Avoiding conjugacy classes on the 5-letter alphabet
Golnaz Badkobeh, Pascal Ochem

To cite this version:
Golnaz Badkobeh, Pascal Ochem. Avoiding conjugacy classes on the 5-letter alphabet. RAIRO - Theoretical Informatics and Applications (RAIRO: ITA), EDP Sciences, 2020, 54 (2), pp.1-4. 10.1051/ita/2020003. hal-02903902

HAL Id: hal-02903902
https://hal.archives-ouvertes.fr/hal-02903902
Submitted on 21 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
AVOIDING CONJUGACY CLASSES ON THE 5-LETTER ALPHABET

Golnaz Badkobeh* and Pascal Ochem†,‡

Abstract. We construct an infinite word \(w\) over the 5-letter alphabet such that for every factor \(f\) of \(w\) of length at least two, there exists a cyclic permutation of \(f\) that is not a factor of \(w\). In other words, \(w\) does not contain a non-trivial conjugacy class. This proves the conjecture in Gamard et al. [Theoret. Comput. Sci. 726 (2018) 1–4].

Mathematics Subject Classification. 68R15.

Received November 14, 2018. Accepted February 18, 2020.

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 \(h: \Delta^* \rightarrow \Sigma^*\) 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 over \(\Sigma\) containing no occurrence of \(p\). Bean et al. [2] and Zimin [8] characterized unavoidable patterns, i.e., such that \(\lambda(p) = \infty\). However, determining the exact avoidability index of an avoidable pattern requires more work. Although patterns with index 4 [2] and 5 [4] have been found, the existence of an avoidable pattern with index at least 6 is an open problem since 2001.

Some techniques in pattern avoidance start by showing that the considered word avoids other structures, such as generalized repetitions [6, 7]. Let us say that a word has property \(P_i\) if it does not contain all the conjugates of the same word \(w\) with \(|w| \geq i\). Recently, in order to study the avoidance of a kind of patterns called circular formulas, Gamard et al. [5] obtained that there exists

- a morphic binary word satisfying \(P_5\),
- a morphic ternary word satisfying \(P_3\),
- a morphic word over the 6-letter alphabet satisfying \(P_2\).

Recall that a pure morphic word is of the form \(m^w(0)\) and a morphic word is of the form \(h(m^w(0))\) for some morphisms \(m\) and \(h\). Independently, Bell and Madill [3] obtained a pure morphic word over the 12-letter alphabet that also satisfies \(P_2\) and some other properties.

Keywords and phrases: Combinatorics on words, conjugacy classes.

1 Golsmiths, University of London.
2 LIRMM, Université de Montpellier, CNRS, Montpellier, France.
* Golnaz Badkobeh is supported by the Leverhulme Trust on the Leverhulme Early Career Scheme.
† This work is supported by the ANR project CoCoGro (ANR-16-CE49-0005).
‡ Corresponding author: ochem@lirmm.fr

© The authors. Published by EDP Sciences, 2020

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
It is conjectured that the smallest alphabet allowing an infinite word satisfying $P_2$ has 5 letters [5], which is best possible. In this paper, we prove this conjecture using a morphic word. This settles the topic of the smallest alphabet needed to satisfy $P_1$.

2. Main result

Let $\varepsilon$ denote the empty word. We consider the morphic word $w_5 = G(F^\omega(0))$ defined by the following morphisms.

\[
\begin{align*}
F(0) &= 01, & G(0) &= abcd, \\
F(1) &= 2, & G(1) &= \varepsilon, \\
F(2) &= 03, & G(2) &= eacd, \\
F(3) &= 24, & G(3) &= becd, \\
F(4) &= 23. & G(4) &= be.
\end{align*}
\]

**Theorem 2.1.** The morphic word $w_5 \in \Sigma_5^\omega$ avoids every conjugacy class of length at least 2.

In order to prove this theorem, it is convenient to express $w_5$ with the larger morphisms $f = F^3$ and $g = G \circ F^2$ given below. Clearly, $w_5 = g(f^\omega(0))$.

\[
\begin{align*}
f(0) &= 01203, & g(0) &= abcdeacd, \\
f(1) &= 0124, & g(1) &= abcdbecebd, \\
f(2) &= 0120323, & g(2) &= abcdbeacdbe, \\
f(3) &= 01240324, & g(3) &= abcdbeacdbed, \\
f(4) &= 01240323. & g(4) &= abcdbeacdb.
\end{align*}
\]

2.1. Avoiding conjugacy classes in $F^\omega(0)$

Here we study the pure morphic word and the conjugacy classes it contains.

