Doubled patterns are 3-avoidable
Pascal Ochem

To cite this version:
Pascal Ochem. Doubled patterns are 3-avoidable. The Electronic Journal of Combinatorics, 2016, 23 (1), pp.P1.19. 10.37236/5618. lirmm-01375763
Doubled patterns are 3-avoidable

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

Submitted: June, 2015; Accepted: XX; Published: XX
Mathematics Subject Classifications: 68R15

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 said to be doubled if no variable occurs only once. Doubled patterns with at most 3 variables and patterns with at least 6 variables are 3-avoidable. We show that doubled patterns with 4 and 5 variables are also 3-avoidable.

Keywords: Word; Pattern avoidance.

1 Introduction

A pattern $p$ is a non-empty 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 $\Sigma$ such that there exists an infinite word $w$ over $\Sigma$ containing no occurrence of $p$. Bean, Ehrenfeucht, and McNulty [2] and Zimin [13] characterized unavoidable patterns, i.e., such that $\lambda(p) = \infty$. We say that a pattern $p$ is $t$-avoidable if $\lambda(p) \leq t$. For more informations on pattern avoidability, we refer to Chapter 3 of Lothaire’s book [8].

It follows from their characterization that every unavoidable pattern contains a variable that occurs once. Equivalently, every doubled pattern is avoidable. Our result is that:

Theorem 1. Every doubled pattern is 3-avoidable.

Let $v(p)$ be the number of distinct variables of the pattern $p$. For $v(p) \leq 3$, Cassaigne [5] began and I [9] finished the determination of the avoidability index of every
pattern with at most 3 variables. It implies in particular that every avoidable pattern with at most 3 variables is 3-avoidable. Moreover, Bell and Goh [3] obtained that every doubled pattern $p$ such that $v(p) \geq 6$ is 3-avoidable.

Therefore, as noticed in the conclusion of [10], there remains to prove Theorem 1 for every pattern $p$ such that $4 \leq v(p) \leq 5$. In this paper, we use both constructions of infinite words and a non-constructive method to settle the cases $4 \leq v(p) \leq 5$.

Recently, Blanchet-Sadri and Woodhouse [4] and Ochem and Pinlou [10] independently obtained the following.

**Theorem 2 ([4, 10])**. Let $p$ be a pattern.

(a) If $p$ has length at least $3 \times 2^{v(p) - 1}$ then $\lambda(p) \leq 2$.

(b) If $p$ has length at least $2^{v(p)}$ then $\lambda(p) \leq 3$.

As noticed in these papers, if $p$ has length at least $2^{v(p)}$ then $p$ contains a doubled pattern as a factor. Thus, Theorem 1 implies Theorem 2.(b).

## 2 Extending the power series method

In this section, we borrow an idea from the entropy compression method to extend the power series method as used by Bell and Goh [3], Rampersad [12], and Blanchet-Sadri and Woodhouse [4].

Let us describe the method. Let $L \subset \Sigma_m^*$ be a factorial language defined by a set $F$ of forbidden factors of length at least 2. We denote the factor complexity of $L$ by $n_i = L \cap \Sigma_i^m$. We define $L'$ as the set of words $w$ such that $w$ is not in $L$ and the prefix of length $|w| - 1$ of $w$ is in $L$. For every forbidden factor $f \in F$, we choose a number $1 \leq s_f \leq |f|$. Then, for every $i \geq 1$, we define an integer $a_i$ such that

$$a_i \geq \max_{u \in L} \left| \{ v \in \Sigma_m^i \mid uv \in L', uv = bf, f \in F, s_f = i \} \right|.$$

We consider the formal power series $P(x) = 1 - mx + \sum_{i \geq 1} a_i x^i$. If $P(x)$ has a positive real root $x_0$, then $n_i \geq x_0^{-i}$ for every $i \geq 0$.

