On some interesting ternary formulas

Pascal Ochem*
LIRMM, CNRS
Université de Montpellier
France
ochem@lirmm.fr

Matthieu Rosenfeld
LIP, ENS de Lyon, CNRS, UCBL
Université de Lyon
France
matthieu.rosenfeld@ens-lyon.fr

Submitted: May 30, 2018; Accepted: Dec 28, 2018; Published: Jan 25, 2019

Abstract

We obtain the following results about the avoidance of ternary formulas. Up to renaming of the letters, the only infinite ternary words avoiding the formula $ABCAB.ABCBA.ACB.BAC$ (resp. $ABCA.BCAB.BCB.CBA$) are the ones that have the same set of recurrent factors as the fixed point of `0` $\mapsto$ `012`, `1` $\mapsto$ `02`, `2` $\mapsto$ `1`. The formula $ABAC.BACA.ABCA$ is avoided by polynomially many binary words (w.r.t. to their lengths) and there exist arbitrarily many infinite binary words with different sets of recurrent factors that avoid it. If every variable of a ternary formula appears at least twice in the same fragment, then the formula is 3-avoidable. The pattern $ABACADABCA$ is unavoidable for the class of C_4-minor-free graphs with maximum degree 3. This disproves a conjecture of Grytczuk. The formula $ABCA.ACBA$, or equivalently the palindromic pattern $ABCADACBA$, has avoidability index 4.

Mathematics Subject Classifications: 68R15

1 Introduction

A *pattern* p is a non-empty finite word over an alphabet $\Delta = \{A, B, C, \dots\}$ of capital letters called *variables*. An *occurrence* of p in a word w is a non-erasing morphism $h : \Delta^* \to \Sigma^*$ such that $h(p)$ is a factor of w. The *avoidability index* $\lambda(p)$ of a pattern p is the size of the smallest alphabet Σ such that there exists an infinite word over Σ containing no occurrence of p.

A variable that appears only once in a pattern is said to be *isolated*. Following Cassaigne [4], we associate a pattern p with the *formula* f obtained by replacing every isolated variable in p by a dot. For example, the pattern $AABCABBDBBAA$ gives the formula

*The authors were partially supported by the ANR project CoCoGro (ANR-16-CE40-0005).

$AAB.ABB.BBAA$. The factors that are separated by dots are called *fragments*. So AAB, ABB, and $BBAA$ are the fragments of $AAB.ABB.BBAA$.

An *occurrence* of a formula f in a word w is a non-erasing morphism $h : \Delta^* \to \Sigma^*$ such that the h-image of every fragment of f is a factor of w. As for patterns, the avoidability index $\lambda(f)$ of a formula f is the size of the smallest alphabet allowing the existence of an infinite word containing no occurrence of f. Clearly, if a formula f is associated with a pattern p, every word avoiding f also avoids p, so $\lambda(p) \leqslant \lambda(f)$. Recall that an infinite word is *recurrent* if every finite factor appears infinitely many times and that any infinite factorial language contains a recurrent word (see Proposition 5.1.13 of [8] for instance). Thus, if there exists an infinite word over Σ avoiding p, then there exists an infinite recurrent word over Σ avoiding p. This recurrent word avoiding p also avoids f, so that $\lambda(p) = \lambda(f)$. Without loss of generality, a formula is such that no variable is isolated and no fragment is a factor of another fragment. We say that a formula f is *divisible* by a formula f' if f does not avoid f', that is, there is a non-erasing morphism h such that the image of any fragment of f' under h is a factor of a fragment of f. If f is divisible by f', then every word avoiding f' also avoids f. Let $\Sigma_k = \{0, 1, \ldots, k-1\}$ denote the k-letter alphabet. We denote by Σ_k^n the k^n words of length n over Σ_k.

A formula is *binary* if it has at most 2 variables. We have recently determined the avoidability index of every binary formula [14]. This exhaustive study led to the discovery of some binary formulas that are avoided by only a few binary words. Determining the avoidability index of every ternary formula would be a huge task. However, we have identified some interesting ternary formulas and this paper describes their properties.

