Exponential lower bounds for the number of words of uniform length avoiding a pattern

Jason P. Bell

Department of Mathematics
Simon Fraser University
8888 University Dr.
Burnaby, BC V5A 1S6. CANADA

Teow Lim Goh

Department of Mathematics
University of Michigan
East Hall, 525 East University Ave.
Ann Arbor, MI 48109-1109. USA

Abstract

We study words on a finite alphabet avoiding a finite collection of patterns. Given a pattern p in which every letter that occurs in p occurs at least twice, we show that the number of words of length n on a finite alphabet that avoid p grows exponentially with n as long as the alphabet has at least 4 letters. Moreover, we give lower bounds describing this exponential growth in terms of the size of the alphabet and the number of letters occurring in p. We also obtain analogous results for the number of words avoiding a finite collection of patterns. We conclude by giving some questions.

Key words: pattern avoidance, combinatorics on words, avoidable patterns.

---

1 Introduction

Let $\mathcal{X}$ be a finite alphabet and let p be a word on some other alphabet $\mathcal{Y}$. We say that a word in $\mathcal{X}$ *avoids the pattern* p if it contains no subword of the form $h(\mathsf{p})$, where h is a nonerasing homomorphism from the free monoid

Email addresses: belljp@umich.edu (Jason P. Bell), gteowlim@umich.edu (Teow Lim Goh).

$\mathcal{Y}^*$ generated by $\mathcal{Y}$ to the free monoid $\mathcal{X}^*$ generated by $\mathcal{X}$. We say that p is *avoidable on* $\mathcal{X}$ if there are infinitely many words that avoid the pattern p. We say that a pattern p is *avoidable* if it is avoidable on some finite alphabet. The Zimin algorithm [8, §3.2] is a recursive algorithm that determines if a given pattern is avoidable or unavoidable. Given a pattern p on an alphabet $\mathcal{Y}$, we define

$$\mathcal{S}(\mathsf{p}, \mathcal{X}) \; = \; \{h(\mathsf{p}) \mid h : \mathcal{Y}^* \to \mathcal{X}^* \text{ is a nonerasing homomorphism}\}. \quad (1.1)$$

We say that a word W on the alphabet $\mathcal{X}$ is *of the form* p if $W \in \mathcal{S}(\mathsf{p}, \mathcal{X})$.

The study of pattern avoidance began in the early 1900s with Thue's work on squarefree words [14], [15] (see also Nagell et al. [11]). A word on a finite alphabet $\mathcal{X}$ is *squarefree* if it contains no subword of the form ww, where w is a nonempty word on $\mathcal{X}$. Equivalently, a word is squarefree if it avoids the pattern t^2. Squarefree words have seen numerous applications over the years. In group theory, they have been used in giving a counter-example to the unrestricted Burnside problem [1]. An interesting application to unending chess appears in Morse and Hedlund [10]. Much work has been done on counting squarefree words over various alphabets [3], [12]. More generally, one can look at words that avoid the pattern t^j for some $j \geq 2$. Work on this problem has been done by Brandenburg [4] and Bean et al. [2]. An excellent survey of pattern avoidance is found in Chapter 3 of Lothaire [8]. In addition to this, Currie [5], [6] has given many interesting open problems on the topic of pattern avoidance.

We note that if $a(n)$ denotes the number of words of length n on an alphabet $\mathcal{X}$ which avoid some pattern or collection of patterns, then $a(n)$ is *sub-multiplicative*; that is, $a(n+m) \leq a(n)a(m)$. Since $a(n)$ is a submultiplicative sequence of natural numbers, Fekete's lemma (cf. Madras and Slade [9, Lemma 1.2.2]) shows that

$$\lim_{n\to\infty} a(n)^{1/n}$$

exists and is equal to some nonnegative real number. When this limit is greater than 1, the number of words $a(n)$ of length n that avoid our given collection of patterns grows exponentially. In this paper, we show that for a large class of avoidable patterns, this exponential growth phenomenon occurs. Specifically, we look at patterns p on some finite alphabet with the property that every letter in the alphabet occurs at least twice in p. Such patterns are known to be avoidable [8, Cor. 3.2.10]. We are able to give exponential lower bounds and thus obtain a stronger result.

Theorem 1 *Let* p *be a pattern on* $k \geq 2$ *letters in which every symbol occurs at least twice. Then for* $m \geq 4$ *and* $(k, m) \neq (2, 4)$ *there are at least* $\lambda(k, m)^n$

words of length n on an m letter alphabet that avoid p*, where*

$$\lambda(k,m) \ := m\Big(1+\frac{1}{(m-2)^k}\Big)^{-1} \tag{1.2}$$

We note that this theorem does not consider the case that the pattern $\mathsf{p} = t^i$ for some $i \geq 2$. Exponential lower bounds for the number of words on a ternary alphabet that avoid such a pattern have been given [12]. Thus in the case when $\mathsf{p} = t^i$ with $i \geq 2$ we have an exponential lower bound whenever $m \geq 3$.

