RAIRO-Theor. Inf. Appl. 54 (2020) 2
https://doi.org/10.1051/ita/2020003

RAIRO - Theoretical Informatics and Applications
www.rairo-ita.org

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 $h : \Delta^* \to \Sigma^*$ 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 *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 $|w| \geqslant i$. 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 $m^\omega(\mathtt{0})$ and a morphic word is of the form $h(m^\omega(\mathtt{0}))$ 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] Goldsmiths, 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-CE40-0005).
‡ Corresponding author: ochem@lirmm.fr

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_i.

2. Main result

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

$$\begin{aligned} F(\mathtt{0}) &= \mathtt{01}, & G(\mathtt{0}) &= \mathtt{abcd},\\ F(\mathtt{1}) &= \mathtt{2}, & G(\mathtt{1}) &= \varepsilon,\\ F(\mathtt{2}) &= \mathtt{03}, & G(\mathtt{2}) &= \mathtt{eacd},\\ F(\mathtt{3}) &= \mathtt{24}, & G(\mathtt{3}) &= \mathtt{becd},\\ F(\mathtt{4}) &= \mathtt{23}. & G(\mathtt{4}) &= \mathtt{be}. \end{aligned}$$

Theorem 2.1. *The morphic word $w_5 \in \Sigma_5^*$ 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, $w_5 = g(f^\omega(\mathtt{0}))$.

$$\begin{aligned} f(\mathtt{0}) &= \mathtt{01203}, & g(\mathtt{0}) &= \mathtt{abcdeacd},\\ f(\mathtt{1}) &= \mathtt{0124}, & g(\mathtt{1}) &= \mathtt{abcdbecd},\\ f(\mathtt{2}) &= \mathtt{0120323}, & g(\mathtt{2}) &= \mathtt{abcdeacdbe},\\ f(\mathtt{3}) &= \mathtt{01240324}, & g(\mathtt{3}) &= \mathtt{abcdbecdeacdbecd},\\ f(\mathtt{4}) &= \mathtt{01240323}. & g(\mathtt{4}) &= \mathtt{abcdbecdeacdbe}. \end{aligned}$$

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

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 $C = \{F(2), F^2(2), F^d(4), f^d(0)\}$, for all $d \geqslant 1$.*

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

Now we show that $F(\mathtt{2})$ does not generate larger conjugacy classes in $F^\omega(\mathtt{0})$. We thus have to exhibit a conjugate of $f(F(\mathtt{2})) = F^4(\mathtt{2}) = \mathtt{0120301240324}$ that is not a factor of $F^\omega(\mathtt{0})$. A computer check shows that the conjugate $\mathtt{4012030124032}$ is not a factor of $F^\omega(\mathtt{0})$. Similarly, $F^2(\mathtt{2})$ does not generate larger conjugacy classes in $F^\omega(\mathtt{0})$ since the conjugate $\mathtt{301203012401203230124032}$ of $f(F^2(\mathtt{2})) = F^5(\mathtt{2}) = \mathtt{012030124012032301240323}$ is not a factor of $F^\omega(\mathtt{0})$. □

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 $\mathtt{ab}$. Notice that $\mathtt{ab}$ only appears in w_5 as the prefix of the g-image of every letter. Since $|w| \geqslant 41$, w contains at least 2 occurrences of $\mathtt{ab}$ and we write $w = ps$ such that the rightmost occurrence of $\mathtt{ab}$ in w is the prefix of s. Then the conjugate sp 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 $F^\omega(\mathtt{0})$, and so does every conjugate of v. Thus, $F^\omega(\mathtt{0})$ contains a conjugacy class c' such that the elements of c with prefix $\mathtt{ab}$ 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 $C = \{F(\mathtt{2}), F^2(\mathtt{2}), F^d(\mathtt{4}), f^d(\mathtt{0})\}$ for all $d \geqslant 1$. The computer check mentioned above settles the case of $F(\mathtt{2})$ and $F^2(\mathtt{2})$ since $|g(F(\mathtt{2}))| < |g(F^2(\mathtt{2}))| = 40 < 1000$. It also settles the case of $f(\mathtt{4})$ and $f(\mathtt{0})$ since $|g(f(\mathtt{0}))| < |g(f(\mathtt{4}))| = 90 < 1000$.

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

- $g(f^d(F(\mathtt{4}))) = g(f^d(\mathtt{23}))$
- $g(f^d(F^2(\mathtt{4}))) = g(f^d(\mathtt{0324}))$
- $g(f^{d+1}(\mathtt{4})) = g(f^d(\mathtt{01240323}))$
- $g(f^{d+1}(\mathtt{0})) = g(f^d(\mathtt{01203}))$

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

Lemma 2.3. *Let $p_{\mathtt{23}} = \mathtt{e}.g(\mathtt{3}f(\mathtt{3})\dots f^{d-1}(\mathtt{3}).f^d(\mathtt{3}))$ and $s_{\mathtt{23}} = g(f^{d-1}(\mathtt{01203}).f^{d-2}(\mathtt{01203})\dots f(\mathtt{01203})\mathtt{01203}).\mathtt{abcdeacdb}$. For every $d \geqslant 0$, the word $T_{\mathtt{23}} = p_{\mathtt{23}}s_{\mathtt{23}}$ is a conjugate of $g(f^d(\mathtt{23}))$ that is not a factor of w_5.*

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

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

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

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

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

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

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

[1]See the program at http://www.lirmm.fr/~ochem/morphisms/conjugacy.htm

Lemma 2.5. *Let* $p_{01240323} = \mathtt{ecdeacdbe}.g(0323f(0323)\cdots f^{d-1}(0323).f^{d}(0323))$ *and* $s_{01240323} = g(f^{d}(012)f^{d-1}(012).\cdots f(012)012).\mathtt{abcdb}$*. For every* $d \geqslant 0$*, 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 $f^{\omega}(0)$ appears after either 2 or 4. However ecdeacdbe is a suffix of $g(4)$ and not of $g(2)$. Therefore ecdeacdbe.$g(0323)$ 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 $f^{\omega}(0)$ appears before either 0 or 4. However abcdb is a prefix of $g(4)$ and not of $g(0)$. Thus $g(012)$.abcdb 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 $f^{\omega}(0)$. □

Lemma 2.6. *Let* $p_{01203} = \mathtt{d}.g(3f(3)\ldots f^{d-1}(3).f^{d}(3))$ *and* $s_{01203} = g(f^{d}(012)f^{d-1}(012).f^{d-2}(012)\ldots f(012)012).\mathtt{abcdeac}$*. For every* $d \geqslant 0$*, 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 $f^{\omega}(0)$ 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 $f^{\omega}(0)$ appears before either 1 or 4, however abcdeac is only a prefix of $g(1)$ and not of $g(4)$. Thus $g(012)$.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 $f^{\omega}(0)$. □

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.