4^{k+1}-1)$ as we like.

On the other hand, suppose that we had chosen $a(4^k - 1) = 2\cdot 4^k - 2$. Then, we must choose the value of $a(2\cdot 4^k - 1)$ to be either $4^{k+1}-3$ or $4^{k+1}-2$. However, if we choose $a(2\cdot 4^k - 1) = 4^{k+1}-3$, we would have the following:

$$\begin{aligned} a(2\cdot 4^k - 2) &= 4^{k+1} - 4 \\ a(2\cdot 4^k - 1) &= 4^{k+1} - 3 \ , \end{aligned}$$

and the iterative rule $a(a(n)) = 4n$ would yield

$$\begin{aligned} a(4^{k+1}-4) &= 2\cdot 4^{k+1} - 2\cdot 4 \\ a(4^{k+1}-3) &= 2\cdot 4^{k+1} - 1\cdot 4 \ , \end{aligned}$$

thus forcing

$$\begin{aligned} a(4^{k+1}-2) &= 2\cdot 4^{k+1} - 3 \\ a(4^{k+1}-1) &= 2\cdot 4^{k+1} - 2 \ , \end{aligned}$$

and we have lost our freedom to choose the value of $a(4^{k+1}-1)$. Therefore, we must choose $a(2\cdot 4^k - 1) = 4^{k+1}-2$, which gives the following:

$$\begin{aligned} a(2\cdot 4^k - 2) &= 4^{k+1} - 4 \\ a(2\cdot 4^k - 1) &= 4^{k+1} - 2 \ , \end{aligned}$$

and the iterative rule $a(a(n)) = 4n$ yields

$$\begin{aligned} a(4^{k+1}-4) &= 2\cdot 4^{k+1} - 2\cdot 4 \\ a(4^{k+1}-2) &= 2\cdot 4^{k+1} - 1\cdot 4 \ . \end{aligned}$$

Here we can choose the value of $a(4^{k+1}-3)$ to be one of $\{2\cdot 4^{k+1}-7, 2\cdot 4^{k+1}-6, 2\cdot 4^{k+1}-5\}$, and our freedom to choose $a(4^{k+1}-1)$ is preserved.

Hence, we have shown that given a partial sequence a of length 4^k, we can extend the domain of a such that, for both choices of $a(4^{k+1}-1) \in \{2\cdot 4^k - 3, 2\cdot 4^k - 2\}$, a is a partial sequence of length 4^{k+1}. The result now follows by induction. □

For the general case, where $a(a(n)) = dn$ for some $d \geq 4$, if we set

$$a(0) = 0, \quad a(1) = 2, \quad a(2) = d$$

then we have $a(2\cdot d^i) = d^{i+1}$. Moreover, if we choose the value of each term $a(d^i - 1)$ from $\{2\cdot d^i - 3, 2\cdot d^i - 2\}$, then (6.2)–(6.8) also hold with the constant 4 replaced by d. The proof then follows essentially as above, though with d instead of 4 throughout. We thus have the following corollary.

Corollary 6.3. *Let d be an integer such that $d \geq 4$. Then there are uncountably many strictly increasing sequences $\mathbf{a} = (a(n))_{n\geq 0}$ satisfying $a(a(n)) = dn$ for all $n \geq 0$.*

In addition to this result, Allouche, Rampersad, and Shallit [3] also give the lexicographically least monotone sequence satisfying $a(a(n)) = dn$ for all $n \geq 0$ and prove that it is d-regular (for the notion of k-regularity see Allouche and Shallit [4, 6]).

Chapter 7

Conclusion and Open Problems

In this chapter we review the work presented in this thesis and state some open problems.

In Chapter 2 we used uniform morphisms to construct an infinite binary word that contains no cubes xxx and no squares yy with $|y| \geq 4$, thus giving a simpler proof than that given by Dekking [28]. We also used uniform morphisms to construct an infinite binary word avoiding all squares except 0^2, 1^2, and $(01)^2$, thus giving a simpler proof than that given by Fraenkel and Simpson [38]. We gave an enumeration of the words satisfying each of these properties; however, the bounds given were not tight, and so we have the following problem.

