

EXISTENCE OF AN INFINITE TERNARY 64-ABELIAN SQUARE-FREE WORD

MARI HUOVA

Abstract. We consider a recently defined notion of *k-abelian equivalence* of words by concentrating on avoidance problems. The equivalence class of a word depends on the numbers of occurrences of different factors of length k for a fixed natural number k and the prefix of the word. We have shown earlier that over a ternary alphabet k -abelian squares cannot be avoided in pure morphic words for any natural number k . Nevertheless, computational experiments support the conjecture that even 3-abelian squares can be avoided over ternary alphabets. In this paper we establish the first avoidance result showing that by choosing k to be large enough we have an infinite k -abelian square-free word over three letter alphabet. In addition, this word can be obtained as a morphic image of a pure morphic word.

Mathematics Subject Classification. 68R15.

INTRODUCTION

The theory of avoidance is one of the oldest and most studied topics in combinatorics on words. The first results were obtained by Axel Thue already at the beginning of the 20th century [17, 18]. He showed among other things the existence of an infinite binary word which does not contain any factor three times consecutively, *i.e.*, it avoids cubes. Similarly, he showed that squares can be avoided in infinite ternary words.

Since the late 1960's *abelian*, *i.e.*, commutative, variants of the above problems have been studied. Apparently, the first nontrivial results were obtained by Evdokimov [8] who showed that commutative squares can be avoided in infinite words over a 25-letter alphabet. The size of the alphabet was reduced to 5 by

Keywords and phrases. Combinatorics on words, k -abelian equivalence, square-freeness.

¹ Department of Mathematics and Statistics & TUCS, University of Turku, 20014 Turku, Finland. mari.huova@utu.fi

Pleasant [16], until the optimal value, 4, was found by Keränen [12]. Dekking [7] managed to prove already earlier that the optimal value for the size of the alphabet, in which abelian cubes are avoidable, is 3.

Our focus is on a new variant of the square-freeness. Repetitions are defined via equivalence relations which lie properly in between equality and abelian equality. For this relation we use the notion of *k-abelian equivalence*, where $k \geq 1$ is a natural number. Natural variants of the above Thue's problems ask what the size are for the smallest alphabets where *k-abelian* squares and cubes can be avoided. Because *k-abelian* equivalence lies between usual equivalence and abelian equivalence these two already examined equivalences give us bounds for our study. For *k-abelian* square-freeness we know that the interesting size of an alphabet is 3. In [11] it was shown that 2-abelian square-free words over ternary alphabets are finite, the longest being 537 letters. We have computational results which support the conjecture of the existence of an infinite ternary 3-abelian square-free word, see [9]. On the other hand, we have shown in the same paper [9] that an infinite ternary *k-abelian* square-free word cannot be obtained by iterating a morphism for any $k \geq 1$.

Although, iterated morphisms constitute a common tool in avoidability questions, there also exist patterns for usual word powers that can be avoided in binary words but not in words produced by only iterating a morphism. Cassaigne gives a classification of binary patterns according to avoidability in binary words, in binary pure morphic words and in ternary pure morphic words [3]. The patterns $\alpha^2\beta^2\alpha$, $\alpha\beta\alpha^2\beta$ and $\alpha\beta\alpha^2\beta\alpha$ are such that they can be avoided over a binary alphabet but not in infinite binary pure morphic words. Similarly, it seems that 3-abelian squares can be avoided over a ternary alphabet but not in infinite ternary pure morphic words. A related well-known example is given by the famous (cube-free) Kolakoski word: it is not pure morphic, see [5], but it is unknown whether it is morphic. On the other hand Currie has conjectured (see [6, 14], Problem 3.1.5, p. 132), that if a pattern p is avoidable over an alphabet Σ , then there exist an alphabet Σ' , two morphisms $f : \Sigma'^* \rightarrow \Sigma^*$ and $g : \Sigma'^* \rightarrow \Sigma'^*$ and a letter $a \in \Sigma'$ such that the infinite word $f(g^\infty(a))$ avoids p , that is, p is avoidable in morphic words.

