17e Journées Montoises d'Informatique Théorique

10-14 Sep 2018
Talence
France

Table of contents

Part I

Exposés invités

Graphs for Genomic Sequences
Raluca Uricaru (LaBRI, Bordeaux)

In this presentation, after an introduction to bioinformatics and to the latest challenges in the field, I will focus on problems concerning data (genomic sequences in a four letter alphabet, i.e., A, T, G, C) generated by Next Generation Sequencing technologies. More specifically, I will dwell on the modélisation of problems like the problem of assembly of genomes and on solutions based on graphs (like overlap graphs and De Bruijn graphs).

About the generalised star-height problem
Laure Daviaud (Warwick)

In this talk, I will discuss about the generalised star-height problem: given a rational language L and an integer k, is L characterised by a rational expression using union, concatenation, complement and at most k nested stars? For $k = 0$, this problem is known to be decidable, and the languages satisfying this condition form the well-known class of star-free languages. I will explain how this decidability result works, describing the notion of identities for rational languages.

Rigidity of square-tiled interval exchange transformations and translation flows
Pascal Hubert (Marseille)

In this talk, we will discuss a joint work with Sébastien Ferenczi about rigidity of a class of interval exchange transformations (iet). Veech proved that almost every interval exchange transformation is rigid. It is difficult to provide examples of non rigid iets. Square-tiled surfaces form a very natural class of translation surfaces those whose periods belong to the integer lattice. We will give a necessary and sufficient condition to get rigidity of a linear flow on such a surface and of the associated iet.

Nyldon words
Émilie Charlier (Liège)

The theorem of Chen-Fox-Lyndon states that every finite word w over a fixed alphabet A can be uniquely factorized as $w = l_1 \cdots l_k$, where $(l_1, \ldots, l_k)$ is a nonincreasing sequence of Lyndon words with respect to the lexicographic order. This theorem can be used to define the family of Lyndon words over A in a recursive way: 1) the letters are Lyndon; 2) a finite word of length greater than one is Lyndon if it cannot be factorized into a nonincreasing sequence of shorter Lyndon words. In a post on Mathoverflow in November 2014, Darij Grinberg defines a variant of Lyndon words, which he calls Nyldon words, by reversing the lexicographic order in the previous recursive definition. The class of words so obtained is not, as one might first think, the class of maximal words in their conjugacy classes. Gringberg asks three questions: 1) How many Nyldon words of length n are there? 2) Is there an equivalent to the Chen-Fox-Lyndon theorem for Nyldon words? 3) Is it true that every primitive words admits exactly one Nyldon word in his conjugacy class? In this talk, I will discuss these questions in the more general context of Lazard factorizations of the free monoid and show that each of Grinberg's questions has a positive answer. This is a joint work with Manon Philibert (ENS Lyon) and Manon Stipulanti (ULiège).

Simple algorithms and fast-growing complexity for well-structured systems
Philippe Schnoebelen (LSV, CNRS, ENS Paris-Saclay)

Well-quasi-orderings (WQOs) are a fundamental tool in logic and computer science. They are the basis of a large number of finiteness/regularity/termination/... results. In constraint solving, automated deduction, program analysis, and many more fields, wqos usually appear under the guise of specific tools, like Dickson's Lemma (for tuples of integers), Higman's Lemma (for words and their subwords), Kruskal's Tree Theorem and its variants (for finite trees with embeddings), and recently the Robertson-Seymour Theorem (for graphs and their minors). What is not very well known is how to analyze the complexity of wqo-based algorithms.

In this talk we survey two lines of recent developments in well-structured systems: (1) generic algorithms and data structures for computing with upward-closed and downward-closed sets in wqos, and (2) generic tools for demonstrating complexity bounds in wqo-based algorithms and systems.

Part II

Exposés courts

Finding Short Synchronizing Words for Prefix Codes

Andrew Ryzhikov

LIGM, Université Paris-Est, Marne-la-Vallée, France

This is joint work with Marek Szykuła.

Introduction Prefix codes are a simple and powerful class of variable-length codes that are widely used in information compression and transmission. A famous example of prefix codes are Huffman's codes [Huf52]. In general, variable length codes are not resistant to errors, since one deletion, insertion or change of a symbol can desynchronize the decoder causing incorrect decoding of the whole remaining part of the message. However, in a large class of codes called synchronizing codes resynchronization of the decoder is possible in such situations. It is known that almost all maximal finite prefix codes are synchronizing [FJTZ03]. Synchronization of finite prefix codes has been investigated a lot [Ba16, Bis08, BP09, CDSGV92, Sch64, Sch67], see also the book [BPR10] and references therein. For efficiency reasons it is important to use as short words resynchronizing the decoder as possible to decrease synchronization time. However, despite the interest to synchronizing prefix codes, the computational complexity of finding short synchronizing words for them has not been studied so far. We provide a systematic investigation of this topic.

Each recognizable (by a finite automaton) maximal prefix code can be represented by an automaton decoding the star of this code. For a finite code, this automaton can be exponentially smaller than the representation of the code by listing all its words (consider, for example, the code of all words of some fixed length). This can of course happen even if the code is synchronizing. In different applications the first or the second way of representing the code can be useful. In some cases large codes having a short description may be represented by a minimized decoder, while in other applications the code can be described by simply providing the list of all codewords. We study the complexity of problems for both arbitrary and literal decoders of finite prefix codes.

Huffman decoders There is a strong relation between partial automata and prefix codes [BPR10]. A set X of words is called a *prefix code* if no word in X is a prefix of another word. The class of recognizable (by an automaton) prefix codes can be described as follows. Take a strongly connected partial automaton A and pick a state r in it. Then the set of all *first return words* of r (that is, words mapping r to itself such that each non-empty prefix does not map r to itself) is a recognizable prefix code. Moreover, each recognizable prefix code can be obtained this way. A prefix code is called *maximal* if it is not a subset of another prefix code. The class of maximal recognizable prefix codes corresponds to the class of complete automata. If a state r can be picked in an automaton in such a way that the set of all first return words is a finite prefix code, we call the automaton a *partial Huffman decoder*. If such automaton is complete (and thus the finite prefix code is maximal), we call it simply a *Huffman decoder*.

For the mentioned classes of decoders we obtain the following inapproximability result.

Theorem 1. *Unless P = NP, the problem of finding a shortest synchronizing words cannot be approximated in polynomial time within a factor of*

(i) $n^{1-\varepsilon}$ *for any* $\varepsilon > 0$ *for* n*-state binary strongly connected automata;*

(ii) $c \log n$ *for some* $c > 0$ *for binary* n*-states Huffman decoders;*

(iii) $n^{\frac{1}{2}-\varepsilon}$ *for any* $\varepsilon > 0$ *for binary* n*-state partial Huffman decoders.*

We remark that for strongly connected automata the bound is optimal because there exists a $O(n)$-approximation algorithm [GH11]. We also conjecture that for binary n-states Huffman decoders the $c \log n$-inapproximability bound is optimal.

Literal decoders The literal automaton of a maximal finite prefix code X is defined as follows. The set of its states is the set of proper prefixes of the words in X and the transitions are naturally defined to concatenate letters to the prefixes (or to map to the empty prefix if the resulting word is in X). The number of states of a literal automaton is polynomially equivalent to the total length of all words in the corresponding finite prefix code. Thus, it is a natural model for the problems where the code is provided by simply enumerating all its codewords.

Theorem 2. *For literal* n*-state decoders there exist*

(i) a polynomial time $O(\log n)$*-approximation algorithm*

and

(ii) for any $\varepsilon > 0$ *a quasi-polynomial time* $(1 + \varepsilon)$*-approximation algorithm*

for the problem of finding a shortest synchronizing word.

We conjecture that there exists a polynomial time exact algorithm for this problem.

Mortal words A word is called *mortal* for a partial automaton A if its mapping is undefined for all the states of A. The techniques that we develop can be easily adapted to get the same inapproximability for the problem of finding a shortest mortal word. This problem is connected for instance to the famous Restivo's conjecture [Res81].

Theorem 3. *Unless P = NP, the problem of finding a shortest mortal word cannot be approximated in polynomial time within a factor of*

(i) $n^{1-\varepsilon}$ *for any* $\varepsilon > 0$ *for* n*-state binary strongly connected partial automata;*

(ii) $c \log n$ *for some* $c > 0$ *for* n*-state binary partial Huffman decoders.*

We also propose a simple polynomial time $O(\log n)$-approximation algorithm for this problem in partial literal decoders.

2

References

[Ba16] Mikhail V. Berlinkov and Marek Szykuła. Algebraic synchronization criterion and computing reset words. *Information Sciences*, 369:718 – 730, 2016.

[Bis08] Marek Biskup. *Error Resilience in Compressed Data – Selected Topics*. PhD thesis, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, 2008.

[BP09] Marek Tomasz Biskup and Wojciech Plandowski. Shortest synchronizing strings for huffman codes. *Theoretical Computer Science*, 410(38):3925 – 3941, 2009.

[BPR10] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Encyclopedia of Mathematics and its Applications 129. Cambridge University Press, 2010.

[CDSGV92] Renato M. Capocelli, A. A. De Santis, Luisa Gargano, and Ugo Vaccaro. On the construction of statistically synchronizable codes. *IEEE Transactions on Information Theory*, 38(2):407–414, 1992.

[FJTZ03] Christopher F Freiling, Douglas S Jungreis, François Théberge, and Kenneth Zeger. Almost all complete binary prefix codes have a self-synchronizing string. *IEEE Transactions on Information Theory*, 49(9):2219–2225, 2003.

[GH11] Michael Gerbush and Brent Heeringa. *Approximating Minimum Reset Sequences*, pages 154–162. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[Huf52] David A. Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101, 1952.

[Res81] Antonio Restivo. Some remarks on complete subsets of a free monoid. *Quaderni de La ricerca scientifica, CNR Roma*, 109:19–25, 1981.

[Sch64] Marcel-Paul Schützenberger. On the synchronizing properties of certain prefix codes. *Information and Control*, 7(1):23 – 36, 1964.

[Sch67] Marcel-Paul Schützenberger. On synchronizing prefix codes. *Information and Control*, 11(4):396 – 401, 1967.

3

FIXED POINTS OF STURMIAN MORPHISMS AND THEIR DERIVATED WORDS

ABSTRACT

KAREL KLOUDA, KATEŘINA MEDKOVÁ, EDITA PELANTOVÁ, AND ŠTĚPÁN STAROSTA

1. Introduction

Sturmian words are probably the most studied object in combinatorics on words. They are aperiodic words over a binary alphabet having the least factor complexity possible, in other words, their factor complexity satisfies $\mathcal{C}_{\mathbf{u}}(n) = n+1$ for each $n \in \mathbb{N}$. Many properties, characterizations and generalizations are known, see for instance [4, 3, 2].

One of their characterizations is in terms of return words to their factors. Let $\mathbf{u} = u_0u_1u_2\cdots$ be a binary infinite word with $u_i \in \{0,1\}$. Let $w = u_iu_{i+1}\cdots u_{i+n-1}$ be its factor. The integer i is called an *occurrence* of the factor w. A return word to a factor w is a word $u_iu_{i+1}\cdots u_{j-1}$ with i and j being two consecutive occurrences of w such that $i < j$. In [13], Vuillon showed that an infinite word $\mathbf{u}$ is Sturmian if and only if each nonempty factor w has exactly two distinct return words. A straightforward consequence of this characterization is that if w is a prefix of $\mathbf{u}$, we may write

$$\mathbf{u} = r_{s_0}r_{s_1}r_{s_2}r_{s_3}\cdots$$

with $s_i \in \{0,1\}$ and r_0 and r_1 being the two return words to w. The coding of these return words, the word $d_{\mathbf{u}}(w) = s_0s_1s_2\cdots$ is called the *derivated word of* $\mathbf{u}$ *with respect to* w, introduced in [6]. A simple corollary of the characterization by return words and a result of [6] is that the derivated word $d_{\mathbf{u}}(w)$ is also a Sturmian word. This simple corollary follows also from other results. For instance, it follows from [1], where the authors investigate the derivated word of a standard Sturmian word and give its precise description. It also follows from the investigation of a more general setting in [5], which may in fact be used to describe derivated words of any episturmian word — generalized Sturmian words [7].

By the main result of [6], if $\mathbf{u}$ is a fixed point of a primitive morphism, the set of all derivated words of $\mathbf{u}$ is finite (the result also follows from [8]). In this case, again by [6], a derivated word itself is a fixed point of a primitive morphism.

In this article we study derivated words of fixed points of primitive Sturmian morphisms. By the results of [10], any primitive Sturmian morphism may be decomposed using elementary Sturmian morphisms — generators of the Sturmian monoid. We user the elementary Sturmian morphisms to describe the relation between the derivated words of a Sturmian sequence. The main result of our article is an exact description of the morphisms fixing the derivated words $d_{\mathbf{u}}(w)$ of $\mathbf{u}$, where $\mathbf{u}$ is fixed by a Sturmian morphism ψ and w is its prefix. For this purpose, we introduce an operation Δ acting on the set of Sturmian morphisms with unique fixed point, see Definition 4. Iterating this operation we create the desired list of the morphisms as stated in Theorem 5. The Sturmian morphisms with two fixed points are treated separately, see Proposition 6.

We continue our study by counting the number of derivated words, in particular by counting the distinct elements in the sequence $\left(\Delta^k(\psi)\right)_{k\geq 1}$. This number depends on the decomposition of ψ into the generators of the special Sturmian monoid, see below in Section 2. Using this decomposition, Propositions 8 and 9 provide the exact number of derivated words for two specific classes of Sturmian morphisms.

(K. Medková, E. Pelantová) Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 7, 115 19, Prague 1, Czech Republic

(K. Klouda, Š. Starosta) Department of Applied Mathematics, Faculty of Information Technology, Czech Technical University in Prague, Thákurova 9, 160 00, Prague 6, Czech Republic

E-mail: stepan.starosta@fit.cvut.cz (Štěpán Starosta).

For a general Sturmian morphism ψ, Proposition 7 gives a sharp upper bound on their number. The upper bound depends on the number of the elementary morphisms in the decomposition of ψ.

For our purposes, we do not fix the alphabet of a derivated word; two derivated words which differ only by a permutation of letters are identified one with another. Moreover, in the sequel, we work only with infinite words which are *uniformly recurrent*, i.e. each prefix w of $\mathbf{u}$ occurs in $\mathbf{u}$ infinitely many times and the set of all return words to w is finite (and, thus, the alphabet of the derivated word of $\mathbf{u}$ with respect to w is finite). Our aim is to describe the set

$$\mathrm{Der}(\mathbf{u}) = \{d_{\mathbf{u}}(w) \colon w \text{ is a prefix of } \mathbf{u}\}.$$

2. Sturmian morphisms

Let $\mathcal{A}$ be a finite alphabet. A *morphism* over $\mathcal{A}^*$ is a mapping $\psi : \mathcal{A}^* \mapsto \mathcal{A}^*$ such that $\psi(vw) = \psi(v)\psi(w)$ for all $v, w \in \mathcal{A}^*$. The domain of the morphism ψ can be naturally extended to $\mathcal{A}^{\mathbb{N}}$ by

$$\psi(u_0u_1u_2\cdots) = \psi(u_0)\psi(u_1)\psi(u_2)\cdots .$$

A morphism ψ is *primitive* if there exists a positive integer k such that the letter a occurs in the word $\psi^k(b)$ for each pair of letters $a, b \in \mathcal{A}$. A *fixed point* of a morphism ψ is an infinite word $\mathbf{u}$ such that $\psi(\mathbf{u}) = \mathbf{u}$.

A morphism ψ is a *Sturmian morphism* if $\psi(\mathbf{u})$ is a Sturmian word for any Sturmian word $\mathbf{u}$. The set of Sturmian morphisms together with composition forms the so-called *Sturmian monoid* usually denoted *St*. We work with these four elementary Sturmian morphisms:

$$\varphi_a : \begin{cases} 0 \to 0 \\ 1 \to 10 \end{cases} \qquad \varphi_b : \begin{cases} 0 \to 0 \\ 1 \to 01 \end{cases} \qquad \varphi_\alpha : \begin{cases} 0 \to 01 \\ 1 \to 1 \end{cases} \qquad \varphi_\beta : \begin{cases} 0 \to 10 \\ 1 \to 1 \end{cases}$$

and with the monoid $\mathcal{M}$ generated by them, i.e. $\mathcal{M} = \langle \varphi_a, \varphi_b, \varphi_\alpha, \varphi_\beta \rangle$. The monoid $\mathcal{M}$ is also called *special Sturmian monoid*. For a nonempty word $u = u_0 \cdots u_{n-1}$ over the alphabet $\{a, b, \alpha, \beta\}$ we put

$$\varphi_u = \varphi_{u_0} \circ \varphi_{u_1} \circ \cdots \circ \varphi_{u_{n-1}}.$$

The monoid $\mathcal{M}$ is not free. It is easy to show that for any $k \in \mathbb{N}$ we have

$$\varphi_{\alpha a^k \beta} = \varphi_{\beta b^k \alpha} \quad \text{and} \quad \varphi_{a \alpha^k b} = \varphi_{b \beta^k a}.$$

We can equivalently say that the following rewriting rules hold on the set of words from $\{a, b, \alpha, \beta\}^*$:

$$\alpha a^k \beta = \beta b^k \alpha \quad \text{and} \quad a \alpha^k b = b \beta^k a \qquad \text{for any } k \in \mathbb{N}\,. \tag{1}$$

In [12], the author reveals a presentation of the Sturmian monoid which includes the special Sturmian monoid $\mathcal{M} = \langle \varphi_a, \varphi_b, \varphi_\alpha, \varphi_\beta \rangle$. A presentation of the special Sturmian monoid follows from this result. It is also given explicitly in [9]:

Theorem 1. *Let $w, v \in \{a, b, \alpha, \beta\}^*$. The morphism φ_w is equal to φ_v if and only if the word v can be obtained from w by applying the rewriting rules (1).*

Note that the presentation of a generalization of the Sturmian monoid, the so-called *episturmian monoid*, is also known, see [11]. The next lemma summarizes several simple and well-known properties of Sturmian morphisms we exploit in the sequel.

Lemma 2. *Let $w \in \{a, b, \alpha, \beta\}^+$.*

(i) The morphism φ_w is primitive if and only if w contains at least one Greek letter α or β and at least one Latin letter a or b.
(ii) If φ_w is primitive, then each of its fixed points is aperiodic and uniformly recurrent.
(iii) If φ_w is primitive, then it has two fixed points if and only if w belongs to $\{a, \alpha\}^*$.

For $w \in \{a, b, \alpha, \beta\}^*$ the rules (1) preserve positions in w occupied by Latin letters $\{a, b\}$ and positions occupied by Greek letters $\{\alpha, \beta\}$. We define that $a < b$ and $\alpha < \beta$ which allows the following definition.

Definition 3. *Let $w \in \{a,b,\alpha,\beta\}^*$. The lexicographically greatest word in $\{a,b,\alpha,\beta\}^*$ which can be obtained from w by application of rewriting rules (1) is denoted $N(w)$. If $\psi = \varphi_w$, then the word $N(w)$ is the* normalized name *of the morphism ψ and it is also denoted by $N(\psi) = N(w)$.*

3. Derivated words of fixed points of Sturmian morphisms

Let $\psi \in \langle \varphi_a, \varphi_b, \varphi_\alpha, \varphi_\beta \rangle$ and $N(\psi) = w \in \{a,b,\alpha,\beta\}^* \setminus \{a,\alpha\}^*$ be the normalized name of the morphism ψ. The word w has a prefix $a^k\beta$ or $\alpha^k b$ for some $k \in \mathbb{N}$. This property enables us to define a transformation on the set of morphisms from $\mathcal{M} \setminus \langle \varphi_a, \varphi_\alpha \rangle$. This transformation is in fact the desired algorithm returning the morphisms $\psi_1, \psi_2, \ldots, \psi_\ell$ mentioned above.

Definition 4. *Let $w \in \{a,b,\alpha,\beta\}^* \setminus \{a,\alpha\}^*$ be the normalized name of a morphism ψ, i.e., $\psi = \varphi_w$. We put*

$$\Delta(w) = \begin{cases} N(w'a^k\beta) & \text{if } w = a^k\beta w', \\ N(w'\alpha^k b) & \text{if } w = \alpha^k b w' \end{cases}$$

and, moreover, $\Delta(\psi) = \varphi_{\Delta(w)}$.

Theorem 5. *Let $\psi \in \langle \varphi_a, \varphi_b, \varphi_\alpha, \varphi_\beta \rangle$ be a primitive morphism and $N(\psi) = w \in \{a,b,\alpha,\beta\}^* \setminus \{a,\alpha\}^*$ be its normalized name. Denote $\mathbf{u}$ the fixed point of ψ. The word $\mathbf{x}$ is (up to a permutation of letters) a derivated word of $\mathbf{u}$ with respect to one of its prefixes if and only if $\mathbf{x}$ is the fixed point of the morphism $\Delta^j(\psi)$ for some $j \geq 1$.*

Given a finite word u, we define the *cyclic shift* of $u = u_0u_1\cdots u_{n-1}$ to be the word

$$\mathrm{cyc}(u) = u_1u_2\cdots u_{n-1}u_0.$$

Proposition 6. *Let $w \in \{a,\alpha\}^*$ be the normalized name of a primitive morphism ψ and let a be its first letter.*

- *(i) Let $\mathbf{u}$ be the fixed point of ψ starting with 0. Denote $v = b^{-1}N(wb) \in \{a,\beta\}^*$. We have $\mathrm{Der}(\mathbf{u}) = \{\mathbf{v}\} \cup \mathrm{Der}(\mathbf{v})$, where $\mathbf{v}$ is the unique fixed point of the morphism φ_v.*
- *(ii) Let $\mathbf{u}$ be the fixed point of ψ starting with 1. Put $v = \mathrm{cyc}(w)$. We have $\mathrm{Der}(\mathbf{u}) = \mathrm{Der}(\mathbf{v})$, where $\mathbf{v}$ is the fixed point of the morphism φ_v.*

4. The number of derivated words

Proposition 7. *If $w \in \{a,b,\alpha,\beta\}^* \setminus \{a,\alpha\}^*$ it the normalized name of a primitive Sturmian morphism $\psi = \varphi_w$ and $\mathbf{u}$ is a fixed point of ψ, then*

$$1 \leq \#\mathrm{Der}(\mathbf{u}) \leq 3|w| - 4\,. \tag{2}$$

Moreover, for any length $n \geq 2$ there exist normalized names $w', w'' \in \{a,b,\alpha,\beta\}^* \setminus \{a,\alpha\}^*$ of length n such that

- *(i) $\varphi_{w'}$ and $\varphi_{w''}$ are not powers of other Sturmian morphisms,*
- *(ii) for the fixed points $\mathbf{u}'$ and $\mathbf{u}''$ of the morphism $\varphi_{w'}$ and $\varphi_{w''}$, the lower resp. the upper bound in (2) is attained.*

We also provide precise numbers of distinct derivated words for these three types of morphisms:

1. ψ is a standard morphism from $\mathcal{M}$, i.e. $\psi \in \langle \varphi_b, \varphi_\beta \rangle$,
2. ψ is a standard morphism from $\mathcal{M} \circ E$, i.e. $\psi \in \langle \varphi_b, \varphi_\beta \rangle \circ E$,
3. ψ is a morphism from $\langle \varphi_a, \varphi_\alpha \rangle$.

To describe these numbers, we introduce the following morphism $F : \{a,b,\alpha,\beta\}^* \mapsto \{a,b,\alpha,\beta\}^*$ determined by

$$F(a) = \alpha, \quad F(\alpha) = a, \quad F(b) = \beta, \quad F(\beta) = b,$$

and we set

$$\mathrm{cyc_F}(w_1w_2w_3\cdots w_n) = w_2w_3\cdots w_nF(w_1)$$

for a finite word $w_1w_2w_3\cdots w_n$.

Proposition 8. *Let* $\mathbf{u}$ *be a fixed point of a standard Sturmian morphism* ψ *which is not a power of any other Sturmian morphism.*

(i) *If* $\psi = \varphi_w$*, then* $\mathbf{u}$ *has* $|w|$ *distinct derivated words, each of them (up to a permutation of letters) is fixed by one of the morphisms*

$$\varphi_{v_0}, \varphi_{v_1}, \varphi_{v_2}, \ldots, \varphi_{v_{|w|-1}}, \quad \text{where } v_k = \mathrm{cyc}^k(w) \text{ for } k = 0, 1, \ldots, |w|-1.$$

(ii) *If* $\psi = \varphi_w \circ E$*, then* $\mathbf{u}$ *has* $|w|$ *distinct derivated words, each of them (up to a permutation of letters) is fixed by one of the morphisms*

$$\varphi_{v_0} \circ E, \varphi_{v_1} \circ E, \varphi_{v_2} \circ E, \ldots, \varphi_{v_{|w|-1}} \circ E, \quad \text{where } v_k = \mathrm{cyc}_F^k(w) \text{ for } k = 0, 1, \ldots, |w|-1.$$

Proposition 9. *Let* $w \in \{\alpha, a\}^*$ *be the normalized name of a primitive morphism* ψ *such that the letter* a *is a prefix of* w*. Moreover, assume that* ψ *is not a power of any other Sturmian morphism.*

(i) *The fixed point of* ψ *starting with* 0 *has exactly* $1 + |w|_\alpha$ *distinct derivated words.*
(ii) *The fixed point of* ψ *starting with* 1 *has exactly* $1 + |w|_a$ *distinct derivated words.*

Acknowledgements

K.M., E.P. and Š.S. acknowledge financial support by the Czech Science Foundation grant GAČR 13-03538S. K. M. also acknowledges financial support by the Czech Technical University in Prague grant SGS17/193/OHK4/3T/14.

