NEW PROBLEMS OF PATTERN AVOIDANCE

JOHN LOFTUS, JEFFREY SHALLIT, AND MING-WEI WANG

Department of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

E-mail: {jaloftus,shallit,m2wang}@uwaterloo.ca

Let $\Sigma_k := \{0, 1, \ldots, k-1\}$ for an integer $k \geq 2$. Define $\sigma(a) = (a+1) \mod k$ for $a \in \Sigma_k$. In this paper we consider several new pattern avoidance problems, of which the following is a typical example: what is the smallest $k$ for which one can simultaneously avoid the patterns $xx$ and $\sigma(x)$ over $\Sigma_k$?

1 Introduction and definitions

Pattern avoidance problems have long been studied in formal language theory, and have interesting applications to group theory, universal algebra, and other areas. For example, Axel Thue\cite{1,2} constructed an infinite squarefree word over $\{0, 1, 2\}$; i.e., a word that contains no subword of the form $xx$, where $x$ is a nonempty word.

Eventually, generalizations of Thue’s problem were considered. Erdős, for example, suggested trying to find infinite words containing no subword of the form $xy$, where $y$ is a permutation of the letters of $x$. Such words are now sometimes called “abelian squarefree”\cite{3}. For other papers on pattern avoidance, the reader can fruitfully consult, for example,\cite{1,5,4}.

In this paper, we consider some new generalizations of Thue’s problem. We start with some notation. Let $\Sigma, \Gamma$ be finite alphabets. A morphism is a map $h : \Gamma^* \rightarrow \Sigma^*$ such that $h(xy) = h(x)h(y)$ for all $x, y \in \Gamma^*$. We let $\Sigma^\omega$ denote the set of all one-sided infinite words over $\Sigma$, and we let $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$. If $x \in \Sigma^+$, then by $x^\omega$ we mean the one-sided infinite word $xxx \cdots$.

If there exist words $x, y \in \Sigma^\omega$, $w, z \in \Sigma^\infty$ such that $w = xyz$, then we say $y$ is a finite subword of $w$. Suppose we are given a finite or infinite subset $P \subseteq \Sigma^\omega$. Then we say a word $w \in \Sigma^\infty$ avoids $P$ if we cannot write $w = xyz$ such that $y \in P$. We say $P$ is avoidable over $\Sigma$ if it is possible to construct an infinite word $w \in \Sigma^\infty$ which avoids $P$.

Sometimes we employ a common abuse of notation. For example, instead of saying that the infinite word $w$ avoids $\{xx : x \in \Sigma^+\}$, we will instead simply say that $w$ avoids the pattern $xx$. When we use this formulation, we always assume the strings in the pattern are nonempty.

We define $\Sigma_k = \{0, 1, 2, \ldots, k-1\}$ for some integer $k \geq 2$, and we define the morphism $\sigma_k(a) = (a+1) \mod k$. If the subscript $k$ is clear from the context, we omit it. In this paper, we consider avoiding patterns of the form $x\sigma^i(x)$.

We use two notational conventions that may be somewhat confusing. First, we think of the elements of $\Sigma_k$ as residue class representatives so that, for example, $-1$ and $2$ denote the same element of $\Sigma_3$. Second, since we allow negative numbers in words, we sometimes use the notation $\{a_1, a_2, a_3, \ldots, a_n\}$ to denote the word $a_1a_2a_3\cdots a_n$. Thus, for example, $012$ and $(0, -2, -1)$ denote the same element of
\( \Sigma_3 \).

Some of the infinite words we construct arise from iterated morphisms. Call a morphism \( h : \Gamma^* \to \Sigma^* \) non-erasing if \( h(a) \neq \epsilon \) for all \( a \in \Gamma \). Let \( h : \Sigma^* \to \Sigma^* \) be a non-erasing morphism, and let \( a \in \Sigma \) be a letter such that \( h(a) = ax \). Then we define \( h^n(a) = a x h(x) h^2(x) h^3(x) \cdots \). Note that \( h^n(a) \) is a fixed point of the map \( h \) extended to \( \Sigma^* \).

2 Avoiding \( x\sigma(x) \)

It is clear that over \( \Sigma_2 = \{0, 1\} \), there are only two infinite words avoiding the pattern \( x\sigma(x) \), namely \( 0^\omega \) and \( 1^\omega \). However, we have the following result:

**Theorem 1** Over \( \Sigma_k \) for \( k \geq 3 \), there are uncountably many infinite words avoiding \( x\sigma(x) \).

**Proof.** Define \( a_1 = 1 \), and set \( a_{i+1} = a_i + 1 \) or \( a_i + 2 \), according to choice. Then

\[
   w = \prod_{i \geq 1} ((-i) \mod 3)^{a_i} = 2^{a_1} 1^{a_2} 0^{a_3} 2^{a_4} 1^{a_5} 0^{a_6} \cdots
\]

avoids the pattern \( x\sigma(x) \), and there are uncountably many such words. \( \blacksquare \)

3 Avoiding \( xx, x\sigma(x), \ldots, x\sigma^j(x) \) simultaneously

The following theorem constitutes our main result. It characterizes, for each integer \( j \geq 0 \), the smallest integer \( k \) for which we can avoid the \( j + 1 \) patterns \( xx, x\sigma(x), \ldots, x\sigma^j(x) \) simultaneously over \( \Sigma_k = \{0, 1, \ldots, k-1\} \).

**Theorem 2**

(a) One can avoid the pattern \( xx \) over \( \Sigma_3 \), and \( 3 \) is best possible.

(b) One can avoid the patterns \( xx \) and \( x\sigma(x) \) simultaneously over \( \Sigma_5 \), and \( 5 \) is best possible.