In this paper we give the first result showing that *k-abelian* squares can be avoided in morphic words by choosing the value of k to be large enough. At the moment we have a proof for the cases where $k \geq 64$. The cases $3 \leq k \leq 63$ are still open. In addition, we give an example of a tool to show that a morphic word cannot be obtained by iterating a single morphism, *i.e.*, it is not pure morphic. In general, rather little is known about avoiding *k-abelian* repetitions and avoiding them even in morphic words. Results in [10, 15] are examples of the case in which the *k-abelian* cube-freeness over a binary alphabet is obtained in morphic words. The results cover the cases $k \geq 5$ and rest of the cases $2 \leq k \leq 4$ are under research. Based on our results and intuition we do not dare to make a conjecture related to Currie's for *k-abelian* repetitions even in the case that a pattern would be an integer power.

1. PRELIMINARIES

For the basic terminology of words as well as avoidability we refer to [4,13]. The basic notion in this paper, *k-abelian equivalence* of words, is defined as follows.

Definition 1.1. Let $k \geq 1$ be a natural number. We say that words u and v in Σ^+ are *k-abelian equivalent*, in symbols $u \equiv_k v$, if

1. $\text{pref}_{k-1}(u) = \text{pref}_{k-1}(v)$, and $\text{suf}_{k-1}(u) = \text{suf}_{k-1}(v)$; and
2. for all $w \in \Sigma^k$, the number of occurrences of w in u and v coincide, *i.e.* $|u|_w = |v|_w$.

Different words of length at most $k - 1$ are defined to be in-equivalent as well as all the words of different length.

Here pref_{k-1} (resp. suf_{k-1}) is used to denote the prefix (resp. suffix) of length $k - 1$ of a word.

It is straightforward to see that \equiv_k is an equivalence relation and, moreover,

$$u = v \Rightarrow u \equiv_k v \Rightarrow u \equiv_a v,$$

where \equiv_a denotes the abelian equivalence, same as \equiv_1 , and that

$$u = v \Leftrightarrow u \equiv_k v \quad \forall k \geq 1.$$

Now, notions like *k-abelian repetitions* are naturally defined. For instance $w = uv$ is a *k-abelian square* if and only if $u \equiv_k v$.

Example 1.2. Words *abaab* and *aabab* are 2-abelian equivalent and thus *abaabaabab* is a 2-abelian square. Those words are not 3-abelian equivalent but *ababbaabb* and *abbaababb* are 3-abelian equivalent and also 2-abelian equivalent.

A *prefix preserving* (or *prolongable*) *morphism* is a morphism $h : \Sigma^* \rightarrow \Sigma^*$ for which there exists a letter $a \in \Sigma$ and a word $\alpha \in \Sigma^*$ such that $h(a) = a\alpha$ and $h^n(\alpha) \neq \epsilon$ for every $n \geq 0$. We call an infinite word a *pure morphic word* if it is obtained by iterating a prefix preserving morphism. A *morphic word* is obtained from a pure morphic word by taking an image of it by a morphism or equivalently under a coding, see [1].

As mentioned in the introduction we have proved the following result in [9].

Theorem 1.3. *Every ternary infinite pure morphic word contains a k-abelian square for any $k \geq 1$.*

In our construction for an infinite 64-abelian square-free word we use two morphisms, one defined by Keränen and the other defined by Badkobeh and Crochemore. Keränen has proved that there exists an 85-uniform morphism over a four letter alphabet that generates an infinite abelian square-free word, [12].

For convenience, let us denote with g the morphism from $\{a, b, c, d\}^*$ into $\{0, 1, 2\}^*$ defined as follows:

$$\left\{ \begin{array}{l} a \mapsto 0102101202102010210121020120210120102120121020120210121 \\ \quad 0212010210121020102120121020120210121020102101202102012 \\ \quad 10212010210121020120210120102120121020102101210212, \\ b \mapsto 0102101202102010210121020120210120102120121020120210121 \\ \quad 0201021012021020121021201021012102010212012102012021012 \\ \quad 10212010210121020120210120102120121020102101210212, \\ c \mapsto 0102101202102010210121020120210120102120121020102101202 \\ \quad 1020121021201021012102010212012102012021012102010210120 \\ \quad 21020102120121020120210120102120121020102101210212, \\ d \mapsto 0102101202102010210121020120210120102120121020102101202 \\ \quad 1020102120121020120210121020102102102012102120102101 \\ \quad 21020102120121020120210120102120121020102101210212. \end{array} \right.$$

