Publications du **La**boratoire de
Combinatoire et d'
Informatique
Mathématique

36

Srečko Brlek
Christophe Reutenauer
(eds.)

Words 2005
5^{th} International Conference on Words

13–17 septembre 2005
Actes

Université du Québec à Montréal

Publications du **La**boratoire de **C**ombinatoire
et d'**I**nformatique **M**athématique

Responsable des Publications du LaCIM
Srecko Brlek courriel: `brlek@lacim.uqam.ca`

Éditeur technique
André Lauzon

LaCIM, Université du Québec à Montréal
C.P. 8888, Succ. Centre-Ville
Montréal (Québec) Canada
H3C 3P8

Volumes parus

1 *Parallélisme: modèles et complexité*, S. Brlek (éd.), ACFAS'89, Actes, **1990**
2 *Séries indicatrices d'espèces pondérées et q-analogues*, H. Décoste, **1990**
3 *Étude des arbres hyperquaternaires*, L. Laforest, **1990**
4 *Contribution à l'étude des empilements*, P. Lalonde, **1991**
5 *Calcul combinatoire de séries indicatrices de cycles*, I. Constantineau, **1991**
6 *Notes On Schubert Polynomials*, I. G. Macdonald, **1991**
7 *Combinatoire des tableaux oscillants et des polynômes de Bessel*, L. Favreau, **1991**
8 *Réécritures dans l'algèbre de Lie libre, le groupe libre et l'algèbre associative libre*, G. Melançon, **1991**
9 *q-énumération de polyominos convexes*, M. Bousquet-Mélou, **1991**
10 *Atelier de Combinatoire franco-québécois*, J. Labelle, J.-G. Penaud (éd.), Mai 1991, Actes, **1992**
11 *Séries formelles et combinatoire algébrique*, P. Leroux, C. Reutenauer (éd.), Juin 1992, Actes, **1992**
12 *Structures combinatoires et calcul symbolique*, Y. Chiricota, **1993**
13 *Aspects combinatoires des nombres de Stirling, des polynômes orthogonaux de Sheffer et de leurs q-analogues*, A. de Médicis, **1993**
14 *A theory of noncommutative determinants and characteristic functions of graphs I*, I. M. Gelfand and V. S. Rethak; *Matroids on chamber systems*, I. M. Gelfand and A. V. Borovik, **1993**
15 *Modèles mathématiques pour la synthèse des systèmes informatiques*, S. Brlek (éd.), ACFAS'94, Actes, **1994**
16 *Produits et coproduits des fonctions quasi-symétriques et de l'algèbre des descentes*, C. Malvenuto, **1994**
17 *Une interprétation combinatoire des approximants de Padé*, Emmanuel Roblet, **1994**

.... *suite en fin de volume*

Publications du **La**boratoire de
Combinatoire et d'
Informatique
Mathématique

36

Srečko Brlek
Christophe Reutenauer
(eds.)

Words 2005
5^{th} International Conference on Words

13–17 septembre 2005
Actes

Université du Québec à Montréal

La conférence Words 2005 a été organisée par le Laboratoire de combinatoire et d'informatique mathématique (LaCIM), Université du Québec à Montréal du 13 au 17 septembre 2005, au Pavillon Sherbrooke, 200, rue Sherbrooke Ouest, Montréal (Québec), Canada.

Comité d'organisation

Srecko Brlek,	LaCIM, UQAM, président
Cédric Chauve,	LaCIM, UQAM
Annie Lacasse,	LaCIM, UQAM
André Lauzon,	LaCIM, UQAM, webmaster
Geneviève Paquin,	LaCIM, UQAM
Lise Tourigny,	LaCIM, UQAM, secrétaire

Comité scientifique

Christophe Reutenauer,	UQAM, président
Jean Berstel,	Universtié Marne-la-Vallée
James Currie,	Uninversity of Winnipeg
Clelia De Felice,	Università di Salerno
Aldo de Luca,	Università degli Studi Napoli "Federico II"
Juhani Karhumäki,	University of Turku
Jean Néraud,	Université de Rouen.

Avec le soutien financier des organismes suivants :

UQÀM Faculté des sciences
Université du Québec à Montréal

Éducation
Québec

Chaire de recherche du Canada
Algèbre, combinatoire et informatique mathématique

PIMS

CRM

Dépôt légal, troisième trimestre 2005, Bibliothèque nationale du Québec

ISBN 2-89276-376-2 LaCIM Montréal

Laboratoire de combinatoire et d'informatique mathématique
Université du Québec à Montréal
C.P. 8888, Succ. Centre-Ville
Montréal (Québec) Canada
H3C 3P8

LACIM UQA

Table des matières

Avant-propos

La conférence WORDS 2005 était la cinquième édition d'une série qui débuta en 1997 à Rouen (France). Le sujet de la conférence est l'étude des mots avec une emphase sur le point de vue théorique. En particulier les aspects combinatoires algébriques et algorithmiques sont privilégiés bien que les motivations puissent provenir d'autres domaines tels que l'informatique théorique. La conférence consista en six conférences plénières et également 28 communications sélectionnées par le comité de programme. Nous aimerions remercier leurs membres ainsi que les arbitres qui les ont assistés dans cette tâche.

Les conférences antérieures ont eu lieu à Rouen (France, 1997 et 1999), Palermo (Italie, 2001) et Turku (Finlande, 2003).

L'organisation a bénéficié du support financier du Centre de Recherches Mathématiques (CRM), le programme de Chaires de recherche du Canada (CRSNG), Pacific Institute for the Mathematical Siences (PIMS), du ministère de l'éducation du Québec (MEQ) et de la faculté des Sciences de l'Université du Québec à Montréal (UQAM).

Finalement, le Laboratoire de Combinatoire et d'Informatique Mathématique (LaCIM) a la chance de compter sur la secrétaire de la conférence Lise Tourigny, sur le webmestre et TEXmestre André Lauzon, sur nos étudiantes Annie Lacasse et Geneviève Paquin. Nous leur adressons nos remerciements chaleureux pour leur apport inestimable à la tenue de cette conférence.

Srečko Brlek et Christophe Reutenauer
Montréal et Le Rasinel, 25 août 2005

Foreword

The WORDS 2005 Conference was the fifth of a series initiated in 1997 in Rouen (France). The conference topic is the study of words with a focus on the theoretical point of view. In particular, the combinatorial, algebraic and algorithmic aspects are emphasized. Motivations may come from other domains such as theoretical computer science. The conference program included six plenary talks and also 28 contributed communications selected by the program committee. We would like to thank their members as well as the referees who assisted them in this task.

Previous conferences were held in Rouen (France, 1997 and 1999), Palermo (Italy, 2001) and Turku (FInland, 2003).

The organisation was sponsored by the "Centre de Recherches Mathématiques (CRM)", the Canadian Research Chair program (NSERC), the Pacific Institute for the Mathematical Siences (PIMS), the ministry of education of Quebec (MEQ) and the "Faculté des Sciences de l'Université du Québec à Montréal (UQAM)".

Finally, the "Laboratoire de Combinatoire et d'Informatique Mathématique (LaCIM)" is fortunate to count on the conference secretary Lise Tourigny, the webmaster and TEXmaster André Lauzon and our students Annie Lacasse and Geneviève Paquin. Our warmest thanks for their invaluable contribution in the organisation of the event.

Srečko Brlek and Christophe Reutenauer
Montréal and Le Rasinel, August 25, 2005

Subwords: repetitions, frequency, uniformity

Arturo Carpi[*]

[*] University degli studi di Perugia, Dipartimento di Matematica e Informatica, Via Vanvitelli 1, 06123 Perugia, Italy. carpi@dipmat.unipg.it

Constrained Coding Systems with Unconstrained Positions

Maxime Crochemore[*]

Abstract

We give a polynomial-time construction of the set of sequences that satisfy a finite-memory constraint defined by a finite list of forbidden blocks, with a specified set of bit positions unconstrained. Such a construction can be used to build modulation/error-correction codes (ECC codes) like the ones defined by the Immink-Wijngaarden scheme in which certain bit positions are reserved for ECC parity. We give a linear-time construction of a finite-state presentation of a constrained system defined by a periodic list of forbidden blocks. These systems, called periodic-finite-type systems, were introduced by Moision and Siegel. Finally, we present a linear-time algorithm for constructing the minimal periodic forbidden blocks of a finite sequence for a given period.

[*] Institut Gaspard-Monge, Laboratoire d'informatique, Université de Marne-la-Vallée, Cité Descartes, 5 Bd Descartes, Champs-sur-Marne, F-77454, Marne-la-Vallée, CEDEX 2 (France), `Maxime.Crochemore@univ-mlv.fr`

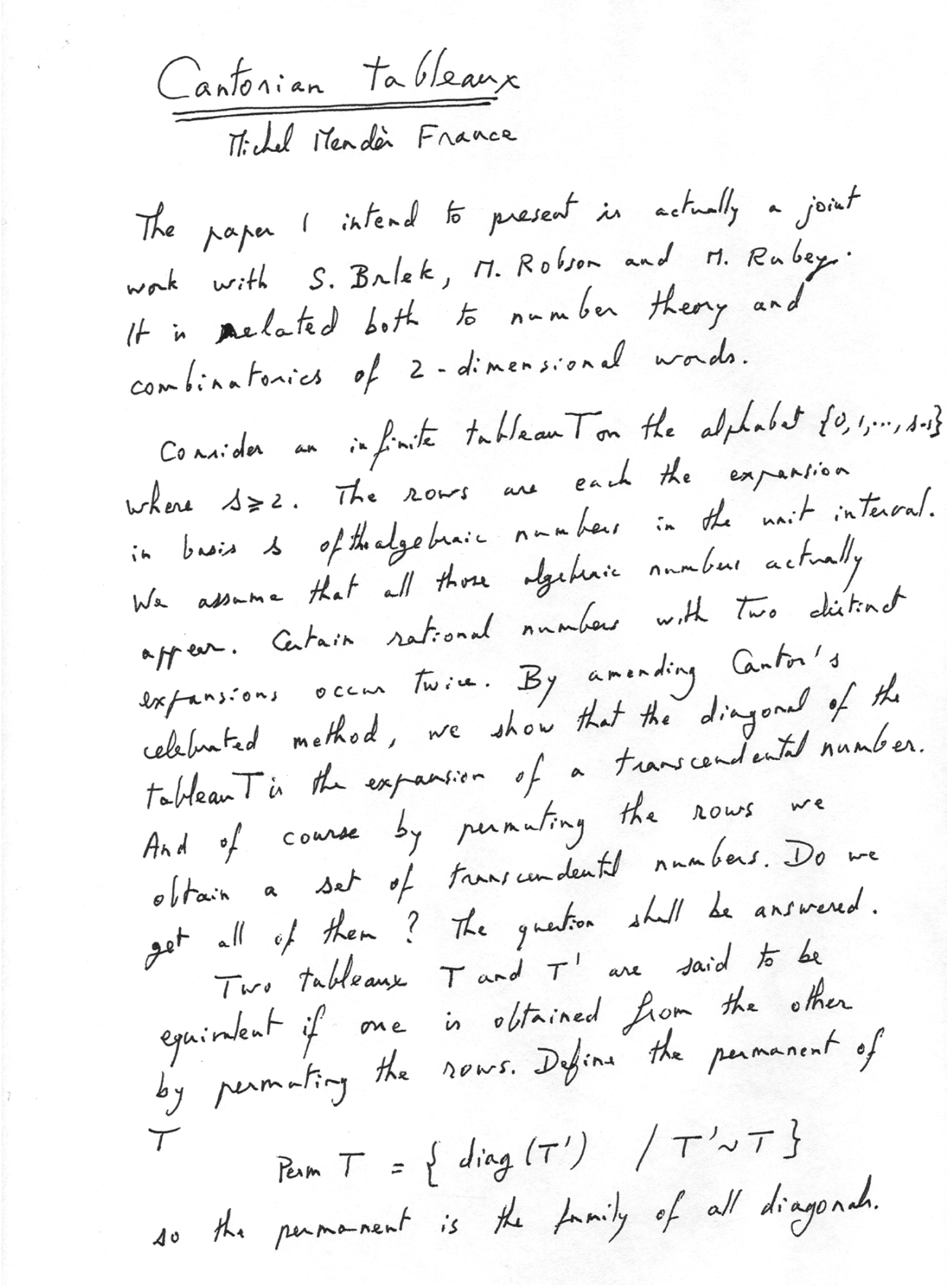
A survey on algebraic characterisations for temporal logics over traces

Volker Diekert[*]

[*] Institut für Formale Methoden der Informatik, Universität Stuttgart, FMI Universitätsstr. 38 D-70569 Stuttgart, diekert@informatik.uni-stuttgart.de

Cantorian Tableaux

Michel Mendès France



Cantorian tableaux

Michel Mendès France

The paper I intend to present is actually a joint work with S. Brlek, M. Robson and M. Rubey. It is related both to number theory and combinatorics of 2-dimensional words.

Consider an infinite tableau T on the alphabet $\{0,1,\dots,s-1\}$ where $s \geq 2$. The rows are each the expansion in basis s of the algebraic numbers in the unit interval. We assume that all those algebraic numbers actually appear. Certain rational numbers with two distinct expansions occur twice. By amending Cantor's celebrated method, we show that the diagonal of the tableau T is the expansion of a transcendental number. And of course by permuting the rows we obtain a set of transcendental numbers. Do we get all of them? The question shall be answered.

Two tableaux T and T' are said to be equivalent if one is obtained from the other by permuting the rows. Define the permanent of T

$$\text{Perm } T = \{ \text{diag}(T') \ / \ T' \sim T \}$$

so the permanent is the family of all diagonals.

The above result states that if $L(T)$ is the set of rows (i.e. algebraic numbers in the unit interval), then

$$L(T) \cap \mathrm{Perm}(T) = \emptyset .$$

This observation becomes now a definition. Let T be a $n \times n$ square tableau on a finite alphabet with s letters. As above, put

$$\mathrm{Perm}(T) = \{ \mathrm{diag}(T') \mid T' \sim T \} .$$

The tableau is Cantorian if $L(T) \cap \mathrm{Perm}(T) = \emptyset$.

For example $\begin{pmatrix} a & b \\ b & a \end{pmatrix}$ is Cantorian whereas

$\begin{pmatrix} a & a & a \\ b & b & b \\ a & a & b \end{pmatrix}$ is not ($a^2 b \in L(T) \cap \mathrm{Perm}(T)$).

Let $C(n,s)$ be the number of $n \times n$ Cantorian tableaux on an s-letter alphabet. Obviously $C(n,s) < s^{n^2}$.

Theorem

(i) If $s = s(n) < (1-\varepsilon)\, n / \log n$, then

$$\lim_{n \to \infty} C(n,s) / s^{n^2} = 0 .$$

For large n, almost no tableaux are Cantorian.

(ii) If $s = s(n) > (1+\varepsilon)\, n / \log n$, then

$$\lim_{n \to \infty} C(n,s) / s^{n^2} = 1 .$$

For large n, almost all tableaux are Cantorian.

The Burrows-Wheeler Transform: a new tool in Combinatorics on Words.

Antonio Restivo[*]

Abstract

Michael Burrows and David Wheeler introduced in 1994 a reversible transformation on words (BWT from now on) that arouses considerable interest and curiosity in the field of Data Compression.

The present contribution discusses the close relation of BWT with some fundamental notions and results in Combinatorics on Words.

We first consider the connections between BWT and a very important theorem of Gessel and Reutenauer that states a bijection between the words on a given alphabet A and the multisets of the conjugacy classes of primitive words over A. Such a connection allows to introduce an extension of BWT, that has several interesting applications to Data Compression and Sequence Comparison.

We then show that BWT is a fundamental tool to investigate the close connections between some important properties of finite binary words, expressed in terms of balance, ordering and conjugacy. This leads in particular to new characterizations of balanced words. Standard words play a central role in this investigation.

[*] University of Palermo, Dipartimento di Matematica ed Applicazioni, Via Archirafi 34, 90123 Palermo, Italy, `restivo@math.unipa.it`

My Favorite Open Problems in Words

Jeffrey Shallit[*]

Abstract

In this talk I will discuss some of my favorite open problems dealing with words and automata. For some of these problems progress has been made, but others have proven remarkably resistant to attack.

[*] School of Computer Science, University of Waterloo, Waterloo, Ontario, N2L 3G1 Canada, `shallit@graceland.math.uwaterloo.ca`

An innocent-looking formula for continued fractions

Boris Adamczewski[*], *Jean-Paul Allouche*[†]

1 Introduction

Looking for patterns in decimal expansions of real numbers is a rewarding hobby that can lead to theorems or, even better, to intractable conjectures: knowing whether the digit 4 occurs infinitely often in the decimal expansions of $\sqrt{7}$ or π seems far out of reach. Alternatively, continued fraction expansions provide another source of questions about patterns that occur in real numbers and several still open conjectures can be stated, e.g., that continued fraction expansions of algebraic numbers of degree at least 3 cannot have bounded partial quotients. In the same spirit several recent papers give transcendence results for real numbers whose base b or continued fraction expansion contain specific patterns like repetitions or palindromes (see for example the survey [7], and the references therein).

In this survey paper we will focus on continued fraction expansions (for real numbers but also for formal Laurent series with coefficients in a finite field) and reversals of patterns or palindromic patterns that occur in such expansions.

We will use the classical notations for finite or infinite continued fractions

$$\frac{p}{q} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}} = [a_0; a_1, \cdots, a_n]$$

[*] CNRS, Institut Camille Jordan, Université Claude Bernard Lyon 1, Bât. Braconnier, 21 avenue Claude Bernard, 69622 VILLEURBANNE Cedex (FRANCE), Boris.Adamczewski@math.univ-lyon1.fr

[†] CNRS, LRI, Université Paris-Sud, Bât. 490, 91405 Orsay Cedex (FRANCE), Jean-Paul.Allouche@lri.fr

resp.

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n + \cfrac{1}{\ddots}}}}} = [a_0; a_1, \cdots, a_n, \cdots]$$

where p/q is a positive rational number, resp. α is a positive irrational real number, n is a nonnegative integer, and the a_i's are positive integers for $i \geq 0$. We will also have continued fractions for formal Laurent series over a field K: in this case, p/q is a rational function (p and q are two polynomials in $K[X]$), resp. α is a Laurent series $\sum_{n \geq t} r_j X^{-j}$, n is a nonnegative integer, and the a_i's are nonzero polynomials in $K[X]$.

2 Two fundamental lemmas

For $0 \leq k \leq n$, let us denote by p_k/q_k the k-th convergent to p/q, i.e., $p_k/q_k = [a_0; a_1, \cdots, a_k]$. In particular, we have $p/q = p_n/q_n = [a_0; a_1, \cdots, a_n]$. The sequence of denominators of the convergents to p/q satisfies, for n such that $0 \leq 1 \leq n$, the relation $q_k = a_k q_{k-1} + q_{k-2}$, with the convention that $q_{-1} = 0$ and $q_0 = 1$.

A pleasant formalism for continued fractions is the matrix formalism that we borrow from papers of van der Poorten (see for example [50, 53]), who says that it goes back at least to [32]: we have that

$$\forall n \geq 0, \ \ [a_0; a_1, \cdots, a_n] = \frac{p_n}{q_n}, \text{ with } \gcd(p_n, q_n) = 1$$

if and only if

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix}.$$

Taking the transpose of this equality easily yields the following lemma.

Lemma 2.1 *Let $a_0, a_1, \ldots$ be positive integers. Let $[a_0; a_1, \cdots, a_n] = \frac{p_n}{q_n}$. Then*

$$\frac{q_n}{q_{n-1}} = a_n + \cfrac{1}{a_{n-1} + \cfrac{1}{\ddots + \cfrac{1}{a_0}}} = [a_n; a_{n-1}, \cdots, a_0]\cdot \tag{2.1}$$

Quite curiously, the innocent-looking Equality (2.1), that will be referred as the *mirror formula* all along this paper, appears to be somewhat ubiquitous. Many occurrences of this formula can be found in the combinatorial study of Sturmian words as well as in the Diophantine property of Sturmian real numbers. Very recently the mirror formula has appeared as being the key point in several studies on simultaneous rational approximation: simultaneous rational approximation of a real number and of its square (see the nice work of Roy), results on the Littlewood conjecture in simultaneous Diophantine approximation, and various transcendence criteria for continued fractions.

A variation on the mirror formula is known as the *folding lemma* (see [46, 50, 53]) whose proof is an easy consequence of the matrix formalism and of the mirror formula.

Lemma 2.2 (Folding lemma) *Let $c, a_0, a_1, \ldots$ be positive integers. Let*

$$[a_0; a_1, \cdots, a_n] = \frac{p_n}{q_n}\,.$$

Then

$$\frac{p_n}{q_n} + \frac{(-1)^n}{cq_n^2} = [a_0; a_1, a_2, \cdots, a_n, c, -a_n, -a_{n-1}, \cdots, -a_1] \tag{2.2}$$

Remark 2.3 In Equality (2.2) negative partial quotients occur. An easy transformation permits to get rid of these forbidden partial quotients (see, e.g., [53]). Note that the terminology "folding lemma" comes from the fact that, defining the word $w := a_1a_2\cdots a_n$ and noting $\overline{w} := a_na_{n-1}\cdots a_1$, we go from $\frac{p_n}{q_n}$ to $\frac{p_n}{q_n} + \frac{(-1)^n}{cq_n^2}$ (up to the first partial quotient a_0) by means of the "perturbed symmetry" $w \longrightarrow w\ c\ \overline{(-w)}$: iterating this operation in the case $w = +1$ and $c = +1$ gives a sequence of ± symbols that is the sequence of creases in a strip of paper repeatedly folded in half (see for example [29]).

This paper surveys various occurrences and uses of the mirror formula and of the folding lemma in combinatorics on words and number theory. Most of them correspond to very recent works. The article is organized as follows. The first part, divided into three sections, is devoted to the well-known Sturmian infinite words and some of their combinatorial properties. The recurrence function of an infinite word is introduced in Section 2. Section 3 is devoted to the study of asymptotic repetitions occurring in infinite words, whereas Section 4 deals with the palindrome density in infinite words. In the second part of Section 4, we investigate Diophantine questions, with a special focus on simultaneous Diophantine approximation, in which the mirror formula plays a central rôle. We begin in Section 5 with rational approximations of real numbers. Section 6 is devoted to uniform simultaneous rational approximation of a real number and of its square. Then, Section 7 deals with another old problem in simultaneous

Diophantine approximation, namely the Littlewood conjecture. Finally, we give in Section 8 some transcendence criteria for continued fractions obtained *via* the subspace Schmidt's theorem. The last section (Section 9) will allude to the use of mirroring and folding for continued fractions of formal Laurent series.

Part I
Sturmian sequences

Sturmian sequences can be defined in several ways. We choose the arithmetic definition. (For a general overview on Sturmian sequences, see for example [12].)

Definition 2.4 A sequence $(u_n)_{n\geq 0}$ is called Sturmian if there exists a positive irrational number α and a real number $\beta \in [0,1)$ such that

either $\forall n \geq 0$, $u_n = \lfloor \alpha(n+1)+\beta \rfloor - \lfloor \alpha n+\beta \rfloor - \lfloor \alpha \rfloor$,

or $\forall n \geq 0$, $u_n = \lceil \alpha(n+1)+\beta \rceil - \lceil \alpha n+\beta \rceil - \lceil \alpha \rceil$.

The number α is called the *slope* of the Sturmian sequence.

A sequence $(u_n)_{n\geq 0}$ is called Sturmian characteristic (or simply characteristic) if it is of the form above with $\beta = 0$.

Remark 2.5 Note that, from the definition, a Sturmian sequence takes its values in $\{0,1\}$.

The following proposition shows a link between Sturmian sequences and continued fractions. (We denote as usual by w^a, where w is a finite word and a a positive integer, the concatenation of a copies of the word w.)

Proposition 2.6 *Let α be an irrational number in $[0,1)$ such that*

$$\alpha = [0; a_1, a_2, \cdots] .$$

Define the sequence of words $(s_j)_{j\geq -1}$ by $s_{-1} := 1$, $s_0 := 0$, $s_1 := s_0^{a_1-1}s_{-1}$, and $s_j := s_{j-1}^{a_j}s_{j-2}$ for $j \geq 2$. Then the sequence $(s_j)_{j\geq 0}$ tends to an infinite word which is equal to the characteristic Sturmian word of slope α.

Definition 2.7 The Fibonacci sequence (or Fibonacci word) on the alphabet $\{0,1\}$ is the characteristic Sturmian sequence defined as $\lim_{j\to+\infty} s_j$ where the words s_j are defined by $s_{-1} := 1$, $s_0 := 0$, and, for all $j \geq 1$, $s_j := s_{j-1}s_{j-2}$. Hence this sequence begins as follows

$$0\,1\,0\,0\,1\,0\,1\,0\,0\,1\,0\,0\,1\,0\,1\,0 \cdots$$

3 Repetitions in Sturmian sequences

Several authors have studied repetitions, i.e., factors (or sub-blocks) of the form w^a) occurring in a Sturmian sequence (see in particular [13, 16, 17, 22, 47, 69]). It happens that the mirror formula is used in these studies. We give, as an example of repetitions in Sturmian words and the mirror formula, a (rephrasing of a) theorem due to Vandeth ([69, Theorem 16]). First recall that the length of a (finite) word w is denoted by $|w|$, and define fractional powers of finite words as follows: if p is a positive real number and w a finite word, then $w^p := w^{\lfloor p \rfloor} u$, where u is the prefix of w of length $\lceil (p - \lfloor p \rfloor)|w| \rceil$. We also define the *critical exponent* of an infinite word as the supremum of all powers occurring in this infinite word.

Theorem 3.1 (Vandeth) *Let α be the real number whose (eventually periodic) continued fraction expansion has the form*

$$\alpha = [0; b_0, b_1, b_2, \cdots, b_m, b_1, b_2, \cdots, b_m \cdots],$$

where the b_i's are positive integers and $b_m \geq b_0$ (in particular α is a quadratic number). Then the critical exponent of the characteristic Sturmian sequence S_α of slope α is

$$\max_{1 \leq t \leq m} [2 + b_t; b_{t-1}, \cdots, b_1, b_m, \cdots, b_1, b_m, \cdots, b_1, \cdots]$$

Sketch of the proof. Vandeth introduces three morphisms X, L, R defined by $X(0) = 1, X(1) = 0$, $L(0) = 0, L(1) = 01$, $R(0) = 0, R(1) = 10$. The proof is then divided into three steps.

– Let T be the "minimal" morphism such that $T(S_\alpha) = S_\alpha$ (the existence of this morphism is granted by [21]). Then T can be written as

$$T = L^{b_0} X L^{b_1} X \cdots L^{b_{m-1}} X L^{b_m - b_0}$$

(m the period of the continued fraction expansion of α has been taken minimal). Define the morphisms F_t, $t \in [-1, m-1]$, by $F_{-1} := Id$, and for $t \in [0, m-1]$

$$F_t := L^{b_0} X L^{b_1} X \cdots L^{b_t} X$$

Then, for $t \in [-1, m-1]$,

$$S_\alpha = \prod_{k=0}^{\infty} F_t(0)^{s_t(k)} F_t(1)$$

where $s_t(k) := \lfloor (k+1)\alpha_t^{-1} \rfloor - \lfloor k \alpha_t^{-1} \rfloor$ and

$$\alpha_t := [0; b_{t+1}, b_{t+2}, \cdots, b_m, b_1, b_2, \cdots, b_m, b_1, b_2, \cdots, b_m, \cdots].$$

– Let $k := 2 + \max_{1 \le i \le m} b_i$. If the word u is such that u^k is a factor of S_α, then u is conjugate to $T^n F_t(0)$ for some $n \ge 0$, and some $t \in [0, m-1]$ (recall that two words are conjugate if they can be written respectively xy and yx).

– The critical exponent of S_α is equal to

$$\max_{1 \le t \le m} [2 + b_t; b_{t-1}, \cdots, b_1, b_m, b_{m-1}, \cdots, b_1, b_m, b_{m-1}, \cdots, b_1, \cdots].$$

Remark 3.2 Note that Vandeth deduces from this theorem the *integer* critical exponent of *any* characteristic Sturmian sequence S_α provided that α has bounded partial quotients (see [69, Theorem 17]).

4 Recurrence function, the Cassaigne spectrum

The *recurrence function* of an infinite sequence describes the size of maximal gaps between two occurrences of a same factor (sub-block) in the sequence. More formally

Definition 4.1 The recurrence function $R_{\mathbf{u}}(n)$ of a sequence $\mathbf{u} = (u_k)_{k \ge 0}$ is defined by: $R(n)$ is the smallest integer such that each factor of length n in the sequence $\mathbf{u}$ contains all factors of length n of the sequence $\mathbf{u}$.

Note that $R(n) \le +\infty$. If $R(n) < +\infty$ for all n, the sequence is said to be *uniformly recurrent.*

The *recurrence quotient* $\rho = \rho_{\mathbf{u}}$ of the sequence $\mathbf{u}$ is defined by $\rho_{\mathbf{u}} := \limsup\limits_{n \to +\infty} \dfrac{R_{\mathbf{u}}(n)}{n}$.

Remark 4.2 It is not difficult to prove that $\rho = +\infty$ if the sequence is not uniformly recurrent, that $\rho = 1$ for a periodic sequence, and that $2 \le \rho \le +\infty$ otherwise.

The following result, due to Cassaigne [19], makes use of the mirror formula.

Theorem 4.3 (Cassaigne) *Let* $\mathbf{u}$ *be a Sturmian sequence of slope*

$$\alpha = [a_0; a_1, a_2, \cdots] .$$

Then

$$\rho_{\mathbf{u}} = 2 + \limsup_{i \to +\infty} [a_i; a_{i-1}, \cdots, a_1].$$

Remark 4.4 The set $\{\rho_{\mathbf{u}},\ \mathbf{u} \text{ Sturmian}\}$ is studied in [19] and hence called the Cassaigne spectrum.

5 Palindrome density

In this section, we consider palindromic prefixes of infinite words. Let us recall that a finite word $W = w_1 w_2 \cdots w_n$ is a palindrome if it is invariant under mirror symmetry, i.e., if it is equal to its *reversal*: $W = \overline{W}$, where $\overline{W} := a_n a_{n-1} \cdots a_1$. Let $\mathbf{w} = w_1 w_2 \cdots w_n \cdots$ be an infinite word beginning in arbitrarily long palindromes. For such a word, let us denote by $(n_i)_{i \geq 1}$ the increasing sequence of all lengths of palindromic prefixes of $\mathbf{w}$. By assumption, this sequence is thus infinite. In [31], Fischler considers the quantity

$$\delta(\mathbf{w}) = \limsup_{i \to \infty} \frac{n_{i+1}}{n_i}.$$

If the word $\mathbf{w}$ begins in only finitely many palindromes, then we set $\delta(\mathbf{w}) = +\infty$. We then define the palindrome density of $\mathbf{w}$, denoted by $d_p(\mathbf{w})$, by

$$d_p(\mathbf{w}) = \frac{1}{\delta(\mathbf{w})}.$$

In particular, we always have $0 \leq d_p(\mathbf{w}) \leq 1$ and $d_p(\mathbf{w}) = 0$ if $\mathbf{w}$ begins in only finitely many palindromes. Furthermore, if $\mathbf{w} = WWW \cdots$ is a periodic word, then $d_p(\mathbf{w}) = 1$ if there exist two (possibly empty) palindromes U and V such that $W = UV$, and $d_p(\mathbf{w}) = 0$ otherwise. Thus the palindrome density of periodic infinite words is either maximal or minimal. This naturally leads to the following question: what is the maximal palindrome density that can be attained by an non-periodic infinite word? This problem is solved in [31].

Theorem 5.1 (Fischler) *Let $\mathbf{w}$ be an infinite non-periodic word. Then,*

$$d_p(\mathbf{w}) \leq \frac{1}{\gamma},$$

where $\gamma = \frac{1+\sqrt{5}}{2}$ is the golden ratio.

The bound obtained in Theorem 5.1 is optimal and reached in particular for the Fibonacci word. More generally, it is possible to compute $d_p(\mathbf{w})$ when $\mathbf{w}$ is a characteristic Sturmian word. Indeed, if $\alpha = [0; a_1, a_2, \cdots]$ denotes a real number and if $\mathbf{w}_\alpha$ is the associated characteristic Sturmian sequence, then

$$d_p(\mathbf{w}_\alpha) = \frac{\sigma + 1}{2\sigma + 1},$$

where $\sigma = \limsup_{n \to \infty} [a_n, a_{n-1}, \cdots, a_1]$. In other words, the computation of the palindrome density of a characteristic Sturmian sequence involves the mirror formula, *via* the convergents to its slope.

Remark 5.2 As suggested by a referee, it is interesting to observe that the characteristic Sturmian words associated with irrational numbers beginning in arbitrarily long palindromes are exactly the standard infinite harmonic word introduced in [18].

We introduce now a modification of the Cassaigne spectrum (see Remark 4.4). Let $\mathcal{S}'_c$ be defined by

$$\mathcal{S}'_c = \{d_p(\mathbf{w}_\alpha),\ \alpha \in (0,1) \setminus \mathbb{Q}\} = \left\{ \frac{\sigma+1}{2\sigma+1},\ \sigma \in \mathcal{S}_c \right\}.$$

We also denote by $\mathcal{S}_p$ the set of the real numbers that can be written $d_p(\mathbf{w})$ for some infinite word $\mathbf{w}$. The following interesting result is proved in [31]: if a non-periodic word $\mathbf{w}$ has a palindrome density that is "too large", then there exists an irrational number α such that $d_p(\mathbf{w}) = d_p(\mathbf{w}_\alpha)$. This can be formalized as follows.

Theorem 5.3 (Fischler) *We have*

$$\mathcal{S}_p \cap \left[\frac{1}{\sqrt{3}}, 1\right) = \mathcal{S}'_c \cap \left[\frac{1}{\sqrt{3}}, 1\right).$$

We end this Section by mentioning that the motivation for studying the palindrome density of infinite words comes from a problem of uniform simultaneous rational approximation. This problem will be introduced in Section 7.

Part II
Diophantine approximation

Diophantine approximation is essentially devoted to the following question: how good an approximation of a given real number by rationals p/q as a function of q can be? Continued fractions and Diophantine approximation are of course intimately connected, since the best rational approximations to a real number are produced by truncating its continued fraction expansion. Thus many Diophantine problems can be solved thanks to continued fractions.

It is however much less known, and quite new, that continued fractions can be used in order to study some questions of simultaneous approximation. We recall that simultaneous rational approximation deals with the more general problem of approximating several real numbers by rationals having the same denominators. Thus, given real numbers $\xi_1, \xi_2, \ldots, \xi_n$, the task is to determine how good an approximation by rationals $p_1/q, p_2/q, \ldots, p_n/q$ as a function of q can be. Mainly due to the lack of a suitable multi-dimensional continued fraction algorithm, this kind of problems are generally considered as rather difficult.

The purpose of this second part is to review some old Diophantine questions together with recent developments where continued fractions, thanks to the mirror formula, are used to provide simultaneous rational approximations for some real numbers. In this regard, Section 6 is an exception since it deals with rational approximation of (only) one real number, defined by its binary expansion. However, Section 6 is still concerned by both Diophantine approximation and the mirror formula.

6 Exact irrationality measure

The *irrationality measure* of an irrational real number α, denoted by $\mu(\alpha)$, is defined as the supremum of the positive real numbers τ for which the inequality

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^\tau}$$

has infinitely many solutions $(p,q) \in \mathbb{Z}^2$. Thus, $\mu(\alpha)$ measures the quality of the best rational approximations to α. Let us recall some well-known facts about this notion. The theory of continued fractions ensures that $\mu(\alpha) \geq 2$, for any irrational number α. Algebraic irrational numbers have irrationality measure 2, as follows from Roth's theorem [57]. This is also the case for almost all real numbers (with respect to the Lebesgue measure). This last result is due to Khintchine [34] (see also [36]). Let us also mention that Liouville numbers are defined as the real numbers having an infinite irrationality measure.

It is in general a challenging problem to obtain an irrationality measure, i.e., to bound the irrationality measure of a given real number. In this section we consider a particular class of irrational numbers having the spectacular property that both their b-adic expansion and their continued fraction expansion can be explicitly determined. We will deduce from this last representation the exact value of their irrationality measure.

With an irrational number α and an integer b, both larger than 1, we associate the real number $S_b(\alpha)$ defined by

$$S_b(\alpha) = (b-1)\sum_{n=1}^{+\infty} \frac{1}{b^{\lfloor n\alpha \rfloor}}.$$

The following nice result can be found in [6] (see also [26]).

Theorem 6.1 (Adams and Davison) *Let $\alpha := [a_0; a_1, a_2, \cdots]$ be a positive irrational number and b be an integer, both larger than 1. Let p_n/q_n be the n-th convergent to $1/\alpha$. For $n \geq 1$, set*

$$t_n = (b^{q_n} - b^{q_{n-2}})/(b^{q_{n-1}} - 1).$$

Then,

$$S_b(\alpha) = [0; t_1, t_2, \cdots, t_n, \cdots].$$

We easily deduce from Theorem 6.1 and the mirror formula, the exact irrationality measure for $S_b(\alpha)$ for any irrational α and any integer b, both larger than 1.

Theorem 6.2 *Let $\alpha := [a_0; a_1, a_2, \cdots]$ be a positive irrational number and b be an integer, both larger than 1. Then,*

$$\mu(S_b(\alpha)) = 1 + \limsup_{n\to\infty} \; [a_n; a_{n-1}, \cdots, a_0].$$

Let us remark that, up to a translation, the set $\mathcal{M} = \{\mu(S_b(\alpha)), \;\; \alpha \notin \mathbb{Q}\}$ is equal to the Cassaigne spectrum (see Remark 4.4). As a consequence, we always have that $\mu(S_b(\alpha)) \geq \frac{3+\sqrt{5}}{2} > 2$. In virtue of Roth's theorem, $S_b(\alpha)$ is thus transcendental.

Proof (of Theorem 6.2) We keep the notations of Theorem 6.1. For any nonnegative integer, let us denote by P_n/Q_n the n-th convergent to $S_b(\alpha)$. By Theorem 6.1, we know that $(Q_n)_{n\geq 0}$ is the sequence defined by

$$Q_0 = 1, \; Q_1 = 1, \;\; \text{and for } n \geq 2, \;\; Q_{n+1} = t_{n+1}Q_n + Q_{n-1}.$$

We first observe that, for any nonnegative integer n, $Q_n = (b^{q_n} - 1)/(b-1)$. Namely, for $n = 0$ and $n = 1$, this follows from $q_0 = 0$ and $q_1 = 1$. For $n \geq 2$, this is implied by

$$\begin{aligned} t_{n+1}\left(\frac{b^{q_n}-1}{b-1}\right) + \frac{b^{q_{n-1}}-1}{b-1} &= \left(\frac{b^{q_{n+1}}-b^{q_{n-1}}}{b^{q_n}-1}\right)\left(\frac{b^{q_n}-1}{b-1}\right) + \frac{b^{q_{n-1}}-1}{b-1} \\ &= \frac{b^{q_{n+1}}-b^{q_{n-1}}}{b-1} + \frac{b^{q_{n-1}}-1}{b-1} = \frac{b^{q_{n+1}}-1}{b-1}\cdot \end{aligned}$$

On the other hand, the theory of continued fractions gives

$$\frac{1}{2Q_nQ_{n+1}} < \left| S_b(\alpha) - \frac{P_n}{Q_n} \right| < \frac{1}{Q_nQ_{n+1}}. \tag{6.1}$$

This can be expressed as follows

$$\frac{1}{Q_n^{1+(\log Q_{n+1}/\log Q_n)+(\log 2/\log Q_n)}} < \left| S_b(\alpha) - \frac{P_n}{Q_n} \right| < \frac{1}{Q_n^{1+(\log Q_{n+1}/\log Q_n)}}.$$

Furthermore, we have that $\log Q_n = \log(b^{q_n} - 1) - \log(b-1)$, which implies

$$\limsup_{n\to\infty} \frac{\log Q_{n+1}}{\log Q_n} = \limsup_{n\to\infty} \frac{q_{n+1}}{q_n}.$$

We thus can precisely estimate the quality of approximations of $S_b(\alpha)$ by the rationals P_n/Q_n. From (6.1) and the mirror formula, we deduce that the inequality

$$\left| S_b(\alpha) - \frac{P_n}{Q_n} \right| < \frac{1}{Q_n^\tau}$$

has infinitely many solutions as soon as

$$\tau < 1 + \limsup_{n\to\infty}[a_n; a_{n-1}, \cdots, a_0],$$

whereas it has only finitely many solutions if

$$\tau > 1 + \limsup_{n\to\infty}[a_n; a_{n-1}, \cdots, a_0].$$

Since the rationals P_n/Q_n are by definition the best rational approximations to $S_b(\alpha)$, we get that $\mu(S_b(\alpha)) = 1 + \limsup_{n\to\infty}[a_n; a_{n-1}, \cdots, a_0]$, concluding the proof. □

7 Simultaneous approximation for a number and its square

The study of approximations to a real number by algebraic numbers of bounded degree was initiated in 1960 by Wirsing [70]. He proved that if n is an integer at least equal to 2 and if ξ is not an algebraic number of degree at most n, there are infinitely many algebraic numbers α of degree at most n satisfying

$$|\xi - \alpha| \ll H(\alpha)^{-(n+3)/2} \tag{7.1}$$

where $H(\alpha)$ denotes the height of α. (Recall that the *height* of an algebraic number is defined as the sum of its degree and of the absolute values of the coefficients of its minimal polynomial with relatively prime integer coefficients.) The constant implied by the notation $\ll$ depends here on n and ξ. A famous conjecture, due to Wirsing [70] and generally referred as Wirsing's conjecture, claims that the right exponent in (7.1) is equal to $n+1$ instead of $(n+3)/2$. Up to now, the Wirsing conjecture is only known to be true for $n = 2$; this is a result of Davenport and Schmidt [24].

In 1969, Davenport and Schmidt [25] investigated the similar question where algebraic numbers are replaced by algebraic integers. In the rest of this section, we will focus on the approximation to a real number by cubic integers, i.e., on a question related to the case $n = 3$ in (7.1). In this direction, Davenport and Schmidt [25] proved the following result.

Theorem 7.1 (Davenport and Schmidt) *Let $\gamma = \frac{1+\sqrt{5}}{2}$. Let ξ be a real number that is neither rational nor quadratic. Then, there exist a positive constant c_1 and infinitely many algebraic integers α of degree at most 3 such that*

$$|\xi - \alpha| \leq c_1 H(\alpha)^{-\gamma^2},$$

where $H(\alpha)$ denotes the height of the algebraic number α.

By duality, approximation to a real number by algebraic numbers of bounded degree is also intimately connected with simultaneous uniform rational approximation to successive powers of a real number. In particular, approximation to a real number ξ by algebraic cubic integers is related to simultaneous uniform rational approximation to ξ and ξ^2, and the authors of [25] derive Theorem 7.1 from the following result.

Theorem 7.2 (Davenport and Schmidt) *Let* $\gamma = \frac{1+\sqrt{5}}{2}$. *Let* ξ *be a real number that is neither rational nor quadratic. Then, there exist a positive constant* c_2 *and arbitrarily large values of* X *such that the inequalities*

$$|x_0| \leq X,\ |x_0\xi - x_1| \leq c_2 X^{-1/\gamma},\ |x_0\xi^2 - x_2| \leq c_2 X^{-1/\gamma},$$

do not have any nonzero solution $(x_0, x_1, x_2) \in \mathbb{Z}^3$.

For a long time it was believed that the value γ^2 in Theorem 7.1 could be improved to 3. This is however not true, as recently discovered by Roy [59]. Actually, Roy proves that the value γ^2 in Theorem 7.1 is optimal.

Theorem 7.3 (Roy) *Let* $\gamma = \frac{1+\sqrt{5}}{2}$. *There exist a positive constant* c_3 *and a real number* ξ *that is neither rational nor quadratic, such that for any algebraic integer* α *of degree at most* 3, *we have*

$$|\xi - \alpha| \geq c_3 H(\alpha)^{-\gamma^2}.$$

To obtain this result, Roy [58,60] proves that the value γ is in fact optimal in Theorem 7.2, against the natural conjecture that the value γ could be improved to 2.

Theorem 7.4 (Roy) *Let* $\gamma = \frac{1+\sqrt{5}}{2}$. *There exist a positive constant* c_4 *and a real number* ξ *that is neither rational nor quadratic, such that the inequalities*

$$|x_0| \leq X,\ |x_0\xi - x_1| \leq c_4 X^{-1/\gamma},\ |x_0\xi^2 - x_2| \leq c_4 X^{-1/\gamma},$$

have a nonzero solution $(x_0, x_1, x_2) \in \mathbb{Z}^3$ *for any real number* $X > 1$.

Following Roy, a real number satisfying the exceptional Diophantine conditions of Theorem 7.4 is called an extremal number. It is proved in [60] that the set of extremal numbers is countable. Surprisingly, Roy provides the following "natural" example of an extremal real number. Let a and b be two distinct positive integers. Let

$$\xi := [a; b, a, a, b, a, b, a, a, b, \cdots],$$

where $abaababaab\cdots$ denotes the Fibonacci word over the alphabet $\{a, b\}$ (see Definition 2.7 in Part I). Then, Roy proves [60] that ξ is an extremal number.

Of course the attractive work of Roy lead to many stimulating questions. For a real number ξ we define, following [14], the exponent $\hat{\lambda}_2(\xi)$ as the supremum of the real numbers λ such that the inequalities

$$|x_0| \leq X,\ |x_0\xi - x_1| \leq c_4 X^{-1/\lambda},\ |x_0\xi^2 - x_2| \leq c_4 X^{-1/\lambda},$$

have a nonzero solution $(x_0, x_1, x_2) \in \mathbb{Z}^3$ for any large enough real number X. Bugeaud and Laurent [14] showed how to use Roy's construction to provide explicit real numbers for which the exponent $\hat{\lambda}_2$ takes values beetwen 2 (the expected value) and γ (the optimal value).

Theorem 7.5 (Bugeaud and Laurent) *Let m and n be two distinct positive integers. Let $\alpha := [0; a_1, a_2, \cdots]$ be an irrational real number and let $(b_n)_{n\geq 1}$ be the characteristic Sturmian sequence of slope α defined on the alphabet $\{m, n\}$. Let ξ be the non-quadratic and irrational real number defined by*

$$\xi := [0; b_1, b_2, \cdots].$$

Then,

$$\hat{\lambda}_2(\xi) = \frac{\sigma + 1}{2\sigma + 1},$$

where $\sigma = \limsup_{n\to\infty} [a_n, a_{n-1}, \cdots, a_1]$.

We end this section with a focus on the main steps of the proof of Theorem 7.3. Our purpose is principally to make clear the central rôle played here by the mirror formula. In this respect, our presentation is quite far from the one of the original proof in [60].

Proof (of Theorem 7.3) The proof can essentially be divided into three steps. In the first and more important step, we show how continued fractions can be used for finding simultaneous rational approximations to a real number and its square, *via* palindromes.

Let $\xi = [0; a_1, a_2, \cdots]$ be a positive real irrational number, and denote by p_n/q_n its convergents, i.e., $p_n/q_n = [0; a_1, \cdots, a_n]$. If the word $a_1 \cdots a_n$ is a palindrome, then the mirror formula implies that

$$\frac{q_{n-1}}{q_n} = [0; a_n, a_{n-1}, \cdots, a_1] = [0; a_1, \cdots, a_n] = \frac{p_n}{q_n}.$$

In this case, we have $p_n = q_{n-1}$. By the theory of continued fractions, we get

$$\left|\xi - \frac{p_n}{q_n}\right| < \frac{1}{q_n^2} \quad\text{and}\quad \left|\xi - \frac{p_{n-1}}{q_{n-1}}\right| < \frac{1}{q_{n-1}^2}.$$

We then infer from $0 < \xi < 1$, $a_1 = a_n$ and $q_n \leq (a_n + 1)q_{n-1}$ that

$$\left|\xi^2 - \frac{p_{n-1}}{q_n}\right| \quad \leq \left|\xi^2 - \frac{p_{n-1}}{q_{n-1}} \times \frac{p_n}{q_n}\right| \leq \left|\xi + \frac{p_{n-1}}{q_{n-1}}\right| \times \left|\xi - \frac{p_n}{q_n}\right| + \frac{1}{q_n q_{n-1}}$$

$$\leq 2\left|\xi - \frac{p_n}{q_n}\right| + \frac{1}{q_n q_{n-1}} < \frac{a_1+3}{q_n^2}.$$

Consequently, if the word $a_1a_2\cdots a_n$ is a palindrome, then

$$|q_n\xi - p_n| < \frac{1}{q_n} \quad \text{and} \quad |q_n\xi^2 - p_{n-1}| < \frac{a_1+3}{q_n}. \tag{7.2}$$

Let us summarize our first step. Each time we will find a palindromic convergent p_n/q_n to the real ξ (i.e., $p_n/q_n = [0; a_1, \cdots, a_n]$ and $a_1 \cdots a_n$ is a palindrome), this will provide very good simultaneous rational approximations to ξ and ξ^2, respectively given by p_n/q_n and p_{n-1}/q_n.

An important feature of the problem we are studying is that we have to prove a uniform statement, that is, it deals with uniform simultaneous rational approximation. In the second step, which will now appear as very natural, we show how the palindrome density of the continued fraction expansion of a real ξ is related to such a uniform statement.

First, let us assume that the infinite word $\mathbf{a} = a_1a_2\cdots a_n\cdots$ begins in infinitely many palindromes. We will use the notation introduced in Section 5. We thus denote by $(n_i)_{i\geq 1}$ the increasing sequence of all lengths of palindromic prefixes of $\mathbf{a}$, and by $d_p(\mathbf{a})$ the palindrome density of the word $\mathbf{a}$. Let us assume that the palindrome density of $\mathbf{a}$ is large enough to ensure that $q_{n_i+1} \leq cq_{n_i}^{\tau}$, for some real number τ larger than one and for a positive constant c independent of i. Then, it easily follows from (7.2) that for any real number $X > 1$, the inequalities

$$|x_0| \leq X, \ |x_0\xi - x_1| \leq cX^{-1/\tau}, \ |x_0\xi^2 - x_2| \leq cX^{-1/\tau}, \tag{7.3}$$

have a nonzero solution $(x_0, x_1, x_2) \in \mathbb{Z}^3$. Indeed, given X there always exists a positive integer n such that $q_n \leq X < q_{n+1}$, and the triple (q_n, p_n, p_{n-1}) is a nonzero solution for (7.3).

Thus, if ξ is a real number whose continued fraction expansion begins in many palindromes, then ξ and ξ^2 are uniformly and simultaneously well approximated by rationals. In view of Section 5, the Fibonacci continued fraction thus appears as a natural candidate for our problem. This ends our second step.

From now on, we assume that $\mathbf{a} = abaab\cdots$ denotes the Fibonacci word over the alphabet $\{a, b\}$ and that $\xi := [a; b, a, a, b, \cdots]$. We want to prove that ξ is an extremal number. We thus have first to estimate the growth of the sequence $(n_i)_{i\geq 1}$. Actually, the value of n_i can be computed exactly (see Section 5 for a proof of this result) and we get that

$$n_i = F_{i+1} - 2, \tag{7.4}$$

where F_i denotes the i-th Fibonacci number.

To end the proof, it now suffices (in view of (7.3)) to prove that there exists a positive constant c independent of i such that

$$q_{n_{i+1}} \le c q_{n_i}^{\gamma}, \tag{7.5}$$

where $\gamma = \frac{1+\sqrt{5}}{2}$ denotes as previously the golden ratio. Lemma 10.2 and Equality (7.4) imply that

$$c_5 < \frac{q_{n_{i+1}}}{q_{n_i} q_{n_{i-1}}} < c_6,$$

for any $i \ge 2$ and for some positive constants c_5 and c_6. We set

$$c_7 = \max\left\{ c_6, \frac{(c_5 q_{n_1})^{\gamma}}{q_{n_2}}, \frac{(c_5 q_{n_2})^{1/\gamma}}{q_{n_1}} \right\}.$$

Since $c_7 \ge c_6$, we obviously get that

$$c_5 < \frac{q_{n_{i+1}}}{q_{n_i} q_{n_{i-1}}} < c_7. \tag{7.6}$$

We set $c_8 = c_5^{\gamma}/c_7$ and $c_9 = c_7^{\gamma}/c_5$, and we are now going to prove by induction on i that

$$c_8 q_{n_i}^{\gamma} \le q_{n_{i+1}} \le c_9 q_{n_i}^{\gamma} \tag{7.7}$$

holds for any $i \ge 2$. For $i = 2$, this follows from (7.6) and from the definition of c_7. Let us assume that (7.7) holds for a fixed integer $i \ge 2$. By (7.6), we have

$$c_5 q_{n_i}^{\gamma} \left(q_{n_i}^{1-\gamma} q_{n_{i-1}} \right) < q_{n_{i+1}} < c_7 q_{n_i}^{\gamma} \left(q_{n_i}^{1-\gamma} q_{n_{i-1}} \right)$$

and since $\gamma(\gamma - 1) = 1$, we obtain

$$c_5 q_{n_i}^{\gamma} \left(q_{n_i} q_{n_{i-1}}^{-\gamma} \right)^{1-\gamma} < q_{n_{i+1}} < c_7 q_{n_i}^{\gamma} \left(q_{n_i} q_{n_{i-1}}^{-\gamma} \right)^{1-\gamma}.$$

We thus deduce from (7.7) that

$$\left(c_5 c_9^{1-\gamma} \right) q_{n_i}^{\gamma} < q_{n_{i+1}} < \left(c_8 c_5^{1-\gamma} \right) q_{n_i}^{\gamma}.$$

By definition of c_8 and c_9, and since $\gamma(\gamma - 1) = 1$, this gives

$$c_8 q_{n_i}^{\gamma} < q_{n_{i+1}} < c_9 q_{n_i}^{\gamma}.$$

We thus have shown that (7.7) holds for any integer $i \ge 2$. In virtue of (7.5) and (7.3), ξ is an extremal number, which concludes the proof of Theorem 7.3. □

8 The Littlewood conjecture

It follows from the theory of continued fractions that, for any real number α, there exist infinitely many positive integers q such that

$$q \cdot \|q\alpha\| < 1, \tag{8.1}$$

where $\|\cdot\|$ denotes the distance to the nearest integer. In particular, for any given pair (α, β) of real numbers, there exist infinitely many positive integers q such that

$$q \cdot \|q\alpha\| \cdot \|q\beta\| < 1.$$

A famous open problem in simultaneous Diophantine approximation, called the Littlewood conjecture (see for example [42]), claims that in fact, for any given pair (α, β) of real numbers, a stronger result holds.

Littlewood's conjecture. *For any given pair (α, β) of real numbers,*

$$\inf_{q \geq 1} q \cdot \|q\alpha\| \cdot \|q\beta\| = 0. \tag{8.2}$$

Let us denote by **Bad** the set of badly approximable numbers, i.e.,

$$\mathbf{Bad} := \{\alpha \in \mathbb{R} : \inf_{q \geq 1} q \cdot \|q\alpha\| > 0\}.$$

The set **Bad** is intimately connected with the theory of continued fractions. Indeed, a real number lies in **Bad** if, and only if, it has bounded partial quotients in its continued fraction expansion. It then follows that the Littlewood conjecture holds true for the pair (α, β) if α or β has unbounded partial quotients in its continued fraction expansion. It also holds when the numbers 1, α, and β are linearly dependent over the rational integers, as follows from (8.1).

The first significant contribution towards the Littlewood conjecture goes back to Cassels and Swinnerton-Dyer [20] who showed that (8.2) holds when α and β belong to the same cubic field. However, since it is still not known whether cubic real numbers have bounded partial quotients or not (see the discussion at the beginning of Section 9), their result does not yield examples of pairs of badly approximable real numbers for which the Littlewood conjecture holds.

In view of the above discussion, it is natural to restrict our attention to independent parameters α and β, both lying in **Bad**. This naturally leads to considering the following problem:

Question 8.1 *Given α in* **Bad***, is there any independent β in* **Bad** *so that the Littlewood conjecture is true for the pair (α, β)?*

Apparently, Question 8.1 remained unsolved until 2000. It has then been answered positively by Pollington and Velani [49], who established the following stronger result.

Theorem 8.2 (Pollington and Velani) *Given α in* **Bad***, there exists a subset $A(\alpha)$ of* **Bad** *with Hausdorff dimension one, such that, for any β in $A(\alpha)$, there exist infinitely many positive integers q with*

$$q \cdot \|q\alpha\| \cdot \|q\beta\| \leq \frac{1}{\log q}. \tag{8.3}$$

In particular, the Littlewood conjecture holds for the pair (α, β) for any β in $A(\alpha)$.

The proof of this result depends on sophisticated tools from metric number theory. At the end of [49], Pollington and Velani give an alternative proof of a weaker version of Theorem 8.2, namely with (8.3) replaced by (8.2). However, even for establishing this weaker version, deep tools from metric number theory are still needed, including in particular a result of Davenport, Erdős and LeVeque on uniform distribution [23] and the *Kaufman measure* constructed in [33].

Very recently, Einsiedler, Katok and Lindenstrauss [30] proved the following remarkable statement.

Theorem 8.3 (Einsideler, Katok and Lindenstrauss) *The set of pairs of real numbers for which the Littlewood conjecture does not hold has Hausdorff dimension zero.*

Obviously, this gives a positive answer to Question 1. Actually, the authors established part of the Margulis conjecture on ergodic actions on the homogeneous space $SL_k(\mathbb{R})/SL_k(\mathbb{Z})$, for $k \geq 3$ (see [44]). It was previously known that such a result would have implications to Diophantine questions, including the Littlewood conjecture. Their sophisticated proof used, among others, deep tools from algebra and from the theory of dynamical systems, involving in particular the important work of Ratner (see for example [56]).

De Mathan gave in [45] an explicit construction of pairs of real numbers (α, β) with bounded partial quotients, such that $1, \alpha, \beta$ are linearly independent over the rationals and satisfy Littlewood's conjecture. We do not resist to give a particular case of De Mathan's result ([45, p. 264]). Recall that the Thue-Morse sequence on the alphabet $\{1, 2\}$ is the sequence $(a_n)_{n \geq 0}$ defined by $a_n = 1$ (resp. $a_n = 2$) if the sum of the binary digits of n is even (resp. odd).

Theorem 8.4 (De Mathan) *Let $A := [1; 2, 2, 1, 2, 1, 1, 2, 2, \cdots]$ be the real number whose continued fraction expansion is the Thue-Morse sequence on the alphabet $\{1, 2\}$. Then, $1, A, 1/A$ are linearly independent on the rationals, and $(A, 1/A)$ satisfies the Littlewood conjecture.*

We are now going to show how our favorite formula allows the authors of [3] to provide a short and elementary positive answer to Question 8.1, and even a stronger form of it. Their approach rests on the basic theory of continued fractions.

Theorem 8.5 (Adamczewski and Bugeaud) *Let φ be a positive and non-increasing function defined on the set of positive integers such that $\varphi(1) = 1$, $\lim_{q\to+\infty} \varphi(q) = 0$ and $\lim_{q\to+\infty} q\varphi(q) = +\infty$. Given α in* **Bad***, there exists an uncountable subset $B_\varphi(\alpha)$ of* **Bad** *such that, for any β in $B_\varphi(\alpha)$, there exist infinitely many positive integers q with*

$$q \cdot \|q\alpha\| \cdot \|q\beta\| \leq \frac{1}{q \cdot \varphi(q)}. \tag{8.4}$$

In particular, the Littlewood conjecture holds for the pair (α, β) for any β in $B_\varphi(\alpha)$. Furthermore, the set $B_\varphi(\alpha)$ can be effectively constructed.

It is of interest to compare this result with Theorem 8.2. Regarding the Littlewood conjecture, Theorem 8.2 is stronger since the set $A(\alpha)$ has Hausdorff dimension one whereas the set $B_\varphi(\alpha)$ has only the power of the continuum. On the other hand, one can remark that the Diophantine property in Theorem 8.5 is really stronger than the one of Theorem 8.2. In particular, one can doubt on the truth of a statement analogous to Theorem 8.2, with the Diophantine condition of Theorem 8.5. However that may be, the main interest of Theorem 8.5 surely is that the proof is elementary and gives a generic way to provide explicit examples for the Littlewood conjecture.

Proof (of Theorem 8.5) Let $\alpha = [0; a_1, a_2, \cdots, a_k, \cdots]$ be in **Bad**, and let M be an integer at least equal to 2 such that $a_k \leq M$, for any positive integer k. We first construct inductively a rapidly increasing sequence $(n_j)_{j\geq 1}$ of positive integers. We set $n_1 = 1$ and we proceed with the inductive step. Assume that $j \geq 2$ is such that $n_1, \ldots, n_{j-1}$ have been constructed. Then, we choose n_j sufficiently large in order that

$$\varphi(2^{(m_j-1)/2}) \leq \frac{1}{4} \cdot \left(\frac{1}{(M+3)^{m_{j-1}+1}}\right)^2, \tag{8.5}$$

where $m_j = n_1 + n_2 + \cdots + n_j + (j-1)$. Such a choice is always possible since φ tends to zero at infinity and since the right hand side of (8.5) only depends on $n_1, n_2, \ldots, n_{j-1}$.

Our sequence $(n_j)_{j\geq 1}$ being now constructed, for an arbitrary integer sequence $\mathbf{t} = (t_k)_{k\geq 1}$ with values in $\{M+1, M+2\}$, we set

$$\begin{aligned} \beta_{\mathbf{t}} = \quad & [0; b_1, b_2, \cdots] \\ = \quad & [0; a_{n_1}, \cdots, a_1, t_1, a_{n_2}, \cdots, a_1, t_2, a_{n_3}, \cdots, a_1, \cdots, a_1, t_{j-1}, a_{n_j}, \cdots]. \end{aligned}$$

Then, we introduce the set

$$B_\varphi(\alpha) = \left\{\beta_{\mathbf{t}},\ \mathbf{t} \in \{M+1, M+2\}^{\mathbb{Z}_{\geq 1}}\right\}.$$

Clearly, the set $B_\varphi(\alpha)$ has the power of the continuum.

Let $\beta_{\mathbf{t}} = \beta$ lie in $B_\varphi(\alpha)$. It remains to prove that (8.4) holds for the pair (α, β). Denote by $(p_j/q_j)_{j\geq 1}$ (resp. by $(r_j/s_j)_{j\geq 1}$) the sequence of convergents to α (resp. to β).

The key point in this proof is again a right use of the mirror formula. Namely, it gives

$$\frac{s_{m_j-1}}{s_{m_j}} = [0; a_1, \cdots, a_{n_j}, t_{j-1}, a_1, \cdots, a_{n_{j-1}}, t_{j-2}, \cdots, t_1, a_1, \cdots, a_{n_1}],$$

which implies that $\|s_{m_j}\alpha\|$ is small. More precisely, using Lemma 10.1, we obtain:

$$\|s_{m_j}\alpha\| \leq s_{m_j}\, q_{n_j}^{-2}.$$

On the other hand, we have

$$\|s_{m_j}\beta\| \leq \frac{1}{s_{m_j}},$$

as follows from the theory of continued fractions. We thus derive that

$$s_{m_j} \cdot \|s_{m_j}\alpha\| \cdot \|s_{m_j}\beta\| \leq s_{m_j}\, q_{n_j}^{-2}.$$

In order to satisfy (8.4), it is enough to have

$$s_{m_j}\, q_{n_j}^{-2} \leq s_{m_j}^{-1}\, \varphi(s_{m_j})^{-1},$$

that is,

$$s_{m_j}^2\, \varphi(s_{m_j}) \leq q_{n_j}^2.$$

We infer from Lemma 10.2 that

$$s_{m_j} \leq 2\, K_{m_j-n_j}(b_1, \ldots, b_{m_j-n_j})\, K_{n_j}(b_{m_j-n_j+1}, \ldots, b_{m_j}),$$

and

$$K_{n_j}(b_{m_j-n_j+1}, \ldots, b_{m_j}) = K_{n_j}(a_1, \ldots, a_{n_j}) = q_{n_j}.$$

Consequently, (8.4) holds as soon as

$$4\,\varphi(s_{m_j}) \leq K_{m_j-n_j}(b_1, \ldots, b_{m_j-n_j})^{-2} = K_{m_{j-1}+1}(b_1, \ldots, b_{m_{j-1}+1})^{-2}. \tag{8.6}$$

Since the partial quotients of β are bounded by $M+2$, we obtain by Lemma 10.3

$$K_{m_{j-1}+1}(b_1, \ldots, b_{m_{j-1}+1}) < (M+3)^{m_{j-1}+1}. \tag{8.7}$$

On the other hand, Lemma 10.3 also implies that

$$s_{m_j} \geq \sqrt{2}^{(m_j-1)} \tag{8.8}$$

It thus follows from (8.5) that (8.6) holds, since φ is a non-increasing function. This completes the proof. □

9 Transcendental continued fractions

It is widely believed (the question was first asked by Khintchine [35] cited in [62]) that the continued fraction expansion of any irrational algebraic number α is either eventually periodic (and this is the case if and only if α is a quadratic irrational) or it contains arbitrarily large partial quotients; but we seem to be very far away from a proof (or a disproof). A first step consists in providing explicit examples of transcendental continued fractions. The first result of this type is due to Liouville [41], who constructed real numbers whose sequence of partial quotients grows very fast, too fast for the numbers to be algebraic. Subsequently, various authors used deeper transcendence criteria from Diophantine approximation to construct other classes of transcendental continued fractions. Of particular interest is the work of Maillet [43] (see also Section 34 of Perron [48]), who was the first to give examples of transcendental continued fractions with bounded partial quotients. Further examples were provided by Baker [9, 10], Shallit [61], Davison [27], M. Queffélec [55], Allouche, Davison, Queffélec and Zamboni [8] and Adamczewski and Bugeaud [1], among others. Note that the folding lemma is used by Shallit in [61] (see also Section 11).

In the previous two sections, we have shown how the mirror formula can be used to find simultaneous rational approximations for some real numbers. On the other hand, algebraic numbers cannot be "too well" simultaneously approximated by rationals. This is a multi-dimensional Roth's principle (see Theorem 9.1 below). Such considerations give naturally rise to transcendence statements, as we will see in this section. We will first be interested in a familly of "quasi-periodic" continued fractions introduced by Maillet [43] and studied later by Baker in [9] and [10]. Then, we will investigate real numbers whose sequence of partial quotients enjoys another combinatorial property, namely is "symmetrical", in the sense that it begins in arbitrarily long palindromes or quasi-palindromes.

The transcendence criteria presented in this section rest on the powerful Subspace Schmidt Theorem [64] (see also [65]) that we state now, as well as on a heavy use of the mirror formula.

Theorem 9.1 (W. M. Schmidt) *Let $m \geq 2$ be an integer. Let $L_1, \ldots, L_m$ be linearly independent linear forms in $\mathbf{x} = (x_1, \ldots, x_m)$ with algebraic coefficients. Let ε be a positive real number. Then, the set of solutions $\mathbf{x} = (x_1, \ldots, x_m)$ in $\mathbb{Z}^m$ to the inequality*

$$|L_1(\mathbf{x}) \ldots L_m(\mathbf{x})| \leq (\max\{|x_1|, \ldots, |x_m|\})^{-\varepsilon}$$

lies in the union of finitely many proper subspaces of $\mathbb{Q}^m$.

As an example of a by-product of the Subspace Theorem, we mention a result concerning the simultaneous rational approximation of a real number and its square. It was originally proved in [63].

Theorem 9.2 (W. M. Schmidt) *Let ξ be a real number, which is neither rational, nor quadratic. If there exist a real number $w > 3/2$ and infinitely many triples of integers (p, q, r) such that*

$$\max\left\{\left|\xi - \frac{p}{q}\right|, \left|\xi^2 - \frac{r}{q}\right|\right\} < \frac{1}{|q|^w},$$

then ξ is transcendental.

A direct consequence of this result is that the extremal numbers considered in Section 7 are transcendental. The following dual form of Theorem 9.2, also proved in [63], which limits the approximation of an algebraic non-quadratic number by quadratic numbers, can also be derived from the Suspace Theorem.

Theorem 9.3 (Schmidt) *Let ξ be a real number, which is neither rational, nor quadratic. If there exist a real number $w > 3$ and infinitely many quadratic numbers α such that*

$$|\xi - \alpha| < H(\alpha)^{-w},$$

then ξ is transcendental (as previously $H(\alpha)$ is the height of α).

9.1 Maillet-Baker's continued fractions

As already mentioned, the first examples of transcendental real numbers with bounded partial quotients were constructed by Maillet [43]. More precisely, Maillet proved that if $\mathbf{a} = (a_n)_{n\geq 1}$ is a non-eventually periodic sequence of positive integers, and if there are infinitely many positive integers n such that

$$a_n = a_{n+1} = \ldots = a_{n+\lambda(n)},$$

then the real number $\xi = [0; a_1, a_2, \cdots]$ is transcendental provided that $\lambda(n)$ is larger than a certain function of q_n, the denominator of the nth convergent to ξ. Actually, the result of Maillet is more general and also includes the case of repetitions of a block of consecutive partial quotients. His proof is based on a general form of the Liouville inequality which limits the approximation of algebraic numbers by quadratic irrationals. Indeed, under the previous assumption, the quadratic irrational real numbers $\xi_n = [0; a_1, a_2, \cdots, a_{n-1}, a_n, a_n, \cdots, a_n, \cdots]$ are “too good” approximations to ξ.

It is not very surprising that the breaktrough made by Roth [57] in 1955 lead to an improvement of this result. Thus, Baker [9] used in 1962 the Roth theorem for number fields obtained by LeVeque [38] to strongly improve the results of Maillet and make them more explicit. His main idea was to see that if the quadratic approximations found by Maillet lie in a same quadratic number field, then one can favorably replace the use of the Liouville inequality by the Roth theorem for number fields. In particular, Baker proved the following result.

Theorem 9.4 (Baker) *Let A be a positive integer. Let $\mathbf{a} = (a_n)_{n\geq 1}$ be a non-eventually periodic sequence of positive integers, all bounded by A. Let $(n_k)_{k\geq 1}$ be an increasing sequence of positive integers and let $(\lambda_k)_{k\geq 1}$ be a sequence of positive integers satisfying*

$$\limsup_{k\to\infty} \frac{\lambda_k}{n_k} > B = B(A),$$

where B is defined by

$$B = 2\left(\frac{\log\left(\left(A+\sqrt{A^2+4}\right)/2\right)}{\log\left((1+\sqrt{5})/2\right)}\right) - 1.$$

Let us assume that for any positive integer k, we have

$$a_{n_k} = a_{n_{k+1}} = \ldots = a_{n_k+\lambda_k}.$$

Then, the real number $\xi := [a_0; a_1, a_2, \cdots]$ is transcendental.

In order to improve this result, it is quite tempting to apply the Subspace Theorem instead of LeVeque's theorem. We first mention that a direct use of Theorem 9.3 in Baker's approach leads to a weaker form of Theorem 9.4. It seems however possible to reach a smaller bound than the one obtained in Theorem 9.4 by using the quadratic approximations previously considered by Maillet or by Baker and the ideas of [1], see [28]. In [4], the authors show how the method introduced in Section 9.2 can also be used to improve Theorem 9.4 in some particular cases.

Quite surprisingly, a tricky use of the Subspace Theorem based on the mirror formula allows to considerably relax the transcendence criterion obtained by Baker. In particular, the following result, obtained in [2], does not depend on the values of the partial quotients of the real number under consideration.

Theorem 9.5 (Adamczewski and Bugeaud) *Under the assumptions of Theorem 9.4, the real number ξ is transcendental if*

$$\limsup_{k\to\infty} \frac{\lambda_k}{n_k} > 0. \tag{9.1}$$

Proof By assumption, for any positive integer k, the following equalities hold: $a_{n_k} = a_{n_{k+1}} = \ldots = a_{n_k+\lambda_k}$. Let us denote by $a(k)$ the positive integer satisfying $a(k) = a_{n_k}$. Since the partial quotients of ξ are bounded, the pigeon-hole principle implies that infinitely many of the $a(k)$ take the same value that we denote by a.

Without loss of generality, we thus assume that $a_{n_k} = a_{n_{k+1}} = \ldots = a_{n_k+\lambda_k} = a$, for any positive integer k. Now, let us introduce the real $\alpha = [a; a, a, a, \cdots]$, whose partial quotients are all equal to a. Then, α is a quadratic number, root of the polynomial $X^2 - aX - 1$. For the reader's convenience, we introduce some more notation. Let us denote by p_n/q_n (resp. by r_n/s_n) the n-th convergent to ξ (resp. to α). Then, we set $P_k = p_{n_k+\lambda_k}$, $Q_k = q_{n_k+\lambda_k}$, $P'_k = p_{n_k+\lambda_k-1}$, $Q'_k = q_{n_k+\lambda_k-1}$ and $S_k = s_{\lambda_k}$.

By assumption, we already know that ξ is irrational and not quadratic. Therefore, we assume that ξ is algebraic and we aim at deriving a contradiction.

By the theory of continued fractions, we have

$$|Q_k\xi - P_k| < \frac{1}{Q_k} \text{ and } |Q'_k\xi - P'_k| < \frac{1}{Q'_k}. \tag{9.2}$$

On the other hand, since by assumption

$$\frac{P_k}{Q_k} = [a_0; a_1, \cdots, a_{n_k-1}, a, a, \cdots, a],$$

we get from the mirror formula that

$$\frac{Q_k}{Q'_k} = [a; a, a, \cdots, a, a_{n_k-1}, \cdots, a_0].$$

Then, Lemma 10.1 implies

$$\left|Q'_k\alpha - Q_k\right| < \frac{Q'_k}{S_k^2}. \tag{9.3}$$

Consider now the four linearly independent linear forms with algebraic coefficients:

$$\begin{aligned} L_1(X_1, X_2, X_3, X_4) &= \xi X_1 - X_3, \\ L_2(X_1, X_2, X_3, X_4) &= \xi X_2 - X_4, \\ L_3(X_1, X_2, X_3, X_4) &= \alpha X_2 - X_1, \\ L_4(X_1, X_2, X_3, X_4) &= X_1. \end{aligned}$$

Evaluating them on the quadruple (Q_k, Q'_k, P_k, P'_k), it follows from (9.2) and (9.3) that

$$\prod_{1\le j\le 4} |L_j(Q_k, Q'_k, P_k, P'_k)| < \frac{1}{S_k^2}. \tag{9.4}$$

By assumption the a_k are bounded by A and Lemma 10.3 implies that $Q_k \le A^{n_k+\lambda_k}$ for any positive integer k. On the other hand, Lemma 10.3 also gives that $S_k \ge (\sqrt{2})^{\lambda_k - 1}$, for any positive integer k. It thus follows that

$$S_k \ge A^{\left(\frac{\log\sqrt{2}}{\log M}\right)(\lambda_k - 1)},$$

for any positive integer k. We infer from (9.4) and (9.1) that

$$\prod_{1\le j\le 4} |L_j(Q_k, Q'_k, P_k, P'_k)| \le Q_k^{-\varepsilon}$$

holds for some positive real number ε and for k large enough.

It then follows from Theorem 9.1 that the points (Q_k, Q'_k, P_k, P'_k) lie in the union of a finite number of proper subspaces of $\mathbb{Q}^4$. Thus, there exist a nonzero integer quadruple (x_1, x_2, x_3, x_4) and an infinite set of distinct positive integers $\mathcal{N}$ such that

$$x_1 Q_k + x_2 Q'_k + x_3 P_k + x_4 P'_k = 0, \tag{9.5}$$

for any k in $\mathcal{N}$. Dividing (9.5) by Q'_k, we obtain

$$x_1 \frac{Q_k}{Q'_k} + x_2 + x_3 \frac{P_k}{Q_k} \cdot \frac{Q_k}{Q'_k} + x_4 \frac{P'_k}{Q'_k} = 0. \tag{9.6}$$

By letting k tend to infinity along $\mathcal{N}$ in (9.6), we obtain

$$x_1 \alpha + x_2 + (x_3 \alpha + x_4)\xi = 0.$$

Since ξ is not quadratic, it does not in particular lie in $\mathbb{Q}(\alpha)$. This implies that $x_3 \alpha + x_4 = 0$ and, since α is irrational, it follows that $x_3 = x_4 = 0$. Then, $x_1 = x_2 = x_3 = x_4 = 0$, wich is a contradiction. □

9.2 Palindromic continued fractions

A common feature of the results mentioned at the beginning of this section is that they apply to real numbers whose continued fraction expansions are 'quasi-periodic', in the sense that they contain arbitrarily long blocks of partial quotients which occur precociously at least twice. We now consider real numbers whose sequence of partial quotients enjoys another combinatorial property, namely is 'symmetrical', in the sense that it begins in arbitrarily long palindromes or quasi-palindromes. The results stated below are proved in [4] (see also [5]) and rest on the Subspace Theorem.

We first mention the following simple transcendental criterion for palindromic continued fractions.

Theorem 9.6 (Adamczewski and Bugeaud) *Let* $\mathbf{a} = (a_n)_{n\geq 1}$ *be a sequence of positive integers. If the word* $\mathbf{a}$ *begins in arbitrarily long palindromes, then the real number* $\xi := [0; a_1, a_2, \cdots, a_n, \cdots]$ *is either quadratic or transcendental.*

As shown in [5], given two distinct positive integers a and b, Theorem 9.6 easily implies the transcendence of the real number $[0; a_1, a_2, \cdots]$, whose sequence of partial quotients is the Thue-Morse sequence on the alphabet $\{a, b\}$, i.e., with $a_n = a$ (resp. $a_n = b$) if the sum of the binary digits of n is odd (resp. even). This result is originally due to M. Queffélec [55] who used a different approach. We also point out that, quite surprisingly, there is no assumption on the growth of the sequence $(a_n)_{n\geq 0}$ in Theorem 9.6.

Proof (of Theorem 9.6) We have to prove that if ξ is algebraic, then it is quadratic irrational. Clearly, ξ is not rational. Therefore, we assume that ξ is a non-quadratic algebraic number and we aim at deriving a contradiction.

Let us denote by p_k/q_k the k-th convergent to ξ. By assumption there exists an infinite set of positive integers $\mathcal{N}$ such that the word $a_1a_2\cdots a_n$ is a palindrome for any $n \in \mathcal{N}$. By the mirror formula, we get for such an integer n that

$$q_{n-1}/q_n = [0; a_n, a_{n-1}, \cdots, a_1] = [0; a_1, a_2, \cdots, a_n] = p_n/q_n.$$

This implies $p_n = q_{n-1}$. Recalling that

$$|q_n\xi - p_n| < \frac{1}{q_n} \quad \text{and} \quad |q_{n-1}\xi - p_{n-1}| < \frac{1}{q_{n-1}},$$

we thus obtain

$$|q_n\xi - q_{n-1}| < \frac{1}{q_n} \quad \text{and} \quad |q_{n-1}\xi - p_{n-1}| < \frac{1}{q_{n-1}}. \tag{9.7}$$

Consider now the three linearly independent linear forms with algebraic coefficients:

$$\begin{aligned} L_1(X_1, X_2, X_3) = &\ \ \xi X_1 - X_2, \\ L_2(X_1, X_2, X_3) = &\ \ \xi X_2 - X_3, \\ L_3(X_1, X_2, X_3) = &\ \ X_2. \end{aligned}$$

Evaluating them on the quadruple (q_n, q_{n-1}, p_{n-1}), we derive from (9.7) that

$$\prod_{1\le j\le 3} |L_j(q_n, q_{n-1}, p_{n-1})| < \frac{1}{q_n}.$$

It then follows from Theorem 9.1 that the points (q_n, q_{n-1}, p_{n-1}), $n \in \mathcal{N}$, lie in the union of a finite number of proper subspaces of $\mathbb{Q}^4$. Thus, there exist a nonzero integer triple (x_1, x_2, x_3) and an infinite set of distinct positive integers $\mathcal{N}_1 \subset \mathcal{N}$ such that

$$x_1q_n + x_2q_{n-1} + x_3p_{n-1} = 0, \tag{9.8}$$

for any $n \in \mathcal{N}_1$. Dividing (9.8) by q_n and letting n tend to infinity along $\mathcal{N}_1$, it thus follows from $p_n = q_{n-1}$ that

$$x_1 + x_2\frac{1}{\xi} + x_3\xi = 0.$$

Since (x_1, x_2, x_3) is a nonzero triple of integers, we obtain that ξ is either a quadratic or a rational number, hence a contradiction. □

Let us introduce some more notation. As we have already shown, a palindrome is a finite word invariant under mirror symmetry. In order to relax this property of symmetry, we now introduce the notion of quasi-palindrome. Let U and V be two finite words. The word $UV\overline{U}$ is called a quasi-palindrome of order w, where $w = |V|/|U|$. Following this definition, the larger w, the weaker the symmetry property. In particular, a palindrome is a quasi-palindrome of order

0. We now give a transcendence criterion in which occurrences of arbitrarily long palindromes are replaced by occurrences of arbitrarily long quasi-palindromes of a fixed and finite order. Of course, an extra assumption on the growth of the partial quotients is then needed. This assumption is not very restrictive. In particular, it is always satisfied by real numbers with bounded partial quotients.

Let $\mathbf{a} = (a_n)_{n\geq 1}$ be a sequence over $\mathcal{A}$. Let w be a rational number with $w > 1$. We say that $\mathbf{a}$ begins in arbitrarily long quasi-palindromes of finite order if there exist a nonnegative real number w, and two sequences of finite words $(U_n)_{n\geq 1}$ and $(V_n)_{n\geq 1}$ such that:

(i) For any $n \geq 1$, the word $U_n V_n \overline{U_n}$ is a prefix of the word $\mathbf{a}$;

(ii) The sequence $(|V_n|/|U_n|)_{n\geq 1}$ is bounded by w;

(iii) The sequence $(|U_n|)_{n\geq 1}$ is increasing.

Then, Theorem 9.6 can be extended in the following way.

Theorem 9.7 (Adamczewski and Bugeaud) *Let $\mathbf{a} = (a_n)_{n\geq 1}$ be a sequence of positive integers. Let $(p_n/q_n)_{n\geq 1}$ denote the sequence of convergents to the real number*

$$\xi := [0; a_1, a_2, \cdots, a_n, \cdots].$$

Assume that the sequence $(q_\ell^{1/\ell})_{\ell\geq 1}$ is bounded, which is in particular the case when the sequence $\mathbf{a}$ is bounded. If $\mathbf{a}$ begins in arbitrarily long quasi-palindromes of finite order, then ξ is either quadratic or transcendental.

In the statements of Theorems 9.6 and 9.7 the palindromes or the quasi-palindromes must appear at the very beginning of the continued fraction under consideration. We mention that the ideas used in their proofs also allow to deal with the more general situation where arbitrarily long quasi-palindromes occur not too far from the beginning (see [4]).

Proof Keep the notation and the hypothesis of this theorem. Assume that the parameter w is fixed, as well as the sequences $(U_n)_{n\geq 1}$ and $(V_n)_{n\geq 1}$. Set also $r_n = |U_n|$ and $s_n = |U_n V_n \overline{U}_n|$, for any $n \geq 1$. Let us also assume that the sequence $\mathbf{a}$ is not eventually periodic. It thus follows that that the real number

$$\xi := [0; a_1, a_2, \cdots]$$

is neither rational nor quadratic. We want to prove that it is transcendental. Therefore, we assume that ξ is algebraic of degree at least three and we aim at deriving a contradiction.

Let $(p_\ell/q_\ell)_{\ell\geq 1}$ denote the sequence of convergents to ξ. The key fact for the proof of Theorem 9.7 is "of course" the mirror formula. Indeed, if W_ℓ denotes the prefix of length ℓ of the sequence $\mathbf{a}$, then $q_{\ell-1}/q_\ell = [0; \overline{W_\ell}]$. Since, by assumption, we have

$$\frac{p_{s_n}}{q_{s_n}} = [0; U_n V_n \overline{U_n}],$$

we get from the mirror formula that

$$\frac{q_{s_n-1}}{q_{s_n}} = [0; U_n \overline{V_n}\ \overline{U_n}],$$

and it follows from Lemma 10.1 that

$$|q_{s_n}\xi - q_{s_n-1}| < q_{s_n} q_{r_n}^{-2}. \tag{9.9}$$

This shows in particular that

$$\lim_{n\to+\infty} \frac{q_{s_n-1}}{q_{s_n}} = \xi. \tag{9.10}$$

Furthermore, we have

$$|q_{s_n}\xi - p_{s_n}| < q_{s_n}^{-1} \quad \text{and} \quad |q_{s_n-1}\xi - p_{s_n-1}| < q_{s_n-1}^{-1}. \tag{9.11}$$

Consider now the four linearly independent linear forms with algebraic coefficients:

$$\begin{aligned} L_1(X_1, X_2, X_3, X_4) = {} & \xi X_1 - X_3, \\ L_2(X_1, X_2, X_3, X_4) = {} & \xi X_2 - X_4, \\ L_3(X_1, X_2, X_3, X_4) = {} & \xi X_1 - X_2, \\ L_4(X_1, X_2, X_3, X_4) = {} & X_2. \end{aligned}$$

Evaluating them on the quadruple $(q_{s_n}, q_{s_n-1}, p_{s_n}, p_{s_n-1})$, it follows from (9.9) and (9.11) that

$$\prod_{1\le j\le 4} |L_j(q_{s_n}, q_{s_n-1}, p_{s_n}, p_{s_n-1})| < q_{r_n}^{-2}. \tag{9.12}$$

By assumption, there exists a real number M such that

$$q_\ell^{1/\ell} \le M$$

for any positive integer ℓ. Thus, Lemma 10.3 implies that for any positive integer n, we have

$$q_{r_n} \ge \sqrt{2}^{(r_n-1)} \ge (M^{s_n})^{((r_n-1)\log\sqrt{2})/(s_n \log M)} \ge q_{s_n}^{((r_n-1)\log\sqrt{2})/(s_n\log M)}$$

and we infer from (9.12) and from (ii) that

$$\prod_{1\le j\le 4} |L_j(q_{s_n}, q_{s_n-1}, p_{s_n}, p_{s_n-1})| \ll q_{s_n}^{-\varepsilon}$$

holds for some positive real number ε.

It then follows from Theorem 9.1 that the points $(q_{s_n}, q_{s_n-1}, p_{s_n}, p_{s_n-1})$ lie in the union of a finite number of proper subspaces of $\mathbb{Q}^4$. Thus, there exist a nonzero integer quadruple (x_1, x_2, x_3, x_4) and an infinite set of distinct positive integers $\mathcal{N}_1$ such that

$$x_1 q_{s_n} + x_2 q_{s_n-1} + x_3 p_{s_n} + x_4 p_{s_n-1} = 0, \tag{9.13}$$

for any n in $\mathcal{N}_1$. Dividing (9.13) by q_{s_n}, we obtain

$$x_1 + x_2 \frac{q_{s_n-1}}{q_{s_n}} + x_3 \frac{p_{s_n}}{q_{s_n}} + x_4 \frac{p_{s_n-1}}{q_{s_n-1}} \cdot \frac{q_{s_n-1}}{q_{s_n}} = 0. \tag{9.14}$$

By letting n tend to infinity along $\mathcal{N}_1$ in (9.14), it follows from (9.10) that

$$x_1 + (x_2 + x_3)\xi + x_4 \xi^2 = 0.$$

Since, by assumption, ξ is not a quadratic number, we have $x_1 = x_4 = 0$ and $x_2 = -x_3$. Then, (9.13) implies that

$$q_{s_n-1} = p_{s_n}. \tag{9.15}$$

Consider now the three linearly independent linear forms with algebraic coefficients:

$$L'_1(Y_1, Y_2, Y_3) = \xi Y_1 - Y_2, \quad L'_2(Y_1, Y_2, Y_3) = \xi Y_2 - Y_3, \quad L'_3(Y_1, Y_2, Y_3) = Y_1.$$

Evaluating them on the triple $(q_{s_n}, p_{s_n}, p_{s_n-1})$, we infer from (9.11) and (9.15) that

$$\prod_{1 \le j \le 3} |L'_j(q_{s_n}, p_{s_n}, p_{s_n-1})| < q_{s_n-1}^{-1} \ll q_{s_n}^{-0.9},$$

since we have

$$q_{\ell+1} \ll q_\ell^{1.1}, \qquad \text{for any } \ell \ge 1,$$

by Roth's Theorem. Here, the constants implied by $\ll$ depend only on ξ.

It then follows from Theorem 9.1 that the points $(q_{s_n}, p_{s_n}, p_{s_n-1})$, with n in $\mathcal{N}_1$, lie in the union of a finite number of proper subspaces of $\mathbb{Q}^3$. Thus, there exist a nonzero integer triple (y_1, y_2, y_3) and an infinite set of distinct positive integers $\mathcal{N}_2$ such that

$$y_1 q_{s_n} + y_2 p_{s_n} + y_3 p_{s_n-1} = 0, \tag{9.16}$$

for any n in $\mathcal{N}_2$. Dividing (9.16) by q_{s_n}, we get

$$y_1 + y_2 \frac{p_{s_n}}{q_{s_n}} + y_3 \frac{p_{s_n-1}}{q_{s_n-1}} \cdot \frac{q_{s_n-1}}{q_{s_n}} = 0. \tag{9.17}$$

By letting n tend to infinity along $\mathcal{N}_2$, it thus follows from (9.15) that

$$y_1 + y_2 \xi + y_3 \xi^2 = 0.$$

Since (y_1, y_2, y_3) is a nonzero triple of integers, we have reached a contradiction. Consequently, the real number ξ is transcendental, concluding the proof. □

10 Auxiliary results on continued fractions

For the reader's convenience, we recall some classical results from the theory of continued fractions, whose proofs can be found for example in the book of Perron [48].

Lemma 10.1 *Let $\xi = [0; a_1, a_2, \cdots]$ and $\beta = [0; b_1, b_2, \cdots]$ be real numbers. Let $n \geq 1$ such that $a_i = b_i$ for any $i = 1, \ldots, n$. We then have $|\xi - \beta| \leq q_n^{-2}$, where q_n denotes the denominator of the n-th convergent to ξ.*

For positive integers $a_1, \ldots, a_m$, we denote by $K_m(a_1, \ldots, a_m)$ the denominator of the rational number $[0; a_1, \cdots, a_m]$, usually called *continuant.*

Lemma 10.2 *For any positive integers $a_1, \ldots, a_m$ and any integer k with $1 \leq k \leq m-1$, we have*

$$K_m(a_1, \ldots, a_m) = K_m(a_m, \ldots, a_1).$$

Furthermore, the following inequalities hold

$$K_k(a_1, \ldots, a_k) \cdot K_{m-k}(a_{k+1}, \ldots, a_m) \leq K_m(a_1, \ldots, a_m)$$

and

$$K_m(a_1, \ldots, a_m) \leq 2\, K_k(a_1, \ldots, a_k) \cdot K_{m-k}(a_{k+1}, \ldots, a_m).$$

Lemma 10.3 *Let $(a_i)_{i \geq 1}$ be a sequence of positive integers at most equal to M. For any positive integer n, we have*

$$\sqrt{2}^{(n-1)} \leq K_n(a_1, \ldots, a_n) \leq (M+1)^n.$$

11 Continued fractions of formal Laurent series

As mentioned above formal Laurent series can be expanded into continued fractions whose partial quotients are polynomials. The mirror formula and the folding lemma still hold in this context.

We will only give a theorem due to van der Poorten and Shallit (see [54], see also [39] from a remark of Shallit given in [53]).

Theorem 11.1 *Let F be the formal Laurent series given by*

$$F(X) := X \sum_{h \geq 0} X^{-2^h}.$$

Then its continued fraction expansion is equal to

$$[1, X, -X, -X, -X, X, X, -X, \cdots]$$

where the sequence of partial quotients starting from the first X is obtained by repeatedly iterating the folding rule: $W_0 := X$, $W_{j+1} := W_j (-X) \overline{(-W_j)}$ for $j \geq 0$.

Remark 11.2 As previously $\overline{W}$ stands for the reversal of W, so that $W_1 = X \ \ -X \ \ -X$, $W_2 = X \ \ -X \ \ -X \ \ -X \ X \ X \ \ -X$. Note that the same folding trick permitted to Shallit [61] and independently to Kmošek [37] to give the continued fraction expansion of real numbers with explicit g-adic expansion such as $\sum 2^{-2^n}$. Also note that van der Poorten studies precisely how and when continued fraction expansions of Laurent formal series can be “specialized” (see in particular [50]) or “reduced” modulo a prime number (see in particular [51] where the folding lemma is also alluded to).

12 Conclusion

Several other beautiful results about continued fraction expansions where either the mirror formula or the folding lemma are used can be found in the literature: we refer the reader in particular to papers of van der Poorten, Tamura, Liardet-Stambul, Berstel-de Luca... ([11, 40, 52, 53, 66–68]...).

We do not resist to ending this survey by citing a very nice paper on palindromes and continued fractions by Burger [15], that studies when a real quadratic irrational is a linear fractional transformation of its conjugate.

References

[1] B. Adamczewski, Y. Bugeaud, On the complexity of algebraic numbers II. Continued fractions, *Acta Math.*, to appear.

[2] B. Adamczewski, Y. Bugeaud, On the Maillet-Baker continued fractions. *Preprint*, Institut Camille Jordan, 2005.

[3] B. Adamczewski, Y. Bugeaud, On the Littlewood conjecture in simultaneous Diophantine approximation, *J. London Math. Soc.*, to appear.

[4] B. Adamczewski, Y. Bugeaud, Palindromic continued fractions, *Preprint*, Institut Camille Jordan, Lyon, 2005.

[5] B. Adamczewski, Y. Bugeaud, A short proof of the transcendence of the Thue-Morse continued fraction, *Preprint*, Institut Camille Jordan, Lyon, 2005.

[6] W. W. Adams, J. L. Davison, A remarkable class of continued fractions, *Proc. Amer. Math. Soc.* **65** (1977) 194–198.

[7] J.-P. Allouche, Nouveaux résultats de transcendance de réels à développement non aléatoire, *Gaz. Math.* **84** (2000) 19–34.

[8] J.-P. Allouche, J. L. Davison, M. Queffélec, L. Q. Zamboni, Transcendence of Sturmian or morphic continued fractions, *J. Number Theory* **91** (2001) 39–66.

[9] A. Baker, Continued fractions of transcendental numbers, *Mathematika* **9** (1962) 1–8.

[10] A. Baker, On Mahler’s classification of transcendental numbers, *Acta Math.* **111** (1964) 97–120.

[11] J. Berstel, A. de Luca, Sturmian words, Lyndon words and trees, *Theoret. Comput. Sci.* **178** (1997) 171–203.

[12] J. Berstel, P. Séébold, Sturmian words in M. Lothaire, *Algebraic Combinatorics on Words*, Encyclopedia of Mathematics and its Applications **90**, Cambridge University Press, 2002, pp. 45–110.

[13] V. Berthé, C. Holton, L. Q. Zamboni, Initial powers of Sturmian sequences, *Acta Arith.*, to appear.

[14] Y. Bugeaud, M. Laurent, Exponents of Diophantine and Sturmian continued fractions, *Ann. Inst. Fourier* **55** (2005) 773–804.

[15] E. B. Burger, A tail of two palindromes, *Amer. Math. Monthly* **112** (2005) 311–321.

[16] W.-T. Cao, Z.-Y. Wen, Some properties of the factors of Sturmian sequences, *Theoret. Comput. Sci.* **304** (2003) 365–385.

[17] A. Carpi, A. de Luca, Special factors, periodicity, and an application to Sturmian words, *Acta Inform.* **36** (2000) 983–1006.

[18] A. Carpi, A. de Luca, Harmonic and gold Sturmian words, *European J. Combin.* **25** (2004) 685–705.

[19] J. Cassaigne, Limit values of the recurrence quotient of Sturmian sequences, WORDS (Rouen, 1997), *Theoret. Comput. Sci.* **218** (1999) 3–12.

[20] J. W. S. Cassels, H. P. F. Swinnerton-Dyer, On the product of three homogeneous linear forms and the indefinite ternary quadratic forms, *Philos. Trans. Roy. Soc. London. Ser. A.* **248** (1955) 73–96.

[21] D. Crisp, W. Moran, A. Pollington, P. Shiue, Substitution invariant cutting sequences, *J. Théor. Nombres Bordeaux* **5** (1993) 123–137.

[22] D. Damanik, D. Lenz, Powers in Sturmian sequences, *European J. Comb.* **24** (2003) 377–390.

[23] H. Davenport, P. Erdős, W. J. LeVeque, On Weyl's criterion for uniform distribution, *Michigan Math. J.* **10** (1963) 311–314.

[24] H. Davenport, W. M. Schmidt, A theorem on linear forms, *Acta Arith.* **14** (1967-1968) 209–223.

[25] H. Davenport, W. M. Schmidt, Approximation to real numbers by algebraic integers, *Acta Arith.* **15** (1968-1969) 393–416.

[26] J. L. Davison, A series and its associated continued fraction, *Proc. Amer. Math. Soc.* **63** (1977) 29–32.

[27] J. L. Davison, A class of transcendental numbers with bounded partial quotients, In *Number theory and applications (Banff, AB, 1988)*, pages 365–371, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., **265**, Kluwer Acad. Publ., Dordrecht, 1989.

[28] J. L. Davison, Quasi-Periodic Continued Fractions, *Preprint*, 2005.

[29] F. M. Dekking, M. Mendès France, A. J. van der Poorten, Folds!, *Math. Intelligencer* **4** (1982) 130–138, 173–181, 190–195; Erratum, **5** (1983) 5.

[30] M. Einsiedler, A. Katok, E. Lindenstrauss, Invariant measures and the set of exceptions to the Littlewood conjecture, *Ann. of Math.*, to appear.

[31] S. Fischler, Palindrome prefixes and episturmian words, *Preprint*, 2005
`http://www.arxiv.org/abs/math.CO/0501420`

[32] J. S. Frame, Continued fractions and matrices, *Amer. Math. Monthly* **56** (1949) 98–103.

[33] R. Kaufman, Continued fractions and Fourier transforms, *Mathematika* **27** (1980) 262–267.

[34] A. Y. Khintchine, Einige Sätze über Kettenbrüche mit Anwendungen auf die Theorie der diophantischen Approximationen, *Math Ann.* **92** (1924) 115–125.

[35] A. Y. Khintchine, *Continued Fractions* (in Russian), Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow-Leningrad, 2nd edition, 1949.

[36] A. Y. Khintchine, *Continued fractions*, Translated by Peter Wynn, P. Noordhoff Ltd., Groningen, 1963.

[37] M. Kmošek, Rozwiniecie niektórych liczb niewymiernych na ułamki łańcuchowe, Master thesis, Warsaw, 1979.

[38] W. J. LeVeque, *Topics in number theory. Vol. 1, 2*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1956.

[39] W. Leighton, W. T. Scott, A general continued fraction expansion, *Bull. Amer. Math. Soc.* **45** (1939) 596–605.

[40] P. Liardet, P. Stambul, Séries de Engel et fractions continuées, *J. Théor. Nombres Bordeaux* **12** (2000) 37–68.

[41] J. Liouville, Sur des classes très étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationelles algébriques, *C. R. Acad. Sci. Paris* **18** (1844) 883–885; 910-911.

[42] J. E. Littlewood, *Some problems in real and complex analysis*, D. C. Heath and Co. Raytheon Education Co., Lexington, Mass., 1968.

[43] E. Maillet, *Introduction à la théorie des nombres transcendants et des propriétés arithmétiques des fonctions*, Gauthier-Villars, Paris, 1906.

[44] G. Margulis, Problems and conjectures in rigidity theory, In *Mathematics: frontiers and perspectives*, pages 161–174, Amer. Math. Soc., Providence, RI, 2000.

[45] B. de Mathan, Conjecture de Littlewood et récurrences linéaires, *J. Théor. Nombres Bordeaux* **15** (2003) 249–266.

[46] M. Mendès France, Sur les fractions continues limitées, *Acta Arith.* **23** (1973) 207–215.

[47] F. Mignosi, G. Pirillo, Repetitions in the Fibonacci infinite word, *RAIRO Inform. Théor. Appl.* **26** (1992) 199–204.

[48] O. Perron, *Die Lehre von den Kettenbrüchen*, Teubner, Leibzig, 1929.

[49] A. D. Pollington, S. L. Velani, On a problem in simultaneous Diophantine approximation: Littlewood's conjecture, *Acta Math.* **185** (2000) 287–306.

[50] A. J. van der Poorten, Continued fractions of formal power series, In *Advances in number theory (Kingston, ON, 1991)*, pp. 453–466, Oxford Sci. Publ., Oxford Univ. Press, New York, 1993.

[51] A. J. van der Poorten, Reduction of continued fractions of formal power series, in *Continued fractions: from analytic number theory to constructive approximation (Columbia, MO, 1998)*, pp. 343–355, Contemp. Math., 236, Amer. Math. Soc., Providence, RI, 1999.

[52] A. J. van der Poorten, Beer and continued fractions with periodic periods, in *Number theory (Ottawa, ON, 1996)*, pp. 309–314, CRM Proc. Lecture Notes, 19, Amer. Math. Soc., Providence, RI, 1999.

[53] A. J. van der Poorten, Symmetry and folding of continued fractions, *J. Théor. Nombres Bordeaux* **14** (2002) 603–611.

[54] A. J. van der Poorten, J. Shallit, Folded continued fractions, *J. Number Theory* **40** (1992) 237–250.

[55] M. Queffélec, Transcendance des fractions continues de Thue-Morse, *J. Number Theory* **73** (1998) 201–211.

[56] M. Ratner, Interactions between ergodic theory, Lie groups, and number theory, In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, pages 157–182, Birkhäuser, Basel, 1995.

[57] K. F. Roth, Rational approximations to algebraic numbers, *Mathematika* **2** (1955) 1–20; corrigendum 168.

[58] D. Roy, Approximation simultanée d'un nombre et de son carré, *C. R. Math. Acad. Sci. Paris* **336** (2003) 1–6.

[59] D. Roy, Approximation to real numbers by cubic algebraic integers, II, *Ann. of Math. (2)* **158** (2003) 1081–1087.

[60] D. Roy, Approximation to real numbers by cubic algebraic integers, I, *Proc. London Math. Soc. (3)* **88** (2004) 42–62.

[61] J. Shallit, Simple continued fractions for some irrational numbers, *J. Number Theory* **11** (1979) 209–217.

[62] J. Shallit, Real numbers with bounded partial quotients: a survey, *Enseign. Math.* **38** (1992) 151–187.

[63] W. M. Schmidt, On simultaneous approximations of two algebraic numbers by rationals, *Acta Math.* **119** (1967) 27–50.

[64] W. M. Schmidt, Norm form equations, *Ann. of Math. (2)*, **96** (1972) 526–551.

[65] W. M. Schmidt, *Diophantine approximation*, Lecture Notes in Mathematics **785**, Springer, Berlin, 1980.

[66] J.-i. Tamura, Symmetric continued fractions related to certain series, *J. Number Theory* **38** (1991) 251–264.

[67] J.-i. Tamura, Transcendental numbers having explicit g-adic and Jacobi-Perron expansions, *Sém. Théor. Nombres Bordeaux Sér. 2* **4** (1992) 75–95.

[68] J.-i. Tamura, A class of transcendental numbers with explicit g-adic expansion and the Jacobi-Perron algorithm, *Acta Arith.* **61** (1992) 51–67.

[69] D. Vandeth, Sturmian words and words with a critical exponent, *Theoret. Comput. Sci.* **242** (2000) 283–300.

[70] E. Wirsing, Approximation mit algebraischen Zahlen beschränkten Grades, *J. Reine Angew. Math.* **206** (1960) 67–77.

On the Representation of Numbers in a Rational Base

Shigeki Akiyama[*], *Christiane Frougny*[†], *Jacques Sakarovitch*[‡]

Dedicated to our friend Antonio Restivo
on the occasion of his 60th birthday

Abstract

A new method for representing positive integers and real numbers in a rational base is considered. It amounts to computing the digits from right to left, least significant first. Every integer has a unique such expansion. The set of expansions of the integers is not a regular language but nevertheless addition can be performed by a letter-to-letter finite right transducer. Every real number has at least one such expansion and a countable infinite set of them have more than one. We explain how these expansions can be approximated and characterize the expansions of reals that have two expansions.

These results are not only developped for their own sake but also as they relate to other problems in combinatorics and number theory. A first example is a new interpretation and expansion of the constant $K(p)$ from the so-called "Josephus problem". More important, these expansions in the base $\frac{p}{q}$ allow us to make some progress in the problem of the distribution of the fractional part of the powers of rational numbers.

Extended Abstract

In this paper[1], we introduce and study a new method for representing positive integers and real numbers in the base $\frac{p}{q}$, where $p > q \geqslant 2$ are coprime integers.[2] The idea of non-standard representation systems of numbers is far from being original and there have been extensive studies of these, from a theoretical standpoint as well as for improving computation algorithms. It is worth first (briefly) recalling the main features of these systems in order to clearly put in perspective and in contrast the results we have obtained on rational base systems.

[*] Dept. of Mathematics, Niigata University, akiyama@math.sc.niigata-u.ac.jp

[†] LIAFA, UMR 7089 CNRS, and Université Paris 8, Christiane.Frougny@liafa.jussieu.fr

[‡] LTCI, UMR 5141, CNRS / ENST, sakarovitch@enst.fr

[1] Work partially supported by the CNRS/JSPS contract 13 569.

[2] To keep the length into reasonnable bounds for the proceedings of a conference, proofs are only sketched or even sometimes omitted. A full paper will soon be submitted to a journal.

Many non-standard numeration systems have been considered, [13, Vol. 2, Chap. 4] or [14, Chap. 7], for instance, give extensive references. Representation in integer base with signed digits was popularized in computer arithmetic by Avizienis [3] and can be found earlier in a work of Cauchy [6]. When the base is a real number $\beta > 1$, any non-negative real number is given an expansion on the canonical alphabet $\{0, 1, \ldots, \lfloor\beta\rfloor\}$ by the greedy algorithm of Rényi [19]; a number may have several β-representations on the canonical alphabet, but the greedy one is the greatest in the lexicographical order. The set of greedy β-expansions of numbers of $[0, 1[$ is shift-invariant, and its closure forms a symbolic dynamical system called the *β-shift*. The properties of the β-shift are well understood, using the so-called "β-expansion of 1", see [14, 17].

When β is a Pisot number[3], the β number system shares many properties with the integer base case: the set of greedy representations is recognizable by a finite automaton; the conversion between two alphabets of digits (in particular addition) is realized by a finite transducer [10].

Here, we first define the *$\frac{p}{q}$-expansion* of an integer N: it is a way of writing N in the *base* $\frac{p}{q}$ by an algorithm which produces the digits from right to left, that is *least significant digits first*. We prove:

Theorem 0.1 *Every non-negative integer N has a* $\frac{p}{q}$-expansion *which is an* integer *representation. It is the* unique finite $\frac{p}{q}$-representation *of N.*

The $\frac{p}{q}$-expansions *are not* the $\frac{p}{q}$-representations that would be obtained by the classical "greedy algorithm" in base $\frac{p}{q}$. They are written on the alphabet $A = \{0, 1, \ldots, p-1\}$, but not every word of A^* is admissible. These $\frac{p}{q}$-expansions share some properties with the expansions in an integer base — digit set conversion is realized by a finite transducer for instance — and are completely different as far as other aspects are concerned. Above all, the set $L_{\frac{p}{q}}$ of all $\frac{p}{q}$-expansions is not a regular language (not even a context-free one). To some extend, the study and understanding of this set of words $L_{\frac{p}{q}}$ is what this paper is about.

By construction, the set $L_{\frac{p}{q}}$ is prefix-closed and any of its elements can be extended (to the right) in $L_{\frac{p}{q}}$. Hence, $L_{\frac{p}{q}}$ is the set of labels of the finite paths in an *infinite subtree* $T_{\frac{p}{q}}$ of the infinite full p-ary tree of the free monoid A^*. The tree $T_{\frac{p}{q}}$ contains a maximal infinite word $\mathbf{t}_{\frac{p}{q}}$ — maximal in the lexicographic ordering — whose numerical value is $\boldsymbol{\omega}_{\frac{p}{q}}$. We define the *admissible $\frac{p}{q}$-expansions* of real numbers to be the set $W_{\frac{p}{q}}$ of infinite words that label the *infinite paths* of $T_{\frac{p}{q}}$ and we prove:

Theorem 0.2 *Every real in $[0, \boldsymbol{\omega}_{\frac{p}{q}}]$ has exactly one $\frac{p}{q}$-expansion, but for an infinite countable number of them which have more than one such expansion.*

[3] An algebraic integer whose Galois conjugates are all less than 1 in modulus

If $p \geqslant 2q-1$, then no real has more than two $\frac{p}{q}$-expansions. It is noteworthy as well that no $\frac{p}{q}$-expansion is eventually periodic and thus in particular — and in contrast with the expansion of reals in an integer base — no $\frac{p}{q}$-expansion ends with 0^ω or, which is the same, is finite. This is a very remarkable feature of the $\frac{p}{q}$ number system for reals and we explain how the $\frac{p}{q}$-expansion of a real number can be computed (in fact approximated).

We shall give here two examples of the relations of the $\frac{p}{q}$-expansions of reals with other problems in combinatorics and number theory. The first one is the so-called "Josephus problem" in which a certain constant $K(p)$ is defined (*cf.* [11, 16, 22]) which is a special case of our constant $\boldsymbol{\omega}_{\frac{p}{q}}$ (with $q = p-1$) and this definition yields a new method for computing $K(p)$.

The connection with the second problem, namely the distribution of the powers of a rational number modulo 1, is even more striking. It requires to be presented that the framework of this long standing and deeply intriguing problem be set.[4]

Koksma proved that for almost every real number $\theta > 1$ the sequence $\{\theta^n\}$ is uniformely distributed in $[0,1]$, but very few results are known for specific values of θ. One of these is that *if θ is a Pisot number*, then the above sequence converges to 0 if we identify $[0,1)$ with $\mathbb{R}/\mathbb{Z}$.

The distribution of $\left\{(\frac{p}{q})^n\right\}$ for coprime positive integers $p > q \geqslant 2$ remains an unsolved problem. Experimental results show that this distribution looks more "chaotic" than the distribution of the fractional part of the powers of a transcendental number like e or π (*cf.* [24]). Vijayaraghavan [23] showed that the sequence has infinitely many limits points.

The next step in attacking this problem has been *to fix the rational* $\frac{p}{q}$ and to study the distribution of the sequence

$$f_n(\xi) = \left\{\xi \left(\frac{p}{q}\right)^n\right\}$$

according to the value of the real number ξ. Once again, the sequence $f_n(\xi)$ is uniformely distributed for almost all $\xi > 0$, but nothing is known for specific value of ξ.

In the search for ξ's for which the sequence $f_n(\xi)$ is *not uniformely distributed*, Mahler considered those for which the sequence is eventually contained in $[0, \frac{1}{2}[$. Mahler's notation is generalized as follow: let I be a (strict) subset of $[0,1[$ — indeed I will be a finite union of semi-closed intervals — and write:

$$\mathbf{Z}_{\frac{p}{q}}(I) = \{\xi \in \mathbb{R} \mid \left\{\xi \left(\frac{p}{q}\right)^n\right\}_{n\in\mathbb{N}} \quad \text{belongs eventually to } I\ \} \ .$$

Mahler [15] proved that $\mathbf{Z}_{\frac{3}{2}}\left([0,\frac{1}{2}[\right)$ is at most countable but left open the problem to decide whether it is empty or not. Mahler's work has been developped

[4] This presentation is based on the introduction of [5]. The fractional part of a number x is denoted by $\{x\}$.

in two directions: the search for subsets I as large as possible such that $\mathbf{Z}_{\frac{p}{q}}(I)$ is empty — which amounts to proving that $f_n(\xi)$ is never too much unevenly distributed — and conversely the search for subsets I as small as possible such that $\mathbf{Z}_{\frac{p}{q}}(I)$ is non-empty — which corresponds to establish the existence of ξ for which $f_n(\xi)$ is as unevenly distributed as possible.

Along the first line, a remarkable progress has been made by Flatto *et al.* ([8]) who proved that the set of reals s such that $\mathbf{Z}_{\frac{p}{q}}\left([s, s+\frac{1}{p}[\right)$ is empty is *dense* in $[0, 1-\frac{1}{p}]$, and this has been even improved by Bugeaud [5] who proved that its complement is of Lebesgue measure 0. Along the other line, Pollington [18] showed that $\mathbf{Z}_{\frac{3}{2}}\left([\frac{4}{65}, \frac{61}{65}[\right)$ is non-empty.

Our contribution to the problem can be seen as an improvement of this result.

