Strict Bounds for Pattern Avoidance

By: F. Blanchet-Sadri, Brent Woodhouse

Blanchet-Sadri, F., Woodhouse, B. (2013). Strict Bounds for Pattern Avoidance. *Theoretical Computer Science, 506,* 17-28. doi: 10.1016/j.tcs.2013.08.010

Abstract:

Cassaigne conjectured in 1994 that any pattern with *m* distinct variables of length at least $3(2^{m-1})$ is avoidable over a binary alphabet, and any pattern with *m* distinct variables of length at least 2^m is avoidable over a ternary alphabet. Building upon the work of Rampersad and the power series techniques of Bell and Goh, we obtain both of these suggested strict bounds. Similar bounds are also obtained for pattern avoidance in partial words, sequences where some characters are unknown.

Keywords: Formal languages | Combinatorics on words | Pattern avoidance | Power series | Partial words

Article:

1. Introduction

Let Σ be an alphabet of letters, denoted by a,b,c,… , and Δ be an alphabet of variables, denoted byA,B,C,… . A *pattern p* is a word over Δ . A word *w* over Σ is an *instance* of *p* if there exists a non-erasing morphism $\varphi:\Delta^{\square}\to\Sigma^{\square}$ such that $\varphi(p)=w$. A word *w* is said to *avoid p* if no factor of *w* is an instance of *p* . For example, aa b aa c contains an instance of *ABA* while abaca avoids *AA*.

A pattern *p* is *avoidable* if there exist infinitely many words *w* over a finite alphabet such that *w* avoids *p*, or equivalently, if there exists an infinite word that avoids *p*. Otherwise *p* is *unavoidable*. If *p* is avoided by infinitely many words over a *k*-letter alphabet, *p* is said to be *k-avoidable*. Otherwise, *p* is *k-unavoidable*. If *p*is avoidable, the minimum *k* such that p is k-avoidable is called the *avoidability index* of p. If p is unavoidable, the avoidability index is defined as ∞. For example, *ABA* is unavoidable while *AA* has avoidability index 3.

If a pattern p occurs in a pattern q, we say *p divides q*. For example, p=ABA divides $q = \underline{ABC}\ \underline{BB}\ \underline{ABC}A$, since we can map *A* to *ABC* and *B* to *BB* and this maps *p* to a factor of *q*.
If *p* divides *q* and *p* is *k*-avoidable, there exists an infinite word *w* over a *k*-letter alphabet that avoids *p*; *w* must also avoid *q*, thus *q* is necessarily *k*-avoidable. It follows that the avoidability index of *q* is less than or equal to the avoidability index of *p*. Chapter 3 of Lothaire [6] is a nice summary of background results in pattern avoidance.

It is not known if it is generally decidable, given a pattern p and integer k, whether p is k-avoidable. Thus various authors compute avoidability indices and try to find bounds on them. Cassaigne [5] listed avoidability indices for unary, binary, and most ternary patterns (Ochem [8] determined the remaining few avoidability indices for ternary patterns). Based on this data, Cassaigne conjectured in his 1994 Ph.D. thesis [5, Conjecture 4.1] that any pattern with m distinct variables of length at least $3(2^{m-1})$ is avoidable over a binary alphabet, and any pattern with m distinct variables of length at least 2^m is avoidable over a ternary alphabet. This is also [6, Problem 3.3.2].

The contents of our paper are as follows. In Section 2, we establish that both bounds suggested by Cassaigne are strict by exhibiting well-known sequences of patterns that meet the bounds. Note that the results of Section 2 were proved by Cassaigne in his Ph.D. thesis with the same patterns (see [5, Proposition 4.3]). We recall them here for sake of completeness. In Section 3, we provide foundational results for the power series approach to this problem taken by Bell and Goh [1] and Rampersad [10], then proceed to prove the strict bounds in Section 4. In Section 5, we apply the power series approach to obtain similar bounds for avoidability in partial words, sequences that may contain some do-not-know characters, or holes, which are*compatible* or match any letter in the alphabet. The modifications include that now we must avoid all partial words compatible with instances of the pattern. Lots of additional work with inequalities is necessary. Finally in Section 6, we conclude with various remarks and conjectures.

2. Two sequences of unavoidable patterns

The following proposition allows the construction of sequences of unavoidable patterns.

Proposition 1.

(See [6, Proposition 3.1.3].) Let p be a k-unavoidable pattern over Δ *and* $A \in \Delta$ *be a variable that does not occur in p. Then the pattern pAp is k-unavoidable.*

Let $A_1, A_2, \ldots$ be distinct variables in Δ. Define $Z_0 = \varepsilon$, the empty word, and for all integers $m \geqslant 0, Z_{m+1} = Z_m A_{m+1} Z_m$. The patterns Z_m are called Zimin words. Since ε is k-unavoidable for every positive integer k, Proposition 1 implies Z_m is k-unavoidable for all $m \in \mathbb{N}$ by induction on m. Thus all the Zimin words are unavoidable. Note that Z_m is over m variables and $|Z_m| = 2^m - 1$. Thus there exists a 3-unavoidable pattern over m variables with length $2^m - 1$ for all $m \in \mathbb{N}$.

Likewise, define $R_1 = A_1A_1$ and for all integers $m \geqslant 1$, $R_{m+1} = R_m A_{m+1} R_m$. Since A_1A_1 is 2-unavoidable, Proposition 1 implies R_m is 2-unavoidable for all $m \in \mathbb{N}$ by induction on m. Note that R_m is over m variables; induction also yields $|R_m| = 3(2^{m-1}) - 1$. Thus there exists a 2-unavoidable pattern over m variables with length $3(2^{m-1}) - 1$ for all $m \in \mathbb{N}$.

3. The power series approach

The following theorem was originally presented by Golod (see [12, Lemma 6.2.7]) and rewritten and proven with combinatorial terminology by Rampersad.

Theorem 1.

(See [10, Theorem 2].) Let S be a set of words over a k-letter alphabet with each word of length at least two. Suppose that for each $i \geqslant 2$*, the set S contains at most* c_i *words of length i. If the power series expansion of*

$$B(x) := \left(1 - kx + \sum_{i \geqslant 2} c_i x^i\right)^{-1}$$

has non-negative coefficients, then there are at least $[x^n]B(x)$ *words of length n over a k-letter alphabet that have no factors in S.*

To count the number of words of length n avoiding a pattern p, we let S consist of all instances of p. To use Theorem 1, we require an upper bound c_i on the number of words of length i in S. The following lemma due to Bell and Goh provides a useful upper bound.

Lemma 1.

(See [1, Lemma 7].) Let $m \geqslant 1$ *be an integer and p be a pattern over an alphabet* $\Delta = \{A_1, \ldots, A_m\}$. *Suppose that for* $1 \leqslant i \leqslant m$, *the variable* A_i *occurs* $d_i \geqslant 1$ *times in p. Let* $k \geqslant 2$ *be an integer and let* Σ *be a k-letter alphabet. Then for* $n \geqslant 1$*, the number of words of length n over* Σ *that are instances of the pattern p is no more than* $[x^n]C(x)$*, where*

$$C(x) := \sum_{i_1 \geqslant 1} \cdots \sum_{i_m \geqslant 1} k^{i_1 + \cdots + i_m} x^{d_1 i_1 + \cdots + d_m i_m}.$$

Note that this approach for counting instances of a pattern is based on the frequencies of each variable in the pattern, so it will not distinguish AABB and ABAB, for example.

