Avoidability of Formulas with Two Variables

Pascal Ochem, Matthieu Rosenfeld

Avoidability of Formulas with two Variables

Pascal Ochem[1] and Matthieu Rosenfeld[2]

[1] LIRMM, CNRS, Univ. Montpellier
Montpellier, France
ochem@lirmm.fr
[2] LIP, ENS de Lyon, CNRS, UCBL, Université de Lyon
Lyon, France
matthieu.rosenfeld@ens-lyon.fr

Abstract. In combinatorics on words, a word w over an alphabet Σ is said to avoid a pattern p over an alphabet Δ of variables if there is no factor f of w such that $f = h(p)$ where $h : \Delta^* \to \Sigma^*$ is a non-erasing morphism. A pattern p is said to be k-avoidable if there exists an infinite word over a k-letter alphabet that avoids p. We consider the patterns such that at most two variables appear at least twice, or equivalently, the formulas with at most two variables. For each such formula, we determine whether it is 2-avoidable.

Keywords: Word, 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. Bean, Ehrenfeucht, and McNulty [3] and Zimin [11] characterized unavoidable patterns, i.e., such that $\lambda(p) = \infty$. We say that a pattern p is *t-avoidable* if $\lambda(p) \leqslant t$. For more informations on pattern avoidability, we refer to Chapter 3 of Lothaire's book [6].

A variable that appears only once in a pattern is said to be *isolated*. Following Cassaigne [4], we associate to a pattern p 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 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 an infinite word containing no occurrence of p. Clearly, 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. If there exists an infinite word over Σ avoiding p, then there 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 another fragment.

Cassaigne [4] began and Ochem [7] finished the determination of the avoidability index of every pattern with at most 3 variables. A *doubled* pattern contains every variable at least twice. Thus, a doubled pattern is a formula with exactly one fragment. Every doubled pattern is 3-avoidable [9]. A formula is said to be *binary* if it has at most 2 variables. In this paper, we determine the avoidability index of every binary formula.

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 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 and thus $\lambda(f) \leqslant \lambda(f')$. For example, the fact that $ABA.AABB$ is 2-avoidable implies that $ABAABB$ and $ABAB.BBAA$ are 2-avoidable. Moreover, the reverse f^R of a formula f satisfies $\lambda(f^R) = \lambda(f)$. See Cassaigne [4] and Clark [5] for more information on formulas and divisibility.

First, we check that every avoidable binary formula is 3-avoidable. Since $\lambda(AA) = 3$, every formula containing a square is 3-avoidable. Then, the only square free avoidable binary formula is $ABA.BAB$ with avoidability index 3 [4]. Thus, we have to distinguish between avoidable binary formulas with avoidability index 2 and 3. A binary formula is minimally 2-avoidable if it is 2-avoidable and is not divisible by any other 2-avoidable binary formula. A binary formula f is maximally 2-unavoidable if it is 2-unavoidable and every other binary formula that is divisible by f is 2-avoidable.

Theorem 1. *Up to symmetry, the maximally 2-unavoidable binary formulas are:*

- $AAB.ABA.ABB.BBA.BAB.BAA$
- $AAB.ABBA$
- $AAB.BBAB$
- $AAB.BBAA$
- $AAB.BABB$
- $AAB.BABAA$
- $ABA.ABBA$
- $AABA.BAAB$

Up to symmetry, the minimally 2-avoidable binary formulas are:

- $AA.ABA.ABBA$
- $ABA.AABB$
- $AABA.ABB.BBA$
- $AA.ABA.BABB$
- $AA.ABB.BBAB$
- $AA.ABAB.BB$
- $AA.ABBA.BAB$
- $AAB.ABB.BBAA$
- $AAB.ABBA.BAA$
- $AABB.ABBA$
- $ABAB.BABA$

- $AABA.BABA$
- AAA
- $ABA.BAAB.BAB$
- $AABA.ABAA.BAB$
- $AABA.ABAA.BAAB$
- $ABAAB$

To obtain the 2-unavoidability of the formulas in the first part of Theorem 1, we use a standard backtracking algorithm. In the rest of the paper, we consider the 2-avoidable formulas in the second part of Theorem 1. Fig. 1 gives the maximal length and number of binary words avoiding each maximally 2-unavoidable formula.

