Combinatorics on words and k-abelian equivalence

Juhani Karhumäki
(jointly with M. Huova and A. Saarela)

Department of Mathematics and TUCS,
University of Turku, Finland

BIRS, February 2012

Outline

1. Basics

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 $u \equiv_{a,k} v$, 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.

Here pref_{k-1} (resp. suf_{k-1}) denotes the prefix (resp. suffix) of length $k-1$.
Remarks:

- $\equiv_{a,k}$ is an equivalence relation
- $u = v \Rightarrow u \equiv_{a,k} v \Rightarrow u \equiv_a v$
- $u = v \Leftrightarrow u \equiv_{a,k} v \ \ \forall\, k \geq 1$

The number of the equivalence classes

We can estimate the number of equivalence classes of 2- and 3-abelian words of length n over binary alphabet with the help of characterization of the representatives of equivalence classes, see [HKSS].

- 2-abelian case: $n^2 - n + 2$, i.e. $\Theta\left(n^2\right)$
- 3-abelian case: $\Theta\left(n^4\right)$

In general, we can estimate the number of k-abelian equivalence classes of words of length n with the following result, see [KSZ].

- Let $k \geq 1$ and $m \geq 2$ be fixed numbers and let Σ be an m-letter alphabet. The number of k-abelian equivalence classes of Σ^n is $\Theta(n^{(m-1)m^{k-1}})$.

For example, in a binary alphabet for $k = 4$ the number is $\Theta(n^8)$.

2. Older observations

k-generalized Parikh properties

Problems on 1-free morphisms and k-generalized Parikh properties (k-abelian) can be reduced to problems on 1-free morphisms and usual Parikh properties (abelian) in a bigger alphabet, as stated in [Ka].

- Let $h : \Sigma^* \to \Sigma^*$ be a 1-free morphism and $k \geq 1$. Then there exists a morphism $\hat{h} : \hat{\Sigma}^* \to \hat{\Sigma}^*$ such that $\bigwedge_k h = \hat{h} \bigwedge_k$, i.e., the following diagram holds true for all $x \in \Sigma^*$

$$
\begin{array}{ccc}
x & \xrightarrow{\quad \bigwedge_k \quad} & \hat{x} \\
{\scriptstyle h}\downarrow & & \downarrow{\scriptstyle \hat{h}} \\
h(x) & \xrightarrow{\bigwedge_k} & \widehat{h(x)} = \hat{h}(\hat{x}).
\end{array}
$$

Here $\bigwedge_k$ is a mapping from Σ^* to $\hat{\Sigma}^*$ and $\bigwedge_k(x) = \hat{x}$.

Modification of the PCP

From the result of the previous slide and with the help of earlier results of automata theory it is shown in [Ka] that a modification of the Post Correspondence Problem is decidable.

- Let h and g be 1-free morphisms from Σ^* into Δ^* and $k \geq 0$ and define sets $P_k(h, g)$ of the form

$$P_k(h,g) = \left\{ x \in \Sigma^+ \mid \exists y \in \Sigma^+ : \ x \equiv_k y, \ h(x) = g(y) \right\}.$$

  Then for a given integer k the problem of emptiness of the set $P_k(h, g)$ is decidable.

- It is also decidable whether $E^k(h, g)$ is empty for

$$E^k(h,g) = \left\{ x \in \Sigma^+ \mid h(x) \equiv_k g(x) \right\}.$$

3. k-abelian repetitions

Repetitions and avoidability

Let u, v and w be words over Σ.

- We say that *a repetition of order two*, i.e. $v^2 = vv$, is *a square* and correspondingly $v^3 = vvv$ is *a cube*.
- Similarly vu is *an abelian square* if $u \equiv_a v$ and uvw is *an abelian cube* if $u \equiv_a v \equiv_a w$.
- The word w *contains* a square if it has a square as a factor, i.e. $w = \alpha v^2 \beta$ for some $v \in \Sigma^+$, $\alpha, \beta \in \Sigma^*$.
- If the word w does not contain a square we say that it *avoids* squares and it is *a square-free word*.
- We say that the alphabet Σ *avoids* squares if there exists an infinite word over Σ that avoids squares.