Let us rewrite that $P(x_0) = 1 - mx_0 + \sum_{i \geq 1} a_i x_0^i = 0$ as

$$m - \sum_{i \geq 1} a_i x_0^{i-1} = x_0^{-1} \quad (1)$$

Since $n_0 = 1$, we will prove by induction that $\frac{m}{n_{i-1}} \geq x_0^{-1}$ in order to obtain that $n_i \geq x_0^{-i}$ for every $i \geq 0$. By using (1), we obtain the base case: $\frac{n_1}{n_0} = n_1 = m \geq x_0^{-1}$. Now, for every length $i \geq 1$, there are:

- $m^i$ words in $\Sigma_m^i$,
- $n_i$ words in $L$,
• at most $\sum_{1 \leq j \leq i} n_i - j a_j$ words in $L'$,
• $m(m^{i-1} - n_{i-1})$ words in $\Sigma_m \setminus \{L \cup L'\}$.

This gives $n_i + \sum_{1 \leq j \leq i} n_j a_{i-j} + m(m^{i-1} - n_{i-1}) \geq m^i$, that is, $n_i \geq mn_{i-1} - \sum_{1 \leq j \leq i} n_{i-j} a_j$.

$$\frac{n_i}{n_{i-1}} \geq m - \sum_{1 \leq j \leq i} a_j \frac{n_{i-j}}{n_{i-1}} \geq m - \sum_{1 \leq j \leq i} a_j x_0^{i-j}$$

By induction

$$\geq m - \sum_{j \geq 1} a_j x_0^{i-j} = x_0^{-1}$$

By (1)

The power series method used in previous papers [3, 4, 12] corresponds to the special case such that $s_f = |f|$ for every forbidden factor. Our condition is that $P(x) = 0$ for some $x > 0$ whereas the condition in these papers is that every coefficient of the series expansion of $\frac{1}{P(x)}$ is positive. The two conditions are actually equivalent. The result in [11] concerns series of the form $S(x) = 1 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots$ with real coefficients such that $a_1 < 0$ and $a_i \geq 0$ for every $i \geq 2$. It states that every coefficient of the series $1/S(x) = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \ldots$ is positive if and only if $S(x)$ has a positive real root $x_0$. Moreover, we have $b_i \geq x_0^{-i}$ for every $i \geq 0$.

The entropy compression method as developed by Gonçalves, Montassier, and Pinlou [6] uses a condition equivalent to $P(x) = 0$. The benefit of the present method is that we get an exponential lower bound on the factor complexity. It is not clear whether it is possible to get such a lower bound when using entropy compression for graph coloring, since words have a simpler structure than graphs.

### 3 Applying the method

In this section, we show that some doubled patterns on 4 and 5 variables are 3-avoidable. Given a pattern $p$, every occurrence $f$ of $p$ is a forbidden factor. With an abuse of notation, we denote by $|A|$ the length of the image of the variable $A$ of $p$ in the occurrence $f$. This notation is used to define the length $s_f$.

Let us first consider doubled patterns with 4 variables. We begin with patterns of length 9, so that one variable, say $A$, appears 3 times. We set $s_f = |f|$. Using the obvious upper bound on the number of pattern occurrences, we obtain

$$P(x) = 1 - 3x + \sum_{a,b,c,d \geq 1} 3^{a+b+c+d} x^{3a+2b+2c+2d}$$

$$= 1 - 3x + \sum_{a,b,c,d \geq 1} (3x^3)^a (3x^2)^b (3x^2)^c (3x^2)^d$$

$$= 1 - 3x + \left( \sum_{a \geq 1} (3x^3)^a \right) \left( \sum_{b \geq 1} (3x^2)^b \right) \left( \sum_{c \geq 1} (3x^2)^c \right) \left( \sum_{d \geq 1} (3x^2)^d \right)$$

$$= 1 - 3x + \left( \frac{1}{1-3x} - 1 \right) \left( \frac{1}{1-3x^2} - 1 \right) \left( \frac{1}{1-3x^2} - 1 \right) \left( \frac{1}{1-3x^2} - 1 \right)$$

$$= 1 - 3x + \left( \frac{1}{1-3x} - 1 \right) \left( \frac{1}{1-3x^2} - 1 \right)^3$$

$$= \frac{1 - 3x - 9x^2 + 24x^3 + 54x^4 - 54x^5 - 108x^6 + 243x^8 + 162x^9 - 243x^{10}}{(1 - 3x^2)(1 - 3x^4)^3}.$$

Then $P(x)$ admits $x_0 = 0.3400 \ldots$ as its smallest positive real root. So, every doubled pattern $p$ with 4 variables and length 9 is 3-avoidable and there exist at least $x_0^n > 2.941^n$. 