We say that two infinite words are equivalent if they have the same set of factors. Let b_3 be the fixed point of $\mathtt{0} \mapsto \mathtt{012}$, $\mathtt{1} \mapsto \mathtt{02}$, $\mathtt{2} \mapsto \mathtt{1}$. A famous result of Thue [2, 15, 16] can be stated as follows:

Theorem 1. *[2, 15, 16] Every recurrent ternary word avoiding AA, $\mathtt{010}$, and $\mathtt{212}$ is equivalent to b_3.*

In Section 2, we obtain a similar result for b_3 by forbidding one ternary formula but without forbidding explicit factors in Σ_3^*. In Section 3, we describe the set of binary words avoiding $ABACA.ABCA$ and $ABAC.BACA.ABCA$. We show that these formulas are avoided by polynomially many binary words (w.r.t. to their lengths) and that there exist infinitely many recurrent binary words with different sets of recurrent factors that avoid them. In the terminology of [14], these formulas are not essentially avoided by a finite set of morphic words. In Section 4, we consider *nice* formulas. A formula f is nice if for every variable X of f, there exists a fragment of f that contains X at least twice. This notion generalizes to formulas the notion of a *doubled* pattern (that is, a pattern that contains every variable at least twice). Every doubled pattern is 3-avoidable [13]. We show that every ternary nice formula is 3-avoidable. In Section 5, we show that $ABACADABCA$ is a 2-avoidable pattern that is unavoidable on graphs with maximum degree 3. In Section 6, we show that there exists a palindromic pattern with index 4.

A preliminary version of this paper, without the results in Sections 4 and 6, has been presented at WORDS 2017.

2 Formulas closely related to b_3

For every letter $c \in \Sigma_3$, $\sigma_c : \Sigma_3^* \mapsto \Sigma_3^*$ is the morphism such that $\sigma_c(a) = b$, $\sigma_c(b) = a$, and $\sigma_c(c) = c$ with $\{a, b, c\} = \Sigma_3$. So σ_c is the morphism that fixes c and exchanges the two other letters.

We consider the following formulas.

- $f_b = ABCAB.ABCBA.ACB.BAC$
- $f_1 = ABCA.BCAB.BCB.CBA$
- $f_2 = ABCAB.BCB.AC$
- $f_3 = ABCA.BCAB.ACB.BCB$
- $f_4 = ABCA.BCAB.BCB.AC.BA$

Notice that f_b is divisible by f_1, f_2, f_3, f_4.

Theorem 2. *Let $f \in \{f_b, f_1, f_2, f_3, f_4\}$. Every ternary recurrent word avoiding f is equivalent to b_3, $\sigma_0(b_3)$, or $\sigma_2(b_3)$.*

By considering divisibility, we can deduce that Theorem 2 holds for 72 ternary formulas. Since b_3, $\sigma_0(b_3)$, and $\sigma_2(b_3)$ are equivalent to their reverses, Theorem 2 also holds for the 72 reverse ternary formulas.

Proof. Using Cassaigne's algorithm [3], we have checked that b_3 avoids f_i, for $1 \leqslant i \leqslant 4$. By symmetry, $\sigma_0(b_3)$ and $\sigma_2(b_3)$ also avoid f_i.

Let w be a ternary recurrent word w avoiding f_b. Assume towards a contradiction that w contains a square uu. Then there exists a non-empty word v such that $uuvuu$ is a factor of w. Thus, w contains an occurrence of f_b given by the morphism $A \mapsto u, B \mapsto u, C \mapsto v$. This contradiction shows that w is square-free.

An occurrence h of a ternary formula over Σ_3 is said to be *basic* if $\{h(A), h(B), h(C)\} = \Sigma_3$. As already noticed by Thue [2], no infinite ternary word avoids squares and `012`. So, every infinite ternary square-free word contains the 6 factors obtained by letter permutation of `012`. Thus, an infinite ternary square-free word contains a basic occurrence of f_b if and only if it contains the same basic occurrence of $ABCAB.ABCBA$. Therefore, w contains no basic occurrence of $ABCAB.ABCBA$. A computer check shows that the longest ternary words avoiding f_b, squares, `021020120`, `102101201`, and `210212012` have length 159. So we assume without loss of generality that w contains `021020120`.

