TUCS

Tero Harju | Juhani Karhumäki (Eds.)

Proceedings of the

Workshop on Fibonacci Words

Turku, September 2006

Turku Centre *for* Computer Science

TUCS General Publication
No 43, June 2007

Proceedings of the

Workshop on Fibonacci Word

September 27-30, 2006, Turku, Finland

Editors:

Tero Harju
Juhani Karhumäki

TUCS General Publication
No 43, June 2007

Preface

The workshop on *Fibonacci Words* was a three day meeting held in Turku 27 – 29 September, 2006. It was part of the activities of the thematic year on Algorithmic and Discrete Mathematics and arranged in the Department of Mathematics during the academic year 2006/2007 at Turku University. There were 43 registered participants in the workshop. The topics of the workshop belong to combinatorics on words and combinatorial number theory related to Fibonacci numbers. The talks given by the key speakers covered a wide area of the topic varying from general surveys to applications. In addition to 12 invited talks there were 6 contributed presentations. The workshop was scientifically highly successful, and it was a pleasant gathering of people that have a common interest in combinatorics on words. Selected works of the workshop will be published in a special issue of Theoretical Informatics and Applications (ITA) edited by Jean Berstel together with the editors of this report. The thematic year was sponsored by the Academy of Finland, Finnish Cultural Foundation, Finnish Academy of Science and Letters, Vilho, Yrjö and Kalle Väisälä Fund, and Nokia Foundation. In addition, the workshop was sponsored by the Turku University Foundation, Turku Centre for Computer Science (TUCS), Turku Energia and Centro Hotel. This proceedings contains the extended abstracts of most of the talks given during the workshop.

Acknowledgement. We thank Eero Lehtonen for carefully collecting this report.

Tero Harju and Juhani Karhumäki
February 2007, Turku

Contents

Extremal properties of (epi)sturmian sequences and distribution modulo 1

J.-P. Allouche
CNRS, LRI
Bâtiment 490
F-91405 Orsay Cedex
France
allouche@lri.fr

1 Introduction

This extended abstract is an announcement of a paper with A. Glen in preparation [4].

A few months ago JPA came across a paper of Y. Bugeaud and A. Dubickas [6] where the authors describe all irrational numbers $\xi > 0$ such that the fractional parts $\{\xi b^n\}$, $n \geq 0$, all belong to an interval of length $1/b$, where $b \geq 2$ is a given integer. They also prove that $1/b$ is the minimal length having this property. An interesting and unexpected result in their paper is that, when the interval of length $1/b$ is closed, these irrational numbers are exactly the positive real numbers whose base b expansions are characteristic Sturmian sequences on $\{k, k+1\}$, where $k \in \{0, 1, \ldots, b-2\}$. (Recall that characteristic Sturmian sequences are codings of nonperiodic trajectories on a square billiard that start from a corner.)

2 More on Bugeaud-Dubickas' result

Looking at the proofs in [6] one sees that the core of the result is the following property:

Theorem 1. *A binary sequence $u := (u_n)_{n\geq 0}$ is a characteristic Sturmian sequence if and only if, for all $k \geq 0$,*

$$0u \leq T^k u \leq 1u$$

where T is the shift defined by $T((u_n)_{n\geq 0}) = (u_{n+1})_{n\geq 0}$ and the order is the lexicographical order.

Actually this theorem was known before. It was indicated to JPA by G. Pirillo (who published it in [12]): JPA suggested that this could well be already in a paper by S. Gan [8] under a slightly disguised form (which is indeed the case). Also J.-P. Borel and F. Laubie proved one direction of the above theorem, namely that characteristic Sturmian sequences satisfy the inequalities $0u \leq T^k u \leq 1u$ [5].

3 Generalizations

Two directions for generalizations are possible. One is purely combinatorial and looks at generalizations of Sturmian sequences (in particular episturmian sequences that have some aspects of Sturmian sequences and similar extremal properties) or characterizations of finite words related to (epi)sturmian sequences, see [11, 13, 9, 10]. The other is number-theoretic and looks at distribution modulo 1 from a combinatorial point of view: recent papers of Dubickas go in this direction; we cite one of them showing an unexpected occurrence of the Thue-Morse sequence [7].

4 The Thue-Morse sequence shows up

In the paper of Dubickas [7] the Thue-Morse sequence appears when studying the "small" and "large" limit points of $\|\xi(p/q)^n\|$ the distance to the nearest integer of the product of any nonzero real number ξ by the powers of a rational.

Interestingly enough this sequence appeared in 1983 in another question of distribution as a by-product of the combinatorial study of a set of sequences related to iterating continuous maps of the unit interval, see [1, 2, 3].

Theorem 2. *Define the set* Γ *by*

$$\Gamma := \{x \in [0, 1],\ 1 - x \leq \{2^k x\} \leq x\}.$$

Then the smallest limit point of Γ *is the number* $\alpha := \sum a_n/2^n$*, where* $(a_n)_{n\geq 0}$ *is the Thue-Morse sequence. The set* Γ *contains only countably many elements less than* α *and they are all rational. Furthermore any segment on the right of* α *contains uncountably many elements of* Γ*. This structure around* α *repeats:* Γ *is a fractal set.*

The reader will have guessed that the above theorem is a by-product of the combinatorial study of the set

$$\Gamma := \{u \in \{0, 1\}^{\mathbb{N}},\ \forall k \geq 0,\ \overline{u} \leq T^k u \leq u\}$$

where $\overline{u}$ is the sequence obtained by switching 0's and 1's in u (see [1]).

References

[1] J.-P. Allouche, *Théorie des nombres et automates*, Thèse d'État, 1983, Université Bordeaux I.

[2] J.-P. Allouche, M. Cosnard, Itérations de fonctions unimodales et suites engendrées par automates, *C. R. Acad. Sci. Paris, Série I* **296** (1983), 159–162.

[3] J.-P. Allouche, M. Cosnard, Non-integer bases, iteration of continuous real maps, and an arithmetic self-similar set, *Acta Math. Hungar.* **91** (2001) 325–332.

[4] J.-P. Allouche, A. Glen, Extremal properties of (epi)sturmian sequences and distribution modulo 1, in preparation.

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

[6] Y. Bugeaud, A. Dubickas, Fractional parts of powers and Sturmian words, *C. R. Acad. Sci. Paris, Sér. I* **341** (2005) 69–74.

[7] A. Dubickas, On the distance from a rational power to the nearest integer, *J. Number Theory* **117** (2006) 222–239.

[8] S. Gan, Sturmian sequences and the lexicographic world, *Proc. Amer. Math. Soc.* **129** (2001) 1445–1451.

[9] A. Glen, A characterization of fine words over a finite alphabet, Preprint 2006, submitted.

[10] A. Glen, J. Justin, G. Pirillo, Characterizations of finite and infinite episturmian words via lexicographic orderings, Preprint 2006, submitted.

[11] J. Justin, G. Pirillo, On a characteristic property of Arnoux-Rauzy sequences, *Theor. Inform. Appl.* **36** (2002) 385–388.

[12] G. Pirillo, Inequalities characterizing standard Sturmian words, *Pure Math. Appl.* **14** (2003) 141–144.

[13] G. Pirillo, Inequalities characterizing standard Sturmian and episturmian words, *Theoret. Comput. Sci.* **341** (2005) 276–292.

Some effective tools for determining return words

Lubomíra Balková [(1)] [(2)]

Joined work with Edita Pelantová [(1)] and Wolfgang Steiner [(3)]

(1) Doppler Institute for Mathematical Physics and Applied Mathematics and
Department of Mathematics, FNSPE, Czech Technical University,
Trojanova 13, 120 00 Praha 2, Czech Republic
(2) Laboratoire APC, UMR 7164 CNRS & Université Paris 7
2 place Jussieu, 75251 Paris Cedex 05, France
(3) LIAFA, UMR 7089 CNRS & Université Paris 7
2 place Jussieu, 75251 Paris Cedex 05, France

Abstract

Recently, the notion of return words has appeared in various branches of mathematics, for instance in Symbolic Dynamical Systems and Number Theory. This mathematical tool has been introduced by Durand [2] in order to obtain a nice characterization of primitive substitutive sequences. Roughly speaking, for a given factor w of the infinite word u, a return word of w is a segment between two successive occurrences of the factor w. Using return words, Vuillon in [9] has found a new equivalent definition of sturmian words. He has shown that an infinite word u over a biliteral alphabet is sturmian if and only if any factor of u has two return words.

A natural generalization of sturmian words to a multiliteral alphabet is provided by Arnoux-Rauzy words. Justin and Vuillon have proved that Arnoux-Rauzy words of order m have for every factor exactly m return words.

Here,we present three very simple ideas which enable to determine easily the number of return words in some infinite words.

- The first observation is that for study of return words of an infinite uniformly recurrent word u, it suffices to limit the considerations to bispecial factors of u.
- We show the importance of Rauzy graphs for determination of the number of return words.
- In case of infinite words invariant under a substitution, we will make use of the relation between return words of a factor w and return words of its image.

In this abstract, we give a new and very short proof of the number of return words for Arnoux-Rauzy sequences using the ideas above. Let us mention results concerning return words of infinite words $u_\beta = (u_n)_{n\in\mathbb{N}}$ which are associated with simple Parry numbers. The infinite word u_β codes the sequence of gaps between successive β-integers $\mathbb{Z}_\beta$ and it is the fixed point of the canonical substitution associated with β. Simple Parry number is an algebraic integer $\beta > 1$ having a finite Rényi expansion of unity $d_\beta(1) = t_1 \ldots t_m$. In this case, the alphabet of the infinite word u_β contains m letters. In [4], it has been proven that the infinite word u_β is Arnoux-Rauzy if and only if $t_1 = \cdots = t_{m-1}$ and $t_m = 1$. Consequently, in this case, the number of return words is equal to m for every factor of u_β. We show that the same statement about return words holds if $t_1 > \max\{t_2, \ldots, t_{m-1}\}$ and $t_m = 1$. In the case $t_m \geq 2$, we focus on a simple situation where the coefficients $t_1, \ldots, t_{m-1}$ are either all the same or mutually different. Under these conditions, we show that for each factor of u_β the number of return words is either m or $m+1$ and both of the values are reached.

1 Preliminaries

An *alphabet* $\mathcal{A}$ is a finite set of symbols called *letters*. A concatenation of letters is a *word*. The set $\mathcal{A}^*$ of all finite words (including the empty word ε) provided with the operation of concatenation is a free monoid. The length of a word $w = w_0w_1w_2\cdots w_{n-1}$ is denoted by $|w| = n$. We will deal also with right-infinite words $u = u_0u_1u_2\cdots$. A finite word w is called a *factor* (*subword*) of the word u (finite or infinite) if there exist a finite word $w^{(1)}$ and a word $w^{(2)}$ (finite or infinite) such that $v = w^{(1)}ww^{(2)}$. The word w is a *prefix* of u if $w^{(1)} = \varepsilon$. Analogically, w is a *suffix* of u if $w^{(2)} = \varepsilon$. A concatenation of k letters a will be denoted by a^k, a concatenation of infinitely many letters a by a^ω.

Let w be a factor of an infinite word u and let $a, b \in \mathcal{A}$. If wa is a factor of u, then we call a a *right extension* of w. Analogically, if bw is a factor of u, we call b a *left extension* of w. We will denote by $Rext(w)$ the set of all right extensions of w and by $\deg_+(w)$ the number of right extensions of w. Analogically, the set of left extensions of w will be denoted by $Lext(w)$ and its cardinality by $\deg_-(w)$.

Definition 1.1. *Let w be a factor of an infinite word u. The factor w is called*

- *right special if* $\deg_+(w) > 1$,
- *left special if* $\deg_-(w) > 1$,
- *bispecial if w is right special and left special.*

We will denote by $\mathcal{L}(u)$ (language on u) the set of all factors of a word u. $\mathcal{L}_n(u)$ denotes the set of all factors of length n of the word u, clearly

$$\mathcal{L}(u) = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n(u).$$

Definition 1.2. *Let $w = w_0 w_1 \dots w_l$ be a factor of an infinite word $u = u_0 u_1 u_2 \dots$. Then we call occurrence of w an integer j such that $u_j u_{j+1} \dots u_{j+l} = w_0 w_1 \dots w_l$.*

Definition 1.3. *Let w be a factor of an infinite word u and let i, j be its successive occurrences, then $u_i \dots u_{j-1}$ is a return word of w. The set of all return words of w is denoted by $M(w)$, i.e.*

$$M(w) = \{u_i \dots u_{j-1} \mid i, j \quad \textit{being the successive occurrences of } w\}.$$

Definition 1.4. *An infinite word u is called uniformly recurrent if for any positive integer n there exists a positive integer $R(n)$ such that in any factor of u of length $R(n)$ all factors of length n appear.*

Remark 1. *If u is a uniformly recurrent word, then for any factor w, the set of return words of w is finite.*

The measure of variability of local configurations in u is expressed by the *factor complexity function* (or simply *complexity*) $\mathcal{C}_u : \mathbb{N} \to \mathbb{N}$, which associates with $n \in \mathbb{N}$ the number $\mathcal{C}_u(n) := \#\mathcal{L}_n(u)$. One can show that a word u is aperiodic if and only if $\mathcal{C}_u(n) \geq n + 1$ for all $n \in \mathbb{N}$. Infinite aperiodic words with the minimal complexity $\mathcal{C}_u(n) = n + 1$ for all $n \in \mathbb{N}$ are called *sturmian words*. These words are studied intensively, several different definitions of turmian words can be found in [1].