4. Derivation of the strict bounds

First we prove a technical inequality.

Lemma 2.

Suppose $k\geqslant 2$ *and* $m\geqslant 1$ *are integers and* $\lambda > \sqrt{k}$. *For any integer P and integers* d_j *for* $1\leqslant j\leqslant m$ *such that* $d_j\geqslant 2$ *and* $P=d_1+\cdots+d_m$,

equation(1)

$$\prod_{i=1}^{m}\frac{1}{\lambda^{d_i}-k}\leqslant\left(\frac{1}{\lambda^2-k}\right)^{m-1}\left(\frac{1}{\lambda^{P-2(m-1)}-k}\right).$$

Proof.

The proof is by induction on m . For $m=1$, $d_1=P$ and the inequality is trivially satisfied. Suppose Eq. (1) holds for m and $d_1+d_2+\cdots+d_{m+1}=P$ with $d_j\geqslant 2$ for $1\leqslant j\leqslant m+1$. Note that $P\geqslant 4$.

Letting $P'=P-d_{m+1}=d_1+\cdots+d_m$, the inductive hypothesis implies

equation(2)

$$\prod_{i=1}^{m}\frac{1}{\lambda^{d_i}-k}\leqslant\left(\frac{1}{\lambda^2-k}\right)^{m-1}\left(\frac{1}{\lambda^{P'-2(m-1)}-k}\right).$$

If $d_{m+1}=2$, multiplying both sides by

$$\frac{1}{\lambda^{d_{m+1}}-k}=\frac{1}{\lambda^2-k}$$

yields the desired inequality.

Otherwise, $d_{m+1}>2$. If $P'-2(m-1)=2$, multiplying both sides of Eq. (2) by

$$\frac{1}{\lambda^{d_{m+1}}-k}=\frac{1}{\lambda^{P-2m}-k}$$

yields the desired inequality. In the remaining case, $P'-2(m-1)>2$. Let $c_1=P'-2(m-1)$ and $c_2=d_{m+1}$. Since $\lambda > \sqrt{k}$ and $c_1,c_2>2$,

$(\lambda^{c_1-1}-\lambda)(\lambda^{c_2-1}-\lambda)\geqslant 0$,

$\lambda^{c_1+c_2-2}-\lambda^{c_1}-\lambda^{c_2}+\lambda^2\geqslant 0$,

$\lambda^{c_1+c_2-2}+\lambda^2\geqslant\lambda^{c_1}+\lambda^{c_2}$,

$-k(\lambda^{c_1+c_2-2}+\lambda^2)\leqslant -k(\lambda^{c_1}+\lambda^{c_2})$,

$(\lambda^{c_1}-k)(\lambda^{c_2}-k)\geqslant(\lambda^{c_1+c_2-2}-k)(\lambda^2-k)$,

$$\frac{1}{(\lambda^{c_1}-k)(\lambda^{c_2}-k)} \leqslant \frac{1}{(\lambda^{c_1+c_2-2}-k)(\lambda^2-k)}.$$

Substituting the c_i,

equation(3)

$$\frac{1}{(\lambda^{P'-2(m-1)}-k)(\lambda^{d_{m+1}}-k)} \leqslant \frac{1}{(\lambda^{P'-2m+d_{m+1}}-k)(\lambda^2-k)}.$$

Multiplying Eq. (2) by $\frac{1}{\lambda^{d_{m+1}}-k}$,

$$\prod_{i=1}^{m+1} \frac{1}{\lambda^{d_i}-k} \leqslant \left(\frac{1}{\lambda^2-k}\right)^{m-1}\left(\frac{1}{\lambda^{P'-2(m-1)}-k}\right)\frac{1}{\lambda^{d_{m+1}}-k}.$$

Substituting Eq. (3),

$$\begin{aligned}\prod_{i=1}^{m+1} \frac{1}{\lambda^{d_i}-k} &\leqslant \left(\frac{1}{\lambda^2-k}\right)^{m}\left(\frac{1}{\lambda^{P'+d_{m+1}-2m}-k}\right)\\ &= \left(\frac{1}{\lambda^2-k}\right)^{(m+1)-1}\left(\frac{1}{\lambda^{P-2((m+1)-1)}-k}\right),\end{aligned}$$

as desired. □

Remark 1.

We have written Lemma 2 in terms of partitions of *P* with parts of size at least 2. However, as it will be used with P=|p| for some pattern *p* containing d_j occurrences of variable A_j, its statement and its proof could also be written in terms of patterns defining p' to be *p* without its d_{m+1} instances of the (m+1)th variable. Then using the inductive hypothesis on p', the proof would follow as it is.

The remaining arguments in this section are based on those of [10], but add additional analysis to obtain the optimal bound.

Lemma 3.

Let m be an integer and p be a pattern over an alphabet $\Delta=\{A_1,\ldots,A_m\}$. *Suppose that for* $1\leqslant i\leqslant m$, A_i *occurs* $d_i\geqslant 2$ *times in p.*

1. *If* $m\geqslant 3$ *and* $|p|\geqslant 4m$, *then for* $n\geqslant 0$, *there are at least* $(1.92)^n$ *words of length n over a binary alphabet that avoid p.*

2. *If* $m\geqslant 2$ *and* $|p|\geqslant 12$, *then for* $n\geqslant 0$, *there are at least* $(2.92)^n$ *words of length n over a ternary alphabet that avoid p* (*for* $m\geqslant 6$, *this implies that every pattern with each variable occurring at least twice is 3-avoidable*).

Proof.

Let Σ be an alphabet of size $k\in\{2,3\}$. Define S to be the set of all words in Σ^* that are instances of the pattern p. By Lemma 1, the number of words of length n in S is at most $[x^n]C(x)$, where

$$C(x) := \sum_{i_1\geqslant 1}\cdots\sum_{i_m\geqslant 1} k^{i_1+\cdots+i_m}x^{d_1 i_1+\cdots+d_m i_m}.$$

By hypothesis, $d_i\geqslant 2$ for $1\leqslant i\leqslant m$. In order to use Theorem 1 on Σ, define

$$B(x) := \sum_{i\geqslant 0} b_i x^i = \left(1-kx+C(x)\right)^{-1},$$

and set the constant $\lambda=k-0.08$. Clearly $b_0=1$ and $b_1=k$. We show that $b_n\geqslant\lambda b_{n-1}$ for all $n\geqslant 1$, hence $b_n\geqslant\lambda^n$ for all $n\geqslant 0$. Then all coefficients of B are non-negative, thus Theorem 1 implies there are at least $b_n\geqslant\lambda^n$ words of length n avoiding S. By construction of S, these words all avoid p.