Theorem 0.3 *If* $p \geqslant 2q-1$, *there exists a subset* $Y_{\frac{p}{q}}$ *of* $[0,1[$, *of Lebesgue measure* $\frac{q}{p}$, *such that* $\mathbf{Z}_{\frac{p}{q}}\left(Y_{\frac{p}{q}}\right)$ *is countable infinite.*

The elements of $\mathbf{Z}_{\frac{p}{q}}\left(Y_{\frac{p}{q}}\right)$ are indeed the reals which have *two* $\frac{p}{q}$*-expansions* and this is the reason why the consideration of the $\frac{p}{q}$ number system allowed to make some progress in Mahler's problem.

*

In conclusion, we have introduced and studied here a fascinating family of sets of words which can be seen from many sides, which raises still many difficult questions and whose further study will certainly mix techniques from word combinatorics, automata theory, and number theory.

In order to keep the proceedings into reasonable length bounds we do not include in this extended abstract any preliminary for definition nor notation on (infinite) words and automata but rather follow [7, 12, 14] and we only give sketch of proofs, when we give any. A more complete version will be to be found elsewhere in a journal ([2]).

1 The $\frac{p}{q}$ number system

Let $p > q \geqslant 1$ be two co-prime integers and let U be the sequence defined by:

$$U = \{u_i = \frac{1}{q}\left(\frac{p}{q}\right)^i \mid i \in \mathbb{Z}\}.$$

We will say that U, *together with* the alphabet $A = \{0, \ldots, p-1\}$, is the $\frac{p}{q}$ *number system.* If $q = 1$, it is exactly the classical number system in base p.

A *representation in the system* U of a non-negative real number x on a finite alphabet of digits D is an infinite sequence of digits in D indexed by a section of $\mathbb{Z}$: $(d_i)_{k \geqslant i \geqslant -\infty}$, such that:

$$x = \sum_{-\infty}^{i=k} d_i u_i, \text{ an equation that is written as } \langle x \rangle_{\frac{p}{q}} = d_k \cdots d_0 \centerdot d_{-1} d_{-2} \cdots \ ,$$

most significant digit first. When a representation ends in infinitely many zeroes, it is said to be *finite*, and the trailing zeroes are omitted. When all the d_i with negative index are zeroes, the representation is said to be an *integer representation.* Conversely, the numerical value in the system U of a word on an alphabet of digits D is given by the *evaluation map* π:

$$\pi \colon D^{\mathbb{Z}} \longrightarrow \mathbb{R} \,, \qquad\qquad \mathbf{d} = \{d_i\}_{k \geqslant i \geqslant -\infty} \longmapsto \pi(\mathbf{d}) = \sum_{-\infty}^{i=k} d_i u_i \ .$$

It is important to remark that this definition *is not* the classical one for the numeration system in base $\frac{p}{q}$: U *is not* the sequence of powers of $\frac{p}{q}$ but rather these powers *divided by* q and the digits *are not* the integers smaller than $\frac{p}{q}$ but rather the integers *whose quotient by* q is smaller than $\frac{p}{q}$. These two differences compensate each other and make the developments that follow possible.

2 Representation of the integers

2.1 The $\frac{p}{q}$-expansion of an integer

Let N be any positive integer. Write $N_0 = N$ and, for $i \geqslant 0$, write

$$q\, N_i = p\, N_{i+1} + a_i$$

where a_i is the remainder of the Euclidean division of $q\,N_i$ by p, and thus belongs to A. This is an algorithm that produces the digits of N from right to left, that is to say *least significant digit first*, and stops for some k when $N_{k+1} = 0$. It holds $N = \sum_{i=0}^{k} a_i u_i$ and thus the word $a_k \cdots a_0$ is a $\frac{p}{q}$-representation of N; it will be called *the* $\frac{p}{q}$*-expansion* of N and written $\langle N \rangle_{\frac{p}{q}}$. By convention the $\frac{p}{q}$-expansion of 0 is the empty word.

It can be further proved that $\langle N \rangle_{\frac{p}{q}}$ is the unique *finite* $\frac{p}{q}$-representation of N (under the condition that $a_k \neq 0$). We have thus established:

Theorem 2.1 *Every non-negative integer N has a* $\frac{p}{q}$-expansion *which is an* integer *representation. It is the* unique finite $\frac{p}{q}$-representation *of N.* ■

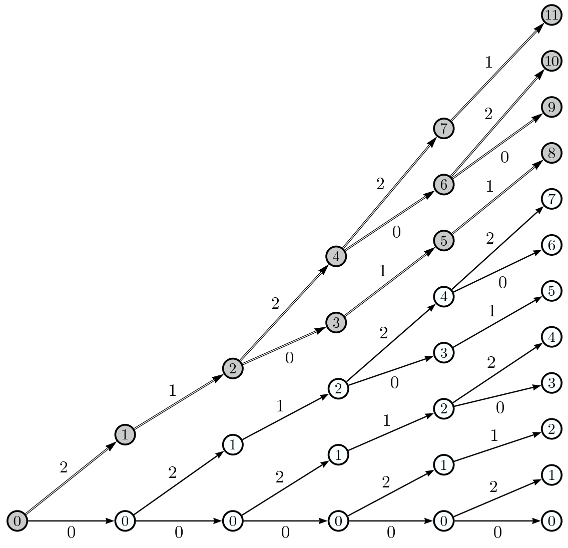
Example 2.1 ex1 Let $p = 3$ and $q = 2$, then $A = \{0, 1, 2\}$ — this will be our main running example. Table 1 in Fig.1 gives the $\frac{3}{2}$-expansions of the twelves first integers (*cf.* Appendix A as well).

Remark 2.2 If $q = 1$, *i.e.* the base is an integer, the above algorithm gives the same representation as the one given by the classical greedy (left-to-right) algorithm (see [19] or [14, Chapter 7]). On the contrary, if $q \neq 1$, the $\frac{p}{q}$-expansion of n *is not* the representation obtained by the greedy algorithm but for $n = 0, 1, \ldots, \lfloor \frac{p}{q} \rfloor$. It follows then from Theorem 2.1 that no integer (but $0, 1, \ldots, \lfloor \frac{p}{q} \rfloor$) is given a finite representation by the greedy algorithm.

2.2 The set of $\frac{p}{q}$-expansions

Let us denote by $L_{\frac{p}{q}}$ the set of $\frac{p}{q}$-*expansions* of the non-negative integers. If $q = 1$ then $L_{\frac{p}{q}}$ is the set of all words of A^* which do not begin with a 0; if we release this last condition, we then get the whole A^*. If $q \neq 1$, $L_{\frac{p}{q}}$ is prefix-closed by construction and the observation of Table 1 shows that it is not suffix-closed.

For each a in A, we define a partial map τ_a from $\mathbb{N}$ into itself: for z in $\mathbb{N}$, $\tau_a(z) = \dfrac{1}{q}(pz + a)$ if the latter is an integer, $\tau_a(z)$ is undefined otherwise. The labelled tree $T_{\frac{p}{q}}$ is then constructed as follows: the root is labelled by 0, the children of a node labelled by z are nodes labelled by the (defined) $\tau_a(z)$, the edge from z to $\tau_a(z)$ being labelled by a. Let us call *word label* of a node s, and write $w(s)$, the label of *the* path from the root to s. By construction the label of s is $\pi(w(s))$. Let us denote $I_{\frac{p}{q}}$ the subtree of $T_{\frac{p}{q}}$ made of nodes whose word label does not begin with a 0. See Fig. 1 for (a part of) $T_{\frac{3}{2}}$ and $I_{\frac{3}{2}}$.



	0
2	1
21	2
210	3
212	4
2101	5
2120	6
2122	7
21011	8
21200	9
21202	10
21221	11

Table 1.

Figure 1: The tree $T_{\frac{3}{2}}$, the tree $I_{\frac{3}{2}}$ in grey and double edge, and a table

Among the digits that label *the edges* that start from a node with label N, there is a minimum one, minDigit(N), which belongs to $\{0, \ldots, q-1\}$ — and this is characteristic of a minimum digit — and there is a maximal one, MaxDigit(N), which belongs to $\{p-q,\ldots,p-1\}$ — and this is characteristic of a maximal digit. If $D = \mathrm{MaxDigit}(N)$ and $d = \mathrm{minDigit}(N+1)$, then $D = d + (p-q)$ and $\tau_d(N+1) = \tau_D(N) + 1$.

It follows that for every integer k, there exists an integer M_k such that the nodes of depth k in $T_{\frac{p}{q}}$ are labelled by the integers from 0 to M_k (see Section 3 for the computation of the M_k's). And the labelling of nodes (in $\mathbb{N}$) gives the ordering in the radix order on $I_{\frac{p}{q}}$. Closer investigations give the following (under the hypothesis that $q \neq 1$):

Proposition 2.3 *No two subtrees of $I_{\frac{p}{q}}$ are isomorphic.*

Corollary 2.4 *$L_{\frac{p}{q}}$ is not a regular language.*

Lemma 2.5 *For any words x and y in $A^* \setminus 0^*$, $y \neq \varepsilon$, there exists an integer constant $K(x,y)$ such that, if xy^n belongs to $L_{\frac{p}{q}}$, then $n \leq K(x,y)$.*

Corollary 2.6 *$L_{\frac{p}{q}}$ is not a context-free language.*

Proposition 2.7 *Every w in A^k is the suffix of the $\frac{p}{q}$-expansion of a unique integer n, $0 \leqslant n < p^k$.*

2.3 Conversion between alphabets

Let D be a finite alphabet of (positive or negative) digits that contains A. The *digit-set conversion* is a map $\chi_D \colon D^* \to A^*$ which commutes to the evaluation map π, that is a map which preserves the numerical value:

$$\forall w \in D^* \qquad \pi(\chi_D(w)) = \pi(w) \ .$$

Proposition 2.8 *For any alphabet D the conversion χ_D is realizable by a finite letter-to-letter sequential right transducer $\mathcal{C}_D$.*

The states of $\mathcal{C}_D$ are *integers*, the state 0 is *initial*, and the final function of a state h (with h positive) is the $\frac{p}{q}$-expansion of h. A transition labelled by $d|a$, with d in D and a in A, goes from h to k if and only if

$$q\,h + d = p\,k + a \ . \tag{2.1}$$

The set of accessible states is finite and this establishes Proposition 2.8. ■

The *integer addition* may be seen — after digit-wise addition — as a particular case of a digit-set conversion χ_D with $D = \{0,1,\ldots,2(p-1)\}$ and Figure 2 shows the converter that realizes addition in the $\frac{3}{2}$-system.

Remark 2.9 Let us stress that χ_D is defined on the whole set D^* even for word v such that $\pi(v)$ is not an integer, and also that, if $\pi(v)$ is in $\mathbb{N}$, then $\chi_D(v)$ is the unique $\frac{p}{q}$-expansion of $\pi(v)$.

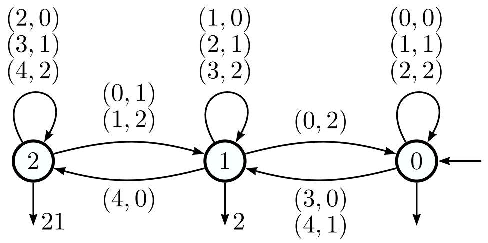


Figure 2: A converter for addition in the $\frac{3}{2}$ number system

Remark 2.10 As $\frac{p}{q}$ is not a Pisot number (when $q \neq 1$), the conversion from any representation onto the expansion computed by the greedy algorithm is not realized by a finite transducer (see [14, Ch. 7]).

Remark 2.11 The conversion χ_D may also be extended to *left infinite* words and it follows from Proposition 2.8 that it is a *continuous fonction.* In particular, the *odometer* is continuous.

3 A remarkable set of infinite words

Let $W_{\frac{p}{q}}$ be the set of labels of infinite paths starting from the root 0 in $T_{\frac{p}{q}}$. It follows from the construction of $T_{\frac{p}{q}}$ by the partial functions $\{\tau_a \mid a \in A\}$ that two nodes with the same label are the root of the same subtree and from Proposition 2.3 that these subtrees are characteristic of the label (in fact, *any* infinite path from a node is characteristic of the label of the node).

This set $W_{\frac{p}{q}}$ will be used in the next section in order to define the representations of real numbers. We first try to describe it and to present some of its properties. In the previous sections, digits in a $\frac{p}{q}$-representation where indexed from left to right by decreasing nonnegative integers for the "integer" part and by decreasing negative integers for the "decimal" part; as we shall now deal mainly with the "decimal" part of the representations, we find it much more convenient *to change the convention of indexing* and use the positive indices *after* the decimal point, in the increasing order.

We denote by $\mathrm{MaxWord}(N)$ (resp. $\mathrm{minWord}(N)$) the label of the infinite path that starts from a node with label N and that follows always the edges with the maximal (resp. minimal) digit label. From our observation in Section 2.2, it follows that for every N, the digit-wise difference between $\mathrm{minWord}(N+1)$ and $\mathrm{MaxWord}(N)$ is $(p-q)^{\omega}$.

Let us note $\mathbf{t}_{\frac{p}{q}} = \mathrm{MaxWord}(0)$ and $\mathbf{g}_{\frac{p}{q}} = \mathrm{minWord}(1)$. The word $\mathbf{t}_{\frac{p}{q}} = (t_i)_{i \geqslant 1}$ belongs to $\{p-q, \ldots, p-1\}^{\mathbb{N}}$ and is the maximal word in $T_{\frac{p}{q}}$ (or $I_{\frac{p}{q}}$) in the lexicographic ordering. The word $\mathbf{g}_{\frac{p}{q}} = (g_i)_{i \geqslant 1}$ belongs to $\{0, \ldots, q-1\}^{\mathbb{N}}$

and the word $q\,\mathbf{g}_{\frac{p}{q}}$ is the minimal word of $I_{\frac{p}{q}}$ in the lexicographic ordering. Remark that when $q = 1$, $\mathrm{MaxWord}(N) = (p-1)^{\omega}$, and $\mathrm{minWord}(N) = 0^{\omega}$ for every N.

Example 3.1 G For $\frac{p}{q} = \frac{3}{2}$, $\mathbf{g}_{\frac{3}{2}} = 101100011010011010100110\ldots.$

For $n \geqslant 1$ let $G_n = \pi(qg_1 \cdots g_{n-1})$ (if $n = 1$, $G_1 = \pi(q) = 1$), and $M_n = \pi(t_1 \cdots t_n)$. Of course $M_n = G_{n+1} - 1$. Let $\boldsymbol{\omega}_{\frac{p}{q}} = \pi(\boldsymbol{.}\mathbf{t}_{\frac{p}{q}})$ the numerical value of the maximal infinite word and it holds $\boldsymbol{\gamma}_{\frac{p}{q}} = \pi(\boldsymbol{.}q\,\mathbf{g}_{\frac{p}{q}}) = \pi(\boldsymbol{.}0\,\mathbf{t}_{\frac{p}{q}}) = \dfrac{q}{p}\boldsymbol{\omega}_{\frac{p}{q}}$. We then have the following results.

Proposition 3.2 *The sequence* $(G_n)_{n\geqslant 1}$ *satisfies the recurrence* $G_n = \lceil \dfrac{p}{q} G_{n-1} \rceil$ *with* $G_1 = 1$*, and for* $n \geqslant 1$ *there exists an integer* e_n*,* $0 \leqslant e_n < (q-1)/(p-q)$*, such that*

$$G_n = \lfloor \boldsymbol{\gamma}_{\frac{p}{q}} \big(\frac{p}{q}\big)^n \rfloor - e_n\,.$$

Corollary 3.3 *If* $p \geqslant 2q - 1$ *then, for* $n \geqslant 1$*,* $G_n = \lfloor \boldsymbol{\gamma}_{\frac{p}{q}} \big(\frac{p}{q}\big)^n \rfloor$.

Remark 3.4 If $p \geqslant 2q - 1$ then for each $n \geqslant 1$, the digit g_n is obtained as follows:
(i) compute $G_{n+1} = \lceil \frac{p}{q} G_n \rceil$ (ii) $g_n = q\,G_{n+1} \mod p\,.$

The definition of the sequence G_n and the computation of $\mathbf{g}_{\frac{p}{q}}$ have been developped not only because they are important for the description of $T_{\frac{p}{q}}$ but also as they relate to a classical problem in combinatorics.

Inspired by the so-called "Josephus problem", Odlyzko and Wilf consider, for a real $\alpha > 1$, the iterates of the function $f(x) = \lceil \alpha x \rceil$: $f_0 = 1$ and $f_{n+1} = \lceil \alpha f_n \rceil$ for $n \geqslant 0$. They show (in [16]) that if $\alpha \geqslant 2$, or $\alpha = 2 - 1/q$ for some integer $q \geqslant 2$, then there exists a constant $H(\alpha)$ such that $f_n = \lfloor H(\alpha)\,\alpha^n \rfloor$ for all $n \geqslant 0$.

We have thus obtain the same result as in [16] for rational $\alpha = \frac{p}{q}$, with $p \geqslant 2q - 1$, and we find $H(\frac{p}{q}) = \frac{p}{q}\boldsymbol{\gamma}_{\frac{p}{q}} = \boldsymbol{\omega}_{\frac{p}{q}}$. Our method does not yield an "independent" way of computing this constant, as was called for in [16], but the writing of $\boldsymbol{\omega}_{\frac{p}{q}}$ in the $\frac{p}{q}$-system gives at least an easy algorithm.

In the case where $q = p - 1$ (the Josephus case), the constant $\boldsymbol{\omega}_{\frac{p}{q}}$ is the constant $K(p)$ in [16]. In this case the integer e_n of Proposition 3.2 is less than $p - 2$, and this is the same bound as in [16].

Example 3.5 gam For $\frac{p}{q} = \frac{3}{2}$, the constant $\boldsymbol{\omega}_{\frac{3}{2}}$ is the constant $K(3)$ already discussed in [11,16,22]. Its decimal expansion $1\boldsymbol{.}62227050288476731595695098\cdots$ is recorded as Sequence A083286 in [21]. Observe that, in the same case, the sequence $(G_n)_{n\geqslant 1}$ is Sequence A061419 in [21].

As a consequence of Lemma 2.5, we have:

Proposition 3.6 *If $q > 1$ then no element of $W_{\frac{p}{q}}$ is eventually periodic, but* 0^ω.

4 Representation of the reals

In order to distinguish between the integer and the real numerical values of some finite word, we put a decimal point **.** indicating the position of the index 0.

4.1 The $\frac{p}{q}$-expansions of reals

We take *as a definition* that the $\frac{p}{q}$-expansions of the real numbers are the elements of $W_{\frac{p}{q}}$. That is, an infinite word $\mathbf{a} = \{a_i\}_{i \geqslant 1}$ in $W_{\frac{p}{q}}$ is *a $\frac{p}{q}$-expansion* of the real number x:

$$x = \pi(\mathbf{.a}) = \frac{1}{q} \sum_{i \geqslant 1} a_i \left(\frac{q}{p}\right)^i .$$

Let $X_{\frac{p}{q}} = \pi(W_{\frac{p}{q}})$. The elements of $X_{\frac{p}{q}}$ are non-negative real numbers less than or equal to $\boldsymbol{\omega}_{\frac{p}{q}}$. Note that $\boldsymbol{\omega}_{\frac{p}{q}} \leqslant \frac{p-1}{p-q}$. The fact that the $\frac{p}{q}$ number system may be used for representing the reals is expressed by the following statement.

Theorem 4.2 *Every real in $[0, \boldsymbol{\omega}_{\frac{p}{q}}]$ has exactly one $\frac{p}{q}$-expansion, but for an infinite countable number of them which have more than one such expansion.*

The proof of the first part of Theorem 4.2, that is to say the proof that $X_{\frac{p}{q}} = [0, \boldsymbol{\omega}_{\frac{p}{q}}]$, relies on three facts. First, $W_{\frac{p}{q}}$ is closed in the compact set $A^{\mathbb{N}}$, hence is compact. Second, the map $\pi \colon W_{\frac{p}{q}} \to X_{\frac{p}{q}}$ is continuous and order-preserving. Hence $X_{\frac{p}{q}}$ is a closed subset of the interval $[0, \boldsymbol{\omega}_{\frac{p}{q}}]$. And finally, properties of the tree $T_{\frac{p}{q}}$ imply that $[0, \boldsymbol{\omega}_{\frac{p}{q}}] \setminus X_{\frac{p}{q}}$ cannot contain any non-empty open interval. From the same properties we deduce that real numbers having more than one expansion correspond to the branching nodes of $T_{\frac{p}{q}}$, hence the second part of Theorem 4.2.

Remark 4.1 In contrast with the classical representations of reals, the finite prefixes of a $\frac{p}{q}$-expansion of a real number, completed by zeroes, *are not* $\frac{p}{q}$-expansions of real numbers (though they can be given a value by the function π of course), that is to say, if a finite word w is in $L_{\frac{p}{q}} \setminus 0^*$, then the word $w0^\omega$ does not belong to $W_{\frac{p}{q}}$.

Corollary 4.2 *If $p \geqslant 2q - 1$ then no real number can have three different expansions.*

4.2 The companion $\frac{p}{q}$-representation and the co-converter

A feature of the $\frac{p}{q}$-expansion of the integers is that it is computed least significant digit first, or *from right to left*. This is quite an accepted process for integers, that becomes problematic when it comes to the reals and that you have to compute from right to left a representation which is *infinite to the right*[5]. This difficulty is somewhat overcome with the definition of *another* $\frac{p}{q}$-representation for the reals; it can be computed with any prescribed precision (provided we can compute in $\mathbb{Q}$ with the same precision) and somehow *from left to right.* The price we have to pay for this is that we use a larger alphabet of digits, containing *negative digits*, exactly as the Avizienis representation of reals allows to perform sequentially addition from left to right [3].

Let $h\colon \mathbb{R}_+ \to \mathbb{Z}$ be the function defined by

$$h(z) = q\left\lfloor (\frac{p}{q})z \right\rfloor - p\left\lfloor z \right\rfloor .$$

The function h is periodic of period q and for all z in $\mathbb{R}_+$, $h(z)$ belongs to the digit alphabet

$$C = \{-(q-1), \ldots, 0, 1, \ldots, (p-1)\} .$$

(If $q = 1$, then $C = A$; $C = A \cup \{-(q-1), \ldots, -1\}$ otherwise.)

Let us write now, for every n in $\mathbb{N}$, $c_n = h\left((\frac{p}{q})^{n-1} z\right)$ which, in turn, defines a map $\varphi(z)\colon \mathbb{R}_+ \to C^{\mathbb{N}}$ by $\varphi(z) = \mathbf{c} = {\bf .}c_1 c_2 \cdots c_n \cdots$. If $q = 1$, c_n is precisely the n-th digit after the decimal point in the expansion of z in base p.

We call the sequence $\varphi(z)$ *the companion representation* of z, and we have:

Proposition 4.3 *For all z in $\mathbb{R}_+$, $\varphi(z)$ is a $\frac{p}{q}$-representation of $\{z\} = z - \lfloor z \rfloor$, the fractional part of z.* ■

Let x be in $[0, \boldsymbol{\omega}_{\frac{p}{q}}]$. Let $\langle x \rangle_{\frac{p}{q}} = \mathbf{a} = {\bf .}a_1 a_2 \cdots$ be a $\frac{p}{q}$-expansion of x and let $\varphi(x) = \mathbf{c} = {\bf .}c_1 c_2 \cdots$ its companion representation. Let us denote by $\rho_n(x)$ the integer part $\lfloor \pi({\bf .}a_{n+1} a_{n+2} \cdots) \rfloor$; easy calculation then shows:

$$c_n + p\,\rho_{n-1}(x) = a_n + q\,\rho_n(x) . \tag{4.1}$$

There are a finite number of possible values for $\rho_n(x)$ since $0 \leq \rho_n(x) < \frac{p-1}{p-q}$, and (4.1) can be seen as the definition of a (left) transducer $\mathcal{A}_{\frac{p}{q}}$: a transition labelled by (c_n, a_n) goes from the state $\rho_{n-1}(x)$ to the state $\rho_n(x)$. We recognize, by comparison with (2.1), that $\mathcal{A}_{\frac{p}{q}}$ is the transposed automaton of the converter $\mathcal{C}_C$ that we have described at Section 2.3. The transducer $\mathcal{A}_{\frac{p}{q}}$ is *co-sequential* (that is *input co-deterministic*) and in substance we have proved:

[5] As W. Allen said: “The infinite is pretty far, especially towards the end”.

Proposition 4.4 *Let x be a real in $[0, \boldsymbol{\omega}_{\frac{p}{q}}]$,* $\mathbf{c}$ *its companion representation and* $\mathbf{a}$ *a $\frac{p}{q}$-expansion of x. Then $(\mathbf{c}, \mathbf{a})$ is the label of an infinite path that begins in the state $\rho_0(x)$ in the transducer $\mathcal{A}_{\frac{p}{q}}$.* ■

If $p \geqslant 2q - 1$, the interesting case which we have already considered, $\mathcal{A}_{\frac{p}{q}}$ has then *only two states.* The transducer $\mathcal{A}_{\frac{3}{2}}$ is drawn at Figure 3.

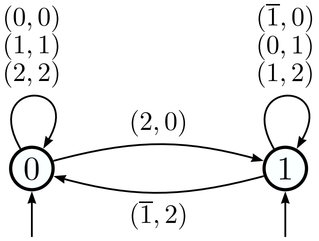


Figure 3: The transducer $\mathcal{A}_{\frac{3}{2}}$

The computation of the companion representation is the first step of the "algorithm" for the computation of $\frac{p}{q}$-expansions of the real numbers.

Let x be in $[0, \boldsymbol{\omega}_{\frac{p}{q}}]$, and let $\mathbf{c}$ be its companion representation. Let n be a fixed large) positive integer and w be the prefix of length n of $\mathbf{c}$. When w is read *from right to left* by the converter $\mathcal{C}_C$ — which is the transposed of $\mathcal{A}_{\frac{p}{q}}$ — and taking a state s as initial state, the output is a word $f^{(s)}$ of length n on the alphabet A and which depends upon s. The maximal common prefix of all these words $f^{(s)}$ is the beginning of all the $\frac{p}{q}$-expansions of x.

To get longer prefixes one has to make again the computation with an n' larger than n, but it is not possible to know in advance how large has to be this n' in order to get a better approximation.

5 On the fractional part of the powers of rational numbers

We are now in a position to give at least a sketch of the proof of the results we have announced in the introduction. In what follows, we suppose, once again, that $p \geqslant 2q - 1$.

For a fixed rational $\dfrac{p}{q}$ we define the subset $Y_{\frac{p}{q}}$ of $[0, 1[$ to be the union of q intervals of length $\dfrac{1}{p}$ by

$$Y_{\frac{p}{q}} = \bigcup_{0 \leqslant c \leqslant q-1} [\frac{1}{p} k_c, \frac{1}{p}(k_c + 1)[$$

where the k_c are such that $k_c \in \{0, \ldots, p-1\}$ and $q\, k_c = c \mod p$. For instance:

$$Y_{\frac{3}{2}} = [0, \frac{1}{3}[\cup[\frac{2}{3}, 1[\,.$$

Theorem 0.3 is then a direct consequence of the following:

Theorem 5.1 *A positive real ξ belongs to $\mathbf{Z}_{\frac{p}{q}}\left(Y_{\frac{p}{q}}\right)$ if and only if ξ has two $\frac{p}{q}$-expansions.*

As one can consider arbitrarily large rationals $\dfrac{p}{q}$, it then comes:

Corollary 5.2 *For any $\varepsilon > 0$, there exists a rational $\frac{p}{q}$ and a subset $Y_{\frac{p}{q}} \subseteq [0, 1[$ of Lebesgue measure smaller than ε such that $\mathbf{Z}_{\frac{p}{q}}\left(Y_{\frac{p}{q}}\right)$ is infinite countable.* ■

The proof of Theorem 5.1 first relies on the characterization of reals with double $\frac{p}{q}$-expansions (Lemma 5.3).

Lemma 5.3 *Let x be in $[0, \boldsymbol{\omega}_{\frac{p}{q}}]$. The following are equivalent:*

(i) *x has more than one expansion;*

(ii) *x has an expansion which is an eventually minimal word;*

(iii) *x has an expansion which is eventually written on the alphabet $\{0, \ldots, q-1\}$;*

(iv) *x has an expansion which is an eventually maximal word;*

(v) *x has an expansion which is eventually written on the alphabet $\{p-q, \ldots, p-1\}$.*

The next step in the proof of Theorem 5.1 is a characterization of the companion representation of the reals that have multiple $\frac{p}{q}$-expansions (and thus two $\frac{p}{q}$-expansions because of the assumption on p and q).

Let us write the digit alphabet $C = \{-(q-1), \ldots, 0, 1, \ldots, (p-1)\}$, the image of the function h, as the union $C = C_1 \cup C_2 \cup C_3$ with $C_1 = \{-(q-1), \ldots, -1\}$, $C_2 = \{0, \ldots, q-1\}$ and $C_3 = \{q, \ldots, p-1\}$.

Proposition 5.4 *A real x has two $\frac{p}{q}$-expansions if and only if its companion representation is eventually in ${C_2}^{\mathbb{N}}$.*

Proof The condition is necessary for if x has two $\frac{p}{q}$-expansions $\mathbf{a}'$ and $\mathbf{a}''$, then $(\mathbf{c}, (\mathbf{a}', \mathbf{a}''))$ must be the label of an infinite path in the *square of the transducer* $\mathcal{A}_{\frac{p}{q}}$ that goes *outside of the diagonal.* Indeed, if $\mathcal{A}$ is an automaton over an alphabet C, two distinct paths in $\mathcal{A}$ with the same label give a path in the

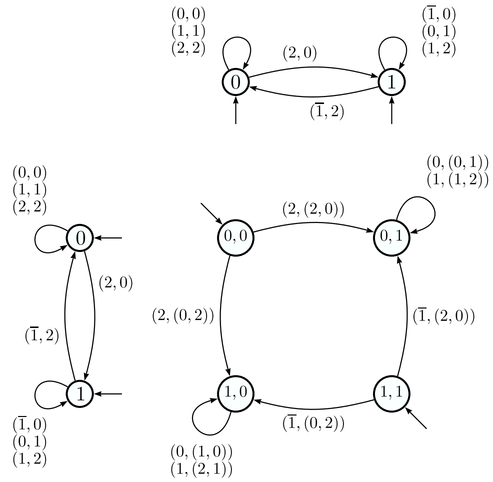


Figure 4: The square of $\mathcal{A}_{\frac{3}{2}}$ (outside of the diagonal)

square of $\mathcal{A}$ that goes outside of the "diagonal". If $\mathcal{T}$ is a transducer, the square $\mathcal{T}^2$ is obtained by constructing the square of the underlying automaton of $\mathcal{T}$ and by giving as output label of each transition of $\mathcal{T}^2$ the pairs of corresponding output labels in $\mathcal{T}$, see [4] or [20].

This implies, as $\mathcal{A}_{\frac{p}{q}}$ has only two states under the current hypothesis, that $\mathbf{c}$ is eventually in C_2 — this can be easily seen on Figure 4 for the case $\frac{p}{q} = \frac{3}{2}$.

Let $\mathbf{c}$ and $\mathbf{a}$ be the companion representation and a $\frac{p}{q}$-representation respectively of a real x. By Proposition 4.4, $(\mathbf{c}, \mathbf{a})$ is the label of an infinite path starting in s in $\mathcal{A}_{\frac{p}{q}}$. Suppose that c_n is the last digit of $\mathbf{c}$ not in C_2 and, by way of example, that it belongs to C_3. Then (c_n, a_n) is the label of a transition that leaves state 0. If $a_n = c_n$, then the infinite word $\mathbf{a}'$ defined by $a'_i = a_i$ for $0 \leqslant i < n$, $a'_n = a_n - q$, and $a'_i = a_i + p - q$ for $n < i$, is such that $(\mathbf{c}, \mathbf{a}')$ is the label of an infinite path in $\mathcal{A}_{\frac{p}{q}}$ with s as initial state — which implies that $\mathbf{a}'$ is a $\frac{p}{q}$-representation of x — and it can be verified that $\mathbf{a}'$ belongs to $W_{\frac{p}{q}}$, which shows that it is a second $\frac{p}{q}$-expansion of x. □

The final step consists in the description of the inverse of the function h. For every c in $C_2 = \{0, \ldots, q-1\}$ let us define the integer k_c in A, *i.e.* $0 \leqslant k_c \leqslant p-1$, by $q\, k_c = c \mod p$.

Lemma 5.5 *For every c in C_2, $h(x) = c$ if and only if* $\left\{\frac{x}{q}\right\} \in [\frac{1}{p}k_c, \frac{1}{p}(k_c+1)[$.

From Proposition 5.4 follows that a real x has two $\frac{p}{q}$-expansions if and only if there exists $M > 0$ such that for any $n > M$,

$$\{\left(\frac{p}{q}\right)^n \frac{x}{q}\} \in Y_{\frac{p}{q}} = \bigcup_{0 \leqslant c \leqslant q-1} [\frac{1}{p}k_c, \frac{1}{p}(k_c+1)[$$

and this concludes the proof of Theorem 5.1.

References

[1] S. Akiyama, Self affine tiling and Pisot numeration system, in *Number theory and its applications*, K. Győry and S. Kanemitsu editors, Kluwer (1999) 7–17.

[2] S. Akiyama, Ch. Frougny and J. Sakarovitch, Powers of rationals modulo 1 and representations in a rational base, *to appear.*

[3] A. Avizienis, Signed-digit number representations for fast parallel arithmetic, *IRE Transactions on electronic computers* **10** (1961) 389–400.

[4] M.-P. Béal, O. Carton, C. Prieur, and J. Sakarovitch, Squaring transducers: An efficient procedure for deciding functionality and sequentiality, *Theoret. Comput. Sci.* **292** (2003) 45–63.

[5] Y. Bugeaud, Linear mod one transformations and the distribution of fractional parts $\{\xi(\frac{p}{q})^n\}$, *Acta Arith.* **114** (2004) 301–311.

[6] A. Cauchy, Sur les moyens d'éviter les erreurs dans les calculs numériques, *C.R. Acad. Sc. Paris* série I **11** (1840) 789–798.

[7] S. Eilenberg, *Automata, Languages and Machines*, Vol. A, Academic Press (1974).

[8] L. Flatto, J.C. Lagarias and A.D. Pollington, On the range of fractional parts $\{\xi(\frac{p}{q})^n\}$, *Acta Arith.* **70** (1995) 125–147.

[9] A.S. Fraenkel, Systems of numeration, *Amer. Math. Monthly* **92** (1985) 105–114.

[10] Ch. Frougny, Representation of numbers and finite automata, *Math. Sys. Th.* **25** (1992) 37–60.

[11] L. Halbeisen and N. Hungerbüler, The Josephus problem, `http://citeseer.nj.nec.com/23586.html`.

[12] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley (1979).

[13] D. Knuth, *The Art of Computer Programming*, Addison Wesley (1969).

[14] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press (2002).

[15] K. Mahler, An unsolved problem on the powers of 3/2, *J. Austral. Math. Soc.* **8** (1968) 313–321.

[16] A. Odlyzko and H. Wilf, Functional iteration and the Josephus problem, *Glasgow Math. J.* **33** (1991) 235–240.

[17] W. Parry, On the β-expansions of real numbers, *Acta Math. Acad. Sci. Hung.*, **11** 401–416.

[18] A.D. Pollington, Progressions arithmétiques généralisées et le problème des $(3/2)^n$. *C.R. Acad. Sc. Paris* série I **292** (1981) 383–384.

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

[20] J. Sakarovitch, *Eléments de théorie des automates*, Vuibert (2003). English translation: *Elements of Automata Theory*, Cambridge University Press, to appear.

[21] N.J.A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, `http://www.research.att.com/~njas/sequences/`.

[22] R. Stephan, On a sequence related to the Josephus problem, `http://arxiv.org/abs/math.CO/0305348` (2003).

[23] T. Vijayaraghavan, On the fractional parts of the powers of a number, I, *J. London Math. Soc.* **15** (1940), 159–160.

[24] E. Weisstein, Power fractional parts, from MathWorld, `http://mathworld.wolfram.com/PowerFractionalParts.html`

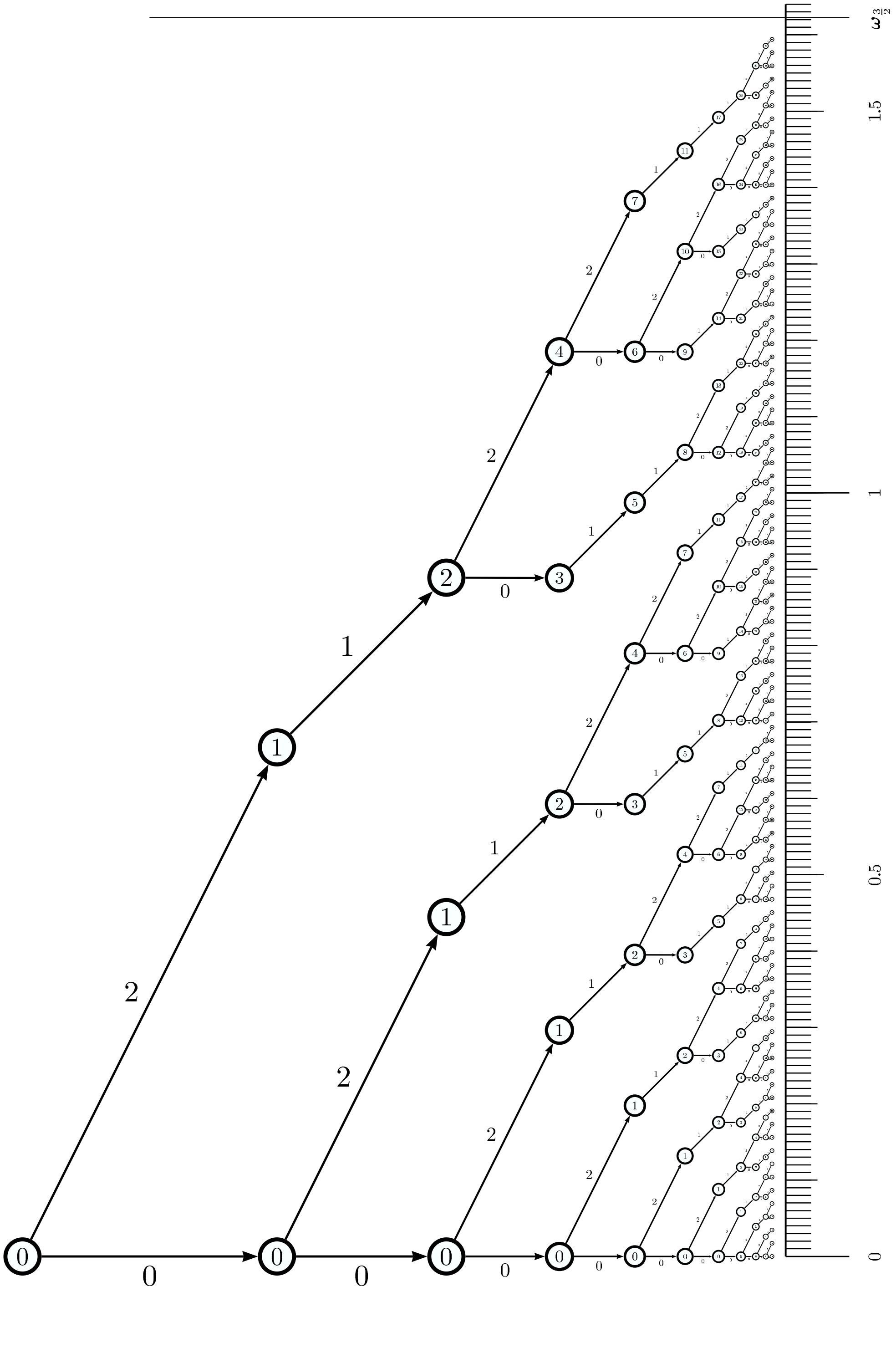
APPENDIX

A The first words in $L_{\frac{3}{2}}$

	0
2	1
21	2
210	3
212	4
2101	5
2120	6
2122	7
21011	8
21200	9
21202	10
21221	11
210110	12
210112	13
212001	14
212020	15
212022	16
212211	17
2101100	18
2101102	19
2101121	20
2120010	21
2120012	22
2120201	23
2120220	24
2120222	25
2122111	26
21011000	27
21011002	28
21011021	29
21011210	30
21011212	31
21200101	32
21200120	33
21200122	34
21202011	35
21202200	36
21202202	37
21202221	38
21221110	39
21221112	40

3/2-expansions of the 41 first integers

B Another view on $T_{\frac{3}{2}}$



Modular and Threshold Subword Counting and Matrix Representations of Finite Monoids

Jorge Almeida[*], *Stuart Margolis*[†], *Benjamin Steinberg*[‡], *Mikhail Volkov*[§]

1 Background and motivation

Recall that a word u over a finite alphabet Σ is said to be a *subword* of a word $v \in \Sigma^*$ if, for some $n \geq 1$, there exist words $u_1, \ldots, u_n, v_0, v_1, \ldots, v_n \in \Sigma^*$ such that $u = u_1 u_2 \cdots u_n$ and

$$v = v_0 u_1 v_1 u_2 v_2 \cdots u_n v_n. \tag{1.1}$$

The subword relation reveals interesting combinatorial properties and plays a prominent role in formal language theory. For instance, recall that languages consisting of all words over Σ having a given word $u \in \Sigma^*$ as a subword serve as a generating system for the Boolean algebra of so-called *piecewise testable* languages. It was a deep study of combinatorics of the subword relation that led Simon [20, 21] to his elegant algebraic characterization of piecewise testable languages. Further, the natural idea to put certain rational constraints on the factors $v_0, v_1, \ldots, v_n$ that may appear in a decomposition of the form (1.1) gave rise to the useful notion of a marked product of languages studied from the algebraic viewpoint by Schützenberger [18], Reutenauer [10], Straubing [23], Simon [22], amongst others.

Yet another natural idea is to count how many times a word $v \in \Sigma^*$ contains a given word u as a subword, that is, to count different decompositions of the form (1.1). Clearly, if one wants to stay within the realm of rational languages, one can only count up to a certain threshold and/or modulo a certain number. For instance, one may consider Boolean combinations of languages consisting of all words over Σ having t modulo p occurrences of a given word $u \in \Sigma^*$ (where p is a given prime number). This class of languages also admits a nice algebraic

[*] Centro de Matemática da Universidade do Porto, Departamento de Matemática Pura, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal, jalmeida@fc.up.pt

[†] Bar Ilan University, 52900 Ramat Gan, Israel, margolis@math.biu.ac.il

[‡] School of Maths & Stats, Carleton University, Herzberg Labs, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6 Canada, bsteinbg@math.carleton.ca

[§] Department of Mathematics and Mechanics, Ural State University, 620083 Ekaterinburg, Russia, Mikhail.Volkov@usu.ru

characterization, see [5, Sections VIII.9 and VIII.10] and also [25]. Combining modular counting with rational constraints led to the idea of marked products with modular counters explored, in particular, by Weil [27] and Peladeau [7].

The most natural version of threshold counting is formalized via the notion of an unambiguous marked product in which one considers words $v \in \Sigma^*$ having exactly one decomposition (1.1) with a given subword u and given rational constraints on the factors $v_0, v_1, \ldots, v_n$. Such unambiguous marked products have been investigated by Schützenberger [19], Pin [8], Pin, Straubing, and Thérien [9], amongst others.

Many known facts on marked products rely on rather difficult techniques from finite monoid theory, namely, on bilateral semidirect product decomposition results of Rhodes *et al.* [14, 16]. These results are proved using Rhodes's classification of maximal proper surmorphisms [6, 11, 15] via case-by-case analysis of the kernel categories of such maps [14, 16]. The aim of the present paper is to give easier and – we hope – more conceptual proofs of several crucial facts about marked products by using matrix representations of finite monoids as a main tool. In particular, we are able to prove the results of Peladeau and Weil in one step, without any case-by-case analysis and without using the machinery of categories. Rather we adapt Simon's analysis of the combinatorics of multiplying upper triangular matrices [22] from the case of Schützenberger products to block upper triangular matrices. We failed to obtain such a purely combinatorial argument for the case of unambiguous products; we still need to use a lemma on kernel categories. Nevertheless we have succeeded in avoiding the decomposition results and case-by-case analysis.

In Section 2 we collect a few facts from the theory of matrix representations of finite monoids. Some of these facts are new; their proofs can be found in the forthcoming paper by the authors [3]. The announced applications to marked products with modular counters and unambiguous marked products are presented in Section 3.

2 Results from Representation Theory

The reader is referred to [4, Chapter 5] and [17] for the basic results of monoid representation theory. All monoids in this paper are assumed to be finite except for the monoid of matrices over an infinite field.

Let M be a monoid and K a field. A (matrix) *representation* of M over K of *degree* n is a homomorphism $\rho : M \to M_n(K)$, where $M_n(K)$ is the monoid of all $n \times n$ matrices over K. Set $V = K^n$. Then a subspace W of V is said to be *M-invariant* if $(M\rho)W \subseteq W$. The representation ρ is said to be *irreducible* if the only M-invariant subspaces are $\{0\}$ and V.

We denote by $K[M]$ the *monoid algebra* of M, that is, the K-algebra with basis M, whose multiplication extends the multiplication of M. Clearly, any representation $\rho : M \to M_n(K)$ uniquely extends to a K-algebra homomorphism

$K[M] \to M_n(K)$. This homomorphism defines a $K[M]$-module structure on the space $V = K^n$. The representation ρ is irreducible if and only if the associated $K[M]$-module is simple. Thus, by choosing a composition series of V, considered as a $K[M]$-module, one can choose a basis for V such that $M\rho$ consists of block upper triangular matrices where the monoids formed by the diagonal blocks are images of M under certain irreducible representations. These irreducible blocks are uniquely determined by ρ and are called the *irreducible constituents* of ρ.

The *regular representation of* M is the representation $\rho_M : M \to M_{|M|}(K)$ on the vector space $K[M]$ extending the homomorphism that maps each element $m \in M$ to the left translation $\lambda_m : m' \mapsto mm'$ of the set M. This is a faithful representation (meaning ρ_M is injective). Moreover, every irreducible representation of M is an irreducible constituent of ρ_M.

If M is a monoid and K is a field, then we define the *Rhodes radical* $\mathrm{Rad}_K(M)$ to be the congruence on M associated to the direct sum of all the irreducible representations of M over K. Equivalently, it is the restriction to M of the congruence on $K[M]$ associated to the Jacobson radical. Alternatively, if we consider the regular representation, placed in block upper triangular form, then the Rhodes radical is the congruence associated to the projection to the block diagonal.

Recall that a *pseudovariety* of monoids (semigroups) is a class of finite monoids (semigroups) closed under the formation of finite direct products, submonoids (subsemigroups) and homomorphic images [1,5]. If $\mathbf{V}$ is a pseudovariety of monoids, then $\mathbf{LV}$ denotes the pseudovariety of semigroups S such that, for each idempotent $e \in S$, the monoid eSe belongs to $\mathbf{V}$. Let $\mathbf{I}$ denote the trivial pseudovariety and $\mathbf{G}_p$ denote the pseudovariety of p-groups for p prime. If $\mathbf{V}$ is a pseudovariety of semigroups, a homomorphism $\varphi : M \to N$ of monoids is called a $\mathbf{V}$-*morphism* if, for each idempotent $f \in N$, one has $f\varphi^{-1} \in \mathbf{V}$.

With this notation, Rhodes showed [12, 17] that if K has characteristic 0, then $\mathrm{Rad}_K(M)$ is the largest congruence $\equiv$ on M such that the quotient $\varphi : M \to M/{\equiv}$ is an $\mathbf{LI}$-morphism. The authors have generalized this [3] to show that if K has characteristic $p > 0$ (a prime), then $\mathrm{Rad}_K(M)$ is the largest congruence $\equiv$ on M such that the quotient $\varphi : M \to M/{\equiv}$ is an $\mathbf{LG}_p$-morphism. Two proofs of these results are given in [3]. The first proof uses the Wedderburn theory of finite dimensional algebras; the second proof uses classical semigroup representation theory and follows along the lines of [12, 17]. One of the key algebraic results used in the first proof, and that we shall use later, is the following, whose proof we include to give the flavor of things. We shall use the fact that a semigroup S is locally a group (in $\mathbf{LG}$ for $\mathbf{G}$ the pseudovariety of groups) if and only if it does not contain a copy of the two element semilattice $\{e, f \mid e = e^2 = ef = fe,\ f = f^2\}$; in this case S is a nilpotent extension of a simple semigroup. By $E(S)$ we denote the set of all idempotents of a semigroup S.

Lemma 2.1 *Let $\varphi : A \to B$ be a morphism of K-algebras with* $\ker\varphi$ *nilpotent. Let S be a finite subsemigroup of A. Then if* $\mathrm{char}K = 0$, *respectively p, then*

$\varphi|_S$ is an $\mathbf{LI}$-morphism, respectively $\mathbf{LG}_p$-morphism.

Proof Without loss of generality, we may assume that S spans A and hence that A is finite dimensional. Let $e_0 \in E(B)$ and $U = e_0\varphi|_S^{-1}$. First we show that U does not contain a copy of the two element semilattice. Indeed, suppose that $e, f \in E(U)$ and $ef = fe = e$. Then

$$(f - e)^2 = f^2 - ef - fe + e^2 = f - e.$$

Since $f - e \in \ker\varphi$, a nilpotent ideal, we conclude $f - e = 0$, that is, $f = e$. As observed before the formulation of the lemma, this means that U is locally a group.

Now let G be a maximal subgroup of U with identity e. Then $g - e \in \ker\varphi$. Since g and e commute, if the characteristic is p, then, for large enough n,

$$0 = (g - e)^{p^n} = g^{p^n} - e$$

and so G is a p-group. Thus $U \in \mathbf{LG}_p$.

If the characteristic is 0, then we observe that $(g-e)^n = 0$ for some n. So by taking the regular representation ρ of G, we see that $g\rho$ is a matrix with minimal polynomial of the form $(x - 1)^n$; that is $g\rho$ is unipotent. A quick consideration of the Jordan canonical form for such $g\rho$ shows that if $g\rho$ is not the identity matrix, then it has infinite order. It follows that $g = e$ and so G is trivial. Thus $U \in \mathbf{LI}$. □

We remark that if A is an algebra of block upper triangular matrices, B is the diagonal block algebra, and φ is the projection to the diagonal block, then the kernel is contained in the algebra of upper triangular matrices with zero diagonal; this algebra is nilpotent and so Lemma 2.1 applies in this context.

We recall that if $\mathbf{V}$ is a pseudovariety of semigroups and $\mathbf{W}$ is a pseudovariety of monoids, then the *Malcev product* $\mathbf{V} \mathbin{\textcircled{m}} \mathbf{W}$ is the pseudovariety generated by all monoids M with a $\mathbf{V}$-morphism to a monoid in $\mathbf{W}$. Given our description of the Rhodes radical, it follows from results of Rhodes and Tilson [6, 13, 26] that $M \in \mathbf{LI} \mathbin{\textcircled{m}} \mathbf{W}$ if and only if $M/\mathrm{Rad}_{\mathbb{Q}}(M) \in \mathbf{W}$ and $M \in \mathbf{LG}_p \mathbin{\textcircled{m}} \mathbf{W}$ if and only if $M/\mathrm{Rad}_{\mathbb{F}_p}(M) \in \mathbf{W}$, where $\mathbb{F}_p$ is the finite field of order p.

3 Applications to Marked Products

In this section we present two applications of representation theory to studying marked products. More can be found in [3].

Recall that Eilenberg established [5, Vol.B, Chap. VII] a correspondence between pseudovarieties of monoids and so-called varieties of languages. If $\mathbf{V}$ is a pseudovariety of monoids and Σ a finite alphabet, then $\mathcal{V}(\Sigma^*)$ denotes the set of all languages over Σ that can be recognized by monoids in $\mathbf{V}$. (Such languages are often referred to as $\mathbf{V}$-*languages*.) The operator $\mathcal{V}$ that assigns each free monoid Σ^* the set $\mathcal{V}(\Sigma^*)$ is said to be *the variety of languages associated to* $\mathbf{V}$. The syntactic monoid [5, loc. cit.] of a rational language L will be denoted M_L. It is known that L is a $\mathbf{V}$-language if and only if $M_L \in \mathbf{V}$.

3.1 Products with Counter

Our first application is to prove the results of Peladeau and Weil [7, 27] on products with counter.

Let $L_0, \ldots, L_m \subseteq \Sigma^*$, $a_1, \ldots, a_m \in \Sigma$ and let n be an integer. Then the *marked product with modulo n counter* $L = (L_0 a_1 L_1 \cdots a_m L_m)_{r,n}$ is the language of all words $w \in \Sigma^*$ with r factorizations modulo n of the form $w = u_0 a_1 u_1 \cdots a_m u_m$ with each $u_i \in L_i$. One can show that L is rational [27] (see also the proof of Theorem 3.2 below). Using a decomposition result of Rhodes and Tilson [14] (see also [16]) based on case-by-case analysis of kernel categories of maximal proper surmorphisms (see [6, 11, 15]), Weil characterized the closure of a variety $\mathcal{V}$ under marked products with modulo p counter. This required iterated usage of the so-called "block product" principle. But Weil missed that the Boolean algebra generated by $\mathcal{V}(\Sigma^*)$ and marked products with modulo p counters of members $\mathcal{V}(\Sigma^*)$ is already closed under marked products with modulo p counters; this was later observed by Peladeau [7]. The difficulty arises because it is not so clear how to combine marked products with modulo p counters into new marked products with modulo p counters.

We use representation theory to prove the result in one fell swoop. Our approach is inspired by a paper of Simon [22] dealing with marked products and the Schützenberger product of finite monoids.

Lemma 3.1 *Let* $\mathbf{V}$ *be a pseudovariety of monoids,* $\varphi : \Sigma^* \to M$ *be a morphism with* M *finite. Let* K *be a field of characteristic* p *and suppose that* M *can be represented faithfully by block upper triangular matrices over* K *so that the monoids formed by the diagonal blocks of the matrices in the image of* M *all belong to* $\mathbf{V}$*. Let* $F \subseteq M$*. Then* $L = F\varphi^{-1}$ *is a Boolean combination of members of* $\mathcal{V}(\Sigma^*)$ *and of marked products with modulo* p *counter* $(L_0 a_1 L_1 \cdots a_n L_n)_{r,p}$ *with the* $L_i \in \mathcal{V}(\Sigma^*)$*.*

Proof Suppose $M \leq M_t(K)$ and $t = t_1 + \cdots + t_k$ is the partition of t giving rise to the block upper triangular form. Let M_i be the monoid formed by the $t_i \times t_i$ matrices over K arising as the i^{th} diagonal blocks of the matrices in the image of M. Given $w \in \Sigma^*$ and $i, j \in \{1, \ldots, k\}$, define $\varphi_{i,j} : \Sigma^* \to M_{t_i, t_j}(K)$ by setting $w\varphi_{i,j}$ to be the $t_i \times t_j$ matrix that is the i, j-block of the block upper triangular form. So in particular $w\varphi_{i,j} = 0$ for $j < i$. Also $\varphi_{i,i}$ is a morphism $\varphi_{i,i} : \Sigma^* \to M_i$ for all i.

First we observe that we may take F to be a singleton $\{u\varphi\}$. For each $1 \leq i \leq j \leq k$, let

$$L_{i,j} = \{w \in \Sigma^* \mid w\varphi_{i,j} = u\varphi_{i,j}\}.$$

Then clearly

$$u\varphi\varphi^{-1} = \bigcap_{1\leq i\leq j\leq k} L_{i,j}.$$

Since $L_{i,i}$ is recognized by M_i, it suffices to show $L_{i,j}$, where $1 \leq i < j \leq k$, can be written as a Boolean combination of marked products with modulo p counter of languages recognized by the M_l. Changing notation, it suffices to show that if $1 \leq i < j \leq k$ and $C \in M_{t_i,t_j}(K)$, then

$$L(C) = \{w \in \Sigma^* \mid w\varphi_{i,j} = C\} \tag{3.1}$$

is a Boolean combination of marked products with modulo p counter of languages recognized by the M_i.

The following definitions are inspired by [22], though what Simon terms an "object", we term a "walk". A *walk* from i to j is a sequence

$$\mathfrak{w} = (i_0, m_0, a_1, i_1, m_1, \ldots, a_r, i_r, m_r) \tag{3.2}$$

where $i = i_0 < i_1 < \cdots < i_r = j$, $a_l \in \Sigma$ and $m_l \in M_{i_l}$. There are only finitely many walks. The set of walks will be denoted $\mathfrak{W}$. Given a walk $\mathfrak{w}$, we define its *value* to be

$$\mathsf{v}(\mathfrak{w}) = m_0(a_1\varphi_{i_0,i_1})m_1 \cdots (a_r\varphi_{i_{r-1},i_r})m_r \in M_{t_i,t_j}(K).$$

If $\mathfrak{w}$ is a walk, we define the *language* of $\mathfrak{w}$ to be the marked product

$$L(\mathfrak{w}) = (m_0\varphi^{-1}_{i_0,i_0})a_1(m_1\varphi^{-1}_{i_1,i_1}) \cdots a_r(m_r\varphi^{-1}_{i_r,i_r}).$$

If $w \in \Sigma^*$ and $\mathfrak{w}$ is a walk of the form (3.2), we define $w(\mathfrak{w})$ to be the *multiplicity* of w in $L(\mathfrak{w})$, that is, the number of factorizations $w = u_0a_1u_1 \cdots a_ru_r$ with $u_l\varphi_{i_l,i_l} = m_l$; this number is taken to be 0 if there are no such factorizations. If $0 \leq n < p$, we establish the shorthand

$$L(\mathfrak{w})_{n,p} = \Big((m_0\varphi^{-1}_{i_0,i_0})a_1(m_1\varphi^{-1}_{i_1,i_1}) \cdots (a_rm_r\varphi^{-1}_{i_r,i_r})\Big)_{n,p}.$$

Notice that $L(\mathfrak{w})_{n,p}$ consists of all words w with $w(\mathfrak{w}) \equiv n \bmod p$ and is a marked product with modulo p counter of $\mathcal{V}(\Sigma^*)$ languages.

The following is a variant of [22, Lemma 7].

Claim 3.1 *Let $w \in \Sigma^*$. Then*

$$w\varphi_{i,j} = \sum_{\mathfrak{w}\in\mathfrak{W}} w(\mathfrak{w})\mathsf{v}(\mathfrak{w}). \tag{3.3}$$

Proof Let $w = b_1 \cdots b_r$ be the factorization of w in letters. Then the formula for matrix multiplication gives

$$w\varphi_{i,j} = \sum (b_1\varphi_{i_0,i_1})(b_2\varphi_{i_1,i_2}) \cdots (b_r\varphi_{i_{r-1},i_r}) \tag{3.4}$$

where the sum extends over all i_l such that $i_0 = i$, $i_r = j$ and $i_l \in \{1, \ldots, k\}$ for $0 < l < r$. Since $v\varphi_{l,n} = 0$ for $l > n$, it suffices to consider sequences such that $i = i_0 \leq i_1 \leq \cdots \leq i_r = j$. For such a sequence, we may group together neighboring indices that are equal. Then using that the $\varphi_{n,n}$ are morphisms, we see that each summand in (3.4) is the value of a walk $\mathfrak{w}$ and that $\mathfrak{w}$ appears exactly $w(\mathfrak{w})$ times in the sum. □

To complete the proof, we observe that $L(C)$ (defined in (3.1)) is a Boolean combination of languages of the form $L(\mathfrak{w})_{n,p}$. Let X be the set of all functions $f : \mathfrak{W} \to \{0, \ldots, p-1\}$ such that

$$\sum_{\mathfrak{w} \in \mathfrak{W}} f(\mathfrak{w})\mathsf{v}(\mathfrak{w}) = C.$$

It is then immediate from (3.3) and $\mathrm{char}K = p$ that

$$L(C) = \bigcup_{f \in X} \bigcap_{\mathfrak{w} \in \mathfrak{W}} L(\mathfrak{w})_{f(\mathfrak{w}),p}$$

completing the proof. □

Theorem 3.2 *Let $L \subseteq \Sigma^*$ be a rational language, $\mathbf{V}$ be a pseudovariety of monoids and K be a field of characteristic p. Then the following are equivalent.*

(1) $M_L \in \mathbf{LG}_p \circledm \mathbf{V}$;

(2) $M_L/\mathrm{Rad}_K(M_L) \in \mathbf{V}$;

(3) *M_L can be faithfully represented by block upper triangular matrices over K so that the monoids formed by the diagonal blocks of the matrices in the image of M_L all belong to $\mathbf{V}$;*

(4) *L is a Boolean combination of members of $\mathcal{V}(\Sigma^*)$ and languages $(L_0a_1L_1 \cdots a_nL_n)_{r,p}$ with the $L_i \in \mathcal{V}(\Sigma^*)$.*

Proof The equivalence of (1) and (2) follows from the results of [3] cited in Section 2.

For (2) implies (3), take a composition series for the regular representation of M_L over K: it is then in block upper triangular form and, by (2) and the comments from Section 2, the monoids formed by diagonal blocks of matrices in the image of M_L all belong to $\mathbf{V}$.

(3) implies (4) is immediate from Lemma 3.1.

For (4) implies (1), it suffices to deal with a marked product with counter $L = (L_0a_1L_1 \cdots a_nL_n)_{r,p}$. Let $\mathcal{A}_i$ be the minimal deterministic automaton for L_i. Let $\mathcal{A}$ be the non-deterministic automaton obtained from the disjoint union of the $\mathcal{A}_i$ by attaching an edge labelled a_i from each final state of $\mathcal{A}_{i-1}$ to the initial state of $\mathcal{A}_i$. To each letter $a \in \Sigma$, we associate the matrix $a\varphi$ of the relation that a induces on the states. Since $a\varphi$ is a $0,1$-matrix, we can view it as a matrix over $\mathbb{F}_p$. In this way we obtain a morphism $\varphi : \Sigma^* \to M_k(\mathbb{F}_p)$ where k is the number of states of $\mathcal{A}$. Let $M = \Sigma^*\varphi$. Trivially, M is finite. We observe that M is block upper triangular with diagonal blocks the syntactic monoids M_{L_i} (the partition of k arises from taking the states of each $\mathcal{A}_i$). Notice that M recognizes L, since L consists of all words w such that $(w\varphi)_{s,f} = r$ where s is the start state of $\mathcal{A}_0$ and f is a final state of $\mathcal{A}_n$. Applying Lemma 2.1 to the projection to the diagonal blocks gives that M and its quotient M_L belong to $\mathbf{LG}_p \mathbin{\textcircled{m}} \mathbf{V}$. □

The proof of (4) implies (1) gives a fairly easy argument that marked products of rational languages with mod p counter are rational.

Since the operator $\mathbf{LG}_p \mathbin{\textcircled{m}} (\ )$ is idempotent, we immediately obtain the following result of [7, 27].

Corollary 3.3 *Let* $\mathbf{V}$ *be a pseudovariety of monoids and* $\mathbf{W} = \mathbf{LG}_p \mathbin{\textcircled{m}} \mathbf{V}$. *Then*

1. $\mathcal{W}(\Sigma^*)$ *is the smallest class of languages containing* $\mathcal{V}(\Sigma^*)$, *which is closed under Boolean operations and formation of marked products with modulo* p *counters.*

2. $\mathcal{W}(\Sigma^*)$ *consists of all Boolean combinations of elements of* $\mathcal{V}(\Sigma^*)$ *and marked products with modulo* p *counters of elements of* $\mathcal{V}(\Sigma^*)$.

Some special cases are the following. If $\mathbf{V}$ is the trivial variety of monoids, then $\mathbf{LG}_p \mathbin{\textcircled{m}} \mathbf{V} = \mathbf{G}_p$ and we obtain Eilenberg's result [5, Section VIII.10] that the $\mathbf{G}_p$ languages consist of the Boolean combinations of languages of the form $(\Sigma^* a_1 \Sigma^* \cdots a_n \Sigma^*)_{r,p}$. Notice that $\mathbf{G}_p$ consists of the groups unitriangularizable over characteristic p. The languages over Σ^* associated to $\mathbf{LG}_p \mathbin{\textcircled{m}} \mathbf{Sl}$ (as observed in [2], this pseudovariety consists of the unitriangularizable monoids over characteristic p) are the Boolean combinations of languages of the forms

$$\Sigma^* a \Sigma^* \quad \text{and} \quad (\Sigma_0^* a_1 \Sigma_1^* \cdots a_n \Sigma_n^*)_{r,p}$$

where $\Sigma_i \subseteq \Sigma$.

We remark that Weil shows [27] that closing $\mathcal{V}(\Sigma^*)$ under marked products with modulo p^n counters, for $n > 1$, does not take you out of the $\mathbf{LG}_p \mathbin{\textcircled{m}} \mathbf{V}$-languages.

3.2 Unambiguous Products

Our next application is to recover results of Schützenberger, Pin, Straubing, and Thérien concerning unambiguous products. Our proof of one direction is along the lines of [9] but our usage of representation theory allows us to avoid using results relying on case-by-case analysis of maximal proper surmorphisms and the block product principle.

Let Σ be a finite alphabet, $L_0, \ldots, L_n \subseteq \Sigma^*$ be rational languages and $a_1, \ldots, a_n \in \Sigma$. Then the *marked product* $L = L_0 a_1 L_1 \cdots a_n L_n$ is called *unambiguous* if each word $w \in L$ has exactly one factorization of the form

$$u_0 a_1 u_1 \cdots a_n u_n \,,$$

where each $u_i \in L_i$. We also allow the degenerate case $n = 0$.

We shall need to use a well-known and straightforward consequence of the distributivity of concatenation over union (cf. [9]), namely, that if $L_0, \ldots, L_n$ are disjoint unions of unambiguous marked products of elements of $\mathcal{V}(\Sigma^*)$, then the same is true for any unambiguous product $L_0 a_1 L_1 \cdots a_n L_n$. We also need a lemma about languages recognized by finite monoids of block upper triangular matrices in characteristic 0.

Lemma 3.4 *Let* $\mathbf{V}$ *be a pseudovariety of monoids,* $\varphi : \Sigma^* \to M$ *be a morphism with* M *finite. Let* K *be a field of characteristic* 0 *and suppose that* M *can be represented faithfully by block upper triangular matrices over* K *so that the monoids* $M_1, \ldots, M_k$ *formed by diagonal blocks of matrices in the image of* M *all belong to* $\mathbf{V}$*. Let* $F \subseteq M$*. Then* $L = F\varphi^{-1}$ *is a disjoint union of unambiguous marked products* $L_0 a_1 L_1 \cdots a_n L_n$ *with the* $L_i \in \mathcal{V}(\Sigma^*)$*.*

Proof We induct on the number k of diagonal blocks. If there is only one block we are done.

Now let $k > 1$. We can repartition n into two blocks, one corresponding to the union of the first $k - 1$ of our original blocks and the other corresponding to the last block. The first diagonal block, call it N, is block upper triangular with diagonal blocks $M_1, \ldots, M_{k-1}$; the second is just M_k. By induction, any language recognized by N is a disjoint union of unambiguous marked products $L_0 a_1 L_1 \cdots a_r L_r$ with the $L_i \in \mathcal{V}(\Sigma^*)$. Thus to prove the result, it suffices to show that L is a disjoint union of unambiguous marked products $L_0 a_1 L_1 \cdots a_n L_n$ with the L_i recognized by $N \times M_k$. It is shown in [3] that the projection from M to $N \times M_k$ has locally trivial kernel category (see [14] for the definition). Then [9, Proposition 2.2] shows us that L is a disjoint union of such unambiguous marked products. □

We ask whether there is a simple combinatorial proof of this lemma that avoids the use of [9, Proposition 2.2] along the lines of the proof of Lemma 3.1.

Theorem 3.5 *Let* $L \subseteq \Sigma^*$ *be a rational language,* $\mathbf{V}$ *be a pseudovariety of monoids and* K *a field of characteristic* 0*. Then the following are equivalent.*

(1) $M_L \in \mathbf{LI} \textcircled{m} \mathbf{V}$*;*

(2) $M_L/\mathrm{Rad}_K(M_L) \in \mathbf{V}$*;*

(3) *M_L can be faithfully represented by block upper triangular matrices over K so that the monoids formed by the diagonal blocks of the matrices in the image of M_L all belong to $\mathbf{V}$.*

(4) *L is a disjoint union of unambiguous products $L_0a_1L_1 \cdots a_nL_n$ with the $L_i \in \mathcal{V}(\Sigma^*)$.*

Proof The equivalence of (1) and (2) follows from the results of [3] quoted in Section 2.

For (2) implies (3), take a composition series for the regular representation of M_L over K: it is then in block upper triangular form and by (2) monoids formed by diagonal blocks of matrices in the image of M_L all belong to $\mathbf{V}$.

(3) implies (4) is immediate from Lemma 3.4.

For (4) implies (1), it suffices to deal with a single unambiguous marked product $L = L_0a_1L_1 \cdots a_nL_n$. Let $\mathcal{A}_i$ be the minimal trim [5] deterministic automaton for L_i and let $\mathcal{A}$ be the non-deterministic automaton obtained from the disjoint union of the L_i by attaching an edge labelled a_i from each final state of $\mathcal{A}_{i-1}$ to the initial state of $\mathcal{A}_i$. To each letter $a \in A$, we associate the matrix $a\varphi$ of the relation that a induces on the states. In this way we obtain a morphism $\varphi : \Sigma^* \to M_k(\mathbb{Q})$ where k is the number of states of $\mathcal{A}$. Let $M = \Sigma^*\varphi$. We observe that M is block upper triangular with diagonal blocks the syntactic monoids M_{L_i} (the partition of k arises from taking the states of each $\mathcal{A}_i$). Notice that M recognizes L, since L consists of all words w such that $(w\varphi)_{s,f} > 0$ where s is the start state of $\mathcal{A}_0$ and f is a final state of $\mathcal{A}_n$. First we show that M is finite. In fact, we claim M contains only $0,1$-matrices (and hence must be finite). Indeed, suppose $(w\varphi)_{i,j} > 1$ some i, j. Since each M_{L_i} consists of $0,1$-matrices, we must have that i is a state of some $\mathcal{A}_l$ and j a state of some $\mathcal{A}_r$ with $l < r$. But $(w\varphi)_{i,j}$ is the number of paths labelled by w from i to j in $\mathcal{A}$. Thus if u, v are words reading respectively from the start state of $\mathcal{A}_0$ to i and from j to a final state of $\mathcal{A}_n$ (such exist since the $\mathcal{A}_i$ are trim), then uwv has at least two factorizations witnessing membership in L, contradicting that L was unambiguous. Since the collection of all block upper triangular matrices is an algebra over $\mathbb{Q}$, as is the collection of block diagonal matrices, an application of Lemma 2.1 to the projection to the diagonal blocks gives that $M \in \mathbf{LI} \circledcirc \mathbf{V}$ and so, since $M \twoheadrightarrow M_L$, we have $M_L \in \mathbf{LI} \circledcirc \mathbf{V}$. □

Since the operator $\mathbf{LI} \circledcirc (\ )$ is idempotent, we immediately obtain the following result of [8, 9].

Corollary 3.6 *Let $\mathbf{V}$ be a pseudovariety of monoids and $\mathbf{W} = \mathbf{LI} \circledcirc \mathbf{V}$. Then*

1. *$\mathcal{W}(\Sigma^*)$ is the smallest class of languages containing $\mathcal{V}(\Sigma^*)$, which is closed under Boolean operations and formation of unambiguous marked products.*

2. *$\mathcal{W}(\Sigma^*)$ consists of all finite disjoint unions of unambiguous marked products of elements of $\mathcal{V}(\Sigma^*)$.*

Recall that the Malcev product of the pseudovariety **LI** with the pseudovariety **Sl** of semilattices (idempotent-commutative monoids) is equal to the famous pseudovariety **DA** of all finite monoids whose regular $\mathcal{D}$-classes are idempotent subsemigroups (see [24] for a nice survey of combinatorial, logical and automata-theoretic characterizations of **DA**). Applying the above corollary, one obtains the classical result of Schützenberger [19] that $\mathcal{DA}(\Sigma^*)$ consists of disjoint unions of unambiguous products of the form $\Sigma_0^* a_1 \Sigma_1^* \cdots a_n \Sigma_n^*$ with $\Sigma_i \subseteq \Sigma$ for all i. It is shown in [3], using representation theory, that **DA** consists of precisely those monoids that can be faithfully represented by upper triangular matrices with zeroes and ones on the diagonal over $\mathbb{Q}$.

Acknowledgments

The work of J. Almeida was (partially) supported by the Centro de Matemtica da Universidade do Porto (CMUP), financed by FCT (Portugal) through the programmes POCTI (Programa Operacional "Cincia, Tecnologia, Inovao") and POSI (Programa Operacional "Sociedade da Informao"), with national and European Community structural funds. This work was done in part while J. Almeida was visiting, on a sabbatical fellowship from FCT, the Chair of Algebra and Discrete Mathematics at the Ural State University, whose hospitality is gratefully acknowledged.

S. Margolis was supported by the Excellency Center "Group Theoretic Methods in the Study of Algebraic Varieties" of the Israeli Science Foundation and by the Binational Science Foundation of the USA and Israel, grant 1999298/1.

B. Steinberg was supported in part by NSERC.

M. Volkov was supported by the Federal Education Agency of Russia, grants 49123 and 04.01.437, the Russian Foundation for Basic Research, grant 05-01-00540, and the Federal Science and Innovation Agency of Russia, grant 2227.2003.01.

References

[1] J. Almeida, Finite Semigroups and Universal Algebra, World Scientific, Singapore, 1994.

[2] J. Almeida, S. W. Margolis and M. V. Volkov, *The pseudovariety of semigroups of triangular matrices over a finite field*, RAIRO - Inf. Theor. Appl. **39** (2005), 31–48.

[3] J. Almeida, S. W. Margolis, B. Steinberg and M. V. Volkov, *Representation theory of finite semigroups, semigroup radicals and formal language theory*, in preparation.

[4] A. H. Clifford and G. B. Preston, The Algebraic Theory of Semigroups, Mathematical Surveys No. 7, AMS, Providence, RI, Vol. 1, 1961.

[5] S. Eilenberg, Automata, Languages and Machines, Academic Press, New York, Vol A, 1974; Vol B, 1976.

[6] K. Krohn, J. Rhodes and B. Tilson, *Lectures on the algebraic theory of finite semigroups and finite-state machines*, Chapters 1, 5-9 (Chapter 6 with M. A. Arbib) of The Algebraic Theory of Machines, Languages, and Semigroups, (M. A. Arbib, ed.), Academic Press, New York, 1968.

[7] P. Peladeau, *Sur le produit avec compteur modulo un nombre premier*, RAIRO - Inf. Theor. Appl. **26** (1992), 553–564.

[8] J.-E. Pin, *Propriétés syntactiques du produit non ambigu*, 7^{th} ICALP, Lect. Notes Comp. Sci. **85**, Springer Verlag, Berlin, Heidelberg, New York, (1980), 483–499.

[9] J.-E. Pin, H. Straubing and D. Thérien, *Locally trivial categories and unambiguous concatenation*, J. Pure Applied Algebra **52** (1988), 297–311.

[10] C. Reutenauer, *Sur les variétés de langages et de monoïdes*, Proc. GI Conf. [Lect. Notes Comp. Sci. **67**], Springer-Verlag, 1979, 260–265.

[11] J. Rhodes, *A homomorphism theorem for finite semigroups*, Math. Systems Theory **1** (1967), 289–304.

[12] J. Rhodes, *Characters and complexity of finite semigroups*, J. Combinatorial Theory **6**, (1969), 67–85.

[13] J. Rhodes, *Algebraic theory of finite semigroups: Structure numbers and structure theorems for finite semigroups*, in: Semigroups, ed. K. Folley, Academic Press, New York, 1969, 125–162.

[14] J. Rhodes and B. Tilson, *The kernel of monoid morphisms*, J. Pure Appl. Algebra **62** (1989), 227–268.

[15] J. Rhodes and P. Weil, *Decomposition techniques for finite semigroups using categories, I*, J. Pure Applied Algebra **62** (1989), 269 – 284.

[16] J. Rhodes and P. Weil, *Decomposition techniques for finite semigroups using categories, II*, J. Pure Applied Algebra **62** (1989), 285 – 312.

[17] J. Rhodes and Y. Zalcstein, *Elementary representation and character theory of finite semigroups and its application* in: Monoids and semigroups with applications (Berkeley, CA, 1989), 334–367, World Sci. Publishing, River Edge, NJ, 1991.

[18] M. P. Schützenberger, *On finite monoids having only trivial subgroups*, Inf. Control **8** (1965), 190–194.

[19] M. P. Schützenberger, *Sur le produit de concatenation non ambigu*, Semigroup Forum **13** (1976), 47–75.

[20] I. Simon, Hierarchies of Events of Dot-Depth One, Ph. D. Thesis, University of Waterloo, 1972.

[21] I. Simon, *Piecewise testable events*, Proc. 2nd GI Conf. [Lect. Notes Comp. Sci. **33**], Springer-Verlag, 1975, 214–222.

[22] I. Simon, *The product of rational languages*, in: “Automata, Languages and Programming,” Andrzej Lingas, Rolf Karlsson, and Svante Carlsson (eds.), Berlin, 1993. Springer-Verlag. Lecture Notes in Computer Science **700**, 430-444.

[23] H. Straubing, *A generalization of the Schützenberger product of finite monoids*, Theor. Comp. Sci. **13** (1981), 137–150.

[24] P. Tesson and D. Thérien, *Diamonds are forever: the variety* **DA**, in: Semigroups, Algorithms, Automata and Languages, eds. G. M. S. Gomes, J. É. Pin and P. V. Silva, World Scientific, Singapore, 2002, 475–499.

[25] D. Thérien, *Subword counting and nilpotent groups*, in: Combinatorics on Words, Progress and Perspectives, ed. L. J. Cummings, Academic Press, New York, 1983, 297–305.

[26] B. Tilson, *Appendix to "Algebraic theory of finite semigroups: Structure numbers and structure theorems for finite semigroups": On the p length of p-solvable semigroups: Preliminary results*, in: Semigroups, ed. K. Folley, Academic Press, New York, 1969, 163–208.

[27] P. Weil, *Closure of varieties of languages under products with counter*, J. Comput. System Sci. **45** (1992), 316–339.

On the tau-adic expansions of real numbers

Petr Ambrož[*]

1 Introduction

The most usual way of representation of numbers are the positional numeration systems, that is, the representation of numbers in the form of finite or infinite words over a given alphabet of digits. Several different concepts of these systems have been studied in the past, e.g. usual representations in an integer base (and its generalizations such as p-adic numeration or systems using signed digits), representations in a real base, based on the so-called beta-expansions (introduced by Rényi [13]), or representations with respect to a sequence of integers. A survey of most of these concepts was given by Frougny in Chapter 7 of [9].

In this paper we study another way of representation of numbers, strongly connected with the representations based on β-expansions, when β is an algebraic integer. In a general sense it could be called either β-adic or α-adic expansion – where α is an algebraic conjugate of β of modulus less than 1 – however we deal in the whole paper only with the "simplest" irrational base, namely with the golden mean $\tau = (1+\sqrt{5})/2$, and hence we use the term τ-adic expansion.

The term τ-adic expansion itself reveals our sources of inspiration – the τ-numeration system on one hand and p-adic numbers (representations of numbers in the form of left infinite power series in a prime p) on the other hand. In fact the τ-adic expansion is a representation of a real number x in the form of a (possibly) left infinite power series. However, contrary to p-adic numbers the base of the τ-adic system is not the number τ itself but its algebraic conjugate τ'. This difference implies an important advantage over the usual p-adic expansions. Since the number τ' is in modulus smaller than one we do not have to introduce any special valuation for the series to converge.

Although our use of τ-adic expansions may be new, the deployment of left infinite power series in an irrational number β is not new at all. It has been used by several authors for different purposes. Vershik [17] (probably the first use of the term fibadic expansion) and Sidorov and Vershik [16] use two-sided expansions to show a connection between symbolic dynamics of toral automorphisms and arithmetic expansions associated with their eigenvalues and for study of the Erdös measure (more precisely two-sided generalization of Erdös measure).

[*] Department of Mathematics, FNSPE, Czech Technical University, Prague *and* LIAFA, CNRS UMR 7089, Paris, ampy@linux.fjfi.cvut.cz

Two-sided beta-shifts have been studied in full generality by Schmidt in [15]. Ito and Rao [8], and Berthé and Siegel [4] use representations of two-sided β-shift in their study of purely periodic expansions with Pisot unit and non-unit base.

Left-sided extensions of numeration systems defined by a sequence of integers, like the Fibonacci numeration system, have been introduced by Grabner, Liardet and Tichy [7], and studied from the point of view of the odometer function. The use (at least implicit) of representations infinite to the left is contained in every study of the Rauzy fractal [12], especially in a study of its border, see e.g. Akiyama [1], Akiyama and Sadahiro [2] or Messaoudi [10].

This contribution is organized as follows. In the first part we recall known facts about the τ-numeration, we define τ-adic expansions, and show a connection between these two notions. Further on, we give an algorithm for computing the τ-adic expansion of an integer. As a consequence we show that every integer has its τ-adic expansion eventually periodic to the left with a finite fractional part (Corollary 3.5). Then we study τ-adic expansions of rational numbers. We give an algorithm for computing the τ-adic expansion of a rational number, which implies that any rational number has its τ-adic expansion eventually periodic to the left with a finite fractional part (Corollary 4.6).

Recall that, by the results of Bertrand [5] and Schmidt [14], a positive real number belongs to the field $\mathbb{Q}(\tau)$ if and only if its τ-expansion (which is right infinite) is eventually periodic. Thus it is natural to try to get a similar result for τ-adic expansions. We prove that a real number belongs to the field $\mathbb{Q}(\tau')$ if and only if its τ-adic expansion is eventually periodic to the left with a finite fractional part (Theorem 5.5). Note that the fields $\mathbb{Q}(\tau)$ and $\mathbb{Q}(\tau')$ are identical, but our result includes also negative numbers that means one can represent by τ-adic expansions with positive digits also negative numbers without utilization of the sign.

2 Tau-expansions

Let $\beta > 1$ be a real number. A representation in base β (or simply *β-representation*) of a real number $x \in \mathbb{R}_+$ is an infinite sequence $(x_i)_{n\geq i>-\infty}$, $x_i \in \mathbb{Z}$ such that $x = x_n\beta^n + x_{n-1}\beta^{n-1} + \cdots + x_1\beta + x_0 + x_{-1}\beta^{-1} + \cdots$ for certain $n \in \mathbb{Z}$. It is denoted by $(x)_\beta = x_n x_{n-1} \dots x_1 x_0 \bullet x_{-1} x_{-2} \dots$, most significant digit first.

Among all β-representations of a number x there is one particular – called *β-expansion* – for which the coefficients x_i are non-negative integers and

$$\sum_{i=-\infty}^{N} x_i\beta^i < \beta^{N+1} \qquad \text{for all } -\infty < N < n\,.$$

Every $x \in \mathbb{R}_+$ has a unique β-expansion which is found by the greedy algorithm [13]. The β-expansion of a number x is denoted by $\langle x\rangle_\beta$.

The *normalization* is the function ν which maps an infinite sequence of digits $(x_i)_{n\geq i>-\infty}$ onto the β-expansion $\langle \pi((x_i)_{n\geq i>-\infty})\rangle_\beta$.

If the β-expansion of x ends in infinitely many zeros, it is said to be finite and the ending zeros are omitted. The set of all real numbers x for which the β-expansion of $|x|$ is finite is denoted by $\mathrm{Fin}(\beta)$. Moreover, if the β-expansion of $|x|$ is of the form $|x| = \sum_{i=0}^{n} x_i \beta^i$ we say that x is a *β-integer*. The set of all beta-integers is denoted by $\mathbb{Z}_\beta$.

A sequence of coefficients which corresponds to some β-expansion is sometimes called *admissible* in the beta-numeration system. For the characterization of admissible sequences of coefficients one needs to introduce the so-called *Rényi expansion of* 1,

$$d_\beta(1) := t_1 t_2 t_3 \ldots, \quad \text{where } t_1 = \lfloor \beta \rfloor \text{ and } \sum_{n=2}^{\infty} \frac{t_n}{\beta^n} \text{ is the } \beta\text{-expansion of } 1 - \frac{t_1}{\beta}\,.$$

The β-expansions (or β-admissible sequences) are then characterized by the *Parry condition* [11]: the sequence $(x_i)_{n \geq i \geq m}$ with $x_i \in \mathcal{A}_\beta \equiv \{0, 1, \ldots, \lfloor \beta \rfloor\}$ is a β-expansion of some $x > 0$ if and only if $x_{n-p} x_{n-p-1} \ldots x_m$ is lexicographically smaller than $d_\beta(1)$ for all $0 \leq p \leq n - m$.

Recall that Pisot number is an algebraic integer $\beta > 1$ with all its algebraic conjugates in modulus smaller than 1. The golden mean τ is the smallest one among all totally real Pisot numbers, i.e. Pisot numbers with all their algebraic conjugates real. It is the root of the polynomial $x^2 - x - 1$, its conjugate is denoted τ'. Obviously $\tau' = -\frac{1}{\tau}$.

The Rényi development of one in the τ-numeration system is $d_\tau(1) = 11$, hence admissible τ-expansions are sequences over the alphabet $\mathcal{A}_\tau = \{0, 1\}$ not containing the word 11 as a factor.

From now on the symbol τ will stand for the golden mean and τ' for its algebraic conjugate.

3 Tau-adic expansions of integers

Definition 3.1 A *τ-adic representation* of a real number $x \in \mathbb{R}$ is a left-infinite sequence $(d_i)_{-n \leq i < \infty}$, $d_i \in \mathbb{Z}$, $n \in \mathbb{Z}_+$ such that

$$x = \sum_{-n \leq i}^{\infty} d_i (\tau')^i\,.$$

It is denoted ${}_{\tau'}(x) := \cdots d_1 d_0 \bullet d_{-1} \cdots d_{-n}$. The *value of a τ-adic representation* is obtained by the function $\pi : \mathbb{Z}^* \mapsto \mathbb{R}$ given by $\pi((d_i)_{-n \leq i < \infty}) := \sum_{i=-n}^{\infty} d_i (\tau')^i$.

If all finite factors of the sequence $(d_i)_{-n \leq i < \infty}$ are admissible in the τ-numeration system, the sequence $(d_i)_{-n \leq i < \infty}$ is said to be the *τ-adic expansion* of the number x, and it is denoted ${}_{\tau'}\langle x \rangle$.

In this section we are concerned with the τ-adic expansions of integers. We know [6] that $\mathrm{Fin}(\tau) = \mathbb{Z}[\tau] = \{a + b\tau \mid a, b \in \mathbb{Z}\}$. Hence for any $z \in \mathbb{Z}_+$ there exist $m, n \in \mathbb{Z}$, $m \leq n$ such that $z = \sum_{i=m}^{n} z_i \tau^i$ thus $z = \sum_{i=m}^{n} z_i (\tau')^i$, i.e. for each $z \in \mathbb{Z}_+$ the τ-adic expansion ${}_{\tau'}\langle z \rangle = \langle z \rangle_\tau$ is a finite word with some possible fractional part.

Example 3.2 The number -1 has two τ-adic expansions

$$^\omega(10)10 \bullet 0 = {}_{\tau'}\langle -1 \rangle \qquad \text{and} \qquad ^\omega(01)00 \bullet 1 = {}_{\tau'}\langle -1 \rangle\,, \tag{3.1}$$

but the number 1 has only one τ-adic expansion which is the same as its expansion in the τ-numeration system, obviously it is 1.

One can easily see that τ-adic expansions of the number -2 are

$$^\omega(01)000 \bullet 1 = {}_{\tau'}\langle -2 \rangle \qquad \text{and} \qquad ^\omega(10)010 \bullet 1 = {}_{\tau'}\langle -2 \rangle\,,$$

whereas the number 2 has only one expansion (τ-adic expansion as well as expansion in the τ-numeration system), $\langle 2 \rangle_\tau = {}_{\tau'}\langle 2 \rangle = 10 \bullet 01$.

Using the fact that the set $\mathrm{Fin}(\tau)$ forms a ring [6], the following proposition can be proved by the induction.

Proposition 3.3 *Let $z \in \mathbb{Z}_-$ be a negative integer. Its τ-adic expansion ${}_{\tau'}\langle z \rangle$ is a left infinite, eventually periodic word of the form ${}_{\tau'}\langle z \rangle = {}^\omega(10)v$, where $v = \langle x \rangle_\tau$, $x \in \mathrm{Fin}(\tau)$.*

The proof of Proposition 3.3 gives us an algorithm for the computation of the τ-adic expansion of an integer $z \in \mathbb{Z}_-$ by successively subtracting 1 from the τ-adic expansion of -1. Indeed, this is not a very efficient algorithm. We give here another one, based on the τ-expansion of the number $-z$ obtained by the greedy algorithm.

Algorithm. Let $z \in \mathbb{Z}_-$ be an integer. We find its τ-adic expansion by the following algorithm.

1. Use the greedy algorithm to find the τ-expansion of the number $-z$.

2. Let us assume $\langle -z \rangle_\tau = \sum_{i=-k}^{l} d_i \tau^i$ such that $d_l \neq 0$ and $d_{-k} \neq 0$. Then ${}_{\tau'}\langle -z \rangle = \sum_{i=-k}^{l} d_i (\tau')^i$.

3. By the transformation $d_i \mapsto -d_i$ for $i = -k, \ldots, l$ we obtain a τ-adic representation of z in the form of a finite word over the alphabet $\{-1, 0\}$.

4. Replace the rightmost occurrence of a coefficient -1 by its own τ-adic expansion ${}^\omega(10)$. Afterward, the representation has a periodic part ${}^\omega(10)$ (to the left) and a pre-period – a finite word over the alphabet $\{-1, 0, 1\}$.

5. Finally, the τ-adic expansion of $-z$ is simply found by the normalization of the pre-period.

Example 3.4 We will find the τ-adic expansion of -4. The τ-expansion of the number 4 is $101 \bullet 01$, so we have[1] $\bar{1}0\bar{1} \bullet 0\bar{1}$ as a τ-adic representation of the number -4. After replacing the rightmost coefficient -1 according to the step 4, we obtain ${}_{\tau'}(-4) = {}^{\omega}(10)1\bar{1}1\bar{1} \bullet 10$. Finally, by normalizing the pre-period (recall that $1\bar{1}\bar{1}$ is a representation of 0)

$$\begin{array}{cccccccc}
1 & \bar{1} & 1 & \bar{1} & \bullet & 1 & 0 & 0 \\
\bar{1} & 1 & 1 & & & & & \\
 & & \bar{1} & 1 & \bullet & 1 & & \\
\hline
0 & 0 & 1 & 0 & \bullet & 2 & 0 & 0 \\
 & & & 1 & \bullet & \bar{1} & \bar{1} & \\
 & 1 & \bar{1} & \bar{1} & & \bar{1} & 1 & 1 \\
\hline
0 & 1 & 0 & 0 & \bullet & 0 & 0 & 1
\end{array}$$

we find the wanted expansion ${}_{\tau'}\langle -4 \rangle = {}^{\omega}(10)0100 \bullet 001$.

Since most of the expansions we will be dealing with will be left infinite eventually periodic, we define the following two sets of numbers.

$$\mathcal{I}_{ep}(\tau') := \left\{ x \in \mathbb{R} \;\middle|\; {}_{\tau'}\langle x \rangle = {}^{\omega}(d_{k+l} \dots d_{k+1}) d_k \dots d_1 d_0 \bullet \right\},$$
$$\mathcal{F}_{ep}(\tau') := \left\{ x \in \mathbb{R} \;\middle|\; {}_{\tau'}\langle x \rangle = {}^{\omega}(d_{k+l} \dots d_{k+1}) d_k \dots d_1 d_0 \bullet d_{-1} \dots d_{-m} \right\},$$

where $k, l, m \in \mathbb{N}$.

Corollary 3.5 *Since any integer $z \in \mathbb{Z}$ has its τ-adic expansion ${}_{\tau'}\langle z \rangle$ eventually periodic to the left with a finite fractional part, we have $\mathbb{Z} \subset \mathcal{F}_{\mathrm{ep}}(\tau')$.*

4 Tau-adic expansions of rational numbers

In this section we will inspect the τ-adic expansions of rational numbers. At first, we restrict ourselves to the case where the number is in modulus smaller than one – we give an algorithm to find the τ-adic expansion of such rational number $q \in \mathbb{Q}$, $|q| < 1$. The given algorithm is a sort of a right to left normalization.

After each step the so far obtained representation of the number q is of the form $0x_2x_1v$ such that $\pi(x_2x_1v) = q$, $v = \langle y \rangle_\tau$, $y \in \mathrm{Fin}(\tau)$ and x_1 is not an integer. The algorithm modifies the prefix $0x_2x_1$ of this representation by adding to it or subtracting from it the word $(1)(-1)(-1)$ multiplied by a suitable rational number. Note that the value of $(1)(-1)(-1)$ is zero. In each step this transformation shifts the last non-integer coefficient more far to the left. Since the starting point of the whole process is a single rational number, after each step there will be at most two non-integer coefficients in the front of the representation.

There are six possible combinations (with the respect to the sign and to the size) of those non-integer coefficients and thus also six possible transformations of the prefix of the so far obtained representation:

[1] the symbol $\bar{1}$ is sometimes used to denote the number -1.

1. $\boldsymbol{x_1 < 0, x_2 \leq 0}$. Then the prefix $(0)(x_2)(x_1)$ is transformed into the prefix $(x_1)(x_2 - x_1)(0)$ by adding the word $(x_1)(-x_1)(-x_1)$.
2. $\boldsymbol{x_1 < 0}$, $\boldsymbol{x_2 > 0}$. The prefix $(0)(x_2)(x_1)$ is transformed into $(x_1)(x_2 - x_1)(0)$ by adding the word $(x_1)(-x_1)(-x_1)$.
3. $\boldsymbol{x_1 > 0}$, $\boldsymbol{x_2 < 0}$. The prefix $(0)(x_2)(x_1)$ is transformed into the prefix $(x_1 - 1)(x_2 + 1 - x_1)(1)$ by adding the word $(x_1 - 1)(1 - x_1)(1 - x_1)$.
4. $\boldsymbol{x_1 \in (0, 1/2]}$, $\boldsymbol{x_2 = 0}$. The prefix $(0)(0)(x_1)$ is transformed into $(x_1)(-x_1)(0)$ by adding the word $(x_1)(-x_1)(-x_1)$.
5. $\boldsymbol{x_1 > 1/2}$, $\boldsymbol{x_2 = 0}$. The prefix $(0)(0)(x_1)$ is transformed into $(x_1 - 1)(1 - x_1)(1)$ by adding the word $(x_1 - 1)(1 - x_1)(1 - x_1)$.
6. $\boldsymbol{x_1 > 0}$, $\boldsymbol{x_2 > 0}$. The prefix $(0)(x_2)(x_1)$ is transformed into $(x_1)(x_2 - x_1)(0)$ by adding the word $(x_1)(-x_1)(-x_1)$.

Example 4.1 We compute the τ-adic expansion of the number $q = \frac{1}{2}$.

$$\begin{array}{ccccccccc}
 & & & & & & 0 & 0 & \frac{1}{2} \\
 & & & & & & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\
\hline
 & & & & & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\
 & & & & & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \\
\hline
 & & & 0 & 0 & -\frac{1}{2} & 1 & 0 & 0 \\
 & & & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & & & \\
\hline
 & & 0 & -\frac{1}{2} & \frac{1}{2} & 0 & 1 & 0 & 0 \\
 & & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & & & & \\
\hline
0 & 0 & -\frac{1}{2} & 0 & 1 & 0 & 1 & 0 & 0 \\
\hline\hline
 & & & {}^{\omega}(0 & 1 & 0) & 1 & 0 & 0
\end{array}$$

The prefix $(0)(0)(-1/2)$ which arose from the second step re-occurred after the fourth step. Indeed, using the same operations will induce the reappearance of this prefix after each even step. The periodicity of the τ-adic expansion is obvious.

Proposition 4.2 *Let $q \in \mathbb{Q}$, $|q| < 1$. Then the non-integer coefficients in each step of the algorithm generating ${}_{\tau'}\langle q \rangle$ are in modulus smaller than or equal to 1.*

Proof We will prove the proposition by the induction on the number of steps of the algorithm. In the first step the statement is valid due to the assumption $|q| < 1$.

Let $(0)(x_2^{(k)})(x_1^{(k)})$ be the prefix of the representation before the k-th step (i.e. before the first step we have $x_2^{(1)} = 0$, $x_1^{(1)} = q$), where $|x_2^{(k)}| < 1$, $|x_1^{(k)}| < 1$

and $x_1^{(k)}$ is the first (from the right) non-integer coefficient. The prefix after the k-th step will be denoted $(x_3^{(k+1)})(x_2^{(k+1)})(x_1^{(k+1)})$. Hence, in the six cases from the description of the algorithm we have

1. $x_1^{(k)} \in (-1,0)$, $x_2^{(k)} \in (-1,0] \;\Rightarrow$
   $x_3^{(k+1)} = x_1^{(k)} \in (-1,0)$, $x_2^{(k+1)} = x_2^{(k)} - x_1^{(k)} \in (-1,1)$, $x_1^{(k+1)} = 0$.
2. $x_1^{(k)} \in (-1,0)$, $x_2^{(k)} \in (0,1) \;\Rightarrow$
   $x_3^{(k+1)} = x_1^{(k)} \in (-1,0)$, $x_2^{(k+1)} = x_2^{(k)} - x_1^{(k)} \in (0,2)$, $x_1^{(k+1)} = 0$
3. $x_1^{(k)} \in (0,1)$, $x_2^{(k)} \in (-1,0) \;\Rightarrow$
   $x_3^{(k+1)} = x_1^{(k)} - 1 \in (-1,0)$, $x_2^{(k+1)} = x_2^{(k)} + 1 - x_1^{(k)} \in (-1,1)$, $x_1^{(k+1)} = 1$.
4. $x_1^{(k)} \in (0,\frac{1}{2}]$, $x_2^{(k)} = 0 \;\Rightarrow$
   $x_3^{(k+1)} = x_1^{(k)} \in (0,\frac{1}{2}]$, $x_2^{(k+1)} = -x_1^{(k)} \in [-\frac{1}{2},0)$, $x_1^{(k+1)} = 0$.
5. $x_1^{(k)} \in (\frac{1}{2},1)$, $x_2^{(k)} = 0 \;\Rightarrow$
   $x_3^{(k+1)} = x_1^{(k)} - 1 \in (-\frac{1}{2},0)$, $x_2^{(k+1)} = 1 - x_1^{(k)} \in (0,\frac{1}{2})$, $x_1^{(k+1)} = 1$.
6. $x_1^{(k)} \in (0,1)$, $x_2^{(k)} \in (0,1) \;\Rightarrow$
   $x_3^{(k+1)} = x_1^{(k)} \in (0,1)$, $x_2^{(k+1)} = x_2^{(k)} - x_1^{(k)} \in (-1,1)$, $x_1^{(k+1)} = 0$.

Therefore, the only case which does not directly fulfill the proposition is the coefficient $x_2^{(k+1)}$ in Case 2. Let us inspect it more closely. Since $x_2^{(k)} > 0$ and $x_1^{(k)} < 0$ the step $(k-1)$ was either of Type 4 or 6:

a. Step $(k-1)$ was of Type 4. This fact gives us more accurate bounds on the prefix before the Step k: $x_2^{(k)} \in (0,\frac{1}{2}]$, $x_1^{(k)} \in [-\frac{1}{2},0)$. Therefore $x_3^{(k+1)} = x_1^{(k)} \in [-\frac{1}{2},0)$, $x_2^{(k+1)} = x_2^{(k)} - x_1^{(k)} \in (0,1]$, $x_1^{(k+1)} = 0$.

b. Step $(k-1)$ was of Type 6. Note that the starting configuration for a step of Type 6, i.e. $x_1^{(k-1)} \in (0,1)$, $x_2^{(k-1)} \in (0,1)$, can be only generated by another preceding step of Type 6. Indeed the step $(k-2)$ was of Type 6. Consequently, all preceding steps up to the first one were of Type 6. This is in the contradiction with the fact, that the algorithm starts in the configuration $(0)(0)(q)$, $q \in (-1,1)$. Therefore the step of Type 6 will never occur in the algorithm.

□

Corollary 4.3 *Since the prefix $(x_3^{(k+1)})(x_2^{(k+1)})(x_1^{(k+1)})$ of the representation after the k-th step of the algorithm is uniquely determined by the prefix $(x_2^{(k)})(x_1^{(k)})$ in the previous step and since according to Lemma 4.2 the values of $x_2^{(k)}$ and $x_1^{(k)}$ have only finitely many possibilities, the τ-adic expansion generated by the algorithm is eventually periodic. This also implies that the algorithm will stop after finitely many steps.*

Corollary 4.4 *Since any rational number $q \in \mathbb{Q}$ which is in modulus smaller than one has its τ-adic expansion ${}_{\tau'}\langle q\rangle$ eventually periodic to the left starting with the coefficient d_0, we have $\mathbb{Q} \cap (-1, 1] \subset \mathcal{I}_{\mathrm{ep}}(\tau')$.*

Now, let us assume that we want to find the τ-adic expansion of a rational number q such that $|q| > 1$. We will discuss separately the cases $q > 1$ and $q < -1$.

Case $q > 1$. Indeed, there are numbers $z \in \mathbb{Z}_+$ and $\hat{q} \in \mathbb{Q} \cap (0, 1)$ such that $q = z + \hat{q}$. We know how to find the τ-adic expansion ${}_{\tau'}\langle z\rangle$, which is a finite word over the alphabet $\mathcal{A}_\tau = \{0, 1\}$ admissible in the τ-numeration system, as well as the τ-adic expansion ${}_{\tau'}\langle \hat{q}\rangle$ which is a word over $\mathcal{A}_\tau$, eventually periodic to the left, admissible in the τ-numeration system and with no fractional part.

Let us assume that ${}_{\tau'}\langle z\rangle = \sum_{i=-k}^{l} d_i(\tau')^i$. If the period of ${}_{\tau'}\langle \hat{q}\rangle$ starts with a power of (τ') less than or equal to l, we shift it to the left so is starts with the power $(\tau')^{l+1}$ and then we decompose this τ-adic expansion in two parts – the pre-period $\mathrm{PreP}({}_{\tau'}\langle \hat{q}\rangle)$ and the period $\mathrm{InP}({}_{\tau'}\langle \hat{q}\rangle)$.

Since $\mathrm{Fin}(\tau)$ is a ring, the τ-adic expansion of the sum ${}_{\tau'}\langle z\rangle + \mathrm{PreP}({}_{\tau'}\langle \hat{q}\rangle)$ is a finite word. To obtain the final result it is enough to concatenate (or add) the period $\mathrm{InP}({}_{\tau'}\langle \hat{q}\rangle)$ to it, with some possible normalization on the point of concatenation.

Example 4.5 We will find the τ-adic expansion of the number $\frac{11}{2}$. We have $z = 5$, ${}_{\tau'}\langle 5\rangle = 1000 \bullet 1001$ and $\hat{q} = \frac{1}{2}$, ${}_{\tau'}\langle \frac{1}{2}\rangle = {}^{\omega}(010)100$. Since the τ-adic expansion of 5 and the periodic part of ${}_{\tau'}\langle \frac{1}{2}\rangle$ are overlapping, we have to shift the period of ${}_{\tau'}\langle \frac{1}{2}\rangle$ one coefficient to the left. We obtain $\mathrm{PreP}({}_{\tau'}\langle \frac{1}{2}\rangle) = 0100$ and $\mathrm{InP}({}_{\tau'}\langle \frac{1}{2}\rangle) = {}^{\omega}(001)0000$. The expansion of $\mathrm{PreP}({}_{\tau'}\langle \frac{1}{2}\rangle) + {}_{\tau'}\langle z\rangle$ is equal to $10000 \bullet 1001$. Hence by adding of $\mathrm{InP}({}_{\tau'}\langle \frac{1}{2}\rangle)$ to it

$$\begin{array}{r}
1\ 0\ 0\ 0\ 0 \bullet 1\ 0\ 0\ 1 \\
{}^{\omega}(0\ 0\ 1)0\ 0\ 0\ 0 \\
\hline
{}^{\omega}(0\ 1\ 0)0\ 2\ 0\ 0\ 0\ 0 \bullet 1\ 0\ 0\ 1 \\
1\ \bar{2}\ 0\ 1 \\
\hline
{}_{\tau'}\langle \frac{11}{2}\rangle = {}^{\omega}(0\ 1\ 0)1\ 0\ 0\ 1\ 0\ 0 \bullet 1\ 0\ 0\ 1
\end{array}$$

we obtain the wanted τ-adic expansion.

Case $q < -1$. In this case we can find $z \in \mathbb{Z}_-$ and $\hat{q} \in \mathbb{Q} \cap (-1, 0)$ such that $q = z + \hat{q}$. The τ-adic expansion ${}_{\tau'}\langle \hat{q}\rangle$ is again a word over $\mathcal{A}_\tau = \{0, 1\}$, eventually periodic to the left, admissible in the τ-numeration system and with no fractional part, whereas the τ-adic expansion ${}_{\tau'}\langle z\rangle$ is a word over the alphabet $\mathcal{A}_\tau$ admissible in the τ-numeration system, eventually periodic to the left with the period ${}^{\omega}(10)$ in this case.

We decompose ${}_{\tau'}\langle z\rangle$ into two parts – the pre-period $\mathrm{PreP}({}_{\tau'}\langle z\rangle)$ and the period $\mathrm{InP}({}_{\tau'}\langle z\rangle)$. Then we obtain the τ-adic expansion of $\mathrm{PreP}({}_{\tau'}\langle z\rangle) + {}_{\tau'}\langle \hat{q}\rangle$

as in the previous Case. Let us assume that the period of ${}_{\tau'}\langle z\rangle$ started with the k-th element of the expansion, then $\mathrm{InP}({}_{\tau'}\langle z\rangle)$ is in fact the τ-adic expansion of the number $-(\tau')^{k-1}$. Therefore, to find the final result it suffices to subtract $(\tau')^{k-1}$ from ${}_{\tau'}\langle \mathrm{PreP}({}_{\tau'}\langle z\rangle) + {}_{\tau'}\langle\hat{q}\rangle\rangle$.

Corollary 4.6 *Since any rational number $q \in \mathbb{Q}$ has its τ-adic expansion eventually periodic to the left with only finite fractional part, we have $\mathbb{Q} \subset \mathcal{F}_{\mathrm{ep}}(\tau')$.*

5 Properties of the set $\mathcal{F}_{\mathrm{ep}}$

We have the following result concerning algebraic properties of the set $\mathcal{F}_{\mathrm{ep}}$. The proof consisting of the construction of an automaton performing deterministic normalization of τ-adic representation over the alphabet $\{0, 1, 2\}$ can be found in [3].

Theorem 5.1 *The set $\mathcal{F}_{\mathrm{ep}}(\tau')$ is closed under addition of positive elements.*

Corollary 5.2 *The set $\mathcal{F}_{\mathrm{ep}}(\tau')$ is closed under addition of any two elements (hence it is closed under addition and subtraction).*

Proof Indeed, it is enough to show that $\mathcal{F}_{\mathrm{ep}}(\tau')$ is closed under subtraction of positive elements. Let $x, y \in \mathcal{F}_{\mathrm{ep}}(\tau')$, $x > y > 0$. We want to find the τ-adic expansion of $x - y$.

The first step is the same as for the addition of positive elements – by simple digit wise addition we find a τ-adic representation of $x - y$, we will denote it by $z = {}_{\tau'}(x - y)$. Obviously, the coefficients of z are from the alphabet $\{-1, 0, 1\}$. Without loss of generality, we can suppose that z has no fractional part.

We define a partition of the representation z into three other representations u, v_{odd} and v_{even}, such that this partition preserves the numerical value $\pi(z) = \pi(u) + \pi(v_{\mathrm{odd}}) + \pi(v_{\mathrm{even}})$ and

- u is obtained from z by putting all the negative coefficients equal to zero
- v_{odd} is obtained from z by keeping only negative coefficients which belong to the odd powers of τ'
- v_{even} is obtained from z by keeping only negative coefficients which belong to the even powers of τ'

Indeed, $\pi(u)$ is a non-negative number belonging to the set $\mathcal{F}_{\mathrm{ep}}(\tau')$. We modify v_{odd} and v_{even} by $v'_{\mathrm{odd}} = v_{\mathrm{odd}} + {}^{\omega}(10) \bullet 11$, $v'_{\mathrm{even}} = v_{\mathrm{even}} + {}^{\omega}(01) \bullet 011$. Hence $\pi(v'_{\mathrm{odd}}) > 0$, $\pi(v'_{\mathrm{even}}) > 0$ and $\pi(v'_{\mathrm{odd}}), \pi(v'_{\mathrm{even}}) \in \mathcal{F}_{\mathrm{ep}}(\tau')$.

Finally, the τ-adic expansion of $x - y$ is obtained by performing two consecutive additions $(u + v'_{\mathrm{odd}}) + v'_{\mathrm{even}}$. The result is an element of $\mathcal{F}_{\mathrm{ep}}(\tau')$ by virtue of Theorem 5.1. □

Example 5.3 Let $z = {}^{\omega}(0\bar{1}\bar{1}1)0\bar{1}01\bar{1}$. Then

$$\begin{aligned} u &= {}^{\omega}(0001)00010\bullet \\ v_{\text{odd}} &= {}^{\omega}(0\bar{1}00)0\bar{1}000\bullet \\ v'_{\text{odd}} &= {}^{\omega}(0001)00010\bullet 11 \\ v_{\text{even}} &= {}^{\omega}(00\bar{1}0)0000\bar{1}\bullet \\ v'_{\text{even}} &= {}^{\omega}(1000)10100\bullet 011 \end{aligned}$$

Corollary 5.4 *The proof of Corollary 5.2 gives an algorithm to compute the τ-adic expansion from a τ-adic representation over any finite alphabet of digits.*

Theorem 5.5 *The field $\mathbb{Q}(\tau')$ (and therefore also the field $\mathbb{Q}(\tau)$) is equal to the set $\mathcal{F}_{\text{ep}}(\tau')$ of all real numbers having eventually periodic τ-adic expansion with a finite fractional part.*

Proof Let $x \in \mathcal{F}_{\text{ep}}(\tau')$, say ${}_{\tau'}\langle x\rangle = {}^{\omega}(d_{k+p}\dots d_{k+1})d_k\dots d_0 \bullet d_{-1}\dots d_{-m}$. The numerical value of ${}_{\tau'}\langle x\rangle$ is

$$\pi({}_{\tau'}\langle x\rangle) = x_1 + \frac{x_2}{1-(\tau')^p} \tag{5.1}$$

where $x_1, x_2 \in \mathbb{Z}[\tau']$, $x_1 = d_{-m}(\tau')^{-m} + \cdots + d_k(\tau')^k$ and $x_2 = d_{k+1}(\tau')^{k+1} + \cdots + d_{k+p}(\tau')^{k+p}$. Equation (5.1) implies $x \in \mathbb{Q}(\tau')$ and hence $\mathcal{F}_{\text{ep}}(\tau') \subset \mathbb{Q}(\tau')$.

Conversely, let $x \in \mathbb{Q}(\tau')$. Say $x = p + q\tau'$ where $p, q \in \mathbb{Q}$. Due to Corollary 4.6 we know that $p, q \in \mathcal{F}_{\text{ep}}$. Since the multiplication by τ' only shifts a τ-adic expansion we have also $q\tau' \in \mathcal{F}_{\text{ep}}$. Finally, as a consequence of Theorem 5.1 (and Corollary 5.2) we have $x = p + q\tau' \in \mathcal{F}_{\text{ep}}(\tau')$ and so $\mathbb{Q}(\tau') \subset \mathcal{F}_{\text{ep}}(\tau')$. □

Corollary 5.6 *The set $\mathcal{F}_{\text{ep}}(\tau')$ is a commutative ring.*

References

[1] S. Akiyama. *Self affine tiling and Pisot numeration system.* In 'Number theory and its applications (Kyoto, 1997)', K. Györy and S. Kanemitsu, editors, volume 2 of *Dev. Math.*, Kluwer Acad. Publ. (1999), 7–17.

[2] S. Akiyama and T. Sadahiro. A self-similar tiling generated by the minimal Pisot number. In 'Proceedings of the 13th Czech and Slovak International Conference on Number Theory (Ostravice, 1997)', volume 6, 9–26, (1998).

[3] P. Ambrož. *Addition of eventually periodic tau-adic expansions.* preprint CTU Prague, (2005).

[4] V. Berthé and A. Siegel. *Purely periodic beta-expansions in the Pisot non-unit case.* Arxiv math. DS/0407282, (2002).

[5] A. Bertrand. *Développements en base de Pisot et répartition modulo* 1. C. R. Acad. Sci. Paris **285** (1977), 419–421.

[6] C. Frougny and B. Solomyak. *Finite beta expansions.* Ergod. Th. and Dynam. Sys. **12** (1992), 713–723.

[7] P. Grabner, P. Liardet, and R. Tichy. *Odometers and systems of numeration.* Acta Arith. **80** (1995), 103–123.

