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: {ja3loftus,shallit,m2wang}@uwaterloo.ca

Let $\Sigma_k := \{0, 1, \ldots, k-1\}$ for an integer $k \geq 2$. Define $\sigma(a) = (a+1) \bmod k$ 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 $x\sigma(x)$ over Σ_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[6,2] 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"[3]. For other papers on pattern avoidance, the reader can fruitfully consult, for example, [1,5,4].

In this paper, we consider some new generalizations of Thue's problem. We start with some notation. Let Σ, Γ 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 Σ^ω denote the set of all one-sided infinite words over Σ, and we let $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$. If $x \in \Sigma^+$, then by x^ω we mean the one-sided infinite word $xxx\cdots$.

If there exist words $x, y \in \Sigma^*$, $w, z \in \Sigma^\infty$ such that $w = xyz$, then we say y is a finite *subword* of w. Suppose we are given a finite or infinite subset $P \subseteq \Sigma^*$. 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 Σ if it is possible to construct an infinite word $\mathbf{w} \in \Sigma^\omega$ which avoids P.

Sometimes we employ a common abuse of notation. For example, instead of saying that the infinite word $\mathbf{w}$ avoids $\{xx : x \in \Sigma^+\}$, we will instead simply say that $\mathbf{w}$ *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 $\sigma_k(a) = (a+1) \bmod k$. 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 Σ_k as residue class representatives so that, for example, -1 and 2 denote the same element of Σ_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

Σ_3^*.

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

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^ω and 1^ω. However, we have the following result:

Theorem 1 *Over Σ_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

$$\mathbf{w} = \prod_{i \geq 1} ((-i) \bmod 3)^{a_i} = 2^{a_1}\, 1^{a_2}\, 0^{a_3}\, 2^{a_4}\, 1^{a_5}\, 0^{a_6} \cdots$$

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

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 Σ_3, and 3 is best possible.*

(b) *One can avoid the patterns xx and $x\sigma(x)$ simultaneously over Σ_5, and 5 is best possible.*

(c) *One can avoid the patterns xx, $x\sigma(x)$, $x\sigma^2(x)$ simultaneously over Σ_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 Σ_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 Σ_{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 Σ_k, we explicitly construct an infinite word over Σ_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 Σ_k. We maintain a queue, Q, and initialize it with the empty word ϵ. 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 it. 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 Σ_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, $x\sigma(x)$, ..., $x\sigma^j(x)$ 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 $0 \le i \le j$. 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 [6,2] 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 ≥ 4 over $\Sigma_2 = \{0, 1\}$ contains an occurrence of the pattern xx. More generally, we have

Proposition 3 *Let $k \ge 2$ be an integer, and let r be an integer with $1 \le r < k$. Then any word of length ≥ 4 over Σ_k contains an occurrence of the pattern $x\sigma^a(x)$ for some $a \not\equiv r$ (mod k).*

Proof. We use the tree traversal algorithm. Assume the first letter is 0; then if the next letter is $a \ne r$, 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 $x\sigma^{2r}(x)$ for $x = (0, r)$. Since $r \ne 0$, we have $2r \not\equiv r$ (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 Σ_3. We also have

Proposition 4 *Every word of length ≥ 24 over Σ_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. ∎

Figure 1: Leaves of the tree giving a proof of Proposition 4.

