Theoretical Computer Science 572 (2015) 83–91

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Avoiding 2-binomial squares and cubes

Michaël Rao [a], Michel Rigo [b,*], Pavel Salimov [b,c,1]

[a] CNRS, LIP, ENS Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France
[b] Dept of Math., University of Liège, Grande traverse 12 (B37), B-4000 Liège, Belgium
[c] Sobolev Institute of Math., 4 Acad. Koptyug avenue, 630090 Novosibirsk, Russia

A R T I C L E I N F O

Article history:
Received 10 September 2014
Received in revised form 15 January 2015
Accepted 17 January 2015
Available online 21 January 2015
Communicated by M. Crochemore

Keywords:
Combinatorics on words
Binomial coefficient
Binomial equivalence
Avoidance
Squarefree
Cubefree

A B S T R A C T

Two finite words u, v are 2-binomially equivalent if, for all words x of length at most 2, the number of occurrences of x as a (scattered) subword of u is equal to the number of occurrences of x in v. This notion is a refinement of the usual abelian equivalence. A 2-binomial square is a word uv where u and v are 2-binomially equivalent.
In this paper, considering pure morphic words, we prove that 2-binomial squares (resp. cubes) are avoidable over a 3-letter (resp. 2-letter) alphabet. The sizes of the alphabets are optimal.

1. Introduction

A *square* (resp. *cube*) is a non-empty word of the form xx (resp. xxx). Since the work of Thue, it is well-known that there exists an infinite squarefree word over a ternary alphabet, and an infinite cubefree word over a binary alphabet [13,14]. A main direction of research in combinatorics on words is about the avoidance of a pattern, and the size of the alphabet is a parameter of the problem.

A possible and widely studied generalization of squarefreeness is to consider an abelian framework. A non-empty word is an *abelian square* (resp. *abelian cube*) if it is of the form xy (resp. xyz) where y is a permutation of x (resp. y and z are permutations of x). Erdös raised the question whether abelian squares can be avoided by an infinite word over an alphabet of size 4 [3]. Keränen answered positively to this question, with a pure morphic word [9]. Moreover Dekking has previously obtained an infinite word over a 3-letter alphabet that avoids abelian cubes, and an infinite binary word that avoids abelian 4-powers [2]. (Note that in all these results, the size of the alphabet is optimal.)

In this paper, we are dealing with another generalization of squarefreeness and cubefreeness. We consider the 2-binomial equivalence which is a refinement of the abelian equivalence, i.e., if two words x and y are 2-binomially equivalent, then x is a permutation of y (but in general, the converse does not hold, see Example 1 below). This equivalence relation is defined thanks to the binomial coefficient $\binom{u}{v}$ of two words u and v which is the number of times v occurs as a subsequence of u

* Corresponding author.
E-mail addresses: michael.rao@ens-lyon.fr (M. Rao), M.Rigo@ulg.ac.be (M. Rigo).

1 The third 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.2015.01.029

(meaning as a "scattered" subword). For more on these binomial coefficients, see for instance [10, Chap. 6]. Based on this classical notion, the m-binomial equivalence of two words has been recently introduced [12].

Definition 1. Let $m \in \mathbb{N} \cup \{+\infty\}$ and u, v be two words over the alphabet A. We let $A^{\leq m}$ denote the set of words of length at most m over A. We say that u and v are *m-binomially equivalent* if

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

We simply write $u \sim_m v$ if u and v are m-binomially equivalent. The word u is obtained as a permutation of the letters in v if and only if $u \sim_1 v$. In that case, we say that u and v are *abelian equivalent* and we write instead $u \sim_{\mathrm{ab}} v$. Note that if $u \sim_{k+1} v$, then $u \sim_k v$, for all $k \geq 1$.

Example 1. The four words 0101110, 0110101, 1001101 and 1010011 are 2-binomially equivalent. Let u be any of these four words. We have

$$\binom{u}{0} = 3, \quad \binom{u}{1} = 4, \quad \binom{u}{00} = 3, \quad \binom{u}{01} = 7, \quad \binom{u}{10} = 5, \quad \binom{u}{11} = 6.$$

For instance, the word 0001111 is abelian equivalent to 0101110 but these two words are not 2-binomially equivalent. Let a be a letter. It is clear that $\binom{u}{aa}$ and $\binom{u}{a}$ carry the same information, i.e., $\binom{u}{aa} = \binom{|u|_a}{2}$ where $|u|_a$ is the number of occurrences of a in u.