References

[1] I. M. Araújo and V. Bruyère, *Words derivated from Sturmian words*, Theoret. Comput. Sci., 340 (2005), pp. 204–219.
[2] L. Balková, E. Pelantová, and Š. Starosta, *Sturmian jungle (or garden?) on multiliteral alphabets*, RAIRO-Theor. Inf. Appl., 44 (2010), pp. 443–470.
[3] J. Berstel, *Sturmian and Episturmian words*, in Algebraic Informatics, S. Bozapalidis and G. Rahonis, eds., vol. 4728 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp. 23–47.
[4] J. Berstel and P. Séébold, *Sturmian words*, in Algebraic Combinatorics on Words, M. Lothaire, ed., Cambridge University Press, 2002, pp. 45–110.
[5] M. Bucci and A. De Luca, *On a family of morphic images of Arnoux-Rauzy words*, in LATA '09: Proceedings of the 3rd International Conference on Language and Automata Theory and Applications, Springer-Verlag, 2009, pp. 259–266.
[6] F. Durand, *A characterization of substitutive sequences using return words*, Discrete Math., 179 (1998), pp. 89–101.
[7] A. Glen and J. Justin, *Episturmian words: a survey*, Theor. Inf. Appl., 43 (2009), pp. 403–442.
[8] C. Holton and L. Q. Zamboni, *Descendants of primitive substitutions*, Theoret. Comput. Syst., 32 (1999), pp. 133–157.
[9] C. Kassel and C. Reutenauer, *Sturmian morphisms,the braid group* B_4*, Christoffel words and bases of* F_2, Annali di Matematica Pura ed Applicata, 186 (2007), pp. 317–339.
[10] F. Mignosi and P. Séébold, *Morphismes sturmiens et règles de Rauzy*, J. Théor. Nombres Bordeaux, 5 (1993), pp. 221–233.
[11] G. Richomme, *Conjugacy and episturmian morphisms*, Theoret. Comput. Sci., 302 (2003), pp. 1–34.
[12] P. Séébold, *Fibonacci morphisms and Sturmian words*, Theoret. Comput. Sci., 88 (1991), pp. 365–384.
[13] L. Vuillon, *A characterization of Sturmian words by return words*, Eur. J. Combin., 22 (2001), pp. 263–275.

k-Spectra of Strictly Balanced Words

Joel D. Day and Pamela Fleischmann

Department of Computer Science,
Kiel University, Germany
{jda,fpa}@informatik.uni-kiel.de

Abstract. A word u is a scattered factor of w if there exist $u_1, u_2, ..., u_n$, and $v_0, v_1, .., v_n$ such that $u = u_1 u_2 ... u_n$ and $w = v_0 u_1 v_1 u_2 v_2 ... u_n v_n$. We consider the set of length-k scattered factors (k-spectrum) of a given word w, denoted $\mathrm{ScatFact}_k(w)$. We prove several properties of the sets $\mathrm{ScatFact}_k(w)$ in the case of words w over a binary alphabet of length $2k$ for which the number of occurrences of each letter is equal. Such words are called strictly balanced. In particular, motivated by the task of recognising whether a set of words is a k-spectra of some word w, we consider the question of which cardinalities $n = |\,\mathrm{ScatFact}_k(w)|$ are obtainable for each k. We also consider the task of reconstructing words from their strictly balanced scattered factors.

1 Introduction

A scattered factor of w can be thought of as a representation of w in which some parts are missing. As such, there is considerable interest in the relationship of a word and its scattered factors from both a theoretical and practical point of view. For an introduction, see [3]. On the one hand, it is easy to imagine how, in any situation where discrete, linear data is read from an imperfect input – such as when sequencing DNA or during the transmission of a digital signal – scattered factors form a natural model, as multiple parts of the input may be missed, but the rest will remain unaffected and in-sequence. On the other hand, from a more theoretical perspective, there have been efforts to bridge the gap between the non-commutative field of combinatorics on words with traditional commutative mathematics via Parikh matricies (cf. e.g., [5, 6]) which are closely related to, and influenced by the topic of scattered factors.

One of the most fundamental questions about scattered factors of words and sets of scattered factors in general, is: given a set S of words (of length k), is S the set of scattered factors (or a k-spectrum) of some word w. In general, it remains a long standing goal of the theory to give a "nice" descriptive characterisation of scattered factor sets (and similarly, k-spectra), and to better understand their structure [3]. Another fundamental question concerning k-spectra, and one well motivated in several applications, is the question of reconstruction: given a word w of length n, for what values k does the k-spectrum of w uniquely determined? This question has generally had more success with definitive answers in a variety of cases. In particular, in [1], the exact bound of $\frac{n}{2}+1$ is given in the general case.

Other variations, including for the definition of k-spectra where multiplicities are also taken into account, are considered in [4], while [2] considers the question of reconstructing words from their palindromic scattered factors.

In the current work, we consider the restricted setting of strictly balanced words: words over a binary alphabet $\{\mathtt{a}, \mathtt{b}\}$ with equal numbers of $\mathtt{a}$s and $\mathtt{b}$s. We show that the cardinality of their scattered factor sets ranges between $k+1$ and 2^k and we prove for every $k+1 \leq i \leq 3k-2$ whether a k-spectrum of cardinality i exists. Moreover some results between $3k-1$ and 2^k are given. In Section 4 we approach the question of reconstructing strictly balanced words from k-spectra in the specific case that the spectra are also limited to strictly balanced words only. While we are not able to resolve the question completely, we conjecture that the situation is similar to the general case; we show that this bound holds in the case that w contains at most two blocks of $\mathtt{b}$s.

Before we are able to present our results, we need to define the setting of strictly balanced words. We consider words w over an alphabet $\Sigma = \{\mathtt{a}, \mathtt{b}\}$. The number of occurrences of a letter $\mathtt{a} \in \Sigma$ in a word $w \in \Sigma^*$ is denoted by $|w|_{\mathtt{a}}$. The subset of Σ^* which contains only words with equal numbers of occurrences of letters is defined by $\Sigma^*_{sb} = \{w \in \Sigma^* | \forall x, y \in \Sigma : |w|_x = |w|_y\}$ and these words are called *strictly balanced*. For example, $\mathtt{abaa}$ is not strictly balanced, while $\mathtt{abbaba}$ is.

Definition 1. *A word $u = a_1 \ldots a_n \in \Sigma^n$, for $n \in \mathbb{N}$, is a* scattered factor *of a word $w \in \Sigma^+$ if there exists $v_0, \ldots, v_n \in \Sigma^*$ with $w = v_0 a_1 v_1 \ldots v_{n-1} a_n v_n$. Let* $\mathrm{ScatFact}(w)$ *denote the set of w's scattered factors and consider additionally* $\mathrm{ScatFact}_k(w)$ *(full k-spectrum) and* $\mathrm{ScatFact}_{\leq k}(w)$ *(k-spectrum) as the two subsets of* $\mathrm{ScatFact}(w)$ *which contain only the scattered factors of length $k \in \mathbb{N}$ or the ones up to length $k \in \mathbb{N}$.*

We note two obvious, but important symmetries regarding k-spectra: for$w \in \Sigma^*$. $\mathrm{ScatFact}(w^R) = \{u^R \mid u \in \mathrm{ScatFact}(w)\}$ and $\mathrm{ScatFact}(\overline{w}) = \{\overline{u} \mid u \in \mathrm{ScatFact}(w)\}$ hold with the renaming morphism $\bar{\cdot}$. Thus, from a structural point of view, it is sufficient to consider only one representative (here the lexicographically smallest with $\mathtt{a} < \mathtt{b}$) from the equivalence classes.

2 Cardinalities of k-Spectra of Strictly Balanced Words

In the current section, we are interested in the cardinalties of the k spectra, and in the question: which cardinalities are not possible? It is a straightforward observation that not every subset of Σ^k is a k-spectrum of some word w. For example $\mathtt{aa}$ and $\mathtt{bb}$ can only be scattered factors of a word containing both $\mathtt{a}$s and $\mathtt{b}$s, and therefore having either $\mathtt{ab}$ or $\mathtt{ba}$ as a scattered factor. In fact, for $k = 2$, the sets $\{\mathtt{aa}, \mathtt{ab}, \mathtt{bb}\}$ and $\{\mathtt{aa}, \mathtt{ba}, \mathtt{bb}\}$ are the smallest possible k-spectra of words of length $2k$ in both the general case, and when restricted to strictly balanced words only. Moreover these sets are equivalent in the sense that one is a renaming (or a reversal) of the other. Note that the largest possible set in this case is $\{\mathtt{aa}, \mathtt{ab}, \mathtt{ba}, \mathtt{bb}\}$ which has size $4 = 2k = 2^k$. Our first result generalises the previous observation about minimal-size and maximal-size k-spectra.

Lemma 1. *For all $k \in \mathbb{N}$, the smallest reachable cardinality for any $w \in \Sigma_{sb}^{2k}$ is $|\operatorname{ScatFact}_k(w)| = k+1$, reached exactly for $w = \mathtt{a}^k\mathtt{b}^k$ (up to renaming and reversal), and $\operatorname{ScatFact}_k(\mathtt{a}^k\mathtt{b}^k) = \{\mathtt{a}^r\mathtt{b}^s \mid r+s=k, r,s \in [k]_0\}$ holds.*

Lemma 2. *Let $k \in \mathbb{N}$. Then $w \in \{ab, ba\}^k$ if and only if $\operatorname{ScatFact}_k(w) = \Sigma^k$.*

By the Lemmas 1 and 2, the characterisation for the smallest and the largest closure w.r.t. cardinality of the given set S are given. Now the gap in between will be investigated. Since there does not exist a gap for $k=2$, assume $k \in \mathbb{N}_{\geq 3}$. The following two statements show that $2^k - 1$ and $2k$ are always reachable and thus the possible cardinalities for $k=3$ are fully characterised.

1. $|\operatorname{ScatFact}_k(w)| = 2^k - 1$ iff $w \in \{(\mathtt{ab})^i\mathtt{a}^2\mathtt{b}^2(\mathtt{ab})^{k-i-2} \mid i \in [k-2]_0\}$ (in particular $\operatorname{ScatFact}_k(w) = \Sigma^k \setminus \{\mathtt{b}^{i+1}\mathtt{a}^{k-i-1}\}$),
2. $|\operatorname{ScatFact}_k(w)| = 2k$ iff $w \in \{\mathtt{a}^{k-1}\mathtt{bab}^{k-1}, \mathtt{a}^{k-1}\mathtt{b}^k\mathtt{a}\}$

The cardinality of $2k$ is important since there is a gap between $k+1$ and $2k$, i.e. $\forall w \in \Sigma_{sb}^{2k} : |\operatorname{ScatFact}_k(w)| \notin \{k+2, \ldots, 2k-1\}$. This shows that with increasing k the number of possible cardinalities at the *beginning* of the scala from $k+1$ to 2^k decreases: the larger k is the more unlikely it is somehow to find a k-spectrum of a small cardinality. To investigate the second gap we have $|\operatorname{ScatFact}_k(\mathtt{a}^{k-i}\mathtt{b}^k\mathtt{a}^i)| = k(i+1)-i^2+1$ for $i \in \left[\lfloor \frac{k}{2} \rfloor\right]$. It is worth noting that this includes all square numbers being at least four: $|\operatorname{ScatFact}_k(\mathtt{a}^{\frac{k}{2}}\mathtt{b}^k\mathtt{a}^{\frac{k}{2}})| = \left(\frac{k}{2}+1\right)^2$ holds for k even. Moreover $|\operatorname{ScatFact}_k(\mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a}^2)| = 3k-3$ holds. This result is important since it will be shown in the following that the cardinalities $2k+1$ up to $3k-4$ are not reachable. In other words $\mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a}^2$ delivers the third smallest cardinality after $k+1$ and $2k$. Contrarily the cardinality $3k-2$ belongs to the word $\mathtt{a}^{k-1}\mathtt{b}^2\mathtt{ab}^{k-2}$.

Proposition 1. *For $k \geq 5$, no word $w \in \Sigma_{sb}^{2k}$ has k-spectrum of cardinality $2k+i$ for $i \in [k-4]$, i.e. between $2k+1$ and $3k-4$ is a cardinality-gap.*

We will end this analysis with the conjecture that in contrast to the first gap, the last gap ends earlier the larger k is. More precisely, if for $k \in \mathbb{N}_{\geq 4}$ and $i \in [k-2]_0$, $w = \mathtt{a}^2\mathtt{b}^2(\mathtt{ab})^{k-3-i}\mathtt{ba}(\mathtt{ab})^i$ holds then $|\operatorname{ScatFact}_k(w)| = 2^k - 2 - i$ follows. Notice that this conjecture implies that indeed similar to the second gap here $4k-4$ is always reached. On the other hand, in contrast to the second gap, the third gap is not of the form $4k-4-i$ for $i \in [k-4]$.

3 Reconstructing Strictly Balanced Words from their k-Spectra

As with the general case, it is easy to see that strictly balanced words of length $2k$ are not uniquely determined by their scattered factors of length k. In the current section we discuss the question of when a strictly balanced word w of length $2k$ is uniquely identified by the set $\operatorname{ScatFact}_{k'}(w) \cap \Sigma_{sb}^{k'}$ for $2k > k' > k$.

Of course if k' is odd then $\mathrm{ScatFact}_{k'}(w) \cap \Sigma_{sb}^{k'} = \emptyset$ for all words w, so in these cases the answer is trivially negative. In the general case, Dress and Erdös [1] showed, that if $\mathrm{ScatFact}_{k+1}(w) = \mathrm{ScatFact}_{k+1}(w')$ holds for $w, w' \in \Sigma^{2k}$ then $w = w'$ follows. If w is strictly balanced we found a straightforward proof for their proposition. However, in both proofs, there is a necessity in some cases to consider scattered factors u consisting mostly of as or mostly of bs – i.e., that do not belong to Σ_{sb}^*. Thus it remains an open problem whether the same bound of $k+1$ (or in the case that k is even, $k+2$) is sufficient. While we do not resolve the question completely, we conjecture that these bounds do still hold.

Conjecture 1. Let $k \in \mathbb{N}$. Let $k' = k+1$ if k is odd, and $k' = k+2$ if k is even. Let $w, w' \in \Sigma_{sb}^{2k}$ such that $\mathrm{ScatFact}_{k'}(w) = \mathrm{ScatFact}_{k'}(w')$. Then $w = w'$.

It is possible to show that the conjecture holds when there are at most two blocks of bs (by symmetry at most two blocks of as), i.e. $w \in \mathtt{a}^*\mathtt{b}^*\mathtt{a}^*\mathtt{b}^*\mathtt{a}^* \cap \Sigma_{sb}^{2k}$:

- for k odd, w is uniquely determined by $\mathrm{ScatFact}_{k+1}(w) \cap \Sigma_{sb}^{k}$,
- for k even, w is uniquely determined by $\mathrm{ScatFact}_{k+2}(w) \cap \Sigma_{sb}^{k}$.

References

1. A. W.M. Dress and P. Erdös. Reconstructing words from subwords in linear time. *Annals of Combinatorics*, 8(4):457–462, 2004.
2. Š. Holub and K. Saari. On highly palindromic words. *Descrete Applied Mathematics*, 157:953–959, 2009.
3. M. Lothaire. *Combinatorics on Words.* Cambridge University Press, 1997.
4. J. Manǔch. Characterization of a word by its subwords. In *Developments in Language Theory*, pages 210–219. World Scientific, 1999.
5. A. Mateescu, A. Salomaa, and S. Yu. Subword histories and parikh matrices. *Journal of Computer and System Sciences*, 68(1):1–21, 2004.
6. A. Salomaa. Connections between subwords and certain matrix mappings. *Theoretical Computer Science*, 340(2):188–203, 2005.

Multidimensional subshifts of finite type are discrete dynamical systems as a set of colorings of an infinite regular grid with elements of a finite set $\mathcal{A}$ together with the shift action. The set of colorings is defined by forbidding a finite set of patterns all over the grid (also called local rules). The most simple and most considered grids of this type are $\mathbb{Z}^2$ and more generally $\mathbb{Z}^d$ for $d \geq 1$. In this case, one can consider a coloring as a bi-dimensional and infinite word on the alphabet $\mathcal{A}$.

They are notably involved in statistical physics in the study of so-called lattice models. These models are often simple to describe : for instance, the hard square model is defined on alphabet $\mathcal{A} = \{0, 1\}$ and by forbidding two 1 to appear on horizontally or vertically adjacent positions of the lattice. However, these models and their physical constants, such as the entropy are difficult to apprehend with general methods, and involve specific properties of the considered model.

Although it is known that it is possible to compute the entropy of one-dimensional version of these models by computing the greatest eigenvalue of a matrix which derives from the description of the subshift, this is not possible for multidimensional subshifts. This is the consequence of a result by M. Hochman and T. Meyerovitch in 2010, which states that the possibles values of the entropy for multidimensional subshifts of finite type are the Π_1-computable numbers, including in particular non-computable numbers.

The method developed for this purpose originates in the work of R. Berger and R. Robinson. It has been developed further in order to characterize other dynamical aspects of SFT with computability conditions, with similar constructions. It consists in the implementation of Turing machines in hierarchical structures that emerge from the local rules.

However, models studied in statistical physics obey to strong dynamical constraints and there is still hope to include them into a sub-class of subshifts of finite type for which the entropy is uniformly computable (this means that there is an algorithm which can provide arbitrarily precise approximations of the entropy, provided the precision and the local rules of the subshift). An example of a constraint defining a class where this is verified is the block gluing : this was proved by R. Pavlov and M. Schraudner. This property means that two square blocks can be viewed in any relative positions in some element of the subshift provided that the distance between the two blocks is sufficiently large, with minimal distance not depending on the size of the blocks) is a computable real number. Although they provided a construction to realize some class numbers as entropy of block gluing SFT, they did not prove a characterization, and this problem seems difficult. However, it could be possible to find broader class for which the entropy is still computable.

A strategy to understand the limit between the general regime where Hochman and Meyerovitch's result holds and this restricted block gluing class is to quantify this property. This means imposing that two patterns can be glued in any two positions in a configuration of the subshift, provided that the distance is great enough, where the minimal distance is a linear function of the size of these patterns.

In a work with Mathieu Sablik, we made a step towards the limit, proving that the result of Hochman and Meyerovitch is robust under the linear version of this property (where the minimal distance function is $O(n)$ where n is the size of the two square blocks). The aim of this talk would be, after a presentation of the problem, to give an insight on the obstacles to this property in the initial construction of Hochman and Meyerovitch, using a construction slightly simpler to present, and on the methods used to overcome the obstacles. These methods involves in particular a modification of the Turing machine model and an operator on subshifts that acts by distortion.

Unary Patterns of Size Four with Morphic Permutations

Kamellia Reshadi

Institut für Informatik, Christian-Albrechts-Universität zu Kiel.
kre@informatik.uni-kiel.de

Abstract. We investigate the avoidability of unary patterns of size of four with morphic permutations. More precisely, we show how to identify precisely, given the positive integers i, j, k, the alphabets over which a pattern $x\pi^i(x)\pi^j(x)\pi^k(x)$ is avoidable, where x is a word variable and π is a function variable with values in the set of all morphic permutations of the respective alphabets. This continues the work of [Manea et al., 2015], where a complete characterisation of the avoidability of cubic patterns with permutations was given.

1 Introduction

The avoidability of patterns in infinite words is an old area of interest with a first systematic study going back to Thue. In these initial papers it was shown that there exist a binary infinite morphic word and a ternary infinite morphic word that avoid cubes and squares, respectively. That is, these infinite words do not contain instances of the patterns xxx and xx, respectively.

In this article, we are studying the avoidability of repetitions in a generalised setting. Namely, we are interested in the avoidability of unary patterns with functional dependencies between variables. We are considering patterns like $x\pi^i(x)\pi^j(x)\pi^k(x)$, where x is a word variable while π is function variable, which can be replaced by bijective morphisms only. The instances of such patterns over an alphabet Σ are obtained by replacing x with a concrete word, and π by a morphic permutation of Σ. For example, an instance of the pattern $x\pi(x)x\pi(x)$ over $\Sigma = \{a, b\}$ is the word $uvuv$ such that $|u| = |v|$, and v is the image of u under any permutation on the alphabet. Considering the permutation $a \to b$, and $b \to a$, then $aba|bab|aba|bab$ is an instance of $x\pi(x)x\pi(x)$.

In this setting, we continue the work of [Manea et al., 2015] as follows. In that paper, a complete characterisation of the avoidability of cubic patterns with permutations $x\pi^i(x)\pi^j(x)$ was given. Furthermore, it was shown that there exists a ternary word that avoids all patterns $\pi_{i_1}(x)\dots\pi_{i_r}(x)$ where $r \geq 4$, x a word variable over some alphabet Σ, with $|x| \geq 2$ and $|\Sigma| \geq 3$, and the π_{i_j} function variables that may be replaced by anti-/morphic permutations of Σ. However, this result only holds when the length of x is restricted to be at least 2. Also, it was shown that all patterns $\pi^{i_1}(x)\dots\pi^{i_n}(x)$ with $n \geq 4$ under morphic permutations are avoidable in alphabets of size $2, 3$, and 4, but there exist patterns which are unavoidable in alphabets of size 5. We extend these results by showing how to determine exactly, for a given unary pattern $\mathcal{P}$ of size four with permutations, which are the alphabets in which it is avoidable.

The main result of our paper is that given i, j, k, we show how to compute the value m such that the pattern $x\pi^i(x)\pi^j(x)\pi^k(x)$, with $i, j, k \geq 0$, is unavoidable in alphabets of size at least m and avoidable in alphabets of size $2, 3, 4, \ldots, m-1$. To achieve this, we define a series of parameters that allow us to characterize, for each alphabet, what form the instances of the pattern may have over a certain alphabet. Then, we show that for each pattern there exists an interval (whose left end is 2 and right end is defined based on the respective parameters) such that over each alphabet whose size is in the respective interval, there exists an infinite word that does not contain instances of the given pattern. The structure of the paper is as follows: we first give a series of basic definitions and preliminary results. Then we define the aforementioned parameters, and show how to use them to compute, for a given pattern, the minimum size σ of an alphabet over which the respective pattern is unavoidable. Finally, we show the correctness of the computation done in the previous step: for alphabets with less than σ symbols the pattern is avoidable.

2 Preliminaries

We define $\Sigma_k = \{0, \ldots, k-1\}$ to be an alphabet with k letters; the empty word is denoted by ε. For words u and w, we say that u is a prefix (resp. suffix) of w, if there exists a word v such that $w = uv$ (resp. $w = vu$). If $f : \Sigma_k \to \Sigma_k$ is a permutation, we say that the order of f, denoted $\mathbf{ord}(f)$, is the minimum value $m > 0$ such that f^m is the identity. If $a \in \Sigma_k$ is a letter, the order of a with respect to f, denoted $\mathbf{ord}_f(a)$, is the minimum number m such that $f^m(a) = a$.

In this paper, we consider only unary patterns (i.e., containing only one variable) with morphic permutations, that is, all function variables are unary and are substituted by morphic permutations only.

The infinite Hall word h is defined as $h = \lim_{n\to\infty} \phi_h^n(0)$, for the morphism $\phi_h : \Sigma_3^* \to \Sigma_3^*$ where $\phi_h(0) = 012$, $\phi_h(1) = 02$ and $\phi_h(2) = 1$. The infinite word h avoids the pattern xx (squares).

3 Avoidability of patterns under permutations

In this section we try to identify an upper bound on the size of the alphabets Σ_m in which a patterns $x\pi^i(x)\pi^j(x)\pi^k(x)$, with $i, j, k \geq 0$ is unavoidable, when π is substituted by a morphic permutation.

In the pattern $x\pi^i(x)\pi^j(x)\pi^k(x)$, the factors x, $\pi^i(x)$, π^j, or $\pi^k(x)$ are called x-items in the following. Our analysis is based on the relation between the possible images of the four x-items occurring in a pattern, following the ideas of [**?**]. For instance, we want to check whether in a possible image of our pattern, all four x-items can be mapped to a different word, or whether the second and the last x-items can be mapped to the same word, etc.

To achieve this, we define in Table 1 the parameters α_i, with $1 \leq i \leq 14$.

$\alpha_1 = \inf\{t : t \nmid i, t \nmid j, t \nmid k, t \nmid \lvert i-j \rvert, t \nmid \lvert i-k \rvert, t \nmid \lvert j-k \rvert\}$	0123
$\alpha_2 = \inf\{t : t \mid i, t \nmid j, t \nmid k, t \nmid \lvert j-k \rvert\}$	0012
$\alpha_3 = \inf\{t : t \nmid i, t \mid j, t \nmid k, , t \nmid \lvert i-k \rvert\}$	0102
$\alpha_4 = \inf\{t : t \nmid i, t \nmid j, t \mid \lvert i-k \rvert\}$	0121
$\alpha_5 = \inf\{t : t \nmid i, t \nmid j, t \nmid \lvert i-j \rvert, t \nmid \lvert i-k \rvert, t \mid \lvert j-k \rvert\}$	0122
$\alpha_6 = \inf\{t : t \mid i, t \mid j, t \nmid k\}$	0001
$\alpha_7 = \inf\{t : t \mid i, t \nmid j, t \mid k\}$	0010
$\alpha_8 = \inf\{t : t \nmid i, t \mid j, t \mid k\}$	0100
$\alpha_9 = \inf\{t : t \nmid i, t \mid \lvert i-j \rvert, t \mid \lvert i-k \rvert\}$	0111
$\alpha_{10} = \inf\{t : t \mid i, t \nmid j, t \mid \lvert j-k \rvert\}$	0011
$\alpha_{11} = \inf\{t : t \nmid i, t \mid j, t \mid \lvert i-k \rvert\}$	0101
$\alpha_{12} = \inf\{t : t \nmid i, t \mid k, t \mid \lvert i-j \rvert\}$	0110
$\alpha_{13} = \inf\{t : t \nmid i, t \nmid k, t \mid \lvert i-j \rvert\}$	0112
$\alpha_{14} = \inf\{t : t \nmid i, t \nmid j, t \mid \lvert i-j \rvert\}$	0120

Table 1. Definition of the values α_i, with $1 \leq i \leq 14$.

Now based on combinatorial relations, we define the some collections of sets. The idea behind all these collections is to generate sets of parameters α_is that cannot be avoided and have a minimal cardinality. No matter what will be added to these sets will preserve their unavoidability, while erasing something from them will make them avoidable. To obtain these collections we used a computer program and randomly generated some unavoidable sets of parameters of size five. Using the similarities between the instances modelled by these sets, defined in terms of (gapped) squares and cubes occurring in their digit representation, we developed an algorithm to generate more sets of patterns. Based on these relations we constructed fourteen sets $\mathcal{S}_i$s . We just define one of them as an example.

Let $\mathcal{S}_1$ be the collection of sets (each with five elements) that contain α_1 and:

- one of the α_is whose representation has a prefix or a suffix square, but no gapped cube. That is: α_2 or α_5.
- one of the α_is that has a gapped square, but does not have two gapped squares. These are α_3 or α_4.
- one of the α_is that contain cubes or two squares: α_6 or α_9 or α_{10}.
- one of the α_is that contain gapped cubes: α_7 or α_8.

For example, one possible set from $\mathcal{S}_1$ is $\{\alpha_1, \alpha_2, \alpha_4, \alpha_6, \alpha_7\}$.