0202	0203210202	031321310310	021020320210313	02102032021032020	0313213103132103131
0213	0210203203	031321031320	021020320210310	02102032021032021	0313213103132103132
0310	0210313210	031321031321	021020320210321	02103132131031320	0320210203202103203
02031	0210320213	032021032020	021031321310310	02103132131032132	0321310313213103202
03131	0313213102	032021032021	031321310321313	02103202102032020	0321310313213103203
03203	0313210202	032131031320	031321310321310	02103202102032131	0321310313213103210
021021	0313210203	032131032131	032021020320213	02103202102032132	02032102032021020320
031320	0313210310	032131032132	032021020321021	02103202102032103	02032102032021032020
032020	0320210202	032102032020	032131031321313	03132131031321020	02032102032021032021
032132	0320210313	032102032102	032131031321020	03132131031321021	02103132131031321020
032103	0320210310	032102032103	032131031321021	03132131031321032	02103132131031321021
0203203	0320210321	0203202102031	032131031321032	03213103132131031	02103132131031321032
0210202	0321310310	0203202103203	032102032021021	03213103132103131	03202102032021032020
0210310	0321310320	0203210203203	0203202102032020	03213103132103132	03202102032021032021
0210321	02032021021	0210203210202	0203202102032021	03210203202102031	03213103132131032132
0320213	02032102031	0210203210203	0203202102032131	03210203202103203	03210203202102032131
0321313	02102032020	0210313213102	0203202102032132	020320210203210202	03210203202102032132
0321021	02102032131	0210320210202	0203202102032103	020320210203210203	03210203202102032103
02032020	02102032132	0313213103131	0210203202102031	020321020320210202	021031321310313210310
02032132	02102032103	0313213103202	0210203202102032	020321020320210313	021032021020320210313
02032103	02103132132	0313213103203	0210203202103203	020321020320210310	021032021020320210310
02102031	03132103131	0313213103210	0210313213103131	020321020320210321	021032021020320210321
02103131	03202102031	0320210203203	0210313213103202	021031321310313213	032131031321310321313
02103203	03202103203	0321020320213	0210313213103203	021031321310321313	032131031321310321310
03132132	03213103131	02032021032020	0210313213103210	021031321310321310	032102032021020321021
03213102	03213103210	02032021032021	0210320210203203	021032021020320213	0203210203202102032131
03210202	03210203203	02032102032020	0313213103132131	021032021020321021	0203210203202102032132
020320213	03210203213	02102032021021	0313213103132132	031321310313210310	0203210203202102032103
020321313	020320210202	02103202102031	0320210203202102	032021020320210313	0210313213103132103131
020321021	020320210313	03132131031320	0320210203210202	032021020320210310	0210313213103132103132
021031320	020320210310	03132131032132	0320210203210203	032021020320210321	0210320210203202103203
021032020	020320210321	03202102032020	0321310313213102	032102032021020320	0321020320210203210202
031321313	020321020321	03202102032131	0321310313210310	032102032021032020	0321020320210203210203
031321021	021020320213	03202102032132	0321020320210202	032102032021032021	02032102032021020321021
031321032	021020321021	03202102032103	0321020320210313	0203210203202102031	02103202102032021032020
032021021	021031321313	03213103132132	0321020320210310	0203210203202103203	02103202102032021032021
032102031	021032021021	020320210203203	0321020320210321	0210320210203202102	020321020320210203210202
0203213102	021032021031	020321020320213	02032021020321021	0210320210203210202	020321020320210203210203
0203213103	021032021032	021020320210202	02032102032021021	0210320210203210203	

Thus we cannot avoid the patterns xx and $x\sigma(x)$ simultaneously over Σ_4. However, we can avoid the patterns xx and $x\sigma(x)$ simultaneously over Σ_5. This will follow from Theorem 5 below.

Next, let us prove assertion (c). As we have seen in Proposition 3 above, every word of length ≥ 4 over Σ_4 contains an occurrence of one of the patterns xx, $x\sigma(x)$, or $x\sigma^2(x)$. We now show

Theorem 5 *It is possible to simultaneously avoid the patterns xx, $x\sigma(x)$, and $x\sigma^2(x)$ over Σ_5.*

Before starting the proof, we introduce some notation. If $\mathbf{w} = a_1a_2a_3\cdots$ is a word over Σ_k, then

$$\Delta(\mathbf{w}) := (a_2 - a_1, a_3 - a_2, a_4 - a_3, \ldots),$$

where the differences are, of course, taken mod k. Similarly, we write

$$S(\mathbf{w}) = (0, a_1, a_1 + a_2, a_1 + a_2 + a_3, \ldots),$$

where the sums are, of course, taken mod k. Note that $\Delta(S(\mathbf{w})) = \mathbf{w}$, and if $a_1 = 0$, then $S(\Delta(\mathbf{w})) = \mathbf{w}$. Finally, if $x = a_1 \cdots a_m \in \Sigma_k^*$, we define $s_k(x) = (\Sigma_{1 \leq i \leq m} a_i) \bmod k$.

The following lemma relates occurrences of patterns of the form $x\sigma^a(x)$ in $\mathbf{w}$ to other, easier-to-study patterns in $\Delta(\mathbf{w})$.

