Avoidability of Formulas with Two Variables

Pascal Ochem, Matthieu Rosenfeld

Avoidability of formulas with two variables

Pascal Ochem[*] Matthieu Rosenfeld[†]

Submitted: Oct 12, 2016; Accepted: Oct 21, 2017; Published: Nov 3, 2017
Mathematics Subject Classifications: 68R15

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, and if it is 2-avoidable, we determine whether it is avoided by exponentially many binary words. [1]

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. Let $\Sigma_k = \{0, 1, \ldots, k-1\}$ denote the k-letter alphabet. The *avoidability index* $\lambda(p)$ of a pattern p is the size of the smallest integer k such that there exists an infinite word in Σ_k^* containing no occurrence of p. Bean, Ehrenfeucht, and McNulty [3] and Zimin [15] 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 [9].

A variable that appears only once in a pattern is said to be *isolated*. Following Cassaigne [5], 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

[*]LIRMM, CNRS, Université de Montpellier, France. ochem@lirmm.fr
[†]LIP, ENS de Lyon, CNRS, UCBL, Université de Lyon, France. matthieu.rosenfeld@ens-lyon.fr
[1]A previous version of this paper, without Theorem 2, has been presented at DLT 2016.
This work was partially supported by the ANR project CoCoGro (ANR-16-CE40-0005).

$\lambda(f)$ of a formula f is the size of the smallest alphabet allowing an infinite word containing no occurrence of f. 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 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 [5] began and Ochem [10] 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 [12]. 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 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 and thus $\lambda(f) \leqslant \lambda(f')$. Moreover, the reverse f^R of a formula f satisfies $\lambda(f^R) = \lambda(f)$. For example, the fact that $ABA.AABB$ is 2-avoidable implies that $ABAABB$ and $BAB.AABB$ are 2-avoidable. See Cassaigne [5] and Clark [6] for more information on formulas and divisibility. For convenience, we say that an avoidable formula f is *exponential* (resp. *polynomial*) if the number of words of length n avoiding f over $\lambda(f)$ letters is exponential (resp. polynomial) in n.

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 [5, 7]. 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:

- $AA.ABA.ABBA$ *(polynomial)*
- $ABA.AABB$ *(polynomial)*
- $AABA.ABB.BBA$ *(polynomial)*
- $AA.ABA.BABB$ *(exponential)*
- $AA.ABB.BBAB$ *(exponential)*
- $AA.ABAB.BB$ *(exponential)*
- $AA.ABBA.BAB$ *(exponential)*
- $AAB.ABB.BBAA$ *(exponential)*
- $AAB.ABBA.BAA$ *(exponential)*
- $AABB.ABBA$ *(exponential)*
- $ABAB.BABA$ *(exponential)*
- $AABA.BABA$ *(exponential)*
- AAA *(exponential)*
- $ABA.BAAB.BAB$ *(exponential)*
- $AABA.ABAA.BAB$ *(exponential)*
- $AABA.ABAA.BAAB$ *(exponential)*
- $ABAAB$ *(exponential)*

Given a binary formula f, we can use Theorem 1 to find $\lambda(f)$. Now, we also consider the problem whether an avoidable binary formula is polynomial or exponential. If $\lambda(f) = 3$, then either f contains a square and is thus exponential, or $f = ABA.BAB$. We will see in Section 5 that $ABA.BAB$ is exponential too. Thus, we consider only the case $\lambda(f) = 2$. If f is divisible by an exponential 2-avoidable formula given in Theorem 1, then f is known to be exponential. This leaves open the case such that f is only divisible by polynomial 2-avoidable formulas. The next result settles every open case.

Theorem 2.
The following formulas are polynomial:

- $BBA.ABA.AABB$
- $AABA.AABB$

The following formulas are exponential:

- $BAB.ABA.AABB$

- $AAB.ABA.ABBA$
- $BAA.ABA.AABB$
- $BBA.AABA.AABB$

1.1 Structure of the proofs

First, we check by computer that Theorem 1 is exhaustive, that is, for every avoidable binary formula f, either f or f^R divides at least one formula in the first list and thus f is 2-unavoidable, or f or f^R is divisible by at least one formula in the second list and thus f is 2-avoidable. Every binary pattern of length 6 (or its reverse) is divisible by at least one formula in the second list. So we only need to consider formulas such that the length of every fragment is at most 5.