A mapping φ on the free monoid $\mathcal{A}^*$ is called a morphism if $\varphi(vw) = \varphi(v)\varphi(w)$ for all $v, w \in \mathcal{A}^*$. Obviously, for determining the morphism, it suffices to give $\varphi(a)$ for all $a \in \mathcal{A}$. The action of a morphism can be naturally extended on right-sided infinite words by the prescription

$$\varphi(u_0 u_1 u_2 \cdots) := \varphi(u_0)\varphi(u_1)\varphi(u_2) \cdots$$

A non-erasing morphism φ, for which there exists a letter $a \in \mathcal{A}$ such that $\varphi(a) = aw$ for some non-empty word $w \in \mathcal{A}^*$, is called a substitution. An infinite word u such that $\varphi(u) = u$ is called a fixed point of the substitution φ. Obviously, every substitution has at least one fixed point, namely

$$\lim_{n \to \infty} \varphi^n(a).$$

2 Return words and Rauzy graphs

Let us explain at first that in order to study return words $M(w)$ of factors w of an infinite word u, it is possible to limit our considerations to bispecial factors w. If a factor w has a unique right extension, say a, then the sets of occurrences of w and wa coincide. Therefore

$$M(w) = M(wa).$$

If a factor w has a unique left extension, say b, then for $i \geq 1$ the occurrence of w in the infinite word u is i if and only if $i - 1$ is the occurrence of bw. This statement does not hold for $i = 0$, but if u is a uniformly recurrent infinite word, the set $M(w)$ of return words of w stays the same no matter whether we include the return word corresponding to the prefix w of u or not. Consequently, it holds

$$M(bw) = bM(w)b^{-1} := \{bvb^{-1} \mid v \in M(w)\},$$

where the expression bvb^{-1} denotes that the word v is prolonged to the left by the letter b and it is shortened from the right by erasing the letter b.

If w is not right special, then the cardinalities of sets $M(w)$ and $M(wa)$ are the same. Let u be a uniformly recurrent infinite word, analogically as in the previous case, if w is not left special, then $M(w)$ and $M(bw)$ has the same cardinality.

For an aperiodic uniformly recurrent infinite word u, each factor w is a subword of a bispecial factor. To describe cardinality of $M(w)$, it suffices to consider bispecial factors w.

Return words and the role of bispecial factors can be well visualized by means of the so-called Rauzy graphs. *Rauzy graph* Γ_n is an oriented graph with the set of vertices $\mathcal{L}_n$ and the set of edges $\mathcal{L}_{n+1}$. The edge $w_0 w_1 \dots w_n \in \mathcal{L}_{n+1}$ goes from the vertex $w_0 \dots w_{n-1}$ to the vertex $w_1 \dots w_n$. The factor $w = w_0 \dots w_{n-1}$ is left special if and only if at least two edges ends in the vertex w and w is right special if and only if at least two edges begins in w. Each return word of a factor w of the length n is visualized as an oriented walk in Γ_n, which begins and ends in the vertex w, and such that the vertex w is not entered in the course of this walk.

If there exists only one vertex w with $\deg_+(w) > 1$ in Γ_n, then the choice of the edge starting the walk determines the walk uniquely. Consequently, there are exactly $\deg_+(w)$ oriented walks beginning and ending in w and, thus, $\#M(w) = \deg_+(w)$. Analogically, if there exists a unique vertex $\hat{w}$ with $\deg_-(\hat{w}) > 1$ in Γ_n, then $\#M(\hat{w}) = \deg_-(\hat{w})$.

These simple considerations prove the following theorem.

Theorem 2.1. *Let $m \in \mathbb{N}$ and let u be a uniformly recurrent infinite word such that for every $n \in \mathbb{N}$, the Rauzy graph Γ_n contains either only one left special factor $\hat{w}$*

with $\deg_-(\hat{w}) = m$ *or only one right special factor* w *with* $\deg_+(w) = m$*. Then the number of return words is equal to* m *for any factor of the infinite word* u*.*

Let us remind that Arnoux-Rauzy words of order m are defined as infinite words which have for every $n \in \mathbb{N}$ exactly one right special factor w with $\deg_+(w) = m$ and exactly one left special factor $\hat{w}$ $\deg_-(\hat{w}) = m$.

Corollary 2.1. *The number of return words of any factor of an Arnoux-Rauzy word is equal to* m*. In particular, the number of return words of any factor of a sturmian word is two.*

Corollary 2.2. *Let* $d_\beta(1) = t_1 \dots t_m$ *be a Rényi expansion of unity. If the coefficients satisfy both of the following conditions*

$$\begin{array}{ll} (a) & t_m = 1 \\ (b) & t_1 = \cdots = t_{m-1} \quad \textit{or} \quad t_1 > \max\{t_2, \dots, t_{m-1}\} \end{array} \tag{1}$$

then $\#M(w) = m$ *for each factor* $w \in \mathcal{L}(u_\beta)$*.*

References

[1] J. Berstel, *Recent Results on Extension of Sturmian Words*, Int. J. Algebra Comput. **12** (2002), 371-385

[2] F. Durand, *A characterization of substitutive sequences using return words*, Discrete Math. **179** (1998), 89-101

[3] S. Fabre, *Substitutions et β-systèmes de numération* Theoret. Comput. Sci. **137** (1995), 219-236

[4] Ch. Frougny, Z. Masáková, E. Pelantová, *Complexity of infinite words associated with beta-expansions*, RAIRO Theor. Inform. Appl. **38** (2004), 163-185; Corrigendum, RAIRO Theor. Inform. Appl. **38** (2004), 269-271

[5] J. Justin, G. Pirillo, *On a combinatorial property of sturmian words*, Theor. Comput. Sci. **154(2)** (1996), 387-394

[6] W. Parry, *On the beta-expansions of real numbers*, Acta Math. Acad. Sci. Hungar.**11** (1960), 401-416

[7] A. Rényi, *Representations for real numbers and their ergodic properties,* Acta Math. Acad. Sci. Hungar. **8** (1957), 477-493

[8] W. P. Thurston, *Groups, tilings, and finite state automata*, Geometry supercomputer project research report GCG1, University of Minnesota (1989)

[9] L. Vuillon, *A characterization of sturmian words by return words*, Eur. J. Comb. 22(2), 263-275 (2001)

Irreducible balanced pairs on substitutive languages

Bernat Julien

Abstract

Let $\mathcal{L}$ be a language. A balanced pair (u, v) consists of two words u and v in $\mathcal{L}$ which have the same number of occurences of each letter. It is irreducible if the pairs of strict prefixes of u and v of the same length do not form balanced pairs.

In this article, we are interested in computing the set of irreducible balanced pairs on several examples mainly issued from substitutive systems, and making connections with the balanced pairs algorithm and discrete geometrical constructions.

Introduction

In this article, we are interested in the set of irreducible balanced pairs associated with a language. Irreducible balanced pairs enables the definition of the balanced pairs algorithm (BPA), introduced in [L87] and studied in [SS02, M04].

A substitution σ naturally defines a symbolic dynamical system $(\mathcal{X}_\sigma, S)$, which may be splitted in topological factors. When σ is a d-letter Pisot type substitution, it is possible to construct a geometrical representation of $(\mathcal{X}_\sigma, S)$ known as the Rauzy fractal of the substitution. The Rauzy fractal satisfies various topological properties, notably it is a compact subset of $\mathbb{R}^{d-1}$ which is equal to the adherence of its inner points, and which has positive Lebesgue measure [SW02]. Also, the corresponding substitutive dynamical system $(\mathcal{X}_\sigma, S)$ admit in this case a minimal translation on the torus $\mathbb{T}^{d-1}$ as a topological factor ([CS01]).

When the substitution σ is unimodular, several combinatorial conditions known as coincidence properties provide additional knowledge on $(\mathcal{X}_\sigma, S)$. Under the strong coincidence property, $(\mathcal{X}_\sigma, S)$ is measure-theoretically isomorphic to the exchange of d domains defined almost everywhere on the associated Rauzy fractal ([AI01]). When the super-coincidence property holds, the Rauzy fractal generates a periodic tiling of the space, that is, there exists a lattice Λ such that $(\mathcal{X}_\sigma, S)$ admits a toral representation as a fundamental domain for $\mathbb{R}^d$ / Λ, see [AI01]. Among unimodular Pisot type substitutions, the super-coincidence property is also equivalent to $(\mathcal{X}_\sigma, S)$ having a discrete spectrum. At the moment, we do not know any Pisot type unimodular substitution which does not satisfy either the super-coincidence or the strong coincidence property.

It is proven in the forthcoming study [BBR06] that, for any Pisot type substitution, the super-coincidence holds if and only if the strong coincidence holds and the BPA terminates. Hence the study of irreducible balanced pairs, and of the action of the balanced pair algorithm, is related to the coincidence conjecture and may provide additional knowlegde on associated substitutive dynamical systems.

The article is structured in the following form. In Section 1, we introduce the definitions and notation. In Section 2, we study the set of irreducible balanced pairs for the two-letter full shift, for the Fibonacci substitution, and for a class of substitutions which naturally generalize the Fibonacci case. We prove the following results:

Proposition 1. *The number of irreducible balanced pairs in the two-letter full shift case corresponds to the Motzkin sequence.*

Proposition 2. *For the Fibonacci language, the irreducible balanced pairs are exactly the pairs* $\{(aub, bua), u$ *bispecial factor*$\}$.

Although we are not able to compute the whole set of irreducible balanced pairs for confluent Parry unit languages, we also compute a particular class of irreducible balanced pairs.

In Section 3, we introduce the balanced pair algorithm. We study the action of the balanced pair algorithm on the Fibonacci example:

Proposition 3. *For the Fibonacci case, the balanced pair algorithm associated with any prefix* w *terminates and contains at most 4 elements.*

Then, we establish a connection with the discrete geometrical representation of Rauzy fractals in Section 4, and we prove:

Theorem 1. *For any primitive unimodular irreducible substitution, there exist infinitely many irreducible balanced pairs.*

We end our study by a non-exhaustive list of open questions.

References

[AI01] Arnoux, P. and Ito, S., Pisot substitutions and Rauzy fractals, *Bull. Belg. Math. Soc. Simon Stevin*, **8/2**, 2001, 181–207.

[BBR06] Bernat, J. and Berthé, V. and Rao, H., On super-coincidence condition of substitutions Preprint, 2006.

[CS01] Canterini, V. and Siegel, A. Geometric representation of substitutions of Pisot type, *Trans. Amer. Math. Soc.*, **353/12**, 2001, 5121–5144.

[L87] Livshits, A. N., On the spectra of adic transformations of Markov compact sets, *Uspekhi Mat. Nauk*, **42/3(255)**, 1987, 189–190.

[M04] Martensen, B. F., Generalized balanced pair algorithm, Spring Topology and Dynamical Systems Conference, *Topology Proc.*, **28**, 2004, 163–178.

[SS02] Sirvent, V. F. and Solomyak, B., Pure discrete spectrum for one-dimensional substitution systems of Pisot type, *Canad. Math. Bull.*, **45/4**, 2002, 697–710.

[SW02] Sirvent, V. F. and Wang, Y., Self-affine tiling via substitution dynamical systems and Rauzy fractals, *Pacific J. Math.*, **206/2**, 2002, 465–485.

Factorizations of infinite words

Jean Berstel
Institut Gaspard Monge (IGM)
Université de Marne-la-Vallée
berstel@univ-mlv.fr

Abstract

In this talk, we present old and new results concerning factorizations of infinite words, with special consideration of the Fibonacci word. The older factorizations are the Lyndon factorization, the singular factorization of Wen and Wen; newer factorizations are Ziv-Lempel and Crochemore factorizations.

1 Introduction

M.-P. Schützenberger introduced an approach to the study of combinatorial identities based on natural bijections between combinatorial objects. One simple and rich family of identities concerns Fibonacci numbers: there are several dozens of such identities. Here is one of them (I write F_n for the n-th Fibonacci number, starting with $F_0 = 1, F_1 = 2$ as is done when number systems are considered):

$$F_{n+2} = 2 + F_0 + F_1 + \cdots + F_n \tag{1}$$

(e.g. $21 = 2 + 1 + 2 + 3 + 5 + 8$.) Now consider the (finite) Fibonacci words defined by $f_0 = a$, $f_1 = ab$ and $f_{n+2} = f_{n+1}f_n$.

$$\begin{aligned} f_0 &= a \\ f_1 &= ab \\ f_2 &= aba \\ f_3 &= abaab \\ f_4 &= abaababa \end{aligned}$$

With our numbering, we have $|f_n| = F_n$. In order to prove (1), we look for factorizations of f_{n+2}. Here are three of them:

$$f_{n+2} = abf_0f_1 \cdots f_n$$
$$f_{n+2} = \tilde{f}_0\tilde{f}_1 \cdots \tilde{f}_n(ab)^{(n)}$$
$$f_{n+2} = aw_0w_1w_2 \cdots w_n(b)^{(n)}$$

where $(w)^{(n)} = w$ $(= \overline{w})$ if n is odd(even) and $\overline{w}$ is the opposite of w obtained by exchanging all a's and b's. Here

$$\begin{aligned} w_0 &= b \\ w_1 &= aa \\ w_2 &= bab \\ w_3 &= aabaa \\ &\cdots \end{aligned}$$

are the *singular factors* of the Fibonacci word. The third of these factorization is the *singular factorization* introduced by Wen and Wen ([9], see also [1]).

The singular factorization is closely related to the Lyndon factorization ([7]) of the Fibonacci word, since

$$\begin{aligned} f_8 &= abaababaabaababaababaabaababaabaababaababaabaababaababa \\ &= aw_0w_1w_2 \cdots w_8a \\ &= \ell_1\ell_3\ell_5\ell_7\ell_9a \end{aligned}$$

where

$$\begin{aligned} \ell_1 &= ab = aw_0 \\ \ell_3 &= aabab = w_1w_2 \\ \ell_5 &= aabaababaabab = w_3w_4 \end{aligned}$$

and more generally $\ell_{2n+1} = w_{2n-1}w_{2n}$.