Problem 7.1. *Can the bounds (the lower bounds in particular) given in Theorems 2.8 and 2.12 be improved?*

In Section 2.5 we stated two open problems (Problems 2.20 and 2.21) regarding abelian squares and abelian cubes. We give another related problem below.

Problem 7.2. *Can the upper bound of $k^2 + 6k$ given in Theorem 2.19 be improved?*

In Chapter 2 we also solved an open problem due to Prodinger and Urbanek [63] by constructing two infinite binary words $\mathbf{x}$ and $\mathbf{y}$ such that neither $\mathbf{x}$ nor $\mathbf{y}$ contain arbitrarily large squares, but their perfect shuffle, $\mathbf{x} \amalg \mathbf{y}$, does contain arbitrarily large squares.

In Chapter 3 we characterized all words that can be avoided in infinite squarefree ternary words, and in Chapter 4 we explored words w satisfying the property that for any subword w' of w, w does not contain the reversal of w' as a subword.

In Chapter 5 we studied non-repetitive colourings of graphs. We examined some open problems posed by Alon *et al.* [7] and Grytczuk [40], and we gave partial results for some of these problems. The main motivating problems of Chapter 5 were given as Problems 5.3 and 5.4. We give an additional problem below.

Problem 7.3. *Can a constructive proof of Theorem 5.2 be found?*

In Chapter 6 we studied non-negative integer solutions to the functional equation $a(a(n)) = dn$ (Equation 6.1). In particular, we showed that for $d \geq 4$ there are uncountably many strictly increasing sequences $(a(n))_{n\geq 0}$ that satisfy (6.1). Some open problems regarding such functional equations are given in Allouche, Rampersad, and Shallit [3]. We give some additional problems below.

Problem 7.4. *What is the lexicographically greatest strictly increasing sequence* $(a(n))_{n\geq 0}$ *that satisfies (6.1)?*

Problem 7.5. *Consider generalizations of (6.1). For example, what can be said about solutions to the equation* $a(a(a(n))) = dn$*?*

Bibliography

[1] S. I. Adjan. *The Burnside Problem and Identities in Groups.* Springer-Verlag, 1979.

[2] M. O. Albertson, G. G. Chappell, H. A. Kierstead, A. Kündgen, and R. Ramamurthi. Coloring with no 2-colored P_4's. *Electronic J. Combinatorics* **11** (2004), #R26. Available electronically at http://www.combinatorics.org/Volume_11/Abstracts/v11i1r26.html.

[3] J.-P. Allouche, N. Rampersad, and J. Shallit. On integer sequences whose first iterates are linear. To appear in *Aequationes Math.*

[4] J.-P. Allouche and J. Shallit. The ring of k-regular sequences. *Theoret. Comput. Sci.* **98** (1992), 163–197.

[5] J.-P. Allouche and J. Shallit. The ring of k-regular sequences, II. *Theoret. Comput. Sci.* **307** (2003), 3–29.

[6] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations.* Cambridge University Press, 2003.

[7] N. Alon, J. Grytczuk, M. Haluszczak, and O. Riordan. Nonrepetitive colorings of graphs. *Random Structures and Algorithms* **21** (2002), 336–346.

[8] N. Alon and J. Spencer. *The Probabilistic Method.* Wiley–Interscience series in discrete mathematics and optimization. John Wiley & Sons, second edition, 2000.

[9] S. E. Aršon. Demonstration de l'existence des suites asymétriques infinies. *Mat. Sbornik* **44** (1937), 769–779.

[10] D. Bean, A. Ehrenfeucht, and G. McNulty. Avoidable patterns in strings of symbols. *Pacific J. Math.* **85** (1979), 261–294.

