Strongly k-Abelian Repetitions[*]

Mari Huova and Aleksi Saarela

Department of Mathematics and Statistics & TUCS
University of Turku, Finland
{mahuov,amsaar}@utu.fi

Abstract. We consider with a new point of view the notion of nth powers in connection with the k-abelian equivalence of words. For a fixed natural number k, words u and v are k-abelian equivalent if every factor of length at most k occurs in u as many times as in v. The usual abelian equivalence coincides with 1-abelian equivalence. Usually k-abelian squares are defined as words w for which there exist non-empty k-abelian equivalent words u and v such that $w = uv$. The new way to consider k-abelian nth powers is to say that a word is *strongly k-abelian nth power* if it is k-abelian equivalent to an nth power. We prove that strongly k-abelian nth powers are not avoidable on any alphabet for any numbers k and n. In the abelian case this is easy, but for $k > 1$ the proof is not trivial.

Keywords: k-abelian equivalence, nth powers, avoidability

1 Introduction

In combinatorics on words the theory of avoidability is one of the oldest and most studied topics. Axel Thue, who proved at the beginning of 20th century the existence of an infinite binary cube-free word and an infinite square-free ternary word, can be referred to as the initiator of this area[12, 13]. Corresponding avoidability questions for *abelian equality*, the commutative variant of equality where only the number of each letter counts and not their order, have been studied since late 1960s. Dekking [3] has proved that the optimal value for the size of the alphabet where abelian cubes are avoidable is three. The problem of the minimal size of the alphabet in which abelian squares can be avoided was an open question for a long time until the optimal value, four, was found by Keränen [8].

Lately, new variants of the avoidability problems have been introduced by defining repetitions via *k-abelian equivalence*, see e.q. [4]. This new equivalence relation, where $k \geq 1$ is a natural number, lies properly in between equality and abelian equality. The obvious modifications of the above Thue's problems ask for what are the smallest alphabets where k-abelian squares and cubes can be avoided. It is known that for $k \geq 3$ k-abelian cubes can be avoided over a binary alphabet [10]. In a case of square-freeness it is known that 2-abelian squares can not be avoided over a ternary alphabet but for large enough values of k

[*] Supported by the Academy of Finland under grant 257857

avoidability is achieved [4, 6]. In [5] it is shown that k-abelian square-free word cannot be obtained by iterating a single prefix preserving morphism.

In this note we consider abelian and k-abelian avoidability with a new perspective. We say that a word w is *a strongly abelian nth power* if it is abelian equivalent to a word which is a usual nth power, i.e., concatenation of n equivalent words. Now if an abelian equivalence class contains a word which is a usual nth power then all the words in this equivalence class are strongly abelian nth powers. So we consider the word more like a representative of its equivalence class than a single word. Corresponding notion of *a strongly k-abelian nth power* can be introduced similarly. We prove that every infinite word contains strongly k-abelian nth powers for all values of k and n.

2 Preliminaries

For the basic terminology of words as well as avoidability we refer to [9] and [2]. Here we define only our basic notions for this note.

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

1. $pref_{k-1}(u) = pref_{k-1}(v)$ *and* $suf_{k-1}(u) = suf_{k-1}(v)$, *and*
2. *for all $w \in \Sigma^k$, the number of occurrences of w in u and v coincide, i.e.* $|u|_w = |v|_w$.
3. *Different words of length at most k are not k-abelian equivalent.*

The k-abelian equivalence is like a sharpening of abelian equivalence and for the value $k = 1$ these define the same equivalence relation. For more about this notion, see [7]. In fact, k-abelian equivalence is a congruence of words, i.e. an equivalence relation R such that $uvRu'v'$ whenever uRu' and vRv'. We are interested in the products of words which are k-abelian equivalent but we will first define squares for all congruences R. Higher powers can be defined analogously.

If u, v are congruent words, then their product uv is an *R-square*. This definition has been used in the study of abelian and k-abelian repetition-freeness. In this article, however, we concentrate on another definition:

Definition 2. *A word w is a* strongly R-square *if it is congruent to a square of some non-empty word v, i.e. $wRvv$.*

For example, $aabb$ is not an abelian square because aa and bb are not abelian equivalent, but it is a strongly abelian square because it is abelian equivalent to $(ab)^2$.