(c) One can avoid the patterns \( xx, x\sigma(x), x\sigma^2(x) \) simultaneously over \( \Sigma_5 \), and \( 5 \) is best possible.

(d) One can avoid the patterns \( xx, x\sigma(x), x\sigma^2(x), x\sigma^3(x) \) simultaneously over \( \Sigma_6 \), and \( 6 \) is best possible.

(e) For \( j \geq 4 \), one can avoid the \( j + 1 \) patterns \( xx, x\sigma(x), \ldots, x\sigma^j(x) \) simultaneously over \( \Sigma_{j+4} \), and \( j + 4 \) is best possible.

**Remark.** Our proofs of these facts are of two different types. First, in order to show that it is possible to avoid a certain set of patterns over \( \Sigma_k \), we explicitly construct an infinite word over \( \Sigma_k \) having the desired property. Second, to show
that \( k \) is optimal for a certain set of patterns, we use a classical breadth-first tree traversal technique, as follows:

Suppose we wish to avoid a given set of words \( P \) over \( \Sigma_k \). We maintain a queue, \( Q \), and initialize it with the empty word \( \epsilon \). If the queue is empty, we are done. Otherwise, we take the next element \( w \) from the queue, and form \( k \) new words by appending \( 0, 1, \ldots, k - 1 \) to \( w \). For each new word \( wa \), we check to see whether some suffix of \( wa \) occurs in \( P \). If it does, we discard it; otherwise we add it to the queue.

If this algorithm terminates, we have proved that it is not possible to avoid \( P \) over \( \Sigma_k \). The resulting proof may be represented in the form of a tree, with the leaves representing minimal length prefixes that contain an occurrence of one of the patterns as a suffix.

In the particular case of the patterns we discuss in this section, two additional efficiencies are possible. First, since a word \( w \) simultaneously avoids the patterns \( xx, x\sigma(x), \ldots, x\sigma^j(x) \) iff \( \sigma(w) \) does, we may without loss of generality consider only the words that begin with the letter 0. Second, if the last letter was \( a \), then the next letter must be contained in the set \{\( a + j + 1, \ldots, a + k - 1 \}\}, for otherwise our word would contain a length-2 subword of the form \( x\sigma^i(x) \) for \( 0 \leq i \leq j \). This observation significantly cuts down on the branching factor of the trees we generate.

**Proof of Theorem 2.** Let us start with assertion (a). As already noted, a classical result due to Thue\(^1\)\(^2\) shows that one can avoid the pattern \( xx \) over \( \Sigma_3 = \{0, 1, 2\} \). Furthermore, it is an old and easy observation that any word of length \( \geq 4 \) over \( \Sigma_2 = \{0, 1\} \) contains an occurrence of the pattern \( xx \). More generally, we have

**Proposition 3** Let \( k \geq 2 \) be an integer, and let \( r \) be an integer with \( 1 \leq r < k \). Then any word of length \( \geq 4 \) over \( \Sigma_k \) contains an occurrence of the pattern \( xx, x\sigma(x) \) for some \( a \neq r \) (mod \( k \)).

**Proof.** We use the tree traversal algorithm. Assume the first letter is 0; then if the next letter is \( a \neq r \), we are done. Hence assume the next letter is \( r \). Then, by a similar argument, the next letter must be \( 2r \) and the next \( 3r \). However, the word \( (0, r, 2r, 3r) \) contains the pattern \( x\sigma^2(x) \) for \( x = (0, r) \). Since \( r \neq 0 \), we have \( 2r \neq r \) (mod \( k \)).

Now let us prove assertion (b) of Theorem 2. By Proposition 3 with \( r = 2 \), one cannot avoid the patterns \( xx \) and \( x\sigma(x) \) simultaneously over \( \Sigma_3 \). We also have

**Proposition 4** Every word of length \( \geq 24 \) over \( \Sigma_4 \) contains an occurrence of either \( xx \) or \( x\sigma(x) \).

**Proof.** We use the tree traversal algorithm. The resulting tree has depth 24 and contains 233 leaves. Figure 1 below lists these leaves in breadth-first order.
Theorem 5. It is possible to simultaneously avoid the patterns \( xx \) and \( x^2(x) \) over \( \Sigma_k \). How-

ever, we cannot avoid the patterns \( xx \) and \( x^2(x) \) simultaneously over \( \Sigma_k \). This follow from Theorem 5 below.

Next, let us prove assertion (c). As we have seen in Proposition 3 above, every word of length \( \geq 4 \) over \( \Sigma_k \) contains an occurrence of one of the patterns \( xx, x^2(x) \). We now show or \( x^2(x) \). We now show

Before starting the proof, we introduce some notation. If \( w = a_1a_2a_3\ldots \) is a word over \( \Sigma_k \), then

\[
\Delta(w) = (\sigma_1 - \sigma_1, \sigma_2 - \sigma_2, \ldots)
\]

where the sums are, of course, taken mod. \( k \). Note that \( \Delta(w) \) is \( \sigma \) and if \( \Delta(w) = 0 \), then \( \Delta(w) = k \). The following lemma relates occurrences of patterns of the form \( xx \) or \( x^2(x) \) in \( w \) to

\[
S(w) = (a_1 - a_1, a_2 - a_2, \ldots)
\]
Lemma 6 Let \( w \in \Sigma_k^\infty \) and let \( a \in \Sigma_k \). Then \( w \) avoids the pattern \( x \sigma^\delta(x) \) iff \\
\( \Delta(w) \) avoids \( \{ ycy : y \in \Sigma_k^\infty, ~ c \in \Sigma_k, \text{ and } s_k(yc) = a \} \).