Then, to obtain the 2-unavoidability of the formulas in the first part of Theorem 1, we use a standard backtracking algorithm. Figure 1 gives the maximal length and number of binary words avoiding each maximally 2-unavoidable formula.

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

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

There remain to show that the formulas in the second part of Theorem 1 are 2-avoidable. These formulas, together with the formulas in Theorem 2, characterize the frontier between polynomial and exponential 2-avoidable binary formulas.

The proof for polynomial formulas is in Section 3. It uses a technical lemma given in Section 2. The proof for exponential formulas is in Section 4.

Similarly, a computer check shows that for every 2-avoidable binary formula f, either f or f^R corresponds to one of the 5 polynomial formulas in Lemma 5, or f or f^R is divisible by at least one exponential formula appearing in Section 4.

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. Given an alphabet Σ and a set S of forbidden structures, 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 word over Σ avoiding S is equivalent to one of the words in W. If W contains only one word w, we denote the set W by w instead of $\{w\}$. A famous result of Thue can then be stated as follows:

Theorem 3 ([4, 14])**.** *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 3. We need the following stronger property of b_3:

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

Proof. We start by checking by computer that b_3 has the same set of factors of length exactly 100 as every bi-infinite ternary word avoiding 010, 212, XX with $1 \leqslant |X| \leqslant 3$, and 2$YY$ with $|Y| \geqslant 4$. We will use the set $F = \{\texttt{00}, \texttt{11}, \texttt{22}, \texttt{010}, \texttt{212}, \texttt{0202}, \texttt{2020}, \texttt{1021}, \texttt{1201}\}$ of ternary words of length at most 4 that are not factors of b_3.

To finish the proof, we use Theorem 3 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 $M = \texttt{0}N$. Then w contains $MM = \texttt{0}N\texttt{0}N$. Since $\texttt{00} \in F$ and 2YY is forbidden, w contains $\texttt{10}N\texttt{0}N$. Since $\{\texttt{11}, \texttt{010}\} \subset F$, w contains $\texttt{210}N\texttt{0}N$. If $N = P\texttt{1}$, then w contains $\texttt{210}P\texttt{10}P\texttt{1}$, which contains 2$YY$ with $Y = 10P$. So $N = P\texttt{2}$ and w contains $\texttt{210}P\texttt{20}P\texttt{2}$. If $P = Q\texttt{1}$, then w contains $\texttt{210}Q\texttt{120}Q\texttt{12}$. Since $\{\texttt{11}, \texttt{212}\} \subset F$, the factor $Q\texttt{12}$ implies that $Q = R\texttt{0}$ and w contains $\texttt{210}R\texttt{0120}R\texttt{012}$. Moreover, since $\{\texttt{00}, \texttt{1201}\} \subset F$, the factor $\texttt{120}R$ implies that $R = \texttt{2}T$ and w contains $\texttt{2102}T\texttt{01202}T\texttt{012}$. Then there is no possible prefix letter for S: 0 gives 2020, 1 gives 1021, and 2 gives 22. This rules out the case $P = Q\texttt{1}$. So $P = Q\texttt{0}$ and w contains $\texttt{210}Q\texttt{020}Q\texttt{02}$. The factor $Q\texttt{020}Q$ implies that $Q = 1R1$, so that w contains $\texttt{2101}R\texttt{10201}R\texttt{102}$. Since $\{\texttt{11}, \texttt{010}\} \subset F$, the factor $\texttt{01}R$ implies that $R = \texttt{2}T$, so that w contains $\texttt{21012}T\texttt{102012}T\texttt{102}$. The only possible right extension with respect to F of 102 is 102012. So w contains $\texttt{21012}T\texttt{102012}T\texttt{102012}$, which contains 2$YY$ with $Y = T\texttt{102012}$.

- Case $M = \mathtt{1}N$. Then w contains $MM = \mathtt{1}N\mathtt{1}N$. In order to avoid $\mathtt{11}$ and $\mathtt{2}YY$, w must contain $\mathtt{01}N\mathtt{1}N$. If $N = P\mathtt{0}$, then w contains $\mathtt{01}P\mathtt{01}P\mathtt{0}$. So w contains the large square $\mathtt{01}P\mathtt{01}P$ and this case is covered by the previous case. So $N = P\mathtt{2}$ and w contains $\mathtt{01}P\mathtt{21}P\mathtt{2}$. Then there is no possible prefix letter for P: $\mathtt{0}$ gives $\mathtt{010}$, $\mathtt{1}$ gives $\mathtt{11}$, and $\mathtt{2}$ gives $\mathtt{212}$.