We show by induction on n that $b_n\geqslant\lambda b_{n-1}$ for all $n\geqslant 1$. We can easily verify $b_1\geqslant(k-0.08)(1)=\lambda b_0$. Now suppose that for all $1\leqslant j<n$, we have $b_j\geqslant\lambda b_{j-1}$. By definition of B, $B(x)(1-kx+C(x))=1$, hence for $n\geqslant 1$, $[x^n]B(1-kx+C)=0$. Expanding the left-hand side,

$$B(1-kx+C) = \left(\sum_{i\geqslant 0} b_i x^i\right)\left(1-kx+\sum_{i_1\geqslant 1}\cdots\sum_{i_m\geqslant 1} k^{i_1+\cdots+i_m}x^{d_1 i_1+\cdots+d_m i_m}\right),$$

thus

$$[x^n]B(1-kx+C) = b_n - kb_{n-1} + \sum_{i_1\geqslant 1}\cdots\sum_{i_m\geqslant 1} k^{i_1+\cdots+i_m} b_{n-(d_1 i_1+\cdots+d_m i_m)} = 0.$$

Rearranging and adding and subtracting λb_{n-1},

$$b_n = \lambda b_{n-1} + (k-\lambda)b_{n-1} - \sum_{i_1\geqslant 1}\cdots\sum_{i_m\geqslant 1} k^{i_1+\cdots+i_m} b_{n-(d_1 i_1+\cdots+d_m i_m)}.$$

To complete the induction, it thus suffices to show

equation(4)

$$(k-\lambda)b_{n-1} - \sum_{i_1\geqslant 1}\cdots\sum_{i_m\geqslant 1} k^{i_1+\cdots+i_m} b_{n-(d_1 i_1+\cdots+d_m i_m)} \geqslant 0.$$

Because $b_j \geqslant \lambda b_{j-1}$ for $1 \leqslant j<n$, $b_{n-i} \leqslant b_{n-1}/\lambda^{i-1}$ for $1 \leqslant i \leqslant n$. Note that the bound on b_{n-i} is stated for $1 \leqslant i \leqslant n$, but actually it is used also for $i>n$, with the implicit convention that $b_{n-i}=0$ in this case. Therefore,

$$\sum_{i_1 \geqslant 1} \cdots \sum_{i_m \geqslant 1} k^{i_1+\cdots+i_m} b_{n-(d_1 i_1+\cdots+d_m i_m)} \leqslant \sum_{i_1 \geqslant 1} \cdots \sum_{i_m \geqslant 1} \frac{k^{i_1+\cdots+i_m}}{\lambda^{d_1 i_1+\cdots+d_m i_m}} \lambda b_{n-1} = \lambda b_{n-1} \sum_{i_1 \geqslant 1} \frac{k^{i_1}}{\lambda^{d_1 i_1}} \cdots \sum_{i_m \geqslant 1} \frac{k^{i_m}}{\lambda^{d_m i_m}}.$$

Since $d_j \geqslant 2$ for $1 \leqslant j \leqslant m$, $k \leqslant 3$, and $\lambda > \sqrt{3}$,

$$\frac{k}{\lambda^{d_j}} \leqslant \frac{3}{\lambda^2} < 1,$$

thus all the geometric series converge. Computing the result, for $1 \leqslant j \leqslant m$,

$$\sum_{i_j \geqslant 1} \frac{k^{i_j}}{\lambda^{d_j i_j}} = \frac{k/\lambda^{d_j}}{1-k/\lambda^{d_j}} = \frac{k}{\lambda^{d_j}-k}.$$

Thus

$$\sum_{i_1 \geqslant 1} \cdots \sum_{i_m \geqslant 1} k^{i_1+\cdots+i_m} b_{n-(d_1 i_1+\cdots+d_m i_m)} \leqslant k^m \lambda b_{n-1} \prod_{i=1}^{m} \frac{1}{\lambda^{d_i}-k}.$$

Applying Lemma 2 to $P=|p|$,

equation(5)

$$\sum_{i_1 \geqslant 1} \cdots \sum_{i_m \geqslant 1} k^{i_1+\cdots+i_m} b_{n-(d_1 i_1+\cdots+d_m i_m)} \leqslant k^m \lambda b_{n-1} \left(\frac{1}{\lambda^2-k}\right)^{m-1} \left(\frac{1}{\lambda^{|p|-2(m-1)}-k}\right).$$

It thus suffices to show

equation(6)

$$(k-\lambda) \geqslant \lambda k^m \left(\frac{1}{\lambda^2-k}\right)^{m-1} \left(\frac{1}{\lambda^{|p|-2(m-1)}-k}\right),$$

since multiplying this by b_{n-1} and using Eq. (5) derives Eq. (4).

To show Statement 1, let $k=2$ and recall we restricted $m \geqslant 3$ and $|p| \geqslant 4m$. Note that the right-hand side of Eq. (6) decreases as $|p|$ increases, thus it suffices to verify the case $|p|=4m$.
Taking $m=3$, $|p|=12$ and

$$k-\lambda=0.08\geqslant 0.02956\ldots=1.92\frac{2^3}{((1.92)^2-2)^2(1.92^{12-2(3-1)}-2)}$$
$$=\lambda k^m\left(\frac{1}{\lambda^2-k}\right)^{m-1}\left(\frac{1}{\lambda^{|p|-2(m-1)}-k}\right).$$

Now consider an arbitrary $m'\geqslant 3$ and p' with $|p'|=4m'$. Substituting $\lambda=1.92$ and $k=2$, it follows that

$$c:=\left(\frac{k}{\lambda^2-k}\right)^{m'-m}\left(\frac{\lambda^{|p|-2(m-1)}-k}{\lambda^{|p'|-2(m'-1)}-k}\right)$$
$$\leqslant(1.19)^{m'-m}\left(\frac{1}{\lambda^{|p'|-2(m'-1)-(|p|-2(m-1))}}\right)=(1.19)^{m'-m}\left(\frac{1}{\lambda^{2(m'-m)}}\right)<1.$$

Thus we conclude

$$k-\lambda\geqslant c\lambda k^m\left(\frac{1}{\lambda^2-k}\right)^{m-1}\left(\frac{1}{\lambda^{|p|-2(m-1)}-k}\right)$$
$$=\lambda k^{m'}\left(\frac{1}{\lambda^2-k}\right)^{m'-1}\left(\frac{1}{\lambda^{|p'|-2(m'-1)}-k}\right).$$

Likewise for Statement 2, for any $m\geqslant 2$, it suffices to verify Eq. (6) for $|p|=\max\{12,2m\}$ (clearly every pattern in which each variable occurs at least twice satisfies $|p|\geqslant 2m$).
For $m=2$ through $m=5$ and $|p|=12$, the equation is easily verified. For $m\geqslant 6$, $|p|=2m$ and

$$\lambda k^m\left(\frac{1}{\lambda^2-k}\right)^{m-1}\left(\frac{1}{\lambda^{|p|-2(m-1)}-k}\right)=2.92\left(\frac{3}{(2.92)^2-3}\right)^m$$
$$\leqslant 2.92(0.5429)^m\leqslant 2.92(0.5429)^6=0.07476\ldots<0.08=k-\lambda.$$

This completes the induction and the proof of the lemma. □

Remark 2.

