Another generalization of abelian equivalence: binomial complexity of infinite words

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

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

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.

Keywords: combinatorics on words, abelian equivalence, binomial coefficient, factor complexity, Sturmian words, Thue-Morse word.

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,

*Corresponding author

[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.

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 $|x| \leqslant k$. 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 $\varphi(a) \sim_{\mathrm{ab}} \varphi(b)$, for all $a, b \in A$, have a bounded m-binomial complexity. In particular, our result can be applied to the Thue–Morse word t: For all $m \geqslant 1$, there exists a constant $C_{t,m}$ such that

$$\mathbf{b}_t^{(m)}(n) \leqslant C_{t,m} \quad \text{for all } n \geqslant 0.$$

The arguments presented here are much simpler than those sketched in [21]. Note that binomial coefficients of t were also considered in [4]. In Section 6, we study frequencies of symbols occurring in a recurrent infinite words with bounded 2-binomial complexity. Since such a word has a bounded abelian complexity, it is well known that these frequencies exist [1]. We show that they are moreover rational. Finally, in the last section of this paper, we consider the links existing with the well-known Parikh matrices [15, 16]. The quotient set $A^*/\sim_m$ can be equipped with a monoid structure which is isomorphic to a finitely generated monoid of integer matrices.

Note that questions about avoidance of 2-binomial squares and cubes are considered in a separate paper [18].

2. First properties of binomial equivalence

We denote by $\mathbf{B}^{(m)}(v)$ the equivalence class of words m-binomially equivalent to v. Binomial coefficients have a nice behavior with respect to the concatenation of words.

Proposition 1. *Let p, s and $e = e_0e_1 \cdots e_{n-1}$ be finite words. We have*

$$\binom{ps}{e} = \sum_{i=0}^{n} \binom{p}{e_0e_1 \cdots e_{i-1}} \binom{s}{e_ie_{i+1} \cdots e_{n-1}}.$$

We can also mention some other basic facts on m-binomial equivalence.

Lemma 2. *Let u, u', v, v' be finite words and $m \geqslant 1$.*

- *If $u \sim_m v$, then $u \sim_\ell v$ for all $\ell \leqslant m$.*
- *If $u \sim_m v$ and $u' \sim_m v'$, then $uu' \sim_m vv'$.*

Proof. Simply note for the second point that, for all $x = x_0 \cdots x_{\ell-1}$ of length $\ell \leqslant m$, $\binom{uu'}{x}$ is equal to

$$\sum_{i=0}^{\ell} \binom{u}{x[0, i-1]} \binom{u'}{x[i, \ell-1]} = \sum_{i=0}^{\ell} \binom{v}{x[0, i-1]} \binom{v'}{x[i, \ell-1]} = \binom{vv'}{x}.$$

□

Remark 2. Thanks to the above lemma, we can endow the quotient set $A^*/\sim_m$ with a monoid structure using an operation

$$\circ : A^*/\sim_m \times A^*/\sim_m \to A^*/\sim_m$$

defined by

$$\mathbf{B}^{(m)}(p) \circ \mathbf{B}^{(m)}(q) = \mathbf{B}^{(m)}(r)$$

if the concatenation (of languages) $\mathbf{B}^{(m)}(p).\mathbf{B}^{(m)}(q)$ is a subset of $\mathbf{B}^{(m)}(r)$. In particular, one can take $r = pq$. More details are given in Section 7.

We will often make use of the following fact: If a word v is factorized as $v = pus$, then the equivalence class $\mathbf{B}^{(m)}(v)$ is completely determined by p, s and $\mathbf{B}^{(m)}(u)$. Indeed, if $u \sim_m u'$ then $pus \sim_m pu's$.

3. On the number of m-binomial equivalence classes

For 2- and 3-abelian equivalence, the number of equivalence classes for words of length n over a binary alphabet are respectively $n^2 - n + 2$ and $\Theta(n^4)$. In general, for k-abelian equivalence, the number of equivalence classes for words of length n over a ℓ-letter alphabet is $\Theta(n^{(\ell-1)\ell^{k-1}})$ [11].

We consider similar results for m-binomial equivalence.

Lemma 3. *Let $u \in A^*$, $a \in A$ and $\ell \geqslant 0$. We have*

$$\binom{u}{a^\ell} = \binom{|u|_a}{\ell} \quad \text{and} \quad \sum_{|v|=\ell} \binom{u}{v} = \binom{|u|}{\ell}.$$

We let $\#B$ denote the cardinality of a set B.

Lemma 4. *Let A be a binary alphabet, we have*

$$\#\left(A^n/\!\sim_2\right) = \sum_{j=0}^{n}((n-j)j+1) = \frac{n^3+5n+6}{6}.$$

Proof. Let $A = \{a, b\}$. The set A^n is split into $n+1$ equivalence classes for the abelian equivalence. Let $j \in \{0, \ldots, n\}$. Consider a representative u of such an abelian equivalence class characterized by $|u|_a = j$ and $|u|_b = n - j$. The extremal values taken by $\binom{u}{ab}$ are

