RAIRO-Theor. Inf. Appl. 48 (2014) 467–478
DOI: 10.1051/ita/2014020

Available online at:
www.rairo-ita.org

5-ABELIAN CUBES ARE AVOIDABLE ON BINARY ALPHABETS *, **

Robert Mercaş[1] and Aleksi Saarela[2]

Abstract. A k-abelian cube is a word uvw, where the factors u, v, and w are either pairwise equal, or have the same multiplicities for every one of their factors of length at most k. Previously it has been shown that k-abelian cubes are avoidable over a binary alphabet for $k \geq 8$. Here it is proved that this holds for $k \geq 5$.

Mathematics Subject Classification. 68Q70, 68R15.

1. Introduction

The concept of repetition-freeness has been investigated in Combinatorics on Words since the beginning of the area, see [15,16]. In a series of papers Axel Thue manages to show that 3 letters are enough to construct an infinite word that does not contain two consecutive repetitions of the same factor, while with 2 letters one can construct an infinite number of words whose factors are consecutively repeated at most twice. For a word uvw, we say that uv is a square if $u = v$ and both words are non-empty, and we say that uvw is an overlap if all words are non-empty, and either $u = v$ and w is a prefix of u, or $v = w$ and u is a suffix of w. For example, the word *banana* contains the squares *anan* and *nana* and the overlap *anana*. A

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

* *Supported by the Alexander von Humboldt Foundation while author was affiliated at Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik, PSF 4120, 39016 Magdeburg, Germany.*

** *Supported by the Academy of Finland under grants 137991 (FiDiPro) and 257857.*

[1] Christian-Albrechts-Universität zu Kiel, Institut für Informatik, 24098 Kiel, Germany. rgm@informatik.uni-kiel.de

[2] Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland. amsaar@utu.fi

Article published by EDP Sciences

© EDP Sciences 2014

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press

word is called square (overlap)-free if it does not contain any square (overlap) as a factor. Due to their initial obscure publication, these results have been later on rediscovered several times.

In 1961 Erdős [5] raised the question whether abelian squares can be avoided in infinitely long words, *i.e.*, whether there exist words over a given alphabet that do not contain two consecutive permutations of the same factor. An abelian square is a non-empty word uv, where u and v have the same number of occurrences of each symbol. For example, *intestines* is an abelian square where *intes* is a permutation of the word *tines*. A word is abelian square-free, if it does not contain any abelian square as a factor. It is easily seen that abelian squares cannot be avoided over a three-letter alphabet, each word of length eight over three letters containing an abelian square. Dekking [4] proved that over a binary alphabet there exists a word that is abelian 4-free. Moreover, using $\mathbb{Z}_7$ instead of $\mathbb{Z}_5$ in the proof of this result, we get that over a ternary alphabet an abelian 3-free infinite word is constructible. The problem of whether abelian squares can be avoided over a four-letter alphabet remained open for a long time. In [11], using an interesting combination of computer checking and mathematical reasoning, Keränen proves that abelian squares are avoidable on four letters.

The problem of repetition-freeness was studied in the last decade also from the point of view of partial words, words that beside regular symbols from some alphabet Σ also contain "hole" symbols that match all other symbols from Σ. It was shown that there exist infinite ternary words with an infinite number of holes whose factors are not matching any squares (overlaps) of words of length greater than one [2,12]. For the abelian case an alphabet with as few as 5 letters is enough in order to construct a word with an infinite number of holes such that none of its factors matches any abelian square word of length greater than two [1].

Recently, the study of abelian repetition-freeness was extended to k-abelian repetition-freeness. That is a word is k-abelian ℓ-free if it does not contain ℓ consecutive occurrences of a factor, nor does it contain ℓ consecutive factors of length at least k, all sharing the same multiplicities for each of their factors of length at most k. In particular, the word *abcacbacabcb* is a 2-abelian square as both the first half and the second half of it share the same numbers for each of the letters, as well as one and only one occurrence of each of the factors *ab*, *ac*, *bc*, *ca*, and *cb*. Thus, the k-abelian concept is situated in between the equality and the usual abelian equivalence (looking for an ℓ-repetition inside a word of length n is equivalent to looking for an $\lceil n/\ell \rceil$-abelian repetition, while looking for an abelian ℓ-repetition is equivalent to looking for a 1-abelian ℓ-repetition).

In [9], the authors show that 2-abelian squares are avoidable only on a four letter alphabet. For $k \geq 3$, the question of avoiding k-abelian squares remained open, the minimal alphabet size assumed to be either three or four, but it is known that there are no pure morphic k-abelian square-free ternary words for any k [7]. It was proved in [8] that for avoiding k-abelian cubes a binary alphabet suffices whenever $k \geq 8$. However, the same work conjectures that a binary alphabet might be enough even for a k as low as 2, since computer generated words of length 100 000 having

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press

the property have been found. We also mention that other research regarding the k-abelian concept has been done, and includes for example a version of the Fine and Wilf's periodicity theorem [10].