In a seminal paper, Ziv and Lempel [8] defined several factorizations of finite words related to information theory and text processing. Several years later, Crochemore ([2, 4, 3]) introduced a similar factorization of words as a key tool in the design of a linear algorithm checking words for square freeness. The Ziv-Lempel factorization of the Fibonacci word will be shown to be the singular factorization in a completely different context.

Ziv-Lempel and Crochemore factorizations have similar properties. Both can be computed in linear time by preprocessing the suffix tree of the word. Furthermore, the number of factors in both factorizations are closely related: the number of factors of the Crochemore factorization is at most twice the number of factors of the Ziv-Lempel factorization. However, there are examples of factorizations which differ significantly infinitely many times.

In this paper we study the behavior of the Crochemore factorization in the case of some of the most known classes of words, i.e., characteristic Sturmian words, the Thue-Morse word, and the period doubling sequence.

We study here both Ziv-Lempel and Crochemore factorization of special classes of infinite words, such as Sturmian words and some automatic words. It appears that these factorizations can be expressed by a closed formula in many significant examples. The proof of these formulas require some insight in the combinatorial structure of the infinite words considered.

As we shall see, the Crochemore factorization (or c-factorization for short) of special infinite words can be described explicitly, and it reflects the structure of these words.

The c-factorization $c(x)$ of a word x is defined as follows. Each factor of $c(x)$ is either a fresh letter, or it is a maximal factor of x already occurring in the prefix of the word; more formally, the *c-factorization* $c(x)$ of a word x is

$$c(x) = (x_1, x_2, \ldots, x_m, x_{m+1}, \ldots)$$

where x_m is the longest prefix of $x_m x_{m+1} \cdots$ occurring twice in $x_1 x_2 \cdots x_m$, or x_m is a letter a if a does not occur in $x_1 \cdots x_{m-1}$. For example, the c-factorization of $x = ababaab$ is (a, b, aba, ab), since aba occurs twice in $ababa$.

Note that the the c-factorization of a word differs slightly from the well known Ziv-Lempel factorization [8] (or z-factorization), so that these two factorizations are in general not comparable. The *z-factorization* $z(x)$ of a word x is

$$z(x) = (y_1, y_2, \ldots, y_m, y_{m+1}, \ldots)$$

where y_m is the shortest prefix of $y_m y_{m+1} \cdots$ which occurs only once in the word $y_1 y_2 \cdots y_m$.

For example, let x be the word $x = aabaaccbaabaabaa$. The c-factorization and the z-factorization of x are:

$$c(x) = (a, a, b, aa, c, c, baa, baabaa)$$
$$z(x) = (a, ab, aac, cb, aabaab, aa).$$

We shall discuss the relation between these factorizations in detail.

The c-factorization has an interesting behavior in all of the well known infinite words we have considered. The c-factorization of the infinite Fibonacci word f is

$$c(f) = (a, b, a, aba, baaba, \ldots) = (a, b, a, \widetilde{f_2}, \widetilde{f_3}, \ldots)$$

Observe that each of the factors (except the first three) is the reverse the finite Fibonacci word f_n. A similar result holds for characteristic Sturmian words.

The three factorizations of the Fibonacci word can be visualized through the following scheme:

$$\begin{array}{rcccccccccccccc} h: & a & b & a & a & b & a & b & a & a & b & a & a & b & \cdots \\ w: & a & b & a & a & b & a & b & a & a & b & a & a & \cdots & \\ c: & a & b & a & a & b & a & b & a & a & b & a & \cdots & & \end{array}$$

The relation between these factorizations is the following. Factors of h and w satisfy

$$bf_{2i} = w_{2i}a \text{ and } af_{2i+1} = w_{2i+1}b,$$

while factors of w and c satisfy

$$aw_{2i} = \widetilde{f}_{2i}b \text{ and } bw_{2i+1} = \widetilde{f}_{2i+1}a.$$

The c-factorization on Fibonacci word is a particular case of a more general result we obtained for the c-factorization on standard Sturmian words.

Theorem *Let s be the standard Sturmian word defined as the limit of*

$$s_{-1} = b, s_0 = a, \textit{ and } s_n = s_{n-1}^{d_n} s_{n-2},$$

where $d_i > 0$ for each i. Then

$$c(s) = (a, a^{d_1 - 1}, b, a^{d_1} \widetilde{s}_1^{\,d_2 - 1}, \widetilde{s}_2^{\,d_3}, \widetilde{s}_3^{\,d_4}, \ldots, \widetilde{s}_n^{\,d_{n+1}}, \ldots)$$

Similar results hold for other familiar infinite words, such as the Thue-Morse and the period doubling sequence. However we do not yet have a full characterization of the c-factorization of automatic words.

References

[1] J.-P. Allouche and J. Shallit, *Automatic Sequences*, Cambridge University Press, 2003.

[2] M. Crochemore, Recherche linéaire d'un carré dans un mot, *Comptes Rendus Sci. Paris Sér. I Math.*, 1983, **296**, 781–784.

[3] M. Crochemore, C. Hancart, and T. Lecroq, *Algorithmique du texte*, Vuibert, 2001.

[4] M. Crochemore and W. Rytter, *Text Algorithms*, The Clarendon Press Oxford University Press, 1994.

[5] A. de Luca, A division property of the Fibonacci word, *Information Processing Letters*, 1995, **54**, 307–312.

[6] R. Kolpakov and G. Kucherov, Periodic structures on words, in Lothaire, *Applied Combinatorics on Words*, Cambridge University Press, 2005.

[7] G. Melançon, Lyndon words and singular factors of Sturmian words, *Theoret. Comput. Sci.*, 1999, **218**, 41–59.

[8] A. Lempel and J. Ziv, On the complexity of finite sequences, *IEEE Transactions in Information Theory*, 1976, **IT-22**, 75–81.

[9] Z.-X. Wen and Z.-Y. Wen, Some properties of the singular words of the Fibonacci word. *European Journal of Combinatorics* 1994, **15**, 587–598.

Sturmian fixed points of morphisms

V. Berthé

February 9, 2007

Sturmian words are infinite words that have exactly $n + 1$ factors of length n for every positive integer n. A Sturmian word $s_{\alpha,\rho}$ is also defined as a coding over a two-letter alphabet of the orbit of the point ρ under the action of the irrational rotation $R_\alpha : x \mapsto x + \alpha \pmod 1$.

We will focus here on fixed points of Sturmian words by investigating the so-called *Rauzy fractals*. Rauzy fractals (first introduced in the case of the Tribonacci morphism $a \mapsto ab$, $b \mapsto ac$, $c \mapsto a$ [5]) are compact attractors of a graph-directed iterated function system associated with a primitive morphism with prescribed algebraic properties.

Yasutomi characterized in [6] all the pairs (α, ρ) such that the Sturmian word $s_{\alpha,\rho}$ is a fixed point of some non-trivial substitution. By studying the Rauzy fractals associated with Sturmian morphisms, we give here an alternative geometric proof [2] of Yasutomi's characterization; its characterization involves the conjugates of the quadratic real number α and can be compared to Galois' theorem for classical continued fractions describing numbers having a purely periodic continued fraction expansion:

Theorem 1 (Yasutomi [6])**.** *Let* $0 < \alpha < 1$ *and* $0 \le \rho \le 1$. *Then* $s_{\alpha,\rho}$ *is substitution invariant if and only if the following two conditions are satisfied:*

(i) α *is an irrational quadratic number and* $\rho \in \mathbb{Q}(\alpha)$*;*

(ii) $\alpha' > 1,\ 1 - \alpha' \le \rho' \le \alpha'$ *or* $\alpha' < 0,\ \alpha' \le \rho' \le 1 - \alpha'$.

Rauzy fractals associated with Sturmian morphisms are easily seen to be intervals, and a notion of dual morphism can thus be defined provided by the associated iterated function system. This duality acts as a natural involution on Christoffel words: it extends to Sturmian morphisms, and preserves conjugation classes [3]. We will investigate the properties of this involution, from a combinatorial point of view and in arithmetical terms, with respect to the continued fraction algorithm.

More generally, there is a natural involution on Christoffel words, originally studied in [4]. We will show that it has several equivalent definitions: one of them uses the slope of the word, and changes the numerator and the denominator respectively in their inverses modulo the length; another one uses the cyclic graph allowing the construction of the word, by interpreting it in two ways (one as a permutation and its ascents and descents, coded by the two letters of the word, the other in the setting of the Fine and Wilf periodicity theorem); a third one uses central words and generation through iterated palindromic closure, by reversing the directive word. We will show further that this involution coincides with the above mentioned duality acting on Sturmian morphisms. The involution on morphisms is the restriction of some conjugation of the automorphisms of the free group. Finally, we show that, through the geometrical interpretation of substitutions of [1], our involution is the same thing as duality of endomorphisms (modulo some conjugation).

This lecture is based on the results of [2] and [3].

References

[1] P. Arnoux, S. Ito, Pisot substitutions and Rauzy fractals, *Bulletin of the Belgian Mathematical Society Simon Stevin* **8** (2001), 181-207.

[2] V. Berthé, H. Ei, S. Ito, H. Rao, On substitution invariant Sturmian words: an application of Rauzy fractals, *Theoret. Informatics Appl.*, to appear.

[3] V. Berthé, A. de Luca, C. Reutenauer, On an involution of Christoffel words and Sturmian morphisms, preprint 2006.

[4] de Luca, A., Combinatorics of standard Sturmian words, *Lecture Notes in Computer Science* **1261** (1997), 249-267.

[5] G. Rauzy, Nombres algebriques et substitutions, *Bull. Soc. Math. France* **110** (1982), 147-178.

[6] S.-I. Yasutomi, On Sturmian sequences which are invariant under some substitutions, in *Number theory and its applications (Kyoto, 1997)*, pp. 347–373, Kluwer Acad. Publ., Dordrecht, 1999.

Weak mixing of Arnoux-Rauzy sequences (Mélange faible des suites d'Arnoux-Rauzy)

Extended abstract

Julien Cassaigne, Sébastien Ferenczi, and Ali Messaoudi

The first attempt to generalize the famous Sturmian words/rotations interaction led to the definition of the *Arnoux-Rauzy* systems, whose complexity is $2n + 1$ and which satisfy an extra combinatorial condition [2]. One of these systems, the *Tribonacci system*, is indeed a natural coding of a rotation of $\mathbb{T}^2$ ([13], see also [1]), and this gives a good (the best possible in some sense [6]) simultaneous approximation for a pair of algebraic numbers. Every Arnoux-Rauzy system defines an algorithm of simultaneous approximation for some pair of numbers, and was conjectured to be also a natural coding of a rotation of $\mathbb{T}^2$; this conjecture was disproved in [5], by exhibiting an Arnoux-Rauzy system whose language is unbalanced, see Corollary below. This example is quite elaborate, and leaves many open questions, which are asked at the end of [5]; the system considered there could still have some weaker properties than being a natural coding: equipped with its unique invariant probability measure, it could still be measure-theoretically isomorphic to a rotation of $\mathbb{T}^2$ or at least admit a rotation of $\mathbb{T}^2$ as a measure-theoretic factor.

In the present paper, we prove a stronger result for a much larger class of systems, by proving that any Arnoux-Rauzy system for which the inverses of the partial quotients (defined in Definition 1 below and linked to the approximation algorithm) form a convergent series is (measure-theoretically) *weakly mixing*, see Definition 2 below. As a consequence, all these systems (which include the example in [5]) have an unbalanced language and do not satisfy any of the weaker properties mentioned above.

To show that our condition on the partial quotients is optimal, we give a completely new example of Arnoux-Rauzy system, with partial quotients equal to $2n + 1$, having rotations as continuous factors.

1 Arnoux-Rauzy systems

We take here as a definition of Arnoux-Rauzy systems their constructive characterization, derived in [2] from the original definition.

Definition 1. *An* Arnoux-Rauzy system *is a symbolic system defined by three families of words A_k, B_k, C_k, build recursively from $A_0 = 1$, $B_0 = 12$, $C_0 = 13$, by using a sequence of combinatorial rules a, b, c, such that each one of the three rules is used infinitely many times, where*

- *by rule a, $A_{k+1} = A_k$, $B_{k+1} = A_k B_k$, $C_{k+1} = A_k C_k$;*
- *by rule b, $A_{k+1} = B_k A_k$, $B_{k+1} = B_k$, $C_{k+1} = B_k C_k$;*
- *by rule c, $A_{k+1} = C_k A_k$, $B_{k+1} = C_k B_k$, $C_{k+1} = C_k$.*

The system (X, T) is the one-sided shift on sequences $(x_n, n \in \mathbb{N})$ such that for each $0 \le s \le t$ there exists k such that $x_s...x_t$ is a subword of A_k.
(X, T) is minimal [2] and uniquely ergodic (by Boshernitzan's result [4] using the fact that the complexity is $2n + 1$) with a unique invariant probability measure μ.
We consider the measure-theoretic system (X, T, μ).
If the sequence of rules a, b, c is r_1 iterated k_1 times, ... r_i iterated k_i times, ..., $r_{i+1} \neq r_i$, $k_i \ge 1$, the k_i are the partial quotients *of the system.*

We can then define our system in another way, which corresponds to a *multiplicative* form of the approximation algorithm.

Lemma 1. *Let an Arnoux-Rauzy system (X, T, μ) be defined as in Definition 1, and k_n be its partial quotients.*
Then (X, T) is the one-sided shift on sequences $(x_n, n \in \mathbb{N})$ such that for each $0 \le s \le t$ there exists n such that $x_s...x_t$ is a subword of H_n, where the three words H_n, G_n, J_n are built from H_0, G_0, J_0 by two families of rules:

- *by a rule of type 1, $H_{n+1} = H_n^{k_{n+1}} G_n$, $G_{n+1} = H_n^{k_{n+1}} J_n$, $J_{n+1} = H_n$;*
- *by a rule of type 2, $H_{n+1} = H_n^{k_{n+1}} G_n$, $G_{n+1} = H_n$, $J_{n+1} = H_n^{k_{n+1}} J_n$.*

And rules of type 1 are used infinitely many times.