$$\binom{b^{n-j}a^j}{ab} = 0 \text{ and } \binom{a^j b^{n-j}}{ab} = j(n-j).$$

Now we show that, for all $k \in \{0, \ldots, j(n-j)\}$, there exists a word w abelian equivalent to u and such that $\binom{w}{ab} = k$. One has simply to consider the $j(n-j)+1$ words $b^{n-j}a^j$, $b^{n-j-1}aba^{j-1}$, $b^{n-j-1}a^2ba^{j-2}$, $\ldots$, $b^{n-j-1}a^jb$, $b^{n-j-2}aba^{j-1}b$, $b^{n-j-2}a^2ba^{j-2}b$, $\ldots$, $b^{n-j-2}a^jb^2$, $\ldots$, a^jb^{n-j}. To conclude the proof, since $\binom{w}{aa}$ and $\binom{w}{bb}$ are determined by the abelian class of w and using Lemma 3

$$\binom{w}{ba} = \binom{|w|}{2} - \binom{w}{aa} - \binom{w}{bb} - \binom{w}{ab}$$

then, the coefficient $\binom{w}{ba}$ for a word w abelian equivalent to u is deduced from $\binom{w}{ab}$. Hence the abelian equivalence class containing $b^{n-j}a^j$ is split into $(n-j)j+1$ classes for the 2-binomial equivalence. □

Proposition 5. *Let $m \geqslant 2$. Let A be a binary alphabet, we have*

$$\#\left(A^n/\!\sim_m\right) \in \mathcal{O}(n^{2((m-1)2^m+1)}).$$

Proof. Let u be a word of length n. Let $\ell \leqslant m$. The number of subwords of length ℓ occurring in u is $\binom{n}{\ell}$. There are exactly 2^ℓ words of length ℓ enumerated lexicographically: $v_{\ell,1}, \dots, v_{\ell,2^\ell}$. Consider the vector $\Psi_\ell(u)$ of size 2^ℓ given by

$$\Psi_\ell(u) := \left(\binom{u}{v_{\ell,1}} \quad \cdots \quad \binom{u}{v_{\ell,2^\ell}}\right).$$

If u and u' are two words of length n such that $\Psi_\ell(u) \neq \Psi_\ell(u')$, then $u \not\sim_\ell u'$ and thus $u \not\sim_m u'$. The number of values taken by $\Psi_\ell(u) \in \mathbb{N}^{2^\ell}$ is bounded by the number of ways to partition the integer $\binom{n}{\ell}$ as a sum of 2^ℓ non-negative integers, that is $(\binom{n}{\ell} + 1)^{2^\ell - 1}$. Hence, we get

$$\#\left(A^n/\sim_m\right) \leqslant \prod_{\ell=1}^{m} (\binom{n}{\ell} + 1)^{2^\ell - 1}.$$

The upper bound is obtained by replacing $(\binom{n}{\ell} + 1)^{2^\ell - 1}$ with $(n^\ell)^{2^\ell}$. □

We denote by $\mathrm{Fac}_x(n)$ the set of factors of length n occurring in x.

Definition 3. Let $m \geqslant 1$. The *m-binomial complexity* of an infinite word x counts the number of m-binomial equivalence classes of factors of length n occurring in x,

$$\mathbf{b}_x^{(m)} : \mathbb{N} \to \mathbb{N},\ n \mapsto \#(\mathrm{Fac}_x(n)/\sim_m).$$

Note that $\mathbf{b}_x^{(1)}$ corresponds to the usual abelian complexity denoted by ρ_x^{ab}.

If p_x denotes the usual factor complexity, then for all $m \geqslant 1$, we have

$$\mathbf{b}_x^{(m)}(n) \leqslant \mathbf{b}_x^{(m+1)}(n) \quad \text{and} \quad \rho_x^{\mathrm{ab}}(n) \leqslant \mathbf{b}_x^{(m)}(n) \leqslant p_x(n). \tag{1}$$

4. The m-binomial complexity of Sturmian words

Recall that a *Sturmian word* x is a non-periodic word of minimal (factor) complexity, that is, $p_x(n) = n + 1$ for all $n \geqslant 0$. The following characterization is also useful.

Theorem 6. *[13, Theorem 2.1.5] An infinite word $x \in \{0,1\}^\omega$ is Sturmian if and only if it is aperiodic and balanced, i.e., for all factors u, v of the same length occurring in x, we have $||u|_1 - |v|_1| \leqslant 1$.*

The aim of this section is to compute the m-binomial complexity of a Sturmian word as expressed by Theorem 7. We show that any two distinct factors of length n occurring in a Sturmian words are never m-binomially equivalent. First note that Sturmian words have a constant abelian complexity. Hence, if x is a Sturmian word, then $\mathbf{b}_x^{(1)}(n) = 2$ for all $n \geqslant 1$.

Theorem 7. *Let $m \geqslant 2$. If x is a Sturmian word, then $\mathbf{b}_x^{(m)}(n) = n + 1$ for all $n \geqslant 0$.*

Remark 3. If x is a right-infinite word such that $\mathbf{b}_x^{(1)}(n) = 2$ for all $n \geqslant 1$, then x is clearly balanced. If $\mathbf{b}_x^{(2)}(n) = n+1$, for all $n \geqslant 0$, then the factor complexity function p_x is unbounded and x is aperiodic. As a consequence of Theorem 7, an infinite word x is Sturmian if and only if, for all $n \geqslant 1$ and all $m \geqslant 2$, $\mathbf{b}_x^{(1)}(n) = 2$ and $\mathbf{b}_x^{(m)}(n) = n+1$.