Lemma 6 *Let $w \in \Sigma_k^\infty$, and let $a \in \Sigma_k$. Then w avoids the pattern $x\sigma^a(x)$ iff $\Delta(w)$ avoids $\{ycy \ : \ y \in \Sigma_k^*, \ c \in \Sigma_k, \ \text{and} \ s_k(yc) = a\}$.*

Proof. Suppose w contains an occurrence of the pattern $x\sigma^a(x)$. Write $x = b_1 b_2 \cdots b_i$. Then

$$w = w' b_1 b_2 \cdots b_i \sigma^a(b_1) \cdots \sigma^a(b_i) \cdots .$$

Thus

$$\Delta(w) = \Delta(w'b_1), (b_2 - b_1, \ldots, b_i - b_{i-1}, \sigma^a(b_1) - b_i, b_2 - b_1, \ldots, b_i - b_{i-1}, \ldots),$$

and hence contains ycy with

$$y = (b_2 - b_1, \ldots, b_i - b_{i-1}), \quad c = \sigma^a(b_1) - b_i.$$

Also

$$\begin{aligned} s_k(yc) &= (b_2 - b_1) + \cdots + (b_i - b_{i-1}) + \sigma^a(b_1) - b_i \\ &= (b_i - b_1) + (a + b_1 - b_i) \\ &= a. \end{aligned}$$

Now suppose $\Delta(w)$ contains a subword ycy with $y \in \Sigma_k^*$, $c \in \Sigma_k$, and $s_k(yc) = a$. Then $\Delta(w) = xycyz$ 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

$$\begin{aligned} w = (&e, e + b_1, e + b_1 + b_2, \ldots, e + b_1 + b_2 + \cdots + b_j, e + f + d_1, e + f + d_1 + d_2, \ldots, \\ &e + f + d_1 + d_2 + \cdots + d_i, e + f + g + c, e + f + g + c + d_1, \\ &e + f + g + c + d_1 + d_2, \ldots, e + f + g + c + d_1 + d_2 + \cdots + d_i, \ldots) \end{aligned}$$

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 $x\sigma^a(x)$, where $x = (e + f, e + f + d_1, \ldots, e + f + d_1 + d_2 + \cdots + d_i)$ and $a = g + c$. But $g + c = s_k(d_1 d_2 \cdots d_i c) = s_k(yc)$. ■

Now to prove Theorem 5, it suffices to construct an infinite word **v** where **v** avoids

$$P_2 := \{ycy \ : \ y \in \Sigma_5^*, \ c \in \Sigma_5, \ \text{and} \ s_5(yc) \in \{0, 1, 2\}\}.$$

For then we could set $\mathbf{w} = S(\mathbf{v})$, and by Lemma 6, **w** avoids the patterns xx, $x\sigma(x)$, and $x\sigma^2(x)$ over Σ_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 $s_5(yc) \in \{0,1,2\}$. Write $h(w) = z_1 ycy z_2$. 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 ≤ 5 that occur in $\{4433, 44433\}^*$. 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 $h(w)$.

Table 1: Proof that $|y| \geq 5$.

$\|y\|$	yc	$s_5(yc)$	ycy	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	343
	344	1	34434	yes	434
	433	0	43343	yes	343
	443	1	44344	yes	434
	444	2	44444	yes	4444
3	3344	4	3344334	no	
	3443	4	3443344	no	
	3444	0	3444344	yes	434
	4334	4	4334433	no	
	4433	4	4433443	no	
	4443	0	4443444	yes	434
4	33443	2	334433344	yes	333
	33444	3	334443344	no	
	34433	2	344333443	yes	333
	34443	3	344433444	no	
	43344	3	433444334	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 $y'4y'$ 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 $y' = 334t$ and note that $y'3y'$ is a subword of $h(w)$. But $s_5(y'3) = s_5(334t3) = s_5(34t33) = s_5(y3) \in \{0,1,2\}$, so $y'3y' \in P_2$, 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 $|b| = 1$. Then $y = 43t'b$. Then $4ycy = 4\ 43t'b\ 4\ 43t'b$ is a subword of $h(w)$. Let $y' = 443t'$. Then $y'by'$ is a subword of $h(w)$, and $s_5(y'b) = s_5(443t'b) = s_5(43t'b4) = s_5(y4) \in \{0,1,2\}$, so $y'by' \in P_2$, 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 443. There are two subcases to consider:

> *Case 5a:* $c = 3$. Then the last two characters of y must be 43. We have $ycy = 443\cdots 43\ 3\ 443\cdots 43$. 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 $ycy = 443\cdots\ 4\ 443\cdots$, so the last three characters of y must be 433. Since $|y| \geq 5$, we must have $y = 4433\cdots 433$. Write $y = 4433y'$. Then $ycy = 4433\ y'\ 44433\ y'$ is a subword of $h(w)$ and there exists u such that $h(u) = y'$. Then $h(u3u) = y'\ 44433\ y'$. Now $s_5(u3) = s_5(h(u3)) = s_5(y'44433) = s_5(4433y'4) = s_5(y4)$, 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

$$\mathbf{v} = h^\omega(4) = 443344334443344433\cdots.$$

We claim $\mathbf{v}$ avoids P_2. This follows because the word 4 avoids P_2, and by Theorem 7, if w avoids P_2 then so does $h(w)$. Now consider $S(\mathbf{v}) = 0431432032103104314\cdots$. From Lemma 6, it follows that $S(\mathbf{v})$ 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 ≥ 4 over Σ_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 Σ_6.*

Proof. We construct an infinite word $\mathbf{w}$ over Σ_6 such that $\mathbf{w}$ 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 $\mathbf{w} = S(g^\omega(5))$ 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 $j = 4$, we would have $h(6) = 666655$ and $h(5) = 6666655$. By inspection, we see that $h(6) = 6666655$ contains ycy where $y = 66$ and $c = 6$. Hence $s_7(yc) = 4 = j$ 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)

0531	0543205432	0653106542065432	06542065431654320543210	065431654320543216432105321065
0542	0543216431	0654206543165431	06543165432054321643216	0532106421065310654206543165431
0643	0642106420	0654316543205431	053210642106531065420653	0543216432105321064210653106543
05320	0642106421	05321064210653105	054321643210532106421064	0642106531065420654316543205431
06420	0642106532	05432164321053216	064210653106542065431653	0653106542065431654320543216431
06532	0653106531	06421065310654205	065310654206543165432053	0654206543165432054321643210531
054310	0653106532	06531065420654310	065420654316543205432165	0654316543205432164321053210643
064216	0653106543	06542065431654321	065431654320543216432106	05321064210653106542065431654321
065316	0654206542	06543165432054320	0532106421065310654206542	05432164321053210642106531065421
065421	0654316542	053210642106531064	0543216432105321064210654	06421065310654206543165432054320
065432	05321064216	054321643210532105	0642106531065420654316542	06531065420654316543205432164320
0532165	05432164320	064210653106542064	0653106542065431654320542	06542065431654320543216432105320
0543164	06421065316	065310654206543164	0654206543165432054321642	06543165432054321643210532106420
0543206	06531065421	065420654316543206	0654316543205432164321054	053210642106531065420654316543206
0543210	06542065432	065431654320543210	05321064210653106542065432	054321643210532106421065310654205
0653105	06543165431	0532106421065310653	05432164321053210642106532	064210653106542065431654320543210
0654205	053210642105	0543216432105321065	06421065310654206543165431	065310654206543165432054321643216
0654310	054321643216	0642106531065420653	06531065420654316543205431	065420654316543205432164321053216
05321642	064210653105	0653106542065431653	06542065431654320543216431	065431654320543216432105321064216
05321643	065310654205	0654206543165432053	06543165432054321643210531	0532106421065310654206543165432053
05321054	065420654310	0654316543205432165	053210642106531065420654310	0543216432105321064210653106542064
05321065	065431654321	05321064210653106543	054321643210532106421065316	0642106531065420654316543205432165
05431653	0532106421064	05432164321053210643	064210653106542065431654321	0653106542065431654320543216432106
05431654	0543216432106	06421065310654206542	065310654206543165432054320	0654206543165432054321643210532105
05432053	0642106531064	06531065420654316542	065420654316543205432164320	0654316543205432164321053210642105
05432165	0653106542064	06542065431654320542	065431654320543216432105320	05321064210653106542065431654320542
06421053	0654206543164	06543165432054321642	0532106421065310654206543164	05432164321053210642106531065420653
06421054	0654316543206	053210642106531065421	0543216432105321064210653105	06421065310654206543165432054321642
06531064	05321064210654	054321643210532106420	0642106531065420654316543206	06531065420654316543205432164321054
06542064	05432164321054	064210653106542065432	0653106542065431654320543210	06542065431654320543216432105321065