Referring to Statement 2 of Lemma 3 “for $m\geqslant 6$, every pattern with each variable occurring at least twice is 3-avoidable” is mentioned by Bell and Goh (not as a theorem, but as a remark at the end of [1, Section 4]). They provide a slightly better constant 2.9293298 for the exponential growth in this case. As a consequence, Statement 2 is new only for $m\in\{2,3,4,5\}$. For $m\in\{2,3\}$, patterns of length 12 were known to be avoidable [11] and [5] but without an exponential lower bound.

Here are the main results. As discussed in Section 2, both bounds below are strict in the sense that for every positive integer m, there exists a 2-unavoidable pattern with m variables and length $3(2^{m-1})-1$ as well as a 3-unavoidable pattern with m variables and length 2^m-1.

Theorem 2.

Let p be a pattern with m distinct variables.

1. *If* $|p| \geqslant 3(2^{m-1})$*, then p is 2-avoidable.*

2. *If* $|p| \geqslant 2^m$*, then p is 3-avoidable.*

Proof.

For Statement 1, we show by induction on m that if p is 2-unavoidable, $|p|<3(2^{m-1})$. For m=1, note that A^3 is 2-avoidable [6], hence A^ℓ is 2-avoidable for all $\ell \geqslant 3$. Thus if a unary pattern p is 2-unavoidable, $|p|<3=3(2^{1-1})$. For m=2, it is known that all binary patterns of length 6 are 2-avoidable [11], hence all binary patterns of length at least 6 are also 2-avoidable. Thus if a binary pattern p is 2-unavoidable, $|p|<6=3(2^{2-1})$. Now assume the statement holds for $m \geqslant 2$ and suppose p is a 2-unavoidable pattern withm+1 variables. For the sake of contradiction, assume that $|p| \geqslant 3(2^m)$. There are two cases to consider.

First, if p has a variable A that occurs exactly once, let $p=p_1Ap_2$, where p_1 and p_2 are patterns with at most m variables. Without loss of generality, suppose $|p_1| \geqslant |p_2|$. Since $|p| \geqslant 3(2^m)$,

$$|p_1| \geqslant \left\lceil \frac{|p|-1}{2} \right\rceil \geqslant \left\lceil \frac{3(2^m)-1}{2} \right\rceil = 3\left(2^{m-1}\right).$$

By the contrapositive of the inductive hypothesis, p_1 is 2-avoidable. But p_1 divides p, hence p is 2-avoidable, a contradiction.

Alternatively, suppose every variable in p occurs at least twice.
Since $|p| \geqslant 3(2^m) \geqslant 4(m+1)$ for $m \geqslant 2$,Lemma 3 indicates there are infinitely many words over a binary alphabet that avoid p, thus p is 2-avoidable, a contradiction. These contradictions imply $|p|<3(2^{(m+1)-1})$, which completes the induction.

For Statement 2, we show by induction on m that if p is 3-unavoidable, $|p|<2^m$. For m=1, note that A^2 is 3-avoidable [6], hence A^ℓ is 3-avoidable for all $\ell \geqslant 2$. Thus if a unary pattern p is 3-unavoidable, $|p|<2=2^1$. For m=2, it is known that all binary patterns of length greater than or equal to 4 are 3-avoidable [11]. Form=3, it is known that all ternary patterns of length greater than or equal to 8 are 3-avoidable [5]. Now assume the statement holds for $m \geqslant 3$ and suppose p is a 3-unavoidable pattern with $m+1 \geqslant 4$ variables. For the sake of contradiction, assume that $|p| \geqslant 2^{m+1}$. There are two cases to consider.

First, if p has a variable A that occurs exactly once, let $p=p_1Ap_2$, where p_1 and p_2 are patterns with at most m variables. Without loss of generality, suppose $|p_1| \geqslant |p_2|$. Since $|p| \geqslant 2^{m+1}$,

$$|p_1| \geqslant \left\lceil \frac{|p|-1}{2} \right\rceil \geqslant \left\lceil \frac{2^{m+1}-1}{2} \right\rceil = 2^m.$$

By the contrapositive of the inductive hypothesis, p_1 is 3-avoidable. But p_1 divides *p*, hence *p* is 3-avoidable, a contradiction.

Alternatively, suppose every variable in *p* occurs at least twice. Since we have $m+1\geqslant 4$, $|p|\geqslant 2^{m+1}\geqslant 12$. Thus Lemma 3 indicates there are infinitely many words over a ternary alphabet that avoid *p*, so *p* is 3-avoidable, a contradiction. These contradictions imply $|p|<2^{m+1}$, which completes the induction. □

5. Extension to partial words

A partial word over an alphabet Σ is a concatenation of characters from the extended alphabet $\Sigma_\diamond=\Sigma\cup\{\diamond\}$, where $\diamond$ is called the hole character and represents any unknown letter. If *u* and *v* are two partial words of equal length, we say *u* is *compatible* with *v*, denoted $u\uparrow v$, if $u[i]=v[i]$ whenever $u[i],v[i]\in\Sigma$. A partial word *w* over Σ is an instance of a pattern *p* over Δ if there exists a non-erasing morphism $\varphi:\Delta^{□}\to\Sigma^{□}$ such that $\varphi(p)\uparrow w$; the partial word *w* avoids *p* if none of its factors is an instance of *p*. For example, $aab\,a\diamond c$ contains an instance of *ABA* while it avoids *AAA*.

A pattern *p* is called *k-avoidable* in partial words if for every $h\in N$ there is a partial word with *h* holes over a *k*-letter alphabet avoiding *p*. The *avoidability index* for partial words is defined analogously to that of full words. For example, *AA* is unavoidable in partial words since a factor of the form $a\diamond$ or $\diamond a$ must occur, where $a\in\Sigma_\diamond$, while the pattern AABB has avoidability index 3 in partial words. Classification of avoidability indices for unary and binary patterns is complete and the ternary classification is nearly complete [2] and [3].

The power series method previously used for full words can also count partial words avoiding patterns, and similar results are obtained. Before we can use the power series approach to develop bounds for partial words, we must obtain an upper bound for the number of partial words over Σ that are compatible with instances of the pattern. This result is comparable with Lemma 1 for full words.

Lemma 4.

Let $m\geqslant 1$ *be an integer and p be a pattern over an alphabet* $\Delta=\{A_1,\ldots,A_m\}$. *Suppose that for* $1\leqslant i\leqslant m$, *the variable* A_i *occurs* $d_i\geqslant 1$ *times in p. Let* $k\geqslant 2$ *be an integer and let* Σ *be a k-letter alphabet. Then for* $n\geqslant 1$, *the number of partial words of length n over* Σ *that are compatible with instances of the pattern p is no more than* $[x^n]C(x)$, *where*

$$C(x):=\sum_{i_1\geqslant 1}\cdots\sum_{i_m\geqslant 1}\left(\prod_{j=1}^{m}\left(k\left(2^{d_j}-1\right)+1\right)^{i_j}\right)x^{d_1 i_1+\cdots+d_m i_m}.$$

Proof.