Earlier results

Avoidability of squares			Avoidability of cubes		
size of the alph.	type of rep. $=$	$\equiv_a$	size of the alph.	type of rep. $=$	$\equiv_a$
2	$-$	$-$	2	$+$	$-$
3	$+$	$-$	3	$+$	$+$
4	$+$	$+$			

Table: Avoidability of different types of repetitions in infinite words.

Results for equality: A. Thue
Results for abelian equality: (A. A. Evdokimov, P. A. B. Pleasant), V. Keränen and F. M. Dekking

Examples

- $abbabaabb \equiv_{a,2} aabbabbab$
- $abcababb \equiv_{a,3} ababcabb$
- $abcababb \equiv_{a,2} ababcabb$
- $abbabaabb \not\equiv_{a,3} aabbabbab$
- $abca \not\equiv_{a,2} acba$

Questions

What is the size of the smallest alphabet that avoids k-abelian squares (resp. cubes)?

- Difficult even for $k = 2$

Avoidability of squares				Avoidability of cubes			
size of the alph.	type of rep. $=$	$\equiv_{a,2}$	$\equiv_a$	size of the alph.	type of rep. $=$	$\equiv_{a,2}$	$\equiv_a$
2	$-$	$-$	$-$	2	$+$	?	$-$
3	$+$	?	$-$	3	$+$	$+$	$+$
4	$+$	$+$	$+$				

Table: Earlier results give limits for our problems.

Iterating morphisms

The infinite words for the results of the previous slide are obtained by iterating morphisms.

- Infinite cube-free, in fact overlap-free, Thue-Morse word (morphism: $0 \to 01$, $1 \to 10$): $01\,\overbrace{101001}\,\overbrace{100101}\,\overbrace{101001}\,011\ldots$
- Infinite cube-free word (morphism: $0 \to 001$, $1 \to 011$): $001001\,\overbrace{011001}\,\overbrace{001011}\,\overbrace{001011}\,011\ldots$

Unavoidability of 2-abelian squares in ternary alphabets

- The longest ternary word which is 2-abelian square-free has length 537.
- The size of the smallest alphabet avoiding 2-abelian squares is four.

This longest word, given below, is unique up to the permutations of the alphabet, $\Sigma = \{a, b, c\}$.

abcbabcacbacabacbabcbacabcbabcabacabcacbacabacbabcbacbcacbabcacbcabcba

bcabacbabcbacbcacbacabacbabcbacabcbabcabacabcacbacabacbabcbacbcacbacab

acbcabacabcacbcabcbacbcacbacabacbabcbacbcacbabcacbcabcbabcabacbabcbacb

cacbacabacbabcbacabcbabcabacabcacbacabacbabcbacbcacbacabacbcabacabcacb

cabcbabcabacabcacbacabacbabcbacabcbabcabacabcacbcabcbabcabacbabcbacbca

cbabcacbcabcbabcabacabcacbcabcbacbcacbacabacbcabacabcacbcabcbabcabacab

cacbacabacbabcbacabcbabcabacabcacbcabcbabcabacbabcbacbcacbabcacbcabcba

bcabacabcacbacabacbabcbacabcbabcabacabcacbabcba

Unavoidability of 2-abelian squares in ternary alphabets

- The longest ternary word which is 2-abelian square-free has length 537.
- The size of the smallest alphabet avoiding 2-abelian squares is four.

This longest word, given below, is unique up to the permutations of the alphabet, $\Sigma = \{a, b, c\}$.

abcbabcacbacabacbabcbacabcbabcabacabcacbacabacbabcbacbcacbabcacbcabcba

bcabacbabcbacbcacbacabacbabcbacabcbabcabacabcacbacabacbabcbacbcacbacab

acbcabacabcacbcabcbacbcacbacabacbabcbacbcacbabcacbcabcbabcabacbabcbacb

cacbacabacbabcbacabcbabcabacabcacbacabacbabcbacbcacbacabacbcabacabcacb

cabcbabcabacabcacbacabacbabcbacabcbabcabacabcacbcabcbabcabacbabcbacbca

cbabcacbcabcbabcabacabcacbcabcbacbcacbacabacbcabacabcacbcabcbabcabacab

cacbacabacbabcbacabcbabcabacabcacbcabcbabcabacbabcbacbcacbabcacbcabcba

bcabacabcacbacabacbabcbacabcbabcabacabcacbabcba $\genfrac{}{}{0pt}{}{a}{\genfrac{}{}{0pt}{}{b}{c}}$