Square-freeness in partially commutative monoids was studied by Carpi and De Luca in [1]. Their approach to square-freeness is similar but not identical to the one in this paper. Another interesting related concept is that of approximate squares, which can be defined as words of the form uv, where the Hamming distance of u and v is "small enough" (this definition is analogous to the definition of R-squares), or equivalently as words w such that the Hamming distance of w and some square is "small enough" (this definition is analogous to the definition of strongly R-squares). The avoidability of approximate squares has been studied by Ochem, Rampersad and Shallit [11].

Lemma 3. *A word is a strongly R-square if and only if it is congruent to an R-square.*

Proof. The "only if" direction is clear. If w is congruent to an R-square, say $wRuv$ and uRv, then $wRuu$, because uRv implies $uvRuu$ (here the assumption that R is not just an equivalence relation but a congruence is used). $\square$

It could be said that strongly R-squares take the concept of squares farther away from words and closer to the monoid defined by R.

Let us now state the definitions of strongly abelian and k-abelian nth powers for any $n \geq 1$.

Definition 4. *A word w is a strongly abelian nth power if it is abelian equivalent to a word which is an nth power.*

Definition 5. *A word w is a strongly k-abelian nth power if it is k-abelian equivalent to a word which is an nth power.*

The basic problem we are considering is avoidability of strongly abelian and strongly k-abelian nth powers. We prove that, for all k and n, they are unavoidable on all finite alphabets.

3 Unavoidability of Strongly Abelian and k-Abelian n-Powers

First we show that in abelian case it is easy to see that there does not exist infinite word which would avoid a strongly abelian nth power. Recall that two words are abelian equivalent if and only if they have the same Parikh vectors. Parikh vector p is a function from the set of words over m-letter alphabet $\{a_1, a_2, \ldots, a_m\}$ to the set of m-dimensional vectors over natural numbers, where $p(w) = (|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_m})$.

Theorem 6. *Let Σ be an alphabet and let $n \geq 2$. Every infinite word $w \in \Sigma^\omega$ contains a non-empty factor that is abelian equivalent to an nth power.*

Proof. A word is abelian equivalent to an nth power if and only if its Parikh vector is zero modulo n. The number of different Parikh vectors modulo n is finite, so w has two prefixes u and uv such that their Parikh vectors are the same modulo n. Then the Parikh vector of v is zero modulo n, so v is abelian equivalent to an nth power. $\square$

Theorem 6 can be generalized for k-abelian equivalence, but this is not trivial. One important difference between abelian and k-abelian equivalence is that if a vector with non-negative elements is given, then a word having that Parikh vector can be constructed, but if for every $t \in \Sigma^k$ a non-negative number n_t is given, then there need not exist a word u such that $|u|_t = n_t$ for all t (see Example 10).

Perhaps the biggest difficulty in generalizing Theorem 6 lies in finding an analogous version of the fact that a word is abelian equivalent to an nth power if and only if its Parikh vector is zero modulo n. On the one direction we have:

Lemma 7. *If a word v of length at least $k-1$ is k-abelian equivalent to an nth power, then*

$$|v|_t + |\mathrm{suf}_{k-1}(v)\mathrm{pref}_{k-1}(v)|_t \equiv 0 \pmod{n} \tag{1}$$

for all $t \in \Sigma^k$.

Proof. Let v be k-abelian equivalent to u^n. Then

$$\begin{aligned} &|v|_t + |\mathrm{suf}_{k-1}(v)\mathrm{pref}_{k-1}(v)|_t = |v\mathrm{pref}_{k-1}(v)|_t \\ =&|u^n\mathrm{pref}_{k-1}(v)|_t = |u^n\mathrm{pref}_{k-1}(u^n)|_t = n|u\mathrm{pref}_{k-1}(u^n)|_t \equiv 0 \pmod{n} \end{aligned}$$

for all $t \in \Sigma^k$. □

The converse does not hold. For example, $v = babbbbab$ satisfies (1) for $n = 2$ and $k = 3$ but it is not 3-abelian equivalent to any square. However, the converse does hold if $|v|_t$ is either large enough or zero for every t. This is formulated precisely in Lemma 11. To prove this we need the following definitions and Lemma 8. These were used in [7] to estimate the number of k-abelian equivalence classes.