06543164	06421065310653	06531065420654316543 1	06542065431654320543216432 16	06543165432054321643210532106421064
053210531	06531065420653	06542065431654320543 1	06543165432054321643210532 16	05321064210653106542065431654320543 1
053210643	06542065431653	06543165432054321643 1	053210642106531065420654316 53	05321064210653106542065431654320543 2
054320542	06543165432053	053210642106531065420 5	054321643210532106421065310 64	05432164321053210642106531065420654 2
054321642	053210642106532	054321643210532106421 6	064210653106542065431654320 53	05432164321053210642106531065420654 3
064210643	054321643210531	064210653106542065431 0	065310654206543165432054321 65	06421065310654206543165432054321643 1
064210654	064210653106543	065310654206543165432 1	065420654316543205432164321 06	06421065310654206543165432054321643 2
065420653	065310654206542	065420654316543205432 0	065431654320543216432105321 05	06531065420654316543205432164321053 1
065431653	065420654316542	065431654320543216432 0	0532106421065310654206543165 42	06531065420654316543205432164321053 2
0532105320	065431654320542	0532106421065310654206 4	0543216432105321064210653106 53	06542065431654320543216432105321064 2
0532105321	0532106421065316	0543216432105321064210 5	0642106531065420654316543205 42	06542065431654320543216432105321064 3
0532106420	0543216432105320	0642106531065420654316 4	0653106542065431654320543216 42	06543165432054321643210532106421065 3
0543205431	0642106531065421	0653106542065431654320 6	0654206543165432054321643210 54	06543165432054321643210532106421065 4

Using the tree traversal algorithm, we can prove

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

Proof. Consider trying to generate an infinite word $\mathbf{w}$ over $\mathbb{Z}$ starting with 0, subject to two conditions: (1) avoiding the pattern $x\sigma^i(x)$ 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 $x\sigma^i(x)$ 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 $x\sigma^i(x)$, except now i is reduced modulo $j+3$. In order for T to correctly represent a proof that the pattern $x\sigma^i(x)$ cannot be avoided for $0 \leq i \leq j$ we must check that $i \bmod (j+3) \in \{0, 1, \ldots, j\}$ 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 $a_1 a_2 \cdots a_j$ represents the leaf $(-a_1, -a_2, \ldots, -a_j)$. ■

Figure 3: Leaves of the tree giving the proof of Theorem 9.

0246	024568AC	01356789A	0234568ABCE
0235	024568AB	01235789B	01356789BDF
0134	0245679A	01234689B	01246789ACD
02457	02456789	0245679BCE	01235789ABC
01357	0234689B	0245679BCD	01234689ABD
01245	0234689A	0245678ACE	0245678ACDEG
023467	0234579B	0234579ABD	0245678ACDEF
013568	02345689	0234579ABC	0234568ABCDF
012468	0135679B	0234568ABD	0234568ABCDE
012356	0135679A	0135678ACE	01356789BDEG
012345	0124678A	0135678ACD	01356789BDEF
0245689	0123578A	01356789BC	01246789ACEG
023468A	0123468A	01246789BD	01246789ACEF
0234578	0245679BD	01246789BC	01235789ABDF

0234567	0245678AB	01246789AB	01235789ABDE
0124679	0234579AC	01235789AC	01234689ABCE
0123579	0234568AC	01234689AC	01234689ABCD
0123467	0135678AB	0245678ACDF	

We now show it is possible to simultaneously avoid the patterns xx, $x\sigma(x)$, $\ldots$, $x\sigma^j(x)$ on Σ_{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* Card $A \leq k-3$. *Then it is possible to simultaneously avoid the patterns $\{x\sigma^a(x) \ : \ a \in A\}$ over Σ_k.*

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

Lemma 11 *Let $\mathbf{w} = a_1a_2a_3\cdots$ be any squarefree word over Σ_3. Then the word $a_1a_1a_2a_2a_3a_3\cdots$ avoids the pattern ycy for $y \in \Sigma_3^+$ and $c \in \Sigma_3$.*

Proof. Suppose $y = b_1b_2\cdots b_k$, and the pattern ycy occurs in $\mathbf{z} = a_1a_1a_2a_2a_3a_3\cdots$. There are three cases to consider, depending on $|y|$ and where y starts in $\mathbf{z}$.