This work closes on the gap of k-abelian cube-freeness. While finding a good morphism satisfying such properties could be done by using a computer, the difficulty of the problem stands in the proof techniques that one has to employ in proving the validity of such a result. We show that with the help of some morphisms created for the classical avoidability of squares in the context of partial words, we are able to first improve the bound of k to 6. The constructed 6-abelian cube-free word is of the form $h(w)$, where h is a specific morphism and w is an arbitrary abelian square-free word on a four letter alphabet. Even more, we improve the bound of k to 5. The constructed 5-abelian cube-free word is of the form $h(w)$, where h is a specific morphism and w is a specific abelian cube-free word on a three letter alphabet. As an outcome, we bind results regarding classical repetitions within the partial words framework together with results from the abelian equivalence setting, in order to get a result somewhere in between.

Next section provides some preliminaries on Combinatorics on Words and repetition-freeness in particular. In Sections 3 and 4, we present our improvements for the k-abelian cubes bounds.

Please note that between the first presentation of this paper and its publication, the main questions of avoidability for k-abelian squares and cubes were fully settled [6,13,14]. Thus, it is now known that over a binary alphabet 2-abelian cubes are avoidable, while only a ternary alphabet is needed in order to avoid 3-abelian squares [14].

2. Preliminaries

We denote by Σ a finite set of symbols called *alphabet*. A *word* w represents a concatenation of letters from Σ. By ε we denote the *empty symbol*. We denote by $|w|$ the *length* of w and by $|w|_u$ the number of occurrences of u in w. The *concatenation* of two words u and v is the word uv obtained by adding to the right of u the letters of the word v. The set of all words over Σ is denoted by Σ^*, while the set of all words of length n is denoted by Σ^n for any positive integer n.

For a factorization $w = uxv$, we say that x is a *factor* of w, and whenever u is empty x is a *prefix* of w. Analogously, having v empty, makes x a *suffix* of w. If $w = a_0a_1 \dots a_{n-1}$ with $a_i \in \Sigma$ for $0 \leq i < n$ we say that $\mathrm{rfact}_k^i(w) = a_i \dots a_{i+k-1}$ is the *right factor at* i of w with length k and that $\mathrm{lfact}_k^i(w) = a_{i-k} \dots a_{i-1}$ is the *left factor at* i of w with length k. Furthermore, $\mathrm{pref}_k(w) = \mathrm{rfact}_k^0(w)$ is the length k *prefix* of w and $\mathrm{suff}_k(w) = \mathrm{lfact}_k^n(w)$ is the length k *suffix* of w.

The *powers* of a word w are defined recursively, $w^0 = \varepsilon$ and $w^n = ww^{n-1}$ for $n > 0$. We say that w is an ℓth power if there exists a word u such that $w = u^\ell$. When the only possible descriptions of w are for the value 1 for ℓ, we say that w is *primitive*. Second powers are called *squares* and third powers *cubes*. In the end, we denote by $w^\omega = \lim_{n\to\infty} w^n$ the word obtained by infinite concatenations of w.

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press

Words u and v are *abelian equivalent* if $|u|_a = |v|_a$ for all letters $a \in \Sigma$.

We say that u and v are *k-abelian equivalent* if either $u = v$ or $|u|, |v| \geq k-1$, $|u|_t = |v|_t$ for every $t \in \Sigma^k$, $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v)$ and $\mathrm{suff}_{k-1}(u) = \mathrm{suff}_{k-1}(v)$. An equivalent definition would be that u and v are k-abelian equivalent if $|u|_t = |v|_t$ for every word t of length at most k. Obviously, 1-abelian equivalence is the same as abelian equivalence.

A *k-abelian nth power* is a word $u_1u_2 \dots u_n$, where $u_1, u_2, \dots, u_n$ are pairwise k-abelian equivalent. For $k = 1$ this gives the definition of an *abelian nth power*.

We say that $f : A^* \to B^*$ is a *morphism* if $f(xy) = f(x)f(y)$ for any words $x, y \in A^*$. A morphism f is *n-uniform* if $|f(a)| = n$ for all $a \in A$. It is *uniform* if it is n-uniform for some n.

In what follows, $\Sigma_2 = \{a, b\}$, $\Sigma_3 = \{a, b, c\}$, and so on $\Sigma_p = \{a_1, a_2, \dots, a_p\}$.

One initial result regarding repetition-freeness is due to Thue [16]:

Theorem 2.1. *The Thue-Morse word given by the fixed point $t^\omega(a)$ of the morphism $t : \Sigma_2^* \to \Sigma_2^*$, with $t(a) = ab$ and $t(b) = ba$, is overlap-free.*

In [3], in the context of partial words, the authors give the following morphism $\xi : \{a, b\}^* \to \{a, b, \diamond\}^*$ defined by

$$\xi(a) = babaab\diamond abbaba\diamond \qquad \text{and} \qquad \xi(b) = baabba\diamond.$$

This morphism provides a way to construct an infinite cube-free partial word in which each length seven factor contains a hole symbol, denoted $\diamond$.

Theorem 2.2. [3] *Mapping the Thue–Morse word with ξ we obtain a word that remains cube-free no matter how we replace the $\diamond$ symbols by letters from $\{a, b\}$.*

Unlike ordinary cubes, abelian cubes cannot be avoided over a binary alphabet, and unlike ordinary squares, abelian squares are not avoidable over a ternary alphabet. However, Dekking showed in [4] that two letters are sufficient for avoiding abelian fourth powers, and three letters suffice for avoiding abelian cubes. We need the following extension of the latter result.