Before proceeding to the proof of Theorem 7, we first recall some well-known facts about Sturmian words. One of the two symbols occurring in a Sturmian word x over $\{0,1\}$ is always isolated, for instance, 1 is always followed by 0. In that latter case, there exists a unique $k \geqslant 1$ such that each occurrence of 1 is always followed by either 0^k1 or $0^{k+1}1$ and x is said to be of *type* 0. See for instance [17, Chapter 6]. More precisely, we have the following remarkable fact showing that the recoding of a Sturmian sequence corresponds to another Sturmian sequence. Note that $\sigma : A^\omega \to A^\omega$ is the shift operator mapping $(x_n)_{n\geqslant 0}$ to $(x_{n+1})_{n\geqslant 0}$.

Theorem 8. *Let $x \in \{0,1\}^\omega$ be a Sturmian word of type 0. There exists a unique integer $k \geqslant 1$ and a Sturmian word $y \in \{0,1\}^\omega$ such that $x = \sigma^c(\mu(y))$ for some $c \leqslant k+1$ and where the morphism $\mu : \{0,1\}^* \to \{0,1\}^*$ is defined by $\mu(0) = 0^k1$ and $\mu(1) = 0^{k+1}1$.*

Corollary 9. *Let $x \in \{0,1\}^\omega$ be a Sturmian word of type 0. There exists a unique integer $k \geqslant 1$ such that any factor occurring in x is of the form*

$$0^r10^{k+\epsilon_0}10^{k+\epsilon_1}1\cdots 0^{k+\epsilon_{n-1}}10^s \tag{2}$$

where $r, s \leqslant k+1$ and $\epsilon_0\epsilon_1\cdots\epsilon_{n-1} \in \{0,1\}^*$ is a factor of the Sturmian word y introduced in the above theorem.

Let $\epsilon = \epsilon_0\cdots\epsilon_{n-1}$ be a word over $\{0,1\}$. For $m \leqslant n-1$, we define

$$S(\epsilon, m) := \sum_{j=0}^{m}(n-j)\epsilon_j \quad \text{and} \quad S(\epsilon) := S(\epsilon, n-1). \tag{3}$$

Remark 4. Let $v = 0^r10^{k+\epsilon_0}10^{k+\epsilon_1}1\cdots 0^{k+\epsilon_{n-1}}10^s$ of the form (2), we have

$$\binom{v}{01} = r(n+1) + \sum_{j=0}^{n-1}(k+\epsilon_j)(n-j) = r(n+1) + S(\epsilon_0\cdots\epsilon_{n-1}) + k\frac{n(n+1)}{2}.$$

We need a technical lemma on the factors of a Sturmian word.

Lemma 10. *Let $n \geqslant 1$. If u and v are two distinct factors of length n occurring in a Sturmian word over $\{0,1\}$, then $S(u) \not\equiv S(v) \pmod{n+1}$.*

Proof. Consider two distinct factors u, v of length n occurring in a Sturmian word y. For $m < n$, we define $\Delta(m) := |u_0u_1\cdots u_m|_1 - |v_0v_1\cdots v_m|_1$. Due to Theorem 8, we have $|\Delta(m)| \leqslant 1$. Note that, if there exists i such that

$\Delta(i) = 1$ then, for all $j > i$, we have $\Delta(j) \geqslant 0$. Otherwise, we would have $|v[i+1,j]|_1 - |u[i+1,j]|_1 > 1$ contradicting the fact that y is balanced. Similarly, for all $j < i$, we also have $\Delta(j) \geqslant 0$.

Since u and v are distinct, replacing u with v if needed, we may assume that there exists a minimal $i \in \{0, \ldots, n-1\}$ such that $\Delta(i) = 1$. From the above discussion and the minimality of i, $\Delta(j) = 0$ for $j < i$ and $\Delta(j) \in \{0, 1\}$ for $j > i$.

From (3), for any $j < n$, we have

$$\begin{aligned}
\Delta(j+1) > \Delta(j) \quad &\Rightarrow \quad S(u, j+1) - S(v, j+1) = S(u, j) - S(v, j) + (n - j) \\
\Delta(j+1) = \Delta(j) \quad &\Rightarrow \quad S(u, j+1) - S(v, j+1) = S(u, j) - S(v, j) \\
\Delta(j+1) < \Delta(j) \quad &\Rightarrow \quad S(u, j+1) - S(v, j+1) = S(u, j) - S(v, j) - (n - j).
\end{aligned}$$

In view of these observations, the knowledge of $\Delta(0), \Delta(1), \ldots$ permits to compute $(S(u, j) - S(v, j))_{0 \leqslant j < n}$ and we deduce that $0 < S(u) - S(v) < n + 1$ concluding the proof. □

Proof (Proof of Theorem 7). Let x be a Sturmian word of type 0 and $m \geqslant 2$. From (1), we have, for all $\ell \geqslant 0$,

$$\mathbf{b}_x^{(2)}(\ell) \leqslant \mathbf{b}_x^{(m)}(\ell) \leqslant p_x(\ell) = \ell + 1.$$

We just need to show that any two distinct factors of length ℓ in x are not 2-binomially equivalent, i.e., $\ell + 1 \leqslant \mathbf{b}_x^{(2)}(\ell)$.