For the case $(k,m) = (2,4)$ in Theorem 1, it can be shown that an exponential lower bound exists, since if p is a pattern on the letters t_1 and t_2, then p must contain either t_1^2, t_2^2, $(t_1t_2)^2$ or $(t_2t_1)^2$ as a subword and hence any squarefree word on an m-letter alphabet will avoid p. Since the number of squarefree words on a m-letter alphabet has an exponential lower bound for $m \geq 3$, this same bound must apply to the pattern p.

We note that for binary alphabets, we cannot possibly avoid any pattern on a two letter alphabet in which each letter occurs at least twice, since any word of length at least 4 contains a square. Karhumäki and Shallit [7] have looked at binary words that avoid the pattern t^α for various values of $\alpha \geq 2$ and have completely determined when exponential and polynomial growth occurs.

We give another theorem which applies to sets of patterns in which each letter that occurs, occurs at least twice. Furthermore, this result makes no restrictions concerning patterns on one letter.

Theorem 2 *Let* $\mathcal{S}$ *be a finite set of patterns* $\{\mathsf{p}_1,\ldots,\mathsf{p}_d\}$ *such that for each* i*, each letter that occurs in* p_i *occurs at least twice. Let* $\varepsilon = 1$ *if* $\mathsf{p}_i = t^2$ *for some* i *and take* $\varepsilon = 0$ *otherwise. Then:*

- *if* $m \geq 36d^{2/3}$*, there are at least* $(m/2)^n$ *words of length* n *on an* m *letter alphabet that avoid* $\mathcal{S}$*;*
- *there is a constant* $C = C(d)$*, depending on* d*, such that for* $m > C$ *there are at least* $(m-\varepsilon-\frac{5d}{\sqrt{m}})^n$ *words of length* n *on an* m *letter alphabet that avoid* $\mathcal{S}$*.*

From this we obtain the following corollary.

Corollary 3 *Let* $\mathcal{S}$ *be a finite set of patterns* $\{\mathsf{p}_1,\ldots,\mathsf{p}_d\}$ *such that for each* i*, each letter that occurs in* p_i *occurs at least twice. Let* $a(n)$ *denote the number of words of length* n *on an* m *letter alphabet that avoid* $\mathcal{S}$ *and define*

$$\Gamma(m,\mathcal{S}) \ := \ m - \lim_{n\to\infty} a(n)^{1/n}.$$

Then $\Gamma(m,\mathcal{S}) \to \varepsilon$ *as* $m \to \infty$*, where* $\varepsilon = 1$ *if the pattern* $t^2 \in \mathcal{S}$ *and* $\varepsilon = 0$

otherwise.

Our main tool in obtaining these results is a theorem due to Golod. Before stating this theorem we recall that if R is a (not necessarily commutative) ring, then an ideal I of R is a non-empty subset that is closed under addition and under left and right multiplication by elements of R. We let $\mathbb{C}\{X_1,\ldots,X_m\}$ denote the free algebra over $\mathbb{C}$ on m variables; that is, $\mathbb{C}\{X_1,\ldots,X_m\}$ consists of the noncommutative polynomials over $\mathbb{C}$ in the variables $X_1,\ldots,X_m$. We note that a word in $X_1,\ldots,X_m$ has a degree, which is just the length of the word. We say that an element of $\mathbb{C}\{X_1,\ldots,X_m\}$ is *homogeneous* if it a noncommutative polynomial in which every term that occurs with nonzero coefficient has the same degree.

We are now ready to state the theorem. We remark that this theorem holds over any field, but we only need it for the field of complex numbers.

Theorem 4 *(Golod [13, Lemma 6.2.7]) Let I be a homogeneous ideal in $\mathbb{C}\{X_1,\ldots,X_m\}$ generated by a set $\mathcal{S}$ of homogeneous elements, each of degree at least 2. Suppose that $\mathcal{S}$ has at most c_i elements of degree i for each $i \geq 2$. Suppose, further, that the power series expansion of*

$$G(x) \ := \ \left(1 - mx + \sum_{j\geq 2} c_j x^j\right)^{-1}$$

has nonnegative coefficients. Then the dimension of the vector space spanned by the images of words of length n in $\mathbb{C}\{X_1,\ldots,X_m\}/I$ is greater than or equal to the coefficient of x^n in the power series expansion of $G(x)$.

This theorem is proved using simple counting arguments which come from looking at resolutions of modules. It is perhaps strange that a theorem from algebra that looks at long exact sequences can give good asymptotic information; and yet we shall see that the lower bounds we obtain are in some sense very close to the optimal values.

In §2, we consider words avoiding a single pattern. In §3, we look at words that simultaneously avoid multiple patterns. In the case of multiple patterns, to keep the presentation simple we do not use the best estimates possible. Nevertheless, we still obtain strong estimates in this case. In §4, we give tables which give the bounds we have obtained for various patterns over different sized alphabets. In §5, we present some open questions along with some concluding remarks.

2 Proofs for avoidance of a single pattern

In this section, we prove Theorem 1. Throughout this section, we take the function $\lambda(k,m)$ to be the function defined in equation (1.2).

