5-Abelian Cubes Are Avoidable on Binary Alphabets

Robert Mercas*and Aleksi Saarela†

June 2012

Abstract

A k-abelian cube is a word uvw, where u, v, w are either equal or have the same factors of length k with the same multiplicities and the same prefixes and suffixes of length $k-1$. Previously it has been known that k-abelian cubes are avoidable over a binary alphabet for $k \geq 8$. Here it is proved that this holds for $k \geq 5$.

1 Introduction

The concept of repetition-freeness has been investigated in Combinatorics on Words since the beginning of the area, see [5, 6]. 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 a factor, while with two 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 word is called square (overlap)-free if it does not contain any square (overlap) as a factor.

In 1961 Erdös 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. A word is abelian square-free, if it does not contain any abelian square as a factor. Dekking [2] proved that over a binary alphabet exist words that are abelian 4-free, while over a ternary alphabet abelian cubes are avoidable. The problem of whether abelian squares are avoided over a four-letter alphabet was open for a long time. In [4], using a 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. For the abelian case an alphabet with as low as 5 letters is enough in order to construct an infinite word with factors that do not match any abelian square word of length greater than two.

*Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik, PSF 4120, D-39016 Magdeburg, Germany; Supported by the Alexander von Humboldt Foundation

†Corresponding author, Department of Mathematics and Statistics, University of Turku, FI-20014 Turku, Finland; Supported by the Academy of Finland under grant 137991 (FiDiPro)

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, each sharing the same prefixes and suffixes of length $k-1$ as well as the same number of each factor of length k.

In [3], 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 is open, the minimal alphabet size being either three or four. For cubes a binary alphabet suffices whenever $k \geq 8$. However, the same work conjectures that for avoiding k-abelian cubes a binary alphabet might be enough even for a k as low as 2, since computer generated words of length 100000 having the property have been found.

This work intends to close on the gap of k-abelian cube-freeness. With the help of some morphisms created for avoidability of squares in the context of partial words, we improve the bound of k to 6 whenever we make use of some abelian square-free fixed point and, even more, make $k = 5$ when the base word comes from a self generating abelian cube-free ternary word.

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

2 Preliminaries

We denote by Σ a finite set of symbols called *alphabet*. A *word* w is 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. For a factorization $w = uxv$, we say that x is a *factor* of w, and whenever u is empty x is a *prefix* and, respectively, when v is empty x is a *suffix* of w. If $w = a_0 \dots a_{n-1}$ with $a_i \in \Sigma$ for $0 \leq i < n$, we say that $\text{rfact}_k^i(w) = a_i \dots a_{i+k-1}$ is the *right factor at* i of w of length k and $\text{lfact}_k^i(w) = a_{i-k} \dots a_{i-1}$ is the *left factor at* i of w of length k. Moreover, $\text{pref}_k(w) = \text{rfact}_k^0(w)$ is the length k *prefix* of w and $\text{suff}_k(w) = \text{lfact}_k^n(w)$ 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 $w = u^\ell$. Second powers are called *squares* and third powers *cubes*.

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 $t \in \Sigma^k$, $\text{pref}_{k-1}(u) = \text{pref}_{k-1}(v)$ and $\text{suff}_{k-1}(u) = \text{suff}_{k-1}(v)$. 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. Clearly, 1-abelian equivalence is the same as abelian equivalence.

A *k-abelian nth power* is a word $u_1 \dots u_n$, where the words $u_1, \dots, u_n$ are 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.

One of the initial results regarding repetition-freeness is due to Axel Thue [6]:

Theorem 2.1. *The Thue-Morse word given by the fixed point of the iterative morphism t with $t(a) = ab$ and $t(b) = ba$ is overlap-free.*

In [1], in the context of partial words, the morphism $\xi : \{a, b\}^* \to \{a, b, \diamond\}^*$ is 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 factor of length seven contains a hole symbol, denoted by $\diamond$.