2 Finite rank and eigenvalues

Definition 2. *Let (X, T, μ) be a finite measure-preserving dynamical system.*

A real number $0 \leq \theta < 1$ is an eigenvalue *of T (denoted additively) if there exists a non-constant f in $\mathcal{L}^1(X, \mathbb{R}/\mathbb{Z})$ such that $f \circ T = f + \theta$ (in $\mathcal{L}^1(X, \mathbb{R}/\mathbb{Z})$)); f is then an* eigenfunction *for the eigenvalue θ. As constants are not eigenfunctions, $\theta = 0$ is not an eigenvalue if T is ergodic. T is* weakly mixing *if it has no eigenvalue.*

The following propositions give a necessary and a sufficient condition for a number to be an eigenvalue of an Arnoux-Rauzy system.

Proposition 2. *If θ is an eigenvalue for an Arnoux-Rauzy system as described above; then $k_{n+1}||h_n\theta|| \to 0$ when $k \to +\infty$, where $|| \quad ||$ denotes the distance to the nearest integer.*

Proposition 3. *If, for an Arnoux-Rauzy system as described above, $\sum_{n=0}^{+\infty} k_{n+1}||h_n\theta|| < +\infty$, then θ is an eigenvalue of the system, with a continous eigenfunction.*

3 Weak mixing

We are now ready to prove the weak mixing under an additional hypothesis: henceforth we consider an Arnoux-Rauzy system as defined in Definition 1, and the associated quantities defined in Lemma 1 and section 2 above; we assume that

$$\sum_{n=1}^{+\infty} \frac{1}{k_n} < +\infty. \tag{1}$$

Theorem 4. *Under condition (1), (X, T, μ) is weakly mixing.*

4 Eigenvalues

Theorem 5. *Let (X, T, μ) be the Arnoux-Rauzy system such that all rules are of typoe 1 and $k_n = 2n + 1$ for all n; then it has eigenvalues, corresponding to continous eigenfunctions.*

References

[1] P. ARNOUX: Un exemple de semi–conjugaison entre un échange d'intervalles et une translation sur le tore, (in French) *Bull. Soc. Math. France* 116 (1988), p. 489–500.

[2] P. ARNOUX, G. RAUZY: Représentation géométrique de suites de complexité $2n + 1$, *Bull. Soc. Math. France* 119 (1991), p. 199–215.

[3] V. BERTHÉ, S. FERENCZI, L.Q. ZAMBONI: Interactions between dynamics, arithmetics and combinatorics: the good, the bad, and the ugly, *AMS Contemporary Math.* 385 (2005), p. 333–364.

[4] M. BOSHERNITZAN: A unique ergodicity of minimal symbolic flows with linear block growth, *J. Analyse Math.* 44 (1984/85), p. 77–96.

[5] J. CASSAIGNE, S. FERENCZI, L. Q. ZAMBONI: Imbalances in Arnoux-Rauzy sequences, *Ann. Inst. Fourier (Grenoble)* 50 (2000), p. 1265–1276.

[6] N. CHEKHOVA, P. HUBERT, A. MESSAOUDI: Propriétés combinatoires, ergodiques et arithmétiques de la substitution de Tribonacci, *J. Théor. Nombres Bordeaux* 13 (2001), p. 371–394.

[7] J.R. CHOKSI, M.G. NADKARNI: The group of eigenvalues of a rank–one transformation, *Canad. Math. Bull.* 38 (1995), p. 42–54.

[8] E.M. COVEN, G.A. HEDLUND: Sequences with minimal block growth, *Math. Systems Theory* 7 (1972), p. 138–153.

[9] S. FERENCZI: Systems of finite rank, *Colloq. Math.* 73 (1997), p. 35–65.

[10] S. FERENCZI, L. ZAMBONI: Examples of 4-interval exchange transformations, submitted, http://iml.univ-mrs.fr/~ferenczi/fz2.pdf.

[11] M. MORSE, G.A. HEDLUND: Symbolic dynamics II. Sturmian trajectories, *Amer. J. Math.* 62 (1940), p. 1–42.

[12] I. OREN: Admissible functions with multiple discontinuities, *Israel J. Math.* 42 (1982), p. 353–360.

[13] G. RAUZY: Nombres algébriques et substitutions, *Bull. Soc. Math. France* 110 (1982), p. 147–178.

Palindromes in Sturmian and episturmian words

Aldo de Luca
Dipartimento di Matematica e Applicazioni dell'Università di Napoli "Federico II"
Via Cintia, Monte S. Angelo, 80126 Napoli, Italy
aldo.deluca@unina.it

As is well known palindromes play a basic role in the theory of Sturmian and episturmian sequences. A theorem due to Droubay and Pirillo [10] allows one to characterize Sturmian sequences as those infinite sequences having exactly *one* palindromic factor of length n for n *even* and *two* palindromic factors for n *odd*. Moreover, both standard Sturmian and episturmian sequences can be constructed by a suitable iteration of the operator of *right palindromic closure* [5],[11].

The palindromic prefixes of all standard Sturmian sequences coincide with *central words*, where a word w is central if it has two coprime periods p and q and its length is $p + q - 2$ [7]. Central words are of fundamental importance in Sturmian words theory since they satisfy several important combinatorial and structural properties. Let us recall some of them:

1. Any finite factor of a Sturmian sequence is a factor of a central word. In particular, any palindrome is a *median* factor of a central word. In this way one can generate all Sturmian palindromes from central words, by deleting their prefixes and suffixes having the same length. From this one obtains a simple formula counting for any n Sturmian palindromes of length n [8].

2. Central words are strictly related to *finite standard words* and to *Christoffel words*. In particular, one has that if w is a central word, then wab and wba are standard words, so that wab and wba are both *symmetric*, i.e., product of two palindromes [7]. Conversely, any finite standard word, different from a single letter, can be written as wxy, with $\{x, y\} = \{a, b\}$ and w central. Moreover, if w is central, then awb is a Christoffel word which is a Sturmian *Lyndon word* [2]. Any Christoffel word, different from a single letter, can be written in the previous way.

3. Central words can be generated by iteration of the operator $^{(+)}$ of *right palindromic closure* associating to each word $w \in \{a, b\}^*$ the word $w^{(+)}$ defined as the shortest palindrome having w as a prefix. More precisely, one introduces a map $\psi : \{a, b\}^* \to \{a, b\}^*$ defined as: $\psi(\varepsilon) = \varepsilon$ and for any word v and

letter x, $\psi(vx) = (\psi(v)x)^{(+)}$. It has been proved that ψ is injective and that $\psi(\{a,b\}^*)$ is the set of central words [5]. Moreover, if one starts with an infinite word $x = x_1x_2\cdots x_n\cdots$ (*directive word*), where the letters belong to $\{a,b\}$ and there are infinitely many occurrences of both the letters, then one obtains an infinite word $\psi(x) = \lim_{n\to\infty}\psi(x_1\cdots x_n)$ which is a standard Sturmian sequence. Conversely, any standard Sturmian sequence can be obtained in this way.

4. There exists an important involutory map which associates to each central word w another central word w^* [6] . This map can be described by continued fractions or in terms of iterated palindrome closure. Indeed, if $w = \psi(v)$, then $w^* = \psi(\tilde{v})$. If p/q and p^*/q^* are respectively the *slopes* of the two Christoffel words awb and aw^*b, then one has: $p+q = p^*+q^*$, $pp^* \equiv qq^* \equiv 1 \bmod (p+q)$, and p,q (resp. p^*,q^*) are two coprime periods of w^* (resp. w). The previous involution can be suitably extended to *special Sturmian morphisms* in the sense that it preserves conjugacy classes of these morphisms, which are in bijection with central words [12]. A variant of the above involution on central words was considered in [3, 4]. The fix points of this latter involution are the so-called *harmonic words*.

The use of iterated palindromic closure in the case of directive words over alphabets with more than two letters, gives rise to the class of standard *episturmian sequences* [11] including both standard Sturmian sequences and the standard Arnoux-Rauzy sequences [1].

Thus the operators $^{(+)}$ (resp. $^{(-)}$) of right (resp. left) palindromic closure and the iterated closure operators play an essential role in the combinatorics of words and of Sturmian and episturmian sequences. Denoting for a word w over any alphabet by π_w the *minimal period* of w, one can prove that $\pi_{w^{(+)}} = \pi_{w^{(-)}}$. Moreover, $\pi_w = \pi_{w^{(+)}}$ if and only if the *fractional root* of w is symmetric. Some important applications to Sturmian words are the following [9]:

a. If w is a finite Sturmian word, then $\pi_w = \pi_{w^{(+)}}$. Moreover, there exists a standard Sturmian sequence s such that both $\pi_{w^{(+)}}$ and $\pi_{w^{(-)}}$ are factors of s.

b. For a non empty word the following conditions are equivalent i.) w is a prefix of a standard Sturmian sequence, ii.) $w^{(+)}$ is central, iii.) the fractional root of w is a standard word.

c. A word w is a finite Sturmian word if and only if $\pi_{w^{(+)}} = R_{w^{(+)}} + 1$, where for any word w, R_w denotes the minimal integer k such that w has no right special factor of length k.

In the case of standard episturmian words one has that the fractional roots of their nonempty prefixes are symmetric. However, differently from the Sturmian case, the fractional root of a finite episturmian word w can be non-symmetric, so that in general $\pi_w \neq \pi_{w^{(+)}}$. It has been proved [9] that if w is a finite episturmian word, then either $w^{(+)}$ or $w^{(-)}$ are episturmian. One can ask the question whether, similarly to the Sturmian case, both $w^{(+)}$ and $w^{(-)}$ are episturmian. A positive answer to this question was very recently given by Zamboni (cf.[13]).

An extension of the theory can be obtained by considering θ-palindromes, i.e., the fix points of an involutory antimorphism θ of a free monoid, and define accordingly the right (resp. left) θ-palindromic closure of a word w as the shortest θ-palindrome having w as a prefix (resp. suffix). Starting with a directive word, one can iterate the application of θ-palindrome operator generating an infinite word called *θ-standard sequence*. In [9] it has been proved that any θ-standard sequence is a morphic image, by an injective morphism, of the standard episturmian sequence having the same directive word. Some further extensions and generalizations have been considered in [9].

References

[1] P. Arnoux, G. Rauzy, Représentation géometrique des suites de complexité $2n + 1$, Bull. Soc. Math. France **119** (1991) 199-215

[2] J. Berstel, A. de Luca, Sturmian words, Lyndon words, and trees, Theoretical Computer Science **178** (1997) 171-203

[3] A. Carpi, A. de Luca, Harmonic and gold Sturmian words, European Journal of Combinatorics **25** (2004) 685-705

[4] A. Carpi, A. de Luca, Farey codes and languages, European Journal of Combinatorics, to appear

[5] A. de Luca, Sturmian words: Structure, combinatorics, and their arithmetics, Theoretical Computer Science **183** (1997) 45-82

[6] A. de Luca, Combinatorics of standard Sturmian words, Lecture Notes in Computer Science, vol.1261, Springer (Berlin, 1997) pp.249–267

[7] A. de Luca, F. Mignosi, Some combinatorial properties of Sturmian words, Theoretical Computer Science **136**(1994) 361-385

[8] A. de Luca, A. De Luca, Combinatorial properties of Sturmian palindromes, International Journal of Foundations of Computer Science **17**(2006) 557-574

[9] A. de Luca, A. De Luca, Pseudopalindrome closure operators in free monoids, Theoretical Computer Science, to appear

[10] X. Droubay, G. Pirillo, Palindromes and Sturmian words, Theoretical Computer Science, **223** (1999) 73–85

[11] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, Theoretical Computer Science **255** (2001) 539-553

[12] V. Berthé, A. de Luca, C. Reutenauer, On an involution of Christoffel words and Sturmian morphisms, preprint 2006

[13] A. de Luca, A. De Luca, L.Q. Zamboni, On some problems related to palindrome closure, manuscript 2006

The Rat Game and the Mouse Game

Aviezri Fraenkel
Computer Science and Applied Mathematics
Weizmann Institute of Science
Rehovot, Israel
aviezri.fraenkel@weizmann.ac.il

Extended Abstract

The Rat game is played on 3 piles of tokens by 2 players who play alternately. Positions in the game are denoted throughout in the form (x, y, z), with $0 \le x \le y \le z$, and moves in the form $(x, y, z) \to (u, v, w)$, where of course also $0 \le u \le v \le w$. The player first unable to move — because the position is $(0, 0, 0)$ — loses; the opponent wins. There are 3 types of moves:

(I) Take any positive number of tokens from up to 2 piles.

(II) Take $\ell > 0$ from the x pile, $k > 0$ from the y pile, and an arbitrary positive number from the z pile, subject to the constraint $|k - \ell| < a$, where

$$a = \begin{cases} 1 & \text{if } y - x \not\equiv 0 \pmod 7 \\ 2 & \text{if } y - x \equiv 0 \pmod 7). \end{cases}$$

(III) Take $\ell > 0$ from the x pile, $k > 0$ from the z pile, and an arbitrary positive number from the y pile, subject to the constraint $|k - \ell| < b$, where $b = 3$ if $w = u$; otherwise,

$$b = \begin{cases} 5 & \text{if } w - u \not\equiv 4 \pmod 7 \\ 6 & \text{if } w - u \equiv 4 \pmod 7). \end{cases}$$

In a move of type (II) we permit the permutation $x \to v$, $y \to w$, $z \to u$, in addition to $x \to u$, $y \to v$, $z \to w$. No other permutations are allowed for (II), and none (except $x \to u$, $y \to v$, $z \to w$) for (III). For (I), of course, any rearrangement is possible. When we write $(x, y, z) \to (u, v, w)$, we always mean $x \to u, y \to v, z \to w$.