Let $s_1, s_2 \in \Sigma^{k-1}$ and let

$$S(s_1, s_2, n) = \Sigma^n \cap s_1\Sigma^* \cap \Sigma^* s_2$$

be the set of words of length n that start with s_1 and end with s_2. For every word $u \in S(s_1, s_2, n)$ we can define a function

$$f_u : \Sigma^k \to \{0, \dots, n-k+1\}, \ f_u(t) = |u|_t.$$

If $u, v \in S(s_1, s_2, n)$, then $u \sim_k v$ if and only if $f_u = f_v$.

If a function $f : \Sigma^k \to \mathbb{N}_0$ is given, then a directed multigraph G_f can be defined as follows:

- The set of vertices is Σ^{k-1}.
- If $t = s_1a = bs_2$, where $a, b \in \Sigma$, then there are $f(t)$ edges from s_1 to s_2.

If $f = f_u$, then this multigraph is related to the Rauzy graph of u.

As stated above, the following lemma was proved in [7]. The proof is simple, so it is repeated here for completeness. Here $\deg^-$ denotes the indegree and $\deg^+$ the outdegree of a vertex in G_f.

Lemma 8. *For a function $f : \Sigma^k \to \mathbb{N}_0$ and words $s_1, s_2 \in \Sigma^{k-1}$, the following are equivalent:*

(i) there is a number n and a word $u \in S(s_1, s_2, n)$ such that $f = f_u$,
(ii) there is an Eulerian path from s_1 to s_2 in G_f,
(iii) the underlying graph of G_f is connected, except possibly for some isolated vertices, and $\deg^-(s) = \deg^+(s)$ for every vertex s, except that if $s_1 \neq s_2$, then $\deg^-(s_1) = \deg^+(s_1) - 1$ and $\deg^-(s_2) = \deg^+(s_2) + 1$,
(iv) the underlying graph of G_f is connected, except possibly for some isolated vertices, and

$$\sum_{a \in \Sigma} f(as) = \sum_{a \in \Sigma} f(sa) + c_s$$

for all $s \in \Sigma^{k-1}$, where

$$c_s = \begin{cases} -1, & \text{if } s = s_1 \neq s_2, \\ 1, & \text{if } s = s_2 \neq s_1, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. (i) $\Leftrightarrow$ (ii): $u = a_1 \dots a_n \in S(s_1, s_2, n)$ and $f = f_u$ if and only if

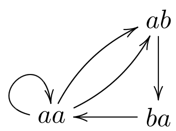
$$s_1 = a_1 \dots a_{k-1} \to a_2 \dots a_k \to \dots \to a_{n-k+2} \dots a_n = s_2$$

is an Eulerian path in G_f.

(ii) $\Leftrightarrow$ (iii): This is well known.

(iii) $\Leftrightarrow$ (iv): (iv) is just a reformulation of (iii) in terms of the function f. $\square$

Example 9. Let $k = 3$ and consider the word $u = aaabaab$. The multigraph G_{f_u} is



The word u corresponds to the Eulerian path

$$aa \to aa \to ab \to ba \to aa \to ab.$$

There is also another Eulerian path from aa to ab:

$$aa \to ab \to ba \to aa \to aa \to ab.$$

This corresponds to the word $aabaaab$, which is 3-abelian equivalent to u.

Example 10. We consider some functions $f : \{a, b\}^2 \to \mathbb{N}_0$.

If $f(aa) = f(bb) = 1$ and $f(t) = 0$ otherwise, then the underlying graph of G_f is not connected, so there does not exist a word u such that $f = f_u$.

If $f(ab) = 2$ and $f(t) = 0$ otherwise, then the indegree of a in G_f is zero but the outdegree is two, so there does not exist a word u such that $f = f_u$.