Lemma 1 *Let $K' \subset K$ be any subset of size at most 4 of K. There exists an infinite word w such that w does not contain 4-powers and if w contains an instance of the pattern $x\pi^i(x)\pi^j(x)\pi^k(x)$ then it can not be modelled by any tuples of the set of patterns K'.*

Theorem 1. *Let i, j, k be positive integers such that $i \neq j \neq k \neq i$, and consider the pattern $p = x\pi^i(x)\pi^j(x)\pi^k(x)$. Let $\sigma = \min\{\max(S) \mid S \in \cup_{1 \leq i \leq 13} S_i\}$. Then $\sigma \geq 5$ and p is unavoidable in Σ_m, for $m \geq \sigma$.*

Proof. We briefly prove this Theorem. The complete proof is in the main paper. We checked with the aid of a computer, by a straightforward backtracking algorithm, that if $m \geq \max(S)$, for some $S \in \cup_{1 \leq i \leq 13} S_i$, then p is unavoidable in Σ_m. Our computer program tries construct a word as long as possible by always adding a letter to the current word it constructed by backtracking; this letter is chosen in all possible ways from the letters contained in the word already, or it may also be a new letter. □

4 Algorithm to generate avoidable cases

Algorithm 1 Algorithm to generate avoidable cases

1: Let $n = 13$. Using the sets $\mathcal{S}_i, (1 \leq i \leq 13)$, generate all sets of α_is of cardinality n, that have no unavoidable sets of patterns as subset; show that they are avoidable;
2: For all n from 12 downto 4, generate all sets of cardinality n that have no unavoidable sets of patterns as subset; these sets should not be subsets of the avoidable sets of α_is of cardinality $n+1$ (to avoid generating repetitive avoidable sets of cases generated in the past step); show that they are avoidable.

Theorem 2. *Given a pattern $p = x\pi^i(x)\pi^j(x)\pi^k(x)$ we can determine effectively the values m, such that the pattern is avoidable in Σ_m.*

Proof. We briefly prove this Theorem. The complete proof is in the main paper. In Theorem 1, we proved that given the pattern$p = x\pi^i(x)\pi^j(x)\pi^k(x)$, for each i, j, and k, we can compute the cardinality of an alphabet over which the pattern is unavoidable. Now to show that this is the minimum cardinality over which the pattern of size four is unavoidable, we proceed as follows. We will show that the subsets and the complements of all the sets $\mathcal{S}_i$,$(1 \leq i \leq 13)$ are avoidable. By complement we mean here all the sets of parameters α_i of which the sets Patterns of Size Four with Morphic Permutations are not subsets of. The reason to define it this way is that if a set of parameters is unavoidable, whatever we add to it remains unavoidable, so this set should not be subset of any avoidable set of patterns. Furthermore, to show that the value $\alpha = \alpha_i$, $1 \leq i \leq 14$ in the set $\mathcal{S}_j, 1 \leq j \leq 13$, is the minimum cardinality of an alphabet over which the pattern of size four is unavoidable, we should prove that α_i is the minimum value such that if we add it to the set $\mathcal{S}_j \backslash \alpha_i$, makes it unavoidable set of patterns. To reach this, we proved that all proper subsets of the sets $\mathcal{S}_j, 1 \leq j \leq 13$ are avoidable sets of patterns. □

On closed and open factors of Arnoux-Rauzy words *

Olga Parshina[1,2] and Luca Zamboni[1]

[1]Université de Lyon, Université Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 boulevard du 11 novembre 1918, F69622 Villeurbanne Cedex, France
[2]Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Science, 4 Acad. Koptyug avenue, 630090 Novosibirsk, Russia, {parshina,zamboni}@math.univ-lyon1.fr

Given a finite non-empty set $\mathbb{A}$, let $\mathbb{A}^{\mathbb{N}}$ denote the set of (right) infinite words $x = x_1x_2x_3\cdots$ with $x_i \in \mathbb{A}$. For each infinite word $x = x_1x_2x_3\cdots \in \mathbb{A}^{\mathbb{N}}$, the factor complexity $p_x(n)$ counts the number of distinct blocks (or factors) $x_ix_{i+1}\cdots x_{i+n-1}$ of length n occurring in x. First introduced by Hedlund and Morse in their seminal 1938 paper [13] under the name of *block growth*, the factor complexity provides a useful measure of the extent of randomness of x. Periodic words have bounded factor complexity while digit expansions of normal numbers have maximal complexity. A celebrated theorem of Morse and Hedlund in [13] states that every aperiodic (meaning not ultimately periodic) word contains at least $n+1$ distinct factors of each length n. Sturmian words are those aperiodic words of minimal factor complexity: $p_x(n) = n+1$ for each $n \geq 1$.

Other notions of complexity have been successfully used in the study of infinite words and their combinatorial properties [1, 5, 6, 7, 15, 16]. In this note, we introduce and study two new complexity functions based on the notions of open and closed words [8]. We recall that a word $u \in \mathbb{A}^+$ is said to be *closed* if either $u \in \mathbb{A}$ or if u is a complete first return to some proper factor $v \in \mathbb{A}^+$, meaning u has precisely two occurrences of v, one as a prefix and one as a suffix. Otherwise, if u is not closed then u is *open*. For example, *abbbab* and *aabaaabaa* are both closed words while *ab* and *abaabbababbaaba* are both open. It is easily seen that all privileged words [15] are closed and hence so are all palindromic factors of rich words [9]. The terminology open and closed was first introduced by the authors in [3] although the notion of a closed word had already been introduced earlier by A. Carpi and A. de Luca in [4]. For a nice overview of open and closed words we refer the reader to the recent survey article by G. Fici [8].

To each infinite word $x \in \mathbb{A}^{\mathbb{N}}$ we consider the functions $f_x^c, f_x^o : \mathbb{N} \to \mathbb{N}$ which count the number of closed and open factors of x of each length $n \in \mathbb{N}$. We study the behaviour of these complexity functions for Arnoux-Rauzy words [2]. Recall an infinite word $x \in \mathbb{A}^{\mathbb{N}}$ is called an *Arnoux-Rauzy word* if it is recurrent and if x contains, for each $n \geq 0$, precisely one right special factor of length n which is a prefix of $|\mathbb{A}|$-many factors of x of length $n+1$ and precisely one left special factor of length n which is a suffix of $|\mathbb{A}|$-many factors of x

*This work was performed within the framework of the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR), and has been supported by RFBS grant 18-31-00009

of length $n+1$. In particular one has $p_x(n) = (|\mathbb{A}|-1)n+1$ and each factor u of x has precisely $|\mathbb{A}|$ distinct complete first returns. Arnoux-Rauzy words were first introduced in [2] in the special case of a 3-letter alphabet. Let us note that in case $|\mathbb{A}| = 2$, then x is Sturmian. Since for any word $x \in \mathbb{A}^{\mathbb{N}}$ we have that $f^c_x(n) + f^o_x(n) = p_x(n)$, it suffices to understand the behaviour of $f^c_x(n)$.

Our main result in Theorem 1 below provides an explicit formula for the closed complexity function $f^c_x(n)$ for an Arnoux-Rauzy word x on a t-letter alphabet $\mathbb{A}$. The formula is expressed in terms of two related sequences associated to x. The first is the sequence $(b_k)_{k\geq 0}$ of the lengths of the bispecial factors $\varepsilon = B_0, B_1, B_2, \ldots$ of x, ordered in increasing length. The second is the sequence $(p^{(k)}_a)^{k\geq 0}_{a\in\mathbb{A}}$ where for each $k \geq 0$, the t coordinates of $(p^{(k)}_a)_{a\in\mathbb{A}}$ are the lengths of the t first returns in x to B_k. More precisely, $p^{(k)}_a = |R^{(k)}_a| - b_k$ where $R^{(k)}_a$ is the complete first return to B_k in x beginning in $B_k a$. Both sequences have already been extensively studied in the literature. In particular, following [11] one has that

$$b_k = \frac{\sum_{a\in\mathbb{A}} p^{(k)}_a - t}{t-1}.$$

Furthermore, for each $k \in \mathbb{N}$, the coordinates of $(p^{(k)}_a)_{a\in\mathbb{A}}$ are coprime and each is a period of the word B_k. Moreover, B_k is an extremal Fine and Wilf word i.e., any word u having periods $(p^{(k)}_a)_{a\in\mathbb{A}}$ and of length greater than b_k is a constant word, i.e., $u = a^n$ for some n (see [17]). The sequence $(p^{(k)}_a)^{k\geq 0}_{a\in\mathbb{A}}$ is computed recursively as follows : $p^{(0)}_a = 1$ for each $a \in \mathbb{A}$. For $k \geq 1$, let $a \in \mathbb{A}$ be the unique letter such that aB_{k-1} is a right special factor of x. Then $p^{(k)}_a = p^{(k-1)}_a$, and $p^{(k)}_b = p^{(k-1)}_b + p^{(k-1)}_a$ for $b \in \mathbb{A} \setminus \{a\}$.

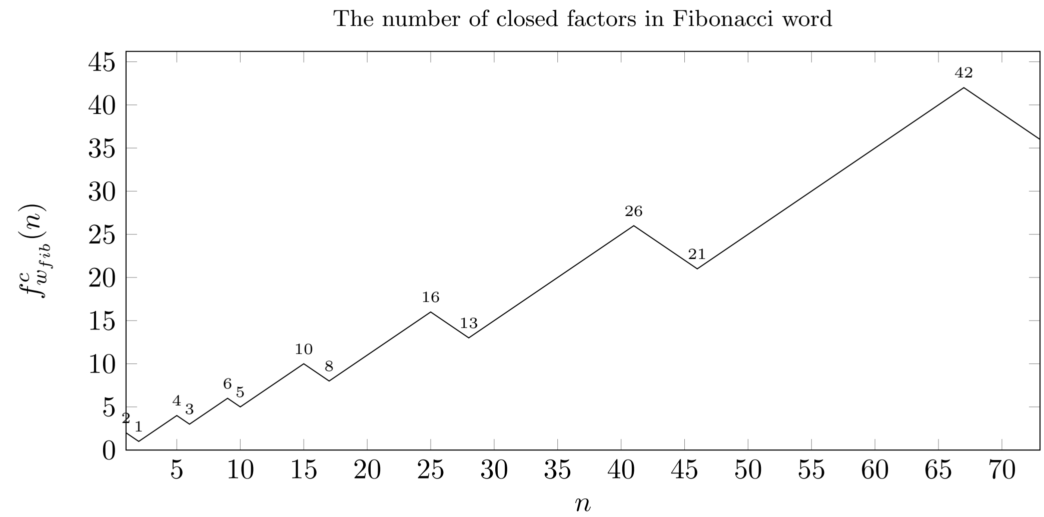
Theorem 1. *Let $x \in \mathbb{A}^{\mathbb{N}}$ be an Arnoux-Rauzy word. For each $k \in \mathbb{N}$ and $a \in \mathbb{A}$ set $I_{k,a} = [b_{k-1} - p_k + p^{(k)}_a + 2, b_k + p^{(k)}_a]$ where $p_k = \min_{b\in\mathbb{A}}\{p^{(k)}_b\}$. Let*

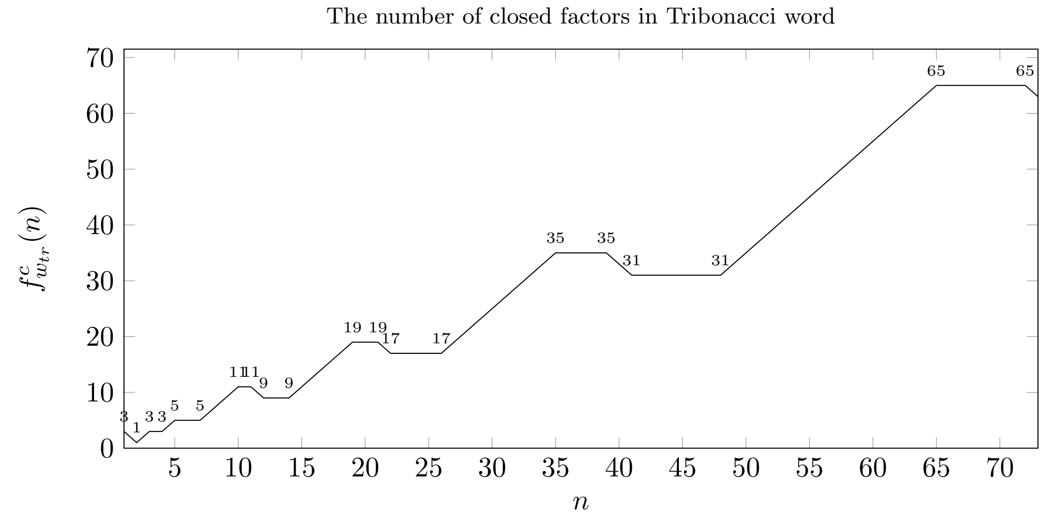
$$F(a,n) = \sum_{\substack{k\in\mathbb{N} \\ n\in I_{k,a}}} (d(n, I_{k,a}) + 1) \tag{1}$$

where for $n \in I_{k,a}$, the quantity $d(n, I_{k,a})$ denotes the minimal distance from n to the endpoints of the interval $I_{k,a}$. Then the number of closed factors of x for each length n is $f^c_x(n) = \sum_{a\in\mathbb{A}} F(a,n)$.

It is easily checked that the length of each interval $I_{k,a}$ is $2p_k - 2$ and that for each fixed n the sum in (1) is actually a finite sum.

The following figures illustrate the behaviour of the closed complexity function f^c_x in the case of the Fibonacci word and the Tribonacci word.





It is evident that in general f_x^c is not monotone. However as a consequence to Theorem 1 we are able to show :

Corollary 2. *Let x be an Arnoux-Rauzy word. Then* $\liminf f_x^c(n) = +\infty$.

In contrast, it is shown in [14] that for any paperfolding word x, $\liminf f_x^c(n) = 0$, in other words, for infinitely many n, x has no closed factors of length n.

References

[1] J.-P. Allouche, M. Baake, J. Cassaigne and D. Damanik, Palindrome complexity, Selected papers in honor of Jean Berstel, Theoret. Comput. Sci., 292 (2003), pp. 9–31.

[2] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité $2n+1$, Bull. Soc. Math.France 119 (2) (1991) pp. 199–215.

[3] M. Bucci, A. De Luca, G. Fici, Enumeration and Structure of Trapezoidal Words, Theoret. Comput. Sci. 468 (2013) pp. 12–22.

[4] A. Carpi, A. de Luca, Periodic-like words, periodicity and boxes, Acta. Inform. 37 (2001), pp. 597–618.

[5] J. Cassaigne, G. Fici, M. Sciortino and L.Q. Zamboni, Cyclic complexity of words, J. Combin. Theory Ser. A, 145 (2017), pp. 36–56.

[6] E. Charlier, S. Puzynina, L.Q. Zamboni, On a group theoretic generalization of the Morse-Hedlund theorem, Proceedings of the AMS, 145 (2017), pp. 3381–3394.

[7] E. M. Coven, G. A. Hedlund, Sequences with minimal block growth, Math. Systems Theory 7 (1973), pp. 138–153.

[8] G. Fici, Open and closed words, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS No. 123 (2017), p. 138–147.

[9] A. Glen, J. Justin, S. Widmer, L.Q. Zamboni, Palindromic richness, European J. Combin. 30 (2009), no. 2, pp. 510–531.

[10] A. de Luca, F. Mignosi, Some combinatorial properties of Sturmian words, Theoret. Comput. Sci., 136 (1994), pp. 361–385

[11] J. Justin, On a paper by Castelli, Mignosi, Restivo, RAIRO: Theoret.Informatics Appl. 34 (2000),pp. 373–377.

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

[13] M. Morse, G. Hedlund, Symbolic dynamics, Amer. J. Math. 60 (1938), pp. 815–866.

[14] O. Parshina, On the number of closed factors in the paperfolding word, preprint in preparation.

[15] J. Peltomäki, Introducing privileged words: privileged complexity of Sturmian words, Theoret. Comput. Sci. 500 (2013), pp. 57–67.

[16] G. Richomme, K. Saari and L.Q. Zamboni, Abelian complexity of minimal subshifts, J. Lond. Math. Soc. (2), 83 (2011), pp. 79–95.

[17] R. Tijdeman, L.Q. Zamboni, Fine and Wilf words for any periods, Indag. Math. (N.S.) 14 (2003), pp. 135–147.

ON THE GROUP OF A RATIONAL MAXIMAL BIFIX CODE

JORGE ALMEIDA, ALFREDO COSTA, REVEKKA KYRIAKOGLOU, AND DOMINIQUE PERRIN

Abstract. We give necessary and sufficient conditions for the group of a rational maximal bifix code Z to be isomorphic with the F-group of $Z \cap F$, when F is recurrent and $Z \cap F$ is rational. The case where F is uniformly recurrent receives special attention.

1. Introduction

In the past few years, special attention has been given to bifix codes which may not be maximal but are maximal within some language, which is usually chosen to be recurrent or uniformly recurrent. This line of research has produced new and strong connections between bifix codes, subgroups of free groups and symbolic dynamical systems (cf. [4] and the sequels [5, 6, 7, 8]).

If Z is a thin maximal bifix code and F is a recurrent set, then $X = Z \cap F$ is an F-maximal bifix code, that is, a maximal bifix code within F (with X finite if F is uniformly recurrent) [4]. This leads to a process of "relativization" of several previously known definitions for maximal codes. An important example is the group $G(Z)$ of a rational code Z, i.e., the Schützenberger group of the minimum ideal $J(Z)$ of the syntactic monoid $M(Z^*)$ of Z^*. In this case, the relativization consists in taking the intersection $X = Z \cap F$ and the Schützenberger group of the minimum $\mathcal{J}$-class $J_F(X)$ that intersects the image of F in $M(X^*)$, when X is rational. This group, denoted by $G_F(X)$, is the *F-group of X*. How are $G(Z)$ and $G_F(X)$ related? They are not always isomorphic, even if Z is a group code (i.e., Z is a code with $M(Z^*)$ a finite group) and F is uniformly recurrent. In [4] it is shown that if Z is a group code and F is Sturmian, then $G(Z)$ and $G_F(X)$ are isomorphic. This was extended to tree sets in the manuscript [10], thanks to a novel approach consisting in exploring links between $G(Z)$, $G_F(X)$ and the Schützenberger (profinite) group $G(F)$ of the minimum $\mathcal{J}$-class $J(F)$ of the topological closure of F within the free profinite monoid generated by the alphabet of F, and, with the help of these links, taking advantage of results on $G(F)$ from [2, 3]. Building on this approach, we get new results about when $G(Z) \simeq G_F(X)$ holds, recovering previous results in the process.

2. Preliminaries on free profinite monoids

Here A is always a finite alphabet. Take $u, v \in A^*$. If $u \neq v$, there is a homomorphism $\varphi\colon A^* \to M$ onto a finite monoid such that $\varphi(u) \neq \varphi(v)$.

2010 *Mathematics Subject Classification.* 20M05, 20E18, 37B10, 68R15.

Key words and phrases. maximal bifix code, rational code, group code, syntactic monoid, F-group, Schützenberger group, uniformly recurrent set, free profinite monoid.

Let $r(u,v)$ be the minimum for $|M|$. We may consider the metric d on A^* such that $d(u,v) = 2^{-r(u,v)}$ if $u \neq v$. The *free profinite monoid* $\widehat{A^*}$ is the compact monoid resulting from the completion of A^* under d, a terminology justified as $\widehat{A^*}$ is the free object in the class of A-generated profinite monoids. For an extended introduction, see [11]. The elements of $\widehat{A^*}$ are called *pseudowords* over A. Words of A^* are topologically isolated in $\widehat{A^*}$. Pseudowords generalize words, but the structure of $\widehat{A^*}$ is richer than that of A^*. Next is a glimpse of that. If F is a factorial subset of A^*, then the topological closure $\overline{F}$ is itself factorial in $\widehat{A^*}$, and when F satisfies the stronger property of being recurrent, there is a minimum $\mathcal{J}$-class $J(F)$ contained in $\overline{F}$, which moreover is regular. Maximal subgroups of $J(F)$ were identified in [2, 3]. We mention that the proper factors (that is, strictly $\mathcal{J}$-above) of $u \in \widehat{A^*} \setminus A^*$ belong to A^* if and only if $u \in J(F)$ for some uniformly recurrent set F [1].

3. Preparatory technical results

In this section we prepare the main results of the next section.

Recall that a *parse* of a word w with respect to a subset X of A^* is a triple (v,x,u) such that $w = vxu$ with $v \in A^* \setminus A^*X$, $x \in X^*$ and $u \in A^* \setminus XA^*$. The number of parses of w with respect to X is denoted by $\delta_X(w)$. The F-*degree* of X is $d_F(X) = \sup\{\delta_X(w) \mid w \in F\}$. The *degree* of X is $d(X) = d_{A^*}(X)$. For F recurrent containing a bifix code X, one has $d_F(X)$ finite if and only if X is F-thin and an F-maximal bifix code [4].

The notion of parse was generalized to pseudowords in [10], and so we may extend δ_X to $\widehat{A^*}$. Since then, we obtained the following useful tool.

Proposition 3.1. *Consider a factorial set F of A^*. Let X be a rational subset of F with finite F-degree d. Then $\delta_X(w) \leq d$ for every $w \in \overline{F}$, and the mapping $\delta_X \colon \overline{F} \to \{1,\dots,d\}$ thus defined is continuous, if we endow $\{1,\dots,d\}$ with the discrete topology.*

A pseudoword u is *forbidden in* $Y \subseteq \widehat{A^*}$ if u is not a factor of an element of Y. The next proposition was deduced with the help of Proposition 3.1. We explain the notation used there. Let L be a rational language of A^*. By the universal property of $\widehat{A^*}$, the syntactic homomorphism $\eta_L \colon A^* \to M(L)$ admits a unique extension to a continuous homomorphism $\hat{\eta}_L \colon \widehat{A^*} \to M(L)$.

Proposition 3.2. *Let Z be a rational maximal bifix code of A^*. Suppose that F is a recurrent subset of A^* and that the intersection $X = Z \cap F$ is rational. The equality $d_F(X) = d(Z)$ holds if and only if the elements of $J(F)$ are forbidden in $\overline{Z}$. Moreover, if $d_F(X) = d(Z)$ then $\hat{\eta}_{Z^*}(J(F)) \subseteq J(Z)$.*

Consider a language L of A^*. Let $u,v \in A^*$. By definition, $\eta_L(u) \leq \eta_L(v)$ if and only if the context of u is contained in the context of v.

Proposition 3.3. *Let Z and F be subsets of A^*, with F factorial, and let $X = Z \cap F$. Suppose Z^* and X^* are rational. Let $e,f \in \widehat{A^*} \setminus \{1\}$ be idempotents, and let $u,v \in \widehat{A^*}$ with $u = euf$, $v = evf$ and $u \in \overline{F}$. Then:*

(1) $\hat{\eta}_{X^*}(u) \leq \hat{\eta}_{X^*}(v) \Rightarrow \hat{\eta}_{Z^*}(u) \leq \hat{\eta}_{Z^*}(v)$, if e and f are forbidden in $\overline{Z}$;
(2) $\hat{\eta}_{Z^*}(v) \leq \hat{\eta}_{Z^*}(u) \Rightarrow \hat{\eta}_{X^*}(v) \leq \hat{\eta}_{X^*}(u)$, if e and f are forbidden in $\overline{X}$.

Applying the preceeding tools, we deduce relationships between the maximal subgroups of $J(F)$, $J(Z)$ and $J_F(Z \cap F)$, for suitable Z and F.

Theorem 3.4. *Let F be a factorial subset of A^*, take a rational prefix code Z of A^*, and suppose $X = Z \cap F$ is a rational F-maximal prefix code. Let H be a maximal subgroup of $\widehat{A^*}$ with $H \subseteq \overline{F}$ and the elements of H being forbidden in $\overline{Z}$. Consider the maximal subgroup H_X of $M(X^*)$ containing $\hat{\eta}_{X^*}(H)$ and the maximal subgroup H_Z of $M(Z^*)$ containing $\hat{\eta}_{Z^*}(H)$. There is an injective homomorphism $\alpha \colon H_X \to H_Z$ such that the diagram*

$$\begin{tikzcd} H \arrow[r, "\hat{\eta}_{Z^*}"] \arrow[d, "\hat{\eta}_{X^*}"'] & H_Z \\ H_X \arrow[ur, "\alpha"'] & \end{tikzcd} \tag{3.1}$$

commutes.

4. Main results

Based on the technical results of the previous section, namely Theorem 3.4, we deduce our main results.

Let F be a recurrent set of A^*. Say that a rational code Z of A^* is F-*charged* if $\hat{\eta}_{Z^*}$ maps maximal subgroups of $J(F)$ onto maximal subgroups of $J(Z)$. A rational code X contained in F is *weakly F-charged* if $\hat{\eta}_{X^*}$ maps maximal subgroups of $J(F)$ onto maximal subgroups of $J_F(X)$.

Theorem 4.1. *Consider a recurrent subset F of A^* and a rational bifix code Z of A^* with finite degree such that $X = Z \cap F$ is rational. Let H be a maximal subgroup of $J(F)$. The following conditions are equivalent:*

(1) Z is F-charged;
(2) $d_F(X) = d(Z)$, $G_F(X) \simeq G(Z)$ and X is weakly F-charged;
(3) $d_F(X) = d(Z)$, $|G_F(X)| = |G(Z)|$ and X is weakly F-charged.

Recall that if $F \subseteq A^*$ is (uniformly) recurrent and Z is a maximal bifix code of A^*, then $Z \cap F$ is an F-maximal bifix (finite) code [4].

We show that a group code of A^* is F-charged when F is an uniformly recurrent *connected* set (that is, with only connected extension graphs, see [5]) with alphabet A. Therefore, we get the following corollary.

Corollary 4.2. *If Z is a group code of A^* and F is a uniformly recurrent connected set of alphabet A, then $d(Z) = d_F(Z \cap F)$ and $G(Z) \simeq G_F(Z \cap F)$.*

We say that a rational code Z is *nil-simple* if all idempotents of $M(Z^*)$ are in $J(Z)$. Group codes and finite codes are nil-simple. If F is uniformly recurrent and Z is nil-simple, the equality $d_F(X) = d(Z)$ in Theorem 4.1 becomes redundant, as seen next.

Theorem 4.3. *Let Z be a uniformly recurrent subset of A^*, and let Z be a nil-simple rational maximal bifix code Z of A^*. The following are equivalent:*

(1) Z is F-charged;
(2) $G_F(Z \cap F) \simeq G(Z)$ and $Z \cap F$ is weakly F-charged;
(3) $|G_F(Z \cap F)| = |G(Z)|$ and $Z \cap F$ is weakly F-charged.

Moreover, if Z is F-charged, then $d(Z) = d_F(Z \cap F)$.