Theorem 2.2. *[1] 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 cannot be avoided over a ternary alphabet. However, Dekking showed in [2] that two letters are sufficient for avoiding abelian fourth powers, and three letters are sufficient for avoiding abelian cubes. We need the following extension of the latter result.

Theorem 2.3. *Let w be the fixed point of morphism $\sigma : \Sigma^* \to \Sigma^*$ with $\Sigma = \{a, b, c\}$ defined by*

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

Then w is abelian cube-free and contains no factor $apbqbrc$ for abelian equivalent words p, q, r.

Proof. The word w was shown to be abelian cube-free in [2]. Similar ideas can be used to show that w avoids the factors $apbqbrc$. Let $f : \Sigma^* \to \mathbb{Z}_7$ be the morphism defined by $f(a) = 1, f(b) = 2, f(c) = 3$ (here $\mathbb{Z}_7$ is the additive group of integers modulo 7). Then $f(\sigma(d)) = 0$ for all $d \in \Sigma$. If $apbqbrc$ is a factor of w, then there are u, s, s' such that $\sigma(u) = sapbqbrcs'$. Consider the values

$$f(s), f(sa), f(sap), f(sapb), f(sapbq), f(sapbqb), f(sapbqbr), f(sapbqbrc). \tag{1}$$

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

$$i, i + 1, i + j + 1, i + j + 3, i + 2j + 3, i + 2j + 5, i + 3j + 5, i + 3j + 1.$$

For all values of i and j, one of these is not 0, 1, 2 or 4. This is a contradiction. □

Abelian squares were proven in 1992 to be avoidable over a four letter alphabet [4].

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.*

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 [3] it was proved that this holds for $k \geq 8$, and conjectured that it holds for $k \geq 2$. Here we prove that this holds for $k \geq 5$. The cases $k \in \{2, 3, 4\}$ remain open.

3 An improvement for k-abelian cube-freeness

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

Lemma 3.1. *Let $w \in \Sigma_p^\omega$ and $(q - 1)n \geq 2k - 3$. Let $\theta : \Sigma_p^* \to \Sigma_2^*$ be an n-uniform morphism. If there exist no $v_0, v_1, v_2, v_3 \in \Sigma_p^q$ and $i_0, i_1, i_2, i_3 \in \{k - 1, \dots, n + k - 2\}$ such that*

$$\begin{gathered} i_1 - i_0 \equiv i_2 - i_1 \equiv i_3 - i_2 \not\equiv 0 \mod n \\ \mathrm{rfact}^{i_0}_{k-1}(\theta(v_0)) = \mathrm{rfact}^{i_1}_{k-1}(\theta(v_1)) = \mathrm{rfact}^{i_2}_{k-1}(\theta(v_2)) \\ \mathrm{lfact}^{i_1}_{k-1}(\theta(v_1)) = \mathrm{lfact}^{i_2}_{k-1}(\theta(v_2)) = \mathrm{lfact}^{i_3}_{k-1}(\theta(v_3)), \end{gathered}$$

then $\theta(w)$ contains no k-abelian cube of a word with length at least $k - 1$ and not divisible by n.

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 all conditions are satisfied.

Let $\text{rfact}^i_{3m}(\theta(w))$ be a k-abelian cube and $i_j \in \{k-1, \dots, n+k-2\}$ be such that $i_j \equiv i + jm \mod n$. It follows that the numbers i_0, i_1, i_2, i_3 satisfy the first condition. Furthermore, we have

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

for some $v_0, v_1, v_2, v_3 \in \Sigma_p^q$. Last two conditions follow from the previous observations and the fact that $\text{rfact}^i_m(\theta(w))$, $\text{rfact}^{i+m}_m(\theta(w))$ and $\text{rfact}^{i+2m}_m(\theta(w))$ are all 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 $l \in \{k-1, \dots, n-k+1\}$ and p words $t_1, \dots, t_p \in \Sigma_2^k$ such that for all $a', b' \in \Sigma_p$ the following conditions hold:*

