Avoidability of long k-abelian repetitions

M. Rao*, M. Rosenfeld†

June 23, 2014

Abstract

We study the avoidability of long k-abelian-squares (resp. cubes) on ternary (resp. binary) alphabet. We show that one cannot avoid abelian cubes of period at least 2 in infinite binary words, answering negatively to one question from Mäkelä. Then we show that one can avoid 3-abelian-squares of period at least 3 in infinite binary words and 2-abelian-squares of period at least 2 in infinite ternary words.

1 Introduction

Avoidability of structures and patterns has been extensively studied in theoretical computer science since the work of Thue on avoidability of repetitions in words [1]. Thue showed that there are infinitely long ternary word without square (a factor of the form ww where w is a word) and infinite binary word without cube (a factor of the form www where w is a word).

The avoidability of abelian repetitions has been studied since a question from Erdős in 1957 [6, 5]. A factor uv is an abelian square if u is a permutation of the letters of v. One cannot avoid abelian squares on infinite ternary words. Erdős asked whether it is possible to avoid abelian squares on an infinite word over an alphabet of size 4 [6, 5]. After some steps (alphabet of size 25 by Evdokimov [7] and size 5 by Pleasant [11]) Keränen answered positively by giving a morphism whose fixed-point is abelian square free. Moreover Dekking showed that it is possible to avoid abelian cubes on a ternary alphabet and abelian 4th power on a binary alphabet [3].

Mäkelä asked the following two questions on the avoidability of long abelian cubes (resp. squares) in a binary (resp. ternary) alphabet:

Question 1 (Mäkelä (see [10])). *Can you avoid abelian cubes of the form uvw where $|u| \geq 2$, over two letters ? - You can do this at least for words of length 250.*

Question 2 (Mäkelä (see [10])). *Can you avoid abelian squares of the form uv where $|u| \geq 2$ over three letters ? - Computer experiments show that you can avoid these patterns at least in words of length 450.*

The notion of k-abelian repetition has been introduced by Karhumäki *et al.* as a generalization of both repetition and abelian repetition [9]. One can avoid 3-abelian-squares (resp. 2-abelian-squares) on ternary (resp. binary) words [12]. Following Mäkelä questions, one can ask whether it is possible to avoid long k-abelian-powers on binary (resp. ternary) words.

In Section 3, we answer negatively to Mäkelä's Question 1. In Section 4, we show that one can avoid 3-abelian-squares of period at least 3 in binary words and 2-abelian-squares of period at least 2 in ternary words.

*CNRS - LIP, ENS Lyon, France
†LIP, ENS Lyon, France

2 Preliminaries

Let A be a finite alphabet. For a word $u \in A^*$ and $a \in A$, we denote by $|u|_a = |\{i : u[i] = a\}|$ the number of occurrences of the letter a in u. For $w \in A^*$, we denote by $|u|_w = |\{i : u[i : i + |w| - 1] = w\}|$ the number of occurrences of the factor w in u.

Two word u and v are said to be *abelian equivalent*, denoted $u \approx_a v$, if for every $a \in \Sigma$, $|u|_a = |v|_a$. A word $u_1 u_2 ... u_p$ is an *abelian-n-power* if $u_1 \approx_a u_2 \approx_a ... \approx_a u_p$. An *abelian-square* (resp. *abelian-cube*) is an abelian 2-power (resp. abelian-3-power).

Two words u and v are said *k-abelian equivalent* (for $k \geq 1$), denoted $u \approx_{a,k} v$, if for every $w \in \Sigma^*$ such that $|w| \leq k$, $|u|_w = |v|_w$. A word $u_1 u_2 ... u_n$ is a *k-abelian-n-power* if $u_1 \approx_{a,k} u_2 \approx_{a,k} ... \approx_{a,k} u_n$. Its *period* is $|u_1|$. Similarly, a *k-abelian-square* (resp. *k-abelian cube*) is a k-abelian 2-power (resp. k-abelian 3-power). Note that the 1-abelian equivalence is exactly the abelian equivalence. A word is said to be *k-abelian-n-power-free* if none of its factor is a k-abelian-n-power.

The *Parikh vector* of a word $w \in A^*$ denoted $\Psi(w)$ is the vector indexed by A such that for every $a \in A$, $\Psi(w)[a] = |w|_a$. Then two words u and v are abelian-equivalent if $\Psi(u) = \Psi(v)$. For a set $S \subset A^*$ and a word $w \in A^*$, $\Psi_S(w)$ is the vector indexed by S such that for every $s \in S$, $\Psi_S(w)[s] = |w|_s$. We may write $\Psi_k(w)$ instead of $\Psi_{B^k}(w)$ for $w \in B^*$.