- Case $M = \mathtt{2}N$. Then w contains $MM = \mathtt{2}N\mathtt{2}N$. If $N = P\mathtt{1}$, then w contains $\mathtt{2}P\mathtt{12}P\mathtt{1}$. This factor cannot extend to $\mathtt{2}P\mathtt{12}P\mathtt{12}$, since this is $\mathtt{2}YY$ with $Y = P\mathtt{12}$. So w contains $\mathtt{2}P\mathtt{12}P\mathtt{10}$. Then there is no possible suffix letter for P: $\mathtt{0}$ gives $\mathtt{010}$, $\mathtt{1}$ gives $\mathtt{11}$, and $\mathtt{2}$ gives $\mathtt{212}$. This rules out the case $N = P\mathtt{1}$. So $N = P\mathtt{0}$ and w contains $\mathtt{2}P\mathtt{02}P\mathtt{0}$. This factor cannot extend to $\mathtt{02}P\mathtt{02}P\mathtt{0}$, since this contains the large square $\mathtt{02}P\mathtt{02}P$ and this case is covered by the first case. 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 $Y = R\mathtt{1202}$. $\square$

3 Polynomial formulas

Let us detail the binary words avoiding the polynomial formulas in Theorems 1 and 2. Lemma 5 will show that they are images of b_3 by the morphisms g_x, g_y, g_z, and g_t defined as follows.

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

Let $\overline{w}$ denote the word obtained from the binary word w by exchanging $\mathtt{0}$ and $\mathtt{1}$. Obviously, if w avoids a given formula, then so does $\overline{w}$. An infinite or b-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 $g_{\overline{z}}(b_3) = \overline{g_z(b_3)}$.

The aim of this section is to prove the following result.

Lemma 5.

- $\{g_x(b_3), g_y(b_3), g_z(b_3), g_{\overline{z}}(b_3)\}$ *essentially avoids $AA.ABA.ABBA$.*

- $g_x(b_3)$ *essentially avoids $AABA.ABB.BBA$.*

- *Let f be either $ABA.AABB$, $BBA.ABA.AABB$, or $AABA.AABB$. Then $\{g_x(b_3), g_t(b_3)\}$ essentially avoids f.*

Let us first state interesting properties of the morphisms and the formulas in Lemma 5.

Lemma 6. *For every $p, s \in \Sigma_3$, $Y \in \Sigma_3^*$ with $|Y| \geqslant 4$, and $g \in \{g_x, g_y, g_z, g_{\overline{z}}, g_t\}$, the word $g(p2YYs)$ contains an occurrence of $AABA.AABBA$.*

Proof.

- Since $\texttt{0}$ is a prefix and a suffix of the g_x-image of every letter, $g_x(p2YYs) = V\texttt{000}U\texttt{00}U\texttt{00}W$ contains an occurrence of $AABA.AABBA$ with $A = \texttt{0}$ and $B = \texttt{0}U\texttt{0}$.

- Since $\texttt{0}$ is a prefix of the g_y-image of every letter, $g_y(2YYs) = \texttt{000}U\texttt{0}U\texttt{0}V$ contains an occurrence of $AABA.AABBA$ with $A = \texttt{0}$ and $B = \texttt{0}U$.

- Since $\texttt{1}$ is a suffix of the g_z-image of every letter, $g_z(p2YY) = V\texttt{111}U\texttt{1}U\texttt{1}$ contains an occurrence of $AABA.AABBA$ with $A = \texttt{1}$ and $B = \texttt{1}U$.

- Since $g_{\overline{z}}(p2YY) = \overline{g_z(p2YY)}$, $g_{\overline{z}}(p2YY)$ contains an occurrence of $AABA.AABBA$.

- Since $\texttt{010}$ is a prefix and a suffix of the g_t-image of every letter, $g_t(p2YYs) = V\texttt{010010010}U\texttt{010010}U\texttt{010010}W$ contains an occurrence of $AABA.AABBA$ with $A = \texttt{010}$ and $B = \texttt{010}U\texttt{010}$. $\square$

The following observation explains why $AABA.AABBA$ is considered in Lemma 6.

