On Some Interesting Ternary Formulas

Pascal Ochem[1(✉)] and Matthieu Rosenfeld[2]

[1] LIRMM, CNRS, Université de Montpellier, Montpellier, France
ochem@lirmm.fr

[2] LIP, ENS de Lyon, CNRS, UCBL, Université de Lyon, Lyon, France
matthieu.rosenfeld@ens-lyon.fr

Abstract. We show that, up to renaming of the letters, the only infinite ternary words avoiding the formula $ABCAB.ABCBA.ACB.BAC$ (resp. $ABCA.BCAB.BCB.CBA$) have the same set of recurrent factors as the fixed point of $\mathtt{0} \mapsto \mathtt{012}$, $\mathtt{1} \mapsto \mathtt{02}$, $\mathtt{2} \mapsto \mathtt{1}$.

Also, we show that the formula $ABAC.BACA.ABCA$ is 2-avoidable. Finally, we show that the pattern $ABACADABCA$ is unavoidable for the class of C_4-minor-free graphs with maximum degree 3. This disproves a conjecture of Grytczuk.

Keywords: Combinatorics on words · Pattern avoidance

1 Introduction

A *pattern* p is a non-empty finite word over an alphabet $\Delta = \{A, B, C, \ldots\}$ 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 [2], we associate a pattern p with the *formula* f obtained by replacing every isolated variable in p by a dot. The factors between the dots are called *fragments*.

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 $\lambda(p) \leq \lambda(f)$. Recall that an infinite word is *recurrent* if every finite factor appears infinitely many times. If there exists an infinite word over Σ avoiding p, then there exists an infinite recurrent word over Σ avoiding p. This recurrent word 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

This work was partially supported by the ANR project CoCoGro (ANR-16-CE40-0005).

S. Brlek et al. (Eds.): WORDS 2017, LNCS 10432, pp. 30–35, 2017.
DOI: 10.1007/978-3-319-66396-8_4

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' by 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.

We say that two infinite words are equivalent if they have the same set of factors. Let b_3 be the fixed point of `0` $\mapsto$ `012`, `1` $\mapsto$ `02`, `2` $\mapsto$ `1`. A famous result of Thue [1,4,5] can be stated as follows:

Theorem 1 [1,4,5]**.** *Every bi-infinite ternary word avoiding AA, `010`, and `212` is equivalent to b_3.*

In Sect. 2, we obtain a similar result for b_3 by forbidding one ternary formula but without forbidding explicit factors in Σ_3^*.

In the remainder of the paper, we discuss a counterexample to a conjecture of Grytczuk stating that every avoidable pattern can be avoided on graphs with an alphabet of size that depends only on the maximum degree of the graph.

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.

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

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 reverse, Theorem 2 also holds for the 72 reverse ternary formulas.

Proof. For $1 \le i \le 4$, f_b contains an occurrence of f_i. Thus, every word avoiding f_i also avoids f_b. Using Cassaigne's algorithm, we have checked that b_3 avoids f_i. 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. Suppose for 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 it is well-known, 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.

Suppose for contradiction that w contains 010. Since w is square-free, w contains 20102. Moreover, w contains the factor of 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.

Suppose for contradiction that w contains 212. Since w is square-free, w contains 02120. Moreover, w contains the factor of 021020 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*

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 $v' \in S$.

Let $c(n) = |S \cup \Sigma_2^n|$ denote the factor complexity of S. By construction of S,

- $c(3n) = 6c(n)$ for $n \geq 3$,
- $c(3n+1) = 4c(n) + 2c(n+1)$ for $n \geq 3$,
- $c(3n+2) = 2c(n) + 4c(n+1)$ for $n \geq 2$.

Thus $c(n) = \Theta\left(n^{\ln 6/\ln 3}\right) = \Theta\left(n^{1+\ln 2/\ln 3}\right)$.

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, 1001101001, and 0110010110 have length 53. So we assume without loss of generality that w contains 1001101001.

Suppose for contradiction that w contains 1100. Since w avoids AAA, w contains 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 1100.

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

Suppose for contradiction that w contains 0101. Since w avoids 01010 and 10101, w contains 001011. Moreover, w avoids AAA, so w contains 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 0101.