[8] S. Ito and H. Rao. *Purely periodic β-expansions with Pisot unit base.* Proc. of Amer. Math. Soc. **133** (2004), 953–964.

[9] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, (2002).

[10] A. Messaoudi. *Frontière du fractal de Rauzy et système de numération complexe.* Acta Arith. **95** (2000), 195–224.

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

[12] G. Rauzy. *Nombres algébriques et substitutions.* Bull. Soc. Math. France **110** (1982), 147–178.

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

[14] K. Schmidt. *On periodic expansions of Pisot numbers and Salem numbers.* Bull. London Math. Soc. **12** (1980), 269–278.

[15] K. Schmidt. *Algebraic coding of expansive group automorphisms and two-sided beta-shifts.* Monatsh. Math. **129** (2000), 37–61.

[16] N. Sidorov and A. Vershik. *Ergodic properties of the Erdős measure, the entropy of the golden shift, and related problems.* Monatsh. Math. **126** (1998), 215–261.

[17] A. M. Vershik. *The fibadic expansions of real numbers and adic transformations.* Prep. Report Inst. Mittag-Leffler (1991/1992), 1–9.

Palindromes and Pseudo-Palindromes in Episturmian and Pseudo-Palindromic Infinite Words

Vyoma Anne, Luca Q. Zamboni, Ioana Zorca[*]

Abstract

Let A be a finite set of cardinality greater or equal to 2. An infinite word $\omega \in A^{\mathbb{N}}$ is called Episturmian if it is closed under mirror image (meaning if $u = u_1u_2 \cdots u_k$ is a subword of ω, then so is $\bar{u} = u_k \cdots u_2u_1$) and if for every $n \geq 1$ there exists at most one subword u of ω of length n which is right special. We show that if u is a subword of an Episturmian word ω which is a palindrome, then every first return to u is also a palindrome. As a consequence, every Episturmian word begins in an infinite number of distinct palindromes. Our methods extend to the context of pseudo-palindromic infinite words: $\omega \in A^{\mathbb{N}}$ is called a *pseudo-palindromic word* if there exists a bijection $\phi : A \to A$ with ϕ^2 the identity such that for each subword u of ω we have that $\phi(\bar{u})$ is also a subword of ω, and for every $n \geq 1$ there exists at most one subword u of ω of length n which is right special. These words arise naturally in the context of the Fine and Wilf Theorem on k-periods. A factor u of ω is called a *pseudo-palindrome* if $u = \phi(\bar{u})$. We deduce that if u is a subword of a pseudo-palindromic word ω which is a pseudo-palindrome, then every first return to u is also a pseudo-palindrome. In particular, every pseudo-palindromic infinite word begins in an infinite number of distinct pseudo-palindromes.

1 Introduction

Given an infinite word $\omega = \omega_0\omega_1\omega_2 \ldots$ on a finite alphabet, denote by $L_n(\omega)$ the set of all subwords of ω of length n, that is $L_n(\omega) = \{\omega_j\omega_{j+1} \ldots \omega_{j+n-1} \mid j \geq 0\}$. The *complexity function* $p(n) = p_\omega(n)$ is defined as the cardinality of $L_n(\omega)$. A celebrated result of Morse and Hedlund states that an infinite word is ultimately periodic if and only if for some n the complexity $p(n) \leq n$. (See [30]). An infinite word ω is called *Sturmian* if $p(n) = n + 1$ for all $n \geq 1$. Thus amongst all aperiodic infinite words, Sturmian words are those having the smallest complexity. Perhaps the most well known example is the Fibonacci word

$$0100101001001010010100100101001001010010100100101001 0\ldots$$

[*]Department of Mathematics, PO Box 311430, University of North Texas, Denton, TX 76203-1430, USA, v0008@yahoo.com, luca@unt.edu, ioanazor@yahoo.com

defined as the fixed point of the morphism $0 \mapsto 01$ and $1 \mapsto 0$.

The study of Sturmian words was originated by M. Morse and G.A. Hedlund in the 1930's. They showed that Sturmian words provide a symbolic coding of the orbit of a point on a circle under an irrational rotation (c.f. [30]). Sturmian words have since been extensively studied from many different points of view: (c.f. [3], [6], [4], [13], [16], [19], [27], [28], [31]). It is well known that if ω is a Sturmian word, then for each factor $u = u_1u_2\cdots u_n$ with $u_i \in \{0,1\}$ the *mirror image* $\bar{u} = u_nu_{n-1}\cdots u_2u_1$ is also a factor of ω, in other words the language of ω is closed under mirror image. Also the condition $p(n+1)-p(n) = 1$ implies that for each n there exists exactly one word $u \in L_n(\omega)$ which is a prefix (respectively suffix) of more than one (in fact two) words in $L_{n+1}(\omega)$. Such a word is said to be *right special* (respectively *left special*). A word which is both right and left special is called *bispecial*.

Let A be a finite set of cardinality greater or equal to two. An infinite word $\omega \in A^{\mathbb{N}}$ is called *Episturmian* if the language of ω is closed under mirror image and for each n there exists at most one right special factor of length n. It follows directly from the definition that ω contains at most one left special factor of every length, and that each bispecial factor of ω is a palindrome. Also the mirror image of every left special factor is right special, and a palindrome is left special if and only if it is right special.

Episturmian words have been extensively studied by a number of people including Droubay, Justin and Pirillo (cf. [17,23–26,35]). Episturmian words are a generalization of Sturmian words, in fact the set of Sturmian words is precisely the set of binary aperiodic Episturmian words. Episturmian words are also a generalization of Arnoux-Rauzy infinite words (c.f. [3, 9, 14, 29, 32, 33, 36, 37]). In fact every Arnoux-Rauzy infinite word is Episturmian but not conversely. Although every aperiodic Episturmian word is the morphic image of an Arnoux-Rauzy word.

Let ω be an aperiodic Episturmian word, and X_ω the shift orbit closure of X. Then X_ω contains an infinite word $\tilde{\omega}$ all of whose prefixes are left special factors. This word is called the associated *standard Episturmian word* (see [24, 25]). The bispecial prefixes of a standard Episturmian word are palindromes, and hence every standard Episturmian word begins in an infinite number of distinct palindromes.

Let ω be an Episturmian word on the alphabet A with $|A| = k \geq 2$. Let $u \in L(\omega)$. A word $w \in L(\omega)$ is called a *first return word* to u if w contains exactly two occurrences of u, one as a prefix and one as a suffix. It is well known that every word $u \in L(\omega)$ has at most k distinct first returns [26]. Our main result is the following:

Theorem 1.1 *Let ω be an Episturmian infinite word, and $x \in L(\omega)$ a palindrome. Then every first return to x is a palindrome.*

As a consequence we deduce that every Episturmian infinite word begins in an infinite number of distinct palindromes.

2 Proof of Theorem 1.1

Let ω be an aperiodic Episturmian word on the alphabet A and $X = X_\omega$ the associated subshift. Let $L(\omega) = \bigcup_{n\geq 0} L_n(\omega)$ be the set of all factors of ω. We denote the length of a word $w \in L(\omega)$ by $|w|$. We regard the empty word, denoted ε, as the unique word in $L(\omega)$ of length zero.

If u and v are non-empty words in $L(\omega)$ we will write $u \vdash uv$ to mean that for each word $w \in L(\omega)$ with $|w| = |u| + |v|$ if u is a prefix of w then $w = uv$. If it is not the case that $u \vdash uv$ then we will write $\neg(u \vdash uv)$. Similarly we will write $vu \dashv u$ to mean that for each word $w \in L(\omega)$ with $|w| = |u| + |v|$ if u is a suffix of w then $w = vu$. Otherwise we write $\neg(vu \dashv u)$.

Lemma 2.1 *Let $a \in A$ and suppose $axa \in L(\omega)$ where x is a bispecial factor of ω. Then ax is a right special factor of ω.*

Proof Suppose to the contrary that ax is not right special. Then $ax \vdash axa$. If no prefix of ax is right special then $a \vdash ax \vdash axa$ which would imply that X contains the periodic infinite word $axaxaxaxax\dots$, a contradiction. Let v (possibly the empty word) be the longest prefix of x such that av is right special. Since v is a prefix of x it follows that v is also left special, hence bispecial. Whence $\bar{v} = v$. Since we are assuming that ax is not right special, it follows that $|v| < |x|$. Equivalently, we can write $ax = avu$ where u is not the empty word. Since av is right special, it follows that $\bar{v}a = va$ is left special and hence the first letter of u must be a. Set $u = au'$. It follows by maximality of v that $ava \vdash avau' \vdash avau'a = axa$. Since av is a suffix of x we have that ava is a suffix of axa. But this implies that ω is eventually periodic, a contradiction. □

Lemma 2.2 *Let $w \in L(\omega)$ be bispecial and $a \in A$.*

(1) Suppose aw is right special and $a \vdash aw$. Then $wa \vdash waw$ and waw is bispecial.

(2) Suppose wa is left special and $wa \dashv a$. Then $waw \dashv aw$ and waw is bispecial.

Proof Since $L(\omega)$ is closed under mirror image, it suffices to prove (1), and (2) will follow. Let us assume that aw is right special and $a \vdash aw$. Clearly $wa \vdash waw$. We show that waw is bispecial. Since wa is left special and $wa \vdash waw$, we have waw is also left special. Since waw is a palindrome (as w is a palindrome), it follows that waw is also right special and hence bispecial. □

Lemma 2.3 *Let $w \in L(\omega)$ be bispecial and $a \in A$.*

(1) *Suppose aw is right special and $\neg(a \vdash aw)$. Let v (possibly empty) be the longest proper prefix of w with the property that av is right special. Thus $w = vau$ for some $u \in L(\omega)$. Then $wa \vdash wau$ and wau is bispecial.*

(2) *Suppose wa is left special and $\neg(wa \dashv a)$. Let v (possibly empty) be the longest proper suffix of w with the property that va is left special. Thus $w = uav$ for some $u \in L(\omega)$. Then $uaw \dashv aw$ and uaw is bispecial.*

Proof Again it suffices to prove (1). We suppose aw is right special, and v (possibly empty) is the longest proper prefix of w with the property that av is right special. Since v is bispecial it follows that v is a palindrome and hence va is left special and hence a prefix of w. We write $w = vau$ for some $u \in L(\omega)$. The maximality of the length of v implies that $ava \vdash avau$. But since av is right special and $|av| \leq |w|$ it follows that av is a suffix of w and hence ava a suffix of wa. Thus $wa \vdash wau$. We now show wau is bispecial. Since wa is left special and $wa \vdash wau$ we have wau is also left special. But $\overline{wau} = (\overline{au})\overline{w} = \overline{au}w = \overline{au}vau = (\overline{au})\overline{v}au = \overline{vau}au = \overline{w}au = wau$, and hence wau is a palindrome. Thus wau is also right special. □

Remark 2.4 It follows from the previous lemmas that if $w \in L(\omega)$ is bispecial and aw is right special (with $a \in A$), then the shortest bispecial word w' containing w as a proper prefix is of the form wau for some suffix u (possibly empty) of w.

Lemma 2.5 *Let $\mathring{a} \in A$ be the unique bispecial factor of ω of length 1. Then the first returns to $\mathring{a}$ are contained in the set $\{\mathring{a}b\mathring{a} \,|\, b \in A, b \neq \mathring{a}\} \cup \{\mathring{a}\mathring{a}\}$. In particular all first returns to $\mathring{a}$ are palindromes.*

Proof Let $b \in A$ with $b \neq \mathring{a}$. If $\mathring{a}b \in L(\omega)$ then it follows that $b\mathring{a} \in L(\omega)$ and since $b \neq \mathring{a}$ we must have that $b \vdash b\mathring{a}$. Thus $\mathring{a}b \vdash \mathring{a}b\mathring{a}$ and hence $\mathring{a}b\mathring{a}$ is a first return to $\mathring{a}$. It follows that a first return to $\mathring{a}$ which is not of the form $\mathring{a}b\mathring{a}$ for some $b \neq \mathring{a}$ must necessarily be of the form $\mathring{a}\mathring{a}$. □

Lemma 2.6 *Let $\mathring{a} \in A$ be the unique bispecial factor of ω of length 1. Let $b \in A$ with $b \neq \mathring{a}$. Then each first return to b is a palindrome.*

Proof Let w be the shortest bispecial factor of ω containing b. Then by Remark 2.4 and Lemma 2.2 we deduce that $w = ubu$ where u is the longest bispecial not containing b. We note that $b \vdash bu$ and $ub \dashv b$. Thus every first return to b begins in bu.

Let $c \in A$ and suppose that $buc \in L(\omega)$. If $c = b$ then bub is a first return to b and is a palindrome. Next suppose that $c \neq b$. If c does not occur in w then $c \vdash cw$ and $wc \dashv c$. In fact, we know that $c \neq \mathring{a}$ and that $c \vdash c\mathring{a}$ and $\mathring{a}c \dashv c$. If $\neg(c \vdash cw)$ then there would be a bispecial proper prefix x of w such that cx is right special. But then xc would be left special and hence a prefix of w

contradicting that c did not occur in w. Thus, in case c does not occur in w we have that $buc \vdash bucw = bucubu$, and hence $bucub$ is a first return to b. Since u is a palindrome, it follows that $bucub$ is also a palindrome.

Next suppose c occurs in w and hence in u. Let x be the longest bispecial proper prefix of w such that cx is right special. Since $b \vdash bu$ it follows that x is a prefix of u. We write $u = xcy$. Then by maximality of the length of x we have that $cxc \vdash cxcybu$. Since cxc is a suffix of buc it follows that $buc \vdash bucybu$ and hence $bucyb$ is a first return to b. To see that it is a palindrome it suffices to show that ucy is a palindrome. Writing $u = xcy$ with u and x both palindromes we see that $\overline{ucy} = \overline{cy}(\overline{u}) = \overline{cy}u = \overline{cy}xcy = \overline{cy}(\overline{x})cy = \overline{xcy}cy = \overline{u}cy = ucy$ as required. Thus all first returns to b are palindromes. □

Lemma 2.7 *Let $a \in A$ and suppose that $axa \in L(\omega)$ where x is a bispecial factor of ω. Then there exists a bispecial factor z containing x as a proper prefix and such that az is also right special.*

Proof Let w be the shortest bispecial factor containing axa as a subword. It follows from Remark 2.4 that $w = zbz'$ where z is bispecial, $b \in A$, bz is right special, and z' is a suffix of z. By minimality of the length of w we have that axa is not a subword of z nor a subword of z'. If axa is a subword of bz' then $a = b$ and hence az is right special and $|z| \geq |z'| \geq |xa| > |x|$. Otherwise there is an occurrence of axa in w which begins in z and terminates in bz'. In this case there exists $u \in L(\omega)$ such that ub is a prefix of x and au a suffix of z. Thus we deduce that au is right special while ub is left special. Hence u is bispecial and $a = b$. Since za is a prefix of w we have that za is left special and hence az is right special. Finally we claim that $z \neq x$. Since $za \vdash w$, if z were equal to x we would have that $xa \vdash w$ which contains axa implying that ω is eventually periodic, a contradiction. Thus $|z| > |x|$ and hence x is a proper prefix of z. □

Lemma 2.8 *Let $a \in A$ and x be bispecial. Suppose that $axa \in L(\omega)$. Then each first return to axa is a palindrome.*

Proof Let z be the shortest bispecial factor containing x as a proper prefix and such that az is right special. The above lemma guarantees the existence of z.. Thus we can write $z = xau$ for some factor $u \in L(\omega)$. By minimality of the length of z we have that $axa \vdash axau = az$ (and $za \dashv axa$) and hence axa is not a subword of z. Hence every first return to axa begins in $axau$. But axa is a suffix of za. Hence $za \vdash zau$ and zau is bispecial. Thus the only occurrence of axa in zau is as a suffix of za.

Let $b \in A$ and suppose that $axaub \in L(\omega)$. Since $za \dashv axa$ it follows that $zaub \in L(\omega)$. If $b = a$, then $axaua = aza$ is a first return to axa (recall that axa is both a prefix and a suffix of aza and axa does not occur in z.) Since z is a palindrome, so is aza. Next suppose $b \neq a$. If b does not occur in zau then $b \vdash bzau$, whence $axaub \vdash axaubzau$ and hence $axaubza$ is a first return to axa. But this is a palindrome since $axaubza = azbza$ and z is a palindrome.

Finally suppose that b occurs in zau. Let x' be the longest bispecial prefix of zau such that bx' is right special. Thus $zau = x'by'$ for some y', and $bx'b \in L(\omega)$ and $bx'b \vdash bx'by'$. Also, since x' is actually a prefix of z we can write $y' = y''au$, in other words $z = x'by''$. Since $bx'b$ is a suffix of $zaub$ it follows that $zaub \vdash zauby'$. Thus $axauby''a$ is a first return to axa. To see that it is a palindrome it suffices to show that $xauby'' = zby'' = (x'by'')by''$ is a palindrome where both z and x' are palindromes. But $\overline{zby''} = \overline{by''}(\overline{z}) = \overline{by''}z = \overline{by''}x'by'' = \overline{by''}(\overline{x'})by'' = \overline{x'by''}by'' = \overline{z}by'' = zby''$ as required. □

Proof of Theorem 1.1: It suffices to prove the result for aperiodic Episturmian infinite words; in fact if ν is a periodic Episturmian word and $u \in L(\nu)$ then there exists an aperiodic Episturmian word ω for which $u \in L(\omega)$. So suppose ω is an aperiodic Episturmian word and $x \in L(\omega)$ a palindrome. We proceed by induction on $|x|$. If $|x| = 1$, then the result follows from lemmas 2.5 and 2.6. If $|x| = 2$, then the result follows from Lemma 2.8. So suppose x is a palindrome and $|x| \geq 3$. Let z be a first return to x. Set $x = ax'a$ with x' a palindrome. If $a^{-1}za^{-1}$ is a first return to x', then by induction hypothesis $a^{-1}za^{-1}$ is a palindrome and hence so is z. Otherwise $a^{-1}za^{-1}$ contains other occurrences of x' other than the one at the beginning and at the end of the word. But since $a^{-1}za^{-1}$ does not contain any occurrences of $x = ax'a$, it follows that there exists $b \in A$, with $b \neq a$ and $a^{-1}za^{-1}$ contains either $x'b$ or bx' as a subword. Thus x' is either right special or left special, and hence bispecial. In this case that z is a palindrome follows from Lemma 2.8. □

It is well known that Episturmian words are linked to the so-called Fine and Wilf words: Given $\{p_1, p_2, \ldots, p_k\}$ with $\gcd(p_1, p_2, \ldots, p_k) = 1$, the longest non-constant word u having periods $p_1, p_2, \ldots, p_k$ on the maximum number of symbols is a bispecial factor of an Episturmian infinite word (see [10, 23, 35]). In particular it is a palindrome. In [8], the authors define and study a class of infinite words on a finite alphabet A they call *pseudo-palindromic words*: A word $\omega \in A^{\mathbb{N}}$ is pseudo-palindromic if there exists a bijection $\phi : A \to A$ with ϕ^2 equal to the identity map on A, such that a) if $u \in L(\omega)$ then so is $\phi(\bar{u})$ and b) for each $n \geq 1$, ω has at most one right special factor of length n. If ϕ is the identity map, then the above definition coincides with that of Episturmian words. Thus pseudo-palindromic words are a generalization of Episturmian words. In [8] it is shown that given $\{p_1, p_2, \ldots, p_k\}$ with $\gcd(p_1, p_2, \ldots, p_k) = 1$, and $m \geq 1$, a non-constant word u of length m having periods $p_1, p_2, \ldots, p_k$ on the most number of symbols is unique up to isomorphism and is a bispecial factor of a pseudo-palindromic word. So while the longest non-constant word with periods $p_1, p_2, \ldots, p_k$ (on the most number of symbols) is a palindrome, shorter words having the requisite periods are only palindromes up to a map $\phi : A \to A$ with ϕ^2 equal to the identity. Given $\{p_1, p_2, \ldots, p_k\}$ with $\gcd(p_1, p_2, \ldots, p_k) = 1$, and $m \geq 1$, there exists an algorithm (arithmetic and combinatorial) similar to the

one in [35] for generating a word u of length m on the most number of symbols having periods $p_1, p_2, \ldots, p_k$ (see [8]). Also in [8] the authors investigate the sub-word complexity and pseudo-palindromic complexity of aperiodic pseudo-palindromic infinite words. A subword u of a pseudo-palindromic infinite word ω is called a *pseudo-palindrome* if $u = \phi(\bar{u})$.
The results and proofs in this paper may be modified in a straightforward fashion to prove the following:

Theorem 2.9 *Let ω be a pseudo-palindromic infinite word on a finite alphabet A with $|A| \geq 2$. Let $x \in L(\omega)$ be a pseudo-palindrome. Then every first return to x is also a pseudo-palindrome.*

As a consequence we deduce that

Corollary 2.10 *Every pseudo-palindromic infinite word begins in an infinite number of distinct pseudo-palindromes.*

It should be noted that in this context a symbol $a \in A$ is a pseudo-palindrome if and only if $\phi(a) = a$. However it is easily shown that every pseudo-palindromic infinite word begins in a pseudo-palindrome, and hence by the above theorem, in an infinite number of pseudo-palindromes.

3 Concluding Remarks

3.1 Codings of 3-Interval Exchange Transformations

Some of the techniques used in the proof of Theorem 1.1 may be extended to prove an analogous result for natural codings of orbits of points under a 3-interval exchange transformation. More precisely given two real numbers $\alpha > 0$ and $\beta > 0$, with $\alpha + \beta < 1$, one considers the dynamical system $\mathcal{I}$ defined on the interval $[0, 1[$ by

- $\mathcal{I}x = x + 1 - \alpha$ if $x \in [0, \alpha[$,
- $\mathcal{I}x = x + 1 - 2\alpha - \beta$ if $x \in [\alpha, \alpha + \beta[$,
- $\mathcal{I}x = x - \alpha - \beta$ if $x \in [\alpha + \beta, 1[$.

A natural coding of the orbit of a point $x \in [0, 1[$ is an infinite word (x_n) taking values $\{1, 2, 3\}$ according to whether the n-th iterate $\mathcal{I}^n(x)$ lies in the first, second, or third interval. The subword complexity of these words is known to be $p(n) = 2n + 1$. As is the case of Episturmian words, they are also closed under mirror image. However, for each $n \geq 1$ there are exactly two right special factors of length n. Various combinatorial properties of the natural codings of 3-interval exchange transformations were studied by Ferenczi, Holton and Zamboni in [21, 22]. In [22] a combinatorial construction is given for generating the two *standard orbits*; a standard orbit is one in which every prefix is left special. As in the Episturmian case, there is a notion of a directive word (see [24,25]). The directive word is obtained via an arithmetic division algorithm applied to the lengths of the intervals. This arithmetic construction was originally introduced by the authors in an earlier paper [20] and may be viewed as a two-dimensional generalization of the regular continued fraction algorithm. Next a combinatorial construction is applied to the directive word to generate the bispecial factors of the associated symbolic subshift as a function of the arithmetic expansion. In [22] the authors obtain a complete characterization of those sequences of block complexity $2n+1$ which are natural codings of orbits of 3-interval exchange transformations. In [2] the authors show that every natural coding of a 3-interval exchange transformation begins in an infinite number of distinct palindromes.

3.2 Avoidance of Large Palindromes

While every uniformly recurrent infinite word which contains arbitrarily large palindromes is necessarily closed under mirror image, we note that Theorem 1.1 is not merely a consequence of the fact that Episturmian words are closed under mirror image. Consider the sequence $\nu = r_1r_2r_3r_4\ldots \in \{1,2,3\}^{\mathbb{N}}$ defined by

$$\nu = 22\tilde{x_1}33\tilde{x_2}22\tilde{x_3}33\tilde{x_4}22\tilde{x_5}33\tilde{x_6}22\tilde{x_7}33\tilde{x_8}\ldots$$

where $x_1x_2x_3x_4\ldots \in \{0,1\}^{\mathbb{N}}$ is a Sturmian word and $\tilde{0} = 11$ and $\tilde{1} = 1$. Let $\omega \in \{0,1\}^{\mathbb{N}}$ be the infinite word whose run length sequence is ν, that is $\omega = 0^{r_1}1^{r_2}0^{r_3}1^{r_4}0^{r_5}1^{r_6}\ldots$. Then ω is a uniformly recurrent infinite word, closed under mirror image, and does not contain any palindromes of length greater than five[1] [12]. While there are numerous examples of uniformly recurrent binary infinite words closed under mirror image (eg. examples generated using paperfolding sequences), the word ω given above is of lowest complexity amongst all such words [12]. The subword complexity of ω is of the form $p(n) = 4n + k$.

References

[1] P. Alessandri, V. Berthé, Three distance theorems and combinatorics on words, *l'Enseignement Mathématique,* **44** (1998), p. 103–132.

[2] V. Anne, I. Zorca, L. Q. Zamboni, Palindromes in codings of 3-interval exchange transformations, preprint (2005).

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

[4] J. Berstel, P. Séébold, Morphismes de Sturm, *Bull. Belg. Math. Soc.* **1** (1994), 175–189.

[1] A binary infinite word which contains no palindromes of length greater than four is necessarily periodic.

[5] J. Berstel, P. Séébold, Sturmian words, in Algebraic Combinatorics on Words (2001), available at http://www-igm.univ-mlv.fr/∼berstel/Lothaire/index.html

[6] V. Berthé, Fréquences des facteurs des suites sturmiennes, *Theoret. Comp. Sci.,* **165** (1996), p. 295–309.

[7] V. Berthé, C. Holton, L. Q. Zamboni, Initial critical exponent, and powers in Sturmian words, *Acta. Arith.*, in press (2005).

[8] D. Bethoney, D. Heckman, K. Schultz, L. Q. Zamboni, Generalized Fine-Wilf words, preprint (2005).

[9] J. Cassaigne, S. Ferenczi, L. Q. Zamboni, Imbalances in Arnoux-Rauzy sequences, *Ann. Inst. Fourier (Grenoble)*, **50**, facs. 4 (2000), p. 1265–1276.

[10] M.G. Castelli, F. Mignosi, A. Restivo, Fine and Wilf's theorem for three periods and a generalization of Sturmian words, *Theoret. Comput. Sci.,* 218 (1999), p. 83–94.

[11] R. V. Chacon, Weakly mixing transformations which are not strongly mixing, *Proc. Amer. Math. Soc.*, **22** (1969), p. 559–562.

[12] E. Patton, V. Orestis, L. Q. Zamboni, Avoidance of large palindromes in binary words, preprint (2005).

[13] E. Coven, G. A. Hedlund, Sequences with minimal block growth, *Math. Syst. Theor.* **7** no. 3 (1973), p. 138–153.

[14] D. Damanik, L. Q. Zamboni, Combinatorial properties of Arnoux-Rauzy subshifts and applications to Schrödinger operators, *Rev. Math. Phys.,* **15** (2003), p. 745–763.

[15] F. M. Dekking, On the Prouhet-Thue-Morse measure, *Acta Universitatis Carolinae, Mathematica et Physica,* **33** (1992), p. 35–40.

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

[17] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* **255** (2001), p. 539–553.

[18] S. Ferenczi, Les transformations de Chacon : combinatoire, structure géométrique, lien avec les systèmes de complexité 2n+1, *Bull. Soc. Math. France*, **123** no. 2 (1995), p. 271–292.

[19] S. Ferenczi, C. Mauduit, Transcendence of numbers with a low complexity expansion, *J. Number Theory,* **67** (1997), p. 146–161.

[20] S. Ferenczi, C. Holton, L. Q. Zamboni, The structure of three-interval exchange transformations I: an arithmetic study, *Ann. Inst. Fourier (Grenoble)* **51** (2001), p. 861–901.

[21] S. Ferenczi, C. Holton, L. Q. Zamboni, Combinatorics of interval exchange transformations, *Twenty - Eighth International Colloquium on Automata, Languages and Programming*, Lectures Notes in Computer Science, Springer - Verlag (2001), p. 567–578.

[22] S. Ferenczi, C. Holton, L. Q. Zamboni, The structure of three-interval exchange transformations II: a combinatorial description of the trajectories, *J. Analyse Math.* **89** (2003), p. 239–276.

[23] J. Justin, Episturmian words and morphisms (results and conjectures), *Algebraic combinatorics and Computer Science*, Springer Italia, Milan 2001, p. 533–539.

[24] J. Justin, G. Pirillo, Episturmian words and Episturmian morphisms, *Theoret. Comput. Sci.* **302** (2003), p. 1–34.

[25] J. Justin, G. Pirillo, Episturmian words: shifts, morphisms and numeration systems, *Internat. J. Found. Comput. Sci.* **15** (2004), p. 329–348.

[26] J. Justin, L. Vuillon, Return words in Sturmian and Episturmian words, *Theor. Inform. Appl.* **34** (2000), p. 343–356.

[27] F. Mignosi, Infinite words with linear subword complexity, *Theoret. Comput. Sci.* **65** (1989), 221–242.

[28] F. Mignosi, On the number of factors of Sturmian words, *Theoret. Comp. Sci.* **82** (1991), 71–84.

[29] F. Mignosi, L. Q. Zamboni, On the number of Arnoux-Rauzy words, *Acta Arith.*,, **101** (2002), p. 21–129.

[30] M. Morse, G. A. Hedlund, Symbolic dynamics II: Sturmian sequences, *Amer. J. Math.,* **62** (1940), p. 1–42.

[31] G. Rauzy, Mots infinis en arithmétique, in Automata on Infinite Words (M. Nivat and D. Perrin eds.,) Lecture Notes in Computer Science, Vol. 192 (Springer, Berlin 1985) 165–171.

[32] G. Rauzy, Nombres algébriques et substitutions, *Bull. Soc. Math. France* **110** (1982), 147–178.

[33] R. N. Risley, L. Q. Zamboni, A generalization of Sturmian sequences; combinatorial structure and transcendence, *Acta Arith.,* **95** (2000), p. 167–184.

[34] N. B. Slater, Gaps and steps for the sequence $n\theta$ (mod1), *Proc. Cambridge Phil. Soc.,* **63** (1967), 1115–1123.

[35] R. Tijdeman, L. Q. Zamboni, Fine and Wilf words for any periods, *Indag. math. (N.S.),* **14** (2003), p. 135–147.

[36] N. Wozny, L. Q. Zamboni, Frequencies of factors in Arnoux-Rauzy sequences and simultaneous rational approximations, *Acta Arith.* **96** (2001), p. 261–278.

[37] L. Q. Zamboni, Une généralisation du théorème de Lagrange sur le développement en fraction continue, *C.R. Acad. Sci. Paris, Série I,* **327** (1998), p. 527–530.

Generalized Substitutions and Stepped Surfaces

Pierre Arnoux[*], *Valérie Berthé, Damien Jamet*[†]

Abstract

A substitution is a non-erasing morphism of the free monoid. The notion of multidimensional substitution of non-constant length acting on multidimensional words introduced in [AI01, ABS04] is proved to be well-defined on the set of two-dimensional words related to discrete approximations of irrational planes. Such a multidimensional substitution can be associated to any usual Pisot unimodular substitution. The aim of this paper is to try to extend the domain of definition of such multidimensional substitutions. In particular, we study an example of a multidimensional substitution which acts on a stepped surface in the sense of [Jam04, JP04].

1 Introduction

Sturmian words are known to be codings of digitizations of an irrational straight line [KR04,LOTH02]. One could expect from a generalization of Sturmian words that they correspond to a digitization of a hyperplane with irrational normal vector. It is thus natural to consider the following digitization scheme corresponding to the notion of arithmetic planes introduced in [Rev91]: this notion consists in approximating a plane in $\mathbb{R}^3$ by selecting points with integral coordinates above and within a bounded distance of the plane; more precisely, given $\mathbf{v} \in \mathbb{R}^3$, $\mu, \omega \in \mathbb{R}$, the lower (resp. upper) discrete hyperplane $\mathfrak{P}(\mathbf{v}, \mu, \omega)$ is the set of points $\mathbf{x} \in \mathbb{Z}^d$ satisfying $0 \leq \langle \mathbf{x}, \mathbf{v} \rangle + \mu < \omega$ (resp. $0 < \langle \mathbf{x}, \mathbf{v} \rangle + \mu \leq \omega$). Moreover, if $\omega = \sum |v_i| = |\mathbf{v}|_1$, then $\mathfrak{P}(\mathbf{v}, \mu, \omega)$ is said to be standard.

In this latter case, one approximates a plane with normal vector $\vec{v} \in \mathbb{R}^3$ by square faces oriented along the three coordinates planes; for each of the three kinds of faces, one defines a distinguished vertex; the standard discrete plane $\mathfrak{P}(\mathbf{v}, \mu, |\mathbf{v}|_1)$ is then equal to the set of distinguished vertices; after projection on the plane $x + y + z = 0$, along $(1, 1, 1)$, one obtains a tiling of the plane with three kinds of diamonds, namely the projections of the three possible faces. One can code this projection over $\mathbb{Z}^2$ by associating to each diamond the name of the

[*] Institut de Mathématique de Luminy, CNRS UMR 6206, 163 avenue de Luminy, 13288 Marseille Cedex 9 (FRANCE), arnoux@iml.univ-mrs.fr

[†] LIRMM-UMR 5506, Université Montpellier II, 161 rue Ada, 34392 Montpellier Cedex 5 (FRANCE), berthe@lirmm.fr, jamet@lirmm.fr

projected face. These words are in fact three-letter two-dimensional Sturmian words (see e.g. [BV00]).

A generalization of the notion of stepped plane, the so-called discrete surfaces, is introduced in [Jam04]. Roughly speaking, a discrete surface is a union of pointed faces such that the orthogonal projection on the plane $x+y+z=0$ induces an homeomorphism from the discrete surface to the plane. As done for stepped planes, one provides any discrete surface with a coding as a two-dimensional word over a three-letter alphabet. In the present paper, we call discrete surfaces *stepped surfaces*, since such objects are not discrete, in the sense, that they are not subsets of $\mathbb{Z}^3$.

Let us recall that a substitution is a non-erasing morphism of the free monoid. A notion of multidimensional substitution of non-constant length acting on multidimensional words is studied in [AI01, AIS01, ABI02, ABS04], inspired by the geometrical formalism of [IO93, IO94]. These multidimensional substitutions are proved to be well-defined on multidimensional Sturmian words. Such a multidimensional substitution can be associated to any usual Pisot unimodular substitution. The aim of the present paper is to explore the domain of definition of such generalized substitutions. For the sake of clarity, we have chosen to work out in full details the example of [ABS04]. We prove that the image of a stepped surface under the action of this multidimensional substitution is well-defined. Our proofs will be based on a geometrical approach. We then use the functionality and the projection on the plane $x+y+z=0$ along $(1,1,1)$ to recover the corresponding results for multidimensional words.

We work here in the three-dimensional space for clarity issues but all the results and methods presented extend in a natural way to $\mathbb{R}^n$.

2 Basic notions

2.1 One-dimensional substitutions

Let $\mathcal{A}$ be a finite alphabet and let $\mathcal{A}^\star$ be the set of finite words over $\mathcal{A}$. The empty word is denoted by ε. A *substitution* is an endomorphism of the free-monoid $\mathcal{A}^\star$ such that the image of every letter of $\mathcal{A}$ is non-empty. Such a definition naturally extends to infinite or biinfinite words in $\mathcal{A}^{\mathbb{N}}$ and $\mathcal{A}^{\mathbb{Z}}$.

We assume $\mathcal{A} = \{1, \ldots, d\}$. Let σ be a substitution over $\mathcal{A}$. The *incidence matrix* of σ, denoted $M_\sigma = (m_{i,j})_{(i,j)\in\{1,\ldots,d\}^2}$, is defined by:

$$M_\sigma = (|\sigma(j)|_i)_{(i,j)\in\{1,\ldots,d\}^2},$$

where $|\sigma(j)|_i$ is the number of occurrences of i in $\sigma(j)$.

Let $\psi : \mathcal{A}^\star \to \mathbb{N}^d$, $w \mapsto (|w|_i)_{i\in\{1,\cdots,d\}}$ be the Parikh mapping, that is, the homomorphism obtained by abelianization of the free monoid. One has for every $w \in \mathcal{A}^\star$, $\psi(\sigma(w)) = M_\sigma\psi(w)$.

Example 2.1 Let $\sigma : \{1,2,3\} \longrightarrow \{1,2,3\}^\star$ be the substitution defined by $\sigma : 1 \mapsto 13$, $2 \mapsto 1$, $3 \mapsto 2$. Then,

$$M_\sigma = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

A substitution σ is said to be a *Pisot substitution* if the characteristic polynomial of its incidence matrix M_σ admits a dominant eigenvalue $\lambda > 1$ such that all its conjugates α satisfy $0 < |\alpha| < 1$. The incidence matrix of a Pisot substitution is primitive [CS01], that is, it admits a power with positive entries.

Finally, a substitution is said to be *unimodular* if $\det M_\sigma = \pm 1$.

From now on, let σ denote a unimodular Pisot substitution over the three-letter alphabet $\mathcal{A} = \{1,2,3\}$.

2.2 Stepped planes

There are several ways to approximate planes by integer points [BCK04]. Usually, these methods consist in selecting integer points within a bounded distance from the considered plane. Such objects are called *discrete planes.*

Let $\{\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}\}$ be the canonical basis of $\mathbb{R}^3$. We call *unit cube* any translate of the *fundamental unit cube* with integral vertices, that is, any set $\mathbf{x} + \mathcal{C}$ where $\mathbf{x} \in \mathbb{Z}^3$ and $\mathcal{C}$ is the fundamental unit cube:

$$\mathcal{C} = \left\{\lambda_1 \mathbf{e_1} + \lambda_3 \mathbf{e_3} + \lambda_3 \mathbf{e_3} \mid (\lambda_1, \lambda_2, \lambda_3) \in [0,1]^3\right\}.$$

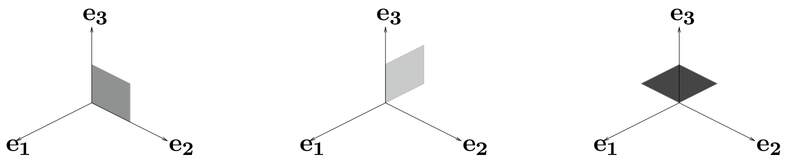
Let $\mathcal{P} : \langle \mathbf{v}, \mathbf{x} \rangle + \mu = 0$, with $\mathbf{v} \in \mathbb{R}^3_+$ and $\mu \in \mathbb{R}$. The *stepped plane* $\mathfrak{P}_\mathcal{P}$ associated to $\mathcal{P}$, also called discrete plane in [ABS04], is defined as the union of the faces of the integral cubes that connect the set $\{\mathbf{x} \in \mathbb{Z}^3 \mid 0 \leq \langle \mathbf{v}, \mathbf{x} \rangle + \mu < \|\mathbf{v}\|_1 = \sum v_i\}$. More precisely:

Definition 2.2 ([IO93, IO94]) We consider the plane $\mathcal{P} : \langle \mathbf{v}, \mathbf{x} \rangle + \mu = 0$, with $\mathbf{v} \in \mathbb{R}^3_+$ and $\mu \in \mathbb{R}$. Let $\mathcal{C}_\mathcal{P}$ be the union of the unit cubes intersecting the open half-space of equation $\langle \mathbf{v}, \mathbf{x} \rangle + \mu < 0$. The *stepped plane* $\mathfrak{P}_\mathcal{P}$ associated to $\mathcal{P}$ is defined by: $\mathfrak{P}_\mathcal{P} = \overline{\mathcal{C}_\mathcal{P}} \setminus \overset{\circ}{\mathcal{C}_\mathcal{P}}$, where $\overline{\mathcal{C}_\mathcal{P}}$ (resp. $\overset{\circ}{\mathcal{C}_\mathcal{P}}$) is the closure (resp. the interior) of the set $\mathcal{C}_\mathcal{P}$ in $\mathbb{R}^3$, provided with its usual topology. The vector $\mathbf{v}$ (resp. μ) is called the *normal vector* (resp. the *translation parameter*) of the stepped plane $\mathfrak{P}_\mathcal{P}$.

It is clear, by construction, that a stepped plane is connected and is a union of faces of unit cubes. In fact, by introducing a suitable definition of faces, we can describe the stepped plane as a partition of such faces, as detailed below.

Let E_1, E_2 and E_3 be the three following *fundamental faces* (see Figure 1):

$$\begin{aligned} E_1 &= \left\{\lambda \mathbf{e_2} + \mu \mathbf{e_3} \mid (\lambda, \mu) \in [0,1[^2\right\}, \\ E_2 &= \left\{-\lambda \mathbf{e_1} + \mu \mathbf{e_3} \mid (\lambda, \mu) \in [0,1[^2\right\}, \\ E_3 &= \left\{-\lambda \mathbf{e_1} - \mu \mathbf{e_2} \mid (\lambda, \mu) \in [0,1[^2\right\}. \end{aligned}$$



(a) E_1.   (b) E_2.   (c) E_3.

Figure 1: The three fundamental faces.

For $\mathbf{x} \in \mathbb{Z}^3$ and $i \in \{1, 2, 3\}$, the *face of type i pointed on* $\mathbf{x} \in \mathbb{Z}^3$ is the set $\mathbf{x} + E_i$. Let us notice that each pointed face includes exactly one integer point, namely, its *distinguished vertex*. As mentioned above we obtain:

Theorem 2.3 ([BV00, ABI02]) *A stepped plane $\mathfrak{P}$ is partitioned by its pointed faces.*

Finally, an easy way to characterize the type of a pointed face included in a stepped plane is given by:

Theorem 2.4 *Let $\mathbf{v} = (v_1, v_2, v_3) \in \mathbb{R}^3_+$ and $\mu \in \mathbb{R}$. Let $\mathfrak{P} = \mathfrak{P}(\mathbf{v}, \mu)$ be the stepped plane with normal vector $\mathbf{v}$ and translation parameter μ. Let $I_1 = [0, v_1[$, $I_2 = [v_1, v_1 + v_2[$ and $I_3 = [v_1 + v_2, v_1 + v_2 + v_3[$. Then,*

$$\forall k \in \{1, 2, 3\}, \forall \mathbf{x} \in \mathfrak{P}, \; x + E_k \subset \mathfrak{P} \iff \langle \mathbf{x}, \mathbf{v} \rangle + \mu \in I_k.$$

Let $\mathcal{P}_0$ be the diagonal plane of equation $x+y+z=0$ and let π be the projection on $\mathcal{P}_0$ along $(1, 1, 1)$.

Theorem 2.5 ([ABI02]) *Let $\mathfrak{P}$ be a stepped plane. The restriction $\pi_{\mathfrak{P}}$ of π from $\mathfrak{P}$ onto $\mathcal{P}_0$ is a bijection. Furthermore, the set of points of $\mathfrak{P}$ with integer coordinates is in one-to-one correspondance with the lattice $\mathbb{Z}\pi(\mathbf{e_1}) + \mathbb{Z}\pi(\mathbf{e_2})$.*

This theorem allows us to code a stepped plane $\mathfrak{P}$ as a two-dimensional word $u \in \{\psi, 2, 3\}^{\mathbb{Z}^2}$ as follows: for all $(m, n) \in \mathbb{Z}^2$, for $i = 1, 2, 3$, then

$$u(m, n) = i \Longleftrightarrow \pi_{\mathfrak{P}}^{-1}(m\pi(\mathbf{e_1}) + n\pi(\mathbf{e_2})) + E_i \subset \mathfrak{P}.$$

2.3 Stepped surfaces

It is thus natural to try to extend the previous definitions and results to more general objects:

Definition 2.6 ([Jam04]) A connected union $\mathfrak{S}$ of pointed faces $\mathbf{x} + E_k$, where $\mathbf{x} \in \mathbb{Z}^3$ and $i \in \{1, 2, 3\}$, is called a *stepped surface* if the restriction $\pi_{\mathfrak{S}} : \mathfrak{S} \longrightarrow \mathcal{P}_0$ of π is a bijection.

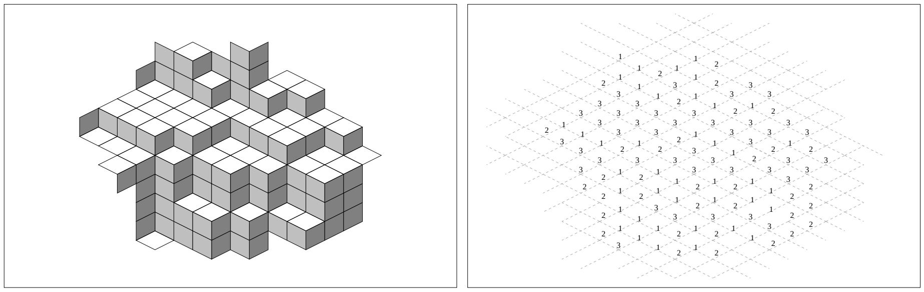


Figure 2: A piece of a stepped surface and its two-dimensional coding.

A two-dimensional word $u \in \{1,2,3\}^{\mathbb{Z}^2}$ is said to be a coding of the stepped surface $\mathfrak{S}$ if for all $(m,n) \in \mathbb{Z}^2$, for $i = 1,2,3$, then

$$u(m,n) = i \Longleftrightarrow \pi_{\mathfrak{S}}^{-1}(m\pi(\mathbf{e_1}) + n\pi(\mathbf{e_2})) + E_i \subset \mathfrak{S}.$$

In particular, a stepped plane is a stepped surface, according to what precedes.

3 Generalized substitutions acting on faces of a stepped plane

The aim of this section is to recall the notion of *generalized substitution* acting on faces of a stepped plane [AI01, AIS01, Pyt02].

Let σ denote a unimodular Pisot substitution over the three-letter alphabet $\mathcal{A} = \{1,2,3\}$. Let M_σ be its incidence matrix, and let $\alpha, \lambda_1, \lambda_2$ denote its eigenvalues with $\alpha > 1 > |\lambda_1| \geq |\lambda_2| > 0$. Let $\mathfrak{P}$ be the *contracting plane* of M_σ, that is, the real plane generated by the eigenvectors associated to λ_1, λ_2.

Since the incidence matrix of a Pisot substitution is primitive [CS01], then, according to Perron-Frobenius Theorem, the eigenvalue α admits a positive eigenvector $\mathbf{v}$. Let us denote by $\mathfrak{P}_\sigma$ the stepped plane with normal vector $\mathbf{v}$ and translation parameter $\mu = 0$.

Example 3.1 We continue Example 2.1. The characteristic polynomial of M_σ is $x^3 - x^2 - 1$; it admits one eigenvalue $\alpha > 1$ (which is known as the second smallest Pisot number), and two complex conjugate eigenvalues of modulus strictly smaller than 1. The contracting plane of M_σ is the plane with equation $\alpha^2 x + \alpha y + z0$.

Definition 3.2 ([IO93, IO94, ABI02, ABS04]) Let σ be a unimodular substitution over the three-letter alphabet $\mathcal{A} = \{1,2,3\}$. Let $\mathfrak{P}_\sigma$ be the stepped plane associated to σ. The *generalized substitution* Σ_σ associated to σ is defined as follows:

$$\Sigma_\sigma(\mathbf{x}+E_i) = \bigcup_{k=1}^{3} \bigcup_{\substack{P \\ \sigma(k)=PiS}} \left(M_\sigma^{-1} \left[\mathbf{x} - \psi(P) - \sum_{j=1}^{i} \mathbf{e_j} \right] \right) + \sum_{j=1}^{k} \mathbf{e_j} + E_k$$

Example 3.3 Let $\sigma : 1 \mapsto 13,\ 2 \mapsto 1,\ 3 \mapsto 2$. Then,

$$\begin{array}{llll} \Sigma_\sigma & : & \mathbf{x}+E_1 & \mapsto & \big(M_\sigma^{-1}\mathbf{x}+\mathbf{e_1}-\mathbf{e_2}+E_1\big) \cup \big(M_\sigma^{-1}\mathbf{x}+\mathbf{e_1}+E_2\big), \\ & & \mathbf{x}+E_2 & \mapsto & M_\sigma^{-1}\mathbf{x}+\mathbf{e_1}+E_3, \\ & & \mathbf{x}+E_3 & \mapsto & M_\sigma^{-1}\mathbf{x}-\mathbf{e_2}-\mathbf{e_3}+E_1. \end{array}$$

In combinatorial terms, Σ_σ can be coded as

$$1 \mapsto \begin{matrix} 2 \\ 1 \end{matrix} \qquad 2 \mapsto 3 \qquad 3 \mapsto 1.$$

Let $r = r(m,n) = -\lceil (\alpha^2 m + \alpha n)/(\alpha^2+\alpha+1) \rceil + 1$. One has:

$$\begin{array}{lll} ((m,n),1) & \mapsto & ((1-n, m-n-r(m,n)-1),1) \\ & & \quad + ((1-n, m-n-r(m,n)),2) \\ ((m,n),2) & \mapsto & ((1-n, m-n-r(m,n)),3) \\ ((m,n),3) & \mapsto & ((1-n, m-n-r(m,n)),1). \end{array}$$

Theorem 3.4 ([AI01]) *Let σ be a unimodular Pisot substitution over the three-letter alphabet $\mathcal{A} = \{1,2,3\}$, let $\mathfrak{P}_\sigma$ be the stepped plane associated to σ and let Σ_σ be the generalized substitution associated to σ.*

i) Two distinct faces have disjoint images under Σ_σ.

ii) The generalized substitution Σ_σ maps any pattern of $\mathfrak{P}_\sigma$ (that is, any finite union of faces of $\mathfrak{P}_\sigma$) on a pattern of $\mathfrak{P}_\sigma$.

iii) $\Sigma_\sigma(\mathfrak{P}_\sigma) \subseteq \mathfrak{P}_\sigma$.

Since Σ_σ is well-defined on $\mathfrak{P}_\sigma$ (according to Theorem 3.4 i)), and since $\mathfrak{P}_\sigma$ is invariant under the action of Σ_σ, it is natural to investigate the action of Σ_σ on any stepped plane. More precisely, given a stepped plane $\mathfrak{P}(\mathbf{v},\mu)$, can we extend the domain of definition of the generalized substitution Σ_σ to the patterns of $\mathfrak{P}(\mathbf{v},\mu)$? In fact:

Theorem 3.5 *Let σ be a unimodular Pisot substitution, let M_σ be its incidence matrix, and let Σ_σ be the generalized substitution associated to σ.*

For any stepped plane $\mathfrak{P}(\mathbf{v},\mu)$ with $\mathbf{v} \in \mathbb{R}^3_+$, one has:

i) The images of two distinct pointed faces of $\mathfrak{P}(\mathbf{v},\mu)$ by Σ_σ are disjoint.

ii) The image of $\mathfrak{P}(\mathbf{v},\mu)$ is included in the stepped plane $\mathfrak{P}({}^tM \cdot \mathbf{v},\mu)$:

$$\Sigma_\sigma(\mathfrak{P}(\mathbf{v},\mu)) \subseteq \mathfrak{P}({}^tM \cdot \mathbf{v},\mu)$$

Proof (Sketch) The proof is based on the same ideas as in the proof of Lemma 2 and 3 in [AI01]. It mainly uses the following geometric interpretation of Theorem 2.4: a pointed face $\mathbf{x} + E_i$ is included in $\mathfrak{P}(\mathbf{v}, \mu)$ if and only the point $\mathbf{x} + \sum_{k=1}^{i-1} \mathbf{e_k}$ is above the plane $\langle \mathbf{v}, \mathbf{x} \rangle + \mu = 0$ while the point $\mathbf{x} + \sum_{k=1}^{i} \mathbf{e_k}$ is below the latter. □

4 Generalized substitutions acting on faces of a stepped surface

4.1 The general case

Since the image of a stepped plane by a generalized substitution is a subset of a stepped plane, it is interesting to investigate the action of generalized substitutions over a more general class of stepped objets, namely, the stepped surfaces. In fact,

Theorem 4.1 *Let $\mathfrak{S}$ be a stepped surface. Let σ be a unimodular Pisot substitution over the three-letter alphabet $\{1, 2, 3\}$ and let Σ_σ be the associated generalized substitution. Then, the image of two distinct pointed faces of $\mathfrak{S}$ are disjoint. Furthermore, the restriction $\pi_{\Sigma_\sigma(\mathfrak{S})} : \Sigma_\sigma(\mathfrak{S}) \longrightarrow \mathcal{P}_0$ is 1-1.*

Proof (Sketch) We first notice that given two faces $\mathbf{x} + E_i$ and $\mathbf{y} + E_j$, then there exists a stepped plane $\mathfrak{P}$ with positive normal vector containing simultaneously $\mathbf{x} + E_i$, $\mathbf{y} + E_j$ and $\mathbf{z} + E_k$. We then apply Theorem 3.5. □

In other words, it remains to prove that $\pi_{\Sigma_\sigma(\mathfrak{S})}$ is onto and that $\Sigma_\sigma(\mathfrak{S})$ is a connected union of faces to deduce that $\Sigma_\sigma(\mathfrak{S})$ is a stepped surface according to Definition 2.6. Let us investigate this problem in the particular case of the generalized substitution Σ_σ associated to the substitution $\sigma : 1 \mapsto 13$, $2 \mapsto 1$, $3 \mapsto 2$.

4.2 The particular case of $\sigma : 1 \mapsto 13$, $2 \mapsto 1$, $3 \mapsto 2$.

In the present section, σ denotes the substitution $\sigma : 1 \mapsto 13$, $2 \mapsto 1$, $3 \mapsto 2$ whereas Σ_σ is the generalized substitition associated to σ:

$$\begin{array}{rrcl} \Sigma_\sigma : & \mathbf{x} + E_1 & \mapsto & \left(M_\sigma^{-1}\mathbf{x} + \mathbf{e_1} - \mathbf{e_2} + E_1\right) \cup \left(M_\sigma^{-1}\mathbf{x} + \mathbf{e_1} + E_2\right), \\ & \mathbf{x} + E_2 & \mapsto & M_\sigma^{-1}\mathbf{x} + \mathbf{e_1} + E_3, \\ & \mathbf{x} + E_3 & \mapsto & M_\sigma^{-1}\mathbf{x} - \mathbf{e_2} - \mathbf{e_3} + E_1. \end{array}$$

Let us show that for this substitution, then the image of a stepped surface is still a stepped surface. First, given a two-dimensional word $u \in \{1, 2, 3\}^{\mathbb{Z}^2}$, we call *hook-word* a factor of u with the following shape (see Fig. 3):

The set of hook-words of u with a hook-shape is called the hook-language of u. In [Jam04, JP04], the authors reduced the recognition problem of the two-dimensional words coding discrete surfaces to a hook recognition problem. More precisely,

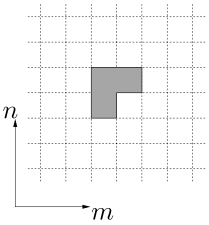


Figure 3: Hook-shape.

Theorem 4.2 ([Jam04, JP04]) *Let $u \in \{1,2,3\}^{\mathbb{Z}^2}$. Then u is a coding of a discrete surface in the sense of Definition 2.6 if and only if the hook-language of u is included in the following set of patterns (see Fig. 4).*

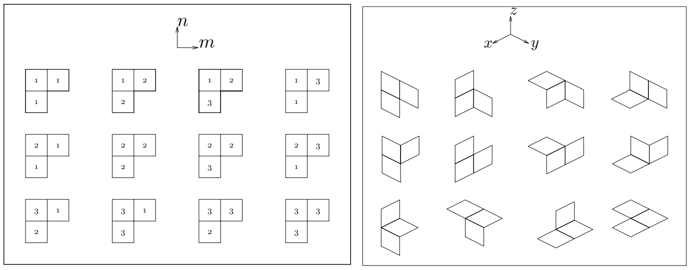


Figure 4: Left: The permitted hook-words. Right: The 3-dimensional representation of the permitted hook-words.

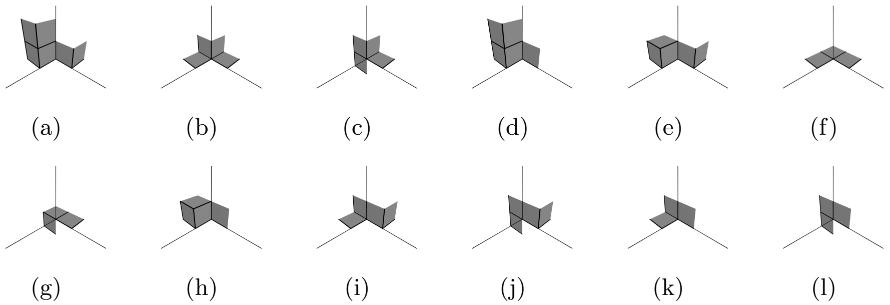
We conversely associate to each permitted hook-word its 3-dimensional representation as a connected union of faces as depicted in Figure 4: the coding of any occurrence of this 3-dimensional representation in a stepped surface is equal to the corresponding hook-word.

Proposition 4.3 *The image by Σ_σ of all the 3-dimensional representations of the permitted hooks (see Fig. 5) are connected in $\mathbb{R}^3$.*

We then deduce that:

Theorem 4.4 *The image of a stepped surface $\mathfrak{S}$ by Σ_σ is connected and the restriction of the projection map π to the latter is injective. Furthermore, all the hook-words occurring in the coding with respect to the injective projection $\pi_{\Sigma_\sigma(\mathfrak{S})}$ (see Theorem 4.1) are permitted hook-words.*

Proof (Sketch) According to Theorem 4.1, the image of a stepped surface by Σ_σ is well-defined. The connectedness follows from Proposition 4.3. Consider now a union H of three faces whose coding according to the injective projection $\pi_{\Sigma_\sigma}(\mathfrak{S})$ (see Theorem 4.1) is a hook-word U_H. There exist (at most) three faces



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

(g) (h) (i) (j) (k) (l)

Figure 5: The image of the permitted hooks by Σ_σ.

of which the union of the images by Σ_σ contains H. One checks that the distance (defined as $d(\mathbf{v}, \mathbf{w}) = |\mathbf{w} - \mathbf{v}|_1$) between the distinguished vertices of those faces is uniformly bounded. By performing a finite case study, one checks that the hook-word U_H is permitted. □

4.2.1 Remarks.

Given a stepped surface $\mathfrak{S}$ containing the unit cube

$$\{\mathbf{e_1} + E_1,\, \mathbf{e_1} + \mathbf{e_2} + E_2,\, \mathbf{e_1} + \mathbf{e_2} + \mathbf{e_3} + E_3, \}\ ,$$

then the sequence of stepped surfaces $(\Sigma_\sigma^n(\mathfrak{S}))_{n\in\mathbb{N}}$ seems to converge towards the stepped plane $\mathcal{P}_\sigma$ (see Fig. 6); to be more precise, the limit points of the sequence $(\Sigma_\sigma^n(\mathfrak{S}))_{n\in\mathbb{N}}$ are subsets of $\mathcal{P}_\sigma$. We will investigate these convergence results and more generally, the possibility of extension of the domain of definition of these multidimensional substitutions to any stepped surface in a subsequent paper. Let us note that this study can also be applied to obtain an efficient generation methods of stepped planes and surfaces.

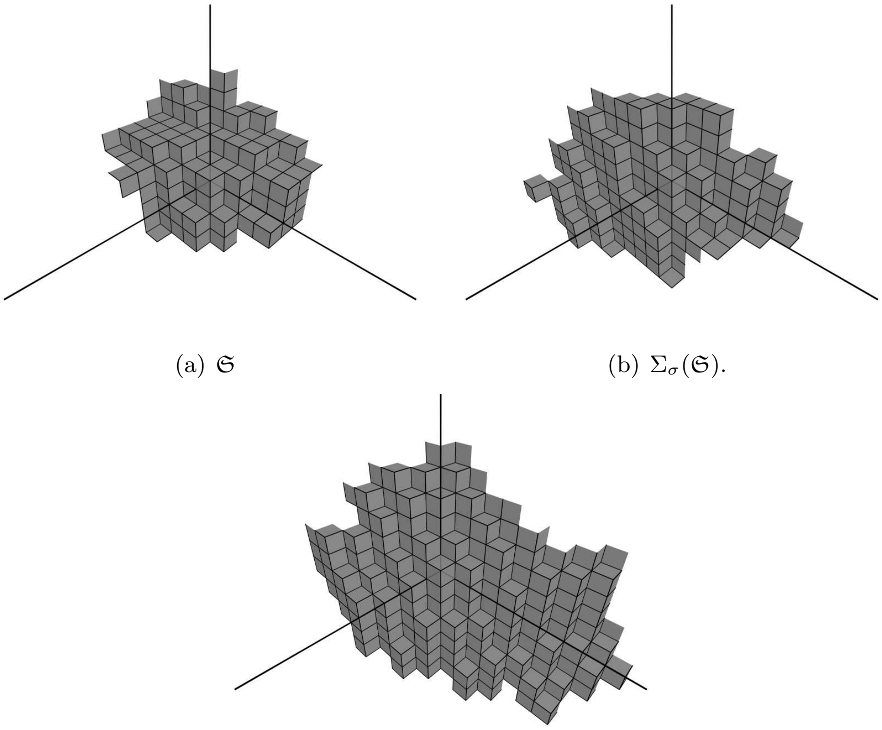
References

[AI01] P. Arnoux and S. Ito. Pisot substitutions and Rauzy fractals. *Bull. Belg. Math. Soc. Simon Stevin*, 8(2):181–207, 2001. Journées Montoises d'Informatique Théorique (Marne-la-Vallée, 2000).

[ABI02] P. Arnoux, V. Berthé, and S. Ito. Discrete planes, $\mathbb{Z}^2$-actions, Jacobi-Perron algorithm and substitutions. *Ann. Inst. Fourier (Grenoble)*, 52(2):305–349, 2002.

[ABS04] P. Arnoux, V. Berthé, and A. Siegel. Two-dimensional iterated morphisms and discrete planes. *Theor. Comput. Sci.*, 319(1-3):145–176, 2004.

[AIS01] P. Arnoux, S. Ito, and Y. Sano. Higher dimensional extensions of substitutions and their dual maps. *J. Anal. Math.*, 83:183–206, 2001.



(a) $\mathfrak{S}$

(b) $\Sigma_\sigma(\mathfrak{S})$.

(c) $\Sigma_\sigma^2(\mathfrak{S})$.

Figure 6: A piece of a non-planar stepped surface $\mathfrak{S}$ and 2 iterations by Σ_σ.

[BCK04] V. Brimkov, D. Coeurjolly, and R. Klette. Digital planarity - a review. Technical Report RR2004-24, Laboratoire LIRIS - Université Lumière Lyon 2, may 2004.

[BV00] V. Berthé and L. Vuillon. Tilings and rotations on the torus: a two-dimensional generalization of sturmian sequences. *Discrete Math.*, 223(1-3):27–53, 2000.

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

[IO93] S. Ito and M. Ohtsuki. Modified Jacobi-Perron algorithm and generating markov partitions for special hyperbolic toral automorphisms. *Tokyo J. Math.*, 16:441–472, 1993.

[IO94] S. Ito and M. Ohtsuki. Parallelogram tilings and Jacobi-Perron algorithm. *Tokyo J. Math.*, 17:33–58, 1994.

[Jam04] D. Jamet. On the Language of Discrete Planes and Surfaces. In *Proceedings of the Tenth International Workshop on Combinatorial Image Analysis*, pages 227-241. Springer-Verlag, 2004.

[JP04] D. Jamet, G. Paquin. Discrete surfaces and infinite smooth words *FP-SAC'05.*

[KR04] R. Klette, A . Rosenfeld, *Digital straightness-A review*, Discrete Applied Mathematics, 139:197–230, 2004.

[LOTH02] N. Lothaire. Algebraic combinatorics on words. *Cambridge University Press*, 2002.

[Pyt02] N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*, volume 1794 of *Lecture Notes in Mathematics.* Springer Verlag, 2002.

[Rev91] J.-P. Reveillès. *Géométrie discrète, calcul en nombres entiers et algorithmique.* Thèse de Doctorat, Université Louis Pasteur, Strasbourg, 1991.

Palindromic complexity of infinite words coding interval exchange transformations

Peter Baláži, Edita Pelantová[*]

Abstract

We study palindromes of infinite words coding r-interval exchange transformations. If the permutation $\pi : \{1, 2, \ldots, r\} \to \{1, 2, \ldots, r\}$ connected with this transformation is given by $\pi(k) = r+1-k$ for all k, then we prove that there is exactly one palindrome of length n in the infinite word, if n is even, and there are exactly r palindromes of length n, if n is odd. For an infinite word coding an interval exchange transformation with another permutation, the length of palindromes is bounded.

1 Introduction

The paper is devoted to palindromes of infinite words coding r-interval exchange transformations. We are generalizing results known for $r = 2$ (i.e. for Sturmian words, see [3]) and for $r = 3$ (see [2]). Let us recall the definition of an interval exchange map. It can be found together with some properties in [3], [5].

Given r positive numbers $\alpha_1, \alpha_2, \ldots, \alpha_r$ such that $\sum_{i=1}^{r} \alpha_i = 1$. They define a partition of the interval $I = [0, 1)$ into r intervals

$$I_k = \left[\sum_{i=1}^{k-1} \alpha_i, \; \sum_{i=1}^{k} \alpha_i \right), \qquad k = 1, 2, \ldots, r\,.$$

Let π denote a permutation of the set $\{1, 2, \ldots, r\}$. The interval exchange transformation associated with $\alpha_1, \ldots, \alpha_r$ and π is defined as the map $T : I \to I$ which exchanges the intervals I_k according to the permutation π,

$$T(x) = x + \sum_{j<\pi(k)} \alpha_{\pi^{-1}(j)} - \sum_{j<k} \alpha_j\,, \qquad \text{for } x \in I_k\,.$$

For $x_0 \in I$, the sequence $(T^n(x_0))_{n\in\mathbb{Z}}$ is called the orbit of x_0 under T. The infinite bidirectional word $(u_n)_{n\in\mathbb{Z}}$ over the alphabet $\mathcal{A} = \{1, \ldots, r\}$ associated to the orbit $(T^n(x_0))_{n\in\mathbb{Z}}$ is defined as

$$u_n = k \in \mathcal{A} \quad \Leftrightarrow \quad T^n(x_0) \in I_k\,.$$

[*] Department of Mathematics, FNSPE, Czech Technical University Trojanova 13, 120 00 Praha 2, Czech Republic, peter_balazi@centrum.cz

Let $(u_n)_{n\in\mathbb{Z}}$ be an infinite word. A finite word $w = w_0w_1\dots w_{n-1}$ of length n is called a factor of $(u_n)_{n\in\mathbb{Z}}$, if there exists $i \in \mathbb{Z}$ such that $w = u_iu_{i+1}\dots u_{i+n-1}$. We denote by $\mathcal{L}_n$ the set of factors of $(u_n)_{n\in\mathbb{Z}}$ of length n, and by $\mathcal{L}$ the language of $(u_n)_{n\in\mathbb{Z}}$, i.e. the set $\mathcal{L} = \bigcup_{n\in\mathbb{N}}\mathcal{L}_n$. The function $\mathcal{C} : \mathbb{N} \to \mathbb{N}$ that assigns to an integer n the number of factors of $(u_n)_{n\in\mathbb{Z}}$ of length n is called the subword complexity of $(u_n)_{n\in\mathbb{Z}}$; we have $\mathcal{C}(n) = \#\mathcal{L}_n$.

The complexity of the word corresponding to any r-interval exchange transformation satisfies $\mathcal{C}(n) \le n(r-1)+1$, for all $n \in \mathbb{N}$. In this paper we focus on mappings T for which the complexity of the word associated to the orbit of arbitrary $x_0 \in I$ satisfies $\mathcal{C}(n) = n(r-1)+1$, for all $n \in \mathbb{N}$. This property is ensured by additional conditions (denoted by $\mathfrak{P}$) on the parameters of the map T.

$$(\mathfrak{P}) \qquad \begin{array}{ll} 1. & \alpha_1,\dots,\alpha_r \text{ are linearly independent over } \mathbb{Q}, \\ 2. & \pi\{1,\dots,k\} \neq \{1,\dots,k\} \text{ for each } k = 1,2,\dots,r-1. \end{array}$$

If the conditions $(\mathfrak{P})$ are fulfilled, then the set $\{T^n(x_0)\}_{n\in\mathbb{Z}}$ is dense in I for each $x_0 \in I$ and the dynamical system associated to the transformation T is minimal. It implies that the infinite word corresponding to the sequence $(T^n(x_0))_{n\in\mathbb{Z}}$ is uniformly recurrent, i.e. any factor of $(u_n)_{n\in\mathbb{Z}}$ appears in $(u_n)_{n\in\mathbb{Z}}$ with bounded gaps.

Another important consequence of $(\mathfrak{P})$ is that the language of the word $(u_n)_{n\in\mathbb{Z}}$ corresponding to $(T^n(x_0))_{n\in\mathbb{Z}}$ does not depend on the position of the starting point x_0, but only on the transformation T. Therefore the notation $\mathcal{L}(T)$, which we adopt here, is justified. Also, it is not difficult to verify that the language of the transformation $\hat{T}$, arising from T by interchanging left semi-closed intervals by right semi-closed intervals, coincides with $\mathcal{L}(T)$. We know that $\mathcal{L}(T_1) = \mathcal{L}(T_2)$ only if T_1 and T_2 coincide, up to the values in the discontinuity points.

Our aim is to describe the palindromic complexity of the infinite word $(u_n)_{n\in\mathbb{Z}}$ coding an r-interval exchange transformation. A factor w of $(u_n)_{n\in\mathbb{Z}}$ is called a palindrome of $(u_n)_{n\in\mathbb{Z}}$ if $w = w_0w_1\dots w_{n-1}$ coincides with its reversal $\overline{w} = w_{n-1}\dots w_1w_0$. The palindromic complexity of the infinite word $(u_n)_{n\in\mathbb{Z}}$ is a function $\mathcal{P} : \mathbb{N} \mapsto \mathbb{N}$ which to an integer n assigns the number $\mathcal{P}(n)$ of palindromes of length n.

2 Invariance of the language under reversal

Let us first explain why the invariance of the language under reversal is important for the study of palindromic complexity of the infinite word $(u_n)_{n\in\mathbb{Z}}$. Assume that the conditions $(\mathfrak{P})$ are satisfied. Then, as we already seen, the infinite word $(u_n)_{n\in\mathbb{Z}}$ coding the r-interval exchange transformation is uniformly recurrent. This implies that for all $n \in \mathbb{N}$ there exists a number $R(n)$ such that all factors of length n (all elements of the set $\mathcal{L}_n$) are included in an arbitrary factor of length $R(n)$. If we suppose that the length of palindromes in the word $(u_n)_{n\in\mathbb{Z}}$ is not bounded, then for each n there exists a palindrome P of length $\geq R(n)$. Since the palindrome P contains all words of $\mathcal{L}_n$, the set $\mathcal{L}_n$ contains with each factor w also its reversal $\overline{w}$. Therefore

$$\overline{\mathcal{L}(T)} := \{\overline{w} \mid w \in \mathcal{L}(T)\} = \mathcal{L}(T)\,,$$

and we say that the language of $(u_n)_{n\in\mathbb{Z}}$ is closed under reversal. As a consequence the palindromic complexity of $(u_n)_{n\in\mathbb{Z}}$ is interesting only in the case that language of $(u_n)_{n\in\mathbb{Z}}$ is closed under reversal, because otherwise $\mathcal{P}(n) = 0$ for sufficiently large n.

Let us therefore study the conditions under which the language $\mathcal{L}(T)$ is closed under reversal, more formally when the identity $\mathcal{L}(T) = \overline{\mathcal{L}(T)}$ holds. Consider a factor $w = w_0w_1\dots w_{n-1} \in \mathcal{L}(T)$ and assume that w codes the trajectory $x, T(x), \dots,\ T^{n-1}(x)$. If we consider the transformation T^{-1} and apply it on the starting point $T^n(x)$ we obtain the sequence $T^n(x), T^{n-1}(x), \dots, T(x)$. The decomposition of the interval $I = [0, 1)$ corresponding to the transformation T^{-1} is the following

$$[0,1) = T(I_{\pi^{-1}(1)}) \cup T(I_{\pi^{-1}(2)}) \cup \dots \cup T(I_{\pi^{-1}(r)}),$$

where π is the permutation associated with the transformation T. If $T^k(x) \in T(I_{\pi^{-1}(j)})$, i.e. the k-th letter is j when coding by T^{-1}, then $T^{k-1}(x) \in I_{\pi^{-1}(j)}$, i.e. the $(k-1)$-st letter is $\pi^{-1}(j)$ when coding by the map T. Therefore if we use the notation $\pi^{-1}(w_0w_1\dots w_n) = \pi^{-1}(w_0)\pi^{-1}(w_1)\dots\pi^{-1}(w_n)$, for each word w over the alphabet $\mathcal{A} = \{1, \dots, r\}$, we can write

$$\overline{\mathcal{L}(T)} = \pi^{-1}\mathcal{L}(T^{-1})\,.$$

Thus in order to obtain the invariance under reversal, it suffices to study the validity of

$$\pi^{-1}(\mathcal{L}(T^{-1})) = \mathcal{L}(T)\,. \tag{2.1}$$

Let us study when there exists a permutation σ of letters $\{1, \dots, r\}$ such that $\sigma(\mathcal{L}(T_1)) = \mathcal{L}(T_2)$. If we use the symmetry of the interval $[0, 1)$ around the point $1/2$, the map

$$\tilde{T}(x) = 1 - T(1-x)\,, \quad x \in [0,1)\,,$$

is again an r-interval exchange transformation, but with the domain $(0, 1]$. This interval exchange transformation corresponds to the lengths of intervals $\alpha_r, \dots, \alpha_1$, respectively, and hence

$$\mathcal{L}(T) = \sigma(\mathcal{L}(\tilde{T})), \tag{2.2}$$

where $\sigma(i) = r+1-i$, for all $i = 1, \dots, r$. From the geometrical interpretation of the r-interval exchange transformation it is clear that any other transformations T_1 and T_2 of the interval I and any other permutation σ do not fulfill the equation

(2.2). Comparing this result with (2.1), it implies that $\pi = \sigma^{-1} = \sigma$ and that $T^{-1} = \tilde{T}$, up to the discontinuity points of T^{-1}. Let us summarize the previous considerations into the following theorem.

Theorem 2.1 *Let $\mathcal{L}(T)$ be the language of the infinite word coding the r-interval exchange transformation given by a positive vector $(\alpha_1, \ldots, \alpha_r)$ and by the permutation π on the set $\{1, \ldots, r\}$. Assume that the conditions $(\mathfrak{P})$ are satisfied. The language $\mathcal{L}(T)$ is closed under reversal if and only if*

$$\pi(1) = r,\ \pi(2) = r - 1, \ldots, \pi(r) = 1. \tag{2.3}$$

Corollary 2.2 *Let $(u_n)_{n\in\mathbb{Z}}$ be an infinite word coding the r-interval exchange transformation with parameters $(\alpha_1, \ldots, \alpha_r)$ and π be a permutation satisfying the conditions $(\mathfrak{P})$. If the permutation π does not fulfill (2.3) then $\mathcal{P}(n) = 0$ for each sufficiently large n.*

3 Palindromic complexity

In this section we will be dealing only with such transformations T of r-intervals for which the permutation π satisfies (2.3). In this case the transformation has the form of

$$T(x) = x + \sum_{j>k} \alpha_j - \sum_{j<k} \alpha_j \quad \text{for } x \in I_k.$$

It is known that there exists an interval $I_w \subset I_{w_0}$ for every word $w = w_0 w_1 \ldots w_{n-1} \in \mathcal{L}(T)$ such that the sequence of points $x, T(x), \ldots, T^{n-1}(x)$ is coded by the same word w for each $x \in I_w$. Note that the boundaries of the interval I_w belong to the set $\mathbb{Z}[\alpha_1, \ldots, \alpha_r] = \{\sum k_i \alpha_i \mid k_i \in \mathbb{Z}\}$.

Let us denote the decomposition of the interval $I = [0, 1)$ by the transformation T^{-1} by $\tilde{I}_1, \tilde{I}_2, \ldots, \tilde{I}_r$ and analogously $\tilde{I}_w$ for an arbitrary $w \in \mathcal{L}(T^{-1})$. We obtain the identity $\tilde{I}_{\pi^{-1}(j)} = T(I_j)$ for each $j \in \{1, \ldots, r\}$ and the following equality holds for the interiors of the intervals for each $w \in \mathcal{L}(T)$

$$\tilde{I}^{\circ}_{\pi^{-1}(w)} = 1 - I^{\circ}_w.$$

Note that the interval I_w is always opened from the right and closed from the left and for the interval $1 - I_w$ it is vice versa.

Now we have everything prepared for the proof of the main theorem of the paper.

Theorem 3.1 *Let $\alpha_1, \ldots, \alpha_r$ be positive real numbers, linearly independent over $\mathbb{Q}$ and π a permutation satisfying (2.3). Then*

$$\mathcal{P}(n) = \begin{cases} 1 & \textit{for each } n \textit{ even}, \\ r & \textit{for each } n \textit{ odd}. \end{cases}$$

Proof Consider the palindrome of even length in the form of

$$w_{n-1}w_{n-2}\dots w_0 w_0 \dots w_{n-2}w_{n-1} \in \mathcal{L}(T).$$

It means that there exists $x \in [0,1)$ such that

$$\begin{array}{lllll} x \in I_{w_0} & , & T(x) \in I_{w_1} & , \ \dots \ , & T^{n-1}(x) \in I_{w_{n-1}}, \\ T^{-1}(x) \in I_{w_0} & , & T^{-2}(x) \in I_{w_1} & , \ \dots \ , & T^{-n+1}(x) \in I_{w_{n-1}}. \end{array}$$

Hence $x \in I_w$, where $w = w_0, \dots w_{n-1}$ and on the other side

$$\begin{array}{lcl} x & \in & T(I_{w_0}) = \tilde{I}_{\pi^{-1}(w_0)}, \\ T^{-1}(x) & \in & T(I_{w_1}) = \tilde{I}_{\pi^{-1}(w_1)}, \\ \vdots & & \\ T^{-n}(x) & \in & T(I_{w_{n-1}}) = \tilde{I}_{\pi^{-1}(w_{n-1})}. \end{array}$$

It follows that $x \in \tilde{I}_{\pi^{-1}(w_0 w_1 \dots w_{n-1})}$. Thus $x \in I_w \cap I_{\pi^{-1}(w)}$ because I_w and $I_{\pi^{-1}(w)}$ are intervals opened from the right and there exists y such that

$$y \in I_w^\circ \cap \tilde{I}^\circ_{\pi^{-1}(w)} = I_w^\circ \cap (1 - \tilde{I}_w^\circ).$$

Now we use the simple fact that $J \cap (s - J) \neq \emptyset \Leftrightarrow \frac{s}{2} \in J$ for an arbitrary interval J. Therefore

$$\frac{1}{2} \in I_w \cap I_{\pi^{-1}(w)}. \tag{3.1}$$

We have shown that every palindrome of even length arises from the coding of

$$T^{-n}\left(\frac{1}{2}\right), \dots, T^{-1}\left(\frac{1}{2}\right), \frac{1}{2}, T\left(\frac{1}{2}\right), \dots, T^{n-1}\left(\frac{1}{2}\right).$$

The proof in the case when the palindrome has an odd length uses the same ideas. □

Note that according to the previous theorem the interval exchange transformation with a permutation satisfying (2.3) has the same palindromic complexity and also subword complexity as Arnoux-Rauzy words over r letters [2], [4].

4 Acknowledgment

The authors acknowledge the financial support of Czech Science Foundation GAČR 201/05/0169 and of the internal grant IG ČVUT 10-85081.

References

[1] M. Boshernitzan, C.R. Carroll, *An extension of Lagrange's theorem to interval exchange transformations over quadratic fields*, J. Anal. Math., **72** (1997), 21-44.

[2] D. Damanik, L.Q. Zamboni, *Arnoux-Rauzy subshifts: Linear recurrences, powers, and palindromes*, arXiv: math.CO/0208137v1.

[3] X. Droubay, G. Pirillo, *Palindromes and Sturmian words*, Theoret. Comput. Sci. **223** (1999), 73-85.

[4] J. Justin, G. Pirillo, *Episturmian words and episturmian morphisms*, Theoret. Comput. Sci. **276** (2002), 281-313.

[5] M.S. Keane, *Interval exchange transformations*, Math. Zeit. **141** (1975), 25–31.

[6] G. Rauzy, *Échanges d'intervalles et transformations induites*, Acta Arith., **34** (1979), 315-328.

Exhaustive generation of some regular languages by using numeration systems

Elena Barcucci, Renzo Pinzani, Maddalena Poneti[*]

Abstract

In this paper we determine an exhaustive generation algorithm for classes of combinatorial objects satisfying some particular recurrence relations having the form $x_n = ax_{n-1} + bx_{n-2}$. In order to achieve our goal, we code the elements of such classes in terms of some regular languages and using numeration systems. Finally, we prove that the proposed generation algorithm runs in constant amortized time.

1 Preliminaries

The aim of exhaustive generation is the development of algorithms to list all the objects of a certain class. Generating algorithms find applications in various areas such as hardware and software testing, combinatorial chemistry, coding theory, and computational biology. Moreover, such algorithms can give more information about the mathematical properties of the class of objects under consideration.

Indeed many scientific and mathematical investigations begin with an exhaustive examination of all possible cases of small instances of a problem. These sets often have some combinatorial structure; this means that they can be modelled by simple discrete structures, as paths, graphs, words of a language or permutations. There is a growing and maturing methodology for attacking such problems in a systematic manner.

In the context of generating combinatorial objects, the primary performance goal is that the amount of computation be proportional to the number of generated objects. An algorithm for the exhaustive generation will then be considered *efficient* if it uses only a constant amount of computation per object, in an amortized sense. Such algorithms are said to be CAT (Constant Amortized Time, for instance see `http://www.cs.uvic.ca/~fruskey/`). Many papers dealt with regular languages related to finite state automata [L, R, S], but without considering exhaustive generation and CAT property.

In this paper we will define a CAT algorithm for the exhaustive generation of a languages class connected with numeration systems studied by A. Fraenkel

[*] Dipartimento di Sistemi e Informatica, Viale Morgagni 65, 50134 Firenze, Italy, phone: +39 055 4237454, fax: +39 055 4237437, {`barcucci, pinzani, poneti`}`@dsi.unifi.it`

related to several different applications (for instance [F1]). In [F], he proves the following theorem:

Theorem 1.1 *Let $1 = x_0 < x_1 < \ldots < x_n < \ldots$ be an integer sequence. Any nonnegative integer N has precisely one representation in the system $\{x_n\}_{n\geq 0}$, of the form*

$$N = \sum_{i=0}^{n} d_i x_i,$$

where d_i are nonnegative integers satisfying "the greedy condition".

$$d_0 x_0 + \cdots + d_i x_i < x_{i+1}, \qquad i \geq 0.$$

Where $d_0 \cdots d_n$ is the representation of N (of length n) with respect to the system $\{x_n\}_{n\geq 0}$. Notice that in such a representation, we allow leading zeroes. So an integer can have more than one representation but at most one of a given length.
We write $w(N, n)$ to denote the representation of N of length n, being understood that $N < x_{n+1}$ and that the system $\{x_n\}_{n\geq 0}$ is clearly given by the context.

In a successive paper, Barcucci and Rinaldi [BR] apply Theorem 1.1 in order to give a combinatorial interpretation, to some integer sequences, in terms of regular languages. In particular they consider sequences $\{x_n\}_{n\geq 0}$ satisfying a recurrence of the form:

$$x_n = a x_{n-1} + b x_{n-2}, \tag{1.1}$$

with initial conditions $x_0 = 1$, $x_1 = c$, where $a, c \in \mathbb{N} \setminus \{0\}$, $b \in \mathbb{Z}$, and such that $a > |b|$. Assume that $x_{-1} = 0$, $x_0 = 1$, to have $x_1 = c = a$ and $\Sigma = \{0, 1, \ldots, a\}$. In this case for $b > 0$ the digit a never compares like last digit of a code since the sequence is $1, a, \ldots$ so, for convenience, we assume $c = a + 1$.

To ensure that $\{x_n\}_{n\geq 0}$ is a non negative sequence, we consider that $a^2 + 4b \geq 0$ as a consequence of the first assumption $a > |b|$; this condition also implies that the defined sequence is strictly increasing, except for the trivial case $x_n = x_{n-1}$.

Some studies on the connections between numeration systems, regular languages and linear recurrences have been presented also in [L, S].

Throughout the paper we will use the approach presented in [BR], and deal with sequences satisfying (1.1), so let us briefly recall some basic ideas and results from [BR]. In a few words, the authors define a language $\mathcal{L}$, on the alphabet $\Sigma = \{0, 1, 2, \ldots, \alpha\}$, with

$$\alpha = \max\left\{a - 1, \left\lfloor \frac{a^2 + b - 1}{a} \right\rfloor\right\}.$$

The set $\mathcal{L}$ is defined as the union of the mutually disjoint sets $\mathcal{L}_n$, where $\mathcal{L}_0 = \{\epsilon\}$ (ϵ is the empty word), and, for $n \geq 1$, $\mathcal{L}_n = \{w(N, n) : m < x_n\}$.

Basically, $\mathcal{L}_n$ is the set of all n-length words in $\mathcal{L}$, and by construction we have that the cardinality of $\mathcal{L}_n$ ($|\mathcal{L}_n|$) is x_n, for all $n \in \mathbb{N}$.

The language $\mathcal{L}$ is regular, and in order to characterize its words we need to distinguish two cases:

1. $a \geq b \geq 0$. In this case the set of terminal symbols is $\Sigma = \{0, 1 \ldots, a\}$; the language $\mathcal{L}$ can be described as the set of all the words $w \in \Sigma^\star$, $w = d_0 \cdots d_r$, with $d_i \in \Sigma$, and such that if $d_i = a$, then $d_{i+1} < b$, for each $i = 0, \ldots, r$.

2. $a > -b \geq 0$, $c = a$. In this case the set of terminal symbols is $\Sigma = \{0, 1 \ldots, a-1\}$, and the language $\mathcal{L}$, is the set of words $w = u_0 \cdots u_r \in \Sigma^\star$ such that if $u_i = a - 1$ then $u_{i+1} \leq a - b - 1$, and if $u_{i+1} = u_{i+2} = \ldots = u_j = a - b - 1$, $j > 0$, then $u_{j+1} \leq a - b - 1$.

To fully understand the heart of the matter, let us present the following example.

Example 1.2 *Consider the Pell numbers,* $1, 3, 7, 17, 41, 99, 239, \ldots,$ *(sequence M1413 in [SP]) defined by the recurrence relation:*

$$\begin{cases} x_n = 2x_{n-1} + x_{n-2} \\ x_0 = 1 \\ x_1 = 3 \end{cases} \tag{1.2}$$

According to Theorem 1.1 each nonnegative integer has its representation in this system, as shown in Table 1. The table below will be used for examples throughout all the paper. ¿From this recurrence we obtain a language $\mathcal{L}$ *on the alphabet* $\Sigma = \{0, 1, 2\}$*, since* $\alpha = max\{1, 2\}$*. We will often refer to* $\mathcal{L}$ *as the set of* Pell words*. The recurrence relation defined in (1) fits into case 1., and the language is* $\mathcal{L} = \{1 \cup 0 \cup 20\}^* \{2 \cup \epsilon\}$*.*

The long term goal of [BR] is to set up a general methodology for systematically generating several classes of objects, based on the numeration systems approach.

The main result of this paper is the definition of an algorithm for the exhaustive generation of each subset $\mathcal{L}_n$ of a language $\mathcal{L}$, once the coefficients a, b have been fixed, and the numbers $\{x_n\}_{n \geq 0}$ satisfy (1.1). Then we study the average cost of the procedure, proving that it runs in constant amortized time.

In particular, this result provides the exhaustive generation of each class of objects $\mathcal{O}$, according to a parameter p on $\mathcal{O}$, such that

$$|\mathcal{O}_n| = |\{O \in \mathcal{O} : p(O) = n\}| = x_n.$$

2 The generating algorithms

17	**7**	**3**	**1**	**i**
0	0	0	0	0
0	0	0	1	1
0	0	0	2	2
0	0	1	0	3
0	0	1	1	4
0	0	1	2	5
0	0	2	0	6
0	1	0	0	7
0	1	0	1	8
0	1	0	2	9
0	1	1	0	10
0	1	1	1	11
0	1	1	2	12
0	1	2	0	13
0	2	0	0	14
0	2	0	1	15
0	2	0	2	16
1	0	0	0	17

Table 1: The codings of the integers from 0 to 17 in the numeration system defined by Pell numbers.

In this section we present a generation algorithm, based on numeration systems, in order to produce all the elements of $\mathcal{L}_n$, for a fixed n , in lexicographical order. We will use the following notation: for any $m \geq 0$, $n \geq 1$, $w(m, n)$ will denote the word $w \in \mathcal{L}_n$ that encodes the integer m. For simplicity sake, when there is no ambiguity, such a word will often be denoted by $w(m)$. Similarly, for any $w \in \mathcal{L}$, $m(w) \in \mathbb{N}$ will denote the integer represented by the word w. For instance, referring to Example 1, $w(53, 5) = 10112$, and $m(10020) = 47$.

The words of $\mathcal{L}_n$ are naturally ordered as follows:

$$w(0) < w(1) < w(2) < \ldots < w(m) < w(m+1) < \ldots < w(x_n - 1).$$

The reader can easily verify that this order coincides with the lexicographical order.

Also in this section we consider the two following different cases:

$$\begin{cases} x_n = ax_{n-1} + bx_{n-2}, b \geq 0 \\ x_0 = 1 \\ x_1 = a + 1, \end{cases} \tag{2.1}$$

$$\begin{cases} x_n = ax_{n-1} - bx_{n-2}, b \geq 0 \\ x_0 = 1 \\ x_1 = a. \end{cases} \tag{2.2}$$

The reason why in (2.1) and (2.2) we choose these special initial conditions is that they allow a to appear in all the positions of the codings of some integer (in particular also as the last term).

The main problem in generating the words of $\mathcal{L}_n$ in the lexicographical order is to determine the successor of a certain word $w \in \mathcal{L}$ without using its decode $m(w)$. Essentially, we look for an efficient method to pass from $w(m)$ to $w(m+1)$, without using heavy encoding and decoding operations.

The basic idea of these algorithms is that, because of the definition of $\mathcal{L}$, there are sub-strings that cannot appear in the coding of any integer m.

- **Recurrence (2.1):** let $\mathcal{L}$ be the language associated with the recurrence (2.1). Let $w \in \mathcal{L}_n$, $w = w[1]\cdots w[n]$, $w[i] \in \Sigma$, $i = 1, \ldots, n$; we observe that w cannot contain the sub-string ab.

  Hereafter we describe an algorithm to determine the successor $s(w)$ of w, assuming that $m(w) \neq x_n - 1$. The procedure starts by checking the rightmost position of w, i.e., $w[n]$:

  1. If $w[n] \neq (b-1)$, and $w[n] \neq a$ then easily,

     $$s(w) = w[1]\cdots(w[n]+1);$$

     let us consider Example 1.2, where $a = 2$, $b = 1$, and let $n = 5$.

     If $w = 10101$, with $m(w) = 49$, then $s(w) = 10102$ $(= 50)$.

  2. let $w[n] = (b-1)$; there are two possible cases:

     a. $w[n-1] = a$; if we set $s(w) = w[1]\cdots(w[n]+1)$, we obtain the “ forbidden” sub-string ab. Then we set $s(w)[n-1] = s(w)[n] = 0$, and increase by one the element in position $n-2$. Again, this element could be equal to $b-1$, so in this case we need to check the previous position. This procedure goes on until we find a position $i_0 > 1$ such that $w[i_0] \neq b-1$, or $w[i_0] = b-1$, but $w[i_0 - 1] \neq a$. If this position does not exist, then we set $i_0 = 1$. Once we have determined i_0,

     $$s(w) = w[1]\cdots w[i_0 - 1]\,(w[i_0]+1)\,0\,\cdots\,0;$$

     referring to Example 1.2:

     Let $w = 01120$, with $m(w) = 30$; we have $w[5] = 0 = b-1$, $w[4] = 2 = a$; since $w[3] = 1 \neq b-1$, then $s(w) = 01200$ $(= 31)$.

     b. if $w[n-1] \neq a$, simply

     $$s(w) = w[1]\cdots w[n-1]\,(w[n]+1).$$

     Let $w = 11110$, with $m(w) = 68$; we have $w[5] = 0 = b-1$, $w[4] = 1 \neq a$, then $s(w) = 11111$ $(= 69)$.

  3. let $w[n] = a$; again there are two possible cases:

a. $w[n-1] = (b-1)$ and $w[n-2] = a$. Using the same argument as for the case 2.a, we look for the rightmost position $i_0 > 1$, such that $w[i_0] \neq (b-1)$ or $w[i_0 - 1] \neq a$. If this position does not exist, then we set $i_0 = 1$. Once we have found it,

$$s(w) = w[1] \cdots w[i_0 - 1]\,(w[i_0] + 1)\,0 \,\cdots\, 0.$$

Let $w = 10202$, with $m(w) = 57$; we have $w[5] = 2 = a$, $w[4] = 0 = b-1$, $w[3] = 2 = a$, $w[2] = 0 = b-1$; then $i_0 = 1$, and $s(w) = 11000$ $(= 58)$.

b. $w[n-1] \neq (b-1)$ or $w[n-2] \neq a$,

$$s(w) = w[1] \cdots (w[n-1] + 1)\,0.$$

Let $w = 11112$, with $m(w) = 70$; we have $w[5] = 2 = a$, $w[4] = 1 \neq b-1$; then $s(w) = 11120$ $(= 71)$.

- **Recurrence (2.2):** Let $\mathcal{L}$ be the language associated with the recurrence (2.1). Let $w \in \mathcal{L}_n$; we observe that $w = w[1] \cdots w[n]$, $w[i] \in \Sigma$, $i = 1, \ldots, n$, cannot contain any sub-string of the form $(a-1)(a-b-1)^j(a-b)$, $j \geq 0$.

  Hereafter we give an algorithm, analogous to the previous one, in order to determine the successor $s(w)$ of w, again assuming that $n(w) \neq x_n - 1$. As usual, we scan the word w from right to left starting from $w[n]$.

1. If $w[n] \neq (a-b-1)$, and $w[n] \neq (a-1)$

$$s(w) = w[1] \,\cdots\, w[n-1]\,(w[n] + 1);$$

2. if $w[n] = (a-b-1)$ we need to check position $n-1$: if $w[n-1] = (a-b-1)$ we find the first position $i_0 > 2$, if it exists, such that $w[i_0] \neq (a-b-1)$. Otherwise we set $i_0 = 1$. Then there are the following cases:

   a. if $w[i] = (a-1)$,

$$s(w) = w[1] \,\cdots\, (w[i-1] + 1)\,0 \,\cdots\, 0;$$

   b. otherwise:

$$s(w) = w[1] \,\cdots\, w[i-1]\,w[i]\,w[i+1] \cdots\, (w[n] + 1);$$

3. if $w[n] = (a-1)$, then

$$s(w) = w[1] \,\cdots\, (w[n-1] + 1)\,0.$$

Example 2.1 Let us consider the sequence of odd index Fibonacci numbers, $1, 3, 8, 21, 55, \ldots$, (sequence M2741 in [SP]) defined by the recurrence relation:

$$\begin{cases} x_n = 3x_{n-1} - x_{n-2} \\ x_0 = 1 \\ x_1 = 3. \end{cases}$$

As usual, we code nonnegative integers using the sequence $\{x_n\}_{n \geq 0}$ as numeration system (see also Table 2).

As an example, let us consider the word $w = 0202$ $(= 18)$, with $n = 4$; since $w[4] = 2 (= a - 1)$, we are considering situation 3, thus $s(w) = 0210$ $(= 19)$. Instead, for $w = 1211$ $(= 41)$, we have $w[4] = 1 = a - b - 1$, $w[3] = 1 = a - b - 1$; we are in case $2.a.1$, and then we set $i_0 = 1$, obtaining $s(w) = 2000$ $(= 42)$.

3 The analysis of the cost of the algorithm

We are now able to generate all the words of $\mathcal{L}_n$, starting from $w(0)$ to $w(x_n - 1)$. Basically, each word is obtained from the previous one making use of both of generating algorithms.

In this section we will prove that the average computational cost of each of the two algorithms is bounded by a constant. As a consequence, we can generate all the elements of $\mathcal{L}_n$ in constant amortized time.

For any $n \geq 1$, let C_n be the number of all the changes of digits necessary to generate all the elements of $\mathcal{L}_n$, and P_n the number of comparisons needed to generate all these elements. Moreover, let $\overline{C_n}$ $(\overline{P_n})$ be the average number of changes (comparisons) needed to generate all the words having length n.

We analyze separately the cases of recurrences (2.1) and (2.2).

21	**8**	**3**	**1**	**i**
0	0	0	0	0
0	0	0	1	1
0	0	0	2	2
0	0	1	0	3
⋮	⋮	⋮	⋮	⋮
0	2	0	2	18
0	2	1	0	19
⋮	⋮	⋮	⋮	⋮
1	2	1	1	41
2	0	0	0	42

Table 2: The coding of some integers using the sequence of odd index Fibonacci numbers.

Recurrence (2.1): referring to the table of the codings, see for instance Table 1, we consider the leftmost $n-1$ columns. The number of changes necessary to generate all the words having length n is given by the number of changes needed to generate the words with length $n-1$ plus the changes performed on the last column of the table, i.e., at most $x_n - 1$. So we have:

$$C_n = C_{n-1} + x_n - 1.$$

By a simple computation we get $C_n = \sum_{i=0}^{n} x_i - n$. Since the number of words with length n is equal to x_n, we have

$$\overline{C_n} = \frac{1}{x_n} C_n = 1 + \sum_{i=0}^{n-1} \frac{x_i}{x_n} - \frac{n}{x_n}.$$

We consider two possible cases:

1. $a \geqslant 2$, and $b \geqslant 0$, we have

$$\frac{x_i}{x_{i+1}} = \frac{1}{a} - \frac{bx_{i-1}}{ax_{i+1}} \leq \frac{1}{a} \leq \frac{1}{2}.$$

Finally, we have:

$$\frac{x_i}{x_n} \leq 2^{-(n-i)},$$

and then

$$\overline{C_n} < 1 + \sum_{i=1}^{n} 2^{-i} < 2.$$

The average number of changes to pass from an element to the successive one is then at most two.

2. $a = 1$, and $b = 1$, we have

$$\frac{x_i}{x_{i+1}} = \frac{1}{\phi},$$

where

$$\phi = \frac{-1 + \sqrt{5}}{2}.$$

As in the first case we have that $\overline{C_n}$ is bounded by a constant.

Making use of the same reasoning used for C_n we then have:

$$P_n = C_n + x_n - 1;$$

indeed the number of comparisons to be made is equal to the number of released changes plus a comparison for each word. When we change a digit in a word, we have to control, at most, the previous digit. Therefore,

$$\overline{P_n} = \overline{C_n} + 1 - \frac{1}{x_n} < \overline{C_n} + 1.$$

We conclude that the algorithm for generating the words of $\mathcal{L}_n$ in the case of recurrence (2.1) runs in constant amortized time.

Recurrence (2.2): The computation is similar to the previous case. Both the average number of changes and the average number of comparisons are bounded by constants, so the exhaustive generation has the CAT property.

4 Concluding remarks

In this paper we have just studied very simple recurrence relations, i.e., those in (2.1), and (2.2). However, we believe that it would be interesting to consider other classes of recurrence relations, and the associated languages by using numeration systems method.

We have recently studied the case of recurrence relations of the form

$$a_n = k_1 a_{n-1} + k_2 a_{n-2} + \cdots + k_r a_{n-r}, \quad n > r, \tag{4.1}$$

satisfying the condition

$$k_1 \geq k_2 \geq \cdots \geq k_r > 0, \tag{4.2}$$

and the obtained results are similar to those in [BR] for recurrence (1.1). Things become much more difficult by removing or slightly modifying the condition in (4.2).

Moreover, we have considered some particular polynomial recurrence relations, without obtaining any substantial result.

In particular, are there special classes of polynomial recurrence relations that lead to regular languages? On the contrary, are there recurrence relations with constant coefficients not leading to regular languages?

References

[BR] E. Barcucci, S. Rinaldi, Some linear recurrences and their combinatorial interpretation by regular languages, *Theoret. Comput. Sci. 255*,(2001) 679–686.

[F] A. S. Fraenkel, Systems of numeration, *Amer. Math. Monthly 92*, (1985) 105–114.

[F1] A. S. Fraenkel, Arrays, numeration systems and Frankenstein games, *Theoret. Comput. Sci. 282/2*,(2002) 271–284.

[L] P. B. A. Lecomte, M. Rigo, Numeration systems on a regular language, *Theory of Comput. Syst. 34*, (2001) 27–44.

[R] M. Rigo, Numeration systems on a regular language: arithmetic operations, recognizability and formal power series, *Theoret. Comput. Sci. 269*,(2001) 469–498.

[S] J. Shallit, Numeration systems, Linear Recurrences, and Regular Sets, *Inform. and Comput. 113 No 2*, (1994) 331–347.

[SP] N. J. A. Sloane and S. Plouffe, *The encyclopedia of integer sequences*, Academic press, (1996).

Infinite Unfair Shuffles and Associativity*

Maurice H. ter Beek[†], *Jetty Kleijn*[‡]

Abstract

We consider a general shuffling operation for finite and infinite words which is not necessarily fair. This means that it may be the case that in a shuffle of two words, from some point onwards, one of these words prevails *ad infinitum* even though the other word still has letters to contribute. Prefixes and limits of shuffles are investigated, leading to a characterization of general shuffles in terms of shuffles of finite words, a result which does not hold for fair shuffles. Associativity of shuffling is an immediate corollary.

1 Introduction

Shuffling two words is usually defined as arbitrarily interleaving subwords in such a way that the resulting word contains all letters of both words, like shuffling two decks of cards. Shuffling is a well-known operation—sometimes referred to as interleaving, weaving, or merging—that, in many variants, has been extensively studied. Its popularity comes from purely mathematical interest [5, 7, 8, 10–13, 15–17] and from its significance as a semantics for concurrent systems consisting of several components [2, 4, 6, 14, 18–20, 22, 23].

When systems may be iteratively composed, the modularity of the chosen semantics becomes important. In particular, when a form of shuffling is used to combine behaviours, this operation should be commutative and associative. In addition, systems—in particular reactive systems—may exhibit ongoing, infinite behaviours, represented by infinite words. While it is in general not difficult to prove the commutativity and associativity of shuffling operations in case only finite words are involved [2, 4, 7, 10, 13, 17, 20, 22, 23], this changes when infinite words are allowed or certain variants of shuffling are considered. Mostly it is still easy to prove commutativity, but it may be quite challenging to prove associativity [2, 4, 19]. There even exist variants of shuffling for which associativity does not hold [5, 8, 15–17] contrary to the intuition.

In this paper we consider shuffles of possibly infinite words which are not necessarily fair in the sense that one of the two words may be delayed indefinitely, while for each position in the shuffle an occurrence of a letter from the

[*] This work was partially supported by the EU IST-3-016004-IP-09 project SENSORIA.

[†] ISTI-CNR, Via G. Moruzzi 1, 56124 Pisa, Italy, maurice.terbeek@isti.cnr.it

[‡] LIACS, Universiteit Leiden, P.O. Box 9512, 2300 RA Leiden, Netherlands, kleijn@liacs.nl

other word is chosen. Note that with this definition, a shuffle of two finite words is always a standard—fair—shuffle. The motivation for this particular shuffle operation stems from our attempts to describe the behaviour of a certain type of team automaton as a language composed of the languages of its constituting component automata [2–4]. These languages are prefix-closed and may contain infinite words. The composed behaviour as exhibited by the team is not necessarily fair in the sense that any individual component is allowed to execute its behaviour *ad infinitum*, without giving other components a fair turn to continue. This leads to a language consisting of potentially unfair shuffles of words representing behaviours of the various components. Since team automata consist in general of two or more components and may also be defined in an iterative fashion, an associativity result for this generalized form of shuffling is needed to establish the compositionality of the semantics. As demonstrated in the Ph.D. thesis [2] of the first author, this associativity result can also be used for proving the associativity of other more involved—synchronized—shuffle operations, relevant when describing the behaviour of team automata cooperating under different synchronization strategies.

Unfortunately we were unable to find in the literature explicit results concerning the associativity of the shuffle operation as considered here, although there exist many references to the associativity of related shuffle operations [7, 10, 13, 17, 20, 22, 23]. We could thus try and adapt existing results to the general case when the words that are shuffled may be finite or infinite and the shuffle does not have to be fair. However, rather than focussing on the single property of associativity, we propose to investigate here the more general issue of the relationship between shuffles of (finite or infinite) words and the shuffles of their finite prefixes. This should shed more light on the relationships between the finite and the infinite behaviours of the composed system, and contribute to the general knowledge of shuffling in the context of infinite words. The associativity of shuffling follows as a corollary. Hence it is our aim to give a self-contained exposition, elaborating the limit behaviour of shuffles with infinite words and leading to a characterization of shuffles in terms of their prefixes.

The organization of the paper is as follows. In Section 2 we introduce the necessary notations and definitions and establish some basic properties. Also proved here is the important result that the prefixes of the shuffles of two words are exactly the shuffles of the prefixes of these words. Next, in Section 3, we separately consider fair shuffles. Using an established technique, it is proved directly that fair shuffling is associative, also when the words involved may be infinite. Consequently, in the main Section 4, we consider general shuffles. As a main result we demonstrate that a word must be a shuffle of two given words whenever all its prefixes are shuffles of the prefixes of these two words. This result does not hold if only fair shuffles are allowed. Together with the earlier result from Section 2 this leads to a characterization of shuffles, and associativity follows.

2 Basic Definitions and Observations

Let Δ be an alphabet, *i.e.* a (possibly empty, possibly infinite) set of symbols or letters. A word over Δ is a sequence $a_1a_2\cdots$ with each $a_i \in \Delta$. A word may be finite or infinite. The empty word is denoted by λ. For a finite word w, we use the notation $|w|$ to denote its length. Hence $|\lambda| = 0$ and if $w = a_1a_2\cdots a_n$, with $n \geq 1$ and $a_i \in \Delta$, for all $1 \leq i \leq n$, then $|w| = n$. For a word w and an integer $j \geq 1$ such that $j \leq |w|$ if w is finite, we use $w(j)$ to denote the symbol occurring at the jth position in w.

The set of all finite words over Δ (including λ) is denoted by Δ^*. The set $\Delta^+ = \Delta^* \setminus \{\lambda\}$ consists of all nonempty finite words. By convention $\Delta \subseteq \Delta^+$. The set of all infinite words over Δ is denoted by Δ^ω. By Δ^∞ we denote the set of all words over Δ. Hence $\Delta^\infty = \Delta^* \cup \Delta^\omega$. A language (over Δ) is a set of words (over Δ). A language consisting solely of finite words is called finitary. If $L \subseteq \Delta^\omega$, *i.e.* all words of L are infinite, then L is called an infinitary language. When dealing with singleton languages, we often omit brackets and write w rather than $\{w\}$.

Given two words $u, v \in \Delta^\infty$, their concatenation $u \cdot v$ is defined as follows. If $u, v \in \Delta^*$, then $u \cdot v(i) = u(i)$ for $1 \leq i \leq |u|$ and $u \cdot v(|u| + i) = v(i)$ for $1 \leq i \leq |v|$. If $u \in \Delta^*$ and $v \in \Delta^\omega$, then $u \cdot v(i) = u(i)$ for $1 \leq i \leq |u|$ and $u \cdot v(|u| + i) = v(i)$ for $i \geq 1$. If $u \in \Delta^\omega$ and $v \in \Delta^\infty$, then $u \cdot v(i) = u(i)$ for all $i \geq 1$. Note that $u \cdot \lambda = \lambda \cdot u = u$, for all $u \in \Delta^\infty$. The concatenation of two languages K and L is the language $K \cdot L = \{u \cdot v : u \in K,\ v \in L\}$. We will mostly write uv and KL rather than $u \cdot v$ and $K \cdot L$, respectively.

A word $u \in \Delta^*$ is a (finite) prefix of a word $w \in \Delta^\infty$ if there exists a $v \in \Delta^\infty$ such that $w = uv$. In that case we write $u \leq w$. If $u \leq w$ and $u \neq w$, then we may use the notation $u < w$. Moreover, if $|u| = n$, for some $n \geq 0$, then u is the prefix of length n of w, denoted by $w[n]$. Note that $w[0] = \lambda$. The set of all prefixes of a word w is $\mathrm{pref}(w) = \{u \in \Delta^* : u \leq w\}$. For a language K, $\mathrm{pref}(K) = \bigcup\{\mathrm{pref}(w) : w \in K\}$.

Both finite and infinite words can be defined as the limit of their prefixes. Let $v_1, v_2, \cdots \in \Delta^*$ be an infinite sequence of words such that $v_i \leq v_{i+1}$, for all $i \geq 1$. Then $\lim_{n\to\infty} v_n$ is the unique word $w \in \Delta^\infty$ defined by $w(i) = v_j(i)$, for all $i, j \in \mathbb{N}$ such that $i \leq |v_j|$. Hence $v_i \leq w$ for all $i \geq 1$ and $w = v_k$ whenever there exists a $k \geq 1$ such that $v_n = v_{n+1}$ for all $n \geq k$. For an infinite sequence of finite words $u_1, u_2, \ldots \in \Delta^*$ we use the notation $u_1u_2\cdots$ to denote the word $\lim_{n\to\infty} u_1u_2\cdots u_n$.

We now move to shuffles. We define a *shuffle* of two words as an interleaving of consecutive finite subwords of these words which stops (is finite) only if both words have been used completely. This implies that one (infinite) word may prevail when the other word, from some point onwards, contributes nothing anymore but the trivial subword λ.

Definition 2.1 Let $u, v \in \Delta^\infty$. Then

(1) $w \in \Delta^\infty$ is a *fair shuffle* of u and v if $w = u_1 v_1 u_2 v_2 \cdots$, where $u_i, v_i \in \Delta^*$, for all $i \geq 1$, are such that $u = u_1 u_2 \cdots$ and $v = v_1 v_2 \cdots$, and

(2) $w \in \Delta^\infty$ is a *shuffle* of u and v if either

 (a) w is a fair shuffle of u and v, or

 (b) $w = u_1 v_1 u_2 v_2 \cdots$, where $u_i, v_i \in \Delta^*$, for all $i \geq 1$, and either $u_1 u_2 \cdots \in \mathrm{pref}(u)$ and $v = v_1 v_2 \cdots \in \Delta^\omega$, or $u = u_1 u_2 \cdots \in \Delta^\omega$ and $v_1 v_2 \cdots \in \mathrm{pref}(v)$.

For $u, v \in \Delta^\infty$, the set of all fair shuffles of u and v is denoted by $u \mid\mid\mid v$ and the set of all shuffles of u and v is denoted by $u \mid\mid v$. Thus, $u \mid\mid\mid v = \{w \in \Delta^\infty : w \text{ is a fair shuffle of } u \text{ and } v\}$ and $u \mid\mid v = \{w \in \Delta^\infty : w \text{ is a shuffle of } u \text{ and } v\}$. Note that, as defined by the fair shuffle operator $\mid\mid\mid$ and the shuffle operator $\mid\mid$, both fair shuffling and shuffling yield languages.

Shuffling two languages is defined element-wise: The *fair shuffle* of two languages L_1 and L_2 is denoted by $L_1 \mid\mid\mid L_2$ and is defined as the set of all words which are a fair shuffle of a word from L_1 and a word from L_2. Hence $L_1 \mid\mid\mid L_2 = \{w \in u \mid\mid\mid v : u \in L_1,\ v \in L_2\}$. Similarly, the *shuffle* of L_1 and L_2 is denoted by $L_1 \mid\mid L_2$ and is defined as $L_1 \mid\mid L_2 = \{w \in u \mid\mid v : u \in L_1,\ v \in L_2\}$.

Note that by definition a shuffle of two finite words is always fair: $u \mid\mid v = u \mid\mid\mid v$ whenever u and v are finite words. On the other hand, if at least one among u and v is infinite, then $u \mid\mid\mid v \subseteq u \mid\mid v$ and this inclusion may be strict, as can be concluded from the following example.

Example 2.2 The word ab is a shuffle of a and b and $a \mid\mid b = \{ab, ba\}$, $a^2 \mid\mid b = \{a^2b, aba, ba^2\}$; in general $a^n \mid\mid b = \{a^i b a^j : i, j \geq 0,\ i + j = n\}$. Note that every shuffle in $a^n \mid\mid b$ is fair. Also $a^\omega \mid\mid\mid b = \{a^i b a^\omega : i \geq 0\}$ consists of fair shuffles only, but $a^\omega \mid\mid b = (a^\omega \mid\mid\mid b) \cup a^\omega$. Note that also for infinite words it may be the case that all shuffles are fair shuffles: $a^\omega \mid\mid\mid a = a^\omega \mid\mid a = a^\omega$.