For all $u \in \Sigma^*$, $i \leq |u|$, let $\text{pref}_i(u)$ be the prefix of size i of u and $\text{suf}_i(u)$ be the suffix of size i of u. There are equivalent definitions of k-abelian equivalence (see [9]). For every two words u and v, the following conditions are equivalent:

- u and v are k-abelian equivalent (*i.e.* $u \approx_{a,k} v$),
- $\Psi_k(u) = \Psi_k(v)$ and $\text{pref}_{k-1}(u) = \text{pref}_{k-1}(v)$,
- $\Psi_k(u) = \Psi_k(v)$ and $\text{suf}_{k-1}(u) = \text{suf}_{k-1}(v)$.

3 Abelian cubes and Mäkelä's question

Dekking showed that it is possible to avoid abelian cubes in an infinite word over a ternary alphabet [3]. More recently Rao showed that one can avoid 2-abelian-cubes on a binary alphabet [12] and one can check that every word over a binary alphabet of length greater than 9 contains an abelian cube. When wondering about the avoidability of long k-abelian-cubes on infinite words, the only question left is the avoidability of long abelian-cubes on binary words. This is subject of the Question 1 from Mäkelä. He asked whether one can avoid every abelian cubes of period at least 2 in binary words. We answer negatively to this question.

For this, we used a property about Lyndon words that made the exhaustive search much faster. A word $w \in \Sigma^*$ is a *Lyndon word* if for all $u, v \in \Sigma^+$ such that $w = uv$, $w <_{lex} vu$, where $<_{lex}$ is the lexicographic order. The well known Chen-Fox-Lyndon Theorem states that every word may be written uniquely as a concatenation of non-increasing Lyndon words. In the following proof we refer to this decomposition as the Lyndon factorization. A language $L \subset \Sigma^*$ is a *factorial* if for every w in L, every factor of w is in S.

Lemma 1. *Any factorial language L with arbitrarily long words contains arbitrarily large power or arbitrarily long Lyndon words.*

Proof. Let assume that there is no arbitrarily long Lyndon words. This implies that there is a finite number n of Lyndon word in L and $s \in \mathbb{N}$ such that for every Lyndon word w in L,

$|w| \leq s$. Let $w_1, ..., w_n \in L^n$ be the Lyndon words of L ordered by decreasing lexicographic order.

Then using the Lyndon factorization they are for every $w \in L$ some Lyndon words $L_1 \geq_{lex} L_2 \geq_{lex} ... \geq_{lex} L_d$ such that $w = L_1 ... L_d$. The fact that our language is factorial give us that all the L_i are in L. We get that for every $w \in L$ there are $\alpha_1, ..., \alpha_n \in \mathbb{N}$ such that $w = w_1^{\alpha_1} ... w_n^{\alpha_n}$. Then $|w| \leq s \times \sum_i \alpha_i \leq s \times t \times \max_i(\alpha_i)$.

If $|w| \geq n \times s \times t$, at least one of the α_i is greater than n and $w_i^n \in L$. By assumption for any $n \in \mathbb{N}$ there is a $w \in L$ such that $|w| \geq n \times s \times t$ so we have arbitrarily long power in L. □

A set of word that avoid certain kind of abelian repetitions is a factorial language and does not contains arbitrarily large power. Thus we just need to check that there is no arbitrarily long Lyndon word in it to deduce that this set does not contain arbitrarily long words. The tree of the possible words to check in the exhaustive search is then shorter and it makes the computation much faster. The next claim answers negatively to Mäkelä's Question 1.

Claim 1. *There is no infinite word over a binary alphabet avoiding abelian cubes of period at least* 2*.*

We checked using a computer program that there is only finitely many Lyndon words over a binary alphabet avoiding abelian cubes of the form uvw such that $|u| \geq 2$. The program took approximately 3 hours to find all 2732711352 such Lyndon words. The longest word has length 290. Using Lemma 1 we deduced that there is no infinite binary word avoiding abelian cubes of size greater or equal to two. Then we can ask:

Question 3. *Is there* $p \in \mathbb{N}$ *such that one can avoid abelian cubes of period at least* p *over two letters ?*

For $p = 3$, we found a word of length 2500.

4 k-Abelian squares

A word is said *(p,k)-abelian-square-free* if it avoid k-abelian-squares of period at least p. A morphism h is said *(p,k)-abelian-square-free* if for every abelian-square-free word w, $h(w)$ is (p,k)-abelian-square-free.

