Doubled patterns with reversal are 3-avoidable

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

January 7, 2022

Abstract

In combinatorics on words, a word $w$ over an alphabet $\Sigma$ is said to avoid a pattern $p$ over an alphabet $\Delta$ 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 $p$. A pattern is **doubled** if every variable occurs at least twice. Doubled patterns are known to be 3-avoidable. Currie, Mol, and Rampersad have considered a generalized notion which allows variable occurrences to be reversed. That is, $h(V^R)$ is the mirror image of $h(V)$ for every $V \in \Delta$. We show that doubled patterns with reversal are 3-avoidable. We also show that for every doubled pattern $p$, the growth rate of ternary words avoiding $p$ is at least the growth rate of ternary square-free words. A previous version of this paper containing only the first result has been presented at WORDS 2021.

1 Introduction

The **mirror image** of the word $w = w_1w_2\ldots w_n$ is the word $w^R = w_nw_{n-1}\ldots w_1$. A pattern with reversal $p$ is a non-empty word over an alphabet $\Delta = \{A, A^R, B, B^R, C, C^R, \ldots\}$ such that $\{A, B, C, \ldots\}$ are the **variables** of $p$. An **occurrence** of $p$ in a word $w$ is a non-erasing morphism $h : \Delta^* \to \Sigma^*$ satisfying $h(X^R) = (h(X))^R$ for every variable $X$ and such that $h(p)$ is a factor of $w$. The avoidability index $\lambda(p)$ of a pattern with reversal $p$ is the
size of the smallest alphabet $\Sigma$ such that there exists an infinite word $w$ over $\Sigma$ containing no occurrence of $p$. A pattern $p$ such that $\lambda(p) \leq k$ is said to be $k$-avoidable. To emphasize that a pattern is without reversal (i.e., it contains no $X^R$), it is said to be classical. A pattern is doubled if every variable occurs at least twice.

Our aim is to strengthen the following result.

**Theorem 1.** [1, 6, 7] *Every doubled pattern is 3-avoidable.*

First, we extend it to patterns with reversal.

**Theorem 2.** *Every doubled pattern with reversal is 3-avoidable.*

Then we obtain a lower bound on the number of ternary words avoiding a doubled pattern. The factor complexity of a factorial language $L$ over $\Sigma$ is $f(n) = |L \cap \Sigma^n|$. The growth rate of $L$ over $\Sigma$ is $\lim_{n \to \infty} f(n)^{1/n}$. We denote by $GR_3(p)$ the growth rate of ternary words avoiding the doubled pattern $p$.

**Theorem 3.** *For every doubled pattern $p$, $GR_3(p) \geq GR_3(AA)$.*

Let $v(p)$ be the number of distinct variables of the pattern $p$. In the proof of Theorem 1, the set of doubled patterns is partitioned as follows:

1. Patterns with $v(p) \leq 3$: the avoidability index of every ternary pattern has been determined [6].

2. Patterns shown to be 3-avoidable with the so-called power series method:
   - Patterns with $v(p) \geq 6$ [1]
   - Patterns with $v(p) = 5$ and prefix $ABC$ or length at least 11 [7]
   - Patterns with $v(p) = 4$ and prefix $ABCD$ or length at least 9 [7]

3. Ten sporadic patterns with $4 \leq v(p) \leq 5$ whose 3-avoidability cannot be deduced from the previous results: they have been shown to be 2-avoidable [7] using the method in [6].

The proof of Theorems 2 and 3 use the same partition. Sections 3 to 5 are each devoted to one type of doubled pattern with reversal. Theorem 3 is proved in Section 6.
2 Preliminaries

A word \( w \) is \( d \)-directed if for every factor \( f \) of \( w \) of length \( d \), the word \( f^R \) is not a factor of \( w \).

Remark 4. If a \( d \)-directed word contains an occurrence \( h \) of \( X.X^R \) for some variable \( X \), then \( |h(X)| \leq d - 1 \).