Theorem 2.3. *Let w be a fixed point of morphism $\sigma : \Sigma_3^* \to \Sigma_3^*$ defined by*

$$\sigma(a) = aabc, \qquad \sigma(b) = bbc, \qquad \sigma(c) = acc.$$

Then w is abelian cube-free and contains no factor of the form $apbqbrc$ for some abelian equivalent words $p, q, r \in \Sigma_3^*$.

Proof. The word w was shown to be abelian cube-free in [4]. Similar ideas can be used to show that w avoids the factors $apbqbrc$.

Let $f : \Sigma_3^* \to \mathbb{Z}_7$ be the morphism defined by

$$f(a) = 1, \qquad f(b) = 2, \qquad f(c) = 3$$

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press

(here $\mathbb{Z}_7$ is the additive group of integers modulo 7). It follows that $f(\sigma(d)) = 0$ for all $d \in \Sigma_3$. If $apbqbrc$ is a factor of w, then there exist some words $u, s, s' \in \Sigma_3^*$ such that $\sigma(u) = sapbqbrcs'$. Consider the values

$$f(s), f(sa), f(sap), f(sapb), f(sapbq), f(sapbqb), f(sapbqbr), f(sapbqbrc). \quad (2.1)$$

These elements are of the form $f(\sigma(u')v') = f(v')$, where v' is a prefix of one of the words $aabc, bbc, acc$. Hence, the possible values for $f(v')$ are 0, 1, 2 and 4. However, if p, q and r are abelian equivalent, then $f(p) = f(q) = f(r)$. Thus, if we denote $i = f(s)$ and $j = f(p) = f(q) = f(r)$, then the corresponding values for the elements in (2.1) are

$$i, i+1, i+j+1, i+j+3, i+2j+3, i+2j+5, i+3j+5, i+3j+1. \quad (2.2)$$

For all $i \in \{0, 1, 2, 4\}$ and all $j \in \mathbb{Z}_7$, one of the values in (2.2) is not in $\{0, 1, 2, 4\}$. This is a contradiction. □

In [11], abelian squares were proven to be avoidable over a four letter alphabet.

Theorem 2.4. *Over a four letter alphabet there exists an infinite word that contains no two consecutive factors that are permutations of each other.*

We know that if abelian cubes are avoidable over some alphabet, then so are k-abelian cubes. Therefore, k-abelian cubes are avoidable over a ternary alphabet for all k. But for which k are they avoidable over a binary alphabet? In [8] it was proved that this holds for $k \geq 8$, and conjectured that it holds for $k \geq 2$. In this article we prove that this holds for k as low as 5.

3. 6-ABELIAN CUBE-FREENESS

The following result proves quite useful in our investigation.

Lemma 3.1. *Let $w \in \Sigma_p^\omega$ and $n \geq 2k-3$ for some positive integers k, n, p. Let $\theta : \Sigma_p^* \to \Sigma_2^*$ be an n-uniform morphism. If there do not exist $v_0, v_1, v_2, v_3 \in \Sigma_p^2$ and $i_0, i_1, i_2, i_3 \in \{k-1, k, \ldots, n+k-2\}$ such that*

$$i_1 - i_0 \equiv i_2 - i_1 \equiv i_3 - i_2 \not\equiv 0 \mod n \quad (3.1)$$

and

$$\begin{aligned} &\mathrm{rfact}_{k-1}^{i_0}(\theta(v_0)) = \mathrm{rfact}_{k-1}^{i_1}(\theta(v_1)) = \mathrm{rfact}_{k-1}^{i_2}(\theta(v_2)) \qquad \textit{and} \\ &\mathrm{lfact}_{k-1}^{i_1}(\theta(v_1)) = \mathrm{lfact}_{k-1}^{i_2}(\theta(v_2)) = \mathrm{lfact}_{k-1}^{i_3}(\theta(v_3)), \end{aligned} \quad (3.2)$$

then $\theta(w)$ does not contain any k-abelian cube of a word with length at least $k-1$ and not divisible by n.

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press

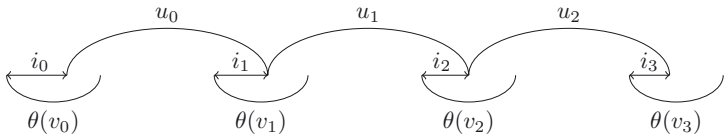


FIGURE 1. A diagram related to the proof of Lemma 3.1.

Proof. We assume towards a contradiction that $\theta(w)$ contains a k-abelian cube of length $3m$, where $m \geq k-1$ and $n \nmid m$, and show that there exist words v_0, v_1, v_2, v_3 and numbers i_0, i_1, i_2, i_3 such that both conditions (3.1) and (3.2) are satisfied.