The special case of Corollary 4.2 in which Z is a group code and F is Sturmian was first proved in [4].

We also studied F-groups as permutation groups acting in a natural manner. In what follows, Q_Y is the set of vertices of the minimal automaton of Y^*, and i_Y is the corresponding initial state.

Theorem 4.4. *Let F be a recurrent subset of A^*. Suppose that Z is a rational bifix code of finite degree d. Let $X = Z \cap F$ and suppose that X is rational. Let H be a maximal subgroup of $J(F)$ such that $H_Z = \hat{\eta}_{Z^*}(H)$ is a maximal subgroup of $J(Z)$, and let $H_X = \hat{\eta}_{X^*}(H)$. Take the map $f\colon Q_X \cdot H_X \to Q_Z \cdot H_Z$ given by $f(i_X \cdot u) = i_Z \cdot u$, for $u \in H$, and take the unique group isomorphism $\alpha\colon H_X \to H_Z$ such that Diagram (3.1) commutes. Then the pair (f, α) is an equivalence of permutation groups with degree d.*

References

1. J. Almeida, *Profinite groups associated with weakly primitive substitutions*, Fundamentalnaya i Prikladnaya Matematika (Fundamental and Applied Mathematics) **11** (2005), 13–48, In Russian. English version in J. Math. Sciences **144**, No. 2 (2007) 3881–3903.
2. J. Almeida and A. Costa, *Presentations of Schützenberger groups of minimal subshifts*, Israel J. Math. **196** (2013), 1–31.
3. ______, *A geometric interpretation of the Schützenberger group of a minimal subshift*, Ark. Mat. **54** (2016), 243–275.
4. J. Berstel, C. De Felice, D. Perrin, Ch. Reutenauer, and G. Rindone, *Bifix codes and Sturmian words*, J. Algebra **369** (2012), 146–202.
5. V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone, *Acyclic, connected and tree sets*, Monatsh. Math. **176** (2015), 521–550.
6. ______, *Bifix codes and interval exchanges*, J. Pure Appl. Algebra **219** (2015), 2781–2798.
7. ______, *The finite index basis property*, J. Pure Appl. Algebra **219** (2015), 2521–2537.
8. ______, *Maximal bifix decoding*, Discrete Math. **338** (2015), no. 5, 725–742.
9. A. Costa and B. Steinberg, *Profinite groups associated to sofic shifts are free*, Proc. London Math. Soc. **102** (2011), 341–369.
10. R. Kyriakoglou and D. Perrin, *Profinite semigroups*, Tech. report, 2017, arXiv:1703.10088 [math.GR].
11. J.-E. Pin, *Profinite methods in automata theory*, vol. 26th International Symposium on Theoretical Aspects of Computer Science (STACS 2009), Internationales Begegnungs-Und Forschungszentrum für Informatik (IBFI), 2009, pp. 31–50.

CMUP, Departamento de Matemática, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal.
E-mail address: jalmeida@fc.up.pt

CMUC, Department of Mathematics, University of Coimbra, Apartado 3008, EC Santa Cruz, 3001-501 Coimbra, Portugal.
E-mail address: amgc@mat.uc.pt

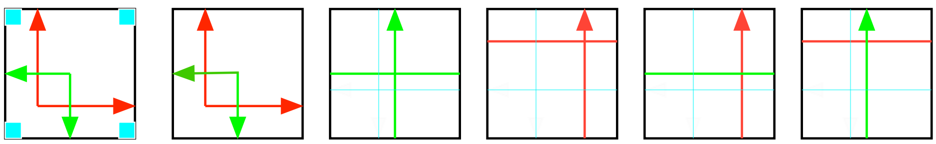
LIGM, Université Paris-Est
E-mail address: krevekka@hotmail.gr

LIGM, Université Paris-Est
E-mail address: dominique.perrin@esiee.fr

Complexity of Robinson tiling
Abstract

Galanov Ilya
galanov@lipn.univ-paris13.fr

Raphael Robinson in his work on undecidability of the domino problem [1] introduced the set of six tiles depicted in Figure 1. The tiles can be rotated and reflected, one tile can fit to the other only in such a way that the arrow head matches arrow tail and each 2×2 block must contain exactly one *bumpy corner*, the leftmost in Figure 1.



(a) (b) (c) (d) (e) (f)

Figure 1: Tiles of type (a) are called *bumpy corners*, tiles of type (b) are called *corners*, all the other tiles are called *arms*.

It is possible to tile the Euclidean plane with copies of this six tiles, but only in an *aperiodic* way. The key to this result is that any Robinson tiling has a hierarchical structure (see Figure 2). For all n, it is possible to define the *supertiles* of rank n inductively. Bumpy corner tiles are said to be supertiles of the first rank, supertiles of second and third rank are shown in Figure 2. An increasing union of supertiles of rank n is called an infinite order supertile. The Robinson tiling can be either made of only *one* infinite order supertile or contain *two* or *four* infinite order supertiles.

We will prove that the number of distinct blocks of size $n \times n$ (with $n \geq 2$) that could appear in a Robinson tiling made of one infinite order supertile is defined by the formula

$$A(n) = 32n^2 + 72n \cdot 2^{\lfloor \log_2 n \rfloor} - 48 \cdot 2^{2\lfloor \log_2 n \rfloor}.$$

Similar method for counting the number of factors in the *paper folding sequence* was used by Jean-Paul Allouche in [2].

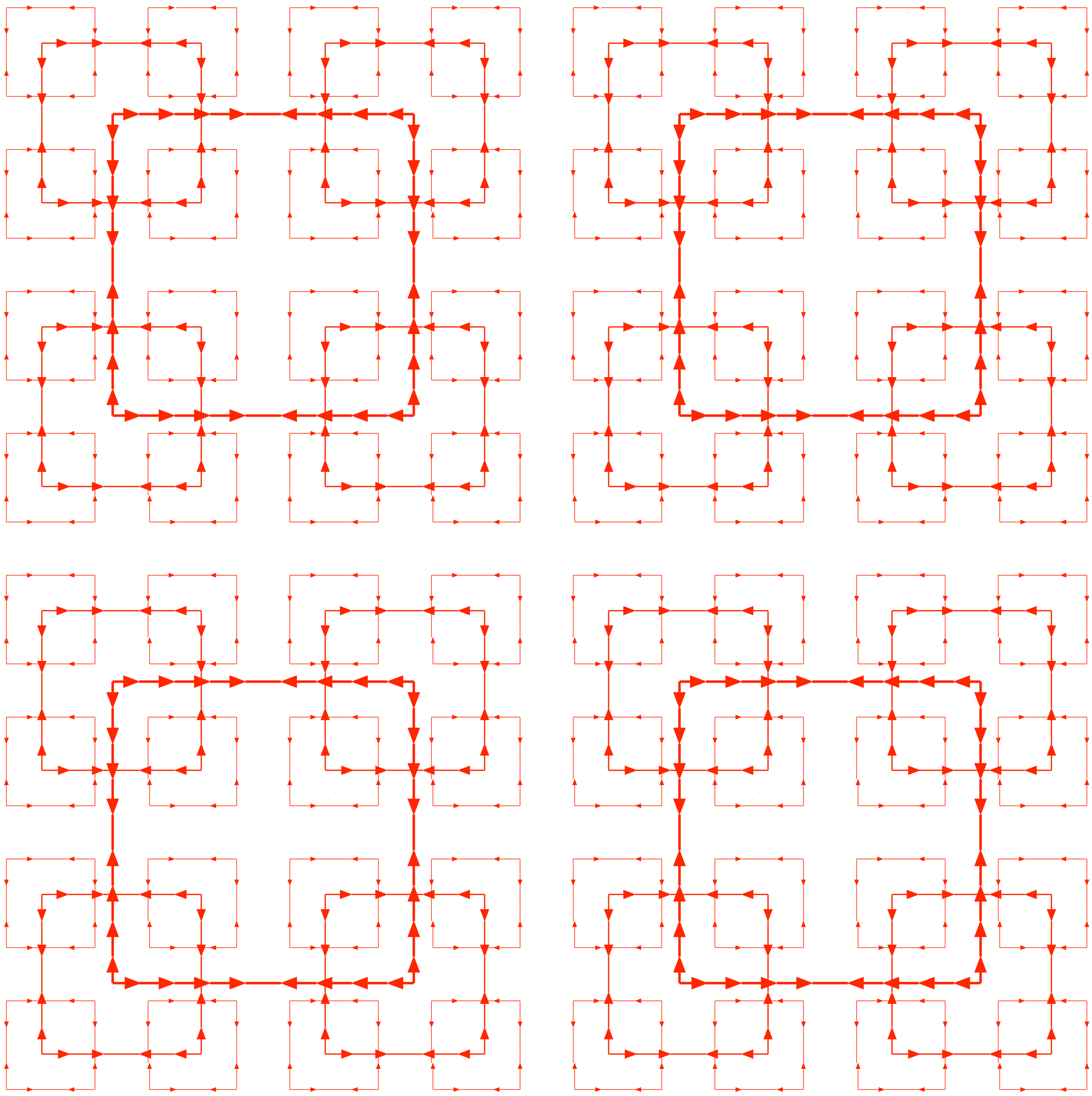
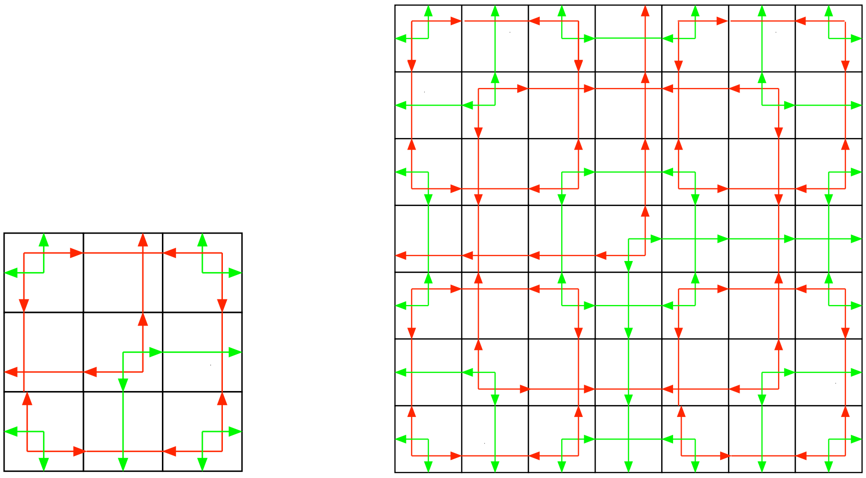


Figure 2: Hierarchy.

2



(a) (b)

Figure 3: Supertiles of second and third rank.

References

[1] R. Robinson, *Undecidability and nonperiodicity for tilings of the plane*, Invent. Math. **12** (1971) 177–209.

[2] J.-P. Allouche, *The number of factors in a paperfolding sequence*, Bull. Austral. Math. Soc. **46**, 23–32 (1992).

Abelian Anti-Powers in Infinite Words

Gabriele Fici
Dipartimento di Matematica e Informatica, Università di Palermo, Italy
gabriele.fici@unipa.it

Mickael Postic
Institut Camille Jordan, Université Claude Bernard Lyon 1, Lyon, France
postic@math.univ-lyon1.fr

Manuel Silva
Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Lisbon, Portugal
mnasilva@gmail.com

Abstract

We introduce and study the notion of an abelian anti-power in the context of combinatorics on words. An abelian anti-power of order k (or simply an abelian k–anti-power) is a concatenation of k consecutive words of the same length having pairwise distinct Parikh vectors. This definition therefore generalizes to the abelian setting the notion of a k–anti-power, as introduced in [G. Fici et al., *Anti-powers in infinite words*, J. Comb. Theory, Ser. A, 2018], that is a concatenation of k pairwise distinct words of the same length. In particular, we deal with the question to determine whether a word contains abelian k–anti-powers for arbitrarily large k. A word with bounded abelian complexity clearly cannot contain abelian anti-powers of arbitrary order. We show that the Sierpiǹski word (whose abelian complexity grows logarithmically) does not contain abelian 11–anti-powers. Another question is to find words with low factor complexity that contain both abelian powers and abelian anti-powers of arbitrary order. We show that all paperfolding words have this property.

1 Introduction

Many of the classical definitions in combinatorics on words (e.g., period, run, power, factor complexity, etc.) have a counterpart in the abelian setting, though they may not enjoy the same properties.

Recall that the Parikh vector $P(w)$ of a word w over a finite ordered alphabet $\mathbb{A} = \{a_1, a_2, \ldots, a_{|\mathbb{A}|}\}$ is the vector whose i-th component is equal to the number of occurrences of the letter a_i in w, $1 \leq i \leq |\mathbb{A}|$. For example, the Parikh vector of $w = abbca$ over $\mathbb{A} = \{a, b, c\}$ is $P(w) = (2, 2, 1)$. This notion is at the basis of the abelian combinatorics on words, where two words are considered equivalent if and only if they have the same Parikh vector.

The fundamental result of Morse and Hedlund [4] (an infinite word is aperiodic if and only if its factor complexity is unbounded) does not hold anymore in the case of the abelian complexity (the function that counts the number of distinct Parikh vectors of factors of length n for each n), as there exist aperiodic words with bounded abelian complexity. In fact, Richomme et al. [5] have observed that if a word has bounded abelian complexity, then it contains abelian powers of any order — an abelian power of order k is a concatenation of k words having the same Parikh vector. However, this is not a characterization of words with bounded abelian complexity. Madill and Rampersad proved that the regular paperfolding word has unbounded abelian complexity [3], and Štěpán Holub proved that it contains abelian powers of any order [2].

In a recent paper [1], the first and the third author, together with Antonio Restivo and Luca Zamboni, introduced the notion of an anti-power. An *anti-power of order k*, or simply a *k–anti-power*, is a concatenation of k consecutive pairwise distinct words of the same length. E.g., $aabaaabbbaba$ is a 4–anti-power.

In [1], it is proved that the existence of powers of any order or anti-powers of any order is an unavoidable regularity for infinite words:

Theorem 1. *[1] Every infinite word contains powers of any order or anti-powers of any order.*

Note that in the previous statement there is no hypothesis on the alphabet size.

In this paper, we extend the notion of an anti-power to the abelian setting.

Definition 2. An *abelian anti-power of order k*, or simply an *abelian k–anti-power*, is a concatenation of k consecutive words of the same length having pairwise distinct Parikh vectors.

For example, $aabaaabbbabb$ is an abelian 4–anti-power. Notice that an abelian k–anti-power is a k–anti-power but the converse does not necessarily holds (which is dual to the fact that a k–power is an abelian k–power but the converse does not necessarily holds).

We think that an analogous of Theorem 1 may still hold in the case of abelian anti-powers, but unfortunately the proof of Theorem 1 does not seem to be generalizable to the abelian setting.

Problem 1. Does every infinite word contain abelian powers of any order or abelian anti-powers of any order?

Clearly, if a word has bounded abelian complexity, then it cannot contain abelian anti-powers of arbitrary order. However, we show in this paper that the converse is not true. Indeed, we prove that the Sierpiǹski word does not contain abelian 11–anti-powers. The Sierpiǹski word has logarithmic abelian complexity (by construction) and contains abelian powers of any order (since it contains arbitrarily long blocks of bs).

An infinite word can contain both abelian powers of any order and abelian anti-powers of any order. This is the case, for example, of any word with full factor complexity. However, finding a class of words with low factor complexity satisfying this property seems a more difficult task. Indeed, most of the well-known examples of aperiodic words (Thue-Morse, Sturmian words, etc.) have bounded abelian complexity, hence they cannot contain abelian anti-powers of any order — whereas, by the aforementioned remark of Richomme et al. [5], they contain abelian powers of any order. Building upon the theory that Štěpán Holub developed to prove that all paperfolding words contain abelian powers of any order [2], we prove that all paperfolding words contains also abelian anti-powers of any order.

References

[1] G. Fici, A. Restivo, M. Silva, and L. Q. Zamboni. Anti-powers in infinite words. *J. Comb. Theory, Ser. A*, 157:109–119, 2018.

[2] S. Holub. Abelian powers in paper-folding words. *J. Comb. Theory, Ser. A*, 120(4):872–881, 2013.

[3] B. Madill and N. Rampersad. The abelian complexity of the paperfolding word. *Discr. Math.*, 313(7):831–838, 2013.

[4] M. Morse and G. A. Hedlund. Symbolic dynamics. *Amer. J. Math.*, 60:1–42, 1938.

[5] G. Richomme, K. Saari, and L. Zamboni. Abelian complexity of minimal subshifts. *J. Lond. Math. Soc.*, 83(1):79–95, 2011.

Inflation of digitally convex polyominoes

Jean-Pierre Borel[a], Lama Tarsissi[b], Laurent Vuillon[c]

[a]*Mathis-XLim, UMR 6172, Université de Limoges - CNRS, 123 avenue Albert Thomas F-87060 LIMOGES CEDEX - FRANCE*
[b]*Laboratoire I3S, Université Nice Sophia Antipolis*
Euclide B, 2000 Route des Lucioles, 06900 Sophia Antipolis, France
[c]*Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LAMA, 73000 Chambéry, France*

Keywords: Digital convexity, Discrete Geometry, Christoffel words, step by step inflation

1. Introduction

In the literature, many papers introduced and studied different convexity notions. For example, Kim and Rosenfeld in [1] investigated different notions of discrete convex sets, where a set in Euclidean geometry is convex if and only if for any pair of points p_1, p_2 in a region R, the line segment joining them is completely included in R. In discrete geometry on square grids, this notion refers to the digitally convex convexity. We recall that a *polyomino* is a finite 4-connected set of unit squares in the lattice $\mathbb{Z}^2$. If P is a polyomino and if for all p_1, p_2 inside P and such that the discret segment joining them is completely included in P then P is a digitally convex polyomino. Digitally convex polyominoes are also the discretization of convexes of $\mathbb{R}^2$, except when this discretization is not 4-connected. It implies that the intersection of two such polyominoes is also digitally convex, as soon as it is 4-connected.

In this talk, we would like to study discrete geometrical constructions to deflate or inflate digitally convex polyominoes.

2. How to deflate a digitally convex polyomino?

We first study how to deflate a digitally convex polyomino P. This polyomino P can be decomposed into four paths. Indeed, a polyomino P is a finite set thus we define the minimal bounded box which is a rectangle such that P touches the four sides of the rectangle. Those paths start at the first unit square of intersection with each side of the rectangle. We denote the unit squares of intersection by W (the lowest unit square on the leftmost side), N (the leftmost unit square in the top side), E (the highest unit square on the rightmost side) and S (the rightmost unit square on the bottom side). The contour of a convex polyomino is then the union of the four (clockwise) paths WN, NE, ES and SW.

We use the result given by Brlek et al. [2] where they introduced a link between convexity and combinatorics on words by encoding the contour of the convex polyomino. Their result was based on *Christoffel* and *Lyndon words.* They considered each path of the convex polyomino and used the alphabet $A = \{0, 1, \overline{0}, \overline{1}\}$ to code the boundary of each polyomino where $0, 1, \overline{0}, \overline{1}$ encode $\rightarrow, \uparrow, \leftarrow, \downarrow$ respectively.

The main result in [2] states that a convex polyomino is characterized by the fact that the WN path admits a unique *Lyndon factorization* $\ell_1^{n_1} \ldots \ell_k^{n_k}$ where the ℓ_i's have decreasing slopes and they are primitive Christoffel words, and similar results for the three other paths.

It is not difficult to deflate the digitally convex polyomino. Indeed, we look at the set $\{s_1, s_2, \cdots, s_{k+1}\}$ of unit squares (called corners) in the boundary of the polyomino such that each s_i are exactly the unit square corresponding to the extremities of each factor $\ell_j^{n_j}$ in the factorization given before (see Figure 1), or equivalently have a common unit square with the convex hull of the polyomino. Remove any square s_i, then the new polyomino is still digitally convex, and this process can be iterated.

This property has two consequences.

Email addresses: `jean-pierre.borel@unilim.fr` (Jean-Pierre Borel), `lama.tarsissi@unice.fr` (Lama Tarsissi), `laurent.vuillon@univ-smb.fr` (Laurent Vuillon)

Preprint submitted to Elsevier *May 28, 2018*

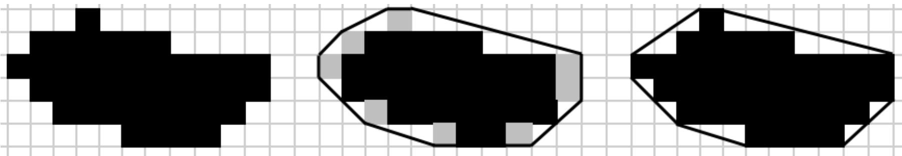


Figure 1: A digitally convex polyomino, its convex hull and its corners (grey unit squares) and a smaller digitally convex polyomino.

1. Given two digitally convex polyomino $C_1 \subset C_2$, C_2 can be deflated to C_1 in such a way that at each step one unit square only is canceled and such that at each step we have a digitally convex polyomino: as the polyomino C_2 is bigger than C_1 it contains a corner which is not in C_1, then cancel this corner.
2. It is possible to inflate step by step from C_1 to C_2, by the reverse process. However, it does not give a practical way to choose the unit square that we must add at each step.

3. How to inflate a digitally convex polyomino?

It is more difficult to get effective methods to inflate a convex polyomino C_1 to a convex polyomino C_2 where $C_1 \subset C_2$ with the constraint that we must add a single unit square at each step and maintain the digitally convex property at each step until reaching C_2.

3.1. The spiral and strate constructions

First of all we take C_1 as a unit square anywhere inside C_2.

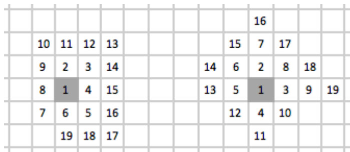


Figure 2: The spiral construction

We first investigate the spiral construction by adding a corner around the polyomino from a unit square polyomino by adding at each step a corner in a clockwise order (see Figure 2). This construction leads to an octogonal shape digitally convex polyomino. However, by keeping only those unit squares which are contained in C_2 (see Figure 3 left and center), we get a global construction of C_2 such that we maintain the convexity property at each step.

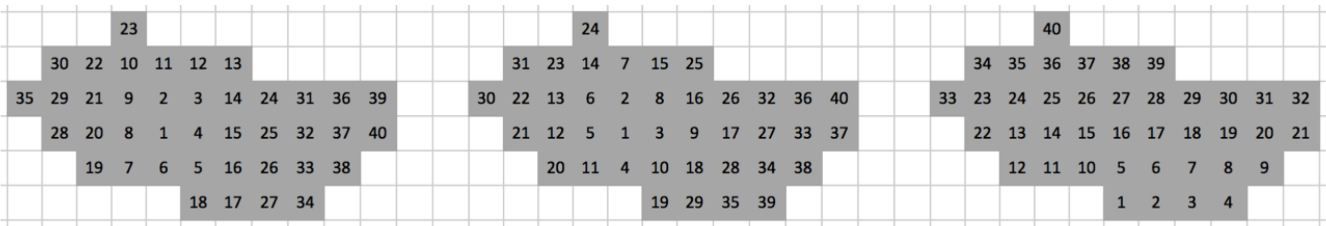


Figure 3: Two spiral and a strate construction of a digitally convex polyomino

In fact, we have many variants of this construction, by considering 4-connected spiral or 8-connected spiral or strate construction (see Figure 3). The strate construction consists in taking at first the lowest unit squares from the left to the right, and to continue to the second row in a correct order, and so on (see Figure 3, right). It has many variants.

If we try to inflate some digitally convex polyomino by the spiral method, it works only when we take special octogones (see Figure 4, left). In the general case, the convexity property disappears at some steps (see Figure 4, right, unit square number 8).

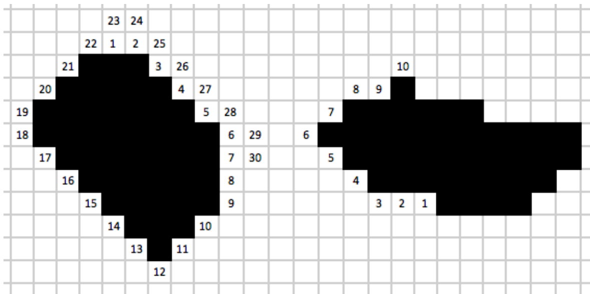


Figure 4: The spiral inflation

Thus now we must add unit squares using one construction of the paper of Dulio et al.[4], the authors introduces the *split* operator (based on the Borel Laubie standard factorization of Christoffel words [3]) in order to add unit square to the border of a digitally convex polyomino. By this operation the Christoffel property of words along the four paths is maintained and thus the difficult part of the inflation remains to maintain the monotony of the slopes of each Christoffel word along the four paths.

3.2. The split operator

From now on, we are working on the WN path. The split operator, as mentioned in [4], considers the furthest point of the Christoffel word ℓ_i with respect to its line segment and switches the factor 01 to 10. Using this operator and replacing the factors is equivalent to adding a unit square to the polyomino. We obtain in this case two new Christoffel words ℓ_i^+ and ℓ_i^-, where $\ell_i = \ell_i^- \ell_i^+$ is the standard factorization of the Christoffel word, and $\ell_i^- < \ell_i < \ell_i^+$ both for slopes and lexical order. The split operator consists in replacing ℓ_i by $\ell_i^+ \ell_i^-$ in the coding word of the path. Then the convexity property remains true in the simple case :

$$\ell_{i+1} \leq \ell_i^- < \ell_i^+ \leq \ell_{i-1}.$$

but these inequalities can fail.

So we might have different cases and situations, that we will present in the next section depending on the property of each unit square to add.

3.3. Adding one unit square:

We start by considering the case, where we will add one unit square to the convex $C_1(j)$ for a given j ($C_1(j)$ represents the jth step of the construction that is C_1 with j added unit squares). Which means that with this case we reach the step $C_1(j+1)$. In fact, three different cases can occur:

1. The first case, where we add a unit square on the Christoffel word ℓ_i with $n_i = 1$ of a given path using the split operator and no problems are faced. Which means the monotone order of slopes is maintained, i.e., as the split operator consists in replacing ℓ_i by $\ell_i^+ \ell_i^-$ in the coding word of the path:

   $$\ell_{i+1} \leq \ell_i^- < \ell_i^+ \leq \ell_{i-1}.$$

   and the convexity is conserved.

2. The second case also needs $n_i = 1$. It is if we add a unit square to ℓ_i of a given path we keep the Lyndon factorization property, but this factorization is not as in the first case

   $$\ell_1^{n_1} \ldots (\ell_{i-1}^{n_{i-1}} \ell_i^+) \ell_i^- \ell_{i+1}^{n_{i+1}} \ldots \ell_k^{n_k}.$$

   The convexity is conserved in this second case.