**Lemma 2.2.** The infinite word $F^\omega(0)$ contains only the conjugacy classes listed in $C = \{F(2), F^2(2), F^3(4), f^d(0)\}$, for all $d \geq 1$.

**Proof.** Notice that the factor 01 only occurs as the prefix of the $f$-image of every letter in $F^\omega(0)$. Moreover, every letter 1 only occurs in $F^\omega(0)$ as the suffix of the factor 01. Let us say that the index of a conjugacy class is the number of occurrences of 1 in any of its elements. An easy computation shows that the set of complete conjugacy classes in $F^\omega(0)$ with index at most one is $C_1 = \{F(2), F^2(2), F(4), F^3(4), f(4), f(0)\}$. Let us assume that $F^\omega(0)$ contains a conjugacy class $c$ with index at least two. Let $w \in c$ be such that 01 is a prefix of $w$. We write $w = ps$ such that the leftmost occurrence of 01 in $w$ is the prefix of $s$. Then the conjugate $sp$ of $w$ also belongs to $c$ and thus is a factor of $F^\omega(0)$. This implies that the pre-image $v = f^{-1}(w)$ is a factor of $F^\omega(0)$, and so does every conjugate of $v$. Thus, $F^\omega(0)$ contains a conjugacy class $c'$ such that the elements of $c$ with prefix 01 are the $f$-images of the elements of $c'$. Moreover, the index of $c'$ is strictly smaller than the index of $c$.

Using this argument recursively, we conclude that every complete conjugacy class in $F^\omega(0)$ has a member of the form $f^k(x)$ such that $x$ is an element of a conjugacy class in $C_1$.

Now we show that $F(2)$ does not generate larger conjugacy classes in $F^\omega(0)$. We thus have to exhibit a conjugate of $f(F(2)) = F^3(2) = 0120301240324$ that is not a factor of $F^\omega(0)$. A computer check shows that the conjugate 40120301240323 is not a factor of $F^\omega(0)$. Similarly, $F^2(2)$ does not generate larger conjugacy classes in $F^\omega(0)$ since the conjugate 3012030124012032301240323 of $f(F^2(2)) = F^5(2) = 012030124012032301240323$ is not a factor of $F^\omega(0)$.\□
2.2. Avoiding conjugacy classes in $w_5$

We are ready to prove Theorem 2.1. A computer check\(^1\) shows that $w_5$ avoids every conjugacy class of length at most 1000. Let us assume that $w_5$ contains a conjugacy class $c$ of length at least 41. Consider a word $w \in c$ with prefix $ab$. Notice that $ab$ only appears in $w_5$ as the prefix of the $g$-image of every letter. Since $|w| \geq 41$, $w$ contains at least 2 occurrences of $ab$ and we write $w = psq$ such that the rightmost occurrence of $ab$ in $w$ is the prefix of $s$. Then the conjugate $sq$ of $w$ also belongs to $c$ and thus is a factor of $w_5$. This implies that the pre-image $v = g^{-1}(w)$ is a factor of $F^w(0)$, and so does every conjugate of $v$. Thus, $F^w(0)$ contains a conjugacy class $c'$ such that the elements of $c$ with prefix $ab$ are the $f$-images of the elements of $c'$.

To finish the proof, it is thus sufficient to show that for every $c' \in C$, there exists a conjugate of $g(c')$ that is not a factor of $w_5$. Recall that $C = \{F(2), F^2(2), F^3(4), F^4(0)\}$ for all $d \geq 1$. The computer check mentioned above settles the case of $F(2)$ and $F^2(2)$ since $|g(F(2))| < |g(F^2(2))| = 40 < 1000$. It also settles the case of $f(4)$ and $f(0)$ since $|g(f(0))| < |g(f(4))| = 90 < 1000$.

The next four lemmas handle the remaining cases (with $d \geq 1$):

\[
\begin{align*}
&g(f^d(F(4))) = g(f^d(23)) \\
g(f^d(F^2(4))) = g(f^d(0324)) \\
g(f^d+1(4)) = g(f^d(01240324)) \\
g(f^d+1(0)) = g(f^d(01203))
\end{align*}
\]

Notice that for technical reasons, we do not consider $g(f(4))$ and $g(f(4))$, which are also covered by the computer check.

