International Journal of Foundations of Computer Science


# UNARY PATTERNS WITH INVOLUTION

Bastian Bischoff
*Institute for Formal Methods in Computer Science, Universität Stuttgart*
*Universitätsstr. 38, D-70569 Stuttgart, Germany*
*bastian.bischoff@googlemail.com*

James Currie*
*Department of Mathematics, University of Winnipeg*
*Winnipeg, Manitoba, Canada R3B 2E9*
*ac.gepinniwu@eirruc.j*

Dirk Nowotka†
*Department of Computer Science, Christian-Albrechts-Universität zu Kiel*
*D-24098 Kiel, Germany*
*dn@informatik.uni-kiel.de*

Received (Day Month Year)
Accepted (Day Month Year)
Communicated by (xxxxxxxxxx)

An infinite word **w** avoids a pattern $p$ with the involution if there is no substitution for the variables in $p$ and no involution on substituted variables such that the resulting word is a factor of **w**. An avoidance index of pattern $p$ is the smallest alphabet size for which a word exists such that $p$ is avoided. A pattern is called unary, if only one variable occurs in it. In this paper, we give the avoidance indices for all unary patterns with involution.

## 1. Introduction

In this article, we are concerned with avoidable patterns with involution. An involution $f$ is a mapping such that $f^2$ is the identity. Consider morphic, where $f(uv) = f(u)f(v)$, and antimorphic involutions, where $f(uv) = f(v)f(u)$. We fully characterize the avoidance indices for all unary patterns with both morphic and antimorphic involutions. As it turns out, unary patterns exhibit the same avoidance indices for both morphic and antimorphic involutions. The arguments for showing those are different however.

The subject of this article draws some motivation from applications in biology where the Watson-Crick complement corresponds to an antimorphic involution in

*The author is supported by an NSERC Discovery grant.
†The author is supported by the DFG Heisenberg grant 590179.

our case; see for example [2]. Our considerations are more general, however, by considering any alphabet size and also morphic involutions.

## 2. Preliminaries

Our notation is guided by what is commonly found in the literature, see for example the first chapter of [3] as a reference. Let $\Sigma_k = \{0, 1, \ldots, k-1\}$ be a finite alphabet of $k$ *letters* for some $k \geq 1$. Let $\Sigma_k^*$ denote all *finite* and $\Sigma_k^\omega$ denote all (right-) *infinite* words over $\Sigma_k$. Let $\varepsilon$ denote the empty word. The $i$-th letter of a word $w$ is denoted by $w_{[i]}$ that is, $w = w_{[1]}w_{[2]}\cdots w_{[n]}$ for finite words, and $\mathbf{w} = \Pi_{i=1}^{\infty}\mathbf{w}_{[i]}$ for infinite words. The length $n$ of $w$ is denoted by $|w|$ as usual. The number of occurrences of a letter $a$ in $w$ is denoted by $|w|_a$. Let $w^R = w_{[n]}w_{[n-1]}\cdots w_{[1]}$ denote the reversal of $w$. Let $w = uv$, then $u$ is called *prefix* of $w$, denoted by $u \leq_{\mathrm{p}} w$, and *proper prefix*, if $v = \varepsilon$, denoted by $u <_{\mathrm{p}} w$.

Besides $\Sigma_k$ we need one more finite set $E$ of symbols. The elements of $E$ are called *variables* and we usually denote them by $\alpha$, $\beta$, or $\gamma$. Words in $E^*$ are called *patterns*. For example $\alpha\beta\alpha \in E^*$ is a pattern consisting of the variables $\alpha$ and $\beta$ in $E$. We assign to every pattern a *pattern language* over an alphabet $\Sigma_k$. This language contains every word that can be generated by substituting all variables in the pattern by non-empty words in $\Sigma_k^*$. For example the pattern language of the pattern $\alpha\alpha$ over $\Sigma_2$ is $\{00, 11, 0000, 0101, 1010, 1111, \ldots\}$, the set of all squares in $\Sigma_2^*$.

We say that a word $\mathbf{w}$ *avoids* a pattern, if no factor of $\mathbf{w}$ exists that is in the pattern language. On the other hand, if a factor of $\mathbf{w}$ is an element of the pattern language, we say $\mathbf{w}$ *contains* the pattern. If for a given pattern $p \in E^*$ and an alphabet $\Sigma_k$ a word $\mathbf{w} \in \Sigma_k^\omega$ exists that avoids $p$, then we say that $p$ is *$k$-avoidable*. Otherwise we call $p$ *$k$-unavoidable*. We call $k \in \mathbb{N}$ the *avoidance index* $\mathcal{V}(p)$ of a pattern $p \in E^*$, if $p$ is $k$-avoidable and $k$ is minimal. If no such $k$ exists, we define $\mathcal{V}(p) = \infty$. A pattern is called *unary*, if it contains exactly one variable. In this paper, we restrict ourselves to unary patterns.

Let $f\colon \Sigma_2^* \to \Sigma_2^*$ with $0 \mapsto 01$ and $1 \mapsto 10$. The fix point $\mathbf{t} = \lim_{k\to\infty} f^k(a)$ exists and is called *Thue–Morse word*. The following result is a classical one.

**Theorem 1 ([4, 5])** *The Thue–Morse word* $\mathbf{t}$ *avoids the patterns* $\alpha\alpha\alpha$ *and* $\alpha\beta\alpha\beta\alpha$.

## 3. Patterns with Involution

For introducing patterns with involution, we extend the set of pattern variables $E$ to $E_\theta$ by adding $\theta(\alpha)$ for all variables $\alpha \in E$. Given a morphic or antimorphic involution, we build the corresponding pattern language by replacing the variables by non-empty words and, for variables of the form $\theta(\alpha)$, by applying the involution after the substitution.