The morphism g is the one for which Badkobeh and Crochemore proved the following result in [2]:

Theorem 1.4 ([2]). *The morphism g translates any infinite $7/5^+$ -free word on the alphabet $\{a, b, c, d\}$ into a $7/4^+$ -free ternary word containing only two $7/4$ -powers, the fewest possible.*

We conclude this section by defining the notions *an identifying factor*, *an identifying prefix* and *an identifying suffix*. These all are defined with respect to a morphism. We denote the set of factors of a word w with $F(w)$ and the set of its prefixes with $\text{pref}(w)$.

Definition 1.5. An identifying factor with respect to a morphism $h : \Sigma_0^* \rightarrow \Sigma_1^*$ is such a factor $f \in \Sigma_1^+$ that

- (a) there exists a unique $a \in \Sigma_0$ such that $f \in F(h(a))$; and
- (b) for this a , $|h(w)|_f = |w|_a \cdot |h(a)|_f$ for all $w \in \Sigma_0^*$.

An identifying prefix with respect to a morphism $h : \Sigma_0^* \rightarrow \Sigma_1^*$ is such a prefix $p \in \Sigma_1^+$ of some word $h(i)$ where $i \in \Sigma_0$ that

- (a) there exists a unique $a \in \Sigma_0$ such that $p \in \text{pref}(h(a))$; and
- (b) for this a , $|h(w)|_p = |w|_a$ for all $w \in \Sigma_0^*$.

An identifying suffix is defined correspondingly.

These identifying objects with respect to a morphism $h : \Sigma_0^* \rightarrow \Sigma_1^*$ can be used to track down properties of a word $w \in \Sigma_0^*$ by analysing the factors of the word $h(w)$. If $h(a)$ for some $a \in \Sigma_0$ contains an identifying factor f then we can find out $|w|_a$ by counting $|h(w)|_f$ and $|h(a)|_f$. Identifying prefixes and suffixes can be used to locate letters in w . We give a short example to illustrate these notions and to show how we will use them.

Example 1.6. Let $h : \{a, b, c\}^* \rightarrow \{a, b, c\}^*$ be a morphism defined by

$$\begin{cases} a \mapsto abcab, \\ b \mapsto bcb, \\ c \mapsto cacb. \end{cases}$$

Now the six shortest identifying factors with respect to h are ab, ac, acb, bcb, cab and cac . From these six factors ab and ac are not identifying prefixes or suffixes but cac is an identifying prefix, acb and cab are identifying suffixes and bcb is both an identifying prefix and an identifying suffix. Consider a word $w' = bcbabcabcacbabcb$ which is obtained by taking a h -morphic image of some word w over an alphabet $\{a, b, c\}$. By counting the occurrences of the identifying factor ab in the word w' we can determine that w contains the letter a twice. Because ab is not an identifying prefix or suffix we cannot locate the positions of these a 's by looking for the positions of ab 's in w' . Instead, cab is an identifying suffix related to $h(a)$ and we can determine that $w' = bcbh(a)cacbh(a)$. By continuing analysis with identifying prefixes bcb and cac we can find out that $w = baca$.

2. A 64-ABELIAN SQUARE-FREE WORD

In this section we will prove the existence of a ternary infinite 64-abelian square-free word by taking a suitable morphic image of an infinite abelian square-free word.