We show in Section 3 that the first three of these formulas are avoided by polynomially many binary words only. The proof uses a technical lemma given in Section 2. Then we show in Section 4 that the other formulas are avoided by exponentially many binary words.

Fig. 1: The number and maximal length of binary words avoiding the maximally 2-unavoidable formulas.

Formula	Maximal length of a binary word avoiding this formula	Number of binary words avoiding this formula
$AAB.BBAA$	22	1428
$AAB.ABA.ABB.BBA.BAB.BAA$	23	810
$AAB.BBAB$	23	1662
$AABA.BAAB$	26	2124
$AAB.ABBA$	30	1684
$AAB.BABAA$	42	71002
$AAB.BABB$	69	9252
$ABA.ABBA$	90	31572

2 The Useful Lemma

Let us define the following words:

- b_2 is the fixed point of $\texttt{0} \mapsto \texttt{01}$, $\texttt{1} \mapsto \texttt{10}$.
- b_3 is the fixed point of $\texttt{0} \mapsto \texttt{012}$, $\texttt{1} \mapsto \texttt{02}$, $\texttt{2} \mapsto \texttt{1}$.
- b_4 is the fixed point of $\texttt{0} \mapsto \texttt{01}$, $\texttt{1} \mapsto \texttt{03}$, $\texttt{2} \mapsto \texttt{21}$, $\texttt{3} \mapsto \texttt{23}$.
- b_5 is the fixed point of $\texttt{0} \mapsto \texttt{01}$, $\texttt{1} \mapsto \texttt{23}$, $\texttt{2} \mapsto \texttt{4}$, $\texttt{3} \mapsto \texttt{21}$, $\texttt{4} \mapsto \texttt{0}$.

Let w and w' be infinite (right infinite or bi-infinite) words. We say that w and w' are equivalent if they have the same set of finite factors. We write $w \sim w'$ if w and w' are equivalent. A famous result of Thue [10] can be stated as follows:

Theorem 2. *[10] Every bi-infinite ternary word avoiding* 010*,* 212*, and squares is equivalent to* b_3*.*

Given an alphabet Σ and forbidden structures S, we say that a finite set W of infinite words over Σ *essentially avoids* S if every word in W avoids S and every bi-infinite words over Σ avoiding S is equivalent to one of the words in S. If W contains only one word w, we denote the set W by w instead of $\{w\}$. Then we can restate Theorem 2: b_3 essentially avoids 010, 212, and squares

The results in the next section involve b_3. We have tried without success to prove them by using Theorem 2. We need the following stronger property of b_3:

Lemma 3. b_3 *essentially avoids* 010*,* 212*,* XX *with* $1 \leqslant |X| \leqslant 3$*, and* $2YY$ *with* $|Y| \geqslant 4$*.*

Proof. We start by checking by computer that b_3 has the same set of factors of length 100 as every bi-infinite ternary word avoiding 010, 212, XX with $1 \leqslant |X| \leqslant 3$, and $2YY$ with $|Y| \geqslant 4$. The set of the forbidden factors of b_3 of length at most 4 is $F = \{00, 11, 22, 010, 212, 0202, 2020, 1021, 1201\}$. To finish the proof, we use Theorem 2 and we suppose for contradiction that w is a bi-infinite ternary word that contains a large square MM and avoids both F and large factors of the form $2YY$.