There are only two morphic involutions over $\Sigma_2$: the identity and the complement, where $f(0) = 1$ and $f(1) = 0$.

We get the pattern language $\{01, 10, 0011, 0110, 1001, 1100, \ldots\}$ for $\alpha\,\theta(\alpha)$ and the complement. Every word in $\Sigma_2^\omega \setminus (0^\omega \cup 1^\omega)$ contains the pattern $\alpha\,\theta(\alpha)$ for the complement.

**Observation 2.** *Let $f$ be a morphic or antimorphic involution and distinct from the identity or reversal mapping. Then every pattern that contains variables of the form $\alpha$ and $\theta(\alpha)$, is avoidable for $f$.*

Indeed, since $f$ is not the identity or reversal mapping, a letter $a \in \Sigma_k$ with $f(a) \neq a$ exists. Therefore $\mathbf{w} = a^\omega$ avoids every pattern that includes variables $\alpha$ and $\theta(\alpha)$.

Because of this observation we do not have to examine, if patterns are avoidable or unavoidable for a given involution. Changing the point of view, we look at all morphic or all antimorphic involutions $\Sigma_k^* \to \Sigma_k^*$ at the same time for a given pattern $p \in E_\theta^*$.

**Definition 3.** *A pattern $p \in E_\theta^*$ is (anti-) morphically* $\theta$-avoidable *in $\Sigma_k$, if there exists an infinite word $\mathbf{w} \in \Sigma_k^\omega$ such that $\mathbf{w}$ does not contain a factor that is in the pattern language of $p$ for any (anti-) morphic involution.*

**Definition 4.** *Let $p \in E_\theta^*$ be a pattern. The minimal $k \in \mathbb{N}$ such that $p$ is morphically (antimorphically) $\theta$-avoidable in $\Sigma_k$, is called the morphic (antimorphic) $\theta$-avoidance index $\mathcal{V}_\mathrm{m}(p)$ ($\mathcal{V}_\mathrm{a}(p)$) of $p$, if such a $k$ exists. If $p$ is unavoidable, then we define $\mathcal{V}_\mathrm{m}(p) = \infty$ ($\mathcal{V}_\mathrm{a}(p) = \infty$).*

We establish the first facts about avoidance of pattern $\alpha\,\theta(\alpha)\,\alpha$.

**Lemma 5.** $\mathcal{V}_\mathrm{m}(\alpha\,\theta(\alpha)\,\alpha) > 2$.

**Proof.** Suppose there exists a word $\mathbf{w} \in \Sigma_2^\omega$ that avoids $\alpha\,\theta(\alpha)\,\alpha$ for all morphic involutions. This word must not contain 000, 111, 010, or 101 as a factor. Without loss of generality $\mathbf{w}$ begins with 0.

Assume $\mathbf{w}$ begins with 01. Then this prefix must be followed by 1, $011 <_\mathrm{p} \mathbf{w}$. The next letter must be 0, the fifth must be 0 too. So we have $01100 <_\mathrm{p} \mathbf{w}$. If the following letter is 0, then 000 is a factor of $\mathbf{w}$. So the next letter must be 1. But for the morphic involution $f$ with $0 \mapsto 1$ and $1 \mapsto 0$ the word $01f(01)01$ is a factor of $\mathbf{w}$; a contradiction. The argument for the case $00 \leq_\mathrm{p} \mathbf{w}$ is analogous. □

The proof of the following lemma is analogous to the previous one.

**Lemma 6.** $\mathcal{V}_\mathrm{a}(\alpha\,\theta(\alpha)\,\alpha) > 2$.

## 4. Patterns of length 3

In this section, we establish the following theorem.

**Theorem 7.** *Let $p \in E_\theta^3$ be a pattern of length 3 in which $\alpha$ and $\theta(\alpha)$ both occur. Then*

$$\mathcal{V}_\mathrm{m}(p) = \mathcal{V}_\mathrm{a}(p) = 3 \ .$$

Recall that $\mathcal{V}_\mathrm{m}(\alpha^3) = \mathcal{V}_\mathrm{a}(\alpha^3) = \mathcal{V}_\mathrm{m}(\theta(\alpha)^3) = \mathcal{V}_\mathrm{a}(\theta(\alpha)^3) = 2$ due to Theorem 1.

We start with the morphic case.

**Lemma 8.** $\mathcal{V}_\mathrm{m}(\alpha\theta(\alpha)\alpha) = 3$.

**Proof.** Let $\mathbf{x} \in \Sigma_3^\omega$ be the image of the Thue-Morse word $\mathbf{t}$ by the morphism defined by $0 \mapsto 0021$ and $1 \mapsto 0221$. We will show, that $\mathbf{x}$ avoids the pattern $\alpha\theta(\alpha)\alpha$ for all morphic involutions. For better readability, we define $x = 0021$ and $y = 0221$.