A 2-*binomial square* (resp. 2-*binomial cube*) is a non-empty word of the form xy where $x \sim_2 y$ (resp. $x \sim_2 y \sim_2 z$). For instance, the prefix of length 12 of the Thue–Morse word: 011010011001 is a 2-binomial cube. Squares are avoidable over a 3-letter alphabet and abelian squares are avoidable over a 4-letter alphabet. Since 2-binomial equivalence lies between abelian equivalence and equality, the question is to determine whether or not 2-binomial squares are avoidable over a 3-letter alphabet. We answer positively to this question in Section 2. The fixed point of the morphism $g : 0 \mapsto 012$, $1 \mapsto 02$, $2 \mapsto 1$ avoids 2-binomial squares.

In a similar way, cubes are avoidable over a 2-letter alphabet and abelian squares are avoidable over a 3-letter alphabet. The question is to determine whether or not 2-binomial cubes are avoidable over a 2-letter alphabet. We also answer positively to this question in Section 3. The fixed point of the morphism $h : 0 \mapsto 001$, $1 \mapsto 011$ avoids 2-binomial cubes.

Remark 1. The m-binomial equivalence is not the only way to refine the abelian equivalence. Recently, a notion of m-abelian equivalence has been introduced [8]. To define this equivalence, one counts the number $|u|_x$ of occurrences in u of all factors x of length up to m (it is meant factors made of consecutive letters). That is, u and v are m-abelian equivalent if $|u|_x = |v|_x$ for all $x \in A^{\leq m}$. In that context, the results on avoidance are quite different. Over a 3-letter alphabet 2-abelian squares are unavoidable: the longest ternary word which is 2-abelian squarefree has length 537 [6], and pure morphic words cannot avoid k-abelian-squares for every k [7]. On the other hand, it has been shown that there exists a 3-abelian squarefree morphic word over a 3-letter alphabet [11]. Moreover 2-abelian-cubes can be avoided over a binary alphabet by a morphic word [11].

The number of occurrences of a letter a in a word u will be denoted either by $\binom{u}{a}$ or $|u|_a$. Let $A = \{0, 1, \dots, k\}$ be an alphabet. The *Parikh map* is an application $\Psi : A^* \to \mathbb{N}^{k+1}$ such that $\Psi(u) = (|u|_0, \dots, |u|_k)^T$. Note that we will deal with column vectors (when multiplying a square matrix with a column vector on its right). In particular, two words are abelian equivalent if and only if they have the same Parikh vector. The mirror of the word $u = u_1 u_2 \cdots u_k$ is denoted by $\widetilde{u} = u_k \cdots u_2 u_1$.

2. Avoiding 2-binomial squares over a 3-letter alphabet

Let $A = \{0, 1, 2\}$ be a 3-letter alphabet. Let $g : A^* \to A^*$ be the morphism defined by

$$g : \begin{cases} 0 \mapsto 012 \\ 1 \mapsto 02 \\ 2 \mapsto 1 \end{cases} \text{ and thus, } \quad g^2 : \begin{cases} 0 \mapsto 012021 \\ 1 \mapsto 0121 \\ 2 \mapsto 02. \end{cases}$$

It is prolongable on 0: $g(0)$ has 0 as a prefix. Hence the limit $\mathbf{x} = \lim_{n \to +\infty} g^n(0)$ is a well-defined infinite word

$$\mathbf{x} = g^\omega(0) = 012021012102012021020121 \cdots$$

which is a fixed point of g. Since the original work of Thue, this word $\mathbf{x}$ is well-known to avoid (usual) squares. It is sometimes referred to as the *ternary Thue–Morse word*. We will make use of the fact that $X = \{012, 02, 1\}$ is a prefix-code and thus an ω-code: Any finite word in X^* (resp. infinite word in X^ω) has a unique factorization as a product of elements in X. Let us make an obvious but useful observation.