Observation 7. *$AABA.AABBA$ is divisible by every formula in Lemma 5.*

We are now ready to prove Lemma 5. To prove the avoidability, we have implemented Cassaigne's algorithm that decides, under mild assumptions, whether a morphic word avoids a formula [5]. We have to explain how the long enough binary words avoiding a formula can be split into 4 or 2 distinct incompatible types. A similar phenomenon has been described for $AABB.ABBA$ [11].

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

Also, consider any infinite binary word w avoiding f, where f is either $ABA.AABB$, $BBA.ABA.AABB$, or $AABA.AABB$. Notice that the formulas $BBA.ABA.AABB$ and $AABA.AABB$ are divisible by $ABA.AABB$. We check by backtracking that w cannot avoid both $\texttt{0010}$ and $\texttt{00110}$. A word containing both $\texttt{0010}$ and $\texttt{00110}$ contains an occurrence of $AABA.AABBA$, and by Observation 7, it also contains an occurrence of f. So w does not contain both $\texttt{0010}$ and $\texttt{00110}$. Thus, there are two possibilities for w depending on whether it contains $\texttt{0010}$ or $\texttt{00110}$.

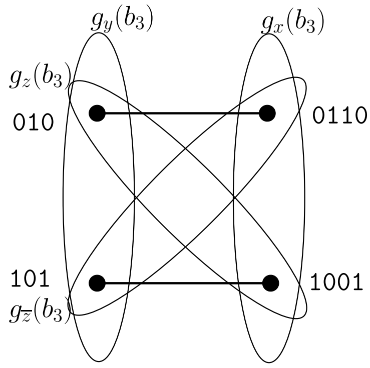

Figure 2: The four infinite binary words avoiding $AA.ABA.ABBA$.

Now our tasks of the form "prove that a set of morphic words essentially avoids one formula" are reduced to (more) tasks of the form "prove that one morphic word essentially avoids one formula and a finite set of factors".

Since all the proofs of such reduced tasks are very similar, we only detail the proof 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 [5], 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 $w_3 \not\sim b_3$. By Lemma 4, w_3 contains a factor $\mathtt{2}YY$ with $|Y| \geqslant 4$. Since w_3 is bi-infinite, w_3 even contains a factor $p\mathtt{2}YYs$ with $p, s \in \Sigma_3$. By Lemma 6, $g_y(w_3)$ contains an occurrence of $AABA.AABBA$. Then by Observation 7, $g_y(w_3)$ contains an occurrence of $AA.ABA.ABBA$. This contradiction shows that $w_3 \sim b_3$. So $g_y(b_3)$ essentially avoids $AA.ABA.ABBA$, 0110, and 1001.

4 Exponential formulas

Given a morphism $g : \Sigma_3^* \to \Sigma_2^*$, an sqf-$g$-image is the image by g of a (finite or infinite) ternary square-free word. With an abuse of language, we say that g avoids a set of formulas if every sqf-g-image avoids every formula in the set. For every 2-avoidable exponential formula f in Theorems 1 and 2, we give below a uniform morphism g that avoids 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 avoids SQ_t with t as small as possible. Since $\lambda(SQ_2) = 3$, a binary word avoiding SQ_3 is necessarily best possible in terms of length of avoided squares.

- $f = AA.ABA.BABB$. This 22-uniform morphism avoids $\left\{f, f^R, SQ_6\right\}$:

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

This 44-uniform morphism avoids $\{f, SQ_5\}$:

$$\begin{aligned}
0 &\mapsto \texttt{00010010011000111001001100010011100100100111}\\
1 &\mapsto \texttt{00010010011000100111001001100011100100100111}\\
2 &\mapsto \texttt{00010010011000100111001001001100011100100111}
\end{aligned}$$

Notice that $\left\{f, f^R, SQ_5\right\}$ is 2-unavoidable and $\{f, SQ_4\}$ is 2-unavoidable.

- $f = AA.ABB.BBAB$. This 60-uniform morphism avoids $\left\{f, f^R, SQ_{11}\right\}$:

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

Notice that $\{f, SQ_{10}\}$ is 2-unavoidable.

- $f = AA.ABAB.BB$ is self-reverse. This 11-uniform morphism avoids $\{f, SQ_4\}$:

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

Notice that $\{f, SQ_3\}$ is 2-unavoidable.