Proof. Suppose \( w \) contains an occurrence of the pattern \( x \sigma^\delta(x) \). Write \( x = b_1 b_2 \cdots b_i \). Then \\
\[ w = w'b_1 b_2 \cdots b_i \sigma^\delta(b_1) \cdots \sigma^\delta(b_i) \cdots \] \\
Thus \\
\[ \Delta(w) = \Delta(w'b_1), \ (b_2 - b_1, \ldots, b_i - b_{i-1}, \sigma^\delta(b_1) - b_i, b_2 - b_1, \ldots, b_i - b_{i-1}) \], \\
and hence contains \( ycy \) with \\
\[ y = (b_2 - b_1, \ldots, b_i - b_{i-1}), \quad c = \sigma^\delta(b_1) - b_i. \] \\
Also \\
\[ s_k(yc) = (b_2 - b_1) + \cdots + (b_i - b_{i-1}) + \sigma^\delta(b_1) - b_i \] \\
\[ = (b_i - b_1) + (a + b_1 - b_i) \] \\
\[ = a. \] \\
Now suppose \( \Delta(w) \) contains a subword \( ycy \) with \( y \in \Sigma_k^\infty, ~ c \in \Sigma_k, \) and \( s_k(yc) = a \). Then \( \Delta(w) = x ycyz \) for some \( x = b_1 b_2 \cdots b_j \) and \( y = d_1 d_2 \cdots d_i \). Then \\
\[ \Delta(w) = b_1 b_2 \cdots b_j d_1 d_2 \cdots d_i c d_1 d_2 \cdots d_i \cdots \]. \\
Then if \( e \) is the first letter of \( w \), we have \\
\[ w = (e, e + b_1, e + b_1 + b_2, \ldots, e + b_1 + b_2 + \cdots + b_j, e + f + d_1, e + f + d_1 + d_2, \ldots, \] \\
\[ e + f + d_1 + d_2 + \cdots + d_i, e + f + g + c, e + f + g + c + d_1, \] \\
\[ e + f + g + c + d_1 + d_2, \ldots, e + f + g + c + d_1 + d_2 + \cdots + d_i, \ldots \] \\
where \( f := b_1 + b_2 + \cdots + b_j \) and \( g := d_1 + d_2 + \cdots + d_i \). It follows that \( w \) contains an occurrence of \( x \sigma^\delta(x) \), where \( x = (e + f, e + f + d_1, \ldots, e + f + d_1 + d_2 + \cdots + d_i) \) and \( a = g + c \). But \( g + c = s_k(d_1 d_2 \cdots d_i c) = s_k(yc) \). \( \blacksquare \) \\

Now to prove Theorem 5, it suffices to construct an infinite word \( v \) where \( v \) avoids \\
\[ P_2 := \{ ycy : y \in \Sigma_5^\infty, \ c \in \Sigma_5, \text{ and } s_5(yc) \in \{0, 1, 2\} \}. \]

For then we could set \( w = S(v) \), and by Lemma 6, \( w \) avoids the patterns \( xx, x \sigma(x), \) and \( x \sigma^2(x) \) over \( \Sigma_5 \). We construct such a \( v \) using the following theorem.

Theorem 7 Let \( h \) be the morphism over \( \{3, 4\} \) defined by \( h(4) = 4433 \) and \( h(3) = 44433 \). Let \( w \) be a finite word. If \( w \) avoids \( P_2 \), then \( h(w) \) avoids \( P_2 \).
Proof. We prove the contrapositive.
Suppose $h(w)$ contains an occurrence of the pattern $ycy$ with $y \in \Sigma_5^*$, $c \in \Sigma_5$, and $s_5(ye) \in \{0, 1, 2\}$. Write $h(w) = z_1 yczy_2$. Without loss of generality, we may assume that $|z_1|$ is as small as possible, or, in other words, that the occurrence of $ycy$ we are dealing with lies as far to the left as possible within $h(w)$.

Also note that $s_5(i) = s_5(h(i))$ for $i \in \{3, 4\}$, and so it follows that $s_5(w) = s_5(h(w))$ for all finite strings $w \in \{3, 4\}^*$.

We claim that if $ycy$ is a subword of $h(w)$ for some $w$ such that $y, c$ obey the given conditions, then $|y| \geq 5$. Table 1 below suffices to prove this.

The explanation of the table is as follows. We examine all possible subwords $yc$ of length $\leq 5$ that occur in $\{443, 4443\}^*$. For each such subword, it suffices to show that either $s_5(yc) \notin \{0, 1, 2\}$, or $ycy$ cannot occur as a subword of $h(w)$ for any $w \in \{3, 4\}^*$. For this last check, it suffices to observe that if $ycy$ contains any of the subwords 434, 343, 333, or 4444, then it cannot occur as a subword of $h(w)$.