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

$$\begin{array}{ccccccccc} b_1 & b_2 & \cdots & b_{2j} & c & b_1 & b_2 & \cdots & b_{2j} \\ & & & & = & & & & \\ a_i & a_i & \cdots & a_{i+j-1} & a_{i+j} & a_{i+j} & a_{i+j+1} & \cdots & a_{i+2j} \end{array}$$

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

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

$$\begin{array}{ccccccccc} b_1 & b_2 & \cdots & b_{2j} & c & b_1 & b_2 & \cdots & b_{2j} \\ & & & & = & & & & \\ a_i & a_{i+1} & \cdots & a_{i+j} & a_{i+j} & a_{i+j+1} & a_{i+j+1} & \cdots & a_{i+2j} \end{array}$$

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

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

$$\begin{array}{ccccccccccc} b_1 & b_2 & \cdots & b_{2j} & b_{2j+1} & c & b_1 & b_2 & \cdots & b_{2j} & b_{2j+1} \\ & & & & & = & & & & & \\ a_i & a_i & \cdots & a_{i+j-1} & a_{i+j} & a_{i+j} & a_{i+j+1} & a_{i+j+1} & \cdots & a_{i+2j} & a_{i+2j+1} \end{array}$$

or

$$\begin{array}{cccccccccc} b_1 & b_2 & \cdots & b_{2j} & b_{2j+1} & c & b_1 & b_2 & \cdots & b_{2j} & b_{2j+1} \\ & & & & & = & & & & & \\ a_i & a_{i+1} & \cdots & a_{i+j} & a_{i+j} & a_{i+j+1} & a_{i+j+1} & a_{i+j+2} & \cdots & a_{i+2j+1} & a_{i+2j+1} \end{array}$$

In either case we find

$$\begin{aligned} a_i = & \quad b_1 & = a_{i+j+1} \\ a_{i+1} = & \quad b_3 & = a_{i+j+2} \\ & \quad \vdots & \\ a_{i+j} = & \; b_{2j+1} & = a_{i+2j+1} \end{aligned}$$

It follows that $a_i a_{i+1} \cdots a_{i+j} = a_{i+j+1} a_{i+j+2} \cdots a_{i+2j+1}$ and so **w** contains the square $a_i a_{i+1} \cdots a_{i+2j+1}$, 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 ≥ 7 over $\{0, 1\}$ contains an occurrence of ycy.

Now we can complete the proof of Theorem 10. Let $\mathbf{x}$ be any squarefree word over $\{0, 1, 2\}$. Since 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 $\varphi : \Sigma_{j+4}^* \to \Sigma_{j+4}^*$ defined as follows:

$$\begin{aligned} 0 &\to dd \\ 1 &\to ee \\ 2 &\to ff \end{aligned}$$

We claim $S(\varphi(\mathbf{x}))$ avoids the patterns xx, $x\sigma(x)$, $\ldots$, $x\sigma^j(x)$.

Let $\mathbf{v} = S(\varphi(\mathbf{x}))$. Then $\Delta(\mathbf{v}) = \varphi(\mathbf{x})$ clearly avoids ycy by Lemma 11, and it also avoids all the letters in A by construction. Then by Lemma 6, $\mathbf{v}$ 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 Σ_{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 Σ_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 Σ_{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 $x\sigma^{-j}(x)$, $\ldots$, $x\sigma^{-1}(x)$, xx, $x\sigma(x)$, $\ldots$, $x\sigma^{j}(x)$ over Σ_{2j+4}.

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

To prove that one cannot simultaneously avoid the patterns $x\sigma^{-j}(x)$, $\ldots$, $x\sigma^{-1}(x)$, xx, $x\sigma(x)$, $\ldots$, $x\sigma^{j}(x)$ over Σ_{2j+3}, we use the tree traversal algorithm. Then every word of length ≥ 8 over Σ_{2j+3} contains an occurrence of $x\sigma^{l}(x)$ 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 $t = j + 1$. ■

Figure 4: Leaves of the tree giving a proof of Theorem 13.