Proceed by contradiction. Assume that x contains two distinct factors u and v that are 2-binomially equivalent. In particular, $\binom{u}{00} = \binom{v}{00}$ and $\binom{u}{11} = \binom{v}{11}$. Hence we get $|u| = |v|$ and $|u|_1 = |v|_1 = n$. From Corollary 9, there exist $k \geqslant 1$ and a Sturmian word y such that

$$u = 0^r 1 0^{k+\epsilon_0} 1 0^{k+\epsilon_1} 1 \cdots 0^{k+\epsilon_{n-1}} 1 0^s, \quad v = 0^{r'} 1 0^{k+\epsilon'_0} 1 0^{k+\epsilon'_1} 1 \cdots 0^{k+\epsilon'_{n-1}} 1 0^{s'}$$

where $\epsilon = \epsilon_0 \epsilon_1 \cdots \epsilon_{n-1}$ and $\epsilon' = \epsilon'_0 \epsilon'_1 \cdots \epsilon'_{n-1}$ are both factors of y.

Since $u \sim_2 v$, it follows $\binom{u}{01} = \binom{v}{01}$. From Remark 4, we get

$$r(n+1) + S(\epsilon) + k\frac{n(n+1)}{2} = r'(n+1) + S(\epsilon') + k\frac{n(n+1)}{2}.$$

Otherwise stated, we get $S(\epsilon) - S(\epsilon') = (r' - r)(n+1)$ contradicting the previous lemma. □

5. Fixed points of Parikh-constant morphisms

The *Thue–Morse* word $t = 0110100110010110100101100110100 1 \cdots$ is the infinite word $\lim_{n\to\infty} \mu^n(a)$ where $\mu : 0 \mapsto 01,\ 1 \mapsto 10$. The factor complexity of the Thue–Morse word is well-known [3, 5, 7]: $p_t(0) = 1$, $p_t(1) = 2$, $p_t(2) = 4$ and

$$p_t(n) = \begin{cases} 4n - 2 \cdot 2^m - 4 & \text{if } 2 \cdot 2^m < n \leqslant 3 \cdot 2^m \\ 2n + 4 \cdot 2^m - 2 & \text{if } 3 \cdot 2^m < n \leqslant 4 \cdot 2^m. \end{cases}$$

Since t is abelian periodic, i.e., t is the concatenation of abelian equivalent factors 01 and 10, then the abelian complexity of t is obvious.

Lemma 11. *We have* $\mathbf{b}_t^{(1)}(2n) = 3$ *and* $\mathbf{b}_t^{(1)}(2n+1) = 2$ *for all* $n \geqslant 1$.

On the other hand, its arithmetical complexity is maximal which means that we have $a_t(n) = 2^n$ for all $n \geqslant 1$. See [2].

The results of this section are quite contrasting with the Sturmian case discussed in Section 4. We will see that some infinite words like the Thue–Morse word exhibit a bounded m-binomial complexity. We will consider not only the Thue–Morse word but a family of morphic words defined as follows.

Definition 4. Let $\varphi : A^* \to A^*$ be a morphism. If $\varphi(a) \sim_{\mathrm{ab}} \varphi(b)$, for all $a, b \in A$, then φ is said to be *Parikh-constant*. In particular, a Parikh-constant morphism is ℓ*-uniform* for some ℓ, i.e., there exists ℓ such that, for all $a \in A$, $|\varphi(a)| = \ell$.

It is clear that the morphism μ generating the Thue–Morse word is Parikh-constant. So the following results can be applied to the Thue–Morse word.

The key statement is the following one: for a Parikh-constant morphism φ, we have $\varphi^k(a) \sim_k \varphi^k(b)$ for all $a, b \in A$ and $k \geqslant 1$.

Lemma 12. *Let* $\varphi : A^* \to A^*$ *be a Parikh-constant morphism. Let* $k \geqslant 1$. *We have* $\mathbf{B}^{(k)}(\varphi^k(a)) = \mathbf{B}^{(k)}(\varphi^k(b))$ *for all* $a, b \in A$.

PROOF. We proceed by induction on k. The case $k = 1$ is simply a reformulation of the definition of a Parikh-constant morphism. Assume that, for some $k \geqslant 1$, $\mathbf{B}^{(k)}(\varphi^k(a)) = \mathbf{B}^{(k)}(\varphi^k(b))$ for all $a, b \in A$. We have to prove that, for all $a, b \in A$,

$$\mathbf{B}^{(k+1)}(\varphi^{k+1}(a)) = \mathbf{B}^{(k+1)}(\varphi^{k+1}(b)).$$

Since φ is Parikh-constant, there exists ℓ such that φ is ℓ-uniform. Let $a \in A$. There exists a word $v = v_0 \cdots v_{\ell-1}$ of length ℓ such that $\varphi(a) = v$. The main argument is the following one. Let u be a word of length at most $k+1$. We have

$$\begin{aligned} \binom{\varphi^{k+1}(a)}{u} &= \binom{\varphi^k(v)}{u} \\ &= \sum_{0 \leqslant i \leqslant \ell-1} \binom{\varphi^k(v_i)}{u} + \sum_{\substack{e_1 \cdots e_p = u \\ e_1, \ldots, e_p \in A^+ \\ p \geqslant 2}} \sum_{0 \leqslant i_1 < \cdots < i_p \leqslant \ell-1} \prod_{j=1}^{p} \binom{\varphi^k(v_{i_j})}{e_j}. \end{aligned}$$