Let $\mathrm{rfact}^{i}_{3m}(\theta(w))$ be a k-abelian cube. Then the words $u_j = \mathrm{rfact}^{i+jm}_{m}(\theta(w))$ $(j = 0, 1, 2)$ are pairwise k-abelian equivalent. Let $i_j \in \{k-1, k, \ldots, n+k-2\}$ be such that $i_j \equiv i + jm \mod n$. It follows that the numbers i_0, i_1, i_2, i_3 satisfy (3.1). Furthermore, there exist $v_0, v_1, v_2, v_3 \in \Sigma_p^2$ such that

$$\begin{aligned}
\mathrm{pref}_{k-1}(u_0) &= \mathrm{rfact}^{i}_{k-1}(\theta(w)) = \mathrm{rfact}^{i_0}_{k-1}(\theta(v_0)),\\
\mathrm{suff}_{k-1}(u_0) &= \mathrm{lfact}^{i+m}_{k-1}(\theta(w)) = \mathrm{lfact}^{i_1}_{k-1}(\theta(v_1)),\\
\mathrm{pref}_{k-1}(u_1) &= \mathrm{rfact}^{i+m}_{k-1}(\theta(w)) = \mathrm{rfact}^{i_1}_{k-1}(\theta(v_1)),\\
\mathrm{suff}_{k-1}(u_1) &= \mathrm{lfact}^{i+2m}_{k-1}(\theta(w)) = \mathrm{lfact}^{i_2}_{k-1}(\theta(v_2)),\\
\mathrm{pref}_{k-1}(u_2) &= \mathrm{rfact}^{i+2m}_{k-1}(\theta(w)) = \mathrm{rfact}^{i_2}_{k-1}(\theta(v_2)),\\
\mathrm{suff}_{k-1}(u_2) &= \mathrm{lfact}^{i+3m}_{k-1}(\theta(w)) = \mathrm{lfact}^{i_3}_{k-1}(\theta(v_3)).
\end{aligned}$$

Condition (3.2) is a consequence of the previous observations and the fact that u_0, u_1, and u_2 are k-abelian equivalent. □

The following observation investigates properties of k-abelian cube-free words obtained by the application of some morphism on an abelian square-free word:

Proposition 3.2. *Let $w \in \Sigma_p^\omega$ be an abelian square-free word and $\theta : \Sigma_p^* \to \Sigma_2^*$ an n-uniform morphism. Furthermore, assume there exist number $l \in \{k-1, k, \ldots, n-k+1\}$ and p words $t_1, t_2, \ldots, t_p \in \Sigma_2^k$ such that for all letters $a', b' \in \Sigma_p$ the following conditions hold:*

1) *for every $j \in \{k-1, k, \ldots, l\}$, if $\mathrm{lfact}^{j}_{k-1}(\theta(a')) = \mathrm{lfact}^{j}_{k-1}(\theta(b'))$, then $\mathrm{pref}_j(\theta(a')) = \mathrm{pref}_j(\theta(b'))$;*
2) *for every $j \in \{l+1, l+2, \ldots, n-k+1\}$, if If $\mathrm{rfact}^{j}_{k-1}(\theta(a')) = \mathrm{rfact}^{j}_{k-1}(\theta(b'))$, then $\mathrm{suff}_{n-j}(\theta(a')) = \mathrm{suff}_{n-j}(\theta(b'))$;*
3) *the words $t_1, t_2, \ldots, t_p$ are not factors of $\mathrm{suff}_{k-1}(\theta(a'))\mathrm{pref}_{k-1}(\theta(b'))$;*
4) *the p vectors $(|\theta(a_1)|_{t_j}, |\theta(a_2)|_{t_j}, \ldots, |\theta(a_p)|_{t_j})$ are linearly independent for all $j \leq p$.*

Then $\theta(w)$ does not contain the k-abelian cube of a word whose length is divisible by n.

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press

Proof. For clarity we start by mentioning that the first two Conditions say that if the images of the letters have the same factor of length $k-1$ ending, respectively, starting, at some position, then they have the same prefix up to that position, respectively, the same suffix starting at that position. Furthermore, the last two Conditions say that there exist at least as many factors as the size of the alphabet, whose occurrences in the two images differ.

Assume now towards contradiction that $\theta(w)$ contains such a k-abelian cube. That is, for some position i, the words $u_j = \mathrm{rfact}_m^{i+jm}(\theta(w)) \in \Sigma_2^*$ are k-abelian equivalent for $j \in \{0,1,2\}$ and, furthermore, $m = m'n$. Let $i' \equiv i \mod n$ be such that $i' \in \{0,1,\ldots,n-1\}$. There are two cases to be analysed.

First, we consider the case when $i' \leq l$. Let

$$u'u_0u_1u_2 = \theta(v_0v_1v_2)u'', \tag{3.3}$$

where $|u'| = |u''| = i'$, $u', u'', u_1, u_2, u_3 \in \Sigma_2^*$, $v_0, v_1, v_2 \in \Sigma_p^{m'}$ and the word $v_0v_1v_2$ is a factor of w. By (3.3), it follows that

$$\theta(v_1) = \mathrm{suff}_{i'}(u_0)\mathrm{pref}_{m-i'}(u_1)$$

and

$$\theta(v_2) = \mathrm{suff}_{i'}(u_1)\mathrm{pref}_{m-i'}(u_2),$$

and thus, for every $t \in \Sigma_2^k$ we have

$$|\theta(v_1)|_t = |\mathrm{suff}_{i'}(u_0)\mathrm{pref}_{k-1}(u_1)|_t + |u_1|_t - |\mathrm{suff}_{i'+k-1}(u_1)|_t$$

and

$$|\theta(v_2)|_t = |\mathrm{suff}_{i'}(u_1)\mathrm{pref}_{k-1}(u_2)|_t + |u_2|_t - |\mathrm{suff}_{i'+k-1}(u_2)|_t.$$