It follows immediately from Definition 2.1 that both fair shuffling and shuffling are commutative operations.

Theorem 2.3 *Let $u, v \in \Delta^\infty$. Then $u \mid\mid\mid v = v \mid\mid\mid u$ and $u \mid\mid v = v \mid\mid u$.*

Also the next observation is easily proved. It describes the structure of (fair) shuffles and it can be used as a recursive definition for the shuffles of finite words (see, *e.g.*, [5, 17, 21]).

Lemma 2.4 *Let $u, v \in \Delta^\infty$ and $a, b \in \Delta$. Then*

(1) $u \mid\mid \lambda = u \mid\mid\mid \lambda = u = \lambda \mid\mid\mid u = \lambda \mid\mid u$ *and*

(2) $au \mid\mid\mid bv = a(u \mid\mid\mid bv) \cup b(au \mid\mid\mid v)$ *and* $au \mid\mid bv = a(u \mid\mid bv) \cup b(au \mid\mid v)$.

As an intermediate result we obtain that any concatenation of (fair) shuffles is a (fair) shuffle of a concatenation. In particular, any shuffle of prefixes of two words is a prefix of the (fair) shuffle of these words.

Lemma 2.5 *Let $u, v \in \Delta^\infty$ and $z, u', v' \in \Delta^*$. Then*

(1) $z(u \mathbin{|||} v) \subseteq zu \mathbin{|||} v$ *and* $z(u \mathbin{||} v) \subseteq zu \mathbin{||} v$, *and*

(2) $(u' \mathbin{||} v')(u \mathbin{|||} v) \subseteq u'u \mathbin{|||} v'v$ *and* $(u' \mathbin{||} v')(u \mathbin{||} v) \subseteq u'u \mathbin{||} v'v$.

Proof (1) We only prove the first inclusion. The other proof is analogous. Let $w \in z(u \mathbin{|||} v)$. Then $w = zw'$ for some $w' \in u \mathbin{|||} v$. By Definition 2.1(1), $w' = u_1v_1u_2v_2\cdots$, with $u_i, v_i \in \Delta^*$ for all $i \geq 1$, $u = u_1u_2\cdots$, and $v = v_1v_2\cdots$. Thus $w = zw' = zu_1v_1u_2v_2\cdots$ with $zu_1u_2\cdots = zu$. Hence $w \in zu \mathbin{|||} v$.

(2) We only prove the first inclusion. The other proof is analogous. First assume $u' = \lambda$. Then $u' \mathbin{||} v' = v'$ by Lemma 2.4(1). From Theorem 2.3 and (1) we have $v'(u \mathbin{|||} v) \subseteq u \mathbin{|||} v'v$. The case that $v' = \lambda$ is symmetric. We proceed by induction on $|u'|+|v'|$. The cases that $|u| = 0$ or $|v| = 0$ have already been dealt with. We thus assume that $u' = au_1$ and $v' = bv_1$ with $a, b \in \Delta$ and $u_1, v_1 \in \Delta^*$. Then, by Lemma 2.4(2), $u' \mathbin{||} v' = au_1 \mathbin{||} bv_1 = a(u_1 \mathbin{||} bv_1) \cup b(au_1 \mathbin{||} v_1)$. This yields

$$\begin{aligned}(u' \mathbin{||} v')(u \mathbin{|||} v) &= a(u_1 \mathbin{||} bv_1)(u \mathbin{|||} v) \cup b(au_1 \mathbin{||} v_1)(u \mathbin{|||} v)\\ &\subseteq a(u_1u \mathbin{|||} bv_1v) \cup b(au_1u \mathbin{|||} v_1v)\\ &\subseteq (au_1u \mathbin{|||} bv_1v) \cup (au_1u \mathbin{|||} bv_1v)\\ &= (u'u \mathbin{|||} v'v)\end{aligned}$$

by applying the induction hypothesis and Lemma 2.4(2) twice. □

In addition, as we prove next, every prefix of a shuffle of two words is a fair shuffle of prefixes of these words. Consequently, the shuffles and the fair shuffles of two words determine the same set of prefixes.

Theorem 2.6 *Let $u, v \in \Delta^\infty$. Then*

$$\mathit{pref}(u) \mathbin{||} \mathit{pref}(v) = \mathit{pref}(u \mathbin{|||} v) = \mathit{pref}(u \mathbin{||} v) = \mathit{pref}(u) \mathbin{|||} \mathit{pref}(v)\,.$$

Proof From Lemma 2.5(2) we know that $\mathrm{pref}(u) \mathbin{||} \mathrm{pref}(v) \subseteq \mathrm{pref}(u \mathbin{|||} v)$. Since $u \mathbin{|||} v \subseteq u \mathbin{||} v$ by Definition 2.1, it follows that $\mathrm{pref}(u \mathbin{|||} v) \subseteq \mathrm{pref}(u \mathbin{||} v)$ and $\mathrm{pref}(u) \mathbin{|||} \mathrm{pref}(v) \subseteq \mathrm{pref}(u) \mathbin{||} \mathrm{pref}(v)$. Hence the proof is complete once we have shown that $\mathrm{pref}(u \mathbin{||} v) \subseteq \mathrm{pref}(u) \mathbin{|||} \mathrm{pref}(v)$. Let $z \in \mathrm{pref}(u \mathbin{||} v)$. This implies that there exist an $n \geq 1$ and $u_1, u_2, \ldots, u_n, v_1, v_2, \ldots, v_n \in \Delta^*$ such that $z = u_1v_1u_2v_2\cdots u_{n-1}v_{n-1}x$ with $x \in \mathrm{pref}(u_nv_n)$, $u_1u_2\cdots u_n \in \mathrm{pref}(u)$, and $v_1v_2\cdots v_n \in \mathrm{pref}(v)$. It is now immediately clear that $z \in \mathrm{pref}(u) \mathbin{|||} \mathrm{pref}(v)$.
□

Example 2.7 Although $a^\omega \mathbin{|||} b \neq a^\omega \mathbin{||} b$, we have

$$\mathrm{pref}\,(a^\omega \mathbin{|||} b) = \mathrm{pref}\,(a^\omega \mathbin{||} b) = \{a^i b a^\omega : i \geq 0\} \cup a^* .$$

3 Associativity of Fair Shuffling

In this section the associativity of fair shuffling is proved: $u \mathbin{|||} (v \mathbin{|||} w) = (u \mathbin{|||} v) \mathbin{|||} w$ for all words u, v, and w. Extending a technique known from, *e.g.*, [13,17,21], to infinite words makes it possibly to prove rather directly that fair shuffling is associative. This technique is based on renaming and inserting: with each word we associate its own (indexed) alphabet and rename its letters accordingly. Next arbitrary (finite) subwords over the other indexed alphabet are inserted to simulate shuffles with arbitrary words over the other indexed alphabet. Then we intersect the resulting sets: all words in the intersection are (fair) shuffles of the renamed words. Hence to obtain all (fair) shuffles, it is sufficient to ultimately simply go back to the original alphabets.

To formalize all this, we use homomorphisms and their extension to infinite words. Let $h : \Sigma \to \Gamma^*$ be a function assigning to each letter of alphabet Σ a finite word over Γ. The homomorphic extension of h to Σ^*, also denoted by h, is defined in the usual way by $h(\lambda) = \lambda$ and $h(xy) = h(x)h(y)$ for all $x, y \in \Sigma^*$. We extend h to Σ^∞ by setting $h(\lim_{n\to\infty} v_n) = \lim_{n\to\infty} h(v_n)$, for all $v_1, v_2, \ldots \in \Sigma^*$ such that for all $i \geq 1$, $v_i \leq v_{i+1}$. Note that this is well-defined, since $v_i \leq v_{i+1}$ implies $h(v_i) \leq h(v_{i+1})$.

Let Δ be an alphabet. For each integer $i \in \mathbb{N}$ and each $a \in \Delta$ we let $[a, i]$ be a distinct symbol. Let $[\Delta, i] = \{[a, i] : a \in \Delta\}$. Thus for all $i, j \in \mathbb{N}$ such that $i \neq j$, $[\Delta, i]$ and $[\Delta, j]$ are disjoint. We moreover assume that Δ and $[\Delta, i]$ are disjoint for all i. The homomorphisms $\beta_i : \Delta^* \to [\Delta, i]^*$ and $\overline{\beta}_i : [\Delta, i]^* \to \Delta^*$ are defined by $\beta_i(a) = [a, i]$ and $\overline{\beta}_i([a, i]) = a$, respectively. Note that β_i and $\overline{\beta}_i$ are renamings (bijections): β_i uniquely labels every letter in a word with i and $\overline{\beta}_i$ can be used to remove this label again. Now let $i \in \mathbb{N}$ and $J \subseteq \mathbb{N}$ be such that $i \notin J$. We define $\varphi_{i,J} : (\bigcup\{[\Delta, j] : j \in \{i\} \cup J\})^* \to \Delta^*$ by $\varphi_{i,J}([a, i]) = a$ and $\varphi_{i,J}([a, j]) = \lambda$, for all $j \in J$. Furthermore, we have $\psi_J : (\bigcup\{[\Delta, j] : j \in J\})^* \to \Delta^*$ defined by $\psi_J([a, j]) = a$, for all $j \in J$. Note that $\varphi_{i,\varnothing} = \overline{\beta}_i$ and $\psi_{\{j\}} = \overline{\beta}_j$. Intuitively, $\varphi_{i,J}$ is used to remove the label i from every letter in a word that is labelled by i and to erase every other symbol from that word, whereas ψ_J simply removes all labels in J from every letter in a word that is labelled by such a label from J.

We begin with the result announced above, which provides an alternative definition for the fair shuffle.

Theorem 3.1 *Let $u, v \in \Delta^\infty$. Then, for all $i, j \in \mathbb{N}$ such that $i \neq j$, $u \mathbin{|||} v = \psi_{\{i,j\}}(\varphi^{-1}_{i,\{j\}}(u) \cap \varphi^{-1}_{j,\{i\}}(v))$.*

Proof Without loss of generality we assume that $i = 1$ and $j = 2$.

($\subseteq$) Let $w \in u \mid\mid\mid v$. Then $w = u_1v_1u_2v_2 \cdots$ with $u_1, u_2, \ldots, v_1, v_2, \ldots \in \Delta^*$ such that $u = u_1u_2 \cdots$ and $v = v_1v_2 \cdots$. Now consider

$$\overline{w} = \beta_1(u_1)\beta_2(v_1)\beta_1(u_2)\beta_2(v_2) \cdots .$$

It follows immediately that $\varphi_{1,\{2\}}(\overline{w}) = u$. Likewise, $\varphi_{2,\{1\}}(\overline{w}) = v$. Hence $\overline{w} \in \varphi^{-1}_{1,\{2\}}(u) \cap \varphi^{-1}_{2,\{1\}}(v)$. Since $\psi_{\{1,2\}}(\overline{w}) = w$, we are done.

($\supseteq$) We only prove the case that $u, v \in \Delta^\omega$. The proofs of the other cases are similar. Let $w \in \psi_{\{1,2\}}(\varphi^{-1}_{1,\{2\}}(u) \cap \varphi^{-1}_{2,\{1\}}(v))$ and $\overline{w} \in \varphi^{-1}_{1,\{2\}}(u) \cap \varphi^{-1}_{2,\{1\}}(v)$ be such that $\psi_{\{1,2\}}(\overline{w}) = w$. As $\varphi_{1,\{2\}}(\overline{w}) = u$ there exist $x_1, x_2, \ldots \in \Delta^*$ and $u_1, u_2, \ldots \in \Delta^+$ such that $\overline{w} = \beta_2(x_1)\beta_1(u_1)\beta_2(x_2)\beta_1(u_2) \cdots$ and $u = u_1u_2 \cdots$. Similarly, $\varphi_{2,\{1\}}(\overline{w}) = v$ implies that there exist $y_1, y_2, \ldots \in \Delta^*$ and $v_1, v_2, \ldots \in \Delta^+$ such that $\overline{w} = \beta_1(y_1)\beta_2(v_1)\beta_1(y_2)\beta_2(v_2) \cdots$ and $v = v_1v_2 \cdots$. Hence

$$\beta_2(x_1)\beta_1(u_1)\beta_2(x_2)\beta_1(u_2) \cdots = \beta_1(y_1)\beta_2(v_1)\beta_1(y_2)\beta_2(v_2) \cdots .$$

Since $[\Delta, 1] \cap [\Delta, 2] = \varnothing$ it must be the case that either $\beta_2(x_1) = \lambda$ or $\beta_1(y_1) = \lambda$.

First assume that $\beta_2(x_1) = \lambda$, *i.e.* $x_1 = \lambda$. Hence

$$\beta_1(u_1)\beta_2(x_2)\beta_1(u_2)\beta_2(x_3) \cdots = \beta_1(y_1)\beta_2(v_1)\beta_1(y_2)\beta_2(v_2) \cdots .$$

Again by $[\Delta, 1] \cap [\Delta, 2] = \varnothing$ and from the fact that $u_i, v_i \in \Delta^+$ for all $i \geq 1$, we know that $\beta_1(u_i) = \beta_1(y_i)$ and $\beta_2(v_i) = \beta_2(x_{i+1})$ for all $i \geq 1$. Thus $w = \psi_{\{1,2\}}(\overline{w}) = u_1v_1u_2v_2 \cdots \in u \mid\mid\mid v$.

The case that $\beta_1(y_1) = \lambda$ is treated analogously. □

This alternative definition makes it possible to derive a symmetric description for the case that a word u is fairly shuffled with the fair shuffles $v \mid\mid\mid w$ of words v and w.

Lemma 3.2 *Let $u, v, w \in \Delta^\infty$. Let $i_1, i_2, i_3 \in \mathbb{N}$ be three different integers and let $j \in \mathbb{N}$ be such that $j \neq i_1$. Then*

$$\begin{aligned}&\psi_{\{i_1,j\}}(\varphi^{-1}_{i_1,\{j\}}(u) \cap \varphi^{-1}_{j,\{i_1\}}(\psi_{\{i_2,i_3\}}(\varphi^{-1}_{i_2,\{i_3\}}(v) \cap \varphi^{-1}_{i_3,\{i_2\}}(w))))\\ &\qquad = \psi_{\{i_1,i_2,i_3\}}(\varphi^{-1}_{i_1,\{i_2,i_3\}}(u) \cap \varphi^{-1}_{i_2,\{i_1,i_3\}}(v) \cap \varphi^{-1}_{i_3,\{i_1,i_2\}}(w)) .\end{aligned}$$

Proof Without loss of generality we assume that $i_k = k$, for $1 \leq k \leq 3$, and $j \neq 1$.

($\subseteq$) Let $z \in \psi_{\{1,j\}}(\varphi^{-1}_{1,\{j\}}(u) \cap \varphi^{-1}_{j,\{1\}}(\psi_{\{2,3\}}(\varphi^{-1}_{2,\{3\}}(v) \cap \varphi^{-1}_{3,\{2\}}(w))))$ and $\overline{z} \in \varphi^{-1}_{1,\{j\}}(u) \cap \varphi^{-1}_{j,\{1\}}(\psi_{\{2,3\}}(\varphi^{-1}_{2,\{3\}}(v) \cap \varphi^{-1}_{3,\{2\}}(w)))$ be such that $\psi_{\{1,j\}}(\overline{z}) = z$. Let $x \in \psi_{\{2,3\}}(\varphi^{-1}_{2,\{3\}}(v) \cap \varphi^{-1}_{3,\{2\}}(w))$ be such that $\overline{z} \in \varphi^{-1}_{1,\{j\}}(u) \cap \varphi^{-1}_{j,\{1\}}(x)$. Let $\overline{x} \in \varphi^{-1}_{2,\{3\}}(v) \cap \varphi^{-1}_{3,\{2\}}(w)$ be such that $\psi_{\{2,3\}}(\overline{x}) = x$. Hence $\overline{x}$ is of the form $\overline{x} = b_1c_1b_2c_2 \cdots$ such that for all $i \geq 1$, $b_i \in [\Delta, 2] \cup \{\lambda\}$ and $c_i \in [\Delta, 3] \cup$

$\{\lambda\}$, $\overline{\beta}_2(b_1b_2\cdots) = v$, and $\overline{\beta}_3(c_1c_2\cdots) = w$. Furthermore $\overline{z}$ is of the form $\overline{z} = a_1\overline{b}_1\overline{c}_1a_2\overline{b}_2\overline{c}_2\cdots$ such that for all $i \geq 1$, $a_i \in [\Delta,1] \cup \{\lambda\}$ and $\overline{b}_i, \overline{c}_i \in [\Delta, j] \cup \{\lambda\}$, $\overline{\beta}_1(a_1a_2\cdots) = u$, and $\overline{\beta}_j(\overline{b}_1\overline{c}_1\overline{b}_2\overline{c}_2\cdots) = \psi_{\{2,3\}}(b_1c_1b_2c_2\cdots)$ is such that $\overline{\beta}_j(\overline{b}_1\overline{b}_2\cdots) = \overline{\beta}_2(b_1b_2\cdots) = v$ and $\overline{\beta}_j(\overline{c}_1\overline{c}_2\cdots) = \overline{\beta}_3(c_1c_2\cdots) = w$. Now consider that $\overline{\overline{z}} = a_1\beta_2(\overline{\beta}_j(\overline{b}_1))\beta_3(\overline{\beta}_j(\overline{c}_1))a_2\beta_2(\overline{\beta}_j(\overline{b}_2))\beta_3(\overline{\beta}_j(\overline{c}_2))\cdots$. Since $\overline{\beta}_1(a_1a_2\cdots) = u$, $\overline{\beta}_2(\beta_2(\overline{\beta}_j(\overline{b}_1))\beta_2(\overline{\beta}_j(\overline{b}_2))\cdots) = \overline{\beta}_j(\overline{b}_1\overline{b}_2\cdots) = v$, and $\overline{\beta}_3(\beta_3(\overline{\beta}_j(\overline{c}_1))\beta_3(\overline{\beta}_j(\overline{c}_2))\cdots) = \overline{\beta}_j(\overline{c}_1\overline{c}_2\cdots) = w$, we know that $\varphi_{1,\{2,3\}}(\overline{\overline{z}}) = u$, $\varphi_{2,\{1,3\}}(\overline{\overline{z}}) = v$, and $\varphi_{3,\{1,2\}}(\overline{\overline{z}}) = w$. Hence $\overline{\overline{z}} \in \varphi^{-1}_{1,\{2,3\}}(u) \cap \varphi^{-1}_{2,\{1,3\}}(v) \cap \varphi^{-1}_{3,\{1,2\}}(w)$ and $\psi_{\{1,2,3\}}(\overline{\overline{z}}) = \psi_{\{1,j\}}(\overline{z}) = z$.

($\supseteq$) Let $z \in \psi_{\{1,2,3\}}(\varphi^{-1}_{1,\{2,3\}}(u) \cap \varphi^{-1}_{2,\{1,3\}}(v) \cap \varphi^{-1}_{3,\{1,2\}}(w))$ and $\overline{z} \in \varphi^{-1}_{1,\{2,3\}}(u) \cap \varphi^{-1}_{2,\{1,3\}}(v) \cap \varphi^{-1}_{3,\{1,2\}}(w)$ be such that $\psi_{\{1,2,3\}}(\overline{z}) = z$. Hence $\overline{z}$ is of the form $\overline{z} = a_1b_1c_1a_2b_2c_2\cdots$ such that for all $i \geq 1$, $a_i \in [\Delta,1] \cup \{\lambda\}$, $b_i \in [\Delta,2] \cup \{\lambda\}$, and $c_i \in [\Delta,3] \cup \{\lambda\}$, $\overline{\beta}_1(a_1a_2\cdots) = u$, $\overline{\beta}_2(b_1b_2\cdots) = v$, and $\overline{\beta}_3(c_1c_2\cdots) = w$. Let $\overline{u} = a_1\alpha_1a_2\alpha_2\cdots$, with $\alpha_i \in ([\Delta, j] \cup \{\lambda\})^*$, be such that for all $i \geq 1$, $\overline{\beta}_j(\alpha_i) = \psi_{\{2,3\}}(b_ic_i)$. Then clearly $\overline{u} \in \varphi^{-1}_{1,\{j\}}(u)$. Next let $\overline{x} = b_1c_1b_2c_2\cdots$. Then $\overline{x} \in \varphi^{-1}_{2,\{3\}}(v) \cap \varphi^{-1}_{3,\{2\}}(w)$. Since for all $i \geq 1$, $\varphi_{j,\{1\}}(\alpha_i) = \overline{\beta}_j(\alpha_i) = \psi_{\{2,3\}}(b_ic_i)$ and $a_i \in [\Delta,1] \cup \{\lambda\}$, it follows that $\overline{u} \in \varphi^{-1}_{j,\{1\}}(\psi_{\{2,3\}}(\overline{x}))$. Thus $\overline{u} \in \varphi^{-1}_{1,\{j\}}(u) \cap \varphi^{-1}_{j,\{1\}}(\psi_{\{2,3\}}(\overline{x}))$. Finally, the fact that for all $i \geq 1$, $\overline{\beta}_j(\alpha_i) = \psi_{\{2,3\}}(b_ic_i)$ now implies that $\psi_{\{1,j\}}(\overline{u}) = \psi_{\{1,2,3\}}(\overline{z}) = z$. □

With this lemma it is now straightforward to prove that fair shuffling of possibly infinite words is associative, a result which is mentioned in [19] (where fair shuffling is called fair merge) but which is not proved there due to the complications caused by a different setting.

Theorem 3.3 *Let $u, v, w \in \Delta^\infty$. Then $u \mathrel{|||} (v \mathrel{|||} w) = (u \mathrel{|||} v) \mathrel{|||} w$.*

Proof By Theorem 3.1 and Lemma 3.1,

$$\begin{aligned} u \mathrel{|||} (v \mathrel{|||} w) &= \psi_{\{1,4\}}(\varphi^{-1}_{1,\{4\}}(u) \cap \varphi^{-1}_{4,\{1\}}(\psi_{\{2,3\}}(\varphi^{-1}_{2,\{3\}}(v) \cap \varphi^{-1}_{3,\{2\}}(w)))) \\ &= \psi_{\{1,2,3\}}(\varphi^{-1}_{1,\{2,3\}}(u) \cap \varphi^{-1}_{2,\{1,3\}}(v) \cap \varphi^{-1}_{3,\{1,2\}}(w))\,. \end{aligned}$$

Similarly, we have

$$\begin{aligned} (u \mathrel{|||} v) \mathrel{|||} w &= \psi_{\{3,4\}}(\varphi^{-1}_{4,\{3\}}(\psi_{\{1,2\}}(\varphi^{-1}_{1,\{2\}}(u) \cap \varphi^{-1}_{2,\{1\}}(v))) \cap \varphi^{-1}_{3,\{4\}}(w)) \\ &= \psi_{\{1,2,3\}}(\varphi^{-1}_{1,\{2,3\}}(u) \cap \varphi^{-1}_{2,\{1,3\}}(v) \cap \varphi^{-1}_{3,\{1,2\}}(w))\,. \end{aligned}$$

Hence $u \mathrel{|||} (v \mathrel{|||} w) = (u \mathrel{|||} v) \mathrel{|||} w$. □

Since for finite words shuffles and fair shuffles are the same, this theorem implies that shuffling is associative for finite words. This is a well-known fact (see, *e.g.*, [7, 10, 13, 17, 20, 22]) which we state here explicitly for completeness' sake and for future reference.

Corollary 3.4 *Let $u, v, w \in \Delta^*$. Then $u \mid\mid (v \mid\mid w) = (u \mid\mid v) \mid\mid w$.*

Theorem 3.1 supplies an alternative definition for *fair* shuffles only, since the inverse homomorphisms used to insert subwords are applied to the complete words to be shuffled. To extend this theorem to the general case we would have to consider also the prefixes of one word in case the other word is infinite. Because of this case distinction, this would lead to a less uniform description for shuffles than we now have for fair shuffles. Rather than proving associativity on basis of such an alternative definition or by further investigating the implications of the associativity of fair shuffling, we will present in the next section a more general approach based on prefix properties. We will express shuffles as limits of shuffles of finite words, which should then allow us to apply the associativity of the shuffling of finite words (Corollary 3.4).

4 General Shuffles

In this section we will prove that a word is a shuffle of two given words if and only if each of its prefixes is a shuffle of prefixes of these two words. We begin by introducing the concept of *decomposition* as an explicit description of how a shuffle is obtained from two given finite words.

Definition 4.1 Let $w \in \Delta^*$. A *decomposition* of w is a sequence $d = (u_1, v_1, u_2, v_2, \ldots, u_n, v_n)$ with $n \geq 1$, $u_1 \in \Delta^*$, $u_2, u_3, \ldots, u_n, v_1, v_2, \ldots, v_{n-1} \in \Delta^+$, $v_n \in \Delta^*$, and $w = u_1 v_1 u_2 v_2 \cdots u_n v_n$. If $u_1 u_2 \cdots u_n = u$ and $v_1 v_2 \cdots v_n = v$, then d is called a (u, v)-*decomposition* of w. The *norm* of d, denoted by $\mid\mid d \mid\mid$, is n.

Note that decompositions — apart from the first and the last subword mentioned — only refer to nonempty subwords of the words that are shuffled. This provides us with a normal form for the description of finite shuffles.

Lemma 4.2 *Let $u, v, w \in \Delta^*$. Then there exists a (u, v)-decomposition of w if and only if $w \in u \mid\mid v$.*

Proof (Only if) Immediate from Definitions 2.1 and 4.1.

(If) Let $w \in u \mid\mid v$. Then by Definition 2.1 we have $w = u_1 v_1 u_2 v_2 \cdots$, with $u_i, v_i \in \Delta^*$ for all $i \geq 1$, $u = u_1 u_2 \cdots$, and $v = v_1 v_2 \cdots$. Let $\rho_1 = (u_1, v_1)$ and if $\rho_k = (\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_\ell, \beta_\ell)$ for some $\ell \geq 1$ and $\alpha_j, \beta_j \in \Delta^*$, for all $1 \leq j \leq \ell$, then

$$\rho_{k+1} = \begin{cases} (\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_\ell u_{k+1}, v_{k+1}) & \text{if } \beta_\ell = \lambda, \\ (\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_\ell, \beta_\ell v_{k+1}) & \text{if } \beta_\ell \neq \lambda \text{ and } u_{k+1} = \lambda, \text{ and} \\ (\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_\ell, \beta_\ell, u_{k+1}, v_{k+1}) & \text{if } \beta_\ell \neq \lambda \text{ and } u_{k+1} \neq \lambda. \end{cases}$$

Thus ρ_{k+1} is obtained from ρ_k by adding the words u_{k+1} and v_{k+1}. These are added in such a way that only the first and the last element of ρ_{k+1} are allowed to equal λ. In general, if $\rho_k = (\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_\ell, \beta_\ell)$, then $\alpha_1, \beta_\ell \in$

Δ^*, $\alpha_j \in \Delta^+$, for all $1 < j \leq \ell$, and $\beta_j \in \Delta^+$, for all $1 \leq j < \ell$. Furthermore, $\alpha_1\beta_1\alpha_2\beta_2\cdots\alpha_\ell\beta_\ell = u_1v_1u_2v_2\cdots u_kv_k$, $\alpha_1\alpha_2\cdots\alpha_\ell = u_1u_2\cdots u_k$, and $\beta_1\beta_2\cdots\beta_\ell = v_1v_2\cdots v_k$. Since w is finite, there must exist an $m \geq 1$ such that for all $n > m$, $u_n = v_n = \lambda$. Then $\rho_m = (\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_\ell, \beta_\ell)$ is such that $\alpha_1\beta_1\alpha_2\beta_2\cdots\alpha_\ell\beta_\ell = w$, $\alpha_1 \in \Delta^*$, $\beta_1, \alpha_2, \beta_2, \alpha_3, \ldots, \beta_{\ell-1}, \alpha_\ell \in \Delta^+$, $\beta_\ell \in \Delta^*$, $\alpha_1\alpha_2\cdots\alpha_\ell = u$, and $\beta_1\beta_2\cdots\beta_\ell = v$. Hence ρ_m is a (u, v)-decomposition of w.

□

It is not difficult to see that a shuffle may have several decompositions. In a series of papers (see, *e.g.*, [16,17]) Mateescu *et al.* use so-called 'trajectories' to describe shuffles. A trajectory defines, in a binary fashion, when to switch from one word to another. When applied, a trajectory thus defines a unique decomposition. Associativity is consequently discussed per set of trajectories. However, associativity of the shuffle as investigated here is not considered.

To be able to describe extensions of shuffles explicitly, we introduce a precedence relation for decompositions.

Definition 4.3 Let $d = (x_1, y_1, x_2, y_2, \ldots, x_k, y_k)$ and $d' = (u_1, v_1, u_2, v_2, \ldots, u_n, v_n)$ be two decompositions of $x_1y_1x_2y_2\cdots x_ky_k \in \Delta^*$ and $u_1v_1u_2v_2\cdots u_nv_n \in \Delta^*$, respectively. Then

(1) *d directly precedes* d' if $k \leq n$ and for all $1 \leq j \leq k-1$, $x_j = u_j$ and $y_j = v_j$, and—moreover—either

  (a) $k = n$, $x_k = u_k$, and $y_ka = v_k$, for some $a \in \Delta$, or

  (b) $k = n$, $y_k = v_k = \lambda$, and $x_ka = u_k$, for some $a \in \Delta$, or

  (c) $k = n-1$, $y_k \neq \lambda$, $v_{k+1} = \lambda$, and $u_{k+1} = a$, for some $a \in \Delta$, and

(2) *d precedes* d' if there exist decompositions $d_0, d_1, \ldots, d_\ell$ such that $\ell \geq 0$, $d = d_0$, $d' = d_\ell$, and for all $0 \leq j \leq \ell - 1$, d_j directly precedes d_{j+1}.

Note that if d and d' are two decompositions such that d directly precedes d', then $||\, d' \,|| = ||\, d \,||$ or $||\, d' \,|| = ||\, d \,|| + 1$. Hence if d precedes d', then $||\, d' \,|| \geq ||\, d \,||$.

It is easy to see that whenever a decomposition d precedes a decomposition d', then d decomposes a prefix of the word that d' decomposes. In fact, we have the following result.

Lemma 4.4 *Let* $d = (x_1, y_1, x_2, y_2, \ldots, x_k, y_k)$ *and* $d' = (u_1, v_1, u_2, v_2, \ldots, u_n, v_n)$ *be two decompositions such that d precedes* d'. *Then*

$$x_1x_2\cdots x_k \in \mathit{pref}(u_1u_2\cdots u_n)\,,$$

$$y_1y_2\cdots y_k \in \mathit{pref}(v_1v_2\cdots v_n)\,,$$

and

$$x_1y_1x_2y_2\cdots x_ky_k \in \mathit{pref}(u_1v_1u_2v_2\cdots u_nv_n)\,.$$

Proof If $d = d'$ there is nothing to prove, so let us assume that $d \neq d'$. From Definition 4.3 it is clear that the statement holds in case d immediately precedes d'.

If d precedes d', then there exist (s_j, t_j)-decompositions d_j of words $w_j \in \Delta^*$ with $0 \leq j \leq \ell$, for some $\ell \geq 1$, such that $d_0 = d$, $d_\ell = d'$, and d_j immediately precedes d_{j+1}, for all $0 \leq j < \ell$. Hence, for all $0 \leq j < \ell - 1$, $s_j \in \text{pref}(s_{j+1})$, $t_j \in \text{pref}(t_{j+1})$, and $w_j \in \text{pref}(w_{j+1})$. Thus $s_0 = x_1 x_2 \cdots x_k \in \text{pref}(s_\ell) = \text{pref}(u_1 u_2 \cdots u_n)$, $t_0 = y_1 y_2 \cdots y_k \in \text{pref}(t_\ell) = \text{pref}(v_1 v_2 \cdots v_n)$, and $w_0 = x_1 y_1 x_2 y_2 \cdots x_k y_k \in \text{pref}(w_\ell) = \text{pref}(u_1 v_1 u_2 v_2 \cdots u_n v_n)$. □

Given this lemma it can be proved that the limit of the shuffles defined by an ordered sequence of (u_i, v_i)-decompositions is a shuffle of the limits of the u_i and the v_i.

Lemma 4.5 *For all $i \geq 0$, let d_i be a (u_i, v_i)-decomposition of a word w_i over Δ such that d_i precedes d_{i+1}. Then $u = \lim_{i\to\infty} u_i$, $v = \lim_{i\to\infty} v_i$, and $w = \lim_{i\to\infty} w_i$ exist, and $w \in u \mid\mid v$.*

Proof By Lemma 4.4 it follows that $u_i \leq u_{i+1}$, $v_i \leq v_{i+1}$, and $w_i \leq w_{i+1}$, for all $i \geq 0$, so indeed u, v, and w exist and we only have to prove that $w \in u \mid\mid v$. We distinguish two cases.

First we consider the case that there exists an $N \in \mathbb{N}$ such that $\mid\mid d_i \mid\mid = \mid\mid d_N \mid\mid$ for all $i \geq N$. Let $N_0 \in \mathbb{N}$ be such an N. Again we distinguish two cases.

Let us assume first that, for all $i \geq N_0$, if $d_i = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)$, then $y_n = \lambda$. Consequently, for all $i \geq N_0$, $v_i = v_{N_0}$. From $u_i \leq u_{i+1}$, for all $i \geq 0$, we infer that for all $i > N_0$ there exist $z_{i-N_0} \in \Delta^*$ such that $u_{i+1} = u_i z_{i-N_0}$. Observe that $u = \lim_{i\to\infty} u_i = u_{N_0} \lim_{i\to\infty} z_1 z_2 \cdots z_{i-N_0}$. We thus obtain that for all $i > N_0$ we have $w_i = w_{N_0} z_1 z_2 \cdots z_{i-N_0}$. Since $w_{N_0} \in u_{N_0} \mid\mid v_{N_0}$ by Lemma 4.2, we conclude that $w = \lim_{i\to\infty} w_i \in (u_{N_0} \mid\mid v_{N_0}) \lim_{i\to\infty} z_1 z_2 \cdots z_{i-N_0} = (u_{N_0} \mid\mid v_{N_0})(\lim_{i\to\infty} z_1 z_2 \cdots z_{i-N_0} \mid\mid \lambda) \subseteq u \mid\mid v_{N_0} \subseteq u \mid\mid v$ by Lemma 2.5(2) and the definition of u.

Next assume there exist an $i \geq N_0$ such that $d_i = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)$ with $y_n \neq \lambda$. Let ℓ_0 be the smallest such i. Thus, for all $i \geq \ell_0$, $u_i = u_{\ell_0}$. From $v_i \leq v_{i+1}$, for all $i \geq 0$, we infer that for all $i > \ell_0$ there exist $z_{i-\ell_0} \in \Delta^*$ such that $v_{i+1} = v_i z_{i-\ell_0}$. Observe that $v = \lim_{i\to\infty} v_i = v_{\ell_0} \lim_{i\to\infty} z_1 z_2 \cdots z_{i-\ell_0}$. Thus for all $i > \ell_0$ we have $w_i = w_{\ell_0} z_1 z_2 \cdots z_{i-\ell_0}$. Since $w_{\ell_0} \in u_{\ell_0} \mid\mid v_{\ell_0}$ by Lemma 4.2, we conclude that $w = \lim_{i\to\infty} w_i \in (u_{\ell_0} \mid\mid v_{\ell_0}) \lim_{i\to\infty} z_1 z_2 \cdots z_{i-\ell_0} = (u_{\ell_0} \mid\mid v_{\ell_0})(\lambda \mid\mid \lim_{i\to\infty} z_1 z_2 \cdots z_{i-\ell_0}) \subseteq u_{\ell_0} \mid\mid v \subseteq u \mid\mid v$ by Lemma 2.5(2) and the definition of u.

Now we move to the case that for all $N \in \mathbb{N}$ there exists a $k \in \mathbb{N}$ such that $\mid\mid d_k \mid\mid \geq N$. Let $j_1, j_2, \ldots \in \mathbb{N}$ be the (unique) infinite sequence of integers such that for all $i \in \mathbb{N}$, $\mid\mid d_{j_i} \mid\mid < \mid\mid d_{j_{i+1}} \mid\mid$ and $\mid\mid d_\ell \mid\mid = \mid\mid d_{j_i} \mid\mid$ for all $j_i \leq \ell < j_{i+1}$. Since $\mid\mid d_0 \mid\mid \leq \mid\mid d_1 \mid\mid \leq \cdots$ is an unbounded sequence of integers we know that the j_i as just described exist. Since each d_{j_i} precedes $d_{j_{i+1}}$, Definition 4.3 implies that there exist $x_1, x_2, \ldots, y_1, y_2, \ldots, s_1, s_2, \ldots, t_1, t_2, \cdots \in \Delta^*$ such that $d_{j_i} = (x_1, y_1, x_2, y_2, \ldots, x_{||\,d_{j_i}||-1}, y_{||\,d_{j_i}||-1}, s_i, t_i)$, for all $i \geq 1$. By Lemma 4.4, $u_{j_i} = x_1 x_2 \cdots x_{||\,d_{j_i}||-1} s_i \in \mathrm{pref}\,(u_{j_{i+1}}) = \mathrm{pref}\,(x_1 x_2 \cdots x_{||\,d_{j_{i+1}}||-1} s_{i+1})$, for all $i \geq 1$, and thus $u = \lim\limits_{n\to\infty} x_1 x_2 \cdots x_n$. Analogously, $v = \lim\limits_{n\to\infty} y_1 y_2 \cdots y_n$, and $w = \lim\limits_{n\to\infty} x_1 y_1 x_2 y_2 \cdots x_n y_n$. Thus $w = x_1 y_1 x_2 y_2 \cdots$ with $x_1 \in \Delta^*$, $x_i \in \Delta^+$ for all $i \geq 2$, $y_i \in \Delta^+$ for all $i \geq 1$, $u = x_1 x_2 \cdots$, and $v = y_1 y_2 \cdots$. Hence $w \in u \mid\mid v$. □

On the other hand, we would now like to show that whenever every prefix of a word w can be obtained as a shuffle of a prefix of a word u and a prefix of a word v, then w is indeed a shuffle of u and v. To prove this it would be convenient if the decompositions describing the prefixes of w as shuffles of prefixes of u and v would precede each other and ultimately lead to w as a shuffle of u and v. As the next lemma demonstrates, this can be achieved by requiring that u and v have no letters in common. We write alph(w) to denote the alphabet of a word w, *i.e.* the set of all letters that actually occur in w.

Lemma 4.6 *Let $u, v \in \Delta^\infty$ be such that alph$(u) \cap$ alph$(v) = \varnothing$ and let $w \in \Delta^\omega$. Then pref$(w) \subseteq$ pref(u) $\mid\mid$ pref(v) implies $w \in u \mid\mid v$.*

Proof Let $\mathrm{pref}\,(w) \subseteq \mathrm{pref}\,(u) \mid\mid \mathrm{pref}\,(v)$. Now consider two arbitrary consecutive prefixes of w. Thus for some $n \geq 0$ we have $w[n]$ and $w[n+1] = w[n]a$ with $a \in \mathrm{alph}(u)$ or $a \in \mathrm{alph}(v)$. Since $\mathrm{pref}\,(w) \subseteq \mathrm{pref}\,(u) \mid\mid \mathrm{pref}\,(v)$, there are prefixes u_n and u_{n+1} of u, and prefixes v_n and v_{n+1} of v such that $w[n] \in u_n \mid\mid v_n$ and $w[n+1] \in u_{n+1} \mid\mid v_{n+1}$. Consequently, $u_{n+1} = u_n a$ and $v_{n+1} = v_n$ if $a \in \mathrm{alph}(u)$, and $v_{n+1} = v_n a$ and $u_{n+1} = u_n$ if $a \in \mathrm{alph}(v)$. Now let d_n be a (u_n, v_n)-decomposition of $w[n]$ with $d_n = (x_1, y_1, x_2, y_2, \ldots, x_k, y_k)$ for some $k \geq 0$. Then we obtain a (u_{n+1}, v_{n+1})-decomposition of $w[n+1]$ as follows.

First assume that $a \in \mathrm{alph}(u)$. If $y_k = \lambda$, then $d_{n+1} = (x_1, y_1, x_2, y_2, \ldots, x_k a, y_k)$, whereas if $y_k \neq \lambda$, then $d_{n+1} = (x_1, y_1, x_2, y_2, \ldots, x_k, y_k, a, \lambda)$. In both cases we have $x_1 x_2 \cdots x_k a = u_n a = u_{n+1}$ and $y_1 y_2 \cdots y_k = v_n = v_{n+1}$. Moreover $x_1 y_1 x_2 y_2 \cdots x_k y_k a = w[n]a = w[n+1]$. Thus d_{n+1} is a (u_{n+1}, v_{n+1})-decomposition of $w[n+1]$ and d_n precedes d_{n+1}.

Secondly, let $a \in \mathrm{alph}(v)$. Now $d_{n+1} = (x_1, y_1, x_2, y_2, \ldots, x_k, y_k a)$. Since $x_1 x_2 \cdots x_k = u_n = u_{n+1}$ and $y_1 y_2 \cdots y_k a = v_n a = v_{n+1}$ are such that $x_1 y_1 x_2 y_2 \cdots x_k y_k a = w[n]a = w[n+1]$ we thus know that d_{n+1} is a (u_{n+1}, v_{n+1})-decomposition of $w[n+1]$, which is preceded by d_n.

Observe that the only decomposition of $w[0] = \lambda$ is $d_0 = (\lambda, \lambda)$. Hence we have defined an infinite (and unique) sequence of (u_i, v_i)-decompositions d_i of $w[i]$, with $i \geq 0$, such that d_i precedes d_{i+1} for all $i \geq 0$. From Lemma 4.5 it thus follows that $w = \lim\limits_{n\to\infty} w[n] \in (\lim\limits_{n\to\infty} u_n) \mid\mid (\lim\limits_{n\to\infty} v_n) = u \mid\mid v$. □

Note that this proof uses the observation that—thanks to the disjointness of the alphabets—any decomposition of a prefix of w into prefixes of u and v, has a (unique) successor describing a decomposition of the next prefix. This ultimately leads to a description of w as a shuffle of u and v. Unfortunately, in general, it is not true that decompositions of prefixes can be extended to decompositions of the next prefix. This is shown in the following example, which even shows that an infinite word may have infinitely many prefixes with non-extendable prefixes.

Example 4.7 Let $u = (a^3b)^\omega$ and $v = b^\omega$. Clearly $\{a^3, a^3b\} \subseteq \text{pref}(u)$, $\{b^2, b^3\} \subseteq \text{pref}(v)$, and $w = a^3b^3 \in \text{pref}(u) \mid\mid \text{pref}(v)$. Note that $d_1 = (a^3, b^3)$ and $d_2 = (a^3b, b^2)$ are two decompositions of w.

Next consider $w' = wa = a^3b^3a \in \text{pref}(u) \mid\mid \text{pref}(v)$. The only decompositions of w' which are directly preceded by a decomposition of prefixes of u and v are $d' = (a^3b, b^2, a, \lambda)$ and $d'' = (a^3, b^2, ba, \lambda)$. Clearly, d_1 neither precedes d' nor d''. Note, however, that d_2 precedes d'.

Finally, let $j \geq 0$, $u_j = a^3(ba^3)^j \in \text{pref}(u)$, and $v_j = b^3(b^3)^j \in \text{pref}(v)$. Then clearly both $w_j = (a^3b^4)^j a^3b^3 \in \text{pref}(u) \mid\mid \text{pref}(v)$ and $w'_j = w_j a = (a^3b^4)^j a^3b^3a \in \text{pref}(u) \mid\mid \text{pref}(v)$. Note that $d_j = (x_0, y_0, x_1, y_1, \ldots, x_j, y_j, a^3, b^3)$, where $x_i = a^3b$ and $y_i = b^3$ for all $0 \leq i \leq j$, is a (u_j, v_j)-decomposition of w_j. Reasoning as for $j = 0$ it is however clear that there does not exist a decomposition of w'_j based on prefixes of u and v that is preceded by d_j.

Despite this example, it can however be shown that for all words $u, v \in \Delta^\infty$ and $w \in \Delta^\omega$, whenever $\text{pref}(w) \subseteq \text{pref}(u) \mid\mid \text{pref}(v)$ then $w \in u \mid\mid v$, even when u and v have letters in common. We do this by establishing the existence of an infinite sequence of (u_n, v_n)-decompositions of $w[n]$, with $n \geq 0$, preceding each other. With this in mind we now recall *König's Lemma.*

Lemma 4.8 (König's Lemma) *If G is an infinite finitely-branching rooted tree, then there exists an infinite path through G, starting in the root.*

For later use we prove a more general result, by not just considering words, but *limit-closed* languages. Limit-closedness guarantees that the infinitary part of a language is characterized by its finite prefixes. This notion has been defined in many disguises throughout the literature on theoretical computer science. The oldest reference we found is [1], where the terminology used is 'a closed process', while the term *limit closure* was coined in [9]—after initially referring to the same concept as 'König closure' in its preceding technical report.

Definition 4.9 Let $K \subseteq \Delta^\infty$. K is *limit-closed* if for all $w_1 \leq w_2 \leq \cdots \in \text{pref}(K)$, $\lim\limits_{n\to\infty} w_n \in K \cup \text{pref}(K)$.

Example 4.10 All singleton languages $\{u\}$ and all finitary languages $L = \{\lambda, a, \ldots, a^n : n \geq 1\}$ over a unary alphabet are limit-closed, whereas a^* is not as $\lim\limits_{n\to\infty} a^n = a^\omega \notin a^* \cup L$. However, $a^* \cup a^\omega$ and a^ω are limit-closed.

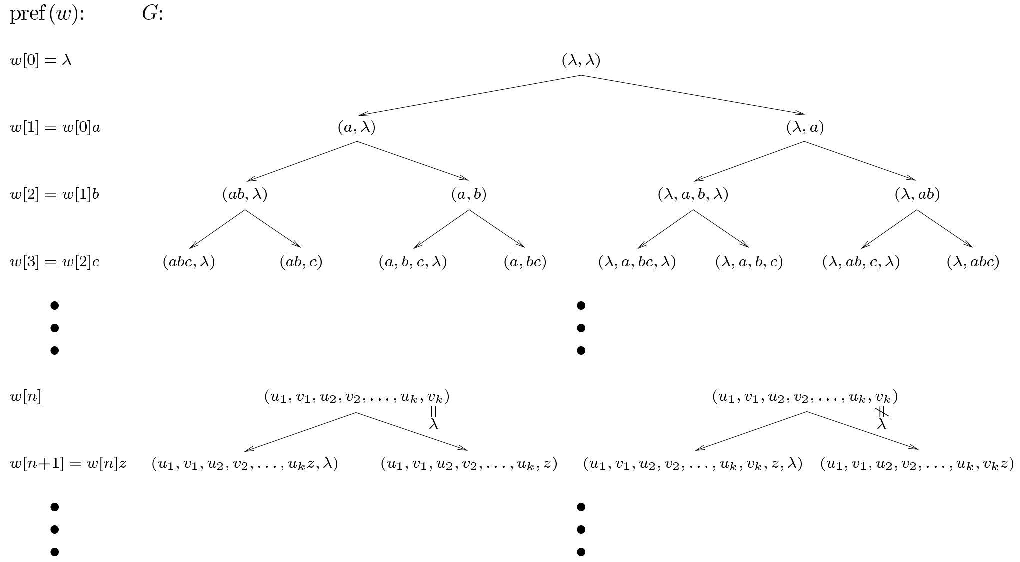
Lemma 4.11 *Let $K, L \subseteq \Delta^\infty$ be limit-closed and let $w \in \Delta^\omega$. Then pref$(w) \subseteq$ pref(K) || pref(L) implies $w \in K$ || L.*

Proof Let pref$(w) \subseteq$ pref(K) || pref(L). For $n \geq 0$, let

$$\begin{aligned} V_n \;=\; \{d : d \text{ is a } (u_n, v_n)\text{-decomposition of } w[n], \\ u_n \in \text{pref}(K), \text{ and } v_n \in \text{pref}(L)\} \end{aligned}$$

be the set of all possible decompositions of the prefix $w[n]$ of w. Note that $V_0 = \{(\lambda, \lambda)\}$. Note furthermore that each V_n is finite, for $n \geq 0$, and that $V_n \cap V_{n'} = \varnothing$, for all $n > n' \geq 0$.

Consider the directly precedes relation $E = \{(d, d') : d \text{ directly precedes } d'\}$. Thus $E \subseteq \bigcup_{n \geq 1}(V_{n-1} \times V_n)$. Note that $G = (\bigcup_{n \geq 0} V_n, E)$ is a directed acyclic graph. It is sketched in Figure 1.



pref (w):
G:
$w[0] = \lambda$
(λ, λ)
$w[1] = w[0]a$
(a, λ)
(λ, a)
$w[2] = w[1]b$
(ab, λ)
(a, b)
(λ, a, b, λ)
(λ, ab)
$w[3] = w[2]c$
(abc, λ)
(ab, c)
(a, b, c, λ)
(a, bc)
$(\lambda, a, bc, \lambda)$
(λ, a, b, c)
$(\lambda, ab, c, \lambda)$
(λ, abc)
$w[n]$
$(u_1, v_1, u_2, v_2, \ldots, u_k, v_k)$
$v_k = \lambda$
$(u_1, v_1, u_2, v_2, \ldots, u_k, v_k)$
$v_k \neq \lambda$
$w[n{+}1] = w[n]z$
$(u_1, v_1, u_2, v_2, \ldots, u_k z, \lambda)$
$(u_1, v_1, u_2, v_2, \ldots, u_k, z)$
$(u_1, v_1, u_2, v_2, \ldots, u_k, v_k, z, \lambda)$
$(u_1, v_1, u_2, v_2, \ldots, u_k, v_k z)$

Figure 1: Sketch of tree $G = (\bigcup_{n \geq 0} V_n, E)$.

Except for (λ, λ), every vertex of G has precisely one incoming edge. This can be seen as follows. The fact that pref$(w) \subseteq$ pref(K) || pref(L) implies that every vertex has at least one incoming edge, whereas the fact that for every decomposition of a prefix $w[n]$, with $n \geq 1$, we can immediately distinguish the unique last symbol of $w[n]$, implies that every vertex has at most one incoming edge. Furthermore, from Definition 4.3 it follows that every vertex has at most two outgoing edges, depending on whether the symbol added to $w[n]$, with $n \geq 0$, to obtain $w[n+1]$ 'belongs' to a prefix from K or to a prefix from L. Hence G is an infinite finitely-branching rooted tree with root (λ, λ).

We can thus use König's Lemma to conclude that there exists an infinite path π through G, starting in the root (λ, λ). Let $\pi = (d_0, d_1, \ldots)$. Then for all

$n \geq 0$, d_n is a (u_n, v_n)-decomposition of $w[n]$ and $(d_n, d_{n+1}) \in E$. Hence from Lemma 4.5 it follows that $u = \lim_{n\to\infty} u_n$, $v = \lim_{n\to\infty} v_n$, and $w = \lim_{n\to\infty} w_n$ exist, and $w \in u \mathbin{||} v$. Since K and L are limit-closed this implies that $w \in K \mathbin{||} L$. □

The statement of this lemma in general does not hold when either K or L is not limit-closed.

Example 4.12 Let $K = a^*$ and $L = \{\lambda\}$. Then

$$\text{pref}\,(a^\omega) = a^* = \text{pref}\,(K) \mathbin{||} \text{pref}\,(L)\,,$$

but $a^\omega \notin a^* = K \mathbin{||} L$.

Since singleton languages are limit-closed, we directly obtain as a corollary the desired result.

Corollary 4.13 *Let $u, v \in \Delta^\infty$ and $w \in \Delta^\omega$. Then pref$(w) \subseteq$ pref$(u) \mathbin{||}$ pref(v) implies $w \in u \mathbin{||} v$.*

It must be noted here that this result does not hold for fair shuffles.

Example 4.14 Consider a^ω. We have $\text{pref}\,(a^\omega) = a^*$ and

$$a^* \subseteq \text{pref}\,(a^\omega) \mathbin{|||} \text{pref}\,(b) = \text{pref}\,(a^\omega) \mathbin{||} \text{pref}\,(b)\,.$$

However, as we have seen in Example 2.2, $a^\omega \in a^\omega \mathbin{||} b$, but $a^\omega \notin a^\omega \mathbin{|||} b$.

Theorem 2.6 and Lemma 4.11 together characterize the shuffles of two words (limit-closed languages) as exactly the limits of the shuffles of the prefixes of these words (languages).

Theorem 4.15 *Let $u, v \in \Delta^\infty$, let $K, L \subseteq \Delta^\infty$ be limit-closed, and let $w \in \Delta^\omega$. Then*
(1) $w \in u \mathbin{||} v$ if and only if pref$(w) \subseteq$ pref$(u) \mathbin{||}$ pref(v), and
(2) $w \in K \mathbin{||} L$ if and only if pref$(w) \subseteq$ pref$(K) \mathbin{||}$ pref(L).

We need one more observation in order to conclude that shuffling is associative.

Corollary 4.16 *Let $v, w \in \Delta^\infty$. Then $v \mathbin{||} w$ is limit-closed.*

Proof Let $y_1 \leq y_2 \leq \cdots \in \text{pref}\,(v \mathbin{||} w)$ and let $y = \lim_{n\to\infty} y_n$. Since for all $x \in \text{pref}\,(y)$, there exists an $i \geq 0$ such that $x \in \text{pref}\,(y_i) \in \text{pref}\,(\text{pref}\,(v \mathbin{||} w)) = \text{pref}\,(v \mathbin{||} w)$, it follows that $\text{pref}\,(y) \subseteq \text{pref}\,(v \mathbin{||} w)$. We distinguish two cases. If $y \in \Delta^*$, then $y \in \text{pref}\,(v \mathbin{||} w)$. If $y \in \Delta^\omega$, then by Theorem 4.15(1), $y \in v \mathbin{||} w$. Hence $y \in v \mathbin{||} w \cup \text{pref}\,(v \mathbin{||} w)$ and $v \mathbin{||} w$ is thus limit-closed. □

Theorem 4.17 *Let $u, v, w \in \Delta^\infty$. Then $u \mathbin{||} (v \mathbin{||} w) = (u \mathbin{||} v) \mathbin{||} w$.*

Proof If u, v, w are finite words, we have Corollary 3.4. If at least one of them is infinite, then both $u \mid\mid (v \mid\mid w)$ and $(u \mid\mid v) \mid\mid w$ consist of infinite words only. Let $x \in u \mid\mid (v \mid\mid w)$. Then Theorem 4.15(2) implies that $\mathrm{pref}\,(x) \subseteq \mathrm{pref}\,(u) \mid\mid \mathrm{pref}\,(v \mid\mid w)$. Thus, by Theorem 2.6,

$$\mathrm{pref}\,(x) \subseteq \mathrm{pref}\,(u) \mid\mid (\mathrm{pref}\,(v) \mid\mid \mathrm{pref}\,(w))\,.$$

Consequently $\mathrm{pref}\,(x) \subseteq (\mathrm{pref}\,(u) \mid\mid \mathrm{pref}\,(v)) \mid\mid \mathrm{pref}\,(w)$ by Corollary 3.4 and $\mathrm{pref}\,(x) \subseteq \mathrm{pref}\,(u \mid\mid v) \mid\mid \mathrm{pref}\,(w)$ by Theorem 2.6. Finally, since $u \mid\mid v$ and $\{w\}$ are limit-closed, Theorem 4.15(2) implies that $x \in (u \mid\mid v) \mid\mid w$. The converse inclusion follows from the above and Theorem 2.3. □

5 Discussion

In this paper we have considered a general shuffling operation for possibly infinite words, which is not necessarily fair, and we have studied its limit behaviour. This has led to a characterization of shuffles in terms of the shuffles of their prefixes, with the associativity of shuffling as an immediate corollary. This proof of the associativity of shuffling is fully self-contained and it does not rely on the sometimes vague or not substantiated claims made in the literature for related operations.

Associativity is of interest not only from a purely mathematical point of view. In fact, as mentioned in the Introduction, our motivation to study the associativity of shuffling stems from the use of shuffling and some of its variants to prove compositionality for different types of team automata [2, 4]. Team automata consist of component automata that collaborate through synchronizations. These synchronizations can be freely chosen depending on the specific protocol of collaboration to be modelled. In [3] we have defined different strategies for choosing the synchronizations of a team automaton. To describe the behaviours of these team automata in terms of the behaviours of their components, several types of ‘synchronized shuffling’ have been introduced in [2, 4]. The associativity of shuffling as defined in this paper, is the basis for proofs of the associativity of some variants of synchronized shuffling in the Ph.D. thesis of the first author [2]. The associativity of these variants, in their turn, is crucial to prove that several types of team automata satisfy compositionality in [2, 4] (in the latter only finitary behaviours are considered).

Since the behaviours of team automata and their components are prefix-closed languages representing ongoing behaviours, we have focussed on the prefix properties of shuffles. As follows from Theorem 2.6, the shuffle operation is sound in the sense that indeed all prefixes of an infinite shuffle appear as shuffles of finite words (behaviours). In addition, the key Lemma 4.11 and its Corollary 4.13 show that every word which is represented through its finite prefixes in the shuffles of finite words is a shuffle of their limits (component behaviours). Together they provide a tool to investigate infinite shuffles as limits of finite shuffles. In a forthcoming paper we intend to address similar issues for the more involved shuffles with synchronization.

Acknowledgement

This paper is dedicated to the memory of Alexandru Mateescu, the founding father of the theory of trajectories and a constant source of information on shuffling for the first author.

References

[1] K. Abrahamson, *Decidability and Expresiveness of Logics of Processes*. Ph.D. thesis, University of Washington, Seattle, 1980.

[2] M.H. ter Beek, *Team Automata—A Formal Approach to the Modeling of Collaboration Between System Components*. Ph.D. thesis, Leiden Institute of Advanced Computer Science, Leiden University, 2003.

[3] M.H. ter Beek, C.A. Ellis, J. Kleijn, and G. Rozenberg, Synchronizations in team automata for groupware systems. *Computer Supported Cooperative Work—The Journal of Collaborative Computing* 12, 1 (2003), 21–69.

[4] M.H. ter Beek and J. Kleijn, Team Automata Satisfying Compositionality. In *Proceedings of FME 2003: Formal Methods—the Twelfth International Symposium of Formal Methods Europe, Pisa, Italy* (K. Araki, S. Gnesi, and D. Mandrioli, eds.), *Lecture Notes in Computer Science* 2805, Springer-Verlag, Berlin, 2003, 381–400.

[5] M.H. ter Beek, C. Martín-Vide, and V. Mitrana, Synchronized Shuffles. Accepted for publication in *Theoretical Computer Science*, 2005. *Cf.* also Technical Report 2003-TR-40, Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, 2003.

[6] J.A. Bergstra, A. Ponse, and S.A. Smolka (eds.), *Handbook of Process Algebra.* Elsevier Science Publishers, Amsterdam, 2001.

[7] S.L. Bloom and Z. Ésik, Free Shuffle Algebras in Language Varieties. *Theoretical Computer Science* 163 (1996), 55–98.

[8] R. De Simone, Langages Infinitaires et Produit de Mixage. *Theoretical Computer Science* 31 (1984), 83–100.

[9] E.A. Emerson, Alternative Semantics for Temporal Logics. *Theoretical Computer Science* 26, 1-2 (1983), 121–130. *Cf.* also Technical Report TR-182, Department of Computer Sciences, University of Texas, Austin, 1981.

[10] S. Ginsburg, *Algebraic and Automata-Theoretic Properties of Formal Languages. Fundamental Studies in Computer Science* 2, North-Holland Publishing Company, Amsterdam, 1975.

[11] S. Ginsburg and E.H. Spanier, Mappings of Languages by Two-Tape Devices. *Journal of the ACM* 12, 3 (1965), 423–434.

[12] J.L. Gischer, Shuffle Languages, Petri Nets, and Context Sensitive Grammars. *Communications of the ACM* 24 (1981), 597–605.

[13] M. Jantzen, The Power of Synchronizing Operations on Strings. *Theoretical Computer Science* 14 (1981), 127–154.

[14] T. Kimura, An Algebraic System for Process Structuring and Interprocess Communication. In *Proceedings of the 8th ACM SIGACT Symposium on Theory of Computing, Hershey, Pennsylvania*, ACM Press, New York, 1976, 92–100.

[15] M. Latteux and Y. Roos, Synchronized Shuffle and Regular Languages. In *Jewels are Forever—Contributions on Theoretical Computer Science in Honor of Arto Salomaa* (J. Karhumäki, H.A. Maurer, Gh. Păun, and G. Rozenberg, eds.), Springer-Verlag, Berlin, 1999, 35–44.

[16] A. Mateescu and G.D. Mateescu, Fair and Associative Infinite Trajectories. In *Jewels are Forever—Contributions on Theoretical Computer Science in Honor of Arto Salomaa*, Springer-Verlag, Berlin, 1999, 327–338.

[17] A. Mateescu, G. Rozenberg, and A. Salomaa, Shuffle on Trajectories: Syntactic Constraints. *Theoretical Computer Science* 197, 1-2 (1998), 1–56.

[18] W.F. Ogden, W.E. Riddle, and W.C. Rounds, Complexity of Expressions Allowing Concurrency. In *Conference Record of the 5th Annual ACM Symposium on Principles of Programming Languages, Tucson, Arizona*, ACM Press, New York, 1978, 185–194.

[19] D. Park, On the semantics of fair parallelism. In *Proceedings of the Copenhagen Winter School on Abstract Software Specifications* (D. Bjørner, ed.), *Lecture Notes in Computer Science* 86, Springer-Verlag, Berlin, 1979, 504–526.

[20] A.W. Roscoe, *The Theory and Practice of Concurrency.* Prentice Hall International Series in Computer Science, London, 1997.

[21] G. Rozenberg and A. Salomaa (eds.), *Handbook of Formal Languages*. Springer-Verlag, Berlin, 1997.

[22] A.C. Shaw, Software Descriptions with Flow Expressions. *IEEE Transactions on Software Engineering* SE-4, 3 (1978), 242–254.

[23] J.L.A. van de Snepscheut, *Trace Theory and VLSI Design. Lecture Notes in Computer Science* 200, Springer-Verlag, Berlin, 1985.

Applying Mealy machine to D0L and u-u words

Aleksandrs Belovs, Jānis Buls[*]

1 Introduction

The theory of finite automata has preserved from its origins a great diversity of aspects. From one point of view, it is a branch of mathematics connected with algebra. From another viewpoint, it is a branch of algorithm design concerned with string manipulation and sequence processing.

A finite automaton can be viewed as a machine model which is as elementary as possible in the sense that the machine has a memory size which is fixed and bounded. The number of possible states of such a machine is itself bounded, whence the notion of a finite-state machine.

A Mealy machine [17,19] is a finite state machine that acts, taking a string on an input alphabet and producing a string of equal length on an output alphabet. This model, namely, Mealy machine, is being investigated intensively since the nineteen fifties (cf. [7, 10, 18, 23, 24]).

A word is a sequence of symbols, finite or infinite, taken from a finite alphabet. Words are central objects of automata theory, and in fact in any standard model of computing. During the last two decades research on combinatorics on words has grown enormously (cf. [2, 13–15]).

The subject of finite automata on infinite words was established in the sixties by Büchi [3] and McNaughton [16]. From this core the theory has developed into many directions. The classification theory of sequence properties is one of these.

In different areas of mathematics, people consider a lot of hierarchies which are typically used to classify some objects according to their complexity. We investigate the lattice of machine invariant classes [4]. This is an infinite completely distributive lattice but it is not a Boolean lattice. The design of stream ciphers motives this report too [5]. It is worth to mention the idea that a lattice would serve as a measure of quality comes from fuzzy mathematics [9].

We concentrate our attention to bi-ideal sequences. Bi-ideal sequences have been considered, with different names, by several authors in algebra and combinatorics [1, 6, 11, 20, 21]. This paper is organized as follows. It is proved that so called u-u words create the machine invariant class, but bounded bi-ideals do not. Also the recurrence function is estimated for transformed u-u words. This

[*] Department of Mathematics, University of Latvia, Raiņa bulvāris 19, Rīga, Latvia, `stiboh@inbox.lv`, `buls@fmf.lu.lv`

estimate is not improvable for ultimately periodic words (as a corollary from Yablonski's theorem [22]).

There are examined D0L, HD0L and CD0L words in this paper too. It is proved that CD0L words create the machine invariant class. Lastly there are analyzed connections between one-sided symbolic dynamics and uniformly recurrent words. So the alternative proof is given for u-u words, namely, those create the machine invariant class.

2 Preliminaries

In this section we present most of the notations and terminology used in this paper. Our terminology is more or less standard (cf. [8,13–15]) so that a specialist reader may wish to consult this section only if need arise.

Let A be a finite non-empty set, we will call *alphabet*, and A^* be a free monoid generated by A. It contains finite sequences of A elements with a concatenation operation. The identity element of A^*, designated as ε, is called the *empty word*.

If $w = w_0 w_1 \dots w_{l-1} \in A^*$ (here $w_i \in A$), then l is called the *length* of w and is denoted as $|w|$. The length of ε is 0. We set $w^0 = \varepsilon$ and $w^{i+1} = w^i w$.

The word w' is a *factor* of w (notation: $w' \backslash w$), if $w = uw'v$ for some u and v. If $u = \varepsilon$ or $v = \varepsilon$, then w' is called, respectively, a *suffix* of a *prefix* of w. We denote, respectively, by $\mathrm{F}(w)$, $\mathrm{Pref}(w)$ and $\mathrm{Suff}(w)$ the sets of all factors, prefixes and suffixes of w.

If $w = w_0 w_1 \dots w_l$, then $w[i, j]$ (with $0 \le i \le j \le l$) stands for a factor $w_i w_{i+1} \dots w_j$. We will use notation $w[i]$ instead of $w[i, i]$. An *occurrence* of a factor v in w is such a pair (i, j) that $v = w[i, j]$.

An (one-sided) infinite word x on the alphabet A is any map $x : \mathbb{N} \to A$. $x = x_0 x_1 x_2 \dots$. All definitions made before can be applied to this case also, only, concatenation xy is defined, if x is finite. Hence prefixes and factors of infinite words are finite, but suffixes are infinite. Suffix $x_i x_{i+1} \dots$ is denoted by $x[i, \infty]$.

A sequence of finite words $\{w_i\}$ is said to converge to the infinite word $y = \lim_{n\to\infty} w_n$, if

$$\forall i \in \mathbb{N} \exists N \in \mathbb{N} \forall m > N : w_m[i] = y[i].$$

As a special case we have $u^\omega = \lim_{n\to\infty} u^n$ for any non-empty u.

A function $\mu : A^* \to B^*$ is called a *morphism*, if $\mu(\varepsilon) = \varepsilon$ and $\mu(uv) = \mu(u)\mu(v)$. The morphism is uniquely defined by its values on A elements. The morphism μ is called *non-erasing*, if $\forall a \in A : \mu(a) \neq \varepsilon$. It is called *uniform*, if $\forall a, b \in A : |\mu(a)| = |\mu(b)|$. It is called *literal*, if $\forall a \in A : |\mu(a)| = 1$. It is clear that a literal morphism is an uniform and non-erasing one. The morphism μ can be applied to an infinite word, defining

$$\mu(x) = \lim_{n\to\infty} \mu(x[0, n]),$$

if this limit exists. It does, if μ is non-erasing.

A 3-sorted algebra $V = \langle Q, A, B, \circ, *\rangle$ is called a *Mealy machine* if Q, A, B are finite non-empty sets and $\circ : Q \times A \to Q$, $* : Q \times A \to B$ are functions. The sets Q, A and B are called respectively *state set*, *input alphabet* and *output alphabet.*

The mappings $\circ$ and $*$ can be extended to $Q \times A^*$ by defining

$$\begin{aligned} q \circ \varepsilon &= q, & q \circ (ua) &= (q \circ u) \circ a, \\ q * \varepsilon &= \varepsilon, & q * (ua) &= (q * u)((q \circ u) * a), \end{aligned}$$

for all $q \in Q, u \in A^*$ and $a \in A$. Henceforth we will omit parentheses, assuming that $\circ$ and $*$ have equal priorities, that is higher than priority of concatenation and lower than one of taking factors. So $q \circ u * x[5,6]y = ((q \circ u) * (x[5,6]))y$. If x is an infinite word and $q \in Q$, we have $q * x = \lim_{n\to\infty} q * x[0,n]$.

A 3-sorted algebra $V_0 = \langle Q, A, B, q_0, \circ, *\rangle$ is called an *initial Mealy machine*, if $\langle Q, A, B, \circ, *\rangle$ is a Mealy machine and *initial state* $q_0 \in Q$. We say a machine V_0 *transforms* x to y (notation: $x \overset{V}{\rightharpoondown} y$) if $y = q_0 * x$. We will write $x \rightharpoondown y$, if there exists V such that $x \overset{V}{\rightharpoondown} y$.

Note that if $y = \mu(x)$, where μ is a literal morphism, then there exists a Mealy machine V with exactly one state such that $x \overset{V}{\rightharpoondown} y$, and vice versa. It is also known that if $x \rightharpoondown y$ and $y \rightharpoondown z$, then $x \rightharpoondown z$.

Our main object of investigation is machine invariant sets. In order to avoid some set-theoretic problems, we will make some assumptions. Let us take the set

$$\mathfrak{N} = \bigcup_{k=0}^{\infty} \{0, 1, \ldots, k\}^{\omega}.$$

We will assume that all states of Mealy machines, input and output alphabets are from the set $\mathbb{N}$. *But if we will use some other input or output alphabet O, we will assume that there is fixed a bijection $k : O \to \{0, 1, \ldots, |O| - 1\}$, and this bijection is applied to the input or output word respectively.*

A non-empty subset $K \in \mathfrak{N}$ is called a *machine invariant set*, if for any word $x \in \mathfrak{N}$ and any Mealy machine V with initial state q_0 the result word $q_0 * x$ also belong to K (of course, in the case when the alphabet of x is a subset of the input alphabet of V). It is proved that *ultimately periodic words* (these are words of a form uv^{ω}) form a machine invariant class [22]. Moreover, it is the minimal machine invariant class [4].

3 D0L words

D0L words is an infinite case of a specific class of L systems of parallel derivation grammar. L systems were introduced by Lindenmayer in [12] as the parallel variant of Chomsky grammars. D0L systems are possibly the most simple case of L systems, and the following definition is widely used in the theory of infinite words.

Definition 3.1 Let A be an alphabet, and $\mu : A^* \to A^*$ be such a morphism, and $a \in A$ such a symbol, that $\mu(a) = au$ and $\mu^k(u) \neq \varepsilon$ for all $k \in \mathbb{N}$. Then a word

$$\mu^\omega(a) = \lim_{k\to\infty} \mu^k(a)$$

is called a D0L word.

It is clear that

$$\mu^k(a) = a \prod_{i=0}^{k-1} \mu^i(u) = au\mu(u)\mu^2(u)\ldots\mu^{k-1}(u).$$

Hence a condition on the non-emptiness of $\mu^k(u)$ is equivalent with the condition of the infinite growth of the length of $\mu^k(a)$. So, the limit exists, if both requests are satisfied. This limit can be also described as the unique fixed point of μ beginning with a.

In some other papers a D0L word is defined by the morphism μ if μ^l for some $l \in \mathbb{N}$ matches both requests. Of course, such a kind of definition leads to the easier description of the corresponding morphism, but it is not important in our investigation.

It is used to name properties of D0L words in the terms of properties of the corresponding morphisms. So, $\mu^\omega(x)$ is uniform if μ is such.

Our main aim is to describe the minimal machine invariant class that contains all D0L words. It is clear that the set of all D0L words does not form one, because it does not even contain an ultimately periodic word 110^ω.

Definition 3.2 We shall call a word $y \in B^\omega$ a HD0L word, if there exists a D0L word $x \in A^\omega$ and a *non-erasing* morphism $\chi : A^* \to B^*$ such that $y = \chi(x)$.

It is not obvious a machine invariant closure of the class of D0L words has to contain all HD0L words, but it immediately follows from the next lemma.

Lemma 3.3 *Any HD0L word $y = \chi(x)$, where $x = \mu^\omega(a)$ is a D0L word, can be expressed in the form $y = \psi(z)$, where $z \in C^\omega$ is a D0L word and $\psi : C \to B$ is a literal morphism.*

In order to stress the importance of the literal morphism in this lemma, we shall use the name of CD0L word, assuming the morphism applied to the D0L word is a literal one. It is possible, in this case, to name properties of the D0L word as the corresponding properties of CD0L word. For example, a CD0L word is an uniform one, if at least one of corresponding D0L words is uniform.

Proof The idea is to construct ν as transforming $\chi(b)$ into $\chi\mu(b)$, dividing the last into parts and matching a part of $\chi\mu(b)$ for a symbol of $\chi(b)$. As one symbol can appear in several $\chi(b)$, literal morphism is needed.

The symbol a is the first symbol of x. First, let us note that

$$|\chi\mu(a)| \geq 2 \tag{3.1}$$

as $\mu(a)$ is of a length at least 2 and χ is a non-erasing morphism.

Let us take a finite alphabet

$$C = \bigcup_{b \in A} \{(b, k) | 0 \leq k < |\chi(b)|\} \subset A \times \mathbb{N}$$

and build a morphism ν in a following way.

For a symbol $b \in A$ we can build such a sequence of $\{u_i^b\}$ that

$$\prod_{i=0}^{|\mu(b)|-1} \prod_{j=0}^{|\chi(\mu(b)[i])|-1} (\mu(b)[i], j) = \prod_{i=0}^{|\chi(b)|-1} u_i^b. \tag{3.2}$$

Define ν with $(b, i) \mapsto u_i^b$ for $0 \leq i < |\chi(b)|$. If we will take a literal morphism $\psi : C \to B$, defined with $\psi(b, i) = \chi(b)[i]$, it follows from the construction that

$$\psi\nu^k \prod_{i=0}^{|\chi(a)|-1} (a, i) = \chi\mu^k(a).$$

Additionally, $\nu(a, 0)$ begins with $(a, 0)$ and as (3.1) holds, we can take $|\nu(a, 0)| > 1$. So, the morphism ν and the symbol $(a, 0)$ matches the first request. A satisfiability of the second one in the general case can be proved taking $u_i^a = \varepsilon$ for all $i > 0$, so that

$$|\nu^k(a, 0)| = |\chi\mu^k(a)| \geq |\mu^k(a)| \underset{k\to\infty}{\longrightarrow} \infty.$$

So $z = \nu^\omega(a, 0)$ exists, and $\chi(x) = \psi(z)$. □

Let us make two useful remarks about lemma 3.3.

1. If the morphism μ erases a symbol $b \in B$ (respectively, $\mu(b) = \varepsilon$), then the morphism χ can also erase this symbol, and the resulting HD0L word will still be a CD0L word. It is so, because we can erase this symbol from all images of the morphism μ, getting the same HD0L word.

2. If both morphisms μ and χ are uniform ones, then we can take ν also uniform. Indeed, all products of u_i^b in (3.2) are of the equal length and of the same number of elements. So we can take all of them of equal size. The second request in the end of the proof can be also shown easy.

Theorem 3.4 *Any initial Mealy machine transforms a CD0L word into a CD0L word.*

Proof We have to prove that a resulting word z is a CD0L word, though we may assume that the source word x is a D0L word, because any CD0L word can be got from a corresponding D0L word using a Mealy machine.

Let $x \in A^\omega$, μ be a morphism such that $x = \mu^\omega(a)$, and $V = \langle Q, A, B, q_0, \circ, * \rangle$ be a machine transforming x into z.

Using a pigeonhole Principe it can be stated that there exists k and l $(1 \leq k < l)$ such that

$$\forall q \in Q \forall \alpha \in A : q \circ \mu^k(\alpha) = q \circ \mu^l(\alpha). \tag{3.3}$$

Moreover, we can assume that $l - k \geq k$.

The idea is to interpret the word x as $\mu^k(\mu'^\omega(a))$, where $\mu' = \mu^{l-k}$, assuming that V transforms $\mu'^\omega(a)$. So let us define $q \bullet v = q \circ \mu^k(v)$, where $v \in A^*$. The $\bullet$ is the analog of $\circ$ in our interpretation. Using (3.3) we have $q \bullet v = q \bullet (\mu'(v))$.

We will take a morphism $\nu : (Q \times A)^* \to (Q \times A)^*$ defined with

$$(q, a) \mapsto \prod_{i=0}^{|\mu'(a)|-1} \left\langle q \bullet (\mu'(a)[0, i-1]),\ \mu'(a)[i] \right\rangle .$$

As $\mu'^\omega(a)$ exists, then $\nu^\omega(q_0, a)$ also exists and can be got adding corresponding states to the symbols.

Using the induction it is easy to see that

$$\mu'^i(a) \overset{W}{\rightharpoondown} \nu^i(q_0, a), \tag{3.4}$$

where $W = \langle Q, A, Q \times A, q_0, \bullet, \diamond \rangle$ with $q \diamond b = (q, b)$.

As we have said before, the morphism $\chi : (Q \times A)^* \to B^*$ will be defined as $(q, a) \mapsto q * \mu^k(a)$. As μ' erases symbols μ^k does (as $l - k \geq k$), we can use the first remark after lemma 3.3 with ν and χ.

The fact that $\chi(\nu^\omega(q_0, a)) = q_0 * x$ can be checked using (3.4). □

It can be also noticed (using the second remark after lemma 3.3) that if the source word x is an uniform CD0L word, then the resulting word $q_0 * x$ is also an uniform one.

4 Recurrent and uniformly recurrent words

This section continues work had begun in the paper [4]. The following theorem has been proved there:

Theorem 4.1 *Every initial Mealy machine transforms an ultimately recurrent word to an ultimately recurrent word.*

Let us recall that a word x is called *recurrent* if any its factor has an infinite number of occurrences in it. So, denoting by $\mathrm{F}^\infty(x)$ the set of all factors of x with the infinite number of occurrences in x, we have that a word x is recurrent if and only if $\mathrm{F}^\infty(x) = \mathrm{F}(x)$. Any word of a form ux, where x is recurrent, is called an *ultimately recurrent* one. Following definition narrows this class.

Definition 4.2 We say a factor u of the infinite word x occurs synthetically in x, if there exists $k \in \mathbb{N}$ that in any factor of x of length k there is at least one occurrence of u. Infinite word x is called uniformly recurrent, if all its factors occur synthetically in x. We say a word x is u-u word if $x = uy$, where y is an uniformly recurrent word.

This definition imply the appearance of the following function

Definition 4.3 The recurrence function of the u-u word x is the function $K_x : \mathbb{N}^+ \to \mathbb{N}^+$ defined with

$$K_x(n) = \max_{u \in F^\infty(x) \cap A^n} k_x(u),$$

where $k_x(u) = \max\{|w| | w \in F^\infty(x) \wedge u \notin F(w)\} + 1$.

The simplest properties of this function are counted in the following lemma

Lemma 4.4 *Any u-u word x and its recurrence function $K_x(n)$ satisfy properties:*

1. *The word x is ultimately recurrent.*
2. *If $x = uy$, then $\forall n \in \mathbb{N} : K_x(n) = K_y(n)$.*
3. *Any recurrent suffix t of x simultaneously is an ultimately recurrent suffix. The function $k_t(u)$ can be described as the minimal k from the definition of the uniformly recurrent word for the factor u. Function $K_t(n)$ is k for the rearrest factor of length n.*
4. *The function $K_x(n)$ is monotonically increasing one and $K_x(n) \geq n$.*

The following theorem is important for us as the revision of the theorem 4.1 .

Theorem 4.5 *The class of u-u words is machine invariant*

Proof Let x be a u-u word and $V_0 = \langle Q, A, B, q_0, \circ, * \rangle$ be an initial Mealy machine.

Since any u-u word is also an ultimately recurrent one, by the conclusion of the theorem 4.1, $y = q_0 * x$ is an ultimately recurrent word. This means that there exists $u \in B^*$ such that $y = uy'$, and y' is recurrent. We can choose u so large, that $x' = x[|u|, \infty]$ is uniformly recurrent.

Let us choose an arbitrary factor $w = y'[m, n]$ of y'. We will use the following inductive construction

Base Let us take $v_1 = x'[m, n]$ and $Q_1 = \{q\}$, where q is the state of the machine V_0 before the transforming of $x'[m, n]$.

Step Let us assume, we have $v_k \in F(x)$, the set Q_k with cardinality k, and $\forall q \in Q_k : w \in \mathrm{F}(q * v_k)$. There are two alternatives

(a) For all occurrences of $v_k = x'[i, j]$ in x', the word $y'[i, j]$ has an occurrence of w. In this case we stop further construction.

(b) There exists the occurrence $v_k = x'[i, j]$, such that $w \notin F(y'[i, j])$. So before the transforming of it, machine was in the state $q \notin Q_k$.

The word y' is recurrent, hence there is occurrence $w = y'[k, l]$ with $k \geq i$. We will take $v_{k+1} = x'\left[i, \max\{j, l\}\right]$ and $Q_{k+1} = Q_k \cup \{q\}$. It is easy to see that inductive assumption holds.

For all k $|Q_k| < |Q| < \infty$, so after the finite number of steps the construction stops. So there exists such a v_k that satisfy the alternative (a). Then w occurs in y' synthetically, because v_k does so in x'. □

The question about some numerical relations between the source and resulting words can arise. The following theorem presents one.

Theorem 4.6 *For any u-u word x and Mealy machine $V_1 = \langle Q, A, B, q_0, \circ, * \rangle$ inequality holds:*

$$K_{q_0 * x}(n) \leq T^{|Q|}(n) - 1,$$

where $T(n) = K_x(n) + 1$. Here T^k means function's T k-th iteration.

Proof At first, construct machine V_0 that differ from V_1 with output: $V_0 = \langle Q, A, Q \times A, q_0, \circ, \acute{*} \rangle$, where $q \acute{*} a = (q, a)$ for all q and a. It is easy to see that recurrence function of V_0 dominates (is larger on all elements of $\mathbb{N}$) the one for V_1. So further we will consider only the case of V_0.

Let us return to the proof of the theorem 4.5. We will define x', y' and $w = y'[m, n+m-1]$ in a same way as in that proof. Again, $v_1 = x'[m, n+m-1]$ and $Q_1 = \{q\}$. The old inductive assumption

$$(v_k \in \mathrm{F}(x')) \wedge (|Q_k| = k) \wedge (\forall q \in Q_k : w \in \mathrm{F}(q \acute{*} v_k)). \tag{4.1}$$

will be completed with the estimation on the length of the factor v_k. It is clear that $|v_1| = n$. For larger k we will suppose that $|v_k| = T^{k-1}(n)$.

Let us change the inductive construction in the alternative (b). All occurrences of v_k belong to one of two groups:

- Occurrences $x'[i, j]$ of the first kind, where $y'[i, j]$ has an occurrence of w.
- Occurrences of the second kind, where $y'[i, j]$ does not contain w.

As the word y' is recurrent, assuming the specific way of output of the machine V_0, we have that there are the infinite number of both. So there exists such an occurrence of v_k in $x'[i, j]$, that is of the second kind, but the next one is of the first kind. The former belongs to the section $x'\left[i + 1, i + K_x(|v_k|)\right]$.

As in the previous proof, the transform of $x'[i,j]$ starts in the state $q \notin Q_k$. Let us take $v_{k+1} = x'\,[i, i + K_x(|v_k|)]$ and $Q_{k+1} = Q_k \cup \{q\}$. It remains to proof estimation of the length, but

$$|v_{k+1}| = K_x(|v_k|) + 1 = T(T^{k-1}(n)) = T^k(n).$$

Inductive construction is over. For some $l \leq |Q|$ alternative (a) holds. In any section of y' of the length $K_x(|v_l|)$ there will be an occurrence of w. And

$$K_x(|v_l|) = T(T^{l-1}(n)) - 1 \leq T^{|Q|}(n) - 1.$$

As it follows from the lemma 4.4, this estimation holds also for the source word y. □

Let us take the periodic word $x = (a_1a_2\dots a_l)^\omega$ with the period l, where all a_i are pairwise distinct. As easy to see, $K_x(n) = n + l - 1$. Using Mealy machine with q states, it is possible to convert it to the periodic word with ql distinct symbols in period. Hence $\forall n : K_y(n) = T^q(n) - 1$.

So in a general case the inequality of the theorem 4.6 is tight. The only problem is that periodic words seems too simple, and the question could be interesting for words with faster growing recurrence functions. The complete and exact answer to this question seems technically difficult. We will describe (without a proof) an example for Mealy machine with three states. Consider such three morphisms

$$\psi : \begin{cases} 0 \mapsto 100000 \\ 1 \mapsto 1000 \end{cases} \qquad \chi : \begin{cases} 0 \mapsto \alpha\theta\theta\theta \\ 1 \mapsto \alpha\theta\theta\theta\theta \end{cases} \qquad \varphi : \begin{cases} \alpha \mapsto 2000110 \\ \theta \mapsto 2110001 \end{cases}$$

and the word $x = \varphi\chi\psi^\omega(1)$. With the Mealy machine on the figure 1 (the initial state is marked with black and the output is universal, i.e. for a state q and an input symbol a it outputs a pair (q, a)), which transforms x to y, we have $K_y(1) = T^3(1) - 1 = 217$.

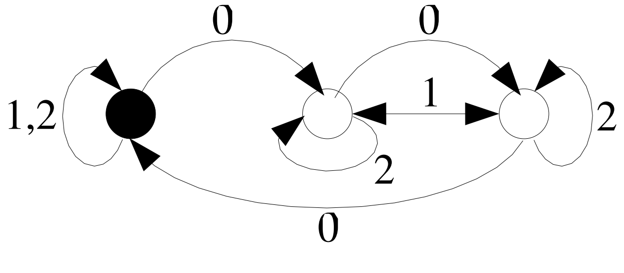


Figure 1: Mealy machine

5 Bounded Bi-ideals

Definition 5.1 Let $\{u_i\}_{i\in\mathbb{N}}$ be a sequence of words from A^*, and $u_0 \neq \varepsilon$. Consider a recurrent sequence

$$v_0 = u_0, \qquad v_{i+1} = v_i u_{i+1} v_i.$$

The word $\lim\limits_{i\to\infty} v_i$ is called a *bi-ideal.*

It can be proved that a word is bi-ideal in that and only in that case, when it is recurrent. So bi-ideals form machine invariant class. In the previous section we have considered one possible narrowing of this class. Now we will consider another one.

Definition 5.2 Let x be a bi-ideal word for a sequence $\{u_i\}_{i\in\mathbb{N}}$. If all u_i lengths are not larger than l, we say the word x is *bounded (with a constant l) bi-ideal.*

As an example one can consider a word s with $u_i = i \bmod 4$. This word begins with

010201030102010001020103010201010102010301020100010201030102010

Using a Mealy machine this word can be translated into a word r with

$$r_i = \left(s_i, (\sum_{k=0}^{i-1} s_i) \bmod 4 \right).$$

Theorem 5.3 *The word r cannot be expressed in the form uw, where w is a bounded bi-ideal.*

So the class of all ultimately bounded bi-ideals is not machine invariant.

Proof We will divide the proof into following parts.

1. Let v be a prefix of any bounded with a constant l bi-ideal z. We will consider a position $\mathrm{SO}(v)$ where the second occurrence of v in z begins. The following estimation holds

   $$\mathrm{SO}(v) \leq 2 \cdot (|v| + l).$$

   Indeed let i be such that (in notations from definition) $|v_i| < |v| \leq |v_{i+1}|$. The word z begins with $v_i u_{i+1} v_i u_{i+2} v_i u_{i+1} v_i$. Hence the desired inequality is obvious.

   So for any bounded bi-ideal

   $$\limsup_{|v|\to\infty} \frac{\mathrm{SO}(v)}{|v|} \leq 2. \tag{5.1}$$

2. Now we will concentrate our attention on s.

   For a prefix $v_{i+1} = v_{i-1}u_iv_{i-1}u_{i+1}v_{i-1}u_iv_{i-1}$:

   - symbols u_i and u_{i+1} will be called *central symbols*,
   - we will say a suffix w to be a *large suffix of* v_{i+1} if $|w| > |v_i|$.

   The following two assertions can be proved:

3. *The word v_i is a prefix of s. If $a \neq u_i$ then v_i has only two occurrences in $\xi = v_iav_i$.*

   This can be proved by the induction on i. For $i = 0$ it can be checked directly.

   Let us assume that for v_i the assertion is proved and prove it for v_{i+1}. At first
   $$v_{i+1}av_{i+1} = v_ibv_iav_ibv_i \tag{5.2}$$
   with $b = u_{i+1}$. It is obvious that for any occurrence of v_{i+1} in ξ either a prefix v_i of v_{i+1} is a factor of a prefix v_ibv_i of ξ or a suffix v_i of v_{i+1} is a factor of a suffix v_ibv_i of ξ.

   Using inductive assumption one can see that there are only 3 possible occurrences of v_{i+1} in ξ. First one is the leftmost, another one is the rightmost and one is exactly in the middle. But the last one does not suite because $a \neq b$.

4. *If w is a large suffix of v_{i+1} and a is one of 2 no-central symbols of v_{i+1} then w has only two occurrences in $v_{i+1}av_{i+1}$.*

   As in the previous point (5.2) holds. The suffix v_i of w can have an occurrence only in a subword v_iav_i or in a subword v_ibv_i (the last one is presented in two copies). Using the previous point one can see that v_i can have only 4 occurrences in $v_{i+1}av_{i+1}$. As $a \neq b$ it can be easy seen that only two of them are suitable for the occurrence of w.

5. Consider a beginning of a word s with $i \equiv 0 \pmod 4$:
   $$v_i1v_i\,2\,v_i1v_i\,3\,v_i1v_i.$$
   Symbols 2 and 3 are not central symbols of $v_{i+1} = v_i1v_i$. Using the assertion from the point 4 any large suffix of v_{i+1} have only 3 occurrences in this beginning fragment. But
   $$\Sigma(v_i1v_i2), \quad \Sigma(v_i1v_i2v_i1v_i3)$$
   are both odd numbers (here Σ means the sum of all symbols in the word).

6. Now we can prove that the word r is not a ultimately bounded bi-ideal. Let us assume that $r = xy$ and y is a bounded bi-ideal. We can take such a big prefix v_i that $2|x| + 2 < |v_i|$ and $i \equiv 0 \pmod 4$, and let w to be such a large suffix of v_i that $|x| + |w| = |v_i|$.

   On the same position as w stands in s a subword stands in r. We will call it w'. It is the prefix of y. Considering the said in the previous point one can see that $\mathrm{SO}(w') \geq 3|w'|$. As we can take as large w' as we like it is in a contradiction with (5.1).

□

6 One-sided Symbolic Dynamics

Let us remind that a system $\langle X, d\rangle$ where $d : X \times X \to \mathbb{R}$ is called a *metric space*, if four axioms holds: $d(x, y) \geq 0$, $(d(x, y) = 0) \Leftrightarrow (x = y)$, $d(x, y) = d(y, x)$ and $d(x, z) \leq d(x, y) + d(y, z)$. A subset A of X is an *open subset*, iff with any element x it contains also $U_\varepsilon(x) = \{y \in X | d(x, y) < \varepsilon\}$ for some $\varepsilon > 0$. A subset A is *closed*, iff $X \setminus A$ is open.

We will say that the set A is closed under the operation f, if $f(A) \subset A$.

It is possible to define a metric d on the set A^ω putting

$$d(x, y) = \begin{cases} 2^{-\min\{k \geq 0 | x[k] \neq y[k]\}}, & \text{if} \quad x \neq y; \\ 0\,, & \text{if} \quad x = y\,. \end{cases}$$

The subspace $T \subset A^\omega$ is closed, if and only if for any sequence $x_{i i \in \mathbb{N}}$ such that $\forall i : x_i \in T$, its limit $y = \lim_{i \to \infty}$ belongs to T (in the case it exists). The space $\langle A^\omega, d\rangle$ is compact. It is so called Koning's Lemma.

We will consider also the function σ, called *shift* function, $\sigma : x \mapsto y$, where $y[i] = x[i + 1]$.

Definition 6.1 A subset of $\langle A^\omega, d\rangle$ both closed in the metric space and under the operation σ is called a *subshift* or *symbolic dynamical system.*

Brief introduction to symbolic dynamical systems can be found in [14]. We will count some main properties of the subshifts

1. A set S is a subshift, if and only if it can be expressed in a form

   $$S = S_X = \{y \in A^\omega | \mathrm{F}(y) \cap X = \emptyset\} = \{y \in A^\omega | \mathrm{F}(y) \subset A^+ \setminus X\}.$$

   for some $X \subset A^*$.

2. An image and a full preimage of a subshift S with a literal morphism is a subshift.

3. An intersection $\bigcap_{S \in T} S$ of a set of subshifts and a union $\bigcup_{S \in T} S$ of a *finite* set of subshifts T are subshifts.

Further with $S(x)$ we will denote the smallest subshift containing a word x. It is easy to prove that $S(x) = S_{A^*\setminus \mathrm{F}(x)}$. Moreover, it can be shown that any non-empty subshift contains a minimal subshift.

The following result is one of the earliest in the symbolic dynamics.

Theorem 6.2 *The word x is uniformly recurrent, if and only if the subshift $S(x)$ is minimal (for any subshift $T \subset S$: $T = S$ or $T = \emptyset$).*

It is also important that any non-empty subshift contains a minimal one – assume it is T. Then, as easy to see, $\forall x \in T : T = S(x)$, so all words from T are uniformly recurrent!

It is possible to use this result in our investigation. *Up to the end of this paragraph we will assume that all machines are with universal output $q{*}a = (q, a)$ and $B = Q \times A$.* So we will omit the last describing Mealy machines.

Theorem 6.3 *For an uniformly recurrent word $x \in A^\omega$ and Mealy machine $\langle Q, A, q_0, \circ, * \rangle$ there exists such a state $q \in Q$, that $q * x \in (Q \times A)^\omega$ is uniformly recurrent and* $\mathrm{F}(q * x) \subset \mathrm{F}(q_0 * x)$.

Proof At first, consider a set $W = \{q * y | (q \in Q) \wedge (y \in S(x)\}$. It is easy to see that it is a subshift. Let us select a subshift $S(q_0 * x) \subset W$. It contains a minimal subshift $V \subset S(q_0 * x)$. Consider a literal morphism $h : (q, a) \mapsto a$. So an image $h(V)$ is a non-empty subshift and $h(V) \subset S(x)$. Hence $h(V) = S(x)$ and $\exists y \in V : x = h(y)$, or, in other words, $\exists q \in Q : q * x = y$. The only left, is to notice that $V = S(y)$ is a minimal subshift, so y is uniformly recurrent and $y \in S(q_0 * x)$, hence $\mathrm{F}(y) \subset \mathrm{F}(q_0 * x)$. □

The result about u-u words (theorem 4.5) follows from this theorem and from the next one

Theorem 6.4 *If a set $C \subset \mathfrak{N}$ is closed under operation σ and literal morphisms and for any Mealy machine $V = \langle Q, A, q_0, \circ, * \rangle$ and $x \in C \cap A^\omega$ holds*

$$\exists q \in Q : (q * x \in C) \wedge (\mathrm{F}(q * x) \subset \mathrm{F}(q_0 * x))$$

then there exists such $u \in (Q \times A)^*$ and $w \in C$ that $q_0 * x = uw$.

We assumed here that there exists a bijection $k : Q \times A \to \{0, 1, \ldots, |Q \times A| - 1\}$ and in fact $k(q * x), k(w) \in C$.

Proof Let us invent an equivalence relation $\sim$ on the set Q defined with

$$(q \sim r) \Leftrightarrow (\exists n \in \mathbb{N} : q \circ x[0, n] = r \circ x[0, n]).$$

Let $A_1, \ldots A_k$ be equivalence classes of this relation, and a_i be an arbitrary element from A_i.

We will construct a Mealy machine $V_k = \langle Q_k, A, q, \bullet, \diamond \rangle$ where

$$Q_k = \{(q_1, \dots q_k) \in Q^k | \forall i \neq j : q_i \neq q_j\} \cup \{\Omega\}$$

.

$$(q_1, \dots q_k) \bullet a = \begin{cases} (q_1 \circ a, \dots q_k \circ a) & , \quad \forall i \neq j : q_i \circ a \neq q_j \circ a \\ \Omega & , \quad \text{otherwise} \end{cases} .$$

and $\Omega \bullet a = \Omega$ for all a. We will take the initial state q equal with $(a_1, \dots, a_k)$. Finally, as usual $q \diamond a = (q, a)$.

As a_i are from distinct equivalence classes, $q \diamond x$ does not contain a factor of a form (Ω, a). So, using our assumption, we can state that there is such $r \in Q_k$ that $r \diamond x \in C$ and it does not contain a factor of a form (Ω, a). Hence $r = (r_1, \dots, r_k)$ and all r_i are from different equivalence classes. Let us take one (assume it is r_1) that $r_1 \sim q_0$.

Using a literal morphism $((q_1, \dots, q_k), a) \mapsto (q_1, a)$ we can state that $r_1 * x \in C$. But $\exists n : (q_0 * x)[n, \infty] = (r_1 * x)[n, \infty]$ and the last suffix of $r_1 * x$ also belongs to C, due it is closed under σ. □

7 Future Work

At the present moment, besides exploring some other machine invariant classes, there are some questions that seems interesting. At first, whether the inequality from the theorem 4.6 is tight in the general case? Also, is it possible to estimate the non-recurrent prefix in the image of ultimately recurrent word with a Mealy machine?

Another interesting problem is to generalize the usage of symbolical dynamics in the proof of the theorem 6.3 to another classes of words.

Acknowledgements

We would like to thank the anonymous reviewer for suggesting a name of CD0L words for the corresponding class of words.

The second author also would like to thank participants of the 4th International Conference on Combinatorics on Words (WORDS'03) for stimulating discussions.

References

[1] D. B. Bean, A. E. Ehrenfeucht and G. McNulty. (1979) *Avoidable Patterns in Strings of Symbols.* Pacific J. Math. **85**, 261–294.

[2] J. Berstel, J. Karhumäki. (2003) *Combinatorics on Words — A Tutorial.* TUCS Technical Report (No 530, June).

[3] J. R. Büchi. (1960) *On a Decision Method in Restricted Second Order Arithmetic.* In: Proc. Internat. Congr. on Logic, Methodology and Philosophy of Science, E. Nagel et al (Eds.), Stanford Univ. Press, Stanford, CA, 1–11.

[4] J. Buls. (2003) *Machine Invariant Classes.* In: Proceedings of WORDS'03, 4th International Conference on Combinatorics on Words, September 10–13, 2003, Turku, Finland, Tero Harju and Juhani Karhumäki (Eds.), TUCS General Publication (No 27, August), 207–211.

[5] J. Buls. (2005) *The Lattice of Machine Invariant Sets and Subword Complexity.* http://arxiv.org/abs/cs.CR/0502064

[6] M. Coudrain and M. P. Schützenberger. (1966) *Une condition de finitude des monoïdes finiment engendrés.* C. R. Acad. Sc. Paris, Sér. A, **262**, 1149–1151.

[7] J. Dassow. (1981) *Completeness Problems in the Structural Theory of Automata.* Mathematical Research (Band 7), Akademie–Verlag, Berlin.

[8] B. A. Davey, H. A. Priestley. (2002) *Introduction to Lattices and Order.* Cambridge University Press.

[9] J. A. Goguen. (1967) *L-fuzzy sets.* J. Math. Anal. Appl., vol. **8**, 145–174.

[10] J. Hartmanis, R. E. Stearns. (1966) *Algebraic Structure Theory of Sequential Machines.* Prentice–Hall, Inc., Englewood Cliffs, New Jersey.

[11] N. Jacobson. (1964) *Structure of Rings.* American Mathematical Society, Providence, RI.

[12] Lindenmayer. (1968) *Mathematical Models for Cellular Interactions in Development* (two parts), J. Theor. Biol. **18**, 280–315.

[13] M. Lothaire. (1983) *Combinatorics on Words.* Encyclopedia of Mathematics and its Applications, Vol. 17, Addison–Wesley, Reading, Massachusetts.

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

[15] Aldo de Luca, Stefano Varricchio. (1999) *Finiteness and Regularity in Semigroups and Formal Languages.* Springer–Verlag, Berlin, Heidelberg.

[16] R. McNaughton. (1966) *Testing and Generating Infinite Sequences by a Finite Automaton.* Inform. and Control **9**, 521–530.

[17] G. H. Mealy. (1955) *A Method for Synthesizing Sequential Circuits.* Bell System Tech. J. vol **34**, September, 1045–1079.

[18] B. I. Plotkin, I. Ja. Greenglaz, A. A. Gvaramija (1992) *Algebraic Structures in Automata and Databases Theory.* World Scientific, Singapore, New Jersey, London, Hong Kong.

[19] S. Seshu. (1959) *Mathematical Models for Sequential Machines.* IRE Mat. Convent, Rec. 7, N 2, 4–16.

[20] I. Simon. (1988) *Infinite Words and a Theorem of Hindman.* Rev. Mat. Apl. **9**, 97–104.

[21] А. И. Зимин. (1982) *Блокирующие множества термов.* [*Blocking Sets of Terms.*] Матем. сб., т.**119**, № 3, с. 363–375. (Russian)

[22] В. Б. Кудрявцев, С. В. Алешин, А. С. Подколзин. (1985) *Введение в теорию автоматов.* [*An Introduction to the Theory of Automata.*] Москва «Наука». (Russian)

[23] А. А. Курмит. (1982) *Последовательная декомпозиция конечных автоматов.* [*Sequential Decomposition of Finite Automata.*] Рига «Зинатне». (Russian)

[24] Б. А. Трахтенброт, Я. М. Барздинь. (1970) *Конечные автоматы* (*поведение и синтез*). [*Finite Automata* (*Behaviour and Synthesis*).] Москва «Наука». (Russian)

Density of Symbols in Discretized Rotation Configurations.[*]

Valérie Berthé[†], *Bertrand Nouvel*[‡]

Abstract

The aim of this paper is to study local configurations for discrete rotations. The algorithm of discrete rotation we consider is the following: a discretized rotation is defined as the composition of a Euclidean rotation with a rounding operation, as studied in [NR05] and [NR04]. It is possible to encode all the information concerning a discrete rotation as two multidimensional words C_α and C'_α that we call configurations. We introduce here two discrete dynamical systems defined by a $\mathbb{Z}^2$-action on the two-dimensional torus that allow us via a suitable symbolic coding to describe the configurations C_α and C'_α and to deduce the densities of occurrence of the symbols in the configurations.

1 Introduction

Symbolic dynamics and more generally, discrete dynamical systems have natural and deep interactions with combinatorics on words. This interaction is particularly well-illustrated in the Sturmian case, see e.g. [Lot02, Fog02]. The combinatorial objects involved are the Sturmian words, while the dynamical systems are the irrational rotations of the torus $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$. A Sturmian word is indeed a coding with respect to a particular two-interval partition of the one-dimensional torus $\mathbb{T}^1$ of the orbit of a point under the action of an irrational rotation. This point of view allows one to deduce many combinatorial properties of Sturmian words, such as for instance the densities of occurrences of factors that can be computed thanks to the equidistribution properties of irrational rotations.

Several attempts of generalization of this fruitful interaction have been proposed. One of the first idea which comes to mind is a rotation of $\mathbb{T}^2$. As an example, the Tribonacci word, that is, the fixed point of the substitution $1 \mapsto 12,\ 2 \mapsto 13,\ 3 \mapsto 1$ codes the orbit of a point of the torus $\mathbb{T}^2$ under the action of a translation in $\mathbb{T}^2$ with respect to a partition of $\mathbb{T}^2$ into three pieces with fractal boundary [Rau82, Lot05].

A second approach, which is dual to the previous one, consists in working with two rotations of $\mathbb{T}^1$. It is indeed convenient to describe discrete planes

[*] financed by *TF1* via an ANRT-CIFRE convention.

[†] LIRMM (UMR 5506), 161, rue Ada, 34392 Montpellier Cedex 5(France), `berthe@lirmm.fr`

[‡] LIP (UMR 5668), ENS Lyon, 49, Allée d'Italie, 69364 Lyon Cedex 7 (France), `bertrand.nouvel@ens-lyon.fr`

by use of the coding with respect to a three-interval partition of a $\mathbb{Z}^2$-action by two irrational rotations on $\mathbb{T}^1$. One thus gets two-dimensional words over a three-letter alphabet that can be considered as two-dimensional Sturmian words [BV00].

We consider here a further generalization. Indeed, we study configurations associated with a discrete rotation, defined as the composition of a Euclidean rotation with a rounding operation. It is possible to encode all the information concerning a discrete rotation as two multidimensional words C_α and C'_α that we call configurations. The main purpose of the present paper is to prove that both configurations are codings of a $\mathbb{Z}^2$-action by two rotations on $\mathbb{T}^2$ with respect to a partition into a finite number of rectangles. We then deduce results concerning the density of each symbol in C_α and C'_α. As a motivation for this study, let note that we plan to use these results in a next future for an algorithm of randomization of discrete rotations.

2 Conventions

We work in the *discrete plane* $\mathbb{Z}^2$. For each point $\mathbf{v}$, $x_\mathbf{v}$ denotes its horizontal coordinate and $y_\mathbf{v}$ its vertical coordinate.

Let x be a real number. We recall that the floor function $x \mapsto \lfloor x \rfloor$ is defined as the greatest integer less or equal to x. The *rounding function* is defined as $[x] := \lfloor x + 0.5 \rfloor$ and $\{x\} := x - [x]$. These applications can be extended to vectors, by independent application on each component of the vector.

The *discretization cell* of the point $\mathbf{v} \in \mathbb{Z}^2$ is defined as the set of elements $\mathbf{w}$ in $\mathbb{R}^2$ which have the same image by discretization as $\mathbf{v}$, i.e., $[\mathbf{v}] = [\mathbf{w}]$. Hence the discretization cell of $\mathbf{v}$ is defined as the half-opened unit square centered in $[\mathbf{v}]$.

We use the canonical bijection between the torus $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$ and the square $\{\mathbf{v} \in \mathbb{R}^2; x_\mathbf{v} \in [-\frac{1}{2}, \frac{1}{2}[$ and $y_\mathbf{v} \in [-\frac{1}{2}, \frac{1}{2}[\}$, i.e., the discretization cell of 0. By abuse of notation, we also denote by $\{\mathbf{v}\}$ the image under the canonical projection from $\mathbb{R}^2$ onto $\mathbb{T}^2$ of a point $\mathbf{v} \in \mathbb{R}^2$. Hence let us stress the fact that the map $x \mapsto \{x\}$ is an additive morphism from $\mathbb{R}^2$ onto $\mathbb{T}^2$.

Without loss of generality, we assume throughout this paper that $\alpha \in [0, \pi/4]$: the arguments used here can be easily extended to the case of any other octant. We denote by r_α the Euclidean rotation of angle α:

$$r_\alpha : \ \mathbb{R}^2 \to \mathbb{R}^2, \ \mathbf{v} \mapsto \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \mathbf{v}.$$

The discrete rotation $[r_\alpha]$ is defined as

$$[r_\alpha] : \mathbb{Z}^2 \to \mathbb{Z}^2, \ \mathbf{v} \mapsto [r_\alpha(\mathbf{v})].$$

By $\{r_\alpha\}$ we mean the map $\{r_\alpha\} : \ \mathbb{Z}^2 \to \mathbb{T}^2, \ \mathbf{v} \mapsto \{r_\alpha(\mathbf{v})\}$.

We denote by $(\mathbf{i}, \mathbf{j})$ the canonical basis of the Euclidean space $\mathbb{R}^2$. We similarly use the notation $\mathbf{i}_\alpha := r_\alpha(\mathbf{i})$ and $\mathbf{j}_\alpha := r_\alpha(\mathbf{j})$.

Let Q be a finite set called alphabet. A two-dimensional word in $Q^{\mathbb{Z}^2}$ is called a *configuration* over Q. An application from $\{0, 1, \cdots, n-1\} \times \{0, 1, \cdots, m-1\}$ to Q is called a *pattern* of size $[m, n]$. Let C be a configuration in $Q^{\mathbb{Z}^2}$. A pattern χ of size $[m, n]$ occurs at position $\mathbf{p}$ in C if $C(\mathbf{p} + \mathbf{v}) = \chi(\mathbf{v})$, for all $\mathbf{v}$ with $x_\mathbf{v}, y_\mathbf{v} \in \{0, 1, \cdots, n-1\} \times \{0, 1, \cdots, m-1\}$. We define $C^{[m,n]}$ as the configuration with values in the finite alphabet consisting of the patterns of size $[m, n]$ over Q, that is defined as the application that returns the pattern of size $[m, n]$ that occurs at the specified position in the configuration.

The *density* of the symbol $p \in Q$ in the configuration $C \in Q^{\mathbb{Z}^2}$ is defined as the following limit (if it exists):

$$\eta_C(p) = \lim_{n \to \infty} \frac{\#\{\mathbf{v} \in \mathbb{Z}^2,\ x_v, y_v \in \{-n, \cdots, n\} \text{ and } C(\mathbf{v}) = p\}}{(2n+1)^2}.$$

A *dynamical system* (X, T) is defined as the action of a continuous and onto map T on a compact space X. Given two continuous and onto maps T_1 and T_2 acting on X and satisfying $T_1 \circ T_2 = T_2 \circ T_1$, the $\mathbb{Z}^2$*-action* by T_1 and T_2 on X, that we denote (X, T_1, T_2), is defined by

$$\forall (m, n) \in \mathbb{Z}^2,\ \forall x \in X,\quad (m, n) \cdot x = T_1^m \circ T_2^n(x).$$

It is natural to associate a two-dimensional symbolic dynamical system to the triple (X, T_1, T_2) by coding the orbits of the points of X under the $\mathbb{Z}^2$-action as follows: given $x_0 \in X$ and given a *labelling function* l defined on X with values in a finite set Q that takes constant values on the atoms of a finite partition of X, the configuration C defined by

$$\forall (m, n) \in \mathbb{Z}^2,\ C(m, n) = l(T_1^m \circ T_2^n(x_0))$$

is called the coding of the orbit of x_0 under the $\mathbb{Z}^2$-action (X, T_1, T_2) with respect to the labelling function l.

3 Dynamical System Associated to C_α

According to [NR05], we associate a first configuration C_α to the discrete rotation $[r_\alpha]$ that encodes all the information concerning the discrete rotation (there exists indeed a planar transducer that uses the configuration C_α as input and gradually computes the action of the discrete rotation). For a given $\mathbf{v} \in \mathbb{Z}^2$, let $\mathcal{V}_4$ denote the set of 4-neighbours of $\mathbf{v}$, that is, $\mathcal{V}_4 = \{\mathbf{v} + \mathbf{i},\ \mathbf{v} + \mathbf{j},\ \mathbf{v} - \mathbf{i},\ \mathbf{v} - \mathbf{j}\}$. The configuration C_α maps each point $\mathbf{v}$ of $\mathbb{Z}^2$ to the set $[r_\alpha](\mathcal{V}_4) - [r_\alpha][\mathbf{v}]$, that is,

$C_\alpha(\mathbf{v}) := \{\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ with $(a_k = [r_\alpha(\mathbf{v} + r_{\pi/2}^k(\mathbf{i}))] - [r_\alpha(\mathbf{v})]$ for $k = 0, \cdots, 3)$.

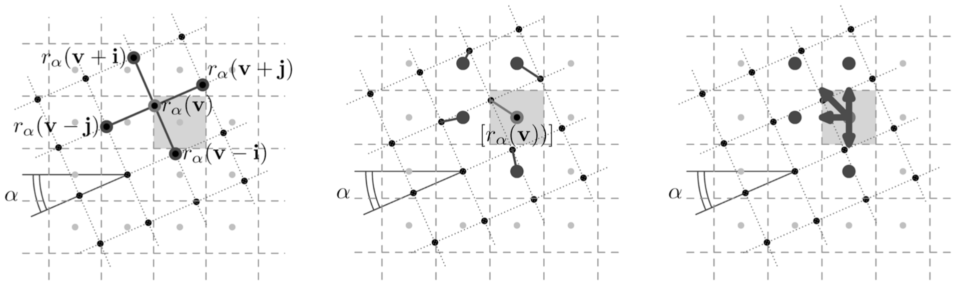

$r_\alpha(\mathbf{v}+\mathbf{i})$
$r_\alpha(\mathbf{v}+\mathbf{j})$
$r_\alpha(\mathbf{v})$
$r_\alpha(\mathbf{v}-\mathbf{j})$
$r_\alpha(\mathbf{v}-\mathbf{i})$
α
$[r_\alpha(\mathbf{v})]$
α
α

Figure 1: A progressive construction of the configuration C_α: we represent the set of vectors that leads to the relative position of the 4-neighbors of $\mathbf{v}$ after the action of the discrete rotation.

Let us note that C_α contains 3 or 4 non-zero elements, according to [NR03]. Let Q_α denote the finite set of values taken by C_α.