It is easy to verify that one cannot avoid squares of period at least 2 on a binary alphabet. Entringer *et al.* showed that it is possible to construct a binary word with no square of period at least 3 [4]. They also showed that every infinite binary word contains arbitrarily long abelian-squares. Thus one can wonder about the avoidability of big k-abelian-squares. Rao asked the following question:

Question 4 (Rao [12])**.** *What is the smallest* k *(if any) such that arbitrarily long* k*-abelian-squares can be avoided over a binary alphabet ?*

In this section, we show that one can avoid long 3-abelian-squares over a binary alphabet, by giving a $(3,3)$-abelian-square-free morphism. Let h be the following morphism:

$$h : \begin{cases} 0 \to & 00001010110 \\ 1 \to & 00111010110 \\ 2 \to & 00011111010 \\ 3 \to & 00011001010. \end{cases}$$

Theorem 1. *The morphism* h *is (3,3)-abelian-square-free.*

Proof. The proof is based on the same idea than the one used by Rao to give sufficient conditions for a morphism to be k-abelian-free [12] which is a generalization of the sufficient conditions given by Carpi for abelian-free-morphisms [2]. In the proof we use the following property:

Proposition 1. *$\forall k, i \in \mathbb{N}$, $\forall u, v \in \Sigma^*$ such that $i < k$, $k - 1 - i \leq |u|$ and $i \leq |v|$:*

$$\Psi_k(uv) = \Psi_k(u \operatorname{pref}_i(v)) + \Psi_k(\operatorname{suf}_{k-1-i}(u)v). \tag{S}$$

Let $A = \{0, 1, 2, 3\}$ and w be an abelian-square-free word in A^*. Let show that $h(w)$ is (3,3)-abelian-square-free. We can check using a computer that $\forall a, b \in A^2$, $a \neq b$, $h(ab)$ is (3,3)-abelian-square-free. So if there is a 3-abelian square it has to be on the image of 3 at least letters. Then there are $a_1, a_2, a_3 \in A$, $x_1, x_2 \in A^*$, $u_1, u_2, v_2, v_3 \in \{0,1\}^*$ and $v_1, u_3 \in \{0,1\}^+$ such that:

- $a_1 x_1 a_2 x_2 a_3$ is a factor of w.
- for every $i \in \{1, 2, 3\}$, $u_i v_i = h(a_i)$
- $v_1 h(x_1) u_2 \approx_{a,3} v_2 h(x_2) u_3$

$|v_1 u_2 v_2 u_3| \geq 13$ so $|v_1 u_2| \geq 7$ or $|v_2 u_3| \geq 7$. And $|v_1 h(x_1) u_2| = |v_2 h(x_2) u_3|$ then $\forall i \in \{1, 2\}$, $|v_i h(x_i) u_{i+1}| \geq 7$. Then $\forall i \in \{1, 2\}$, $|v_i| \geq 3$, $|h(x_i)| \geq 3$ or $|u_{i+1}| \geq 3$.
If $|u_{i+1}| \geq 3$:

$$\begin{aligned}
\Psi_3(v_i h(x_i) u_{i+1}) &= \Psi_3(v_i 00) + \Psi_3(h(x_i) u_{i+1}) \quad \text{(using (S) and } \operatorname{pref}_2(h(x_i) u_{i+1}) = 00) \\
&= \Psi_3(v_i 00) + \Psi_3(h(x_i) 00) + \Psi_3(u_{i+1}) \quad \text{(using (S) and } \operatorname{pref}_2(u_{i+1}) = 00) \\
&= \Psi_3(v_i 00) + \Psi_3(h(x_i) 00) + \Psi_3(01 u_{i+1}) - \Psi_3(0100).
\end{aligned}$$

If $|v_{i+1}| \geq 3$ or $|h(x_i)| \geq 3$ we have the same result. So :

$$\Psi_3(v_i h(x_i) u_{i+1}) = \Psi_3(v_i 00) + \Psi_3(h(x_i) 00) + \Psi_3(01 u_{i+1}) - \Psi_3(0100).$$

Let N be the matrix indexed by $\{0,1\}^3 \times \{0,1,2,3\}$ with $N[w, x] = |h(x)00|_w$. Then :

$${}^tN = \begin{pmatrix} 3 & 1 & 2 & 1 & 1 & 2 & 1 & 0 \\ 1 & 1 & 1 & 2 & 1 & 2 & 2 & 1 \\ 2 & 1 & 1 & 1 & 1 & 1 & 1 & 3 \\ 2 & 2 & 2 & 1 & 2 & 1 & 1 & 0 \end{pmatrix}.$$

$$\Psi_3(v_i h(x_i) u_{i+1}) = \Psi_3(v_i 00) + N\Psi(x_i) + \Psi_3(01 u_{i+1}) - \Psi_3(0100).$$