Lemma 5 *Let $k \geq 2$ and $m \geq 4$. Then $\lambda(k,m) \geq m - \frac{1}{2}$ provided $(k,m) \neq (2,4)$.*

Proof. We have equality when $(k,m) = (2,5)$. For $k \geq 3$ or $m \geq 6$ we have

$$\lambda(k,m) \;=\; m\Big(1 + \frac{1}{(m-2)^k}\Big)^{-1}.$$

Hence

$$\begin{aligned}
\lambda(k,m) &= m\Big(1 + \frac{1}{(m-2)^k}\Big)^{-1} \\
&= m\left(\sum_{i=0}^{\infty}(-1)^i(m-2)^{-ki}\right) \\
&\geq m\Big(1 - 1/(m-2)^k\Big).
\end{aligned}$$

We now divide the proof into two cases.

CASE 1: $k \geq 3$.

In this case, we have

$$\begin{aligned}
\lambda(k,m) &\geq m\Big(1 - 1/(m-2)^k\Big) \\
&\geq m\Big(1 - 1/(m-2)^3\Big) \\
&= m - \frac{m}{(m-2)^3} \\
&\geq m - \frac{4}{2^3} \\
&= m - \frac{1}{2},
\end{aligned}$$

where the penultimate step follows from the fact that $x/(x-2)^3$ is a decreasing function on $[4,\infty)$ and $m \geq 4$. This completes the proof when $k \geq 3$.

CASE 2: $k = 2$, $m \geq 6$.

Here, we use the fact that $x/(x-2)^2$ is decreasing on $[6,\infty)$ to obtain

$$\begin{aligned}\lambda(k,m) &\geq m\Big(1-1/(m-2)^k\Big)\\ &= m\Big(1-1/(m-2)^2\Big)\\ &= m - m/(m-2)^2\\ &\geq m - 6/4^2\\ &\geq m-\frac{1}{2}.\end{aligned}$$

This completes the proof. ∎

Lemma 6 *Let $m \geq 4$, $k \geq 2$, $(k,m) \neq (2,4)$, and let $a_1,\ldots,a_k$ be positive integers each of which is at least 2. Define $b_0, b_1, \ldots$ by*

$$\sum_{i=0}^{\infty} b_i x^i \;=\; \left(1 - mx + \sum_{i_1=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} m^{i_1+\cdots+i_k} x^{a_1 i_1+\cdots+a_k i_k}\right)^{-1}.$$

Then $b_n \geq \lambda(k,m) b_{n-1}$. In particular, $b_n \geq \lambda(k,m)^n$ for all $n \geq 0$.

Proof. We use induction to prove this claim. For the sake of simplicity, we fix k and m and write $\lambda = \lambda(k,m)$.

Since $b_1 = m > \lambda$ and $b_0 = 1$, the claim is true when $n = 1$. Assume that $b_j \geq \lambda b_{j-1}$ for all $j < n$. We now show that $b_n \geq \lambda b_{n-1}$. Computing the coefficient of x^n in both sides of the equation

$$\left(\sum_{i=0}^{\infty} b_i x^i\right)\left(1 - mx + \sum_{i_1=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} m^{i_1+\cdots+i_k} x^{a_1 i_1+\cdots+a_k i_k}\right) \;=\; 1,$$

we see that

$$b_n - m b_{n-1} + \sum_{i_1=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} m^{i_1+\cdots+i_k} b_{n-a_1 i_1-\cdots-a_k i_k} \;=\; 0.$$

Hence

$$b_n \;=\; \lambda b_{n-1} + (m-\lambda) b_{n-1} - \sum_{i_1=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} m^{i_1+\cdots+i_k} b_{n-a_1 i_1-\cdots-a_k i_k}.$$

Thus to show that $b_n \geq \lambda b_{n-1}$, it is sufficient to show that

$$(m-\lambda) b_{n-1} - \sum_{i_1=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} m^{i_1+\cdots+i_k} b_{n-a_1 i_1-\cdots-a_k i_k} \;\geq\; 0. \qquad (2.3)$$

By the inductive hypothesis,

$$b_{n-i} \;\leq\; \frac{b_{n-1}}{\lambda^{i-1}} \qquad \text{for } 1 \leq i \leq n.$$

Thus

$$\sum_{i_1=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} m^{i_1+\cdots+i_k} b_{n-a_1 i_1-\cdots-a_k i_k} \le \sum_{i_1=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} m^{i_1+\cdots+i_k}\frac{b_{n-1}}{\lambda^{a_1 i_1+\cdots+a_k i_k-1}} = \lambda b_{n-1}\sum_{i_1=1}^{\infty}\frac{m^{i_1}}{\lambda^{a_1 i_1}}\cdots\sum_{i_k=1}^{\infty}\frac{m^{i_k}}{\lambda^{a_k i_k}}. \tag{2.4}$$

Since $a_i \ge 2$ for $i \le k$ and $\lambda > \sqrt{m}$, we conclude

$$\lambda b_{n-1}\sum_{i_1=1}^{\infty}\frac{m^{i_1}}{\lambda^{a_1 i_1}}\cdots\sum_{i_k=1}^{\infty}\frac{m^{i_k}}{\lambda^{a_k i_k}} \le \lambda b_{n-1}\sum_{i_1=1}^{\infty}\frac{m^{i_1}}{\lambda^{2 i_1}}\cdots\sum_{i_k=1}^{\infty}\frac{m^{i_k}}{\lambda^{2 i_k}} = \lambda b_{n-1}\left(\sum_{i=1}^{\infty}\frac{m^{i}}{\lambda^{2i}}\right)^k = \lambda b_{n-1}\Big(\frac{m}{\lambda^2-m}\Big)^k. \tag{2.5}$$