[11] J. Beck. An application of Lovász local lemma: there exists an infinite 01-sequence containing no near identical intervals. In A. Hajnal et al., editors, *Finite and Infinite Sets, Vol. I, II (Eger, 1981)*, Vol. 37 of *Colloq. Math. Soc. János Bolyai*, pp. 103–107. North-Holland, 1984.

[12] J. Berstel. Sur les mots sans carré définis par un morphisme. In *Proc. 6th Int'l Conf. on Automata, Languages, and Programming (ICALP)*, Vol. 71 of *Lecture Notes in Computer Science*, pp. 16–25. Springer-Verlag, 1979.

[13] J. Berstel and J. Karhumäki. Combinatorics on words. TUCS Technical Report 530, Turku Centre for Computer Science, June 2003.

[14] B. Bollobás. *Modern Graph Theory*, Vol. 184 of *Graduate Texts in Mathematics.* Springer-Verlag, 1998.

[15] F.-J. Brandenburg. Uniformly growing kth power-free homomorphisms. *Theoret. Comput. Sci.* **23** (1983), 69–82.

[16] J. Brinkhuis. Nonrepetitive sequences on three symbols. *Quart. J. Math. Oxford Ser. (2)* **34** (1983), 145–149.

[17] T. C. Brown. Is there a sequence on four symbols in which no two adjacent segments are permutations of one another? *Amer. Math. Monthly* **78** (1971), 886–888.

[18] W. Burnside. On an unsettled question in the theory of discontinuous groups. *Quart. J. Pure Appl. Math.* **33** (1902), 230–238.

[19] A. Carpi. Multidimensional unrepetitive configurations. *Theoret. Comput. Sci.* **56** (1988), 233–241.

[20] A. Carpi. On abelian power-free morphisms. *Internat. J. Algebra Comput.* **3** (1993), 151–167.

[21] J. Cassaigne. Unavoidable binary patterns. *Acta Informatica* **30** (1993), 385–395.

[22] G. Chartrand and F. Harary. Planar permutation graphs. *Ann. Inst. H. Poincar Sect. B (N.S.)* **3** (1967), 433–438.

[23] C. Choffrut and J. Karhumäki. *Combinatorics of words*, Vol. 1, chapter 6, pp. 329–438. Springer-Verlag, 1997.

[24] H. S. M. Coxeter. The four-color map problem, 1840–1890. *Math. Teacher* **52** (1959), 283–289.

[25] M. Crochemore. Sharp characterizations of squarefree morphisms. *Theoret. Comput. Sci.* **18** (1982), 221–226.

[26] J. Currie. Pattern avoidance: themes and variations. In *Proceedings of WORDS'03 (4th International Conference on Combinatorics on Words)*, Turku Center for Computer Science General Publication Number 27, pp. 14–26, 2003.

[27] J. Currie and J. Simpson. Non-repetitive tilings. *Electronic J. Combinatorics* **9** (2002), #R28. Available electronically at `http://www.combinatorics.org/Volume_9/Abstracts/v9i1r28.html`.

[28] F. M. Dekking. On repetitions of blocks in binary sequences. *J. Combin. Theory. Ser. A* **20** (1976), 292–299.

[29] F. M. Dekking. The spectrum of dynamical systems arising from substitutions of constant length. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **41** (1978), 221–239.

[30] F. M. Dekking. Strongly non-repetitive sequences and progression-free sets. *J. Combin. Theory. Ser. A* **27** (1979), 181–185.

[31] R. Diestel. *Graph Theory*, Vol. 173 of *Graduate Texts in Mathematics*. Springer-Verlag, second edition, 2000.

[32] S. Ekhad and D. Zeilberger. There are more than $2^{n/17}$ n-letter ternary square-free words. *J. Integer Sequences* **1** (1998), Article 98.1.9. Available electronically at `http://www.math.uwaterloo.ca/JIS/zeil.html`.