We define a *frame* of the torus $\mathbb{T}^2 \equiv [-\frac{1}{2}, \frac{1}{2}[\times [-\frac{1}{2}, \frac{1}{2}[$ as a rectangle of the form $[a, b[\times[c, d[$, with $-\frac{1}{2} \leq a \leq b < \frac{1}{2}$ and $-\frac{1}{2} \leq c \leq b < \frac{1}{2}$. The interpretation of C_α as a coding a $\mathbb{Z}^2$-action is based on the following result:

Theorem 3.1 ([NR05]) *There exists a partition P_α of the torus $\mathbb{T}^2$ into a finite number of frames such that for each $p \in Q_\alpha$, there exists a frame I_p such that for all $\mathbf{v} \in \mathbb{Z}^2$, then $C_\alpha(\mathbf{v}) = p$ if and only if $\{r_\alpha(\mathbf{v})\} \in I_p$.*

Consider the following two actions $T_{\mathbf{i}_\alpha} : \mathbb{T}^2 \to \mathbb{T}^2,\ x \mapsto x + \{\mathbf{i}_\alpha\}$, $T_{\mathbf{j}_\alpha} : \mathbb{T}^2 \to \mathbb{T}^2,\ x \mapsto x + \{\mathbf{j}_\alpha\}$. One has for every $\mathbf{v} \in ZZ^2$, $\{r_\alpha(\mathbf{v})\} = T_{\mathbf{i}_\alpha}^{x_\mathbf{v}} \circ T_{\mathbf{i}_\alpha}^{y_\mathbf{v}}(\mathbf{0})$. Let us define l_{C_α} as the labelling function associated to the partition P_α defined by $l_{C_\alpha} : \mathbb{T}^2 \to Q_\alpha,\ \mathbf{v} \mapsto \phi_c(f_{C_\alpha}(\mathbf{v}_x), f_{C_\alpha}(\mathbf{v}_y))$ with f_{C_α} defined as follows:

if $\alpha \in [0, \pi/6]$:

$$\left|\begin{array}{ll} [-\frac{1}{2}, \frac{1}{2} - \cos(\alpha)[& \mapsto 0 \\ [\frac{1}{2} - \cos(\alpha), \sin(\alpha) - \frac{1}{2}[& \mapsto 1 \\ [\sin(\alpha) - \frac{1}{2}, \frac{1}{2} - \sin(\alpha)[& \mapsto 2 \\ [\frac{1}{2} - \sin(\alpha), \cos(\alpha) - \frac{1}{2}[& \mapsto 3 \\ [\cos(\alpha) - \frac{1}{2}, \frac{1}{2}[& \mapsto 4 \end{array}\right|$$

if $\alpha \in [\pi/6, \pi/4]$:

$$\left|\begin{array}{ll} [-\frac{1}{2}, \frac{1}{2} - \cos(\alpha)[& \mapsto 0 \\ [\frac{1}{2} - \cos(\alpha), \frac{1}{2} - \sin(\alpha)[& \mapsto 1 \\ [\frac{1}{2} - \sin(\alpha), \sin(\alpha) - \frac{1}{2}[& \mapsto 5 \\ [\sin(\alpha) - \frac{1}{2}, \cos(\alpha) - \frac{1}{2}[& \mapsto 3 \\ [\cos(\alpha) - \frac{1}{2}, \frac{1}{2}[& \mapsto 4 \end{array}\right|$$

where ϕ_c is described in Figure 2. The values taken by C_α, that is, the elements of Q_α are represented in Figure 2 as sets of vectors.

Theorem 3.1 can then be reformulated as follows: C_α is the coding of the orbit of $\mathbf{0}$ under the $\mathbb{Z}^2$-action $(\mathbb{T}^2, T_{\mathbf{i}_\alpha}, T_{\mathbf{i}_\alpha})$ with respect to the labelling function l_{C_α}.

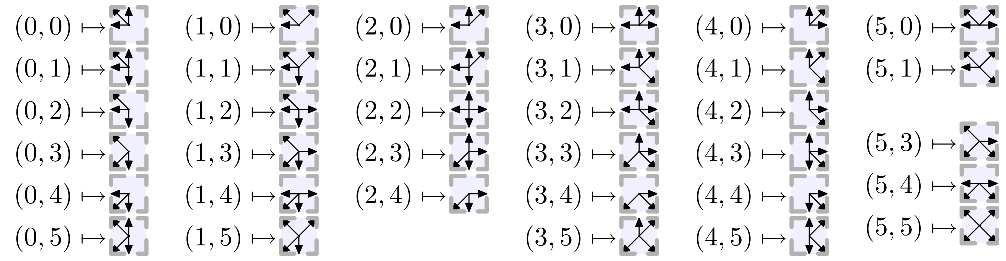


Figure 2: Table describing the action of ϕ_c. The symbols represent the all the vectors of the set.

4 Distribution of Symbols in C_α

We can now deduce from the $\mathbb{Z}^2$-action introduced in Section 3 results concerning the densities of symbols in C_α by using classical tools from symbolic dynamics and ergodic theory.

Let $G_\alpha \subseteq \mathbb{T}^2$ denote the orbit of $\mathbf{0}$ under the $\mathbb{Z}^2$-action $(\mathbb{T}^2, T_{\mathbf{i}_\alpha}, T_{\mathbf{i}_\alpha})$ with respect to the labelling function l_{C_α}: this very orbit is the orbit coded by the configuration C_α. In other words, G_α is the image by the canonical projection $x \mapsto \{x\}$ onto $\mathbb{T}^2$ of the lattice $L_\alpha := \mathbb{Z}\mathbf{i}_\alpha + \mathbb{Z}\mathbf{j}_\alpha$; G_α has a group structure, and is invariant by rotation by $\pi/2$.

Let us recall that an angle α is said *Pythagorean* if $\cos\alpha$ and $\sin\alpha$ are both rational. Let us distinguish two cases according to the fact that α is Pythagorean or not, that is, according to the density of G_α in $\mathbb{T}^2$.

The Dense Case

Lemma 4.1 *We assume that α is not Pythagorean. For every symbol $p \in Q_\alpha$, its density $\eta_{C_\alpha}(p)$ exists and is equal to the area of the frame I_p defined in Theorem 3.1.*

Proof (Sketch) If either $\cos(\alpha)$ or $\sin(\alpha)$ is irrational, then one cannot have simultaneously $p\cos(\alpha) + q\sin(\alpha) \in \mathbb{Z}$ and $-p\sin(\alpha) + q\cos(\alpha) \in \mathbb{Z}$, for any $(p, q) \in \mathbb{Z}^2$. Hence one concludes by using a classical argument on Weyl sums. □

The Pythagorean Case

If α is a Pythagorean angle then G_α is not dense in the torus $\mathbb{T}^2$: indeed, G_α is a finite cyclic group. It has order c where $(a, b, c) \in \mathbb{N}^3$ is the prime Pythagorean triple satisfying $1 \le b \le a \le c$, $a^2 + b^2 = c^2$, $gcd(a, b, c) = 1$ and $c\exp(i\alpha) = a + ib$ that generates the angle α. More information on Pythagorean angles can be found in [NR04].

Lemma 4.2 *Let $\alpha \in [0, ...\pi/4[$ be a Pythagorean angle. Let c denote the order of the cyclic group G_α. The density $\eta_{C_\alpha}(p)$ of the symbol p in C_α satisfies*

$$\forall p \in Q_\alpha, \eta_{C_\alpha}(p) = \frac{Card\ (G'_\alpha \cap I_p)}{c}.$$

Proof (Sketch) By definition,

$$\eta_{C_\alpha}(p) = \lim_{n\to\infty} (\boldsymbol{\{}r_\alpha\boldsymbol{\}}(\{-n, \cdots, n\}^2) \cap I_p)/(2n+1)^2.$$

One first checks that

$$\eta_{C_\alpha}(p) = \lim_{n\to\infty} (\boldsymbol{\{}r_\alpha\boldsymbol{\}}(\{-c\lfloor n/c\rfloor, ..., c\lfloor n/c\rfloor\}^2) \cap I_p)/(2n+1)^2).$$

But as G_α is cyclic and of order c, then

$$\eta_{C_\alpha}(p) = \frac{\boldsymbol{\{}r_\alpha\boldsymbol{\}}(\{0, ..., c-1\}^2) \cap I_p}{c^2} = \frac{\text{Card } (G'_\alpha \cap I_p)}{c}.$$

□

5 Distribution of Symbols in C'_α

Let us define now C'_α:

$$\forall \mathbf{v} \in \mathbb{Z}^2,\ C'_\alpha(\mathbf{v}) := \bigcup_{\mathbf{w} \text{ such that } [r_\alpha(\mathbf{w})]=\mathbf{v}} C_\alpha(\mathbf{w}).$$

Let Q'_α denote the set of values taken by C'_α. We want to state a result analogous to Theorem 3.1 in order, first, to interpret the configuration C'_α as a coding of a symbolic dynamical system, and second, to compute the densities of the symbols in C'_α. Let us note that Corollary 1 in [NR05] does not directly yield a dynamical interpretation of C'_α.

Our strategy in order to describe C'_α as a coding of a $\mathbb{Z}^2$-action is the following. We first create a "block configuration" by working with patterns of size $[2, 2]$ that occur in C'_α. We then introduce a particular domain of $\mathbb{R}^2$ that is a fundamental domain for the lattice $\mathbb{Z}\mathbf{i}_\alpha + \mathbb{Z}\mathbf{j}_\alpha$, such that if we know the projection of a point $\mathbf{p} \in \mathbb{Z}\mathbf{i}_\alpha + \mathbb{Z}\mathbf{j}_\alpha$ in that domain, then we can recover the symbols that appear in the block configuration; therefore we find out what are the symbols that appear in C'_α. We thus deduce a symbolic dynamical system for the block configuration. Finally, we use this dynamical system, in order to get the density of the symbols both in the block configuration and in C'_α.

5.1 Dynamical System for C'_{B_α}

Let $C'_{B_\alpha}(\mathbf{v})$ be defined as the following 2×2-block configuration:

$$\forall \mathbf{v} \in \mathbb{Z}^2,\ C'_{B_\alpha}(\mathbf{v}) = C'^{[2,2]}_\alpha(2\mathbf{v}).$$

Since $C'_{B_\alpha}(\mathbf{v})$ is an application that returns patterns of size $[2,2]$, then $C'_\alpha(\mathbf{v}) = \big(C'_{B_\alpha}(\lfloor x_\mathbf{v}/2\rfloor, \lfloor y_\mathbf{v}/2\rfloor)\big)\,(\mathbf{v}_x \bmod 2, \mathbf{v}_y \bmod 2)$. For any $\mathbf{v} \in \mathbb{Z}^2$, one sets

$$F_B(\mathbf{v}) = [x_\mathbf{v} - \frac{1}{2}, x_\mathbf{v} + \frac{3}{2}[\times[y_\mathbf{v} - \frac{1}{2}, y_\mathbf{v} + \frac{3}{2}[.$$

The introduction of this block configuration is natural, since the intersection between $F_B(\mathbf{v})$ and $r_\alpha(\mathbb{Z}^2) = \mathbb{Z}\mathbf{i}_\alpha + \mathbb{Z}\mathbf{j}_\alpha$ is nonempty for every $\mathbf{v} \in \mathbb{Z}^2$; this is a direct consequence of the fact that two holes (a hole is an element $\mathbf{v} \in \mathbb{Z}^2$ that has no antecedent by $[r_\alpha]$) can never be adjacent (see [NR04]). An example of a hole is depicted in Figure 4 below. Let

$$F_{D_\alpha} := \left([-\frac{1}{2}, \cos\alpha - \frac{1}{2}[\right)^2 \cup \left([\cos\alpha - \frac{1}{2}, \cos\alpha + \sin\alpha - \frac{1}{2}[\times[-\frac{1}{2}, \sin\alpha - \frac{1}{2}[\right).$$

The set F_{D_α} is a fundamental domain for the lattice $L_\alpha = \mathbb{Z}\mathbf{i}_\alpha + \mathbb{Z}\mathbf{j}_\alpha$ (see Figure 3). Hence for any $\mathbf{v} \in \mathbb{Z}^2$, there exists a unique $\mathbf{w} \in \mathbb{Z}^2$ such that $r_\alpha(\mathbf{w}) \in \mathbf{v} + F_{D_\alpha}$. Therefore for all $\mathbf{v} \in \mathbb{Z}^2$, we first define

$$\theta' :\ \mathbb{Z}^2 \to L_\alpha,\ \mathbf{v} \mapsto r_\alpha(\mathbf{w}),$$

where $\mathbf{w}$ is the unique point such that $r_\alpha(\mathbf{w}) \in \mathbf{v} + F_{D_\alpha}$, and then

$$\theta :\ \mathbb{Z}^2 \to F_{D_\alpha},\ \mathbf{v} \mapsto \theta'(\mathbf{v}) - \mathbf{v}.$$

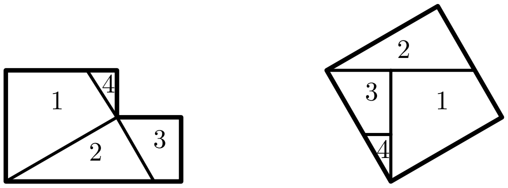


Figure 3: An exchange of pieces between F_{D_α} and the canonical representation of $\mathbb{R}^2/L_\alpha$. This exchange of pieces only requires translations of the form $k\mathbf{i}_\alpha + k'\mathbf{j}_\alpha$, with $k, k' \in \mathbb{Z}$.

Theorem 5.1 *There exists a partition of F_{D_α} into a finite number of frames $J_{p'}$, for p' pattern of size $[2,2]$ that occurs in C'_α, such that for all $\mathbf{v} \in \mathbb{Z}^2$, $\theta(2\mathbf{v}) \in J_{p'}$ if and only if $C'_{B_\alpha}(\mathbf{v}) = p'$.*

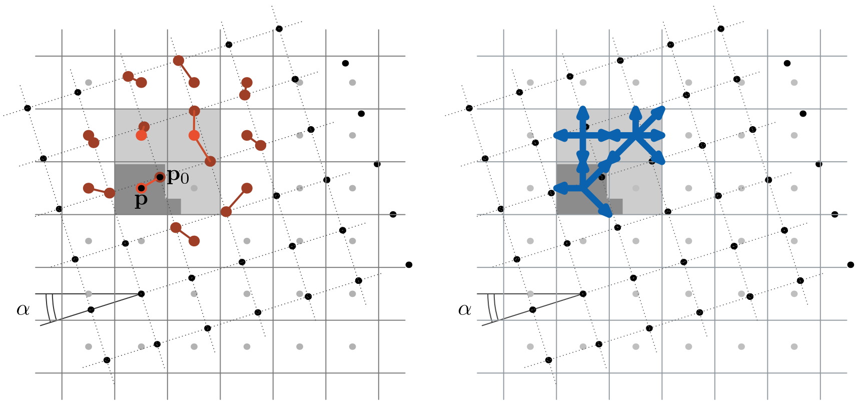


Figure 4: From a point $\mathbf{p}_0 = \theta'(2\mathbf{v}) \in \mathbb{Z}\mathbf{i}_\alpha + \mathbb{Z}\mathbf{j}_\alpha$ that falls into the domain $F_{D_\alpha}(2\mathbf{v})$ (in dark gray), we can recover all the symbols of C'_α that contribute to the block of size $[2, 2]$ whose associated domain is $F_B(2\mathbf{v})$ in light gray.

Proof (Sketch) The proof is based on the following idea: from the location of $\theta(2\mathbf{v})$ in F_{D_α}, it is possible to deduce the value of $C'_{B_\alpha}(\mathbf{v})$. We notice that, for all the points $\mathbf{w}$ of $\mathbb{Z}^2$ that have their image by r_α in $F_B(2\mathbf{v})$ we can compute $C_\alpha(\mathbf{w})$. Indeed we show that if $x_{\theta(2\mathbf{v})} < \frac{1}{2}$, $[\theta(2\mathbf{v})] = 0$, else $[\theta(2\mathbf{v})] = 1$; we thus deduce $C_\alpha(\mathbf{w})$ from $\{\theta'(2\mathbf{v})\}$, according to Theorem 3.1. The same argument applies for all the points $\mathbf{w}' = r_\alpha(\mathbf{w})$ of $\mathbb{Z}_{\mathbf{i}_\alpha} + \mathbb{Z}_{\mathbf{j}_\alpha}$ that are inside $F_B(2\mathbf{v})$; note that $\mathbf{w}' = \theta(2\mathbf{v}) + k\mathbf{i}_\alpha + k'\mathbf{j}_\alpha$, with $k, k' \in \mathbb{Z}$. We thus similarly localize the position in $(2\mathbf{v} + \{0,1\}^2)$ of all the images of points in $\mathbb{Z}\mathbf{i}_\alpha + \mathbb{Z}\mathbf{j}_\alpha \cap F_B(2\mathbf{v})$. This is sufficient to conclude that we can infer the pattern $C'_B(\mathbf{v})$ from $\theta(2\mathbf{v})$. □

Let $l_{C'_{B_\alpha}}$ be the labeling function given by the partition of Theorem 5.1 that associates to a frame in F_{D_α} the corresponding pattern of size $[2, 2]$.

From Theorem 5.1, we deduce that

$$\forall \mathbf{v} \in \mathbb{Z}^2, C'_{B_\alpha}(\mathbf{v}) = l_{C'_{B_\alpha}}(\theta(2\mathbf{v}))$$

Now, let $\mathbb{T}^2_\alpha = \mathbb{R}^2/(\mathbb{Z}\mathbf{i}_\alpha + \mathbb{Z}\mathbf{j}_\alpha)$; we denote as $\mathbf{v} \mapsto \{\mathbf{v}\}_\alpha$ the canonical projection on $\mathbb{T}^2_\alpha$, that is in one-to-correspondence with F_{D_α}. One has

$$\forall \mathbf{v} \in \mathbb{Z}^2,\ \theta(\mathbf{v}) \equiv -\{\mathbf{v}\}_\alpha \text{ modulo } L_\alpha.$$

Finally, the configuration C'_{B_α} is a coding of the orbit 0 under the $\mathbb{Z}^2$-action $(\mathbb{T}^2_\alpha, \mathbf{v} \mapsto \mathbf{v} + \{\mathbf{i}\}_\alpha, \mathbf{v} \mapsto \mathbf{v} + \{\mathbf{j}\}_\alpha)$ with respect to the labelling function $l_{C'_{B_\alpha}}$.

5.2 Application

We assume that α is not a Pythagorean angle. Similarly as in the study of C_α, the orbit of 0 under the $\mathbb{Z}^2$-action is dense and uniformly distributed in $\mathbb{T}^2_\alpha$. We thus deduce that

$$\forall p \in Q'_\alpha,\ \eta_{C'_\alpha}(p) = \sum_{p' \in Q_\alpha^{[2,2]}} n(p', p)\, \mu(f_{p'}),$$

where $Q_\alpha^{[2,2]}$ is the set of patterns of size $[2,2]$ that occur in C'_α, $n(p', p)$ is the function that returns the number of occurrences of p in the pattern p' of size $[2,2]$, and $\mu(J_{p'})$ denotes the area of frame $J_{p'}$ associated to the symbol p' according to Theorem 5.1.

However practically, the computations for these symbolic maps are quite tedious. For each symbol p, there exist 40 patterns p' of size $[2,2]$ to compute. This leads to approximatively 360 inequations... and there are approximatively 25 symbols p to consider! See [BN05] for a program that handles these symbolical expressions. The results describing the densities of the symbols in C'_α have been summarized in Figure 6.

Let us note that in the Pythagorean case, the theory is also similar to the

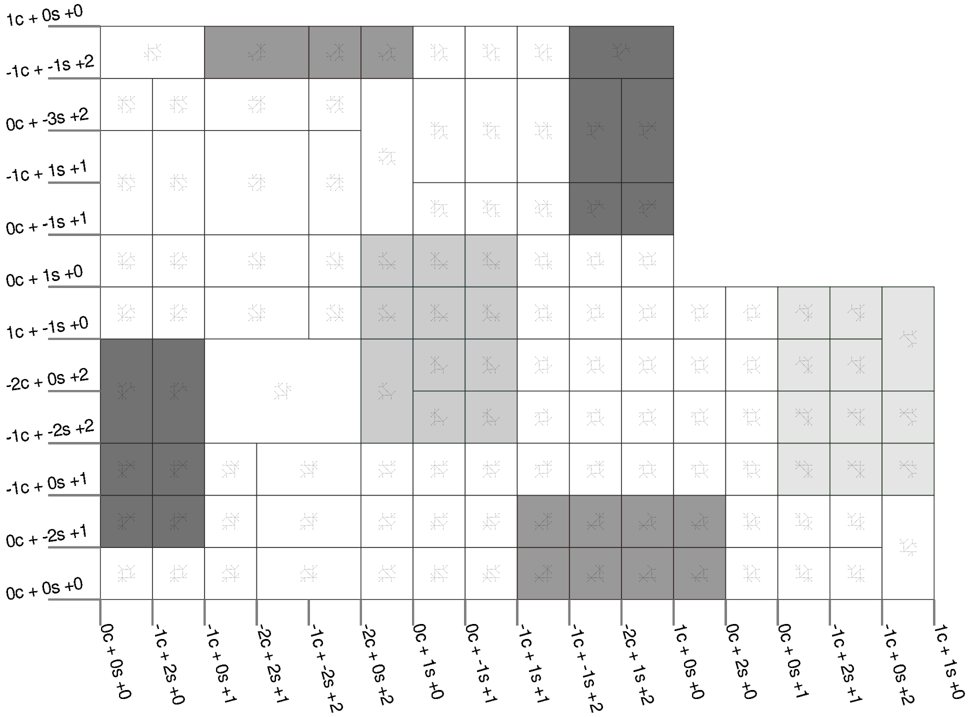
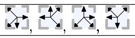
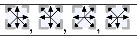


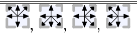
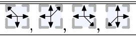
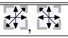
Figure 5: A partition of the domain F_{D_α}, for $\alpha \approx 0.464705$ rad. This partition gives according to the position of $\theta(2\mathbf{v})$ inside that domain the pattern of size $[2,2]$ that appears in $C'_{B_\alpha}(\mathbf{v})$. On the axis the positions are labeled by expressions of the form $kc+k's+k''$, meaning that the corresponding line is located at $k\cos(\alpha)+k'\sin(\alpha)+k''-\frac{1}{2}$ in F_{D_α}. For readability reasons, the scale is monotone but not linear.

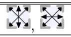
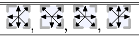
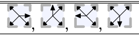
one developed for C_α.

5.2.1 Remark

Let us observe that all the results we have given here for symbols are extendable without major difficulty to patterns of a given size $[m, n]$. Actually, a frame is associated to each pattern, and the same theory can be used.

α			
$[0..\arctan(\sqrt{2}/4)]$	$2\cos(\alpha)sin(\alpha)-2\cos(\alpha)-2\sin(\alpha)+2$	$2\cos(\alpha)sin(\alpha)-2\cos(\alpha)-2\sin(\alpha)+2$	$-(\cos(\alpha))^2-\cos(\alpha)sin(\alpha)+2\cos(\alpha)+\sin(\alpha)-1$
$[\arctan(\sqrt{2}/4)..\arctan(1/2)]$	$2\cos(\alpha)sin(\alpha)-2\cos(\alpha)-2\sin(\alpha)+2$	$2\cos(\alpha)sin(\alpha)-2\cos(\alpha)-2\sin(\alpha)+2$	$-(\cos(\alpha))^2+2\cos(\alpha)sin(\alpha)+\cos(\alpha)-2\sin(\alpha)$
$[\arctan(1/2)..\pi/6]$	$2\cos(\alpha)sin(\alpha)-2\cos(\alpha)-2\sin(\alpha)+2$	$2\cos(\alpha)sin(\alpha)-2\cos(\alpha)-2\sin(\alpha)+2$	0
$[\pi/6..\arctan(3/4)]$	$2\cos(\alpha)sin(\alpha)-2\cos(\alpha)-2\sin(\alpha)+2$	$-2\cos(\alpha)sin(\alpha)+1$	0
$[\arctan(3/4)..\pi/4]$	$2\cos(\alpha)sin(\alpha)-2\cos(\alpha)-2\sin(\alpha)+2$	$-2\cos(\alpha)sin(\alpha)+1$	0

α			
$[0..\arctan(\sqrt{2}/4)]$	$(\cos(\alpha))^2-2\cos(\alpha)+1$	$-2(\sin(\alpha))^2-2\cos(\alpha)sin(\alpha)+\cos(\alpha)+3\sin(\alpha)-1$	$3\cos(\alpha)sin(\alpha)-\cos(\alpha)-3\sin(\alpha)+1$
$[\arctan(\sqrt{2}/4)..\arctan(1/2)]$	$(\cos(\alpha))^2-2\cos(\alpha)+1$	$-2(\sin(\alpha))^2-2\cos(\alpha)sin(\alpha)+\cos(\alpha)+3\sin(\alpha)-1$	0
$[\arctan(1/2)..\pi/6]$	$2\cos(\alpha)sin(\alpha)-\cos(\alpha)-2\sin(\alpha)+1$	$-2(\sin(\alpha))^2-2\cos(\alpha)sin(\alpha)+\cos(\alpha)+3\sin(\alpha)-1$	0
$[\pi/6..\arctan(3/4)]$	0	0	0
$[\arctan(3/4)..\pi/4]$	0	0	$2(\cos(\alpha))^2-\cos(\alpha)sin(\alpha)-3\cos(\alpha)+\sin(\alpha)+1$

α			
$[0..\arctan(\sqrt{2}/4)]$	0	0	0
$[\arctan(\sqrt{2}/4)..\arctan(1/2)]$	$-3\cos(\alpha)sin(\alpha)+\cos(\alpha)+3\sin(\alpha)-1$	0	0
$[\arctan(1/2)..\pi/6]$	$-2(\cos(\alpha))^2+\cos(\alpha)sin(\alpha)+3\cos(\alpha)-\sin(\alpha)-1$	$(\cos(\alpha))^2-2\cos(\alpha)sin(\alpha)-\cos(\alpha)+2\sin(\alpha)$	0
$[\pi/6..\arctan(3/4)]$	$-2(\cos(\alpha))^2+\cos(\alpha)sin(\alpha)+3\cos(\alpha)-\sin(\alpha)-1$	$(\cos(\alpha))^2-2\cos(\alpha)+1$	$-2(\sin(\alpha))^2+2\cos(\alpha)sin(\alpha)-\cos(\alpha)+\sin(\alpha)$
$[\arctan(3/4)..\pi/4]$	0	$-(\cos(\alpha))^2+\cos(\alpha)sin(\alpha)+\cos(\alpha)-\sin(\alpha)$	$-2(\sin(\alpha))^2+2\cos(\alpha)sin(\alpha)-\cos(\alpha)+\sin(\alpha)$

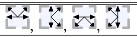
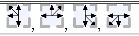
α				
$[0..\arctan(\sqrt{2}/4)]$	0	$-(\cos(\alpha))^2-\cos(\alpha)sin(\alpha)+2\cos(\alpha)+\sin(\alpha)-1$	0	$4(\sin(\alpha))^2-4\sin(\alpha)+1$
$[\arctan(\sqrt{2}/4)..\arctan(1/2)]$	0	$-(\cos(\alpha))^2-\cos(\alpha)sin(\alpha)+2\cos(\alpha)+\sin(\alpha)-1$	0	$4(\sin(\alpha))^2-4\sin(\alpha)+1$
$[\arctan(1/2)..\pi/6]$	0	$-(\cos(\alpha))^2-\cos(\alpha)sin(\alpha)+2\cos(\alpha)+\sin(\alpha)-1$	0	$4(\sin(\alpha))^2-4\sin(\alpha)+1$
$[\pi/6..\arctan(3/4)]$	$-2\cos(\alpha)sin(\alpha)+\cos(\alpha)+2\sin(\alpha)-1$	$-(\cos(\alpha))^2+\cos(\alpha)sin(\alpha)+\cos(\alpha)-\sin(\alpha)$	$4(\sin(\alpha))^2-4\sin(\alpha)+1$	0
$[\arctan(3/4)..\pi/4]$	$-2\cos(\alpha)sin(\alpha)+\cos(\alpha)+2\sin(\alpha)-1$	$-(\cos(\alpha))^2+\cos(\alpha)sin(\alpha)+\cos(\alpha)-\sin(\alpha)$	$4(\sin(\alpha))^2-4\sin(\alpha)+1$	0

Figure 6: Table describing $\eta_{C'_\alpha}(p)$ for each symbol p that appears in C'_α, with respect to the value of α.

References

[BN05] V. Berthé and B. Nouvel. Dynamical systems and distribution of symbols in rotations configurations. `http://perso.ens-lyon.fr/bertrand.nouvel/pub/DSDSRC.pdf`, 2005.

[BV00] V. Berthé and L. Vuillon. Tilings and rotations on the torus: a two-dimensional generalization of Sturmian sequences. *Discrete Math.*, 223:27–53, 2000.

[Fog02] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[Lot02] M. Lothaire. *Algebraic combinatorics on words*. Cambridge University Press, 2002.

[Lot05] M. Lothaire. *Applied combinatorics on words*. Cambridge University Press, 2005.

[NR03] B. Nouvel and E. Rémila. On colorations induced by discrete rotations. In *DGCI*, number 2886 in LNCS, pages 174–183, 2003.

[NR04] B. Nouvel and E. Rémila. Characterization of bijective discretized rotations. In *International Workshop on Combinatorial Images Analysis, 10th International Conference, IWCIA 2004, Auckland, New Zealand, December 1-4, 2004*, number 3322 in LNCS, 2004.

[NR05] B. Nouvel and E. Rémila. Configurations induced by discrete rotations: Periodicity and quasiperiodicity properties. *Discrete Applied Mathematics*, 2-3(147):325–343, 2005.

[Rau82] G. Rauzy. Nombres algébriques et substitutions. *Bull. Soc. Math. France*, 110(2):147–178, 1982.

Complexity and palindromic complexity of billiard words

Jean-Pierre Borel[*]

Abstract

We present some recent results on palindromic factors and prefixes of billiard words in a k-dimensional space, with $k \geq 3$. The language of these words is already known in the usual case. We give a geometrical characterization of factors of billiard words, using some projection on a $(k-1)$-dimensional space. As a consequence, we get some results on the complexity and palindromic complexity of these words, in the non-usual case. For example, we get some billiard words on 3 letters without any palindromic factor of even length, or billiard words on 4 letters whose palindromic factors have a bounded length. All the results are obtained by geometrical methods.

Keywords Languages, Billiard words, Complexity, Palindromic factors.

AMS classification 68R15.

1 k-dimensional billiard words

1.1 Billiard words starting at the origin

Let $\mathcal{D}$ be the half-line of origin O, in the k-dimensional space $\mathbb{R}^k$, and parallel to the positive vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$. Then we define the associated billiard word, or *cutting sequence*, (starting from O) denoted by $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ on the alphabet $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$ as shown in Figure 1. In dimension 2, this can be made using the three following methods:

1. encoding by a_1 the black horizontal unitary segment and by a_2 the black vertical unitary segment (see Figure 1 (a)). Then $a_1 c_{\alpha_1,\alpha_2}$ encodes the discrete path immediately below the half-line, hence

   $$c_{\alpha_1,\alpha_2} = a_2a_1a_2a_1a_2a_2a_1a_2a_1 \ldots$$

   in Figure 1 (a). The infinite word $a_1 c_{\alpha_1,\alpha_2}$ is the well-known *Christoffel word*;

[*] LACO, UMR CNRS 6090 - 123 avenue Albert Thomas, F-87060, LIMOGES CEDEX (FRANCE), borel@unilim.fr, partially supported by Région Limousin

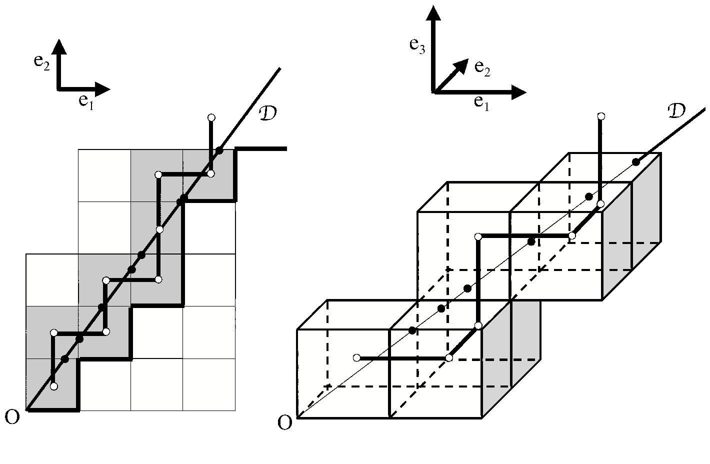


Figure 1: (a) and (b)

2. moving from the origin to infinity, encode the sequence of intercepts between $\mathcal{D}$ and the grid, using a_1 for a vertical line and a_2 for an horizontal one (black points on Figure 1 (a));

3. by looking at the sequence of the centers (white points) of the unit squares crossed by $\mathcal{D}$. Two consecutive centers correspond to joining squares, so that the vector joining these two points is one of the two vectors of the canonical basis (e_1, e_2). Then encode by a_j the vector e_j, $j = 1, 2$.

In higher dimension $k \geq 3$, both methods 2 and 3 can be generalized. We consider now the *facets* of the unit k-cubes crossed by $\mathcal{D}$, instead of the sides of the unit squares (see Figure 1 (b), with $k = 3$). In this figure, the crossing points are the black points and the centers of units k-cubes are the white ones, as in Figure 1 (a). In both cases, we encode the vectors e_j $(1 \leq j \leq k)$ of the canonical basis by the letters a_j, and a crossed facet by its orthogonal direction, and we get the billiard word $c_{\alpha_1,\alpha_2,\alpha_3} = a_1a_2a_3a_1a_2a_3\ldots$. This works as long as the half-line $\mathcal{D}$ crosses each facet in its interior (so we can define *consecutive* crossed unit k-cubes), i.e., $\mathcal{D}$ does not contain any point with two integer coordinates, except for O. This property corresponds to the following condition:

$$\frac{\alpha_i}{\alpha_j} \notin \mathbb{Q} \text{ , } i \neq j. \tag{1.1}$$

This condition already holds in the usual case:

$$\text{the } \alpha_i \text{ are } \mathbb{Q}\text{-linearly independent.} \tag{1.2}$$

Then we say that $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ is *totally irrational.* This strong hypothesis has been made in the former works in this topic. Note that the two conditions (1.1) and (1.2) are the same only in dimension 2.

1.2 Billiard words with intercept

The same construction can be made with the half-line $\mathcal{D}$ starting from any point S and parallel to the positive vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$. For simplicity reasons, we choose S in the subspace $\mathcal{D}^{\perp}$, and by periodicity we can assume that S is in the orthogonal projection $\mathcal{P}$ of the unit k-cube centered at the origin, onto $\mathcal{D}^{\perp}$. As before, we must assume that:

$$\mathcal{D} \text{ does not contain any point with two integer coordinates.} \tag{1.3}$$

This condition depends on both S and $(\alpha_1, \alpha_2, \ldots, \alpha_k)$. This billiard word will be denoted by $c_{\alpha_1,\alpha_2,\ldots,\alpha_k,S}$.

2 Already known results on factors of billiard words

Billiard words and Sturmian words have been intensively studied, see [1], [4] or [5] for general exposures, and many results are known, concerning the language of these words, i.e., the set of all finite factors. The well-known notion of *Rauzy's diagram* gives very nice results. With Hypothesis (1.2) of total irrationality, the complexity function is known, [2], [3], and so for the palindromic complexity function: the complexity function $p_u(n)$ (resp. palindromic complexity function $pal_u(n)$) of an infinite word u is the number of distinct factors (resp. palindromic factors) v of length n of u. It is also possible to look at the first occurence of each palindromic factor, to characterize Sturmian words in dimension 2, [11].

3 Recent results on palindromic factors and prefixes

Theorem 3.1 *With Hypothesis (1.2) of total irrationality, the billiard word* $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ *has:*

- *when n is even, a unique palindromic factor of length n. The center of this palindromic factor is the unique pair of letters aa which belongs to the language of $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$;*
- *when n is odd, and for each letter a of the alphabet $\mathcal{A}$, a unique palindromic factor of length n in which the letter a is in central position.*

This result implies that for two different palindromic factors of odd length of the billiard word with the same central letter, the shortest one is a central factor of the longest. The same result is true for two distinct palindromic factors of even length.

Theorem 3.2 *With Hypothesis (1.2) of total irrationality,*

- *in dimension $k = 2$, the billiard words have infinitely many palindromic prefixes; these factors are related to the continued fraction expansion of the slope ρ of $\mathcal{D}$;*
- *in dimension $k \geq 3$, the set of vectors $(\alpha_1, \alpha_2, \dots, \alpha_k)$ such that the billiard word $c_{\alpha_1,\alpha_2,\dots,\alpha_k}$ has infinitely many palindromic prefix factors is a negligible set, in the sense of the Lebesgue measure on the k-dimensional unit sphere. However, this set is dense on the positive part of this unit sphere.*

These results, in dimension 2, have been stated in some slightly different formulations in [7], [8], [9], [10]. In higher dimension, it has been proved in [6]. The main problem is to *synchronize* the denominators of the convergents of the continued fraction expansion of the ratios $\frac{\alpha_i}{\alpha_j}$.

4 A geometrical characterization of factors of billiard words

4.1 b-walks

We consider the $(k-1)$-dimensional subspace $\mathcal{D}^\perp$ in $\mathbb{R}^k$, and the orthogonal projection $\mathcal{P}$ of the unit k-cube centered at the origin, onto $\mathcal{D}^\perp$. We denote by b_j, $1 \leq j \leq k$, the orthogonal projections onto $\mathcal{D}^\perp$ of the vectors e_j of the canonical basis of $\mathbb{R}^k$.

Definition 4.1 Let H be any point in $\mathcal{P}$. A b-walk in $\mathcal{P}$ starting from H is a finite sequence of points $H_0, H_1, H_2, \dots H_n$ in $\mathcal{P}$, such that $H_0 = H$, and such that there exists, for any $1 \leq i \leq n$, some $j = j(i)$ such that $\overrightarrow{H_{i-1}H_i} = b_j$. The integer n is called the length of the walk.

Such a walk is characterized by its starting point H and its *coding word*, i.e., the finite word of length n on $\mathcal{A}$ obtained by encoding each vector b_j by the letter a_j.

In Figure 2, corresponding to $k = 3$, $\mathcal{P}$ is an hexagon, and the b-walk starting from H and of length 11 is encoded by $a_3a_1a_3a_3a_2a_1a_3a_3a_1a_3a_2$.

4.2 b-walks and factors of billiard words

Theorem 4.2

- *Except for some points H, there exists a unique b-walk starting from H with a given length n.*
- *A finite word on $\mathcal{A}$ is a factor of the billiard word $c_{\alpha_1,\alpha_2,\dots,\alpha_k}$ only if it encodes some b-walk. This condition is also sufficient with Hypothesis (1.2).*

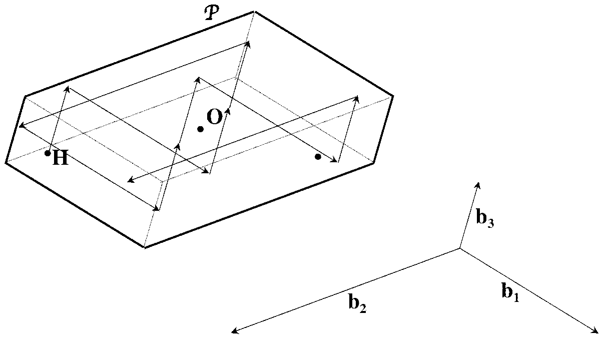


Figure 2:

The set $\mathcal{E}$ of the exceptional points H in this theorem is the set of all points H such that there exists a b-walk starting from H and ending on the boundary of $\mathcal{P}$. It is a negligible set, in the sense of the Lebesgue measure on $\mathbb{R}^{k-1}$. Using the Part 1. of the Theorem, for any $H \notin \mathcal{E}$, there exists a unique infinite b-walk starting from H, which is the limit of the finite b-walks starting from H. It can be proved that:

- with Hypothesis (1.1), $C := O + \frac{1}{2}(\sum_{j=1}^{k} b_j)$ is not in $\mathcal{E}$, and the billiard word (starting at the origin) encodes the infinite b-walk starting from C;
- Hypothesis (1.3) exactly corresponds to $S \notin \mathcal{E}$. In this case, the corresponding billiard word encodes the infinite b-walk starting from $H := O + \vec{SC}$.

In Figure 3, the two translated hexagons are the orthogonal projections $\mathcal{P}_C$ of the usual unit cube in $\mathbb{R}^3$ (all the three coordinates between 0 and 1), whose center is C, and $\mathcal{P}_S$ of the unit cube centered at the starting point S. The projections of all the centers of the unit k-cubes crossed by the half-line $\mathcal{D}$ starting from S are in $\mathcal{P}_S$.

Proposition 4.3 *For any finite word u on $\mathcal{A}$, the set $\mathcal{P}_u$ of the starting points H of all the b-walks encoded by u is a convex polyhedron (except for the points in $\mathcal{E}$), whose diameter tends to* 0 *as the length $|u|$ tends to infinity.*

4.3 An application for palindromic factors

We consider now the closure $\mathcal{H}$ of the set of the orthogonal projections on $\mathcal{D}^{\perp}$ of the centers of the k-cubes crossed by $\mathcal{D}$. Hence $\mathcal{H}$ is a subset of $\mathcal{P}$, and is equal

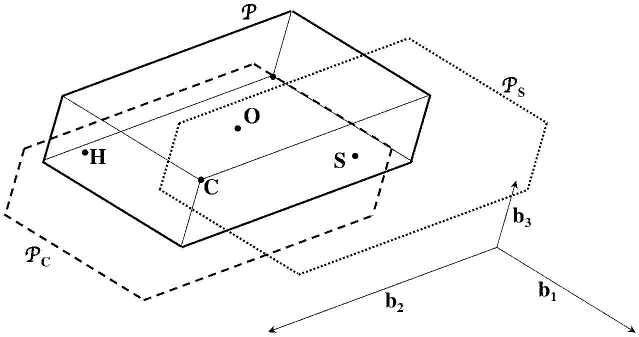


Figure 3:

to $\mathcal{P}$ with Hypothesis (1.2) of total irrationality. More generally, we have:

Theorem 4.4

- *The billiard word $c_{\alpha_1,\alpha_2,\ldots,\alpha_k,S}$ contains arbitrarily long palindromic factors of even length if and only if the origin is in $\mathcal{H}$.*

- *The billiard word $c_{\alpha_1,\alpha_2,\ldots,\alpha_k,S}$ contains arbitrarily long palindromic factors of odd length and central letter a_j if and only if $\frac{1}{2}b_j$ is in $\mathcal{H}$.*

4.4 An example in dimension 3

As an example in dimension $k = 3$, consider the vector $(2, \sqrt{5}, 1 + \sqrt{5})$, which satisfies (1.1), but not the Hypothesis (1.2) of total irrationality : $\alpha_1 + 2\alpha_2 - 2\alpha_3 = 0$. Then $\mathcal{P}$ is the following hexagon:

In Figure 4 are given four of the sets $\mathcal{P}_u$, which are parallelograms in this case (triangles or hexagons may also appear, for some other u), corresponding to some palindromic factors u of length 2 and 7, which may appear in the corresponding billiard words. We consider the billiard word starting at the origin. Then $\mathcal{H}$ is the union of the four parallel thick segments. The factors u which appear in this word, are those such that $\mathcal{H}$ intersects $\mathcal{P}_u$ in its interior. The set $\mathcal{H}$ does not intersects $\mathcal{P}_{a_3a_3}$ in its interior. Hence the word a_3a_3 is not a factor of the billiard word. By this way, we obtain that the palindromic factors of length less or equal to 7 of the billiard word are:

$$a_1 \; ; \; a_2 \; ; \; a_3$$

$$a_3a_1a_3 \; ; \; a_3a_2a_3 \; ; \; a_2a_3a_2$$

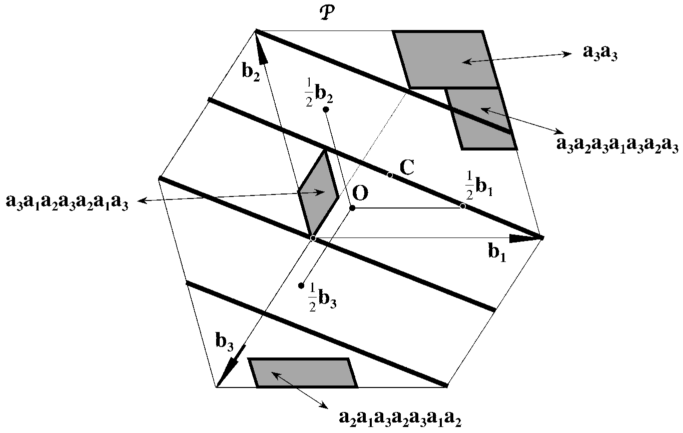


Figure 4:

$$a_2a_3a_1a_3a_2 \; ; \; a_1a_3a_2a_3a_1$$

$$a_3a_2a_3a_1a_3a_2a_3.$$

In this case, we only have $\frac{1}{2}b_1 \in \mathcal{H}$, hence there only exist arbitrarily long palindromic factors with central letter a_1. The factor a_3a_3 does not appear, so that there exists no palindromic factor of even length. More precisely, we prove in the general case:

Proposition 4.5

- *In dimension* 3*, the billiard words starting from the origin always have arbitrarily long palindromic factors. However, for almost all* S*, the billiard word starting from* S *has a finite number of palindromic factors.*
- *In dimension* $k \geq 4$*, the billiard word starting from the origin may have a finite number of palindromic factrors.*

4.5 Complexity of non-usual billiard words

With the usual Hypothesis (1.2) of total irrationality, the complexity and palindromic complexity functions are already known. They depend only on the dimension k.

$$p_2(n) = n + 1$$

$$p_3(n) = n^2 + n + 1$$

$$p_k(n) = \sum_{i=0}^{min(k-1,n)} i! \begin{pmatrix} k-1 \\ i \end{pmatrix} \begin{pmatrix} n \\ i \end{pmatrix}$$

$$pal_k(n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ k & \text{if } n \text{ is odd} \end{cases}$$

These results come from the original works on Sturmian words in dimension $k = 2$, from [2] (dimension $k = 3$), and [3] (for any k); the palindromic complexity can be considered as a classical unwritten result.

Consider now the *non-usual case*: Hypothesis (1.1) is true, but there exist some linear relations over $\mathbb{Q}$ on the coefficients α_j. Let $r \geq 1$ be the maximal number of such relations, linearly independent over $\mathbb{Q}$. This means that the linear space generated by the α_j's over $\mathbb{Q}$ is of dimension $k - r$. Thus Hypothesis (1.1) implies $r \leq k - 2$.

Then we have:

Theorem 4.6 *For each billiard word, the set $\mathcal{H}$ is the union of a finite number of parallel $(k - r - 1)$-dimensional polyhedrons.*

Corollary 4.7 *Consider two billiard words $c_{\alpha_1,\alpha_2,\ldots,\alpha_k,S_1}$ and $c_{\alpha_1,\alpha_2,\ldots,\alpha_k,S_2}$ corresponding to a given positive vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$. Then:*

- *either these two words have the same language,*
- *or the intersection set of their languages is finite.*

The second case is the most usual one.

The first case appears for some special values of the vector $\overrightarrow{S_1S_2}$.

Corollary 4.8 (CONJECTURE) *For each billiard word $c_{\alpha_1,\alpha_2,\ldots,\alpha_k,S}$, the complexity function $p(n)$grows like n^{k-r-1}, and is* **under some strong technical hypothesis** *a polynomial function of n whose degree is $k - r - 1$, for n sufficiently large.*

This result can be proved in some special cases, for example when $k = 3$ or 4, or when $r = k - 2$.

In dimension $k = 3$, the only possibility is $r = 1$, and the complexity function is $p(n) = cn + c + 1$ for sufficiently large n, with $c = |c_1| + |c_2| + |c_3| - d$, where the c_j are the integer coefficients of the unique linear relation $c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3 = 0$ over $\mathbb{Z}$ with coprime coefficients, and $d = 2$ or 1 depends on S. The set $\mathcal{H}$ is the union of $c + 1$ parallel segments. Two examples corresponding to billiard words starting at the origin:

- in the case of Figure 4, $d = 2$ and $c = 3$, and the complexity functions are given by:

n	:	1	2	3	4	5	$n \geq 6$
$p(n)$	:	3	6	9	14	18	$3n + 4$
$pal(n)$	:	3	0	3	0	2	$n \mod 2$

  where $n \mod 2$ is equal to 0 or 1.

- when we have $\alpha_1 + \alpha_2 - \alpha_3 = 0$, the complexity functions are $p(n) = n + 2$ and $pal(n) = 3(n \mod 2)$ for any $n \geq 1$.

References

[1] J.-P. Allouche, J. Shallit, *Automatic sequences: Theory and Applications*, Cambridge University Press, Cambridge, 2003.

[2] P. Arnoux, C. Mauduit, I. Shiokawa, J. I. Tamura, Complexity of sequences defined by billiard in the cube, *Bull. Soc. Math. France*, **122**, 1994, 1-12.

[3] Y. Baryshnikov, Complexity of trajectories in rectangular billiards, *Comm. Math. Phys.*, **174**, 1995, 43-56.

[4] J. Berstel, A. de Luca, Sturmian words, Lyndon words and trees, *Theoret. Comput. Sci.*, **178**, 1997, 171-203.

[5] J. Berstel, P. Sbold, Sturmian words, in M. Lothaire, *Algebraic combinatorics on words*, Cambridge University Press, 2002.

[6] J.-P. Borel, C. Reutenauer, Palindromic factors of billiard words, *Theoret. Comput. Sci.*, **340-2**, 2005, pp. 334-348.

[7] A. de Luca, Sturmian words: structure, combinatorics, and their arithmetics, *Theoret. Comput. Sci.*, **183**, 1997, 45-82.

[8] A. de Luca, Combinatorics of standard Sturmian words, *Structures in Logic and Computer Science, Lecture Notes Comput. Sci.*, **1261**, 1997, 249-267.

[9] X. Droubay, Palindromes in the Fibonacci word, *Inf. Proc. Letters*, **55**, 1995, 217-221.

[10] X. Droubay, G. Pirillo, Palindromes and Sturmian words, *Theoret. Comput. Sci.*, **223**, 1999, 73-85.

[11] G. Pirillo, A new characteristic property of the palindrome prefixes of a standard Sturmian word, *Sm. Lotharingien Combinatoire*, **43**, 1999, (electronic, see http://www.mat.univie.ac.at/∼slc/).

Efficient Word Recognition of Certain Locally Defined Trace Languages[*]

Luca Breveglieri, Stefano Crespi Reghizzi, Alessandra Savelli[†]

Abstract

We present a new cubic time algorithm for solving the word membership problem of certain rational trace languages, namely those defined by local automata whose state graphs are well-nested cycles. For this family our result improves on the known algorithms for rational trace languages, which have a polynomial time bound with a non-fixed exponent, determined by the independence relation. This bound is exceedingly large for practical application such as program parallelization, which motivated this research.

1 Introduction

This work is a first step in the direction of finding more efficient word recognition algorithms for trace languages [3]. As our interest for trace languages comes from their capacity to model dependence relations in computer programs, in order to perform program optimization, we have narrowed our attention to trace languages defined by finite automata of the local type. This because, for the purpose of code parallelization, a program can be conveniently schematized by a local machine such that each instruction is identified by a distinct letter of the partially commutative alphabet. As program loops are the most rewarding regions for parallelization by optimizing compilers, we have focused attention on automata consisting of nested cycles, with the intention to consider more general situations in the future.

The trace language associated to such local automaton represents all valid permutations (i.e. schedules) of the possible runs of the program. Therefore solving the membership problem is an essential step for further work on program scheduling.

The membership problems for trace languages defined by rational (i.e. regular) and context-free languages have been studied in [6], [2] and [1], where they are shown to be solvable in polynomial time. The combinatorial interest of the problem is related to the number of prefixes in traces of given length.

In [5] the best algorithm based on the analysis of prefixes is described. Its worst-case time complexity, although polynomial, is not of fixed degree, but it

[*] Work partially supported by MIUR Project "Linguaggi formali" and by CNR-IEIIT.

[†] Politecnico di Milano, Dipartimento di Elettronica e Informazione, P.za L. da Vinci 32, I-20133 Milano, {brevegli, crespi, savelli}@elet.polimi.it

grows with the size of the cliques of the independence relation (the complement of the dependence relation). A quick analysis of the independence relation of some ordinary programs consisting of some hundredths machine instructions shows that the exponent may be of the same order of magnitude, that is unacceptable for any practical purpose.

By restricting the problem to local automata and by further limiting the topology of the graph to nested cycles, we have been able to significantly reduce the worst case time-performance. Now to test a string for validity is equivalent to computing the number of iterations of the cycles of the automaton, and to check whether iteration counts match. This strategy is very different from prefix analysis of previous algorithms.

The paper proceeds as follows. In Sect. 2 we introduce the basic definitions. In Sect. 3 we state and justify the conditions, based on the the occurrence of factors and substrings, for a string to be in the trace language. In Sect. 4 the recognition algorithm is presented, its correctness is argued by exploiting the above-mentioned conditions, and its time complexity is analyzed.

2 Basic definitions

We list the basic definitions, mostly taken from [3]. Let Σ be an alphabet and I be an arbitrary *independence relation* on Σ which is symmetric and irreflexive. The complementary relation D of I, the *dependence relation*, is symmetric and reflexive. If $(a, b) \in D$ (or $(a, b) \in I$), we say that a and b are dependent (or independent) and we write aDb (or aIb).

Let $\sim_I$ be the smallest congruence on Σ^* such that $ab \sim_I ba$ for all pairs of independent letters a, b. The *trace monoid* $\mathbf{M}(\Sigma, I)$ defined by $\langle \Sigma, I \rangle$ is the quotient $\Sigma^* / \sim_I$. A *trace* $[x]$ is an element of $\mathbf{M}(\Sigma, I)$ represented by the string x, and a *trace language* T is a subset of $\mathbf{M}(\Sigma, I)$. For every $L \subseteq \Sigma^*$, the trace language defined by L is $[L] = \{t \in \mathbf{M}(\Sigma, I) \mid t = [x] \text{ for some } x \in L\}$.

In this work, we only consider trace languages defined by regular languages.

A *local automaton* $\mathcal{A} = (\Sigma, Q, \delta, q_0, F)$ is such that for every $q, q' \in Q$ and $a, a' \in \Sigma$, $\delta(q, a) = \delta(q', a')$ if and only if $a = a'$. Therefore all the arcs of $\mathcal{A}$ labelled by the same letter enter the same state, which no other arc enters. For simplicity, we will name the state with the label of incoming arcs.

A *nested cycle language* is a regular language defined as:

- A letter $a \in \Sigma$ is a nested cycle language.
- If N_1 and N_2 are nested cycle languages, then the concatenation $N_1 N_2$ is a nested cycle language.
- If N is a nested cycle language, then $(N)^+$ is a nested cycle language.[1]
- Nothing else is a nested cycle language.

[1] N^+ is defined as NN^*.

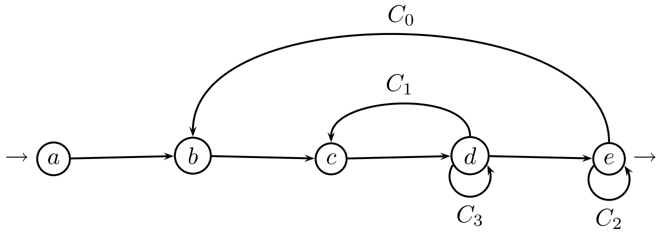


Figure 1: A nested cycle local automaton $\mathcal{A}$.

Since $+$ is idempotent we may assume that it is never applied twice consecutively.

A *nested cycle automaton* is an automaton recognizing a nested cycle language. In this work we only consider nested cycle language which in addition are accepted by local automata. An example of nested cycle and local automaton is in Figure 1, where $\Sigma = \{a, b, c, d, e\}$, $L(\mathcal{A}) = a(b(cd^+)^+e^+)^+$, and C_0, C_1, C_2, C_3 are labels given to cycles. Note that the final state is always unique.

From now on, $\mathcal{A}$ and L will always represent a nested cycle local automaton and the language recognized by $\mathcal{A}$ respectively.

If we look at trace languages as a model of programs, the letters of Σ represent the instructions of the program and two independent letters represent the independence between the instructions, that is, the possibility of commuting or parallelizing their execution. A word of a trace language is then an acceptable order of instructions execution of the program.

The locality of the automaton reflects the fact that the instructions of a program can be considered to be all distinct (i.e. each is identified by its address). By restricting the study to nested cycle automata, the corresponding programs are nested *repeat-until* instructions, without conditional instructions.

Ordering of letters, nesting of cycles, and nested iteration tree

Consider the automaton $\mathcal{A}'$ obtained from $\mathcal{A}$ by removing all back arcs (i.e. erasing all $+$ from the regular expression of L). We define a total ordering relation on Σ: *a precedes b* ($a < b$) if there exists a path in $\mathcal{A}'$ from a to b. In Figure 1, $a < b < c < d < e$. Such relation is surely antisymmetric.

Since the cycles are well nested in $\mathcal{A}$, we can define a partial order on them: we write $C \prec C'$ if C is nested in C'. If there is no cycle C'' such that $C \prec C'' \prec C'$, we say that C' is *immediately nested* in C'. In Figure 1, $C_3 \prec C_1 \prec C_0$, and $C_2 \prec C_0$.

Let u be in L. The *Nested Iteration Tree (NIT)* of u is the tree of iterations of the cycles traversed by the automaton $\mathcal{A}$ recognizing u. The transitive closure of the NIT is denoted by NIT^*.

For example, the NIT of $u = abcdcddebcdee$ is represented in Figure 2. We say that u iterates cycle C_0 twice, and that $C_0(2)$ (i.e. the second iteration of C_0) iterates C_1 once and C_2 twice, etc..

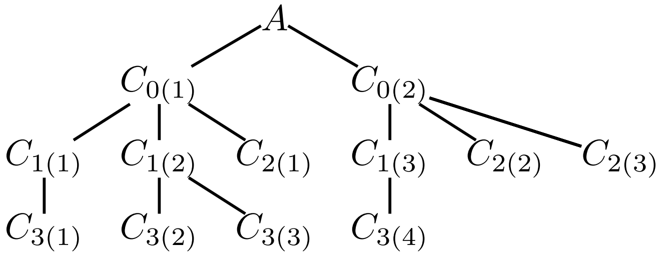


Figure 2: The NIT of *abcdcddebcdee*.

3 Main Theorem

Let $B \subseteq \Sigma$. We denote by $C(B)$ the *innermost cycle* of $\mathcal{A}$ containing all the letters in B. Note: for two sets of letters, B, B', if $B \subseteq B'$, then it holds $C(B) \prec C(B')$. Let $w \in \Sigma^*$ and $B \subseteq \Sigma$. The *projection of w onto B* is denoted by $\pi_B(w)$. Let $b \in \Sigma$. A *run* of b in a word w is a maximal factor of b, that is, a sequence of b not contained in a longer sequence of b.

Lemma 3.1 *Let $w, w' \in \Sigma^*$. Then $w \sim_I w'$ if, and only if, $\pi_{a,b}(w) = \pi_{a,b}(w')$ for each pair of letters $a, b \in \Sigma$ (possibly identical) such that aDb.*

Proof By contradiction, suppose that $w \nsim_I w'$. Since $\pi_a(w) = \pi_a(w')$ for every $a \in \Sigma$, then w and w' have the same number of occurrences of letters. Thus, w' can be obtained by permutating the letters of w, necessarily including permutations between dependent letters, that is, there are $a_1, a_2 \in \Sigma$ such that $a_1 D a_2$ and $\pi_{a_1,a_2}(w) \neq \pi_{a_1,a_2}(w')$, a contradiction. The second implication is obvious. □

Lemma 3.2 *Let $w \in L$, $|w| = n$, and $a \in \Sigma$. Then the number of a in w is equal to the number of times that w iterates $C(a)$.*
Let $a_1, a_2 \in \Sigma$ such that $a_1 < a_2$. Then, the number of factors a_1a_2 in $\pi_{a_1,a_2}(w)$ is equal to the number of times that w iterates $C(a_1, a_2)$.
Moreover, the length of the i-th $(1 \leq i \leq n)$ run of a_1 (resp. a_2) in $\pi_{a_1,a_2}(w)$ is equal to the number of iterations of $C(a_1)$ (resp. $C(a_2)$) inside the i-th iteration of $C(a_1, a_2)$.

Proof Let $p = \pi_{a_1,a_2}(w)$. Since $C(a_1, a_2)$ is the innermost cycle containing both a_1 and a_2, a walk from a_1 to a_2 in the automaton (that is, each factor a_1a_2 in p) represents an iteration of $C(a_1, a_2)$. Thus, the i-th run of a_1 (or a_2) in p refers to the i-th iteration of C. Obviously, each occurrence of a_1 (resp. a_2) is an iteration of $C(a_1)$ (resp. $C(a_2)$), so that, in conclusion, the length of the i-th run of a_1 (a_2) in p is the number of iterations of $C(a_1)$ ($C(a_2)$) in the i-th iteration of $C(a_1, a_2)$. □

By counting the above features of w, we obtain partial information on the NIT of w, namely the number of iterations of some cycles and the existence of some paths in the tree. For instance, the fact that a cycle C' is iterated m times in $C(1)$ (first iteration of a cycle C), implies that there exist paths from node $C(1)$ to the m nodes representing the first m iterations of C'. Such a path is represented by an edge in the NIT^* (transitive closure of NIT), called *hierarchical edge*.

Moreover, the hierarchical edges from the iterations of C to those of C' determine a partition of the iterations of C': all the iterations of C' connected with the same iteration of C are a class.

We will use the next simple fact.

Lemma 3.3 *Let $u, v \in \{a, b\}^*$. Then $u = v$ if, and only if, they have the same initial letter and the sequences of lengthes of runs of a (and b) in u and v coincide.*

Theorem 3.4 *Let Σ be an alphabet, D be an arbitrary dependence relation, L be a language recognized by a nested cycle local automaton, and $w \in \Sigma^*$. Then $w \in T = [L]$ if, and only if:*

1. *for every pair of distinct $a_1, a_2 \in \Sigma$ such that $a_1 D a_2$ and $a_1 < a_2$, it holds $\pi_{a_1,a_2}(w) \in a_1\{a_1, a_2\}^* a_2$, and*
   there exists $w' \in L$ such that
   - *(a) for every $a \in \Sigma$, w' iterates $C(a)$ $\pi_a(w)$ times;*
   - *(b) the number of iterations of $C(a_1, a_2)$ of w' is equal to the number of factors $a_1 a_2$ in $\pi_{a_1,a_2}(w)$;*
   - *(c) the number of iterations of $C(a_1)$ of w' (resp. $C(a_2)$) in the i-th iteration of $C(a_1, a_2)$ is equal to the length of the i-th run of a_1 (resp. a_2) in $\pi_{a_1,a_2}(w)$.*

Proof If $w \in T$, then there exists $w' \in L$ such that $w \sim_I w'$ and, by Lemmas 3.1 and 3.2 we have the first implication. As to the other implication, by Lemmas 3.2, 3.3, and 3.1 it follows that $w \sim_I w'$. □

4 Recognition algorithm and example

For a word $u \in L$ a *Nested Iteration Graph (NIG)* of u is a subgraph of the NIT^* of u. By Theorem 3.4, if a word w belongs to the trace language $[L]$, then we can construct a graph G_{in} which is a NIG of a word $w' \in L$ such that $w \sim_I w'$, in the following way. The nodes representing the iterations of a cycle C are in the level of G_{in} that is the level of nesting of C in the starting automaton. For every $a \in \Sigma$, there are $\pi_a(w)$ nodes representing the iterations of $C(a)$ in G_{in}. For every $a_1, a_2 \in \Sigma$ such that $a_1 D a_2$, the number of nodes representing the iterations of $C(a_1, a_2)$ in the NIG is equal to the number of factors $a_1 a_2$ in $\pi_{a_1,a_2}(w)$. Moreover, there are hierarchical edges between the nodes representing the iterations of $C(a_1, a_2)$ and those representing the iterations of $C(a_1)$ (and $C(a_2)$, respectively) depending on the lengths of runs of a_1 (resp. a_2) in $\pi_{a_1,a_2}(w)$.

We have that $[w]$ belongs to the trace language $[L]$ if, and only if,

- every projection of w on a pair of dependent letters satisfies the first condition of Theorem 3.4, and
- the counters computed at points (a), (b), and (c), permit to construct a graph G_{in}, which is a NIG of a word $w' \in L$ such that $w \sim_I w'$. (Indeed, in general we may obtain contrasting information about the number of iterations of a cycle or about the hierarchical edges connecting two levels in G_{in}.)

Next we illustrate the construction of such a graph G_{in} for the running example. Consider the input string and dependence relation

$$w = cdbadcdeedceebe \qquad D = \{(c,e),(d,e),(a,a),(b,b),(c,c),(d,d),(e,e)\}$$

The values of the counters determine the features of the graph listed in Table 1.

Each edge from a node to an encircled set of nodes represents a set of hierarchical edges.

Sometimes the graph G_{in} thus constructed does not contain all the nodes corresponding to the cycles of the starting automaton. This means that no information is available about the number of iterations of some cycle C. This happens, following Theorem 3.4, if C is such that:

Counter values	*Features of the graph G_{in}*
$C(c) = C_1$ and $\pi_c(w) = ccc$	if $w \in [L]$, then every $w' \in L$ such that $w' \sim_I w$ iterates C_1 three times.
$C(c,e) = C_0$, $C(c) = C_1$, $C(e) = C_2$ and $\pi_{c,e}(w) = cceeceee$	w' iterates C_0 twice (see the factors ce in $\pi_{c,e}(w)$); w' iterates C_1 twice in the first iteration of C_0 and once in the second iteration (see the runs of c in $\pi_{c,e}(w)$); C_2 is iterated twice in the first iteration of C_0 and three times in the second iteration (see the runs of e).
$C(d,e) = C_0$, $C(d) = C_3$, $C(e) = C_2$ and $\pi_{d,e}(w) = dddeedeee$	w' iterates C_0 twice (see the factors de in $\pi_{d,e}(w)$); w' iterates C_3 three times in the first iteration of C_0 and once in the second iteration (see the runs of d in $\pi_{d,e}(w)$); C_2 is iterated twice in the first iteration of C_0 and three times in the second iteration.

Table 1: Counter values and features of the resulting graph G_{in} (depicted in Figure 3).

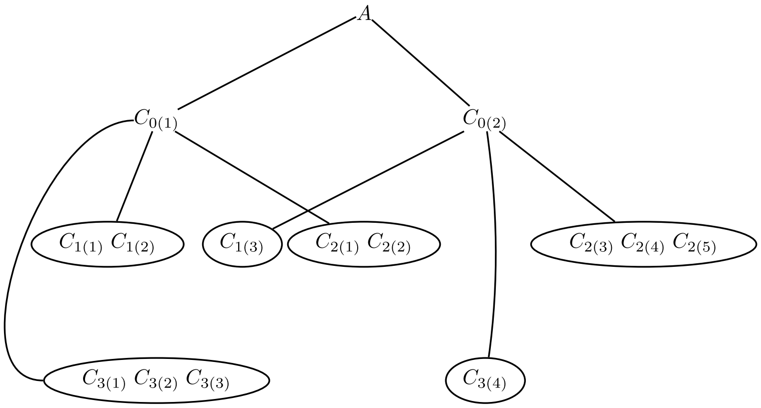


Figure 3: The graph G_{in} of *cdbadcdeedceebe*.

- there exists no letter a such that $C(a) = C$, and
- for no pair of distinct dependent letters a, b, it is $C(a, b) = C$.

To illustrate, imagine to add a self loop to state c in the automaton of Figure 1. Then Theorem 1 would provide no information about the iterations of C_1.

When this happens, in order to construct the graph G_{in} we need the following result, which states that we can assume the undetermined number of iterations of cycle C to be equal to that of its father.

Proposition 4.1 *Let C be a cycle of a nested cycle local automaton such that for every pair a, b of letters belonging to C, $C(a) \not\supseteq C$ and if aDb, $C(a, b) \not\supseteq C$. Let C be immediately nested in C'. Consider a word $u \in L$ such that u iterates C' h times and C k_i times in the i-th iteration of C', $i = 1, \ldots, h$.*

Then there exists a word $v \in L$ such that: $u \sim_I v$, v iterates C' h times, and v iterates C once in every iteration of C'.

Proof (Hints of the complete proof) Since for no letter $a \in \Sigma$ it is $C(a) = C$, then C can be viewed as a sequence $C_1C_2\cdots C_m$ of inner cycles. Since for every pair of distinct dependent letters a, b belonging to C one has $C(a, b) \neq C$, then no inner cycle depends on another one, so that, if we interchange two of these cycles, we obtain an equivalent word. Then we can obtain v from u by a series of interchanges, grouping all iterations of C_1, all iterations of C_2, and so on, inside every sequence of iterations of C. In conclusion, such a v iterates C only once in every iteration of C' but it iterates more times the inner cycles $C_1C_2\cdots C_m$. □

This transformation is applied in the algorithm, to construct G_{in}. We add to G_{in} the nodes representing the iterations of the undetermined cycles (as many as the number of their parents by Proposition 4.1), thus obtaining another NIG

graph, named G_0, which contains all and only the nodes of the NIT to be constructed (if it exists).

The next and last phase of the algorithm computes the missing father-son edges in the graph G_0 producing a graph named G. From Theorem 1 and Proposition 4.1 we draw the following conclusion: $[w] \in T$ if, and only if, G_0 can be "completed" obtaining a graph G with all the father-son edges. Such a completion must of course be consistent with the hierarchical paths of G_0; this implies that, for every edge (n_1, n_2) in G_0, there exists a path in G from n_1 to n_2, made by a chain of father-son edges. Therefore G contains as subgraph the NIT of a word $w' \in L$, $w' \sim_I w$.

Completion algorithm

The algorithm is greedy. At each step we consider the last level l of G_0 missing some father-son connection to level $l-1$ (though G_0 is not a tree, it already has its nodes organized as the final NIT, so that the concept of level is well defined). Then we connect, if possible, each iteration node of level l to a father node.

After a connection has been established, the father inherits the hierarchical paths of its sons. This means that, if C has been connected to the father C' and there is a hierarchical path from C to an ancestor C'', then this path must traverse C'. Consequently the hierarchical edge from C' to C'' is added to the graph.

Clearly the algorithm terminates since the number of cycles in the automaton is finite.

From previous statements we know the algorithm ends with a graph G containing a NIT if, and only if, the given word belongs to the trace language.

Partitioning step Next we specify the step for connecting a level to the one above. This step bears resemblance to the Earley context-free parsing algorithm [4].

We remember that the hierarchical edge in G_0 connecting the iterations of a cycle C'' to those of an inner cycle C determines a partition of the iterations of C.

Let $C \prec C'$ be immediately nested cycles of the automaton at respective levels l and $l-1$, whose iterations are still unconnected in the current graph.

Consider in the graph the finest partitions $P_{C'}$ and P_C of these iterations (indeed there can be more partitions corresponding to hierarchical edges connecting the iterations of a cycle to more than one level upwards). If there are no partitions, then take the trivial partition, i.e. the partition having only one class. The trivial partition can in fact be viewed as a hierarchical edge to the root.

Let h and k be the number of classes of $P_{C'}$ and P_C respectively. In order to find a good connection between the iterations of C' and those of C, we have to find a partition $P'_{C'}$ of the iterations of C' into k classes such that

- $P'_{C'}$ is coarser or finer than $P_{C'}$, according to the fact that $h \geq k$ or $h < k$, and
- the cardinality of every class of $P'_{C'}$ must not exceed that of the corresponding class of P_C.

This inequality follows from the remark that we want to connect the iterations of each class of $P'_{C'}$ to those of the corresponding class of P_C and from $C \prec C'$, so that in every iteration of C' there must be one iteration of C at least.

Of the two cases of the former condition, we only discuss the case $h \geq k$ (the other case being similar).

A coarser partition $P'_{C'}$ consisting of k classes can be represented by an array of k cells, where every cell is a sequence of classes of $P_{C'}$. In order to find a coarser partition $P'_{C'}$, we apply the following procedure, which first constructs an array S with k cells, which are sets of sequences of classes of $P_{C'}$.

We compute S from left to right, starting by the empty cells. We put into each cell the classes that can be associated to the corresponding class of P_C. The last class of the sequence written in a cell S_i determines the first class of a sequence in the next cell S_{i+1}. We proceed in this way until the last cell is assigned.

```
procedure   construction of the array $S$ of sets of class sequences
Initial ← $\{1\}$
NextInitial ← $\emptyset$
for $i := 1$ to $k$ do
    for each $ind$ in Initial do
        for $j := ind$ to indLastClass($ind$, size($P_C(i)$)) do
            $S_i := S_i \cup \langle P_{C'}(ind) P_{C'}(ind+1) \cdots P_{C'}(ind+j) \rangle$
            NextInitial := NextInitial $\cup \{ind + j + 1\}$
        end for
    end for
    Initial ← NextInitial
end for
```

Notes:
size($P_C(i)$) represents the cardinality of the i-th class of P_C.
indLastClass(ind, size($P_C(i)$)) is the last class of a sequence starting with $P_{C'}(ind)$ we can put into S_i, that is, $P_{C'}(ind + size(P_C(i)) - 1)$. Indeed, the i-th cell of S represents the possible fathers of $P_C(i)$, so that the cardinality of a sequence we can put into S_i cannot be greater than size($P_C(i)$).

Now a good connection between the iterations of C' and C is possible if in the last cell the last class of a sequence coincides with the last class of C'.

Having obtained the coarser partition $P'_{C'}$, each node, but the last one, of each class $P'_{C'}(i)$ of $P'_{C'}$ is connected to a single node of $P_C(i)$. The last node

S_1	S_2
$\langle P_{C_1}(1)\rangle$ $\langle P_{C_1}(1)P_{C_1}(2)\rangle$	$\langle P_{C_1}(2)\rangle$

Figure 4: Array S.

is connected to all the remaining nodes of $P_C(i)$ (by construction of $P'_{C'}$ the cardinality of each $P'_{C'}(i)$ is not greater than that of $P_C(i)$).

Example

To illustrate, consider the graph G_0 of Figure 3. The iterations of C_3 have to be connected to those of C_1, the immediately enclosing cycle. The finest partition of C_1 is $P_{C_1} = \{\{C_{1(1)}, C_{1(2)}\}, \{C_{1(3)}\}\}$, while that of C_3 is $P_{C_3} = \{\{C_{3(1)}, C_{3(2)}, C_{3(3)}\}, \{C_{3(4)}\}\}$. Note that this connection case is very simple, as we have the same number of classes for C_1 and C_3, and in fact the missing paths can easily be obtained transitively since there are hierarchical edges from C_0 to both C_1 and C_3. However, we can use this example to illustrate the run of the completion algorithm. We have to find a partition of C_1 in two classes (coarser or finer than P_{C_1}, but we already have a two class partition in this case) such that the former class has cardinality ≤ 3 and the latter class has cardinality ≤ 1 (3 and 1 are the cardinalities of the classes of P_{C_3}).

Array S has a number of cells equal to the number of classes of P_{C_3}, that is 2. We begin assigning sequences of P_{C_1} classes to its cells. Since the first class of P_{C_3} has 3 elements, we can put into S_1 two sequences: $\langle P_{C_1}(1)\rangle$, and $\langle P_{C_1}(1)P_{C_1}(2)\rangle$. Now there is only one possible initial index of P_{C_1} classes for the second cell, that is 2. (From the second sequence in S_1 we cannot get an index, since there are only two classes in P_{C_1}. Then we put into S_2 the sequence $\langle P_{C_1}(2)\rangle$, thus completing array S:

In order to reconstruct a partition P'_f, we examine the last cell of S, S_2, looking for a sequence ending by the last class of P_{C_1}, which is $\langle P_{C_1}(2)\rangle$. This is chosen as the last father's class of P'_{C_1}. Next we visit the preceding cell, looking for a sequence ending by the class of P_{C_1} which is a predecessor of the first class of the sequence already chosen. In this case, the sequence is $\langle P_{C_1}(1)\rangle$.

Here we stop. In general we continue until the first class of P'_{C_1} is computed. The new connections are added to the graph and the iterations of C_1 inherit from their sons the hierarchical edges to the iterations of C_0, drawn as dotted arcs in Figure 5 (in this case, the information provided by the inherited hierarchical edges was already partially known, which is not always the case).

Let n be the length of the word to be tested for membership. Clearly both phase 1 and 2 take time $O(n)$. Phase 3 performs steps 3(a) and 3(b) a constant number of times, equal to the maximal nesting depth of the automaton. Step 3(b) takes time $O(n)$. It remains to analyze step 3(a), which turns out to be the dominating step. The code contains three nested for loops. The number of

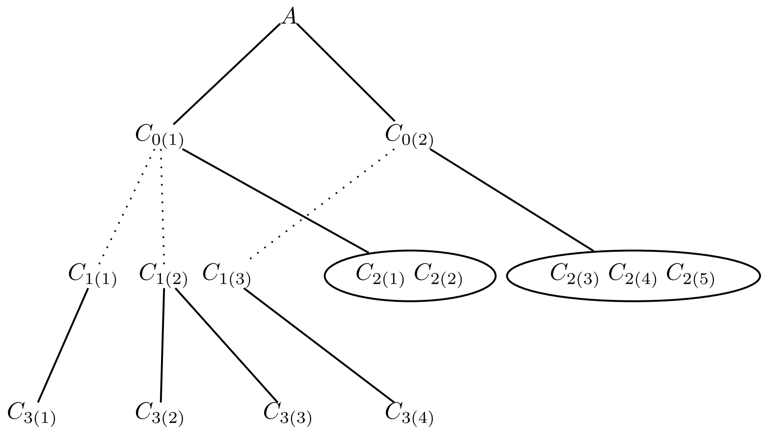


Figure 5: Adding new edges.

iterations of each of them is bounded by the number h of classes of $P_{C'}$, which cannot exceed n, since every cycle contains at least one letter. In conclusion we have:

Proposition 4.2 *The worst case complexity of the word membership problem for a trace language defined by a nested cycle and local automaton is $O(n^3)$.*

Notice that the exponent is independent from any property of the dependence relation, unlike the algorithm of [5].

5 Conclusion

At present it is unclear whether the lower time complexity of our algorithm, with respect to the general approach of [5], is due to the hypothesis of locality, to that of nested cycle restricted topology, or to the different approach based on detecting cycle iterations, instead of prefix analysis. In the future we intend to investigate the possibility of extending the algorithm to the general case of local automata.

Acknowledgements: to Jean Berstel, Christian Choffrut, and Massimiliano Goldwurm for valuable comments.

References

[1] G. Mauri A. Bertoni and N. Sabadini. Membership problems for regular and context free trace languages. *Information and Computation*, 82:135–150, 1989.

[2] M. Clerbout and M. Latteux. Partial commutations and faithful rational transductions. *Theoretical Computer Science*, 34(3):241–254, 1984.

[3] V. Diekert and G. Rozenberg. *The Book of Traces.* World Scientific, Singapore, 1995.

[4] Jay Earley. An efficient context-free parsing algorithm. *Commun. ACM*, 13(2):94–102, 1970.

[5] M. Goldwurm and A. Avellone. Analysis of algorithms for the recognition of rational and context-free trace languages. *RAIRO Theoretical Informatics and Applications*, 32:141–152, 1998.

[6] W. Rytter. Some properties of trace languages. *Fundamenta Informaticae*, VII:117–127, 1984.

On arithmetical complexity of Sturmian words[*]

Julien Cassaigne[†], *Anna E. Frid*[‡]

Abstract

Using the geometric dual technique by Berstel and Pocchiola, we give a uniform $O(n^3)$ upper bound for the arithmetical complexity of a Sturmian word. We also give explicit expressions for the arithmetical complexity of Sturmian words of slope between 1/3 and 2/3 (in particular, of the Fibonacci word). In this case, the difference between the genuine arithmetical complexity function and our upper bound is bounded and ultimately 2-periodic.

1 Introduction

Arithmetical complexity of infinite words, defined by Avgustinovich, Fon-Der-Flaass and Frid in 2000 [2], is the function $a_w(n)$ equal to the number of words of length n which occur in arithmetical subsequences of a word w: for a word $w = w_0w_1 \cdots w_n \cdots$, we by definition have

$$a_w(n+1) = \#\{w_k w_{k+d} \cdots w_{k+nd} | k \geq 0, d \geq 1\}.$$

Nowadays this function is one of the most explored modifications of the classical *subword complexity* function $f_w(n)$ defined by

$$f_w(n+1) = \#\{w_k w_{k+1} \cdots w_{k+n} | k \geq 0\};$$

for other modifications, see e.g. [7–9].

One of the first questions that naturally arose for arithmetical complexity as well as for any other modified complexity function concerns non-periodic words of minimal complexity. The first candidates for minimality were as always Sturmian words. However, Sturmian words do not even fall into the class of uniformly recurrent words whose complexity is linear. Such words have been characterized [6] and are always generated by Toeplitz transforms from a specific family. In particular, frequencies of letters in such words are rational, so they can never be Sturmian. A family of words with lowest possible arithmetical

[*] The research of the second author is supported in part by RFBR grant 03-01-00796 and by Russian Science Support Foundation.

[†] `cassaigne@iml.univ-mrs.fr`

[‡] `frid@math.nsc.ru`

complexity have been distinguished among them; unlike the subword complexity, here no unique “minimal” complexity function but a family of functions with decreasing slopes of asymptotes exist.

But then, what is the arithmetical complexity of Sturmian words? In [5], it has been proved that it grows at least as $O(n^3)$: the lower bound looks as a function of n not depending on α minus a function of α, more precisely, as $n^3/4\pi^2 + O(n^2) - O(1/\alpha^3)$, where α is the slope of the Sturmian word. Here we supplement that result by an upper bound, uniform for all Sturmian words and also equal to $O(n^3)$ (more precisely, to $(1/6 + 1/\pi^2)n^3 + O(n^2)$). So, the upper and the lower bounds differ approximately in 10.58 times, and the upper bound seems to be closer to the genuine arithmetical complexity function.

Unlike the bound, the function itself depends on the choice of the Sturmian word. For some Sturmian words, including the Fibonacci word, we find it explicitly. In the considered cases, the difference between the upper bound and the genuine arithmetical complexity is bounded.

2 Preliminaries

Sturmian words have several equivalent definitions, including the complexity one: a right infinite word s is Sturmian if and only if its subword complexity is $f_s(n) = n + 1$ for all n. (For a detailed presentation on Sturmian words, see [3].) In what follows we use the following representation of Sturmian words. Let $\alpha \in (0, 1)$ be irrational, and $\rho \in [0, 1)$ be arbitrary; the Sturmian word $s_{\alpha,\rho} = s_0 s_1 \cdots s_n \cdots$, where $s_n \in \{0, 1\}$, is defined by

$$s_n = \begin{cases} 1, & \text{if } \{(n+1)\alpha + \rho\} < \alpha, \\ 0, & \text{otherwise.} \end{cases}$$

for all $n \geq 0$. Here α is called the *slope* of $s_{\alpha,\rho}$; formally speaking, $<$ may be substituted by $\leq$ to get another Sturmian word which differs from $s_{\alpha,\rho}$ by at most two symbols. However, since we are interested in arithmetical complexity, we do not need such details. Indeed, the set of factors of a Sturmian word, and thus the set of its arithmetical factors, depend only on its slope. It what follows we denote the set of arithmetical factors of $s_{\alpha,\rho}$ by A_α and $A_\alpha \cap \{0, 1\}^n$ by $A_\alpha(n)$; the arithmetical complexity which we need to find is the number $a_\alpha(n)$ of elements of $A_\alpha(n)$.

For $\beta, \gamma \in \mathbb{R}$, let us denote by $w_\alpha(\beta, \gamma, n) = w_0 w_1 \cdots w_n$ the word of length $n + 1$ defined by

$$w_i = \begin{cases} 1, & \text{if } \{i\beta + \gamma\} < \alpha, \\ 0, & \text{otherwise} \end{cases}$$

for all $i = 0, \ldots, n$. Then clearly $w_\alpha(\{d\alpha\}, \{(k+1)\alpha + \rho\}, n) = s_k s_{k+d} \cdots s_{k+nd}$ for all $k \geq 0$, $d \geq 1$. Since α is irrational, both $\{d\alpha\}_{d=1}^\infty$ and $\{(k+1)\alpha + \rho\}_{k=0}^\infty$ constitute dense sets on $(0, 1)$ depending on independent variables k and d. We see that

$$A_\alpha(n+1) = \bigcup_{\beta,\gamma\in[0,1)} w_\alpha(\beta,\gamma,n). \tag{2.1}$$

In what follows we shall use this representation of $A_\alpha(n+1)$ to estimate and find its cardinality, i. e., $a_\alpha(n+1)$.

Note that given an irrational α, we could use the same arguments considering not only a Sturmian word of the slope α but any infinite word $w = w_0 w_1 \cdots w_n \cdots$ defined by

$$w_n = \begin{cases} 1, & \text{if } \{(n+1)\theta+\rho\} < \alpha, \\ 0, & \text{otherwise,} \end{cases}$$

where θ is irrational. (If θ is rational, the word w is periodic.) Such words were considered e. g. by Rote [10]; the arithmetical complexity of such a word does not depend on anything but α and is also equal to $a_\alpha(n)$. Thus, in fact we study the arithmetical complexity of words from a class wider than Sturmian.

3 Geometric dual method

In this section we describe the technique taken from Berstel and Pocchiola [4] and adopted to our problem. Originally, this technique was used to count the number of all finite words which are factors of Sturmian words. The exposition below in this section follows the line of [4].

Geometrically, a word $w_\alpha(\beta,\gamma,n)$ can be depicted as follows. Let us shadow all strips $k \le y < k+\alpha$ in the quadrant $x \ge 0$, $y \ge 0$. Then the $(l+1)$th symbol w_l of $w_\alpha(\beta,\gamma,n)$ is equal to 1 if and only if the line $y = \beta x + \gamma$ crosses the vertical $x = l$ in the shadowed strip (see Fig. 1). We say that the line l with equation $y = \beta x + \gamma$ *defines* the word $w_\alpha(\beta,\gamma,n)$. Let $\mathcal{L}$ be the set of lines $y = \beta x + \gamma$ with $\beta, \gamma \in [0,1)$; then each line from $\mathcal{L}$ defines one word of each length.

Let us denote by $\mathcal{P}$ the affine plane and by $\mathcal{P}_\alpha(n)$ the set of points p_{ij} with $i = 0,\ldots,n$, $j = 0,\ldots,2i+1$ defined by

$$p_{i,2k} = (i,k),\ p_{i,2k+1} = (i,k+\alpha)$$

for all $i = 0,\ldots,n$, $k = 0,\ldots,i$. Let us denote by $P_\alpha(\beta,\gamma,n)$ the sequence of points $p_{0j_0}, p_{1j_1},\ldots,p_{n,j_n}$ such that each j_i is the maximal integer from $\{0,\ldots,2i+1\}$ with the point p_{ij_i} lying in the half-plane $y \le \beta x + \gamma$. So, the $(i+1)$th symbol w_i of $w_\alpha(\beta,\gamma,n)$ is equal to 1 if and only if j_i is even in $P_\alpha(\beta,\gamma,n)$. In particular we obtain the following lemma:

Lemma 3.1 *If for some $\beta,\beta',\gamma,\gamma' \in [0,1)$ we have $P_\alpha(\beta,\gamma,n) = P_\alpha(\beta',\gamma',n)$, then $w_\alpha(\beta,\gamma,n) = w_\alpha(\beta',\gamma',n)$.*

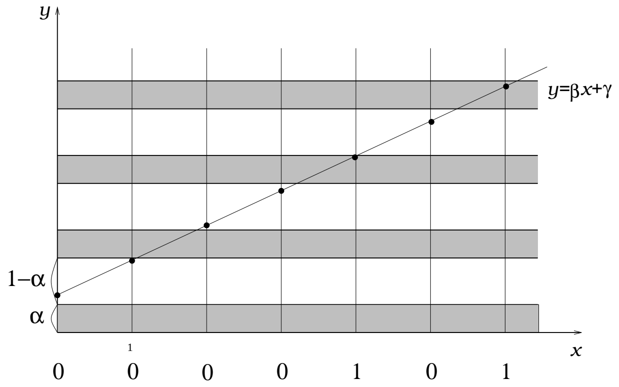


Figure 1: A word from A_α

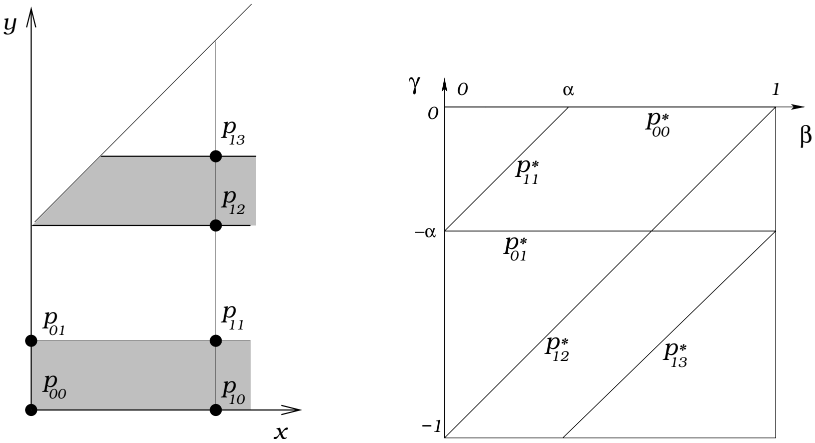


Figure 2: Construction of the arrangement $D_\alpha(1)$

Note that the converse is false: different sequences $P_\alpha(\beta, \gamma, n)$ may give the same words $w_\alpha(\beta, \gamma, n)$. Below we shall partially classify the situations when it happens.

Now let us define the set $\mathcal{P}^*$ of all non-vertical lines $y = \beta x + \gamma$, $\beta, \gamma \in \mathbb{R}$ and introduce the duality transform $x \in \mathcal{P} \cup \mathcal{P}^* \mapsto x^* \in \mathcal{P} \cup \mathcal{P}^*$ from [4] which maps the point p with coordinates (β, γ) to the line p^* with equation $y = \beta x - \gamma$ and the line $l \in \mathcal{P}^*$ with equation $y = \beta x + \gamma$ to the point $l^* \in \mathcal{P}$ with coordinates $(\beta, -\gamma)$. This transform preserves the incidence relation: points lying on the same line are mapped to lines crossing at the same point and vice versa; a point $p = (a, b)$ lies in the half-plane $y \leq cx + d$ below the line l with equation $y = cx + d$ if and only if $l^* = (c, -d)$ lies in the half-plane $y \leq ax - b$ below p^*; etc. The set $\mathcal{L} \subset \mathcal{P}^*$ is mapped onto the square $\mathcal{L}^* = \{(\beta, \gamma) | 0 \leq \beta, -\gamma < 1\}$. For details and pictures concerning the duality transform, see [4].

Let us draw the intersection of the set $\mathcal{P}^*_\alpha(n)$ with the interior of $\mathcal{L}^*$. Adding to the segments of p^*_{ij} the four borders of the square $\mathcal{L}^*$, we obtain the picture

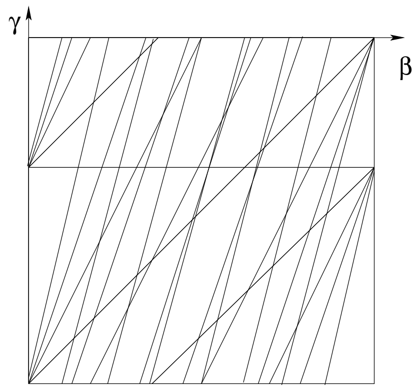


Figure 3: An arrangement $D_\alpha(4)$

called the *arrangement* $D_\alpha(n)$ (see Fig. 2, 3). In what follows, we shall interpret an arrangement as a planar graph with vertices (defined as the intersection points of all the segments of the picture), edges (defined as pieces of segments delimited by two vertices), and faces, defined as the *interiors* of the polygons delimited by the edges. Note that the external face is not counted. For example, $D_\alpha(1)$ has 6 faces (see Fig. 2).

We shall say that a point $l^* \in \mathcal{L}^*$ *defines* a word w if the line $l \in \mathcal{L}$ defines w.

Lemma 3.2 *If two points l^* and l'^* lie in the same face of $D_\alpha(n)$, then l and l' define the same word of $A_\alpha(n+1)$.*

Due to this lemma, we can say that a face f of the arrangement $D_\alpha(n)$ *defines* a word $w \in A_\alpha(n+1)$ if any point l^* lying in it defines w. Recall that the face is defined as the interior of a polygon!

Note that if $l^* = (x, -y)$ lies on an edge of the arrangement $D_\alpha(n)$ which is a part of the line p^*_{ij}, then its dual line l passes via the point p_{ij}. We see that l^* defines the same word as the face below it (more precisely, the face containing points $(x+\varepsilon, -y-\varepsilon)$ for small positive ε). So, points which do not lie in faces of $D_\alpha(n)$ do not define new words of $A_\alpha(n+1)$, and we have the following

Corollary 3.3 *For each α and $n \geq 1$, the arithmetical complexity $a_\alpha(n)$ is not greater than the number of faces of $D_\alpha(n-1)$.*

At the same time, the number of faces of $D_\alpha(n)$ can be computed by just the same technique which has been used by Berstel and Pocchiola [4]. In particular, it does not depend on α, provided that α is irrational.

Lemma 3.4 *For any irrational α, the number of faces of $D_\alpha(n)$ is*

$$d_\alpha(n) = 2 + \frac{n(n+1)(n+2)}{3} + 2\sum_{p=1}^{n}(n-p+1)\varphi(p),$$

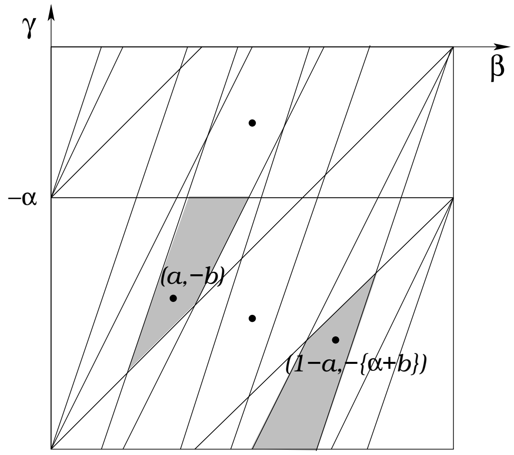


Figure 4: Centers of symmetry and a pair of symmetric points

where $\varphi(p)$ is the Euler function.

4 Symmetry

In fact, the upper bound from Corollary 3.3 can be instantly made two times less when we mention that arrangements are symmetric. It can be easily seen that the part of an arrangement lying above the line $y = -\alpha$ (equal to p^*_{01}) is symmetric about the point $(1/2, -\alpha/2)$, and the part lying below this line is symmetric about the point $(1/2, -(1 + \alpha)/2)$ (see Fig. 4). This symmetry applies not only to faces of the arrangement but to words defined by them.

Lemma 4.1 *If a point $(a, -b) \in \mathcal{L}^*$ lies in a face of $D_\alpha(n)$, then so does the point $(1 - a, -\{\alpha - b\})$. These two points define the same word of length $n + 1$.*

The centers of symmetry $C_1 = (1/2, -\alpha/2)$ and $C_2 = (1/2, -(1 + \alpha)/2)$ always lie inside faces c_1 and c_2 which define respectively the words $10101\cdots$ and $01010\cdots$. Any other face of $D_\alpha(n)$ is symmetric to another one, distinct from it but defining the same word. So, Corollary 3.3 can be improved to

$$a_\alpha(n) \leq d_\alpha(n-1)/2 + 1.$$

To write it more clearly, let us denote

$$g(n) = \frac{n(n+1)(n+2)}{6} + \sum_{p=1}^{n}(n-p+1)\varphi(p) + 2.$$

Then we have the following

Theorem 4.2 *For each irrational $\alpha \in (0, 1)$ the inequality holds $a_\alpha(n + 1) \leq g(n)$.*

Note that $g(n) = (1/6 + 1/\pi^2)n^3 + O(n^2)$, so we have obtained a uniform $O(n^3)$ upper bound for $a_\alpha(n)$. In what follows we shall try to do more, namely, to classify non-symmetric faces of $D_\alpha(n)$ defining the same words and thus to pass from the upper bound to a precise formula for the arithmetical complexity. We shall succeed only for $\alpha \in (1/3, 2/3)$ and discuss the difficulties arising for other α's. Note that for each irrational α, the sets A_α and $A_{1-\alpha}$ can be obtained one from the other by interchanging 0s and 1s. Due to this symmetry, without loss of generality from now on we consider only $\alpha < 1/2$.

5 Inheritance of faces

Let us say that a face of $D_\alpha(n-1)$ is a *splitting* face if two lines p^*_{ni} and $p^*_{n,i+1}$ meet it for some i, that is, if it is split into (at least) three faces of $D_\alpha(n)$.

Suppose that we have two faces f_1 and f_2 of $D_\alpha(n)$ not symmetric to each other and defining the same word $x = x_0 \cdots x_n$. Without loss of generality we assume that both of them intersect with the half-square $\beta < 1/2$ (otherwise due to Lemma 4.1 we could substitute one or both of them by respective symmetrical faces). We shall refer to such faces as to a *non-symmetric pair.* The faces of such pair are subsets either of the same face f of $D_\alpha(n-1)$ which defines the word $x' = x_0 \cdots x_{n-1}$, or of two different faces f'_1 and f'_2 of $D_\alpha(n-1)$, both defining the word x'. In the first case, the face f is a splitting face. In the second case, suppose that f'_1 and f'_2 are symmetric to each other; then both of them meet the line $\beta = 1/2$. Otherwise f'_1 and f'_2 also constitute a non-symmetric pair.

We see that to classify the non-symmetric pairs of $D_\alpha(n)$, it is sufficient to trace what happens in $D_\alpha(n-1)$ to splitting faces, non-symmetric pairs and faces meeting the line $\beta = 1/2$. This task becomes easier when we mention that splitting faces cannot occur from nowhere:

Lemma 5.1 *Suppose that there is a splitting face in $D_\alpha(n)$ with $n \geq 1$ defining a word $x_0 x_1 \cdots x_n$. Then there is a splitting face in $D_\alpha(n-1)$ defining the word $x_1 \ldots x_n$.*

Lemma 5.2 *Suppose that for some $n \geq 4$, the only splitting face(s) in $D_\alpha(n-1)$ are c_1 or/and c_2 (which are faces containing centers of symmetry of the upper and the lower parts of the arrangement). Then there no splitting faces in $D_\alpha(n)$ except possibly c_2 or/and c_1.*

In what follows, we show that the situation of this lemma ultimately appears for $\alpha > 1/3$. To do it, we first show that arrangements of order n corresponding to slopes from the same Farey interval of order n are isomorphic.

Recall that the Farey series of order n is the sequence of all irreducible fractions with denominators not greater than n taken in ascending order. For example, Farey series of order 6 is $0, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, 1$. The open interval between two successive entries of the Farey series is called a *Farey interval*

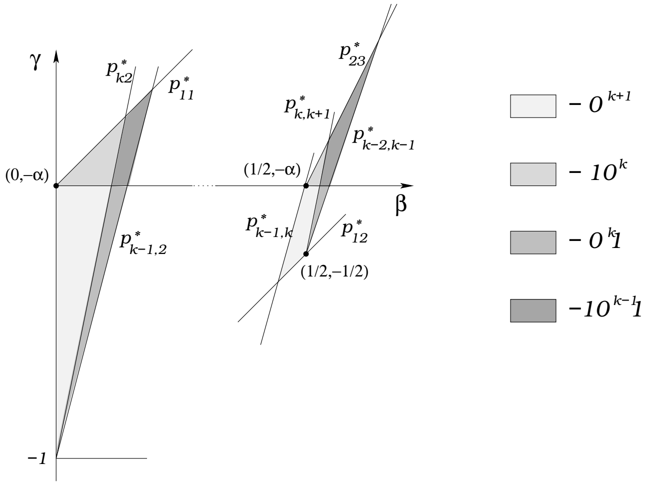


Figure 5: Non-symmetric pairs in $D_\alpha(k)$, k is odd

of order n.

Lemma 5.3 *If α_1 and α_2 lie in the same Farey interval of order n, then $D_{\alpha_1}(n)$ and $D_{\alpha_2}(n)$ are isomorphic.*

Here the isomorphism of arrangements is defined so that it implies a one-to-one correspondence between faces. If $D_{\alpha_1}(n)$ and $D_{\alpha_2}(n)$ are isomorphic, then to each face of $D_{\alpha_1}(n)$ we can relate a face of $D_{\alpha_2}(n)$ so that these faces define equal words and are bounded by sets of similarly denoted lines from $\mathcal{P}^*_{\alpha_1}(n)$ and $\mathcal{P}^*_{\alpha_2}(n)$.

6 Precise formulas

Recall that due to the symmetry, we can assume that $\alpha < 1/2$.

Theorem 6.1 *For each irrational $\alpha \in (0.4, 0.5)$ we have $a_\alpha(1) = 2$, $a_\alpha(2) = 4$, $a_\alpha(3) = 8$, $a_\alpha(4) = 16$, $a_\alpha(5) = 30$, and*

$$a_\alpha(n+1) = \begin{cases} g(n) - 4, & \textit{if } n \textit{ is odd}, \\ g(n) - 3, & \textit{if } n \textit{ is even} \end{cases}$$

for $n + 1 \geq 6$.

Proof (Sketch) Let us fix an irrational $\alpha \in (0.4, 0.5) = (2/5, 1/2)$. The clusters of 4 faces of $D_\alpha(5)$ and 3 faces of $D_\alpha(6)$ which appear in non-symmetric pairs are depicted on Fig. 5 and 6 respectively for $k = 5$ and $k = 6$; there are no more non-symmetric pairs in those arrangements.

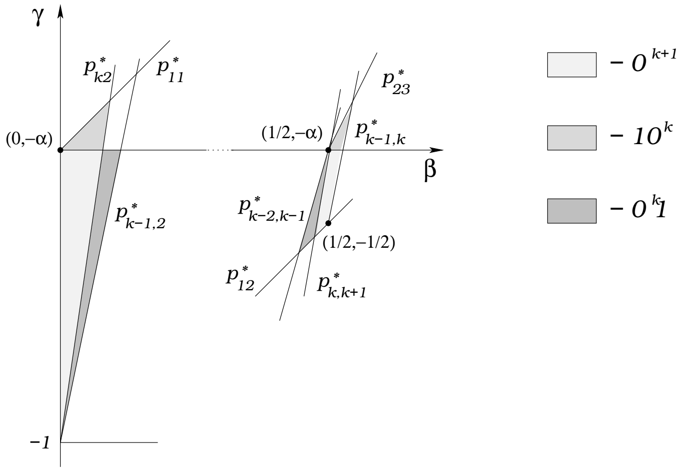


Figure 6: Non-symmetric pairs in $D_\alpha(k)$, k is even

Note that all non-symmetric pairs of $D_\alpha(6)$ occur from non-symmetric pairs of $D_\alpha(5)$. So, there are no splitting faces in $D_\alpha(5)$ which would split to non-symmetric pairs, and the only splitting faces are c_1 and c_2. Due to Lemma 5.2, there are no splitting faces except c_1 and c_2 in arrangements of higher orders. Due to Lemma 5.3, all these arguments apply for all α from the same Farey interval of order 5, that is, from $(2/5, 1/2) = (0.4, 0.5)$. So, it remains to trace what happens to non-symmetric pairs of $D_\alpha(n)$ when passing to $D_\alpha(n+1)$. This can be done by induction, and the general situation can be seen at Fig.. 5 and 6 . We see that non-symmetric pairs are 4 if n is odd and 3 if n is even, which is sufficient for the theorem to be proved. □

Theorem 6.2 *For each irrational* $\alpha \in (0.375, 0.4)$ *we have* $a_\alpha(1) = 2$*,* $a_\alpha(2) = 4$*,* $a_\alpha(3) = 8$*,* $a_\alpha(4) = 16$*,* $a_\alpha(5) = 30$*,* $a_\alpha(6) = 52$*,* $a_\alpha(7) = 83$*,* $a_\alpha(8) = 128$ *and*

$$a_\alpha(n+1) = \begin{cases} g(n) - 8, & \textit{if } n \textit{ is even,} \\ g(n) - 9, & \textit{if } n \textit{ is odd} \end{cases} \tag{6.1}$$

for $n + 1 \geq 9$*.*

In particular, this formula is valid for the Fibonacci word, i. e., for $\alpha = (3 - \sqrt{5})/2 = 0.381966\cdots$.

Proof (Sketch) In this range of α, splitting faces not equal to c_1 and c_2 exist longer and disappear only at $D_\alpha(8)$. To prove it, we consider $D_\alpha(8)$. In this arrangement, there are eight non-symmetric pairs of faces: three of them look as at Fig. 6 and five cross vertical lines $\beta = 1/5$ and $\beta = 2/5$, different faces of the same pair crossing different lines. Faces of these five pairs define the word

100001000 and the four its conjugates. At the same time, at $D_\alpha(9)$ there are nine non-symmetric pairs, four of them look as at Fig. 5 and five are descendants of respective pairs of $D_\alpha(8)$. So, there are no splitting faces in $D_\alpha(8)$; this is valid for all α from the Farey interval $(3/8, 2/5) = (0.375, 0.4)$. The 3 or 4 non-symmetric pairs from Fig.. 5 and 6 behave exactly as for $\alpha \in (0.4, 0.5)$; the new pairs always remain 5. □

We managed also to trace splitting faces remaining in $D_\alpha(8)$ for $1/3 < \alpha < 3/8$. In this case, splitting faces disappear in $D_\alpha(3k-1)$, where k is uniquely defined by $k/(3k-1) < \alpha < (k-1)/(3k-4)$, and the arrangements start to behave as for $\alpha \in (0.375, 0.4)$. So, we can obtain the following

Theorem 6.3 *Suppose that α is irrational and lies in the interval $\left(\frac{k}{3k-1}, \frac{k-1}{3k-4}\right)$ for some integer $k \geq 3$. Then*

- *for $n+1 < 9$, $a_\alpha(n+1)$ behaves as it is described in Theorem 6.2;*
- *for $n+1 = 9, \ldots, 3k-1$, the function $g(n) - a_\alpha(n+1)$ is 6-periodic with the period 10,12,10,11,11,11;*
- *for $n+1 \geq 3k$, (6.1) holds.*

This theorem completes the description of $a_\alpha(n)$ for irrational $\alpha \in (1/3, 1/2)$. The situation for $\alpha < 1/3$ is more complicated since there exist non-central splitting faces in $D_\alpha(n)$ for arbitrarily large n, and the difference between $g(n)$ and $a_\alpha(n+1)$ grows.

References

[1] S. V. Avgustinovich, J. Cassaigne, A. Frid, *Sequences of low arithmetical complexity*, submitted.

[2] S. V. Avgustinovich, D. G. Fon-Der-Flaass, A. E. Frid, *Arithmetical complexity of infinite words*, in: Words, Languages and Combinatorics III, World Scientific, Singapore, 2003. P. 51–62. (Proc. 3rd ICWLC, Kyoto, March 2000).

[3] J. Berstel, P. Séébold, *Sturmian words*, in: M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, 2002. P. 40–97.

[4] J. Berstel, M. Pocchiola, *A geometric proof of the enumeration formula for Sturmian words*, Int. J. Algebra and Comput. **3** (1993), 349-355.

[5] A. E. Frid, *A lower bound for the arithmetical complexity of Sturmian words*, Siberian Electronic Mathematical Reports **2**, 14-22 [Russian, English abstract].

[6] A. Frid, *Sequences of linear arithmetical complexity*, Theoret. Comput. Sci., accepted.

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

[8] I. Nakashima, J.-I. Tamura, S.-I. Yasutomi, **-Sturmian words and complexity*, Journal de Théorie des Nombres de Bordeaux **15** (2003), 767–804.

[9] A. Restivo, S. Salemi, *Binary Patterns in Infinite Binary Words*, in: Formal and Natural Computing, W. Brauer, H Ehrig, J. Karhumaki, A. Salomaa, Eds., Lecture Notes in Computer Science 2300 (2002) 107-116.

[10] G. Rote, *Sequences with subword complexity* $2n$., J. Number Th. **46** (1994), 196–213.

Congruences and the Thue-Morse Sequence

Emeric Deutsch[*], *Bruce E. Sagan*[†]

Abstract

We show that the Thue-Morse sequence arises in the study of various congruences. In particular, it is connected to congruences involving the central binomial coefficients, Motzkin numbers, Motzkin prefix numbers, Riordan numbers, and hex tree numbers.

Résumé

Nous démontrons que la séquence Thue-Morse apparaît dans l'étude des congruences. En particulier, c'est reliée aux congruences qui contiennent les coefficients binomiaux centrals, les nombres de Motzkin, les nombres de Motzkin prefix, les nombres de Riordan, et les nombres d'arbre hex.

1 Introduction

The Thue-Morse sequence is certainly ubiquitous in the study of combinatorics on words [2]. However, it does not seem to have been noticed until now that it is also intimately connected with congruences for certain combinatorial sequences. We will show that it appears in congruences involving the central binomial coefficients, Motzkin numbers, Motzkin prefix numbers, Riordan numbers, and hex tree numbers. Here, only the results are presented. Those wishing details of the proofs should consult [4].

First, let us set some notation. Let $\mathbb{N}$ denote the nonnegative integers and consider an integer sequence

$$\mathbf{a} = (a_0, a_1, a_2, \ldots) = (a_n)_{n\in\mathbb{N}}.$$

Then we will perform operations from linear algebra on such sequences as if they were infinite vectors. When appropriate, we will identify a scalar k with the sequence which is constant at k. So, for example, the sequence $2\mathbf{a}+3$ would have $2a_n + 3$ as it's nth term.

If m is a positive integer then it will be convenient to write $k \equiv_m l$ instead of the more conventional $k \equiv l (\bmod m)$. If the base m expansion of k is

$$k = k_0 + k_1 m + k_2 m^2 + \cdots$$

[*] Department of Mathematics, Polytechnic University, Brooklyn, NY 11201, USA, deutsch@duke.poly.edu

[†] Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027, USA, sagan@math.msu.edu

then we will denote the sequence of digits by

$$(k)_m = (k_0, k_1, k_2, \ldots) = (k_i).$$

We will also let

$$\delta_m(k) = \# \text{ of ones in } (k)_m.$$

Finally, we denote the Thue-Morse sequence by

$$\mathbf{t} = (0,\ 1,\ 1,\ 0,\ 1,\ 0,\ 0,\ 1,\ \ldots).$$

One can define $\mathbf{t}$ in many ways. For example, t_n is the parity of $\delta_2(n)$.

2 Central binomial coefficients

One of the most famous and most useful congruences for binomial coefficients is due to Lucas [10].

Theorem 2.1 (Lucas) *Let p be a prime and let $(n)_p = (n_i)$ and $(k)_p = (k_i)$. Then*

$$\binom{n}{k} \equiv_p \prod_i \binom{n_i}{k_i}. \tag{2.1}$$

Using this theorem, it is not hard to give a simple formula for the congruence class of the central binomial coefficients modulo 3 which settles conjectures of Cloitre and Zumkeller [11, A074938–40]. To state the result, let

$$T(01) = \{n \in \mathbb{N} \ : \ (n)_3 \text{ contains only digits equal to 0 or 1}\}.$$

Theorem 2.2 *The central binomial coefficients satisfy*

$$\binom{2n}{n} \equiv_3 \begin{cases} (-1)^{\delta_3(n)} & \text{if } n \in T(01), \\ 0 & \text{otherwise.} \end{cases}$$

Of course, one could also write down an expression for $\binom{2n}{n}$ modulo any prime. The interest in the modulus 3 case stems from the connection with $\mathbf{t}$. The next result can be derived from Theorem 2.2 and settles further conjectures of Cloitre [11, A074938–9].

Theorem 2.3 *We have*

$$\left(n \ : \ \binom{2n}{n} \equiv_3 1\right) \equiv_3 \mathbf{t}.$$

and

$$\left(n \ : \ \binom{2n}{n} \equiv_3 -1\right) \equiv_3 1 - \mathbf{t}.$$

It is amusing that the Thue-Morse sequence, which is such an essentially modulo 2 object, comes up in this modulo 3 setting.

3 Motzkin numbers

The Motzkin numbers are closely related to the *Catalan numbers*

$$C_n = \frac{1}{n+1}\binom{2n}{n}, \qquad n \in \mathbb{N}.$$

In fact, the nth *Motzkin number* can be defined as

$$M_n = \sum_{k\geq 0}\binom{n}{2k}C_k, \qquad n \in \mathbb{N},$$

where, as usual, a binomial coefficient is zero if the bottom is larger than the top. Just as with C_n, there are also a number of combinatorial ways to define M_n. In particular, they count the number of ordered trees with n edges such that each vertex has at most 2 children.