- Case $M = 0N$. Then w contains $MM = 0N0N$. Since $00 \in F$ and $2YY$ is forbidden, w contains $10N0N$. Since $\{11, 010\} \subset F$, w contains $210N0N$. If $N = P1$, then w contains $210P10P1$, which contains $2YY$ with $Y = 10P$. So $N = P2$ and w contains $210P20P2$. If $P = Q1$, then w contains $210Q120Q12$. Since $\{11, 212\} \subset F$, the factor $Q12$ implies that $Q = R0$ and w contains $210R0120R012$. Moreover, since $\{00, 1201\} \subset F$, the factor $120R$ implies that $R = 2S$ and w contains $2102S01202S012$. Then there is no possible prefix letter for S: 0 gives 2020, 1 gives 1021, and 2 gives 22. This rules out the case $P = Q1$. So $P = Q0$ and w contains $210Q020Q02$. The factor $Q020Q$ implies that $Q = 1R1$, so that w contains $2101R10201R102$. Since $\{11, 010\} \subset F$, the factor $01R$ implies that $R = 2S$, so that w contains $21012S102012S102$. The only possible right extension with respect to F of 102 is 102012. So w contains $21012S102012S102012$, which contains $2YY$ with $Y = S102012$.
- Case $M = 1N$. Then w contains $MM = 1N1N$. In order to avoid 11 and $2YY$, w must contain $01N1N$. If $N = P0$, then w contains $01P01P0$. So w contains the large square $01P01P$ and this case is covered by the previous item. So $N = P2$ and w contains $01P21P2$. Then there is no possible prefix letter for P: 0 gives 010, 1 gives 11, and 2 gives 212.
- Case $M = 2N$. Then w contains $MM = 2N2N$. If $N = P1$, then w contains $2P12P1$. This factor cannot extend to $2P12P12$, since this is $2YY$ with $Y = P12$. So w contains $2P12P10$. Then there is no possible suffix letter for P: 0 gives 010, 1 gives 11, and 2 gives 212. This rules out the case $N = P1$. So $N = P0$ and w contains $2P02P0$. This factor cannot extend to $02P02P0$, since this contains the large square $02P02P$ and this case is

covered by the first item. Thus w contains $\mathtt{12}P\mathtt{02}P\mathtt{0}$. If $P = Q\mathtt{1}$, then w contains $\mathtt{12}Q\mathtt{102}Q\mathtt{10}$. Since $\{\mathtt{22}, \mathtt{1021}\} \subset F$, the factor $\mathtt{102}Q$ implies that $Q = \mathtt{0}R$, so that w contains $\mathtt{120}R\mathtt{1020}R\mathtt{10}$. Then there is no possible prefix letter for R: $\mathtt{0}$ gives $\mathtt{00}$, $\mathtt{1}$ gives $\mathtt{1201}$, and $\mathtt{2}$ gives $\mathtt{0202}$. This rules out the case $P = Q\mathtt{1}$. So $P = Q\mathtt{2}$ and w contains $\mathtt{12}Q\mathtt{202}Q\mathtt{20}$. The factor $Q\mathtt{202}$ implies that $Q = R\mathtt{1}$ and w contains $\mathtt{12}R\mathtt{1202}R\mathtt{120}$. Since $\{\mathtt{00}, \mathtt{1201}\} \subset F$, w contains $\mathtt{12}R\mathtt{1202}R\mathtt{1202}$, which contains $\mathtt{2}YY$ with $Y = R\mathtt{1202}$.

3 Formulas Avoided by Few Binary Words

The first three 2-avoidable formulas in Theorem 1 are not avoided by exponentially many binary words:

- $\{g_x(b_3), g_y(b_3), g_z(b_3), g_{\overline{z}}(b_3)\}$ essentially avoids $AA.ABA.ABBA$.
- $\{g_x(b_3), g_t(b_3)\}$ essentially avoids $ABA.AABB$.
- $g_x(b_3)$ essentially avoids $AABA.ABB.BBA$.