We supposed $v_1 h(x_1) u_2 \approx_{a,3} v_2 h(x_2) u_3$, so:

$$\Psi_3(v_1 00) + N\Psi(x_1) + \Psi_3(01 u_2) = \Psi_3(v_2 00) + N\Psi(x_2) + \Psi_3(01 u_3)$$

Let M be the sub-matrix of N made of its rows 1, 2, 6 and 7 (they correspond to the words $000, 001, 101, 110$).

$$M = \begin{pmatrix} 3 & 1 & 2 & 2 \\ 1 & 1 & 1 & 2 \\ 2 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 \end{pmatrix}$$

Let $\Psi_S(w)$ be the sub-vector of $\Psi_3(w)$ made of the rows 1,2,6 and 7. We can check that M is non-singular. Then we can write:

$$M^{-1}(\Psi_S(v_1 00) + \Psi_S(01 u_2) - \Psi_S(v_2 00) - \Psi_S(01 u_3)) = \Psi(x_2) - \Psi(x_1).$$

Let $\Psi(v_1, u_2, v_2, u_3) = \Psi_S(v_1 00) + \Psi_S(01u_2) - \Psi_S(v_2 00) - \Psi_S(01u_3)$.
$\Psi(v_1, u_2, v_2, u_3)$ is in $Im(N)$ thus $M^{-1}(\Psi(v_1, u_2, v_2, u_3))$ is an integer vector. Moreover, $\text{pref}_2(v_1 00) = \text{pref}_2(v_2 00)$, $\text{suf}_2(10u_2) = \text{suf}_2(10u_3)$ and $|v_1 u_2| \equiv |v_2 u_3| \pmod{11}$.

Claim 2. *For all $a_1, a_2, a_3 \in A$ and $u_1, v_1, u_2, v_2, u_3, v_3 \in \{0,1\}^+$ such that:*

- $\forall i \in \{1, 2, 3\}$, $u_i v_i = h(a_i)$,
- $\text{pref}_2(v_1 00) = \text{pref}_2(v_2 00)$, $\text{suf}_2(10u_2) = \text{suf}_2(10u_3)$ *and* $|v_1 u_2| \equiv |v_2 u_3| \pmod{11}$,
- $\Psi(v_1, u_2, v_2, u_3)$ *is in* $Im(N)$, $M^{-1}(\Psi(v_1, u_2, v_2, u_3))$ *is an integer vector*

There are $(\alpha_1, \alpha_2, \alpha_3) \in \{0, 1\}$ *such that:*

$$M^{-1}(\Psi(v_1, u_2, v_2, u_3)) = \alpha_1 \Psi(a_1) - (2\alpha_2 - 1)\Psi(a_2) - (1 - \alpha_3)\Psi(a_3)$$

Proof. This claim can be verified using a computer program. There are 4^3 values for the a_i and 11^2 way of choosing the u_i, v_i for each of them which makes less than 7744 cases to check (most of them being eliminated by the prefix and suffix conditions). □

From the claim we have $(\alpha_1, \alpha_2, \alpha_3) \in \{0, 1\}$ such that:

$$M^{-1}(\Psi(v_1, u_2, v_2, u_3)) = \alpha_1 \Psi(a_1) - (2\alpha_2 - 1)\Psi(a_2) - (1 - \alpha_3)\Psi(a_3\{0, 1\}).$$

Now we can introduce x'_1, x'_2 such that : $\forall i \in \{1, 2\}$, $x'_i = a_i^{\alpha_i} x_i a_{i+1}^{1-\alpha_{i+1}}$. $x'_1 x'_2$ is a factor of w and

$$\Psi(x'_1) - \Psi(x'_2) = \Psi(x_1) - \Psi(x_2) - \alpha_1 \Psi(a_1) + (2\alpha_2 - 1)\Psi(a_2) + (1 - \alpha_3)\Psi(a_3) = 0.$$

Then there is an abelian-square on w and we have a contradiction. This implies that h is (3,3)-abelian-square-free. □

Theorem 1 gives a partial answer to Rao's Question 4. Moreover, since we know that there are exponentially many square-free words of a given length on four letters, we can deduce that there are exponentially many $(3, 3)$-abelian-square-free infinite words over a binary alphabet. We can then ask the following question:

Question 5. *Can you avoid 2-abelian-squares of period at least p for some $p \in \mathbb{N}$?*

Computer experiments show that you can avoid those patterns for $p = 3$ in words of length 15000.