A variable that appears only once in a pattern is said to be isolated. The formula \( f \) associated to a pattern \( p \) is 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 \( h \) 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 \( f \). Recently, the avoidability of formulas with reversal has been considered by Currie, Mol, and Rampersad [3, 4] and me [8].

Recall that a formula is nice if every variable occurs at least twice in the same fragment. In particular, a doubled pattern is a nice formula with exactly one fragment.

The avoidability exponent \( AE(f) \) of a formula \( f \) is the largest real \( x \) such that every \( x \)-free word avoids \( f \). Every nice formula \( f \) with \( v(f) \geq 3 \) variables is such that \( AE(f) \geq 1 + \frac{1}{2v(f)-3} \) [11].

Let \( \simeq \) be the equivalence relation on words defined by \( w \simeq w' \) if \( w' \in \{w, w^R\} \). Avoiding a pattern up to \( \simeq \) has been investigated for every binary formulas [2]. Remark that for a given classical pattern or formula \( p \), avoiding \( p \) up to \( \simeq \) implies avoiding simultaneously all the variants of \( p \) with reversal.

Recall that a word is \((\beta^+, n)\)-free if it contains no repetition with exponent strictly greater than \( \beta \) and period at least \( n \).

3 Formulas with at most 3 variables

For classical doubled patterns with at most 3 variables, all the avoidability indices are known. There are many such patterns, so it would be tedious to consider all their variants with reversal.

However, we are only interested in their 3-avoidability, which follows from the 3-avoidability of nice formulas with at most 3 variables [10].

Thus, to obtain the 3-avoidability of doubled patterns with reversal with at most 3 variables, we show that every minimally nice formula with at most
3 variables is 3-avoidable up to \( \simeq \).

The minimally nice formulas with at most 3 variables, up to symmetries, are determined in [10] and listed in the following table. Every such formula \( f \) is avoided by the image by a \( q \)-uniform morphism of either any infinite \((\frac{5}{4}^+)\)-free word \( w_5 \) over \( \Sigma_5 \) or any infinite \((\frac{7}{5}^+)\)-free word \( w_4 \) over \( \Sigma_4 \), depending on whether the avoidability exponent of \( f \) is smaller than \( \frac{7}{5} \).

| Formula \( f \)       | \( = f^R \) | \( AE(f) \) | Word | \( q \) | \( d \) | Freeness       |
|-----------------------|-------------|------------|------|-------|-------|----------------|
| \( ABA.BAB \)         | yes         | 1.5        | \( g_a(w_4) \) | 9     | 9     | \( \frac{131}{90}^{-+}, 28 \) |
| \( ABCA.BCAB.CABC \)  | yes         | 1.333333333 | \( g_b(w_5) \) | 6     | 8     | \( \frac{4^+}{7}, 25 \) |
| \( ABCBA.CBABC \)     | yes         | 1.333333333 | \( g_c(w_5) \) | 4     | 9     | \( \frac{30^+}{73}, 18 \) |
| \( ABCA.BCAB.CBC \)   | no          | 1.381966011 | \( g_d(w_5) \) | 9     | 4     | \( \frac{62^+}{45}, 37 \) |
| \( ABA.BCB.CAC \)     | yes         | 1.5        | \( g_e(w_4)^1 \) | 9     | 4     | \( \frac{67^+}{45}, 37 \) |
| \( ABCA.BCAB.CBAC \)  | yes\(^2\)   | 1.333333333 | \( g_f(w_5) \) | 6     | 6     | \( \frac{41^+}{24}, 31 \) |
| \( ABCA.BAB.CAC \)    | yes         | 1.414213562 | \( g_g(w_4) \) | 6     | 8     | \( \frac{89^+}{63}, 61 \) |
| \( ABCA.BAB.CBC \)    | no          | 1.430159709 | \( g_h(w_4) \) | 6     | 7     | \( \frac{17^+}{72}, 61 \) |
| \( ABCA.BAB.CBC \)    | no          | 1.381966011 | \( g_i(w_5) \) | 8     | 7     | \( \frac{127^+}{96}, 41 \) |
| \( ABCBA.CABC \)      | no          | 1.361103081 | \( g_j(w_5) \) | 6     | 8     | \( \frac{4^+}{3}, 25 \) |
| \( ABCBA.CAC \)       | yes         | 1.396608253 | \( g_k(w_5) \) | 6     | 13    | \( \frac{4^+}{3}, 25 \) |