**Lemma 2.3.** Let $p_{23} = e.g(3f(3) \ldots f^d(3), f(4))$ and $s_{23} = g(f^d(01203).f^d(01203) \ldots f(01203).01203).abcedeacdb$. For every $d \geq 0$, the word $T_{23} = p_{23}s_{23}$ is a conjugate of $g(f^d(23))$ that is not a factor of $w_5$.

**Proof.** It is easy to check that $T_{23}$ is indeed a conjugate of $g(f^d(23))$. Let us assume that $T_{23}$ appears in $w_5$.

The letter 3 in $F^w(0)$ appears after either 0 or 2. However $e$ is a suffix of $g(2)$ and not of $g(0)$. Therefore, $e.g(3)$ is a suffix of $g(23)$ only. Since 23 is a suffix of $f(2)$ and not of $f(0)$, then $g(23.f(3))$ is a suffix of $g(f(23))$ only. Using this argument recursively, $p_{23}$ is a suffix of $g(f^d(23))$ only.

Now, the letter 3 in $F^w(0)$ appears before either 0 or 2, however $abcdeacdb$ is a prefix of $g(2)$ and not of $g(0)$. Thus $g(01203).abcdeacdb$ is a prefix of $g(012032)$ only. Since 012032 is a prefix of $f(2)$ and not of $f(0)$, then $g(f(01203).012032)$ only. Using this argument recursively, $s_{23}$ is a prefix of $g(f^d+1(012032))$ only. Thus, if $T_{23}$ is a factor of $w_5$, then $g(f^d(23))$ is a factor of $w_5$. This is a contradiction since 232 is not a factor of $f^w(0)$.

**Lemma 2.4.** Let $p_{0324} = acdbecd.g(24f(24) \ldots f^d(24)).f^d(24))$ and $s_{0324} = g(f^d+1(01240).f^d(01240).01240).abcede$. For every $d \geq 0$, the word $T_{0324} = p_{0324}g(f^d(0)).s_{0324}$ is a conjugate of $g(f^d(0324))$ that is not a factor of $w_5$.

**Proof.** Let us assume that $T_{0324}$ appears in $w_5$.

The letter 2 in $F^w(0)$ appears after either 1 or 3. However $acdbecd$ is a suffix of $g(3)$ and not of $g(1)$. Therefore $acdbecd.g(24)$ is a suffix of $g(324)$ only. Since 324 is a suffix of $f(3)$ and not of $f(1)$, then $g(324.f(24))$ is a suffix of $g(f(23))$ only. Using this argument recursively, $p_{0324}$ is a suffix of $g(f^d(324))$ only.

Now, the letter 0 in $F^w(0)$ appears before either 1 or 3. However $abcdbece$ is a prefix of $g(3)$ and not of $g(1)$. Thus $g(01240).abcdbece$ is a prefix of $g(012403)$ only. Since 012403 is a prefix of $f(3)$ and not of $f(1)$, then $g(f(01240).012403)$ is a prefix of $g(f(012403))$ only. Using this argument recursively, $s_{0324}$ is a prefix of $g(f^d+1(012403))$ only. Thus, if $T_{0324}$ is a factor of $w_5$, then $g(f^d(32403))$ is a factor of $w_5$. This is a contradiction since $32403$ is not a factor of $f^w(0)$.

\(^1\)See the program at http://www.lirmm.fr/~ochem/morphisms/conjugacy.htm
Lemma 2.5. Let \( p_{01240323} = \text{ecdeacdb}.g(0323f(0323) \ldots f^{d-1}(0323).f^{d}(0323)) \) and \( s_{01240323} = g(f(012)f^{d-1}(012) \ldots f(012)012).\text{abcd}. \) For every \( d \geq 0 \), the word \( T_{01240323} = p_{01240323}s_{01240323} \) is a conjugate of \( g(f^{d}(01240323)) \) that is not a factor of \( w_{5} \).

Proof. Let us assume that \( T_{01240323} \) appears in \( w_{5} \).