Note that in (II), the congruence conditions depend only on 2 of the piles moved *from*: the smallest and the intermediate; whereas in (III) they depend only on 2 of the piles moved *to*: the smallest and the largest. The case $w = u$ in (III) is an initial condition, to accommodate the end position $(0, 0, 0)$.

Let $S \subsetneqq \mathbb{Z}_{\geq 1}$.Define the mex_1 operator by $\text{mex}_1(S) = \min(\mathbb{Z}_{\geq 1} \setminus S)$ =the smallest positive integer not in S. The following is a recursive characterization of the P-positions (second player winning positions).

Theorem 1. *The P-positions are given by $(0, 0, 0)$, and for $n \geq 1$, $A_n = \text{mex}_1\{A_i, B_i, C_i : 1 \leq i < n\}$, $B_n = A_n + \lfloor (7n-2)/4 \rfloor$, $C_n = B_n + \lfloor (7n-3)/2 \rfloor$.*

The first few P-positions are displayed in Table 1.

Table 1: The first few P-positions of the Rat game.

n	A_n	B_n	C_n
0	0	0	0
1	1	2	4
2	3	6	11
3	5	9	18
4	7	13	25
5	8	16	32
6	10	20	39
7	12	23	46
8	14	27	53
9	15	30	60
10	17	34	67
11	19	37	74
12	21	41	81
13	22	44	88
14	24	48	95
15	26	51	102

This recursive definition of the P-positions, alas, provides only an exponential-time strategy for winning the game. However, we will provide an explicit (closed formula) winning strategy, which turns out to be polynomial as well.

What's the connection of this game to rats? And where is the mouse game? Are rats and mice connected to the topic of the workshop? Answers—in Turku.

Fibonacci-Padovan Sequences and MacWilliams Transform Matrices

N. Gogin and A. Mylläri
University of Turku, Turku, Finland

Somehow unexpected connection between MacWilliams transform matrices and classic integer Fibonacci, Lucas and Padovan sequences is established, namely, it is proved that summing along some naturally selected planes in the "pyramid" constructed from these matrices leads to a new integer sequence that appears to be a convolution of Fibonacci numbers and (alternating in sign) Padovan numbers. This convolution in turn can be linearly expanded in Lucas and Padovan numbers.

Fibonacci number and words in tilings of the hyperbolic plane

Maurice Margenstern
Laboratoire d'Informatique Théorique et Appliquée, EA 3097,
Université de Metz, I.U.T. de Metz,
Département d'Informatique,
Île du Saulcy,
57045 Metz Cedex, France,
margens@univ-metz.fr

Abstract

There are several connections of terms of the Fibonacci sequence and of Fibonacci words with tilings of the hyperbolic plane. In this paper, investigate some new properties in this direction.

Keywords: hyperbolic plane, tessellations, Fibonacci sequences, Fibonacci words.

1 Introduction

As indicated in the abstract, several connections can be indicated between the Fibonacci sequence and hyperbolic geometry. As far as I know, [2] is the main indication of a property which received not much attention before the series of papers inaugurated by [3].

In our first section, we report the properties which were established in [3] and which concern the **pentagrid**, *i.e.* the tiling $\{5, 4\}$ of the hyperbolic plane. It is generated by the recursive reflection of a fixed regular rectangular pentagon in its sides and of the images in their sides. We also report the extension of this property to the **ternary hepagrid**, see [1] which is the tiling $\{7, 3\}$ of the hyperbolic plane: it is generated as the pentagrid by this time starting from a regular heptagon with a vertex angle of $\frac{2\pi}{7}$.

In the second section, we go back to balls in the hyperbolic plane in both the pentagrid and the ternary heptagrid. We mention a tiling property of the balls. We also indicate how the splitting properties used in the first section allow to find out an equality on terms of the Fibonacci sequence. We look at the border of the balls and derive a result on the language of the words constituted by such borders.

In the third section, we introduce the notion of Fibonacci **carpet**. We derive from this a continuous bijection between the pentagrid and the ternary heptagrid. We also introduce the notion of **levels** and connect them with fixed points of a Fibonacci substitution on words.

We conclude with possible extensions of the results to other families of words attached to other algebraic sequences of numbers.

2 The Fibonacci tree and the tilings $\{5, 4\}$ and $\{7, 3\}$

As proved in [7], a spanning tree can be associated to the restriction of the tiling $\{5, 4\}$ of the hyperbolic plane to a quarter in such a way that the right angle defining the quarter should be a vertex angle of a tile. As indicated in that paper, two kinds of nodes are defined for such a tree: black and white nodes. By definition, a black node has two sons, a black one and a white one; a white node has three sons: a black one and two white ones. Counter-clockwise orienting the plane, the black son is always the first son of a node. The root of this tree, which we call the Fibonacci tree, is a white node.

In [3], a basic property was noticed. Number the nodes of the Fibonacci tree level after level, starting from the root which is numbered by 1 and, on each level, from the left to the right. Next, consider the representation of these numbers as sums of distinct terms of the Fibonacci sequence, when consecutive 1's are ruled out. This uniquely fixes the representation. Then, we have the following property:

Theorem 1. *Among the sons of a node, exactly one of them ends in* **00**. *More precisely, if α is the representation of the number attached to the node, α***00** *is the representation of this son which is called the* **preferred son**. *Moreover, in a black node, the preferred son is the black one and, in a white node, it is the first white node, counter-clockwise counting the nodes.*

This property makes it easy to implement cellular automata in the pentagrid and in the ternary heptagrid. Considering the above representation as **coordinates** for the nodes, the coordinate of the neighbours of a cell are easily computed in the coordinate of the cell.

As indicated in [1], the pentagrid can be represented by a central tile surrounded by five quarters, each one being spanned by a Fibonacci tree as just described. As also indicated in the same paper, a similar representation holds for the ternary heptagrid, this time with seven **angular sectors** spanned by a Fibonacci tree. We cannot better explain the structure of the sectors within the frame of this short note. This simple splitting property explains the result of [2] on the number of tiles on the border of a ball in $\{5, 4\}$ and $\{7, 3\}$ which is $5f_{2n+1}$ and $7f_{2n+1}$i, respectively, for balls of radius n. As proved in [7], the number of tiles of the level n of the Fibonacci tree is f_{2n+1}.

3 Balls in the tilings $\{5, 4\}$ and $\{7, 3\}$ and their borders

Remind that a ball $B_n(T_0)$ around T_0 of radius n in the pentagrid or the ternary heptagrid is the set of tiles T which are at distance n from T_0. This means that there is a path constituted of adjacent tiles going from T_0 to T, that the number of tiles on the path is n, T_0 not being taken into account, and that there is no shorter path joining T_0 to T. We say ball, for short, denoted by B_n, when we do not specify the centre of the ball.

The balls have the following surprising properties:

Theorem 2. *Let n be any natural number. Then both in the pentagrid and the ternary heptagrid, the balls of radius n tile the hyperbolic plane.*

We have a stronger property, from which theorem 2 is an easy corollary:

Theorem 3. *Let $\mathcal{I}$ be an increasing sequence of positive numbers. Then both in the pentagrid and the ternary heptagrid we have the following. It is possible to tile the hyperbolic plane by taking balls B_m with $m \in \mathcal{I}$ and in such a way that for each $n \in \mathcal{I}$, the ball B_n occurs infinitely many often in the tiling.*

As explained in [7, 3], the spanning tree can be generated by a splitting of the quarter of the hyperbolic plane using the basic tile (the pentagon or the heptagon) and two basic regions: the quarter and the strip which is a difference between to quarters, see [3].

Denote by Q_n the trace of a ball of radius n around T in a quarter rooted at T. Similarly, denote by R_n the set of tiles which is obtained by applying the rules of a Fibonacci tree when the root is a black node. It is not difficult to see that in R_n, the number of tiles of R_n at distance n from T is f_{2n}. Then, considering the elementary splitting property established in [7, 3] applied to the difference between B_{n+m} and B_n, we obtain:

Theorem 4. *We can split Q_{n+m} into Q_n, f_{2n+2} copies of Q_{m-1} and f_{2n+1} copies of R_{m-1}. Similarly, R_{n+m} can be split into R_n, f_{2n+1} copies of Q_{m-1} and f_{2n} copies of R_{m-1}.*

From this, we easily again obtain the identities:

$$f_{2n+2m+1} = f_{2n+2}f_{2m-1} + f_{2n+1}f_{2m-2},$$
$$f_{2n+2m} = f_{2n+1}f_{2m-1} + f_{2n}f_{2m-2},$$

In a similar way, we obtain the identities:

$$f_{2n+2m+1} = f_{2n}f_{2m+1} + f_{2n-1}f_{2m},$$
$$f_{2n+2m} = f_{2n-1}f_{2m+1} + f_{2n-2}f_{2m},$$

Now, if we consider the tiles which are at distance n of T in a ball around T of radius n and we colour them in black and white depending on whether there are black or white nodes, this defines a word which we call the **border word** of the ball of radius n (we ignore the centre) and we denote it by w_n. Then:

Theorem 5. *The set of border words w_n, considered either in the pentagrid or in the heptagrid, is a language which is neither regular nor context free.*

For this result, we can take an argument which is similar to that of the proof of the non context-freeness of a language derived in the case of the tiling $\{5, 3, 4\}$ of the hyperbolic $3D$ space, see [8].

4 Fibonacci carpets and infinite words

In [6], a new ingredient is brought in to the notion of Fibonacci tree: the notion of **Fibonacci carpet**. We can define it as the limit of a sequence $\mathcal{S}$ of Fibonacci trees where the root of a tree is the first white son of the root of the previous tree and where $\mathcal{S}$ has no initial tree in the order induced by this definition. The notion holds both for the pentagrid and the ternary heptagrid. Coordinates can be defined involving $\mathcal{S}$ and the coordinates defined in each tree belonging to $\mathcal{S}$. It is worth noticing that the new construction and its connected coordinates induce an identity mapping which is bi-continuous.

Moreover, in this construction, it is possible to define the notion of **levels**. Now, the levels have an interesting connection with Fibonacci words: again consider the tiles as black or white depending on whether they are black or white nodes in any tree of the sequence which contain them. They constitute an infinite word ω. Then, as the substitution can be realized as a transformation of any level into the next one, we get:

Theorem 6. *All the levels define the same word ω which is a fixed point of the substitution defining the Fibonacci trees applied to bi-infinite words.*

5 Conclusion

We conclude with the indication that there are certainly a lot of other properties in the directions just outlined by the previous sections. One of these directions is the application of the same geometrical considerations connected with other tilings of the hyperbolic plane or of the hyperbolic $3D$ space, and other algebaric numbers. I have already a few papers in this direction, namely [8, 4]. The reader may find other references in [5].

References

[1] K. Chelghoum, M. Margenstern, B. Martin, I. Pecci, Cellular automata in the hyperbolic plane: proposal for a new environment, *Lecture Notes in Computer Sciences*, **3305**, (2004), 678-687.

[2] Harborth, H.: Concentric cycles in mosaic graphs, **3**, 1990, 123–128.

[3] M. Margenstern, New Tools for Cellular Automata of the Hyperbolic Plane, *Journal of Universal Computer Science* **6**N°12, 1226–1252, (2000)

[4] Margenstern M., The Tiling of the Hyperbolic $4D$ Space by the 120-cell is Combinatoric, *Journal of Universal Computer Science*, **10**, No9 1212-1238, (2004).

[5] M. Margenstern, Cellular Automata and Combinatoric Tilings in Hyperbolic Spaces, a survey, invited talk to **DMTCS'03**, July 7-12, Dijon, France, Lecture Notes in Computer Sciences, **2731**, (2003), 48-72.

[6] M. Margenstern, A new way to implement cellular automata on the penta- and heptagrids, Journal of Cellular Automata, **1**, N°1, (2006), 1-24.

[7] M. Margenstern, K. Morita, NP problems are tractable in the space of cellular automata in the hyperbolic plane, *Theoretical Computer Science*, **259**, 99–128, (2001)

[8] M. Margenstern, G. Skordev, Tools for devising cellular automata in the hyperbolic 3D space, *Fundamenta Informaticae*, **58**, N°2, (2003), 369-398.

Word Equations, Fibonacci Numbers, and Hilbert's Tenth Problem

Yuri Matiyasevich
Steklov Institute of Mathematics at Saint-Petersburg
27 Fontanka, Saint-Petersburg, 191023, Russia
http://logic.pdmi.ras.ru/ yumat

In his tenth problem D. Hilbert [1] asked for an algorithm for deciding whether an arbitrary Diophantine equation has a solution in integers. As a possible way to establish the undecidability of Hilbert's tenth problem A.A.Markov suggested to prove the undecidability of *word equations*. Any such equation (w.l.o.g, in the two-letter alphabet $B = \{0, 1\}$) can be easily reduced to a Diophantine equation using the fact that every 2×2 matrix with natural number elements and the determinant equal to zero can be represented in a unique way as the product of matrices $B_0 = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ and $B_1 = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. This allows us to treat a quadruple of non-negative integers $\langle a_{11}, a_{12}, a_{21}, a_{22}\rangle$ such that $a_{11}a_{22} - a_{21}a_{12} = 1$ as a word in the alphabet B, the concatenation being just matrix multiplication.

This approach to Hilbert's tenth problem turned out to be fruitless: G. S. Makanin [2] found a decision procedure for word equations. Nevertheless, we could try to revive Markov's idea by considering a wider class of word equations.