Assume towards a contradiction that w contains `010`. Since w is square-free, w contains `20102`. Moreover, w contains the factor `20120` of `021020120`. So w contains the basic occurrence $A \mapsto$ `2`, $B \mapsto$ `0`, $C \mapsto$ `1` of $ABCAB.ABCBA$. This contradiction shows that w avoids `010`.

Assume towards a contradiction that w contains `212`. Since w is square-free, w contains `02120`. Moreover, w contains the factor `02102` of `021020120`. So w contains the basic occurrence $A \mapsto 0$, $B \mapsto 2$, $C \mapsto 1$ of $ABCAB.ABCBA$. This contradiction shows that w avoids 212.

Since w avoids squares, 010, and 212, Theorem 1 implies that w is equivalent to b_3. By symmetry, every ternary recurrent word avoiding f_b is equivalent to b_3, $\sigma_0(b_3)$, or $\sigma_2(b_3)$. □

3 Avoidability of $ABACA.ABCA$ and $ABAC.BACA.ABCA$

Following the terminology in [14], we say that a finite set of infinite words $\mathcal{M}$ *essentially avoids* a formula f if every infinite word over $\Sigma_{\lambda(f)}$ avoiding f has the same set of recurrent factors as a word in $\mathcal{M}$. Let us list all the formulas (up to symetries) from the literature that are known to be essentially avoided by a finite set of words.

- Five binary formulas are known to be essentially avoided by a finite set of binary morphic words [14].
- $\{b_3, \sigma_0(b_3), \sigma_2(b_3)\}$ essentially avoids the ternary formulas in Section 2.
- $\{b_4, b'_4, b''_4\}$ essentially avoids $AB.AC.BA.CA.CB$ [1], where b_4 is the fixed point of $0 \mapsto 01$, $1 \mapsto 21$, $2 \mapsto 03$, $3 \mapsto 23$, b'_4 is obtained from b_4 by exchanging 0 and 1, and b''_4 is obtained from b_4 by exchanging 0 and 3.

The formulas listed above are also the only ones known to be avoided by polynomially many words (w.r.t. to their lengths). In this section, we show that the formulas $ABACA.ABCA$ and $ABAC.BACA.ABCA$ behave differently: they are avoided by polynomially many binary words but they are not essentially avoided by a finite set of morphic words.

We consider the morphisms $m_a : 0 \mapsto 001$, $1 \mapsto 101$ and $m_b : 0 \mapsto 010$, $1 \mapsto 110$. That is, $m_a(x) = x01$ and $m_b(x) = x10$ for every $x \in \Sigma_2$. We construct the set S of binary words as follows:

- $0 \in S$.
- If $v \in S$, then $m_a(v) \in S$ and $m_b(v) \in S$.
- If $v \in S$ and v' is a factor of v, then $v' \in S$.

Theorem 3. *Let $f \in \{ABACA.ABCA, ABAC.BACA.ABCA\}$. The set of words u such that u is recurrent in an infinite binary word avoiding f is S.*

Proof. Let R be the set of words u such that u is recurrent in an infinite binary word avoiding $ABACA.ABCA$. Let R' be the set of words u such that u is recurrent in an infinite binary word avoiding $ABAC.BACA.ABCA$. An occurrence of $ABACA.ABCA$ is also an occurrence of $ABAC.BACA.ABCA$, so that $R' \subseteq R$.

Let us show that $R \subseteq S$. We study the small factors of a recurrent binary word w avoiding $ABACA.ABCA$. Notice that w avoid the pattern $ABAAA$ since it contains the occurrence $A \mapsto A$, $B \mapsto B$, $C \mapsto A$ of $ABACA.ABCA$. Since w contains recurrent factors only, w also avoids AAA.