The words $\mathrm{suff}_{i'}(u_0)$, $\mathrm{suff}_{i'}(u_1)$ and $\mathrm{suff}_{i'}(u_2)$ are prefixes of some of the images of the letters in Σ_p. Because u_0, u_1 and u_2 are k-abelian equivalent, their suffixes of length $k-1$ are the same. From Condition 1 of the Proposition 3.2, it follows that $\mathrm{suff}_{i'}(u_0) = \mathrm{suff}_{i'}(u_1) = \mathrm{suff}_{i'}(u_2)$. But also $\mathrm{pref}_{k-1}(u_1) = \mathrm{pref}_{k-1}(u_2)$, hence

$$|\mathrm{suff}_{i'}(u_0)\mathrm{pref}_{k-1}(u_1)|_t = |\mathrm{suff}_{i'}(u_1)\mathrm{pref}_{k-1}(u_2)|_t.$$

Since u_1 and u_2 are k-abelian equivalent, we have that $|u_1|_t = |u_2|_t$. By Condition 3 of the Proposition, if $t \in \{t_1, t_2, \ldots, t_p\}$, then

$$|\mathrm{suff}_{i'+k-1}(u_1)|_t = |\mathrm{suff}_{i'}(u_1)|_t = |\mathrm{suff}_{i'}(u_2)|_t = |\mathrm{suff}_{i'+k-1}(u_2)|_t.$$

It has been shown that for $i \in \{1,2,\ldots,p\}$ we have

$$|\theta(v_1)|_{t_i} = |\theta(v_2)|_{t_i}. \tag{3.4}$$

Let M be the invertible $p \times p$ matrix whose rows are the vectors from Condition 4. For any $v \in \Sigma_p^*$, if P_v is the Parikh vector of v, then

$$MP_v^T = (|\theta(v)|_{t_1}, |\theta(v)|_{t_2}, \ldots, |\theta(v)|_{t_p})^T. \tag{3.5}$$

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press

From (3.4) and (3.5), we have that $MP_{v_1}^T = MP_{v_2}^T$. However, since M is invertible, it follows that $P_{v_1} = P_{v_2}$. Thus w contains an abelian square v_1v_2, which is a contradiction. This concludes our first case.

The other case to consider is when $i' > l$. As we will see this is similar to the previous case when $i' \leq l$. Let

$$u_0u_1u_2u' = u''\theta(v_0v_1v_2),$$

where now $|u'| = |u''| = n - i'$, and, again, $u', u'', u_1, u_2, u_3 \in \Sigma_2^*$, $v_0, v_1, v_2 \in \Sigma_p^{m'}$, and the word $v_0v_1v_2$ is a factor of w. As above, Conditions 2 and 3 imply that $|\theta(v_0)|_{t_i} = |\theta(v_1)|_{t_i}$ for any $i \in \{1, 2, \ldots, p\}$. Making use of the matrix M, it follows once more that v_0 and v_1 are abelian equivalent, which is a contradiction with our initial assumption. This concludes our proof. □

Further, we shall try to junction different images of ξ, where the $\diamond$'s are replaced by letters of Σ_2, such that applying these newly formed morphisms to some specific words, it will provide us with new words that are k-abelian cube-free for some k as small as possible.

A computer search reveals that for a four letter alphabet, $p = 4$, when we combine two by two images of ξ, there are, up to symmetry, only two morphisms that satisfy the conditions of Lemma 3.1 for Σ_4 and $k = 5$, where abelian square-free words given by Theorem 2.4 are employed. However, unfortunately, both map some abelian square-free words to a word that contains a 5-abelian cube of a word of length 42. On the other hand, many morphisms satisfy both the conditions of Lemma 3.1 for $k = 6$ and those of Proposition 3.2 for $k = 5$.

Theorem 3.3. *There exist infinite binary 6-abelian cube-free words.*

Proof. Let $\rho : \Sigma_4^* \to \Sigma_2^*$ be the 21-uniform morphism defined by

$$\begin{aligned} a &\mapsto babaab\ a\ abbaba\ a\ baabba\ b,\\ b &\mapsto babaab\ a\ abbaba\ b\ baabba\ b,\\ c &\mapsto baabba\ a\ babaab\ a\ abbaba\ b,\\ d &\mapsto baabba\ a\ babaab\ b\ abbaba\ b. \end{aligned}$$

In fact, the first two images are obtained by the concatenation of $\xi(a)$ with $\xi(b)$ and the replacement of the holes with a, a, and b, and, respectively, a, b, and b, while the last two are obtained by replacing the holes with a, a, and b, and, respectively, a, b, and b, within the image of $\xi(ba)$.