We assume that there exists a morphic involution $f$ and a substitution $u$ for $\alpha$, such that $uf(u)u$ is a factor of $\mathbf{x}$. The proof is by contradiction. First, we examine the possibilities of replacing the variable $\alpha$ by words $u \in \Sigma_3^+$ of length $|u| < 7$. The word $u\,f(u)\,u$ has a maximal length of 18. Therefore there must exist a morphic involution so that $u\,f(u)\,u$ is a factor of a word $w \in \{\,x, y\,\}^6$. Because of Theorem 1, none of the words $xxx$, $yyy$, $xyxyx$ and $yxyxy$ is a factor of $w$. A computer program can easily check these finite possibilities with the result, that no words $u$ and $w$ exist, which fulfill the conditions. Now we assume $\alpha$ gets replaced by a word $u \in \Sigma_3^+$ with $|u| \geq 7$. Then, the word $u$ contains 0021 or 0221. Without loss of generality, $u$ contains 0021. Therefore, $f(u)$ contains the factor $f(002) = f(0)\,f(0)\,f(2)$. In addition $f(u)$, and hence, $f(0)\,f(0)\,f(2)$, is a factor of $\mathbf{x}$. There are only two possibilities for two succeeding identical letters in $\mathbf{x}$. Either these letters are two letters 2 followed by the letter 1, but then $f$ is not an involution and we have a contradiction, or two letters 0 are followed by the letter 2. This implies that $f$ is the identity since $u\,f(u)\,u$ is a factor of $\mathbf{x}$. Furthermore, this implies that $|u| = 4 \cdot k$ for $k \in \mathbb{N}$. This is visualized in Fig. 1, where $w_i, w_{i'}, w_{i''} \in \{\,x, y\,\}$ holds for all $0 \leq i \leq k$. If the word $(w_0)_{[2]}(w_0)_{[3]}(w_0)_{[4]}$ or $(w_0)_{[1]}(w_0)_{[2]}(w_0)_{[3]}(w_0)_{[4]} = w_0$ is a prefix of the first $u$ in Fig. 1, then the following equations apply:

$$\begin{array}{ccccc} w_0 & = & w_{0'} & = & w_{0''} \\ w_1 & = & w_{1'} & = & w_{1''} \\ \vdots & & \vdots & & \vdots \\ w_{k-1} & = & w_{k-1'} & = & w_{k-1''} \end{array}$$

The word $w_0w_1\dots w_{k-1}\,w_{0'}w_{1'}\dots w_{k-1'}\,w_{0''}w_{1''}\dots w_{k-1''} = (w_0w_1\dots w_{k-1})^3$ is a factor of $\mathbf{x}$. Because of $w_i \in \{\,x, y\,\}$ for all $0 \leq i \leq k-1$, this is a contradiction to Theorem 1. On the other hand, if only $(w_0)_{[3]}(w_0)_{[4]}$ or $(w_0)_{[4]}$ is a prefix of $u$, then $w_0 \neq w_{0'}$ is possible. But in this case $(w_{k''})_{[1]}(w_{k''})_{[2]}$ or $(w_{k''})_{[1]}(w_{k''})_{[2]}(w_{k''})_{[3]}$ is a suffix of the third $u$. This implies

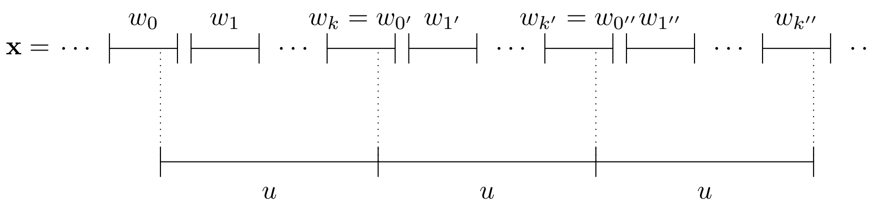


Figure 1. Part of $\mathbf{x}$ to illustrate the factor $uuu$

$$\begin{array}{ccccc} w_1 & = & w_{1'} & = & w_{1''} \\ w_2 & = & w_{2'} & = & w_{2''} \\ \vdots & & \vdots & & \vdots \\ w_k & = & w_{k'} & = & w_{k''} \end{array}$$

and $w_1 w_2 \ldots w_k\, w_{1'} w_{2'} \ldots w_{k'}\, w_{1''} w_{2''} \ldots w_{k''} = (w_1 w_2 \ldots w_k)^3$ is a factor of $\mathbf{x}$. Again, this is a contradiction to Theorem 1. The theorem follows from Lemma 5.□

The result of Lemma 8 is also valid for the antimorphic case.

**Lemma 9.** $\mathcal{V}_{\mathrm{a}}(\alpha\theta(\alpha)\alpha) = 3$.

**Proof.** This proof follows the proof of the previous theorem. Let $\mathbf{x} \in \Sigma_3^\omega$ be the image of the Thue-Morse word $\mathbf{t}$ by the morphism $0 \mapsto 00112$ and $1 \mapsto 00221$. We will show that $\mathbf{x}$ avoids the pattern $\alpha\,\theta(\alpha)\,\alpha$ for all antimorphic involutions. For better readability, we define $x = 00112$ and $y = 00221$.

We assume that there exist an antimorphic involution $f$ and a substitution of $\alpha$ by a word $u \in \Sigma_3^+$ such that $u\,f(u)\,u$ is a factor of $\mathbf{x}$. First we suppose that $|u| < 9$ holds. The word $u\,f(u)\,u$ then has a maximal length of 24 and $u\,f(u)\,u$ is factor of a word $w \in \{x, y\}^6$. The words $xxx$, $yyy$, $xyxyx$, and $yxyxy$ must not be a factor of $w$ because of Theorem 1. A computer program can check these finite possibilities with the result that no words $u$ and $w$ exist that fulfill these conditions for an antimorphic involution $f$. So, $|u| \geq 9$ must hold and $u$ contains at least one word $x$ or $y$ completely. We now look at the first $u$ of the factor $u\,f(u)\,u$ of $\mathbf{x}$. Let $w_1 w_2'$ be a suffix of $u$ with $w_1, w_2 \in \{x, y\}$, $w_2 = w_2' w_2''$ and $|w_2'| < 5$. We get Fig. 2 where $w_3, w_4 \in \{x, y\}$. Without loss of generality, let $w_1 = x = 00112$. Then $f(u)$ and therefore $w_2 w_3 w_4$ contains the word $f(00112) = f(2)\,f(1)\,f(1)\,f(0)\,f(0)$ of length 5 as a factor. Hence we look at the following words:

$$\begin{aligned} xx &= 00112\,00112 \\ xy &= 00112\,00221 \\ yx &= 00221\,00112 \\ yy &= 00221\,00221\ . \end{aligned}$$