A computer check shows that the longest binary words avoiding $ABACA.ABCA$, AAA, $\texttt{1001101001}$, and $\texttt{0110010110}$ have length 53. So we assume without loss of generality that w contains $\texttt{1001101001}$.

Assume towards a contradiction that w contains $\texttt{1100}$. Since w avoids AAA, w contains $\texttt{011001}$. Then w contains the occurrence $A \mapsto \texttt{01}, B \mapsto \texttt{1}, C \mapsto \texttt{0}$ of $ABACA.ABCA$. This contradiction shows that w avoids $\texttt{1100}$.

Since w contains $\texttt{0110}$, the occurrence $A \mapsto \texttt{0}, B \mapsto \texttt{1}, C \mapsto \texttt{1}$ of $ABACA.ABCA$ shows that w avoids $\texttt{01010}$. Similarly, w contains $\texttt{1001}$ and avoids $\texttt{10101}$.

Assume towards a contradiction that w contains $\texttt{0101}$. Since w avoids $\texttt{01010}$ and $\texttt{10101}$, w contains $\texttt{001011}$. Moreover, w avoids AAA, so w contains $\texttt{10010110}$. Then w contains the occurrence $A \mapsto \texttt{10}, B \mapsto \texttt{0}, C \mapsto \texttt{1}$ of $ABACA.ABCA$. This contradiction shows that w avoids $\texttt{0101}$.

So w avoids every factor in $\{\texttt{000}, \texttt{111}, \texttt{0101}, \texttt{1100}\}$. Thus, it is not difficult to check that if we extend any factor $\texttt{01}$ in w to three letters to the right, we get either $\texttt{01001}$ or $\texttt{01101}$, that is, $\texttt{01}x\texttt{01}$ with $x \in \Sigma_2$. This implies that w is the m_a-image of some binary word.

Obviously, the image by a non-erasing morphism of a word containing a formula also contains the formula. Thus, the pre-image of w by m_a also avoids $ABACA.ABCA$. This shows that $R \subseteq S$.

Let us show that $S \subseteq R'$, that is, every word in S avoids $ABAC.BACA.ABCA$. Assume towards a contradiction that a finite word $w \in S$ avoids $ABAC.BACA.ABCA$ and that $m_a(w)$ contains an occurrence h of $ABAC.BACA.ABCA$.

If we write $w = w_0w_1w_2w_3\ldots$, then the word $m_a(w) = w_0\texttt{01}w_1\texttt{01}w_2\texttt{01}w_3\texttt{01}\ldots$ is such that:

- Every factor $\texttt{00}$ occurs at position $0 \pmod 3$.
- Every factor $\texttt{01}$ occurs at position $1 \pmod 3$.
- Every factor $\texttt{11}$ occurs at position $2 \pmod 3$.
- Every factor $\texttt{10}$ occurs at position 0 or $2 \pmod 3$, depending on whether the factor $\texttt{1}w_i\texttt{0}$ is $\texttt{100}$ or $\texttt{110}$.

We say that a factor s is *gentle* if either $|s| \geqslant 3$ or $s \in \{\texttt{00}, \texttt{01}, \texttt{11}\}$. By the previous remarks, all the occurrences of the same gentle factor have the same position modulo 3.

First, we consider the case when $h(A)$ is gentle. This implies that the distance between two occurrences of $h(A)$ is $0 \pmod 3$. Because of the repetitions $h(ABA)$, $h(ACA)$, and $h(ABCA)$ are contained in the formula, we deduce that

- $|h(AB)| = |h(A)| + |h(B)| \equiv 0 \pmod 3$.
- $|h(AC)| = |h(A)| + |h(C)| \equiv 0 \pmod 3$.

- $|h(ABC)| = |h(A)| + |h(B) + |h(C)| \equiv 0 \pmod 3$.

This gives $|h(A)| \equiv |h(B)| \equiv |h(C)| \equiv 0 \pmod 3$. Clearly, such an occurrence of the formula in $m_a(w)$ implies an occurrence of the formula in w, which is a contradiction.