1. *If $\text{lfact}^j_{k-1}(\theta(a')) = \text{lfact}^j_{k-1}(\theta(b'))$, then $\text{pref}_j(\theta(a')) = \text{pref}_j(\theta(b'))$ for all $j \in \{k-1, \dots, l\}$,*

2. *If $\text{rfact}^j_{k-1}(\theta(a')) = \text{rfact}^j_{k-1}(\theta(b'))$, then $\text{suff}_{n-j}(\theta(a')) = \text{suff}_{n-j}(\theta(b'))$ for all $j \in \{l+1, \dots, n-k+1\}$,*

3. *The words $t_1, \dots, t_p$ are not factors of $\text{suff}_{k-1}(\theta(a'))\text{pref}_{k-1}(\theta(b'))$,*

4. *The p vectors $(|\theta(a_1)|_{t_j}, |\theta(a_2)|_{t_j}, \dots |\theta(a_p)|_{t_j})$ are linearly independent for all $j \leq p$.*

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

Proof. Assume towards a contradiction that the words $u_j = \text{rfact}^{i+jm}_m(\theta(w)) \in \Sigma_2^*$ are k-abelian equivalent for $j \in \{0,1,2\}$ and, also, $m = m'n$. Let $i' \equiv i \mod n$ be such that $i' \in \{0, \dots, n-1\}$. Let $i' \leq l$ (the case $i' > l$ is similar). Now for $u', u'', u_1, u_2, u_3 \in \Sigma_2^*$ and $v_0, v_1, v_2 \in \Sigma_p^{m'}$

$$u' u_0 u_1 u_2 = \theta(v_0 v_1 v_2) u'', \tag{2}$$

where $|u'| = |u''| = i'$ and the word $v_1 v_2 v_3$ is a factor of w. By (2), it follows that $\theta(v_1) = \text{suff}_{i'}(u_0)\text{pref}_{m-i'}(u_1), \theta(v_2) = \text{suff}_{i'}(u_1)\text{pref}_{m-i'}(u_2)$, and thus, for every $t \in \Sigma_2^k$ we have

$$\begin{aligned} |\theta(v_1)|_t &= |\text{suff}_{i'}(u_0)\text{pref}_{k-1}(u_1)|_t + |u_1|_t - |\text{suff}_{i'+k-1}(u_1)|_t \qquad \text{and} \\ |\theta(v_2)|_t &= |\text{suff}_{i'}(u_1)\text{pref}_{k-1}(u_2)|_t + |u_2|_t - |\text{suff}_{i'+k-1}(u_2)|_t. \end{aligned}$$

The words $\text{suff}_{i'}(u_0)$, $\text{suff}_{i'}(u_1)$ and $\text{suff}_{i'}(u_2)$ are prefixes of some words of $\theta(\Sigma_p)$. 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, it follows that $\text{suff}_{i'}(u_0) = \text{suff}_{i'}(u_1) = \text{suff}_{i'}(u_2)$. But also $\text{pref}_{k-1}(u_1) = \text{pref}_{k-1}(u_2)$, hence $|\text{suff}_{i'}(u_0)\text{pref}_{k-1}(u_1)|_t = |\text{suff}_{i'}(u_1)\text{pref}_{k-1}(u_2)|_t$. 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 $t \in \{t_1, t_2, \dots, t_p\}$, then $|\text{suff}_{i'+k-1}(u_1)|_t = |\text{suff}_{i'}(u_1)|_t = |\text{suff}_{i'}(u_2)|_t = |\text{suff}_{i'+k-1}(u_2)|_t$.