3. The third case that we can face, is if we add a unit square to ℓ_i of a given path we loose the Lyndon factorization property; which means the monotone order of slopes is no longer correct, and the new polyomino is not digitally convex.

However, we can get a characterization of the first case. We give this characterization in the simplest case, which corresponds to $n_i = 1$. Going back to the factorization $\ell_1^{n_1} \ldots \ell_k^{n_k}$ of the WN path, cases 1 or 2 occur when the following conditions are satisfied.

1. $\ell_{i+1} \leq \ell_i^-$ or $\ell_{i+1} = \ell_i^{-k}\ell$ for some positive integer k and some Christoffel word ℓ;

2. $\ell_{i-1} \geq \ell_i^+$ or $\ell_{i-1} = \ell\ell_i^{+k'}$ for some positive integer k' and some Christoffel word ℓ.

3.4. An inflation method

We consider as before two distinct digitally convex polyominoes $C_1 \subset C_2$. Then there always exists in each of the four paths describing C_1 at least one Christoffel word corresponding to the cases 1 or 2 before. More precisely, we can choose the longest Christoffel word of the path, among those which does not correspond to some side of C_2. It gives an effective method to inflate C_1 to C_2 and conserving the convex property at each step.

Some other methods based on completely different considerations could be investigated. For example we can use a continuous and increasing deformation of the border of the convex hull of C_1 to the border of the convex hull of C_2, such that this deformation contains at any time at most one integer point. An effective construction of such a deformation can be made, using simple arguments but encountering some technical difficulties.

References

[1] Kim, C.E., Rosenfeld, A.: Digital straight lines and convexity of digital regions, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-4 (1982) 149153.

[2] Brlek, S., Lachaud, J.-O., Provençal, X. and Reutenauer, C.: Lyndon + Christoffel=digitally convex. Pattern Recognition, 42-10, 2239-2246 (2009).

[3] Borel, J.P., Laubie, F.: Quelques mots sur la droite projective réelle. J. Théor. Nombres Bordeaux, 5(1), 23-51 (1993).

[4] Dulio, P., Frosini, A., Rinaldi, S., Tarsissi, L., Vuillon, L.: First Steps in the Algorithmic Reconstruction of Digital Convex Sets, Lecture Notes in Computer Science, 10432, 164-176, (2017)

Séries $\mathbb{Q}$-Hadamard et $\mathbb{Q}$-automates circulaires

Louis-Marie Dando *et Sylvain Lombardy *

On étudie ici les séries formelles sur un monoïde libre A^* finiment engendré avec coefficients dans un corps commutatif $\mathbb{K}$, avec $\mathbb{K}$ égal à $\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$. Dans ce cadre, les séries rationnelles et les séries reconnaissables coïncident et le *produit d'Hadamard* (ou produit terme à terme, noté $\odot$) de deux séries rationnelles est rationnel [8].

On considère deux opérations naturelles liées au produit d'Hadamard. Soient s et t deux séries formelles dans $\mathbb{K}\langle\langle A^*\rangle\rangle$.

— si pour tout w dans A^*, $\langle t, w\rangle \neq 0$, alors le quotient d'Hadamard de s et t est défini par :

$$\odot\frac{s}{t} = \sum_{w\in A^*} \frac{\langle s, w\rangle}{\langle t, w\rangle} w;$$

— si pour tout w dans A^*, $\langle s, w\rangle^* = \sum_{k\geqslant 0}\langle s, w\rangle^k$ existe, alors l'itération d'Hadamard de s est défini par :

$$s^{\circledast} = \sum_{w\in A^*} \langle s, w\rangle^* w.$$

Proposition 1 *L'ensemble des séries $\mathbb{K}$-rationnelles n'est clos ni par quotient d'Hadamard (cf. [7]), ni par itération d'Hadamard.*

On définit donc l'ensemble des séries d'Hadamard comme la plus petite famille de séries contenant les séries rationnelles et close pour ces opérations.

Théorème 2 *L'ensemble des séries $\mathbb{K}$-Hadamard sur A^* est définie de manière équivalente comme*
a) l'ensemble des quotients d'Hadamard de séries $\mathbb{K}$-rationnelles ;
b) la clôture des séries $\mathbb{K}$-rationnelles par somme, produit d'Hadamard et itération d'Hadamard.

Remarquons que l'ensemble des séries $\mathbb{K}$-Hadamard n'est pas clos pour les opérations rationnelles (il n'est même pas clos par produit de Cauchy).

On considère par ailleurs des *automates circulaires* à multiplicité dans un corps. Comme un automate pondéré classique, lors de tout calcul, un tel automate lit son entrée de gauche à droite en calculant une valeur comme produit des valeurs associées aux transitions empruntées ; mais en plus, arrivé à la fin de son entrée l'automate peut soit stopper s'il se trouve dans un état final, soit emprunter une transition spéciale qui replace la tête de lecture au début du mot. La valeur associée à un mot est la somme (potentiellement infinie) des valeurs

*LaBRI UMR 5800, Université de Bordeaux, INP Bordeaux, CNRS, Bordeaux, FRANCE

des calculs acceptant ce mot. Si tout mot a une valeur bien définie, l'automate est dit *valide* et son *comportement* est la série dans laquelle le coefficient de chaque mot est la valeur calculée par l'automate.

Ce modèle naturel d'automate – souvent considéré comme une restriction des automates bi-directionnels [3, 5] – permet, dans le cas non pondéré, de reconnaître l'intersection de deux langages avec une machine qui compte un nombre linéaire d'états par rapport aux automates d'entrée et, dans le cas pondéré, de réaliser avec la même complexité le produit d'Hadamard de deux séries (même dans le cas de poids non commutatifs).

Comme on ne peut pas décider si la série réalisée par un $\mathbb{Q}$-automate probabiliste a tous ses coefficients strictement inférieurs à 1/2 [4], si la série réalisée par un $\mathbb{Q}$-automate a tous ses coefficients strictement inférieurs à 1, en faisant boucler un tel automate, on ne peut pas décider si l'automate circulaire obtenu est valide.

Proposition 3 *Il n'est pas décidable si un $\mathbb{Q}$-automate circulaire est valide.*

Ceci est à mettre en lien avec le fait qu'on ne peut pas décider si le support d'une série $\mathbb{Q}$-rationnelle est A^* tout entier ; on ne peut donc pas décider si l'inverse d'Hadamard d'une telle série est défini.

Par contre, si les séries sont bien définies, on a l'équivalence suivante.

Théorème 4 *L'ensemble des séries réalisables par des $\mathbb{K}$-automates circulaires valides est exactement l'ensemble des séries $\mathbb{K}$-Hadamard.*

Corollaire 5 *L'équivalence des $\mathbb{Q}$-automates circulaires valides est décidable.*

De fait, le comportement s de tout $\mathbb{Q}$-automate circulaire peut être représenté par une paire de $\mathbb{Q}$-automates classiques réalisant des séries t_1 et t_2 telles que $s = \odot\frac{t_1}{t_2}$. Savoir si deux comportements $s = \odot\frac{t_1}{t_2}$ et $s' = \odot\frac{t'_1}{t'_2}$ sont égaux revient alors à savoir si $t_1 \odot t'_2$ et $t'_1 \odot t_2$, qui sont des séries $\mathbb{Q}$-rationnelles, sont égales. Or, l'équivalence des séries $\mathbb{Q}$-rationnelles est décidable [1].

Comme nous l'avons dit plus haut, les automates circulaires peuvent être vus comme restriction des automates bi-directionnels (ou boustrophédons) ; ces derniers sont donc au moins aussi puissants. Dans le cas non pondéré, tous ces modèles sont équivalents aux automates classiques [9, 6], alors que dans les cas pondérés les automates bi-directionnels peuvent être strictement plus puissants que les automates circulaires, eux-mêmes plus puissants que les automates classiques (voir par exemple [2] pour le cas des transducteurs, équivalents à des automates sur le semi-anneau $\mathrm{Rat}B^*$).

De manière assez surprenante, dans le cas des corps, les automates bi-directionnels ne sont pas plus puissants que les automates circulaires.

Théorème 6 *Toute série réalisée par un $\mathbb{K}$-automate bi-directionnel peut être réalisée par un $\mathbb{K}$-automate circulaire.*

Références

[1] J. Berstel and C. Reutenauer. *Noncommutative Rational Series with Applications*, volume 137 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2010.

[2] B. Guillon. Sweeping weakens two-way transducers even with a unary output alphabet. In *NCMA'15*, volume 318 of *books@ocg.at*, pages 91–108. Austrian Comp. Soc., 2015.

[3] C. Kapoutsis, R. Královič, and T. Mömke. Size complexity of rotating and sweeping automata. *Journal of Computer and System Sciences*, 78(2) :537–558, 2012.

[4] A. Paz. *Introduction to Probabilistic Automata (Computer Science and Applied Mathematics)*. Academic Press, Inc., Orlando, FL, USA, 1971.

[5] G. Pighizzini. Two-way finite automata : Old and recent results. *Fundam. Inform.*, 126(2-3) :225–246, 2013.

[6] M. O. Rabin and D. Scott. Finite automata and their decision problems. *IBM J. Res. Dev.*, 3(2) :114–125, 1959.

[7] C. Reutenauer. Sur les éléments inversibles de l'algèbre de Hadamard des séries rationnelles. *Bull. Soc. Math. France*, 110 :225–232, 1982.

[8] M.-P. Schützenberger. On a theorem of R. Jungen. *Proc. Amer. Math. Soc.*, 13(6) :885–890, 1962.

[9] J. C. Shepherdson. The reduction of two-way automata to one-way automata. *IBM J. Res. Dev.*, 3(2) :198–200, 1959.

A definition and counting of biperiodic recurrent configurations in the sandpile model on $\mathbf{Z}^2$

Henri Derycke Yvan Le Borgne — Univ. Bordeaux, Bordeaux INP, CNRS, LaBRI, UMR5800, F-33400 Talence, France

Abstract

For the sandpile model on the usual two dimensional grid, we propose a weaker version of Dhar criterion to define recurrent configurations among stable biperiodic configurations. We check this new criterion via an algorithm which auto-stabilises to a canonical ultimately periodic behaviour independent of details in its not fully specified initialisation. This leads to ultimately periodic edge/vertex traversals similar to those of Cori-Le Borgne [2] in the case of finite graphs and then to a bijection with *some* cycle-rooted forests on the torus describing the period. A determinantal formula [5] counts all those forests and the refinement with some monodromy parameters allows to identify in some coefficients the number of recurrent configurations.

The Abelian sandpile model was introduced by physicists Bak, Tang and Wiesenfeld in [1] as a model of self-organized criticality. Given a simple, undirected graph $(V \cup \{s\}, E)$ where we distinguish s as the sink of the graph, we consider *configurations* in this model which are an assignments $\eta : V \mapsto \mathbf{Z}$ of some grains of sand on each vertex. We say that η is stable at $x \in V$ if $\eta(x) < \deg(x)$, and η is stable if it is stable at all $x \in V$. If η is unstable at x, then x is allowed to topple which means that the vertex x sends one grain along each incident edge. This toppling is said *legal*. A toppling is *forced* when it is not necessarily legal. Grains arriving at the sink are lost. Given a configuration η, we define a stabilization as a sequence of allowed topplings until a stable configuration is reached. The result of all stabilizations is unique due to commutations of topplings of unstable vertices and is noted $stab(\eta)$.

Let $P(\eta)$ be the result of a stabilization of $\eta + \mathbf{1}_{s\sim}$, that is η with an extra grain on each neighbour of s, which may be interpreted as a forced toppling of the sink. $P(\eta)$ is also called the Dhar criterion since the set of recurrent configurations is a subset of the stable configurations characterized by Dhar [4] as the fixed points of P. For such a fixed point, each vertex topples exactly once in this process. The notion of recurrence is related to a natural Markov chain in this model not discussed here [3], and it is well studied for its connection with spanning trees [4], uniform spanning tree, the Tutte polynomial on the underlying graph [2]. Also on finite graphs, the set of recurrent configurations equipped with the operation $(\eta, \mu) \mapsto stab(\eta + \mu)$ is an abelian group [3]. When the graph is the grid $\mathbf{Z}^2$, the existence of such a group is open.

One of our motivations is the search of finite groups on a subset of recurrent configurations on $\mathbf{Z}^2$, which may be subgroups of the hypothetical (infinite) group. We focus on the subset of biperiodic configurations on the grid defined as follows. Let $\vec{P}_1, \vec{P}_2 \in \mathbf{Z}^2$ two non collinear vectors. A configuration η of $\mathbf{Z}^2$ is biperiodic of period $(\vec{P}_1, \vec{P}_2)$ if for all $x \in \mathbf{Z}^2$, $\eta(x+\vec{P}_1) = \eta(x+\vec{P}_2) = \eta(x)$.

Several approaches are suggested by litterature to define the notion of recurrence: for example, adding one edge per period to an extra vertex called the sink (dissipative sandpiles [6]), another example merges in one sink vertex all vertices outside a finite polygon and then scale this polygon [7]. Our approach relies on a weaker form of Dhar criterion and leads to a degenerate polygon which is an half-plane. Indeed we place the sink at infinity in a direction by analogy to the projective plane. The sink s is a point at infinity and will be describe by an euclidean vector $\vec{s} = (s_x, s_y) \in \mathbf{Z}^2$ where $gcd(s_x, s_y) = 1$. Note that $\vec{s}$ and $-\vec{s}$ refer to different sinks.

The sink being sent to infinity, the difficulty of toppling it appears. We replace this by a forced toppling of an half-plane of line boundary orthogonal to the sink $\vec{s}$.

Definition 1 (Weak Dhar criterion in a rational direction)**.** *A configuration η is said* recurrent in the direction of the sink $\vec{s}$ *if and only if for any $k \in \mathbf{Z}$ the forced toppling of the vertices of the half-plane $\{(x,y) \in \mathbf{Z}^2 \mid s_x x + s_y y \geq k\}$ leads to the legal toppling of all other vertices.*

Proposition 1. *There exists execution of the weak Dhar criterion on biperiodic configurations that is auto-stabilizing to an ultimately periodical behaviour which does not depend on the position of the half-plane defined by a line colinear to $\vec{s}^{\perp}$ and can be simulated in finite time.*

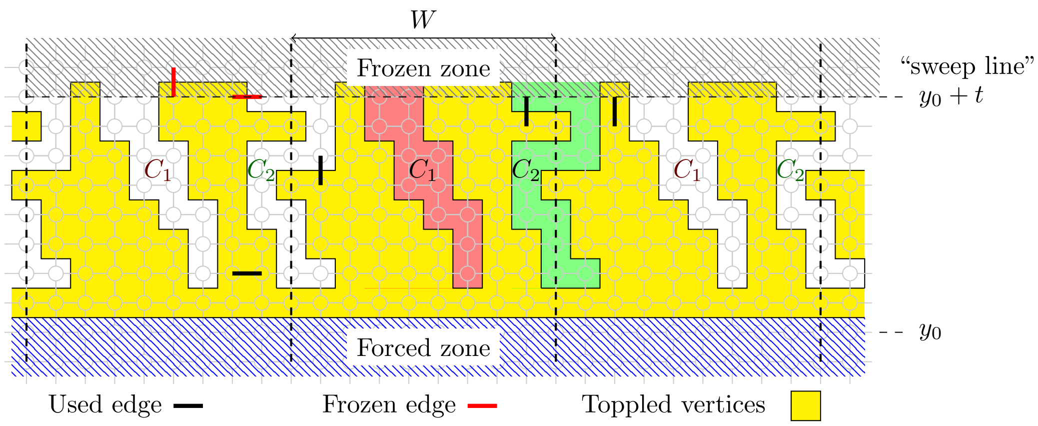


Figure 1: Weak Dhar criterion after step t, C_1 and C_2 are next connected components

Sketch of the proof when $\vec{s} = (0,-1)$ (that can be generalized for all $\vec{s}$). Let η be a recurrent configuration in direction $(0,-1)$ of period $(\vec{P_1}, \vec{P_2})$. Without losing generality, we can assume that $\vec{P_1} = (W, 0)$ and $\vec{P_2} = (0, H)$ where $W, H > 0$. We equip the set of edges with the following order $\prec_{\vec{s}}$. Let e_1 (resp. e_2) be an edge of middle m_1 (resp. m_2) in the usual embedded of $\mathbf{Z}^2$, $e_1 \prec_{\vec{s}} e_2$ if and only if $\vec{s} \cdot \vec{m_1} < \vec{s} \cdot \vec{m_2}$ or ($\vec{s} \cdot \vec{m_1} = \vec{s} \cdot \vec{m_2}$ and $\vec{s}^{\perp} \cdot \vec{m_1} < \vec{s}^{\perp} \cdot \vec{m_2}$) where $\vec{s} \cdot \vec{m_1}$ is the usual scalar product between $\vec{s}$ and $\vec{m_1}$ and $\vec{s}^{\perp}$ is the vector $(-s_y, s_x)$. For $\prec_{(0,-1)}$, edges are ordered increasingly from top to bottom in priority, ties broken from left to right. When a vertex become unstable, it topples and its grains are sent along incident edges which become pending. Such a crossing grain is received at opposite endpoint when this pending edge is activated. We control the process of stabilisation by activating the maximal pending edge according to $\prec_{\vec{s}}$ (see [2] for details).

We start by a forced toppling of the half-plane $\overrightarrow{(x,y)} \cdot (-\vec{s}) = y \leq y_0 \in \mathbf{Z}$, denoted $H_{\leq y_0}$. The remaining legal topplings, more precisely edges allowed to be activated, are enclosed between this half-plane and a "sweep line" $y = y_0 + t$, where $t \geq 0$ is called a step of execution for Dhar criterion. More precisely at each step $t > 0$, we only consider edges which middle has ordinate $y \in [y_0, y_0 + t]$. Thus there is a forced toppled zone $y \leq y_0$, a frozen zone $y > y_0 + t$ and a working zone $y_0 < y \leq y_0 + t$, see Figure 1. Since η is recurrent, at step t there is at least one vertex v_t that topples on line $y = y_0 + t$. The order $\prec_{\vec{s}}$ guarantees that the set of connected components of untoppled vertices in the working zone is periodic of period $(W, 0)$ (C_1 and C_2 on Figure 1) and that these components are finite as long as η is recurrent since enclosed between the sweep line

2

and the sequences of topplings leading to toppled vertices $(v_t + k\vec{P_1})_{k\in\mathbf{Z}}$. Thus the toppling of a vertex v is independent of the toppling of each $v + k(W, 0)$ with $k \in \mathbf{Z}^*$. This observation allows to simulate the criterion on a cylinder $[1, W] \times \mathbf{Z}$ with a periodic configuration of period $(0, H)$ for some H.

The more we force topplings, the more we topple vertices. This and translation symmetry $\vec{P_2}$ implies that if a vertex v topples at step $t + H$, then $v - (0, H) = v - \vec{P_2}$ topples no later than step t. As a corollary, at most one of the vertices $(v + k\vec{P_2})_{k\in\mathbf{Z}}$ can topple legally in any H consecutive steps. Thus for any H consecutive steps, there is at most WH vertices that topple.

We can show by contradiction that each vertex v topples at some step. For any vertex v, either all $(v + k\vec{P_2})_{k\in\mathbf{Z}}$ topples or there exists k_v such that exactly $(v + k\vec{P_2})_{k\geq k_v}$ do not topple. If not all vertices topples at some step, we can define $k_{WH} = \max_v k_v$ where v runs over the subset S of the vertices of $[1, W] \times [1, H]$ for which k_v is defined and $\overline{S} := [1, W] \times [1, H] \setminus S$ the complement subset. By definition, in the half-plane $H_{>y_0+k_{W,H}H}$, the subsets $(S + k\vec{P_2})_{k\geq k_{WH}}$ never topples and all other vertices in $(\overline{S} + k\vec{P_2})_{k\geq k_{W,H}}$ topples. From this, we also deduce the existence of a step $t_{W,H}$ such that no vertex in the half plane $H_{\leq y_0+k_{W,H}H}$ can legally topple at steps $t \geq t_{W,H}$. For steps $t \geq t_{W,H}$, only vertices in $(\overline{S} + k\vec{P_2})_{k\geq k_{W,H}}$ periodically topples. This describes an (infinite) sequence of legal topplings toward a stable configuration where $(S + k\vec{P_2})_{k\geq k_{WH}}$ did not topple, which is a contradiction with the recurrence of η (so $S = \emptyset$).

Hence, let T be the first step when all vertices of $[1, W] \times [1, H]$ has been toppled. Since there is at most WH vertices that topple at step T, the sequence of toppling that destabilizes the last untoppled vertex of $[1, W] \times [1, H]$ starts on line $y = y_0 + T$, is at most of length WH and ends to a line below $y = y_0 + H$ so $T \leq H + WH$. So the criterion is finite and effective.

Moreover, from steps $T + 1$ to $T + H$, we observe the ultimately periodic behaviour of the Dhar criterion: WH vertices topple, all having distinct copies in the fundamental domain $[1, W] \times [1, H]$. In addition, for $k \leq H$, the half-plane $H_{\leq y_0+k}$ has toppled at step T so forcing toppling of this half-plane, instead of $H_{\leq} y_0$, results after $T - k$ steps in the same set of toppled vertices. So the ultimately periodic behaviour of Dhar criterion starting either by $H_{\leq y_0}$ or $H_{\leq y_0+k}$ are the same. $\square$

The proof induces a bijection with a subset of cycle rooted spanning forests [5] on the toroidal grid: from step $T + 1$ to step $T + H$, we attach to each vertex and its repetitions the edge that destabilizes it. In order to respect the order $\prec_{\vec{s}}$ on the cylinder, it is enough to process the toppling in only one repetition of each connected component at a given step. The result is a cycle rooted spanning forest of the toroidal grid with non contractible cycles. These cycles correspond to infinite periodic branches in the plane whose slopes are not orthogonal to $\vec{s}$.

Theorem 1. *Let $\vec{s}$ be a sink. The set of recurrent biperiodic configuration of pattern size $W \times H$ on $\mathbf{Z}^2$ is in bijection with the set of cycle rooted spanning forests of the toroidal grid $W \times H$ whose slope (a, b) is such that $a \cdot s_x + b \cdot s_y \neq 0$.*

We assume in the next part that $W, H \geq 2$. We call *NCRSF* a non-contractible cycle rooted spanning forests. We want to count the recurrent configurations in a direction $\vec{s}$. By definition of homology, the image of one copy of an (oriented) cycle of homology class $(i, j) \in \mathbf{Z}^2$ starting at (x, y) into the plane ends at $(x + jW, y + iH)$. Kenyon [5] gives the following determinantal formula $\sum_{\text{NCRSFs }\gamma} \left(2 - z^i w^j - z^{-i} w^{-j}\right)^k = \det \Delta$ where (i, j) is the homology class of the cycles of γ and k their number, z is the monodromy of cycles with homology class $(1, 0)$, w is the monodromy of cycles with homology class $(0, 1)$, Δ is the Laplacian on the line bundle with connection 1 over all

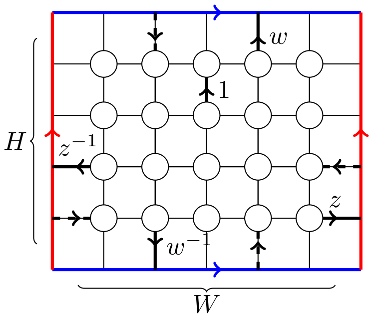


Figure 2: Connection Φ on the torus $W \times H$

kj \ ki	0	1	2	3	4
0		31300528	541732	1528	1
1	31300528	5427200	31232	4	
2	541732	31232	6		
3	1528	4			
4	1				

Figure 3: The number of NCRSF on a toroidal grid 4×4 in each direction where $k = gcd(ki, kj)$

oriented edges except those crossing the blue (connection w or w^{-1}) and red sides (connection z or z^{-1}) of a fix fundamental rectangle as in figure 2. Then $\Delta(v) = \sum_{u \to v} v - \Phi_{u \to v} u$ where $\Phi_{u \to v}$ is the connection value on edge $u \to v$.

Due to planarity of the grid, a cycle cannot cross itself on the toroidal grid. Thus the homology class (i, j) of a cycle in this graph has $gcd(i, j) = 1$. Moreover the length of such a cycle is at least $|Wj| + |Hi|$. That gives the first part of the following proposition.

Proposition 2. *Given γ a NCRSF on a toroidal grid $W \times H$, if γ has k cycles with homology class (i, j) then i and j are co-prime and $|kjW| + |kiH| \leq WH$. Reciprocally for any $(i, j) \in \mathbf{Z}^2$ co-prime and any $k > 0$ such that $|kjW| + |kiH| \leq WH$, there exists a NCRSF of parameters (i, j, k).*

The second part of this proposition is achieved by k repetitions of a digital line from Bresenham's line algorithm (with corners).

We denote $Q_{i,j,k} = \left(2 - z^i w^j - z^{-i} w^{-j}\right)^k$ for any $k > 0$ and any $(i, j) \in S = \{(0, 1)\} \cup \{(a, b) \mid a > 0 \text{ and } gcd(a, b) = 1\}$, the $(Q_{i,j,k})_{i,j,k}$ are linearly independent. So the number of NCRSFs is the sum of the coefficients $(\alpha_{i,j,k})_{i,j,k}$ of $\det \Delta$ in the decomposition in $(Q_{i,j,k})_{i,j,k}$: $\det \Delta = \sum_{i,j,k} \alpha_{i,j,k} Q_{i,j,k}$. One can show that $\alpha_{i,j,k} = \alpha_{i,-j,k}$ for $i > 0$.

Proposition 3. *The number of biperiodic recurrent configurations in direction $\vec{s}$ of size $W \times H$ is $\sum_{(i,j,k) \in S_{\vec{s}}} \alpha_{i,j,k}$ where $S_{\vec{s}} = \{(i, j, k) \mid (i, j) \in S, k > 0, |kjW| + |kiH| \leq WH, iWs_x + jHs_y \neq 0\}$.*

This formula enhances the counting the recurrent configurations in a direction that was limited to enumeration. Some explicit results are given on Figure 3. Some extra results up to $m \times n$ with $m, n \leq 9$ can be found at `https://www.labri.fr/perso/hderycke/biperiodic_recurrent/`.

References

[1] P. Bak, C. Tang, K. Wiesenfeld. Self-organized criticality. Phys. Rev. A 38 (1988) 364–374.

[2] R. Cori, Y. Le Borgne. The sand-pile model and Tutte polynomials. Advances in Applied Mathematics, 30:44-52 (2003).