Combining equations (2.4) and (2.5), we see

$$\sum_{i_1=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} m^{i_1+\cdots+i_k} b_{n-a_1 i_1-\cdots-a_k i_k} \ \le\ \lambda b_{n-1}\Big(\frac{m}{\lambda^2-m}\Big)^k. \tag{2.6}$$

Using equations (2.3) and (2.6), we see that to complete the proof it is sufficient to prove that

$$(m-\lambda) \ge \lambda\Big(\frac{m}{\lambda^2-m}\Big)^k. \tag{2.7}$$

We have $\lambda \ge m - \frac{1}{2}$ by Lemma 5. Hence

$$\lambda\Big(\frac{m}{\lambda^2-m}\Big)^k \le \lambda\left(\frac{m}{(m-1/2)^2-m)}\right)^k = \lambda\Big(\frac{m}{m^2-2m+1/4}\Big)^k \le \lambda\Big(\frac{m}{m^2-2m}\Big)^k = \lambda\Big(\frac{1}{m-2}\Big)^k. \tag{2.8}$$

Observe that

$$\lambda\Big(1+\frac{1}{(m-2)^k}\Big) \ = \ m$$

and hence

$$\lambda\Big(\frac{1}{m-2}\Big)^k \;=\; m-\lambda. \tag{2.9}$$

Combining inequalities (2.8) and (2.9) we deduce that

$$\lambda\Big(\frac{m}{\lambda^2-m}\Big)^k \;=\; m-\lambda.$$

This completes the proof. ∎

Lemma 7 *Let $k \geq 1$ and let p be a pattern on a k-letter alphabet $\mathcal{T} = \{t_1,\ldots,t_k\}$ in which t_i occurs $a_i \geq 1$ times for $1 \leq i \leq k$. Let c_n be the number of words of length n on $\mathcal{X} = \{X_1,\ldots,X_m\}$ of the form p. Then*

$$\sum_{n\geq 2} c_n x^n \;=\; \sum_{i_1=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} m^{i_1+\cdots+i_k} x^{a_1 i_1+\cdots+a_k i_k}.$$

In particular, if $a_i \geq 2$ for $1 \leq i \leq k$, then $c_n \leq 2^{n-1}m^{n/2}$.

Proof. Given a word W, we let $\ell(W)$ denote its length. Let $\mathcal{S}(\mathsf{p},\mathcal{X})$ be as defined in equation (1.1). Observe that since p has k letters that occur, there is a surjection from the ordered k-tuples of non-empty words on the alphabet $\mathcal{X}$ onto the set $\mathcal{S}(\mathsf{p},\mathcal{X})$. Thus

$$\begin{aligned}
\sum_{i\geq 2} c_i x^i &= \sum_{W\in\mathcal{S}(\mathsf{p},\mathcal{X})} x^{\ell(W)} \\
&\leq \sum_{W_1\in\mathcal{X}^*\setminus\{1\}}\cdots\sum_{W_k\in\mathcal{X}^*\setminus\{1\}} x^{a_1\ell(W_1)+a_2\ell(W_2)+\cdots+a_k\ell(W_k)} \\
&= \sum_{W_1\in\mathcal{X}^*\setminus\{1\}} x^{a_1\ell(W_1)} \sum_{W_2\in\mathcal{X}^*\setminus\{1\}} x^{a_2\ell(W_2)}\cdots \sum_{W_k\in\mathcal{X}^*\setminus\{1\}} x^{a_k\ell(W_k)} \\
&= \sum_{i_1=1}^{\infty} m^{i_1}x^{a_1 i_1}\cdots\sum_{i_k=1}^{\infty} m^{i_k}x^{a_k i_k} \\
&= \sum_{i_1=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} m^{i_1+\cdots+i_k}x^{a_1 i_1+\cdots+a_k i_k}.
\end{aligned}$$

Next, observe that if $a_1,\ldots,a_k \geq 2$ and $a_1 i_1+\cdots+a_k i_k = n$, then

$$m^{i_1+\cdots+i_k} \;\leq\; m^{(a_1 i_1+\cdots+a_k i_k)/2} \;=\; m^{n/2}.$$

Hence

$$c_n \;\leq\; m^{n/2}\#\{a_1 i_1+\cdots+a_k i_k = n \mid i_1,\ldots,i_k \geq 1\}.$$

But

$$
\begin{aligned}
\#\{a_1 i_1 + \cdots + a_k i_k = n \mid i_1, \ldots, i_k \geq 1\} &= [x^n] \sum_{i_1=1}^{\infty} \cdots \sum_{i_k=1}^{\infty} x^{a_1 i_1 + \cdots + a_k i_k} \\
&= [x^n] \sum_{i_1=1}^{\infty} x^{a_1 i_1} \cdots \sum_{i_k=1}^{\infty} x^{a_k i_k} \\
&\leq [x^n] \Big(x + x^2 + x^3 + \cdots \Big)^k \\
&= [x^n] x^k / (1-x)^k \\
&= [x^{n-k}] (1-x)^{-k} \\
&= \binom{n-1}{k-1} \\
&\leq 2^{n-1},
\end{aligned}
$$

where $[x^j]F(x)$ represents the coefficient of x^j in the power series expansion of $F(x)$. Hence $c_n \leq 2^{n-1} m^{n/2}$. ∎