- $f = AA.ABBA.BAB$ is self-reverse. This 30-uniform morphism avoids $\{f, SQ_6\}$:

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

Notice that $\{f, SQ_5\}$ is 2-unavoidable.

- $f = AAB.ABB.BBAA$ is self-reverse. This 30-uniform morphism avoids $\{f, SQ_5\}$:

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

Notice that $\{f, SQ_4\}$ is 2-unavoidable.

- $f = AAB.ABBA.BAA$ is self-reverse. This 38-uniform morphism avoids $\{f, SQ_5\}$:

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

Notice that $\{f, SQ_4\}$ is 2-unavoidable.

- $f = AABB.ABBA$. This 193-uniform morphism avoids $\{f, SQ_{16}\}$:

0 $\mapsto$ 0001000101101110110001011011100010110111011100010
110001000101101110110001011011101110001011011 1011
0001011011100010110111011100010110001000101101110
0010110111011100010110111011000101101110001011

1 $\mapsto$ 0001000101101110110001011011100010110111011100010
1100010001011011100010110111011100010110111011000
1011011100010110111011100010110001000101101110110
0010110111011100010110111011000101101110001011

2 $\mapsto$ 0001000101101110001011011101110001011000100010110
1110110001011011101110001011011101100010110111000
1011011101110001011000100010110111011000101101110
0010110111011100010110111011000101101110001011

Notice that $\{f, f^R\}$ is 2-unavoidable and $\{f, SQ_{15}\}$ is 2-unavoidable. Previous papers [10, 11] have considered a 102-uniform morphism to avoid $\{f, SQ_{27}\}$.

- $f = ABAB.BABA$ is self-reverse. This 50-uniform morphism avoids $\{f, SQ_3\}$, see [10]:

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

Notice that a binary word avoiding $\{f, SQ_3\}$ contains only the squares 00, 11, and 0101 (or 00, 11, and 1010).

- $f = AABA.BABA$: A case analysis of the small factors shows that a recurrent binary word avoids $\{f, f^R, SQ_3\}$ if and only if it contains only the squares 00, 11, and 0101 (or 00, 11, and 1010). Thus, the previous 50-uniform morphism that avoids $\{ABAB.BABA, SQ_3\}$ also avoids $\{f, f^R, SQ_3\}$.

- $f = AAA$ is self-reverse. This 32-uniform morphism avoids $\{f, SQ_4\}$:

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

Notice that $\{f, SQ_3\}$ is 2-unavoidable.

- $f = ABA.BAAB.BAB$ is self-reverse. This 10-uniform morphism avoids $\{f, SQ_3\}$:

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

- $f = AABA.ABAA.BAB$ is self-reverse. This 57-uniform morphism avoids $\{f, SQ_6\}$:

$$\begin{aligned}
0 &\mapsto \texttt{000101011100010110010101100010111001011000101011100101011}\\
1 &\mapsto \texttt{000101011100010110010101100010101110010110001011100101011}\\
2 &\mapsto \texttt{000101011100010110010101100010101110010101100010111001011}
\end{aligned}$$

  Notice that $\{f, SQ_5\}$ is 2-unavoidable.

- $f = AABA.ABAA.BAAB$ is self-reverse. This 30-uniform morphism avoids $\{f, SQ_3\}$:

$$\begin{aligned}
0 &\mapsto \texttt{000101110001110101000101011101}\\
1 &\mapsto \texttt{000101110001110100010101110101}\\
2 &\mapsto \texttt{000101110001010111010100011101}
\end{aligned}$$

- $f = ABAAB$. This 10-uniform morphism avoids $\left\{f, f^R, SQ_3\right\}$, see [10]:

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

- $f = BAB.ABA.AABB$ is self-reverse. This 16-uniform morphism avoids $\{f, SQ_5\}$:

$$\begin{aligned}
0 &\mapsto \texttt{0101110111011101}\\
1 &\mapsto \texttt{0100010111010001}\\
2 &\mapsto \texttt{0001010111010100}
\end{aligned}$$

  Notice that $\{f, SQ_4\}$ is 2-unavoidable.