[3] D. Dhar. Self-organized critical state of sandpile automaton models. Phys. Rev. Lett. 64 (1990), no. 14.

[4] D. Dhar, S.N. Majumdar. Equivalence of the Abelian sandpile model and the $q \to 0$ limits of the Potts model. Phys. A 185 (1992) 129–135.

[5] R. Kenyon. Determinantal spanning forests on planar graphs. arXiv:1702.03802 (2017)

[6] C. Maes,, F. Redig, E. Saada. The infinite volume limit of dissipative abelian sandpiles. Comm. Math. Phys. 244 (2004), no. 2, 395–417.

[7] G. Paoletti. Deterministic Abelian Sandpile Models and Patterns. 2012.

RETURN WORDS AND DERIVATED SEQUENCES TO ROTE SEQUENCES

KATEŘINA MEDKOVÁ

1. Introduction

We study complementary symmetric Rote sequences, which are sequences over the binary alphabet $\{0,1\}$ with factor complexity $C(n) = 2n$ and with language closed on exchange of letters $0 \leftrightarrow 1$. We refer about the work in progress [7].

Rote in [8] proved that a sequence $\mathbf{v} = v_0v_1\cdots$ is complementary symmetric Rote sequence if and only if its first difference sequence $\mathbf{u} = u_0u_1\cdots$, which is defined by $u_i = v_i - v_{i+1} \mod 2$, is Sturmian sequence.

Our aim is to describe return words and derivated sequences of complementary symmetric Rote sequences. In the sequel, we will call them simply Rote sequences. The study is based on the link between Rote and Sturmian sequences and recent work about derivated sequences of Sturmian sequences [1, 6].

Let $\mathbf{u} = u_0u_1u_2\cdots$ be an infinite sequence and let $w = u_iu_{i+1}\cdots u_{i+n-1}$ be its factor. The integer i is called an occurrence of the factor w. A return word to a factor w is a word $u_iu_{i+1}\cdots u_{j-1}$ with i and j being two consecutive occurrences of w such that $i < j$.

Let w be a prefix of $\mathbf{u}$ which has k return words $r_0, r_1, \ldots, r_{k-1}$. Then the sequence $\mathbf{u}$ can be written as a concatenation of these return words: $\mathbf{u} = r_{d_0}r_{d_1}r_{d_2}\cdots$ and the derivated sequence of $\mathbf{u}$ to prefix w is the sequence $\mathbf{d_u}(w) = d_0d_1d_2\cdots$. For simplicity, we consider two derivated sequences to be the same if they differ only by a permutation of letters. We work only with sequences which are uniformly recurrent, i.e. each prefix w of $\mathbf{u}$ occurs in $\mathbf{u}$ infinitely many times and the set of all return words to w is finite.

Recall that the factor w of $\mathbf{u}$ is right special, if both words $w0$ and $w1$ are factors of $\mathbf{u}$. Analogously the left special factor is defined. The factor is bispecial, if it is both left and right special. To find all derivated sequence it suffices to study only right special prefixes. Indeed, if the prefix w is not right special, then there is a unique letter a such that wa is a factor of $\mathbf{u}$. Thus the occurrences of w and wa in $\mathbf{u}$ coincides and they have the same return words and derivated sequences. If $\mathbf{u}$ is aperiodic, then w is always a prefix of some right special prefix of $\mathbf{u}$.

We focus on standard Sturmian sequences, i.e. the sequences whose each prefix is left special. In that case, we can take into consideration only bispecial prefixes to find all derivated sequences.

2. Rote sequences and associated Sturmian sequences

We define the mapping $\mathcal{S}$ which maps factors of Rote sequence to factors of associated Sturmian sequence.

Definition 1. *The mapping $\mathcal{S} : \{0,1\}^+ \to \{0,1\}^*$ is for every $v = v_0v_1\cdots v_n \in \{0,1\}^+$ of length at least 2 defined by $\mathcal{S}(v_0v_1\cdots v_n) = u_0u_1\cdots u_{n-1}$, where $u_i = v_i + v_{i+1} \mod 2$ for all $i \in \{0,1,\ldots,n-1\}$, $\mathcal{S}(v_0) = \varepsilon$.*

We can naturally extend the domain of $\mathcal{S}$ to $\{0,1\}^{\mathbb{N}}$ and write the associated Sturmian sequence $\mathbf{u}$ to the Rote sequence $\mathbf{v}$ as $\mathbf{u} = \mathcal{S}(\mathbf{v})$. To each Sturmian sequence, there are two associated Rote sequences $\mathbf{v}$ and $\mathbf{v}'$. Since we have $\mathbf{v}' = E(\mathbf{v})$, we work only with Rote sequences starting with the letter 0 without lose of generality. The factors of Rote sequence $\mathbf{v}$ and associated Sturmian sequence $\mathbf{u}$ are closely related.

Proposition 2. *Let $\mathbf{u}$ be a Sturmian sequence and $\mathbf{v}$ be a Rote sequence such that $\mathbf{u} = \mathcal{S}(\mathbf{v})$. The word u is a factor of $\mathbf{u}$ if and only if both words v, v' such that $u = \mathcal{S}(v) = \mathcal{S}(v')$ are the factors of $\mathbf{v}$. Moreover, for every $i \in \mathbb{N}$, i is an occurrence of u in $\mathbf{u}$ if and only if i is an occurrence of v in $\mathbf{v}$ or an occurrence of v' in $\mathbf{v}$.*

To study return words and derivated sequences of a given Rote sequence, we have to examine these objects in the case of associated Sturmian sequence. In [6] the authors describe derivated sequences of fixed points of Sturmian morphisms. Their basic idea is to suitably decompose a given Sturmian morphism onto some elementary morphisms.

A morphism is a mapping $\psi : \mathcal{A}^* \to \mathcal{B}^*$ such that $\psi(uv) = \psi(u)\psi(v)$ for all $u, v \in \mathcal{A}^*$. If $\mathcal{A} = \mathcal{B}$, ψ is a morphism over $\mathcal{A}^*$. The domain of the morphism ψ can be naturally extended to $\mathcal{A}^{\mathbb{N}}$. The matrix of a morphism ψ over $\mathcal{A}^*$ is a matrix M_ψ defined by $(M_\psi)_{ab} = |\psi(a)|_b$ for all $a, b \in \mathcal{A}$. The morphism is primitive, if there is a positive integer k such that all elements of $(M_\psi)^k$ are positive.

A fixed point of a morphism ψ is a sequence $\mathbf{u}$ such that $\psi(\mathbf{u}) = \mathbf{u}$. The sequence $\mathbf{u}$ is substitutive if $\mathbf{u} = \sigma(\mathbf{v})$ for a morphism σ and a sequence $\mathbf{v}$ which is a fixed point of morphism θ. Moreover, $\mathbf{u}$ is substitutive primitive if θ is primitive. Durand in [3] proved that sequence $\mathbf{u}$ is substitutive primitive if and only if $\mathbf{u}$ has finite number of distinct derivated sequences.

A morphism ψ is Sturmian if $\psi(\mathbf{u})$ is Sturmian sequence for any Sturmian sequence $\mathbf{u}$. Consider the following elementary Sturmian morphisms φ_b and φ_β defined by

$$\varphi_b : \begin{array}{l} 0 \to 0 \\ 1 \to 01 \end{array} \quad \text{with } M_b = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \varphi_\beta : \begin{array}{l} 0 \to 10 \\ 1 \to 1 \end{array} \quad \text{with } M_\beta = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

By [4] to a given standard Sturmian sequence $\mathbf{u}$ we can uniquely assign the pair: directive sequence $\mathbf{z} \in \{b, \beta\}^{\mathbb{N}}$ and the sequence $(\mathbf{u}^{(n)})_{n \geq 0}$, where $\mathbf{u}^{(n)} \in \{0,1\}^{\mathbb{N}}$ is a standard Sturmian sequence, such that for every $n \in \mathbb{N}$ we have

$$\mathbf{u} = \varphi_{z_0 z_1 \dots z_{n-1}}(\mathbf{u}^{(n)}).$$

The directive sequence $\mathbf{z}$ contains infinitely many letters b and infinitely many letters β. In addition, $\mathbf{z}$ is purely periodic, i.e. $\mathbf{z} = z^\infty$, if and only if $\mathbf{u}$ is the fixed point of the morphism φ_z. This fixing morphism φ_z is always primitive.

3. Return words to prefixes of Rote sequences

Vuillon in [13] showed that an infinite sequence is Sturmian if and only if each non-empty factor has exactly two distinct return words. Using some results from [2] we show that all derivated sequences of Rote sequences to non-empty prefixes are over a ternary alphabet.

Theorem 3. *Let $\mathbf{v}$ be a Rote sequence. Then every non-empty prefix x of $\mathbf{v}$ has exactly three distinct return words.*

In addition, we can construct the return words in Rote sequence using the relevant return words in associated Sturmian sequence. We need an auxiliary definition of stability.

Definition 4. *The word $u = u_0 u_1 \cdots u_{n-1} \in \{0,1\}^*$ is called stable (S) if $\sum_{i=0}^{n-1} u_i = 0 \mod 2$. Otherwise, u is unstable (U).*

Definition 5. *Let w be a prefix of a Sturmian sequence $\mathbf{u}$ with return words r, s. Due to Vuillon's result [9] $\mathbf{u}$ is a concatenation of blocks $r^k s$ and $r^{k+1}s$ or blocks sr^k and sr^{k+1} for some positive integer k. With respect to the return words of w we distinguish three cases:*

- i) *w is of type $SU(k)$, if r is stable and s is unstable;*
- ii) *w is of type $US(k)$, if r is unstable and s is stable;*
- iii) *w is of type $UU(k)$, if both r and s are unstable.*

The type of prefix w is denoted $\mathcal{T}_w$.

We do not define the type SS since it cannot appear in Strumian sequences. We use these prefix types to describe the return words to corresponding prefixes in Rote sequences.

Theorem 6. *Let x be a prefix of a Rote sequence $\mathbf{v}$. Denote by $\mathbf{u}$ the Sturmian sequence $S(\mathbf{v})$ and by w its prefix $S(x)$ with return words r and s. Then the prefix x of $\mathbf{v}$ has three return words $A, B, C \in \{0,1\}^+$ such that:*

- i) *$r = \mathcal{S}(A0)$, $sr^{k+1}s = \mathcal{S}(B0)$ and $sr^k s = \mathcal{S}(C0)$ if w is of type $SU(k)$;*

ii) $rr = \mathcal{S}(A0)$, $rsr = \mathcal{S}(B0)$ *and* $s = \mathcal{S}(C0)$ *if* w *is of type* $US(k)$*;*
iii) $rr = \mathcal{S}(A0)$, $rs = \mathcal{S}(B0)$ *and* $sr = \mathcal{S}(C0)$ *if* w *is of type* $UU(k)$*.*

Remark 7. We can also determine the type of w using the matrix P_w composed of the Parikh vectors of return words to w. Let w be a prefix of a Sturmian sequence $\mathbf{u}$ with return words r, s, where r is the more frequent return word. Then the matrix P_w is defined by:

$$P_w = \begin{pmatrix} |r|_0 & |s|_0 \\ |r|_1 & |s|_1 \end{pmatrix} \mod 2\,,$$

where $|u|_a$ denotes the number of letters a in the word u. Then the type $\mathcal{T}_w$ of the prefix w is

i) SU if $P_w = \begin{pmatrix} p & q \\ 0 & 1 \end{pmatrix}$ ii) US if $P_w = \begin{pmatrix} p & q \\ 1 & 0 \end{pmatrix}$ iii) UU if $P_w = \begin{pmatrix} p & q \\ 1 & 1 \end{pmatrix}$

for some numbers $p, q \in \{0, 1\}$.

4. Derivated sequences of Rote sequences

Theorem 8. *Let* $\mathbf{v}$ *be a Rote sequence with non-empty bispecial prefix* x*. Then the derivated sequence* $\mathbf{d_v}(x)$ *is uniquely determined by derivated sequence* $\mathbf{d_u}(w)$ *of* $\mathbf{u} = \mathcal{S}(\mathbf{v})$ *to the prefix* $w = \mathcal{S}(x)$ *and by type* $\mathcal{T}_w$ *of prefix* w*.*

Moreover, we are able to construct derivated sequence $\mathbf{d_v}(x)$ of a given Rote sequence $\mathbf{v}$ to a prefix x if we know the type of $S(x)$ and the derivated sequences of $S(\mathbf{v})$ to $S(x)$.

First, we explain how to determine the type of a given prefix w of a standard Sturmian sequence $\mathbf{u}$. Clearly it suffices to study only bispecial prefixes. Let us order the bispecial prefixes of $\mathbf{u}$ by their length starting from the shortest one. Their types can be determined using results from Section 3 of [6], where the authors explain how prefixes and their return words change under application of morphisms φ_b and φ_β.

Proposition 9. *Let* $\mathbf{u}$ *be a standard Sturmian sequence with directive sequence* $\mathbf{z} \in \{b, \beta\}^{\mathbb{N}}$*. Then the type of its* n*-th bispecial prefix* w *is given by the natural number* k *and the matrix* P_w*, where*

i) $P_w = M_{z_0}M_{z_1}\cdots M_{z_{n-1}}\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \mod 2$ *if the sequence* $z_nz_{n+1}z_{n+2}\cdots$ *has a prefix* $b^k\beta$*,*
ii) $P_w = M_{z_0}M_{z_1}\cdots M_{z_{n-1}}\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) \mod 2$ *if the sequence* $z_nz_{n+1}z_{n+2}\cdots$ *has a prefix* $\beta^k b$*.*

The derivated sequences of a given Sturmian sequence $\mathbf{u}$ are described in detail in [6]. In particular, if $\mathbf{u}$ is a standard Sturmian sequence with directive sequence $\mathbf{z} = z_0z_1z_2\cdots$, then the derivated sequence $\mathbf{d_u}(w)$ to the n-th bispecial prefix w is standard Sturmian sequence with directive sequence $z_nz_{n+1}z_{n+2}\cdots$.

It is well known that all Sturmian sequences are 2iet sequences. We show that derivated sequences of Rote sequences are 3iet sequences. A three interval exchange transformation $T : [0,1) \to [0,1)$ is given by partition of interval $[0,1)$ into three subintervals $I_A = [0,\alpha)$, $I_B = [\alpha, \alpha+\beta)$ and $I_C = [\alpha+\beta, 1)$ and by permutation π on the set $\{1,2,3\}$ which expresses how these subintervals are rearranged, see [5] for more details. The 3iet sequence $\mathbf{u} = u_0u_1u_2\cdots \in \{A,B,C\}^{\mathbb{N}}$ of transformation T with initial point $\rho \in [0,1)$ is defined by $u_n = L$ if $T^n(\rho) \in I_L$ for all $n \in \mathbb{N}$.

Proposition 10. *Let* $\mathbf{v}$ *be a Rote sequence with non-empty prefix* x*, let* $\mathbf{u} = \mathcal{S}(\mathbf{v})$ *be a standard Sturmian sequence with prefix* $w = \mathcal{S}(x)$*. If* $\mathbf{d_u}(w)$ *is 2iet sequence of transformation with intervals* $[0,\alpha)$ *and* $[\alpha, 1)$ *and initial point* ρ*, then the derivated sequence* $\mathbf{d_v}(x)$ *is 3iet sequence of transformation* T *with initial point* ρ*, where* T *is given by:*

i) *intervals* $[0,\alpha)$, $[\alpha, 2\alpha - k(1-\alpha))$, $[2\alpha - k(1-\alpha), 1)$ *and permutation* $(3,2,1)$ *if* w *is of type* $SU(k)$*;*
ii) *intervals* $[0, 2\alpha-1)$, $[2\alpha-1, \alpha)$, $[\alpha, 1)$ *and permutation* $(3,2,1)$ *if* w *is of type* $US(k)$*;*
iii) *intervals* $[0, 2\alpha-1)$, $[2\alpha-1, \alpha)$, $[\alpha, 1)$ *and permutation* $(2,3,1)$ *if* w *is of type* $UU(k)$*.*

Now we suppose that directive sequence $\mathbf{z}$ is purely periodic, i.e. $\mathbf{u}$ is fixed point of morphism φ_z.

By [6], sequence $\mathbf{u}$ has at most $|z|$ distinct derivated sequences. We prove that the associated Rote sequence $\mathbf{v}$ has a finite number of distinct derivated sequences too and so by Durand's result from [3] it is substitutive primitive.

Theorem 11. *Let* $\mathbf{v}$ *be a Rote sequence and let* $\mathcal{S}(\mathbf{v}) = \mathbf{u}$ *be a standard Sturmian sequence which is a fixed point of a morphism* φ_z*. Then*

i) $\mathbf{v}$ *has at most* $3|z|$ *distinct derivated sequences, each of them is fixed by a morphism;*
ii) $\mathbf{v}$ *is substitutive primitive.*

Moreover, using our results and Durand's construction from [3] we can construct the fixing morphisms of these derivated sequences algorithmically.

Acknowledgements

Kateřina Medková acknowledges financial support by the Czech Technical University in Prague grant SGS17/193/OHK4/3T/14.

References

[1] I. M. Araújo, V. Bruyère, *Words derivated from Sturmian words*, Theoret. Comput. Sci., 340 (2005), pp. 204–219.
[2] L. Balková, E. Pelantová, and W. Steiner, *Sequences with constant number of return words*, Monatsh. Math., 155 (2008), pp. 251–263.
[3] F. Durand, *A characterization of substitutive sequences using return words*, Discrete Math., 179 (1998), pp. 89–101.
[4] J.Justin, G. Pirillo, *Episturmian words and episturmian morphisms*,Theoret. Comput. Sci., 276 (2002), pp. 281–313.
[5] M. Keane, *Interval exchange transformations*, Mathematische Zeitschrift, 141 (1975), pp. 25–31.
[6] K. Klouda, K. Medková, E. Pelantová and Š. Starosta, *Fixed points of Sturmian morphisms and their derivated words*, preprint: arXiv:1801.09203.
[7] K. Medková, E. Pelantová, L. Vuillon, *Return words and derivated sequences to Rote sequences*, in preparation.
[8] G. Rote, *Sequences with subword complexity* $2n$, J. of Number Theory, 46 (1993), pp. 196–213.
[9] L. Vuillon, *A characterization of Sturmian words by return words*, Eur. J. Combin., 22 (2001), pp. 263–275.

Formal Intercepts of Sturmian words

Caïus Wojcik

Abstract :

We study Sturmian words, and particularly the second parameter describing this class of words.

Sturmian words are infinite words over a 2-letter alphabet. They are defined as the infinite words having lowest unbounded complexity. Namely, a theorem of Morse and Hedlund states that an infinite word is ultimately periodic if and only if it has bounded complexity. Sturmian words are characterised by the properties of not being ultimately periodic, and being balanced, that is, the number of 1 appearing in factors of a given length only takes two values.

The first parameter decribing a Sturmian word is its slope, defined as the asymptotic proportion of 1, and characterises the set of factors of the sturmian word. Through the continued fraction expansion of this irrational number one can construct a distinguished Sturmian word of the corresponding slope, called the characteristic word, which is the only sturmian word of this slope admitting two sturmian extensions on the left.

Sturmian words are obtained geometrically by drawing a line on the plane and coding the crossing through a vertical line by a 1 and the horizontal line by a 0. The slope of the sturmian word being obtained through the slope of the considered line, and the second parameter, the intercept, is usually presented as the real number on the y-axis intersecting the line.

The characteristic word may be described using so-called standard and central words. The first ones are obtained by a concatenation process similar to Euclide's algorithm on integers. The second ones are obtained by removing the last two letters of a standard word, and are palindromic words with

two relatively prime periods, realizing the sharp bound of Fine and Wilf's theorem.

On the other hand, we study the repetition function of sturmian words, defined as the number of distinct factors of a given length appearing at the beginning of an infinite word. This function is related to the diphantine exponent of infinite words, and in the case of sturmian words, it has been shown that the diophantine exponent is bounded if and only if the partial quotients of the continued fraction expansion of the slope are bounded.

We study the repetition function using so-called Rauzy graphs of infinite words, also called factor graphs. Rauzy graphs are defined as a sequence of directed graphs, with vertexes the factors of a fixed infinite word and arrows linking two factors if one is obtained by the other by a one-letter shift. The Rauzy graphs of arbitrary words are in general difficult to compute, but in the case of sturmian words they are particularly simple.

Indeed, the rauzy graphs of sturmian words are constituted of two cycles, patched together by a common part. The length of these two cycles are linked to the denominator of the sequence of convergents in the continued fraction expansion of the slope. Also, the evolution of the Rauzy graphs in the case of sturmian word may be fully described.

Any infinite word defines an infinite path on its Rauzy graph. Since these paths must be coherent with each others, there are a lot of restrictions on the possible paths taken by a given word. The consideration of these paths for different element of the shift orbit of the base word is a natural point of view of combinatorics on words. The repetition function can be read on those path, since it is the length of the longest hamiltonian path at the beginning of the base word's path.

Among the two cycles of the Rauzy graph of a sturmian word there a distinguished one, that we call the referent cycle, defined as the one through whom the characteristic word first passes. Indeed, we can compute the number of times the characteristic word turns around a cycle, and take it as a reference for other sturmian words. We use this idea to compute the length of cycles in Rauzy graphs, using the fact that the first repeated factor at the beginning of the characteristic word is its prefix.

These description of Rauzy graphs give another understanding of the so-called three-gap theorem, stating that an arithmetic sequence on the circle divides it into arcs with at most three possible lengths, corresponding to the frequencies of factors in the sturmian word. Indeed, two factors belonging to the same branch of the Rauzy graph will have same frequencies, and the frequencies of the factors of the common part will be the sum of the two other frequencies.

The dynamical point of view of sturmian words as coding of rotations on the circle gives an analytic description of the intercept. Namely, an intercept can be expressed as an infinite alternating sum of irrational number obtained by shifting the original sequence of partial quotients, pounded with integer coefficients satisfying Ostrowski conditions. However, computations difficulties aside, there remains the problem of combinatorialy define the intercept, in addition to the non-injectivity of this representation.

Ostrowski conditions may be view as a condition of injectivity of the number system associated to the continuants of the continued fraction expansion. They describe the rule that are being applied when one wants to sum two integers written in Ostrowski expansion, an operation that is very far from being understood.

In the dynamical setting given by an infinite word, and given an element in its dynamical subshift, realized as an approximation by suffixes of the base word, one can naturally consider the sequence of natural integers encoding each of the corresponding suffixes. In the context of Sturmian words, this sequence consists of partial expansions of an infinite Ostrowski expansion, linking the chaos of dynamics to the rigidity of combinatorics.

This infinite Ostrowski expansion is obtained by reading the operation of shifting on the Rauzy graphs of Sturmian words. Namely, the partial sums are obtained by considering the case of a length that is the increment of the length of bispecials that are purely central (that is, coming from standard words that are not semi-standard).

This observation allows us to define what we call the *formal intercept* of a sturmian word. They are formally defined as the projective limit of the system of integers with a given maximal number of terms in their Ostrowski expansion. Our main result is that this definition fully describes in a combinatorial and dynamical point of view the set of Sturmian words, in a bijective way.

To prove this result, we use the particularly simple shape of Rauzy graphs of Sturmian words. We use the computations of lengths of cycles previously obtained to check the congruences relations to build a Sturmian words associated to a given formal intercept. For the converse, we characterize the formal intercept as an exact approximation of a given Sturmian word.

We will compute the formal intercept of the two sturmian extensions of the caracteristic word. They have the particularity of having their support in even places and odd places respectively, the support being defined as the set of indices with non-zero Ostrowski coefficients.

We will present the content of a work in progress, that goes as follows.

The action of the shift on Sturmian words allows us to add 1 to a formal intercept, therefore defining the addition of a formal intercept with a natural number. We will say that two formal intercept are equivalent whenever they coincide up to addition by some integers. For Sturmian words that are not equivalent to the characteristic word, this reduces to equality of almost all Ostrowski coefficients.

On the other hand, given a formal intercept one can use product formulas on reversal of standard words to define a Sturmian word. However this process will not reach the sturmian words that are equivalent to the characteristic word without being one of its suffixes. This operation, presented as a product formula, is convenient when considering questions of computations.

The sturmian words obtained by this process have a formal intercept that is complementary of the base intercept. That is, for a sum law defined on the set of classes of equivalence of Sturmian words, this amounts to the consideration of the opposite of a given Sturmian word.

The sum law defined this way would give us an isomorphism between the set of equivalence classes of Sturmian words and the quotient of the real numbers by the subgroup generated by 1 and the slope of the sturmian word. We will present how certain formulas on continuants of continued fraction expansions of quadratic numbers can be interpreted as elements of torsion in the group of equivalence class of Sturmian words, giving an asymptotic realisation of the operation of division by a natural integer in Ostrowski expansion.

UN ENSEMBLE INFINI DÉCIDABLE DE POINTS FIXES SANS CUBE ADDITIF

DAMIE JAMET, FLORIAN LIETARD, AND THOMAS STOLL

Résumé. Dans ce travail, nous exhibons une classe infinie de points fixes de morphismes pour lesquels le problème d'évitabilité des cubes additifs est décidable. Pour cela, nous étendons l'algorithme proposé par Cassaigne *et al.* [1] à l'ensemble des morphismes semblables (au sens des matrices d'incidence) au morphisme étudié dans leur article. Comme dans l'article de Cassaigne *et al.*, la preuve est en partie informatique : autrement dit, elle nécessite une étude exhaustive, par ordinateur, d'un grand nombre de cas.

Le premières études sur les problèmes d'évitement remontent aux travaux de A. Thue [2, 3]. On sait ainsi qu'il existe un mot infini sans carré sur un alphabet de trois lettres (un *carré* est un mot fini de la forme w_1w_2 où $w_1 = w_2$).

Dans [4], F.M. Dekking montre qu'il est existe un mot infini sur 3 lettres sans cube abélien (un mot fini w_1w_2 sur l'alphabet Σ est un *carré abélien* si w_1 et w_2 sont de même longueur et si w_1 est l'image de w_2 par une permutation de Σ). Dans [5], V. Keränen a montré l'existence d'un mot infini sur 4 lettres sans carré abélien.