Now we consider the case when $h(B)$ is gentle. If $h(CA)$ is also gentle, then the factors $h(BACA)$ and $h(BCA)$ imply that $|h(A)| \equiv 0 \pmod 3$. Thus, $h(A)$ is gentle and the first case applies. If $h(CA)$ is not gentle, then $h(CA) = \mathtt{10}$, that is, $h(C) = \mathtt{1}$ and $h(A) = \mathtt{0}$. Thus, $m_a(w)$ contains both $h(BAC) = h(B)\mathtt{01}$ and $h(BCA) = h(B)\mathtt{10}$. Since $h(B)$ is gentle, this implies that 01 and 10 have the same position modulo 3, which is impossible.

The case when $h(C)$ is gentle is symmetrical. If $h(AB)$ is gentle, then $h(ABAC)$ and $h(ABC)$ imply that $|h(A)| \equiv 0 \pmod 3$. If $h(AB)$ is not gentle, then $h(A) = \mathtt{1}$ and $h(B) = \mathtt{0}$. Thus, $m_a(w)$ contains both $h(ABC) = \mathtt{01}h(C)$ and $h(BAC) = \mathtt{10}h(C)$. Since $h(C)$ is gentle, this implies that 01 and 01 have the same position modulo 3, which is impossible.

Finally, if $h(A)$, $h(B)$, and $h(C)$ are not gentle, then the length of the three fragments of the formula is $2|h(A)| + |h(B)| + |h(C)| \leqslant 8$. So it suffices to consider the factors of length at most 8 in S to check that no such occurrence exists.

This shows that $S \subseteq R'$. Since $R' \subseteq R \subseteq S \subseteq R'$, we obtain $R' = R = S$, which proves Theorem 3. □

Corollary 4. *Neither $ABACA.ABCA$ nor $ABAC.BACA.ABCA$ is essentially avoided by a finite set of morphic words.*

Proof. Let $c(n) = |S \cap \Sigma_2^n|$ denote the number of words of length n in S. By construction of S,

$$c(n) = 2 \sum_{0 \leqslant i \leqslant 2} c\left(\left\lceil \tfrac{n-i}{3} \right\rceil\right) \text{ for every } n \geqslant 8.$$

Thus $c(n) = \Theta\left(n^{\ln 6/\ln 3}\right) = \Theta\left(n^{1+\ln 2/\ln 3}\right)$. Devyatov [7] has recently shown that the factor complexity (i.e. the number of factors of length n) of a morphic word is either $O\left(n \ln(n)\right)$ or $\Theta\left(n^{1+1/k}\right)$ for some integer $k \geqslant 1$. Thus, S cannot be the union of the factors of a finite number of morphic words. □

4 Ternary nice formulas

Clark [5] introduced the notion of *n-avoidance basis* for formulas, which is the smallest set of formulas with the following property: for every $i \leqslant n$, every avoidable formula with i variables is divisible by at least one formula with at most i variables in the n-avoidance basis. See [5, 9] for more discussions about the n-avoidance basis. The avoidability index of every formula in the 3-avoidance basis has been determined:

- AA ($\lambda = 3$ [15])
- $ABA.BAB$ ($\lambda = 3$ [4])
- $ABCA.BCAB.CABC$ ($\lambda = 3$ [9])

- $ABCBA.CBABC$ ($\lambda = 2$ [9])
- $ABCA.CABC.BCB$ ($\lambda = 3$ [9])
- $ABCA.BCAB.CBC$ ($\lambda = 3$, reverse of $ABCA.CABC.BCB$)
- $AB.AC.BA.CA.CB$ ($\lambda = 4$ [1])

Recall that a formula f is *nice* if for every variable X of f, there exists a fragment of f that contains X at least twice. Every formula in the 3-avoidance basis except $AB.AC.BA.CA.CB$ is both nice and 3-avoidable. This raised the question in [9] whether every nice formula is 3-avoidable, which would generalize the 3-avoidability of doubled patterns. In this section, we answer this question positively for ternary formulas.

