Theoretical Computer Science 601 (2015) 47–57

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Another generalization of abelian equivalence: Binomial complexity of infinite words

M. Rigo [a,*], P. Salimov [a,b,1]

[a] Dept. of Math., University of Liège, Grande traverse 12 (B37), B-4000 Liège, Belgium
[b] Sobolev Institute of Math., 4 Acad. Koptyug avenue, 630090 Novosibirsk, Russia

ARTICLE INFO

Article history:
Received 7 January 2014
Accepted 21 October 2014
Available online 14 July 2015

Keywords:
Combinatorics on words
Abelian equivalence
Binomial coefficient
Factor complexity
Sturmian words
Thue–Morse word

ABSTRACT

The binomial coefficient of two words u and v is the number of times v occurs as a subsequence of u. Based on this classical notion, we introduce the m-binomial equivalence of two words refining the abelian equivalence. Two words x and y are m-binomially equivalent, if, for all words v of length at most m, the binomial coefficients of x and v and respectively, y and v are equal. The m-binomial complexity of an infinite word x maps an integer n to the number of m-binomial equivalence classes of factors of length n occurring in x. We study the first properties of m-binomial equivalence. We compute the m-binomial complexity of two classes of words: Sturmian words and (pure) morphic words that are fixed points of Parikh-constant morphisms like the Thue–Morse word, i.e., images by the morphism of all the letters have the same Parikh vector. We prove that the frequency of each symbol of an infinite recurrent word with bounded 2-binomial complexity is rational.

1. Introduction

In the literature, many measures of complexity of infinite words have been introduced. One of the most studied is the factor complexity p_x counting the number of distinct blocks of n consecutive letters occurring in an infinite word $x \in A^{\mathbb{N}}$ [8,9]. In particular, Morse–Hedlund theorem gives a characterization of ultimately periodic words in terms of bounded factor complexity. Sturmian words have a null topological entropy and are characterized by the relation $p_x(n) = n+1$ for all $n \geqslant 0$. Abelian complexity counts the number of distinct Parikh vectors for blocks of n consecutive letters occurring in an infinite word, i.e., factors of length n are counted up to abelian equivalence [19]. Related to Van der Waerden theorem, we can also mention the arithmetic complexity [2] mapping $n \geqslant 0$ to the number of distinct subwords $x_i x_{i+p} \cdots x_{i+(n-1)p}$ built from n letters arranged in arithmetic progressions in the infinite word x, $i \geqslant 0$, $p \geqslant 1$. In the same direction, one can also consider maximal pattern complexity [10].

As a generalization of abelian complexity, the k-abelian complexity was recently introduced through a hierarchy of equivalence relations, the coarsest being abelian equivalence and refining up to equality. We recall these notions.

Let $k \in \mathbb{N} \cup \{+\infty\}$ and A be a finite alphabet. As usual, $|u|$ denotes the length of u and $|u|_x$ denotes the number of occurrences of the word x as a factor of the word u. Karhumäki et al. [11] introduce the notion of *k-abelian equivalence* of finite words as follows. Let u, v be two words over A. We write $u \sim_{\mathrm{ab},k} v$ if and only if $|u|_x = |v|_x$ for all words x of length

* Corresponding author.
E-mail address: M.Rigo@ulg.ac.be (M. Rigo).
[1] The second author is supported by the Russian President's grant no. MK-4075.2012.1 and Russian Foundation for Basic Research grants no. 12-01-00089 and no. 11-01-00997 and by a University of Liège post-doctoral grant.

http://dx.doi.org/10.1016/j.tcs.2015.07.025

$|x| \leqslant k$. In particular, $u \sim_{\mathrm{ab},1} v$ means that u and v are *abelian equivalent*, i.e., u is obtained by permuting the letters in v. In that latter case, we also write $u \sim_{\mathrm{ab}} v$. Also, $u \sim_{ab,k+1} v$ trivially implies that $u \sim_{ab,k} v$.

The aim of this paper is to introduce and study the first properties of a different family of equivalence relations over A^*, called m-binomial equivalence, where the coarsest relation coincide again with the abelian equivalence.