Dans [1], Cassaigne *et al.* montrent que le point fixe $w_0 = \lim_{n\to\infty} \varphi_0^n(0) = 031430\ldots$ du morphisme $\varphi_0 : 0 \mapsto 03,\ 1 \mapsto 43,\ 3 \mapsto 1,\ 4 \mapsto 01$ est sans cube additif : un *cube additif* est un mot fini sur $\Sigma \subset \mathbb{N}$ de la forme $w_1w_2w_3$ où w_1, w_2 et w_3 sont trois mots finis de même longueur et :

$$\sum_{i_1 \in w_1} i_1 = \sum_{i_2 \in w_2} i_2 = \sum_{i_3 \in w_3} i_3.$$

Dans notre exposé nous montrerons dans un premier temps que l'algorithme proposé dans [1] s'étend de manière naturelle à l'ensemble des morphismes semblables (au sens des matrices d'incidence) au morphisme φ_0. En particulier, nous exhibons ainsi une classe potentiellement infinie (à un facteur multiplicatif et à une translation des lettres près sur l'alphabet) de points fixes sur 4 lettres sans cube additif.

Bien que les morphismes considérés soient semblables deux à deux, il est intéressant de constater que leurs points fixes ne possèdent pas tous les mêmes propriétés additives. Par exemple, soit $w_1 = \lim_{n\to\infty} \varphi_1^n(6) = 602106\ldots$ le point fixe du morphisme $\varphi_1 : 0 \mapsto 2,\ 1 \mapsto 62,\ 2 \mapsto 10, 6 \mapsto$

60 semblable au morphisme φ_0. Le mot infini w_1 est l'image du mot infini w_0 par la transformation de $\{0,1,3,4\}$ dans $\{0,1,2,6\}$ comme suit : $0 \mapsto 6$, $3 \mapsto 0$, $1 \mapsto 2$, $4 \mapsto 1$. Notre implémentation de l'algorithme de Cassaigne *et al.*[1] nous donne une preuve informatique du fait que w_1 est sans cube additif. Cependant, alors que nous conjecturons que tous les carrés additifs dans w_0 sont des carrés abéliens, nous remarquons que le mot w_1 possède des carrés additifs propres, autrement dit, non abéliens, comme par exemple

$$w_1 = 6\,0\,2\,1\,0\,6\,2\,2\,6\,0\,1\,0\,1\;\underline{0\,6\,0\,2\,6}\;\underline{2\,2\,6\,2\,2}\;6\,0\,2\cdots$$

Cette partie sera consacrée à une étude numérique détaillée de ce phénomène.

Nous terminerons notre exposé par une présentation des pistes envisagées pour nous attaquer aux problèmes ouverts suivants :

(1) Existe-t-il un mot infini sur l'alphabet $\{0,1,2,3\}$ sans cubes additifs ?

(2) Quel est le plus petit morphisme sur 3 lettres dont le point fixe ne possède pas de cubes additifs ?

(3) Existe-t-il un mot infini sur un alphabet fini sans carrés additifs ?

Références

[1] J. Cassaigne, J. D. Currie, L. Schaeffer, and J. Shallit. *Avoiding Three Consecutive Blocks of the Same Size and Same Sum. J. ACM*, 61, 2, Article 10, 2014.

[2] A. Thue. *Über unendliche Zeichenreihen.* Skrifter udgivne af Videnskabsselskabet i Christiania : Mathematisk-naturvidenskabelig Klasse. 1906.

[3] A. Thue. *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen.* J. Dybwad, 1912.

[4] F.M Dekking. Strongly non-repetitive sequences and progression-free sets. *Journal of Combinatorial Theory, Series A*, 27(2) :181 – 185, 1979.

[5] V. Keränen. *Abelian squares are avoidable on 4 letters*, pages 41–52. Springer Berlin Heidelberg, Berlin, Heidelberg, 1992.

E-mail address: damien.jamet@loria.fr
E-mail address: florian.lietard@loria.fr
E-mail address: thomas.stoll@univ-lorraine.fr

Loria – UMR 7503, Campus Scientifique, BP 239, 54506 Vandoeuvre-lès-Nancy

Loria – UMR 7503, Campus Scientifique, BP 239, 54506 Vandoeuvre-lès-Nancy – Institut Élie Cartan de Lorraine – UMR 7502, Université de Lorraine, Site de Nancy, B.P. 70239, F-54506 Vandoeuvre-lès-Nancy Cedex

Institut Élie Cartan de Lorraine – UMR 7502, Université de Lorraine, Site de Nancy, B.P. 70239, F-54506 Vandoeuvre-lès-Nancy Cedex

1. voir https://members.loria.fr/FLietard/publications/

Autour du déséquilibre des mots C-adiques

Mélodie Andrieu

28 mai 2018

Résumé

Nous étudions une propriété combinatoire, le déséquilibre, d'une classe particulière de mots sur l'alphabet $\{a, b, c\}$: les mots C-adiques. En particulier, nous exhibons des familles de mots C-adiques de déséquilibres arbitrairement grands, et même des mots C-adiques de déséquilibre infini. Ces constructions ont été obtenues par l'exploration d'un automate et l'étude de ses chemins.

1 Motivations

À l'algorithme de fraction continue soustractif décrit par l'itération de l'application

$$\begin{array}{rcll} (\mathbb{R}^+)^2 & \to & (\mathbb{R}^+)^2 & \\ (x, y) & \mapsto & (x - y, y) & \text{si } x \geq y \\ & & (x, y - x) & \text{sinon} \end{array}$$

est associée une classe particulière de mots infinis binaires : les mots sturmiens. Ceux-ci jouissent de deux caractérisations combinatoires : d'une part, ce sont exactement les mots de complexité $n + 1$, c'est-à-dire les mots qui admettent $n + 1$ facteurs de longueur n pour tout entier n ; d'autre part, ce sont les mots apériodiques dont le déséquilibre vaut 1, c'est-à-dire les mots apériodiques dans lesquels chaque lettre apparaît, à une unité près, un même nombre de fois dans tous les facteurs d'une longueur donnée.

Plusieurs tentatives ont été faites pour généraliser les fractions continues à des triplets de réels positifs. Un tel algorithme pourrait permettre d'approcher simultanément deux réels par une suite de couples de nombres rationnels.

Dans ce document, nous nous interrogeons sur les mots C-adiques, qui sont les mots ternaires associés à l'algorithme [3] :

$$\begin{array}{rcll} (\mathbb{R}^+)^3 & \to & (\mathbb{R}^+)^3 & \\ (x, y, z) & \mapsto & (x - z, z, y) & \text{si } x \geq z \\ & & (y, x, z - x) & \text{sinon.} \end{array}$$

Tout comme les mots d'Arnoux-Rauzy, ces mots sont de complexité $2n+1$. L'intérêt de cet algorithme est qu'il admet comme instance n'importe quel triplet de réels positifs, contrairement à l'algorithme d'Arnoux-Rauzy qui n'est défini que sur un ensemble de mesure de Lebesgue nulle.

Aussi, il est naturel de s'interroger sur l'existence d'une borne uniforme pour le déséquilibre. Hélas, comme pour les mots d'Arnoux-Rauzy [2], nous pouvons construire des familles de mots C-adiques de déséquilibre aussi grand que souhaité, et même, par un lemme de pompage, des mots de déséquilibre infini.

2 Déséquilibre, mots C-adiques

Soit u un mot fini sur l'alphabet ternaire $A = \{a, b, c\}$ et $\alpha \in A$ une lettre. On désigne par $|u|_\alpha$ le nombre d'occurrences de la lettre α dans le mot u. Le *vecteur de Parikh* de u est le vecteur $\chi(u) = (|u|_a, |u|_b, |u|_c)$, qui compte les multiplicités de chacune des lettres de l'alphabet. Remarquons que la somme des coordonnées de ce vecteur est égale à la longueur du mot u, que l'on note $|u|$. Étant donné deux mots de même longueur u et v, on appelle *vecteur de déséquilibre de u et v* la différence de leurs vecteurs de Parikh. La somme des coordonnées d'un tel vecteur est nulle. Le déséquilibre d'un mot infini w est la quantité (éventuellement infinie) :

$$d = \sup_{n\in\mathbb{N}} \sup_{u,v\in F_n(w)} ||\chi(u) - \chi(v)||_\infty,$$

qui s'écrit encore :

$$d = \sup_{n\in\mathbb{N}} \sup_{u,v\in F_n(w)} \max_{\alpha\in\{a,b,c\}} ||u|_\alpha - |v|_\alpha|.$$

Elle mesure les iniquités de répartition entre les lettres dans un mot donné.

Dans toute la suite, on s'intéressera aux substitutions $c_1 : a \mapsto a, b \mapsto ac, c \mapsto b$ et $c_2 : a \mapsto b, b \mapsto ac, c \mapsto c$, qui proviennent de l'algorithme de fraction continue 1 [3]. On notera aussi $C = \{c_1, c_2\}$.

Un *mot C-adique* est un mot infini de la forme $w = lim_{n\to\infty} s_1 \circ ... \circ s_n(w')$, où $(s_k)_{k\in\mathbb{N}} \in C^{\mathbb{N}}$ porte le nom de *suite directrice* de w, et où w' est un mot infini quelconque sur A ; avec la condition supplémentaire que chacune des substitutions c_1 et c_2 apparaisse une infinité de fois dans la suite directrice.

Remarque : pour S un ensemble de substitutions, on peut étudier les mots S-adiques dans un cadre plus général [1].

Enfin, pour $s \in C$, nous introduisons les applications ${}^-s$ et s^- qui, à un mot fini non vide u, associent le mot $s(u)$ auquel on efface la première (resp. la dernière) lettre. Ces applications ne sont pas des morphismes.

3 Construction de mots de C-adiques de déséquilibres arbitrairement grands

Pour tout entier n, nous souhaitons exhiber un mot C-adique de déséquilibre supérieur à n. Pour ce faire, nous allons construire par récurrence une suite $(w_n)_{n\in\mathbb{N}}$ de mots C-adiques, et donner sur chacun d'eux deux facteurs u_n et v_n de même longueur, dont le vecteur de déséquilibre est de norme n. Cela assure que w_n a pour déséquilibre au moins n. L'intuition de ces constructions provient de l'exploration d'un automate et de l'étude de ses chemins.

Construction de (w_n).

Soit w_0 n'importe quel mot C-adique, par exemple $c_1 \circ c_2 \circ c_1 \circ c_2 \circ ...(a)$. Posons $w_1 = c_2 \circ c_2 \circ c_2(w_0)$ et pour tout $n \geq 1$:

$$\begin{cases} w_{n+1} = c_1^{2n+2} \circ c_2(w_n) & \text{si } n \text{ est impair} \\ w_{n+1} = c_2^{2n+2} \circ c_1(w_n) & \text{sinon.} \end{cases}$$

Pour tout entier n, w_n est un mot C-adique.

Construction des suites de facteurs (u_n) et (v_n).

La lettre b apparaît dans w_0 (la complexité nous le garantit), donc $acc = c_2 \circ c_2 \circ c_2(b)$ apparaît dans w_1. Posons $u_1 = ac$, $v_1 = cc$ et, pour tout $n \geq 1$:

$$\begin{cases} u_{n+1} = c_1^- \circ (c_1 \circ c_1)^n \circ c_1 \circ c_2(u_n) & \text{si } n \text{ est impair} \\ v_{n+1} = c_1 \circ (c_1 \circ^- c_1)^n \circ c_1 \circ c_2(v_n) \end{cases}$$

$$\begin{cases} u_{n+1} = c_2 \circ (c_2 \circ c_2^-)^n \circ c_2 \circ c_1(u_n) & \text{sinon.} \\ v_{n+1} =^- c_2 \circ (c_2 \circ c_2)^n \circ c_2 \circ c_1(v_n) \end{cases}$$

Les mots u_n et v_n sont bien facteurs de w_n, pour chaque n.

Lemme 1. *Soit $n \geq 1$.*

— *Si n est impair, alors u_n commence par a et termine par c, v_n commence et termine par c, et $\chi(u_n) - \chi(v_n) = (n, 1-n, -1)$.*

— *Sinon, alors u_n commence et termine par a, v_n commence par a et termine par c, et $\chi(u_n) - \chi(v_n) = (1, n-1, -n)$.*

Theorème 1. *Il existe des mots C-adiques de déséquilibre arbitrairement grand.*

4 Construction de mots C-adiques de déséquilibre infini

Nous venons de construire une famille de mots C-adiques pour laquelle le déséquilibre n'est pas borné. Toutefois, le déséquilibre des mots pris individuellement peut l'être. Pour construire un mot dont le déséquilibre est infini, nous allons recourir à un lemme de pompage.

La première étape consiste à montrer que lorsque l'on compose un mot C-adique par l'une ou l'autre des substitutions c_1 et c_2, on ne rééquilibre pas le mot au-delà d'une certaine proportion.

Proposition 1. *Si w est un mot C-adique de déséquilibre $Des(w) \geq 3n$, alors $c_1(w)$ (resp. $c_2(w)$) est un mot C-adique de déséquilibre $Des(w) \geq n$.*

Pour tout entier naturel m, on note désormais $C^m := \{s_0 \circ ... \circ s_{m-1} \in \{c_1, c_2\}^m\}$, et $C^* := \cup_{m\in\mathbb{N}} C^m$.

Corollaire 1 (Lemme de pompage). $\forall s \in C^*, \forall n \in \mathbb{N}, \exists \sigma \in C^*$ *tel que pour tout mot C-adique w, $Des(s \circ \sigma(w)) \geq n$.*

Theorème 2. *Il existe un mot C-adique de déséquilibre infini.*

Démonstration. D'après le corollaire 1, je peux construire une suite $(\sigma_k)_{k\in\mathbb{N}} \in (C^*)^{\mathbb{N}}$ telle que pour tout entier naturel n et pour tout mot C-adique w, le déséquilibre du mot $\sigma_0 \circ ... \circ \sigma_n(w)$ vaut au moins n. Ainsi, le mot limite $w_\infty = \lim_{n\to\infty} \sigma_0 \circ ...\sigma_n(w_0)$ (s'il existe, sinon prendre une valeur d'adhérence de la suite), où w_0 est un mot C-adique quelconque, est lui même un mot C-adique (chaque σ_k fourni par le corollaire contient c_1 et c_2), et son déséquilibre vaut au moins n, pour tout n ; il est donc de déséquilibre infini. ☐

Références

[1] Valérie Berthé, Wolfgang Steiner, and Jörg Thuswaldner. Geometry, dynamics, and arithmetic of s-adic shifts. *preprint https ://arxiv.org/abs/1410.0331*, 2014.

[2] Julien Cassaigne, Sébastien Ferenczi, and Luca Q. Zamboni. Imbalances in arnoux-rauzy sequences. *Annales de l'Institut Fourier*, 50 :1265–1276, 2000.

[3] Julien Cassaigne, Sébastien Labbé, and Julien Leroy. A set of sequences of complexity 2n+1. In *Combinatorics on Words - 11th International Conference, WORDS 2017, Montréal, QC, Canada, September 11-15, 2017, Proceedings*, pages 144–156.

Number of valid decompositions of Fibonacci prefixes

A. E. Frid, P. Bonardo

May 29, 2018

Abstract

We establish several recurrence relations and an explicit formula for the number of factorizations of the length-n prefix of the Fibonacci word into a (not strictly) decreasing sequence of standard Fibonacci words (OEIS sequence A300066).

1 Introduction

Extended Ostrowski numeration systems were introduced in [4] to solve a problem on palindromes in Sturmian words. A representation of n in such a system related to a given Sturmian slope corresponds to a factorization of the prefix of length n of the standard Sturmian word of this slope as a concatenation of finite standard words in a non-strictly decreasing order. Since in this abstract we consider only the Fibonacci case, it is reasonable to give a Fibonacci example: consider the prefix *abaababaabaaba* of the Fibonacci word of length 14 and its decompositions to standard words $s_0 = a$, $s_1 = ab$, $s_2 = aba$, $s_3 = abaab$, $s_4 = abaababa$, $s_5 = abaababaabaab$ in a decreasing order. We see that

$$\begin{aligned} abaababaabaaba &= (abaababaabaab)(a) = s_5 s_0 \\ &= (abaababa)(abaab)(a) = s_4 s_3 s_0 \\ &= (abaababa)(aba)(ab)(a) = s_4 s_2 s_1 s_0 \\ &= (abaababa)(aba)(aba) = s_4 s_2 s_2 \\ &= (abaab)(aba)(aba)(aba) = s_3 s_2 s_2 s_2 \\ &= (abaab)(aba)(aba)(ab)(a) = s_3 s_2 s_2 s_1 s_0. \end{aligned}$$

These six factorizations correspond to six *valid* representations of 14:

$$14 = \overline{100001} = \overline{11001} = \overline{10111} = \overline{10200} = \overline{1300} = \overline{1211}.$$

If we restrict ourselves to representations corresponding to strictly decreasing sequences, or, which is the same, to the representations only containing zeros and ones, their number for each n is equal to the well-studied OEIS sequence A000119 (see, e.g., [2]). In particular, the lower limit of the sequence is 1, and the upper asymptotics grows as $O(\sqrt{n})$. But here we consider the number of all valid representations of n, denoted by $T(n)$, so that, for example, $T(14) = 6$, the obtained sequence is new and was just recently uploaded to the OEIS as A300066. Here we prove a series of recurrence relations and an explicit formula for it.

2 Result

Let φ denote the golden ratio, $\varphi = \frac{1+\sqrt{5}}{2}$. The Fibonacci word is a Sturmian word $s = s[1]s[2]\cdots$ of the slope $1/(\varphi+1) = 1/\varphi^2$ and of zero intercept, that is, for all n, we have

$$s[n] = \begin{cases} a, & \text{if } \{n/\varphi^2\} < 1 - 1/\varphi^2, \\ b, & \text{otherwise.} \end{cases} \tag{1}$$

Here $\{x\}$ denotes the fractional part of x. Another way to construct s is to consider it as a limit $s = \lim s_n$ of finite standard words

$$s_{-1} = b, s_0 = a, s_{n+1} = s_n s_{n-1} \text{ for all } n \geq 0. \tag{2}$$

We write $N = \overline{k_n \cdots k_0}$ and call this representation of N *valid* if $k_i \geq 0$ for all i and $s(0..N] = s_n^{k_n} s_{n-1}^{k_{n-1}} \cdots s_0^{k_0}$, where $s(0..N]$ is the prefix of length N of the Fibonacci word. The number of valid representations of N is denoted by $T(N)$.

Proposition 1. *If $s[n] = a$, all valid representations of n end with an even number of 0s. If $s[n] = b$, all of them end with an odd number of 0s.*

The main result of this abstract is the following

Theorem 1. *If $s[n] = a$, then $T(n) = \lceil n/\varphi^2 \rceil$, or, which is the same, $T(n)$ is equal to the number of occurrences of b to $s(0..n]$ plus one. If $s[n] = b$, then $T(n) = \lceil n/\varphi^3 \rceil$, or, which is the same, $T(n)$ is equal to the number of occurrences of aa to $s(0..n]$ plus one.*

The proof of the theorem is based on several recurrence relations on $T(n)$:

Proposition 2. *For all $\bar{s}$, $T(\overline{r0}) \geq T(\bar{r})$. If $r = r'10^{2k}$ for some $k \geq 0$, then $T(\overline{r0}) = T(\bar{r})$.*

Proposition 3. *For all $z \in \{0,1\}^*$ and for all $k \geq 1$,*

$$T(\overline{z10^{2k}}) = T(\overline{z10^{2k-2}}) + T(\overline{z(01)^k}).$$

Proposition 4. *For all $z \in \{0,1\}^*$ and for all $k \geq 1$,*

$$T(\overline{z10^k1}) = \begin{cases} T(\overline{z10^{k+1}}), & \textit{if } k \textit{ is odd}, \\ T(\overline{z10^k}) + T(\overline{z(01)^{k/2}}), & \textit{if } k \textit{ is even}. \end{cases}$$

Propositions 2 to 4 give a full list of recurrence relations sufficient to compute $T(n)$ for every $n > 1$, starting from $T(1) = 1$. In particular, as corollaries, we get simple formulas on the values of T on Fibonacci numbers and their predecessors: starting from $F_1 = 1$, $F_2 = 2$, $F_{n+2} = F_{n+1} + F_n$, we get

$$T(F_{2n-1}) = T(F_{2n}) = F_{2n-3} + 1$$

and

$$T(F_{2n} - 1) = T(F_{2n+1} - 1) = F_{2n-2}.$$

The same recurrence relations serve to prove Theorem 1.

3 Acknowledgement

We are deeply grateful to J. O. Shallit for computing the first values of the considered sequence and submitting it to the On-Line Encyclopedia of Integer Sequences.

References

[1] J.-P. Allouche, J. Shallit. Automatic Sequences: Theory, Applications, Generalizations. Cambridge University Press, 2003.

[2] J. Berstel. An exercise on Fibonacci representations. Theor. Inform. Appl. 35 (2002) 491–498.

[3] J. Berstel, P. Séébold. Sturmian words. In: M. Lothaire, Algebraic combinatorics on words. Cambridge University Press, 2002. Chapter 2, pp. 45–110.

[4] A. E. Frid, Sturmian numeration systems and decompositions to palindromes. European J. Combin. 71 (2018) 202–212.

[5] C. G. Lekkerkerker. Voorstelling van natuurlijke getallen door een som van getallen van Fibonacci. Simon Stevin 29 (1952) 190–195.

[6] A. Ostrowski. Bemerkungen zur Theorie der diophantischen Approximationen. Hamb. Abh. 1 (1921) 77–98.

[7] E. Zeckendorf. Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas. Bull. Soc. R. Sci. Liège 41 (1972) 179–182.

Repetition avoidance in products of factors

Pamela Fleischmann

Institut für Informatik, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

fpa@informatik.uni-kiel.de

Pascal Ochem

LIRMM, CNRS, Univ. Montpellier, Montpellier, France

ochem@lirmm.fr

Kamellia Reshadi

Institut für Informatik, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

kre@informatik.uni-kiel.de

Abstract

We consider a variation on a classical avoidance problem from combinatorics on words that has been introduced by Mousavi and Shallit at DLT 2013. Let $\texttt{pexp}_i(w)$ be the supremum of the exponent over the products of i factors of the word w. The repetition threshold $\mathrm{RT}_i(k)$ is then the infimum of $\texttt{pexp}_i(w)$ over all words $w \in \Sigma_k^\omega$. Moussavi and Shallit obtained that $\mathrm{RT}_i(2) = 2i$ and $\mathrm{RT}_2(3) = \frac{13}{4}$. We show that $\mathrm{RT}_i(3) = \frac{3i}{2} + \frac{1}{4}$ if i is even and $\mathrm{RT}_i(3) \geqslant \frac{3i}{2} + \frac{1}{6}$ if i is odd and $i \geqslant 3$.

Keywords: Words; Repetition avoidance.

1 Main results

Mousavi and Shallit [2] have considered two generalizations of the avoidance of fractional repetitions in infinite words. A word is circularly r^+-power-free if it does not contain a factor pxs such that sp is a repetition of exponent strictly greater than r. Let $\Sigma_k = \{0, 1, \ldots, k-1\}$. The smallest real number r such that w is r^+-power-free is denoted by $\texttt{cexp}(w)$. Let $\mathrm{RTC}(k)$ denote the minimum of $\texttt{cexp}(w)$ over every $w \in \Sigma_k^\omega$. Similarly, $\texttt{pexp}_i(w)$ is the smallest real number r such that every product of i factors of w is r^+-power-free word and $\mathrm{RT}_i(k)$ is the minimum of $\texttt{pexp}_i(w)$ over every $w \in \Sigma_k^\omega$.

In this paper, we consider the ternary alphabet. We obtain bounds on $\mathrm{RT}_i(3)$ which extend the result of Mousavi and Shallit that $\mathrm{RT}_2(3) = \frac{13}{4}$.

Proposition 1. $\mathrm{RT}_2(k) = \mathrm{RTC}(k)$.

Proof. The language of words in Σ_k^* avoiding circular repetitions of exponent at least e (or strictly greater than e) is a factorial language. As it is well-known [1], if a factorial language is infinite, then it contains a uniformly recurrent word w. By Proposition 14 in [2], $\texttt{pexp}_2(w) = \texttt{cexp}(w)$. This implies that $\mathrm{RT}_2(k) = \mathrm{RTC}(k)$. □

Proposition 2. *If i is even and $i \geqslant 2$, then* $\mathrm{RT}_i(3) \geqslant \frac{3i}{2} + \frac{1}{4}$.

Proof. Mousavi and Shallit [2] have proved that $\mathrm{RT}_2(3) = \frac{13}{4}$, which settles the case $i = 2$. We have double checked their computation of the lower bound $\mathrm{RT}_2(3) \geqslant \frac{13}{4}$. Suppose that i is a fixed even integer and that w_3 is an infinite ternary word. The lower bound for $i = 2$ implies that there exists two factors u and v such that $uv = t^e$ with $e \geqslant \frac{13}{4}$. Thus, the prefix t^3 of uv is also a 2-terms product of factors of w_3. So we can form the i-terms product $(t^3)^{i/2-1}uv$ which is a repetition of the form t^x with exponent $x = 3\left(\frac{i}{2} - 1\right) + e \geqslant 3\left(\frac{i}{2} - 1\right) + \frac{13}{4} = \frac{3i}{2} + \frac{1}{4}$. This is the desired lower bound. □

Proposition 3. *If i is odd and $i \geqslant 3$, then* $\mathrm{RT}_i(3) \geqslant \frac{3i}{2} + \frac{1}{6}$.

Proof. Suppose that $i \geqslant 3$ is a fixed odd integer, that is, $i = 2j+1$. Suppose that w_3 is a recurrent ternary word such that the product of i factors of w_3 is never a repetition of exponent at least $\frac{3i}{2} + \frac{1}{6} = 3j + \frac{5}{3}$. First, w_3 is square-free since otherwise there would exist an i-terms product of exponent $2i$. Also, w_3 does not contain two factors u and v with the following properties:

- $uv = t^3$,
- $u = t^e$ with $e \geqslant \frac{5}{3}$.

Indeed, this would produce the i-terms product $(uv)^j u$ which is a repetition of the form t^x with exponent $x = 3j + e \geqslant 3j + \frac{5}{3}$.

2

So if a, b, and c are distinct letters, then w_3 does not contain both $u = abcab$ and $v = cabc$ and w_3 does not contain both $u = abcbabc$ and $v = babcb$. A computer check shows that no infinite ternary square-free word satisfies this property. This proves the desired lower bound. □

Proposition 4. *If i is even and $i \geqslant 2$, then* $\mathrm{RT}_i(3) \leqslant \frac{3i}{2} + \frac{1}{4}$.