Proof of Theorem 1 Let $\mathcal{X} = \{X_1, X_2, \ldots, X_m\}$. We create the *free algebra* $\mathbb{C}\{X_1, X_2, \ldots, X_m\}$; this is just the algebra of "non-commutative polynomials" in $X_1, X_2, \ldots, X_m$. Let $\mathcal{S}(\mathsf{p}, \mathcal{X})$ be as defined in equation (1.1). and let I be the ideal in $\mathbb{C}\{X_1, X_2, \ldots, X_m\}$ generated by the words in $\mathcal{S}(\mathsf{p}, \mathcal{X})$. We define

$$A \ := \ \mathbb{C}\{X_1, X_2, \ldots, X_m\}/I.$$

We note that a basis for the images in A of the homogeneous elements of $\mathbb{C}\{X_1, \ldots, X_m\}$ of degree n is given by the set of words of length n that avoid p; we denote by a_n the size of this set.

We define c_n to be the number of words of length n in $\mathcal{S}(\mathsf{p}, \mathcal{X})$. By Lemma 7, we have

$$\sum_{i \geq 2} c_i x^i \ = \ \sum_{i_1=1}^{\infty} \cdots \sum_{i_k=1}^{\infty} m^{i_1 + \cdots + i_k} x^{a_1 i_1 + \cdots + a_k i_k}.$$

Hence by Lemma 6 the numbers $b_0, b_1, \ldots$ defined by

$$\sum_{n=i}^{\infty} b_i x^i \ = \ \left(1 - mx + \sum_{i=2}^{\infty} c_i x^i\right)^{-1}$$

satisfy

$$b_n \ \geq \ \lambda(k, m)^n.$$

By Theorem 4 we have $a_n \geq b_n \geq \lambda(k, m)^n$ for all $n \geq 1$. This completes the proof. ∎

3 Simultaneous multiple pattern avoidance

In this section, we extend our results on pattern avoidance to words avoiding multiple patterns simultaneously. Unfortunately, analogues of Lemma 6 for multiple patterns are messy. For this reason, we use cruder estimates which, although not ideal, allow us to prove Theorem 2.

We begin with the following lemma.

Lemma 8 *Let $\varepsilon \in \{0,1\}$ and let m and d be positive integers. Define*

$$F(x) \;:=\; m - x - \frac{m\varepsilon}{x} - \frac{4dm\sqrt{m}}{x^2 - 2\sqrt{m}x}.$$

Then the following hold:

- *if $m \geq 36d^{2/3}$ then $F(m/2) > 0$;*
- *$F(m - \varepsilon - 5d/\sqrt{m}) > 0$ for all m sufficiently large.*

Proof. For the first part of the lemma, notice that if $m \geq 36d^{2/3}$, then

$$\begin{aligned}
F(m/2) &= m - m/2 - m\varepsilon/(m/2) - 16dm\sqrt{m}/(m^2 - 4\sqrt{m}m)\\
&= m/2 - 2\varepsilon - 16d/(\sqrt{m} - 4)\\
&\geq 18d^{2/3} - 2\varepsilon - 16d/(6d^{1/3} - 4)\\
&= 18d^{2/3} - 2\varepsilon - 16d^{2/3}/(6 - 4d^{-1/3})\\
&\geq 18d^{2/3} - 2\varepsilon - 16d^{2/3}/2\\
&= 10d^{2/3} - 2\varepsilon\\
&> 0.
\end{aligned}$$

This establishes the first part of the lemma.

Observe that when $x = m - \varepsilon - 5d/\sqrt{m}$,

$$4dm\sqrt{m}/(x^2 - 2\sqrt{m}x) \;=\; 4d/\sqrt{m} + \mathrm{O}(1/m)$$

and

$$m\varepsilon/x \;=\; \varepsilon + \mathrm{O}(1/m).$$

Hence when $x = m - \varepsilon - 5d/\sqrt{m}$, we have

$$F(x) \;=\; \frac{5d}{\sqrt{m}} - \frac{4d}{\sqrt{m}} + \mathrm{O}(1/m) \;=\; \frac{d}{\sqrt{m}}\Big(1 + \mathrm{O}(m^{-1/2})\Big) \qquad \text{as } m \to \infty.$$

Hence $F(x) > 0$ for all m sufficiently large when $x = m - \varepsilon - 5d/\sqrt{m}$. ∎

Proposition 9 *Let $d \geq 1, m \geq 4$ and let $\varepsilon \in \{0,1\}$ Define $b_0, b_1, \ldots$ by*

$$\sum_{i=0}^{\infty} b_i x^i \;=\; \left(1 - mx + m\varepsilon x^2 + \sum_{i=3}^{\infty} 2^{i-1} d m^{i/2} x^i\right)^{-1}$$

and define

$$F(x) \;:=\; m - x - \frac{m\varepsilon}{x} - \frac{4dm\sqrt{m}}{x^2 - 2\sqrt{m}x}.$$

If $\lambda > 2\sqrt{m}$ and $F(\lambda) > 0$, then $b_n \geq \lambda b_{n-1}$ for all $n \geq 1$.