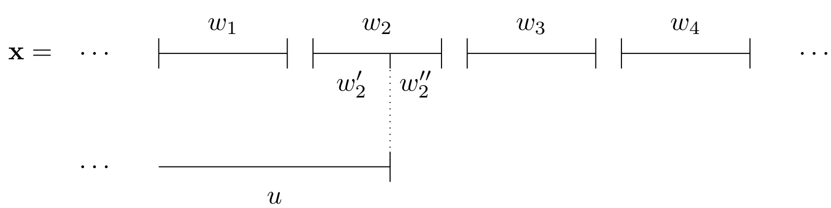


Figure 2. Part of $\mathbf{x}$ and the factor $u$ of $\mathbf{x}$

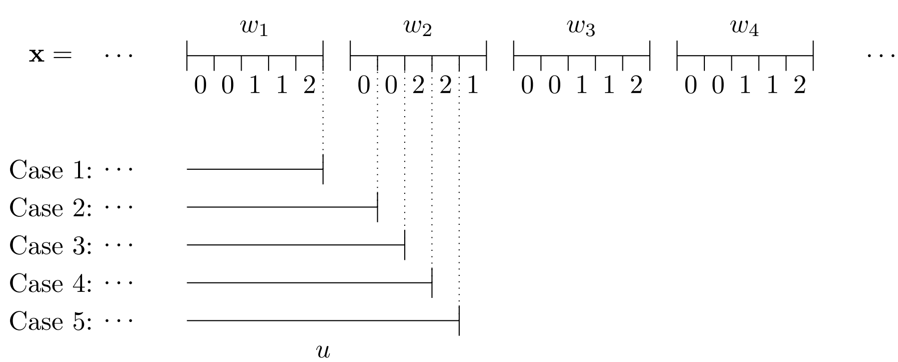


Figure 3. Illustration of possible positions of the factor $u$ of $\mathbf{x}$

Only $xx$ contains $f(2)\,f(1)\,f(1)\,f(0)\,f(0)$ for the antimorphic involution $f$ with $0 \mapsto 1$, $1 \mapsto 0$, and $2 \mapsto 2$. Because of $w_1 = x$, the equation $w_2w_3 = xx$ is a contradiction to Theorem 1. The case $w_2w_3w_4 = yxx$ remains. Now there are five possibilities for the position of $u$, see Fig. 3. It is easy to check that in all five cases $f(u) \leq_{\mathrm{p}} w_2''w_3w_4$ respectively $w_2''w_3w_4 \leq_{\mathrm{p}} f(u)$ does not hold. So our assumption that there exists an antimorphic involution $f$ and a word $u \in \Sigma_3^+$ with $u\,f(u)\,u$ is a factor of $\mathbf{x}$, was wrong. The theorem follows with Lemma 6. □

Patterns similar to $\alpha\,\theta(\alpha)\,\alpha$ are considered next.

For the next lemma we need a further definition. Let $p \in E_\theta^*$ be a pattern consisting of variables of the form $\alpha$ and $\theta(\alpha)$ and $p'$ be the pattern that we get, when all variables $\alpha$ and $\theta(\alpha)$ in $p$ are switched. We call $p' \in E_\theta^*$ the *$\theta$-complementary* pattern of $p$. For example the $\theta$-complementary pattern of $\alpha\,\alpha\,\theta(\alpha)\,\beta$ is $\theta(\alpha)\,\theta(\alpha)\,\alpha\,\theta(\beta)$.

**Lemma 10.** *Let $p \in E_\theta^*$ be a pattern and $p' \in E_\theta^*$ be the $\theta$-complementary pattern of $p$. Then $\mathcal{V}_{\mathrm{a}}(p) = \mathcal{V}_{\mathrm{a}}(p')$ and $\mathcal{V}_{\mathrm{m}}(p) = \mathcal{V}_{\mathrm{m}}(p')$.*

**Proof.** Firstly, we show $\mathcal{V}_{\mathrm{m}}(p) = \mathcal{V}_{\mathrm{m}}(p')$. For better readability, we replace the variable $\alpha$ in the pattern $p'$ by $\alpha'$ and $\theta(\alpha)$ by $\theta(\alpha')$. We assume a word $\mathbf{x} \in \Sigma^\omega$ contains the pattern $p$ for a morphic involution $f$ and a substitution of $\alpha$ by $u \in \Sigma^+$.

Then $\mathbf{x}$ contains the pattern $p'$ for the same morphic involution $f$ by substituting $\alpha'$ by $f(u)$. The equation $\mathcal{V}_{\mathrm{m}}(p) = \mathcal{V}_{\mathrm{m}}(p')$ follows. The proof of $\mathcal{V}_{\mathrm{a}}(p) = \mathcal{V}_{\mathrm{a}}(p')$ is identical. □

Consider the following $\theta$-free patterns; see [1].

**Theorem 11.** *The patterns $\alpha\alpha$, $\alpha\alpha\beta$, $\beta\alpha\alpha$, $\alpha\alpha\beta\alpha$, $\alpha\beta\beta\alpha$, $\alpha\alpha\beta\beta$, $\alpha\beta\alpha\beta$, $\alpha\alpha\beta\alpha\alpha$, and $\alpha\alpha\beta\alpha\beta$ are 2-unavoidable and 3-avoidable.*

**Lemma 12.** *Let $p \in E_\theta^*$ be a pattern that contains the variables $\alpha$ and $\theta(\alpha)$, and $p$ contains no other variable of the form $\theta(\gamma)$. Let $p'$ be the pattern when all occurrences of $\theta(\alpha)$ in $p$ are replaced by $\alpha$. The pattern $p''$ is obtained when all occurrences of $\theta(\alpha)$ in $p$ are replaced by a new variable $\beta$.*