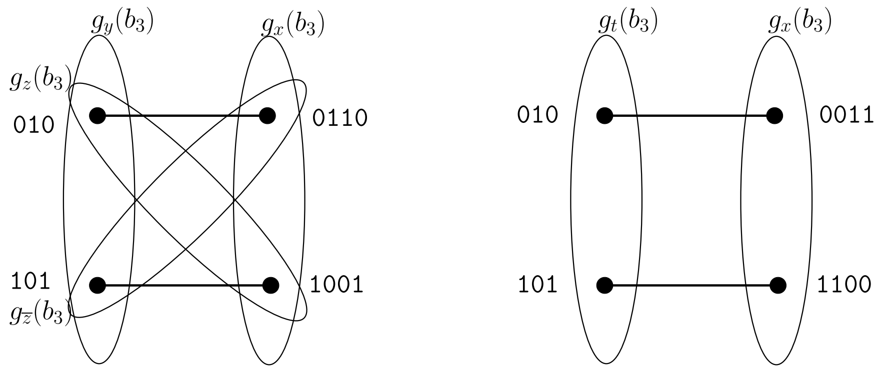
The words avoiding these formulas are morphic images of b_3 by the morphisms given below. Let $\overline{w}$ denote the word obtained from the (finite or bi-infinite) binary word w by exchanging $\mathtt{0}$ and $\mathtt{1}$. Obviously, if w avoids a given formula, then so does $\overline{w}$. A (bi-infinite) binary word w is *self-complementary* if $w \sim \overline{w}$. The words $g_x(b_3)$, $g_y(b_3)$, and $g_t(b_3)$ are self-complementary. Since the frequency of $\mathtt{0}$ in $g_z(b_3)$ is $\frac{5}{9}$, $g_z(b_3)$ is not self-complementary. Then $g_{\overline{z}}$ is obtained from g_z by exchanging $\mathtt{0}$ and $\mathtt{1}$, so that $g_{\overline{z}}(b_3) = \overline{g_z(b_3)}$.

$$
\begin{array}{llll}
g_x(\mathtt{0}) = \mathtt{01110}, & g_y(\mathtt{0}) = \mathtt{0111}, & g_z(\mathtt{0}) = \mathtt{0001}, & g_t(\mathtt{0}) = \mathtt{01011011010}, \\
g_x(\mathtt{1}) = \mathtt{0110}, & g_y(\mathtt{1}) = \mathtt{01}, & g_z(\mathtt{1}) = \mathtt{001}, & g_t(\mathtt{1}) = \mathtt{01011010}, \\
g_x(\mathtt{2}) = \mathtt{0}. & g_y(\mathtt{2}) = \mathtt{00}. & g_z(\mathtt{2}) = \mathtt{11}. & g_t(\mathtt{2}) = \mathtt{010}.
\end{array}
$$

To prove the avoidability, we have implemented Cassaigne's algorithm that decides, under mild assumptions, whether a morphic word avoids a formula [4]. For the first two formulas, we have to explain how the long enough binary words split into 4 or 2 distinct incompatible types. A similar phenomenon has been described for $AABB.ABBA$ [8].

First, consider any infinite binary word w avoiding $AA.ABA.ABBA$. A computer check shows by backtracking that w must contain the factor $\mathtt{01110001110}$. In particular, w contains $\mathtt{00}$. Thus, w cannot contain both $\mathtt{010}$ and $\mathtt{0110}$, since it would produce an occurrence of $AA.ABA.ABBA$. Moreover, a computer check shows by backtracking that w cannot avoid both $\mathtt{010}$ and $\mathtt{0110}$. So, w must contain either $\mathtt{010}$ or $\mathtt{0110}$ (this is an exclusive or). Similarly, w must contain either $\mathtt{101}$ or $\mathtt{1001}$. There are thus at most 4 possibilities for w, depending on which subset of $\{\mathtt{010}, \mathtt{0110}, \mathtt{101}, \mathtt{1001}\}$ appears among the factors of w, see Figure 2a.