A binary word is a factor of the m_a-image of some binary word if and only if it avoids $\{\texttt{000}, \texttt{111}, \texttt{0101}, \texttt{1100}\}$. Indeed, both kinds of binary words are characterized by the same Rauzy graph with vertex set $\Sigma_2^3 \setminus \{\texttt{000}, \texttt{111}\}$. So 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 $ABAC.BACA.ABCA$. We suppose for contradiction that a finite word $w \in S$ avoids $ABAC.BACA.ABCA$ and that $m_a(w)$ contains an occurrence h of $ABAC.BACA.ABCA$.

The word $m_a(w)$ is of the form $\diamond\texttt{01}\diamond\texttt{01}\diamond\texttt{01}\diamond\texttt{01}\ldots$. Thus, in $m_a(w)$:

- Every factor 00 is in position $0 \pmod 3$.
- Every factor 01 is in position $1 \pmod 3$.
- Every factor 11 is in position $2 \pmod 3$.
- Every factor 10 is in position 0 or $2 \pmod 3$, depending on whether a factor $\texttt{1}\diamond\texttt{0}$ is 100 or 110.

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

First, we consider the case such that $h(A)$ is gentle. This implies that the distance between two occurrences of $h(A)$ is $0 \pmod 3$. Because 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 such that $h(B)$ 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 $h(CA) = \texttt{10}$, that is, $h(C) = \texttt{1}$ and $h(A) = \texttt{0}$. Thus, $m_a(w)$ contains both $h(BAC) = h(B)\texttt{01}$ and $h(BCA) = h(B)\texttt{10}$. Since $h(B)$ is gentle, this implies that 01 and 10 have the same position modulo 3, which is impossible.

The case such that $h(C)$ 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 $h(A) = \texttt{1}$ and $h(B) = \texttt{0}$. Thus, $m_a(w)$ contains both $h(ABC) = \texttt{01}h(C)$ and $h(BAC) = \texttt{10}h(C)$. 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 $2|h(A)|+|h(B)|+|h(C)| \le 8$. 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.

4 A Counter-Example to a Conjecture of Grytczuk

Grytczuk [3] has 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, $\lambda(p, G)$ is the smallest number of colors needed to color the vertices of G such that every non-intersecting path in G induces a word avoiding p.

We think that the natural framework is that of directed graphs, and we consider only non-intersecting paths that are oriented from a starting vertex to an ending vertex. 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 $i \ge 0$. The directed graphs that we consider have no loops and no multiple arcs, since they do not modify the set of non-intersecting oriented paths. However, opposite arcs (i.e., digons) are allowed. Thus, an undirected graph is viewed as 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 are nitpicking about 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, $\lambda(ABCACB) = \lambda\left(ABCACB, \overrightarrow{P}\right) = 2$ and $\lambda(ABCACB, P) = 3$.

We do not attempt the hazardous task of defining a notion of avoidance for formulas on graphs.

A conjecture of Grytczuk [3] says that for every avoidable pattern p, there exists a function g such that $\lambda(p, G) \leq g(\Delta(G))$, where G is an undirected graph and $\Delta(G)$ denotes its maximum degree. Grytczuk [3] obtained that his conjecture holds for doubled patterns.

As a counterexample, we consider the pattern $ABACADABCA$ which is 2-avoidable by the result in the previous section. Of course, $ABACADABCA$ is not doubled because of the 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 $i \geq 0$. Notice that $\overrightarrow{G}$ is obtained from $\overrightarrow{P}$ by adding the arcs $\overrightarrow{v_{100i} v_{100i+2}}$. 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 $i \geq 0$. 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 $\overrightarrow{G}$, which has maximum degree 3.

References

1. Berstel, J.: Axel Thue's papers on repetitions in words: a translation, vol. 20. Publications du LACIM. Université du Québec à Montréal (1994)
2. Cassaigne, J.: Motifs évitables et régularité dans les mots. Ph.D. thesis, Université Paris VI (1994)
3. Grytczuk, J.: Pattern avoidance on graphs. Discrete Math. **307**(11–12), 1341–1346 (2007)
4. Thue, A.: Über unendliche Zeichenreihen. Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania **7**, 1–22 (1906)
5. Thue, A.: Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania **10**, 1–67 (1912)