*Then $\mathcal{V}(p') \leq \mathcal{V}_{\mathrm{m}}(p) \leq \mathcal{V}(p'')$ and $\mathcal{V}_{\mathrm{a}}(p) \leq \mathcal{V}(p'')$.*

**Proof.** The relation $\mathcal{V}(p') \leq \mathcal{V}_{\mathrm{m}}(p)$ holds, since the morphic $\theta$-avoidance index involves all morphic involutions, including the identity mapping. Suppose there exists a word $\mathbf{x} \in \Sigma_k^\omega$ that avoids the pattern $p''$.Then this word also avoids the pattern $p$ for all morphic and antimorphic involutions. Therefore the relations $\mathcal{V}_{\mathrm{m}}(p) \leq \mathcal{V}(p'')$ and $\mathcal{V}_{\mathrm{a}}(p) \leq \mathcal{V}(p'')$ hold. □

**Lemma 13.** $\mathcal{V}_{\mathrm{a}}(\alpha\,\alpha\,\theta(\alpha)) = \mathcal{V}_{\mathrm{m}}(\alpha\,\alpha\,\theta(\alpha)) = 3.$

**Proof.** According to Theorem 11 the equation $\mathcal{V}(\alpha\,\alpha\,\beta) = 3$ holds. Lemma 12 implies $\mathcal{V}_{\mathrm{a}}(\alpha\,\alpha\,\theta(\alpha))$, $\mathcal{V}_{\mathrm{m}}(\alpha\,\alpha\,\theta(\alpha)) \leq 3$. We show by contradiction that it holds that $\mathcal{V}_{\mathrm{a}}(\alpha\,\alpha\,\theta(\alpha)) \neq 2$. The proof for the relation $\mathcal{V}_{\mathrm{m}}(\alpha\,\alpha\,\theta(\alpha)) \neq 2$ is analogous. Assuming there exists a word $\mathbf{x} \in \Sigma_2^\omega$ such that it avoids the pattern $\alpha\,\alpha\,\theta(\alpha)$ for all antimorphic involutions. Then $\mathbf{x}$ contains neither 00 nor 11 as a factor. Without loss of generality $\mathbf{x}$ begins with the letter 0. It follows that $\mathbf{x} = (01)^\omega$. But $\mathbf{x} = (01)^\omega$ contains the pattern $\alpha\,\alpha\,\theta(\alpha)$ for $\alpha = 01$ and the antimorphic involution defined by $0 \mapsto 1$ and $1 \mapsto 0$. This is a contradiction to our assumption. Therefore $\mathcal{V}_{\mathrm{a}}(\alpha\,\alpha\,\theta(\alpha)) \neq 2$ holds and analogously $\mathcal{V}_{\mathrm{m}}(\alpha\,\alpha\,\theta(\alpha)) \neq 2$. We get $\mathcal{V}_{\mathrm{a}}(\alpha\,\alpha\,\theta(\alpha)) = \mathcal{V}_{\mathrm{m}}(\alpha\,\alpha\,\theta(\alpha)) = 3$. □

**Lemma 14.** $\mathcal{V}_{\mathrm{a}}(\theta(\alpha)\,\alpha\,\alpha) = \mathcal{V}_{\mathrm{m}}(\theta(\alpha)\,\alpha\,\alpha) = 3.$

**Proof.** The proof is analogous to the proof of Lemma 13. □

**Corollary 15.**

*(1)* $\mathcal{V}_{\mathrm{m}}(\theta(\alpha)\,\alpha\,\theta(\alpha)) = \mathcal{V}_{\mathrm{a}}(\theta(\alpha)\,\alpha\,\theta(\alpha)) = 3$,
*(2)* $\mathcal{V}_{\mathrm{m}}(\theta(\alpha)\,\theta(\alpha)\,\alpha) = \mathcal{V}_{\mathrm{a}}(\theta(\alpha)\,\theta(\alpha)\,\alpha) = 3$, *and*
*(3)* $\mathcal{V}_{\mathrm{m}}(\alpha\,\theta(\alpha)\,\theta(\alpha)) = \mathcal{V}_{\mathrm{a}}(\alpha\,\theta(\alpha)\,\theta(\alpha)) = 3$.

Indeed, (1) follows from Lemma 8, 9 and 10, (2) follows from 13, and (3) follows from Lemma 14.

## 5. Patterns longer than 3

In this section, we establish the following theorem.

**Theorem 16.** *Let $p \in E_\theta^*$ be a pattern longer than 3. Then*

$$\mathcal{V}_{\mathrm{m}}(p) = \mathcal{V}_{\mathrm{a}}(p) = 2 \ .$$

The avoidance indices are clearly 2 when $\alpha\alpha\alpha$ is a factor of $p$ (Theorem 1). We consider words $p \in E_\theta^4$ where $\alpha\alpha\alpha$ is not a factor. Interchanging $\alpha$ and $\theta(\alpha)$'s if necessary, assume that $|p|_\alpha \geq |p|_{\theta(\alpha)}$. Since $\alpha\alpha\alpha$ is not to be a factor of $p$, either $|p|_{\theta(\alpha)} = 1$ or $|p|_{\theta(\alpha)} = 2$. In the first case, our word $p$ is $\alpha\alpha\theta(\alpha)\alpha$ or $\alpha\theta(\alpha)\alpha\alpha$. Since avoidance indices are preserved under reversal, we need only consider the case $p = \alpha\alpha\theta(\alpha)\alpha$ here. If $|p|_{\theta(\alpha)} = 2$, ignoring reversals, we consider $\alpha\theta(\alpha)\alpha\theta(\alpha)$, $\theta(\alpha)\alpha\alpha\theta(\alpha)$, and $\alpha\alpha\theta(\alpha)\theta(\alpha)$. For each of these $p \in E_\theta^4$ we will show that both avoidance indices are 2. This amounts to constructing an infinite binary word with no factor $xxf(x)x$ ($xf(x)xf(x)$, $f(x)xxf(x)$, $xxf(x)f(x)$) where $x$ is non-empty and $f$ is a morphic ($f$ is an antimorphic) involution.