It is easy to see that the image of any word under ρ does not contain a cube of some word of length less than 5. Furthermore, using a computer it can be checked that ρ satisfies the conditions of Lemma 3.1 for $k = 6$. The conditions of Proposition 3.2 are satisfied for $k = 5$ whenever $l = 10$ and $t_1 = aabaa, t_2 = babaa, t_3 = babab, t_4 = aabab$. We conclude that $\rho(w)$ is 6-abelian cube-free for all abelian square-free words w. □

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press

4. 5-ABELIAN CUBE-FREENESS

Let us now consider the 21-uniform morphism $\delta : \Sigma_3^* \to \Sigma_2^*$ defined by

$$\begin{aligned} a &\mapsto babaab\ a\ abbaba\ b\ baabba\ a,\\ b &\mapsto baabba\ a\ babaab\ b\ abbaba\ a,\\ c &\mapsto baabba\ a\ babaab\ b\ abbaba\ b. \end{aligned}$$

Observe that for δ the image of a has been obtained from $\xi(ab)$, while the images of b and c were both obtained from $\xi(ba)$, by replacement of the holes. It is not a difficult exercise to see that for $k = 5$ the conditions of Lemma 3.1 are met by δ. The following remark summarises the first two conditions of Proposition 3.2.

Remark 4.1. Let $d, e \in \Sigma_3$ and $i \in \{0, \ldots, 17\}$. Whenever $\mathrm{rfact}_4^i(\delta(d)) = \mathrm{rfact}_4^i(\delta(e))$, the two images must share the same prefix of length i, that is, $\mathrm{pref}_i(\delta(d)) = \mathrm{pref}_i(\delta(e))$.

The next result provides us a setting in which 5-abelian cube-free words exist.

Theorem 4.2. *Let $w \in \Sigma_3^\omega$ be an abelian cube-free word that does not contain a factor of the form $apbqbrc$ for any abelian equivalent words p, q, r. Then the infinite binary word $\delta(w)$ is 5-abelian cube-free.*

Proof. It is easy to see that the image of any word under δ does not contain a cube of a word of length less than 4. Furthermore, for some abelian cube-free word w, Lemma 3.1 provides a proof of the fact that $\delta(w)$ cannot contain 5-abelian cubes of words whose lengths are not divisible by 21.

We assume toward a contradiction that such a 5-abelian cube exists and take the factors $u_j = \mathrm{rfact}_m^{i+(j-1)m}(\delta(w))$ to be 5-abelian equivalent for $j \in \{1, 2, 3\}$, where $m = 21m'$ for some non-empty integers m and m'. Let $i' \equiv i \mod 21$ be such that $i' \in \{0, 1, \ldots, 20\}$. There are two cases that need consideration.

If we take $i' \leq 17$ for $u', u'', u_1, u_2, u_3 \in \Sigma_2^*$, $v_1, v_2, v_3 \in \Sigma_3^{m'}$ we have

$$u'u_1u_2u_3 = \delta(v_1v_2v_3)u'',$$

where $|u'| = |u''| = i'$ and $v_1v_2v_3$ is a factor of w. Furthermore,

$$\begin{aligned} \delta(v_1) &= u'\mathrm{pref}_{m-i'}(u_1),\\ \delta(v_2) &= \mathrm{suff}_{i'}(u_1)\mathrm{pref}_{m-i'}(u_2),\\ \delta(v_3) &= \mathrm{suff}_{i'}(u_2)\mathrm{pref}_{m-i'}(u_3), \end{aligned} \tag{4.1}$$

and, thus for all $t \in \Sigma_2^5$,

$$\begin{aligned} |\delta(v_1)|_t &= |u'\mathrm{pref}_4(u_1)|_t + |u_1|_t - |\mathrm{suff}_{i'+4}(u_1)|_t,\\ |\delta(v_2)|_t &= |\mathrm{suff}_{i'}(u_1)\mathrm{pref}_4(u_2)|_t + |u_2|_t - |\mathrm{suff}_{i'+4}(u_2)|_t,\\ |\delta(v_3)|_t &= |\mathrm{suff}_{i'}(u_2)\mathrm{pref}_4(u_3)|_t + |u_3|_t - |\mathrm{suff}_{i'+4}(u_3)|_t. \end{aligned}$$

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press

Because $\mathrm{pref}_4(u_1) = \mathrm{pref}_4(u_2) = \mathrm{pref}_4(u_3)$, it follows from (4.1) and Remark 4.1 that $u' = \mathrm{suff}_{i'}(u_1) = \mathrm{suff}_{i'}(u_2)$. Therefore

$$|u'\mathrm{pref}_4(u_1)|_t = |\mathrm{suff}_{i'}(u_1)\mathrm{pref}_4(u_2)|_t = |\mathrm{suff}_{i'}(u_2)\mathrm{pref}_4(u_3)|_t.$$

Because u_1, u_2 and u_3 are 5-abelian equivalent, $|u_1|_t = |u_2|_t = |u_3|_t$. However, the words *babab* and *bbabb* do not appear as factors in $\mathrm{suff}_4(\delta(d))\mathrm{pref}_4(\delta(e))$ for any letters $d, e \in \Sigma_3$. Therefore $|\mathrm{suff}_{i'+4}(u_j)|_t = |\mathrm{suff}_{i'}(u_j)|_t$ for every $t \in \{babab, bbabb\}$ and $j \in \{1,2,3\}$, and these three numbers are equal following Remark 4.1 and the fact that $\mathrm{suff}_4(u_1) = \mathrm{suff}_4(u_2) = \mathrm{suff}_4(u_3)$.