Now, consider any infinite binary word w avoiding $ABA.AABB$. Notice that w cannot contain both $\mathtt{010}$ and $\mathtt{0011}$. Also, a computer check shows by backtracking that w cannot avoid both $\mathtt{010}$ and $\mathtt{1100}$. By symmetry, there are thus at most 2 possibilities for w, depending on which subset of $\{\mathtt{010}, \mathtt{0011}, \mathtt{101}, \mathtt{1100}\}$ appears among the factors of w, see Figure 2b.



(a) The four bi-infinite binary words avoiding $AA.ABA.ABBA$.

(b) The two bi-infinite binary words avoiding $ABA.AABB$.

Fig. 2

Let us first prove that $g_y(b_3)$ essentially avoids $AA.ABA.ABBA$, 0110, and 1001. We check that the set of prolongable binary words of length 100 avoiding $AA.ABA.ABBA$, 0110, and 1001 is exactly the set of factors of length 100 of $g_y(b_3)$. Using Cassaigne's notion of circular morphism [4], this is sufficient to prove that every bi-infinite binary word of this type is the g_y-image of some bi-infinite ternary word w_3. It also ensures that w_3 and b_3 have the same set of small factors. Suppose for contradiction that $w_3 \neq b_3$. By Lemma 3, w_3 contains $2YY$. Then w_3 contains $2YYa$ with $a \in \Sigma_3$. Notice that 0 is a prefix of the g_y-image of every letter. So $g_y(w_3)$ contains $g_y(2YYa) = 000U0U0V$ with $U, V \in \Sigma_3^+$, which contains an occurrence of $AA.ABA.ABBA$ with $A = 0$ and $B = 0U$. This shows that $w_3 \sim b_3$, and thus $g_y(w_3) \sim g_y(b_3)$. Thus $g_y(b_3)$ essentially avoids $AA.ABA.ABBA$, 0110, and 1001. The argument is similar for the other types and we only detail the final contradiction:

- Since 1 is a suffix of the g_z-image of every letter, $g_z(2YY) = 11U1U1$ contains an occurrence of $AA.ABA.ABBA$ with $A = 1$ and $B = 1U$.
- Since 010 is a prefix and a suffix of the g_t-image of every letter, $g_t(u2YY) = V010010010U010010U010$ contains an occurrence of $ABA.AABB$ with $A =$ 010 and $B = 010U010$.
- Since 0 is a prefix and a suffix of the g_x-image of every letter, $g_x(u2YYa) = V000U00U00W$ contains an occurrence of $AABA.AABBA$ with $A = 0$ and $B = 0U0$. Therefore, $g_x(u2YYa)$ contains an occurrence of $AA.ABA.ABBA$, $ABA.AABB$, and $AABA.ABB.BBA$.

4 Formulas Avoided by Exponentially Many Binary Words

The other 2-avoidable formulas in Theorem 1 are avoided by exponentially many binary words. For every such formula f, we give below a uniform morphism g that maps every ternary square free word to a binary word avoiding f. If possible, we simultaneously avoid the reverse formula f^R of f. We also avoid large squares. Let SQ_t denote the pattern corresponding to squares of period at least t, that is, $SQ_1 = AA$, $SQ_2 = ABAB$, $SQ_3 = ABCABC$, and so on. The morphism g produces words avoiding SQ_t with t as small as possible.

- $AA.ABA.BABB$ is avoided with its reverse by the following 22-uniform morphism which also avoids SQ_6:

$$\begin{aligned} 0 &\mapsto \mathtt{0001101101110011100011} \\ 1 &\mapsto \mathtt{0001101101110001100011} \\ 2 &\mapsto \mathtt{0001101101100011100111} \end{aligned}$$

  Notice that $\{AA.ABA.BABB, AA.ABA.BBAB, SQ_5\}$ is 2-unavoidable. However, $\{AA.ABA.BABB, SQ_4\}$ is 2-avoidable:

$$\begin{aligned} 0 &\mapsto \mathtt{0001001001100011100100110001001110010010011\,1} \\ 1 &\mapsto \mathtt{00010010011000100111001001100011100100100111} \\ 2 &\mapsto \mathtt{00010010011000100111001001001100011100100111} \end{aligned}$$