Minimal number of 3-abelian-squares in infinite binary words. Fraenkel and Simpson showed that there is an infinite word containing only the squares 0^2, 1^2, $(01)^2$ [8]. A natural question is if we can extend this property to the 3-abelian case. Using a computer one can check that there is no infinite word with only 3 non-equivalent 3-abelian-squares (the longest is of size 70). Let:

$$h_2 : \begin{cases} 0 \to & 00000101011 \\ 1 \to & 00001101011 \\ 2 \to & 000111 \\ 3 \to & 001010011. \end{cases}$$

Theorem 2. *h_2 is $(3, 3)$-abelian-square-free. Moreover for every abelian-square-free word w, $h(w)$ contains only 5 different 3-abelian-squares: 0^2, 1^2, $(00)^2$ $(01)^2$ and $(10)^2$.*

The proof is similar to the proof of the Theorem 1, and is omitted. One can ask the following questions.

Question 6. *Is there an infinite binary word that contains only 4 different 3-abelian-squares?*

Question 7. *Is there an infinite binary word w and $k \in \mathbb{N}$ such that w contains only 3 different k-abelian-squares?*

2-abelian squares over a ternary alphabet. Rao showed that one can build an infinite word that avoid 3-abelian-squares over a ternary alphabet [12]. (The longest 2-abelian-square-free ternary word has size 537 [9].) Mäkelä (Question 2) asked whether we can avoid abelian-squares of period at least 2 in ternary words. We answer to a weaker version of this question, that is one can avoid (2,2)-abelian squares over three letters. (The proof is similar to the proof of the Theorem 1, and is omitted.) Let :

$$h_3 : \begin{cases} 0 \to & 00021 \\ 1 \to & 00111 \\ 2 \to & 01121 \\ 3 \to & 01221. \end{cases}$$

Theorem 3. *h_3 is $(2,2)$-abelian-square-free.*

5 Conclusion

(l,k)-abelian-squares on binary words

k \ l	1	2	3	≥ 4
1	3	10	18	Finite [4]
2	3	18	≥ 15000 Quest. 5	Quest. 5
3	3	18	∞ Th. 1	∞
∞	3	18	∞ [4]	∞

(l,k)-abelian-squares on ternary words

k \ l	1	2	≥ 3
1	7	≥ 4900 Quest. 2	Quest. 2
2	537 [9]	∞ Th. 3	∞
3	∞ [12]	∞	∞

(l,k)-abelian-cubes on binary words

k \ l	1	2	≥ 3
1	9	Finite Claim 1	Quest. 3
2	∞ [12]	∞	∞

Figure 1: The avoidability of long k-abelian-squares (resp. cubes) on ternary and binary words.

The three tables in Figure 1 recapitulate the results we know about the avoidability of big k-abelian-n-power. The value in each case is the value of the longest word avoiding the corresponding kind of repetition. "∞" means that we can avoid them on arbitrarily long words, and "Finite" means that we cannot but that we do not know the maximum length.

References

[1] A.Thue. Über die gegenseitige lage gleicher teile gewisser zeichenreihen. *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania,*, 10:1–67, 1912.

[2] A. Carpi. On abelian power-free morphisms. *International Journal of Algebra and Computation*, 03(02):151–167, 1993.

[3] F.M Dekking. Strongly non-repetitive sequences and progression-free sets. *Journal of Combinatorial Theory, Series A*, 27(2):181 – 185, 1979.

[4] R.C Entringer, D.E Jackson, and J.A Schatz. On nonrepetitive sequences. *Journal of Combinatorial Theory, Series A*, 16(2):159 – 164, 1974.

[5] P. Erdős. Some unsolved problems. *The Michigan Mathematical Journal*, 4(3):291–300, 1957.

[6] P. Erdős. Some unsolved problems. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 6:221–254, 1961.

[7] A. A. Evdokimov. Strongly asymmetric sequences generated by a finite number of symbols. *Dokl. Akad. Nauk SSSR*, 179:1268–1271, 1968.

[8] Aviezri S. Fraenkel and R. Jamie Simpson. How many squares must a binary sequence contain? *The Electronic Journal of Combinatorics*, 2, 1995.

[9] J. Karhumaki, A. Saarela, and L. Q. Zamboni. On a generalization of Abelian equivalence and complexity of infinite words. *Journal of Combinatorial Theory, Series A*, 120(8):2189 – 2206, 2013.

[10] V. Keränen. New abelian square-free DT0L-languages over 4 letters. *Manuscript*, 2003.

[11] P. A. B. Pleasants. Non-repetitive sequences. *Mathematical Proceedings of the Cambridge Philosophical Society*, 68:267–274, 9 1970.

[12] M. Rao. On some generalizations of abelian power avoidability. *Manuscript*, 2014.