To determine the parity of M_n, we will need a sequence related to $\mathbf{t}$. A *run* in a sequence is a maximal subsequence of consecutive, equal elements. Now define a sequence $\mathbf{c}$ whose nth term is the number of elements in the first n runs of $\mathbf{t}$ (where we consider the initial zero of $\mathbf{t}$ to be the 0th run). Then

$$\mathbf{c} = (1,\ 3,\ 4,\ 5,\ 7,\ \ldots).$$

Various properties of $\mathbf{c}$ were studied in [1]. Using the description of M_n in terms of trees and induction, we are able to prove the following theorem which is also implicit in the work of Klazar and Luca [9].

Theorem 3.1 *The Motzkin number M_n is even if and only if either $n \in 4\mathbf{c}-2$ or $n \in 4\mathbf{c}-1$.*

4 Related sequences

There are various sequences related to the M_n. So, using Theorem 3.1, one can also derive their congruence properties modulo 2 in terms of the sequence $\mathbf{c}$.

A *Motzkin path of length n* is a lattice path in the lattice $\mathbb{N}\times\mathbb{N}$ with steps $(1,1)$, $(1,-1)$, and $(1,0)$ starting at $(0,0)$ and ending at $(n,0)$. It is well known that M_n is the number of Motzkin paths of length n. (Note that we do not need any condition about staying above the x-axis since we are working in $\mathbb{N}\times\mathbb{N}$.) Define a *Motzkin prefix of length n* to be a lattice path which forms the first n steps of a Motzkin path of length $m \geq n$. Equivalently, a Motzkin prefix is exactly like a Motzkin path except that the endpoint is not specified. Let P_n, $n \geq 0$, be the number of Motzkin prefixes of length n. This is sequence A005773 in Sloane's Encyclopedia [11]. The P_n also count directed rooted animals with $n+1$ vertices as proved by Gouyou-Beauchamps and Viennot [7].

Corollary 4.1 *The number P_n is even if and only if $n \in 2\mathbf{c}-1$.*

Next we consider the *Riordan numbers* [11, A005043], γ_n, which count the number of ordered trees with n edges where every nonleaf has at least two children. These are called *short bushes* by Bernhart [3]. If we relax the degree restriction so that the root can have any number of children then the resulting trees are called *bushes.* It is known [5,6] that M_n is the number of bushes with $n+1$ edges. It follows that

$$M_n = \gamma_{n+1} + \gamma_n$$

since every bush with $n+1$ edges is either a short bush or has a root with one child which generates a short bush with n edges. From this one gets the following result.

Corollary 4.2 *The number γ_n is even if and only if $n \in 2\mathbf{c} - 1$.*

Finally, consider the sequence counting *restricted hexagonal polyominos* [11, A002212]. The reader can find the precise definition of these objects in the paper of Harary and Read [8]. We will use an equivalent definition in terms of trees which can be obtained from the polyomino version by connecting the centers of adjacent hexagons. A *ternary tree* is a rooted tree where every vertex has some subset of three possible children: a left child, a middle child, or a right child. A *hex tree* is a ternary tree where no node can have two adjacent children. (A middle child would be adjacent to either a left or a right child but left and right children are not adjacent.) Let H_n, $n \geq 0$, be the number of hex trees having n edges.

Corollary 4.3 *The number H_n is even if and only if $n \in 4\mathbf{c} - 2$ or $n \in 4\mathbf{c} - 1$.*

It would be interesting to understand these results combinatorially. Since the proof of Theorem 3.1 is partly inductive, there is as yet no really good explanation for why the sequence $\mathbf{c}$ enters into these congruences.

References

[1] J.-P. Allouche, A. Arnold, J. Berstel, S. Brlek, W. Jockusch, S. Plouffe, and B. E. Sagan, A relative of the Thue-Morse sequence, *Discrete Math.* **139** (1995), 455–461.

[2] J.-P. Allouche and J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, in "Sequences and their applications, Proceedings of SETA'98," C. Ding, T. Helleseth, and H. Niederreiter eds., Springer-Verlag, 1999, 1–16.

[3] F. Bernhart, Catalan, Motzkin, and Riordan numbers, *Discrete Math.* **204** (1999), 73–112.

[4] E. Deutsch and B. E. Sagan, Congruences for Catalan and Motzkin numbers and related sequences, preprint, available at **http://www.math.msu.edu/~sagan/**.

[5] R. Donaghey, Restricted plane tree representations of four Motzkin-Catalan equations, *J. Combin. Theory, Ser. B* **22** (1977), 114–121.

[6] R. Donaghey and L. W. Shapiro, Motzkin numbers, *J. Combin. Theory, Ser. A* **23** (1977), 291–301.

[7] D. Gouyou-Beauchamps and G. Viennot, Equivalence of the two-dimensional directed animal problem to a one-dimensional path problem, *Adv. Appl. Math.* **9** (1988), 334–357.

[8] F. Harary and R. C. Read, The enumeration of tree-like polyhexes, *Proc. Edinburgh Math. Soc.* **17** (1970), 1–13.

[9] M. Klazar and F. Luca, On integrality and periodicity of the Motzkin numbers, preprint.

[10] E. Lucas, Sur les congruences des nombres eulériens et des coefficients différentiels des fonctions trigonométriques suivant un module premier, *Bull. Soc. Math. France* **6** (1877–1878), 49–54.

[11] N. J. A. Sloane, “The On-Line Encyclopedia of Integer Sequences,” available at **http://www.research.att.com/~njas/sequences/**.

Substitutions on an infinite alphabet: first results

Sébastien Ferenczi[*]

Abstract

We give a few examples of substitutions on an infinite alphabet, and the beginning of a general theory of the associated dynamical systems.

1 Substitutions

Let A be a finite or countable set, called the **alphabet**, and its elements will be called **letters**.

Definition 1.1 A **word** is a finite string $w_1...w_k$ of elements of A ; the concatenation of two words w and w' is denoted multiplicatively, by ww'. A word $w_1...w_k$ is said to **occur** at place i in the infinite sequence or finite word u if $u_i = w_1, ..., u_{i+k-1} = w_k$; when u is finite, we denote by $N(w,u)$ the number of these occurrences.

A **substitution** is an application from an alphabet A into the set $A^\star$ of finite words on A ; it extends to a morphism of $A^\star$ for the concatenation by $\sigma(ww') = \sigma w \sigma w'$.

It is called **primitive** if there exists k such that a occurs in $\sigma^k b$ for any $a \in A$, $b \in A$.

It is called **of constant length** q if σa is of length q for any $a \in A$.

A **fixed point** of σ is an infinite sequence u with $\sigma u = u$.

For any sequence $u = (u_n, n \in \mathbb{N})$ on a finite alphabet A, we can define the (topological) **symbolic dynamical system** associated to u: we first take $\Omega = A^{\mathbb{N}}$, equipped with the product topology (each copy of A being equipped with the discrete topology) and T the one-sided shift

$$T(x_0x_1x_2\ldots) = x_1x_2x_3\ldots$$

then X_u is the closure of the orbit of u under T. The dynamical system associated to a primitive substitution is the symbolic system (X_u, T) associated to any of its fixed points.

[*] Institut de Mathématiques de Luminy, CNRS, UPR 9016, 163 av. de Luminy, F13288, Marseille Cedex 9 (France), Fédération de Recherche des Unités de Mathématiques de Marseille, CNRS - FR 2291, phone: (+33) 491269675, fax (+33) 491269655, ferenczi@iml.univ-mrs.fr

In the usual case when A is finite, the theory is well-established, see for example [QUE], [PYT]: under the (relatively mild) assumption of primitivity, the symbolic system X_u, T is **minimal**: the closed orbit of any point under T is the whole X_u, or, equivalently, for all m there exists n such that every word of length m occurring in u occurs in every word of length n occurring in u. Under the same assumption, the system is **uniquely ergodic**: it admits a unique invariant probability measure μ. The measure-theoretic dynamical systems built from primitive substitutions give many interesting examples in ergodic theory, such as

Example 1.2 (The Morse substitution)

$$\begin{array}{ccc} a & \to & ab \\ b & \to & ba \end{array}$$

For any sequence u, the **language** $L(u)$ is the set of all words occurring in u ; the **complexity** of u is the function $p(n)$ which associates to each $n \in \mathbb{N}$ the number of words of length n in $L(u)$. For fixed points of primitive substitutions, the complexity is always bouded by Cn, and this implies the system has topological (and hence measure-theoretic) **entropy** zero.

2 A fundamental example

The following broad question was asked by C. Mauduit: what can be said of the following substitution on $A = \mathbb{Z}$?

Example 2.1 (The drunken man substitution)

$$n \to (n-1)(n+1)$$

for all $n \in \mathbb{Z}$

The first obstacle is that, if we look at the k-th image of 0, it is made only of even (resp. odd) numbers if k is even (resp. odd); this reflects the fact that the matrix has period two (see section 3 below). Hence the right substitution to consider is

Example 2.2 (The squared drunken man substitution)

$$n \to (n-2)nn(n+2)$$

for all $n \in A = 2\mathbb{Z}$

This substitution, which we denote by σ, has no fixed point; but we can define a subset X of $A^{\mathbb{N}}$ to be the set of all sequences $x = x_0x_1\dots$ such that every word occurring in x occurs also in $\sigma^n 0$ for at least one $n > 0$. X is then a closed subset of the (noncompact) set $\mathbb{Z}^{\mathbb{N}}$ equipped with the product topology (each copy of $\mathbb{Z}$ being equipped with the discrete topology), and is invariant by the shift T. We say that (X, T) is the (non-compact) symbolic system associated to the substitution σ.

It is trivially false that, in any given sequence x of X, for all m there exists n such that every word of length m occurring in x occurs in every word of length n occurring in x; but on an infinite alphabet the minimality of the system (X, T) would be equivalent to a weaker property, namely that any word occurring in one element of X occurs in every element of X. But in fact this property is not satisfied here, as there exist infinite sequences in X without any occurrence of the letter 0: take for example the sequence beginning by $\sigma^n(2n)$ for all n.

Hence

Proposition 2.3 (X, T) *is not minimal.*

Though individual sequences may have strange properties, we are looking at good statistical properties for "typical" sequences of X. This involves looking for invariant measures; but here the situation is also different from the finite case, as

Proposition 2.4 *There is no finite measure on X invariant under T.*

Definition 2.5 For any words v and w, we say that v is an **ancestor** (under σ) of w with multiplicity m if w occurs in σv at m different places. If $w = w_0 \dots w_s$, the **cylinder** $[w]$ is the set $\{x \in X; x_0 = w_0, \dots, x_s = w_s\}$.

We define the **natural measure** μ on (X, T) by assigning to each cylinder $[n]$, $n \in A$, or $T^k[n]$, the measure 1, and to a cylinder $[w]$, or $T^k[w]$, the measure $\frac{1}{4} \sum \mu[v] m(v)$, the sum being taken on all its ancestors v and $m(v)$ denoting their multiplicities.

Proposition 2.6 *μ is an infinite measure on X invariant under T.*

Lemma 2.7 *σ is* left determined*: there exists N such that, if w is a word of length at least N in the language $L(u)$, it has a unique decomposition $w = w_1 \dots w_s$ where each w_i is a σa_i for some $a_i \in A$, except that w_1 may be only a suffix of σa_1 and w_s may be only a prefix of σa_s; furthermore the a_i, $1 \leq i \leq s-1$, are unique.*

Lemma 2.8 *The system (X, T, μ) is generated by a countable family of* **Rokhlin stacks***: namely, for every $n \in \mathbb{N}$, X is, up to sets of μ-measure zero, the disjoint union of the $T^k[\sigma^n j]$, $j \in 2\mathbb{Z}$, $0 \leq k \leq 4^n - 1$.*

Proposition 2.9 *The system (X, T, μ) is* **recurrent***: namely, for every set E with $0 < \mu(E)$, $\mu\{x \in E; T^n x \notin E$ for every $n > 0\} = 0$.*

Proposition 2.10 *The system (X, T, μ) is* **ergodic***: namely, for every set E with $0 < \mu(E)$ and $\mu(E \Delta TE) = 0$, either $\mu(E) = 0$ or $\mu(X/E) = 0$.*

In fact, to prove ergodicity, we prove that though we cannot define frequencies for words, we may define ratios of frequencies: namely, for almost all $x \in X$, and words w and w', $\frac{1}{n}N(w, x_0 \dots x_{n-1})$ has limit zero when $n \to +\infty$, but $\frac{N(w,x_0 \dots x_{n-1})}{N(w',x_0 \dots x_{n-1})}$ does converge to $\frac{\mu[w]}{\mu[w']}$.

Note that there are many T-invariant measures concentrated on the same set as μ; in particular, if in the definition of μ we replace the natural constant 4 by any $C \geq 4$, we get another infinite measure on X invariant under T - necessarily nonergodic.

Because of the recurrence, it makes sense to study the **induced**, or first return, map of (X, T, μ) on the cylinder $[0]$. Let (Y, S, ν) be this system.

Proposition 2.11 *The system (Y, S, ν) is measure-theoretically isomorphic to the symbolic system associated to the substitution τ on $A = \mathbb{N} \times \mathbb{Z}$, equipped with its natural measure, which is an invariant probability measure.*

where τ is the

Example 2.12 (The induced drunken man substitution)

$$(m, n) \to \prod_{j=0}^{n-1+m^+} (j, 1) \quad (m, n+1) \prod_{i=-n+1+m^-}^{-1} (i, 1)$$

for all $m \in \mathbb{Z}$ and $n \geq 1$.

and its natural measure is defined by $\nu[m, n] = \nu(S^k[m, n]) = 2^{-|m|-2n}$ and a cylinder $[w]$, or $S^k[w]$, has measure $\frac{1}{4}\sum \mu[v]m(v)$, the sum being taken on all its ancestors (under τ) v and $m(v)$ denoting their multiplicities.

Proposition 2.13 *The system (Y, S, ν) is not minimal and not uniquely ergodic.*

The system (Y, S, ν) being a finite measure-preserving system, we can compute its measure-theoretic **entropy** $h(S, \nu)$. Note that τ has a fixed point u, which is the infinite sequence beginning by $\tau^n(0, 1)$ for every n.

Lemma 2.14 *If, for given M, the sequence $v(M)$ is deduced from u by replacing each (m, n) with the symbol ω when $|m| > M$ or $n > M$, then its complexity is bounded by $C(M)n^2$.*

Corollary 2.15 $h(S, \nu) = 0$.

3 General theory

Definition 3.1 The **matrix** of a substitution σ is defined by $M = ((m_{ij}))$ where m_{ij} is the number of occurrences of the letter j in the word σi.

We define the substitution σ_l on the alphabet A^l by associated to the l-letter $v_1 \dots v_l$ the l-word made by enumerating all the words of length l occurring in $\sigma(v_1 \dots v_l)$, starting from the first position. We denote by M_l the matrix of this substitution.

Definition 3.2 Let M be a natrix on a countable alphabet. We denote by $m_{ij}(n)$ the coefficients of M^n; M is **irreducible** if for every (i, j) there exists l such that $M_{ij}(l) > 0$. An irreducible M has **period** d if for every i $d = \mathrm{G}CD\{l; m_{ii}(l) > 0\}$, and is aperiodic if $d = 1$.

An irreducible aperiodic matrix admits a **Perron-Frobenius** eigenvalue λ defined as $lim_{n\to+\infty} m_{ij}(n)^{\frac{1}{n}}$. M is **transient** if

$$\sum_n m_{ij}(n)\lambda^{-n} < +\infty,$$

recurrent otherwise. For a recurrent M, we define $l_{ij}(1) = m_{ij}$, $l_{ij}(n+1) = \sum_{r\neq i} l_{ir}(n)m_{rj}$; M is **null recurrent** if

$$\sum_n nl_{ii}(n)\lambda^{-n} < +\infty,$$

and **positive recurrent** otherwise.

The reference for all the definitions and results on infinite matrices above is [KIT]. The vocabulary comes from the theory of random walks: when it is stochastic, a matrix is positive recurrent if it is the matrix of a random walk which returns to each point with probability one and the expectation of the waiting time is finite, it is null recurrent if it is the matrix of a random walk which returns to each point with probability one and the expectation of the waiting time is infinite, and it is transient if it is the matrix of a random walk which does not return to each point with probability one. And of course, the matrix of the drunken man substitution is the matrix of the famous random walk of the same name, though the dynamical systems we can associate to these two objects are completely different.

Now, for a given substitution σ on a countable alphabet A, we define the dynamical system asscociated to σ in the same way as in the previous section.

Proposition 3.3 *If σ is of constant length, left determined, and has an irreducible aperiodic positive recurrent matrix, the associated system (X, T) admits a natural ergodic invariant measure which is a probability.*

The natural measure is defined by taking $(\mu[n], n \in A)$ to be the normalized left eigenvctor of M for its Perron-Frobenius eigenvalue λ, $(\mu[w], w \in A^l)$ to be the normalized left eigenvctor of M_l for its (same) Perron-Frobenius eigenvalue λ, and $\mu(T^k[w]) = \mu[w]$ for all cylinders. When σ is of constant length, λ is the common length of the σn, $n \in A$.

Proposition 3.4 *If σ is of constant length, left determined, and has an irreducible aperiodic null recurrent matrix, the associated system (X, T) admits a natural infinite invariant measure.*

The natural measure is defined as in the previous case by Perron-Frobenius eigenvectors; as for the transient case, such a measure may exist or not.

4 Further examples

Example 4.1 (The one-sided drunken man substitution)

$$n \to (n-1)(n+1)$$

for all $n \geq 1$, and

$$0 \to 1.$$

As for its two-sided counterpart, this substitution has a matrix of period 2, hence we study its square.

Example 4.2 (The squared one-sided drunken man substitution)

$$n \to (n-2)nn(n+2)$$

for all even $n \geq 2$, and

$$0 \to 02.$$

Here the matrix is transient, but there exists an infinite invariant measure, whose value on letters is given by $\mu[2n] = 2n+1$. When we induce it on the cylinder [0] we get:

Example 4.3 (The induced one-sided drunken man substitution)

$$n \to 123\dots(n+1)$$

for all $n \geq 1$.

Turning to positive recurrent examples, we have:

Example 4.4 (The one step forward, two step backwards, substitution)

$$n \to (n-1)(n-1)(n+1)$$

for all $n \geq 1$, and

$$0 \to 111.$$

As its matrix is of period two,

Example 4.5 (The squared one step forward, two step backwards, substitution)

$$n \to (n-2)(n-2)n(n-2)(n-2)nnn(n+2)$$

for all even $n \geq 2$, and

$$0 \to 002002002.$$

The system has a natural invariant ergodic probability measure, which gives measure $\frac{1}{3}$ to $[0]$ and 2^{-2n+1} to $[2n]$, $n \geq 1$. But it is still not minimal, and not uniquely ergodic.

Example 4.6 (The golden ratio substitution)

$$n \to (n-2)(n+1)$$

for all $n \geq 2$,

$$0 \to 01,$$

$$1 \to 02.$$

It has an aperiodic matrix, positive recurrent, and the natural invariant ergodic probability gives to $[n]$ the measure $\frac{2^n(3-\sqrt{5})}{2(1+\sqrt{5})^n}$.

Example 4.7 (The infini-Bonacci substitution)

$$n \to 1(n+1)$$

for all $n \geq 1$.

This is a very special case, as the symbolic system is minimal and uniquely ergodic. Measure-theoretically, the system is isomorphic to the dyadic odometer, with an explicit coding to and from the system generated by the *period-doubling substitution* on two letters, 1 → 12, 2 → 11. From the combinatorial point of view, the infini-Bonacci fixed point was used by Cassaigne to build many interesting new sequences and thus earned the unofficial nickname of *the universal counter-example.*

References

[KIT] B. Kitchens: Symbolic dynamics. One-sided, two-sided and countable state Markov shifts, Universitext. (1998), Springer-Verlag.

[PYT] N. Pytheas Fogg: Substitutions in dynamics, arithmetics and combinatorics, Lecture Notes in Math. vol. 1794 (2002), Springer-Verlag.

[QUE] M. Queffelec: Substitution dynamical systems - Spectral analysis, Lecture Notes in Math. vol. 1294 (1987), Springer-Verlag.

Substitutions on Multidimensional Sequences

Thomas Fernique[*]

Abstract

We provide in this paper a multidimensional generalization of substitutions on words, which is defined as the action on multidimensional sequences of a *non-pointed substitution* endowed with *local rules*. The non-pointed substitutions and the local rules have in the multidimensional case respectively the roles played by the substitutions defined on letters and by the concatenation on words. This definition then allows us to provide a (yet partial) multidimensional generalization of an algebraic characterization of Sturmian words that are fixed-point or morphic image of a fixed-point of a non-trivial substitution on words.

Introduction

A substitution acts on a word in this way: the image of each letter is a word, and the image of the whole word is then just the concatenation of the images of its letters. Substitutions are powerful combinatorical tools, and have natural interactions with language theory, geometry of tilings, automata theory, and many others (see e.g. [14] and the references inside). It thus would be useful to define a similar tool in the more general framework of multidimensonal sequences, that are sequences of letters indexed by $\mathbb{Z}^n$ (whereas words are sequences of letters indexed by $\mathbb{N}$). It is however a difficult problem, mainly for lack of a natural "multidimensional concatenation".

Such a generalization has already been introduced in [15]: for $p_1, \ldots, p_n$ fixed in $\mathbb{N}$, a letter u indexed by $(i_1, \ldots, i_n)$ is mapped to a set $\sigma(u)$ of letters indexed by $\{(j_1, \ldots, j_n) \mid \forall k,\ p_k i_k \leq j_k < p_k(i_k+1)\}$ (that is, a $p_1 \times \ldots \times p_n$-rectangle). But it generalizes in fact only *constant-length* substitutions on words (which map letters to words all of the same length). An algebraic characterization of all the multidimensional sequences which are fixed point of such substitutions is also proved (see again [15]), what generalizes a similar result for words which are fixed-point of a constant-length substitution (see e.g. [1]).

A first aim of this paper is to introduce a notion of multidimensional substitution which generalizes any type of substitutions on words, and not only the

[*] LIRMM CNRS-UMR 5506 and Université Montpellier II, 161 rue Ada 34392 Montpellier Cedex 5 (France), PONCELET Lab. CNRS-UMI 2615 and Independent University of Moscow, Bol'shoj Vlas'evskij per. 11. 119002 Moscow (Russia), thomas.fernique@ens-lyon.org

constant-length ones (or any other particular type). Second, we would like to give an algebraic characterization of the multidimensional sequences which are fixed-point of such a multidimensional substitution. More precisely, Theorem 3.9 generalizes the following result (see e.g. [6, 9]):

Let α be an irrational number in $[0,1]$. *One defines the* Sturmian sequence $u_\alpha = (u_n)$ *over the alphabet* $\{1,2\}$ *by:*

$$\forall n \geq 1, \quad u_n = 1 \;\Leftrightarrow\; (n\alpha) \; mod \; 1 \in I_\alpha,$$

where $I_\alpha = (0, 1-\alpha]$ *or* $I_\alpha = [0, 1-\alpha)$. *Then* u_α *is a fixed point (resp. the morphic image of a fixed point) of a substitution on words if and only if* α *has a purely periodic (resp. eventually periodic) continued fraction expansion.*

Notice that this characterization concerns only Sturmian sequences, that is, a subset of the set of all the sequences. Thus, generalizing this result also requires to define a notion of "multidimensional Sturmian sequence".

The paper is organized as follows. In the first section, we define *non-pointed substitutions* and *local rules*, that are our multidimensional equivalents of the "classic" substitutions defined on letters, and of the concatenation product used to make such substitutions act on sequences. It allows us, under conditions on the local rules, to define our notion of multidimensional substitution. In Section 2, we describe a type of local rules which satisfy the conditions required to define a multidimensional substitution: the local rules derived from a *global rule.* In Section 3, we resume the notion of *generalized substitutions*, define *Sturmian hyperplane sequences* and then we show that these generalized substitutions provide global rules from which we can derive local rules as described in Section 2. It yields multidimensional substitutions on Sturmian hyperplane sequences, and allows us to give (Theorem 3.9) a partial generalization of the algebraic characterization of fixed-points stated above.

1 Non-pointed substitutions and local rules

Let $\mathcal{A}$ be a finite alphabet. A *pointed letter* is an element $L = (x, l)$ of $\mathbb{Z}^n \times \mathcal{A}$, where x is the *location* of the letter l. We denote by $\mathcal{L}$ the set of pointed letters.

A *pointed pattern* is a set of pointed letters with distinct locations. The *support* of a pointed pattern is defined as the set of the locations of its letters. Two pointed patterns are said *consistent* if two letters with the same location are identical. The notions of union, intersection and inclusion are then defined for consistent patterns as for usual sets. We denote by $\mathcal{P}$ the set of pointed patterns.

The lattice $\mathbb{Z}^n$ acts on pointed letters (resp. pointed patterns) by translation on the locations (resp. supports): the classes of equivalence of this action are called *non-pointed letters* and denoted by $\overline{\mathcal{L}}$ (resp. *non-pointed patterns*, denoted by $\overline{\mathcal{P}}$).

Thus, to each pointed pattern P corresponds a unique non-pointed pattern, called its *underlying non-pointed pattern* and denoted $\overline{P}$. Conversely, to each non-pointed pattern $\overline{P}$ corresponds all the *congruent* pointed patterns, called *realizations* of $\overline{P}$, that have $\overline{P}$ as underlying non-pointed pattern. If P and P' are congruent pointed patterns, one denotes $v(P, P') \in \mathbb{Z}^n$ the vector that maps P onto P' by translation.

We are now in a position to give our multidimensional generalization of the definition on letters of a substitution on words:

Definition 1.1 A *non-pointed substitution* is a map from $\overline{\mathcal{L}}$ to $\overline{\mathcal{P}}$.

In what follows, $\overline{\sigma}$ denote a non-pointed substitution. We now define *local rules*, which are the main ingredient of our "multidimensional concatenation".

Definition 1.2 We define two types of *local rules* for $\overline{\sigma}$:

- an *initial rule* λ^* is defined on a set $I(\lambda^*) = \{L\}$ of **one** pointed letter, and maps L to a realization of $\overline{\sigma}(\overline{L})$;
- an *extension rule* λ is defined on a set $E(\lambda) = \{L, L'\}$ of **two** pointed letters with distinct locations, and maps L and L' to disjoint realizations of respectively $\overline{\sigma}(\overline{L})$ and $\overline{\sigma}(\overline{L'})$.

Roughly speaking, an initial rule tells us how to position $\overline{\sigma}(\overline{L})$ for a particular pointed letter L, while an extension rule λ such that $E(\lambda) = \{L, L'\}$ is used, for a pointed pattern $\{A, A'\}$ congruent to $\{L, L'\}$, to position $\overline{\sigma}(\overline{A'})$ relatively to $\overline{\sigma}(\overline{A})$ in the same way $\lambda(L')$ is positioned relatively to $\lambda(L)$. We first define the action of $\overline{\sigma}$ on Λ-*paths*:

Definition 1.3 Let U be a pointed pattern and Λ be a set of local rules for $\overline{\sigma}$. A Λ-*path* of U is a sequence $R = (R_1, \ldots, R_k)$ of pointed letters of U such that:

- there exists an initial rule $\lambda^* \in \Lambda$ such that $I(\lambda^*) = \{R_1\}$;
- for $i = 1 \ldots k-1$, there exist an extension rule $\lambda_i \in \Lambda$ and $x_i \in \mathbb{Z}^n$ such that $E(\lambda_i) = \{L_i, L_i'\}$ with $R_i = L_i + x_i$ and $R_{i+1} = L_i' + x_i$.

One then defines by induction a map denoted by $(\overline{\sigma}, \Lambda, R)$ on the letters of R (see Fig. 1):

- $(\overline{\sigma}, \Lambda, R)(R_1) = \lambda^*(R_1)$;
- for $i = 1 \ldots k-1$, $(\overline{\sigma}, \Lambda, R)(R_{i+1}) = \lambda_i(L_i') + v(\lambda_i(L_i), (\overline{\sigma}, \Lambda, R)(R_i))$.

Notice that, when computing the action of a substitution σ on a word, we proceed in the same way: the image by σ of the first letter of the word (here seen as a path) has a specified position (here given by an initial rule), while the position of the image of a letter follows, by induction, from the position of the concatenation of the images of the previous letters (here, we use extension rules to do that). We then define the action of $\overline{\sigma}$ on pointed patterns:

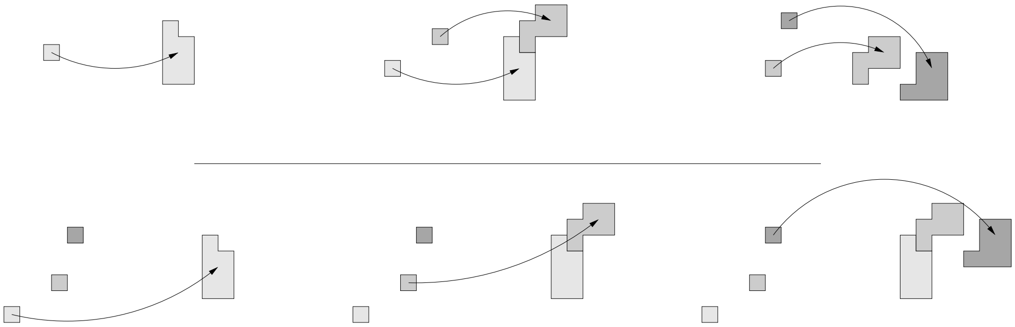


Figure 1: Top: from left to right, an initial rule and two extension rules; bottom: computation of the image of a path using successively the three previous local rules.

Definition 1.4 Let Λ be a set of local rules for $\overline{\sigma}$ and U be a pointed pattern. The set Λ is said to *cover* U if any pointed letter of U belongs to a Λ-path of U and is said to be *consistent* on U if for any two Λ-paths R and R' of U which both contain a pointed letter L, $(\overline{\sigma}, \Lambda, R)(L) = (\overline{\sigma}, \Lambda, R')(L)$.

If Λ covers U and is consistent on U, one then defines the action of $\overline{\sigma}$ endowed with the set of local rules Λ, denoted by $(\overline{\sigma}, \Lambda)$, as follows:

$$(\overline{\sigma}, \Lambda)(U) = \bigcup \{(\overline{\sigma}, \Lambda, R)(L) \mid R \text{ is a } \Lambda\text{-path of } U \text{ and } L \in R\}\,.$$

Thus, $(\overline{\sigma}, \Lambda)$ is our notion of multidimensional substitution on pointed patterns. It can be shown that it generalizes the substitutions on words as well as the multidimensional substitutions described in [15]. The possibilities are much larger, but it is in general not easy to obtain sets of local rules that are consistent on a set of pointed patterns and cover this set: the next section presents a way to obtain such sets of local rules.

2 Local rules derived from a global rule

Let $\overline{\sigma}$ be a non-pointed substitution and $\mathcal{H}$ be a set of pointed patterns. We are here interested in a generic way to obtain sets of local rules for $\overline{\sigma}$ that cover $\mathcal{H}$ and are consistent on it (that is, that cover any pointed pattern of $\mathcal{H}$ and are consistent on any of them). We derive such sets of local rules from *global rules*:

Definition 2.1 A *global rule* on $\mathcal{H}$ for $\overline{\sigma}$ is a map Γ defined on the set of pointed letters $\{L \in U \mid U \in \mathcal{H}\}$ such that:

- a pointed letter L is mapped to a realization of $\overline{\sigma}(\overline{L})$;
- pointed letters with distinct locations are mapped to disjoint pointed patterns.

Let us denote by $d(L, L')$ the distance $\sum |x_i - x'_i|$ between the locations (x_i) and (x'_i) of L and L'. We introduce a notion of weak connexity:

Definition 2.2 The *span* between two pointed letters L and L' of $U \in \mathcal{H}$, denoted by $\mathrm{sp}(L, L')$, is the smallest integer D such that there exists a sequence $(L_1 = L, L_2, \ldots, L_k = L')$ of pointed letters of U which verifies: $\forall j$, $d(L_j, L_{j+1}) \leq D$. The spans of U and $\mathcal{H}$ are then defined by:

$$\mathrm{sp}(U) = \sup_{L,L' \in U} \mathrm{sp}(L, L') \qquad \text{and} \qquad \mathrm{sp}(\mathcal{H}) = \sup_{U \in \mathcal{H}} \mathrm{sp}(U).$$

For example, $\mathrm{sp}(U) = 1$ if and only if U is 4-connected. Let us now derive a set of local rules from a global rule:

Definition 2.3 Let H_0 be a pointed pattern and Γ a global rule on $\mathcal{H}$ for $\overline{\sigma}$. A set Λ of local rules for $\overline{\sigma}$ is said to be *derived* from $(\mathcal{H}, H_0, \Gamma)$ if it verifies:

1. if λ^* is an initial rule of Λ with $I(\lambda^*) = \{L\}$, then $L \in H_0$ and $\lambda^*(L) = \Gamma(L)$;
2. if λ is an extension rule of Λ with $E(\lambda) = \{L, L'\}$, then $d(L, L') \leq \mathrm{sp}(\mathcal{H})$, $\lambda(L) = \Gamma(L)$ and $\lambda(L') = \Gamma(L')$;
3. if λ and λ' are extension rules of Λ, then $E(\lambda)$ and $E(\lambda')$ are not congruent.

Such derived sets of locals rules have interesting properties:

Proposition 2.4 *If H_0 is finite and $sp(\mathcal{H})$ is bounded, then any set of local rules derived from $(\mathcal{H}, H_0, \Gamma)$ is finite.*

Proof Let Λ be derived from $(\mathcal{H}, H_0, \Gamma)$. There is no more than $|H_0|$ initial rules in Λ. There are $|\mathcal{A}|^{|(\mathrm{sp}(\mathcal{H})+1)^n/\mathbb{Z}^n|}$ non-congruent pointed patterns $\{L, L'\}$ that verify $d(L, L') \leq \mathrm{sp}(\mathcal{H})$: it follows that there is a finite number of extension rules in Λ. Thus, Λ is finite. □

Definition 2.5 A global rule Γ on $\mathcal{H}$ is said *context-free* if, for $U \in \mathcal{H}$, $L, L' \in U$ and $x \in \mathbb{Z}^n$ such that $L + x, L' + x \in U$, one has:

$$v(\Gamma(L), \Gamma(L + x)) = v(\Gamma(L'), \Gamma(L' + x)).$$

We present examples of such global rules in Section 3.

Proposition 2.6 *If Γ is a context-free global rule on $\mathcal{H}$, then any set of local rules derived from $(\mathcal{H}, H_0, \Gamma)$ is consistent on $\mathcal{H}$.*

Proof Suppose that Γ is context-free, and let Λ be a set of local rules derived from $(\mathcal{H}, H_0, \Gamma)$. Let $R = (R_1, \ldots, R_k)$ be a Λ-path of $U \in \mathcal{H}$. Let us prove by induction that for all i, $(\overline{\sigma}, \Lambda, R)(R_i) = \Gamma(R_i)$. Since R is a Λ-path, there exists an initial rule $\lambda^* \in \Lambda$ such that $I(\lambda^*) = \{R_1\}$, and since Λ is derived from $(\mathcal{H}, H_0, \Gamma)$, $(\overline{\sigma}, \Lambda_k, R)(R_1) = \lambda^*(R_1) = \Gamma(R_1)$. Suppose now that

$(\overline{\sigma}, \Lambda_k, R)(R_i) = \Gamma(R_i)$. According to Definition 1.3, there exists an extension rule $\lambda_i \in \Lambda$ and $x_i \in \mathbb{Z}^n$ such that $E(\lambda_i) = \{L_i, L'_i\}$ with $R_i = L_i + x_i$ and $R_{i+1} = L'_i + x_i$, and $(\overline{\sigma}, \Lambda_k, R)(R_{i+1}) = \lambda(L'_i) + v(\lambda(L_i), (\overline{\sigma}, \Lambda_k, R)(R_i))$. But Λ is derived from $(\mathcal{H}, H_0, \Gamma)$, hence $\lambda(L_i) = \Gamma(L_i)$ and $\lambda(L'_i) = \Gamma(L'_i)$. Moreover, $(\overline{\sigma}, \Lambda_k, R)(R_i) = \Gamma(R_i) = \Gamma(L_i + x_i)$. Thus, $(\overline{\sigma}, \Lambda_k, R)(R_{i+1}) = \Gamma(L'_i) + v(\Gamma(L_i), \Gamma(L_i + x_i)$. Finally, since Γ is context-free, $(\overline{\sigma}, \Lambda_k, R)(R_{i+1}) = \Gamma(L'_i) + v(\Gamma(L'_i), \Gamma(L'_i + x_i)) = \Gamma(L'_i + x_i) = \Gamma(R_{i+1})$. It yields that Λ is consistent on $\mathcal{H}$. □

Proposition 2.7 *If H_0 intersects any pointed pattern of $\mathcal{H}$, then there exist sets of local rules derived from $(\mathcal{H}, H_0, \Gamma)$ that cover $\mathcal{H}$.*

Proof Let us define $\mathcal{E} = \{\{L, L'\} \mid L, L' \in U,\ U \in \mathcal{H} \text{ and } d(L, L') \leq \mathrm{sp}(\mathcal{H})\}$, and let $\mathcal{E}'$ be a maximal subset of $\mathcal{E}$ that does not contain congruent pointed patterns. Let Λ be the set of the following local rules:

- for each $L \in H_0$, the initial rule λ^* defined on $I(\lambda^*) = \{L\}$ by $\lambda^*(L) = \Gamma(L)$;
- for each $\{L, L'\} \in \mathcal{E}'$, the extension rule λ defined on $E(\lambda) = \{L, L'\}$ by $\lambda(L) = \Gamma(L)$ and $\lambda(L') = \Gamma(L')$.

One easily checks that Λ is derived from $(\mathcal{H}, H_0, \Gamma)$. Let us prove that Λ covers $\mathcal{H}$. Let $U \in \mathcal{H}$ and $L' \in U$. Since H_0 intersects any pointed pattern of $\mathcal{H}$, there exists $L \in U \cup H_0$. By definition, there also exists a sequence of pointed letters $(L_1 = L, L_2, \ldots, L_k = L')$ such that $\forall i,\ d(L_i, L_{i+1}) \leq \mathrm{sp}(\mathcal{H})$. Then, for all i there exists $x_i \in \mathbb{Z}^n$ such that $\{L_i, L_{i+1}\} + x_i \in \mathcal{E}'$, and there exists an initial rule of Λ defined on $\{L_1\}$. It yields that $(L_1, \ldots, L_k)$ is a Λ-path which contains L'. Thus, Λ covers $\mathcal{H}$ □

We can resume the previous propositions in the following theorem:

Theorem 2.8 *Let Γ be a context-free global rule on $\mathcal{H}$ for $\overline{\sigma}$. If $sp(\mathcal{H})$ is bounded and if $H_0 \in \mathcal{P}$ is a finite pointed pattern intersecting any pointed pattern of $\mathcal{H}$, then one can derive from $(\mathcal{H}, H_0, \Gamma)$ a finite set of local rules that is consistent on $\mathcal{H}$ and covers it.*

We thus have a way to derive, from a context-free global rule, local rules consistent on a given set of pointed pattern a covering this set. This result is applied in the next section to a particular type of context-free global rule.

3 Sturmian hyperplane sequences and algebraicity

We first briefly resume the notion of *generalized substitution* (see e.g. [4, 5, 14]). Let $\vec{e_1}, \ldots, \vec{e_n}$ denote the canonical basis of $\mathbb{R}^n$ and let $\langle ., . \rangle$ denote the canonical scalar product on $\mathbb{R}^n$.

A *face* (x, i^*), for $x \in \mathbb{Z}^n$ and $i \in \{1, \ldots, n\}$ is defined by:

$$(x, i^*) = \{x + \sum_{j \neq i} r_j \vec{e_j} \mid 0 \leq r_j \leq 1\}.$$

Such faces generate the $\mathbb{Z}$-module of the formal sums of weighted faces $\mathcal{G} = \{\sum m_{x,i}(x, i^*) \mid m_{x,i} \in \mathbb{Z}\}$, on which the lattice $\mathbb{Z}^n$ acts by translation: $y + (x, i^*) = (y + x, i^*)$. Faces are used to approximate hyperplanes of $\mathbb{R}^n$:

Definition 3.1 Let $\vec{\alpha} \in \mathbb{R}^n_+$, $\vec{\alpha} \neq 0$. The hyperplane $\mathcal{P}_{\vec{\alpha}}$ of $\mathbb{R}^n$ is defined by:

$$\mathcal{P}_{\vec{\alpha}} = \{x \in \mathbb{R}^n \mid \langle x, \vec{\alpha} \rangle = 0\}.$$

The *stepped hyperplane* $\mathcal{S}_{\vec{\alpha}}$ associated to $\mathcal{P}_{\vec{\alpha}}$ is defined by:

$$\mathcal{S}_{\vec{\alpha}} = \{(x, i^*) \mid \langle x, \vec{\alpha} \rangle > 0 \text{ and } \langle x - \vec{e_i}, \vec{\alpha} \rangle \leq 0\},$$

and a *patch* of $\mathcal{S}_{\vec{\alpha}}$ is a finite subset of the set of faces of $\mathcal{S}_{\vec{\alpha}}$.

Notice that a patch of $\mathcal{S}_{\vec{\alpha}}$ belongs to the $\mathbb{Z}$-module $\mathcal{G}$, but is geometric, that is, without multiple faces. Let us recall that the *incidence matrix* M_σ of a substitution on words σ gives at position (i, j) the number of occurences of the letter i in the word $\sigma(j)$. If $detM_\sigma = \pm 1$, then σ is said *unimodular*.

Definition 3.2 The *generalized substitution* associated to the unimodular substitution σ is the endomorphism Θ_σ of $\mathcal{G}$ defined by:

$$\begin{cases} \forall i \in \mathcal{A}, & \Theta_\sigma(0, i^*) = \sum_{j=1}^{3} \sum_{s:\sigma(j)=p \cdot i \cdot s} \left(M_\sigma^{-1}(f(s)), j^*\right), \\ \forall x \in \mathbb{Z}^3,\ \forall i \in \mathcal{A}, & \Theta_\sigma(x, i^*) = M_\sigma^{-1} x + \Theta_\sigma(0, i^*), \\ \forall \sum m_{x,i}(x, i^*) \in \mathcal{G}, & \Theta_\sigma\left(\sum m_{x,i}(x, i^*)\right) = \sum m_{x,i} \Theta_\sigma(x, i^*), \end{cases}$$

where $f(w) = (|w|_1, |w|_2, |w|_3)$ and $|w|_i$ is the number of occurences of the letter i in w.

The following type of substitution is particularly interesting:

Definition 3.3 A substitution σ is of *Pisot* type if its incidence matrix M_σ has eigenvalues $\lambda, \mu_1, \ldots, \mu_{n-1}$ satisfying $0 < |\mu_i| < 1 < \lambda$. The generalized substitution Θ_σ is then also said of Pisot type.

Indeed, the following result is proved in [4, 5]:

Proposition 3.4 ([4, 5]) *If σ is of Pisot type and if $\vec{\alpha}$ is a* left *eigenvector of M_σ for the dominant eigenvalue λ, then $\Theta_\sigma(\mathcal{S}_{\vec{\alpha}}) \subset \mathcal{S}_{\vec{\alpha}}$ and Θ_σ maps distinct faces of the stepped hyperplane $\mathcal{S}_{\vec{\alpha}}$ to disjoint patches of $\mathcal{S}_{\vec{\alpha}}$.*

The stepped hyperplane $\mathcal{S}_{\vec{\alpha}}$ is called the *invariant hyperplane* of Θ_σ. It is also proved in [11]:

Proposition 3.5 ([11]) *If the modified Jacobi-Perron algorithm ([8]) yields a purely periodic (resp. eventually periodic) continued fraction expansion for $\vec{\alpha} \in \mathbb{R}^n$, then the stepped hyperplane $\mathcal{S}_{\vec{\alpha}}$ is a fixed point (resp. the image by a generalized substitution of a fixed point) of a generalized substitution of Pisot type.*

We then define hyperplane sequences, mapping stepped hyperplanes of $\mathbb{R}^n$ to $(n-1)$-dimensional sequences over the alphabet $\{1,\ldots,n\}$. The following proposition (proved in Appendix) resumes a result given in [2,3]:

Proposition 3.6 *Let $\mathcal{V}_{\vec{\alpha}} \subset \mathbb{Z}^n$ be the set of the vertices that belong to the faces of $\mathcal{S}_{\vec{\alpha}}$. Let $v_{\vec{\alpha}}$ and $\pi_{\vec{\alpha}}$ be the maps defined respectively on $\mathcal{S}_{\vec{\alpha}}$ and $\mathcal{V}_{\vec{\alpha}}$ by:*

$$v_{\vec{\alpha}}(x,i^*) = x + \vec{e_1} + \ldots + \vec{e_{i-1}} \quad \textit{and} \quad \pi_{\vec{\alpha}}(x_1,\ldots,x_n) = (x_1 - x_n, \ldots, x_{n-1} - x_n).$$

Then, $v_{\vec{\alpha}}$ (resp. $\pi_{\vec{\alpha}}$) is a bijection from $\mathcal{S}_{\vec{\alpha}}$ onto $\mathcal{V}_{\vec{\alpha}}$ (resp. from $\mathcal{V}_{\vec{\alpha}}$ onto $\mathbb{Z}^{n-1}$).

Let $\phi_{\vec{\alpha}}$ be defined on $\mathcal{S}_{\vec{\alpha}}$ by $\phi_{\vec{\alpha}}(x,i^*) = (\pi_{\vec{\alpha}}(v_{\vec{\alpha}}(x,i^*)), i)$: it maps bijectively the faces of $\mathcal{S}_{\vec{\alpha}}$ to the letters of a $(n-1)$-dimensional sequence over $\{1,\ldots,n\}$. Notice that **not all** these $(n-1)$-dimensional sequences over $\{1,\ldots,n\}$ correspond to a stepped hyperplane. We thus introduce the following definition:

Definition 3.7 An *hyperplane sequence* is an $(n-1)$-dimensional sequence over $\{1,\ldots,n\}$ defined, for $\vec{\alpha} \in \mathbb{R}^n$, by $\phi_{\vec{\alpha}}(\mathcal{S}_{\vec{\alpha}})$. One denotes by $\mathcal{H}_{\vec{\alpha}}$ such an hyperplane sequence. Moreover, if $\vec{\alpha} = (\alpha_1,\ldots,\alpha_n)$ is such that $1,\alpha_1,\ldots,\alpha_n$ are linearly independent over $\mathbb{Q}$, then $\mathcal{H}_{\vec{\alpha}}$ is called a *Sturmian hyperplane sequence.*

For $n = 2$, Sturmian hyperplane sequences are nothing but Sturmian sequences over $\{1,2\}$ (see [12]), and for $n = 3$, one retrieves the notion of two-dimensional Sturmian sequence of [7]. Notice that an hyperplane sequence $\mathcal{H}_{\vec{\alpha}}$ is defined on the whole $\mathbb{Z}^{n-1}$: it yields $\mathrm{sp}(\mathcal{H}_{\vec{\alpha}}) = 1$. Let us now derive, from generalized substitution, context-free global rules on hyperplane sequences:

Proposition 3.8 *Let σ be a Pisot unimodular substitution on words over $\{1,\ldots,n\}$. Let Θ_σ be the associated generalized substitution, and $\mathcal{S}_{\vec{\alpha}}$ its invariant stepped hyperplane. Let $\mathcal{H}_{\vec{\alpha}} = \phi_{\vec{\alpha}}(\mathcal{S}_{\vec{\alpha}})$. We set $\mathcal{L} = \mathbb{Z}^{n-1} \times \{1,\ldots,n\}$ and define:*

$$\Gamma_\sigma = \phi_{\vec{\alpha}} \circ \Theta_\sigma \circ \phi_{\vec{\alpha}}^{-1} \qquad \textit{and} \qquad \overline{\sigma^*} \ : \ \overline{(0,i)} \in \overline{\mathcal{L}} \mapsto \overline{\Gamma_\sigma(0,i)} \in \overline{\mathcal{P}}.$$

Then, Γ_σ is a context-free global rule on $\mathcal{H}_{\vec{\alpha}}$ for the non-pointed substitution $\overline{\sigma^*}$.

Proof For $(x, i) \in \mathcal{H}_{\vec{\alpha}}$ and $y \in \mathbb{Z}^{n-1}$, one computes:

$$\Gamma_\sigma((x,i) + y) = \Gamma_\sigma(x,i) + \pi_{\vec{\alpha}}(M_\sigma^{-1}\pi_{\vec{\alpha}}^{-1}(y)).$$

It follows that $\overline{\Gamma_\sigma(x,i)} = \overline{\Gamma_\sigma(0,i)} = \overline{\sigma^*}\left(\overline{(0,i)}\right)$. Moreover, since Θ_σ maps distinct faces of $\mathcal{S}_{\vec{\alpha}}$ to disjoint patches of $\mathcal{S}_{\vec{\alpha}}$ (see Proposition 3.4) and since $\phi_{\vec{\alpha}}$ maps bijectively the faces of $\mathcal{S}_{\vec{\alpha}}$ to the letters of $\mathcal{H}_{\vec{\alpha}}$, $\Gamma_\sigma = \phi_{\vec{\alpha}} \circ \Theta_\sigma \circ \phi_{\vec{\alpha}}^{-1}$ maps letters with distinct locations to disjoint pointed patterns. Thus, Γ_σ is a global rule on $\mathcal{H}_{\vec{\alpha}}$ for $\overline{\sigma^*}$.

Then, if $(x, i) \in \mathcal{H}_{\vec{\alpha}}$, $(x', i) \in \mathcal{H}_{\vec{\alpha}}$ and $y \in \mathbb{Z}^{n-1}$, one has:

$$v(\Gamma_\sigma(x,i), \Gamma_\sigma((x,i)+y)) = \pi_{\vec{\alpha}}(M_\sigma^{-1}\pi_{\vec{\alpha}}^{-1}(y)) = v(\Gamma_\sigma(x',i), \Gamma_\sigma((x',i)+y)).$$

Hence Γ_σ is context-free, according to Definition 2.5. □

Finally, combining Theorem 2.8 and Proposition 3.5 and 3.8, we obtain:

Theorem 3.9 *If the modified Jacobi-Perron algorithm ([8]) yields a purely periodic (resp. eventually periodic) continued fraction expansion for $\vec{\alpha} \in \mathbb{R}^n$, then the Sturmian hyperplane sequence $\mathcal{H}_{\vec{\alpha}}$ is a fixed point (resp. the image by a multidimensional substitution of a fixed point) of a multidimensional substitution.*

This result can thus be seen as a multidimensional generalization of the algebraic characterization resumed in the introduction, though it provides only a sufficient condition for a Sturmian hyperplane sequence to be a fixed point of a multidimensional substitution or the image by a multidimensional substitution of such a fixed point. In fact, the proof of the algebraic characterization resumed in the introduction uses the notion of *return words* of [10]. This notion has already been generalized, in terms of tilings, in [13]: it thus gives us a possible way to achieve the characterization of Theorem 3.9.

Example 3.10 Let σ be the classic substitution defined on $\{1, 2, 3\}$ by $\sigma(1) = 13$, $\sigma(2) = 1$ and $\sigma(3) = 2$. One computes:

$$M_\sigma^{-1} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \text{and} \qquad \Theta_\sigma \ : \ \begin{array}{l} (0,1^*) \mapsto ((1,-1,0),1^*) + (0,2^*) \\ (0,2^*) \mapsto (0,1^*) \\ (0,3^*) \mapsto (0,2^*) \end{array},$$

which yields the non-pointed substitution:

$$\overline{\sigma^*} \ : \ \overline{1_{0,0}} \mapsto \overline{\{1_{0,0}, 2_{0,1}\}}, \qquad \overline{2_{0,0}} \mapsto \overline{\{3_{0,0}\}}, \qquad \overline{3_{0,0}} \mapsto \overline{\{1_{0,0}\}},$$

which one can also represent as follows:

$$\overline{\sigma^*} \ : \ 1 \mapsto \begin{matrix} 2 \\ 1 \end{matrix}, \quad 2 \mapsto 3, \quad 3 \mapsto 1.$$

Let us define $\mathcal{H} = \{\Gamma_\sigma^n((0,0),1), n \geq 1\}$. One can prove in this particular case that $\mathrm{sp}(\mathcal{H}) = 1$. Thus, one can compute (Theorem 2.8) a finite set of local rules that covers $\mathcal{H}$ and is consistent on it. One obtains for example the initial rule defined by:

$$\lambda^* \ : \ ((0,0),1) \mapsto \{((0,0),1),((0,1),2)\},$$

and five extension rules, represented as follows (the bolded letters are mapped to the bolded letters, so the information about *relative* locations is still conserved):

$$\lambda_1 \ : \ \begin{array}{c} 2 \\ \mathbf{1} \end{array} \mapsto \begin{array}{cc} & \mathbf{2} \\ 3 & \mathbf{1} \end{array}, \qquad \lambda_2 \ : \ 3 \ \ \mathbf{1} \mapsto \begin{array}{c} \mathbf{2} \\ \mathbf{1} \\ 1 \end{array}, \qquad \lambda_3 \ : \ \begin{array}{c} 1 \\ \mathbf{1} \end{array} \mapsto \begin{array}{cc} & \mathbf{2} \\ 2 & \mathbf{1} \\ 1 & \end{array},$$

$$\lambda_4 \ : \ 2 \ \ \mathbf{1} \mapsto \begin{array}{c} \mathbf{2} \\ \mathbf{1} \\ 3 \end{array}, \qquad \lambda_5 \ \begin{array}{c} \mathbf{1} \\ 3 \end{array} \mapsto \begin{array}{cc} \mathbf{1} & 1 \\ \mathbf{2} & \end{array} .$$

For example, computing the sequence $(\overline{\sigma^*}, \{\lambda^*, \lambda_1, \ldots, \lambda_5\})^n((0,0),1)$ for $n = 1, \ldots, 7$ gives (the letter with location $(0,0)$ is bolded):

$$\mathbf{1} \mapsto \begin{array}{c} 2 \\ \mathbf{1} \end{array} \mapsto \begin{array}{cc} & 2 \\ 3 & \mathbf{1} \end{array} \mapsto \begin{array}{cc} & 2 \\ 3 & \mathbf{1} \\ & 1 \end{array} \mapsto \begin{array}{ccc} & & 2 \\ & 2 & 1 \\ 3 & \mathbf{1} & \\ & 1 & \end{array} \mapsto \begin{array}{cccc} & 2 & & 2 \\ 3 & 1 & 2 & 1 \\ & 3 & \mathbf{1} & \\ & & 1 & \end{array} \mapsto \begin{array}{cccc} & 2 & & 2 \\ 3 & 1 & 2 & 1 \\ & 3 & \mathbf{1} & \\ & 2 & 1 & \\ 3 & 1 & & \\ & 1 & & \end{array} \mapsto \ldots$$

We can in this way generate arbitrarely large patches of the hyperplane sequence $\mathcal{H}_{\vec{\alpha}}$, where $\vec{\alpha}$ is a left eigenvector of M_σ. Moreover, $\mathcal{H}_{\vec{\alpha}}$ is a fixed-point of this multidimensional substitution.

Acknowledgements. We would like to thank Valérie Berthé and Pierre Arnoux for many useful suggestions.

References

[1] J.-P. Allouche, J. O. Shallit, *Automatic sequences: Theory and Applications*, Cambridge University Press, 2002.

[2] P. Arnoux, V. Berthé, S. Ito, *Discrete planes, $\mathbb{Z}^2$-actions, Jacobi-Perron algorithm and substitutions.* Ann. Inst. Fourier (Grenoble) **52** (2002), 1001–1045.

[3] P. Arnoux, V. Berthé, A. Siegel, *Two-dimensional iterated morphisms and discrete planes.* Theoret. Comput. Sci. **319** no. 1-3 (2004), 145–176.

[4] P. Arnoux, S. Ito, *Pisot substitutions and Rauzy fractals.* Bull. Belg. Math. Soc. Simon Stevin **8** no. 2 (2001), 181–207.

[5] P. Arnoux, S. Ito, Y. Sano, *Higher dimensional extensions of substitutions and their dual maps.* J. Anal. Math. **83** (2001), 183–206.

[6] V. Berthé, C. Holton, L. Q. Zamboni, *Initial powers of Sturmian sequences.* Acta Arithmetica, to appear.

[7] V. Berthé, L. Vuillon, *Tilings and rotations on the torus: a two-dimensional generalization of Sturmian sequences.* Discrete Math. **223** (2000), 27–53.

[8] A. J. Brentjes, *Multi-dimensional continued fraction algorithms.* Mathematical Centre Tracts 145, Matematisch Centrum, Amsterdam, 1981.

[9] D. Crisp, W. Moran, A. Pollington, P. Shiue, *Substitution invariant cutting sequences.* J. Théor. Nombres Bordeaux **5** (1993), 123–137.

[10] F. Durand, *A characterization of substitutive sequences using return words.* Inventiones Math. **132** (1998), 179–188.

[11] T. Fernique, *Bidimensional Sturmian Sequences and Substitutions.* Research rapport 05024 (2005), LIRMM.

[12] M. Lothaire, *Algebraic combinatorics on words*, Cambridge University Press, 2002.

[13] N. Priebe, *Towards a characterization of self-similar tilings in terms of derived Vorono tessellations.* Geom. Dedicata **79** (2000), 239–265.

[14] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics.* Lecture Notes in Math. **1794** (2002), Springer Verlag.

[15] O. Salon, *Suites automatiques à multi-indices et algébricité.* C. R. Acad. Sci. Paris Sér. I Math **305** (1987), 501–504.

Appendix

Proof (of Proposition 3.6) Let (x, i^*) and (y, j^*) be two faces of $\mathcal{S}_{\vec{\alpha}}$ such that $v_{\vec{\alpha}}(x, i^*) = v_{\vec{\alpha}}(y, j^*)$. If $i < j$, then $x = y + \vec{e_i} + \ldots + \vec{e_{j-1}}$, and $\langle x - \vec{e_i}, \vec{\alpha}\rangle = \langle (y + \vec{e_{i+1}} + \ldots + \vec{e_{j-1}}, \vec{\alpha}\rangle = \langle y, \vec{\alpha}\rangle + \langle \vec{e_{i+1}} + \ldots + \vec{e_{j-1}}, \vec{\alpha}\rangle$. Since $(y, j^*) \in \mathcal{S}_{\vec{\alpha}}$, $\langle y, \vec{\alpha}\rangle > 0$. Moreover, $\langle \vec{e_{i+1}} + \ldots + \vec{e_{j-1}}, \vec{\alpha}\rangle \geq 0$. Thus, $i < j$ would yield $\langle x - \vec{e_i}, \vec{\alpha}\rangle > 0$, what would contradict $(x, i^*) \in \mathcal{S}_{\vec{\alpha}}$. Similarly, $i > j$ is impossible. Hence $i = j$, and $x = y$ follows. It proves that $v_{\vec{\alpha}}$ is one-to-one from $\mathcal{S}_{\vec{\alpha}}$ to $\mathcal{V}_{\vec{\alpha}}$.

If $y \in \mathcal{V}_{\vec{\alpha}}$, then there exist $(x, i^*) \in \mathcal{S}_{\vec{\alpha}}$ and $I \subset \{1, \ldots, n\}$, $i \notin I$, such that $y = x + \sum_{j \in I} \vec{e_j}$. Let us denote $f : k \mapsto \langle x + \sum_{j\in I} \vec{e_j} - \vec{e_1} - \ldots - \vec{e_k}, \vec{\alpha}\rangle$. One has:

$$f(0) = \langle x, \vec{\alpha}\rangle + \sum_{j \in I} \langle \vec{e_j}, \vec{\alpha}\rangle > 0, \qquad f(n) = \langle x - \vec{e_i}, \vec{\alpha}\rangle - \sum_{j \notin I, j \neq i} \langle \vec{e_j}, \vec{\alpha}\rangle \leq 0,$$

and f is decreasing. Let k_0 such that $f(k_0 - 1) > 0$ and $f(k_0) \leq 0$. Let $y_0 = y - \vec{e_1} - \ldots - \vec{e_{k_0-1}}$. Then, $\langle y_0, \vec{\alpha}\rangle = f(k_0 - 1) > 0$, and $\langle y_0 - \vec{e_{k_0}}, \vec{\alpha}\rangle = f(k_0) \leq 0$. Thus, $(y_0, k_0^*) \in \mathcal{S}_{\vec{\alpha}}$. Since $v_{\vec{\alpha}}(y_0, k_0^*) = y$, it proves that $v_{\vec{\alpha}}$ is onto from $\mathcal{S}_{\vec{\alpha}}$ on $\mathcal{V}_{\vec{\alpha}}$.

Let us denote $\vec{\alpha}$ by $(\alpha_1, \ldots, \alpha_n)$. Recall that the α_i are positive and not all equal to zero. Let then $x = (x_1, \ldots, x_n) \in \mathcal{V}_{\vec{\alpha}}$ and $(x', i^*) = v_{\vec{\alpha}}^{-1}(x)$. One has $0 < \langle x', \vec{\alpha}\rangle \leq \langle \vec{e_i}, \vec{\alpha}\rangle = \alpha_i$. Thus:

$$0 < \sum_{j=1}^{n} x_j \alpha_j - \sum_{j=1}^{i-1} \alpha_j \leq \alpha_i.$$

Suppose now $\pi_{\vec{\alpha}}(x) = (y_1, \ldots, y_{n-1})$. The previous formula yields:

$$0 < \sum_{j=1}^{n-1} y_j \alpha_j + x_n \sum_{j=1}^{n} \alpha_j \leq \sum_{j=1}^{i-1} \alpha_j + \alpha_i \leq \sum_{j=1}^{n} \alpha_j,$$

and performing the division by $\sum_{j=1}^{n} \alpha_j > 0$, it then gives:

$$0 < \frac{\sum_{j=1}^{n-1} y_j \alpha_j}{\sum_{j=1}^{n} \alpha_j} + x_n \leq 1,$$

that is, since $x_n \in \mathbb{Z}$:

$$x_n = 1 - \left\lceil \frac{\sum_{j=1}^{n-1} y_j \alpha_j}{\sum_{j=1}^{n} \alpha_j} \right\rceil.$$

Conversely, given $(y_1, \ldots, y_{n-1}) \in \mathbb{Z}^{n-1}$, setting $x_n \in \mathbb{Z}$ as above and then, for $i = 1 \ldots n-1$, $x_i = y_i + x_n$ yields $\pi_{\vec{\alpha}}(x_1, \ldots, x_n) = (y_1, \ldots, y_{n-1})$. Thus, $\pi_{\vec{\alpha}}$ is a bijection from $\mathcal{V}_{\vec{\alpha}}$ to $\mathbb{Z}^{n-1}$ (and the proof provides an explicit formula for $\pi_{\vec{\alpha}}^{-1}$). □

An automaton that recognizes the base of a semiretract[*]

Wit Foryś, Tomasz Krawczyk[†]

1 Introduction

Semiretracts of free monoids were introduced by Jim Anderson [1] and then were investigated in [2], [3], [4], [5], [6], [7], [8], [9], [10], [11].

We present an algorithmic approach to the problem of finding the base of a semiretract.

2 Basic Notions And Definitions

Let A^* denote a free monoid generated by a finite set A. A retraction $r : A^* \longrightarrow A^*$ is a morphism such that $r \circ r = r$. A retract of A^* is an image of A^* by a retraction. A semiretract of A^* is the intersection of a family of retracts of A^*. A word $w \in A^*$ is called a key-word if there is at least one letter in A that occurs exactly once in w and the letter is called a key of w. A set $C \subset A^*$ of key-words is called a key-code if there exists an injection (called key-injection) $key : C \longrightarrow A$ such that

1. for any $w \in C$, $key(w)$ is a key of w,
2. the letter $key(w)$ occurs in no word of C other than w itself.

Obviously any key-code is a code. For any key-word w in a key-code C and fixed mapping key we use the notation $w = l(a)ar(a)$ where $a = key(w)$ is the key of w and $l(a), r(a)$ denote a suitable prefix and sufix of w. Given a key-code C and a fixed key-injection key the set of all keys of words in C is denoted by $key(C)$.

The following characterization of retracts is due to T. Head [10].

Theorem 2.1 *$R \subset A^*$ is a retract of A^* if and only if $R = C^*$ where C is a key-code.*

In [2] T.Anderson proved the following, basic for our considerations theorem.

[*] Supported by KBN grant no 3 T11C 010 27

[†] Jagiellonian University, Institute of Computer Science, Nawojki 11, 30-072 Krakow, Poland, `forysw@ii.uj.edu.pl`, `krawczyk@ii.uj.edu.pl` (corresponding author)

Theorem 2.2 *Let $S = \cap_{i=1}^{n} R_i$ be a semiretract given by retracts R_i with key-codes $C_i \subset A^*$ for $i = 1, \dots, n$. There exist key-codes $D_i \subset A^*$ such that*

1. *$S = \cap_{i=1}^{n} T_i$ where retracts $T_i = D_i^*$ for $i = 1, \dots, n$*
2. *$key(D_1) = key(D_2) = \dots = key(D_n)$*

Hence any semiretract S is an intersection of a family of retracts generated by key-codes having the common set of keys.

An algorithm which produces such key-codes over the common set of keys is a subject of the paper.

For a key word $u = l(u)ar(u)$ with a as the key we enumerate all the positions of letters in u putting 0 for the key letter, numbering sequentially position to the right from the key by positive and to the left by negative integers. Hence any key word u defines a discrete interval of positions $[-p, r]$. The first letter of u has the position $-p$, the key of u is numbered by 0 and the last letter is numbered by r. In the sequel we will use the notation $u = a_{-p} \dots a_0 \dots a_r$ from which it is easy to find the key, the prefix $l(u)$ and the sufix $r(u)$ of the word u.

Any word $u \in A^*$ can be considered as a domino on a plane, where the domino consists of $\mid u \mid$ squares filled up in turn with the letters of u. Let $S = \cap_{i=1}^{n} R_i$ be a semiretract where $C_i \subset A^*$ for $i = 1, \dots, n$ denote key-codes of R_i. We consider any key-code C_i as a set of dominoes and assume that all dominoes in C_i are coloured by the colour i. Hence any domino is identified by a key-word u and a color $i \in \{1, \dots, n\}$ and so can be represented by a pair (i, u). The set of dominoes of key-words in C_i is denoted by C_i or by $\{i\} \times C_i$ if we want to point out the colour. For a semiretract S denote by V the set of all dominoes , that is

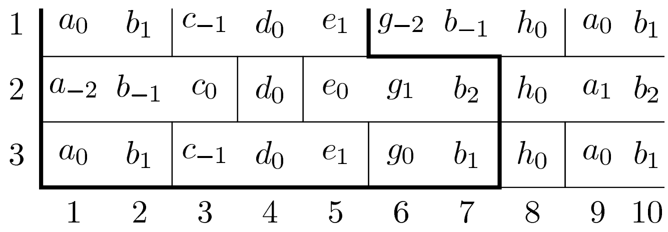
$$V = \bigcup_{i=1}^{n} \{i\} \times C_i.$$

A word w is in S if and only if w is in every R_i, that is the word w is expressible in the words of C_i for $i = 1, \dots, n$. It means that the following equalities are true

$$\begin{array}{ll} w = & u_1^1 .....u_{m_1}^1 \ (1) \\ w = & u_1^2 .....u_{m_2}^2 \ (2) \\ .... & .... \\ w = & u_1^n .....u_{m_n}^n \ (n) \end{array}$$

where $u_j^i \in C_i$ for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$.

Let $T(w)$ be a table of the dimension $\{1, \dots, n\} \times \{1, \dots, |w|\}$. The fact that w satisfies equalities (1)-(n) is equivalent to the possibility of tilling the table $T(w)$ by dominoes according to the rule that in $i-$th row we use $i-$th color domino-code and every square in the j-th column is filled up with j-th letter of the word w, for all $j \in \{1, \dots, \mid w \mid\}$. It means that all the entries $t_{i,j}$ of the table $T(w)$ are identical for a fixed $j \in \{1, \dots, |w|\}$ and $i \in \{1, \dots, n\}$. Any



	1	2	3	4	5	6	7	8	9	10
1	a_0	b_1	c_{-1}	d_0	e_1	g_{-2}	b_{-1}	h_0	a_0	b_1
2	a_{-2}	b_{-1}	c_0	d_0	e_0	g_1	b_2	h_0	a_1	b_2
3	a_0	b_1	c_{-1}	d_0	e_1	g_0	b_1	h_0	a_0	b_1

Figure 1: The table $T(abcdegbhab)$.

domino in the table $T(w)$ is identified by the color i, by the word $u \in C_i$ and by the position x of the key of word u counted from the left in the $i-$th row of $T(w)$. Hence any domino in the table $T(w)$ has its identification triple (i, u, x) where $i \in \{1, \ldots, n\}$, $u \in C_i$ and $x \in \mathcal{Z}$. Note that if $u = a_{-p} \ldots a_0 \ldots a_r$ then the domino (i, u, x) covers the entries (squares) from the set $\{(i, j) \mid x - p \leq j \leq x + r\}$. The set of squares covered by a domino (i, u, x) is denoted shortly by $[(i, u, x)]$.

Example 2.3 Consider the semiretract $\bigcap_{i=1}^{3} C_i^* \subset A^*$ and assume that

$$\begin{array}{lcl} \overline{a}b, c\overline{d}e, g\overline{b}h & \in & C_1, \\ ab\overline{c}, \overline{d}, \overline{e}gb, \overline{h}ab & \in & C_2, \\ \overline{a}b, c\overline{d}e, \overline{g}b, \overline{h} & \in & C_3. \end{array}$$

(the keys of the key words are overlined). Then the set $V = \bigcup_{i=1}^{3} \{i\} \times C_i$ contains dominoes $(1, a_0b_1)$, $(1, c_{-1}d_0e_1)$, $(1, g_{-1}b_0h_1)$, $(2, a_{-2}b_{-1}c_0)$, $(2, d_0)$, $(2, e_0g_1b_2)$, $(2, h_0a_1b_2)$, $(3, a_0b_1)$, $(3, c_{-1}d_0e_1)$, $(3, g_0b_1)$, $(3, h_0)$.

As the word $abcdegbhab$ can be factorized over codes C_1, C_2 and C_3, then $abcdegbhab$ is in $\bigcap_{i=1}^{3} C_i^*$. Hence we can tile the table $T(abcdegbhab)$ of dimension $\{1, 2, 3\} \times \{1, 2, \ldots, 10\}$ (see Figure 1) according to the rules given in the previous paragraph.

Definition 2.4 Let $A \subset V \times \mathbb{Z}$. A is a configuration of dominoes if A fulfills the following conditions:

1. the set of squares $\{[(i, u, x)] \mid (i, u, x) \in A\}$ consists of pairwise disjoint elements for any $i \in \{1, \ldots, n\}$ (the dominoes do not overlap in the table A),
2. for any fixed $j \in \mathbb{Z}$ there is no two squares (i_1, j) and (i_2, j) in A filled up with two different letters for $i_1, i_2 \in \{1, \ldots, n\}$.

Note that however dominoes do not overlap in a configuration A there are possible some gaps between them in A. If A is a configuration of dominoes, then we denote by

- $A^{\mapsto z}$ the shift of A by $z \in \mathbb{Z}$, that is

$$A^{\mapsto z} = \{(i, u, x + z) \mid (i, u, x) \in A\}$$

- $[A]$ the set of squares covered by dominoes from A

$$[A] = \bigcup \{[(i, u, x)] \mid (i, u, x) \in A\}$$

We say that a configuration A is connected if for any colour $i \in \{1, \ldots, n\}$ there are no gaps between dominoes coloured by i. It means that all dominoes from A coloured by i occupy squares indexed in A by $\{i\} \times [l, r]$ for some $l, r \in \mathbb{Z}$.

We say that two configurations A and B are equal with respect to the shift if there exist $z \in \mathbb{Z}$ such that $A^{\mapsto z} = B$.

3 Preliminary Results

Let S be a semiretract and consider a domino (i, u, x) that occurs in the table $T(w)$ for some word $w \in S$. Hence the domino (i, u, x) occurs in the $i-$th row of the table $T(w)$. Assume now that in the word u occurs a letter $key(v)$ which is the key letter for some $v \in C_j$ and in this row there exists a square $(i, z) \in [(i, u, x)]$ which is filled up with a letter $key(v)$. Since $key(v)$ occurs only once in the word $v \in C_j$ and in no other word from C_j, then the only domino in $j-$th row which covers the square (j, z) with a letter $key(v)$ is (j, v, z). Hence, the element (j, v, z) has to be in $T(w)$. In general, a domino (i, u, x) that occurs in $T(w)$ in a $i-$th row enforces in others rows an occurence of these dominoes that have as key-letters the letters occuring in u. To obtain a clear cut picture of those dependiences we introduce the following relation E and a multidigraph associated with S.

Definition 3.1 Let $C_i \subset A^*$ for $i = 1, \ldots, n$ denote key-codes of retracts R_i and let V be the set of all dominoes. Let $(i, u), (j, v) \in V$ be two dominoes such that $u = a_{-p} \ldots a_0 \ldots a_r$ and $v = b_{-s} \ldots b_0 \ldots b_t$. A triple $((i, u), z, (j, v))$ is in the relation $E \subset V \times \mathbf{Z} \times V$ if and only if $a_z = b_0$ for some $z \in \{-p, \ldots, 0, \ldots, r\}$.

We consider the relation E as the set of arrows between nodes in V labeled by integers. We use in the sequel the notation $(i, u) \rightarrow_z (j, v)$ for a triple $((i, u), z, (j, v))$ in E and say that (i, u) binds (j, v).

Definition 3.2 Let C_i be key-codes of retracts R_i for $i \in \{1, \ldots, n\}$ and let $S = \cap_{i=1}^n R_i$ denote a semiretract. A directed multigraph $G = (V, E)$ where $V = \cup_{i=1}^n \{i\} \times C_i$ is the set of all dominoes and E considered as a relation on V with integer labels is called a labeled multidigraph associated with S.

Example 3.3 As the letter $e \in A$ is the key of domino $(2, e_0 g_1 b_2)$, then

$$(1, d_0 e_1) \mapsto_1 (2, e_0 g_1 b_2) \text{ and } (3, c_{-1} d_0 e_1) \mapsto_1 (2, e_0 g_1 b_2)\,.$$

For pathes in $G = (V, E)$ we use the following notation.

Definition 3.4 There is a path $(i_0, v_0) \rightarrow^*_x (i_m, v_m)$ in a multidigraph $G = (V, E)$ associated with a semiretract S if there exist nodes $(i_0, v_0), \ldots, (i_m, v_m) \in V$ and integers $x_1, \ldots, x_m \in \mathbb{Z}$ such that

1. $(v_0, i_0) \rightarrow_{x_1} (v_1, i_1) \rightarrow_{x_2} \ldots.. \rightarrow_{x_m} (i_m, v_m)$,
2. $\sum_{i=1}^{m} x_i = x$.

Using introduced in the above notions our observations done at the begining of this section can be summarized as follows.

Fact 3.5 If (i, u, x) is in $T(w)$ for some $w \in S$, $x \in \{1, \ldots, |w|\}$ and $(i, u) \rightarrow_z (j, v)$ for some $z \in \mathbb{Z}$, then $(j, v, x + z)$ is in $T(w)$.

Definition 3.6 Let S denote a semiretract and a word $w \in S$. Let a domino $(i, u) \in V$ occurs in $T(w)$ at the position x, that is $(i, u, x) \in T(w)$ for some $x \in \{1, \ldots, |w|\}$. The set

$$B(i, u, x) = \{(j, v, x + z) \in T(w) \mid (i, u) \rightarrow^*_z (j, v),\ (j, v) \in V, z \in \mathbb{Z}\}$$

is called a neighbourhood of (i, u, x) in relation to $T(w)$.

In the other words, the neighbourhood $B(i, u, x)$ of (i, u, x) is a part of the table $T(w)$ containing the domino (i, u, x) itself and all dominoes from $T(w)$ that are binded with it.

Example 3.7 The neighbourhood $B(1, c\bar{d}e, 4)$ of dominoe $(1, c_{-1}d_0e_1, 4)$ contains all elements included in the borded area (see Figure 1).

Let us denote by $CC(G)$ the set of all strongly connected components $W \subset V$ of a multidigraph $G = (V, E)$. For any such component $W \in CC(G)$ we fix a node $(i, u) \in W$ and this node is called a representant of W. To express the fact that (i, u) represents W we write $W_{(i,u)}$. The following lemma points out a role of connected components of a multidigraph $G = (V, E)$ associated with a semiretract S.

Lemma 3.8 *Let $W_{(i,u)}$ be a strongly connected component in G associated with a semiretract S and represented by (i, u). Let a domino (i, u, x) occurs in a table $T(w)$ at the position x for some $w \in S$, $x \in \{1, \ldots, |w|\}$. Then*

1. *for any dominoes $(j_1, v_1), (j_2, v_2) \in W_{(i,u)}$ there exists exactly one integer $z \in \mathbb{Z}$ such that $(j_1, v_1) \rightarrow^*_z (j_2, v_2)$. If $(j_1, v_1) \rightarrow^*_z (j_2, v_2)$, then $(j_2, v_2) \rightarrow^*_{-z} (j_1, v_1)$,*
2. *if $(j, v) \in W_{(i,u)}$ and $(i, u) \rightarrow^*_z (j, v)$, then $(j, v, x + z) \in B(i, u, x)$ and the neighbourhoods $B(i, u, x)$ and $B(j, v, x + z)$ are equal;*

3. the sets $Bs(W_{(i,u)})^{\mapsto x}$ *and* $B(W_{(i,u)})^{\mapsto x}$*, where*

$$Bs(W_{(i,u)}) = \{(j,v,z) \mid (i,u) \to_z^* (j,v),\ (j,v) \in W_{(i,u)}, z \in \mathbb{Z}\}$$

and

$$B(W_{(i,u)}) = \{(j,v,z) \mid (i,u) \to_z^* (j,v),\ (j,v) \in V, z \in \mathbb{Z}\}$$

are subsets of $T(w)$*. All neighbourhoods of dominoes from* $Bs(W_{(i,u)})^{\mapsto x}$ *coincides and are equal to* $B(W_{(i,u)})^{\mapsto x}$*.*

The defined in the above sets $Bs(W_{(i,u)})$ and $B(W_{(i,u)})$ are called a base of $W_{(i,u)}$ and a neighbourhood of $W_{(i,u)}$ respectively. Note that these two notions are defined in relation to a multidigraph G and a strongly connected component. Now we define a link between these notions and a table $T(w)$ for a word $w \in S$. If a domino (i,u) occurs in the table $T(w)$ at the position x for some $w \in S$, $x \in \{1, \dots, \mid w \mid\}$, then all dominoes binded by (i,u) occur in the table. In other words the strongly connected component W containing the domino (i,u) shifted by some integer x is a part of the table. Note that this strongly connected component is equal to the base $Bs(W)^{\mapsto x}$. Hence to locate the base $Bs(W_{(i,u)})$ in the table $T(w)$ it is enough to know a position of the representant of $W_{(i,u)}$ which can be for example (i,u). This fact is reflected in the following definition.

Definition 3.9 Let $w \in S$. A strongly connected component $W_{(i,u)} \subset V$ represented by (i,u) occurs in the table $T(w)$ at the position x if and only if $(i,u,x) \in T(w)$. In this case the component is equal to the base $Bs(W_{(i,u)})^{\mapsto x} \subset T(w)$. A neighbourhood of strongly connected component $W_{(i,u)}$ in $T(w)$ is defined as $B(W_{(i,u)})^{\mapsto x} \subset T(w)$.

Note that, in fact we have $B(W_{(i,u)})^{\mapsto x} = B(Bs((W_{(i,u)})))^{\mapsto x}$. Let $CC_S(G) \subset CC(G)$ denote the following set.

$$CC_S(G) = \{W_{(i,u)} \in CC(G) \mid \exists w \in S, x \in \mathbb{N},\ Bs(W_{(i,u)})^{\mapsto x} \subset T(w)\}$$

We partially order the sets $CC(G)$ and $CC_S(G)$ putting

$$W_{(j,v)} \sqsubseteq W_{(i,u)} \ \Leftrightarrow\ \exists y \in \mathbb{Z} :\ (i,u) \to_y^* (j,v).$$

By $maxCC(G)$ and $maxCC_S(G)$ we denote the set of all maximal elements of the poset $(CC(G), \sqsubseteq)$ and $(CC_S(G), \sqsubseteq)$ respectively.

Lemma 3.10 *Let* $W_{(i,u)} \in CC_S(G)$ *for a multidigraph* G *associated with a semiretract* S*. Then*

(1) *the base* $Bs(W_{(i,u)})$ *is a configuration of dominoes; the neighbourhood* $B(W_{(i,u)})$ *is a strongly connected configuration of dominoes;*

(2) *the following equality is true*

$$B(W_{(i,u)}) = Bs(W) \cup \bigcup \{Bs(W_{(j,v)})^{\mapsto y} \mid (i,u) \to_y^* (j,v), y \in \mathbb{Z}\}$$

and elements of the sum are pairwise disjoint;

(3) $W_{(i,u)} \in maxCC_S(G)$ *iff for every domino* $(j,v,z) \in Bs(W_{(i,u)})$ *the column* z *is covered by a dominoes from* $Bs(W_{(i,u)})$*;*

(4) *If* $B(W_1)^{\mapsto x} \in T(w)$ *for some* $w \in S$, $x \in \{1,\ldots,|w|\}$, *then* $B(W_1)^{\mapsto x} \subset B(W_2)^{\mapsto y}$ *for some* $W_2 \in maxCC_S(G)$, $y \in \{1,\ldots,|w|\}$.

Now we present the main result.

Theorem 3.11 *Let* $S = \cap_{i=1}^n R_i$ *be a semiretract given by retracts* R_i *with key-codes* $C_i \subset A^*$ *for* $i = 1,\ldots,n$. *There exist key-codes* $D_i \subset A^*$ *for* $i = 1,\ldots,n$ *such that*

1. $S = \cap_{i=1}^n T_i$ *where retracts* $T_i = D_i^*$ *for* $i = 1,\ldots,n$
2. $key(D_1) = key(D_2) = \ldots = key(D_n)$

Proof For $S = \{1\}$ the conclusion is obvious. Hence assume that a nonempty word w is in the base of a semiretract $S = \cap_{i=1}^n C_i^*$ and consider the table $T(w)$. Let $w = w_1.....w_k$ in the key-code C_1. Denote by $W_1,\ldots,W_m$ all maximal strongly connected components that occur in $T(w)$ and assume that the components are ordered according to the to the order of $w_1,\ldots,w_k$. Note that any w_i for $i = 1,\ldots,k$ occurs in some maximal strongly component W_j, however $m \leq k$ in general. Of course $T(w) \subset \bigcup_{i=1}^m B(W_i)^{\mapsto x_i}$. Now consider any two subsequent component bases $Bs(W_j)$ and $Bs(W_{j+1})$. It may happened that $B(W_j)^{\mapsto x_j} \cap B(W_{j+1})^{\mapsto x_{j+1}}$ is not an empty set. Let l be the smallest index enumerating columns in $T(w)$ such that the $l-$column contains a square covered by a domino from $B(W_{j+1})^{\mapsto x_{j+1}}$. Denote by r the greatest index enumerating columns in $T(w)$ such that the $r-$column contains a square covered by a domino from $B(W_j)^{\mapsto x_j}$. It follows that $l \leq r$. The squares in the table $T(w)$ indexed by $\{1,\ldots,n\} \times [l,r]$ could be divided into three subsets:

- L - entries covered by dominoes from $B(W_j)^{\mapsto x_j} \setminus B(W_{j+1})^{\mapsto x_{j+1}}$;
- R - entries covered by dominoes from $B(W_{j+1})^{\mapsto x_{j+1}} \setminus B(W_j)^{\mapsto x_j}$;
- LR - entries covered by dominoes in $[B(W_j)^{\mapsto x_j} \cap B(W_{j+1})^{\mapsto x_{j+1}}]$.

Notice that the set R can be presented in the form

$$R = (\{1,\ldots,n\} \times [x_j, r]) \setminus [B(W_j)^{\mapsto x_j}].$$

Hence R is fully determined by $B(W_j)$. Observe that the triple (L, LR, R) defined for $B(W_j)$ is determined by R. The triple (L, LR, R) is called right anchor of $B(W_j)$ as it fixes relative positions of $B(W_j)^{\mapsto x_j}$ and $B(W_{j+1})^{\mapsto x_{j+1}}$ in the table $T(w)$. In general, for any two words w_1, w_2 in S such that $B(W_j)$ occurs in $T(w_1)$ and in $T(w_2)$ it follows that the right anchor of $B(W_j)$ is invariant respectively to a shift, that is equal to (L, LR, R). The right anchor of $B(W_m)^{\mapsto x_m}$ is equal to $(\emptyset, \emptyset, \emptyset)$ and we say that W_m is the final strongly connected component. Any maximal strongly connected component with this property is called final. In a quite similar way we define the notion of the left anchor and strongly connected components that has the left anchor equal to $(\emptyset, \emptyset, \emptyset)$ we call an initial components. Hence, W_1 is an initial component. With a neibourhood of a maximal connected component $B(W_j)^{\mapsto x_j}$, $j \in \{1, \ldots, m\}$ we can associate n words $v_1(W_i)$, $\ldots$, $v_n(W_i)$. Any $v_i(W_j)$ is the word consisted of the letters from the i-th row of $B(W_j)^{\mapsto x_j} \setminus B(W_{i+1})^{\mapsto x_{i+1}}$. Hence according to the previous considerations, words determined by $B(W_j)^{\mapsto x_j}$. Moreover, the words $v_i(W_j)$ are key words with a common key letter. We can choose as a common key letter the key of a representant of W_j. Finally, let us define

$$D_i = \{v_i(W) \mid W \in \max(CC_S(G))\}$$

Continuing considerations of the word w we conclude the following equalities

$$\begin{array}{ll} w = & v_1(W_1)....v_1(W_m)\ (1) \\ w = & v_2(W_1)....v_2(W_m)\ (2) \\ .... & .... \\ w = & v_n(W_1)....v_n(W_m)\ (n) \end{array}$$

It implies that $C \subset D_i^*$ for all $i \in \{1, \ldots, n\}$. On the other hand, if $w \in \bigcap_{i=1}^n D_i^*$, then there exist an initial component W_1, final component W_m and components $W_2, \ldots,$ W_m such that the equalities $(1) - (n)$ are true. Hence $w \in C$, since we can reconstruct a table $T(w)$. Finally by the construction in the above, the codes D_i are key-codes and the sets of keys are identical. □

4 An algorithm

The main theorem proved in the previous paragraph points out the importance of strongly connected components in $CC_S(G)$, especially those in $maxCC_S(G)$. Hence, our problem is to check out if $W \in CC(G)$ represented by (i, u) is a member of $maxCC_S(G)$. Basing on the results of the previous section the following conditions W has to fulfill to be in $maxCC_S(G)$:

1. $(W, E_{|W \times \mathbb{Z} \times W})$ is a graph - Lemma 3.6.(1);
2. $Bs(W_{(i,u)})$ is a configuration of dominoes - Lemma 3.8.(1);
3. $B(W_{(i,u)})$ is a connected configuration of dominoes - Lemma 3.8.(1);
4. for every domino $(j, v, z) \in Bs(W_{(i,u)})$ the column z is covered by a dominoes from $Bs(W_{(i,u)})$ - Lemma 3.8.(3).

Let us denote by $MAX(G)$ all components in $CC(G)$ that fulfill the conditions (1),(2),(3) and (4). Of course $maxCC_S(G) \subset MAX(G)$ but generally it may happened that $maxCC_S(G) \neq MAX(G)$. Since the anchors of $W \in maxCC_S(G)$ are fully determined by $B(W)$, then also the left and right anchor for $W \in MAX(G)$ are determined.

Definition 4.1 Let $W_1, W_2 \in MAX(G)$. We say that a component W_1 is initial (final) if the left (right) anchor of W_1 is equal to $(\emptyset, \emptyset, \emptyset)$. We say that W_2 follows W_1 if the right anchor of W_1 is equal the left anchor of W_2 with respect to the shift.

Lemma 4.2 *Let $W_1, \ldots, W_m \in MAX(G)$ be a sequence such that*

- *(i) W_1 is an initial component,*
- *(ii) W_{i+1} follows W_i for $i = 1, \ldots, m-1$,*
- *(iii) W_m is a final component*

Then for $i = 1, \ldots, m$ $W_i \in maxCC_S(G)$ and the word

$$w = v_1(W_1) \ldots v_1(W_m) = ... = v_n(W_1) \ldots v_n(W_m)$$

is in the base of semiretract S. Moreover, for any word w in C there exist a sequence $W_1, \ldots, W_m \in maxCC_S(G)$ such that the above is true.

Proof All the statements follows by Theorem 3.9. □

Any sequence $W_1, \ldots, W_m \in MAX(G)$ fulfilling assumptions (i)-(iii) is called a generating component sequence.

Lemma 4.3 *$W \in MAX(G)$ is in $maxCC_S(G)$ if and only if (5) there exists a generating component sequence $W_1, \ldots, W_m \in MAX(G)$ such that $W = W_i$ for some $i \in \{1, \ldots, m\}$.*

Now we are ready to present a sketch of an algorithm that produces key-codes $D_1, \ldots, D_n$ all with the same key set K and such that $S = \bigcap_{i=1}^{n} D_i^*$.

1. construct the set $CC(G)$; the poset $(CC(G), \sqsubseteq)$ is created automatically;
2. for every W in $CC(G)$ in the order given by poset $(CC(G), \sqsubseteq)$:
   - (a) construct $Bs(W)$ and $B(W)$:
     - i. test if W satisfy the conditions (1),(2);
     - ii. construct $B(W)$ - use sets $B(W_1)$ where $W_1 \sqsubseteq W$ for some $W_1 \sqsubseteq W$ if necessary - Lemma 3.8.(2).
     - iii. test if W satisfy (3);

iv. test if W satisfy (4); if YES then insert W into $MAX(G)$;

(b) for any $W \in MAX(G)$ compute the left and the right anchor of W; insert the left and the right anchor into the set Dic;

(c) for any $W \in MAX(G)$ test if W satisfy the condition (5); if YES then insert W into $maxCC_S(G)$;

(d) for any $W \in CC_S(G)$ produce n key words $v_1(W) \in D_1, \ldots, v_n(W) \in D_n$ by the rule given in the Theorem 3.9.

Observe that if a strongly connected component W does not fulfill the condition (1) or (2) or (3) then we stop the computation for W and abandon computations for all $U \in CC(G)$ such that $W \sqsubseteq U$ and U binds W. Thus, just after verifying that W does not fulfill (1) or (2) or (3) we stop the computations for W and come back to the point (2). Using a dictionary Dic for storing all anchors of W one can easily verify if W_2 follows W_1 for any $W_2, W_1 \in max(G)$. Hence, we can construct a data structure that stores words $v_1(W) \in D_1, \ldots, v_n(W) \in D_n$ for all $W \in maxCC_S(G)$ in time

$$O(max(n, log|A|) * (|C_1| + \ldots + |C_n|)),$$

where $|C_i| = \sum_{w \in C_i} |w|$ and $|A|$ is the number of elements in the alphabet over which the codes $C_1, \ldots, C_n$ were defined. Note that the length of the input is equal to $|C_1| + \ldots + |C_n|$.

5 An automaton that recognizes the base of a semiretract S.

The last problem we want to deal with in this paper is construction of the minimal, deterministic automaton A_S that recognizes the base of a semiretract S. In order to get a characteristics of words in the base of a semiretract, let us introduce two equivalence relations λ and ρ on the set $maxCC_S(G)$.

Definition 5.1 We say that $W_1, W_2 \in maxCC_S(G)$ are in relation λ (ρ) if the left (right) anchors of W_1 and W_2 are equal with respect to the shift.

Note that in $maxCC_S(G)_{/\lambda}$ there exist an equivalence class (block) that consists of all initial keys. This equivalence class is denoted by L_{init}. Dually in $maxCC_S(G)_{/\lambda}$ there exists a block that consists of all final keys. This block is denoted by R_{final}. Suppose now that $maxCC_S(G)_{/\lambda} = \{L_{init}, L_1, \ldots, L_m\}$. Hence, $maxCC_S(G)_{/\rho} = \{R_{final}, R_1, \ldots, R_m\}$ where $R_1, \ldots, R_m \subset CC_S(G)$ are such that W_2 follows W_1 if and only if $W_1 \in R_j$ and $W_2 \in L_j$ for some $j \in \{1, \ldots, m\}$. Now we are ready to describe the procedure that gives generating component sequence $W_1, \ldots, W_p \in maxCC_S(G)$ for a semiretract S:

1. choose a a component W_1 from the block of initial components L_{init};

2. find a block $R \in \{R_{final}, R_1, \ldots, R_m\}$ that contains W_1;
3. if W_1 is not a final component then find a block $L \in \{L_1, \ldots, L_m\}$ that is matched to R;
4. choose a component W_2 from the block L;
5. repeat steps 2-4 until the chosen component is final;
6. write down all the obtained components in the order that they were produced.

This construction justifies the following theorem.

Theorem 5.2 *Any sequence of component obtained by the above procedure is a generating key sequence.*

Assume now that for any $j \in \{1, \ldots, m\}$ a triple (A, AB, B) is the common anchor of components $W_1 \in R_j$ (right) and $W_2 \in L_j$ (left) and denote by $anch_j$ the word consisted of all the letters from, for example, the first row of the rectangle covered by (A, AB, B) (see Theorem 3.9). With any component $W \in maxCC_S(G)$ we associate two words $left(W), right(W) \in A^*$ according to the following rules. Let $key(W)$ denote a common key of words $v_1(W), \ldots, v_n(W)$. If W is an initial component, then for all $i = 1, \ldots, n$ we have $v_i(W) = wkey(W)u_i$ for some $w, u_1, \ldots, u_n \in A^*$ and we put $left(W) = w$. If W is a final component then for $i = 1, \ldots, n$ we have $v_i(W) = u_i key(W)w$ for some $w, u_1, \ldots, u_n \in A^*$ and we put $right(W) = w$. Finally, if $W_1 \in R_j$ and $W_2 \in L_j$ (W_2 follows W_1), then for $i = 1, ..n$ we have $v_i(W_1)v_i(W_2) = u_i key(W_1)w_1 anch_j w_2 v_j$ for some $w_1, w_2 \in A^*$ and $u_1, v_1, \ldots, u_n, v_n \in A^*$. Then we put $right(W_1) = w_1$ and $left(W_2) = w_2$. It is not hard to verify that $left(W), right(W)$ are properly defined for any $W \in maxCC_S(G)$.

Hence the just introduced notions allow us to formulate the following statement.

Lemma 5.3 *Let $W_1, \ldots, W_m \in maxCC_S(G)$ be a generating component sequence. Then the word*

$$w = left(W_1)key(W_1)right(W_1)anch_{i_1} \ldots anch_{i_{p-1}} left(W_p)key(W_p)right(W_p),$$

where for $j = 1, \ldots, p-1$ the index $i_j \in \{1, \ldots, m\}$ is such that $W_i \in R_{i_j}$ and $W_{i+1} \in L_{i_j}$ is in the base of semiretract S.

Moreover, if w is in the base of a semiretract then there exists a generating key sequence $W_1, \ldots, W_m \in maxCC_S(G)$ such that the above is true.

Now we present a construction of an automaton which recognizes the base of a semiretract S. For any $L \in \{L_{init}, L_1, .., L_m\}$ let us consider the language $\{left(W) | \, W \in L\}$. If the word w is a prefix of any word from $\{left(W) | \, W \in L\}$

then w defines a state q_w. The set of all states obtained in this way we denote by $Q(L)$. There is an edge (q_{w_1}, a, q_{w_2}) between states $q_{w_1}, q_{w_2} \in Q(L)$ if and only if $w_1 a = w_2$. The set of all edges obtained in that way we denote by $E(L)$. Note that for any $W \in left(W)$ there exist a state $q_{left(W)} \in Q(L)$ such that there is a path from q_1 (the state for the empty prefix) to $q_{left(W)}$ with $left(W)$ as the label.

For any $R \in \{R_{final}, R_1, .., R_m\}$ let us consider the language $\{right(W) |\, W \in R\}$. If the word w is a suffix of any word from $\{right(W) |\, W \in R\}$ then w defines a state q_w. The set of all states obtained in this way we denote by $Q(R)$. There is an edge (q_{w_1}, a, q_{w_2}) between states $q_{w_1}, q_{w_2} \in Q(L)$ if and only if $w_1 = aw_2$. The set of all edges obtained in this way we denote by $E(R)$. Note that for any $W \in R$ there exist a state $q_{right(W)} \in Q(R)$ such that there is a path from $q_{right(W)}$ to q_1 (q_1 is the state for empty suffix) with $right(W)$ as the label.

Let $j \in \{1, \ldots, m\}$. Suppose now that all states in $Q(L_j)$ and $Q(R_j)$ are distinguishable. If it is necessary we write an upper index R_j or L_j to underline that a state is in $Q(R_j)$ or $Q(L_j)$. Assume that $anch_j = w_1 \ldots w_k$. We define a set

$$Q(R_j, L_j) = Q(R_j) \cup Q_{L_j} \cup \{q_1, \ldots, q_{k-1}\}$$

and suppose that this sum is pairwise disjoint. Then define the set of edges $E(R_j, L_j)$ putting

$$E(R_j, L_j) = E(R_j) \cup E(L_j) \cup \{(q_1^{R_j}, w_1, q_1), (q_1, w_2, q_2), \ldots, (q_{k-1}, w_k, q_1^{L_j})\}.$$

Note that by the construction, for any $W_1 \in R_j, W_2 \in L_j$ there exists a path from $q_{right(W_1)} \in Q(R_j)$ to $q_{left(W_2)} \in Q(L_j)$ with the label $right(W_1) anch_j left(W_2)$.

Finally, assume that all states in $Q(L_{init})$, $Q(R_1, L_1), \ldots, Q(R_m, L_m)$, $Q(R_{final})$ are distinguishable. We define an automaton $A_S = (Q_S, E_S, I_S, T_S)$ that recognizes the base of semiretract S as follows. The set of states is equall to $Q_S = Q(L_{init}) \cup Q(R_{final}) \cup \bigcup_{j=1}^{m} Q(R_j, L_j)$. Let $W \in maxCC_S(G)$ and assume that $W \in R_x$ and $W \in L_y$ for some $x, y \in \{1, \ldots m\}$. Then there exists two states $q_{left(W)} \in Q(L_y)$ and $q_{right(W)} \in Q(R_x)$. Let us connect this states with an edge $(q_{left(W)}, key(W), q_{right(W)})$. We repeat that procedure for any component $W \in maxCC_S(G)$ and we denote by E_1 the set of all edges obtained in this way. Hence, we put $E_S = E_1 \cup E(L_{init}) \cup E(R_{final}) \cup \bigcup_{j=1}^{m} E(R_j, L_j)$. Finally, we can take $q_1 \in Q(L_{init})$ as the only initial state and $q_1 \in Q(R_{final})$ as the only terminal state.

By the construction of the automaton $A_S = (Q_S, E_S, I_S, T_S)$ and by Lemma 5.3 we have the following statement.

Lemma 5.4 *The automaton $A_S = (Q_S, E_S, I_S, T_S)$ described above is minimal, deterministic and recognizes the base of semiretract S.*

Proof By construction, the automaton A is deterministic. It is not hard to verify that the sets of all words $L(q)$ for any $q \in Q$ are pairwise different, where

$L(q)$ denotes the set of all words that occurs as a label on a path from q to the terminal state. Hence, the automaton is minimal. □

To construct the automaton $A_S = (Q_S, E_S, I_S, T_S)$ we can use an algorithm presented in the previous paragraph. Moreover, it is possible to propose a data structure that allows us to construct A_S in time

$$O(max(n, log|A|) * (|C_1| + \ldots + |C_n|)).$$

References

[1] J. A. Anderson, Semiretracts of a free monoid, Theoretical Computer Science 134, 1994

[2] J. A. Anderson, The intersection of retracts of A^*, Theoretical Computer Science, 237, 2000

[3] J. A. Anderson, Code properties of minimal generating sets of retracts and semiretracts, SEA Bull. Math, 18, 1994

[4] J. A. Anderson, W. Forys, Regular languages and semiretracts, Intern. Conference on Words, Kyoto, 2000

[5] J. A. Anderson, T. Head, The lattice of semiretracts of a free monoid, Intern. J. Computer Math. 43, 1992

[6] J. A. Anderson, W. Forys, T. Head, Retracts and semiretracts of free monoids, AMS Meeting, San Francisco 1991

[7] J. Berstel, D. Perrin, Theory of Codes, Academic Press, new York, 1985

[8] W. Forys, T. Head, The poset of retracts of a free monoid, Intern. J. Computer Math. 37, 1990

[9] W. Forys, On the family of retracts of free monoids, Intern. J. Computer Math. 33, 1990

[10] T. Head, Fixed languages and the adult languages of 0L-schemes, Intern. J. Computer Math. 10, 1981

[11] T. Head, Expanded subalphabets in the theories of languages and semigroups, Intern. J. Computer Math. 12, 1982

[12] G. Lallement, Semigroups and combinatorial applications, John Wiley&Sons, 1979

Powers in a class of $\mathcal{A}$-strict standard episturmian words

Amy Glen[*]

1 Introduction

Introduced by Droubay, Justin and Pirillo [8], *episturmian words* are a natural extension of the well-known family of *Sturmian words* (aperiodic infinite words of minimal complexity) to an arbitrary finite alphabet. In this paper, the study of episturmian words is continued in more detail. In particular, for a specific class of episturmian words (a typical element of which we shall denote by $\mathbf{s}$), we will explicitly determine all the integer powers occurring in its constituents. This has recently been done in [6] for Sturmian words, which are exactly the aperiodic episturmian words over a two-letter alphabet.

A finite word w is said to have an integer *power* in an infinite word $\mathbf{x}$ if $w^p = ww \cdots w$ (p times) is a *factor* of $\mathbf{x}$ for some integer $p \geq 2$. Here, our analysis of powers occurring in episturmian words $\mathbf{s}$ hinges on canonical decompositions in terms of their 'building blocks'. Another key tool is a generalization of *singular words*, which were first defined in [17] for the ubiquitous *Fibonacci word*, and later extended to Sturmian words in [15] and the *Tribonacci sequence* in [16]. Our generalized singular words will prove to be useful in the study of factors of episturmian words, just as they have for Sturmian words.

This paper is organized as follows. After some preliminaries (Section 2), we define, in Section 3, a restricted class of episturmian words upon which we will focus for the rest of the paper. A typical element of this class will be denoted by $\mathbf{s}$. In Section 4, we give some simple results which, in turn, lead us to a generalization of *singular words* for episturmian words $\mathbf{s}$. The *index*, i.e., maximal *fractional power*, of the building blocks of $\mathbf{s}$ is then studied in Section 5. Finally, in Section 6, we determine all squares (and subsequently higher powers) occurring in $\mathbf{s}$. The main results are demonstrated via the *k-bonacci word*; a generalization of the Fibonacci word to a k-letter alphabet ($k \geq 2$).

[*] School of Mathematical Sciences, Discipline of Pure Mathematics, University of Adelaide, South Australia, Australia, 5005, amy.glen@adelaide.edu.au

2 Definitions and notations

2.1 Words

Let $\mathcal{A}$ denote a finite alphabet. A (finite) *word* is an element of the *free monoid* $\mathcal{A}^*$ generated by $\mathcal{A}$, in the sense of concatenation. The identity ε of $\mathcal{A}^*$ is called the *empty word*, and the *free semi-group*, denoted by $\mathcal{A}^+$, is defined by $\mathcal{A}^+ := \mathcal{A}^* \backslash \{\varepsilon\}$. Similarly, we define the set $\mathcal{A}^\omega$ of all *infinite words* (or *sequences*) $\mathbf{x} = x_0x_1x_2\cdots$ over $\mathcal{A}$, and set $\mathcal{A}^\infty := \mathcal{A}^* \cup \mathcal{A}^\omega$. If u is a non-empty finite word, then u^ω denotes the *purely periodic* infinite word $uuu\cdots$.

If $w = x_1x_2\ldots x_m \in \mathcal{A}^+$, each $x_i \in \mathcal{A}$, the *length* of w is $|w| = m$ and we denote by $|w|_a$ the number of occurrences of a letter a in w. (Note that $|\varepsilon| = 0$.) The *reversal* of w is $\widetilde{w} = x_mx_{m-1}\ldots x_1$, and if $w = \widetilde{w}$, then w is called a *palindrome*.

A finite word w is a *factor* of $z \in \mathcal{A}^\infty$ if $z = uwv$ for some $u \in \mathcal{A}^*$, $v \in \mathcal{A}^\infty$, and we write $w \prec z$. Further, w is called a *prefix* (resp. *suffix*) of z if $u = \varepsilon$ (resp. $v = \varepsilon$), and we write $w \subseteq_p z$ (resp. $w \subseteq_s z$). An infinite word $\mathbf{x} \in \mathcal{A}^\omega$ is called a *suffix* of $\mathbf{z} \in \mathcal{A}^\omega$ if there exists a word $w \in \mathcal{A}^+$ such that $\mathbf{z} = w\mathbf{x}$. A factor w of a word $z \in \mathcal{A}^\infty$ is *right* (resp. *left*) *special* if wa, wb (resp. aw, bw) are factors of z for some letters $a, b \in \mathcal{A}$, $a \neq b$.

For $\mathbf{x} \in \mathcal{A}^\omega$, $\Omega(\mathbf{x})$ denotes the set of all its factors, and $\Omega_n(\mathbf{x})$ denotes the set of all factors of $\mathbf{x}$ of length $n \in \mathbb{N}$, i.e., $\Omega_n(\mathbf{x}) := \Omega(\mathbf{x}) \cap \mathcal{A}^n$. Moreover, the *alphabet* of $\mathbf{x}$ is $\mathrm{Alph}(\mathbf{x}) := \Omega(\mathbf{x}) \cap \mathcal{A}$, and we denote by $\mathrm{Ult}(\mathbf{x})$ the set of all letters occurring infinitely often in $\mathbf{x}$. An infinite word $\mathbf{y} \in \mathcal{A}^\omega$ is said to be *equivalent* to $\mathbf{x}$ if $\Omega(\mathbf{y}) = \Omega(\mathbf{x})$, i.e., if $\mathbf{y}$ has the same set of factors as $\mathbf{x}$.

Let $w = x_1x_2\cdots x_m \in \mathcal{A}^*$, each $x_i \in \mathcal{A}$, and let $j \in \mathbb{N}$ with $0 \leq j \leq m-1$. The *j-th conjugate* of w is the word $C_j(w) := x_{j+1}x_{j+2}\cdots x_mx_1x_2\cdots x_j$, and we denote by $\mathcal{C}(w)$ the conjugacy class of w, i.e., $\mathcal{C}(w) := \{C_j(w) \ : \ 0 \leq j \leq |w|-1\}$. Observe that if w is primitive (i.e., not a power of a shorter word), then w has exactly $|w|$ distinct conjugates.

The *inverse* of $w \in \mathcal{A}^*$, written w^{-1}, is defined by $ww^{-1} = w^{-1}w = \varepsilon$. It must be emphasized that this is merely notation, i.e., for $u, v, w \in \mathcal{A}^*$, the words $u^{-1}w$ and wv^{-1} are defined only if u (resp. v) is a prefix (resp. suffix) of w.

A *morphism on* $\mathcal{A}$ is a map $\psi : \mathcal{A}^* \to \mathcal{A}^*$ such that $\psi(uv) = \psi(u)\psi(v)$ for all $u, v \in \mathcal{A}^*$. It is uniquely determined by its image on the alphabet $\mathcal{A}$.

2.2 Episturmian words

Let $\mathcal{A}$ be an arbitrary finite alphabet. An infinite word $\mathbf{t} \in \mathcal{A}^\omega$ is *episturmian* if $\Omega(\mathbf{t})$ is closed under reversal and $\mathbf{t}$ has at most one right special factor of length n for each $n \in \mathbb{N}$. Moreover, an episturmian word is *standard* if all of its left special factors are prefixes of it.

Let $\mathbf{t}$ be a standard episturmian word over $\mathcal{A}$ and let $u_1 = \varepsilon$, u_2, u_3, $\ldots$ be the sequence of its palindromic prefixes (which exist by results in [8]). Then there exists an infinite word $\Delta(\mathbf{t}) = x_1x_2x_3\ldots$, each $x_i \in \mathcal{A}$, called the *directive* *word* of $\mathbf{t}$, such that

$$u_{n+1} = (u_n x_n)^{(+)}, \quad n \in \mathbb{N}^+, \tag{2.1}$$

where the *palindromic right-closure* $w^{(+)}$ of a word w is the (unique) shortest palindrome of which w is a prefix (see [7]). The important point here is that a standard episturmian word $\mathbf{t}$ can be constructed as a limit of an infinite sequence of its palindromic prefixes, i.e., $\mathbf{t} = \lim_{n\to\infty} u_n$.

For each letter $a \in \mathcal{A}$, define the morphism Ψ_a on $\mathcal{A}$ by $\Psi_a(a) = a$ and $\Psi_a(x) = ax$ for all $x \in \mathcal{A} \setminus \{a\}$. Further, let us define [12]

$$\mu_n := \Psi_{x_1}\Psi_{x_2}\cdots\Psi_{x_n}, \quad \mu_0 = \mathrm{Id},$$

and

$$h_n := \mu_n(x_{n+1}), \quad n \in \mathbb{N}.$$

Then, we have the following useful formula [12]

$$u_{n+1} = h_{n-1}u_n;$$

and whence, for $n > 1$ and $0 < p < n$,

$$u_n = h_{n-2}h_{n-3}\cdots h_1 h_0 = h_{n-2}h_{n-3}\cdots h_{p-1}u_p. \tag{2.2}$$

Lemma 2.1 [12] *For all $n \in \mathbb{N}$,*

(i) *h_n is a primitive word*;

(ii) *$h_n = h_{n-1}$ if and only if $x_{n+1} = x_n$;*

(iii) *if $x_{n+1} \neq x_n$, then u_n is a proper prefix of h_n.*

Two functions can be defined with regard to positions of letters in a given directive word. For $n \in \mathbb{N}^+$, let $P(n) = \sup\{p < n \ : \ x_p = x_n\}$ if this integer exists, $P(n)$ undefined otherwise. Also, let $S(n) = \inf\{p > n \ : \ x_p = x_n\}$ if this integer exists, $S(n)$ undefined otherwise. By the definitions of palindromic closure and the words u_n, it follows that $u_{n+1} = u_n x_n u_n$ (whence $h_{n-1} = u_n x_n$) if x_n does not occur in u_n, and $u_{n+1} = u_n u_{P(n)}^{-1} u_n$ (whence $h_{n-1}u_{P(n)} = u_n$) if x_n occurs in u_n. Thus, if $P(n)$ exists, then

$$h_{n-1} = h_{n-2}h_{n-3}\cdots h_{P(n)-1}, \quad n \geq 1. \tag{2.3}$$

A standard episturmian word $\mathbf{t}$, or any equivalent (episturmian) word, is said to be $\mathcal{A}$-*strict* (or $|\mathcal{A}|$-*strict*) if $\mathrm{Alph}(\Delta(\mathbf{t})) = \mathrm{Ult}(\Delta(\mathbf{t})) = \mathcal{A}$. The k-strict episturmian words have *complexity* $(k-1)n+1$ for each $n \in \mathbb{N}$ (i.e., $(k-1)n+1$ distinct factors of length n for each $n \in \mathbb{N}$). Such words are exactly the k-letter *Arnoux-Rauzy sequences*, the study of which began in [1].

2.3 Return words

Let $\mathbf{x} \in \mathcal{A}^\omega$ be *recurrent*, i.e., any factor w of $\mathbf{x}$ occurs infinitely often in $\mathbf{x}$. A *return word* [9] of factor w of $\mathbf{x}$ is a factor of $\mathbf{x}$ that begins with w and ends exactly before the next occurrence of w in $\mathbf{x}$. Episturmian words are recurrent and, according to [13, Corollary 4.5], each factor of an $\mathcal{A}$-strict episturmian word has exactly $|\mathcal{A}|$ return words.

3 A class of strict standard episturmian words

Given any infinite sequence $\Delta = x_1x_2x_3\cdots$ over a finite alphabet $\mathcal{A}$, we can define a standard episturmian word having Δ as its directive word (using (2.1)). In this paper, however, we shall only consider a specific family of $\mathcal{A}$-strict standard episturmian words.