- $AA.ABB.BBAB$ is avoided with its reverse, 60-uniform morphism, avoids SQ_{11}:

$$\begin{aligned} 0 &\mapsto \mathtt{000110011100011001110011000111000110011100011100110001110011} \\ 1 &\mapsto \mathtt{000110011100011001110001110011000111000110011100110001110011} \\ 2 &\mapsto \mathtt{000110011100011001110001100111000111001100011100110001110011} \end{aligned}$$

  Notice that $\{AA.ABB.BBAB, SQ_{10}\}$ is 2-unavoidable.
- $AA.ABAB.BB$ is self-reverse, 11-uniform morphism, avoids SQ_4:

$$\begin{aligned} 0 &\mapsto \mathtt{00100110111} \\ 1 &\mapsto \mathtt{00100110001} \\ 2 &\mapsto \mathtt{00100011011} \end{aligned}$$

- $AA.ABBA.BAB$ is self-reverse, 30-uniform morphism, avoids SQ_6:

$$\begin{aligned} 0 &\mapsto \mathtt{000110001110011000110011100111} \\ 1 &\mapsto \mathtt{000110001100111001100011100111} \\ 2 &\mapsto \mathtt{000110001100011001110011100111} \end{aligned}$$

- $AAB.ABB.BBAA$ is self-reverse, 30-uniform morphism, avoids SQ_5:

$$\begin{aligned} 0 &\mapsto \mathtt{000100101110100010110111011101} \\ 1 &\mapsto \mathtt{000100101101110100010111011101} \\ 2 &\mapsto \mathtt{000100010001011101110111010001} \end{aligned}$$

- $AAB.ABBA.BAA$ is self-reverse, 38-uniform morphism, avoids SQ_5:

$$\begin{aligned} 0 &\mapsto \mathtt{00010001000101110111010001011100011101} \\ 1 &\mapsto \mathtt{00010001000101110100011100010111011101} \\ 2 &\mapsto \mathtt{00010001000101110001110100010111011101} \end{aligned}$$

- $AABB.ABBA$ is unavoidable with its reverse, 193-uniform morphism, avoids SQ_{16}:

  0 $\mapsto$ 00010001011011101100010110111000101101110111000101100010001011
  011101100010110111011100010110111011000101101110001011011101110001
  01100010001011011100010110111011100010110111011000101101110001011
  1 $\mapsto$ 00010001011011101100010110111000101101110111000101100010001011
  011100010110111011100010110111011000101101110001011011101110001011
  00010001011011101100010110111011100010110111011000101101110001011
  2 $\mapsto$ 00010001011011100010110111011100010110001000101101110110001011
  011101110001011011101100010110111000101101110111000101100010001011
  01110110001011011100010110111011100010110111011000101101110001011

  Previous papers [7,8] have considered a 102-uniform morphism to avoid $AABB.ABBA$ and SQ_{27}. No infinite binary word avoids $AABB.ABBA$ and SQ_{15}.
- $ABAB.BABA$ is self-reverse, 50-uniform morphism, avoids SQ_3, see [7]:

  0 $\mapsto$ 00011001011000111001011001110001011100101100010111
  1 $\mapsto$ 00011001011000101110010110011100010110001110010111
  2 $\mapsto$ 00011001011000101110010110001110010111000101100111

  Notice that a binary word avoiding $ABAB.BABA$ and SQ_3 contains only the squares 00, 11, and 0101 (or 00, 11, and 1010).
- $AABA.BABA$: A case analysis of the small factors shows that a recurrent binary word avoids $AABA.BABA$, $ABAA.ABAB$, and SQ_3 if and only if it contains only the squares 00, 11, and 0101 (or 00, 11, and 1010). We thus obtain the same morphism as for $ABAB.BABA$.
- AAA is self-reverse, 32-uniform morphism, avoids SQ_4:

  0 $\mapsto$ 00101001101101001011001001101011
  1 $\mapsto$ 00101001101100101101001001101011
  2 $\mapsto$ 00100101101001001101101001011011