Proof. Let i be any even integer at least 2. To prove this upper bound, it is sufficient to construct a ternary word w satisfying $\mathtt{pexp}_i(w) \leqslant \frac{3i}{2} + \frac{1}{4}$. The ternary morphic word used in [2] to obtain $\mathrm{RT}_2(3) \leqslant \frac{13}{4}$ seems to satisfy the property. However, it is easier for us to consider another construction. Let us show that the image of every $7/5^+$-free word over Σ_4 by the following 45-uniform morphism satisfies $\mathtt{pexp}_i \leqslant \frac{3i}{2} + \frac{1}{4}$.

$$\begin{aligned}
0 &\mapsto \mathtt{010201210212021012102010212012101202101210212} \\
1 &\mapsto \mathtt{010201210212012101202101210201021202101210212} \\
2 &\mapsto \mathtt{010201210120212012102120210121021201210120212} \\
3 &\mapsto \mathtt{010201210120210121021201210120212012102010212}
\end{aligned}$$

First, we check that such ternary images are $\left(\frac{202}{135}^+, 36\right)$-free using the method in [3]. Since $\frac{202}{135} < \frac{3}{2}$, the period of every repetition formed from i pieces and with exponent at least $\frac{3i}{2}$ must be at most 35. Then we check exhaustively that the ternary images do not contain two factors u and v such that

- $uv = t^e$,
- $e > 3$,
- $9 \leqslant |t| \leqslant 35$.

Thus, the period of every repetition formed from i pieces and with exponent strictly greater than $\frac{3i}{2}$ must be at most 8. Finally, we check exhaustively that $\mathtt{pexp}_i \leqslant \frac{3i}{2} + \frac{1}{4}$ by considering only i-terms products that are repetitions of period at most 8. □

2 Concluding remarks

We conjecture that $\mathrm{RT}_i(3) = \frac{3i}{2} + \frac{1}{6}$ for every odd $i \geqslant 3$, based on numerical evidence. We hope to get a suitable morphism and a proof of this case in the near future. Then the next step would be to consider the 4-letter alphabet. A quick computer check shows that $\mathrm{RT}_i(4) \geqslant i + \frac{1}{2}$ for every $i \geqslant 2$ and we conjecture that this is best possible. However, a proof of an upper bound of the form $\mathrm{RT}_i(4) \leqslant i + c$ cannot be similar to the proof of Proposition 4. That is because the multiplicative factor of i, which drops from $\frac{3}{2}$ when $k = 3$ to 1 when $k = 4$, forbids that the constructed word is a morphic image of a Dejean word.

References

[1] Pytheas Fogg. Substitutions in Dynamics, Arithmetics and Combinatorics. Springer Science & Business Media, 2002.

[2] H. Mousavi and J. Shallit. Repetition avoidance in circular factors. *Developments in Language Theory 2013*, 384–395.

[3] P. Ochem. A generator of morphisms for infinite words. *RAIRO - Theor. Inform. Appl.*, 40:427–441, 2006.

[4] A. Thue. Über unendliche Zeichenreihen. *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania*, 7:1–22, 1906.

Rigidity for some dynamical systems of arithmetic origin

Sébastien Ferenczi, Pascal Hubert
CNRS - Institut de Mathématiques de Marseille
France

May 29, 2018

The rigidity property for a measure-theoretic dynamical systems is the convergence to the identity of a sequence of powers of the map. We look at examples of rigid and non-rigid systems in the class of interval exchanges. Following those coming from square-tiled surfaces, which will be mentioned in P. Hubert's lecture, we consider the famous Veech example of 1969 and some generalizations, which are finite extensions of rotations of angle α with marked points β_i : by the same word-combinatorial methods as in those previous cases, we can prove they are rigid if α has unbounded partial quotients, non-rigid if the coding by the partition defined by the β_i is linearly recurrent. In the intermediate case when α has bounded partial quotients but the coding is not linearly recurrent, we have partial results using the Ostrowski expansions of the β_i related to α: there are rigid examples, including Veech 1969 in this case, and non-rigid ones providing the first known examples of non-rigid not linearly recurrent interval exchanges.

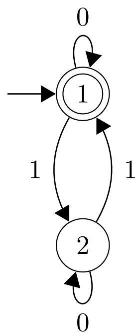
Etude de l'automate minimal des écritures en base 2^p des multiples de l'ensemble de Thue-Morse

Adeline Massuir

30 mai 2018

Ce travail a été réalisé en collaboration avec Emilie Charlier et Célia Cisternino.

L'ensemble de Thue-Morse, $\mathscr{T}$, est l'ensemble des nombres naturels dont l'écriture en base 2 contient un nombre pair de 1. Les premiers éléments sont $0, 3, 5, 6, 9, \ldots$ Cet ensemble est clairement 2-reconnaissable. En effet, l'automate suivant accepte les représentations en base 2 de ces nombres :



De plus, on peut facilement, à partir de cet automate, décrire l'automate qui accepte les représentations de ces nombres dans une base qui est une puissance de 2. Enfin, $\mathscr{T}$ n'étant clairement pas ultimement périodique, on sait par le théorème de Cobham qu'il ne peut être b-reconnaissable pour un b qui n'est pas une puissance (entière positive) de 2.

Nous nous sommes donc intéressées à l'automate minimal de chacun des langages

$$\mathrm{rep}_b\,(m\mathscr{T})$$

où $b = 2^p$ (avec $p > 0$) et $m \in \mathbb{N}_0$.

Pour ce faire, nous avons commencé par construire un automate qui acceptait ce langage. Plusieurs étapes sont nécessaires. Tout d'abord, on construit un automate qui accepte les couples $(\mathrm{rep}_b(n), \mathrm{rep}_b(mn))$ pour tous les naturels n (en prenant pour convention qu'on ajoute des 0 de tête au plus court des deux mots afin que les deux composantes aient la même longueur). Ensuite, on construit l'automate acceptant le langage

$$\{(\mathrm{rep}_b(t), \mathrm{rep}_b(n)) : t \in \mathscr{T}, n \in \mathbb{N}\}$$

(avec la même convention que précédemment). Ensuite, on effectue le produit des deux automates, pour obtenir un automate acceptant le langage suivant (toujours avec la même convention) :

$$\{(\mathrm{rep}_b(t), \mathrm{rep}_b(mt)) : t \in \mathscr{T}\}$$

Finalement, il nous reste à projeter le label de chaque arc sur sa deuxième composante.

A partir de cet automate, nous avons pu prouver que, si on note $b = 2^p$ et $m = k2^i$ où $p > 0$, k impair et $i \geq 0$, alors le nombre d'états de l'automate minimal du langage

$$\mathrm{rep}_b\,(m\mathscr{T})$$

est $2k + \lceil \frac{i}{p} \rceil$.

Ce travail fait écho à l'article de B. Alexeev [1], dans lequel il détermine le nombre d'états de l'automate minimal de chaque langage $\text{rep}_b\,(m\,\mathbb{N})$ pour tous $m, b \in \mathbb{N}_0$. Toutefois, nos procédés de démonstration sont différents des siens.

1. Boris Alexeev, *Minimal DFA for testing divisibility*, Journal of Computer and System Sciences **69** (2004), 2, pp. 235–243.

Generalized Beatty sequences and complementary triples

Abstract

A Beatty sequence is the sequence $A(n) = \lfloor n\alpha \rfloor$ for $n \geq 1$, where α is a positive real number. What Beatty observed is that when B is the sequence $B(n) = \lfloor n\beta \rfloor$, with α and β satisfying

$$\frac{1}{\alpha} + \frac{1}{\beta} = 1, \tag{1}$$

then $(A(n))$ and $(B(n))$ are *complementary* sequences, that is, the sets $\{A(n) : n \geq 1\}$ and $\{B(n) : n \geq 1\}$ are disjoint and their union is the set of positive integers.

A *generalized Beatty sequence* is a sequence v defined by $v(n) = p\lfloor n\alpha \rfloor + qn + r$, where α, p, q, r are real numbers. These occur in several problems, as for instance in homomorphic embeddings of Sturmian languages in the integers ([1]).

Question 1 Let α be an irrational number, and let A defined by $A(n) = \lfloor n\alpha \rfloor$ for $n \geq 1$ be the Beatty sequence of α. Let Id defined by $\mathrm{Id}(n) = n$ be the identity map on the integers. For which sixtuples of integers p, q, r, s, t, u are the two sequences

$$v = pA + q\,\mathrm{Id} + r \text{ and } w = sA + t\,\mathrm{Id} + u$$

complementary sequences?

Question 2 For which nonatuples of integers $(p_1, q_1, r_1, p_2, q_2, r_2, p_3, q_3, r_3)$ the three sequences

$$v_i = p_i A + q_i\,\mathrm{Id} + r_i,\ i = 1, 2, 3$$

are a complementary triple?

Here a *complementary triple* are three sequences, with the property that the sets they determine are disjoint with union the positive integers.

In this talk, based on joint work with Jean-Paul Allouche I give (incomplete) answers to these questions.

[1] Michel Dekking, The Frobenius problem for homomorphic embeddings of languages into the integers, Theoretical Computer Science 732(2018),73-79.

Rigidity and Substitutive Dendric Words

J. Leroy

Département de Mathématique, Université de Liège,
12, Allée de la découverte, 4000 Liège, Belgium.

Joint work with V. Berthé, F. Dolce, F. Durand and D. Perrin

Dendric words are infinite words defined in terms of extension graphs that describe the left and right extensions of their factors. Extension graphs are bipartite graphs that can be roughly described as follows: given an infinite word x, and given a finite factor w of x, one puts an edge between left and right copies of letters a and b such that awb is a factor of x. Dendric words are defined by requiring that the extension graph of each of its factor is a tree. This class of words with linear factor complexity includes classical families of words such as Sturmian words, codings of interval exchanges, or else, Arnoux-Rauzy words. Dendric words have striking combinatorial, ergodic and algebraic properties. This includes in particular algebraic properties of their return words [4], and of maximal bifix codes defined with respect to their languages [2, 5, 6]. They have been introduced in [4] and studied in several papers (as, for instance, [5, 6]), under the name of tree words. We have chosen to call them here dendric words, and the subshifts they generate dendric subshifts, in order to avoid any ambiguity with respect to the notion of tree shift that refers to shifts defined on trees, and not on words (see e.g. [1]).

In this talk, based on [7], I investigate the properties of substitutive dendric words and prove some rigidity properties. Rigidity of an infinite word x has to do with the algebraic properties of its *stabilizer* $\mathrm{Stab}\,(x)$, that is the monoid of substitutions that fix it: an infinite word generated by a substitution is rigid if all the substitutions in $\mathrm{Stab}\,(x)$ are powers of a unique substitution. In this work, we concentrate on the iterative stabilizer according to the terminology of [9]: we focus on non-erasing morphisms and on infinite words generated by iterating a substitution.

There are numerous results on the two-letter case concerning rigidity (see [10, 11] and also [3]). It is indeed well known that Sturmian words generated by substitutions are rigid [10, 11]. The situation is more contrasted as soon as the size of the alphabet increases. For instance, over a ternary alphabet, the stabilizer of a given infinite word can be infinitely generated, even when the word is generated by iterating an invertible primitive morphism (see [8, 9]).

Our main results are the following, where §-adic expansions correspond to the limit of compositions of substitutions of the form $\sigma_1 \circ \cdots \circ \sigma_n$, $\S_e$ stands for

1

the set of elementary positive automorphisms of the free group generated by the alphabet of the word and tame substitutions are elements of $\S_e^*$.

Theorem 1. *A recurrent dendric word over an alphabet A is primitive substitutive if and only if it has an eventually periodic primitive $\S_e$-adic representation.*

Theorem 2. *Let x be a dendric word. Primitive substitutions in the stabilizer Stab (x) of x coincide up to powers. More precisely, if x is a fixed point of both σ and τ primitive substitutions, then there exist $i, j \geq 1$ such that $\tau^i = \sigma^j$.*

Let x be a recurrent substitutive dendric word. There is a primitive tame substitution θ such that any primitive substitution $\sigma \in$ Stab (x) has a power that is (tamely) conjugate to a power of θ, that is, there exists a tame substitution τ such that $\sigma^i = \tau\theta^j\tau^{-1}$, for some $i, j \geq 1$.

In particular, if x is a dendric word, any primitive substitution in Stab (x) is a tame substitution.

Theorem 3. *Let (X, S) be an aperiodic minimal dendric subshift. Then it admits no rational topological eigenvalue.*

Corollary 4. *Let (X, S) be an aperiodic minimal dendric subshift. Then, it can neither be generated by a primitive constant length substitution, nor be a Toeplitz subshift.*

Our proofs rely on the notion of return words and on the so-called Return Theorem [4] that states that for every infinite dendric word defined over the alphabet A, the set of (right) return words is a basis of the free group generated by the alphabet A.

References

[1] N. Aubrun, M.-P. Béal, *Tree-shifts of finite type*, Theoret. Comput. Sci. 459 (2012), 16–25.

[2] J. Berstel, C. De Felice, D. Perrin, C. Reutenauer, G. Rindone, *Bifix codes and Sturmian words*, J. Algebra 369 (2012), 146–202.

[3] V. Berthé, D. Frettlöh, V. Sirvent, *Selfdual substitutions in dimension one*, European J. of Combinatorics 33 (2012), 981–1000.

[4] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, G. Rindone, *Acyclic, connected and tree sets*, Monatsh. Math. 176 (2015), 521–550.

[5] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, G. Rindone, *Maximal bifix decoding*, Discrete Math. 338 (2015), 725–742.

[6] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, G. Rindone, *The finite index basis property*, J. Pure Appl. Algebra 219 (2015), 2521–2537.

[7] V. Berthé, F. Dolce, F. Durand, J. Leroy, D. Perrin, *Rigidity and Substitutive Dendric Words*, Int. J. Found. Comput. Sci. (to appear).

[8] V. Diekert, D. Krieger, *Some remarks about stabilizers*, Theoret. Comput. Sci. 410 (2009), 2935–2946.

[9] D. Krieger, *On stabilizers of infinite words*, Theoret. Comput. Sci. 400 (2008), 169–181.

[10] G. Richomme, P. Séébold, *Completing a combinatorial proof of the rigidity of Sturmian words generated by morphisms*, Theoret. Comput. Sci. 428 (2012), 92–97.

[11] P. Séébold, *On the conjugation of standard morphisms*, Theoret. Comput. Sci. 195 (1998), 91–109.

Resynchronizing Classes of Word Relations*

María Emilia Descotte
LaBRI, Université de Bordeaux

Diego Figueira
CNRS, LaBRI, Université de Bordeaux

Gabriele Puppis
CNRS, LaBRI, Université de Bordeaux

1 Introduction

We study relations of finite words, that is, binary relations $R \subseteq \mathbb{A}^* \times \mathbb{A}^*$ for a finite alphabet $\mathbb{A}$. The study of these relations dates back to the works of Büchi, Elgot, Mezei, and Nivat in the 1960s [3, 6, 11], with much subsequent work done later (*e.g.*, [1, 5]). Most of the investigations focused on extending the standard notion of regularity from languages to relations. This effort has followed the long-standing tradition of using equational, operational, and descriptive formalisms – that is, finite monoids, automata, and regular expressions – for describing relations, and gave rise to three different classes of relations: the *Recognizable*, the *Automatic* (*a.k.a. Regular* [1] or *Synchronous* [5]), and the *Rational* relations.

The above classes of relations can be seen as three particular examples of a much larger (in fact infinite) range of possibilities, where relations are described by special languages over extended alphabets, called *synchronizing languages* [8]. Intuitively, the idea is to describe a binary relation by means of a two-tape automaton with two heads, one for each tape, which can move independently one of the other. In the basic framework of synchronized relations, one lets each head of the automaton to either move right or stay in the same position. In addition, one can constrain the possible sequences of head motions by a suitable regular language $C \subseteq \{1,2\}^*$. In this way, each regular language $C \subseteq \{1,2\}^*$ induces a class of binary relations, denoted $\textsc{Rel}(C)$, which is contained in the class of Rational relations (due to Nivat's Theorem [11]). For example, the class of Recognizable, Automatic, and Rational relations are captured, respectively, by the languages $C_{\text{Rec}} = \{1\}^* \cdot \{2\}^*$, $C_{\text{Aut}} = \{12\}^* \cdot \{1\}^* \cup \{12\}^* \cdot \{2\}^*$, and $C_{\text{Rat}} = \{1,2\}^*$. However, it should be noted that other well-known subclasses of rational relations, such as deterministic or functional relations, are not captured by notion of synchronization. In general, the correspondence between a language $C \subseteq \{1,2\}^*$ and the induced class $\textsc{Rel}(C)$ of synchronized relations is not one-to-one: it may happen that different languages C, D induce the same class of synchronized relations. There are thus fundamental questions that arise naturally in this framework: *When do two classes of synchronized relations coincide, and when is one contained in the other?* Our contribution is a precise algorithmic answer to this type of questions.

More concretely, given a binary alphabet $\mathbb{2} = \{1,2\}$ and another finite alphabet $\mathbb{A}$, a word $w \in (\mathbb{2} \times \mathbb{A})^*$ is said to *synchronize* the pair $(w_1, w_2) \in \mathbb{A}^* \times \mathbb{A}^*$ if, for both $i = 1, 2$, w_i is the projection of w on $\mathbb{A}$ restricted to the positions marked with i. For short, we denote this by $[\![w]\!] = (w_1, w_2)$ —*e.g.*, $[\![(1,a)(1,b)(2,b)(1,a)(2,c)]\!] = (aba, bc)$. According to this definition, every word over $\mathbb{2} \times \mathbb{A}$ synchronizes a pair of words over $\mathbb{A}$, and every pair of words over $\mathbb{A}$ is synchronized by (perhaps many) words over of $\mathbb{2} \times \mathbb{A}$. This notion is readily lifted to languages: a language $L \subseteq (\mathbb{2} \times \mathbb{A})^*$ synchronizes the relation $[\![L]\!] = \{[\![w]\!] \mid w \in L\} \subseteq \mathbb{A}^* \times \mathbb{A}^*$. For example, $[\![((1,a)(2,a) \cup (1,b)(2,b))^*]\!]$ denotes the equality relation over $\mathbb{A} = \{a, b\}$.

* This work is going to appear in the proceedings of ICALP 2018.

In this setup, one can define classes of relations by restricting the set of admitted synchronizations. The natural way of doing so is to fix a language $C \subseteq \mathbb{2}^*$, called *control language*, and let L vary over all regular languages over the alphabet $\mathbb{2} \times \mathbb{A}$ whose projections onto $\mathbb{2}$ are in C. Thus, for every regular $C \subseteq \mathbb{2}^*$, there is an associated class $\textsc{Rel}(C)$ of C-controlled relations, namely, relations synchronized by regular languages $L \subseteq (\mathbb{2}\times\mathbb{A})^*$ whose projection onto $\mathbb{2}$ are in C. Clearly, $C \subseteq D \subseteq \mathbb{2}^*$, implies $\textsc{Rel}(C) \subseteq \textsc{Rel}(D)$, but the converse does not hold: while $\textsc{Rel}(C_{\text{Rec}}) = \textit{Recognizable} \subseteq \textit{Automatic} = \textsc{Rel}(C_{\text{Aut}})$, we have $C_{\text{Rec}} \not\subseteq C_{\text{Aut}}$. Moreover, as we have mentioned earlier, different control languages may induce the same class of synchronized relations. For example, once again, the class of Recognizable relations is induced by the control language $C_{\text{Rec}} = \{1\}^*\{2\}^*$, but also by $C'_{\text{Rec}} = \{1\}^*\{2\}^*\{1\}^*$, and the class of Automatic relations is induced by $C_{\text{Aut}} = \{12\}^* \cdot \{1\}^* \cup \{12\}^* \cdot \{2\}^*$, or equally by $C'_{\text{Aut}} = \{21\}^* \cdot \{1\}^* \cdot \{2\}^*$. This 'mismatch' between control languages and induced classes of relations gives rise to the following algorithmic problem.

Class Containment Problem	
Input:	Two regular languages $C, D \subseteq \mathbb{2}^*$
Question:	Is $\textsc{Rel}(C) \subseteq \textsc{Rel}(D)$?

Note that the above problem is different from the (C, D)-membership problem on synchronized relations, which consists in deciding whether $R \in \textsc{Rel}(D)$ for a given $R \in \textsc{Rel}(C)$, and which can be decidable or undecidable depending on C, D [4]. The Class Containment Problem can be seen as the problem of whether *every* C-controlled regular language L has a D-controlled regular language L' so that $[\![L]\!] = [\![L']\!]$. It was proved in [8] that this problem is decidable for some particular instances of D, namely, for $D =$ *Recognizable*, *Automatic*, *Length-preserving* or *Rational*. More specifically, given a regular language C over the binary alphabet $\mathbb{2}$, it is decidable whether $\textsc{Rel}(C)$ is contained or not in *Recognizable* (respectively, *Automatic*, *Length-preserving* and *Rational*). Our main contribution is a procedure for deciding the Class Containment Problem in full generality, i.e. for arbitrary C and D.

► **Main Theorem.** *The Class Containment Problem is decidable.*

In addition, our results show that, for positive instances (C, D), one can effectively transform any regular C-controlled language L into a regular D-controlled language L' so that $[\![L]\!] = [\![L']\!]$. By 'effectively transform' we mean that one can receive as input an automaton (or a regular expression) for L and produce an automaton (or a regular expression) for L'. In particular, we show a normal form of control languages, implying that every synchronized class can be expressed through a control language of star-height at most 1.

Related work. The formalization of a framework in which one can describe classes of word relations by means of synchronization languages is quite recent [8]. As already mentioned, the class containment problem was only addressed for the classes of Recognizable, Automatic and Rational relations, for which several characterizations have been proposed [8]. The formalism of synchronizations has been extended beyond rational relations by means of semi-linear constraints [7] in the context of path querying languages for graph databases.

The paper [2] studies relations with origin information, as induced by non-deterministic (one-way) finite state transducers. Origin information can be seen as a way to describe a synchronization between input and output words – somehow in the same spirit of our synchronization languages – and was exploited to recover decidability of the equivalence problem for transducers. The paper [9] pursues further this principle by studying "distortions" of the origin information, called resynchronizations. Despite the similar terminology and the connection between origins and synchronizing languages, the problems studied in [2, 9] are of rather different nature than our Class Containment Problem.

2 Characterization of the Class Containment Problem

We give an overview of our decision procedure for class containment.

The main idea is to decompose languages as finite unions of what we call *simple* languages. The definition can be found in the article ([10]) but the interest of these languages lies in the fact that we can prove the following two results.

▶ **Proposition 1.** *Every regular language $C \subseteq \mathbb{2}^*$ is effectively $=_{\text{REL}}$-equivalent to a finite union of simple languages.*

The proof of the above proposition is quite involved and it can be found in full detail in the appendix of the article.

▶ **Proposition 2.** *The class containment problem for simple languages C and D is decidable.*

Moreover, Proposition 2 follows from a characterization of class containment for simple languages in terms of their Parikh-images and cycles which is relatively simple.

The generalization of the previous result to the case where C is a finite union of simple languages is given by the following basic result.

▶ **Lemma 3.** $C_1 \cup C_2 \subseteq_{\text{REL}} D$ *iff* $C_1 \subseteq_{\text{REL}} D$ *and* $C_2 \subseteq_{\text{REL}} D$.

The characterization turns out to be more involved when we have unions on the right hand-side. In particular, The analogous of Lemma 3 for unions on the right hand-side does not hold in general.

The characterization we provide is inductive on the number of languages that are unioned on the right hand-side.

References

1 Jean Berstel. *Transductions and Context-Free Languages*. B. G. Teubner, 1979.

2 Mikołaj Bojańczyk. Transducers with origin information. In *International Colloquium on Automata, Languages and Programming (ICALP)*, volume 8573 of *Lecture Notes in Computer Science*, pages 26–37. Springer, 2014. doi:10.1007/978-3-662-43951-7.

3 Julius Richard Büchi. Weak second-order arithmetic and finite automata. *Mathematical Logic Quarterly*, 6(1-6):66–92, 1960.

4 Olivier Carton, Christian Choffrut, and Serge Grigorieff. Decision problems among the main subfamilies of rational relations. *Informatique Théorique et Applications (ITA)*, 40(2):255–275, 2006. doi:10.1051/ita:2006005.

5 Christian Choffrut. Relations over words and logic: A chronology. *Bulletin of the EATCS*, 89:159–163, 2006.

6 Calvin C. Elgot and Jorge E. Mezei. On relations defined by generalized finite automata. *IBM Journal of Research and Development*, 9(1):47–68, 1965. doi:10.1147/rd.91.0047.

7 Diego Figueira and Leonid Libkin. Path logics for querying graphs: Combining expressiveness and efficiency. In *Annual IEEE Symposium on Logic in Computer Science (LICS)*, pages 329–340. IEEE Computer Society Press, 2015. doi:10.1109/LICS.2015.39.

8 Diego Figueira and Leonid Libkin. Synchronizing relations on words. *Theory of Computing Systems*, 57(2):287–318, 2015. doi:10.1007/s00224-014-9584-2.

9 Emmanuel Filiot, Ismaël Jecker, Christof Löding, and Sarah Winter. On equivalence and uniformisation problems for finite transducers. In *International Colloquium on Automata, Languages and Programming (ICALP)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 125:1–125:14. Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPIcs.ICALP.2016.125.

10 D. Figueira M.E. Descotte and G. Puppis. Resynchronizing classes of word relations. In *International Colloquium on Automata, Languages, and Programming (ICALP)*, 2018. URL: `hal-01721046v2`.

11 Maurice Nivat. Transduction des langages de Chomsky. *Annales de l'Institut Fourier*, 18:339–455, 1968.

List of participants

- Andrieu Mélodie
- Bertrand-Mathis Anne
- Bonardo Pierre
- Borel Jean-Pierre
- Bose Sougata
- Bousquet-Mélou Mireille
- Brlek Srecko
- Cassaigne Julien
- Casteigts Arnaud
- Charlier Emilie
- Cisternino Célia
- Costa Alfredo
- Crochemore Maxime
- Dando Louis-Marie
- Daviaud Laure
- Dekking Michel
- Delecroix Vincent
- Derycke Henri
- Descotte Maria
- Duchon Philippe
- Ferenczi Sébastien
- Fleischmann Pamela
- Galanov Ilya
- Gangloff Silvère
- Guibert Olivier

- Hubert Pascal
- Janin David
- Kachigar Ghazal
- Karaboghossian Théo
- Labbé Sébastien
- Lejeune Marie
- Leroux Jérôme
- Leroy Julien
- Lietard Florian
- Massuir Adeline
- Medkova Katerina
- Ochem Pascal
- Parshina Olga
- Pelantova Edita
- Place Thomas
- Postic Mickaël
- Puppis Gabriele
- Reshadi Kamellia
- Ryzhikov Andrew
- Schnoebelen Philippe
- Stipulanti Manon
- štěpán Starosta
- Tanasa Adrian
- Tarsissi Lama
- Uricaru Raluca
- Weil Pascal
- Wojcik Caius
- Zeitoun Marc