| $|y|$ | $ye$ | $s_5(ye)$ | $ycy$ contains forbidden subword | if so, which one |
|-----|------|----------|----------------------------------|------------------|
| 0   | 3    | 3        | 3                               | no               |
|     | 4    | 4        | 4                               | no               |
| 1   | 33   | 1        | 333                             | yes              | 333 |
|     | 34   | 2        | 343                             | yes              | 343 |
|     | 43   | 2        | 434                             | yes              | 434 |
|     | 44   | 3        | 444                             | no               |     |
| 2   | 334  | 0        | 33433                          | yes              | 334 |
|     | 344  | 1        | 34434                          | yes              | 344 |
|     | 433  | 0        | 43343                          | yes              | 343 |
|     | 443  | 1        | 44344                          | yes              | 443 |
|     | 444  | 2        | 44444                          | yes              | 4444 |
| 3   | 3344  | 4      | 3344334                        | no               |     |
|     | 3443  | 4      | 3443344                        | no               |     |
|     | 3444  | 0      | 3444444                       | yes              | 434 |
|     | 4334  | 4      | 4334433                       | no               |     |
|     | 4433  | 4      | 4433443                       | no               |     |
|     | 4434  | 0      | 4434444                       | yes              | 434 |
| 4   | 33443  | 2     | 334433344                      | yes              | 334 |
|     | 33444  | 3     | 334433434                     | no               |     |
|     | 34433  | 2     | 344333434                     | yes              | 333 |
|     | 34443  | 3     | 344433444                     | no               |     |
|     | 43344  | 3     | 433443443                     | no               |     |
|     | 44334  | 3     | 443344433                     | no               |     |
|     | 44433  | 3     | 444334443                     | no               |     |

It follows that $|y| \geq 5$. There are now several cases to consider.
Case 1: \( y \) starts with 33. Then \( ycy = 33 \cdots c 33 \cdots \). Since \( h(w) \in \{4433, 44433\}^* \), we must have \( c = 4 \). Also, \( y \) must end with 4, and furthermore the letter immediately preceding the occurrence of \( ycy \) in \( h(w) \) must be 4. We can therefore write \( y = 33t4 \) for some string \( t \), and observe that \( 4 \ 33t4 \ 4 \ 33t4 = 4y4y \) is a subword of \( h(w) \). Now let \( y' = 433t \), and note that \( y'y' \) is a subword of \( h(w) \). But \( s_5(y'4) = s_5(433t4) = s_5(33t44) = s_5(y4) \in \{0, 1, 2\} \), so \( y'4y' \in P_2 \), contradicting our assumption that \( ycy \) was the leftmost such occurrence in \( h(w) \).

Case 2: \( y \) starts with 34. Then \( ycy = 34 \cdots c 34 \cdots \), so \( c = 3 \). Thus \( ycy = 34 \cdots 3 \ 34 \cdots \), so \( y \) must end in 4, and further the letter immediately preceding the occurrence of \( ycy \) in \( h(w) \) must be 3. We can therefore write \( y = 34t3 \) for some string \( t \), and observe that \( 3 \ 34t3 \ 3 \ 34t3 = 3y3y \) is a subword of \( h(w) \). Now let \( y' = 33t3 \) and note that \( y'y'y' \) is a subword of \( h(w) \). But \( s_5(y'3) = s_5(33t34) = s_5(34t33) = s_5(y3) \in \{0, 1, 2\} \), so \( y'y'y' \in P_2 \), contradicting our assumption that \( ycy \) was the leftmost such occurrence in \( h(w) \).

Case 3: \( y \) starts with 43. Then \( ycy = 43 \cdots c 43 \cdots \), so \( c = 4 \), and further the letter immediately preceding the occurrence of \( ycy \) in \( h(w) \) must be 4. Thus \( y = 43t \). Write \( t = t'b \), where \( |t'| = 1 \). Then \( y = 43t'b \). Then \( 4yxy = 4 \ 43t'b \ 4 \ 43t'b \) is a subword of \( h(w) \). Let \( y' = 433t' \). Then \( y'y'y' \) is a subword of \( h(w) \), and \( s_5(y'b) = s_5(433t'b) = s_5(43t'4) = s_5(y4) \in \{0, 1, 2\} \), so \( y'y'y' \in P_2 \), contradicting our assumption that \( ycy \) was the leftmost such occurrence in \( h(w) \).

Case 4: \( y \) starts with 444. Then \( ycy = 444 \cdots c 444 \cdots \), so \( c = 3 \), and further, \( y \) ends with 3. Since \( |y| \geq 5 \), we can write \( y = 444t3 \) for some string \( t \). It follows that \( y3y3 = 444t3 \ 3 \ 444t3 \ 3 \) is a subword of \( h(w) \). Hence there exists a string \( u \) such that \( h(3u) = y3 \), and \( 3u3u \) is a subword of \( w \). We have \( s_5(u3) = s_5(3u) = s_5(y3) \in \{0, 1, 2\} \), so \( u3u \) is an occurrence of a string of \( P_2 \) in \( w \), as desired.

Case 5: \( y \) starts with 445. There are two subcases to consider:

Case 5a: \( c = 3 \). Then the last two characters of \( y \) must be 43. We have \( ycy = 443 \cdots 433 \ 443 \cdots 43 \). Then \( y3y3 \) is a subword of \( h(w) \), and there must exist \( u \) such that \( h(4u) = y3 \) and \( u4u \) is a subword of \( w \). Then \( s_5(u4) = s_5(4u) = s_5(y3) \in \{0, 1, 2\} \), so \( u4u \) is an occurrence of a string of \( P_2 \) in \( w \), as desired.

Case 5b: \( c = 4 \). Then \( ycy = 443 \cdots 444 \cdots 443 \), so the last three characters of \( y \) must be 433. Since \( |y| \geq 5 \), we must have \( y = 443 \cdots 433 \). Write \( y = 4433y' \). Then \( ycy = 4433 \ y' \ 4433 \ y' \) is a subword of \( h(w) \) and there exists \( u \) such that \( h(u) = y' \). Then \( h(u3u) = y' \ 4433 \ y' \). Now \( s_5(u3) = s_5(h(u3)) = s_5(y'44433) = s_5(y3) = s_5(y4) \in \{0, 1, 2\} \), so \( u3u \) is an occurrence of a string of \( P_2 \) in \( w \), as desired.