For each partial word w compatible with an instance of the pattern, there exists a map ϕ from Δ^{*} to Σ^{*} such that $w\uparrow\varphi(p)$. For $1\leqslant j\leqslant m$, define $i_j=|\varphi(A_j)|$. Now either the first character of $\varphi(A_j)$ corresponds to ⋄ in w for all occurrences of A_j in p, or there exists some $a\in\Sigma$ such that the first character in $\varphi(A_j)$ corresponds to either a or ⋄ in w (and not to ⋄ for every occurrence of A_j in p). In the latter case, since there are d_j occurrences of A_j in p and k possible values of a, there are $k(2^{d_j}-1)$ possibilities for the assignment of the first characters compatible with all occurrences of A_j. Thus adding in the possibility that the first character of $\varphi(A_j)$ corresponds to ⋄ in w for all occurrences of A_j in p, there are $k(2^{d_j}-1)+1$ possible assignments of the first characters compatible with all occurrences of A_j. The same arguments apply to all i_j characters in $\varphi(A_j)$, and their assignments are independent, yielding $(k(2^{d_j}-1)+1)^{i_j}$ total possible assignments for the characters in w corresponding to $\varphi(A_j)$. These assignments are independent for $1\leqslant j\leqslant m$, thus there are

$$\prod_{j=1}^{m}\left(k\left(2^{d_j}-1\right)+1\right)^{i_j}$$

partial words corresponding to ϕ with $i_j=|\varphi(A_j)|$.

Summing over all lengths i_j of images of ϕ for $1\leqslant j\leqslant m$ and noting that the length of the resulting partial words is $i_1d_1+\cdots+i_md_m$, we see that the number of partial words of length n over Σ that are compatible with instances of p is no more than $[x^n]C(x)$. □

Once again we require a technical inequality.

Lemma 5.

Suppose $(k,\lambda)\in\{(2,2.97),(3,3.88)\}$ *and* $m\geqslant 1$ *is an integer. For any integer P and integers* d_j *for* $1\leqslant j\leqslant m$ *such that* $d_j\geqslant 2$ *and* $P=d_1+\cdots+d_m$,

equation(7)

$$\prod_{i=1}^{m}\frac{k(2^{d_i}-1)+1}{\lambda^{d_i}-(k(2^{d_i}-1)+1)}\leqslant\left(\frac{3k+1}{\lambda^2-(3k+1)}\right)^{m-1}\left(\frac{k}{(\frac{\lambda}{2})^{P-2(m-1)}-k}\right).$$

Proof.

The proof is by induction on m. For $m=1$, $d_1=P$ and the left-hand side is

$$\frac{k(2^P-1)+1}{\lambda^P-(k(2^P-1)+1)}<\frac{k(2^P)}{\lambda^P-k(2^P)}=\frac{k}{(\frac{\lambda}{2})^P-k}.$$

Now suppose Eq. (7) holds for m and $d_1+d_2+\cdots+d_{m+1}=P$ with $d_j\geqslant 2$ for $1\leqslant j\leqslant m+1$. Note that $P\geqslant 4$.

Let $P'=P-d_{m+1}$, so that $P'=d_1+\cdots+d_m$. If $d_j=2$ for $1\leqslant j\leqslant m$, equation(8)

$$\prod_{i=1}^{m}\frac{k(2^{d_i}-1)+1}{\lambda^{d_i}-(k(2^{d_i}-1)+1)}=\left(\frac{3k+1}{\lambda^2-(3k+1)}\right)^m.$$

In this case, $d_{m+1}=P-2m$. Note that

$$\frac{k(2^{d_{m+1}}-1)+1}{\lambda^{d_{m+1}}-(k(2^{d_{m+1}}-1)+1)}\leqslant\frac{k}{(\frac{\lambda}{2})^{d_{m+1}}-k}=\frac{k}{(\frac{\lambda}{2})^{P-2(m+1-1)}-k}.$$

Multiplying Eq. (8) by this inequality on both sides yields the desired result for m+1. Otherwise, $P'-2(m-1)>2$. Since $P'=d_1+\cdots+d_m$, the inductive hypothesis implies

equation(9)

$$\prod_{i=1}^{m}\frac{k(2^{d_i}-1)+1}{\lambda^{d_i}-(k(2^{d_i}-1)+1)}\leqslant\left(\frac{3k+1}{\lambda^2-(3k+1)}\right)^{m-1}\left(\frac{k}{(\frac{\lambda}{2})^{P'-2(m-1)}-k}\right).$$

If $d_{m+1}=2$, multiplying both sides by

$$\frac{k(2^{d_{m+1}}-1)+1}{\lambda^{d_{m+1}}-(k(2^{d_{m+1}}-1)+1)}=\frac{3k+1}{\lambda^2-(3k+1)}$$

yields the desired inequality.

In the remaining case, $d_{m+1}>2$. Let $c_1=P'-2(m-1)$ and $c_2=d_{m+1}$. Note that $c_1,c_2\geqslant 3$ and recall$(k,\lambda)\in\{(2,2.97),(3,3.88)\}$. Define $z=\frac{\lambda}{2}$. We first verify the following inequality by induction on $c_1$and c_2:

equation(10)

$$(k-1)(z^{c_1+c_2-2}-k)\leqslant k(3k+1)(z^{c_1-2}-1)(z^{c_2-2}-1).$$

The base cases $c_i\in\{3,4\}$ are easily verified for the specified k and λ . Now assume Eq. (10) holds for c_1 and c_2. Then $k\leqslant z^{c_2}$, thus

$$z^{c_1-2}(z-1)(k)\leqslant z^{c_1-2}(z-1)(z^{c_2}),$$

$$-kz^{c_1-2}-z^{c_1+c_2-1}\leqslant -kz^{c_1-1}-z^{c_1+c_2-2},$$

$$\frac{z^{c_1+c_2-1}-k}{z^{c_1+c_2-2}-k}\leqslant\frac{z^{c_1-1}-1}{z^{c_1-2}-1}.$$

Multiplying Eq. (10) by this inequality on the left and right yields the desired inequality for c_1+1 and c_2. Symmetry indicates the desired inequality also holds for c_1 and c_2+1, completing the induction.

To complete the main induction step, expanding and rearranging Eq. (10) yields

$(k-1)z^{c_1+c_2}+k(3k+1)(z^{c_1}+z^{c_2})\leqslant(3k+1)kz^{c_1+c_2-2}+4k^2z^2$,

$4kz^{c_1+c_2}+k(3k+1)(z^{c_1}+z^{c_2})\leqslant(3k+1)z^{c_1+c_2}+k(3k+1)z^{c_1+c_2-2}+4k^2z^2$,

$$k^2\left(\lambda^2-(3k+1)\right)\left(\left(\frac{\lambda}{2}\right)^{c_1+c_2-2}-k\right)\leqslant k(3k+1)\left(\left(\frac{\lambda}{2}\right)^{c_1}-k\right)\left(\left(\frac{\lambda}{2}\right)^{c_2}-k\right),$$

$$\left(\frac{k}{(\frac{\lambda}{2})^{c_1}-k}\right)\left(\frac{k}{(\frac{\lambda}{2})^{c_2}-k}\right)\leqslant\left(\frac{3k+1}{\lambda^2-(3k+1)}\right)\left(\frac{k}{(\frac{\lambda}{2})^{c_1+c_2-2}-k}\right).$$

Substituting the c_i,

equation(11)

$$\left(\frac{k}{(\frac{\lambda}{2})^{P'-2(m-1)}-k}\right)\left(\frac{k}{(\frac{\lambda}{2})^{d_{m+1}}-k}\right)\leqslant\left(\frac{3k+1}{\lambda^2-(3k+1)}\right)\left(\frac{k}{(\frac{\lambda}{2})^{P'-2m+d_{m+1}}-k}\right).$$