The number of 2-abelian square-free words

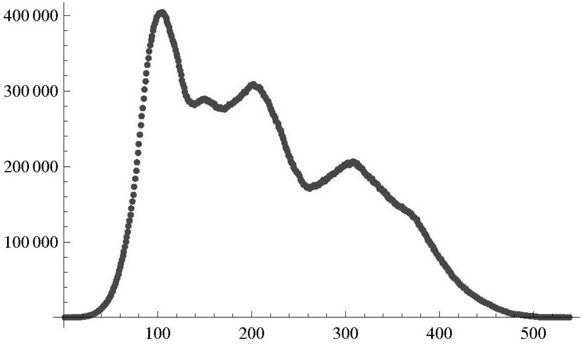

Figure: The number of ternary 2-abelian square-free words with respect to their lengths.

Behaviour of 2-abelian cube-free words in binary alphabets

- There exist binary words of more than 100 000 letters that still avoid 2-abelian cubes.
- The number of words with fixed lengths up to 60 letters grows approximately with factor 1,3 with respect to the lengths.
- The number of binary 2-abelian cube-free words of length 60 is 478 456 030.
- With length 12 there exist more binary 2-abelian cube-free words (254) than ternary 2-abelian square-free words (240).
- Examples of such binary 2-abelian cube-free words that the number of their extensions grows again approximately with factor 1,3 when increasing the length of extensions by one.

The number of 2-abelian cube-free words

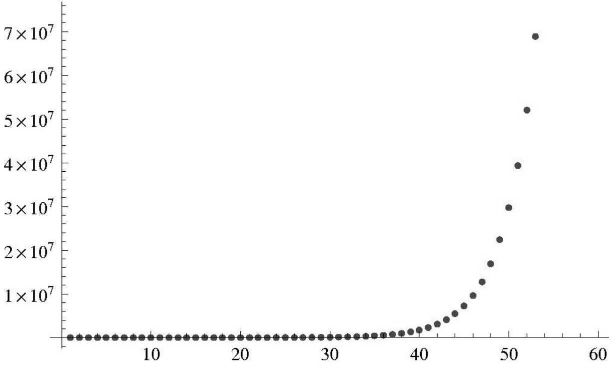

Figure: The number of binary 2-abelian cube-free words with respect to their lengths for small values of length.

The number of extensions

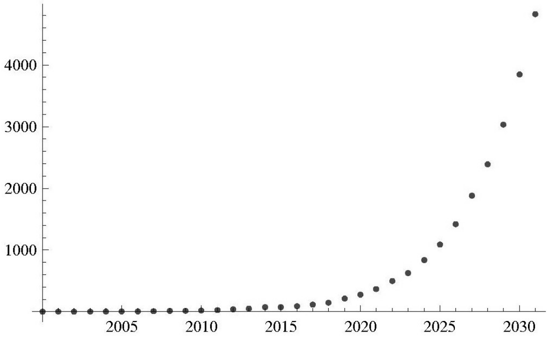

Figure: The numbers of 2-abelian cube-free words of lengths 2 000-2 031 having a fixed prefix of length 2 000.

A cube-free and a square-free word

- The morphism

$$h : \begin{cases} a \mapsto aab \\ b \mapsto abb \end{cases},$$

defines a cube-free word, as is particularly simple to see, if we notice that $h(x) = axb$ for $x \in \{a, b\}$.

- By using corresponding methods to produce an infinite ternary square-free word we end up with a morphic word with morphism

$$g : \begin{cases} a \mapsto abcbacbcabcba \\ b \mapsto bcacbacabcacb \\ c \mapsto cabacbabcabac \end{cases}.$$

It was already proved by Leech [Le] that this word is square-free.

An 8-abelian cube-free binary word

In [HKS] we use similar techniques to prove the existence of an infinite binary 8-abelian cube-free word.

Theorem

Let $w \in \{0,1,2,3\}^\omega$ be an abelian square-free word. Let $k \leq n$ and $h : \{0,1,2,3\}^* \to \{0,1\}^*$ be an n-uniform morphism that satisfies the following three conditions. Then $h(w)$ is k-abelian cube-free.