The Electronic Journal of Combinatorics **XX** (2013), #R00
ternary words avoiding p. Notice that for patterns with 4 variables and length at least 10, every term of \( \sum_{a,b,c,d \geq 1} 3^{a+b+c+d} x^{3a+2b+2c+2d} \) in \( P(x) \) gets multiplied by some positive power of \( x \). Since \( 0 < x < 1 \), every term is now smaller than in the previous case. So \( P(x) \) admits a smallest positive real root that is smaller than 0.3400... Thus, these patterns are also 3-avoidable.

Now, we consider patterns with length 8, so that every variable appears exactly twice. If such a pattern has \( ABCD \) as a prefix, then we set \( s_f = \frac{|f|}{2} = |A| + |B| + |C| + |D| \). So we obtain \( P(x) = 1 - 3x + \sum_{a,b,c,d \geq 1} x^{a+b+c+d+e} x^{3a+2b+2c+2d+2e} = 1 - 3x + \left( \frac{1}{1-x} - 1 \right)^4 \). Then \( P(x) \) admits 0.3819... as its smallest positive real root, so that this pattern is 3-avoidable.

Among the remaining patterns, we rule out patterns containing an occurrence of a doubled pattern with at most 3 variables. Also, if one pattern is the reverse of another, then they have the same avoidability index and we consider only one of the two. Thus, there remain the following patterns: \( ABACBDCD \), \( ABACDBDC \), \( ABACDCBD \), \( ABCA DBDC \), \( ABCADBDC \), and \( ABCBDADC \).

Now we consider doubled patterns with 5 variables. Similarly, we rule out every pattern of length at least 11 by setting \( s_f = |f| \). Then we check that \( P(x) = 1 - 3x + \sum_{a,b,c,d,e \geq 1} 3^{a+b+c+d+e} x^{3a+2b+2c+2d+2e} = 1 - 3x + \left( \frac{1}{1-x} - 1 \right)^3 \left( \frac{1}{1-x} - 1 \right)^2 \) has a positive real root.

We also rule out every pattern of length 10 having \( ABC \) as a prefix. We set \( s_f = |f| - |ABC| = |A| + |B| + |C| + 2|D| + 2|E| \). Then we check that \( P(x) = 1 - 3x + \sum_{a,b,c,d,e \geq 1} 3^{d+e} x^{a+b+c+2d+2e} = 1 - 3x + \left( \frac{1}{1-x} - 1 \right)^3 \left( \frac{1}{1-x} - 1 \right)^2 \) has a positive real root.

Again, we rule out patterns containing an occurrence of a doubled pattern with at most 4 variables and patterns whose reversed pattern is already considered. Thus, there remain the following patterns: \( ABACBDCEDE \), \( ABACBCDEDE \), and \( ABACDBDECE \).

4 Sporadic doubled patterns

In this section, we consider the 10 doubled patterns on 4 and 5 variables whose 3-avoidability has not been obtained in the previous section.

We define the avoidability exponent \( AE(p) \) of a pattern \( p \) as the largest real \( x \) such that every \( x \)-free word avoids \( p \). This notion is not pertinent e.g. for the pattern \( ABWBAXACYCAZBC \) studied by Baker, McNulty, and Taylor [1], since for every \( \epsilon > 0 \), there exists a \((1 + \epsilon)\)-free word containing an occurrence of that pattern. However, \( AE(p) > 1 \) for every doubled pattern. To see that, consider a factor \( A...A \) of \( p \). If an \( x \)-free word contains an occurrence of \( p \), then the image of this factor is a repetition such that the image of \( A \) cannot be too large compared to the images of the variables occurring between the As in \( p \). We have similar constraints for every variable and this set of constraints becomes unsatisfiable as \( x \) decreases towards 1. We present one way of obtaining the avoidability exponent for a doubled pattern \( p \) of length \( 2v(p) \). We construct the \( v(p) \times v(p) \) matrix \( M \) such that \( M_{i,j} \) is the number of occurrences of the variable \( X_j \) between the two occurrences of the variable \( X_i \). We compute the largest eigenvalue \( \beta \) of \( M \) and then we
have $AE(p) = 1 + \frac{1}{\beta+1}$. For example if $p = ABACDCBD$, then we get $M = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$, $\beta = 1.9403\ldots$, and $AE(p) = 1 + \frac{1}{\beta+1} = 1.340\ldots$. The avoidability exponents of the 10 patterns considered in this section range from $AE(ABACDBDC) = 1.292893219$ to $AE(ABACBDCD) = 1.381966011$. For each pattern $p$ among the 10, we give a uniform morphism $m : \Sigma_5^* \to \Sigma_2^*$ such that for every $\left(\frac{5}{4}\right)$-free word $w \in \Sigma_5^*$, we have that $m(w)$ avoids $p$. The proof that $p$ is avoided follows the method in [9]. Since there exist exponentially many $\left(\frac{5}{4}\right)$-free words over $\Sigma_5$ [7], there exist exponentially many binary words avoiding $p$.