Every word $X = \chi_n\chi_{n-1} \dots \chi_1$ in the alphabet B can be viewed as the number

$$x = \chi_n u_n + \chi_{n-1}u_{n-1} + \cdots + \chi_1 u_1 \tag{1}$$

written in positional system with weights of digits being the Fibonacci numbers $u_1 = 1$, $u_2 = 1$, $u_3 = 2$, $u_4 = 3$, $u_5 = 5$, …(rather than traditional 1, 2, 4, 8, 16, …). According to Zeckendorf's theorem, every natural number x can be represented in the form (4) with additional restrictions $\chi_{i+1}\chi_i = 0$ and $\chi_1 = 0$, and in unique way.

Every word $\alpha_{i_n}\alpha_{i_{n-1}} \dots \alpha_{i_1}$ in the infinite alphabet $A = \{\alpha_1, \alpha_2, \dots\}$ can be presented as a word in the alphabet B

$$10^{i_m}10^{i_{m-1}} \dots 10^{i_1} \tag{2}$$

satisfying the restrictions of Zeckendorf's theorem. Thus we get, via Fibonacci numbers, a natural one-to-one correspondence between words in the infinite alphabet A and natural numbers, the kth word under this enumeration will be denoted Z_k.

Now in order to be able to transform an arbitrary word equation into an equivalent Diophantine equation we need only to express the *concatenation relation* $Z_k = Z_{k_1} Z_{k_2}$ by Diophantine equation(s). For this goal it is more convenient to code a word X by a quadruple $\langle v, w, x, y \rangle$ where x is as in (1), and

$$v = u_n, \quad w = u_{n-1}, \quad y = \chi_n u_{n-1} + \chi_{n-1} u_{n-2} + \cdots + \chi_1 u_0. \tag{3}$$

Thanks to the equality $\phi^m = \phi u_m + u_{m-1}$ where ϕ is the golden ratio $\frac{1+\sqrt{5}}{2}$, the number

$$\phi x + y = \chi_n \phi^n + \chi_{n-1} \phi^{n-1} + \cdots + \chi_1 \phi \tag{4}$$

is represented by the word X in the positional system with weights of digits being powers of the golden ratio 1, ϕ, ϕ^2, ϕ^3, ... Now the relation $Z_{\langle v,w,x,y \rangle} = Z_{\langle v_1,w_1,x_1,y_1 \rangle} Z_{\langle v_2,w_2,x_2,y_2 \rangle}$ is just the equality

$$\phi x + y = (\phi x_1 + y_1)(\phi v_2 + w_2) + \phi x_2 + y_2. \tag{5}$$

It still remains to distinguish those quadruples $\langle v, w, x, y \rangle$ which are codes of words. It is not difficult to check that numbers defined by (1) and (3) satisfy the following conditions:

$$(w^2 + vw - v^2)^2 = 1, \tag{6}$$

$$v \leq x < v + w, \tag{7}$$

$$\phi - 2 < y - x/\phi < \phi - 1. \tag{8}$$

On the other hand, according to an old theorem of Wasteels [5] solutions of the equation (6) are exactly consecutive Fibonacci numbers. Thus for a fixed value of x, conditions (6)–(8) uniquely determine the values of v, w, and y.

Conditions (5)–(8) can be easily transformed into desired Diophantine equations.

Now we are capable to transform into Diophantine equations not only word equations but a broader class of conditions on words. In particular, condition $v_1 = v_2$ expresses the equality of the lengths of words $Z_{\langle v_1,w_1,x_1,y_1 \rangle}$ and $Z_{\langle v_2,w_2,x_2,y_2 \rangle}$.

OPEN QUESTION. *Is there an algorithm for deciding whether an arbitrary system of word equations and equalities of length of words has a solution?*

As we have seen, the undecidability of this problem would give a proof of the undecidability of Hilbert's tenth problem very different from known today (see, for example, [4]).

In the above arithmetization of words we had the restriction that a word cannot contain "11" and cannot end by "1". This drawback can be easily eliminated by the following modifications. Having an arbitrary word in the alphabet B, we at first transform it into a "restricted" word by replacing each "1" by"10" and each "0" by "00"; formally, representation (1) is replaced by

$$x = \chi_n u_{2n} + \chi_{n-1} u_{2(n-1)} + \cdots + \chi_1 u_2. \tag{9}$$

Condition (6) is strengthened to

$$w^2 + vw - v^2 = 1 \tag{10}$$

to imply that $v = u_m$ with even m and hence the word $Z_{\langle v,w,x,y\rangle}$ has even length. In order to pass back from numbers to words we cut $Z_{\langle v,w,x,y\rangle}$ into blocks of length 2 and then replace "00" by "0" while both "10" and "01" are replaced by "1". Now instead of one-to-one correspondence between words and numbers we have one-to-many correspondence but this isn't an obstacle for transforming systems of word equations and length equalities into equivalent Diophantine equations.

Besides equalities of lengths of words, we cam impose conditions in the form "the length of the word $Z_{\langle v_1,w_1,x_1,y_1\rangle}$ is divisible by the length of the word $Z_{\langle v_2,w_2,x_2,y_2\rangle}$" (see [3]).

References

[1] David Hilbert, Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker Kongress zu Paris 1900, *Nachr. K. Ges. Wiss., Göttingen, Math.-Phys. Kl.* (1900), 253–297. See also David Hilbert, *Gesammelte Abhandlungen*, Berlin : Springer, vol. 3 (1935), 310 (Reprinted: New York : Chelsea (1965)).

[2] G. S. Makanin. The problem of solvability of equations in a free semigroup (in Russian). *Math. USSR Sbornik*, 32(2):129–198, 1977.

[3] Matiyasevich Yu. Svyaz' sistem uravneniĭ v slovakh i dlinakh s 10-ĭ problemoĭ Gil'berta, *Zap. nauch. seminar. Leningr. otd. Mat. in-ta AN SSSR*, 8:132–144, 1968. English translation: The connection between Hilbert's Tenth Problem and systems of equations between words and lengths. *Seminars in Mathematics, V.A.Steklov Mathematical Institute*, 8:61–67, 1970, available via `http://logic.pdmi.ras.ru/ yumat/Journal`.

[4] Matiyasevich Yu. *Desyataya Problema Gilberta* (in Russian). Moscow, Fizmatlit, 1993. English translation: Hilbert's tenth problem. MIT Press, 1993. French translation: Le dixième problème de Hilbert, Masson, 1995. Homepage of the book: `http://logic.pdmi.ras.ru/ yumat/H10book`.

[5] M. J. Wasteels. Quelques propriété des nombres de Fibonacci. *Mathesis*, troisième série, tome II, 60-62, 1902.

Spelling Fibonacci

Patrice SÉÉBOLD*
LIRMM, UMR 5506 CNRS
161 rue Ada
34392 Montpellier Cédex 5, FRANCE
Patrice.Seebold@lirmm.fr

1 Introduction

Let $A_k = \{1, \ldots, k\}$ be the alphabet whose letters are the first k positive integers.

The *enumeration derivative* of a word u over A_k is the word $d(u)$ obtained by describing (Conway uses the terminology "look and say") the word u. For example, if $u = 3\ 5\ 5\ 2$ then the description of u is "one three, two fives, one two" thus $d(u) = 1\,3\,2\,5\,1\,2$.

In the enumeration derivation it is always understood that repetitions are collected maximally: in the previous example the description cannot be "one three, one five, one five, one two". This is obvious when u is written as $u = 3\ 5^2\ 2$.

Moreover, to be derivable a word needs to have only finite runs of each letter. In the following, all the words will have this property.

The word $d(u)$ is the first *descendant* of u. Applying the derivation process n times gives $d^n(u)$, the n-th descendant of u. As usual, $d^0(u) = u$.

It seems that this operation was first studied by Conway [3]. Other properties were given by Germain-Bonne in a series of three unpublished papers [4, 5, 6]. This operation may also be compared to run-length encoding (see, e.g., the recent work of Brlek *et al.* on smooth infinite words [2]).

The goal of the present paper is to start the study of the behaviour of binary words generated by morphisms under enumeration derivation. As we will see below the case of the Fibonacci word is interesting.

*Also: Département Mathématiques Informatique et Applications - UFR4, Université Paul Valéry, Route de Mende, 34199 Montpellier Cédex 5, France

2 Fixed points

2.1 Finite fixed points

Property 1. *[3, 4] The empty word ε and the word 22 are the only finite fixed points of d.* ■

In the following, ε and 22 will be called *trivial words*. A consequence of the above property is that every non trivial finite word u has an infinite number of descendants that are all different; in particular $d^\omega(u) = \lim_{n\to\infty} d^n(u)$ is an infinite word (more precisely a set of infinite words, see the Remark below).

Property 2. *[3, 5] For every positive integer k and every word u over A_k, the word $d(u) = a_1^{\alpha_1} a_2^{\alpha_2} \ldots a_n^{\alpha_n} \ldots$ is such that $\alpha_i \in \{1, 2, 3\}$ for all $i \geq 1$.* ■

This implies in particular that if u is a non trivial word then $d^\omega(u)$ is a (set of) infinite word over the alphabet $\{1, 2, 3\}$.

2.2 Cycles

Let $x_1 = 1\,3\,2\,1\,1\,3\,2$ and $y_1 = 1\,3\,2\,2\,1\,1\,3\,3\,1\,1\,2$. The word

$$X_1 = x_1\, y_1\, d^3(y_1)\, d^6(y_1) \ldots = x_1 \prod_{i=0}^{\infty} d^{3i}(y_1)$$

is an infinite fixed point of d^3, i.e., $X_1 = d^3(X_1)$.

Let $x_2 = 3\,1\,2$ and $y_2 = 3\,2\,1\,1\,2$. The word

$$X_2 = x_2\, y_2\, d^3(y_2)\, d^6(y_2) \ldots = x_2 \prod_{i=0}^{\infty} d^{3i}(y_2)$$

is an infinite fixed point of d^3, i.e., $X_2 = d^3(X_2)$.

The words $X_3 = 2\,2\,X_1$ and $X_4 = 2\,2\,X_2$ are also fixed points of d^3.

The following result indicates that these four words are essentially the only infinite fixed points of some power of d.

Property 3. *[5] Let k be a positive integer and let u be an infinite word over A_k. If there exists a positive integer n such that $d^n(u) = u$ then $n = 3$ and $u \in \bigcup_{j=1}^{4} \{X_j, d(X_j), d^2(X_j)\}$.* ■

Remark. Each of the four sets $\{X_j, d(X_j), d^2(X_j)\}$ is a *cycle*. We will see below that each non trivial binary word u ultimately tends towards one of these four cycles. A consequence is that, for a word u, the word $d^\omega(u)$ is generally not well defined because it has three different values: $\lim_{n\to\infty} d^{(3n)}(u)$, $\lim_{n\to\infty} d^{(3n+1)}(u)$, and $\lim_{n\to\infty} d^{(3n+2)}(u)$. But if one of these limits is X_j ($j \in \{1,2,3,4\}$) then the two others are necessarily $d(X_j)$ and $d^2(X_j)$ for the same value of j. In what follows we only need to know in which "family" (i.e., the value of j) falls $d^\omega(u)$; thus we will write $d^\omega(u) = X_j$ to indicate that $\{\lim_{n\to\infty} d^{(3n)}(u), \lim_{n\to\infty} d^{(3n+1)}(u), \lim_{n\to\infty} d^{(3n+2)}(u)\}$
$= \{X_j, d(X_j), d^2(X_j)\}$.

3 Descendants of binary words

Let $A_2 = \{1, 2\}$. An exhaustive investigation gives the following.

Property 4. *Let u be a word over A_2.*

- *If u starts with* $1\,1\,1\,2\,2\,2$, $2\,2\,2\,1\,1\,2\,1$, $2\,2\,1\,1\,2\,1\,2$, *or* $2\,2\,1\,1\,2\,1\,1\,1$ *then* $d^\omega(u) = X_2$.
- *If u starts with* $2\,2\,1\,1\,1\,2\,2\,2$ *then* $d^\omega(u) = X_4$.
- *If u starts with* $2\,2\,1\,2$, $2\,2\,1\,1\,1\,1\,1$, $2\,2\,1\,1\,1\,2\,1$, *or* $2\,2\,1\,1\,1\,2\,2\,1$ *then* $d^\omega(u) = X_3$.
- *Otherwise* $d^\omega(u) = X_1$. ■

As an interesting corollary, we obtain that almost all the Sturmian words have the same ultimate descendant.

Corollary 5. *Let u be a Sturmian word over A_2.*

- *If u starts with* $2\,2\,1$ *then* $d^\omega(u) = 2\,2\,X_1$.
- *Otherwise* $d^\omega(u) = X_1$. ■

4 The Fibonacci case

Let φ be the morphism on A_2 defined by $\varphi(1) = 1\,2$, $\varphi(2) = 1$. The *Fibonacci word* is the infinite word $F = \varphi^\omega(1)$.

$$F = 1\,2\,1\,1\,2\,1\,2\,1\,1\,2\,1\,1\,2\,1\,2\,1\,1\,2\,1\,2\,1\,1\,2\,1\,1\,2\,1\,2\,1\,1\,2\,1\,1\,2\ldots$$

From Corollary 5 we know that $d^\omega(F) = X_1$.

Let $F' = d(F)$ be the first descendant of F. The word F' has interesting properties that we start studying here.

$$F' = 1\,1\,1\,2\,2\,1\,1\,2\,1\,1\,1\,2\,2\,1\,1\,2\,2\,1\,1\,2\,1\,1\,1\,2\,2\,1\,1\,2\,1\,1\,1\,2\,2\,1\,1\,2\ldots$$

4.1 Generating F'

The word F can be factorized over the set $\{1\,2, 1\,1\,2\}$. It is clear that each occurrence of the letter 2 (which is always followed by the letter 1) corresponds in F' to an occurrence of $1\,2$ and since in F the letter 2 is preceded either by 1 or by $1\,1$, F' decomposes over $\{1\,1\,1\,2, 2\,1\,1\,2\}$.