Proof. We use induction to prove the claim. We have

$$\sum_{i=0}^{\infty} b_i x^i \left(1 - mx + m\varepsilon x^2 + \sum_{i=3}^{\infty} 2^{i-1} d m^{i/2} x^i\right) \;=\; 1.$$

Equating the coefficients of x^n on both sides of this equation, we see

$$b_n - m b_{n-1} + m\varepsilon b_{n-2} + \sum_{i=3}^{\infty} 2^{i-1} d m^{i/2} b_{n-i} \;=\; 0. \tag{3.10}$$

Observe that if $F(\lambda) > 0$, then $\lambda < m$ and since $b_1 = m$, the claim is true when $n = 1$. Assume that $b_j \geq \lambda b_{j-1}$ for $j < n$ and consider the case when $j = n$. Using equation (3.10), we see

$$b_n \;=\; \lambda b_{n-1} + (m - \lambda) b_{n-1} - m\varepsilon b_{n-2} \sum_{i=3}^{\infty} 2^{i-1} d m^{i/2} b_{n-i}. \tag{3.11}$$

Hence, it is sufficient to show that

$$(m - \lambda) b_{n-1} - m\varepsilon b_{n-2} \sum_{i=3}^{\infty} 2^{i-1} d m^{i/2} b_{n-i} \;\geq\; 0.$$

By the inductive hypothesis,

$$b_{n-i} \leq \frac{b_{n-1}}{\lambda^{i-1}}$$

and hence

$$\sum_{i=3}^{\infty} 2^{i-1} d m^{i/2} b_{n-i} \;\geq\; \sum_{i=3}^{\infty} 2^{i-1} d m^{i/2} b_{n-1} \lambda^{-i+1}.$$

Thus it is sufficient to show

$$(m - \lambda) \geq m\varepsilon\lambda^{-1} + \sum_{i=3}^{\infty} 2^{i-1} d m^{i/2} \lambda^{-i+1}$$

$$
\begin{aligned}
&= m\varepsilon\lambda^{-1} + 4dm\sqrt{m}\lambda^{-2}\frac{1}{1-2\sqrt{m}/\lambda} \\
&= (m-\lambda) - F(\lambda),
\end{aligned}
$$

where the penultimate step uses the fact that $\lambda > 2\sqrt{m}$. By assumption, $F(\lambda) > 0$, and hence we see that this inequality holds. The result now follows by induction. ∎

Proof of Theorem 2. Let $\mathcal{X}$ be an m letter alphabet and let c_n be the number of words on $\mathcal{X}$ of the form p for some $\mathsf{p} \in \mathcal{S}$. By Lemma 7,

$$c_n \ \leq \ 2^{n-1}dm^{n/2} \qquad \text{for } n \geq 3.$$

Notice that if $\mathcal{S}$ does not contain the pattern t^2, then $c_2 = 0$; otherwise, $c_2 = m$. Hence

$$\sum_{n=2}^{\infty} c_n x^n \ \leq \ \varepsilon m x^2 + \sum_{n\geq 3} 2^{n-1}dm^{n/2}x^n,$$

where $\leq$ is taken coefficient-wise and ε is 1 if and only if $\mathcal{S}$ contains the pattern t^2. By Proposition 9 and Lemma 8,

$$\sum b_i x^i \ := \ \Big(1 - mx + m\varepsilon x^2 + \sum_{i\geq 3} 2^{i-1}dm^{i/2}x^i\Big)^{-1}$$

has coefficients satisfying:

- $b_n \geq (m/2)^n$ if $m \geq 36d^{2/3}$;
- $b_n \geq (m - \varepsilon - 5d/\sqrt{m})^n$ for m sufficiently large.

Just as in the proof of Theorem 1, Golod's theorem gives that the number of words of length n on our m letter alphabet which avoid $\mathcal{S}$ is at least b_n. The result now follows. ∎

Proof of Corollary 3. By Fekete's lemma (cf. Madras and Slade [9, Lemma 1.2.2]) the limit giving $\Gamma(m, \mathcal{S})$ exists. Furthermore, since there are m^n words of length n on an m letter alphabet, we see that $\Gamma(m, \mathcal{S}) \geq 0$. By Theorem 2 if the pattern $t^2 \notin \mathcal{S}$, then $\Gamma(m, \mathcal{S}) \leq 5d/\sqrt{m}$ for all m sufficiently large. Hence $\Gamma(m, \mathcal{S}) \to 0$ as $m \to \infty$ if $t^2 \notin \mathcal{S}$.

If, on the other hand, $t^2 \in \mathcal{S}$, then Theorem 2 gives that $\Gamma(m, \mathcal{S}) \leq 1 + 5d/\sqrt{m}$ for m sufficiently large and hence

$$\limsup_{m\to\infty} \Gamma(m, \mathcal{S}) \leq 1. \tag{3.12}$$

It is well-known that the number of squarefree words of length n on an m letter alphabet is at most $m(m-1)^{n-1}$ since consecutive letters must be different.

