k-Abelian Equivalence and Factor Complexity

Juhani Karhumäki
(with A. Saarela and L. Zamboni)

Department of Mathematics and Statistics
and FUNDIM Centre,
University of Turku, Finland

Challenges in Combinatorics on Words
Fields Institute, Toronto
April 2013

Outline

Definition of k-Abelian Equivalence

For words x, y, let $|x|_y$ denote the number of occurrences of y in x.

Let $k \geq 1$. Words u and v are *k-Abelian equivalent* if

- $|u|_w = |v|_w$ for all w such that $|w| \leq k$.

This is denoted by $u \sim_k v$.

If $|u|, |v| \geq k - 1$, then $u \sim_k v$ if and only if

- $|u|_w = |v|_w$ for all w such that $|w| = k$ and
- $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v)$ and $\mathrm{suff}_{k-1}(u) = \mathrm{suff}_{k-1}(v)$.

Examples

Example

$aaabaab$ and $aabaaab$ are 3-Abelian equivalent because

- the factors of length 3 are aaa, aab, aba, baa, aab for both and
- $\mathrm{pref}_2(aaabaab) = aa = \mathrm{pref}_2(aabaaab)$ and $\mathrm{suff}_2(aaabaab) = ab = \mathrm{suff}_2(aabaaab)$.

$aabb$ and $abab$ are not 2-Abelian equivalent because

- $|aabb|_{aa} = 1 \neq 0 = |abab|_{aa}$.

aba and bab are not 2-Abelian equivalent because

- $\mathrm{pref}_1(aba) = a \neq b = \mathrm{pref}_1(bab)$
- (or because they are not Abelian equivalent).

Remarks

- ∞-Abelian equivalence is equality.
- 1-Abelian equivalence is Abelian equivalence.
- $\sim_k$ is an equivalence relation, and also a congruence.
- $u = v \Rightarrow u \sim_k v \Rightarrow u \sim_1 v$
- $u \sim_{k+1} v \Rightarrow u \sim_k v$
- $u = v \Leftrightarrow (u \sim_k v \ \forall k \in \mathbb{N}_1)$

Number of Equivalence Classes

Theorem (Karhumäki, Saarela, Zamboni [3])

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

$$\Theta(n^{m^k - m^{k-1}}).$$

So the number of equivalence classes is polynomial with respect to n, but the degree of the polynomial grows exponentially with respect to k.

Example

The number of 2-Abelian equivalence classes of $\{0,1\}^n$ is $\Theta(n^2)$.
The exact number is $n^2 - n + 2$.

Outline

Factor Complexity

Let $w \in \Sigma^\omega$. The set of factors of w of length n is denoted by $\mathcal{F}_w(n)$. The *factor complexity* of w is the function $\mathcal{P}_w^{(\infty)} : \mathbb{N}_1 \to \mathbb{N}_1$ defined by

$$\mathcal{P}_w^{(\infty)}(n) = \#\mathcal{F}_w(n).$$

In other words, $\mathcal{P}_w^{(\infty)}(n)$ is the number of factors of w of length n.

We are mostly interested in the binary case $\Sigma = \{0, 1\}$.

Classical results:

- w is ultimately periodic iff $\mathcal{P}_w^{(\infty)}$ is bounded; otherwise $\mathcal{P}_w^{(\infty)}(n) \geq n + 1$ for all n (Morse-Hedlund [5]).
- w is Sturmian iff $\mathcal{P}_w^{(\infty)}(n) = n + 1$ for all n (by definition).

k-Abelian Complexity

The *k-Abelian complexity* of w is the function $\mathcal{P}_w^{(k)} : \mathbb{N}_1 \to \mathbb{N}_1$ defined by

$$\mathcal{P}_w^{(k)}(n) = \#(\mathcal{F}_w(n)/\sim_k).$$

In other words, $\mathcal{P}_w^{(k)}(n)$ is the number of nonequivalent factors of w of length n.

Classical results on Abelian complexity:

- w is periodic iff $\mathcal{P}_w^{(1)}(n) = 1$ for some $n \geq 1$.
- w is Sturmian iff it is aperiodic and $\mathcal{P}_w^{(1)}(n) = 2$ for all n (Coven-Hedlund [1]).

k-Abelian Morse-Hedlund

Theorem (Karhumäki, Saarela, Zamboni [3])

$$\exists n : \mathcal{P}_w^{(k)}(n) < \min(n+1, 2k) \Rightarrow w \textit{ ultimately periodic} \Rightarrow \mathcal{P}_w^{(k)} \textit{ bounded}$$

This does not give a characterization of ultimate periodicity.