In the table above, the columns indicate respectively, the considered minimally nice formula \( f \), whether is equivalent to its reversed formula, the avoidability exponent of \( f \), the infinite ternary word avoiding \( f \), the value \( q \) such that the corresponding morphism is \( q \)-uniform, the value such that the avoiding word is \( d \)-directed, and the suitable property of \((\beta^n, n)\)-freeness used in the proof that \( f \) is avoided. We list below the corresponding morphisms.

\(^1\)The formula \( ABA.BCB.CAC \) seems also avoided up to \( \simeq \) by the Hall-Thue word, i.e., the fixed point of \( 0 \to 012; 1 \to 02; 2 \to 1 \).

\(^2\)We mistakenly said in [10] that \( ABCA.BCAB.CBAC \) is different from its reverse.
As an example, we show that $ABCBA.CAC$ is avoided by $g_k(w_5)$. First, we check that $g_k(w_5)$ is $(\frac{4}{3}, 25)$-free using the main lemma in [6], that is, we check the $(\frac{4}{3}, 25)$-freeness of the $g_k$-image of every $(\frac{5}{7})$-free word of length at most $2 \times \frac{4}{3} - 5 = 32$. Then we check that $g_k(w_5)$ is 13-directed by inspecting the factors of $g_k(w_5)$ of length 13. For contradiction, suppose that $g_k(w_5)$ contains an occurrence $h$ of $ABCBA.CAC$ up to $\simeq$. Let us write $a = |h(A)|$, $b = |h(B)|$, $c = |h(C)|$.

Suppose that $a \geq 25$. Since $g_k(w_5)$ is 13-directed, all occurrences of $h(A)$ are identical. Then $h(ABCBA)$ is a repetition with period $|h(ABC)| \geq 25$. So the $(\frac{4}{3}, 25)$-freeness implies the bound $\frac{2a + 2b + c}{a + 2b + c} \leq \frac{4}{3}$, that is, $a \leq b + \frac{1}{2}c$.

In every case, we have

$$a \leq \max \{b + \frac{1}{2}c, 24\}.$$ 

Similarly, the factors $h(BCB)$ and $h(CAC)$ imply

$$b \leq \max \{\frac{1}{2}c, 24\}$$

and

$$c \leq \max \{\frac{1}{2}a, 24\}.$$ 

Solving these inequalities gives $a \leq 36$, $b \leq 24$, and $c \leq 24$. Now we can check exhaustively that $g_k(w_5)$ contains no occurrence up to $\simeq$ satisfying these bounds.
Except for $ABCBA.CBABC$, the avoidability index of the nice formulas in the above table is 3. So the results in this section extend their 3-avoidability up to $\simeq$.

4 The power series method

The so-called power series method has been used [1, 7] to prove the 3-avoidability of many classical doubled patterns with at least 4 variables and every doubled pattern with at least 6 variables, as mentioned in the introduction.

Let $p$ be such a classical doubled pattern and let $p'$ be a doubled pattern with reversal obtained by adding some $-R$ to $p$. Without loss of generality, the leftmost appearance of every variable $X$ of $p$ remains free of $-R$ in $p'$. Then we will see that $p'$ is also 3-avoidable. The power series method is a counting argument that relies on the following observation. If the $h$-image of the leftmost appearance of the variable $X$ of $p$ is fixed, say $h(X) = w_X$, then there is exactly one possibility for the $h$-image of the other appearances of $X$, namely $h(X) = w_X$. This observation can be extended to $p'$, since there is also exactly one possibility for $h(X^R)$, namely $h(X^R) = w_X^R$.