This gives that

$$\Gamma(m,\mathcal{S}) \geq m-(m-1) \ = \ 1.$$

Hence

$$\liminf_{m\to\infty} \Gamma(m,\mathcal{S}) \geq 1. \tag{3.13}$$

Combining inequalities (3.12) and (3.13) we obtain the desired result. ∎

4 Computational results

In this section we give sharper estimates for the number of words of length n on an m letter alphabet that avoid certain patterns on a k letter alphabet for small values of k and m.

We note that in the inductive argument used in the proof of Lemma 6, inequality (2.7)

$$m^k\lambda \leq (\lambda^2-m)^k(m-\lambda)$$

was the important step needed. It is not possible in general to obtain a closed form for the solutions in λ to the equation

$$m^k\lambda \ = \ (\lambda^2-m)^k(m-\lambda). \tag{4.14}$$

Ultimately we used an approximation given by the function $\lambda(k,m)$ given in equation (1.2). It is possible, however, to compute solutions to equation (4.14) for small values of k and m using Maple. Let $\phi(k,m)$ denote the largest real solution in λ to equation (4.14) that is greater than $\sqrt{m}$ (when such a solution exists). Then if p is a pattern on a k letter alphabet in which each letter occurs at least twice, then there are at least $\phi(k,m)^n$ words of length n on an m letter alphabet that avoid p. (This follows from following the induction argument in Lemma 6 and noting that $\lambda = \phi(k,m)$ satisfies the inequality (2.7).)

Table 1 displays the values of $\phi(k,m)$ to 7 decimal places of accuracy for $2 \leq k \leq 10$ and $3 \leq m \leq 8$. Table 2 displays values of $\lambda(k,m)$ to 8 decimal places of accuracy for $2 \leq k \leq 10$ and $4 \leq m \leq 8$. From these tables we see that $\lambda(k,m)$ is not so far from the $\phi(k,m)$, which is the best value that can be obtained by our methods. Certain entries in the tables are blank because the functions $\phi(k,m)$ and $\lambda(k,m)$ are not defined at all values (k,m) listed in the tables. These tables were computed using Maple.

We note that the function

$$F_k(x) \ := 3^kx-(x^2-3)^k(3-x)$$

Table 1
Values of $\phi(k, m)$ for $2 \le k \le 10$, $3 \le m \le 7$

k	$\phi(k,3)$	$\phi(k,4)$	$\phi(k,5)$	$\phi(k,6)$	$\phi(k,7)$	$\phi(k,8)$
2			4.5297921	5.7094634	6.7813902	7.8231030
3		3.7762249	4.9122992	5.9494073	6.9666496	7.9762667
4		3.9432625	4.9797410	5.9902649	6.9945636	7.9966566
5		3.9826299	4.9950613	5.9980732	6.9990985	7.9995237
6	2.9293298	3.9943963	4.9987751	5.9996156	6.9998499	7.9999319
7	2.9716300	3.9981560	4.9996945	5.9999231	6.9999749	7.9999902
8	2.9870526	3.9993884	4.9999236	5.9999846	6.9999958	7.9999986
9	2.9938174	3.9997965	4.9999809	5.9999969	6.9999993	7.9999998
10	2.9969835	3.9999322	4.9999952	5.9999993	6.9999998	7.9999999

Table 2
Values of $\lambda(k, m)$ for $2 \le k \le 10$, $4 \le m \le 8$

k	$\lambda(k,4)$	$\lambda(k,5)$	$\lambda(k,6)$	$\lambda(k,7)$	$\lambda(k,8)$
2		4.50000000	5.64705882	6.73076923	7.78378378
3	3.55555555	4.82142857	5.90769230	6.94444444	7.96313364
4	3.76470588	4.93902439	5.97665369	6.98881789	7.99383192
5	3.87878787	4.97950819	5.99414634	6.99776071	7.99897132
6	3.93846153	4.99315068	5.99853551	6.99955202	7.99982853
7	3.96899224	4.99771480	5.99963381	6.99991040	7.99997142
8	3.98443579	4.99923803	5.99990844	6.99998208	7.99999523
9	3.99220272	4.99974598	5.99997711	6.99999641	7.99999920
10	3.99609756	4.99991532	5.99999427	6.99999928	7.99999986

satisfies $F_k(3) = 3^{k+1}$ and $F_k(2.8) = (2.8)3^k - (.2)(4.84)^k$, which is negative for $k \ge 6$. By the intermediate value theorem, there is a real positive solution to $F_k(x) = 0$ with $x \in (2.8, 3)$. It follows that $\phi(k, 3) > 2.8$ for $k \ge 6$ and hence a pattern on a k letter alphabet in which each letter occurs at least twice is avoidable on a 3 letter alphabet, and furthermore the number of words of length n that avoid the pattern grows exponentially with n. We do not get exponential lower bounds for $m = 3$ and $k < 6$. We cannot get exponential lower bounds in the case that $m = 2$ with our methods. In general, the smallest size m such that an avoidable pattern is avoidable on an m letter alphabet is called the *avoidability index*. It is a notoriously hard problem to compute the avoidability index even in the case that the pattern is on a 3 letter alphabet. For example, it appears that it is unknown whether the avoidability index of $t_1^2t_2^2t_3^2$ is 2 or 3 [8, §3.3.1].