Note that

$$\frac{k(2^{d_{m+1}}-1)+1}{\lambda^{d_{m+1}}-(k(2^{d_{m+1}}-1)+1)}\leqslant\frac{k}{(\frac{\lambda}{2})^{d_{m+1}}-k}.$$

Multiplying Eq. (9) by this inequality on the left and right,

$$\prod_{i=1}^{m+1}\frac{k(2^{d_i}-1)+1}{\lambda^{d_i}-(k(2^{d_i}-1)+1)}\leqslant\left(\frac{3k+1}{\lambda^2-(3k+1)}\right)^{m-1}\left(\frac{k}{(\frac{\lambda}{2})^{P'-2(m-1)}-k}\right)\left(\frac{k}{(\frac{\lambda}{2})^{d_{m+1}}-k}\right).$$

Substituting Eq. (11) to complete the induction,

$$\prod_{i=1}^{m+1}\frac{k(2^{d_i}-1)+1}{\lambda^{d_i}-(k(2^{d_i}-1)+1)}\leqslant\left(\frac{3k+1}{\lambda^2-(3k+1)}\right)^{m}\left(\frac{k}{(\frac{\lambda}{2})^{P-2m}-k}\right).\qquad\square$$

When all variables in the pattern occur at least twice, we obtain the following exponential lower bounds.

Lemma 6.

Let $m\geqslant4$ *be an integer and p be a pattern over an alphabet* $\Delta=\{A_1,\ldots,A_m\}$. *Suppose that for* $1\leqslant i\leqslant m$, A_i *occurs* $d_i\geqslant2$ *times in p.*

1. *If* $|p|\geqslant15(2^{m-3})$, *then for* $n\geqslant0$, *there are at least* $(2.97)^n$ *partial words of length n over a binary alphabet that avoid p.*

2. *If* $|p| \geqslant 2^m$*, then for* $n \geqslant 0$*, there are at least* $(3.88)^n$ *partial words of length n over a ternary alphabet that avoid p.*

Proof.

Let $(k,\lambda) \in \{(2,2.97),(3,3.88)\}$ and Σ be an alphabet of size k. Define S to be the set of all words in $(\Sigma_\diamond)^{\square}$ that are compatible with instances of the pattern p. By Lemma 4, the number of partial words of length n in S is at most $[x^n]C(x)$, where

$$C(x) := \sum_{i_1 \geqslant 1} \cdots \sum_{i_m \geqslant 1} \left(\prod_{j=1}^{m} \left(k\left(2^{d_j}-1\right)+1\right)^{i_j} \right) x^{d_1 i_1 + \cdots + d_m i_m}.$$

Since every variable in p occurs at least twice, $d_i \geqslant 2$ for $1 \leqslant i \leqslant m$. In order to use Theorem 1 on $\Sigma_\diamond$ (which has cardinality $k+1$), define

$$B(x) := \sum_{i \geqslant 0} b_i x^i = \left(1-(k+1)x+C(x)\right)^{-1}.$$

Clearly $b_0=1$ and $b_1=k+1$. We show that $b_n \geqslant \lambda b_{n-1}$ for all $n \geqslant 1$, hence $b_n \geqslant \lambda^n$ for all $n \geqslant 0$. Then all coefficients of B are non-negative, thus Theorem 1 implies there are at least $b_n \geqslant \lambda^n$ words of length$n$ avoiding S. By construction of S, these partial words all avoid p.

We show by induction on n that $b_n \geqslant \lambda b_{n-1}$ for all $n \geqslant 1$. We can easily verify $b_1=(k+1)(1) \geqslant \lambda b_0$. We omit steps very similar to those in the proof of Lemma 3.

To complete the induction, it suffices to show

equation(12)

$$(k+1-\lambda)b_{n-1} - \sum_{i_1 \geqslant 1} \cdots \sum_{i_m \geqslant 1} \left(\prod_{j=1}^{m} \left(k\left(2^{d_j}-1\right)+1\right)^{i_j} \right) b_{n-(d_1 i_1 + \cdots + d_m i_m)} \geqslant 0.$$

Because $b_j \geqslant \lambda b_{j-1}$ for $1 \leqslant j < n$, $b_{n-i} \leqslant b_{n-1}/\lambda^{i-1}$ for $1 \leqslant i \leqslant n$. Therefore,

$$\begin{aligned} &\sum_{i_1 \geqslant 1} \cdots \sum_{i_m \geqslant 1} \left(\prod_{j=1}^{m} \left(k\left(2^{d_j}-1\right)+1\right)^{i_j} \right) b_{n-(d_1 i_1 + \cdots + d_m i_m)} \\ &\quad \leqslant \sum_{i_1 \geqslant 1} \cdots \sum_{i_m \geqslant 1} \left(\prod_{j=1}^{m} \left(k\left(2^{d_j}-1\right)+1\right)^{i_j} \right) \frac{\lambda b_{n-1}}{\lambda^{d_1 i_1 + \cdots + d_m i_m}} \\ &\quad = \lambda b_{n-1} \sum_{i_1 \geqslant 1} \left(\frac{k(2^{d_1}-1)+1}{\lambda^{d_1}} \right)^{i_1} \cdots \sum_{i_m \geqslant 1} \left(\frac{k(2^{d_m}-1)+1}{\lambda^{d_m}} \right)^{i_m}. \end{aligned}$$

Since $d_j \geqslant 2$ for $1 \leqslant j \leqslant m$, $k \leqslant 3$ and $\lambda > 2\sqrt{k}$,

$$\frac{k(2^{d_j}-1)+1}{\lambda^{d_j}} < \frac{k}{(\frac{\lambda}{2})^{d_j}} \leqslant \frac{k}{(\frac{\lambda}{2})^2} < 1,$$

thus all the geometric series converge. Computing the result, for $1\leqslant j\leqslant m$,

$$\sum_{i_j\geqslant 1}\left(\frac{k(2^{d_j}-1)+1}{\lambda^{d_j}}\right)^{i_j} = \frac{k(2^{d_j}-1)+1}{\lambda^{d_j}-(k(2^{d_j}-1)+1)}.$$

Thus

$$\sum_{i_1\geqslant 1}\cdots\sum_{i_m\geqslant 1}\left(\prod_{j=1}^{m}\left(k\left(2^{d_j}-1\right)+1\right)^{i_j}\right)b_{n-(d_1i_1+\cdots+d_mi_m)} \leqslant \lambda b_{n-1}\prod_{j=1}^{m}\frac{k(2^{d_j}-1)+1}{\lambda^{d_j}-(k(2^{d_j}-1)+1)}.$$

Applying Lemma 5 to $P=|p|$,

$$\sum_{i_1\geqslant 1}\cdots\sum_{i_m\geqslant 1}\left(\prod_{j=1}^{m}\left(k\left(2^{d_j}-1\right)+1\right)^{i_j}\right)b_{n-(d_1i_1+\cdots+d_mi_m)} \leqslant \lambda b_{n-1}\left(\frac{3k+1}{\lambda^2-(3k+1)}\right)^{m-1}\left(\frac{k}{(\frac{\lambda}{2})^{|p|-2(m-1)}-k}\right).$$