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

$$|\theta(v_1)|_{t_i} = |\theta(v_2)|_{t_i}. \tag{3}$$

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 $MP_v^T = (|\theta(v)|_{t_1}, |\theta(v)|_{t_2}, \dots |\theta(v)|_{t_p})^T$. From this and Equation (3), we have that $MP_{v_1}^T = MP_{v_2}^T$. Because M is invertible, it follows that $P_{v_1} = P_{v_2}$. Thus w contains an abelian square $v_1 v_2$, which is a contradiction. □

We shall try to junction different images of ξ where the $\diamond$ is replaced by letters of Σ_2 such that we find words that are k-abelian cube-free for as small k as possible. This gives many morphisms that 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

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

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. $\square$

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

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

The image of a has been obtained from $\xi(ab)$, while the images of b and c from $\xi(ba)$ by replacing the holes. The following technical lemma is obtained from the definition of δ.

Lemma 3.4. *Let $d, e \in \Sigma_3$ and $i \in \{0, \ldots, 17\}$. If $\mathrm{rfact}_4^i(\delta(d)) = \mathrm{rfact}_4^i(\delta(e))$, then the two images share the same prefix of length i, that is, $\mathrm{pref}_i(\delta(d)) = \mathrm{pref}_i(\delta(e))$.*

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

Theorem 3.5. *Let $w \in \Sigma_3^\omega$ be an abelian cube-free word that does not contain a factor $apbqbrc$ for any abelian equivalent words p, q, r. 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, 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 $u_j = \mathrm{rfact}_m^{i+(j-1)m}(\delta(w))$ are 5-abelian equivalent for $j \in \{1, 2, 3\}$, where $m = 21m'$ for some non-empty integers m, m'. Let $i' \equiv i \mod 21$ be such that $i' \in \{0, \ldots, 20\}$. 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, $\delta(v_1) = u'\mathrm{pref}_{m-i'}(u_1), \delta(v_2) = \mathrm{suff}_{i'}(u_1)\mathrm{pref}_{m-i'}(u_2), \delta(v_3) = \mathrm{suff}_{i'}(u_2)\mathrm{pref}_{m-i'}(u_3)$, and, thus for all $t \in \Sigma_2^5$,

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

Because $\mathrm{pref}_4(u_1) = \mathrm{pref}_4(u_2) = \mathrm{pref}_4(u_3)$, it follows from the description of the images of the v's and Lemma 3.4 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, we have $|u_1|_t = |u_2|_t = |u_3|_t$. The words $babab$ and $bbabb$ do not appear as factors in $\text{suff}_4(\delta(d))\text{pref}_4(\delta(e))$ for any letters $d, e \in \Sigma_3$, so $|\text{suff}_{i'+4}(u_j)|_t = |\text{suff}_{i'}(u_j)|_t$ for every $t \in \{babab, bbabb\}$ and $j \in \{1, 2, 3\}$, and these three numbers are equal by $\text{suff}_4(u_1) = \text{suff}_4(u_2) = \text{suff}_4(u_3)$ and Lemma 3.4.

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

$$|\delta(v_1)|_t = |\delta(v_2)|_t = |\delta(v_3)|_t. \tag{4}$$

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 $\text{suff}_4(\delta(d))\text{pref}_4(\delta(e))$ for any $d, e \in \Sigma_3$, it follows that

$$MP_v^T = (|\delta(v)|_{babab}, |\delta(v)|_{bbabb}, |\delta(v)|)^T. \tag{5}$$

Following (5) and (4), $MP_{v_1}^T = MP_{v_2}^T = MP_{v_3}^T$. Moreover, since M is invertible we have $P_{v_1} = P_{v_2} = P_{v_3}$. Thus, we conclude that w contains an abelian cube $v_1v_2v_3$, which is a contradiction.

For the case when $i' > 17$, let $i'' = 21 - i' \in \{1, 2, 3\}$. For $u_1, u_2, u_3 \in \Sigma_2^*$ and $v_1, v_2, v_3 \in \Sigma_3^{m'-1}$,

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

where $u' = \text{suff}_{i''}(\delta(a_3))$, $a_0, a_1, a_2, a_3 \in \Sigma_3$ and the word $a_0v_1a_1v_2a_2v_3a_3$ is a factor of w.