1. *If $u \in \{0,1,2,3\}^4$ is square-free, then $h(u)$ is k-abelian cube-free.*
2. *If $u \in \{0,1,2,3\}^*$ and v is a factor of $h(u)$ of length $2k-2$, then every occurrence of v in $h(u)$ has the same starting position modulo n.*

An 8-abelian cube-free binary word

Theorem (Continues)

3. *There is a number i such that $0 \leq i \leq n-k$ and for at least three letters $x \in \{0,1,2,3\}$, $v = h(x)[i..i+k]$ satisfies $|h(u)|_v = |u|_x$ for every $u \in \{0,1,2,3\}^*$.*

An 8-abelian cube-free binary word

By using the previous Theorem and the fact that there exists an infinite abelian square-free word over four letter alphabet, see [Ke], we can construct an 8-abelian cube-free binary word.

Theorem

Let $w \in \{0,1,2,3\}^\omega$ be an abelian square-free word. Let $h : \{0,1,2,3\}^* \to \{0,1\}^*$ be the morphism defined by

$$\begin{aligned} h(0) &= 00101\ 0\ 011001\ 0\ 01011, \\ h(1) &= 00101\ 0\ 011001\ 1\ 01011, \\ h(2) &= 00101\ 1\ 011001\ 0\ 01011, \\ h(3) &= 00101\ 1\ 011001\ 1\ 01011. \end{aligned}$$

Now $h(w)$ is 8-abelian cube-free.

4. Local vs. global regularity

Description of the problem

In [HKSS] we examine the following problem:

- For a given number n, if a binary right-infinite word contains at every position a square of a word of length at most n, is the word necessarily ultimately periodic?

We have nine different variants depending on whether we study the word, the abelian or the 2-abelian case and whether we use a consept of *left square*, *right square* or *centered square*.

Type of a square

A word w contains everywhere a

- *left square* of length at most n, if every factor of w of length $2n$ has a nonempty square as a suffix,
- *right square* of length at most n, if every factor of w of length $2n$ has a nonempty square as a prefix,
- *centered square* of length at most n, if every factor of w of length $2n$ has a nonempty square exactly in the middle, i.e. is of the form $uxxv$, where $|u| = |v|$ and $x \neq 1$.

Results

The following Table presents the minimal values of n for which there are aperiodic right-infinite words containing an ordinary (or 2-abelian or abelian) left (or right or centered) square of length at most n everywhere.

	words	2-abelian	abelian
left	5	5	3
right	5	5	3
centered	∞	12	8

Table: Optimal values for local regularity which does not imply global regularity in our problems.

About general k-abelian case

There are two remarks on general k-abelian case.

- For left and right squares the values of Table would remain as 5.
- For the centered variant of the problem the exact borderline for k-abelian repetitions when $k \geq 3$ is unknown.

References

References

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

A. A. Evdokimov: *Strongly asymmetric sequences generated by a finite number of symbols*. Dokl. Akad. Nauk SSSR 179, 1268-1271 (1968); English translation in Soviet Math. Dokl. 9, 536-539 (1968).

M. Huova, J. Karhumäki, A. Saarela: *Problems in between words and abelian words: k-abelian avoidability*. In: G. Rozenberg, A.Salomaa (eds.) Special issue of Theoretical Computer Science, to be published.

References

- M. Huova, J. Karhumäki, A. Saarela, K. Saari: *Local squares, periodicity and finite automata*. In: C. Calude, G. Rozenberg, A. Salomaa (eds.) Rainbow of Computer Science, 90-101, Springer, 2011.
- J. Karhumäki: *Generalized Parikh Mappings and Homomorphisms*. Information and control 47, 155-165 (1980).
- J. Karhumäki, A. Saarela, L. Zamboni: *Generalized abelian equivalence*. Manuscript to be published.
- V. Keränen: *Abelian squares are avoidable on 4 letters*. In: W. Kuich (ed.) ICALP 1992. LNCS, vol. 623, 41-52. Springer, Heidelberg, 1992.

References

- J. Leech: *A problem on strings of beads*. Math. Gazette 41, 277-278 (1957).
- P. A. B. Pleasant: *Non-repetitive sequences*. Proc. Cambridge Philos. Soc. 68, 267-274 (1970).
- 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 For Your Attention!