(0, -t, -2t, -3t)	(0, -t, -2t, -t, 0, t)	(0, -t, -2t, -t, 0, -t, -2t, -3t)
(0, -t, 0, -t)	(0, -t, 0, t, 0, t)	(0, -t, -2t, -t, 0, -t, -2t, -t)
(0, t, 0, t)	(0, t, 0, -t, 0, -t)	(0, -t, 0, t, 0, -t, 0, -t)
(0, t, 2t, 3t)	(0, t, 2t, t, 0, -t)	(0, -t, 0, t, 0, -t, 0, t)
(0, -t, -2t, -t, -2t)	(0, -t, -2t, -t, 0, -t, 0)	(0, t, 0, -t, 0, t, 0, -t)
(0, -t, 0, t, 2t)	(0, -t, 0, t, 0, -t, -2t)	(0, t, 0, -t, 0, t, 0, t)
(0, t, 0, -t, -2t)	(0, t, 0, -t, 0, t, 2t)	(0, t, 2t, t, 0, t, 2t, t)
(0, t, 2t, t, 2t)	(0, t, 2t, t, 0, t, 0)	(0, t, 2t, t, 0, t, 2t, 3t)

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

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

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

Proposition 14 *Over* Σ_2*, every word of length* ≥ 8 *contains a subword of the form* $x\sigma^{i}(x)$ *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. ■

Figure 5: Leaves of the tree giving a proof of Proposition 14.

0000	000101	00010000
0011	001001	00010001
0101	010000	00100010
0110	011100	00100011

00011	0001001	01000100
00101	0010000	01000101
01001	0100011	01110110
01111	0111010	01110111

However, it is possible to weakly avoid the patterns $x\sigma^i(x)$ for all $i \geq 0$ over Σ_3. Let $\mathbf{w}$ be any squarefree word over $\{0,1,2\}$, and consider the morphism f which maps

$$\begin{aligned} 0 &\to 00 \\ 1 &\to 10 \\ 2 &\to 20. \end{aligned}$$

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

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

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

$$\begin{aligned} z &= \overbrace{d_1 0 d_2 0 \cdots d_j 0}^{x} \mid \overbrace{d_{j+1} 0 \cdots d_{2j} 0}^{\sigma^i(x)} \\ z &= 0 d_1 0 d_2 \cdots 0 d_j \mid 0 d_{j+1} \cdots 0 d_{2j} \end{aligned}$$

where $d_t = c_{t+k}$ 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 $d_t = d_{j+t}$ 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 $\mathbf{w}$ was squarefree.

Case 2: $|x| \equiv 1 \pmod 2$.

$$\begin{aligned} z &= \overbrace{d_1 0 d_2 \cdots}^{x} \mid \overbrace{0 d_j 0 \cdots}^{\sigma^i(x)} \\ z &= 0 d_1 0 \cdots \mid d_j 0 d_{j+1} \cdots \end{aligned}$$

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 $\mathbf{w}$, a contradiction. ■

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

Theorem 16 *If $k = 4$, one can, over Σ_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 Σ_4 as follows: let $\mathbf{w}$ be any squarefree word over $\{1, 2, 3\}$, and consider the morphism f which maps

$$1 \to 10$$
$$2 \to 20$$
$$3 \to 30.$$

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

On the other hand, the tree traversal algorithm shows that over Σ_3, any word of length ≥ 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.

0101	010202	01020101
0120	012102	01020102
0202	020101	01210120
0210	021201	01210121
01021	0102012	02010201
01212	0121010	02010202
02012	0201021	02120210
02121	0212020	02120212

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

1. D. A. Bean, G. Ehrenfeucht, and G. McNulty. Avoidable patterns in strings of symbols. *Pacific J. Math.* **85** (1979), 261–294.
2. J. Berstel. *Axel Thue's Papers on Repetitions in Words: a Translation.* Number 20 in Publications du Laboratoire de Combinatoire et d'Informatique Mathématique. Université du Québec à Montréal, February 1995.
3. T. C. Brown. Is there a sequence on four symbols in which no two adjacent segments are permutations of one another? *Amer. Math. Monthly* **78** (1971), 886–888.
4. J. Cassaigne. Unavoidable binary patterns. *Acta Informatica* **30** (1993), 385–395.
5. J. D. Currie. Open problems in pattern avoidance. *Amer. Math. Monthly* **100** (1993), 790–793.

6. A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **1** (1912), 1–67. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 413–478.