Indeed, the first term of the sum counts the occurrences of u as a subword of $\varphi^k(v_i)$ for some i. The second term of the sum counts the occurrences of u that are split amongst several of the $\varphi^k(v_i)$'s as sketched in Figure 1: we consider all the factorizations of u into an arbitrary number of $p \geqslant 2$ non-empty factors $e_1, \ldots, e_p$ and how these factors can occur as a subword in the factors $\varphi^k(v_0), \ldots, \varphi^k(v_{\ell-1})$. Note that since $p \geqslant 2$, $|e_i| \leqslant k$ for all i. Let $b \in A$. There

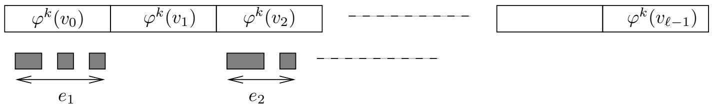

Figure 1: One of the possible occurrence of u in $\varphi^{k+1}(a)$.

exists a word $v' = v'_0 \cdots v'_{\ell-1}$ of length ℓ such that $\varphi(b) = v'$. To conclude with the proof, observe that the induction hypothesis implies that

$$\binom{\varphi^k(v_{i_j})}{e_j} = \binom{\varphi^k(v'_{i_j})}{e_j}$$

for any word e_j of length at most k. Moreover, since $v_0 \cdots v_{\ell-1} \sim_{\mathrm{ab}} v'_0 \cdots v'_{\ell-1}$, then

$$\sum_{0 \leqslant i \leqslant \ell-1} \binom{\varphi^k(v_i)}{u} = \sum_{0 \leqslant i \leqslant \ell-1} \binom{\varphi^k(v'_i)}{u}.$$

We conclude that

$$\binom{\varphi^{k+1}(a)}{u} = \binom{\varphi^{k+1}(b)}{u}.$$

□

Theorem 13. *Let x be an infinite word that is a fixed point of a Parikh-constant morphism. Let $m \geqslant 2$. There exists a constant $C_{x,m} > 0$ (depending only on x and m) such that the m-binomial complexity of x satisfies*

$$\mathbf{b}_x^{(m)}(n) \leqslant C_{x,m} \quad \textit{for all } n \geqslant 0.$$

PROOF. Let $\varphi : A^* \to A^*$ be a Parikh-constant morphism that is ℓ-uniform for some ℓ. Let x be an infinite word that is a fixed point of φ. Let $m \geqslant 2$. Notice that

$$|\varphi^m(a)| = \ell^m, \quad \forall a \in A.$$

Since $x = x_0x_1x_2 \cdots$ is also a fixed point of φ^m. It can be factorized as

$$x = \varphi^m(x_0)\varphi^m(x_1)\varphi^m(x_2) \cdots .$$

Let $n \geqslant 1$. Any factor v of length n occurring in x can therefore be written as

$$v = p\,\varphi^m(y)\,s$$

where p is a suffix of some $\varphi^m(x_i)$, $y = x_{i+1} \cdots x_j$ and s is a prefix of $\varphi^m(x_{j+1})$. In particular, we have $0 \leqslant |p| < \ell^m$, $0 \leqslant |s| < \ell^m$. Also, it is important to notice that, for a given n, in such a factorization $|y|$ can take at most two values: $\lfloor n/\ell^m \rfloor$ or $\lfloor n/\ell^m \rfloor - 1$.

Let $y = y_1 \cdots y_r$ and $y' = y'_1 \cdots y'_r$ be two words of the same length r. We have $\varphi^m(y) = \varphi^m(y_1) \cdots \varphi^m(y_r)$ and $\varphi^m(y') = \varphi^m(y'_1) \cdots \varphi^m(y'_r)$. Since the morphism φ is Parikh-constant, we can apply the previous lemma:

$$\varphi^m(y_i) \sim_m \varphi^m(y'_i)$$

for all i. Hence, we conclude with Lemma 2 that $\varphi^m(y) \sim_m \varphi^m(y')$ for any two words y, y' of the same length.

As already observed in Remark 2, the equivalence class of $v = p\,\varphi^m(y)\,s$ depends only on p, s and the equivalence class of $\varphi^m(y)$. But p and s belong to some finite set (they have a length bounded by a constant). For a given n, with the above discussion, the number of possible equivalence classes for $\varphi^m(y)$ is at most 2.

In particular, x_i (resp. x_{j+1}) can take $\#A$ values, p which is a suffix of $\varphi^m(x_i)$ (resp. s which is a prefix of $\varphi^m(x_{j+1})$) has a length between 0 and $\ell^m - 1$. Hence, the constant $C_{x,m}$ is less or equal to $2(\#A.\ell^m)^2$. □

To conclude this section, the next corollary is just a special case of Theorem 13.

Corollary 14. *Let $m \geqslant 2$. There exists a constant $C_{t,m} > 0$ such that the m-binomial complexity of the Thue–Morse word satisfies $\mathbf{b}_t^{(m)}(n) \leqslant C_{t,m}$ for all $n \geqslant 0$.*

Remark 5. By computer experiments, $\mathbf{b}_t^{(2)}(n)$ is equal to 9 if $n \equiv 0 \pmod 4$ and to 8 otherwise, for $10 \leqslant n \leqslant 1000$. Moreover, $\mathbf{b}_t^{(3)}(n)$ is equal to 21 if $n \equiv 0 \pmod 8$ and to 20 otherwise, for $8 \leqslant n \leqslant 500$.