- $f = AAB.ABA.ABBA$ is avoided with its reverse. This 84-uniform morphism avoids $\left\{f, f^R, SQ_5\right\}$:

$$\begin{aligned}
0 \mapsto \; & \texttt{000100010111000111010001000101110111010001}\\
& \texttt{011100011101000101110111010001110001011101}\\
1 \mapsto \; & \texttt{000100010111000111010001000101110100011100}\\
& \texttt{010111011101000101110001110100010111011101}\\
2 \mapsto \; & \texttt{000100010111000111010001000101110100011100}\\
& \texttt{010111010001000101110001110100010111011101}
\end{aligned}$$

  Notice that $\{f, SQ_4\}$ is 2-unavoidable.

- $f = BAA.ABA.AABB$. This 304-uniform morphism avoids $\{f, SQ_7\}$:

$0 \mapsto$ 000110001100111000111001100011001110011100110001100011001110
1100011100011001110011100110001100111000111001100011000110011
1001100011100011001110011100110001100011001110001110011000110
0111001110011000111000110011100110001100011001110011100110001
100111000111001100011000110011100111001100011100011001110011
$1 \mapsto$ 000110001100111000111001100011001110011100110001100011001110
1100011100011001110011100110001100111000111001100011000110011
1001100011100011001110011100110001100011001110001110011000110
0111001110011000110001100111001100011100011001110011100110001
100111000111001100011000110011100111001100011100011001110011
$2 \mapsto$ 000110001100111000111001100011001110011100110001100011001110
1100011100011001110011100110001100011001110001110011000110011
1001110011000111000110011100110001100011001110011100110001100
1110001110011000110001100111001100011100011001110011100110001
100011001110001110011000110011100111001100011100011001110011

Using the morphism g_w below and the technique in [1], we can show that $g_w(b_3)$ essentially avoids $\{f, SQ_6\}$:

$g_w(0) =$ 01110011100111000110011100110001100011 0
$g_w(1) =$ 0111001110011000110001 10
$g_w(2) =$ 001110011000110

Notice that $\left\{f, f^R\right\}$ is 2-unavoidable and $\{f, SQ_5\}$ is 2-unavoidable.

- $f = BBA.AABA.AABB$. This 160-uniform morphism avoids $\left\{f, f^R, SQ_{21}\right\}$:

$0 \mapsto$ 00010110010111000101110010110001011100010110010111001 0
11000101110010110001011001011100101100010111000101100 1
01110001011100101100010110010111001011000101110010 11
$1 \mapsto$ 00010110010111000101110010110001011100010110010111001 0
11000101110010110001011001011100010110010111001011000 1
01110001011100101100010110010111001011000101110010 11
$2 \mapsto$ 00010110010111000101100101110010110001011100010110010 1
11000101110010110001011001011100101100010111000101110 0
10110001011001011100010110010111001011000101110010 11

This 202-uniform morphism avoids $\{f, SQ_5\}$:

$0 \mapsto$ 0001101001110110100011010100011101101001101101010001110110100011010100011101101010001101001110110100110110101000110100111011010100011101101000110101000111011010100011010011101101010001110110100110110101
$1 \mapsto$ 0001101001110110100011010100011101101001101101010001101001110110101000111011010001101010001110110101000110100111011010100011101101001101101010001101001110110100110110101000111011010001101010001110110101
$2 \mapsto$ 0001101001110110100011010100011101101001101101010001101001110110101000111011010001101010001110110101000110100111011010011011010100011101101000110101000111011010100011010011101101010001110110100110110101

Notice that $\{f, f^R, SQ_{20}\}$ is 2-unavoidable and $\{f, SQ_4\}$ is 2-unavoidable.

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 q-uniform morphism g, the sqf-g-images are claimed to avoid SQ_t with $2t < q$. Let us prove that SQ_t is avoided. We check exhaustively that the sqf-g-images contain no square uu such that $t \leqslant |u| \leqslant 2q - 2$. 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.