This completes the proof of Theorem 7.

Proof of Theorem 5. Define
\[
v = h^a(4) = 44344334443344444333 \cdots
\]
We claim \( v \) avoids \( P_2 \). This follows because the word 4 avoids \( P_2 \), and by Theorem 7, if \( w \) avoids \( P_3 \) then so does \( h(w) \). Now consider \( S(v) = 0431432032103104314 \cdots \). From Lemma 6, it follows that \( S(v) \) avoids the patterns \( xx, x\sigma(x), \) and \( x\sigma^2(x) \).

This completes the proof of Theorem 5, and hence assertion (c) of Theorem 2.

We now turn to assertion (d) of Theorem 2. From Proposition 4 with \( r = 4 \), we know any word of length \( \geq 4 \) over \( \Sigma_5 \) contains an occurrence of one of the patterns \( xx, x\sigma(x), x\sigma^2(x), \) or \( x\sigma^3(x) \). The methods of Theorem 7 and Lemma 6 lead immediately to

**Theorem 8** It is possible to simultaneously avoid the patterns \( xx, x\sigma(x), x\sigma^2(x), \) and \( x\sigma^3(x) \) over \( \Sigma_6 \).

**Proof.** We construct an infinite word \( w \) over \( \Sigma_6 \) such that \( w \) simultaneously avoids the patterns \( xx, x\sigma(x), x\sigma^2(x), \) and \( x\sigma^3(x) \). Let \( g \) be the morphism over \( \{4,5\} \) defined by \( g(5) = 55544 \) and \( g(4) = 555544 \). We claim that \( w = S(g^r(5)) \) simultaneously avoids the patterns \( xx, x\sigma(x), x\sigma^2(x), \) and \( x\sigma^3(x) \). The proof follows exactly the same plan as that of Theorem 7. We omit it here.

**Remark.** We note that the morphisms used in the proof of Theorem 7 and Theorem 8 do not generalize to any other \( j \). For example, if we were to define \( h \) analogously for \( j \neq 4 \), we would have \( h(6) = 666655 \) and \( h(5) = 6666555 \). By inspection, we see that \( h(6) = 6666555 \) contains \( gcy \) where \( y = 66 \) and \( c = 6 \). Hence \( \sigma(y) = 4 \) and so \( S(h(6)) \) does not avoid the pattern \( x\sigma^4(x) \).

Finally, we turn to assertion (e). First, we show it is not possible to avoid the patterns \( xx, x\sigma(x), \ldots, x\sigma^4(x) \) on 7 letters. Here the corresponding tree has 215 leaves, and the longest leaf has length 36. See Figure 2 below.

**Figure 2:** Leaves of the tree giving the proof of assertion (e)

```plaintext
011  1842143432  184120542041  1440214201420142  0431440241012101
012  1842143431  184120542041  1440214201420142  0431440241012101
013  1842143431  184120542041  1440214201420142  0431440241012101
014  1842143431  184120542041  1440214201420142  0431440241012101
015  1842143431  184120542041  1440214201420142  0431440241012101
016  1842143431  184120542041  1440214201420142  0431440241012101
017  1842143431  184120542041  1440214201420142  0431440241012101
018  1842143431  184120542041  1440214201420142  0431440241012101
019  1842143431  184120542041  1440214201420142  0431440241012101
020  1842143431  184120542041  1440214201420142  0431440241012101
021  1842143431  184120542041  1440214201420142  0431440241012101
022  1842143431  184120542041  1440214201420142  0431440241012101
023  1842143431  184120542041  1440214201420142  0431440241012101
024  1842143431  184120542041  1440214201420142  0431440241012101
025  1842143431  184120542041  1440214201420142  0431440241012101
026  1842143431  184120542041  1440214201420142  0431440241012101
027  1842143431  184120542041  1440214201420142  0431440241012101
028  1842143431  184120542041  1440214201420142  0431440241012101
029  1842143431  184120542041  1440214201420142  0431440241012101
030  1842143431  184120542041  1440214201420142  0431440241012101
031  1842143431  184120542041  1440214201420142  0431440241012101
032  1842143431  184120542041  1440214201420142  0431440241012101
033  1842143431  184120542041  1440214201420142  0431440241012101
034  1842143431  184120542041  1440214201420142  0431440241012101
035  1842143431  184120542041  1440214201420142  0431440241012101
036  1842143431  184120542041  1440214201420142  0431440241012101
```

\[
avoid2: \text{submitted to World Scientific on November 29, 1999}
\]
Using the tree traversal algorithm, we can prove

**Theorem 9** One cannot avoid the patterns $xx$, $x\sigma(x)$, $\ldots$, $x^{i}(x)$ on $j+3$ letters, for $j \geq 5$.

**Proof.** Consider trying to generate an infinite word $w$ over $\mathbb{Z}$ starting with 0, subject to two conditions: (1) avoiding the pattern $x^{i}(x)$ for all $i$, where $|x| \geq 2$, and (2) avoiding all subwords of length 2 that are not of the form $(n, n - 1)$ or $(n, n - 2)$ for $n \in \mathbb{Z}$.

Let us now apply the tree traversal algorithm to this avoidance problem. The tree $T$ so produced has 71 leaves and the longest leaf has length 12. All the occurrences of $x^{i}(x)$ found at the leaves of $T$, for $|x| \geq 2$, satisfy $i \in X = \{-3, -4, -6, -7, -8\}$.