Let $\mathcal{A}_k$ denote a k-letter alphabet, say $\mathcal{A}_k = \{a_1, a_2, \ldots, a_k\}$, and suppose $\mathbf{t}$ is a standard episturmian word over $\mathcal{A}_k$. Then the directive word of $\mathbf{t}$ can be expressed as:

$$\Delta(\mathbf{t}) = a_1^{d_1}a_2^{d_2}\cdots a_k^{d_k}a_1^{d_{k+1}}a_2^{d_{k+2}}\cdots a_k^{d_{2k}}a_1^{d_{2k+1}}\cdots ,$$

where the d_i are non-negative integers. In what follows, we restrict our attention to the case when all $d_i > 0$; that is, we shall only study the class of k-strict standard episturmian words $\mathbf{s} \in \mathcal{A}_k^\omega$ with directive words of the form:

$$\Delta = a_1^{d_1}a_2^{d_2}\cdots a_k^{d_k}a_1^{d_{k+1}}a_2^{d_{k+2}}\cdots a_k^{d_{2k}}a_1^{d_{2k+1}}\cdots , \quad d_i > 0. \tag{3.1}$$

This definition of $\mathbf{s}$ will be kept throughout the rest of this paper.

Let us define a sequence $(s_n)_{n\geq 1-k}$ of words associated with $\mathbf{s}$ as follows:

$$\begin{aligned} &s_{1-k} = a_2, \quad s_{2-k} = a_3, \quad \ldots\ , \quad s_{-1} = a_k, \quad s_0 = a_1, \\ &s_n = s_{n-1}^{d_n}s_{n-2}^{d_{n-1}}\cdots s_0^{d_1}a_{n+1}, \quad 1 \leq n \leq k-1, \\ &s_n = s_{n-1}^{d_n}s_{n-2}^{d_{n-1}}\cdots s_{n-k+1}^{d_{n-k+2}}s_{n-k}, \quad n \geq k. \end{aligned} \tag{3.2}$$

Clearly, s_n is a prefix of s_{n+1} for all $n \geq 0$ (and hence $(|s_n|)_{n\geq 0}$ is a strictly increasing sequence of positive integers).

Example 3.1 It is well-known that the standard Sturmian word c_α of irrational *slope* $\alpha = [0; 1+d_1, d_2, d_3, \ldots]$, $d_1 \geq 1$, (see [3] for definition) is the standard episturmian word over $\mathcal{A} = \{a, b\}$ with directive word $\Delta(c_\alpha) = a^{d_1}b^{d_2}a^{d_3}b^{d_4}a^{d_5}\cdots$. We have $c_\alpha = \lim_{n\to\infty} s_n$, where $(s_n)_{n\geq -1}$ is the *standard sequence* associated with c_α, defined by

$$s_{-1} = b, \quad s_0 = a, \quad s_n = s_{n-1}^{d_n}s_{n-2}, \quad n \geq 1.$$

This coincides with our definition (3.2) above. Observe that, for all $n \geq 0$, $|s_n| = q_n$, where q_n is the denominator of the n-th convergent to $[0; 1 + d_1, d_2, d_3, \ldots]$.

For all $m \geq 1$, let $L_m := d_1+d_2+\cdots+d_m$. Then, writing $\Delta(c_\alpha) = x_1x_2x_3\cdots$ with each $x_i \in \mathcal{A}$, we have $x_{n+1} \neq x_n$ if and only if n is equal to some L_m. One easily deduces that $S(L_m) = L_{m+1} + 1$ and $P(L_{m+1} + 1) = L_m$, and it can also be shown that the h_{L_m} satisfy the same recurrence relation as the q_m. Hence, $|h_{L_m}| = q_m$, and clearly we have $h_{L_m} = s_m$ (see Proposition 3.2, to follow).

Notation 3.1 Hereafter, let $L_n := d_1 + d_2 + \cdots + d_n$ for each $n \geq 1$.

Proposition 3.2 *For any $n \geq 1$, $s_n = h_{L_n}$. Moreover,* $\mathbf{s} = \lim_{n\to\infty} s_n$.

Accordingly, the words $(s_n)_{n\geq 1}$ can be viewed as 'building blocks' of $\mathbf{s}$.

Example 3.3 The *Tribonacci sequence* is the standard episturmian word over $\{a, b, c\}$ directed by $(abc)^\omega$. Since all $d_i = 1$, we have $L_n = n$, and hence $h_n = s_n = s_{n-1}s_{n-2}s_{n-3}$, for all $n \geq 1$.

4 Generalized singular words

Recall the standard Sturmian word c_α of slope $\alpha = [0; 1 + d_1, d_2, d_3, \ldots]$, $d_1 \geq 1$ (Example 3.1). Melançon [15] (also see [4]) introduced the singular words $(w_n)_{n\geq 1}$ of c_α defined by

$$w_n = \begin{cases} as_nb^{-1} & \text{if } n \text{ is odd,} \\ bs_na^{-1} & \text{if } n \text{ is even,} \end{cases}$$

with the convention $w_{-2} = \varepsilon$, $w_{-1} = a$, $w_0 = b$. Let us remark that $s_n = u_{L_n}ab$ (resp. $s_n = u_{L_n}ba$) if n is odd (resp. even).

Singular words are profoundly useful in studying properties of factors of c_α (e.g., [4,10,11,14,15,17]). It is for this very reason that we now generalize these words to the case of standard episturmian words $\mathbf{s}$. Firstly, however, we state some basic results concerning the words s_n and u_{L_n}, as detailed in the next section. (Proofs will appear in the extended version of this paper.)

4.1 Useful results

For each $n \geq 0$, set $D_n := u_{L_{n+1}}$. Observe that, for any $m \geq 1$,

$$|D_m| = (d_{m+1} - 1)|s_m| + \sum_{j=0}^{m-1} d_{j+1}|s_j|. \tag{4.1}$$

Indeed, using (2.2), one finds that

$$\begin{aligned} D_m = u_{L_{m+1}} &= h_{L_{m+1}-2}h_{L_{m+1}-3}\cdots h_1h_0 \\ &= h_{L_m}^{d_{m+1}-1}h_{L_{m-1}}^{d_m}h_{L_{m-2}}^{d_{m-1}}\cdots h_{L_1}^{d_2}h_0^{d_1} \\ &= s_m^{d_{m+1}-1}s_{m-1}^{d_m}s_{m-2}^{d_{m-1}}\cdots s_1^{d_2}s_0^{d_1}. \end{aligned} \tag{4.2}$$

Also note that $D_0 = a_1^{d_1-1}$ since $D_0 = u_{d_1} = h_{d_1-2}h_{d_1-3}\cdots h_1h_0 = h_0^{d_1-1}$. For technical reasons, we shall set $D_{-j} := a_{k+1-j}^{-1}$ and $|D_{-j}| = -1$ for $1 \le j \le k$.

Proposition 4.1 *Let $1 \le i \le k$. For all $n \ge 1-k$, $a_i \subseteq_s s_n$ if $n \equiv i-1 \pmod{k}$.*

Proposition 4.2 *For all $n \ge 0$, $s_{n+1}D_{n-k+1} = s_nD_n$, and hence $|D_n| - |D_{n-k+1}| = |s_{n+1}| - |s_n|$.*

Proposition 4.3 *For all $n \ge 1$, $|s_n| > |D_{n-1}|$.*

Recall that the words D_n and s_n are prefixes of $\mathbf{s}$ for all $n \in \mathbb{N}$. Thus, according to Proposition 4.3, the palindromes $D_0, D_1, \ldots, D_{n-1}$ are prefixes of s_n. In fact, the maximal index i such that D_i is a proper prefix of s_n is $i = n-1$, which is evident from the following result.

Proposition 4.4 *For all $n \ge 0$, $D_n = s_n^{d_{n+1}}D_{n-k}$.*

Proposition 4.5 *For all $n \ge 0$, $s_n = D_{n-k}\widetilde{s}_nD_{n-k}^{-1}$.*

Remark 4.6 This result shows, in particular, that $\widetilde{s}_n = D_{n-k}^{-1}s_nD_{n-k}$, i.e., $\widetilde{s}_n$ is the $|D_{n-k}|$-th conjugate of s_n for each $n \ge k$. (For $0 \le n \le k-1$, $\widetilde{s}_n$ is the $(|s_n|-1)$-st conjugate of s_n since $\widetilde{s}_n = a_{n+1}s_na_{n+1}^{-1}$.) The following two corollaries are direct results of the above proposition.

Corollary 4.7 *For any $n \ge 0$, the word $\widetilde{s}_nD_{n-k}^{-1}$ is a palindrome. In particular, let $U_n = D_{n-k}$ and $V_n = \widetilde{s}_nD_{n-k}^{-1}$. Then $s_n = U_nV_n$ is the unique factorization of s_n as a product of two palindromes.*

Corollary 4.8 *For all $n \ge 0$, $s_n = D_n\widetilde{s}_nD_n^{-1}$.*

Now, for each $n \in \mathbb{N}$, we define the words $G_{n,r}$ by

$$s_n = D_{n-r}G_{n,r}, \quad 1 \le r \le k-1.$$

For example, in the case of Sturmian words c_α, $r = 1$ and $s_n = u_{L_n}G_{n,1}$ for all $n \ge 1$, where $G_{n,1} = ab$ or ba, according to n odd or even, respectively.

Let us note that since $D_{n-r} = a_{k+1+n-r}^{-1}$ for $0 \le n < r$, we also set

$$G_{n,r} = a_{k+1+n-r}s_n, \quad 0 \le n < r. \tag{4.3}$$

Proposition 4.9 *For all $n \ge 1$, $s_ns_{n-1}G_{n-1,k-1}^{-1} = s_{n-1}s_nG_{n,1}^{-1}$.*

Remark 4.10 Recall Example 3.1. For c_α with $\alpha = [0; 1+d_1, d_2, d_3\ldots]$, it is well-known that, for all $n \ge 2$, $s_ns_{n-1}(xy)^{-1} = s_{n-1}s_n(yx)^{-1}$, where $x, y \in \{a, b\}$, $x \ne y$, and $xy \subseteq_s s_{n-1}$. This is known as the *Near-Commutative Property* of the words s_n and s_{n-1}. Because $s_ns_{n-1}(xy)^{-1} = s_nD_{n-2}$ and $s_{n-1}s_n(yx)^{-1} =$

$s_{n-1}D_{n-1}$, Proposition 4.9 is merely an extension of this property to standard episturmian words $\mathbf{s}$. It is also worthwhile noting that Proposition 4.9 shows that s_n is a prefix of $s_{n-1}s_n$.

Proposition 4.2 implies that $|s_{n+1}| - |D_n| = |s_n| - |D_{n-k+1}|$, and hence $|G_{n+1,1}| = |G_{n,k-1}|$. In fact:

Proposition 4.11 *For all $n \geq 1$, $G_{n,1} = \widetilde{G}_{n-1,k-1}$.*

Proposition 4.12 *Let $1 \leq i \leq k$ and $1 \leq r \leq k-1$. For all $n \geq 0$,*

(i) *$a_i \subseteq_p G_{n,r}$ if $n \equiv i + r - 1 \pmod{k}$;*

(ii) *$a_i \subseteq_s G_{n,r}$ if $n \equiv i - 1 \pmod{k}$.*

Hereafter, we set $d_{-j} = 0$ for $j \geq 0$.

4.2 Singular n-words of the r-th kind

By definition of the words $(s_n)_{n\geq 1-k}$ (see (3.2)) and the fact that $\mathbf{s} = \lim_{n\to\infty} s_n$, one deduces that, for any $n \geq 0$, $\mathbf{s}$ can be written as a concatenation of blocks of the form s_n, s_{n-1}, ... , s_{n-k+1}, i.e.,

$$\begin{aligned}\mathbf{s} = [((&s_n^{d_{n+1}} s_{n-1}^{d_n} \cdots s_{n-k+2}^{d_{n-k+3}} s_{n-k+1})^{d_{n+2}} s_n^{d_{n+1}} \cdots s_{n-k+3}^{d_{n-k+4}} s_{n-k+2})^{d_{n+3}} \\ &(s_n^{d_{n+1}} s_{n-1}^{d_n} \cdots s_{n-k+2}^{d_{n-k+3}} s_{n-k+1})^{d_{n+2}} s_n^{d_{n+1}} \cdots s_{n-k+4}^{d_{n-k+5}} s_{n-k+3}]^{d_{n+4}} \cdots . \qquad (4.4)\end{aligned}$$

We shall call this unique decomposition the *n-partition* of $\mathbf{s}$. This will be a useful tool in our subsequent analysis of powers of words occurring in $\mathbf{s}$ (Section 6, to follow).

Remark 4.13 Since each factor of $\mathbf{s}$ has exactly k different return words, two consecutive s_{n+1-i} blocks ($1 \leq i \leq k$) of the n-partition are separated by a word V, of which there are k different possibilities. From now on, it is advisable to keep this observation in mind.

Lemma 4.14 *Let $1 \leq r \leq k-1$. For any $n \in \mathbb{N}^+$, a factor u of length $|s_n|$ of $\mathbf{s}$ is a factor of at least one of the following words:*

- *$C_j(s_n)$, $0 \leq j \leq |s_n| - 1$;*
- *$s_{n-r}^{d_{n-r+1}-1} \cdots s_{n-k+1}^{d_{n-k+2}} s_{n-k} s_{n-1}^{d_n} \cdots s_{n-r+1}^{d_{n-r+2}} s_{n-r} s_n$ if $n \geq r$;*
- *$a_{n+1} s_n a_{n+1}^{-1} a_{n-r+k+1} s_n$ if $n < r$.*

Remark 4.15 The word $s_{n-r}^{d_{n-r+1}-1} \cdots s_{n-k+1}^{d_{n-k+2}} s_{n-k} s_{n-1}^{d_n} \cdots s_{n-r+1}^{d_{n-r+2}} s_{n-r}$ ($1 \leq r \leq k-1$) has length $|s_n|$.

Lemma 4.16 *Let $1 \leq r \leq k-1$. For any $n \geq r$, we have*

$$s_{n-r}^{d_{n-r+1}-1}\cdots s_{n-k+1}^{d_{n-k+2}}s_{n-k}s_{n-1}^{d_n}\cdots s_{n-r+1}^{d_{n-r+2}}s_{n-r} = D_{n-r}\widetilde{G}_{n,r},$$

and for $1 \leq n < r$, $a_{n+1}s_n a_{n+1}^{-1}a_{n-r+k+1} = \widetilde{G}_{n,r}$.

Whence, it is now plain to see that each word $\widetilde{G}_{n,r}s_n = \widetilde{G}_{n,r}D_{n-r}G_{n,r}$ is a factor of $\mathbf{s}$. We will now partition the set of factors of length $|s_n|$ of $\mathbf{s}$ into k disjoint classes.

Theorem 4.17 *Let $1 \leq r \leq k-1$. For any $n \in \mathbb{N}^+$, the set of factors of length $|s_n|$ of $\mathbf{s}$ can be partitioned into the following k disjoint classes:*

- $\Omega_n^0 := \mathcal{C}(s_n) = \{C_j(s_n) \ : \ 0 \leq j \leq |s_n| - 1\}$;
- $\Omega_n^r := \{w \in \mathcal{A}_k^* \ : \ |w| = |s_n|$ *and* $w \prec x^{-1}\widetilde{G}_{n,r}D_{n-r}G_{n,r}x^{-1}\}$*, where x is the last letter of $G_{n,r}$.*

That is, $\Omega_{|s_n|}(\mathbf{s}) = \Omega_n^0 \,\dot{\cup}\, \Omega_n^1 \,\dot{\cup} \cdots \dot{\cup}\, \Omega_n^{k-1}$.

Let us remark that $\widetilde{\Omega}_n^r := \{\widetilde{w} \ : \ w \in \Omega_n^r\} = \Omega_n^r$ since $x^{-1}\widetilde{G}_{n,r}D_{n-r}G_{n,r}x^{-1}$ is a palindrome. We shall call the factors of $\mathbf{s}$ in Ω_n^r the *singular n-words of the r-th kind*. Such words will play a key role in our study of powers of words occurring in $\mathbf{s}$.

Evidently, for Sturmian words c_α, $\Omega_n^1 = \{w_n\}$ and we have $\Omega_{|s_n|}(c_\alpha) = \mathcal{C}(s_n) \cup \{w_n\}$.

5 Index

A word of the form $w = (uv)^n u$ is written as $w = z^r$, where $z = uv$ and $r := n + |u|/|z|$. The rational number r is called the *exponent* of z, and w is said to be a *fractional power*.

Now suppose $\mathbf{x}$ is an infinite word. For any $w \prec \mathbf{x}$, the *index* of w in $\mathbf{x}$ is given by the number

$$\operatorname{ind}(w) = \sup\{r \in \mathbb{Q} \ : \ w^r \prec \mathbf{x}\},$$

if such a number exists; otherwise, w is said to have infinite index in $\mathbf{x}$. Furthermore, the greatest number r such that w^r is a prefix of $\mathbf{x}$ is called the *prefix index* of w in $\mathbf{x}$. Obviously, the prefix index is zero if the first letter of w differs from that of $\mathbf{x}$, and it is infinite if and only if $\mathbf{x}$ is purely periodic.

The next two results extend those of Berstel [2].

Lemma 5.1 *For all $n \geq 1$, the prefix index of s_n in $\mathbf{s}$ is $1 + d_{n+1} + |D_{n-k}|/|s_n|$.*

Lemma 5.2 *For all $n \geq 1$, the index of s_n as a factor of $\mathbf{s}$ is $\operatorname{ind}(s_n) = 2 + d_{n+1} + |D_{n-k}|/|s_n|$, and hence $\mathbf{s}$ contains cubes.*

6 Powers occurring in s

For each $m,\ l \in \mathbb{N}$ with $l \geq 2$, let us define the following set of words:

$$\mathcal{P}(m;l) := \{w \in \mathcal{A}_k^* \ : \ |w| = m,\ w^l \prec \mathbf{s}\},$$

where $\mathbf{s}$ is the k-strict standard episturmian word over $\mathcal{A}_k = \{a_1, a_2, \ldots, a_k\}$ with directive word Δ given by (3.1). Also, let $p(m;l) := |\mathcal{P}(m;l)|$.

The next theorem is a generalization of Theorem 1 in [6]. It gives all the lengths m such that there is a non-trivial power of a word of length m in $\mathbf{s}$. Firstly, let us define the following k sets of lengths for fixed $n \in \mathbb{N}^+$:

$$\begin{aligned}
\mathcal{D}_1(n) &:= \{r|s_n| \ : \ 1 \leq r \leq d_{n+1}\},\\
\mathcal{D}_i(n) &:= \{|s_n^r s_{n-1}^{d_n} \cdots s_{n+2-i}^{d_{n+3-i}} s_{n+1-i}| \ : \ 1 \leq r \leq d_{n+1}\}, \quad 2 \leq i \leq k-1,\\
\mathcal{D}_k(n) &:= \{|s_n^r s_{n-1}^{d_n} \cdots s_{n+2-k}^{d_{n+3-k}} s_{n+1-k}| \ : \ 1 \leq r \leq d_{n+1} - 1\}.
\end{aligned}$$

Theorem 6.1 *Let $m,\ n \in \mathbb{N}^+$ be such that $|s_n| \leq m < |s_{n+1}|$ and suppose $m \notin \bigcup_{i=1}^k \mathcal{D}_i(n)$. Then $p(m;l) = 0$ for all $l \geq 2$.*

Remark 6.2 Put simply, the above theorem states that if $m \notin \bigcup_{i=1}^k \mathcal{D}_i(n)$ for some n, then there are no l-th powers of words of length m in $\mathbf{s}$ for any $l \geq 2$. Equivalently, if $w^l \prec \mathbf{s}$ with $|s_n| \leq |w| < |s_{n+1}|$, then $|w| \in \bigcup_{i=1}^k \mathcal{D}_i(n)$. For instance, if $k = 3$ and $|s_n| \leq m < |s_{n+1}|$ with

$$m \notin \{|s_n^r|, |s_n^r s_{n-1}| : 1 \leq r \leq d_{n+1}\} \cup \{|s_n^r s_{n-1}^{d_n} s_{n-2}| : 1 \leq r \leq d_{n+1} - 1\},$$

then $p(m;l) = 0$ for all $l \geq 2$. For the particular case of the Tribonacci sequence, this implies that if w^l is a factor, then $|w| \in \{|s_n|, |s_n| + |s_{n-1}|\}$ for some n, where the lengths $(|s_i|)_{i\geq 0}$ are the *Tribonacci numbers*: $T_0 = 1$, $T_1 = 2$, $T_2 = 4$, $T_i = T_{i-1} + T_{i-2} + T_{i-3}$, $i \geq 3$.

The proof of Theorem 6.1 requires several lemmas (Lemmas 6.3–6.5 below). Let us first observe that in the n-partition of $\mathbf{s}$ (see (4.4)) to the left of each s_n block, there is an s_{n+1-j} block for some $j \in [1,k]$. Also note that each s_{n+1-j} is a prefix of s_n. Furthermore, to the left of each s_{n+1-i} block is another s_{n+1-i} block or an s_{n+2-i} block, for each $i \in [2,k]$.

Lemma 6.3 *Let $n \in \mathbb{N}^+$. Consider a word $w \prec \mathbf{s}$ of the form $w = us_nv$ for some words $u,\ v \in \mathcal{A}_k^*$, $u \neq \varepsilon$.*

(i) *If $w = u_1u_2$, where $u_1 \subseteq_s s_{n+1-i}$ for some $i \in [1,k]$ and $u_2 \subseteq_p s_n$, then $u_1 = u$.*

(ii) *If $w = u_1 s_{n+1-i} u_2$ for some $i \in [2,k]$, where $u_1 \subseteq_s s_{n+2-i}$ and $u_2 \subseteq_p s_n$, then $u_1 = u$ or $u_1 s_{n+1-i} = u$.*

(iii) *If $w = u_1 s_{n+1-i} u_2$ for some $i \in [2, k-1]$, where $u_1 \subseteq_s s_{n+1-i}$ and $u_2 \subseteq_p s_n$, then $u_1 = u$ or $u_1 s_{n+1-i} = u$.*

Lemma 6.4 *Let $c \in \mathcal{A}_k$ and $n \in \mathbb{N}$ be fixed. Consider an occurrence of cs_n in $\mathbf{s}$. Then the letter c is the last letter of a block s_{n+1-i} of the n-partition of $\mathbf{s}$, for some $i \in [1, k]$, and the integer i (equiv. the block s_{n+1-i}) is uniquely determined by c. In particular, in every occurrence of $s_{n+1-i}s_n$ in $\mathbf{s}$, the word s_{n+1-i} is a block in the n-partition of $\mathbf{s}$.*

That is, occurrences of words w containing cs_n ($c \in \mathcal{A}_k$) must be aligned to the n-partition of $\mathbf{s}$. Now we have an analogue of Lemma 3.5 in [5]:

Lemma 6.5 *Let $n \in \mathbb{N}^+$ and suppose $u \prec \mathbf{s}$ with $|s_n| \leq |u| < |s_{n+1}|$. Then the following assertions hold.*

(1) *For all $i \in [1, k]$, there is at most one position in s_{n+1-i} such that any occurrence of u in $\mathbf{s}$ which starts in some s_{n+1-i} block of the n-partition of $\mathbf{s}$ must start at this particular position in s_{n+1-i}.*

(2) *For all $i \in [1, k-1]$, if u can start at position l in s_{n+1-i} and at position m in s_{n-i}, then $l = m$.*

Notation 6.1 Given $l \in \mathbb{N}$ and $w \in \mathcal{A}_k^*$, denote by $\mathrm{Pref}_l(w)$ the prefix of w of length l if $|w| \geq l$, w otherwise. Likewise, denote by $\mathrm{Suff}_l(w)$ the suffix of w of length l if $|w| \geq l$, w otherwise. Recall that Ω_n^r denotes the set of singular n-words of the r-th kind ($1 \leq r \leq k-1$), as defined in Theorem 4.17.

Lemma 6.6 *Let $n \in \mathbb{N}^+$ and suppose $w \in \Omega^1_{n+1-i}$ for some $i \in [1, k-1]$. Then w begins with $v := \mathrm{Suff}_l(x^{-1}\widetilde{G}_{n+1-i,1})$ for some $l \in \mathbb{N}$ with $1 \leq l \leq |G_{n+1-i,1}|-1$. Moreover, the word vs_{n+1-i} occurs at position p in $\mathbf{s}$ if and only if the n-partition of $\mathbf{s}$ contains an s_n starting at postion $p+l$ and an s_{n-i} ending at position $p+l-1$. In particular, w occurs at exactly those positions where vs_{n+1-i} occurs in $\mathbf{s}$.*

Note 6.7 It is assumed that $n \geq i$.

Consider two distinct occurrences of a factor w in $\mathbf{s}$, say

$$\mathbf{s} = uw\mathbf{v} = u'w\mathbf{v}', \quad |u'| > |u|,$$

where $\mathbf{v}$, $\mathbf{v}' \in \mathcal{A}_k^\omega$. These two occurrences of w in $\mathbf{s}$ are said to be *positively separated* (or *disjoint*) if $|u'| > |uw|$, in which case $u' = uwz$ for some $z \in \mathcal{A}_k^+$, and hence $\mathbf{s} = uwzw\mathbf{v}'$.

Lemma 6.8 *For any $n \in \mathbb{N}^+$, successive occurrences of a singular word $w \in \bigcup_{j=1}^{k-1} \Omega_n^j$ in $\mathbf{s}$ are positively separated.*

The next lemma follows from Lemmas 5.2, 6.6 and 6.8.

Lemma 6.9 *Let $n \in \mathbb{N}^+$ and suppose $u \prec \mathbf{s}$ with $|u| = |s_n|$. Then $u^2 \prec \mathbf{s}$ if and only if $u \in \mathcal{C}(s_n)$. In particular, if u is a singular word of any kind of $\mathbf{s}$, then $u^2 \not\prec \mathbf{s}$.*

More generally, we have the following result.

Lemma 6.10 *Let $n \in \mathbb{N}^+$ and suppose $u^2 \prec \mathbf{s}$ with $|s_n| \le |u| < |s_{n+1}|$. Then u does not contain a singular word from the set Ω^1_{n+1-i} for any $i \in [1, k-1]$.*

6.1 Squares

The next two main theorems concern squares of factors of $\mathbf{s}$ of length $m < d_1 + 1 = |s_1|$ and length $m \ge |s_1|$, respectively.

A letter a in a finite or infinite word w is *separating for w* if any factor of length 2 of w contains the letter a. For example, a is separating for the infinite word $(aaba)^\omega$. If a is separating for an infinite word $\mathbf{x}$, then it is clearly separating for any factor of $\mathbf{x}$. According to [8, Lemma 4], since the standard episturmian word $\mathbf{s}$ begins with a_1, the letter a_1 is separating for $\mathbf{s}$ and its factors. Moreover, a_1 occurs in runs of length d_1 or $d_1 + 1$ in $\mathbf{s}$ (inspect the 0-partition of $\mathbf{s}$), and the following is deduced:

Theorem 6.11 *For $1 \le r \le d_1$, we have*

$$p(r;2) = \begin{cases} 1 & \text{if } r \le (d_1+1)/2, \\ 0 & \text{if } r > (d_1+1)/2. \end{cases}$$

In particular, $\mathcal{P}(r;2) = \{(a_1^r)^2\}$ for $r \le (d_1+1)/2$, and $\mathcal{P}(r;2) = \emptyset$ for $r > (d_1+1)/2$.

Theorem 6.12 *Let n, $r \in \mathbb{N}^+$.*

(i) *For $1 \le r \le d_{n+1}$,*

$$p(|s_n^r|;2) = \begin{cases} |s_n| & \text{if } 1 \le r < 1 + d_{n+1}/2, \\ |D_{n-k}| + 1 & \text{if } d_{n+1} \text{ is even and } r = 1 + d_{n+1}/2, \\ 0 & \text{if } 1 + d_{n+1}/2 < r \le d_{n+1}. \end{cases} \tag{6.1}$$

That is,

$$\mathcal{P}(|s_n^r|;2) = \begin{cases} \{C_j(s_n^r) : 0 \le j \le |s_n| - 1\} & \text{if } 1 \le r < 1 + d_{n+1}/2, \\ \{C_j(s_n^r) : 0 \le j \le |D_{n-k}|\} & \text{if } d_{n+1} \text{ is even} \\ & \text{and } r = 1 + d_{n+1}/2, \\ \emptyset & \text{if } 1 + d_{n+1}/2 < r \le d_{n+1}. \end{cases} \tag{6.2}$$

(ii) *For* $1 \le r \le d_{n+1}$ *and* $i \in [2,k]$ *(with* $r \neq d_{n+1}$ *if* $i = k$*), we have*

$$p(|s_n^r s_{n-1}^{d_n} \cdots s_{n+2-i}^{d_{n+3-i}} s_{n+1-i}|; 2) = |D_{n+1-i}| + 1. \tag{6.3}$$

That is,

$$\begin{aligned} &\mathcal{P}(|s_n^r s_{n-1}^{d_n} \cdots s_{n+2-i}^{d_{n+3-i}} s_{n+1-i}|; 2) \\ &\quad = \{C_j(s_n^r s_{n-1}^{d_n} \cdots s_{n+2-i}^{d_{n+3-i}} s_{n+1-i}) : 0 \le j \le |D_{n+1-i}|\}. \end{aligned} \tag{6.4}$$

Remark 6.13 For standard Sturmian words c_α, we have $s_n = D_{n-1}xy$, where x, $y \in \{a, b\}$ $(x \neq y)$, and hence $|D_{n-1}| = q_n - 2$ for all $n \ge 1$. Accordingly, Theorem 6.12 agrees with Theorem 3 in [6] for the case of a two-letter alphabet.

Theorem 6.11 is trivial, whereas the proof of Theorem 6.12 requires the following two lemmas.

Lemma 6.14 *Let* $n \in \mathbb{N}^+$ *and let* $u^2 = u^{(1)}u^{(2)}$ *be an occurrence of* u^2 *in* $\mathbf{s}$*, where* $|s_n| \le |u| < |s_{n+1}|$.

(i) *For all* $n \ge 1$*, if* $|u| = |s_n^r|$ *with* $1 \le r \le d_{n+1}$*, then* $u^{(1)}$ *begins in an* s_n *block of the n-partition of* $\mathbf{s}$*. Moreover,* u^2 *is a factor of* $s_n^{d_{n+1}+2} s_n v^{-1} = s_n^{d_{n+1}+2} D_{n-k}$*, where* $|v| = |s_n| - |D_{n-k}|$.

(ii) *Let* $i \in [2, k-1]$*. For all* $n \ge i - 1$*, if* $|u| = |s_n^r s_{n-1}^{d_n} \cdots s_{n+2-i}^{d_{n+3-i}} s_{n+1-i}|$ *with* $1 \le r \le d_{n+1}$*, then* $u^{(1)}$ *starts in an* s_n *block and contains an* s_{n+1-i} *block that is followed by an* s_n *block in the n-partition of* $\mathbf{s}$*. Moreover,* u^2 *is a factor of* $(s_n^r s_{n-1}^{d_n} \cdots s_{n+2-i}^{d_{n+3-i}} s_{n+1-i})^2 D_{n+1-i}$*, which is a factor of*

$$(s_n^{d_{n+1}} s_{n-1}^{d_n} \cdots s_{n+2-i}^{d_{n+3-i}} s_{n+1-i})^2 D_{n+1-i}.$$

(iii) *For all* $n \ge k - 1$*, if* $|u| = |s_n^r s_{n-1}^{d_n} \cdots s_{n+2-k}^{d_{n+3-k}} s_{n+1-k}|$ *with* $1 \le r \le d_{n+1} - 1$*, then* $u^{(1)}$ *starts in an* s_n *block and contains an* s_{n+1-k} *block of the n-partition of* $\mathbf{s}$*. Moreover,* u^2 *is a factor of*

$$(s_n^r s_{n-1}^{d_n} \cdots s_{n+2-k}^{d_{n+3-k}} s_{n+1-k})^2 D_{n+1-k}\,,$$

which is a factor of

$$s_{n+1}^2 = (s_n^{d_{n+1}} s_{n-1}^{d_n} \cdots s_{n+2-k}^{d_{n+3-k}} s_{n+1-k})^2.$$

Lemma 6.15 *For all* n, $r \in \mathbb{N}^+$ *and* $i \in [2, k]$*, the word*

$$v := s_n^r s_{n-1}^{d_n} \cdots s_{n+2-i}^{d_{n+3-i}} s_{n+1-i}$$

is primitive.

6.2 Cubes and higher powers

Our subsequent analysis of cubes and higher powers occurring in $\mathbf{s}$ is now an easy task due to the above consideration of squares. Extending Theorem 6.12 (see Theorem 6.17 below), only requires the following lemma, together with arguments used in the proof of Theorem 6.12.

Lemma 6.16 *Let $n \in \mathbb{N}^+$ and suppose $u^3 \prec \mathbf{s}$ with $|s_n| \le |u| < |s_{n+1}|$. Then u^3 does not contain a singular word from the set Ω^1_{n+1-i} for any $i \in [1, k-1]$.*

Theorem 6.17 *Let n, r, $l \in \mathbb{N}^+$, $l \ge 3$.*

(i) *For $1 \le r \le d_{n+1}$,*

$$p(|s_n^r|; l) = \begin{cases} |s_n| & \textit{if } 1 \le r < (d_{n+1}+2)/l, \\ |D_{n-k}| + 1 & \textit{if } r = (d_{n+1}+2)/l, \\ 0 & \textit{if } (d_{n+1}+2)/l < r \le d_{n+1}. \end{cases} \tag{6.5}$$

That is,

$$\mathcal{P}(|s_n^r|; l) = \begin{cases} \{C_j(s_n^r) : 0 \le j \le |s_n| - 1\} & \textit{if } 1 \le r < (d_{n+1}+2)/l, \\ \{C_j(s_n^r) : 0 \le j \le |D_{n-k}|\} & \textit{if } r = (d_{n+1}+2)/l, \\ \emptyset & \textit{if } (d_{n+1}+2)/l < r \le d_{n+1}. \end{cases} \tag{6.6}$$

(ii) *For $1 \le r \le d_{n+1}$ and $i \in [2, k]$ (with $r \ne d_{n+1}$ if $i = k$), we have*

$$p(|s_n^r s_{n-1}^{d_n} \cdots s_{n+2-i}^{d_{n+3-i}} s_{n+1-i}|; l) = 0. \tag{6.7}$$

Example 6.18 Let us define the *k-bonacci word* to be the standard episturmian word $\boldsymbol{\eta}_k \in \mathcal{A}_k^\omega$ with directive word $(a_1a_2\cdots a_k)^\omega$. Since all $d_i = 1$, we have $s_n = s_{n-1}s_{n-2}\cdots s_{n-k}$ for all $n \ge 1$ (and the lengths $|s_n|$ are the k-bonacci numbers). Thus, for fixed $n \in \mathbb{N}^+$ and $l \ge 2$, if $w^l \prec \boldsymbol{\eta}_k$ with $|s_n| \le |w| < |s_{n+1}|$, then we necessarily have $|w| = |s_n| + |s_{n-1}| + \cdots + |s_{n+1-i}|$ for some $i \in [1, k-1]$ (by Theorem 6.1). The preceding theorems reveal that

$$\mathcal{P}(1; 2) = \{a_1\}, \quad \mathcal{P}(|s_n|; 2) = \mathcal{C}(s_n) = \Omega_n^0$$

and

$$\mathcal{P}(|s_n|; 3) = \{C_j(s_n) \ : \ 0 \le j \le |D_{n-k}|\}.$$

Furthermore, for each $i \in [2, k-1]$, we have

$$\mathcal{P}(|s_ns_{n-1}\cdots s_{n+1-i}|; 2) = \{C_j(s_ns_{n-1}\cdots s_{n+1-i}) \ : \ 0 \le j \le |D_{n+1-i}|\}.$$

All other $\mathcal{P}(|w|; l) = \emptyset$, $l \ge 2$. In particular, k-bonacci words are 4-power free.

7 Concluding remarks

Theorems 6.11, 6.12 and 6.17 also suffice to describe all integer powers occurring in any (episturmian) word $\mathbf{t} \in \mathcal{A}_k^\omega$ that is equivalent to $\mathbf{s}$. (See [12, Theorem 3.10] for a definition of such $\mathbf{t}$.) The problem of determining all integer powers occurring in general standard episturmian words (with not all d_i necessarily positive) remains open.

References

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

[2] J. Berstel, On the index of Sturmian words, in *Jewels Are Forever*, *Springer-Verlag*, Berlin, 1999, pp. 287–294.

[3] J. Berstel, P. Séébold, Sturmian words, in: *M. Lothaire, Algebraic Combinatorics On Words, Encyclopedia of Mathematics and its Applications*, vol. 90, *Cambridge University Press*, U.K., 2002, pp. 45–110.

[4] W.-T. Cao, Z.-Y. Wen, Some properties of the factors of Sturmian sequences, *Theoret. Comput. Sci.* **304** (2003), 365–385.

[5] D. Damanik, D. Lenz, The index of Sturmian sequences, *European J. Combin.* **23** (2002), 23–29.

[6] D. Damanik, D. Lenz, Powers in Sturmian sequences, *European J. Combin.* **24** (2003) 377–390.

[7] A. de Luca, Sturmian words: structure, combinatorics and their arithmetics, *Theoret. Comput. Sci.* **183** (1997), 45–82.

[8] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* **255** (2001), 539–553.

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

[10] A. Glen, Occurrences of palindromes in characteristic Sturmian words, *Theoret. Comput. Sci.*, to appear.

[11] A. Glen, Conjugates of characteristic Sturmian words generated by morphisms, *European J. Combin.* **25** (7) (2004), 1025–1037.

[12] J. Justin, G. Pirillo, Episturmian words and episturmian morphisms, *Theoret. Comput. Sci.* **276** (2002), 281–313.

[13] J. Justin, L. Vuillon, Return words in Sturmian and episturmian words, *Theor. Inform. Appl.* **34** (5) (2000), 343–356.

[14] F. Levé, P. Séébold, Conjugation of standard morphisms and a generalization of singular words, *Bull. Belg. Math. Soc. Simon Stevin* **10** (5) (2003), 737–747.

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

[16] B. Tan, Z.-Y. Wen, Some properties of the Tribonacci sequence. Preprint.

[17] Z.-X. Wen, Z.-Y. Wen, Some properties of the singular words of the Fibonacci word, *European J. Combin.* **15** (6) (1994), 587–598.

Post Correspondence Problem for morphisms with unique blocks

Vesa Halava, Tero Harju, Juhani Karhumäki[*], *Michel Latteux*[†]

Abstract

In the Post Correspondence Problem (PCP) an instance (h, g) consists of two morphisms h and g, and the problem is to determine whether or not there exists a word w such that $h(w) = g(w)$. Here we prove that the PCP is decidable for instances with unique blocks and that the infinite PCP is decidable for instance with unique continuation in the construction of the solution. These results establish a new larger class of decidable instances of the PCP, including the class of marked instances.

1 Introduction

In the *Post Correspondence Problem* (PCP, for short), we are given two morphisms $h, g\colon A^* \to B^*$, where A and B are finite alphabets, and we are asked whether or not there exists a nonempty word $w \in A^*$ such that $h(w) = g(w)$. The pair (h, g) is called an *instance* of the PCP and a word $w \in A^+$ is a *solution* of the instance (h, g) if $h(w) = g(w)$. The set of all solutions,

$$E(h, g) = E(I) = \{w \in A^+ \mid h(w) = g(w)\},$$

is called the *equality set* of the instance $I = (h, g)$. The *size* of an instance I is $|A|$, that is, the cardinality of the domain alphabet of the morphisms.

The PCP is undecidable in this general form (see [12]). The borderline between decidable and undecidable sets of instances has been investigated in several occasions by restricting the instances of the PCP. For example, it is an easy exercise to show that the *unary PCP*, where the domain alphabet has only one letter, is decidable. An instance (h, g) of the PCP, where $h, g\colon A^* \to B^*$, is *binary* if $|A| = 2$. It was proved in [1] that the PCP is decidable for binary instances; see also [5] or [6] for a somewhat simpler proof. On the other hand, the PCP is undecidable for instances with domain alphabets A satisfying $|A| \geq 7$ (see [11]).

[*] Department of Mathematics and TUCS - Turku Centre for Computer Science, University of Turku, FIN-20014 Turku, Finland, vehalava@utu.fi, harju@utu.fi, karhumak@staff.cs.utu.fi. Supported by the Academy of Finland under grant 44087.

[†] Université des Sciences et Technologies de Lille, Bâtiment M3, 59655 Villeneuve d'Ascq Cédex, France, latteux@lifl.fr

Another known borderline between decidability and undecidability is provided by *marked* and *prefix* morphisms. A morphism $h\colon A^* \to B^*$ is said to be marked if the images $h(a)$ and $h(b)$ of any two different letters $a, b \in A$ begin with a different letter. The problem where the instances are pairs of marked morphisms is called the *marked PCP* (consisting of *marked instances*). It was proved in [9], that the marked PCP is decidable in general. On the other hand, in [13] it was shown that the PCP is undecidable for instances of prefix morphisms. A morphism is called prefix if no image of a letter is a prefix of an image of another letter.

In this paper we study instances of the PCP, which are not necessary marked, but they can be reduced to instances of the marked PCP. This reduction is due to the unique condition satisfied by the original instance.

We also study the infinite PCP in the infinite solutions of the instances (h, g), which satisfy the condition of the unique continuation. Two (finite) words u and v are said to be *comparable*, if one is a prefix of the other. Let $\omega = a_1a_2\cdots$ be an infinite word over A where $a_i \in A$ for each index $i = 1, 2, \cdots$. Note that $h(\omega) = g(\omega)$ if the morphisms h and g *agree* on ω, that is, if $h(u)$ and $g(u)$ are comparable for all finite prefixes u of ω. We also say that such an infinite word ω is an *infinite solution* of the instance $I = (h, g)$.

The problem whether or not a given instance of the PCP has an infinite solution is called naturally the *infinite PCP*, or *ωPCP*, for short. It was shown by Ruohonen [13] that there is no algorithm to determine whether a general instance of the PCP has an infinite solution. I

It was proved in [3] that the ωPCP is decidable for marked instances of the PCP. Later, using the previous result, it was shown in [7] that the ωPCP is decidable for all binary instances. Recently, it was proved in [4] that the ωPCP is undecidable for instances of size 9.

We shall now fix some notation. The *empty word* is denoted by ε. The length of a word u is denoted by $|u|$. A word $u \in A^*$ is said to be a *prefix* of a word $v \in A^*$, denoted by $u \le v$, if $v = uw$ for some $w \in A^*$. Also, if $u \neq \varepsilon$ and $w \neq \varepsilon$ in $v = uw$, then u is a *proper* prefix of v, and this is denoted by $u < v$. Recall that u and v are comparable, $u \bowtie v$, if $u \le v$ or $v \le u$. The longest common prefix of the words u and v is denoted by $u \wedge v$. If $v = uw$ then we also write $u = vw^{-1}$ and $w = u^{-1}v$.

A word $u \in A^*$ is said to be a *suffix* of a word $v \in A^*$, if $v = wu$ for some $w \in A^*$. If $u \neq \varepsilon$ and $u \neq v$, then the suffix u is *proper.*

Finally, a morphism h is called *non-erasing* if $h(w) = \varepsilon$ implies that $w = \varepsilon$, i.e., no image of a letter is empty for h.

2 Unique block instances

The basic result on which we build the results of this paper is the following, see [9] or [2] for proofs.

Theorem 2.1 *The PCP is decidable for marked instances of any size.*

The basic idea of the proofs is the concept of *blocks*. We aim at a decision method for the PCP, but we start with the following simpler problem:

Problem 2.1 *Given an instance $I = (h, g)$ of the marked PCP, where $h, g\colon A^* \to B^*$, and $b \in B$. Does there exist $x, y \in A^+$ such that $h(x) = g(y)$ and $b \leq h(x)$?*

We do not look for solutions for $h(x) = g(x)$ here, but only for $h(x) = g(y)$, and we additionally require that $h(x)$ begins with a specific letter b. This problem is known to be decidable for two morphisms in general, the reasoning being that the language $h(A^*) \cap bB^*$ is regular, and there exist such words x and y if and only if

$$\big(h(A^*) \cap bB^*\big) \cap \big(g(A^*) \cap bB^*\big) \neq \emptyset, \tag{2.1}$$

and the emptiness problem is decidable for regular languages. If $h(u) = g(v)$ and $h(u') \neq g(v')$ for all $\varepsilon < u' \leq u$ and $\varepsilon < v' \leq v$ with $(u', v') \neq (u, v)$, and $b \leq h(u)$, then the pair (u, v) is called a *block* or a *a block for the letter* b, and u and v are called the *block words*. Denote by $S_b(h, g)$ or S_b, for short, the set of all blocks for letter b.

If (u, v) is a solution of the equation $h(x) = g(y)$, then there exist decompositions $u = u_1u_2 \cdots u_k$ and $v = v_1v_2 \cdots v_k$ of u and v such that $(u_i, v_i) \in S_{b_i}$ for $b_i \in B$ for $i = 1, \ldots, k$. Thus taking $u = w = v$, where w is a solution of the marked instance (h, g), there exists a *block decomposition* of w,

$$w = u_1u_2 \cdots u_k = v_1v_2 \cdots v_k, \tag{2.2}$$

where $(u_i, v_i) \in S_{b_i}$ for $b_i \in B$ for $i = 1, \ldots, k$. This means that each solution is a concatenation of blocks.

For marked instance (h, g) the block for every letter a is unique, and, therefore, every solution has unique block decomposition, see [9] or [2] for proofs.

Let us now define the first property of unique continuation:

> **UC1**. The instance (h, g), where $h, g\colon A^* \to B^*$, of the PCP, is called *unique block instance* if, for every letter $a \in A$,
>
> 1. the block (au, v) is unique, and
> 2. the block (u, av) is unique,
>
> if they exist.

Note that in (UC1) we assume that every letter of $a \in A$ is the first letter of at most one block in block words of both h and g. Clearly, the condition (UC1) implies that h and g are non-erasing. If an instance I satisfies (UC1), and assume that (u, v) is a block and $a \leq u$, $a \in A$, then we denote $\beta(a) = (u, v)$.

Example 2.2 We give an example of a non-marked unique block instance.

	a	b	c	d
h	ab	abc	$cccc$	b
g	a	$babc$	cc	bbc

has unique blocks, the blocks are (ab, ab), (c, cc) and (db, b). For example, $h(adc^i) \bowtie g(adc^j)$ and $h(aac^i) \bowtie g(abc^j)$ for all i and j, but there is no block of the form (adc^i, adc^j) or (aac^i, abc^j). Clearly, $w = ab$ is a solution of the PCP and $\omega = adc^\omega$ is a solution of ωPCP for this instance.

Note that the instances satisfying the condition that, for all $b \in B$,

$$|S_b| \leq 1,$$

is a subclass of (UC1). We leave the details for the reader.

Now the condition (UC1) ensures the reduction to marked PCP.

Theorem 2.3 *The PCP is decidable for unique block instances.*

Proof Assume that $I = (h, g)$, where $h, g\colon A^* \to B^*$, is a unique block instance. Now define the instance $I' = (h', g')$ of the marked PCP, where $h', g'\colon C^* \to A^*$ and

$$C = \{a \in A \mid \beta(a) \text{ exists}\}.$$

We set for $a \in C$ and (the unique) block $\beta(a) = (u, v)$

$$h'(a) = u \quad \text{ and } g'(a) = v.$$

Clearly, h' and g' are marked by condition (UC1). Note that $a \leq h'(a)$ for all $a \in C$.

We prove that I' has a solution if and only I has. Indeed, assume, that I has a solution w, and let

$$u_1 u_2 \cdots u_k = w = v_1 v_2 \cdots v_k$$

be its block decomposition, where $(u_i, v_i) = \beta(a_i)$. The block decomposition is unique, since the first block (u_1, v_1) is unique for the letter $a_1 \leq w$ etc. Now, for $w' = a_1 a_2 \cdots a_k$,

$$h'(w') = u_1 u_2 \cdots u_k = w = v_1 v_2 \cdots v_k = g'(w'),$$

and w' is a solution of I'.

For the other direction, assume that I' has a solution $w' = a_1 a_2 \cdots a_k$. Now $w = h'(w') = g'(w')$ is a solution of I, since by the definition of I'

$$h(h'(w')) = h(u_1 u_2 \cdots u_k) = g(v_1 v_2 \cdots v_k) = g(g'(w')),$$

where $(u_i, v_i) = \beta(a_i)$ for some letters $a_i \in C$ for all i, since $h(u_i) = g(v_i)$ for all i.

The result follows by Theorem 2.1. □

An effective decision procedure uses exactly the same technique as the algorithm for marked PCP in [9]. Indeed, for an instance $I = (h, g)$ of the marked PCP, the *successor* $I' = (h', g')$ is build as in the proof of Theorem 2.3. This successor construction is iterated and a *successor sequence* $I^{(i)} = (h^{(i)}, g^{(i)})$ is defined. The conclusion is that this sequence is ultimately periodic, that is, there exist numbers n and d such that $I^{(i)} = I^{(i+d)}$ for all $i \geq n$. Finally, there exists a solution beginning with a letter a for I if and only if $h^{(i)}(a) = a = g^{(i)}(a)$ for all $i \geq n$. See [9] for more detailed study and proofs. Note that for an instance I of the marked PCP the letters for which a minimal solution appears can be detected and the minimal solution for each letter is unique, that is,

$$\begin{aligned} E_{\min}(I) &= E(I) \setminus E(I)^2 \\ &= \{w_1, w_2, \ldots, w_k \mid w_i \text{ is the minimal solution for letter } a_i\}\,. \end{aligned}$$

Now by the proof of Theorem 2.3, we obtain

Corollary 2.4 *Let $I = (h, g)$ be a unique block instance of the PCP and assume that the domain alphabet is A. The following sets can be effectively found:*

1. $S = \{a \in A \mid$ *there exists a solution w for $I, a \leq w$*$\}$
2. $E_{\min}(I)$ *is a finite marked set effectively computable. Marked here means that every element of $E(I)$ begins with different letter.*

In order to make use of Theorem 2.3, we must be able to prove that we may detect the unique block instances. Therefore, we need to prove

Theorem 2.5 *It is decidable, whether or not an instance (h, g) of the PCP is a unique block instance.*

Proof We establish a procedure for deciding whether or not an instance is a unique block instance.

Let $I = (h, g)$, where $h, g\colon A^* \to B^*$, be an instance of the PCP. For a letter $a \in A$, construct the minimal deterministic finite automata for the regular language

$$H_a = h(aA^*) \cap g(A^*).$$

This can be done by the usual tricks in the theory of finite automata, by first defining automata for languages $h(aA^*)$ and $g(A^*)$, and then using the construction in [10] for intersection. The minimal automaton can be found with so called *The Method of Quotients* in [10]. Let $\mathcal{A}$ be the automaton. Now I satisfies condition 1 in (UC1) for the letter a only if $\mathcal{A}$ is not *branching*, that is, there is a unique path from initial state to the final state in $\mathcal{A}$ reading a word $w_a \in H_a$. Indeed, then $H_a = w_a^*$. We still need to check that the word au such that $h(au) = w_a$ is unique, but this can be checked simply by trying all possible coverings of w_a by images of h. Now, only if au is unique, the condition 1 in

(UC1) is satisfied. We still need to decide whether or not the condition 2 holds, but this can done similarly to the check of uniqueness of au for all w_a, that is, check that the words v such that $g(v) = w_a$ are unique. Still, we need to check that all these (now unique) v's for w_a's begin with different letters.

□

Now the property (UC1) does not help in the ωPCP, since no reduction to marked infinite PCP can be established. Indeed, unique block instance can have an infinite solution without block decomposition and still be non-ultimately periodic.

For the ωPCP we need a more stronger unique continuation condition to have a decidable ωPCP.

3 Unique continuation instances

> **UC2**. The instance (h, g), where $h, g\colon A^* \to B^*$, of the PCP, is called *unique continuation instance*, if, for $u, v \in A^*$, $h(u) < g(v)$ or $g(v) < h(u)$, then there exists at most one letter x such that $h(ux) \bowtie g(v)$ or $h(u) \bowtie g(vx)$, respectively.

First of all,

Theorem 3.1 *It is decidable, whether or not an instance of the PCP is a unique continuation instance.*

Proof Let $I = (h, g)$, where $h, g\colon A^* \to B^*$, be an instance of the PCP. We define the following procedure called Continuation. The input is (a, h, g) for $a \in A$:

(1) Set $i = 1$, $x_1 = a$ and $y_1 = \varepsilon$, $S_g = S_h = \emptyset$.

(2) If $h(x_i) = g(y_i)$, then return Unique, case 1.

(3) Else if $g(y_i) < h(x_i)$, then if $s_i = g(y_i)^{-1}h(x_i) \in S_h$ return Unique, case 2. Else set $S_h := S_h \cup \{s_i\}$.

If the letter b such that $g(y_i b) \bowtie h(x_i)$ is unique, then set $x_{i+1} = x_i$, $y_{i+1} = y_i b$ and $i = i + 1$, GOTO 2. If no such b exists, return No block. Else return Not unique.

(4) Else if $h(x_i) < g(y_i)$, then if $s_i = h(x_i)^{-1}g(y_i) \in S_g$ return Unique, case 2. Else set $S_g := S_g \cup \{s_i\}$.

If the letter b such that $h(x_i b) \bowtie g(y_i)$ is unique, then set $y_{i+1} = y_i$, $x_{i+1} = x_i b$ and $i = i + 1$, GOTO 2. If no such b exists, return No block. Else return Not unique.

Now the instance satisfies condition (UC2) if and only if, for all $a \in A$, the procedure Continuation returns unique for both inputs (a, h, g) and (a, g, h). □

The following theorem follows, since a unique continuation instance is unique block instance. Indeed, the procedure Continuation can be transformed into a procedure which determines the unique block also, simply by returning the pair (x_i, y_i) if the procedure stop in the command line (2). Note that the words s_i in the procedure Continuation are called *suffix overflows.*

Theorem 3.2 *The PCP is decidable for unique continuation instances.*

The difference between unique block instances and unique continuation instances is that for unique block instance we may find several letters each overflow in steps command lines (3) and (4) both the equality $h(x_i) = g(y_i)$ is eventually satisfied only for one choice of the letter b. But in the unique continuation morphisms the choice of the next letter is always deterministic according to the overflow. Using this determinism we are able to prove that the ωPCP is decidable for unique continuation instances. The proof of the decidability of the ωPCP for unique continuation instances uses the idea of the proof of the next theorem proved in [3].

Theorem 3.3 *The ωPCP is decidable for marked instances.*

We are ready to prove our main theorem on ωPCP.

Theorem 3.4 *The ωPCP is decidable for unique continuation instances.*

Proof Assume that $I = (h, g)$, where $h, g\colon A^* \to B^*$, is an instance of the ωPCP, and I is a unique continuation instance. First of all assume that $E(I) = \emptyset$, since for any nonempty $w \in E(I)$, w^ω is an infinite solution and we are done.

The infinite solutions of I are of two type, either they have a block decomposition, or they do not have a block decomposition. We prove that the infinite solutions of both type can be detected. Note that we say that an infinite solution $\omega \in A^\omega$ has a block decomposition, if

$$\omega = u_1 u_2 \cdots = v_1 v_2 \cdots , \tag{3.1}$$

and (u_i, v_i) is block for some letter a_i for $i = 1, 2, \ldots$. Assume that $I' = (h', g')$ constructed as in the proof of Theorem 2.3. It is easy to prove that there exists an infinite solution ω with block decomposition in (3.1) for I if and only if $\omega' = a_1 a_2 \cdots$ is an infinite solution of the instance I' of the marked ωPCP. Since the marked ωPCP is decidable (see [3]), the solution with block decomposition can be detected. Note that, for any infinite solution ω' of the instance I', $h'(\omega')$ is an infinite solution of I.

The harder case is the infinite solution without block decomposition. Assume that ω is such an infinite solution, that is,

$$\omega = u_1 u_2 \cdots u_k \omega_1 = v_1 v_2 \cdots v_k \omega_2, \tag{3.2}$$

where (u_i, v_i) are blocks for $i = 1, \ldots, k$ and $\omega_1, \omega_2 \in A^\omega$, and k is maximal. The maximality here means, that ω_1 and ω_2 do not have block words as a prefix. We shall describe a procedure which detects these infinite solutions.

Since there is no block as a prefix of ω_1 and ω_2, it is necessary, that they both begin with letters such that no block exist for these letters. Otherwise, by unique continuation, there would be a whole block. Clearly, such letters are those for which the procedure Continuation returns "unique case 2". In this case, a suffix overflow appears again, and, since the instance is unique continuation instance, necessarily the overflows appear cyclically. Assume, that $(x_i, y_i) = (u, u')$ is the pair when the first repeated overflow appears the first time, and $(x_j, y_j) = (uv, u'w)$ when the same overflow appears the second time. It is immediate that $h(uv^\omega) = g(u'w^\omega)$. Therefore, $\omega_1 = uv^\omega$ and $\omega_2 = u'w^\omega$ for some letters $b \leq u$ and $c \leq u'$ and for any block (x, y), $b \not\leq x$ and $c \not\leq y$. We achieve that all possible ω_1 and ω_2 can be determined. For each letter that *disappears*, that is, there is no block for them, we construct the words ω_1 and ω_2 if they exist. Note that we check also whether or not $\omega_1 = \omega_2$ which would immediately imply that ω_1 is an infinite solution.

What we still need is to prove that for all a_1 the block part in (3.2) can be effectively found. Note that we know that the number k is finite, since, otherwise, there would be an infinite solution with block decomposition beginning with a_1, which would be already be detected by the first case.

We construct the word $u_1 \ldots u_k$ and $v_1 \ldots v_k$ using the same idea as in procedure Continuation. Clearly, there exists a solution (3.2) if and only if either

$$(u_1 \cdots u_k)^{-1}(v_1 \cdots v_k) = cz \text{ or } (v_1 \cdots v_k)^{-1}(u_1 \cdots u_k) = bz$$

for some word z, according to whether or not $|u_1 \cdots u_k| < |v_1 \cdots v_k|$ or not. Our algorithm works as the Continuation for (a_1, h, g), but the sequence (x_i, y_i) is constructed so that for each step we check that $x_i \bowtie y_i$. If at some step i, x_i and y_i are not comparable, there is no infinite solution for a_1. Similarly, if Continuation returns "no block", that is no next letter for x_i or y_i exists and $h(x_i) \neq g(y_i)$. Now if Continuation stops in the case $h(x_i) = g(y_i)$ (and $x_i \bowtie y_i$), we have that $x_i = y_i dz$ or $y_i = x_i dz$, for some letter d and word z. In the first case, if there is a block word for g beginning with d, we set $x_{i+1} = x_i$ and $y_{i+1} = y_i d$, and continue according to the procedure Continuation. Otherwise, there is no block word for g beginning with d, and $x_i = u_1 \cdots u_k$ and $y_i = v_1 \cdots v_k$. In the other case, where $y_i = x_i dz$, we reason similarly. Note that $dz \neq \varepsilon$, since $E(I) = \emptyset$. Since there is no infinite solution with block decomposition, this algorithm necessarily stops.

Finally, we need to check whether or not for the letter d there exists the infinite words $\omega_1 = uv^\omega$ and $\omega_2 = u'w^\omega$ as above, and that

$$u_1u_2 \cdots u_k uv^\omega = v_1v_2 \cdots v_k u'w^\omega.$$

These words are ultimately periodic and, therefore, their equality can be determined in a trivial way. □

The structure of the solutions of the marked infinite PCP was studied in [8], and the infinite solutions of the unique continuation instances have the same structure. Indeed, it was proved that the set of infinite solutions for the marked instances i is of the form

$$E_{\min}(I)^\omega \cup E_{\min}(I)^* \left(P \cup F\right), \tag{3.3}$$

where P is a finite set of ultimately periodic words, and F is a finite set of morphic images of fixed points of D0L systems. In the proof of Theorem 3.4 it was proved that the solutions with block decomposition are morphic images of solutions of a marked instance, and the solution without block decomposition are ultimately periodic. Since the morphic images of ultimately periodic words in P are ultimately periodic and the morphic images of morphic images of F are of the type F, we get that

Theorem 3.5 *The infinite solutions of the instance with unique continuation property have the structure of* (3.3)*.*

Finally, we give another uniqueness property for the instances of the PCP.

> **UC3**. The instance (h, g), where $h, g\colon A^* \to B^*$, of the PCP, is called *unique equality continuation instance*, if, for $u \in A^*$ and $a, b \in A$, $h(ua) \bowtie g(ua)$ and $h(ub) \bowtie g(ub)$, then either $h(u) = g(u)$ or $a = b$.

We leave the following two questions as open problems: Is the PCP is decidable for unique equality continuation instances or not? Is it decidable whether or not an instance of the PCP satisfies the property (UC3)?

Note that for a unique equality continuation instance I, $E_{\min}(I)$ is finite and marked as well, and the set of infinite solutions of I has the structure as in (3.3).

References

[1] A. Ehrenfeucht, J. Karhumäki, and G. Rozenberg. The (generalized) Post Correspondence Problem with lists consisting of two words is decidable. *Theoret. Comput. Sci.*, 21:119–144, 1982.

[2] V. Halava. *The Post Correspondence Problem for Marked Morphisms.* PhD thesis, Department of Math., Univ. of Turku. TUCS Dissertations no. 37, 2002.

[3] V. Halava and T. Harju. Infinite solutions of the marked Post Correspondence Problem. In J. Karhumäki, W. Brauer, H. Ehrig and A. Salomaa, editors, *Formal and Natural Computing*, volume 2300 of *Lecture Notes in Comput. Sci.*, pages 57–68. Springer-Verlag, 2002.

[4] V. Halava and T. Harju. Infinite Post Correspondence is Undecidable for Instance of Size 9. submitted, also as TUCS Tech. Report 662, 2005.

[5] V. Halava, T. Harju, and M. Hirvensalo. Generalized Post correspondence problem for marked morphisms. *Internat. J. Algebra Comput.*, 10(6):757–772, 2000.

[6] V. Halava, T. Harju, and M. Hirvensalo. Binary (generalized) Post Correspondence Problem. *Theoret. Comput. Sci.*, 276:183–204, 2002.

[7] V. Halava, T. Harju, and J. Karhumäki. Decidability of the binary infinite Post Correspondence Problem. *Discrete Appl. Math.*, 130:521–526, 2003.

[8] V. Halava, T. Harju, and J. Karhumäki. The Structure of Infinite Solutions of Marked and Binary Post Correspondence Problems. *Theory Comput. Syst.*, to appear.

[9] V. Halava, M. Hirvensalo, and R. de Wolf. Marked PCP is decidable. *Theoret. Comput. Sci.*, 255(1-2):193–204, 2001.

[10] M. V. Lawson, *Finite Automata*, Chapman & Hall, 2004.

[11] Y. Matiyasevich and G. Sénizergues. Decision problems for semi-Thue systems with a few rules. *Theor. Comput. Sci.* 330(1):145-169, 2005.

[12] E. Post. A variant of a recursively unsolvable problem. *Bull. of Amer. Math. Soc.*, 52:264–268, 1946.

[13] K. Ruohonen. Reversible machines and Post's correspondence problem for biprefix morphisms. *Elektron. Informationsverarb. Kybernet. (EIK)*, 21(12):579–595, 1985.

On systems of word equations with simple loop sets

Štěpán Holub[*], *Juha Kortelainen*[†]

Abstract

Consider the infinite system S of word equations

$$\{\mathbf{x}_0\mathbf{u}_1^i\mathbf{x}_1\mathbf{u}_2^i\mathbf{x}_2\cdots\mathbf{u}_m^i\mathbf{x}_m = \mathbf{y}_0\mathbf{v}_1^i\mathbf{y}_1\mathbf{v}_2^i\mathbf{y}_2\cdots\mathbf{v}_n^i\mathbf{y}_n \mid i\in\mathbb{N}\}$$

For each $k\in\mathbb{N}_+$, let T_k be the subsystem of S given by $i\in\{k,k+1,k+2\}$. We prove two properties of the above system.

1) Let $k\geq 1$. If φ is a solution of T_k such that primitive roots of $\varphi(\mathbf{u}_1),\varphi(\mathbf{u}_2),\ldots,\varphi(\mathbf{u}_m)$ are conjugated, as well as primitive roots of $\varphi(\mathbf{v}_1),\varphi(\mathbf{v}_2),\ldots,\varphi(\mathbf{v}_n)$, then φ is a solution of the whole S.
2) If $n=1$ then, for any $k\geq 2$, a solution φ of T_k is also a solution of S.

1 Introduction

Classical examples of language families whose elements possess some kind of a pumping property are regular, context-free, bounded, and commutative languages. When considering, for instance, the decidability of morphism (or some other mapping) equivalence or effective existence of a test set for those languages, we are led to systems of word equations where pumping in one or several points in an equation can appear.

Throughout the paper we will study the infinite system S of word equations:

$$\{\mathbf{x}_0\mathbf{u}_1^i\mathbf{x}_1\mathbf{u}_2^i\mathbf{x}_2\cdots\mathbf{u}_m^i\mathbf{x}_m = \mathbf{y}_0\mathbf{v}_1^i\mathbf{y}_1\mathbf{v}_2^i\mathbf{y}_2\cdots\mathbf{v}_n^i\mathbf{y}_n \mid i\in\mathbb{N}\}$$

Its subsystem of cardinality three, given by $i\in\{k,k+1,k+2\}$, with $k\in\mathbb{N}$, will be denoted T_k.

By the validity of Ehrenfeucht Conjecture [2], [4], [11], the system S has a finite subsystem that is equivalent to S. Let us briefly survey what is known about our system up to now.

In [1] it is shown that the single equation

$$\mathbf{u}_1^n = \mathbf{v}_1^n\mathbf{v}_2^n\cdots\mathbf{v}_n^n$$

[*] Department of Algebra, Charles University, Prague, Czech Republic, holub@karlin.mff.cuni.cz

[†] Department of Information Processing Science, University of Oulu, Oulu, Finland, jkortela@cc.oulu.fi

is equivalent to

$$\{\mathbf{u}_1^i = \mathbf{v}_1^i \mathbf{v}_2^i \cdots \mathbf{v}_n^i \mid i \in \mathbb{N}\}.$$

This fact was generalized in [6]. The results of [7] imply that if all the midwords x_i and y_i are empty, the system S is equivalent to its subsystem T_k, whenever $k \geq 2$. The paper [5] considers S when $\max\{m, n\} = 3$. It is proved that in such a case it is equivalent to the subsystem induced by $i = 0, 1, 2, 3, 4, 5$. Finally, in [9] it is shown that S is equivalent to its subsystem induced by $i = 0, 1, 2, \ldots m + n + 2$.

It is not known (see *Open Problem 1*), whether S has an equivalent subsystem of a constant size, i.e., a size independent of m and n. In this paper we give small equivalent subsystems in two special cases. It is organized as follows.

In the second section some preliminaries, definitions and well-known results in the theory of combinatorics on words are given.

In the third section, the first of our cases is studied. We impose an additional condition on the structure of loops, and prove that if for some $k \geq 1$ there is a solution φ of T_k such that the primitive roots of $\varphi(u_1), \varphi(u_2), \ldots, \varphi(u_m)$ are conjugated, and also the primitive roots of $\varphi(v_1), \varphi(v_2), \ldots, \varphi(v_n)$ are conjugated, then φ is a solution of whole S.

Section four explains in some generality a method, which in section five is used to prove that if $n = 1$ then S is equivalent to T_k for any $k \geq 2$. That is our second special case, in which the system contains just one loop on one side. Note that the number of loops on the other side is arbitrary.

In the sixth section some open problems and topics of further investigation are presented.

2 Preliminaries

We suppose that the reader is familiar with basic concepts of combinatorics on words as it can be found in [10], where also a proof is given for the following two results belonging to the folklore of combinatorics on words.

Lemma 2.1 *Let x and y be nonempty words. The following three conditions are equivalent.*

1. *The words x and y are conjugate ;*
2. *The words x and y are of equal length and there exist unique words t_1, and t_2, with t_2 nonempty, such that $t = t_1 t_2$ is primitive and $x \in (t_1 t_2)^+$ and $y \in (t_2 t_1)^+$;*
3. *There exists a word z such that $xz = zy$.*

Furthermore, assume that any of the three conditions above holds and that t_1 and t_2 are as in condition (2). Then, for each word w, we have $xw = wy$ if and only if $w \in (t_1 t_2)^* t_1$.

In the setting of the previous lemma we say that x and y are conjugated by z.

Lemma 2.2 *Two nonempty words commute if and only if they are powers of the same (primitive) word, i.e., they have the same primitive root.*

Recall that the primitive root of a word u is the shortest word r such that $u = r^i$ for some integer $i \geq 1$.

One of the strongest results in the elementary theory of combinatorics on words is the Periodicity Lemma. A slight modification of it can be stated as follows (for the proofs, see for instance [3], [8] and [10]).

Lemma 2.3 *If two powers u^m and v^n of nonempty words u and v have a common subword of length at least $|u| + |v| - d$ (d being the greatest common divisor of $|u|$ and $|v|$), then the primitive roots of u and v are conjugated.*

Note that if in the previous lemma u^m and v^n have a common prefix of length at least $|u| + |v| - d$, then u and v have the same primitive root, so they are powers of the same (primitive) word.

For each word w, the infinite word $w\,w \cdots$ is denoted by w^ω. In our considerations we will also need the following lemma.

Lemma 2.4 *Let u and v be words such that $|u| \leq |v|$ and each factor of v of length $|u|$ is conjugated with u. Then v is a factor of u^ω.*

Proof Let arb be a factor of v of length $|u|+1$, where a and b are letters. Since both ar and rb are conjugated with u, we deduce $a = b$ from $|ar|_a = |rb|_a = |u|_a$. The claim follows. □

3 Conjugated primitive roots

In this section we prove the result announced in the introduction. To simplify notation, we like to formulate it in the following way.

Theorem 3.1 *Let m, n be positive integers, and $x_0, \ldots, x_m$, $y_0, \ldots, y_n$, $u_1, \ldots,$ u_m, $v_1, \ldots, v_n$ words such that for each $i, j \in \{1, 2, \ldots, m\}$ the primitive roots of u_i and u_j are conjugated, and similarly for each $i, j \in \{1, 2, \ldots, n\}$ the primitive roots of v_i and v_j are conjugated. Let $k \geq 1$ be a positive integer. If*

$$x_0 u_1^i x_1 u_2^i x_2 \cdots u_m^i x_m = y_0 v_1^i y_1 v_2^i y_2 \cdots v_n^i y_n \quad (i = k, k+1, k+2) \tag{3.1}$$

then also

$$x_0 u_1^i x_1 u_2^i x_2 \cdots u_m^i x_m = y_0 v_1^i y_1 v_2^i y_2 \cdots v_n^i y_n \quad (i = 0, 1, 2, 3, \ldots)\ . \tag{3.2}$$

Proof We first introduce several additional assumptions which do not harm generality.

Clearly, we may suppose that the words u_i, $i \in \{1, \ldots, m\}$, and v_i, $i \in \{1, 2, \ldots, n\}$, are non-empty. Assume also that $y_0 = \epsilon$ and that either $x_m = \epsilon$ or $y_n = \epsilon$.

Let $i \in \{1, \ldots, m-1\}$ be such that x_i is empty. We then may suppose that u_i and u_{i+1} do not commute, since otherwise we merge them by writing $(u_i u_{i+1})^j$ instead of $u_i^j u_{i+1}^j$.

We say that two nonempty words u and v are *marked* if they do not begin with the same symbol.

Let $i \in \{1, 2, \ldots, m\}$ be such that x_i is nonempty. The reasoning below verifies that we may consider, without loss of generality, only cases in which u_i and x_i, are marked.

Suppose that z is the longest nonempty prefix of x_i, which is also a prefix of $u_i x_i$. Let $x'_{i-1} = x_{i-1} z$, $u'_i = z^{-1} u_i z$, and $x'_i = z^{-1} x_i$. It is not difficult to see that x'_{i-1}, x'_i and u'_i are well defined, and for any j the word

$$x_0 u_1^j x_1 u_2^j x_2 \cdots u_m^j x_m$$

does not change if we substitute x_{i-1}, x_i, and u_i by x'_{i-1}, x'_i, and u'_i. Repeating the procedure finitely many times we shall obtain the markedness.

Analogously we assume that for each $i \in \{1, 2, \ldots, n-1\}$ such that $y_i = \epsilon$, the words v_i and v_{i+1} do not commute and that for each $j \in \{1, 2, \ldots, n\}$ such that $y_j \neq \epsilon$, the words v_j and y_j are marked.

The proof of the theorem will now proceed by induction with respect to the number $m + n$.

Suppose that $m+n \leq 2$. An obvious length argument yields that $m = n = 1$, $x_1 = \epsilon$, $|x_0| = |y_1|$, and $|u_1| = |v_1|$. From equalities

$$x_0 u_1^i = v_1^i y_1 \quad (i = k, k+1) \tag{3.3}$$

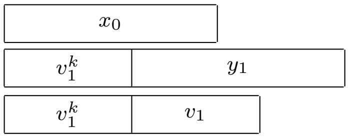
one obtains that v_1 and u_1 are conjugated by $x_0 u_1^k = v_1^k y_1$. Lemma 2.1 now easily implies that (3.2) holds.

Suppose that $m + n > 2$. We distinguish two main cases:

1° $|x_0| > |v_1^k|$;

2° $|x_0| \leq |v_1^k|$.

Consider the first case. If $|y_1| > 0$ then the words x_0, v_1 and y_1 begin with the same symbol, and v_1, y_1 are not marked, which is against our assumptions.

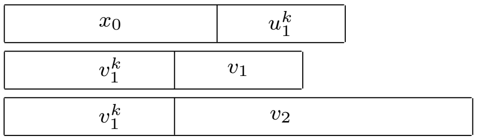


Let therefore $y_1 = \epsilon$.

If

$$|x_0 u_1^k| \geq \min\{|v_1^{k+1}|, |v_1^k v_2|\}$$

then the words v_1 and v_2 are comparable, i.e., one of them is a prefix of the other. Since the primitive roots of v_1 and v_2 are conjugated, they coincide, and v_1 and v_2 commute, again a contradiction with the global assumption.



Suppose, on the other hand, that

$$|x_0 u_1^k| < \min\{|v_1^{k+1}|, |v_1^k v_2|\}.$$

Then the word $d = v_1^{-k} x_0$ is a prefix of both v_1 and v_2. Surely

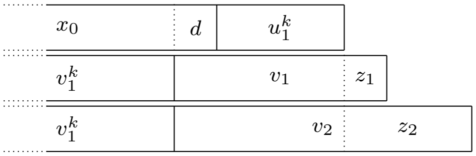
$$|u_1^k| < \min\{|v_1|, |v_2|\}.$$

From (3.1) we have

$$d u_1^i x_1 u_2^i x_2 \cdots u_m^i x_m = v_1^{i-k} v_2^i y_2 \cdots v_n^i y_n \quad (i = k, k+1, k+2)\ . \tag{3.4}$$

Let z_1 and z_2 be words such that

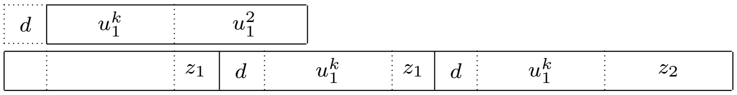
$$v_1 \ = \ d u_1^k z_1 \qquad \text{and} \qquad v_2 \ = \ d u_1^k z_2.$$



By (3.4),

$$u_1^i x_1 u_2^i \cdots u_m^i x_m = (u_1^k z_1 d)^{i-k} (u_1^k z_2 d)^{i-1} u_1^k z_2 y_2 \cdots v_n^i y_n \tag{3.5}$$

for $i = k, k+1, k+2$.



Consider the common prefix of u_1^{k+2} and $(u_1^k z_1 d)^2 u_1^k$.

If

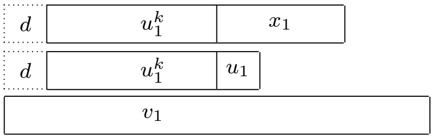
$$|u_1^{k+2}| > |u_1| + |u_1^k z_1 d| - 1$$

then, by the Periodicity Lemma, the words u_1 and $u_1^k z_1 d$ have the same primitive root t. Since v_1 and v_2 have conjugated primitive roots and their common prefix is longer than t, they commute.

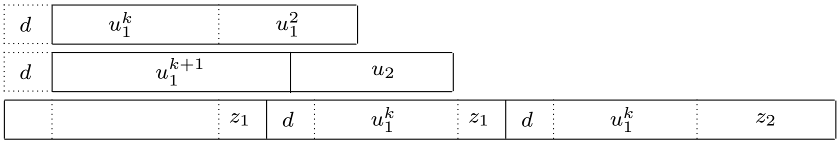
Assume that

$$|v_1| = |u_1^k z_1 d| > |u_1^{k+1}| + 1.$$

If $x_1 \neq \epsilon$, then the words u_1 and x_1 are not marked.



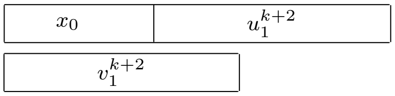
Suppose therefore that $x_1 = \epsilon$. Then (3.5) implies that u_1^{k+2} is a prefix of $u_1^k z_1 d u_1$, and $u_1^k z_1 d u_1$ is comparable with $u_1^{k+1} u_2$. Therefore also u_1^{k+2} and $u_1^{k+1} u_2$ are comparable, and since primitive roots of u_1 and u_2 are conjugated, they commute.



The second main case was $|x_0| \leq |v_1^k|$. If

$$|u_1^{k+2}| \geq |u_1| + |v_1| - 1 \qquad \text{and} \qquad |v_1^{k+2}| - |x_0| \geq |u_1| + |v_1| - 1 \tag{3.6}$$

then, by the Periodicity Lemma, primitive roots of u_1 and v_1 are conjugated. Clearly they are conjugated by x_0.



Now the number $n + m$ can be decreased by eliminating u_1 (if $|u_1| > |v_1|$) or v_1 (if $|v_1| > |u_1|$) or both (if $|u_1| = |v_1|$), and we are through by induction.

Let us be more more rigorous. Using Lemma 2.1, let $t = t_1 t_2$ be a primitive word, and q, r and s positive integers such that $u_1 = (t_1 t_2)^q$, $v_1 = (t_2 t_1)^r$ and $x_0 = t_2 (t_1 t_2)^s$. Suppose that $q + s \geq r$ (the opposite case being similar). Now (3.1) allows us to deduce that

$$t_2 (t^{q+s-r})^i x_2 u_2^i \cdots u_m^i x_m = y_1 v_2^i y_2 \cdots v_n^i y_n \quad (i = k, k+1, k+2)\ . \tag{3.7}$$

By induction, we deduce that

$$t_2 (t^{q+s-r})^i x_2 u_2^i \cdots u_m^i x_m = y_1 v_2^i y_2 \cdots v_n^i y_n \quad (i = 0, 1, 2, \dots) \tag{3.8}$$

is true. Obviously, also

$$t_2 (t^{q+s})^i x_2 u_2^i \cdots u_m^i x_m = ((t_2 t_1)^r)^i y_1 v_2^i y_2 \cdots v_n^i y_n \quad (i = 0, 1, 2, \dots) \tag{3.9}$$

holds, and we are done.

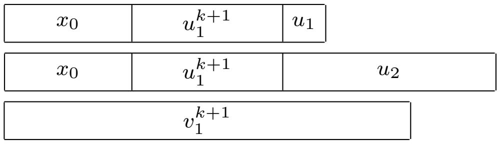
Assume that (3.6) does not hold because of

$$|v_1| > |u_1^{k+1}| + 1. \tag{3.10}$$

If $x_1 \neq \epsilon$, the words u_1 and x_1 are not marked. Suppose that $x_1 = \epsilon$. If

$$|v_1^{k+1}| \geq |x_0| + |u_1^{k+2}|$$

then u_1 and u_2 commute.

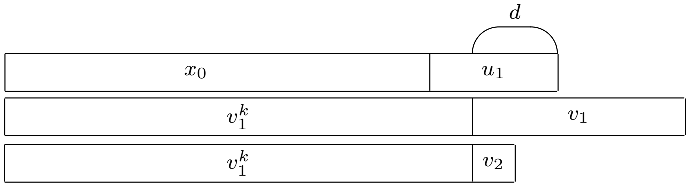


Suppose

$$|v_1^{k+1}| < |x_0| + |u_1^{k+2}|.$$

This implies, together with (3.10), that $|v_1^k| < |x_0u_1|$. Let $d = v_1^{-k}x_0u_1$. Note that d is a prefix of v_1.

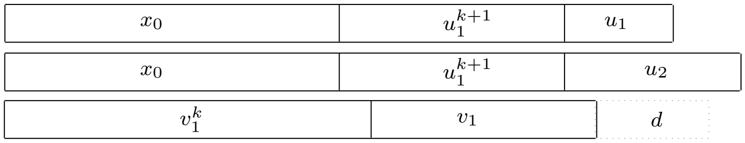
If $y_1 \neq \epsilon$ then v_1 and y_1 are not marked. Therefore $y_1 = \epsilon$ and d is comparable with v_2. If $|d| \geq |v_2|$ then v_1 and v_2 commute.



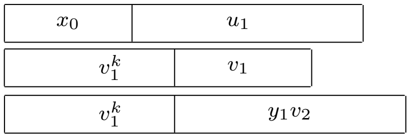
Suppose the contrary, which implies that d is a prefix of v_2 (as well as of v_1). Then both $x_0u_1^{k+2}$ and $x_0u_1^{k+1}u_2$ are comparable with $v_1^{k+1}d$. Since

$$|v_1^{k+1}d| = |x_0u_1| + |v_1| > |x_0u_1^{k+2}|,$$

the words u_1 and u_2 are comparable, and therefore commute.



Suppose then that the second inequality of (3.6) is not true, that is $|v_1^{k+1}| < |x_0u_1| - 1$. Then either v_1 and y_1 are not marked, or (if $y_1 = \epsilon$) the words v_1 and v_2 commute.



The proof is now complete. □

In the rest of the paper we shall prove our second result:

Theorem 3.2 *Let $k \geq 2$ be a positive integer. The system of equations S:*

$$\{\mathbf{x}_0\mathbf{u}_1^i\mathbf{x}_1\mathbf{u}_2^i\mathbf{x}_2\cdots\mathbf{u}_m^i\mathbf{x}_m = \mathbf{y}_0\mathbf{v}^i\mathbf{y}_1 \mid i \in \mathbb{N}\} \tag{3.11}$$

is equivalent to its subsystem T_k given by $i \in \{k, k+1, k+2\}$.

4 Characteristic equation

In this section we explain the method to be used in the proof of the Theorem 3.2. It was first introduced in [7].

Let $X = \{x_1, \ldots, x_k\}$ be a set of unknowns, and $e = (w_1, w_2) \in X^* \times X^*$ an equation, such that $\mathrm{alph}(e) = X$.

Consider a non-erasing morphism $\varphi : X^* \to \Sigma^*$ solving e, i.e., $\varphi(w_1) = \varphi(w_2)$, and denote $d_i = |\varphi(x_i)|$.

Having got such a solution we choose a new alphabet of unknowns H, construct a new equation $\overline{e} = (\overline{w}_1, \overline{w}_2) \in H^* \times H^*$, and define a length-preserving morphism $\overline{\varphi} : H^* \to \Sigma^*$.

The set H consists of letters $x_{i,j}$, for $i = 1, \ldots, k$ and $j = 1, \ldots, d_i$. Informally, alphabet H is a set of names of all positions in images of φ. This naturally induces the morphism $\psi : X^* \to H^*$ defined by

$$\psi(x_i) = x_{i,1} \cdots x_{i,d_i}.$$

With help of that morphism, the equation $\overline{e}$ is given by

$$\overline{w}_i = \psi(w_i),$$

$i = 1, 2$. The equation $\overline{e} = (\overline{w}_1, \overline{w}_2)$ is called the *characteristic equation* of e with respect to the morphism φ. Clearly the characteristic equation only depends on the values $d_i, \ldots, d_k$.

Finally, the morphism $\overline{\varphi}$ is defined by

$$\overline{\varphi} \circ \psi = \varphi.$$

It should be clear that $\overline{\varphi}$ is well defined and length-preserving. Indeed, it maps H into Σ, since $\overline{\varphi}(x_{i,j})$ is the jth letter of $\varphi(x_i)$ for each j and i.

The definition also immediately implies that $\overline{\varphi}$ is a solution of $\overline{e}$:

$$\overline{\varphi}(\overline{w}_1) = \overline{\varphi} \circ \psi(w_1) = \varphi(w_1) = \varphi(w_2) = \overline{\varphi} \circ \psi(w_2) = \overline{\varphi}(\overline{w}_2).$$

Example 4.1 Consider equation $\mathbf{yzxy} = \mathbf{xyyz}$ and its solution φ:

$$\varphi(\mathbf{x}) = ab, \qquad \varphi(\mathbf{y}) = a, \qquad \varphi(\mathbf{z}) = ba.$$

Then we may denote letters in the new alphabet by

$$H = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{z}_1, \mathbf{z}_2\},$$

the characteristic equation is

$$\mathbf{y}_1\mathbf{z}_1\mathbf{z}_2\mathbf{x}_1\mathbf{x}_2\mathbf{y}_1 = \mathbf{x}_1\mathbf{x}_2\mathbf{y}_1\mathbf{y}_1\mathbf{z}_1\mathbf{z}_2,$$

morphism ψ is defined by

$$\psi(\mathbf{x}) = \mathbf{x}_1\mathbf{x}_2, \qquad \psi(\mathbf{y}) = \mathbf{y}_1, \qquad \psi(\mathbf{z}) = \mathbf{z}_1\mathbf{z}_2,$$

and $\overline{\varphi}$ by

$$\begin{aligned} \overline{\varphi}(\mathbf{x}_1) &= a, & \overline{\varphi}(\mathbf{x}_2) &= b, \\ \overline{\varphi}(\mathbf{z}_1) &= b, & \overline{\varphi}(\mathbf{z}_2) &= a, \\ \overline{\varphi}(\mathbf{y}_1) &= a. \end{aligned}$$

The reason for introducing the characteristic equation $\bar{e}$ is that it allows to produce linear equalities, which can yield—as we shall see in the next section—important information about e.

The linear equalities are obtained in the following way. Let p be a factor of the word $w = \varphi(w_1) = \varphi(w_2)$. The number of occurrences of p in w can be expressed in two different ways, using $\overline{w}_1$ and $\overline{w}_2$, respectively.

Let's first introduce some more notation. By $\mathbf{F}(w)$ denote the set of all factors of a word w, and by $|w|_p$ the number of occurrences of the word p in w.

Now, given an arbitrary word $p \in \Sigma^*$, we have

$$\sum_{\overline{\varphi}(\alpha)=p} |\overline{w}_1|_\alpha = \sum_{\overline{\varphi}(\alpha)=p} |\overline{w}_2|_\alpha = |w|_p. \tag{4.1}$$

Proof (Proof of (4.1)) Fix $i \in \{1, 2\}$. Recall that $\varphi = \overline{\varphi} \circ \psi$ and $\overline{w}_i = \psi(w_i)$. Each occurrence of p in $\varphi(w_i)$ is therefore an image of some $\alpha \in \mathbf{F}(\psi(w_i))$ mapped by $\overline{\varphi}$. The number $|w|_p$ is given by the number of such preimages α in $\overline{w}_i$. □

Example 4.2 (Example continued) Consider $p = aa$. The word

$$w = \varphi(\mathbf{yzxy}) = \varphi(\mathbf{xyyz}) = abaaba$$

contains one occurrence of p. There are nine words $\alpha \in H^*$ satisfying $\overline{\varphi}(\alpha) = aa$, namely

$$\begin{aligned} \alpha_1 &= \mathbf{x}_1\mathbf{x}_1, & \alpha_2 &= \mathbf{x}_1\mathbf{y}_1, & \alpha_3 &= \mathbf{x}_1\mathbf{z}_2, \\ \alpha_4 &= \mathbf{y}_1\mathbf{x}_1, & \alpha_5 &= \mathbf{y}_1\mathbf{y}_1, & \alpha_6 &= \mathbf{y}_1\mathbf{z}_2, \\ \alpha_7 &= \mathbf{z}_2\mathbf{x}_1, & \alpha_8 &= \mathbf{z}_2\mathbf{y}_1, & \alpha_9 &= \mathbf{z}_2\mathbf{z}_2. \end{aligned}$$

Therefore (4.1) has the form

$$\sum_{i=1}^{9} |\psi(\mathbf{yzxy})|_{\alpha_i} = \sum_{i=1}^{9} |\psi(\mathbf{xyyz})|_{\alpha_i}. \tag{4.2}$$

The equality holds, since $|\psi(\mathbf{yzxy})|_{\alpha_i}$ is equal to one for $i = 7$, and is zero otherwise, while $|\psi(\mathbf{xyyz})|_{\alpha_i}$ is one just for $i = 5$.

Informally, we can say that the factor aa comes on the left side of the equation from a different source than on the right side. The formalism of the characteristic equation is designed to express and exploit that fact.

5 One loop systems

We are now ready to prove Theorem 3.2. The theorem deals with the systems S and T_k when $n = 1$, hence we define

$$X = \{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \mathbf{x}_0, \ldots, \mathbf{x}_m, \mathbf{y}_0, \mathbf{y}_1\}.$$

Fix $k \geq 2$, and a morphism φ, which solves the system T_k. Define H, morphisms ψ and $\overline{\varphi}$ as in the previous section. Our task is to show that φ solves S as well. It will be done by showing that the primitive roots of all $\varphi(u_1), \varphi(u_2), \ldots, \varphi(u_m)$ are conjugated. Theorem 2.1 then applies.

Denote

$$\begin{aligned} \ell_i &= \mathbf{x}_0\mathbf{u}_1^i\mathbf{x}_1\mathbf{u}_2^i\mathbf{x}_2 \cdots \mathbf{u}_m^i\mathbf{x}_m, \\ r_i &= \mathbf{y}_0\mathbf{v}^i\mathbf{y}_1 \end{aligned}$$

for $i = k, k+1, k+2$. Recall that, for each i,

$$(\overline{\ell}_i, \overline{r}_i) = (\psi(\ell_i), \psi(r_i))$$

is the characteristic equation of (ℓ_i, r_i) with respect to φ.

Define the word p, whose number of occurrences will be counted. Let t be the shortest among the primitive roots of words $\varphi(\mathbf{u}_1), \varphi(\mathbf{u}_2), \ldots, \varphi(\mathbf{u}_m)$. Then p is defined by the following conditions:

(i) The word p is a factor of t^ω.

(ii) There exist a word $\alpha \in H^+$ such that

- α is a factor of $\overline{\ell}_{k+2}$,
- $\psi(\mathbf{u}_j^{k+2})$ is a factor of α for some $j \in \{1, 2, \ldots, m\}$; and
- $\overline{\varphi}(\alpha) = p$;

(iii) If p' satisfies (i) and (ii) then $|p| \geq |p'|$.

Proof (Example) We shall illustrate the definition of p. Let $k = 2$, $m = 2$ and

$$\varphi(\mathbf{x}_0) = b \qquad \varphi(\mathbf{x}_1) = aba^6b \qquad \varphi(\mathbf{x}_2) = b$$
$$\varphi(\mathbf{u}_1) = a \qquad \varphi(\mathbf{u}_1) = ab.$$

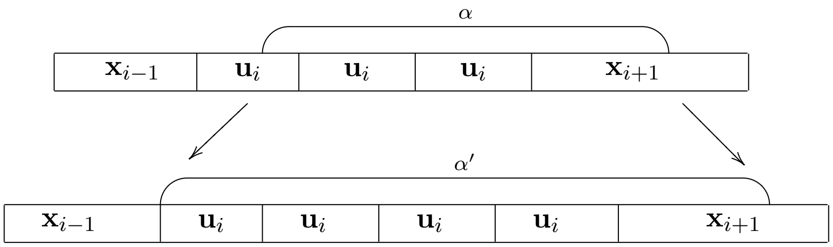
Note that φ is not a solution of the considered system, but it is not important for the definition of p.

In this case $t = a$, and we look for the largest power of a in ℓ_4, which covers the image of some $\varphi(\mathbf{u}_i)$ as required by the condition (ii). Therefore $p = a^5$. Although a^6 is also a factor of ℓ_4, it does not satisfy (ii). □

Lemma 5.1 *Let i be in $\{k, k+1\}$ and $\alpha \in \mathbf{F}(\overline{\ell}_i)$ be a word such that $\overline{\varphi}(\alpha) = p$. Then*

$$|\overline{\ell}_{k+1}|_\alpha - |\overline{\ell}_k|_\alpha = |\overline{\ell}_{k+2}|_\alpha - |\overline{\ell}_{k+1}|_\alpha. \tag{5.1}$$

Proof We first show that no factor of $\psi(u_j^{k+1})$, longer than $|\psi(u_j^2)|$, is a factor of α. Suppose the contrary. Then, by the Periodicity Lemma, the word $\varphi(u_j)$ commutes with a conjugate of t. This implies that we can find a factor α' of $\overline{\ell}_{k+2}$, longer than α, such that $\overline{\varphi}(\alpha')$ is also a factor of t^∞. It is enough, informally speaking, to extend in α each factor of $\psi(u_j^{k+1})$, longer than $|\psi(u_j^2)|$, to $\psi(u_j^{k+2})$. This is a contradiction with the maximality of p.



This implies that α hits some $\psi(x_j)$ for at most one j. Therefore if it contains at least one letter of some $\psi(x_j)$ then it occurs exactly once in all ℓ_k, ℓ_{k+1} and ℓ_{k+2}, i.e.,

$$|\overline{\ell}_k|_\alpha = |\overline{\ell}_{k+1}|_\alpha = |\overline{\ell}_{k+2}|_\alpha = 1.$$

Note that the previous argument would not work for $k = 1$.

The only remaining possibility is that α is a factor of $\psi(u_j^{k+1})$, shorter than $\psi(u_j^2)$.

Then it is easy to see that

$$|\overline{\ell}_{k+1}|_\alpha - |\overline{\ell}_k|_\alpha = |\overline{\ell}_{k+2}|_\alpha - |\overline{\ell}_{k+1}|_\alpha = 1.$$

The proof is now complete. □

Equality (4.1) yields

$$|\varphi(\ell_i)|_p = \sum_{\overline{\varphi}(\alpha)=p} |\overline{\ell}_i|_\alpha$$

for $i = k, k+1, k+2$. In this point we shall exploit the requirement (ii) in the definition of p. The condition guarantees that there is at least one word $\alpha \in H^+$ satisfying $\overline{\varphi}(\alpha) = p$, which is a factor of $\overline{\ell}_{k+2}$ and is neither a factor of $\overline{\ell}_k$ nor of $\overline{\ell}_{k+1}$. That implies, together with (5.1), that

$$|\varphi(\ell_{k+2})|_p - |\varphi(\varphi(\ell_{k+1})|_p > |\varphi(\ell_{k+1})|_p - |\varphi(\ell_k)|_p. \tag{5.2}$$

Confronting the last inequality with the structure of the right side of our equations we get the following claim.

Lemma 5.2 *The primitive root of $\varphi(\mathbf{v}_1)$ is conjugated with t.*

Proof Let α be a word from $\mathbf{F}(\overline{r}_k) \cup \mathbf{F}(\overline{r}_{k+1})$ satisfying $\overline{\varphi}(\alpha) = p$. In a similar manner as in Lemma 5.1 one can show that for our α the equality

$$|\overline{r}_{k+1}|_\alpha - |\overline{r}_k|_\alpha = |\overline{r}_{k+2}|_\alpha - |\overline{r}_{k+1}|_\alpha. \tag{5.3}$$

holds. Then from (5.2) we deduce that there must exist at least one factor α of $\overline{r}_{k+2}$, which is neither a factor of $\overline{r}_k$ nor of $\overline{r}_{k+1}$, such that $\overline{\varphi}(\alpha) = p$. Such an α necessarily contains the factor $\psi(v^{k+1})$. The Periodicity Lemma concludes the proof. □

Now, it can be intuitively clear that there cannot exist a loop, the primitive root of which is not conjugated with t. A proof of this fact is given in the following lemma.

Lemma 5.3 *For any $i \in \{1, 2, \ldots, m\}$ the primitive root of $\varphi(\mathbf{u}_i)$ is conjugated with t.*

Proof Let $j \in \{1, 2, \ldots m\}$ and γ be a factor of $\psi(\mathbf{u}_j^2)$ of length $|t|$ such that $s = \overline{\varphi}(\gamma)$ is not conjugated with t. From the structure of $\overline{\ell}_k$ and $\overline{\ell}_{k+1}$ it is straightforward to see that $|\varphi(\ell_k)|_s < |\varphi(\ell_{k+1})|_s$.

Let us now turn our attention to $\overline{r}_k$ and $\overline{r}_{k+1}$. Their structure clearly implies that the equality $|\varphi(\ell_k)|_s = |\varphi(\ell_{k+1})|_s$ holds; the preimages of s have to hit either y_0 or y_1, otherwise s is conjugated with t. We have achieved a contradiction, therefore for any $j \in \{1, 2, \ldots, m\}$ all factors of $\varphi(\mathbf{u}_j^2)$ of length $|t|$ are conjugated with t. Lemma 2.4 and the Periodicity Lemma conclude the proof. □

By the results above, the primitive roots of words $u_1, u_2, \ldots, u_m$ are conjugated and, by Theorem 2.1, we are done.

6 Some open problems

The following problem still remains open:

Open Problem 1. Does there exist $q \in \mathbb{N}$ such that (for any m and n in $\mathbb{N}$), the system S is equivalent to the subsystem induced by $i = 0, 1, 2, \ldots, q$?

We also wish to mention

Open Problem 2. Is the system $\{u_1^i = v_1^i v_2^i \cdots v_n^i \mid i \in \mathbb{N}\}$ equivalent to the subsystem induced by $i = 1, 2, 3$?

References

[1] K. I. Appel and F. M. Djorup, On the equation $z_1^n z_2^n \cdots z_k^n = y^n$ in a free semigroup. *Trans. Amer. Math. Soc.* **134** (1968), 461–470.

[2] M. H. Albert and J. Lawrence, A proof of Ehrenfeucht‘s Conjecture. *Theoret. Comput. Sci.* **41** (1985), 121–123.

[3] N. J. Fine and H. S. Wilf, Uniqueness theorem for periodic functions. *Proc. Amer. Math. Soc.* **16** (1965), 109–114.

[4] V. S. Guba, Equivalence of infinite systems of equations in free groups and semigroups to finite subsystems, *Mat. Zametki* **40** (1986), 321-324.

[5] I. Hakala and J. Kortelainen, On the system of word equations $x_0 u_1^i x_1 u_2^i x_2 u_3^i x_3 = y_0 v_1^i y_1^i v_2^i y_2 v_3^i y_3$ $(i = 0, 1, 2, \ldots)$ in a free monoid. *Theoret. Comput. Sci.* **225** (1999), 149–161.

[6] T. Harju, D. Nowotka, On the Equation $x_k = z_1^{k_1} z_2^{k_2} \ldots z_n^{k_n}$ in a free femigroup. *TUCS Tecnical report* **602**, 2004.

[7] Š. Holub, Local and global cyclicity in free semigroups. *Theoret. Comput. Sci.* **262** (2001), 25–36.

[8] V. Keränen, On the k-freeness of morphisms on free monoids. *Ann. Acad. Sci. Fenn. Math. Diss.* **61**, 1986.

[9] J. Kortelainen, On the system of word equations
$x_0 u_1^i x_1 u_2^i x_2 \cdots u_m^i x_m = y_0 v_1^i y_1 v_2^i y_2 \cdots v_n^i y_n \quad (i = 0, 1, 2, ...)$
in a free monoid. *J. Autom. Lang. Comb.* **3** (1998), 43–357.

[10] M. Lothaire, *Combinatorics on Words*, Addison-Wesley, Reading Massachusetts, 1983.

[11] A. Salomaa, The Ehrenfeucht conjecture: A proof for language theorist. *Bull. EATCS* **27** (1985),

A note on the number of distinct squares in a word[*]

Lucian Ilie[†]

Abstract

Fraenkel and Simpson [2] proved that the number of distinct squares in a word of length n is at most $2n$. Based on the numerical evidence, it has been conjectured that this number is actually less than n; see also [6]. We improve here this bound to $2n - \Theta(\log n)$.

1 Introduction and basic definitions

Fraenkel and Simpson [2] investigated the number of distinct squares in a word and showed it to be at most twice the length of the word. Based on the numerical evidence, it has been conjectured that this number is actually less than the length; see also [6]. We improve here this bound slightly.

Let us fix first some notation. For an alphabet A, A^* denotes the set of finite words over A; ε is the empty word. The length of $w \in A^*$ is denoted $|w|$. For $x, y, w \in A^*$, if $w = xy$, then x is a prefix of w, denoted $x \leq w$; if also $x \neq w$, then x is called proper prefix, denoted $x < w$. If $w = xyz$, then y is a factor of w; if $y = xx$, for some word $x \neq \varepsilon$, then y is called a square. For notions and results from combinatorics on words, we refer to [4, 5].

We next recall briefly the approach in [2]. Fraenkel and Simpson counted each square at the beginning of its last occurrence in the word. For a word $w \in A^*$ of length n, consider the sequence $\mathbf{s}(w) = s_1 s_2 \ldots s_n$, where s_i is the number of squares whose last occurrence in w starts at i. Recall that we count distinct squares. Without this restriction the problem is trivial. They proved that no three words can have the last occurrence starting at the same position, that is, $s_i \leq 2$, for all i; put otherwise, $\mathbf{s}(w) \in \{0, 1, 2\}^*$. This obviously implies the result.

The proof in [2] uses a rather intricate combinatorial result of Crochemore and Rytter [1] concerning the lengths of three squares which are prefixes of each other. The same proof is included also in Lothaire's second book [5, p.281-2]. A very short proof of is given in [3].

The main idea here is to look closer at consecutive 2s in $\mathbf{s}(w)$. We prove some upper bounds on the lengths of such runs and then use those to improve slightly the bound of [2].

[*] Research supported in part by NSERC.

[†] Department of Computer Science, University of Western Ontario, London, ON, N6A 5B7, Canada, ilie@csd.uwo.ca

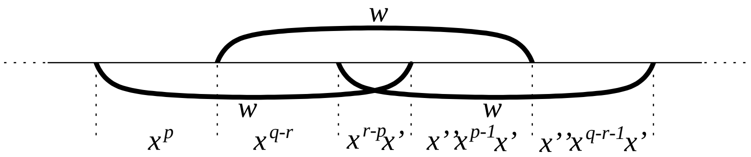


Figure 1: Three mutually overlapping occurrences of w

2 Runs of 2s in s(w)

We start with a useful technical lemma which concerns three occurrences of the same word which mutually overlap.

In what follows we shall need the following synchronization property for primitive words: a word w is primitive if and only if w has exactly two occurrences as factor of ww, namely as a prefix and as a suffix.

Lemma 2.1 *Let w be a word and let three mutually overlapping occurrences of w be $z_1z_2z_3 = z_2z_3z_4 = z_3z_4z_5 = w$, for some $z_i \in A^* - \{\varepsilon\}$. Then there exist $x \in A^*$ primitive and $1 \leq p \leq r < q$ such that $x = x'x''$, for $x', x'' \in A^*$, x'' non-empty, and $z_1 = x^p$, $z_2 = x^{q-r}$, $z_3 = x^{r-p}x'$, $z_4 = x''x^{p-1}x'$, and $z_5 = x''x^{q-r-1}x'$; see Fig. 1.*

Proof For simplicity, let us denote $w_i = z_iz_{i+1}z_{i+2}$, $1 \leq i \leq 2$. Put $z_1 = x^p$, x primitive, $p \geq 1$. The overlap between w_1 and w_2 gives $w = x^qx'$, $q \geq p$, x' a proper prefix of x; put also $x = x'x''$. Then, the overlap between w_2 and w_3 is longer than $|z_1|$ and so longer than $|x|$. Thus z_3 contains a full x which has to synchronize with the ones in w_3. Hence, $z_3z_4 = x^rx'$, for some $r \geq p$. This implies the claimed values of z_i, $2 \leq i \leq 5$. □

We say that u^2 is a "square at position i" if the last occurrence of u^2 starts at i. The next lemma gives a relation between the lengths of squares at neighboring positions.

Lemma 2.2 *If $v^2 < u^2$ are two squares at i and w^2 is a square at $i+1$, then either $|w| \in \{|v|, |u|\}$ or $|w| \geq 2|v|$.*

Proof Asssume first that $|w| < |v|$. Because these are the last occurrences of the three squares in w, we get that $aw < v < u < w^2 < v^2 < u^2$, for some letter $a \in A$; see Fig. 2. We have then three occurences of w as follows: the second w in w^2, the one at the beginning of the second v (in the prefix aw of v) and the one at the beginning of the second u – shown with bold lines in Fig. 2.

We apply Lemma 2.1 to these there occurrences of w. Using the notations there, (the first) w has a prefix x^p and (the last) w has a suffix $x''x^{q-r-1}x'$. We can write the second v as $v = awt$, for a prefix t of w such that $|t| = |x^p| - 1$. Denoting by b the letter following v^2 in u^2, we obtain that tb and the suffix

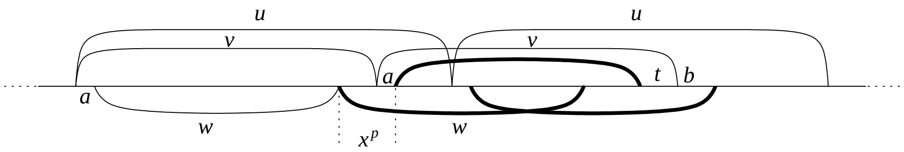


Figure 2: $v^2 < u^2$ squares at i and w^2 square at $i+1$ with $|w| < |v|$

$x''x^{q-r-1}x'$ of w are prefixes of each other. If $p \geq 2$, then t contains as prefix one full x and using synchronization, we have $x' = \varepsilon$, $x'' = X$, and so $w = x^q$ appears also $|x|$ positions later, a contradiction. If $p = 1$, then $|tb| = |x|$ and so $a = b$, $x' = \varepsilon$, and we obtain the same contradiction.

Consider next the case $|v| < |w| < |u|$. The reasoning is similar with the above. Denoting $v = av'$, we obtain three mutually overlapping occurrences of v': the suffix of the second v, the prefix of the second w, and the one occurring in the prefix av' of the second u. Applying Lemma 2.1 for these three v's and using synchronization we can show as above that v^2 appears later, a contradiction. (Two cases are distinguished, depending on which of aw^2 and uv is longer, but they are treated similarly.)

If $|u| < |w| < 2|v|$, we obtain a similar contradiction concerning v^2. This proves the lemma. □

We now consider the impact of Lemma 2.2 on the lengths of runs of consecutive 2s in $\mathbf{s}(w)$. It is clear intuitively that the more such 2s the longer they should become and ultimately impact on the length of w but we shall make this precise. Lemma 2.2 says that the squares at any position have either the same lengths as the ones at the previous position or at least twice larger. The next result investigates the case when the lengths of the squares are preserved.

Lemma 2.3 *Let $m \geq 1$ such that, for any $i, 1 \leq i \leq m$, we have $s_i = 2$, $v_i^2 < u_i^2$ are the squares at i, and $|u_i| = p$, $|v_i| = q$. Then (i) $|u| + m \leq 2|v|$, (ii) $|u| \geq |v| + m + 1$, and (iii) $p \geq 3m + 2$, $q \geq 2m + 1$.*

Proof It is clear that $m \leq |v|$ as otherwise v^2 appears again $|v|$ positions later. Denote u_1 by u, v_1 by v and let z be the common prefix of length $m - 1$ of u and v. (All squares are circular shifts of u and v respectively.)

For (i), we must have that $|u| + |z| < 2|v|$ since otherwise v^2 appears later as prefix of uz.

On the other hand, if $|u| \leq |v| + |z|$, then v is a factor of vz which is neither a prefix nor a suffix. The synchronization property implies that v is not primitive and also it appears later, a contradiction. In fact $|u| = |v| + |z| + 1$ still implies the same thing. (If we denote by b the letter following v^2z in u^2, then $v = zay = yzb$, for some word y, and so $a = b$ and we finally obtain the same contradiction.) Therefore, we must have $|u| \geq |v| + |z| + 2$, which proves (ii).

From the two inequalities we obtain $|v| \geq 2|z| + 3$ and $|u| \geq 3|z| + 5$ which gives (iii). □

Remark 2.4 The inequalities in Lemma 2.3 are very close to optimal. Consider for example $z = (ab)^n a$, $v = zazbaa$, $u = vzaab$; here $|v| = 2|z| + 4$ and $|u| = 3|z| + 7$.

Finally, we formulate the impact of runs of 2s in $\mathbf{s}(w)$ on the length of w in the following result.

Lemma 2.5 *If $\mathbf{s}(w)$ has a prefix of length m such that $s_i = 2$ for all $1 \leq i \leq m$, then $|w| > 2m$.*

Proof Let us write the prefix of length m of $\mathbf{s}(w)$ as

$$s_1 \dots s_{i_1} s_{i_1+1} \dots s_{i_1+i_2} \dots s_{i_1+\cdots+i_k},$$

where $i_1 + \cdots + i_k = m$ and, for any $i_\ell < i < j \leq i_{\ell+1}$, the lengths of the squares at i are the same as those at j. Denote the lengths of the square at i_ℓ by q_ℓ, p_ℓ, with $q_\ell < p_\ell$. By Lemma 2.2, we have $p_{\ell+1} \geq q_\ell$.

Using Lemma 2.3(iii), we obtain first $q_1 \geq 2i_1 + 1$, $p_1 \geq 3i_1 + 2$. Then using (i)-(iii) in the same lemma, we have

$$\begin{aligned}
&p_\ell \geq 2q_{\ell-1},\\
&q_\ell \geq \frac{1}{2}(p_\ell + i_\ell) \geq q_{\ell-1} + \frac{1}{2}i_\ell,\\
&p_\ell \geq q_\ell + i_\ell + 1 \geq q_{\ell-1} + \frac{3}{2}i_\ell + 1.
\end{aligned}$$

Therefore

$$p_k \geq \frac{1}{2}\sum_{j=1}^{k} i_j + \frac{3}{2}i_1 + i_k + 2,$$

which implies

$$|w| \geq \sum_{j=1}^{k-1} i_j + 2p_k + i_k > 2m.$$

□

Remark 2.6 The result in Lemma 2.5 can be improved but the proof becomes more complicated and our main result in the next section does not change much.

I think the optimal bound is essentially the one given by Lemma 2.3(iii). However, as discussed in the next section, we need more than that to prove the conjecture.

3 The improved bound

We can consider now the impact of the above results on the number of distinct squares.

Theorem 3.1 *The number of distinct squares in a word of length n is at most $2n - \Theta(\log n)$.*

Proof By Lemma 2.5, the number of 2s at the beginning of w cannot exceed $n/2$ since otherwise the squares would fall off the end of w. Therefore, we need a 1 in the first half of $\mathbf{s}(w)$. Similarly, we need another 1 in the first half of what is left of $\mathbf{s}(w)$, and so on. The result follows. □

The improvement we obtained is not very big. Essentially we proved that the number of distinct squares in w is bounded away from $2n$. Still, it is the only non-trivial improvement we know of. Moreover, Fraenkel and Simpson [2] considered of importance even improving the bound by a constant amount: they spent some effort to obtain the bound $2n-8$, for $n \geq 5$, and $2n-29$, for $n \geq 22$. Sgnificant improvements of this bound seem difficult.

To prove the conjecture by the above idea, or at least to improve it, some bounds for arbitrary sequences $\mathbf{s}(w)$ would have to be found. Computing such bounds seems difficult. What might help is the fact that the 2s in $\mathbf{s}(w)$ seem to decrease the number of distinct squares by the repetitions they introduce. In the example with $n - o(n)$ squares from [2], the sequence $\mathbf{s}(w)$ consist almost entirely of 1s.

References

[1] M. Crochemore and W. Rytter, Squares, cubes, and time-space efficient string searching, *Algorithmica* **13** (1995) 405–425.

[2] A. S. Fraenkel and J. Simpson, How many squares can a string contain?, *J. Combin. Theory, Ser. A,* **82** (1998) 112–120.

[3] L. Ilie, A simple proof that a word of length n has at most $2n$ distinct squares, *J. Combin. Theory, Ser. A*, to appear.

[4] M. Lothaire, *Combinatorics on Words*, Addison-Wesley, Reading, Mass., 1983.

[5] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge Univ. Press, 2002.

[6] M. Lothaire, *Applied Combinatorics on Words*, Cambridge Univ. Press, to appear.

Isomorphism Between Classes Counted by Fibonacci Numbers

Asep Juarna[*], *Vincent Vajnovszki*[†]

Abstract

An isomorphism between two combinatorial classes is a *closeness* preserving bijection, that is, two objects in a class are close if and only if their images by this bijection are also close. Isomorphism allows us to find out some results concerning a class X if similar results are found for the preimage of X. Results that are usually to be found out are exhaustive and random generation, ranking and unranking algorithms or diameter and Hamiltonicity of the graph induced by the combinatorial class.