Let us start with the morphic case. Let $\mathbf{w}$ be the infinite word

$$\mathbf{w} = \Pi_{i=0}^{\infty} 0^2 1^{\mathbf{t}[i]+2} \ .$$

We see that $\mathbf{w}$ is concatenated from blocks of two 0's alternated with blocks of either two or three 1's.

**Lemma 17.** *Word $\mathbf{w}$ morphically avoids the pattern $\alpha\alpha\theta(\alpha)\alpha$.*

**Proof.** Suppose there exist a substitution $x$ for $\alpha$ and a morphic involution $f$ such that $xxf(x)x$ is a factor of $\mathbf{w}$.

If $|x|_0 = 0$, then $x = 1^m$ for some $m$. If $f$ is the identity, this makes 1111 a factor of $\mathbf{w}$, which is impossible. If $f$ is the complement, then $m \leq 2$, since $f(x) = 0^m$ is a factor of $\mathbf{w}$. Then, however, $xxf(x)x = 1101$ or 11110011, neither of which is a factor of $\mathbf{w}$. If $|x|_1 = 0$, then $x = 0$ or $x = 00$. If $f$ is the identity, this makes 0000 a factor of $\mathbf{w}$, which is impossible. If $f$ is the complement morphism, then 0010 or 00001100 is a factor of $\mathbf{w}$ neither of which is possible. We conclude that $|x|_0$, $|x|_1 \geq 1$.

Suppose that $f$ is the complement morphism. Word $\mathbf{w}$ has factors $f(x)x$ and $xx$, hence factors $0x$ and $1x$. This means that $x$ cannot start with 01, 10 or 00, since none of 101, 010 or 000 are factors of $\mathbf{w}$. We deduce that $x$ commences 11. Similarly, $x$ ends 11. Now, however, $xx$ has 1111 as a factor, which is impossible.

Suppose then that $f$ is the identity morphism, so that $xxxx$ is a factor of $\mathbf{w}$. Let $s \geq 0$ be maximal so that $0^s$ is a prefix of $x$. Let $t \geq 0$ be maximal such that $0^t$ is a suffix of $x$. Since $|x|_1 \geq 1$, $x$ has prefix $0^s1$ and suffix $10^t$, and $10^{t+s}1$ is a factor of $xx$, implying $t + s = 0$ or $t + s = 2$.

**Case 1:** Suppose $t + s = 0$. If $|x|_0 = 2$, write $x = 1^r0^21^q$, $r, q \geq 1$. Then $xxxx = 1^r0^21^{q+r}0^21^{q+r}0^21^{q+r}0^21^q$, and $\mathbf{t}$ contains the overlap $(q + r - 2)(q +$

$r-2)(q+r-2)$, which is impossible. Thus assume $|x|_0 > 2$, and write $x = 1^r 0^2 1^{\mathbf{t}_{[i]}+2} 0^2 \cdots 1^{\mathbf{t}_{[j]}+2} 0^2 1^q$, $r, q \geq 1$, $i \leq j$. Then $xxxx$ is

$$1^r 0^2 1^{\mathbf{t}_{[i]}+2} \cdots 1^{\mathbf{t}_{[j]}+2} 0^2 1^{q+r} 0^2 1^{\mathbf{t}_{[i]}+2} \cdots 1^{\mathbf{t}_{[j]}+2} 0^2 1^{q+r} 0^2 1^{\mathbf{t}_{[i]}+2} \cdots \\ \cdots 1^{\mathbf{t}_{[j]}+2} 0^2 1^{q+r} 0^2 1^{\mathbf{t}_{[i]}+2} 0^2 \cdots 1^{\mathbf{t}_{[j]}+2} 0^2 1^q \ ,$$

and $\mathbf{t}$ contains the overlap of powers $(q+r-2)\mathbf{t}_{[i]} \cdots \mathbf{t}_{[j]}(q+r-2)\mathbf{t}_{[i]} \cdots \mathbf{t}_{[j]}(q+r-2)$, which is again impossible.

**Case 2:** Suppose $t+s=2$. If $|x|_0 = 2$, write $x = 0^s 1^{\mathbf{t}_{[i]}+2} 0^t$, some $i$. Then $xxxx = 0^s 1^{\mathbf{t}_{[i]}+2} 0^2 1^{\mathbf{t}_{[i]}+2} 0^2 1^{\mathbf{t}_{[i]}+2} 0^2 1^{\mathbf{t}_{[i]}+2} 0^t$, and $\mathbf{t}$ contains the overlap $\mathbf{t}_{[i]}\mathbf{t}_{[i]}\mathbf{t}_{[i]}$, which is impossible. Thus assume $|x|_0 > 2$, and write $x = 0^s 1^{\mathbf{t}_{[i]}+2} 0^2 \cdots 1^{\mathbf{t}_{[j]}+2} 0^t$, $i \leq j$. Then $xxxx$ is

$$0^s 1^{\mathbf{t}_{[i]}+2} \cdots 1^{\mathbf{t}_{[j]}+2} 0^2 1^{\mathbf{t}_{[i]}+2} \cdots 1^{\mathbf{t}_{[j]}+2} 0^2 1^{\mathbf{t}_{[i]}+2} \cdots 1^{\mathbf{t}_{[j]}+2} 0^2 1^{\mathbf{t}_{[i]}+2} \cdots 1^{\mathbf{t}_{[j]}+2} 0^t,$$

and $\mathbf{t}$ contains the overlap of powers $\mathbf{t}_{[i]} \cdots \mathbf{t}_{[j]} \mathbf{t}_{[i]} \cdots \mathbf{t}_{[j]} \mathbf{t}_{[i]}$, which is again impossible. □

Let $\mathbf{v}$ be the infinite word

$$\mathbf{v} = \Pi_{i=0}^{\infty} 0 1^{2\mathbf{t}_{[i]}+1} \ .$$

We see that $\mathbf{v}$ is concatenated from 0's alternated with blocks of either one or three 1's.

