Strongly k-abelian repetitions

M. Huova and A. Saarela

Department of Mathematics and Statistics & TUCS,
University of Turku, Finland
Supported by the Academy of Finland under grant 257857.

WORDS 2013, Turku

Outline

Introduction

Preliminaries

Main results

References

1. Introduction

Avoidability

- infinite binary cube-free and ternary square-free words [4, 5]
- an infinite ternary abelian cube-free word [2]
- an infinite abelian square-free word over four letter alphabets [1]
- *k*-abelian repetitions
  - for $k \geq 3$: k-abelian cubes can be avoided over binary alphabets [2]
  - 2-abelian squares can not be avoided over a ternary alphabet [3]
  - for any k, k-abelian squares can not be avoided in ternary pure morphic words [4]
- strongly abelian repetitions
- strongly *k*-abelian repetitions

2. Preliminaries

Definition of k-abelian equivalence

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

   Different words of length at most k are not k-abelian equivalent as well as words of different lengths.

Remarks:

- $\sim_k$ is an equivalence relation, and also a congruence
- $u = v \Rightarrow u \sim_k v \Rightarrow u \equiv_a v$
- $u = v \Leftrightarrow (u \sim_k v \quad \forall\, k \geq 1)$

Congruence

- Congruence is an equivalence relation R s.t. $uvRu'v'$ whenever uRu' and vRv'.
- If uRv then the product uv is an R-square.

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

Example:
A word *aabb* is not an abelian square but it's a strongly abelian square.

Strongly R-squares

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

- $\Rightarrow$: Clear.
- $\Leftarrow$: If $wRuv$ and uRv, then $wRuu$, because uRv implies $uuRuv$.

$\square$

An example

Example:
Consider words $w \in \{a, b\}^6$ having $|w|_a = 4$ and $|w|_b = 2$.

$aabaab$	$aababa$	$aaaabb$
$abaaba$	$aabbaa$	$aaabab$
$baabaa$	$abaaab$	$aaabba$
	$ababaa$	$abbaaa$
	$baaaab$	$babaaa$
	$baaaba$	$bbaaaa$

All these words are strongly abelian squares.

Definitions

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

Related works

- Square freeness in partially commutative monoids was studied by Carpi and De Luca in [1].
- Avoidability of approximate squares has been studied by Ochem, Rampersad and Shallit [3].
  - uv, where the Hamming distance of u and v is "small enough" or equivalently as words w such that the Hamming distance of w and some square is "small enough"

3. Main results

Strongly abelian nth powers

Theorem:
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.

Parikh vector:
Let $\Sigma = \{a_1, a_2, \ldots, a_m\}$, then Parikh vector is a function $p : \Sigma^* \to \mathbb{N}^m$ where $p(w) = (|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_m})$.

Strongly abelian nth powers

Proof:

- A word x is abelian equivalent to an nth power $\Leftrightarrow p(x) \equiv 0 \pmod{n}$.
- The number of different Parikh vectors modulo n is finite.
- $\Rightarrow$ The infinite word w has two prefixes u and uv s.t. $p(u) \equiv p(uv) \pmod{n}$.
- $\Rightarrow p(v) \equiv 0 \pmod{n}$.
- $\therefore$ The word v is abelian equivalent to an nth power. $\square$

Strongly abelian nth powers

Example:

Consider the fixed point $h^\infty(a)$ of the morphism

$$h : \{a, b, c\}^* \to \{a, b, c\}^* \text{ defined by } \begin{cases} h(a) = aabc \\ h(b) = bbc \\ h(c) = acc. \end{cases}$$

The fixed point $h^\infty(a) = aabcaabcbbcacc\ldots$ is known to be abelian cube-free but it's not strongly abelian cube-free.

The prefixes of length 3 and 12 of this word are $u = aab$ and $uv = aabcaabcbbca$. Now $p(u) = (2, 1, 0)$ and $p(uv) = (5, 4, 3) \equiv (2, 1, 0) \pmod 3$. Thus $p(v) = (3, 3, 3) \equiv (0, 0, 0) \pmod 3$ and clearly $v = caabcbbca$ is abelian equivalent to abc^3.

Strongly k-abelian nth powers

Lemma:
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

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

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

Strongly k-abelian nth powers

The difficulty in generalizing result from strongly abelian nth powers to strongly k-abelian nth powers lies in the fact that the converse of the previous Lemma does not hold.