Referencing Eq. (12), it thus suffices to show

equation(13)

$$(k+1-\lambda) \geqslant \lambda\left(\frac{3k+1}{\lambda^2-(3k+1)}\right)^{m-1}\left(\frac{k}{(\frac{\lambda}{2})^{|p|-2(m-1)}-k}\right).$$

To show Statement 1, let $(k,\lambda)=(2,2.97)$ and recall we restricted $m\geqslant 4$ and $|p|\geqslant 15(2^{m-3})$. Eq. (13)is easily verified for $m=4$ and $|p|=30$. Clearly if Eq. (13) holds for $|p|$, it will hold for p' with $|p'|>|p|$. Thus it suffices to check the general case $m'>4$ and $|p'|=15(2^{m'-3})$. We define

$$c := \left(\frac{3k+1}{\lambda^2-(3k+1)}\right)^{m'-m}\left(\frac{(\frac{\lambda}{2})^{|p|-2(m-1)}-k}{(\frac{\lambda}{2})^{|p'|-2(m'-1)}-k}\right)$$
$$\leqslant (3.85)^{m'-m}\left(\frac{1}{(\frac{\lambda}{2})^{|p'|-2(m'-1)-(|p|-2(m-1))}}\right) \leqslant (3.85)^{m'-m}\left(\frac{1}{(\frac{\lambda}{2})^{2^{m'-1}}}\right) < 1.$$

Thus we conclude

$$(k+1-\lambda) \geqslant c\lambda\left(\frac{3k+1}{\lambda^2-(3k+1)}\right)^{m-1}\left(\frac{k}{(\frac{\lambda}{2})^{|p|-2(m-1)}-k}\right) = \lambda\left(\frac{3k+1}{\lambda^2-(3k+1)}\right)^{m'-1}\left(\frac{k}{(\frac{\lambda}{2})^{|p'|-2(m'-1)}-k}\right).$$

Verification of Eq. (13) for Statement 2 is similar, so it is omitted. □

Thus for certain patterns, there exist λ^n partial words of length *n* that avoid the pattern, for some λ. It is not immediately clear that this is enough to prove the patterns are avoidable in partial words. The next lemma asserts this count is so large that it must include partial words with arbitrarily many holes, thus the patterns are 2-avoidable or 3-avoidable in partial words.

Lemma 7.

Suppose $k \geqslant 2$ *is an integer,* $k<\lambda<k+1$*,* Σ *is an alphabet of size k, and S is a set of partial words over* Σ *with at least* λ^n *words of length n for each* $n>0$*. For all integers* $h \geqslant 0$*, S contains a partial word with at least h holes.*

Proof.

To count length *n* partial words with exactly $h \leqslant n$ holes, note that there are $\binom{n}{h}$ choices for hole positions, then k^{n-h} choices for the remaining letters in the word, so $\binom{n}{h}k^{n-h}$ total partial words of length *n* with *h* holes. Suppose for the sake of contradiction that there exists an integer $\bar{h}$ such that *S* contains no partial words with more than $\bar{h}$ holes. Then the number of partial words of length *n* in *S* cannot exceed the number of partial words of length *n* with no more than $\bar{h}$ holes, so for any length $n \geqslant \bar{h}$,

$$T(n) := \sum_{h=0}^{\bar{h}} k^{n-h}\binom{n}{h} \geqslant \lambda^n.$$

Rewriting in terms of factorials, for any $h \leqslant \bar{h}$,

$$\frac{\binom{n+1}{h}}{\binom{n}{h}} = \frac{\frac{(n+1)!}{(n+1-h)!h!}}{\frac{n!}{(n-h)!h!}} = \frac{(n+1)}{(n+1-h)} \leqslant \frac{(n+1)}{(n+1-\bar{h})}.$$

We estimate

$$\frac{T(n+1)}{T(n)} = \frac{\sum_{h=0}^{\bar{h}} k^{n+1-h}\binom{n+1}{h}}{\sum_{h=0}^{\bar{h}} k^{n-h}\binom{n}{h}} \leqslant k\left(\max_{h \leqslant \bar{h}} \frac{\binom{n+1}{h}}{\binom{n}{h}}\right) \leqslant k\frac{(n+1)}{(n+1-\bar{h})}.$$

The term on the right tends to *k* from above as $n \to \infty$, thus we may choose $N > \bar{h}$ so that for $n \geqslant N$,

$$\frac{T(n+1)}{T(n)} < k + \frac{\lambda - k}{2} = \frac{k+\lambda}{2}.$$

Then for $n \geqslant N$,

$$\lambda^n \leqslant T(n) < T(N)\left(\frac{k+\lambda}{2}\right)^{n-N}.$$

Since T(N) is constant and $k<\lambda$, $\frac{k+\lambda}{2} < \lambda$, which is a contradiction for large enough *n*. □

Unfortunately, the pattern $A^2BA^2CA^2$ of length $8=2^3$ is unavoidable in partial words, thus to obtain the2^m bound for avoidability as in the full word case, we require information about quaternary patterns of length$16=2^4$. Fortunately, for certain patterns, constructions can be made from full words avoiding a pattern to partial words avoiding a pattern that provide upper bounds on avoidability indices. We obtain the following bounds.

Theorem 3.

Let p be a pattern with m distinct variables.

1. *If* $m\geqslant 3$*and* $|p|\geqslant 15(2^{m-3})$, *then p is 2-avoidable in partial words.*

2. *If* $m\geqslant 3$*and* $|p|\geqslant 5(2^{m-2})$, *then p is 3-avoidable in partial words.*

3. *If* $m\geqslant 4$*and* $|p|\geqslant 2^m$, *then p is 4-avoidable in partial words.*

Proof.

For Statement 1, we prove by induction on m that if p is 2-unavoidable, $|p|<15(2^{m-3})$. The base case of ternary patterns (m=3) is handled by a list of over 800 patterns in the appendix of [2]. The maximum length 2-unavoidable ternary pattern in partial words is $A^2BA^2CA^2BA^2$, length $11<15=15(2^{3-3})$.

Now suppose the result holds for m and let p be a pattern with $m+1\geqslant 4$ distinct variables. If every variable in p is repeated at least twice, Statement 1 of Lemma 6 implies there exists a set S of partial words with at least $(2.97)^n$ binary words of length n that avoid p for each $n\geqslant 0$. Applying Lemma 7 to S , we find that for each $h\geqslant 0$, there exists a partial word with at least *h* holes that avoids *p*. Thus *p* is 2-avoidable. If *p* has a variable that occurs exactly once, we reason as in the proof of Theorem 2 to complete the induction.

For Statement 2, we prove by induction on m that if p is 3-unavoidable, $|p|<5(2^{m-2})$. For m=3, all patterns of length $10=5(2^{3-2})$ are shown to be 3-avoidable in [2]. For $m\geqslant 4$, Statement 2 of Lemma 6 andLemma 7 imply that every pattern of length at least 2^m in which each variable appears at least twice is 3-avoidable. If *p* has a variable that occurs exactly once, we reason as in the proof of Theorem 2 to complete the induction.