- $ABA.BAAB.BAB$ is self-reverse, 10-uniform morphism, avoids SQ_3:

  0 $\mapsto$ 0001110101
  1 $\mapsto$ 0001011101
  2 $\mapsto$ 0001010111

- $AABA.ABAA.BAB$ is self-reverse, 57-uniform morphism, avoids SQ_6:

  0 $\mapsto$ 000101011100010110010101100010111001011000101011100101011
  1 $\mapsto$ 000101011100010110010101100010101110010110001011100101011
  2 $\mapsto$ 000101011100010110010101100010101110010101100010111001011

- $AABA.ABAA.BAAB$ is self-reverse, 30-uniform morphism, avoids SQ_3:

  0 $\mapsto$ 000101110001110101000101011101
  1 $\mapsto$ 000101110001110100010101110101
  2 $\mapsto$ 000101110001010111010100011101

– $ABAAB$ is avoided with its reverse, 10-uniform morphism, avoids SQ_3, see [7]:

$$\begin{aligned}0 &\mapsto \texttt{0001110101}\\ 1 &\mapsto \texttt{0000111101}\\ 2 &\mapsto \texttt{0000101111}\end{aligned}$$

For every q-uniform morphism g above, we say that a binary word is an sqf-g-image if it is the g-image of a ternary square free word. Let us show that for every minimally 2-avoidable formula f and corresponding morphism g, every sqf-g-image avoids f.

We start by checking that every morphism is synchronizing, that is, for every letters $a, b, c \in \Sigma_3$, the factor $g(a)$ only appears as a prefix or a suffix in $g(bc)$.

For every morphism g, the sqf-g-images are claimed to avoid SQ_t with $2t < q$. Let us prove that SQ_t is avoided. We first check exhaustively that the sqf-g-images contain no square uu such that $t \leqslant |u| < 2q - 1$. Now suppose for contradiction that an sqf-g-image contains a square uu with $|u| \geqslant 2q - 1$. The condition $|u| \geqslant 2q - 1$ implies that u contains a factor $g(a)$ with $a \in \Sigma_3$. This factor $g(a)$ only appears as the g-image of the letter a because g is synchronizing. Thus the distance between any two factors u in an sqf-g-image is a multiple of q. Since uu is a factor of an sqf-g-image, we have $q \mid |u|$. Also, the center of the square uu cannot lie between the g-images of two consecutive letters, since otherwise there would be a square in the pre-image. The only remaining possibility is that the ternary square free word contains a factor $aXbXc$ with $a, b, c \in \Sigma_3$ and $X \in \Sigma_3^+$ such that $g(aXbXc) = bsYpsYpe$ contains the square $uu = sYpsYp$, where $g(X) = Y$, $g(a) = bs$, $g(b) = ps$, $g(c) = pe$. Then, we also have $a \neq b$ and $b \neq c$ since $aXbXc$ is square free. Then abc is square free and $g(abc) = bspspe$ contains a square with period $|s| + |p| = |g(a)| = q$. This is a contradiction since the sqf-g-images contain no square with period q.

Notice that f is not square free, since the only avoidable square free binary formula is $ABA.BAB$, which is not 2-avoidable. Now, we distinguish two kinds of formula. A formula is *easy* if every appearing variable is contained in at least one square. Every potential occurrence of an easy formula then satisfies $|A| < t$ and $|B| < t$ since SQ_t is avoided. The longest fragment of every easy formula has length 4. So, to check that the sqf-g-images avoids an easy formula, it is sufficient to consider the set of factors of the sqf-g-images with length at most $4(t - 1)$.