Example:

The word $v = babbbbab$ satisfies (1) for $n = 2$ and $k = 3$ but it is not 3-abelian equivalent to any square.

Auxiliary notions

We need a few definitions and one Lemma that were used in [5].

- For $s_1, s_2 \in \Sigma^{k-1}$ let $S(s_1, s_2, n) = \Sigma^n \cap s_1\Sigma^* \cap \Sigma^* s_2$.
- For all $u \in S(s_1, s_2, n)$ define $f_u : \Sigma^k \to \{0, \dots, n-k+1\},\ f_u(t) = |u|_t$.
- 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 $t = s_1 a = b s_2$, 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.

Example of functions and their directed multigraphs

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

1. If $f(aa) = f(bb) = 1$ and $f(t) = 0$ otherwise.
2. If $f(ab) = 2$ and $f(t) = 0$ otherwise.

Auxiliary Lemma

Lemma: 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 $u \in S(s_1, s_2, n)$ s.t. $f = f_u$,

(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 $\deg^-(s_2) = \deg^+(s_2) + 1$,

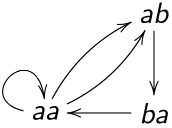
(iv) the underlying graph of G_f is connected, except possibly for some isolated vertices, and $\sum_{a \in \Sigma} f(as) = \sum_{a \in \Sigma} f(sa) + c_s$ for all $s \in \Sigma^{k-1}$, where

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

Example of two words and corresponding Eulerian paths

Example:

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 of the use of the auxiliary Lemma

Example:

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 $f = f_u$.

Strongly k-abelian nth powers

For a word v it's enough to require $|v|_t$ to be large enough to force the word to be a strongly k-abelian nth power.

Lemma:

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 $|v\mathrm{pref}_{k-1}(v)|_t = 0$ for all $t \in \Sigma^k$, then v is k-abelian equivalent to an nth power.

Strongly k-abelian nth powers

Proof:
Let $s_1 = \mathrm{pref}_{k-1}(v)$ and $s_2 = \mathrm{suf}_{k-1}(v)$. By auxiliary Lemma, the underlying graphs of G_{f_v} and $G_{f_{s_2 s_1}}$ are connected and

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

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

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

Strongly k-abelian nth powers

Proof continues:

By the assumptions of the Lemma, we can define a function $f : \Sigma^k \to \mathbb{N}_0$ by

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

- Clearly $f(t) \in \mathbb{N}_0$ for all $t \in \Sigma^k$.
- If $|\mathrm{vpref}_{k-1}(v)|_t = 0$ then $f_v(t) = f_{s_2 s_1}(t) = 0$, thus $f(t) = 0$.
- Otherwise $|v|_t > 0$, and then

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

and thus $f(t) > 0$.

Strongly k-abelian nth powers

Proof continues:

Now the underlying graph of G_{f_v} is connected and if $f_v(t) > 0$ then $f(t) > 0$, thus the underlying graph of G_f must be connected, too. By using (3), we get

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

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

So by auxiliary Lemma, 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

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

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

Strongly k-abelian nth powers

Theorem:
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 nth power.

Strongly k-abelian nth powers

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

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

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

Numerical results

Example:

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: $a^{12}, a^{10}b^2, a^8b^4, a^6b^6, a^4b^8, a^2b^{10}, b^{12}$.

Conclusion and further questions

- For $n \geq 2$ every infinite word contains a non-empty factor which is strongly abelian nth power.
- For $k, n \geq 2$ every infinite word contains a non-empty factor which is strongly k-abelian nth power.
- 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?

References

References

A. Carpi and A. De Luca: *Square-free words on partially commutative free monoids*. Information Processing Letters 22(3), 125-131 (1986).

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

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).

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).

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.

References

V. Keränen: *Abelian squares are avoidable on 4 letters*. In: W. Kuich (ed.) ICALP 1992, LNCS, vol. 623, 41–52. Springer, Heidelberg (1992).

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).

P. Ochem, N. Rampersad and J. Shallit: *Avoiding approximate squares*. International Journal of Foundations of Computer Science 19(3), 633-648 (2008).

A. Thue: *Über unendliche Zeichenreihen*. Norske vid. Selsk. Skr. Mat. Nat. Kl. 7, 1–22 (1906).

A. Thue: *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*. Norske vid. Selsk. Skr. Mat. Nat. Kl. 1, 1–67 (1912).

Thank You!