[33] R. C. Entringer, D. E. Jackson, and J. A. Schatz. On nonrepetitive sequences. *J. Combin. Theory. Ser. A* **16** (1974), 159–164.

[34] P. Erdős. Some unsolved problems. *Magyar Tud. Akad. Kutató Int. Közl.* **6** (1961), 221–254.

[35] P. Erdős and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In A. Hajnal et al., editors, *Infinite and Finite Sets, Vol. II*, Vol. 10 of *Colloq. Math. Soc. János Bolyai*, pp. 609–628. North-Holland, 1975.

[36] A. A. Evdokimov. Strongly asymmetric sequences generated by a finite number of symbols. *Soviet Math. Dokl.* **9** (1968), 536–539.

[37] H. J. Fleischner, D. P. Geller, and F. Harary. Outerplanar graphs and weak duals. *J. Indian Math. Soc.* **38** (1974), 215–219.

[38] A. S. Fraenkel and J. Simpson. How many squares must a binary sequence contain? *Electronic J. Combinatorics* **2** (1995), #R2. Available electronically at `http://www.combinatorics.org/Volume_2/Abstracts/v2i1r2.html`.

[39] U. Grimm. Improved bounds on the number of ternary square-free words. *J. Integer Sequences* **4** (2001), Article 01.2.7. Available electronically at `http://www.math.uwaterloo.ca/JIS/VOL4/GRIMM/words.html`.

[40] J. Grytczuk. Thue-like sequences and rainbow arithmetic progressions. *Electronic J. Combinatorics* **9** (2002), #R44. Available electronically at `http://www.combinatorics.org/Volume_9/Abstracts/v9i1r44.html`.

[41] M. Hall. Generators and relations in groups—the Burnside problem. In T. L. Saaty, editor, *Lectures on Modern Mathematics*, Vol. 2, chapter 2, pp. 42–92. Wiley, 1964.

[42] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979.

[43] S. Istrail. On irreducible languages and nonrational numbers. *Bull. Math. Soc. Sci. Math. R. S. Roumanie (N.S.)* **21** (1977), 301–308.

[44] J. Justin. Characterization of the repetitive semigroups. *J. Algebra* **21** (1972), 87–90.

[45] J. Justin. Généralization du théorème de van der Waerden sur les semigroupes répétitifs. *J. Combin. Theory. Ser. A* **12** (1972), 357–367.

[46] J. Karhumäki and J. Shallit. Polynomial versus exponential growth in repetition-free binary words. *J. Combin. Theory. Ser. A* **105** (2004), 335–347.

[47] V. Keränen. Abelian squares are avoidable on 4 letters. In *Proc. 19th Int'l Conf. on Automata, Languages, and Programming (ICALP)*, Vol. 623 of *Lecture Notes in Computer Science*, pp. 41–52. Springer-Verlag, 1992.

[48] V. Keränen. On abelian square-free DT0L-languages over 4 letters. In *Proceedings of WORDS'03 (4th International Conference on Combinatorics on Words)*, Turku Center for Computer Science General Publication Number 27, pp. 95–109, 2003.

[49] D. König. Sur les correspondances multivoques des ensembles. *Fundamenta Math.* **8** (1926), 114–134.

[50] D. König. Über eine Schlußweise aus dem Endlichen ins Unendliche. *Acta Lit. Sci. (Szeged)* **3** (1927), 121–130.

[51] A. Kündgen and M. Pelsmajer. Nonrepetitive colorings of graphs of bounded treewidth. Manuscript.

[52] J. Leech. A problem on strings of beads. *Math. Gazette* **41** (1957), 277–278.

[53] M. Lothaire. *Combinatorics on Words*, Vol. 17 of *Encyclopedia of Mathematics*. Addison-Wesley, 1983. Reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, 1997.

[54] S. Mäkelä. Patterns in words. Master's thesis, University of Turku, 2002.

[55] M. Morse and G. A. Hedlund. Unending chess, symbolic dynamics and a problem in semigroups. *Duke Math. J.* **11** (1944), 1–7.