Compare with the classical result:

$$\exists n : \mathcal{P}_w^{(\infty)}(n) < n+1 \Leftrightarrow w \text{ ultimately periodic} \Leftrightarrow \mathcal{P}_w^{(\infty)} \text{ bounded}$$

Sturmian Words

Theorem (Karhumäki, Saarela, Zamboni [3])

$$w \textit{ Sturmian} \Leftrightarrow w \textit{ aperiodic and } \forall n : \mathcal{P}_w^{(k)}(n) = \min(n+1, 2k)$$

This gives a characterization of Sturmian words among aperiodic words. Sturmian words have the smallest possible k-Abelian complexity among aperiodic words.

Compare with the classical result:

$$w \text{ Sturmian} \Leftrightarrow \forall n : \mathcal{P}_w^{(\infty)}(n) = n+1$$

Sturmian Words and Ultimately Periodic Words

For any finite k, there are ultimately periodic words having the same k-Abelian complexity as Sturmian words.

Example

If $w = 0^{2k-1}1^\omega$, then $\mathcal{P}_w^{(k)}(n) = \min(n+1, 2k)$ for all n.

On the other hand, for any ultimately periodic word w there is a k such that $\mathcal{P}_w^{(k)}(n) < \min(n+1, 2k)$ for all sufficiently large n.

Outline

Thue-Morse Word

Let

$$T = 0110100110010110\ldots$$

be the Thue-Morse word, which is a fixed point of the morphism defined by $0 \mapsto 01, 1 \mapsto 10$. It is known that

$$\mathcal{P}_T^{(\infty)}(n) = \Theta(n) \qquad \text{and} \qquad \mathcal{P}_T^{(1)}(n) = \begin{cases} 2 & \text{if } n \text{ is odd} \\ 3 & \text{if } n \text{ is even} \end{cases}.$$

What about $\mathcal{P}_T^{(2)}$? The first few values are

$$2, 4, 6, 8, 6, 8, 10, 8, 6, 8, 8, 10, 10, 10, 8, 8, 6, 8, 10, 10.$$

Logarithmic Growth

i	$n = (2 \cdot 4^i + 4)/3$	$\mathcal{P}_T^{(2)}(n)$
0	2	4
1	4	8
2	12	10
3	44	14
4	172	16
5	684	20
6	2732	22

So is $\mathcal{P}_T^{(2)}(n) = \Theta(\log n)$?

Constant Values

i	$n = 2^i + 1$	$\mathcal{P}_T^{(2)}(n)$
1	3	6
2	5	6
3	9	6
4	17	6
5	33	6
6	65	6
7	129	6

So is $\mathcal{P}_T^{(2)}(n) = \Theta(1)$?
It can't be both $\Theta(\log n)$ and $\Theta(1)$.

Upper and Lower Complexities

Factor complexity functions are always increasing.
For k-Abelian complexity this is not true.

We define *upper k-Abelian complexity* $\mathcal{U}_w^{(k)}$ and *lower k-Abelian complexity* $\mathcal{L}_w^{(k)}$ by

$$\mathcal{U}_w^{(k)}(n) = \max_{m \leq n} \mathcal{P}_w^{(k)}(m) \qquad \text{and} \qquad \mathcal{L}_w^{(k)}(n) = \min_{m \geq n} \mathcal{P}_w^{(k)}(m).$$

These quantities might deviate from one another quite drastically.
This is the case for the Thue-Morse word.

Complexity of the Thue-Morse Word

Theorem (Karhumäki, Saarela, Zamboni [4])

Let $k \geq 2$. Then

$$\mathcal{U}_T^{(k)}(m) = \Theta(\log n), \qquad \mathcal{L}_T^{(k)}(m) = \Theta(1), \qquad \mathcal{L}_T^{(2)}(m) = 6.$$

Idea of the proof:

- Prove a similar result about the Abelian complexity of the fixed point S of the morphism defined by $0 \mapsto 01, 1 \mapsto 00$.
- S is closely related to T.
- Prove that the 2-Abelian complexity of T is of the same order as the Abelian complexity of S.
- Prove that the k-Abelian complexity of T is of the same order as the 2-Abelian complexity of T for all $k \geq 2$.