**Lemma 18.** *Word* $\mathbf{v}$ *morphically avoids the pattern* $\theta(\alpha)\alpha\alpha\theta(\alpha)$.

**Proof.** Suppose there exist a substitution $x$ for $\alpha$ and a morphic involution $f$ such that $f(x)xxf(x)$ is a factor of $\mathbf{v}$.

Since 00 is not a factor of $\mathbf{v}$ but $xx$ is a factor, $|x|_1 \geq 1$. If $|x|_0 = 0$, then $x = 1^m$ for some $m$. If $f$ is the identity, this makes 1111 a factor of $\mathbf{v}$, which is impossible. If $f$ is the complement morphism, then $m = 1$, since $f(x) = 0^m$ is a factor of $\mathbf{v}$. Then, however, $f(x)xxf(x) = 0110$, which is not a factor of $\mathbf{v}$. We conclude that $|x|_0, |x|_1 \geq 1$.

Suppose that $f$ is the complement morphism. If $x$ begins and ends with different letters, then one of $f(x)x$ and $xf(x)$ has 00 as a factor, which is impossible. Therefore the first and last letters of $x$ are the same. They must both be 1; otherwise $xx$ would contain 00. Again 11 cannot be a factor of $x$; otherwise 00 would be a factor of $f(x)$. It follows that $x$ begins with 10 and ends with 01. Now, however, $xx$ has the factor 0110, which is impossible.

Suppose then that $f$ is the identity, so that $xxxx$ is a factor of $\mathbf{v}$. If $|x|_0 = 1$, write $x = 1^q 0 1^r$, some $q, r \geq 0$. We must have $q + r \geq 1$, since $|x|_1 \geq 1$. Now $xxxx = 1^q 0 1^{r+q} 0 1^{r+q} 0 1^{r+q} 0 1^t$. This implies the existence of an overlap of powers $\frac{r+q-1}{2}\frac{r+q-1}{2}\frac{r+q-1}{2}$ in $\mathbf{t}$, which is impossible.

Assume then that $|x|_0 \geq 2$. Write $x = 1^q 0 1^{2\mathbf{t}_{[i]}+1} \cdots 1^{2\mathbf{t}_{[j]}+1} 0 1^r$ for some $i \leq j$, some $q, r \geq 0$. Then $xxxx$ has the factor

$$1^{r+q} 0 1^{2\mathbf{t}_{[i]}+1} \cdots 1^{2\mathbf{t}_{[j]}+1} 0 1^{r+q} 0 1^{2\mathbf{t}_{[i]}+1} \cdots 1^{2\mathbf{t}_{[j]}+1} 0 1^{r+q} 0$$

and $\mathbf{t}$ contains the overlap of powers

$$\frac{r+q-1}{2} \mathbf{t}_{[i]} \cdots \mathbf{t}_{[j]} \frac{r+q-1}{2} \mathbf{t}_{[i]} \cdots \mathbf{t}_{[j]} \frac{r+q-1}{2} \ .$$

This is impossible. □

Let $\mathbf{u}$ be the infinite word

$$\mathbf{u} = \Pi_{i=0}^{\infty} 0 1^{t_i+2} \ .$$

We see that $\mathbf{u}$ is concatenated from 0's alternated with blocks of either two or three 1's.

**Lemma 19.** *Word* $\mathbf{u}$ *morphically avoids the patterns* $\alpha\alpha\theta(\alpha)\theta(\alpha)$ *and* $\alpha\theta(\alpha)\alpha\theta(\alpha)$.

**Proof.** Suppose for the sake of getting a contradiction that there exists a substitution $x$ for $\alpha$ and a morphic involution $f$ such that $xxf(x)f(x)$ or $xf(x)xf(x)$ is a factor of $\mathbf{u}$.

First suppose that $f$ is the complement morphism. Since $\mathbf{u}$ contains a factor $f(x)$, but no factor 00, word $x$ cannot contain 11 as a factor. Similarly, $\mathbf{u}$ does not contain a factor 010, so that $x$ cannot contain a factor 101. The only possibilities for $x$ are then 0, 1, 01 and 10. The resulting values for $xxf(x)f(x)$ (resp. $xf(x)xf(x)$) would be 0011, 1100, 01011010, 10100101 (resp. 0101, 1010, 01100110, 10011001) which all contain either 00 or 010 and are thus impossible.

Suppose then that $f$ is the identity morphism. Thus $xxf(x)f(x) = xf(x)xf(x) = xxxx$. Since 00 is not a factor of $\mathbf{u}$ but $xx$ is a factor, $|x|_1 \geq 1$. If $|x|_0 = 0$, then $x = 1^m$ for some $m$, and 1111 is a factor of $\mathbf{u}$. This is impossible. It follows that $|x|_0$, $|x|_1 \geq 1$. If $|x|_0 = 1$, write $x = 1^q 0 1^r$, some $q, r \geq 0$. Then $xxxx = 1^q 0 1^{r+q} 0 1^{r+q} 0 1^{r+q} 0 1^t$. This implies the existence of an overlap $(r+q-2)(r+q-2)(r+q-2)$ in $\mathbf{t}$, which is impossible.