[56] J. Noonan and D. Zeilberger. The Goulden-Jackson cluster method: extensions, applications and implementations. *J. Differ. Equations Appl.* **5** (1999), 355–377.

[57] P. S. Novikov. On periodic groups. *Dokl. Akad. Nauk SSSR* **127** (1959), 749–752.

[58] P. S. Novikov and S. I. Adjan. Infinite periodic groups, I. *Izv. Akad. Nauk SSSR Ser. Mat.* **32** (1968), 212–244.

[59] P. S. Novikov and S. I. Adjan. Infinite periodic groups, II. *Izv. Akad. Nauk SSSR Ser. Mat.* **32** (1968), 251–524.

[60] P. S. Novikov and S. I. Adjan. Infinite periodic groups, III. *Izv. Akad. Nauk SSSR Ser. Mat.* **32** (1968), 709–731.

[61] P. Ochem, 2003. Personal communication with Jeffrey Shallit.

[62] P. A. B. Pleasants. Non-repetitive sequences. *Proc. Cambridge Phil. Soc.* **68** (1970), 267–274.

[63] H. Prodinger and F. J. Urbanek. Infinite 0–1-sequences without long adjacent identical blocks. *Discrete Math.* **28** (1979), 277–289.

[64] J. Propp. Problem 474. *Crux Math.* **5** (1979), 229. Solution by G. Patruno, **6** (1980), 198.

[65] N. Rampersad and J. Shallit. Words avoiding reversed subwords. Preprint available at `http://www.arxiv.org/abs/math.CO/0311121`.

[66] N. Rampersad, J. Shallit, and M.-w. Wang. Avoiding large squares in infinite binary words. In *Proceedings of WORDS'03 (4th International Conference on Combinatorics on Words)*, Turku Center for Computer Science General Publication Number 27, pp. 185–197, 2003.

[67] C. Richard and U. Grimm. On the entropy and letter frequencies of ternary square-free words. *Electronic J. Combinatorics* **11** (2004), #R14. Available electronically at `http://www.combinatorics.org/Volume_11/Abstracts/v11i1r14.html`.

[68] P. Roth. l-occurrences of avoidable patterns. In *STACS 91, Proc. 8th Symp. Theoretical Aspects of Comp. Sci.*, Vol. 480 of *Lecture Notes in Computer Science*, pp. 42–29. Springer-Verlag, 1991.

[69] P. Roth. Every binary pattern of length six is avoidable on the two-letter alphabet. *Acta Informatica* **29** (1992), 95–107.

[70] G. Rozenberg and A. Salomaa. *The Mathematical Theory of L Systems*, Vol. 90 of *Pure and Applied Mathematics*. Academic Press, 1980.

[71] N. J. A. Sloane. On-line encyclopedia of integer sequences. Published electronically at `http://www.research.att.com/~njas/sequences/`.

[72] J. Spencer. Asymptotic lower bounds for Ramsey functions. *Discrete Math.* **20** (1976), 69–76.

[73] X. Sun. New lower bound on the number of ternary square-free words. *J. Integer Sequences* **6** (2003), Article 03.3.2. Available electronically at `http://www.math.uwaterloo.ca/JIS/VOL6/Sun/sun.html`.

[74] A. Szilard, 2003. Personal communication with Jeffrey Shallit.

[75] A. Thue. Über unendliche Zeichenreihen. *Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl.* **7** (1906), 1–22. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell *et al.*, editors, Universitetsforlaget, Oslo, 1977, pp. 139–158.

[76] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl.* **1** (1912), 1–67. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell *et al.*, editors, Universitetsforlaget, Oslo, 1977, pp. 413–477.

[77] H. Whitney. Non-separable and planar graphs. *Trans. Amer. Math. Soc.* **34** (1932), 339–362.

[78] A. I. Zimin. Blocking sets of terms. *Mat. Sbornik* **119** (1982), 363–375.