Now consider the labels of this tree reduced modulo $j + 3$. The patterns at the leaves are still of the form $x^{i}(x)$, except now $i$ is reduced modulo $j + 3$. In order for $T$ to correctly represent a proof that the pattern $x^{i}(x)$ cannot be avoided for $0 \leq i \leq j$ we must check that $i \mod (j + 3) \in \{0, 1, \ldots, j\}$ for all $i \in X$. But this is clearly true for $j \geq 5$.

Figure 3 lists the leaves of $T$ in coded form. We use the letters $A, B, C, D, E, F, G$ to represent 10, 11, 12, 13, 14, 15, 16 respectively, and the word $a_{1}a_{2}\ldots a_{j}$ represents the leaf $(-a_{1}, -a_{2}, \ldots, -a_{j})$. ■

**Figure 3:** Leaves of the tree giving the proof of Theorem 9.
We now show it is possible to simultaneously avoid the patterns $xx, x\sigma(x), \ldots, x\sigma^j(x)$ on $\Sigma_{j+4}$ for $j \geq 4$. Actually, we prove a more general result from which this result will follow.

**Theorem 10** Let $k \geq 4$ be an integer, and let $A \subset \Sigma_k$ such that $|\text{Card } A| \leq k - 3$. Then it is possible to simultaneously avoid the patterns $\{x\sigma^a(x) : a \in A\}$ over $\Sigma_k$.

**Proof.** Once again, the idea is to consider the first differences of words, modulo $k$. Suppose we can construct a word $w$ over $\Sigma_k$ such that $w$ avoids both (i) the pattern $yey$, where $|y| \geq 1$ and $|e| = 1$, and (ii) the letters $a \in A$. Then it follows from Lemma 6 that $S(w)$ avoids the pattern $x\sigma^a(x)$.

**Lemma 11** Let $w = a_1a_2a_3 \cdots$ be any squarefree word over $\Sigma_3$. Then the word $a_1a_1a_2a_2a_3a_3 \cdots$ avoids the pattern $yey$ for $y \in \Sigma_3^*$ and $e \in \Sigma_3$.

**Proof.** Suppose $y = b_1b_2b_3$, and the pattern $yey$ occurs in $z = a_1a_1a_2a_2a_3a_3 \cdots$. There are three cases to consider, depending on $|y|$ and where $y$ starts in $z$.

**Case 1:** $|y|$ is even and $y$ starts with $ai_i$. Let $k = 2j$. Then we have

$$b_1 \ b_2 \ \cdots \ b_{2j} \ \ c \ \ b_1 \ b_2 \ \cdots \ b_{2j}$$

$$= \ a_i \ a_i \ \cdots \ a_{i+j-1} \ a_{i+j} \ a_{i+j} \ a_{i+j+1} \ \cdots \ a_{i+2j}$$

and so $a_{i+j} = b_1 = a_i = b_2 = a_{i+j+1}$. It follows that $w$ contains the square $a_ia_{i+1}a_{i+1}$, a contradiction.

**Case 2:** $|y|$ is even and $y$ starts with $ai_{i+1}$. Let $k = 2j$. Then we have

$$b_1 \ b_2 \ \cdots \ b_{2j} \ \ c \ \ b_1 \ b_2 \ \cdots \ b_{2j}$$

$$= \ a_i \ a_{i+1} \ \cdots \ a_{i+j} \ a_{i+j} \ a_{i+j+1} \ a_{i+j+1} \ \cdots \ a_{i+2j}$$

and so $a_i = b_1 = a_{i+j+1} = b_2 = a_{i+1}$. It follows that $w$ contains the square $a_ia_{i+1}$, a contradiction.

**Case 3:** $|y|$ is odd. Let $k = 2j + 1$. Then either

$$b_1 \ b_2 \ \cdots \ b_{2j} \ b_{2j+1} \ c \ \ b_1 \ b_2 \ \cdots \ b_{2j} \ b_{2j+1}$$

$$= \ a_i \ a_i \ \cdots \ a_{i+j-1} \ a_{i+j} \ a_{i+j} \ a_{i+j+1} \ a_{i+j+1} \ \cdots \ a_{i+2j} \ a_{i+2j+1}$$

**avoid2: submitted to World Scientific on November 29, 1999**
or
\[
\begin{array}{ccccccc}
  b_1 & b_2 & \cdots & b_j & b_{j+1} & c & b_1 & b_2 & \cdots & b_j & b_{j+1} \\
  a_i & a_{i+1} & \cdots & a_{i+j} & a_{i+j+1} & a_{i+j+2} & a_{i+j+3} & a_{i+j+4} & \cdots & a_{i+j+2j} & a_{i+j+2j+1}
\end{array}
\]

In either case we find
\[
\begin{align*}
  a_i &= b_1 = a_{i+j+1} \\
  a_{i+1} &= b_3 = a_{i+j+2} \\
  &\vdots \\
  a_{i+j} &= b_{2j+1} = a_{i+j+2j+1}
\end{align*}
\]

It follows that \( a_i a_{i+1} \cdots a_{i+j} = a_{i+j+1} a_{i+j+2} \cdots a_{i+j+2j+1} \) and so \( w \) contains the square \( a_i a_{i+1} \cdots a_{i+j} \), a contradiction. The proof of the Lemma is complete. ■

**Remark.** One cannot avoid the pattern \( ycy \), with \(|y| \geq 1 \) and \(|c| = 1 \), over an alphabet of 2 letters. As the tree traversal algorithm shows, any word of length \( \geq 7 \) over \( \{0,1\} \) contains an occurrence of \( ycy \).

Now we can complete the proof of Theorem 10. Let \( x \) be any squarefree word over \( \{0,1,2\} \). Since \( \text{Card} A = k - 3 \), we have \( \Sigma_k - A = \{d, e, f\} \) for some distinct integers \( 0 \leq d, e, f < k \).