Let us show that for every formula f above and corresponding morphism g, g avoids f. Notice that f is not square-free, since the only avoidable square-free binary formula is $ABA.BAB$, which is not 2-avoidable. 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 g 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 formula is *tough* if 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 divides $|BAA|$ and the factor BAB implies that q divides $|BA|$. This implies that q divides $|A|$, which contradicts $|A| < t$. For the other tough 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 $\boldsymbol{B}A\boldsymbol{B}$ for $AABA.ABAA.BAB$, $\boldsymbol{B}AA\boldsymbol{B}$ for $AABA.ABAA.BAAB$, and $\boldsymbol{AB}A\boldsymbol{AB}$ 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 ABA.BAB is exponential

We know that $\lambda(ABA.BAB) = 3$. Cassaigne [5] shows that the image of b_4 by $\mathtt{0} \mapsto \mathtt{01}$, $\mathtt{1} \mapsto \mathtt{02}$, $\mathtt{2} \mapsto \mathtt{12}$, $\mathtt{3} \mapsto \mathtt{21}$ avoids $ABA.BAB$. Gamard et al. [7] show that the image of b_4 by $\mathtt{0} \mapsto \mathtt{0010}$, $\mathtt{1} \mapsto \mathtt{1122}$, $\mathtt{2} \mapsto \mathtt{0200}$, $\mathtt{3} \mapsto \mathtt{1212}$ avoids $ABA.BAB$.

By the results in the last two sections, we know the status (polynomial or exponential) of every 2-avoidable binary formula. As already mentioned, every binary formula f such that $\lambda(f) = 3$ either contains a square and is thus exponential, or is $ABA.BAB$. To finish the determination of the status of every binary formula, we show that $ABA.BAB$ is exponential.

Recall that a word is (α^+, ℓ)-free if it contains no repetition with period at least ℓ and exponent strictly greater than α. Using the general method described in [10], we check that the image of every $(7/4^+)$-free word over Σ_4 by the following 26-uniform morphism is $(3/2^+, 2)$-free and $(113/78^+, 3)$-free.

$$
\begin{aligned}
\mathtt{0} &\mapsto \mathtt{00121102200112021100220121} \\
\mathtt{1} &\mapsto \mathtt{00112200211001202210122021} \\
\mathtt{2} &\mapsto \mathtt{00112022110012200210120221} \\
\mathtt{3} &\mapsto \mathtt{00112002110012202101200221}
\end{aligned}
$$

By the $(113/78^+, 3)$-freeness, we only have to check that these binary words contain no occurrence h of $ABA.BAB$ such that $|h(A)| = |h(B)| = 1$. These binary words are interesting with respect to generalized repetition thresholds [8]. Since they are $(3/2^+, 2)$-free, they provide an alternative proof that $R(3, 2) = \frac{3}{2}$.

6 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. Let us summarize the known types of structure to forbid to have a set of essentially avoiding words.

- one pattern and two factors:
  - b_3 essentially avoids AA, `010`, and `212` [14].
  - A morphic image of b_5 essentially avoids AA, `010`, and `020` [1, 14].
  - A morphic image of b_5 essentially avoids AA, `121`, and `212` [1, 14].
  - b_2 essentially avoids $ABABA$, `000`, and `111` [14].
- two patterns: b_2 essentially avoids $ABABA$ and AAA [14].
- one formula over three variables:
  - b_3 and two words obtained from b_3 by letter permutation essentially avoid $ABCAB.ABCBA$ [13].
  - b_4 and two words obtained from b_4 by letter permutation essentially avoid $AB.AC.BA.BC.CA$ [2].
- one formula over two variables (see Lemma 5):
  - $g_x(b_3)$ 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 $ABA.AABB$).

Notice that every binary formula that admits only polynomially many avoiding words has a set of essentially avoiding words. We show that this is not the case for the ternary formula $ABACA.ABCA$ [13].

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. Berstel. Axel Thue's Papers on Repetitions in Words: a Translation. *Publications du Laboratoire de Combinatoire et d'Informatique Mathématique. Université du Québec à Montréal*, Number 20, February 1995.

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

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

[7] G. Gamard, P. Ochem, G. Richomme, and P. Séébold. Avoidability of circular formulas. arXiv:1610.04439

[8] L. Ilie, P. Ochem, and J.O. Shallit. A generalization of repetition threshold, *Theoret. Comput. Sci.*, 92(2):71–76, 2004.

[9] M. Lothaire. *Algebraic Combinatorics on Words.* Cambridge Univ. Press, 2002.

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

[11] P. Ochem. Binary words avoiding the pattern AABBCABBA. *RAIRO - Theoret. Informatics Appl.*, 44(1):151–158, 2010.

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

[13] P. Ochem and M. Rosenfeld. On some interesting ternary formulas. *WORDS 2017. LNCS 10432.* arXiv:1706.03233

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

[15] A. I. Zimin. Blocking sets of terms. *Math. USSR Sbornik*, 47(2):353–364, 1984.