Thus F' is obtained from F in the following way. Let d' be the morphism on A_2 defined by $d'(1) = 2\,1\,1\,2$, $d'(2) = 1\,1\,1\,2 : F' = d'(2\,F)$.

Proposition 6. *The word F' cannot be generated by a morphism, but it is generated by a tag-system.*

Sketch of proof. The only cube at the beginning of F' is $1\,1\,1$. If F' was generated by a morphism g then it would start with $g(1\,1\,1)$, a cube different from $1\,1\,1$.

Let a, b be two new letters and let $A'_2 = \{1, 2, a, b\}$. Let h be the morphism defined on A'_2 by $h(1) = 1\,a\,a\,b\,2\,a\,a\,b$, $h(2) = 1\,a\,a\,b\,2\,a\,a\,b\,2\,a\,a\,b$, and $h(a) = h(b) = \varepsilon$. Let c be the morphism from A'_2 onto A_2 defined by $c(1) = c(a) = 1$ and $c(2) = c(b) = 2$. The word F' is generated by the tag-system $< A'_2, h, 1, A_2, c >$, i.e., $F' = c(h^\omega(1))$. ■

4.2 The complexity of F'

The *complexity function* of an infinite word u is the function P_u which gives for each non negative integer n the number $P_u(n)$ of different factors of length n in u.

Proposition 7. $P_{F'}(0) = 1$, $P_{F'}(1) = 2$, $P_{F'}(2) = 4$, $P_{F'}(3) = 6$, $P_{F'}(0) = n + 4, n \geq 4$.

Sketch of proof. This is a direct consequence of the fact that F' contains exactly one right special factor of each length greater than or equal to 4. Such a factor necessarily ends with $2\,1\,1\,2$ and is the end of the image by d' of the unique right special factor of each length in F. ■

4.3 The word F' is a Lyndon infinite word

Here we suppose A_2 is totally ordered by $1 < 2$. A finite nonempty word over A_2 is a *finite Lyndon word* if it is smaller than all its proper suffixes. An infinite word over A_2 is an *infinite Lyndon word* if it has an infinite number of prefixes being finite Lyndon words.

Theorem 8. *The word F' is an infinite Lyndon word.*

The following lemma, which is central in the proof of Theorem 8, is a well known property of the Fibonacci word F. It is a particular case of the same result (explicitly proved in [1]) about characteristic words.

Lemma 9. *The word* $2\,F$ *is lexicographically greater than all its proper suffixes.*

■

5 Further investigations

Many other properties of the enumeration derivation should be investigated. We only mention here the following one which concerns the Fibonacci word. One can remark that F has the particularity that its first descendant F' is also a word over A_2. This is of course not the case in general since, for a word $u \in A_2$, $d(u) \in A_2$ iff u does not contain any factor $1\,1\,1$, nor $2\,2\,2$. The word F is a particular case between Sturmian words because the morphism φ generates a word which contains no $1\,1\,1$, nor $2\,2\,2$.

Let $St = \{\varphi, \tilde{\varphi}, E\}$ be the set of all the Sturmian morphisms ($E : 1 \mapsto 2\,, 2 \mapsto 1$ and $\tilde{\varphi} : 1 \mapsto 2\,1\,, 2 \mapsto 1$). Except φ, are there some other morphisms from St generating words without $1\,1\,1$ and $2\,2\,2$? The answer is of course yes since it is already known that the morphisms from the set $\{\varphi E, \tilde{\varphi} E\}^+ \cup \{E\varphi, E\tilde{\varphi}\}^+$ generate words with arbitrarily large powers of one single letter (see, e.g., [7]). First investigations gave the following.

Conjecture 10. *A morphism from* St *generates words containing* $1\,1\,1$ *or* $2\,2\,2$ *iff*

$$f \in \{\varphi E, \tilde{\varphi} E, E\varphi, E\tilde{\varphi}, g, Eg\} \textit{ where } g \in \{\varphi E\varphi, \varphi E\tilde{\varphi} = \tilde{\varphi} E\varphi, \tilde{\varphi} E\tilde{\varphi}\}.St$$

References

[1] J. BERSTEL, P. SÉÉBOLD, A characterization of Sturmian morphisms, *MFCS'93*, Gdansk (Pologne), Lect. Notes Comp. Sci. **711** (1993), 281–290.

[2] S. BRLEK, S. DULUCQ, A. LADOUCEUR, L. VUILLON, Combinatorial properties of smooth infinite words, *Theoret. Comput. Sci.* **352** (2006), 306–317.

[3] J. H. CONWAY, The weird and wonderful chemistry of audioactive decay, *in*: T. M. Cover, B. Gopinath (eds), *Open problems in communication and computation*, Springer-Verlag, New-York 1987, 173–188.
See also *Eureka* **46** (1986), 5–18.

[4] B. GERMAIN-BONNE, À propos d'une itération sur chaînes de caractères numériques, *Laboratoire d'Analyse Numérique et d'Optimisation, Université Lille 1*, Research Report ANO 293, january 1993, 8 pp.

[5] B. GERMAIN-BONNE, Chaînes alphanumériques ; cycles et points fixes, *Laboratoire d'Analyse Numérique et d'Optimisation, Université Lille 1*, Research Report ANO 301, march 1993, 23 pp.

[6] B. GERMAIN-BONNE, Mots autodescriptifs et co-descriptifs, *Laboratoire d'Analyse Numérique et d'Optimisation, Université Lille 1*, Research Report ANO 332, june 1994, 32 pp.

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

Fibonacci Sequence and Beyond

Wen Zhi-Ying

1 Introduction

We use the following terminology.

Let S be an alphabet with finite elements. Let S^* and $\tilde{S}$ stand respectively for the free monoid and the free group generated by S. The empty word ε is their neutral element. Let S^ω be the set of sequences (or infinite words), indexed by $\mathbb{N}$ ($0 \in \mathbb{N}$ by convention), on S.

If $w \in S^*$ is a word, we denote by $|w|$ its length and, for a letter $a \in S$, by $|w|_a$ the number of occurrences of the letter a in it. Let $P(w)$ stand for the vector $(|w|_a)_{a\in S}$, called the Parikh vector of w.

A word v is a factor of a word w, written $v \prec w$, if there exist $u, u' \in S^*$, such that $w = uvu'$. We say that v is a prefix (resp. suffix) of a word w, and then we write $v \lhd w$ (resp. $v \rhd w$), if there exists $u \in S^*$ such that $w = vu$ (resp. $w = uv$). The notions of prefix and factor extend in a natural way to infinite words. The language of length n of w, denoted by $\Omega_n(w)$ (Ω_n for short if no confusion), is the set of all factors of length n of w.

If $v \prec w$, where $w = w_0w_1 \cdots w_n \cdots$ (with $w_i \in S$) is a finite word or a sequence, v is said to occur at place m in w if $w_m w_{m+1} \cdots w_{m+|v|-1} = v$. In this case, we also say that m is a place where u occurs in w.

Let $w = w_0w_1 \cdots w_{n-1} \in S^*$, where $w_i \in S$. The mirror word $\overline{w}$ of w is defined to be $\overline{w} = w_{n-1} \dots w_1w_0$. A word w is called a palindrome if $w = \overline{w}$.

A morphism $\tau : S^* \to S^*$ is called a substitution of S^*. Since τ is determined by its images on letters, we always denote the substitution by the vector $(\tau(a))_{a\in S}$. We denote by F_τ any one of the fixed points of τ (i.e. $\tau(F_\tau) = F_\tau$), if it exists, and by M_τ the matrix $(P(\tau(a))^t)_{a\in S} = (|\tau(b)|_a)_{a,b\in S}$ (where the superscript ^{t} means the transposition of a vector) called the matrix of the substitution τ. A substitution is said to be primitive if its matrix is.

Let $w = w_0w_1 \cdots w_{n-1} \in S^*$ ($w_i \in S$), we denote by w^{-1} the inverse word of w, that is $w^{-1} = w_{n-1}^{-1} \cdots w_1^{-1} w_0^{-1}$. Let $w = uv$, then $wv^{-1} := u$ and $u^{-1}w := v$ by convention.

Let $w = w_0w_1 \cdots w_{n-1} \in S^*$ (with $w_i \in S$) and $0 \leq k < |w|$, we define the k-th conjugate of w by $C_k(w) := x_k \cdots x_{n-1}x_0x_1 \cdots x_{k-1}$. The set of conjugates of w is defined by $C(w) := \{C_k(w);\ 0 \leq k < |w|\}$. A word $w \in S^*$ is said to be primitive if $w = u^p$ with $p \in \mathbb{N}$ implies $p = 1$, in other words, if the conjugates of w are distinct (See [2] for example).

Let $w \in \tilde{S}$, we denote by $\imath_w$ the inner isomorphism $u \mapsto wuw^{-1}$, $u \in S^*$. If there exists a $w \in S^*$ such that $\phi = \imath_w\tau$ or $\tau = \imath_w\phi$, we say that ϕ is conjugate to τ and write $\phi \sim \tau$.

Two sequences s and t are said to be locally isomorphic if, for any factor w of s, either w or its mirror $\overline{w}$ is a factor of t and vice versa.

2 Two-Letter Case

We consider the two-letter alphabet $S = \{a, b\}$ case in this section.

2.1 The Fibonacci sequence and substitution

Define $\sigma = (ab, a)$, called the Fibonacci substitution. Its fixed point

$$\xi = abaababa \cdots$$

is called the Fibonacci sequence.

For $n \geq 0$, define $F_n = \sigma^n(a)$. Thus $F_0 = a, F_1 = ab, \cdots$ And $F_{-1} = b$ by convention.

Easy to see that $|F_n| = l_n$, where $\{l_n\}$ is the Fibonacci numbers, defined by $l_{n+2} = l_{n+1} + l_n$ and $l_0 = l_{-1} = 1$.

Put, for $n \geq -1$, $w_n = \alpha F_n \beta^{-1}$, where β is the last letter of F_n, and α is another letter in the alphabet. Thus $w_{-1} = a, w_0 = b, w_1 = aa, \cdots$

Proposition 2.1. *For $n \geq -1$, w_n is a palindrome, and*

$$\Omega_{l_n} = C(F_n) \cup \{w_n\}.$$

The word w_n is called the n-th singular word of the Fibonacci sequence.

Theorem 2.2. *We have*

$$\xi = w_{-1}w_0w_1w_2 \cdots := \prod_{j=-1}^{\infty} w_j.$$

Let $s = s_0s_1s_2s_3\cdots$ be a sequence, u be a factor of s. Suppose that u occurs in s at places p and q ($p < q$), then the distance between the two occurrences is defined to be the difference $q - p$. If the distance between any two (distinct) occurrences of u is larger than or equal to (resp. strictly larger than) the length of u, we say that u is separated (resp. positively separated) in s.

Theorem 2.3. *For any $n \geq 0$,*
1. w_n is positively separated in ξ;
2. Let

$$\xi = z_0w_nz_1w_nz_2w_n\cdots$$

be the factorization of ξ which exhausts all the occurrence of w_n in ξ, then $z_0 = \prod_{j=-1}^{n-1} w_j$, and $z = z_1z_2\cdots$ is the Fibonacci sequence over the new alphabet $\{w_{n+1}, w_{n-1}\}$.

The above factorization is called the singular factorization of the Fibonacci sequence.

As applications, the properties of the singular word-in particular, the positively separated property of the singular words-can be used in the study of the power of the factors, the local isomorphism problem, the study of the special words, the overlap of the factors, *etc*. The reader is referred to [13] for more details.

2.2 Invertible substitutions

We denote by Aut($\tilde{S}$) the group of automorphisms of $\tilde{S}$. It is known [5, 7] that Aut($\tilde{S}$) is generated by the following three special automorphisms

$$\sigma = (ab, a), \quad \pi = (b, a), \quad \phi = (a, b^{-1}).$$

If a substitution is also in Aut($\tilde{S}$), it is called an invertible substitution. The set of invertible substitutions is a monoid, and the monoid is denoted by IS(S).

Theorem 2.4. ([12]) IS*(S) is generated by the three substitutions $\pi = (b, a), \sigma = (ab, a)$, and $\rho = (ba, a)$:* IS*(S)*=$\langle (b, a), (ab, a), (ba, a) \rangle$.

Likewise, for an invertible substitution, one can define the singular words, and then consider the singular factorizations. The reader is referred to [11] for more information.

Also the reader can find the applications of singular factorizations in [6].

On the local isomorphism problem of invertible substitutions, one has

Theorem 2.5. ([14]) *Let φ_1 and φ_2 be two invertible primitive substitutions having ξ_1 and ξ_2 as fixed points respectively. Then the following assertions are equivalent;*

1. ξ_1 and ξ_2 are locally isomorphic;

2. there exist a primitive substitution φ, and integers m, n such that $M_{\varphi_1} = M_{\varphi}^m$ and $M_{\varphi_2} = M_{\varphi}^n$.

2.3 Sturmian sequences

A sequence s is called Sturmian if, for any $n \in \mathbb{N}$, the number of its factors of length n is exactly $n+1$. The sequence s is said to be balanced if, for any pair u, v of factors of s with the same length, one has $|\ |u|_a - |v|_a\ | \leq 1$. The sequence s is ultimately periodic, if there exist $u, v \in S^*$ such that $s = uvvvv \cdots = uv^\infty$.

Proposition 2.6. *A sequence is Sturmian if and only if it is balanced and not ultimately periodic.*

There are many other equivalent definitions of Sturmian sequences. An excellent description of Sturmian sequences can be found in Chapter 2 in the book [3].