Theorem 2.1. *Let $\Omega \in \{a, b, c, d\}^\omega$ be an infinite abelian square-free word and $g : \{a, b, c, d\}^* \rightarrow \{0, 1, 2\}^*$ the morphism introduced in the previous section. The infinite word $\omega = g(\Omega)$ over $\{0, 1, 2\}$ is 64-abelian square-free.*

Proof. First, we make some remarks about the morphism g and the word Ω . First of all, the word Ω exists and for example, the word constructed by Keränen could be chosen to be Ω . The morphism g is 160-uniform and synchronizing. It is of the form

$$g : \begin{cases} a \mapsto w\alpha yz \\ b \mapsto uv\beta yz \\ c \mapsto uw\gamma xz \\ d \mapsto uw\delta xz \end{cases}, \text{ where } \begin{cases} |u| = 47, |z| = 40, |v| = |x| = 10, |w| = |y| = 13, \\ |\alpha| = |\beta| = |\gamma| = |\delta| = 50. \end{cases}$$

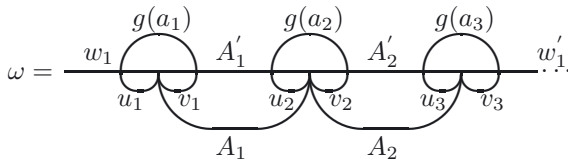
The word Ω can only contain factors ij where $i, j \in \{a, b, c, d\}$ and $i \neq j$. It can be easily checked by computer that now each word $g(ij)$ contains each of its factors of length 63 at most once except in cases $g(bd) = pfqrf s$ and $g(db) = rfs p f q$ where $|p| = 32, |f| = 78, |q| = 50, |r| = 57, |s| = 25$. In these cases there are 16 factors of length 63 that occur twice, namely the factors of f . Thus each $g(i)$ for $i \in \{a, b, c, d\}$ contains at least one identifying factor of each length from 63 up to 160 and each $g(i)$ contains also an identifying prefix and suffix of length 63

for any $i \in \{a, b, c, d\}$. In addition, the factor $\text{pref}_{30}(u)$ occurs only as a prefix of $g(a), g(b), g(c)$ and $g(d)$. Respectively for the factor $\text{suf}_{30}(z)$. If the word ω contains a factor of length 63 twice they cannot overlap by the observations above, so $\omega = \omega_0 t \omega_1 t \omega_2$ where $|t| = 63$. Now $|t\omega_1| = n \cdot 160 + A_1 \cdot 135 + A_2 \cdot 25$, where $n \in \{0, 1, 2, \dots\}$ and $(A_1, A_2) \in \{(0, 0), (0, 1), (1, 0)\}$. Here $A_1 \neq A_2$ if and only if $t \in F(f)$. In most of the cases the two occurrences of t originate from the images of two occurrences of the same letter in Ω . Consequently, the coefficient 160 comes from the length of the morphism. In the case that the first occurrence of t belongs to f in $g(d)$ and the second occurrence of t belongs to f in $g(b)$, *i.e.* $A_1 = 1$, there exists extra factor of 135 referring to $|fsp|$. If the occurrences of $t \in F(f)$ are in the opposite order, thus first in $g(b)$ and then in $g(d)$, then we have $A_2 = 1$ and $|fqr| = 185 = 160 + 25$.

We proceed by showing that if ω contained a 64-abelian square then Ω would not be abelian square-free or $g(\Omega)$ would not be square-free which would give a contradiction. We use identifying factors and prefixes to return our analysis to the properties of Ω . So assume that the word ω contains a 64-abelian square, *i.e.*, $\omega = wA_1A_2w'$ and $A_1 \equiv_{64} A_2$. If $|A_1| \leq 64$ then $A_1 = A_2$ and some of words $g(ij)$, where $ij \in \{a, b, c, d\} (\{a, b, c, d\} \setminus \{i\})$, should contain a square which contradicts the result of Theorem 1.4. So we may assume that $|A_1| > 64$.