Theorem 5. *Every nice formula with at most* 3 *variables is* 3*-avoidable.*

We say that a nice formula is minimal if it is not divisible by another nice formula with at most the same number of variables. The following property of every minimal nice formula is easy to derive. If a variable V appears as a prefix of a fragment ϕ, then

- V is also a suffix of ϕ,
- ϕ contains exactly two occurrences of V,
- V is neither a prefix nor a suffix of any fragment other than ϕ,
- Every fragment other than ϕ contains at most one occurrence of V.

Thus, if f is a minimal nice formula with $n \geqslant 2$ variables, then f has at most n fragments. Moreover, every fragment has length at most $2 + 2^{n-1} - 1 = 2^{n-1} + 1$, since otherwise it would contain a doubled pattern as a factor.

This implies an algorithm to list the minimal nice formulas with at most n variables. The table below lists the formulas that need to be shown 3-avoidable, that is, the minimal nice formulas with at most 3 variables that do not belong to the 3-avoidance basis. Also, if two distinct formulas are the reverse of each other, then only one of them appears in the table and the given avoiding word avoids both formulas. Some of these formulas are avoided by b_3 and the proof uses Cassaigne's algorithm [3] as in Section 2. The other formulas are each avoided by the image by a uniform morphism of either any infinite $\left(\frac{5}{4}^+\right)$-free word w_5 over Σ_5 or any infinite $\left(\frac{7}{5}^+\right)$-free word w_4 over Σ_4. We refer to [12, 13] for details about the technique to prove avoidance with morphic images of (α^+)-free words.

Formula	Closed under reversal?	Avoidability exponent	Avoiding word
$ABA.BCB.CAC$	yes	1.5	b_3
$ABCA.BCAB.CBAC$	no	1.333333333	b_3
$ABCA.BAB.CAC$	yes	1.414213562	$g_v(w_4)$
$ABCA.BAB.CBC$	no	1.430159709	$g_w(w_4)$
$ABCA.BAB.CBAC$	no	1.381966011	$g_x(w_5)$
$ABCBA.CABC$	no	1.361103081	$g_y(w_5)$
$ABCBA.CAC$	yes	1.396608253	$g_z(w_5)$

$$g_v \quad \begin{array}{l} 0 \to \texttt{01220}, \\ 1 \to \texttt{01110}, \\ 2 \to \texttt{00212}, \\ 3 \to \texttt{00112}. \end{array} \qquad g_w \quad \begin{array}{l} 0 \to \texttt{02111}, \\ 1 \to \texttt{01121}, \\ 2 \to \texttt{00222}, \\ 3 \to \texttt{00122}. \end{array} \qquad g_x \quad \begin{array}{l} 0 \to \texttt{021110}, \\ 1 \to \texttt{012221}, \\ 2 \to \texttt{011120}, \\ 3 \to \texttt{002211}, \\ 4 \to \texttt{001122}. \end{array} \qquad g_y \quad \begin{array}{l} 0 \to \texttt{022}, \\ 1 \to \texttt{021}, \\ 2 \to \texttt{012}, \\ 3 \to \texttt{011}, \\ 4 \to \texttt{000}. \end{array} \qquad g_z \quad \begin{array}{l} 0 \to \texttt{120201}, \\ 1 \to \texttt{100002}, \\ 2 \to \texttt{022221}, \\ 3 \to \texttt{011112}, \\ 4 \to \texttt{001122}. \end{array}$$

5 A counter-example to a conjecture of Grytczuk

Grytczuk [10] considered the notion of pattern avoidance on graphs. This generalizes the definition of nonrepetitive coloring, which corresponds to the pattern AA. Given a pattern p and a graph G, the avoidability index $\lambda(p, G)$ is the smallest number of colors needed to color the vertices of G such that every path in G induces a word avoiding p.

We think that the natural framework is that of directed graphs with no loops and no multiple arcs, but such that opposite arcs (i.e., digons) are allowed. An oriented path in a directed graph $\overrightarrow{G}$ is a sequence of distinct vertices $v_1, v_2, \ldots, v_k$ such that $\overrightarrow{G}$ contains all the arcs $\overrightarrow{v_iv_{i+1}}$ such that $1 \leqslant i \leqslant k-1$.