Assume then that $|x|_0 \geq 2$. Write $x = 1^q 0 1^{\mathbf{t}_{[i]}+2} \cdots 1^{\mathbf{t}_{[j]}+2} 0 1^r$ for some $i \leq j$, some $q, r \geq 0$. Then $xxxx$ has the factor

$$1^{r+q} 0 1^{\mathbf{t}_{[i]}+2} \cdots 1^{\mathbf{t}_{[j]}+2} 0 1^{r+q} 0 1^{\mathbf{t}_{[i]}+2} \cdots 1^{\mathbf{t}_{[j]}+2} 0 1^{r+q}$$

and $\mathbf{t}$ contains the overlap of powers

$$(r+q-2)\mathbf{t}_{[i]} \cdots \mathbf{t}_{[j]} (r+q-2) \mathbf{t}_{[i]} \cdots \mathbf{t}_{[j]} (r+q-2) \ .$$

This is impossible. □

We continue with the antimorphic case. There are only two antimorphisms over $\Sigma_2$: the reversal $x \mapsto x^R$ and the reverse complement $x \mapsto \bar{x}^R$ with $\bar{0} = 1$ and $\bar{1} = 0$.

**Lemma 20.** *Word* $\mathbf{w}$ *antimorphically avoids the pattern* $\alpha\theta(\alpha)\alpha$.

**Proof.** Suppose for the sake of getting a contradiction that there exists a substitution $x$ for $\alpha$ and an antimorphic involution $f$ such that $xf(x)x$ is a factor of $\mathbf{w}$.

By Lemma 17 we may assume that $f(x) \neq x$, since we have shown that $\mathbf{w}$ has no factor $xxxx$ with $x$ non-empty. Similarly, we may assume that $f(x) \neq \bar{x}$. These conditions together imply that $x$ is not a palindrome, and that $x^R \neq \bar{x}$. Suppose, for example that $x$ is a palindrome. If $f$ is the reversal, then $f(x) = x$, which we have forbidden. If $f$ is the reverse complement, then $f(x) = (x^R) = \bar{x}$, again forbidden. Similarly one checks that $x^R \neq \bar{x}$. To continue with our proof, suppose that $f$ is the reverse complement. Since $\mathbf{w}$ contains a factor $f(x)$, but no factor 000, word $x$ cannot contain 111 as a factor. Also, $\mathbf{w}$ does not contain 010 or 101 as a factor. It follows that $x$ is a factor of $(0011)^\omega$. Since $xf(x)$ and $f(x)x$ are factors of $\mathbf{w}$, $x$ cannot begin or end with 01 or 10. It therefore begins and ends with 00 or 11. The length 2 prefix and length 2 suffix of $x$ must differ, since otherwise $xx$ would have 0000 or 1111 as a factor. We conclude that $x = (0011)^n$ or $x = (1100)^n$ for some $n$. But then $x$ is the complem ent of its reverse, contradicting our previous assumption.

Suppose then that $f$ is the reversal. Since $xf(x)$ and $xx$ are both factors of $\mathbf{w}$ but 010, 101 are not, $x$ cannot end in 01 or 10. Then $x$ ends in 00 or 11, and $xf(x)$ contains 0000 or 1111 as a factor. This is impossible. □

**Lemma 21.** *Word* $(0001)^\omega$ *antimorphically avoids the patterns* $\alpha\alpha\theta(\alpha)\theta(\alpha)$, $\alpha\theta(\alpha)\alpha\theta(\alpha)$, *and* $\theta(\alpha)\alpha\alpha\theta(\alpha)$.

**Proof.** Suppose for the sake of getting a contradiction that there exists a substitution $x$ for $\alpha$ and an antimorphic involution $f$ such that $xxf(x)f(x)$, $xf(x)xf(x)$, or $f(x)xxf(x)$ is a factor of $(0001)^\omega$.

If $f$ is the reversal, then $x$ cannot end in 01 or 10; this would imply 0110 or 1001 as a factor of $xf(x)$; however these are not factors of $(0001)^\omega$. It follows that if $|x| > 1$ then $x$ ends in 00, since 11 is not a factor of $(0001)^\omega$. Then, however 0000 is a factor of $xf(x)$, which is impossible. We conclude that $|x| = 1$, and $xxf(x)f(x)$, $xf(x)xf(x)$, $f(x)xxf(x) \in \{1111, 0000\}$. This is impossible.

If $\theta$ is the reverse complement, 00 cannot be a factor of $x$; otherwise 11 is a factor of $\theta(x)$. However, $x$ cannot end in 01 or 10, or $x\theta(x)$ would have 0101 or 1010 as a factor. We conclude that $|x| = 1$, and $xx\theta(x)\theta(x) = x\theta(x)x\theta(x) = \theta(x)xx\theta(x) \in \{0011, 0101, 1001\}$, which are impossible. □

## 6. Result

The above lemmas imply the following conclusion.

**Theorem 22.**

$$\mathcal{V}_{\mathrm{m}}(p) = \mathcal{V}_{\mathrm{a}}(p) = \begin{cases} 3 & \textit{if } p \in E^3 \setminus \{\alpha\alpha\alpha, \theta(\alpha)\theta(\alpha)\theta(\alpha)\} \\ \infty & \textit{if } p \in \{\alpha, \theta(\alpha), \alpha\theta(\alpha), \theta(\alpha)\alpha\} \\ 2 & \textit{otherwise} \end{cases}$$

## Acknowledgement

We would like to thank the anonymous referees for their careful reviews and constructive comments.

## Bibliography

[1] J. Cassaigne. *Unavoidable Patterns*, chapter 3, pages 111–134. In [3], 2002.
[2] Ehsan Chiniforooshan, Lila Kari, and Zhi Xu. Pseudopower avoidance. *Fundamenta Informaticae*, 114:1–18, 2012.
[3] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, Cambridge, UK, 2002.
[4] A. Thue. Über unendliche Zeichenreihen. *Norske Vid. Skrifter I. Mat.-Nat. Kl., Christiania*, 7:1–22, 1906.
[5] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Skrifter I. Mat.-Nat. Kl., Christiania*, 1:1–67, 1912.