Definition 1. Let $u = u_0 \cdots u_{n-1}$ be a word of length n over A. Let $\ell \leqslant n$. Let $s : \mathbb{N} \to \mathbb{N}$ be an increasing map such that $s(\ell - 1) < n$. Then the word $u_{s(0)} \cdots u_{s(\ell-1)}$ is a *subword* of length ℓ of u. Note that what we call subword is also called scattered subword (or scattered factor) in the literature. The notion of *binomial coefficient* of two finite words u and v is well-known, $\binom{u}{v}$ is defined as the number of times v occurs as a subword of u. In other words, the binomial coefficient of u and v is the number of times v appears as a subsequence of u.

Properties of these coefficients are presented in the chapter of Lothaire's book written by Sakarovitch and Simon [12, Section 6.3]. Let $a, b \in A$, $u, v \in A^*$ and p, q be integers. We set $\delta_{a,b} = 1$ if $a = b$, and $\delta_{a,b} = 0$ otherwise. We just recall that

$$\binom{a^p}{a^q} = \binom{p}{q}, \ \binom{u}{\varepsilon} = 1, \ |u| < |v| \Rightarrow \binom{u}{v} = 0, \ \binom{ua}{vb} = \binom{u}{vb} + \delta_{a,b}\binom{u}{v}$$

and the last three relations completely determine the binomial coefficient $\binom{u}{v}$ for all $u, v \in A^*$.

Remark 1. Note that we have to make a distinction between subwords and factors. A factor is a particular subword made of consecutive letters. Factors of u are denoted either by $u_i \cdots u_j$ or $u[i, j]$, $0 \leqslant i \leqslant j < |u|$.

Definition 2. Let $m \in \mathbb{N} \cup \{+\infty\}$ and u, v be two words over A. We say that u and v are *m-binomially equivalent* if

$$\binom{u}{x} = \binom{v}{x}, \ \forall x \in A^{\leqslant m}.$$

Since the main relation studied in this paper is the m-binomial equivalence, we simply write in that case: $u \sim_m v$.

Since $\binom{u}{a} = |u|_a$ for all $a \in A$, it is clear that two words u and v are abelian equivalent if and only if $u \sim_1 v$. As for abelian equivalence, we have a family of refined relations: for all $u, v \in A^*$, $m \geqslant 0$, $u \sim_{m+1} v \Rightarrow u \sim_m v$.

Example 1. For instance, the four words *ababbba*, *abbabab*, *baabbab* and *babaabb* are 2-binomially equivalent. For any w amongst these words, we have the following coefficients

$$\binom{w}{a} = 3, \ \binom{w}{b} = 4, \ \binom{w}{aa} = 3, \ \binom{w}{ab} = 7, \ \binom{w}{ba} = 5, \ \binom{w}{bb} = 6.$$

Let us show that the first two words are not 3-binomially equivalent. As an example, we have

$$\binom{ababbba}{aab} = 3 \text{ but } \binom{abbabab}{aab} = 4.$$

Indeed, for this last binomial coefficient, *aab* appears as subwords $w_0w_3w_4$, $w_0w_3w_6$, $w_0w_5w_6$ and $w_3w_5w_6$.

We now show that m-binomial equivalence and m-abelian equivalence are two different notions. Considering again the first two words (which are 2-binomially equivalent), we find $|ababbba|_{ab} = 2$ and $|abbabab|_{ab} = 3$, showing that these two words are not 2-abelian equivalent. Conversely, the words *abbaba* and *ababba* are 2-abelian equivalent but are not 2-binomially equivalent:

$$\binom{abbaba}{ab} = 4 \text{ but } \binom{ababba}{ab} = 5.$$

This paper is organized as follows. In the next section, we present some straightforward properties of binomial coefficients and m-binomial equivalence. In Section 3, we give upper bounds on the number of m-binomial equivalence classes partitioning A^n. Section 3 ends with the introduction of the m-binomial complexity $\mathbf{b}_x^{(m)} : \mathbb{N} \to \mathbb{N}$ of an infinite word x. In Section 4, we prove that if x is a Sturmian word then, for any $m \geqslant 2$,

$$\mathbf{b}_x^{(m)}(n) = n + 1 \quad \text{for all } n \geqslant 0.$$

This paper is an updated and extended version of the paper [21] presented during the WORDS conference in Turku, September 2013. The second half of this paper contains new material: In Section 5, we compute the m-binomial complexity of a family of (pure) morphic words. Namely, we show that fixed points of a morphism $\varphi : A^* \to A^*$ satisfying $\varphi(a) \sim_{\mathrm{ab}} \varphi(b)$,