- $AE(ABACBDCD) = 1.381966011$, 17-uniform morphism

\[
\begin{align*}
0 & \mapsto 000001111101010110 \\
1 & \mapsto 0000011010010011011 \\
2 & \mapsto 000000111001101111 \\
3 & \mapsto 000000110101011111 \\
4 & \mapsto 000000110010011111
\end{align*}
\]

- $AE(ABACDBDC) = 1.333333333$, 33-uniform morphism

\[
\begin{align*}
0 & \mapsto 00000010110100011111101100101111 \\
1 & \mapsto 00000010011010000111111101001111 \\
2 & \mapsto 00000001011010001111111101001111 \\
3 & \mapsto 0000000100110100011111111101001111 \\
4 & \mapsto 0000000100110011000011111101001111
\end{align*}
\]

- $AE(ABACDCBD) = 1.340090632$, 28-uniform morphism

\[
\begin{align*}
0 & \mapsto 0000101010001111101000011111 \\
1 & \mapsto 0000001111101000110100111111 \\
2 & \mapsto 0000001101000111110100111111 \\
3 & \mapsto 0000001011110000110100111111 \\
4 & \mapsto 00000010101111000011111101001111
\end{align*}
\]

- $AE(ABCADBDC) = 1.292893219$, 21-uniform morphism

\[
\begin{align*}
0 & \mapsto 000011111110101111111111 \\
1 & \mapsto 00001011010010011111111111 \\
2 & \mapsto 0000011111101000110100111111 \\
3 & \mapsto 0000011011110000110100111111 \\
4 & \mapsto 00000111011011011111111111
\end{align*}
\]
\[
AE(ABCADCB) = 1.295597743, \text{ 22-uniform morphism}
\]
\[
0 \mapsto 000001101101010010011111 \\
1 \mapsto 000001101010100111111111 \\
2 \mapsto 000001101101001111111111 \\
3 \mapsto 000001101101001011111111 \\
4 \mapsto 000001101010100111111111
\]
\[
AE(ABCADCDB) = 1.327621756, \text{ 26-uniform morphism}
\]
\[
0 \mapsto 00000011110010101011000111 \\
1 \mapsto 000000110111111010011111 \\
2 \mapsto 000000110111111010011111 \\
3 \mapsto 000000110111111010011111 \\
4 \mapsto 000000110111111010011111
\]
\[
AE(ABCBDADEC) = 1.302775638, \text{ 33-uniform morphism}
\]
\[
0 \mapsto 000000101111110011000110011111101 \\
1 \mapsto 000000101111001000001100111111101 \\
2 \mapsto 000000101111001000001100111111101 \\
3 \mapsto 000000101111001000001100111111101 \\
4 \mapsto 000000101111100101011111101
\]
\[
AE(ABACBDCEDE) = 1.366025404, \text{ 15-uniform morphism}
\]
\[
0 \mapsto 001011011110000 \\
1 \mapsto 001010100111111 \\
2 \mapsto 000110010011000 \\
3 \mapsto 000011111111100 \\
4 \mapsto 000011010101110
\]
\[
AE(ABACBDCEDE) = 1.302775638, \text{ 18-uniform morphism}
\]
\[
0 \mapsto 000010110100100111 \\
1 \mapsto 000010100111111111 \\
2 \mapsto 000000110110011111 \\
3 \mapsto 000000101010101111 \\
4 \mapsto 000000000111100111
\]
\[
AE(ABACDBCEDE) = 1.320416579, \text{ 22-uniform morphism}
\]
\[
0 \mapsto 0000001111110011001110111 \\
1 \mapsto 0000001111110011001110111 \\
2 \mapsto 0000001111110011001110111 \\
3 \mapsto 0000001111110011001110111 \\
4 \mapsto 0000001111110011001110111
\]
5 Simultaneous avoidance of doubled patterns