So far we have shown that for $t \in \{babab, bbabb\}$ we have

$$|\delta(v_1)|_t = |\delta(v_2)|_t = |\delta(v_3)|_t. \tag{4.2}$$

Let M be the invertible 3×3 matrix

$$\begin{pmatrix} |\delta(a)|_{babab} & |\delta(b)|_{babab} & |\delta(c)|_{babab} \\ |\delta(a)|_{bbabb} & |\delta(b)|_{bbabb} & |\delta(c)|_{bbabb} \\ |\delta(a)| & |\delta(b)| & |\delta(c)| \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 21 & 21 & 21 \end{pmatrix}.$$

For any $v \in \Sigma_3^*$, if $P_v = (|v|_a, |v|_b, |v|_c)$ is the Parikh vector of v, since *babab* and *bbabb* are not factors of $\mathrm{suff}_4(\delta(d))\mathrm{pref}_4(\delta(e))$ for any $d, e \in \Sigma_3$, we have

$$MP_v^T = (|\delta(v)|_{babab}, |\delta(v)|_{bbabb}, |\delta(v)|)^T. \tag{4.3}$$

Putting together the equations (4.3) and (4.2), we get that $MP_{v_1}^T = MP_{v_2}^T = MP_{v_3}^T$. Moreover, since M is invertible we have $P_{v_1} = P_{v_2} = P_{v_3}$. Thus, we conclude that w contains the abelian cube $v_1v_2v_3$, which is a contradiction.

Next we consider the case when $i' > 17$. Let $i'' = 21 - i' \in \{1,2,3\}$. Now

$$u_1u_2u_3u' = u''\delta(v_1a_1v_2a_2v_3a_3),$$

where $u' = \mathrm{suff}_{i''}(\delta(a_3))$ and $u'' = \mathrm{suff}_{i''}(\delta(a_0))$ for $a_0, a_1, a_2, a_3 \in \Sigma_3$, $u_1, u_2, u_3 \in \Sigma_2^*$, $v_1, v_2, v_3 \in \Sigma_3^{m'-1}$ and $a_0v_1a_1v_2a_2v_3a_3$ is a factor of w.

Note that $\delta(a_1)$ ends with $\mathrm{suff}_4(u_1)\mathrm{pref}_{i''}(u_2)$, while $\delta(a_2)$ has as a suffix $\mathrm{suff}_4(u_2)\mathrm{pref}_{i''}(u_3)$, which is the same word. Following Remark 4.1 we have $\delta(a_1) = \delta(a_2)$ and denoting $d = a_1 = a_2$ we get from the same remark that

- whenever $d = a$, we have $a_3 = a$;
- whenever $d = b$, we have $a_0 \in \{a, b\}$ and $a_3 \in \{b, c\}$;
- whenever $d = c$, we have $a_0 = c$.

Let us first analyse the case when $u' = \mathrm{suff}_{i''}(\delta(d))$. We have that

$$u_1u_2u_3u' = \mathrm{suff}_{i''}(\delta(a_0))\delta(v_1dv_2dv_3d).$$

Consequently, we get that

$$\begin{aligned} \delta(v_1d) &= \mathrm{suff}_{m-i''}(u_1)\mathrm{pref}_{i''}(u_2), \\ \delta(v_2d) &= \mathrm{suff}_{m-i''}(u_2)\mathrm{pref}_{i''}(u_3), \\ \delta(v_3d) &= \mathrm{suff}_{m-i''}(u_3)u', \end{aligned}$$

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press

and thus for every $t \in \Sigma_2^5$

$$|\delta(v_1)d|_t = |\mathrm{suff}_4(u_1)\mathrm{pref}_{i''}(u_2)|_t + |u_1|_t - |\mathrm{pref}_{i''+4}(u_1)|_t,$$
$$|\delta(v_2)d|_t = |\mathrm{suff}_4(u_2)\mathrm{pref}_{i''}(u_3)|_t + |u_2|_t - |\mathrm{pref}_{i''+4}(u_2)|_t,$$
$$|\delta(v_3)d|_t = |\mathrm{suff}_4(u_3)u'|_t + |u_3|_t - |\mathrm{pref}_{i''+4}(u_3)|_t.$$

Furthermore, we have $\mathrm{pref}_{i''}(u_2) = \mathrm{pref}_{i''}(u_3) = u'$ and, therefore,

$$|\mathrm{suff}_4(u_1)\mathrm{pref}_{i''}(u_2)|_t = |\mathrm{suff}_4(u_2)\mathrm{pref}_{i''}(u_3)|_t = |\mathrm{suff}_4(u_3)u'|_t.$$

Because u_1, u_2 and u_3 are 5-abelian equivalent, $|u_1|_t = |u_2|_t = |u_3|_t$. However, the words *babab* and *bbabb* do not appear as factors in $\mathrm{suff}_4(\delta(d))\mathrm{pref}_4(\delta(e))$ for any letters $d, e \in \Sigma_3$. Therefore if $t \in \{babab, bbabb\}$, then for every $j \in \{1, 2, 3\}$ we have $|\mathrm{pref}_{i''+4}(u_j)|_t = 0$. It follows that for all $t \in \{babab, bbabb\}$ we have

$$|\delta(v_1d)|_t = |\delta(v_2d)|_t = |\delta(v_3d)|_t.$$

As above, one can show using the matrix M that $v_1dv_2dv_3d$ is an abelian cube.