Note that $\delta(a_1)$ ends in $\text{suff}_4(u_1)\text{pref}_{i''}(u_2)$, while $\delta(a_2)$ ends in $\text{suff}_4(u_2)\text{pref}_{i''}(u_3)$, which is the same word. From Lemma 3.4 it follows that $\delta(a_1) = \delta(a_2)$, denoted $d = a_1 = a_2$, and:
– 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 $a_0 = c$.

Consider the case when $u' = \text{suff}_{i''}(\delta(d))$. Thus $u_1u_2u_3u' = \text{suff}_{i''}(\delta(a_0))\delta(v_1dv_2dv_3d)$. We have $\delta(v_1d) = \text{suff}_{m-i''}(u_1)\text{pref}_{i''}(u_2)$, $\delta(v_2d) = \text{suff}_{m-i''}(u_2)\text{pref}_{i''}(u_3)$, $\delta(v_3d) = \text{suff}_{m-i''}(u_3)u'$, and

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

for every $t \in \Sigma_2^5$. Furthermore, $\text{pref}_{i''}(u_2) = \text{pref}_{i''}(u_3) = u'$ and, hence $|\text{suff}_4(u_1)\text{pref}_{i''}(u_2)|_t = |\text{suff}_4(u_2)\text{pref}_{i''}(u_3)|_t = |\text{suff}_4(u_3)u'|_t$. Because u_1, u_2 and u_3 are 5-abelian equivalent, $|u_1|_t = |u_2|_t = |u_3|_t$. The words $babab$ and $bbabb$ do not appear as factors in $\text{suff}_4(\delta(d))\text{pref}_4(\delta(e))$ for any letters $d, e \in \Sigma_3$, so if $t \in \{babab, bbabb\}$, then for $j \in \{1, 2, 3\}$ we have $|\text{pref}_{i''+4}(u_j)|_t = 0$. Furthermore for all $t \in \{babab, bbabb\}$ we have $|\delta(v_1d)|_t = |\delta(v_2d)|_t = |\delta(v_3d)|_t$. It can be shown with the help of 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 $v_1av_2av_3a$ is an abelian cube. If $d = c$, then $cv_1cv_2cv_3$ is an abelian cube. If $d = b$, then either $v_1bv_2bv_3b$ or $bv_1bv_2bv_3$ are abelian cubes, or $a_0v_1dv_2dv_3a_3 = av_1bv_2bv_3c$, is the word form in our Theorem's statement. □

To see that abelian cube-freeness is not enough and that the assumptions of Theorem 3.5 are needed to prove that 5-abelian cubes are avoidable over a binary alphabet, consider the word

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

Given the previous observations, Theorem 2.3 settles the question.

Theorem 3.6. *Over a binary alphabet there exist infinitely many 5-abelian cube-free words.*

References

[1] Francine Blanchet-Sadri, Kevin Black, and Andrew Zemke. Unary pattern avoidance in partial words dense with holes. In Adrian Horia Dediu, Shunsuke Inenaga, and Carlos Martín-Vide, editors, *LATA*, volume 6638 of *Lecture Notes in Computer Science*, pages 155–166. Springer, 2011.

[2] F. M. Dekking. Strongly nonrepetitive sequences and progression-free sets. *J. Combin. Theory Ser. A*, 27(2):181–185, 1979.

[3] Mari Huova, Juhani Karhumäki, and Aleksi Saarela. Problems in between words and abelian words: k-abelian avoidability. *Theoretical Computer Science*, 2012. in press.

[4] Veikko Keränen. Abelian squares are avoidable on 4 letters. In *Proceedings of the 19th International Colloquium on Automata, Languages and Programming*, pages 41–52, 1992.

[5] Axel Thue. Über unendliche Zeichenreihen. *Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. Christiana*, 7:1–22, 1906. (Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, Norway (1977), pp. 139–158).

[6] Axel Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. Christiana*, 1:1–67, 1912. (Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, Norway (1977), pp. 413–478).