Because $A_1 \equiv_{64} A_2$ we have $\text{pref}_{63}(A_1) = \text{pref}_{63}(A_2)$ and ω has the same factor of length 63 twice. Similarly, $\text{suf}_{63}(A_1) = \text{suf}_{63}(A_2)$ and thus $|A_1| = n_p \cdot 160 + A_{1p} \cdot 135 + A_{2p} \cdot 25$ and $|A_2| = n_s \cdot 160 + A_{1s} \cdot 135 + A_{2s} \cdot 25$. Now $|A_1| = |A_2|$ and the only possibility is that $n_p = n_s, A_{1p} = A_{1s}$ and $A_{2p} = A_{2s}$. If $A_{1p} = 1$, then both $\text{suf}_{63}(A_1) \in F(f)$ and $\text{pref}_{63}(A_2) \in F(f)$. In fact, $\text{suf}_{63}(A_1)\text{pref}_{63}(A_2) \in F(f)$ should be a factor of f which is not possible because $|\text{suf}_{63}(A_1)\text{pref}_{63}(A_2)| > |f|$. Similar reasoning holds if $A_{2p} = 1$. So we have $A_{1p} = A_{2p} = A_{1s} = A_{2s} = 0$ and $|A_1| = |A_2| = n_p \cdot 160$.

Let $\omega = wA_1A_2w' = w_1g(a_1)A'_1g(a_2)A'_2g(a_3)w'_1$, where $a_1, a_2, a_3 \in \{a, b, c, d\}$, $g(a_1) = u_1v_1, g(a_2) = u_2v_2, g(a_3) = u_3v_3$ and $v_1A'_1u_2 = A_1$ and $v_2A'_2u_3 = A_2$. The following graph illustrates the situation.



Because $|A_1| = |A_2| = n_p \cdot 160$ we have $|u_1| = |u_2| = |u_3|$ and $|v_1| = |v_2| = |v_3|$. Now we can divide the study into two cases.

If $|u_1| = 0$ then $A_1 = g(\alpha_1), A_2 = g(\alpha_2)$ for some $\alpha_1\alpha_2 \in F(\Omega)$. Because each $g(i)$ for any $i \in \{a, b, c, d\}$ contains at least one identifying factor of length 64 and $A_1 \equiv_{64} A_2$ so α_1 should be abelian equivalent to α_2 . This gives a contradiction because Ω is abelian square-free.

If $|u_1| = m > 0$ then $A_1 = v_1g(\alpha_1)u_2, A_2 = v_2g(\alpha_2)u_3$ for some $\alpha_1a_2\alpha_2 \in F(\Omega)$. We may assume $|u_2| > 63$, the case $|v_2| > 63$ would be similar. Now $\text{suf}_{63}(u_2) = \text{suf}_{63}(A_1) = \text{suf}_{63}(A_2) = \text{suf}_{63}(u_3)$ and $|u_2v_2g(\alpha_2)| = (|\alpha_2| + 1) \cdot 160$ so u_2 have to be equal to u_3 and $a_2 = a_3$, too. Because each $g(i)$ for any $i \in \{a, b, c, d\}$ has an identifying prefix of length 64 and $A_1 \equiv_{64} A_2$, so $g(\alpha_1a_2)$ and $g(\alpha_2a_3) = g(\alpha_2a_2)$ have to have those same identifying factors, so $\alpha_1 \equiv_a \alpha_2$, too. This gives a contradiction because now $\alpha_1a_2\alpha_2a_3 \in F(\Omega)$ and $\alpha_1a_2 \equiv_a \alpha_2a_3$. \square

Now we have proved the existence of a 64-abelian square-free word. By choosing the initial abelian square-free word to be morphic we get that the 64-abelian square-free word constructed as in Theorem 2.1 is also morphic. Though, by Theorem 1.3 this word cannot be pure morphic.

Example 2.2. Let Ω_K be the abelian square-free word over four letter alphabet constructed by Keränen in [12]. The word is produced by iterating an 85-uniform morphism, *i.e.*, Ω_K is pure morphic. By taking a morphic image of it with the morphism g we have a 64-abelian square-free word, which is morphic. Now by Theorem 1.3 $g(\Omega_K)$ cannot be pure morphic. In general, this can be formulated as follows: let Ω_m be a morphic word over $\{a, b, c, d\}$ then there does not exist a single morphism that would generate the morphic word $g(\Omega_m)$ directly.

3. CONCLUSIONS

We know that k -abelian squares are avoidable either over a ternary alphabet or over a four letter alphabet. The avoidability behavior seems to depend on the value of k . An earlier result shows that 2-abelian squares are not avoidable over ternary alphabets. Computational evidence suggests that it would be enough to increase the value of k by one to achieve avoidability. On the other hand, we have shown that k -abelian squares are not avoidable in pure morphic words over ternary alphabet for any value of k .