Lemma 11. *If*

$$|v|_t + |\mathrm{suf}_{k-1}(v)\mathrm{pref}_{k-1}(v)|_t \equiv 0 \pmod{n} \tag{2}$$

and either $|v|_t > (n-1)(k-1)$ *or* $|v|_t = 0$ *for all* $t \in \Sigma^k$*, then* v *is* k*-abelian equivalent to an* n*th power.*

Proof. Let $s_1 = \mathrm{pref}_{k-1}(v)$ and $s_2 = \mathrm{suf}_{k-1}(v)$. By Lemma 8,

$$\sum_{a \in \Sigma} f_v(as) = \sum_{a \in \Sigma} f_v(sa) + c_s \qquad \text{and} \qquad \sum_{a \in \Sigma} f_{s_2 s_1}(as) = \sum_{a \in \Sigma} f_{s_2 s_1}(sa) - c_s \tag{3}$$

for all $s \in \Sigma^{k-1}$, where

$$c_s = \begin{cases} -1, & \text{if } s = s_1 \neq s_2, \\ 1, & \text{if } s = s_2 \neq s_1, \\ 0, & \text{otherwise.} \end{cases}$$

By (2), a function $f : \Sigma^k \to \mathbb{N}_0$ can be defined by

$$f(t) = \frac{f_v(t) - (n-1) f_{s_2 s_1}(t)}{n}.$$

By (3),

$$\sum_{a \in \Sigma} f(as) = \sum_{a \in \Sigma} f(sa) + c_s$$

for all $s \in \Sigma^{k-1}$. If $f_v(t) > 0$, then

$$f_v(t) = |v|_t > (n-1)(k-1) \geq (n-1) f_{s_2 s_1}(t)$$

and thus $f(t) > 0$. This means that since the underlying graph of G_{f_v} is connected, also the underlying graph of G_f must be connected. By Lemma 8, there is a word $u \in S(s_1, s_2, |u|)$ such that $f = f_u$. Then u^n begins with s_1 and ends with s_2 and

$$|u^n|_t = n|u|_t + (n-1)|s_2 s_1|_t = n f(t) + (n-1) f_{s_2 s_1}(t) = f_v(t) = |v|_t$$

for all $t \in \Sigma^k$, so u^n is k-abelian equivalent to v. □

Now we are ready to express the main result of strongly k-abelian avoidability.

Theorem 12. *Let* Σ *be an alphabet and let* $k, n \geq 2$*. Every infinite word* $w \in \Sigma^\omega$ *contains a non-empty factor that is* k*-abelian equivalent to an* n*th power.*

Proof. For a prefix u of w, consider the pair $(f_u \bmod n, \mathrm{suf}_{k-1}(u))$. The number of different pairs is finite, so w has infinitely many prefixes $u_1, u_2, \ldots$ such that their pairs are the same. Let i be such that no factor of length k appearing only finitely many times in w appears after u_i. Let $j > i$ be such that if $u_j = u_i v$, then every other factor of length k appears at least $(n-1)(k-1)$ times in v. Then

$$\begin{aligned}&|v|_t + |\text{suf}_{k-1}(v)\text{pref}_{k-1}(v)|_t = |\text{suf}_{k-1}(v)v|_t = |\text{suf}_{k-1}(u_i)v|_t\\ =&|u_i v|_t - |u_i|_t = f_{u_j}(t) - f_{u_i}(t) \equiv 0 \pmod{n}\end{aligned}$$

for all $t \in \Sigma^k$. Thus v satisfies the conditions of Lemma 11 and v is k-abelian equivalent to an nth power. □

4 Further Questions

Some further questions that might be asked on strongly k-abelian powers are:

- How many k-abelian equivalence classes of words of length l contain an nth power?
- How many words there are in those equivalence classes, i.e. how many words of length l are strongly k-abelian nth powers?
- What is the length of the longest word avoiding strongly k-abelian nth powers?
- How many words avoid strongly k-abelian nth powers?

The answers depend on k, n, l and the size of the alphabet. The analysis of these questions is outside the scope of this extended abstract, but a few remarks can be made.