5 Open problems and concluding remarks

We conclude by making a few simple remarks and giving some questions which we are unable to solve.

Lothaire [8, Cor. 3.2.11] notes that patterns on a k letter alphabet that have length at least 2^k are avoidable. This follows from an induction argument noting that patterns in which each letter that occurs, occurs at least twice are avoidable. It is well-known that the number of words of length n on an m letter alphabet that avoid a pattern of the form t^j with $j \geq 2$ grows exponentially with n if $m \geq 3$. From this we see that both Theorem 1 and 2 apply to patterns on a k letter alphabet that have length at least 2^k.

We also make the remark that in the statement of Theorem 2 there is a constant C which depends on d. We did not specify C, instead preferring to use O-notation in the proof of Lemma 8 to avoid complicating the exposition. We note, however, that in the proof of Lemma 8, if one is more careful with the estimates, then the conclusion that $F(m - \varepsilon - 5d/\sqrt{m}) > 0$ for $m > 200d$ can be obtained. In fact, for d large, one can take m much smaller. From this we see that we can take $C = 200d$ in the statement of Theorem 2.

Next, let $\mathcal{S}$ be a finite set of avoidable patterns and let m be a positive integer. Let $\Gamma(m, \mathcal{S})$ be defined as in the statement of Corollary 3. Corollary 3 shows that $\Gamma(m, \mathcal{S})$ is well-behaved when $\mathcal{S}$ consists of patterns in which each letter that occurs, occurs at least twice. It is natural to ask if $\Gamma(m, \mathcal{S})$ is well-behaved in general.

Question 1 *Let $\mathcal{S}$ be a finite set of avoidable patterns. Is it true that $\lim_{m\to\infty} \Gamma(m, \mathcal{S})$ exists? Is it finite? Is it an integer?*

Question 2 *Let $\mathcal{S}$ be a finite set of patterns in which every letter that occurs, occurs at least twice. What can be said about the rate at which $\Gamma(m, \mathcal{S})$ tends to its limit?*

We note that Theorem 2 shows that if $L = \lim_{m\to\infty} \Gamma(m, \mathcal{S})$, then $\Gamma(m, \mathcal{S}) = L + \mathrm{O}(m^{-1/2})$ as m tends to infinity. This, however, is a crude estimate and it would be interesting to get an exact asymptotic estimate for $\Gamma(m, \mathcal{S}) - L$. In the case that $\mathcal{S}$ consists of a single pattern on a k letters in which each letter occurs at least twice and $k \geq 2$, Theorem 1 shows that $\Gamma(m, \mathcal{S}) = O(1/m^k)$ as $m \to \infty$.

Question 3 *Is it true that the collection of patterns in which every letter that occurs, occurs at least twice are* all *simultaneously avoidable on an m letter alphabet for some m? Can one in fact take $m = 4$?*

Currie has an interesting conjecture about patterns on 4 letter words. He speculates [5, p. 791] that an avoidable pattern is in fact avoidable on a 4 letter alphabet. Theorem 1 along with facts about squarefree words show that this is true for patterns in which each letter that occurs, occurs at least twice.

6 Acknowledgments

We thank James Currie and the referee for many valuable comments.

References

[1] S. Adjan. The Burnside problem and identities in groups. *Ergeb. Math. Grenzgeb.* vol. 95, Springer 1979.

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

[3] J. Berstel. Some recent results on squarefree words. STACS 84 (Paris, 1984), 14–25, *Lecture Notes in Comput. Sci.,* 166, Springer, Berlin, 1984.

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

[5] J. D. Currie. Open problems in pattern avoidance. *Amer. Math. Monthly* **100** (1993), 790–793.

[6] J. D. Currie. Pattern Avoidance: Themes and Variations. *Theoret. Comput. Sci.* **339** (2005) no. 1, 7–18.

[7] J. Karhumäki and J. Shallit. Polynomial versus exponential growth in repetition-free binary words. *J. Comb. Theory Ser. A* **105** (2004) no. 2, 335–347.

[8] M. Lothaire. *Algebraic combinatorics on words.* Encyclopedia of mathematics and its applications v. 90, Cambridge University press, Cambridge, New York.

[9] N. Madras and G. Slade. *The self-avoiding walk.* Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA, 1993.

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

[11] T. Nagell, A. Selberg, S. Selberg, and K. Thalberg (Editors). *Selected Mathematical Papers of Axel Thue.* Oslo, Norway, Universitetsforlaget, 1977.

[12] J. Noonan and D. Zeilberger. The Goulden-Jackson Cluster Method: Extensions, Applications, and Implementations. *J. Differ. Eq. Appl.* **5** (1999), 355–377.

[13] L. Rowen. *Ring theory. Vol. II.* Pure and Applied Mathematics, 128. Academic Press, Inc., Boston, MA, 1988.

[14] A. Thue. Über unendliche Zeichenreihen. *Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. Christiana* **7** (1906), 1–22.

[15] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. Christiana* **1** (1912), 1–67.