Notice that this straightforward generalization of the power series method from classical doubled patterns to doubled patterns with reversal cannot be extended to avoiding a doubled pattern up to $\simeq$. Indeed, if $h(X) = w_X$ for the leftmost appearance of the variable $X$ and $w_X$ is not a palindrome, then there exist two possibilities for the other appearances of $X$, namely $w_X$ and $w_X^R$.

5 Sporadic patterns

Up to symmetries, there are ten doubled patterns whose 3-avoidability cannot be deduced by the previous results. They have been identified in [7] and are listed in the following table.
Let \( w_5 \) be any infinite \( \left( \frac{5^+}{4} \right) \)-free word over \( \Sigma_5 \) and let \( h \) be the following 9-uniform morphism.

\[
\begin{align*}
    h(0) &= 020022221 \\
    h(1) &= 011111221 \\
    h(2) &= 010202110 \\
    h(3) &= 010022112 \\
    h(4) &= 000022121
\end{align*}
\]

First, we check that \( h(w_5) \) is 7-directed and \( \left( \frac{139}{108}, 46 \right) \)-free. Then, using the same method as in Section 3, we show that \( h(w_5) \) avoids up to \( \simeq \) these ten sporadic patterns simultaneously.

### 6 Growth rate of ternary words avoiding a doubled pattern

Theorem 1 obviously holds for \( p = AA \). Without loss of generality, we do not need to consider a doubled pattern \( p \) that contains an occurrence of another doubled pattern. In particular, \( p \) is square-free. So we need to show that \( GR_3(p) \) is at least \( GR_3(AA) \), which is close to 1.30176 \[12\].

If \( p \) is 2-avoidable, then \( p \) is avoided by sufficiently many ternary words. By Lemma 4.1 in \[6\], \( \lambda(p) = 2 \) implies that \( GR_3(p) \geq 2^\frac{1}{7} > GR_3(AA) \).

Moreover, for every doubled pattern \( p \) whose 3-avoidability has been obtained via the power series method, we even get \( GR_3(p) > 2 > GR_3(AA) \).
According to the partition of the set of doubled patterns mentioned in the introduction, the two remarks above handle the case $v(p) \geq 4$. To handle the case $v(p) \leq 3$, we explore (by manual backtracking) the space of square-free doubled patterns, using the 2-avoidability of $ABACBC$ [6], $ABCACB$ [6], and $ABCBABC$ [5].

\begin{itemize}
  \item $A$ is unavoidable
  \item $AB$ is unavoidable
  \item $ABA$ is unavoidable
  \item $ABAC$ is unavoidable
  \item $ABACA$ is unavoidable
  \item $ABACAB$ is unavoidable
  \item $ABACABA$ is unavoidable
  \item $ABACABC$ is 2-avoidable ($BACABC$ is the reverse of $ABCACB$)
  \item $ABACB$ is unavoidable
  \item $ABACBA$ is unavoidable
  \item $ABACBAB$ is not doubled (it is the formula $ABA.BAB$)
  \item $ABACBABC$ is 2-avoidable ($ACBABC$ is $ABCACB$)
  \item $ABACBC$ is 2-avoidable
  \item $ABC$ is unavoidable
  \item $ABCAB$ is unavoidable
  \item $ABCABA$ is unavoidable
  \item $ABCA$ is unavoidable
  \item $ABCAB$ is unavoidable
  \item $ABCABA$ is unavoidable
  \item $ABCABAC$ is 2-avoidable ($BCABAC$ is $ABCACB$)
  \item $ABCAC$ is unavoidable
  \item $ABCACB$ is 2-avoidable
  \item $ABCB$ is unavoidable
  \item $ABCBAB$ is unavoidable
  \item $ABCBABC$ is 2-avoidable
  \item $ABCBAC$ is 2-avoidable (it is the reverse of $ABCACB$)
\end{itemize}

7 Conclusion

Unlike classical formulas, we know that there exist avoidable formulas with reversal of arbitrarily high avoidability index [8]. Maybe doubled patterns and nice formulas are easier to avoid. We propose the following open problems.
• Are there infinitely many doubled patterns up to \(\simeq\) that are not 2-avoidable?