Now, the word $a_0v_1dv_2dv_3a_3$ is a factor of w. If $d = a$, then $a_3 = a$ and $v_1av_2av_3a$ is an abelian cube. If $d = c$, then $a_0 = c$ and $cv_1cv_2cv_3$ is an abelian cube. If $d = b$, then $a_0 \in \{a, b\}$ and $a_3 \in \{b, c\}$ and $v_1bv_2bv_3a_3$ is an abelian cube, $a_0v_1bv_bdv_3$ is an abelian cube, or $a_0v_1dv_2dv_3a_3 = av_1bv_2bv_3c$ has the word form present in the statement of our theorem. □

To see that the abelian cube-freeness of w is not enough and that further assumptions similar to those presented in Theorem 4.2 are needed in order to prove that $\delta(w)$ avoids 5-abelian cubes, consider the word *abbc* and its image (note that here p, q, and r are all the empty word):

$$\delta(abbc) = babaabaabbbabbabbaabba(abaabbaababaabbabbaba)^3b.$$

Given the previous observations, we have the following result:

Theorem 4.3. *Over a binary alphabet there exist infinitely many 5-abelian cube-free words.*

Proof. Use the abelian cube-free word w defined in [4], together with the previously constructed morphism δ. Theorems 2.3 and 4.2 settle the question. □

References

[1] F. Blanchet-Sadri, J.I. Kim, R. Mercaş, W. Severa, S. Simmons and D. Xu, Avoiding abelian squares in partial words. *J. Combin. Theory Ser. A* **119** (2012) 257–270.

[2] F. Blanchet-Sadri, R. Mercaş and G. Scott, A generalization of Thue freeness for partial words. *Theoret. Comput. Sci.* **410** (2009) 793–800.

[3] F. Blanchet-Sadri, K. Black and A. Zemke, Unary pattern avoidance in partial words dense with holes. In *Proc. of Language and Automata Theory and Applications – 5th International Conference, LATA 2011, Tarragona, Spain, May 26-31, 2011.* Edited by A.H. Dediu, S. Inenaga and C. Martín-Vide. Vol. 6638 of *Lect. Notes Comput. Science.* Springer (2011) 155–166.

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press

[4] F.M. Dekking, Strongly nonrepetitive sequences and progression-free sets. *J. Combin. Theory Ser. A* **27** (1979) 181–185.
[5] P. Erdős, Some unsolved problems. *Magyar Tudományos Akadémia Matematikai Kutató Intézete* **6** (1961) 221–254.
[6] M. Huova, Existence of an infinite ternary 64-abelian square-free word. *RAIRO: ITA* **48** (2014) 307–314.
[7] M. Huova and J. Karhumäki, On the unavoidability of k-abelian squares in pure morphic words. *J. Integer Seq.* **16** (2013).
[8] M. Huova, J. Karhumäki and A. Saarela, Problems in between words and abelian words: k-abelian avoidability. *Theoret. Comput. Sci.* **454** (2012) 172–177.
[9] M. Huova, J. Karhumäki, A. Saarela and K. Saari, Local squares, periodicity and finite automata. In *Rainbow of Computer Science*, edited by C. Calude, G. Rozenberg and A. Salomaa. Springer (2011) 90–101.
[10] J. Karhumäki, S. Puzynina and A. Saarela, Fine and Wilf's theorem for k-abelian periods. *Internat. J. Found. Comput. Sci.* **24** (2013) 1135–1152.
[11] V. Keränen, Abelian squares are avoidable on 4 letters. In *Proc. of the 19th International Colloquium on Automata, Languages and Programming* (1992) 41–52.
[12] F. Manea and R. Mercaş, Freeness of partial words. *Theoret. Comput. Sci.* **389** (2007) 265–277.
[13] R. Mercaş and A. Saarela, 3-abelian cubes are avoidable on binary alphabets. In *Proc. of the 17th International Conference on Developments in Language Theory*, edited by M.-P. Béal and O. Carton, vol. 7907 of *Lect. Notes Comput. Science*. Springer (2013) 374–383.
[14] M. Rao, On some generalizations of abelian power avoidability. preprint (2013).
[15] A. Thue, Über unendliche Zeichenreihen. *Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. Christiania* **7** (1906) 1–22. (Reprinted in *Selected Mathematical Papers of A. Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, Norway (1977) 139–158).
[16] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. Christiania* **1** (1912) 1–67. (Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, Norway (1977) 413–478).

Communicated by Ch. Choffrut.
Received March 14, 2014. Accepted March 19, 2014.

https://doi.org/10.1051/ita/2014020 Published online by Cambridge University Press