The factor 03 in \( f^{d}(0) \) appears after either 2 or 4. However \( \text{ecdeacdb} \) is a suffix of \( g(f(4)) \) and not of \( g(2) \). Therefore \( \text{ecdeacdb}.g(0323) \) is a suffix of \( g(40323) \) only. Since 40323 is a suffix of \( f(4) \) and not of \( f(2) \), then \( g(40323f(0323)) \) is a suffix of \( g(f(40323)) \), using this argument recursively, \( p_{01240323} \) is a suffix of \( g(f^{d}(40323)) \) only.

Now, the factor 12 in \( f^{d}(0) \) appears before either 0 or 4. However \( \text{abcd} \) is a prefix of \( g(4) \) and not of \( g(0) \). Thus \( g(012).\text{abcd} \) must only be a prefix of \( g(0124) \) and since 0323 is a prefix of \( f(4) \) and not of \( f(0) \) then \( g(f(012)0124) \) is a prefix of \( g(f(0124)) \) only. Using this argument recursively, \( s_{01240323} \) is a prefix of \( g(f^{d}(0124)) \) only. Thus, if \( T_{01240323} \) is a factor of \( w_{5} \), then \( g(f^{d}(403230124)) \) is a factor of \( w_{5} \). This is a contradiction since 403230124 is not a factor of \( f^{d}(0) \).

Lemma 2.6. Let \( p_{01203} = d.g(3f(3) \ldots f^{d-1}(3).f^{d}(3)) \) and \( s_{01203} = g(f^{d}(012).f^{d-1}(012).f^{d-2}(012) \ldots f(012)012).\text{abcdeac} \). For every \( d \geq 0 \), the word \( T_{01203} = p_{01203}s_{01203} \) is a conjugate of \( g(f^{d}(01203)) \) that is not a factor of \( w_{5} \).

Proof. Let us assume that \( T_{01203} \) appears in \( w_{5} \).

The letter 3 in \( f^{d}(0) \) appears after either 0 or 2. However \( d \) is a suffix of \( g(0) \) and not of \( g(2) \). Therefore \( d.g(2) \) is a suffix of \( g(12) \) only. Since 12 is a suffix of \( f(1) \) and not of \( f(3) \), then \( g(12f(2)) \) is a suffix of \( g(f(12)) \) only. Using this argument recursively, \( p_{01203} \) is a suffix of \( g(f^{d}(12)) \) only.

Now, 012 in \( f^{d}(0) \) appears before either 1 or 4, however \( \text{abcdeac} \) is only a prefix of \( g(1) \) and not of \( g(4) \). Thus \( g(012).\text{abcdeac} \) is a prefix of \( g(0120) \) only. Since 0120 is a prefix of \( f(1) \) and not of \( f(4) \), then \( g(f(012)0120) \) is a prefix of \( g(f(0120)) \) only. Using this argument recursively, \( s_{01203} \) is a prefix of \( g(f^{d}(0120)) \). Thus, if \( T_{01203} \) is a factor of \( w_{5} \), then \( g(f^{d}(030120)) \) is a factor of \( w_{5} \). This is a contradiction since 030120 is not a factor of \( f^{d}(0) \).

\[\text{References}\]

[1] K.A. Baker, G.F. McNulty and W. Taylor, Growth problems for avoidable words. Theoret. Comput. Sci. 69 (1989) 19–345.
[2] D.R. Bean, A. Ehrenfeucht and G.F. McNulty, Avoidable patterns in strings of symbols. Pacific J. Math. 85 (1979) 261–294.
[3] J.P. Bell and B.W. Madill, Iterative Algebras. Algebr. Represent. Theor. 18 (2015) 1533–1546.
[4] R.J. Clark, Avoidable formulas in combinatorics on words. Ph.D. thesis, University of California, Los Angeles. Available at http://www.lirmm.fr/~ochem/morphisms/clark_thesis.pdf (2001).
[5] G. Gamard, P. Ochem, G. Richomme and P. Séébold, Avoidability of circular formulas. Theoret. Comput. Sci. 726 (2018) 1–4.
[6] L. Ilie, P. Ochem and J.O. Shallit, A generalization of repetition threshold. Theoret. Comput. Sci. 92 (2004) 71–76.
[7] P. Ochem, A generator of morphisms for infinite words. RAIRO: ITA 40 (2006) 427–441.
[8] A.I. Zimin, Blocking sets of terms. Math. USSR Sbornik 47 (1984) 353–364.