6. Frequencies of symbols

In this section, we prove the following result: If an infinite recurrent word has bounded 2-binomial complexity, then the frequency of each symbol occurring in x exists and is a rational number.

Let $C > 0$. Recall that an infinite word x over A is *C-balanced* if, for all factors u, v in x of the same length, $|\,|u|_a - |v|_a\,| \leqslant C$ for all letters $a \in A$.

Lemma 15. *[20] An infinite word has bounded abelian complexity if and only if it is C-balanced for some $C > 0$.*

Definition 5. Recall that the *frequency* of occurrence of a letter $a \in A$ in an infinite word $x = x_0x_1x_2\cdots$ is defined as

$$\lim_{n\to+\infty} \frac{|x_0\cdots x_{n-1}|_a}{n}.$$

But such a limit does not always exists.

Remark 6. In [1, Prop. 7], Adamczewski shows that if a word is C-balanced for some $C > 0$, then the frequency of each symbol exists. Consequently, if an infinite word x over A has bounded abelian complexity, then the frequency of each symbol exists and is denoted by Λ_a.

From (1), we know that $\rho_x^{\mathrm{ab}}(n) \leqslant \mathbf{b}_x^{(2)}(n)$ for all n. Hence, if the 2-binomial complexity of x is bounded, this implies, on one hand, that the frequency of each symbol exists, but, on the other hand, also offers some more information:

Theorem 16. *If an infinite recurrent word x has bounded 2-binomial complexity, then the frequency of each symbol occurring in x exists and is rational.*

Proof. Without loss of generality, we can restrict ourselves to the case of a binary alphabet. Indeed, assume that x is an infinite word over an arbitrary alphabet A. If we are interested in the frequency Λ_a of a particular letter $a \in A$, we can apply the morphism $h : A^* \to \{0,1\}^*$ which maps a to 0 and all the other letters to 1. It is clear that

$$\mathbf{b}_{h(x)}^{(2)}(n) \leqslant \mathbf{b}_x^{(2)}(n), \quad \forall n \geqslant 0$$

and the frequencies of a in x or 0 in $h(x)$ are the same and will be denoted Λ_0.

From now on, we assume that x is an infinite word over $\{0,1\}$. Assume that the 2-binomial complexity $\mathbf{b}_x^{(2)}$ is bounded by c. Let pvp be a prefix of x where p is assumed to be long enough to guarantee $|p| > c$. Such a prefix pvp always occur because x is recurrent and thus p occurs twice. We set $n = |pv|$.

For $i = 0, \ldots, |p| - 1$, we define the words

$$u^{(i)} = p_i p_{i+1} \cdots p_{|p|-1} v p_0 p_1 \cdots p_{i-1}.$$

In particular, we have $u^{(0)} = pv$. Since $u^{(i+1)}$ is a cyclic shift of $u^{(i)}$ by one letter, we have the following relations:

$$\begin{aligned}\binom{u^{(i+1)}}{01} &= \binom{u^{(i)}}{01} - |pv|_1, \text{ if } p_i = 0;\\ \binom{u^{(i+1)}}{01} &= \binom{u^{(i)}}{01} + |pv|_0, \text{ if } p_i = 1.\end{aligned}$$

For $i > 0$, applying i times these relations yields

$$\binom{u^{(i)}}{01} = \binom{u^{(0)}}{01} + c_{0,i}|pv|_0 - c_{1,i}|pv|_1$$

for some integers $c_{0,i}, c_{1,i} \in \{0, \ldots, i\}$. Note that $c_{0,i} + c_{1,i} = i$. Since the 2-binomial complexity is bounded by c, thanks to Pigeonhole principle, there exist $j, k \leqslant c$ with $j \neq k$ such that

$$\binom{u^{(j)}}{01} = \binom{u^{(k)}}{01}.$$

Therefore, we can write

$$\binom{u^{(0)}}{01} + c_{0,j}|pv|_0 - c_{1,j}|pv|_1 = \binom{u^{(j)}}{01} = \binom{u^{(k)}}{01} = \binom{u^{(0)}}{01} + c_{0,k}|pv|_0 - c_{1,k}|pv|_1.$$

Otherwise stated, there exist integers $c_0, c_1 \leqslant c$ such that

$$c_0|pv|_0 + c_1|pv|_1 = 0. \qquad (4)$$

Note that $c_0 \neq c_1$. Recall that $|pv| = n$. Obviously, we have

$$|pv|_0 + |pv|_1 - n(\Lambda_0 + 1 - \Lambda_0) = 0.$$

From Lemma 15 and Remark 6, we already know that the frequency Λ_0 exists. This means that $|pv|_0/n$ tends to Λ_0 as n tends to infinity. Hence, we can write $|pv|_0 - n\Lambda_0 = \delta_n$ and $|pv|_1 - n(1 - \Lambda_0) = -\delta_n$ where δ_n/n tends to zero as n tends to infinity. Substituting $|pv|_0$ and $|pv|_1$ in (4), we get

$$c_0(n\Lambda_0 + \delta_n) + c_1(n(1 - \Lambda_0) - \delta_n) = 0.$$

Finally, we get

$$\Lambda_0 = \frac{c_1}{c_1 - c_0} - \frac{\delta_n}{n}\,.$$

Letting n tends to infinity, leads us to the conclusion: Λ_0 is equal to the rational $c_1/(c_1 - c_0)$. □

Remark 7. For instance, the Tribonacci word r (which is the fixed point of the primitive morphism $a \mapsto ab, b \mapsto ac, c \mapsto a$) has frequencies of letters that are not rational [14, Chapter 10]. Therefore, this implies that $\mathbf{b}_r^{(2)}$ cannot be bounded by a constant.