In this paper we proved that 64-abelian squares can be avoided over three letter alphabet. This is the first result for avoidability concerning k -abelian squares in ternary alphabets. In addition, the 64-abelian square-free word can be chosen to be morphic. It cannot be pure morphic but it is enough to take a morphic image of a pure morphic word. On the other hand, this shows how we can use Theorem 1.3 as a tool for proving that a word is not achievable by generating a single morphism. The further aim of the research is to decrease the size of k and find the bound for the value of k to determine avoidability and unavoidability, if it is possible. This bound would divide the k -abelian equivalence classes to those that behave like usual equality and to those that behave like abelian equality.

Acknowledgements. The author would like to thank the anonymous referees for constructive comments.

REFERENCES

- [1] J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press (2003).
- [2] G. Badkobeh and M. Crochemore, Finite-Repetition threshold for infinite ternary words, in *8th International Conference WORDS 2011*, edited by P. Ambroz, S. Holub and Z. Masáková. *EPTCS* **63** (2011) 37–43.
- [3] J. Cassaigne, Unavoidable binary patterns. *Acta Informatica* **30** (1993) 385–395.
- [4] C. Choffrut and J. Karhumäki, Combinatorics of words, in vol. 1 of *Handbook of Formal Languages*, edited by G. Rozenberg and A. Salomaa. Springer (1997) 329–438.
- [5] K. Culik, J. Karhumäki and A. Lepistö, Alternating iteration of morphisms and the Kolakoski sequence, in *Lindenmayer Systems*, edited by G. Rozenberg and A. Salomaa. Springer (1992) 93–106.
- [6] J.D. Currie, Open problems in pattern avoidance. *Amer. Math. Monthly* **100** (1993) 790–793.
- [7] F.M. Dekking, Strongly non-repetitive sequences and progression-free sets. *J. Combin. Theory Ser. A* **27** (1979) 181–185.
- [8] A.A. Evdokimov, Strongly asymmetric sequences generated by a finite number of symbols. *Dokl. Akad. Nauk SSSR* **179** (1968) 1268–1271; English translation in *Soviet Math. Dokl.* **9** (1968) 536–539.
- [9] M. Huova and J. Karhumäki, On Unavoidability of k -abelian Squares in Pure Morphic Words. *J. Integer Sequences*, being published.
- [10] M. Huova, J. Karhumäki and A. Saarela, Problems in between words and abelian words: k -abelian avoidability, in *Formal and Natural Computing Honoring the 80th Birthday of Andrzej Ehrenfeucht*, edited by G. Rozenberg and A. Salomaa. *Theoret. Comput. Sci.* **454** (2012) 172–177.
- [11] M. Huova, J. Karhumäki, A. Saarela and K. Saari, Local squares, periodicity and finite automata, in *Rainbow of Computer Science*, edited by C.S. Calude, G. Rozenberg and A. Salomaa. Vol. 6570 of *Lect. Notes Comput. Sci.* Springer (2011) 90–101.
- [12] V. Keränen, Abelian squares are avoidable on 4 letters, in *Proc. ICALP 1992*, edited by W. Kuich. Vol. 623 of *Lect. Notes Comput. Sci.* Springer (1992) 41–52.
- [13] M. Lothaire, *Combinatorics on words*, Addison-Wesley (1983).
- [14] M. Lothaire, *Algebraic combinatorics on words*. Cambridge University Press (2002).
- [15] R. Mercas and A. Saarela, 5-abelian cubes are avoidable on binary alphabets, in *the conference the 14th Mons Days of Theoretical Computer Science* (2012).
- [16] P.A.B. Pleasant, Non-repetitive sequences. *Proc. Cambridge Philos. Soc.* **68** (1970) 267–274.
- [17] A. Thue, Über unendliche Zeichenreihen. *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **7** (1906) 1–22.
- [18] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **1** (1912) 1–67.

Communicated by Ch. Choffrut.

Received March 14, 2014. Accepted March 14, 2014.