First, it is easy to prove that two squares uu and vv are k-abelian equivalent if and only if u and v are. Thus the number of k-abelian equivalence classes of words of length $2l$ containing a square is the number of k-abelian equivalence classes of words of length l. This number has been estimated in [7] and is polynomial with respect to l.

Second, some of the equivalence classes contain exponentially many words. For example, a word on the alphabet $\{a, b\}$ is 2-abelian equivalent to $(a^m(ab)^m)^2$ if and only if it has the same length, begins with a, ends with b, contains no two consecutive b's and contains $2m$ b's. The number of such words is exponential with respect to m.

Example 13. In $\{a, b\}^{12}$ there are

- 64 squares,
- 168 2-abelian squares,
- 924 abelian squares,
- 1024 strongly 2-abelian squares,
- 2048 strongly abelian squares,
- 4096 words.

Those 1024 strongly 2-abelian squares belong to 32 different equivalence classes and strongly abelian squares belong to 7 different equivalence classes. Representatives for each of these seven classes over a binary alphabet are as follows: $a^{12}, a^{10}b^2, a^8b^4, a^6b^6, a^4b^8, a^2b^{10}, b^{12}$.

5 Conclusion

We have shown that for $k, n \geq 2$ every infinite word contains a non-empty factor which is strongly abelian nth power as well as a non-empty factor which is strongly k-abelian nth power. As is known, usual abelian nth powers can be avoided depending on the value of n and the size of the alphabet. Corresponding results are also known for k-abelian powers. Other questions arising from the notion of strongly k-abelian equivalence are, for example, counting the number of words of length l that contain strongly k-abelian nth powers, or counting the number of strongly k-abelian nth powers of length l.

References

1. A. Carpi and A. De Luca: *Square-free words on partially commutative free monoids.* Information Processing Letters 22(3), 125-131 (1986).
2. C. Choffrut and J. Karhumäki: *Combinatorics of words.* In: G. Rozenberg, A. Salomaa (eds.) Handbook of Formal Languages, vol. 1, 329-438. Springer, Heidelberg (1997)
3. F. M. Dekking: *Strongly non-repetitive sequences and progression-free sets.* J. Combin. Theory Ser. A 27(2), 181–185 (1979).
4. M. Huova, J. Karhumäki, A. Saarela and K. Saari: *Local squares, periodicity and finite automata.* In: C.S. Calude, G. Rozenberg, A. Salomaa (eds.) Rainbow of Computer Science, LNCS, vol. 6570, 90-101. Springer, Heidelberg (2011).
5. M. Huova and J. Karhumäki: *On the unavoidability of k-abelian squares in pure morphic words.* Journal of Integer Sequences 16, article 13.2.9 (2013).
6. M. Huova: *Existence of an infinite ternary 64-abelian square-free word.* Special Issue of the journal RAIRO - Theoretical Informatics and Applications dedicated to "Journees Montoises d'Informatique Theorique 2012". Submitted.
7. J. Karhumäki, A. Saarela and L. Zamboni: *On a generalization of Abelian equivalence and complexity of infinite words.* Submitted, arXiv preprint at http://arxiv.org/abs/1301.5104.
8. V. Keränen: *Abelian squares are avoidable on 4 letters.* In: W. Kuich (ed.) ICALP 1992, LNCS, vol. 623, 41–52. Springer, Heidelberg (1992).
9. M. Lothaire: *Combinatorics on words.* Addison-Wesley, Reading (1983).
10. R. Mercas and A. Saarela: *3-abelian cubes are avoidable on binary alphabets.* Proceedings of the 17th International Conference on Developments in Language Theory, to appear (2013).
11. P. Ochem, N. Rampersad and J. Shallit: *Avoiding approximate squares.* International Journal of Foundations of Computer Science 19(3), 633-648 (2008).
12. A. Thue: *Über unendliche Zeichenreihen.* Norske vid. Selsk. Skr. Mat. Nat. Kl. 7, 1–22 (1906).
13. A. Thue: *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen.* Norske vid. Selsk. Skr. Mat. Nat. Kl. 1, 1–67 (1912).