A pattern occurs in a vertex-colored directed graph $\overrightarrow{G}$ if the sequence of colors on a directed path of $\overrightarrow{G}$ induces an occurrence of the pattern. Informally, the orientation of the path corresponds to the reading direction. We define $\lambda\left(p, \overrightarrow{G}\right)$ as the smallest number of colors such that there exists a vertex coloring avoiding p. This way, $\lambda(p) = \lambda\left(p, \overrightarrow{P}\right)$, where $\overrightarrow{P}$ is the infinite oriented path with vertices v_i and arcs $\overrightarrow{v_iv_{i+1}}$, for every $i \geqslant 0$.

Thus, an undirected graph corresponds to a symmetric directed graph: for every pair of distinct vertices u and v, either there exists no arc between u and v, or there exist both the arcs $\overrightarrow{uv}$ and $\overrightarrow{vu}$. Let P denote the infinite undirected path. We prefer the framework of directed graphs because, even though $\lambda\left(AA, \overrightarrow{P}\right) = \lambda(AA, P) = 3$, there exist patterns such that $\lambda\left(p, \overrightarrow{P}\right) < \lambda(p, P)$. For example, $\lambda(ABCACB) = \lambda\left(ABCACB, \overrightarrow{P}\right) = 2$ [12], whereas $\lambda(ABCACB, P) = 3$ since a computer check shows that the longest binary words avoiding both $ABCACB$ and its reverse $ABCBAC$ have length 23. The equivalence between avoiding a pattern and its corresponding formula holds for $\overrightarrow{P}$ but does not generalize to other directed graphs. So we do not try to define a notion of avoidance for formulas on graphs or directed graphs.

A conjecture of Grytczuk [10] says that for every avoidable pattern p, there exists a function g such that $\lambda(p, G) \leqslant g(\Delta(G))$, where G is an undirected graph and $\Delta(G)$ denotes its maximum degree. Grytczuk [10] obtained that his conjecture holds for doubled patterns.

As a counterexample, we consider the pattern $ABACADABCA$ which is 2-avoidable by the result in Section 3. Of course, $ABACADABCA$ is not doubled because of the isolated variable D. Let us show that $ABACADABCA$ is unavoidable on the infinite oriented graph $\overrightarrow{G}$ with vertices v_i and arcs $\overrightarrow{v_i v_{i+1}}$ and $\overrightarrow{v_{100i} v_{100i+2}}$, for every $i \geqslant 0$. Notice that $\overrightarrow{G}$ is obtained from $\overrightarrow{P}$ by adding the arcs $\overrightarrow{v_{100i} v_{100i+2}}$. The constant 100 in the construction is arbitrary and can be replaced by any constant.

Suppose that $\overrightarrow{G}$ is colored with k colors. Consider the factors in the subgraph $\overrightarrow{P}$ induced by the paths from $v_{300ik+1}$ to $v_{300ik+200k+1}$, for every $i \geqslant 0$. Since these factors have bounded length, the same factor appears on two disjoint such paths p_l and p_r (such that p_l is on the left of p_r). Notice that p_l contains $2k+1$ vertices with index $\equiv 1 \pmod{100}$. By the pigeon-hole principle, p_l contains three such vertices with the same color a. Thus, p_l contains an occurrence of $ABACA$ such that $A \mapsto a$ on vertices with index $\equiv 1 \pmod{100}$. The same is true for p_r. In $\overrightarrow{G}$, the occurrences of $ABACA$ in p_l and p_r imply an occurrence of $ABACADABCA$ since we can skip an occurrence of the variable A in p_l thanks to some arc of the form $\overrightarrow{v_{100j} v_{100j+2}}$.