About rational frequencies, it is also well-known that if the frequencies of symbol exist for a k-automatic sequence, i.e., images under a coding of the fixed point of a k-uniform morphism, then these frequencies are rational [6].

7. Parikh matrices

Parikh matrices are a well-know tools to deal with subwords [15, 16]. Such a matrix associated with a word v contains information on the number of occurrences of some subwords of v. There is a vast literature on the subject. We recall the definition of these matrices. Let $A = \{a_1, \dots, a_k\}$ be a finite alphabet. We define a morphism Ψ_{M_k} from A^* to the set of square matrices of size $k+1$ as follows.

$$\Psi_{M_k}(a_q) = (m_{i,j})_{1 \leqslant i,j \leqslant k+1}$$

where $m_{i,i} = 1$ for $1 \leqslant i \leqslant k+1$, $m_{q,q+1} = 1$ and all the other elements in $\Psi_{M_k}(a_q)$ are zero. Therefore, $\Psi(v_0 \cdots v_{\ell-1}) = \Psi(v_0) \cdots \Psi(v_\ell - 1)$ where on the right hand side, we consider the usual matrix multiplication.

Example 2. Consider a 3-letter alphabet $A = \{a_1, a_2, a_3\}$. We have

$$\Psi_{M_3} : a_1 \mapsto \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad a_2 \mapsto \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad a_3 \mapsto \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

For a word w, one can prove by induction on the length of w that

$$\Psi_{M_3}(w) = \begin{pmatrix} 1 & |w|_{a_1} & \binom{w}{a_1a_2} & \binom{w}{a_1a_2a_3} \\ 0 & 1 & |w|_{a_2} & \binom{w}{a_2a_3} \\ 0 & 0 & 1 & |w|_{a_3} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

As an example, we have

$$\Psi_{M_3}(a_1a_1a_2a_3a_2a_1a_2) = \begin{pmatrix} 1 & 3 & 7 & 2 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In the following lines, we consider a seemingly similar construction. Let $u = u_0u_1 \dots u_{m-1}$ be a word of length m. For any word v, we define a $(m+1)\times(m+1)$ matrix $M_u(v)$ given by

$$M_u(v) = \begin{pmatrix} 1 & \binom{v}{u_0u_1\cdots u_{m-1}} & \binom{v}{u_0u_1\cdots u_{m-2}} & \dots & \binom{v}{u_0u_1} & \binom{v}{u_0} \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & \binom{v}{u_{m-1}} & 1 & \ddots & 0 & 0 \\ 0 & \binom{v}{u_{m-2}u_{m-1}} & \binom{v}{u_{m-2}} & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \binom{v}{u_2u_3\cdots u_{m-1}} & \binom{v}{u_2u_1\cdots u_{m-2}} & \dots & 1 & 0 \\ 0 & \binom{v}{u_1u_2\cdots u_{m-1}} & \binom{v}{u_1u_1\cdots u_{m-2}} & \dots & \binom{v}{u_1} & 1 \end{pmatrix}.$$

Proposition 17. *For all words v, w, we have $M_u(vw) = M_u(v)M_u(w)$.*

Proof. It follows directly from Lemma 2 and its proof. □

Example 3. Consider the word $u = 01$. We get

$$M_u(v)M_v(w) = \begin{pmatrix} 1 & \binom{v}{01} & \binom{v}{0} \\ 0 & 1 & 0 \\ 0 & \binom{v}{1} & 1 \end{pmatrix} \begin{pmatrix} 1 & \binom{w}{01} & \binom{w}{0} \\ 0 & 1 & 0 \\ 0 & \binom{w}{1} & 1 \end{pmatrix}$$

and

$$M_u(v)M_v(w) = \begin{pmatrix} 1 & \binom{w}{01} + \binom{v}{01} + \binom{v}{0}\binom{w}{1} & \binom{v}{0} + \binom{w}{0} \\ 0 & 1 & 0 \\ 0 & \binom{v}{1} + \binom{w}{1} & 1 \end{pmatrix} = \begin{pmatrix} 1 & \binom{vw}{01} & \binom{vw}{0} \\ 0 & 1 & 0 \\ 0 & \binom{vw}{1} & 1 \end{pmatrix}.$$

Let m be an integer. For any word v, we define the block diagonal matrix

$$M_m(v) = \begin{pmatrix} M_{u^{(1)}}(v) & 0 & \dots & 0 \\ 0 & M_{u^{(2)}}(v) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & M_{u^{(\#A^m)}}(v) \end{pmatrix},$$

where $u^{(1)}, u^{(2)}, \ldots, u^{(\#A^m)}$ are all the words of length m over A that have been lexicographically ordered.