Consider the morphism \( \phi : \Sigma_{j+4}^* \rightarrow \Sigma_{j+4}^* \) defined as follows:
\[
\begin{align*}
  0 &\rightarrow dd \\
  1 &\rightarrow ee \\
  2 &\rightarrow ff
\end{align*}
\]

We claim \( S(\phi(x)) \) avoids the patterns \( xx, x\sigma(x), \ldots, x\sigma^j(x) \).

Let \( v = S(\phi(x)) \). Then \( \Delta(v) = \phi(x) \) clearly avoids \( ycy \) by Lemma 11, and it also avoids all the letters in \( A \) by construction. Then by Lemma 6, \( v \) avoids the patterns \( x\sigma^a(x) \) for \( a \in A \). ■

As a consequence we get

**Corollary 12** It is possible to simultaneously avoid the patterns \( xx, x\sigma(x), \ldots, x\sigma^j(x) \) on \( \Sigma_{j+4}^* \) for \( j \geq 4 \).

The proof of Theorem 2 is now complete. ■

4 Even more results

One may also consider the problem of avoiding other sets of patterns of the form \( x\sigma^a(x) \). In this section, we let \( j \geq 1 \) be an integer, and consider avoiding the \( 2j+1 \) patterns \( x\sigma^{-j}(x), \ldots, x\sigma^{-1}(x), xx, x\sigma(x), \ldots, x\sigma^j(x) \) simultaneously over the alphabet \( \Sigma_k \).

**Theorem 13** For \( j \geq 1 \), one can simultaneously avoid the patterns \( x\sigma^{-j}(x), \ldots, x\sigma^{-1}(x), xx, x\sigma(x), \ldots, x\sigma^j(x) \) over \( \Sigma_{2j+4}^* \), and this is best possible.
Proof.
By Theorem 10 with $A = \{-j, 1 - j, \ldots, -1, 0, 1, 2, \ldots, j\}$, we see that we can simultaneously avoid the patterns $x\sigma^{-j}(x)$, ..., $xx\sigma^{-1}(x)$, $xx$, $xx\sigma(x)$, ..., $xx\sigma^i(x)$ over $\Sigma_{2j+4}$.

It follows from Proposition 3 that one cannot avoid $x\sigma^{-j}(x)$, ..., $xx\sigma^{-1}(x)$, $xx$, $xx\sigma(x)$, ..., $xx\sigma^i(x)$ over $\Sigma_{2j+3}$ or smaller alphabet.

To prove that one cannot simultaneously avoid the patterns $x\sigma^{-j}(x)$, ..., $xx\sigma^{-1}(x)$, $xx$, $xx\sigma(x)$, ..., $xx\sigma^i(x)$ over $\Sigma_{2j+3}$, we use the tree traversal algorithm. Then every word of length $\geq 8$ over $\Sigma_{2j+3}$ contains an occurrence of $xx\sigma^i(x)$ for some $l$ with $-j \leq l \leq j$. Figure 4 below gives the output of the tree traversal algorithm, showing that there are 24 leaves. Here $t = j + 1$.

<table>
<thead>
<tr>
<th>$0$, $-t$, $-2t$, $-3t$</th>
<th>$0$, $-t$, $-2t$, $-t$, $0$, $t$</th>
<th>$0$, $-t$, $-2t$, $-t$, $0$, $-t$, $-2t$, $-3t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0$, $-t$, $0$, $-t$</td>
<td>$0$, $-t$, $0$, $t$, $0$, $t$</td>
<td>$0$, $-t$, $-2t$, $-t$, $0$, $-t$, $-2t$, $-t$</td>
</tr>
<tr>
<td>$0$, $t$, $0$, $t$</td>
<td>$0$, $t$, $0$, $-t$, $0$, $-t$</td>
<td>$0$, $t$, $0$, $-t$, $0$, $-t$, $-2t$, $-t$</td>
</tr>
<tr>
<td>$0$, $t$, $2t$, $3t$</td>
<td>$0$, $t$, $2t$, $t$, $0$, $-t$</td>
<td>$0$, $t$, $0$, $t$, $0$, $t$, $-2t$, $0$</td>
</tr>
<tr>
<td>$0$, $-t$, $-2t$, $-t$, $-2t$</td>
<td>$0$, $-t$, $-2t$, $-t$, $-t$, $0$</td>
<td>$0$, $t$, $0$, $-t$, $0$, $t$, $0$, $-t$</td>
</tr>
<tr>
<td>$0$, $-t$, $0$, $t$, $2t$</td>
<td>$0$, $-t$, $0$, $t$, $0$, $-t$, $-2t$</td>
<td>$0$, $t$, $0$, $-t$, $0$, $t$, $0$, $t$</td>
</tr>
<tr>
<td>$0$, $t$, $0$, $-t$, $-2t$</td>
<td>$0$, $t$, $0$, $-t$, $0$, $t$, $2t$</td>
<td>$0$, $t$, $0$, $t$, $0$, $t$, $2t$, $2t$, $t$</td>
</tr>
<tr>
<td>$0$, $t$, $2t$, $2t$, $2t$</td>
<td>$0$, $t$, $2t$, $t$, $0$, $t$, $2t$, $0$</td>
<td></td>
</tr>
</tbody>
</table>

5 Avoiding $xx\sigma^i(x)$ for all $i$

Generalizing the results of the previous section, we may ask if it is possible to avoid the patterns $xx\sigma^i(x)$ for all $i$. Unfortunately, this is clearly impossible, for if a word $z$ begins with $ij$, then it contains a subword of the form $i\sigma^{i-1}(i)$.