This shows that $ABACADABCA$ is unavoidable on $\overrightarrow{G}$. So Grytczuk's conjecture is disproved since $\overrightarrow{G}$ has maximum degree 3. It is also a counterexample to Conjecture 6 in [6] which states that every avoidable pattern is avoidable on the infinite graph with vertices $\{v_0, v_1, \ldots\}$ and the arcs $\overrightarrow{v_i v_{i+1}}$ and $\overrightarrow{v_i v_{i+2}}$ for every $i \geqslant 0$.

6 A palindrome with index 4

Mikhailova [11] considered the largest avoidability index $\mathcal{P}$ of an avoidable pattern that is a palindrome. She proved that $\mathcal{P} \leqslant 16$. An obvious lower bound is $\mathcal{P} \geqslant \lambda(AA) = 3$. For a better lower bound, we consider the palindromic pattern $ABCADACBA$ or, equivalently, the ternary formula $f = ABCA.ACBA$. Since it is a ternary formula, f is 4-avoidable. More precisely, f is not nice because of the variable C, so the only formula in the 3-avoidance basis that divides f is $AB.AC.BA.CA.CB$, which is avoided by b_4.

Let us show that f is not 3-avoidable. Let w be a ternary recurrent word avoiding f. Assume towards a contradiction that w contains a square uu. Then there exists a non-empty word v such that $uuvuu$ is a factor of w. Thus, w contains an occurrence of f given by the morphism $A \mapsto u, B \mapsto u, C \mapsto v$. This contradiction shows that w is square-free. A computer check shows that no infinite ternary square-free word avoids f. This holds even if we forbid only squares and every occurrence h of f such that $|h(A)| = 1$ and $|h(B)| + |h(C)| \leqslant 5$. Thus, $\mathcal{P} \geqslant \lambda(ABCADACBA) = \lambda(ABCA.ACBA) = 4$.

References

[1] K. A. Baker, G. F. McNulty, and W. Taylor. Growth problems for avoidable words. *Theoret. Comput. Sci.*, 69(3):319–345, 1989.

[2] J. Berstel. *Axel Thue's papers on repetitions in words: a translation*, volume 20 of *Publications du LACIM*. Université du Québec à Montréal, 1994.

[3] J. Cassaigne. *An Algorithm to Test if a Given Circular HDOL-Language Avoids a Pattern. IFIP Congress*, pages 459–464, 1994.

[4] J. Cassaigne. *Motifs évitables et régularité dans les mots.* PhD thesis, Université Paris VI, 1994.

[5] R. J. Clark. *Avoidable formulas in combinatorics on words.* PhD thesis, University of California, Los Angeles, 2001. Available at http://www.lirmm.fr/~ochem/morphisms/clark_thesis.pdf

[6] M. Debski, U. Pastwa, and K. Wesek. Grasshopper avoidance of patterns. *Electron. J. Combinatorics.*, 23(4):#P4.17, 2016.

[7] R. Devyatov. On subword complexity of morphic sequences. *Math. USSR Sbornik*, 2015. arXiv:1502.02310

[8] Pytheas Fogg. Substitutions in Dynamics, Arithmetics and Combinatorics. Springer Science & Business Media, 2002.

[9] G. Gamard, P. Ochem, G. Richomme, and P. Séébold. Avoidability of circular formulas. *Theor. Comput. Sci.*, 726:1–4, 2018.

[10] J. Grytczuk. Pattern avoidance on graphs. *Discrete Math.*, 307(11-12):1341–1346, 2007.

[11] I. Mikhailova. On the avoidability index of palindromes. *Matematicheskie Zametki.*, 93(4):634–636, 2013.

[12] P. Ochem. A generator of morphisms for infinite words. *RAIRO - Theoret. Informatics Appl.*, 40:427–441, 2006.

[13] P. Ochem. Doubled patterns are 3-avoidable. *Electron. J. Combinatorics.*, 23(1):#P1.19, 2016.

[14] P. Ochem and M. Rosenfeld. Avoidability of formulas with two variables. *Electron. J. Combin.*, 24(4):#P4.30, 2017.

[15] A. Thue. Über unendliche Zeichenreihen. *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania*, 7:1–22, 1906.

[16] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania,*, 10:1–67, 1912.