Example 4. For $m = 2$, $M_2(v)$ is a 12×12 block diagonal matrix made of the following four diagonal blocks:

$$\begin{pmatrix} 1 & \binom{v}{00} & \binom{v}{0} \\ 0 & 1 & 0 \\ 0 & \binom{v}{0} & 1 \end{pmatrix}, \begin{pmatrix} 1 & \binom{v}{01} & \binom{v}{0} \\ 0 & 1 & 0 \\ 0 & \binom{v}{1} & 1 \end{pmatrix}, \begin{pmatrix} 1 & \binom{v}{10} & \binom{v}{1} \\ 0 & 1 & 0 \\ 0 & \binom{v}{0} & 1 \end{pmatrix}, \begin{pmatrix} 1 & \binom{v}{11} & \binom{v}{1} \\ 0 & 1 & 0 \\ 0 & \binom{v}{1} & 1 \end{pmatrix}$$

Corollary 18. *For all words v, w, we have $M_m(vw) = M_m(v)M_m(w)$. The monoid $\langle A^*/\sim_m, \circ\rangle$ introduced in Remark 2 is isomorphic to the submonoid of matrices generated by $\{M_m(a) \mid a \in A\} \cup \{I\}$.*

Recall that it is licit to define a product $\circ$ over the quotient set $A^*/\sim_m$ by setting

$$\mathbf{B}^{(m)}(p) \circ \mathbf{B}^{(m)}(q) = \mathbf{B}^{(m)}(pq).$$

The map that sends $\mathbf{B}^{(m)}(v)$ to $M_m(v)$ is clearly an isomorphism. Indeed, all the coefficients $\binom{v}{u}$ appear at least once as an element of $M_m(v)$ for all words u of length at most m.

Acknowledgments

The idea of this binomial equivalence came after the meeting "Representing streams" organized at the Lorentz center in December 2012 where Jean-Eric Pin presented a talk, *Noncommutative extension of Mahlers theorem on interpolation series*, involving binomial coefficients on words. Jean-Eric Pin and the first author proposed independently to introduce this new relation.

References

[1] B. Adamczewski, Balances for fixed points of primitive substitutions, *Theoret. Comput. Sci.* **307** (2003), 47-75.

[2] S.V. Avgustinovich, D.G. Fon-Der-Flaass, A.E. Frid, Arithmetical complexity of infinite words, in *Words, Languages & Combinatorics III*, M. Ito and T. Imaoka (Eds.), World Scientific Publishing (2003) 51–62.

[3] S. Brlek, Enumeration of factors in the Thue-Morse word, *Discrete Appl. Math.* **24** (1989), 83–96.

[4] J. Berstel, M. Crochemore, J.-E. Pin, Thue–Morse sequence and p-adic topology for the free monoid, *Disc. Math.* **76** (1989), 89–94.

[5] J. Cassaigne, Complexité et facteurs spéciaux, *Bull. Belg. Math. Soc. Simon Stevin* **4** (1997), 67-88.

[6] A. Cobham, Uniform tag sequences, *Math. Systems Theory* **6** (1972), 164192.

[7] A. de Luca, S. Varricchio, On the factors of the Thue-Morse word on three symbols, *Inform. Process. Lett.* **27** (1988), 281–285.

[8] A. Ehrenfeucht, G. Rozenberg, A limit theorem for sets of subwords in deterministic TOL languages, *Information Processing Lett.* **2** (1973), 70-73.

[9] A. Ehrenfeucht, K. P. Lee, G. Rozenberg, Subword complexities of various classes of deterministic developmental languages without interactions, *Theor. Comput. Sci.* **1** (1975), 59-75.

[10] T. Kamae, L. Zamboni, Sequence entropy and the maximal pattern complexity of infinite words, *Ergodic Theory Dynam. Systems* **22** (2002), 1191–1199.

[11] J. Karhumäki, A. Saarela, L. Q. Zamboni, On a generalization of Abelian equivalence and complexity of infinite words, *J. Combin. Theory Ser. A* **120** (2013), 2189-2206.

[12] M. Lothaire, *Combinatorics on Words*, Cambridge Mathematical Library, Cambridge University Press, (1997).

[13] M. Lothaire, *Algebraic Combinatorics on Words*, Encyclopedia of Mathematics and its Applications **90**, Cambridge University Press (2002).

[14] M. Lothaire, *Applied Combinatorics on Words*, Encyclopedia of Mathematics and its Applications **105**, Cambridge University Press (2005).

[15] A. Mateescu, A. Salomaa, K. Salomaa, and S. Yu. A sharpening of the Parikh mapping, *Theor. Inform. Appl.* **35** (2001), 551-564.

[16] A. Mateescu, A. Salomaa, and S. Yu, Subword histories and Parikh matrices, *J. Comput. Syst. Sci.* **68** (2004), 1-21.

[17] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel (Eds.), Lecture Notes in Mathematics **1794**, Springer-Verlag, Berlin, 2002.

[18] M. Rao, M. Rigo, P. Salimov, Avoiding 2-binomial squares and cubes, `arXiv:1310.4743`.

[19] G. Richomme, K. Saari, L. Q. Zamboni, Balance and abelian complexity of the Tribonacci word, *Adv. in Appl. Math.* **45** (2010), 212-231.

[20] G. Richomme, K. Saari, L. Q. Zamboni, Abelian complexity of minimal subshifts, *J. Lond. Math. Soc.* **83** (2011), 79-95.

[21] M. Rigo, P. Salimov, Another Generalization of Abelian Equivalence: Binomial Complexity of Infinite Words, *Lect. Notes in Comput. Sci.* **8079**, 217–228 Springer Verlag (2013), J. Karhumäki, L. Zamboni (Eds.).