• Is there a nice formula up to \(\simeq\) that is not 3-avoidable?

A first step would be to improve Theorem 2 by generalizing the 3-avoidability of doubled patterns with reversal to doubled patterns up to \(\simeq\). Notice that the results in Sections 3 and 5 already consider avoidability up to \(\simeq\). However, the power series method gives weaker results. Classical doubled patterns with at least 6 variables are 3-avoidable because

\[
1 - 3x + \left( \frac{3x^2}{1 - 3x^2} \right)^v
\]

has a positive real root for \(v \geq 6\). The (basic) power series for doubled patterns up to \(\simeq\) with \(v\) variables would be

\[
1 - 3x + \left( \frac{6x^2}{1 - 3x^2} - \frac{3x^2 + 3x^4}{1 - 3x^4} \right)^v.
\]

The term \(\frac{6x^2}{1 - 3x^2}\) counts for twice the term \(\frac{3x^2}{1 - 3x^2}\) in the classical setting, for \(h(V)\) and \(h(V)^R\). The term \(\frac{3x^2 + 3x^4}{1 - 3x^4}\) corrects for the case of palindromic \(h(V)\), which should not be counted twice. This power series has a positive real root only for \(v \geq 10\). This leaves many doubled patterns up to \(\simeq\) whose 3-avoidability must be proved with morphisms.

Looking at the proof of Theorem 2, we may wonder if a doubled pattern with reversal is always easier to avoid than the corresponding classical pattern. This is not the case: backtracking shows that \(\lambda(ABCA^R C^R B) = 3\), whereas \(\lambda(ABCACB) = 2\) [6].

The proof of Theorem 3 suggests the following open problem:

• Is every square-free doubled pattern 2-avoidable?

It would imply Theorem 3 and it resembles the conjecture that there exist only finitely many 2-unavoidable doubled patterns [7, 9].

References

[1] J. Bell, T. L. Goh. Exponential lower bounds for the number of words of uniform length avoiding a pattern. Inform. and Comput. 205 (2007), 1295-1306.
[2] J. Currie and L. Mol. The undirected repetition threshold and undirected pattern avoidance. *Theor. Comput. Sci.* **866** (2021), 56–69.

[3] J. Currie, L. Mol, and N. Rampersad. A family of formulas with reversal of high avoidability index. *International Journal of Algebra and Computation* **27**(5) (2017), 477–493.

[4] J. Currie, L. Mol, and N. Rampersad. Avoidance bases for formulas with reversal. *Theor. Comput. Sci.* **738** (2018), 25–41.

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

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

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

[8] P. Ochem. A family of formulas with reversal of arbitrarily high avoidability index. *Theoret. Comput. Sci.* **896** (2021), 168–170.

[9] P. Ochem and A. Pinlou. Application of entropy compression in pattern avoidance. *Electron. J. Combinatorics.* **21**(2) (2014), #RP2.7.

[10] P. Ochem and M. Rosenfeld. On some interesting ternary formulas. *Electron. J. Combinatorics.* **26**(1) (2019), #P1.12.

[11] P. Ochem and M. Rosenfeld. Avoidability of palindrome patterns. *Electron. J. Combinatorics.* **28**(1) (2021), #P1.4.

[12] A. Shur. Growth rates of complexity of power-free languages. *Theoret. Comput. Sci.* **411**(34-36) (2010), 3209–3223.