Outline

Arbitrarily Slowly Growing Complexities

If $\mathcal{P}_w^{(\infty)}(n) < n+1$ for some n, then $\mathcal{P}_w^{(\infty)}(n) = O(1)$, so there is a gap between complexity $n+1$ and bounded complexities.

For k-Abelian complexities there is no such gap.

Theorem (Karhumäki, Saarela, Zamboni [4])

For every increasing unbounded function $f : \mathbb{N}_1 \to \mathbb{N}_1$ there is a uniformly recurrent word $w \in \{0,1\}^\omega$ such that $\mathcal{P}_w^{(k)}(n) = O(f(n))$ but $\mathcal{P}_w^{(k)}(n)$ is not bounded.

Construction

- Let $n_1, n_2, \dots$ be a sequence of integers greater than 1.
- Let $m_j = n_1 \dots n_j$ for $j = 0, 1, 2, \dots$.
- Let $a_i = 0$ if the greatest j such that $m_j | i$ is even and $a_i = 1$ otherwise.
- Let $w = a_1 a_2 a_3 \dots$.
- The faster the sequence $n_1, n_2, \dots$ grows, the slower $\mathcal{U}_w^{(k)}(n)$ grows, but it is still unbounded.

Example

If $n_i = 2$ for every i, then $w = 010001010100\dots$ and $\mathcal{U}_w^{(2)}(n) = O(\log n)$.
If $n_i = 2^{2^i}$ for every i, then $\mathcal{U}_w^{(2)}(n) = O(\log \log n)$.

Outline

Constant Abelian Complexity

- There are uniformly recurrent words with constant Abelian complexity 3 (Richomme, Saari, Zamboni [6]).
- There are no recurrent words with constant Abelian complexity 4 (Currie, Rampersad [2]).
- For every c, there are uniformly recurrent words with ultimately constant Abelian complexity c (Saarela [7]).

Bounded Complexity and Powers

Theorem (Richomme, Saari, Zamboni [6])

An infinite word with bounded Abelian complexity contains arbitrarily high Abelian powers.

Theorem (Karhumäki, Saarela, Zamboni [3])

An infinite word with bounded k-Abelian complexity contains arbitrarily high k-Abelian powers.

- Can be proved using van der Waerden's theorem.
- Stronger version can be proved using Szemerédi's theorem.

Summary

- $\exists n : \mathcal{P}_w^{(k)}(n) < \min(n+1, 2k) \Rightarrow w$ ultimately periodic
- w Sturmian $\Leftrightarrow w$ aperiodic and $\forall n : \mathcal{P}_w^{(k)}(n) = \min(n+1, 2k)$
- For the Thue-Morse word T, $\mathcal{U}_T^{(2)}(m) = \Theta(\log n)$ and $\mathcal{L}_T^{(2)}(m) = 6$.
- There are uniformly recurrent words with arbitrarily slowly growing upper k-Abelian complexities.

Examples of open problems:

- For a morphic word w, how slowly can $\mathcal{U}_w^{(k)}(n)$ grow without being bounded? Can it grow slower than logarithmically?
- How high can the $(k+1)$-Abelian complexity of a k-Abelian periodic word be?

References I

Ethan M. Coven and G. A. Hedlund.
Sequences with minimal block growth.
Math. Systems Theory, 7:138–153, 1973.

James Currie and Narad Rampersad.
Recurrent words with constant Abelian complexity.
Adv. in Appl. Math., 47(1):116–124, 2011.

Juhani Karhumäki, Aleksi Saarela, and Luca Zamboni.
On a generalization of Abelian equivalence and complexity of infinite words, Submitted, arXiv preprint at http://arxiv.org/abs/1301.5104.

Juhani Karhumäki, Aleksi Saarela, and Luca Zamboni.
Variations of the Morse-Hedlund theorem for k-Abelian equivalence, Submitted, arXiv preprint at http://arxiv.org/abs/1302.3783.

References II

Marston Morse and Gustav A. Hedlund.
Symbolic dynamics.
Amer. J. Math., 60(4):815–866, 1938.

Gwenael Richomme, Kalle Saari, and Luca Q. Zamboni.
Abelian complexity of minimal subshifts.
J. Lond. Math. Soc. (2), 83(1):79–95, 2011.

Aleksi Saarela.
Ultimately constant abelian complexity of infinite words.
J. Autom. Lang. Comb., 14(3–4):255–258, 2009.

Welcome to the

9th international conference on
WORDS

September 16–20, 2013
Turku, Finland