Bell and Goh [3] have also considered the avoidance of multiple patterns simultaneously and ask (question 3) whether there exist an infinite word over a finite alphabet that avoids every doubled pattern. We give a negative answer.

A word \( w \) is \( n \)-splitted if \( |w| \equiv 0 \pmod{n} \) and every factor \( w_i \) such that \( w = w_1w_2\ldots w_n \) and \( |w_i| = \frac{|w|}{n} \) for \( 1 \leq i \leq n \) contains every letter in \( w \). An \( n \)-splitted pattern is defined similarly. Let us prove by induction on \( k \) that every word \( w \in \Sigma^n_k \) contains an \( n \)-splitted factor. The assertion is true for \( k = 1 \). Now, if the word \( w \in \Sigma^n_k \) is not itself \( n \)-splitted, then by definition it must contain a factor \( w_i \) that does not contain every letter of \( w \). So we have \( w_i \in \Sigma^{n-1}_{k-1} \). By induction, \( w_i \) contains an \( n \)-splitted factor, and so does \( w \).

This implies that for every fixed \( n \), every infinite word over a finite alphabet contains \( n \)-splitted factors. Moreover, an \( n \)-splitted word is an occurrence of an \( n \)-splitted pattern such that every variable has a distinct image of length 1. So, for every fixed \( n \), the set of all \( n \)-splitted patterns is not avoidable by an infinite word over a finite alphabet.

Notice that if \( n \geq 2 \), then an \( n \)-splitted word (resp. pattern) contains a 2-splitted word (resp. pattern) and a 2-splitted word (resp. pattern) is doubled.

6 Conclusion

Our results answer settles the first of two questions of our previous paper [10]. The second question is whether there exists a finite \( k \) such that every doubled pattern with at least \( k \) variables is 2-avoidable. As already noticed [10], such a \( k \) is at least 5 since, e.g., \( ABCCBADD \) is not 2-avoidable.

Acknowledgments

I am grateful to Narad Rampersad for comments on a draft of the paper, to Vladimir Dotsenko for pointing out the result in [11], and to Andrei Romashchenko for translating this paper.

References

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

[2] D.R. Bean, A. Ehrenfeucht, and G.F. McNulty. Avoidable patterns in strings of symbols. *Pacific J. of Math.* **85** (1979) 261–294.

[3] 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.
[4] F. Blanchet-Sadri, B. Woodhouse. Strict bounds for pattern avoidance. *Theor. Comput. Sci.* **506** (2013), 17–27.

[5] J. Cassaigne. Motifs évitables et régularité dans les mots. Thèse de Doctorat, Université Paris VI, Juillet 1994.

[6] D. Gonçalves, M. Montassier, and A. Pinlou. Entropy compression method applied to graph colorings. arXiv:1406.4380

[7] R. Kolpakov and M. Rao: On the number of Dejan words over alphabets of 5, 6, 7, 8, 9 and 10 letters *Theor. Comput. Sci.* **412(46)** (2011), 6507–6516.

[8] M. Lothaire. Algebraic Combinatorics on Words. *Cambridge Univ. Press* (2002).

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

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

[11] D. I. Piotkovskii. On the growth of graded algebras with a small number of defining relations. *Uspekhi Mat. Nauk.* **48:3(291)** (1993), 199–200.

[12] N. Rampersad. Further applications of a power series method for pattern avoidance. *Electron. J. Combinatorics.* **18(1)** (2011), #P134.

[13] A.I. Zimin. Blocking sets of terms. *Math. USSR Sbornik* **47(2)** (1984) 353–364. English translation.