However, we can relax our conditions for avoidance, as follows: we say an infinite word weakly avoids the patterns $xx\sigma^i(x)$ if it contains no subwords of the form $xx\sigma^i(x)$ with $|x| \geq 2$. (In contrast, our previous notion of avoidability we will call strong.)

**Proposition 14** Over $\Sigma_2$, every word of length $\geq 8$ contains a subword of the form $xx\sigma^i(x)$ for some $i \geq 0$, with $|x| \geq 2$.

**Proof.** Our simple tree traversal algorithm proves this. The tree generated has 24 leaves, and the leaves are given in Figure 5.

<table>
<thead>
<tr>
<th>$0000$</th>
<th>$00101$</th>
<th>$00010000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0011$</td>
<td>$001001$</td>
<td>$00010001$</td>
</tr>
<tr>
<td>$0101$</td>
<td>$010000$</td>
<td>$00100010$</td>
</tr>
<tr>
<td>$0110$</td>
<td>$011000$</td>
<td>$00100011$</td>
</tr>
</tbody>
</table>
\[
\begin{array}{ccc}
00011 & 001001 & 0100010 \\
00101 & 0010000 & 01000101 \\
01001 & 0100011 & 01110110 \\
01111 & 0111010 & 01110111
\end{array}
\]

However, it is possible to weakly avoid the patterns \(x\sigma^i(x)\) for all \(i \geq 0\) over \(\Sigma_3\). Let \(w\) be any squarefree word over \(\{0, 1, 2\}\), and consider the morphism \(f\) which maps

\[
\begin{align*}
0 & \rightarrow 00 \\
1 & \rightarrow 10 \\
2 & \rightarrow 20.
\end{align*}
\]

**Theorem 15** The infinite word \(f(w)\) weakly avoids the patterns \(x\sigma^i(x)\) for all \(i \geq 0\).

**Proof.** Let \(w = c_1c_2c_3 \cdots\), and \(f(w)\) contains a subword of the form \(z = x\sigma^i(x)\) for some \(i\) and \(|x| \geq 2\). There are two cases, depending on \(|x| \mod 2\).

Case 1: \(|x| \equiv 0 \pmod{2}\). In this case, there are two possibilities, depending where \(x\) starts in \(f(w)\):

\[
\begin{align*}
\sigma^i(x) & \\
z & = d_1d_2\cdots d_j0 \\
\sigma^i(x) & \\
z & = 0d_1d_2\cdots d_j0
\end{align*}
\]

where \(d_i = c_{i+k}\) for some integer \(k \geq 0\). Comparing the second symbol in the first case, or the first symbol in the second case, we see that if \(z = x\sigma^i(x)\), then \(i = 0\). Hence \(d_i = d_{j+t}\) for \(1 \leq t \leq j\), and so \(c_{k+t} = c_{k+j+t}\) for \(1 \leq t \leq j\), contradicting the assumption that \(w\) was squarefree.

Case 2: \(|x| \equiv 1 \pmod{2}\).

\[
\begin{align*}
\sigma^i(x) & \\
z & = d_10d_2\cdots 0d_j0 \\
\sigma^i(x) & \\
z & = 0d_10\cdots d_j0d_{j+1}\cdots
\end{align*}
\]

If \(z = x\sigma^i(x)\), then, in the first case, we must have \(d_1 = d_2\), and in the second \(d_j = d_{j+1}\). Both correspond to a square in \(w\), a contradiction. \(\blacksquare\)

We might also try weakly avoiding \(x\sigma^i(x)\) for \(0 < i < k\) over \(\Sigma_k\), while simultaneously (strongly) avoiding \(xx\).

**Theorem 16** If \(k = 4\), one can, over \(\Sigma_k\), simultaneously weakly avoid \(x\sigma^i(x)\) for \(0 < i < k\) and strongly avoid \(xx\). Here \(k\) is best possible.
Proof. We can weakly avoid \(x\sigma^i(x)\) for \(0 < i < k\) and strongly avoid \(xx\) over \(\Sigma_4\) as follows: let \(w\) be any squarefree word over \(\{1, 2, 3\}\), and consider the morphism \(f\) which maps

\[
\begin{align*}
1 & \rightarrow 10 \\
2 & \rightarrow 20 \\
3 & \rightarrow 30.
\end{align*}
\]

Then it follows from the same method of the proof of Theorem 15 that \(f(w)\) weakly avoids \(x\sigma^i(x)\) for all \(i\). However, it is clear from the construction that \(f(w)\) has no subword of the form \(cc\) for \(c \in \Sigma_4\), so \(f(w)\) also strongly avoids \(xx\).

On the other hand, the tree traversal algorithm shows that over \(\Sigma_3\), any word of length \(\geq 8\) has a (weak) occurrence of \(x\sigma^i(x)\) with \(0 < i < 3\), or a strong occurrence of \(xx\). The tree generated has 24 leaves, and the leaves are given in Figure 6. □

Figure 6: Leaves of the tree giving a proof of Theorem 16.

\[
\begin{array}{ccc}
0101 & 010202 & 01200101 \\
0120 & 012102 & 0120102 \\
0202 & 020101 & 0121020 \\
0210 & 021201 & 0121012 \\
01021 & 0120102 & 0201020 \\
01212 & 0121010 & 02010202 \\
02012 & 0201021 & 02120210 \\
02121 & 0212020 & 02120212 \\
\end{array}
\]

Acknowledgments

We acknowledge with thanks conversations with James Currie. Jean-Paul Allouche read an early draft of this paper and made several useful suggestions.

References