A substitution τ is called Sturmian if $\tau(\xi)$ is a Sturmian sequence for any Sturmian sequence ξ. And all the Sturmian substitutions form a monoid, and this monoid coincides with the monoid of invertible substitutions [4, 12]. This implies, in particular, that the fixed point (if any) of an invertible substitution is a Sturmian sequence (except for the following substitutions: (ab^n, b) which fixes ab^∞, and (a, ba^n), which fixes ba^∞; these substitutions are not primitive and their fixed points are ultimately periodic).

The singular words and singular factorization of general Sturmian sequences are studied in [1].

2.4 Invertible substitutions and Sturmian sequences

As we mentioned above, there are many equivalent definitions of Sturmian sequences. One of them is the following: Consider the intersections of a line $y = \alpha x + \beta$ ($\alpha > 0, \beta \in [-\alpha, 1]$) with the lines of grid with nonnegative coordinates. Code the intersection points one by one in the following way: code it by 0 if the $y-$coordinate of the point is an integer, code it by 1 otherwise. If the point is a lattice point, we agree to code it by 01 or 10 (See [3]). Then we get a so-called cutting sequence.

On the other hand, a fixed point of an invertible substitution is Sturmian, thus there is a line $y = \alpha x + \beta$ associated with this substitution. In [9], given an invertible substitution, we determine the line.

3 Three-Letter Case

Now we consider the three letter case: $S = \{a, b, c\}$.

3.1 The Tribonacci sequence and substitution

The Tribonacci sequence is a natural generalization of the Fibonacci sequence.

The substitution $\sigma = (ab, ac, a)$ (recall that this means $\sigma(a) = ab, \sigma(b) = ac, \sigma(c) = a$) is called the Tribonacci substitution and its fixed point

$$\xi = abacabaabacababacabaabac \cdots$$

is called the Tribonacci sequence.

Denote the words

$$A_n = \sigma^n(a); \qquad B_n = \sigma^n(b); \qquad C_n = \sigma^n(c).$$

By convention, $A_0 = a$ and $A_{-1} = c$ (thus $\sigma(A_{n-1}) = A_n$ for $n \geq 0$). Define the number sequence $\{f_n\}_{n \geq -1}$ as $f_n = f_{n-1} + f_{n-2} + f_{n-3}$ $(n \geq 2)$ with $f_{-1} = f_0 = 1$ and $f_1 = 2$. Then $|A_n| = f_n$.

Define the words D_n, E_n and F_n as follows.

$$D_n = A_{n-1}A_{n-2} \cdots A_2A_1A_0, \quad A_n = D_{n-1}E_n = D_{n-2}F_n.$$

Then one shows that D_n is a palindrome, and

Theorem 3.1. *The set of factors of length f_n $(n \geq 2)$ can be divided into the following three classes:*

- Ω_n^0={ *conjugates of* A_n }.
- Ω_n^1={ *factor of length* f_n *of the word* $\alpha^{-1}\overline{E_n}D_{n-1}E_n\alpha^{-1}$ }.
- Ω_n^2={ *factor of length* f_n *of the word* $\beta^{-1}E_{n+1}D_{n-2}\overline{E_{n+1}}\beta^{-1}$ }.

Where α is the last letter of E_n and β is the first letter of E_{n+1}. Moreover, the 3 classes are pairwise disjoint, and each Ω_n^i $(i = 0, 1, 2)$ is closed under the operation of taking mirror, i.e. $\overline{\Omega_n^i} := \{\overline{w};\, w \in \Omega_n^i\} = \Omega_n^i$.

The words in Ω_n^1 and Ω_n^2 are called singular words. In Fibonacci case, given a length, there is only one singular word, but in Tribonacci case, there are two classes of singular words, thus the situation is much more complicated.

Theorem 3.2. *For $n \geq 1$, any factor in Ω_n^1 or Ω_n^2 is positively separated.*

One can also consider the singular factorization and the applications, see [10] for more details.

3.2 Invertible substitutions

The structure of the monoid of invertible substitution over 3 letters is much more complicated. We know that in 2-letter case, the monoid is finitely generated. But in 3-letter case, there are infinitely many so-called indecomposable substitutions, thus

Proposition 3.3. ([15]) *The monoid of invertible substitution over 3 letters is not finitely generated.*

Fortunately, one can show that

Theorem 3.4. ([8]) *1. Let $\sigma \in \mathrm{IS}(S^*)$ be an invertible substitution. Then there exist $w \in S^*$ or $w^{-1} \in S^*$, $\sigma_1, \cdots, \sigma_k \in \{\pi_1, \pi_2, \phi_l, \phi_r\}$, where $\pi_1 = (b, a, c), \pi_2 = (c, b, a), \phi_l = (ba, b, c), \phi_r = (ab, b, c)$, such that*

$$\sigma = \imath_{w^{-1}} \circ \sigma_1 \circ \cdots \circ \sigma_k.$$

2. Let M be a 3×3-matrix of non-negative integral coefficients. M is the substitution matrix of some invertible substitution if and only if it is a finite product of non-negative elementary matrices.

3.3 Open problems

QUESTION 1 The complexity of the fixed point of an invertible substitution.

QUESTION 2 The singular words for general invertible substitutions.

QUESTION 3 Local isomorphism problem for 3 letters.

QUESTION 4 Generalization of Sturmian sequences in high dimension.

QUESTION 5 The structure of monoid of invertible substitutions over 4 or more letters.

References

[1] W.T. Cao and Z.Y. Wen, Some properties of the factors of Sturmian sequences. *Theoret. Comput. Sci.* 304 (2003), no. 1-3, 365–385.

[2] M. Lothaire, *Combinatorics on words.* Encyclopedia of Mathematics and its applications Vol. 17, Addison-Wesley, (1983).

[3] M. Lothaire, *Algebraic combinatorics on words.* Encyclopedia of Mathematics and its Applications Vol. 90, Cambridge, (2002).

[4] F. Mignosi and P. Séébold, Morphismes sturmiens et règles de Rauzy. *J. Théor. Nombres Bordeaux* 5 (1993), no. 2, 221–233.

[5] J. Nielsen, Die Isomorphismengruppen der freien Gruppen, *Math. Ann.* 91 (1924) 169–209. Available at http://mathlib.sub.uni-goettingen.de/JFM/digit.php?an= JFM+50.0078.04.

[6] J. Peyrière, B. Tan, Z.X. Wen and J. Wu, The factor composition matrix of sequences. *Theoret. Comput. Sci.* 329 (2004), no. 1-3, 251–269.

[7] J. Peyrière, Z.X. Wen and Z.Y. Wen, Polynômes associés aux endomorphismes de groupes libres, *Enseign. Math.* 39 (2) (1993) 153–175.

[8] B. Tan, Z.X. Wen and Y.P. Zhang, The structure of invertible substitutions on a three-letter alphabet. *Adv. in Appl. Math.* 32 (2004), no. 4, 736–753.

[9] B. Tan and Z.Y. Wen, Invertible substitutions and Sturmian sequences. *European J. Combin.* 24 (2003), no. 8, 983–1002.

[10] B. Tan and Z.Y. Wen, Some Properties of the Tribonacci Sequence. Submitted to *European J. Combin.*

[11] Z.X. Wen and Z.Y. Wen, Some studies of factors of infinite words generated by invertible substitutions, in: Fifth Conference on Formal Power Series and Algebraic Combinatorics, Florence, 1993, 455–466.

[12] Z.X. Wen and Z.Y. Wen, Local isomorphisms of invertible substitutions. *C. R. Acad. Sci. Paris Sér. I Math.* 318 (1994), no. 4, 299–304.

[13] Z.X. Wen and Z.Y. Wen, Some properties of the singular words of the Fibonacci word. *European J. Combin.* 15 (1994), no. 6, 587–598.

[14] Z.X. Wen, Z.Y. Wen and J. Wu, Invertible substitutions and local isomorphisms. *C. R. Math. Acad. Sci. Paris* 334 (2002), no. 8, 629–634.

[15] Z.X. Wen and Y.P. Zhang, Some remarks on invertible substitutions on three letter alphabet. *Chinese Sci. Bull.* 44 (1999), no. 19, 1755–1760.

Combinatorial and Arithmetic Properties of Symmetric k-IET languages

Luca Zamboni

Abstract

We begin with a brief discussion of the rich interaction between the combinatorics of Sturmian words on one hand, and the dynamics of circle rotations on the other hand, and the connection between the two as given by the continued fraction algorithm. In this context the continued fraction algorithm is described as a natural process for coding the evolution of the bispecial factors of a Sturmian language. There have been numerous attempts to generalize this theory to a broader setting, by either generalizing the combinatorial description, as in the study of Episturmian words, or by generalizing the arithmetic description, as in the theory of multi-dimensional continued fractions, or by generalizing the dynamical description, as in the case of the symbolic dynamics of interval exchange transformations. Our talk will primarily focus on the latter generalization. We will describe the very rich combinatorial structure behind the symbolic codings of a 'symmetric' exchange of k-intervals. We give a combinatorial characterization of such systems which reduces to the usual characterization of Sturmian words (in terms of the subword complexity) in the binary case $k = 2$. Analogous to the Sturmian case, we describe a natural way to code the evolution of the bispecial factors which in turn defines a new multi- dimensional arithmetic algorithm. This coding process may be described in terms of an infinite path in a finite graph G_k. Each vertex of G_k, i.e., each allowable state, may be viewed as a tree of relations possessing a certain 'circular' structure which is also inherent in the structure of RNA. This is based on joint work with J. Cassaigne and S. Ferenczi.

Turku Centre for Computer Science
TUCS General Publications

20. **Reima Suomi, Jarmo Tähkäpää (Eds.)**, Health and Wealth trough Knowledge
21. **Johan Lilius, Seppo Virtanen (Eds.)**, TTA Workshop Notes 2002
22. **Mikael Collan**, Investment Planning – An Introduction
23. **Mats Aspnäs, Christel Donner, Monika Eklund, Pia Le Grand, Ulrika Gustafsson, Timo Järvi, Nina Kivinen, Maria Prusila, Thomas Sund (Eds.)**, Turku Centre for Computer Science, Annual Report 2000-2001
24. **Ralph-Johan Back and Victor Bos**, Centre for Reliable Software Technology, Progress Report 2003
25. **Pirkko Walden, Stina Störling-Sarkkila, Hannu Salmela and Eija H. Karsten (Eds.)**, ICT and Services: Combining Views from IS and Service Research
26. **Timo Järvi and Pekka Reijonen (Eds.)**, People and Computers: Twenty-one Ways of Looking at Information Systems
27. **Tero Harju and Juhani Karhumäki (Eds.)**, Proceedings of WORDS'03
28. **Mats Aspnäs, Christel Donner, Monika Eklund, Pia Le Grand, Ulrika Gustafsson, Timo Järvi and Nina Kivinen (Eds.)**, Turku Centre for Computer Science, Annual Report 2002
29. **João M. Fernandes, Johan Lilius, Ricardo J. Machado and Ivan Porres (Eds.)**, Proceedings of the 1st International Workshop on Model-Based Methodologies for Pervasive and Embedded Software
30. **Mats Aspnäs, Christel Donner, Monika Eklund, Ulrika Gustafsson, Timo Järvi and Nina Kivinen (Eds.)**, Turku Centre for Computer Science, Annual Report 2003
31. **Andrei Sabelfeld (Editor)**, Foundations of Computer Security
32. **Eugen Czeizler and Jarkko Kari (Eds.)**, Proceedings of the Workshop on Discrete Models for Complex Systems
33. **Peter Selinger (Editor)**, Proceedings of the 2nd International Workshop on Quantum Programming Languages
34. **Kai Koskimies, Johan Lilius, Ivan Porres and Kasper Østerbye (Eds.)**, Proceedings of the 11th Nordic Workshop on Programming and Software Development Tools and Techniques, NWPER'2004
35. **Kai Koskimies, Ludwik Kuzniarz, Johan Lilius and Ivan Porres (Eds.)**, Proceedings of the 2nd Nordic Workshop on the Unified Modeling Language, NWUML'2004
36. **Franca Cantoni and Hannu Salmela (Eds.)**, Proceedings of the Finnish-Italian Workshop on Information Systems, FIWIS 2004
37. **Ralph-Johan Back and Kaisa Sere**, CREST Progress Report 2002-2003
38. **Mats Aspnäs, Christel Donner, Monika Eklund, Ulrika Gustafsson, Timo Järvi and Nina Kivinen (Eds.)**, Turku Centre for Computer Science, Annual Report 2004
39. **Johan Lilius, Ricardo J. Machado, Dragos Truscan and João M. Fernandes (Eds.)**, Proceedings of MOMPES'05, 2nd International Workshop on Model-Based Methodologies for Pervasive and Embedded Software
40. **Ralph-Johan Back, Kaisa Sere and Luigia Petre**, CREST Progress Report 2004-2005
41. **Tapio Salakoski, Tomi Mäntylä and Mikko Laakso (Eds.)**, Koli Calling 2005 - Proceedings of the Fifth Koli Calling Conference on Computer Science Education
42. **Petri Paju, Nina Kivinen, Timo Järvi and Jouko Ruissalo (Eds.)**, History of Nordic Computing - HiNC2
43. **Tero Harju and Juhani Karhumäki (Eds.)**, Proceedings of the Workshop on Fibonacci Words 2006

TURKU
CENTRE *for*
COMPUTER
SCIENCE

Joukahaisenkatu 3-5 B, 20520 Turku, Finland | www.tucs.fi

University of Turku
- Department of Information Technology
- Department of Mathematics

Åbo Akademi University
- Department of Information Technologies

Turku School of Economics
- Institute of Information Systems Sciences

ISBN 978-952-12-1910-8
ISSN 1239-1905

Harju and Karhumäki (Eds.) Proceedings of the Workshop on Fibonacci Words 2006

Harju and Karhumäki (Eds.) Proceedings of the Workshop on Fibonacci Words 2006

Harju and Karhumäki (Eds.) Proceedings of the Workshop on Fibonacci Words 2006