For Statement 3, we show by induction on m that if p is 4-unavoidable, $|p|<2^m$. We first establish the base case m=4 by showing that every pattern p of length $16=2^4$ is 4-avoidable. Using the data in [2], the ternary patterns of length at least 7 which have avoidability index greater than 4 are AABAACAA of length 8
and AABCABA, ABACAAB, ABACBAA, ABBCBAB, ⋯ of length 7 (up to reversal and renaming of variables).

Consider any p with $|p|=16$. If every variable in p occurs at least twice, Statement 2 of Lemma 6 implies there exists a set S with at least $(3.88)^n$ ternary partial words of length n that avoid p for each $n\geqslant 0$. Applying Lemma 7 to S, we find that for each $h\geqslant 0$, there exists a ternary partial word with at least h holes that avoids p. Thus p is 3-avoidable. Otherwise, p contains a variable α that occurs exactly once and $p=p_1\alpha p_2$ for patterns p_1 and p_2 with at most 3 distinct variables. Note that $|p_1|+|p_2|=15$. If p_1 has length at least 9, then p_1 is 4-avoidable, hence p is 4-avoidable by divisibility (likewise for p_2).

Thus the only remaining cases are when $|p_1|=8$ and $|p_2|=7$ or vice versa. Suppose $|p_1|=8$ and $|p_2|=7$ (the other case is similar). If p_1 or p_2 is not in the list of ternary patterns above, it is 4-avoidable, hence p is 4-avoidable. Otherwise $p_1=A^2BA^2CA^2$ up to a renaming of the variables. Note that p_1 contains a factor of the form A^2BA, which fits the form of [2, Theorem 6(2)] for $q_1=B$. All of the possible values of p_2 are on three variables, so they must contain B. Thus setting $q_2=B$, [2, Theorem 6(2)] implies p is 4-avoidable.

For $m\geqslant 5$, Lemma 6 and Lemma 7 imply that every pattern with length at least 2^m in which each variable appears at least twice is 3-avoidable. If p has a variable that occurs exactly once, we reason as in the proof of Theorem 2 to complete the induction. □

Note that Theorem 3(3) gives a strict bound for 4-avoidability in partial words, using one of the sequences of patterns given for full words in Section 2.

6. Concluding remarks and conjectures

Overall, the power series method is a useful way to show existence of infinitely many words avoiding patterns in full words and partial words. It is mainly helpful to obtain upper bounds as derived here, since it utilizes the frequencies of each variable in the pattern and not their placement relative to one another. Only patterns where each variable occurs at least twice can be investigated in this way, but induction arguments as in Theorem 2 then imply bounds for all patterns. For patterns with a variable that appears exactly once, the counts used in Lemma 1 and Lemma 4 grow too quickly, thus the power series method is not applicable.

It would be nice to attain strict bounds for 2-avoidability and 3-avoidability in partial words. Statement 1 of the following conjecture appears in [2], and we add Statement 2.

Conjecture 1.

Let p be a pattern with m distinct variables.

1. *If* $|p|\geqslant 3(2^{m-1})$, *then p is 2-avoidable in partial words.*

2. *If* $m\geqslant 4$ *and* $|p|\geqslant 2^m$, *then p is 3-avoidable in partial words.*

Both bounds would then be strict, using the same sequences of patterns given for full words in Section 2.

To show Statement 1 using the power series method, we require either an improvement of the bound $15(2^{m-3})$ to $3(2^{m-1})$ in Statement 1 of Lemma 6 or some additional data about avoidability indices of patterns over 4 variables. It may be possible to improve the count used in Lemma 4 to improve this bound. To show Statement 2 using the power series method, we require additional data about avoidability indices of patterns over 4 variables. Unfortunately, finding avoidability indices using HD0L systems as in [2] is likely infeasible for patterns over 4 variables. Perhaps some constructions can be made from words avoiding long enough 2-avoidable or 3-avoidable patterns in full words to prove there exist infinitely many partial words that avoid the pattern over 2 or 3 letters.

Finally, it may be possible to make better approximations than Theorem 1 and Lemma 1 based on the Goulden–Jackson method for avoiding a finite number of words [7]. The method works better when the growth rate of words avoiding a k -avoidable pattern is close to k , whereas it is known that for the pattern AABBCABBA, where k=2, the growth rate is close to 1. There is no hope for the pattern ABWACXBCYBAZCA, where k=4, since only polynomially many words over 4 letters avoid it (here the growth rate is 1). Perhaps the method could handle the cases where each variable of the pattern occurs at least twice, but even the case of the pattern AA , where k=3, seems to be challenging with a 1.31 growth rate.

Note that part of this paper was presented at DLT 2013 [4]. A preliminary version of this paper was submitted to DLT 2013 on January 2, 2013. Some referees made us aware that Theorem 2 has also been found, completely independently and almost simultaneously, by Pascal Ochem and Alexandre Pinlou [9]. Their proof of Statement 1 uses Bell and Goh's method, while their proof of Statement 2 uses the entropy compression method.

References

[1] J. Bell, T.L. Goh, Exponential lower bounds for the number of words of uniform length avoiding a pattern, Inf. Comput. 205 (2007) 1295–1306.

[2] F. Blanchet-Sadri, A. Lohr, S. Scott, Computing the partial word avoidability indices of ternary patterns, in: S. Arumugam, B. Smyth (Eds.), IWOCA

2012, 23rd International Workshop on Combinatorial Algorithms, in: Lect. Notes Comput. Sci., vol. 7643, Springer-Verlag, Berlin, Heidelberg, 2012,

pp. 206–218 (expanded version to appear in Journal of Discrete Algorithms).

[3] F. Blanchet-Sadri, R. Merca ¸s, S. Simmons, E. Weissenstein, Avoidable binary patterns in partial words, Acta Inform. 48 (2011) 25–41.

[4] F. Blanchet-Sadri, B. Woodhouse, Strict bounds for pattern avoidance, in: M.-P. Béal, O. Carton (Eds.), DLT 2013, 17th International Conference on

Developments in Language Theory, in: Lect. Notes Comput. Sci., vol. 7907, Springer-Verlag, Berlin, Heidelberg, 2013, pp. 106–117.

[5] J. Cassaigne, Motifs évitables et régularités dans les mots, Ph.D. thesis, Paris VI, 1994.

[6] M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, Cambridge, 2002.

[7] J. Noonan, D. Zeilberger, The Goulden–Jackson cluster method: Extensions, applications, and implementations, Int. J. Differ. Equ. Appl. 5 (1999) 355–377.

[8] P. Ochem, A generator of morphisms for infinite words, RAIRO Theor. Inform. Appl. 40 (2006) 427–441.

[9] P. Ochem, A. Pinlou, Application of entropy compression in pattern avoidance, arXiv:1301.1873, January 9, 2013.

[10] N. Rampersad, Further applications of a power series method for pattern avoidance, Electron. J. Comb. 18 (2011) P134.

[11] P. Roth, Every binary pattern of length six is avoidable on the two-letter alphabet, Acta Inform. 29 (1992) 95–106.

[12] L. Rowen, Ring Theory, vol. II, Pure Appl. Math., vol. 128, Academic Press, Boston, 1988.