A *tough* formula is such that one of the variables is not contained in any square. The tough formulas have been named so that this variable is B. The tough formulas are $ABA.BAAB.BAB$, $ABAAB$, $AABA.ABAA.BAAB$, and $AABA.ABAA.BAB$. As before, every potential occurrence of a tough formula satisfies $|A| < t$ since SQ_t is avoided. Suppose for contradiction that $|B| \geqslant 2q-1$. By previous discussion, the distance between any two occurrences of B in an sqf-g-image is a multiple of q. The case of $ABA.BAAB.BAB$ can be settled as follows. The factor $BAAB$ implies that $q \mid |BAA|$ and the factor BAB implies that $q \mid |BA|$. This implies that $q \mid |A|$, which contradicts $|A| < t$. For the other formulas, only one fragment contains B twice. This fragment is said to be *important*. Since $|A| < t$, the important fragment is a repetition which is "almost" a square. The important fragment is $\boldsymbol{BAB}$ for $AABA.ABAA.BAB$, $\boldsymbol{BAAB}$ for $AABA.ABAA.BAAB$, and $\boldsymbol{ABAAB}$ for $ABAAB$. Informally, this almost square implies a factor $aXbXc$ in the ternary pre-image, such that $|a| = |c| = 1$ and $1 \leqslant |b| \leqslant 2$. If $|X|$ is small, then $|B|$ is small and we check exhaustively that there exists no small occurrence of f. If $|X|$ is large, there would exist a ternary square free factor $aYbYc$ with $|Y|$ small, such that $g(aYbYc)$ contains the important fragment of an occurrence of f if and only if $g(aXbXc)$ contains the important fragment of a smaller occurrence of f.

5 Concluding Remarks

From our results, every minimally 2-avoidable binary formula, and thus every 2-avoidable binary formula, is avoided by some morphic image of b_3.

What can we forbid so that there exists only few infinite avoiding words ? The known examples from the literature [1,2,10] are:

- one pattern and two factors:
  - b_3 essentially avoids AA, `010`, and `212`.
  - A morphic image of b_5 essentially avoids AA, `010`, and `020`.
  - A morphic image of b_5 essentially avoids AA, `121`, and `212`.
  - b_2 essentially avoids $ABABA$, `000`, and `111`.
- two patterns: b_2 essentially avoids $ABABA$ and AAA.
- one formula over three variables: b_4 and two words from b_4 obtained by letter permutation essentially avoid $AB.AC.BA.BC.CA$.

Now we can extend this list:

- one formula over two variables:
  - $g_x(b_3)$ essentially avoids $AAB.BAA.BBAB$.
  - $\{g_x(b_3), g_t(b_3)\}$ essentially avoids $ABA.AABB$.
  - $\{g_x(b_3), g_y(b_3), g_z(b_3), g_{\overline{z}}(b_3)\}$ essentially avoids $AA.ABA.ABBA$.
- one pattern over three variables: $ABACAABB$ (same as $ABA.AABB$).

References

1. G. Badkobeh and P. Ochem. Characterization of some binary words with few squares. *Theoret. Comput. Sci.*, 588:73–80, 2015.
2. K. A. Baker, G. F. McNulty, and W. Taylor. Growth problems for avoidable words. *Theoretical Computer Science*, 69(3):319 – 345, 1989.
3. D.R. Bean, A. Ehrenfeucht, and G.F. McNulty. Avoidable patterns in strings of symbols. *Pacific J. of Math.*, 85:261–294, 1979.
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.
6. M. Lothaire. *Algebraic Combinatorics on Words.* Cambridge Univ. Press, 2002.

7. P. Ochem. A generator of morphisms for infinite words. *RAIRO - Theoret. Informatics Appl.*, 40:427–441, 2006.
8. P. Ochem. Binary words avoiding the pattern AABBCABBA. *RAIRO - Theoret. Informatics Appl.*, 44(1):151–158, 2010.
9. P. Ochem. Doubled patterns are 3-avoidable. *Electron. J. Combinatorics.*, 23(1), 2016.
10. A. Thue. Über unendliche Zeichenreihen. *'Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania*, 7:1–22, 1906.
11. A.I. Zimin. Blocking sets of terms. *Math. USSR Sbornik*, 47(2):353–364, 1984.