Simion and Schmidt in 1985 presented a constructive bijection between two combinatorial classes counted with the Fibonacci numbers: the set F_{n-1} of length $(n-1)$ binary strings with no *two* consecutive 1s, and the set $S_n(\tau_3)$ of length n permutations avoiding the patterns $\tau_3 = \{123, 132, 213\}$. In 2003, by rather analytical methods, Egge and Mansour generalized this result showing that $S_n(\tau_p)$, the set of permutations avoiding the patterns $\tau_p = \{12\ldots p, 132, 213\}$, is counted by $(p-1)$th order Fibonacci numbers.

In this paper we show that Simion-Schmidt's bijection can be extended to $S_n(\tau_p)$ by giving a constructive bijection from the set of Fibonacci strings $F_{n-1}^{(p-1)}$, i.e. the set of length $(n-1)$ binary strings with no $(p-1)$ consecutive 1s, to the set $S_n(\tau_p)$. Moreover, we show that this bijection is a combinatorial isomorphism. Furthermore we illustrate how this allows to obtain a Gray code (or equivalently Hamiltonian path) and exhaustive generating algorithm for the set of permutations avoiding given patterns from known similar results for Fibonacci strings.

This is also the first paper which deals with isomorphism on combinatorial classes or with Gray codes and exhaustive generation of pattern avoiding permutations.

Keywords: Pattern(s) avoiding permutations, generalized Fibonacci strings, Gray codes, combinatorial isomorphism.

1 Introduction and motivation

Let S_ℓ be the set of all permutations of $\{1, 2, \ldots, \ell\}$. Let $\pi \in S_n$ and $\tau \in S_k$ be two permutations, $k \leq n$. We say that π *contains* τ if there exists a subsequence

[*] LE2I - UMR CNRS, Université de Bourgogne, B.P. 47 870, 21078 DIJON - Cedex (France), akang92@yahoo.com

[†] vincent.vajnovszki@ubourgogne.fr, Corresponding author

$1 \leq i_1 < i_2 < \ldots < i_k \leq n$ such that $(\pi(i_1)\ldots\pi(i_k))$ has all of pairwise comparisons the same as τ; in this context τ is usually called a *pattern*. We say that π *avoids* τ, or is τ*-avoiding*, if such a subsequence does not exist. The set of all τ-avoiding permutations in S_n is denoted by $S_n(\tau)$ and $|S_n(\tau)|$ is its cardinality. For an arbitrary finite collection of patterns T, we say that π avoids T if π avoids any $\tau \in T$; the corresponding subset of S_n is denoted by $S_n(T)$ and $|S_n(T)|$ is its cardinality.

The fundamental questions of the problems of pattern avoiding permutations are to determine $|S_n(T)|$, viewed as a function of n for given T, to find an explicit bijection between $S_n(T)$ and $S_n(T')$ if $|S_n(T)| = |S_n(T')|$, and to find relations between $S_n(T)$ and other combinatorial structures. By determining $|S_n(T)|$ we mean finding explicit formula, or ordinary or exponential generating functions. From these researches, a number of enumerative results have been proved, new bijections found, and connections to other fields established.

The problems of pattern avoiding permutations were appeared since Knuth [1] posed, in his text book, the problem of sorting with a single stack. The problem actually was the 312-pattern avoiding permutations. In the other section of his book, he showed that the cardinality of all three length patterns avoiding permutations is the Catalan numbers. The investigations of the problems of pattern avoiding permutations then become wider to some set of patterns of length three, four, five, and so on, some combinations of these patterns, generalized patterns, and permutations avoiding some patterns while in the same time containing exactly a numbers of other patterns.

Pattern avoiding permutations has proved to be useful language in a variety of seemingly unrelated problems, from theory of Kazhdan-Lusztig polynomials, to singularities of Schubert varieties, to Chebyshev polynomials, to rook polynomials for a rectangular board, to various sorting algorithms, sorting stacks and sortable permutations [2].

The first systematic study of patterns avoiding permutations was undertaken in 1985 when Simion and Schmidt [3] solved the problem for every set of patterns in S_3. Two of their propositions are:

1. For every $n \geq 1$, $|S_n(\tau_3)| = f_{n+1}$, where $\tau_3 = \{123, 132, 213\}$ and $\{f_n\}_{n\geq 0}$ is the Fibonacci numbers, initialized by $f_0 = 0$, $f_1 = 1$.

2. For each $n \geq 1$, there is a constructive bijections between $S_n(\tau_3)$ and the set $F^{(2)}_{n-1}$ of binary strings of length $(n-1)$ having no *two* consecutive ones.

In 2004, Egge and Mansour [4] generalized the first point above and showed that for all integers n and $p \geq 2$, $|S_n(\tau_p)| = f^{(p-1)}_{n+1}$, where $\tau_p = \{12\ldots p, 132, 213\}$ and $f^{(p)}_n$ is the p-th order Fibonacci number.

In this paper we extend the Simion-Schmidt bijection (the second point), from $F^{(2)}_{n-1} \rightarrow S_n(\tau_3)$ to $F^{(p-1)}_{n-1} \rightarrow S_n(\tau_p)$, where $F^{(p)}_n$ is the set of length n binary strings with no p consecutive ones. We also show that this bijection is actually an isomorphism and we give a Gray code for the set $S_n(\tau_p)$ which is the image through this isomorphism of a known Gray code for $F_{n-1}^{(p-1)}$ [5]. Finally, we give some graph theoretic and algorithmic considerations.

2 SSEM (Simon-Schmidt-Egge-Masour) bijection

The p-th order n-th Fibonacci set, $F_n^{(p)}$, is the set of all length n binary strings with no p consecutive ones, and it can be defined recursively by [5]:

$$F_n^{(p)} = \begin{cases} \{\lambda\} & n = 0 \\ \{0,1\}^n & 1 \leq n < p \\ 0 \cdot F_{n-1}^{(p)} \cup 10 \cdot F_{n-2}^{(p)} \cup \ldots \cup 1^{p-1}0 \cdot F_{n-p}^{(p)} & n \geq p \end{cases} \tag{2.1}$$

where λ is the empty string, and for arbitrary string α and set of strings F we mean $\alpha \cdot F$ as concatenation of α to each string of F. It is easy to show that:

$$|F_n^{(p)}| = f_{n+p}^{(p)} \tag{2.2}$$

where $f_{n+p}^{(p)}$ is p-right shifting of $f_n^{(p)}$, and the latter is the pth order n-th Fibonacci numbers defined by:

$$f_n^{(p)} = \begin{cases} 0 & 0 \leq n < p-1 \\ 1 & n = p-1 \\ \sum_{j=n-p}^{n-1} f_j^{(p)} & n \geq p. \end{cases} \tag{2.3}$$

For $p = 2$ the above relation gives the well-known Fibonacci integer sequence $f_n^{(2)}$, which is usually (and also in this paper) written just as f_n.

Simion and Schmidt [3] showed that the cardinality of $S_n(\tau_3)$ is given by

$$|S_n(\tau_3)| = f_{n+1}. \tag{2.4}$$

Here is their nice proof. Let $\pi \in S_n(\tau_3)$ and π^{-1} its inverse, that is $\pi^{-1}(\pi(i)) = i$. It is trivial to verify the result for $n \leq 2$. Indeed, $S_1(\tau_3)$ consists of a single length one permutation and $S_2(\tau_3)$ contains two permutations, namely (12) and (21). If $n \geq 3$, then $\pi^{-1}(n) \leq 2$, else either 123 or 213 could not be avoided. If $\pi(1) = n$, then $(\pi(2), \ldots, \pi(n)) \in S_{n-1}(\tau_3)$. If $\pi(2) = n$, then we must have $\pi(1) = n-1$, else 132 could not be avoided; thus $(\pi(3), \ldots, \pi(n)) \in S_{n-2}(\tau_3)$. Hence, for $n \geq 3$, $|S_n(\tau_3)| = |S_{n-1}(\tau_3)| + |S_{n-2}(\tau_3)|$; this recurrence relation is satisfied by the Fibonacci numbers given in (2.3) with $p = 2$.

Simion and Schmidt also given a constructive bijection between the set $F_{n-1}^{(2)}$ and $S_n(\tau_3)$. Their construction is as follows. Let $s = s_1 s_2 \ldots s_{n-1} \in F_{n-1}^{(2)}$; its corresponding permutation $\pi \in S_n(\tau_3)$ is obtained by determining $\pi(i)$, $1 \leq i < n$, as follows: if $X_i = \{1, 2, \ldots, n\} - \{\pi(1), \ldots, \pi(i-1)\}$ then

$$\pi(i) = \begin{cases} \text{largest element in } X_i \text{ if } s_i = 0 \\ \text{second largest element in } X_i \text{ if } s_i = 1. \end{cases} \tag{2.5}$$

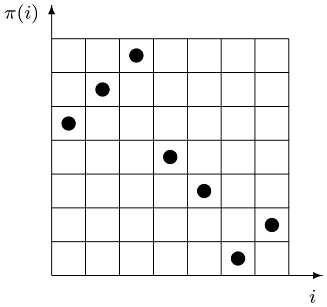


Figure 1: The permutation $(5674312) \in S_8(\tau_4)$ corresponding to the binary string $110001 \in F_7^{(3)}$.

Finally, $\pi(n)$ is the unique element in X_n.

Example 2.1 By this bijection $(564231) \in S_6(\tau_3)$ corresponds to $10010 \in F_5^{(2)}$ and $(7563412) \in S_7(\tau_3)$ corresponds to $010101 \in F_6^{(2)}$.

Simion-Schmidt counting relation (2.4) has been generalized by Egge and Mansour [4] to $S_n(\tau_p) = S_n(\{12\ldots p, 132, 213\})$:

$$|S_n(\tau_p)| = f_{n+p-2}^{(p-1)}. \tag{2.6}$$

But $f_{n+p-2}^{(p-1)}$ is the cardinality of $F_{n-1}^{(p-1)}$, the set of all binary strings of length $(n-1)$ having no $(p-1)$ consecutive ones. (2.6) is the generalization of (2.4) and now we extend (2.5) by giving a constructive bijection between $F_{n-1}^{(p-1)}$ and $S_n(\tau_p)$.

Let $s_1s_2\ldots s_{n-1} \in F_{n-1}^{(p-1)}$, X_i as above, and let π be the length n permutation defined by:

$$\pi(i) = \begin{cases} \text{largest element in } X_i \text{ if } s_i = 0 \\ 2^{nd} \text{ largest element in } X_i \text{ if } s_i = 1 \text{ and} \\ \quad \text{(either } s_{i+1} = 0 \text{ or } i = n-1) \\ 3^{rd} \text{ largest element in } X_i \text{ if } s_i = s_{i+1} = 1 \text{ and} \\ \quad \text{(either } s_{i+2} = 0 \text{ or } i = n-2) \\ \vdots \\ (p-2)^{th} \text{ largest element in } X_i \text{ if } s_i = s_{i+1} = \ldots = s_{i+p-4} = 1 \\ \quad \text{and (either } s_{i+p-3} = 0 \text{ or } i = n-p+3) \\ (p-1)^{th} \text{ largest element in } X_i \text{ if } s_i = s_{i+1} = \ldots = s_{i+p-3} = 1 \end{cases} \tag{2.7}$$

and $\pi(n)$ is the unique element in X_n. It is routine to verify that this construction yields a bijection from $F_{n-1}^{(p-1)}$ to $S_n(\tau_p)$ and in the following we refer it as the *SSEM bijection.*

Example 2.2

- SSEM bijection maps 110001 $\in F_7^{(3)}$ into $(5674312) \in S_8(\tau_4)$.
- SSEM bijection between $F_6^{(4)}$ and $S_7(\tau_5)$ maps 011101 into (7345612), and 011001 into (7456312).

See Figures 1 and 2 and Table 1.

Now, we will show that the SSEM bijection is a *combinatorial isomorphism*, that is, it and its inverse map *close* objects onto *close* objects.

In a permutation, we define a *left block* as a sequence of *increasing consecutive integers* which can not be extended at left. For instance, in $(56734128) \in S_8$, the sequences 56, 567, 34, 12, and 8 are left blocks. Notice that 67 is not a left block since it can be extended at left as 567. *Right block* is defined similarly. Also notice that 8 is a left block and at once also a right block.

Definition 2.3

1. Two permutations in $S_n(\tau_p)$ are close if one is obtained from the other by a transposition of two adjacent blocks of total length less than p, one a left block and the other a right block;
2. Two binary strings are close if they differ in a single position.

Example 2.4 The permutations $\pi = (7\underline{3}\,\underline{456}12)$ and $\pi' = (7\underline{456}\,\underline{3}12)$ in $S_7(\tau_5)$ are close since π is obtained from π' by transposing the right block 3 with the left block 456 in π'; or conversely, π' is obtained from π by transposing the right block 456 with the left block 3 in π; see Figure 2.

Theorem 2.5 *The SSEM bijection is a combinatorial isomorphism, that is, two binary strings in $F_{n-1}^{(p-1)}$ are close if and only if their images under this bijection are close in $S_n(\tau_p)$.*

Proof Let $s, s' \in F_{n-1}^{(p-1)}$ which differ just in position i, like the following scheme:

$$\begin{aligned} s &= s_1 s_2 \ldots s_{t-2} 0 \quad \underbrace{1\ldots1}_{i-t} s_i \underbrace{1\ldots1}_{u-i-1} 0 \quad s_{u+1} \ldots s_{n-1} \\ s' &= s_1 s_2 \ldots s_{t-2} 0 \quad \underbrace{1\ldots1}_{i-t} s'_i \underbrace{1\ldots1}_{u-i-1} 0 \quad s_{u+1} \ldots s_{n-1} \end{aligned} \tag{2.8}$$

where $s_t \ldots s_{i-1}$ and $s_{i+1} \ldots s_{u-1}$ are, possibly empty, contiguous sequences of 1s and $s'_i = 1 - s_i$. Without any loss of generality suppose $s_i = 1$ (and therefore $s'_i = 0$) and in this case $u - t$ (the length of contiguous sequence of 1s , including s_i, in s) is less than or equal to $p - 1$.

The shape of π and π', the images of s and s' through the SSEM bijection, are:

$$\begin{array}{rlll} \pi = & (\pi(1)\ldots\pi(t-1) & \pi(t)\ldots\pi(i-1)\pi(i)\pi(i+1)\ldots\pi(u) & \pi(u+1)\ldots\pi(n)) \\ \pi' = & (\pi(1)\ldots\pi(t-1) & \pi'(t)\ldots\pi'(i-1)\pi'(i)\pi'(i+1)\ldots\pi'(u) & \pi(u+1)\ldots\pi(n)) \,. \end{array} \tag{2.9}$$

By relation (2.7), $\pi(t)\ldots\pi(i)\ldots\pi(u)$ is at once left and right block in π and so are $\pi'(t)\ldots\pi'(i)$ and $\pi'(i+1)\ldots\pi'(u)$ in π'. Since $\{\pi(t),\ldots,\pi(i),\ldots,\pi(u)\}$ and $\{\pi'(t),\ldots,\pi'(i),\ldots,\pi'(u)\}$ are equal (as sets, but different as sequences) and $\pi'(u)<\pi'(t)$ (actually $\pi'(u)=\pi'(t)-1$) we have $\pi(t)\ldots\pi(i)\pi(i+1)\ldots\pi(u)=\pi'(i+1)\ldots\pi'(u)\pi'(t)\ldots\pi'(i)$. □

3 Gray code for $S_n(\tau_p)$

In this section we show how the SSEM isomorphism transforms a known Gray code for Fibonacci strings into a Gray code for the set of permutations $S_n(\tau_p)$.

By definition, a Gray code for a combinatorial family is a listing of objects in the family so that successive objects differ in some pre-specified, usually small, way [6]. In [5] is given a Gray code list for the set of Fibonacci strings defined by (2.1). In this list successive strings differ in a single position and its definition is:

$$\mathcal{F}_n^{(p)} = \begin{cases} \lambda & if \quad n=0 \\ 0,1 & if \quad n=1 \\ 0\cdot\overline{\mathcal{F}}_{n-1}^{(p)}\circ 10\cdot\overline{\mathcal{F}}_{n-2}^{(p)}\circ\ldots\circ 1^{p-1}0\cdot\overline{\mathcal{F}}_{n-p}^{(p)} & if \quad n>1 \end{cases} \tag{3.1}$$

where $\circ$ is the operator of concatenation of two lists, $\overline{\mathcal{F}}$ is the list obtained by reversing $\mathcal{F}$, and with two conventions: (1) the list $\alpha\cdot\mathcal{F}_{-1}^{(p)}$ consists of the single string list obtained from α by deleting its last bit, and (2) $\mathcal{F}_{-t}^{(p)}$ is the empty list for $t>1$.

By applying the SSEM bijection to each binary string in the list $\mathcal{F}_n^{(p)}$ one obtains a list for the set $S_{n+1}(\tau_{p+1})$; or equivalently, by the SSEM bijection, the Gray code $\mathcal{F}_{n-1}^{(p-1)}$ is transformed into the list $\mathcal{S}_n(\tau_p)$ for the set $S_n(\tau_p)$ defined

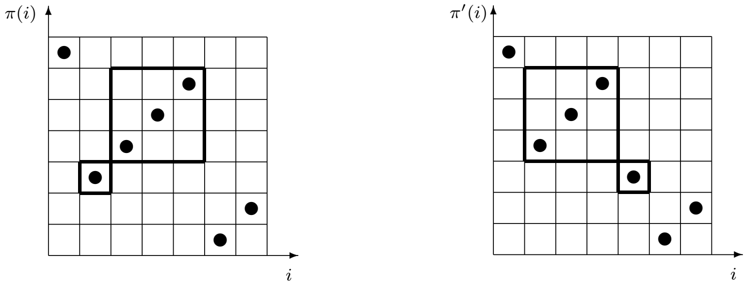


Figure 2: $\pi=(7345612)$ and $\pi'=(7456312)$ in $S_7(\tau_5)$ are the images of 011101 and 011001 in $F_6^{(4)}$, respectively. π' is obtained from π by transposing the left block 3 with the right block 456 in π.

by:

$$\mathcal{S}_n(\tau_p) = \begin{cases} (1) & \text{if} \quad n = 1 \\ (21), (12) & \text{if} \quad n = 2 \\ n \cdot \overline{\mathcal{S}}_{n-1}(\tau_p) \circ (n-1)n \cdot \overline{\mathcal{S}}_{n-2}(\tau_p) \circ & \\ \quad \cdots \circ (n-p+2) \dots n \cdot \overline{\mathcal{S}}_{n-p+1}(\tau_p) & \text{if} \quad n > 2. \end{cases} \tag{3.2}$$

with the conventions: (1) the list $\alpha \cdot \mathcal{S}_0(\tau_p) = \alpha$, and (2) $\mathcal{S}_{-t}(\tau_p)$ is the empty list for $t > 0$. Table 1 shows for the lists $\mathcal{F}_5^{(2)}$ and $\mathcal{F}_4^{(3)}$ with their images $\mathcal{S}_6(\tau_3)$ and $\mathcal{S}_5(\tau_4)$, respectively.

Since any two consecutive strings in $\mathcal{F}_{n-1}^{(p-1)}$ are two close, by Theorem 2.5, so are their images through SSEM bijection, hence the Hamming distance between consecutive permutations in $\mathcal{S}_n(\tau_p)$ is less than p. The following lemma formalizes this result using different approach from Theorem 2.5.

Lemma 3.1 *The Hamming distance between any two consecutive elements of $\mathcal{S}_n(\tau_p)$ is upper bounded by the minimum between $(p-1)$ and n.*

Proof We consider p fixed and for notational convenience we omit '(τ_p)' in this proof. Obviously, the Hamming distance between two consecutive elements in $\mathcal{S}_n$ is less than or equal to n. Suppose $n \geq p$ and let

$$(n-k)(n-k+1) \dots n \cdot \overline{\mathcal{S}}_{n-k-1} \tag{3.3}$$

and

$$(n-k-1)(n-k) \dots n \cdot \overline{\mathcal{S}}_{n-k-2} \tag{3.4}$$

be two (not empty) consecutive sublists in the definition (3.2), for $n > 2$, with $0 \leq k \leq p-3$. We show that the Hamming distance between the last element in the list (3.3) and the first one in (3.4) is less than p.

$$\begin{aligned} & \mathit{last}\left((n-k)(n-k+1) \dots n \cdot \overline{\mathcal{S}}_{n-k-1}\right) \\ & \qquad = \quad (n-k)(n-k+1) \dots n \cdot \mathit{last}(\overline{\mathcal{S}}_{n-k-1}) \\ & \qquad = \quad (n-k)(n-k+1) \dots n \cdot \mathit{first}(\mathcal{S}_{n-k-1}) \\ & \qquad = \quad (n-k)(n-k+1) \dots n(n-k-1) \cdot \mathit{first}(\overline{\mathcal{S}}_{n-k-2}). \end{aligned}$$

So, the last element in the list (3.3) differs from the first element in the list (3.4) in exactly $k + 2 \leq p - 1$ positions. The induction on n completes the proof. □

The following remark is useful to the generating algorithm sketched in the last section.

Remark 3.2 Let s and s' be two binary strings in the list $\mathcal{F}_{n-1}^{(p-1)}$ with s' the successor of s and π and π' their images by SSEM bijection as in schemes (2.8) and (2.9). If s differs from s' in position i then either $i = n-1$ or $s_{i+1} = s'_{i+1} = 0$, see [5] . With the notations in the proof of Theorem 2.5

Table 1: (a) The list $\mathcal{F}_5^{(2)}$ and its image $\mathcal{S}_6(\tau_3) = \mathcal{S}_6(123, 132, 213)$, and (b) The list $\mathcal{F}_4^{(3)}$ and its image $\mathcal{S}_5(\tau_4) = \mathcal{S}_5(1234, 132, 213)$ together with the Hamming distances between consecutive permutations. Notice that the Hamming distance between any two consecutive elements of $\mathcal{S}_6(\tau_3)$ is two.

(a)		(b)		
$\mathcal{F}_5^{(2)}$	$\mathcal{S}_6(\tau_3)$	$\mathcal{F}_4^{(3)}$	$\mathcal{S}_5(\tau_4)$	distance
01001	645312	0110	52341	
01000	645321	0100	53421	3
01010	645231	0101	53412	2
00010	654231	0001	54312	2
00000	654321	0000	54321	2
00001	654312	0010	54231	2
00101	653412	0011	54123	3
00100	653421	1011	45123	2
10100	563421	1010	45231	3
10101	563412	1000	45321	2
10001	564312	1001	45312	2
10000	564321	1101	34512	3
10010	564231	1100	34521	2

- if $s_i = 0$ then π' is obtained from π by transposing the block $\pi_t \dots \pi_i$ (with $t = i$ if $i = 1$ or $s_{i-1} = 0$) with the single element block π_{i+1},
- if $s_i = 1$ then π' is obtained from π by transposing the single element block π_t with the block $\pi_{t+1} \dots \pi_{i+1}$.

See Figure 2.

Notice that when $p = 3$ consecutive permutations in $\mathcal{S}_n(\tau_3)$ differ by the transposition of two adjacent elements.

4 Graph theoretic and algorithmic considerations

The isomorphism shown by Theorem 2.5 also has a graph theoretical meaning. Let X be a class of combinatorial objects and $G(X)$ be the graph induced by X, i.e., the graph with vertex set X, and edges connecting *close* vertices. With this terminology, a Gray code for X is a Hamiltonian path for $G(X)$.

Theorem 2.5 implies that the SSEM bijection is a graph isomorphism between $G(F_{n-1}^{(p-1)})$ and $G(S_n(\tau_p))$; this isomorphism transforms the Hamiltonian path $\mathcal{F}_{n-1}^{(p-1)}$ defined by (3.1) into the Hamiltonian path $\mathcal{S}_n(\tau_p)$ defined by (3.2). Figure 3 shows the graphs $G(F_4^{(3)})$ and $G(S_5(\tau_4))$ where the Hamiltonian paths $\mathcal{F}_4^{(3)}$ and $\mathcal{S}_5(\tau_4)$ are in bold.

Now, we explain how a slight modification of an efficient exhaustive generation algorithm for the list $\mathcal{F}_{n-1}^{(p-1)}$ transforms it into a similar algorithm for

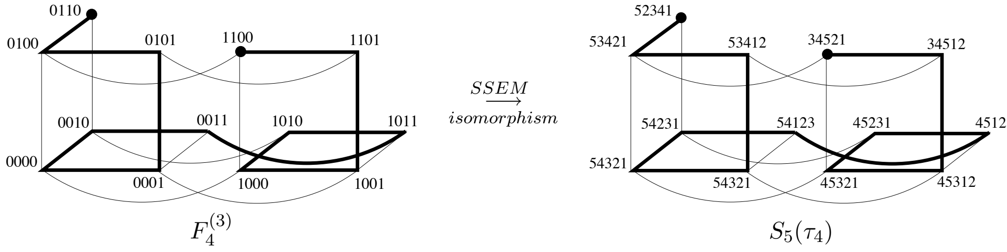


Figure 3: Graph isomorphism between $F_4^{(3)}$ and $S_5(\tau_4)$. Two vertices in $F_4^{(3)}$ are connected if their Hamming distance is one, while two vertices in $S_5(\tau_4)$ are connected if one is obtained from the other by transposing two adjacent blocks of size at most three. Bold lines are Hamiltonian paths.

$\mathcal{S}_n(\tau_p)$. In [5] is presented the loopless procedure *next* which after a linear-time precomputation step (and using additional data structures) computes, in constant time, the position i where the current string belonging to $\mathcal{F}_{n-1}^{(p-1)}$ must be changed in order to obtain the next one. *next* subsequently computes the length of the contiguous sequence of 1s ending in position $i-1$ (that is, $i-t$ with the notations in the proof of Theorem 2.5).

The following scheme yields a generating algorithm for $\mathcal{S}_n(\tau_p)$. Initialize s by the first string in $\mathcal{F}_{n-1}^{(p-1)}$ as in [5] and π by its image through the SSEM bijection; then, run *next* and update π as in Remark 3.2. The time complexity of the obtained algorithm is given by the second step—the blocks transposition—and it is $O(p)$ per permutation, independent on n. A linked representation for π can down this complexity to $O(1)$; see [7] for a detailed explanation of this technique.

References

[1] Donald E. Knuth, *The Art of Programming, Vol. 1 Fundamental Algorithms, Section 2.2.1.*, Addison Wesley, Reading Massachusetts, 1973.

[2] Sergey Kitaev and Toufik Mansour, *A Survey on Certain Pattern Problems*, www.ms.uky.edu/˜math/MAreport/ survey.ps (2003)

[3] Rodica Simion and Frank W. Schmidt, *Restricted Permutations*, Europ. J. Combinatorics (1985) **6**, pp. 383-406.

[4] Eric S. Egge and Toufik Mansour, *Restricted Permutations, Fibonacci Numbers, and k-generalized Fibonacci Numbers*, www.gettysburg.edu/˜eegge/kgenfib.pdf (2003).

[5] Vincent Vajnovszki, *A Loopless Generation of Bitstring without p Consecutive Ones*, Proceeding of 3-rd Discrete Mathematics and Theoretical Computer Science (2001), pp.227-239.

[6] J. T. Joichi, D. E. White, and S.G. Williamson, *Combinatorial Gray Codes*, SIAM Journal on Computing (1980) **9**, No. 1, pp. 130-141

[7] J. F. Korsh and S. Lipschutz, *Generating multiset permutations in constant time*, J. Algorithms (1997) **25**, No. 1, pp. 321–335

Transposition Invariant Words[*]

Arto Lepistö,[†]*Kalle Saari*[‡]

Abstract

We define a special type of finite words, so-called transposition invariant words. After presenting a few basic results from the point of view of combinatorics on words, and thus attempting to justify our interest in this concept, we give a full characterization of these words in terms of number theory. The characterization reveals a rather surprising connection to prime numbers; it divides prime numbers into two sets which we call favorable and unfavorable primes. We establish the infiniteness of favorable primes. Finally, we give two conjectures concerning unfavorable prime numbers and transposition invariant words.

Keywords: transposition invariant word, partition generated by a subgroup, favorable prime number.

1 Introduction

Let $w = w_0 w_1 \cdots w_n$ be a word, that is, a string of symbols over a finite alphabet Σ. The length of w is denoted by $|w|$, so that $|w| = n+1$. We will use the same notation for the cardinality of a set later in this paper, but this should cause no ambiguity. Assume then that $|w| = n+1 = pq$ for some integers $p, q > 0$, and consider the $p \times q$-matrix

$$A = \begin{pmatrix} w_0 & w_1 & \cdots & w_{q-1} \\ w_q & w_{q+1} & \cdots & w_{2q-1} \\ \vdots & & & \vdots \\ w_{(p-1)q} & w_{(p-1)q+1} & \cdots & w_{pq-1} \end{pmatrix}.$$

By reading the entries of this matrix row by row starting from the upper left corner, we obtain the word w. When reading the entries column by column, we get another word

$$w^T = w_0 w_q \cdots w_{(p-1)q}\, w_1 w_{q+1} \cdots w_{(p-1)q+1} \cdots w_{q-1} w_{2q-1} \cdots w_{pq-1}\,.$$

Equivalently, we obtain w^T by reading the transpose matrix A^T row by row.

[*]Supported by the Academy of Finland under grant 203354.

[†]Department of Mathematics, University of Turku, 20014 Turku, Finland, alepisto@utu.fi

[‡]Turku Centre for Computer Science, 20520 Turku, Finland, kasaar@utu.fi

If $w^T = w$, we say that w is *$p\times q$–invariant.* The word w is *$(|w|)$–transposition invariant*, or just *$(|w|)$–invariant*, if it is $p \times q$–invariant for all integers $p, q > 0$ for which $pq = |w|$. If the subword $w_1 w_2 \cdots w_{n-1}$ of w is unary, then w is trivially invariant. In this case we say that w is *trivial.*

Note that, in classic cryptography, transposition invariant words are the words that are immune to the rail fence cipher without padding.

For example, the Finnish word *möhömahat* —the people with a big belly— is 3×3–invariant word. Moreover, as $3 \cdot 3$ is the only proper factorization of 9, it follows that *möhömahat* is a non-trivial transposition invariant word. Similar examples in English are *Malayalam* and *votometer.* The last example is the Latin sentence

$$\begin{array}{ccccc} \mathrm{S} & \mathrm{A} & \mathrm{T} & \mathrm{O} & \mathrm{R} \\ \mathrm{A} & \mathrm{R} & \mathrm{E} & \mathrm{P} & \mathrm{O} \\ \mathrm{T} & \mathrm{E} & \mathrm{N} & \mathrm{E} & \mathrm{T} \\ \mathrm{O} & \mathrm{P} & \mathrm{E} & \mathrm{R} & \mathrm{A} \\ \mathrm{R} & \mathrm{O} & \mathrm{T} & \mathrm{A} & \mathrm{R}, \end{array}$$

which translates roughly as “The seed man Arepo uses wheels in his work” (thanks to M. Hirvensalo for pointing out this example).

From now on, we use notations $w = w_0 w_1 \cdots w_n$ and $w = w(0)w(1)\cdots w(n)$ interchangeably.

The following lemma is a direct consequence of the definition:

Lemma 1.1 *The word w is $p \times q$-invariant if and only if*

$$w(ip + j) = w(jq + i) \tag{1.1}$$

for all $0 \leq i < q$ and $0 \leq j < p$.

2 Basic Results

We begin by showing that certain finite prefixes of the celebrated Thue-Morse word t are invariant. The Thue-Morse word is generated by iterating the morphism $\mu : 0 \mapsto 01$, $1 \mapsto 10$ on 0 ad infinitum. Thus

$$t = \lim_{n\to\infty} \mu^n(0) = 0110100110010110100\cdots .$$

Proposition 2.1 *For all positive integers k, the prefix of length 2^k of the Thue-Morse word is transposition invariant.*

Proof Let $t = t(0)t(1)t(2)\cdots$ be the Thue-Morse word. It can be proved that $t(i)$ is the number $\pmod 2$ of symbols 1 in the binary expansion of i (see [6]). Using this property, it is easy to see that

$$t(2^e i + j) = t(i) + t(j) \pmod 2$$

for all integers $i \geq 0$ and $0 \leq j < 2^e$.

Let u be the prefix of t of length 2^k. We have to show u is $p \times q$–invariant whenever $pq = |u| = 2^k$, that is, whenever $p = 2^e$ and $q = 2^f$ with $e + f = k$.

Assume $0 \leq i < q = 2^f$ and $0 \leq j < p = 2^e$. Then

$$u(ip+j) = t(ip+j) = t(i2^e+j) = t(i)+t(j) \pmod 2 = t(j2^f+i) = u(jq+i)\,.$$

Thus by Lemma 1.1 the word u is $p \times q$–invariant. Since this is true for all appropriate p and q, the word u is 2^k–invariant. □

Next we show that there exist periodic non-trivial transposition invariant words.

Proposition 2.2 *For all integers $k, m \geq 0$, the word $w = 1(0^k1)^m$ is transposition invariant.*

Proof Assume $|w| = 1+m(k+1) = pq$. We will show that w is $p \times q$–invariant. To do this, let $0 \leq i < q$ and $0 \leq j < p$. Then $w(ip+j) = 1$ implies

$$ip + j = l(k+1) \tag{2.1}$$

for some $0 \leq l \leq m$. After multiplying the equation (2.1) by q, using the identity $pq = 1 + m(k+1)$, and rearranging the terms, we get $qj + i = (lq - im)(k+1)$ and hence $w(qj+i) = 1$. Using symmetry, we may deduce that $w(ip+j) = 1$ if and only if $w(jq+i) = 1$. By Lemma 1.1 w is $p \times q$–invariant. Since p and q were arbitrary, w is transposition invariant. □

Corollary 2.3 *There exists a non-trivial transposition invariant word of length $l > 4$ whenever integer l is odd, a perfect square, or $l \equiv 4 \pmod 6$.*

Proof In view of Proposition 2.2, take the transposition invariant word $1(0^k1)^m$. Then, for every integer $n > 0$, we have

$$|1(0^k1)^m| = 1 + m(k+1) = \begin{cases} 1+2n & \text{if } k = 1 \text{ and } m = n, \\ n^2 & \text{if } k = n \text{ and } m = n-1, \\ 4+6n & \text{if } k = 2n \text{ and } m = 3. \end{cases}$$

So assume that l is an integer greater than 4. Now if l is odd, perfect square, or congruent to $4 \pmod 6$, that is, if $l = 2n+1$, $l = n^2$, or $l = 4+6n$, respectively, then $n \geq 2$, $n \geq 3$, or $n \geq 1$, respectively. In each case, the corresponding word $1(0^k1)^m$ is non-trivial. □

Proposition 2.4 *If the word w is non-unary, then there exist a rational number $\alpha \geq 1$ and integers $p, q > 1$ such that w^α is a non-trivial $p \times q$–invariant word.*

Proof Assume $|w| = n$. Let two positive integers k and p satisfy the conditions $kn+1 = p$ and p prime. The existence of such integers is guaranteed because, by Dirichlet's Theorem (see [1]), every arithmetic progression with relatively prime coefficients contains infinitely many primes.

Now set $\alpha = p^2/n$, so that $|w^\alpha| = \alpha|w| = p^2$. We show that w^α is $p \times p$–invariant. To do so, assume $0 \leq i, j < p$. Using the fact that w^α has a period n, we get

$$w^\alpha(ip+j) = w^\alpha\,(ikn+i+j) = w^\alpha(i+j) = w^\alpha(jkn+j+i) = w^\alpha(jp+i)\,.$$

Again, by Lemma 1.1 w^α is $p \times p$–invariant. Moreover, it is non-trivial because $\alpha > 1$ and w is not unary. □

3 A Characterization of Transposition Invariant Words

Throughout this section, we assume that $w = w_0 w_1 \cdots w_n$ and $|w| = n+1 = pq$, where $p, q > 0$ are integers. Moreover, $<\!p\!>$ is the cyclic subgroup generated by p in the multiplicative group Z_n^*. Note that $<\!p\!> = <\!q\!>$ because $p = q^{-1}$ in $\mathbb{Z}_n^*$. The notation $k\!<\!p\!>$, where $k \in \mathbb{Z}_n$, is understood as the set of positions, or indices, in w obtained by multiplying the elements in $<\!p\!>$ by k and taking the positive residue $\pmod n$.

Proposition 3.1 *The word w is $p \times q$-invariant if and only if*

$$w(h) = w(k) \tag{3.1}$$

for all $k \in \mathbb{Z}_n$ and for all $h \in k\!<\!p\!>$.

Proof First, we show that the Condition (1.1) in Lemma 1.1 is equivalent to the condition

$$w(k) = w\big(kp \pmod n\big) \tag{3.2}$$

for all $k \in \mathbb{Z}_n$. This is true because $k = jq + i$, with $0 \leq i < q$ and $0 \leq j < p$, if and only if

$$j = \left\lfloor \frac{k}{q} \right\rfloor \qquad \text{and} \qquad i = k \pmod q\,.$$

Furthermore, in this case

$$\begin{aligned} ip+j &= (k - \left\lfloor \frac{k}{q} \right\rfloor q)p + \left\lfloor \frac{k}{q} \right\rfloor \\ &= kp - (pq-1)\left\lfloor \frac{k}{q} \right\rfloor \\ &= kp \pmod{(pq-1)} \\ &= kp \pmod n\,. \end{aligned}$$

Thus the Condition (1.1) is equivalent to the Condition (3.2).

Using the identity in (3.2) repeatedly, we see that

$$\begin{aligned} w(k) &= w(\,kp \pmod{n}\,) = w(\,kp^2 \pmod{n}\,) \\ &= w(\,kp^3 \pmod{n}\,) = w(\,kp^4 \pmod{n}\,) = \cdots \end{aligned}$$

for all $k \in \mathbb{Z}_n$. But this is exactly the requirement in Condition (3.1). Thus it follows that (3.1) is equivalent to (3.2), which concludes our proof. □

Now we are ready to establish a number theoretic characterization for transposition invariant words. With the same considerations we can prove a somewhat more general result. To do that, let S_n be the set of all positive divisors of $n+1$, that is,

$$S_n = \{d > 0 \,:\, d \,|\, (n+1)\}\ .$$

Let $S \subseteq S_n$. We say that the word w is *S–invariant* if it is $p \times (n+1)/p$–invariant for all $p \in S$. Then the concepts $p \times q$–invariant and transposition invariant coincide with $\{p\}$–invariant and S_n–invariant, respectively.

Proposition 3.2 *Let $S \subseteq S_n$. Then the word w is S–invariant if and only if, for every $k \in \mathbb{Z}_n$, all letters at positions indicated by the set $k{<}S{>}$ are the same.*

Proof Assume w is S–invariant, that is, w is $p \times (n+1)/p$–invariant for every $p \in S$. Let $r, s \in S$. Using the condition (3.1), we see that $kr^is^j \in kr^i{<}s{>}$ implies $w(kr^is^j) = w(kr^i)$, and, moreover, $kr^i \in k{<}r{>}$ implies $w(kr^i) = w(k)$. Thus, for all elements $h \in k{<}r{>}{<}s{>} = k{<}r,s{>}$, we have $w(h) = w(k)$. It follows by induction that, for every $h \in k{<}S{>}$, $w(h) = w(k)$.

Conversely, assume that $w(h) = w(k)$ for every $h \in k{<}S{>}$. Then, because ${<}p{>} \subseteq {<}S{>}$ for all $p \in S$, it certainly holds that $w(h) = w(k)$ for every $h \in k{<}p{>}$. According to Proposition 3.1, the word w is $p \times (n+1)/p$–invariant for every $p \in S$, that is, S–invariant. □

Next we will find the maximal number of distinct letters in S–invariant words. But first we need a lemma. The following lemma has a short and straightforward proof, so we omit it.

Lemma 3.3 *Let $S \subseteq \mathbb{Z}_n^*$ and $k, h \in \mathbb{Z}_n$. Then either*

$$k{<}S{>} = h{<}S{>} \qquad \textit{or} \qquad k{<}S{>} \cap h{<}S{>} = \emptyset\,.$$

This Lemma implies that every subset $S \subseteq \mathbb{Z}_n^*$ induces a partition of $\mathbb{Z}_n$ by means of ${<}S{>} \leq \mathbb{Z}_n^*$, the subgroup $\mathbb{Z}_n^*$ generated by S. More precisely, there exist integers $k_1, k_2, \ldots, k_r \in \mathbb{Z}_n$ such that

$$\mathbb{Z}_n = \bigcup_{1 \leq i \leq r} k_i{<}S{>}$$

and the sets $k_i{<}S{>}$ and $k_j{<}S{>}$ are disjoint if $i \neq j$. We call this partition *the partition of $\mathbb{Z}_n$ generated by the set $S \subseteq \mathbb{Z}_n^*$* and denote it by $\Pi_n(S)$, that is,

$$\Pi_n(S) = \{k_1{<}S{>}, \ldots, k_r{<}S{>}\}\,.$$

Note that by Proposition 3.2 the maximal number of distinct letters in an S–invariant word is the number of elements in $\Pi_n(S)$ plus one (remember that $|w| = n+1$ and that Proposition 3.2 does not put any constraint to the last letter in w).

Assume that $S \subseteq \mathbb{Z}_n^*$ and that $d > 1$ is a divisor of n. In what follows, we use the notation ${<}S{>}_d$ for the subgroup generated by S in the group $\mathbb{Z}_d^*$. Clearly this is a sound definition since, for all $a \in S$, $(a,n) = 1$ and $d\,|\,n$ implies $(a,d) = 1$.

Proposition 3.4 *Let $S \subseteq \mathbb{Z}_n^*$. Then*

$$|\,\Pi_n(S)| \;=\; \sum_{d|n}\, [\mathbb{Z}_d^* \,:\, {<}S{>}_d] \;=\; \sum_{d|n} \frac{\varphi(d)}{|{<}S{>}_d|}\,,$$

where φ denotes the Euler totient function.

Proof For all $k_i{<}S{>} \in \Pi_n(S)$, write $k_i = a_i d_i$, where $d_i = \gcd(k_i, n)$ and $a_i = k_i/d_i$, so that

$$\Pi_n(S) = \{a_1 d_1{<}S{>}, \ldots, a_r d_r{<}S{>}\}\,.$$

We need a few auxiliary results to prove the statement. They are numbered accordingly.

If $\gcd(a,n) = 1$, then $adi \equiv adj \pmod{n}$ if and only if $di \equiv dj \pmod{n}$, and thus

$$1) \qquad |ad{<}S{>}| = |d{<}S{>}|\,.$$

Consider next the mapping $d{<}S{>} \longrightarrow {<}S{>}_{n/d}$ defined by $di \mapsto i$. First of all, this mapping is well-defined because $\gcd(i,n) = 1$ implies $\gcd(i, n/d) = 1$. Moreover, it is injective, which is easily seen by using the equivalence $di \equiv dj \pmod{n}$ if and only if $i \equiv j \pmod{n/d}$. Consequently,

$$2) \qquad |d{<}S{>}| \leq |{<}S{>}_{n/d}|\,.$$

Now let us define the mapping β as follows:

$$\beta \,:\, \Pi_n(S) \longrightarrow \bigcup_{d\,|\,n} Z_{n/d}^*/{<}S{>}_{n/d}\,, \qquad ad{<}S{>} \mapsto a{<}S{>}_{n/d}\,.$$

We leave it to the reader to verify that β is both well-defined and injective. Consider the number α_d of the sets in $\Pi_n(S)$ of the form $ad{<}S{>}$ with $\gcd(a,n) = 1$. Clearly

$$\alpha_d = \{i \,:\, d = \frac{k_i}{a_i}\}\,,$$

so that

$$\sum_{d|n} \alpha_d = |\,\Pi_n(S)|\,. \tag{3.3}$$

It follows from the definition and injectivity of β that α_d is at most the number of elements in the quotient group $\mathbb{Z}^*_{n/k}/{<}S{>}_{n/d}$, that is to say,

$$3) \qquad \alpha_d \leq [Z^*_{n/d} : {<}S{>}_{n/d}]\,.$$

Now we are ready to employ these observations. Recall that $\Pi_n(S)$ is a partition of $\mathbb{Z}_n$, and hence

$$\begin{aligned} n &= \sum_{ad<S>\in\Pi(S)} |ad{<}S{>}| \overset{1)}{=} \sum_{ad<S>\in\Pi(S)} |d{<}S{>}| \\ &= \sum_{d\,|\,n} \alpha_d\,|d{<}S{>}| \overset{2)\,,3)}{\leq} \sum_{d\,|\,n} [\mathbb{Z}^*_{n/d} : {<}S{>}_{n/d}]\,|{<}S{>}_{n/d}| \qquad (3.4) \\ &= \sum_{d\,|\,n} \varphi(\frac{n}{d}) \;=\; n \qquad \text{(see, e.g., [1], Theorem 2.2).} \end{aligned}$$

Thus Inequality (3.4) is actually an equality. Consequently, inequalities in 2) and 3) also are equalities. Hence $\alpha_d = [\mathbb{Z}^*_{n/d} : {<}S{>}_{n/d}]$, and this, together with (3.3), attests the statement. □

We introduce the notation

$$\iota_n(S) = |\,\Pi_n(S)| = \sum_{d|n} \frac{\varphi(d)}{|{<}S{>}_d|}\,.$$

Also, we say that an S–invariant word is *alphabetically maximal* if there does not exist another S–invariant word with more distinct letters occurring in it.

By merging the previous considerations to Proposition 3.2, we can sum up our conversation about S–invariant words as follows:

Theorem 3.5 *Let $S \subseteq S_n$ and $\Pi_n(S)$ be the partition of $\mathbb{Z}_n$ generated by S. Then word w of length $n+1$ is S–invariant if and only if, for every set $P \in \Pi_n(S)$, we have $|w(P)| = 1$. Hence there exists, up to renaming, a unique alphabetically maximal S–invariant word, and it has $\iota_n(S)+1$ distinct letters. Furthermore, there are exactly*

$$k^{\iota_n(S)+1}$$

S–invariant words over k-letter alphabet.

4 Favorable Prime Numbers

We begin with an example: let the word $w = w_0 w_1 \cdots w_n$ of length $n+1$ over four letter alphabet $A = \{a, b, c, d\}$ be determined by the condition

$$w_i = \begin{cases} a & \text{if } i = 0, \\ b & \text{if } \gcd(i,n) = 1, \\ c & \text{if } \gcd(i,n) > 1 \text{ and } i < n, \\ d & \text{if } i = n \end{cases}$$

for all $0 \le i \le n$. Then w is transposition invariant, for if $0 \le i, j \le n$, where $\gcd(i,n) = 1$ and $\gcd(j,n) > 1$, then $i{<}S{>} \cap j{<}S{>} = \emptyset$. Moreover, if n is composite, then both letters b and c occur in w, and thus w is non-trivial. We conclude that if a positive integer n is composite, then there always exist non-trivial S–invariant words of length $n+1$ for every $S \subseteq S_n$.

This leads us to the next result:

Theorem 4.1 *Let $S \subseteq S_n$. Then there exist only trivial S–invariant words of length $n+1$ if and only if n is prime and ${<}S{>} = \mathbb{Z}_n^*$.*

Proof Assume there exist only trivial S–invariant words of length $n+1$. This is equivalent to the condition that the partition of $\mathbb{Z}_n$ generated by ${<}S{>}$ has only two elements, $\{0\}$ and $\{1, 2, \ldots, n-1\}$, which then have to be $0{<}S{>}$ and $1{<}S{>}$, respectively. This is equivalent to ${<}S{>} = \mathbb{Z}_n \setminus \{0\}$, which, in turn, happens exactly when n is prime and ${<}S{>} = \mathbb{Z}_n \setminus \{0\} = \mathbb{Z}_n^*$. □

Motivated by Theorem 4.1, we say that a prime number n is *favorable (for the existence of non-trivial invariant words)* if there exists a non-trivial $(n+1)$–invariant word, that is, if ${<}S_n{>} \neq \mathbb{Z}_n^*$. Next we will prove that there exist infinitely many favorable primes. This is done number theoretically by using *quadratic residues* (see, e.g., [1] or [5]). First we need the following lemma:

Lemma 4.2 *If a positive integer n is prime and $n \equiv 7 \pmod 8$, then every integer dividing $n+1$ is a quadratic residue $\pmod n$.*

Proof Since the product of two quadratic residues $\pmod n$ is a quadratic residue, it is enough to show that each prime divisor of $n+1$ is a quadratic residue $\pmod n$. So assume that p is prime and $p \mid n+1$. We have to consider the case $p = 2$ separately. Since $n \equiv 7 \pmod 8$, we have

$$\left(\frac{2}{n}\right) = 1\,,$$

where $\left(\frac{2}{n}\right)$ is the Legendre's symbol. Thus 2 is a quadratic residue $\pmod n$. Now we may assume that $p > 2$. Then, by using the basic principles of residue computing, we get

$$\begin{aligned}
\left(\frac{p}{n}\right) &= (-1)^{\frac{p-1}{2}\cdot\frac{n-1}{2}}\left(\frac{n}{p}\right) && \big(\text{law of quadratic reciprocity}\big)\\
&= (-1)^{\frac{p-1}{2}}\left(\frac{n}{p}\right) && \Big(\, n \equiv 7 \pmod 8 \text{ implies } \frac{n-1}{2} \text{ odd }\Big)\\
&= (-1)^{\frac{p-1}{2}}\left(\frac{-1}{p}\right) && \big(\, n \equiv -1 \pmod p \,\big)\\
&= (-1)^{\frac{p-1}{2}}\cdot(-1)^{\frac{p-1}{2}}\\
&= 1\,.
\end{aligned}$$

Hence p is a quadratic residue $\pmod n$. This completes our proof. □

Theorem 4.3 *There exist infinitely many favorable primes.*

Proof Let n be a prime number with $n \equiv 7 \pmod 8$. Dirichlet's classic theorem about arithmetic progressions (see [1]) says that there exist infinitely many such primes n. By Lemma 4.2, every integer in the set S_n is a quadratic residue $\pmod n$. Thus every integer in the set $<S_n>$ also is a quadratic residue. But exactly half of all the integers in $\mathbb{Z}_n^*$ are quadratic residues $\pmod n$; the other half is the set of quadratic non-residues $\pmod n$. Therefore, $<S_n> \neq \mathbb{Z}_n^*$, and non-trivial $(n+1)$–invariant words exist. Now, by definition, n is favorable. □

The question as to whether or not there exist infinitely many unfavorable primes seems to be more difficult. Computer aided calculations encourages us to formulate the following conjecture (see Table 1 for the unfavorable primes up to 1000):

Conjecture 4.4 *There exist infinitely many unfavorable primes.*

3	5	11	13	17	19	29	37	41	53	59	61	67	83	89
97	101	107	109	113	131	137	139	149	163	173	179	181	197	211
227	229	233	251	257	269	281	293	307	317	347	349	353	373	379
389	401	419	421	433	443	449	461	467	491	499	509	521	523	541
547	557	563	569	571	587	593	601	613	617	619	641	643	653	659
661	677	683	701	709	757	761	769	773	787	797	809	821	827	829
853	857	859	877	881	883	907	929	937	941	947	953	971	977	

Table 1: The unfavorable primes less than 1000

By Theorem 4.1 there only exist trivial $p \times q$–invariant words if and only if $n = pq - 1$ is prime number and p its primitive root. Since primitive roots are of general interest in number theory, we also give the following conjecture:

Conjecture 4.5 *There exist infinitely many primes n having a primitive root that divides $n+1$.*

Using the terminology of transposition invariant words, Conjecture 4.5 says that there are infinitely many positive integers p, q for which all $p \times q$–invariant words are trivial.

Remark 4.6 All the primes in Table 1 satisfy the condition of Conjecture 4.5 except the prime 571: it does not have a primitive root that divides 572. However $<S_{571}> = \mathbb{Z}^*_{571}$, so by the definition, 571 is an unfavorable prime.

5 Notes

There is no reason to limit the concept of invariant words to matrices, i.e., to two dimensions. In fact, there is a natural generalization to the third dimension, and it gives rise to new questions and problems.

In general, questions concerning primitive roots modulo a prime number are extremely difficult. However, our conjectures are considerably less strict than Artin's Conjecture for primitive roots (see [2] and [4]), which says that, given any non-zero integer a other than $1, -1$, or a perfect square, there exist infinitely many primes p for which a is a primitive root (mod p). In our case, it would be enough to show that Artin's conjecture holds true for $a = 2$. It is known that one of the primes 2, 3, or 5 is a primitive root (mod p) for infinitely many primes p (see [3]).

6 Acknowledgment

The second author wishes to thank Prof. Jeffrey Shallit for numerous discussions and the School of Computing at the University of Waterloo for the hospitality.

References

[1] T. M. Apostol, *Introduction to Analytic Number Theory*, New York: Springer-Verlag, 1976.

[2] L. Goldstein, Density questions in algebraic number theory, *Amer. Math. Monthly* **78**, 342–351, 1971.

[3] D. R. Heath-Brown, Artin's conjecture for primitive roots, *Quart. J. Math. Oxford* (2) **37**, 27-38, 1986.

[4] C. Hooley, On Artin's Conjecture, *J. Reine Angew. Math.* **225**, 209–220, 1967.

[5] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory (2nd ed.)*, Graduate texts in mathematics **84**, New York: Springer-Verlag, 1990.

[6] M. Lothaire, *Combinatorics on Words*, Vol. 17 of *Encyclopedia of Mathematics and Its Applications.* Addison-Wesley, 1983.

On the equation $XL = LX$[*]

Paolo Massazza[†]

Abstract

Given a language $L \subseteq \Sigma^\star$, the centralizer of L is the largest subset of $\Sigma^\star$ commuting with L, that is, the maximal solution of the equation $XL = LX$. In this paper we show that if the smallest word (with respect to a lexicographic order) of a language L is primitive and prefix distinguishable in L, then the centralizer of L is $L^\star$.

1 Introduction

Equations with languages as variables have been considered with particular interest since the work of Chomsky and Schützenberger [4]. For example, in [2] special systems of equations (called left language equations) have been studied for their relation with boolean automata and sequential networks. The equation $X^p = L$, for instance, has been studied in [11], where the problem of the existence of the solution has been shown to be decidable when L is a recognizable set over a free monoid. In [1] it has been studied how to decompose a language by writing it as union of powers of a simpler language, showing that a rational language admits a rational decomposition.

The equation $XL = LX$ has been deeply investigated since 1971, when Conway raised a problem concerning the commutation with rational sets ([5]). More precisely, his question was about the centralizer of a rational set L, that is, the maximal solution of the equation $XL = LX$: given a rational set L, is the centralizer of L rational?

Several interesting results concerning the commutation of languages have been presented since then. In particular, in the case of codes, in [10] it has been shown that if a code L and a circular code X commute, then $L = X^k$ for a suitable integer k. Moreover, for each prefix code L (no word is a prefix of another) its centralizer is always $L^\star$. Ratoandromanana also raised a conjecture that was solved several years later: in fact, in [7] it has been shown that a language commutes with a code X if and only if it is a union of powers of the primitive root of X, $\rho(X)$ (this was done by first proving that the centralizer of a code X is always $\rho(X)^*$).

[*] Partially supported by the Project M.I.U.R. COFIN 2003-2005 "Formal languages and automata: methods, models and applications".

[†] Università degli Studi dell'Insubria, Dipartimento di Informatica e Comunicazione, Via Mazzini 5, 21100 Varese (Italy), `paolo.massazza@uninsubria.it`

Partial answers to the Conway's problem have been given recently. In [3] the commutation with a two-word set L has been studied, showing that the centralizer is A^+ where either $A = L$ (if L consists of two noncommuting words) or $A = \{t\}$ and t is a primitive word (if L consists of two commuting words $x = t^r$, $y = t^s$). Later on, a similar result has been given for three-word sets (see [6], [7]).

More recently, a definitive and negative answer to the question raised by Conway has been found. In fact, a finite language whose centralizer is not recursively enumerable is shown in [8].

So, since the centralizer of a very simple language (like a finite language) can be very complex, it is interesting to look for sufficient conditions that guarantee that the centralizer is as simple as possible. According to this goal, in this paper we show that if the smallest word (with respect to a lexicographic order) of a language L is primitive and prefix distinguishable in L (i.e. it is not a prefix of other words in L) then the centralizer of L is $L^\star$, that is, the submonoid generated by L.

2 Preliminaries

Let Σ be a finite alphabet and $\leqslant$ a linear order on Σ. For $\sigma, \tau \in \Sigma$ we write $\sigma < \tau$ if $\sigma \leqslant \tau$ and $\sigma \neq \tau$. A language on Σ is a subset of the free monoid generated by Σ, $L \subseteq \Sigma^\star$. Given a word $w \in \Sigma^\star$ we denote its length by $|w|$. The word of length 0 is the *empty* word ϵ. For two words $x, w \in \Sigma^\star$, we say that x is a *prefix* of w, and write $x \leq w$, if and only if $w = xy$ for a suitable $y \in \Sigma^\star$. If $y \neq \epsilon$ we say that x is a *proper prefix* of w and write $x < w$. Analogously, a word y is a *suffix* of w if and only if $w = xy$ for a suitable word x.

Given a word w and an integer e, $0 \leq e \leq |w|$, we denote by $\text{pref}_e(w)$ $(\text{suf}_e(w))$ the prefix (suffix) of w having length e, that is, the string $x \in \Sigma^e$ such that $w = xy$ $(w = yx)$ for a suitable $y \in \Sigma^{|w|-e}$. Two words x, y are said to be *prefix incomparable* if $x \not\leq y$ and $y \not\leq x$. A word $w \in L$ is said *prefix distinguishable* in L if and only if for any $y \in L \setminus \{w\}$, w and y are prefix incomparable.

We can extend the relation $\leqslant$ on Σ in order to define a lexicographic order $\leq_{\text{lex}}$ on $\Sigma^\star$. Given $x, y \in \Sigma^\star$, we write $x \leq_{\text{lex}} y$ if and only if either $x \leq y$ or there exist $\alpha, u, v \in \Sigma^\star$ and $\sigma, \tau \in \Sigma$ such that $x = \alpha\sigma u$ and $y = \alpha\tau v$ with $\sigma < \tau$. We write $x <_{\text{lex}} y$ if $x \leq_{\text{lex}} y$ and $x \neq y$. We denote by $\min_{\text{lex}}(L)$ the smallest word of L with respect to $\leq_{\text{lex}}$, that is, the word $x \in L$ such that $x \leq_{\text{lex}} y$ for all $y \in L$. Some useful and elementary properties of the relation $\leq_{\text{lex}}$ are listed below.

(a0) $wu \leq_{\text{lex}} wv$ implies $u \leq_{\text{lex}} v$;

(a1) $u \leq_{\text{lex}} v$ implies $wu \leq_{\text{lex}} wv$ for all w;

(b) $u \leq_{\text{lex}} v$ and $u \not\leq v$ implies $uw \leq_{\text{lex}} vz$ for all w, z;

(c) if $uw \leq_{\text{lex}} vz$ for some u, v, w, z with $u \neq v$ and $|u| = |v|$, then $u <_{\text{lex}} v$;

(d) if $w_0 = \min_{\text{lex}}(L)$ is prefix distinguishable in L then for each $Y \subseteq \Sigma^\star$ the condition $w_0 y \in LY$ implies $y \in Y$.

We say that $X \subseteq \Sigma^\star$ commutes with L if and only if $XL = LX$. Note that if S and R commute with L then $S \cup R$ commutes with L. So, we define the *centralizer* of $L \subseteq \Sigma^\star$ as the largest subset of $\Sigma^\star$ that commutes with L, that is, the maximal solution of the equation $XL = LX$. We indicate by $\mathcal{C}(L)$ the centralizer of L; note that if A commutes with L then $A \subseteq \mathcal{C}(L)$. In particular, since for any L we have $L^\star L = LL^\star$, the submonoid generated by L is always contained in $\mathcal{C}(L)$.

It is immediate to see that if ϵ belongs to L then $\mathcal{C}(L)$ is $\Sigma^\star$. Henceforth, we are interested in the centralizer of languages that do not contain ϵ.

A word w is called *primitive* if $w = u^r$ implies $u = w$ and $r = 1$. We recall here one of the oldest results in combinatorics on words regarding the commutation of words (see, for example, [9]).

Theorem 2.1 *Let $u, v \in \Sigma^\star$. The following properties are equivalent*

1. $uv = vu$,

2. there exist $t \in \Sigma^\star$ and $r, s \in \mathbb{N}$ such that $u = t^r, v = t^s$.

3 Auxiliary results

In this section we consider some lemmata that state useful properties associated with commuting languages. We point out that all the results we present are obtained by purely combinatorial methods based on Properties (a0)-(d) shown above and on Theorem 2.1. A first lemma regards the smallest words of L and $\mathcal{C}(L) \setminus \{\epsilon\}$.

Lemma 3.1 *Let $L \subseteq \Sigma^+$ and $\alpha_0 = \min_{lex}(L)$, $w_0 = \min_{lex}(\mathcal{C}(L) \setminus \{\epsilon\})$. If α_0 is prefix distinguishable in L then*

$$\alpha_0 w_0 = w_0 \alpha_0 = \min_{lex}((C(L) \setminus \{\epsilon\})L) = \min_{lex}(L(\mathcal{C}(L) \setminus \{\epsilon\})).$$

Proof Let $\beta_1 = \min_{\text{lex}}(L(\mathcal{C}(L) \setminus \{\epsilon\}))$ and $\beta_2 = \min_{\text{lex}}((\mathcal{C}(L) \setminus \{\epsilon\})L)$. Since α_0, the smallest word of L, is prefix distinguishable in L, we have $\alpha_0 \leq_{\text{lex}} \alpha_i$ and $\alpha_0 \not\leq \alpha_i$ for all $\alpha_i \in L$, $i \neq 0$. So, by Property (b) we have $\alpha_0 w_0 \leq_{\text{lex}} \alpha_i w_j$ for all $\alpha_i \in L$ and $w_j \in \mathcal{C}(L) \setminus \{\epsilon\}$, that is,

$$\beta_1 = \alpha_0 w_0.$$

Note that β_1 belongs to $\mathcal{C}(L)L$ (since $L\mathcal{C}(L) = \mathcal{C}(L)L$) and $\beta_1 \neq \alpha$ for all $\alpha \in L$ (recall that α_0 is prefix distinguishable in L). This implies

$$\beta_1 = w_r \alpha_s$$

for suitable $w_r \in \mathcal{C}(L) \setminus \{\epsilon\}, \alpha_s \in L$. Now, since $\alpha_0 \leq_{\text{lex}} \alpha_i$, by Property (a1) we obtain

$$\beta_2 = w_p \alpha_0$$

for a suitable $w_p \in \mathcal{C}(L) \setminus \{\epsilon\}$. Note that we cannot have $\beta_2 \in L$ since otherwise we would have $\beta_2 <_{\text{lex}} \alpha_0$ because $\beta_2 \leq_{\text{lex}} \beta_1 = \alpha_0 w_0$ and $\alpha_0 \not< \alpha$ for all $\alpha \in L$. So, β_2 belongs to $L(\mathcal{C}(L) \setminus \{\epsilon\})$ and we immediately get

$$\beta_2 = \beta_1 = \alpha_0 w_0.$$

Finally, the condition $\beta_2 = w_p \alpha_0$ implies $|w_p| = |w_0|$: this and the inequality $w_p \alpha_0 \leq_{\text{lex}} w_0 \alpha_0$ imply $w_p = w_0$ (see Property (c)). □

The following lemma is of technical nature and it is mainly thought to help the proof of the main result in Section 4.

Lemma 3.2 *Let $L \subseteq \Sigma^+$ such that $\alpha_0 = \min_{lex}(L)$ is primitive and prefix distinguishable in L. Then, for any $A \subseteq \Sigma^\star$ commuting with L, if there are $w_1, w_2 \in A$ such that $w_1 \leq_{lex} w_2$ and $w_2 = \alpha_0 u$, $w_1 = u\alpha$ for suitable $u \in \Sigma^+$, $\alpha \in L$, then $w_1 \leq w_2$.*

Proof (By contradiction) We suppose $w_1 \not\leq w_2$ and show that we can find a word in AL that is not in LA.
Let us consider $\gamma, t_1, t_2 \in \Sigma^\star$ and $\sigma, \tau \in \Sigma$, with $\sigma < \tau$, such that

$$w_1 = \gamma\sigma t_2, \quad w_2 = \gamma\tau t_1.$$

Since $w_1 \leq_{\text{lex}} w_2$ and $w_2 = \alpha_0 u$, $w_1 = u\alpha$, we can find $k \in \mathbb{N}$, $e \in \{0, \ldots, |\alpha_0| - 1\}$ and $y \in \Sigma^\star$ such that

$$w_1 = \alpha_0^k \text{pref}_e(\alpha_0)\sigma t_2, \quad w_2 = \alpha_0^k \text{pref}_{e+1}(\alpha_0) y$$

with

$$\text{pref}_e(\alpha_0)\sigma <_{\text{lex}} \text{pref}_{e+1}(\alpha_0) \leq_{\text{lex}} \alpha_0.$$

Now, we distinguish two cases.

$(k = 0)$ We observe that for any $\beta_1, \beta_2 \in L$ and $z \in A$ we have

$$w_1\beta_1 = \text{pref}_e(\alpha_0)\sigma t_2 \beta_1 \neq \beta_2 z$$

because for each $v \leq w_1\beta_1$ we have $v <_{\text{lex}} \alpha_0$. This means that there is a word in AL that is not in LA.

$(k > 0)$ Since $AL = LA$ and $\alpha_0 < w_1$, for each $\beta \in L$ we have

$$w_1\beta = \alpha_0 z_1 \in LA.$$

Then, by Property (d) it follows that $z_1 \in A$,

$$z_1 = \alpha_0^{k-1}\text{pref}_e(\alpha_0)\sigma t_2\beta.$$

If $k-1>0$ we proceed by considering the word $z_1\beta = \alpha_0 z_2$ with $z_2 \in A$ (by Property (d)),

$$z_2 = \alpha_0^{k-2}\text{pref}_e(\alpha_0)\sigma t_2\beta^2.$$

Then, after k iterations, we get a word $z_k \in A$,

$$z_k = \text{pref}_e(\alpha_0)\sigma t_2\beta^k,$$

such that for any $\delta \in L$ and $z \in A$ we have

$$z_k\beta = \text{pref}_e(\alpha_0)\sigma t_2\beta^{k+1} \neq \delta z,$$

that is, $z_k\beta \in AL$ and $z_k\beta \notin LA$.

□

The following lemma states an important property associated with the left-product by a language whose smallest word is prefix distinguishable.

Lemma 3.3 *Let $L \subseteq \Sigma^+$ and $A_1, A_2 \subseteq \Sigma^\star$ such that $LA_1 = LA_2$. If $\alpha_0 = \min_{lex}(L)$ is prefix distinguishable in L then $A_1 = A_2$.*

Proof (By contradiction) Suppose $A_1 \neq A_2$ and, without loss of generality, let $\tilde{w} = \min_{\text{lex}}(A_2 \setminus A_1)$. Let us consider the word $w = \alpha_0\tilde{w}$ and note that $w \in LA_1$ because $LA_1 = LA_2$. Since α_0 is prefix distinguishable in L, by Property (d) we have $\tilde{w} \in A_1$. □

4 Main result

In this section we prove our main result about the centralizer of languages whose smallest word is primitive and prefix distinguishable. For the sake of clarity, we first prove in the following theorem a fundamental property on which our result is based.

Theorem 4.1 *Let $L \subseteq \Sigma^+$ such that $\alpha_0 = \min_{lex}(L)$ is primitive and prefix distinguishable in L. Then, for any $A_1, A_2 \subseteq \Sigma^\star$ commuting with L such that $L \cup \{\epsilon\} \subseteq A_1 \subset A_2$ we have*

$$min_{lex}(A_2L \setminus A_1L) = \alpha_0 min_{lex}(A_2 \setminus A_1) = min_{lex}(A_2 \setminus A_1)\alpha_0.$$

Proof We first observe that $A_2L \setminus A_1L = LA_2 \setminus LA_1 \neq \emptyset$ because Lemma 3.3 states that $A_2 = A_1$ if $LA_2 = LA_1$. So, let $w = \min_{\text{lex}}(A_2L \setminus A_1L)$ and $\tilde{w} = \min_{\text{lex}}(A_2 \setminus A_1)$. Since $A_2L = LA_2$, we have

$$w = w_1\alpha_s = \alpha_t w_2$$

for suitable $w_1, w_2 \in A_2 \setminus A_1$, and $\alpha_s, \alpha_t \in L$ with $\tilde{w} \leq_{\text{lex}} w_1, w_2$. Note that $\alpha_t = \alpha_0$ since otherwise $\alpha_0 w_2 <_{\text{lex}} w$ and then $\alpha_0 w_2$ would belong to LA_1, that is, w_2 would belong to A_1 (by Property (d)).
A similar reasoning shows that $w_2 = \tilde{w}$ and then

$$w_1\alpha_s = \alpha_0\tilde{w}.$$

Since α_0 is primitive, Lemma 3.1 and Theorem 2.1 imply

$$w_0 = \min_{\text{lex}}(\mathcal{C}(L) \setminus \{\epsilon\}) = \alpha_0^k.$$

Now, let $n = |\alpha_0|$ and note that $|w_1| \geq n$ since otherwise $w_1 < \alpha_0$ and then $w_1 <_{\text{lex}} w_0$. Moreover, $w_1 \neq \alpha_0$ because $w_1 \in A_2 \setminus A_1$ and $\alpha_0 \in L \subseteq A_1$. Hence, there exists $u \in \Sigma^+$ such that

$$w_1\alpha_s = \alpha_0 u\alpha_s = \alpha_0\tilde{w},$$

that is, $w_1 = \alpha_0 u$ and $\tilde{w} = u\alpha_s$ with $\tilde{w} \leq_{\text{lex}} w_1$. So, by Lemma 3.2 we obtain $\tilde{w} \leq w_1$ and then there are $e \in \{0, \ldots, n-1\}$ and $j > 0$ such that

$$w_1 = \alpha_0^j \text{pref}_e(\alpha_0), \quad \tilde{w} = \alpha_0^{j-1}\text{pref}_e(\alpha_0)\alpha_s.$$

Note that if $e = 0$, from $\tilde{w} = \alpha_0^{j-1}\alpha_s \leq_{\text{lex}} w_1 = \alpha_0^j$, by Property (a0) we obtain $\alpha_s \leq_{\text{lex}} \alpha_0$. Therefore, in this case we have $\alpha_s = \alpha_0$ and $w_1 = \tilde{w}$. Similarly, if $\alpha_0 = \alpha_s$ we conclude by observing that from $w_1\alpha_0 = \alpha_0\tilde{w}$ we get $|w_1| = |\tilde{w}|$ and then $w_1 = \tilde{w}$ since $\tilde{w} \leq w_1$.

Finally, we show that if $e \neq 0$ and $\alpha_0 <_{\text{lex}} \alpha_s$ we can find a word in A_1 that is smaller than w_0.
Recalling that $\tilde{w} \leq w_1$, from

$$\alpha_0^{j-1}\text{pref}_e(\alpha_0)\alpha_s \leq \alpha_0^j\text{pref}_e(\alpha_0),$$

we get

$$\text{pref}_e(\alpha_0)\alpha_s \leq \alpha_0\text{pref}_e(\alpha_0). \tag{4.1}$$

From $\alpha_0 <_{\text{lex}} \alpha_s$ and Property (a1) we obtain $\text{pref}_e(\alpha_0)\alpha_0 <_{\text{lex}} \text{pref}_e(\alpha_0)\alpha_s$ and then, by (4.1) and Property (b),

$$\text{pref}_e(\alpha_0)\alpha_0 <_{\text{lex}} \alpha_0\text{pref}_e(\alpha_0). \tag{4.2}$$

Now, we consider the word $w_1\alpha_0 = \alpha_0^j\text{pref}_e(\alpha_0)\alpha_0$ and observe that it belongs to A_1L (and to LA_1) since $w_1\alpha_0 <_{\text{lex}} w$. So, by Property (d) there exists $y_1 \in A_1$,

$$y_1 = \alpha_0^{j-1}\text{pref}_e(\alpha_0)\alpha_0,$$

such that $w_1\alpha_0 = \alpha_0 y_1$. Then, we distinguish two cases.

($j-1=0$) By (4.2) we have $y_1 = \mathrm{pref}_e(\alpha_0)\alpha_0 <_{\mathrm{lex}} \alpha_0\mathrm{pref}_e(\alpha_0)$ and so there are $\gamma, x_1, x_2 \in \Sigma^\star$ and $\sigma, \tau \in \Sigma$ with $\sigma < \tau$, such that

$$y_1 = \mathrm{pref}_e(\alpha_0)\alpha_0 = \gamma\sigma x_1, \quad \alpha_0\mathrm{pref}_e(\alpha_0) = \gamma\tau x_2.$$

If $|\gamma| < n$ then we have found a word y_1 such that $y_1 <_{\mathrm{lex}} \alpha_0 \leq_{\mathrm{lex}} w_0$, otherwise we have $y_1 = \alpha_0\mathrm{suf}_e(\alpha_0)$ with

$$\mathrm{suf}_e(\alpha_0) <_{\mathrm{lex}} \mathrm{pref}_e(\alpha_0). \tag{4.3}$$

In this case we consider the word

$$y_1\alpha_0 = \alpha_0\tilde{y}$$

and, by Property (d), we obtain a word $\tilde{y} \in A_1$, $\tilde{y} = \mathrm{suf}_e(\alpha_0)\alpha_0$. Then, (4.3) and Property (b) imply $\tilde{y} <_{\mathrm{lex}} \alpha_0 \leq_{\mathrm{lex}} w_0$.

($j-1>0$) We consider the word $y_1\alpha_0 \in A_1L$. Since $A_1L = LA_1$, we have $y_1\alpha_0 = \alpha_0 y_2$ with $y_2 = \alpha_0^{j-2}\mathrm{pref}_e(\alpha_0)\alpha_0^2 \in A_1$ (by Property (d)). By iterating the process, we eventually get a word $y_j \in A_1$,

$$y_j = \mathrm{pref}_e(\alpha_0)\alpha_0^j.$$

Lastly, as shown in the previous case, we obtain a word $\hat{y} \in A_1$ that satisfies $\hat{y} <_{\mathrm{lex}} \alpha_0 \leq_{\mathrm{lex}} w_0$ by setting either $\hat{y} = y_j$ if ($\alpha_0 \not\leq y_j$) or $\hat{y} = \bar{y}$ (if $y_j = \alpha_0\bar{y}$).

□

It is now easy to prove that if the smallest word of a language L is prefix distinguishable in L and the centralizer of L properly contains $L^\star$, then the smallest word of L is not primitive. More formally, we have the following:

Theorem 4.2 *Let $L \subseteq \Sigma^+$, $\mathcal{U} = \mathcal{C}(L) \setminus L^\star$ and $\alpha_0 = \min_{lex}(L)$. If α_0 is prefix distinguishable in L and $\mathcal{U} \neq \emptyset$ then α_0 is not primitive.*

Proof We trivially have $L^\star \subseteq \mathcal{C}(L)$ and $L^\star L = LL^\star$. So, if $\mathcal{U} \neq \emptyset$ let $\tilde{w} = \min_{\mathrm{lex}}(\mathcal{U})$ and consider the languages (commuting with L) $A_1 = L^\star$, $A_2 = \mathcal{C}(L)$. Since $L \cup \{\epsilon\} \subseteq A_1 \subset A_2$ then, by Theorem 4.1, we have

$$\min_{\mathrm{lex}}(A_2L \setminus A_1L) = \alpha_0\tilde{w} = \tilde{w}\alpha_0.$$

Now, by Theorem 2.1 there are $z \in \Sigma^+$ and two integers a, b such that $\alpha_0 = z^a$ and $\tilde{w} = z^b$. Note that if $a = 1$ then $\tilde{w} = \alpha_0^b \in L^b$ and so $\tilde{w}$ cannot belong to $\mathcal{U}$. Hence $a > 1$ and α_0 is not primitive. □

Finally, we have:

Theorem 4.3 *Let $L \subseteq \Sigma^+$ and $\alpha_0 = \min_{lex}(L)$. If α_0 is primitive and prefix distinguishable in L then $\mathcal{C}(L) = L^\star$.*

Proof Trivial. Suppose $\mathcal{C}(L) \neq L^\star$. Then we have $\mathcal{C}(X) \setminus X^\star \neq \emptyset$ and so Theorem 4.2 states that α_0 is not primitive. □

5 Conclusions and open problems

Given a language L, let $\mathcal{P}(L)$ be true if and only if the smallest word of L (with respect to $\leq_{\mathrm{lex}}$) is primitive and prefix distinguishable in L. We have shown that $\mathcal{P}$ is a sufficient condition that a language L might satisfy in order to admit as centralizer the submonoid $L^\star$. Obviously, $\mathcal{P}$ is not necessary, as we can easily see, for instance, by constructing a three-word code that does not satisfy $\mathcal{P}$ and recalling that the centralizer of every three-word code L is $L^\star$ (see [6]). So, it is an open problem to characterize the class of languages L with $\mathcal{C}(L) = L^\star$.

References

[1] J. M. Autebert, L. Boasson and M. Latteux, *Motifs et bases de langages*, R.A.I.R.O. Infor. Theo. et Appl., vol. 23 n.4, (1989), p. 379–393.

[2] J. A. Brzozowski and E. Leiss, *On equations for regular languages, finite automata, and sequential networks*, Theor. Comp. Sc., 10 (1980), p. 19–35.

[3] C. Choffrut, J. Karhumäki and N. Ollinger, *The commutation of finite sets: a challenging problem*, Theor. Comp. Sc., 273 (2002), p. 69–79.

[4] N. Chomsky and M. P. Schützenberger, *The algebraic theory of context-free languages*, Computer Programming and Formal Systems, P. Braffort and D. Hirschberg eds., North-Holland, Amsterdam, 1963, 118.

[5] J. H. Conway, *Regular Algebra and Finite Machines*, Chapman & Hall, London, 1971.

[6] J. Karhumäki and I. Petre, *Conway's problem for three-word sets*, Theor. Comp. Sc., 289 (2002), p. 705–725.

[7] J. Karhumäki, A. Latteux and I. Petre, *The commutation with codes and ternary sets of words*, Proc. of STACS 2003, LNCS 2607, Springer 2003, p. 74–84.

[8] M. Kunc, *The power of commuting with finite sets of words*, Proc. of STACS 2005, LNCS 3404, Springer 2005, p. 569–580.

[9] R. C. Lyndon and M. P. Schützenberger, *The equation $a^m = b^n c^p$ in a free group*, Michigan Math J. , 9, (1962), p. 289–298.

[10] B. Ratoandromanana, *Codes et motifs*, R.A.I.R.O. Infor. Theo. et Appl., vol. 23, n.4, 1989, p. 425–444.

[11] A. Restivo, *Some decision results for recognizable sets in arbitrary monoids*, Proc. of the fifth colloquium automata, languages and programming, LNCS 62, Springer 1978, p. 363–371.

Parkization of words and its algebraic applications

Jean-Christophe Novelli, Jean-Yves Thibon[*]

Abstract

The notion of parkization of a word, a variant of the classical standardization, allows one to define a Hopf algebra based on parking functions, and to introduce an internal product on it. A subalgebra based on a Catalan set is stable under this operation and contains the descent algebra as a left ideal.

1 Introduction

The notion of *standardization* of a word over an ordered alphabet is of fundamental importance in the study of various sorting algorithms. It is also essential to the understanding of the Robinson-Schensted correspondence and of most its generalizations, and provides a simple explanation of the existence of a Hopf algebra based on permutations.

Recall that the standardized $\mathrm{Std}(w)$ of a word w is the permutation obtained by iteratively scanning w from left to right, and labelling $1, 2, \ldots$ the occurrences of its smallest letter, then numbering the occurrences of the next one, and so on. Alternatively, $\sigma = \mathrm{Std}(w)^{-1}$ can be characterized as the unique permutation of minimal length such that $w\sigma$ is a nondecreasing word. For example, $\mathrm{Std}(bbacab) = 341625$.

This characterizes completely the sequences of transpositions effected by the bubble sort algorithm on w. An elementary observation, which is at the basis of the constructions of [1], is that the noncommutative polynomials

$$\mathbf{G}_\sigma(A) = \sum_{w\in A^*;\mathrm{Std}(w)=\sigma} w \tag{1.1}$$

span a subalgebra of $\mathbb{Z}\langle A\rangle$. Moreover, if A is infinite, this subalgebra admits a natural Hopf algebra structure. This is **FQSym**, the algebra of *Free Quasi-Symmetric Functions.*

In the sequel, we will describe a similar construction, in which permutations are replaced by another class of special words, known as *parking functions.* Before going into details, we need to review some history of the subject.

[*] Institut Gaspard Monge, Université de Marne-la-Vallée, 5 Boulevard Descartes, Champs-sur-Marne, 77454, Marne-la-Vallée cedex 2 (France), `{novelli,jyt}@univ-mlv.fr`

In 1976, Solomon [14] constructed for each finite Coxeter group a remarkable subalgebra of its group algebra, now called its descent algebra.

For the infinite series of Weyl groups, the direct sums of descent algebras can be endowed with some interesting extra structure. This is most particularly the case for symmetric groups (type A), where the direct sum $\Sigma = \bigoplus_{n\geq 0} \Sigma_n$ (Σ_n being the descent algebra of $\mathfrak{S}_n$) builds up a Hopf algebra, isomorphic to **Sym** (noncommutative symmetric functions) and dual to $QSym$ (quasi-symmetric functions).

It has been understood by Reutenauer [12] and Patras [10] that Σ could be interpreted as a subalgebra of the direct sum $\mathfrak{S} = \sum_{n\geq 0} \mathbb{Z}\mathfrak{S}_n$ for the *convolution product* of permutations, which arises when permutations are regarded as graded endomorphisms of a free associative algebra. Indeed, Σ is then just the convolution subalgebra generated by the homogeneous components of the identity map. Further understanding of the situation has been provided by Malvenuto and Reutenauer [8], who gave a complete description of the Hopf algebra structure of $\mathfrak{S}$, and by Poirier-Reutenauer [11], who discovered an interesting subalgebra based on standard Young tableaux.

Finally, the introduction of **FQSym** clarified the picture and brought up a great deal of simplification. Indeed, **FQSym** is an algebra of noncommutative polynomials over some auxiliary set of variables a_i, which is isomorphic to $\mathfrak{S}$, and is mapped onto ordinary quasi-symmetric function $QSym$ when the a_i are specialized to commuting variables x_i, the natural basis $\mathbf{F}_\sigma$ of **FQSym** going to Gessel's fundamental basis F_I. At the level of **FQSym**, the coproduct has a transparent definition (ordered sum of alphabets), and most of its properties become obvious.

There is at least one point, however, on which this construction does not shed much light. It is the original product of the descent algebras Σ_n, which gives rise on **Sym** to a noncommutative analogue of the internal product of symmetric functions (see [7] for the classical case). The introduction of the Hopf structure of Σ = **Sym** was extremely useful, thanks to the so-called *splitting formula* [2, 5], a compatibility property between all operations (internal and external product, coproduct). But the embedding of **Sym** in **FQSym** does not seem to bring new information. In particular, the coproduct dual to the composition of permutations has no nice definition in terms of product of alphabets, and the splitting formula is no more valid in general. Hopf subalgebras in which it remains valid have been studied by Schocker (Lie idempotent algebra, [13]) and by Patras-Reutenauer [11], this last one being maximal with respect to this property.

There are many combinatorial objects which can be regarded, in one way or another, as generalizations of permutations. Among them are *parking functions*, on which a Hopf algebra structure **PQSym**, very similar to that of **FQSym**, can be defined [9]. Actually, **FQSym** is a Hopf subalgebra of **PQSym**.

We shall briefly review the construction of **PQSym**, and then proceed to show that it is possible to define on **PQSym** an internal product, dual to a natural coproduct corresponding to the Cartesian product of ordered alphabets, exactly as in Gessel's construction of the descent algebra [3]. This product is very different from the composition of permutations or endofunctions, and looks actually rather strange. It can be characterized in terms of the fundamental notion of *parkization* of words defined over a totally ordered alphabet in which each element has a successor.

In [9], various Hopf subalgebras of **PQSym** have been introduced. We shall show that the Catalan subalgebra **CQSym** (based on the Catalan family of non-decreasing parking functions, or equivalently, non-crossing partitions) is stable under this new internal product, and contains the descent algebra as a left ideal. Moreover, the splitting formula remains valid for it.

Acknowledgements.- This project has been partially supported by EC's IHRP Programme, grant HPRN-CT-2001-00272, "Algebraic Combinatorics in Europe". The authors would also like to thank the contributors of the MuPAD project, and especially of the combinat part, for providing the development environment for their research.

2 Parking functions and parkization

A *parking function* on $[n] = \{1, 2, \ldots, n\}$ is a word $\mathbf{a} = a_1a_2\cdots a_n$ of length n on $[n]$ whose non-decreasing rearrangement $\mathbf{a}^{\uparrow} = a'_1a'_2\cdots a'_n$ satisfies $a'_i \le i$ for all i. Let PF_n be the set of such words.

One says that $\mathbf{a}$ has a *breakpoint* at b if $|\{\mathbf{a}_i \le b\}| = b$. Then, $\mathbf{a} \in \mathrm{PF}_n$ is said to be *prime* if its only breakpoint is $b = n$. Let $\mathrm{PPF}_n \subset \mathrm{PF}_n$ be the set of prime parking functions on $[n]$.

For a word w on the alphabet $1, 2, \ldots$, denote by $w[k]$ the word obtained by replacing each letter i by $i + k$. If u and v are two words, with u of length k, one defines the *shifted concatenation*

$$u \bullet v = u \cdot (v[k]) \tag{2.1}$$

and the *shifted shuffle*

$$u \Cup v = u ⧢ (v[k])\,. \tag{2.2}$$

The set of permutations is closed under both operations, and the subalgebra spanned by this set is isomorphic to $\mathfrak{S}$ [8] or to **FQSym** [1].

Clearly, the set of all parking functions is also closed under these operations. The prime parking functions exactly are those which do not occur in any non-trivial shifted shuffle of parking functions. These properties allowed us to define a Hopf algebra of parking functions in [9].

This algebra, denoted by **PQSym**, for *Parking Quasi-Symmetric functions*, is spanned as a vector space by elements $\mathbf{F_a}$ ($\mathbf{a} \in \mathrm{PF}$), the product being defined by

$$\mathbf{F}_{\mathbf{a}'}\mathbf{F}_{\mathbf{a}''} := \sum_{\mathbf{a}\in\mathbf{a}'\Cup\mathbf{a}''} \mathbf{F_a}\,. \tag{2.3}$$

For example,

$$\mathbf{F}_{12}\mathbf{F}_{11} = \mathbf{F}_{1233} + \mathbf{F}_{1323} + \mathbf{F}_{1332} + \mathbf{F}_{3123} + \mathbf{F}_{3132} + \mathbf{F}_{3312}\,. \tag{2.4}$$

The coproduct on **PQSym** is a natural extension of that of **FQSym**. Recall (see [1,8]) that if σ is a permutation,

$$\Delta\mathbf{F}_\sigma = \sum_{u\cdot v=\sigma} \mathbf{F}_{\mathrm{Std}(u)} \otimes \mathbf{F}_{\mathrm{Std}(v)}, \tag{2.5}$$

where Std denotes the usual notion of standardization of a word.

For a word w over a totally ordered alphabet in which each element has a successor, we defined in [9] a notion of *parkized word* Park(w), a parking function which reduces to Std(w) when w is a word without repetition.

For $w = w_1 w_2 \cdots w_n$ on $\{1, 2, \ldots\}$, we set

$$d(w) := \min\{i | \#\{w_j \le i\} < i\}\,. \tag{2.6}$$

If $d(w) = n + 1$, then w is a parking function and the algorithm terminates, returning w. Otherwise, let w' be the word obtained by decrementing all the elements of w greater than $d(w)$. Then $\mathrm{Park}(w) := \mathrm{Park}(w')$. Since w' is smaller than w in the lexicographic order, the algorithm terminates and always returns a parking function.

For example, let $w = (3, 5, 1, 1, 11, 8, 8, 2)$. Then $d(w) = 6$ and the word $w' = (3, 5, 1, 1, 10, 7, 7, 2)$. Then $d(w') = 6$ and $w'' = (3, 5, 1, 1, 9, 6, 6, 2)$. Finally, $d(w'') = 8$ and $w''' = (3, 5, 1, 1, 8, 6, 6, 2)$, that is a parking function. Thus, $\mathrm{Park}(w) = (3, 5, 1, 1, 8, 6, 6, 2)$.

The coproduct on **PQSym** is defined by

$$\Delta\mathbf{F}_{\mathbf{a}} := \sum_{u\cdot v=\mathbf{a}} \mathbf{F}_{\mathrm{Park}(u)} \otimes \mathbf{F}_{\mathrm{Park}(v)}, \tag{2.7}$$

For example,

$$\Delta\mathbf{F}_{3132} = 1 \otimes \mathbf{F}_{3132} + \mathbf{F}_1 \otimes \mathbf{F}_{132} + \mathbf{F}_{21} \otimes \mathbf{F}_{21} + \mathbf{F}_{212} \otimes \mathbf{F}_1 + \mathbf{F}_{3132} \otimes 1\,. \tag{2.8}$$

The product and the coproduct of **PQSym** are compatible, so that **PQSym** is a graded bialgebra, connected, hence a Hopf algebra. Let $\mathbf{G}_{\mathbf{a}} = \mathbf{F}_{\mathbf{a}}^* \in \mathbf{PQSym}^*$ be the dual basis of $(\mathbf{F}_{\mathbf{a}})$. If $\langle\, , \rangle$ denotes the duality bracket, the product on $\mathbf{PQSym}^*$ is given by

$$\mathbf{G}_{\mathbf{a}'}\mathbf{G}_{\mathbf{a}''} = \sum_{\mathbf{a}} \langle\, \mathbf{G}_{\mathbf{a}'} \otimes \mathbf{G}_{\mathbf{a}''}, \Delta\mathbf{F}_{\mathbf{a}}\,\rangle\, \mathbf{G}_{\mathbf{a}} = \sum_{\mathbf{a}\in\mathbf{a}'*\mathbf{a}''} \mathbf{G}_{\mathbf{a}}\,, \tag{2.9}$$

where the *convolution* $\mathbf{a}'*\mathbf{a}''$ of two parking functions is defined as

$$\mathbf{a}'*\mathbf{a}'' = \sum_{u,v;\mathbf{a}=u\cdot v,\mathrm{Park}(u)=\mathbf{a}',\mathrm{Park}(v)=\mathbf{a}''} \mathbf{a}\,. \tag{2.10}$$

For example,

$$\begin{aligned}\mathbf{G}_{12}\mathbf{G}_{11} &= \mathbf{G}_{1211} + \mathbf{G}_{1222} + \mathbf{G}_{1233} + \mathbf{G}_{1311} + \mathbf{G}_{1322}\\ &\quad + \mathbf{G}_{1411} + \mathbf{G}_{1422} + \mathbf{G}_{2311} + \mathbf{G}_{2411} + \mathbf{G}_{3411}\,.\end{aligned} \tag{2.11}$$

When restricted to permutations, it coincides with the convolution of [8, 12].

The coproduct of a $\mathbf{G}_\mathbf{a}$ is

$$\Delta\mathbf{G}_\mathbf{a} := \sum_{u,v;\mathbf{a}\in u\Cup v} \mathbf{G}_u \otimes \mathbf{G}_v\,. \tag{2.12}$$

For example,

$$\begin{aligned}\Delta\mathbf{G}_{41252} &= 1\otimes\mathbf{G}_{41252} + \mathbf{G}_1\otimes\mathbf{G}_{3141} + \mathbf{G}_{122}\otimes\mathbf{G}_{12}\\ &\quad + \mathbf{G}_{4122}\otimes\mathbf{G}_1 + \mathbf{G}_{41252}\otimes 1\,.\end{aligned} \tag{2.13}$$

3 Polynomial realization of PQSym*

We shall need the following definitions: given a totally ordered alphabet A, the *evaluation vector* $\mathrm{Ev}(w)$ of a word w is the sequence of number of occurrences of all the elements of A in w. The *packed evaluation vector* $c(w)$ of w is obtained from $\mathrm{Ev}(w)$ by removing all its zeros. The *fully unpacked evaluation vector* $d(w)$ of w is obtained from $c(w)$ by inserting $i-1$ zeros after each entry i of $c(w)$ except the last one. For example, if $w = 3117291781329$, $\mathrm{Ev}(w) = (4,2,2,0,0,0,2,1,2)$, $c(w) = (4,2,2,2,1,2)$, and $d(w) = (4,0,0,0,2,0,2,0,2,0,1,2)$.

The algebra $\mathbf{PQSym}^*$ admits a simple realization in terms of noncommutative polynomials, which is reminescent of the construction of **FQSym**. If A is a totally ordered infinite alphabet, one can define the following polynomial in $\mathbb{K}\langle A\rangle$

$$\mathbf{G}_\mathbf{a}(A) = \sum_{w\in A^*,\mathrm{Park}(w)=\mathbf{a}} w, \tag{3.1}$$

$\mathbf{a}$ being a parking function. Then

Theorem 3.1 *The $\mathbf{G}_\mathbf{a}(A)$ span a subalgebra of $\mathbb{K}\langle A\rangle$, and the product*

$$\mathbf{G}_{\mathbf{a}'}(A)\mathbf{G}_{\mathbf{a}''}(A)$$

is given by Formula (2.9), so that $\phi_A : \mathbf{G}_\mathbf{a} \to \mathbf{G}_\mathbf{a}(A)$ induces an isomorphism of algebras. Moreover, if one denotes by A' and A'' two mutually commuting alphabets isomorphic to A as ordered sets, and $A'\dotplus A''$ the ordered sum, the coproduct is given by

$$\Delta(\mathbf{G})(A) = \mathbf{G}(A'\dotplus A'') \tag{3.2}$$

under the identification $U(A)\otimes V(A) = U(A')V(A'')$.

Proof First, note that the parkization algorithm is compatible with deconcatenation in the following sense: if $u = v\cdot w$ is a word and $\mathrm{Park}(u) = v'\cdot w'$ with v and v' of the same size, then $\mathrm{Park}(v) = \mathrm{Park}(v')$ and $\mathrm{Park}(w) = \mathrm{Park}(w')$. Assume that a word u_1 appears in $\mathbf{G}_{\mathbf{a}'}(A)\mathbf{G}_{\mathbf{a}''}(A)$, and let u_2 be a word with same parkized word as u_1. Then u_2 also appears in this product thanks to the previous remark. Since all words appearing in the product $\mathbf{G}_{\mathbf{a}'}\mathbf{G}_{\mathbf{a}''}$ appear with multiplicity one, this proves that the $\mathbf{G}_{\mathbf{a}}(A)$ span a subalgebra of $\mathbb{K}\langle A\rangle$.

Now, since each word appears in exactly one $\mathbf{G}_{\mathbf{a}}(A)$, there are no non-trivial linear relations between the $\mathbf{G}_{\mathbf{a}}(A)$. Let ϕ be the linear isomorphism sending the basis $\mathbf{G}_{\mathbf{a}}$ to the basis $\mathbf{G}_{\mathbf{a}}(A)$. We will now prove that ϕ is an algebra morphism, that is, that the product of $\mathbf{G}_{\mathbf{a}}$ is the same as the product of $\mathbf{G}_{\mathbf{a}}(A)$. Thanks to Formula (2.10), if $\mathbf{G}_{\mathbf{a}}$ appears in the product $\mathbf{G}_{\mathbf{a}'}\mathbf{G}_{\mathbf{a}''}$, then $\mathbf{G}_{\mathbf{a}}(A)$ appears in $\mathbf{G}_{\mathbf{a}'}(A)\mathbf{G}_{\mathbf{a}''}(A)$. Conversely, any word appearing in $\mathbf{G}_{\mathbf{a}'}(A)\mathbf{G}_{\mathbf{a}''}(A)$ has a parkized word in $\mathbf{a}' * \mathbf{a}''$.

Clearly, Δ as defined by Formula (3.2) is an algebra morphism. Let us show that $\mathbf{G}_{\mathbf{a}}(A'\dot{+}A'')$ is given by Formula (2.12). Indeed, if a parking function $\mathbf{a}$ can be written as $u \Cup v$, then $\mathbf{G}_u(A) \otimes \mathbf{G}_v(A)$ belongs to $\Delta\mathbf{G}_{\mathbf{a}}(A)$ since u can be in $(A')^*$ and v in $(A'')^*$. Conversely, if $\mathbf{G}_u(A) \otimes \mathbf{G}_v(A)$ appears in $\Delta\mathbf{G}_{\mathbf{a}}(A)$, there exist two words u' and v' respectively in $(A')^*$ and $(A'')^*$ with respective parkized u and v. So $\mathbf{a}$ is in $u' \sqcup\!\sqcup v'$, and hence in $u \Cup v$. □

Recall from [9] that the sums

$$\mathbf{P}^{\pi} := \sum_{\mathbf{a};\mathbf{a}^{\uparrow}=\pi} \mathbf{F}_{\mathbf{a}} \qquad (3.3)$$

where $\mathbf{a}^{\uparrow}$ means the non-decreasing reordering and π runs over non-decreasing parking functions, span a cocommutative Hopf subalgebra **CQSym** of **PQSym**.

As with **FQSym**, one can take the commutative image of the $\mathbf{G}_{\mathbf{a}}$, that is, replace the alphabet A by an alphabet X of commuting variables (endowed with an isomorphic ordering). Then, $\mathbf{G}_{\mathbf{a}'}(X) = \mathbf{G}_{\mathbf{a}''}(X)$ iff $\mathbf{a}'$ and $\mathbf{a}''$ have the same non-decreasing reordering π, and both coincide with the generalized quasi-monomial function $\mathcal{M}_{\pi} = (\mathbf{P}^{\pi})^*$ of [9], that is, the natural basis of the commutative Catalan algebra $\mathbf{CQSym}^* = CQSym$.

Actually, $CQSym$ contains $QSym$ as a subalgebra, the quasi-monomial functions being obtained as $M_I = \sum_{c(\pi)=I} \mathcal{M}_{\pi}$.

As a first application of the polynomial realization, we can quantize $CQSym$. Indeed, we can proceed as for the quantization of $QSym$ [15], that is, we map the a_i on q-commuting variables x_i, that is, $x_j x_i = q x_i x_j$ for $i < j$, $\mathbf{G}_{\mathbf{a}'}(X)$ and $\mathbf{G}_{\mathbf{a}''}(X)$ are equal only up to a power of q when $\mathbf{a}'$ and $\mathbf{a}''$ have the same non-decreasing reordering π, and the resulting algebra is not commutative anymore. Deforming the coproduct so as to maintain compatibility with the product, one may obtain a self-dual Hopf algebra isomorphic to the Loday-Ronco algebra of planar binary trees [6], but the natural structure of this q-deformation is rather that of a twisted Hopf algebra (see [4]).

However, our main application will be the definition of an internal product on **PQSym**.

4 The internal product

Let us first recall some standard notations about biwords. Let $x_{ij} = \begin{pmatrix} i \\ j \end{pmatrix}$ be commuting indeterminates, and $a_{ij} = \begin{bmatrix} i \\ j \end{bmatrix}$ be noncommuting ones. We shall denote by $\begin{pmatrix} i_1\, i_2 \cdots i_r \\ j_1\, j_2 \cdots j_r \end{pmatrix}$ the monomial $\begin{pmatrix} i_1 \\ j_1 \end{pmatrix} \begin{pmatrix} i_2 \\ j_2 \end{pmatrix} \cdots \begin{pmatrix} i_r \\ j_r \end{pmatrix}$ and by $\begin{bmatrix} i_1, i_2, \cdots i_r \\ j_1, j_2, \cdots j_r \end{bmatrix}$ the word $\begin{bmatrix} i_1 \\ j_1 \end{bmatrix} \begin{bmatrix} i_2 \\ j_2 \end{bmatrix} \cdots \begin{bmatrix} i_r \\ j_r \end{bmatrix}$. Such expressions will be referred to respectively as *bimonomials* and *biwords*.

Recall that Gessel constructed the descent algebra by extending to $QSym$ the coproduct dual to the internal product of symmetric functions. That is, if X and Y are two totally and isomorphically ordered alphabets of commuting variables, we can identify a tensor product $f \otimes g$ of quasi-symmetric functions with $f(X)g(Y)$. Denoting by XY the Cartesian product $X \times Y$ endowed with the lexicographic order, Gessel defined for $f \in QSym_n$

$$\delta(f) = f(XY) \in QSym_n \otimes QSym_n \,. \tag{4.1}$$

The dual operation on $\mathbf{Sym}_n$ is the internal product $*$, for which $\mathbf{Sym}_n$ is anti-isomorphic to the descent algebra Σ_n.

This construction can be extended to the commutative Catalan algebra $CQSym = \mathbf{CQSym}^*$, and in fact, even to $\mathbf{PQSym}^*$.

Let A' and A'' be two totally and isomorphically ordered alphabets of noncommuting variables, but such that A' and A'' commute with each other. We denote by $A'A''$ the Cartesian product $A' \times A''$ endowed with the lexicographic order. This is a total order in which each element has a successor, so that $G_{\mathbf{a}}(A'A'')$ is a well defined polynomial. Identifying tensor products of words of the same length with words over $A'A''$, we have

$$\mathbf{G}_{\mathbf{a}}(A'A'') = \sum_{\mathrm{Park}(u \otimes v) = \mathbf{a}} u \otimes v \,. \tag{4.2}$$

For example, writing tensor products as biwords, one has

$$\mathbf{G}_{4121}(A'A'') = \sum_{a,b,c,d} \begin{bmatrix} b & a & a & a \\ d & c & c+1 & c \end{bmatrix} \tag{4.3}$$

with $b > a$, or $b = a$ and $d \geq c + 3$. Our main result is the following

Theorem 4.1 *The formula* $\delta(\mathbf{G_a}) = \mathbf{G_a}(A'A'')$ *defines a coassociative coproduct on each homogeneous component* $\mathbf{PQSym}_n^*$. *Actually,*

$$\delta(\mathbf{G_a}) = \sum_{\mathrm{Park}(\mathbf{a}'\otimes\mathbf{a}'')=\mathbf{a}} \mathbf{G}_{\mathbf{a}'} \otimes \mathbf{G}_{\mathbf{a}''}\,, \tag{4.4}$$

where $\mathbf{a}'$ *and* $\mathbf{a}''$ *are parking functions. By duality, the formula*

$$\mathbf{F}_{\mathbf{a}'} * \mathbf{F}_{\mathbf{a}''} = \mathbf{F}_{\mathrm{Park}(\mathbf{a}'\otimes\mathbf{a}'')} \tag{4.5}$$

defines an associative product on each $\mathbf{PQSym}_n$.

Proof The lexicographic product of ordered alphabets being associative, $\delta(\mathbf{G_a}) = \mathbf{G_a}(A'A'')$ is coassociative. If the parkized of $u \otimes v$ is $\mathbf{a}$ then the parkized of $\mathrm{Park}(u) \otimes \mathrm{Park}(v)$ is also $\mathbf{a}$. This implies Formula (4.4). □

Since A is infinite, δ is compatible with the product of $\mathbf{PQSym}^*$.

Example 4.2

$$\begin{aligned}\delta\mathbf{G}_{4121} =& (\mathbf{G}_{2111} + \mathbf{G}_{3111} + \mathbf{G}_{4111}) \otimes (\mathbf{G}_{1232} + \mathbf{G}_{1121} + \mathbf{G}_{2121} + \mathbf{G}_{3121} + \mathbf{G}_{4121})\\ &+ \mathbf{G}_{1111} \otimes \mathbf{G}_{4121}.\end{aligned} \tag{4.6}$$

Example 4.3

$$\mathbf{F}_{211} * \mathbf{F}_{211} = \mathbf{F}_{311}; \qquad \mathbf{F}_{211} * \mathbf{F}_{112} = \mathbf{F}_{312}; \tag{4.7}$$

$$\mathbf{F}_{211} * \mathbf{F}_{121} = \mathbf{F}_{321}; \qquad \mathbf{F}_{112} * \mathbf{F}_{312} = \mathbf{F}_{213}; \tag{4.8}$$

$$\mathbf{F}_{31143231} * \mathbf{F}_{23571713} = \mathbf{F}_{61385451}. \tag{4.9}$$

The main tool for handling internal products of non-commutative symmetric functions is the splitting formula (see [2], Proposition 5.2). It does not hold in **PQSym**, but one can find subalgebras of **PQSym** larger than **Sym** in which it remains true.

5 Subalgebras of $(\mathbf{PQSym}_n, *)$

In [9], we provide two subalgebras of the Hopf algebra **PQSym**, namely the Catalan algebra **CQSym** and the Schröder algebra **SQSym**. It turns out that the former is stable under $*$, whereas the latter is not.

Define the parkized word of a bimonomial as the non-decreasing parking function obtained by parkizing its lexicographically sorted biword. Recall that bimonomials can be encoded as matrices, the entry A_{ij} being the number of bi-letters (ij) in the biword, so that it makes sense to speak of the parkized word of a matrix.

Theorem 5.1 *The homogeneous components* $\mathbf{CQSym}_n$ *of the Catalan algebra are stable under the internal product* $*$*. More precisely, one has*

$$\mathbf{P}^{\pi'} * \mathbf{P}^{\pi''} = \sum_{\pi} \mathbf{P}^{\pi} \tag{5.1}$$

where π *runs over the parkized words of all non-negative integer matrices with row sum* $\mathrm{Ev}(\pi')$ *and column sum* $\mathrm{Ev}(\pi'')$*.*

Proof Let $u \otimes v$ be a biword. Then the parkized of a permutation of $u \otimes v$ is the same permutation of the parkized of $u \otimes v$. So one can regroup all biwords appearing in $\mathbf{P}^{\pi'} * \mathbf{P}^{\pi''}$ into rearrangement classes. By definition of the basis $\mathbf{P}$, each class contributes one term in the product. This term is labelled by the non-decreasing parking function obtained by sorting the parkized word of any biword in the class. Thanks to the definition of a parkized word of a bimonomial, we then obtain the label of any term as the parkized word of the corresponding matrix. □

Example 5.2

$$\mathbf{P}^{1123} * \mathbf{P}^{1111} = \mathbf{P}^{1134}; \qquad \mathbf{P}^{1111} * \mathbf{P}^{1123} = \mathbf{P}^{1123}. \tag{5.2}$$

$$\mathbf{P}^{1123} * \mathbf{P}^{1112} = 2\mathbf{P}^{1134} + \mathbf{P}^{1234}; \qquad \mathbf{P}^{1122} * \mathbf{P}^{1224} = \mathbf{P}^{1134} + \mathbf{P}^{1233} + 2\mathbf{P}^{1234}. \tag{5.3}$$

$$\mathbf{P}^{1123} * \mathbf{P}^{1224} = 2\mathbf{P}^{1134} + 5\mathbf{P}^{1234}. \tag{5.4}$$

The matrices appearing in the last product are

$$\begin{pmatrix} 1 & 1 & . & . \\ . & 1 & . & . \\ . & . & . & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & . & . \\ . & . & . & 1 \\ . & 1 & . & . \end{pmatrix} \begin{pmatrix} 1 & . & . & 1 \\ . & 1 & . & . \\ . & 1 & . & . \end{pmatrix} \begin{pmatrix} . & 2 & . & . \\ 1 & . & . & . \\ . & . & . & 1 \end{pmatrix}$$
$$\begin{pmatrix} . & 2 & . & . \\ . & . & . & 1 \\ 1 & . & . & . \end{pmatrix} \begin{pmatrix} . & 1 & . & 1 \\ 1 & . & . & . \\ . & 1 & . & . \end{pmatrix} \begin{pmatrix} . & 1 & . & 1 \\ . & 1 & . & . \\ 1 & . & . & . \end{pmatrix} \tag{5.5}$$

the fourth and the fifth matrices having 1134 as parkized word whereas the other ones yield 1234.

It is interesting to observe that these algebras are non-unital. Indeed, it follows from Formula (5.1) that

Corollary 5.3 *The element* $\mathbf{J}_n = \mathbf{P}^{(1^n)}$ *is a left unit for* $*$*, but not a right unit.*

The description of $\mathbf{P}^{\pi'} * \mathbf{P}^{\pi''}$ in terms of integer matrices being essentially identical to that of $S^I * S^J$ in **Sym**, the same argument as in [2], proof of Proposition 5.2, shows that the splitting formula remains valid in $\mathbf{CQSym}_n$:

Proposition 5.4 *Let μ_r denote the r-fold product map from* $\mathbf{CQSym}^{\otimes r}$ *to* $\mathbf{CQSym}$*, Δ^r the r-fold coproduct with values in* $\mathbf{CQSym}^{\otimes r}$*, and $*_r$ the internal product of the r-fold tensor product of algebras* $\mathbf{CQSym}^{\otimes r}$*. Then, for $f_1, \ldots, f_r, g \in$* $\mathbf{CQSym}$*,*

$$(f_1 \cdots f_r) * g = \mu_r[(f_1 \otimes \cdots \otimes f_r) *_r \Delta^r(g)] \,. \tag{5.6}$$

This is indeed the same formula as with the internal product of **Sym**, actually, an extension of it, since we have

Corollary 5.5 *The Hopf subalgebra of* **CQSym** *generated by the elements $\mathbf{J}_n$, which is isomorphic to* **Sym** *by $j : S_n \mapsto \mathbf{J}_n$, is stable under $*$, and thus also $*$-isomophic to* **Sym***. Moreover, the map $f \mapsto f * \mathbf{J}_n$ is a projector onto $\mathbf{Sym}_n$, which is therefore a left $*$-ideal of $\mathbf{CQSym}_n$.*

More precisely, if $i < j < \ldots < r$ are the letters occuring in π, so that as a word $\pi = i^{m_i} j^{m_j} \cdots r^{m_r}$, then

$$\mathbf{P}^{\pi} * \mathbf{J}_n = \mathbf{J}_{m_i} \mathbf{J}_{m_j} \cdots \mathbf{J}_{m_r} \,. \tag{5.7}$$

In the classical case, the non-commutative complete fonctions split into a sum of ribbon Schur functions, using a simple order on compositions. To get an analogous construction in our case, we have defined a partial order on non-decreasing parking functions.

Let π be a non-decreasing parking function and $\mathrm{Ev}(\pi)$ be its evaluation vector. The successors of π are the non-decreasing parking functions whose evaluations are given by the following algorithm: given two non-zero elements of $\mathrm{Ev}(\pi)$ with only zeros between them, replace the left one by the sum of both and the right one by 0. For example, the successors of 113346 are 111146, 113336, and 113344.

By transitive closure, the successor map gives rise to a partial order on non-decreasing parking functions. We will write $\pi \preceq \pi'$ if π' is obtained from π by successive applications of successor maps.

The Catalan ribbon functions [9] are defined by

$$\mathbf{P}^{\pi} =: \sum_{\pi' \succeq \pi} \mathbf{R}_{\pi'} \,. \tag{5.8}$$

It follows from Theorem 5.1 that the $\mathbf{R}_{\pi}$ are the pre-images of the ordinary ribbons under the projection $f \mapsto f * \mathbf{J}_n$:

Corollary 5.6 *Let I be the composition obtained by discarding the zeros of the evaluation of an non-decreasing parking function π. Then*

$$\mathbf{R}_{\pi} * \mathbf{J}_n = j(R_I). \tag{5.9}$$

More precisely, if $I = (i_1, \ldots, i_p)$, this last element is equal to $\mathbf{R}_{1^{i_1} \bullet 1^{i_2} \bullet \cdots \bullet 1^{i_p}}$, that is, the Catalan ribbon indexed by the only non-decreasing word of evaluation $d(\pi)$.

The internal product of **CQSym** is dual to the coproduct $\delta f = f(XY)$ on the commutative algebra $CQSym$, quotient of $\mathbf{PQSym}^*$. For example, we have

$$\begin{aligned}\mathcal{M}_{113}(XY) \;\; = \;\; & (\mathcal{M}_{112}(X)+\mathcal{M}_{113}(X))(\mathcal{M}_{111}(Y)+\mathcal{M}_{112}(Y) \\ & +\mathcal{M}_{113}(Y)+\mathcal{M}_{122}(Y))+\mathcal{M}_{111}(X)\mathcal{M}_{113}(Y). \qquad (5.10)\end{aligned}$$

$$\mathcal{M}_{112}(XY) = +\mathcal{M}_{111}(X)\mathcal{M}_{112}(Y). \qquad (5.11)$$

References

[1] G. Duchamp, F. Hivert, and J.-Y. Thibon, *Noncommutative symmetric functions VI: free quasi-symmetric functions and related algebras*, Internat. J. Alg. Comput. **12** (2002), 671–717.

[2] I. M. Gelfand, D. Krob, A. Lascoux, B. Leclerc, V. S. Retakh, and J.-Y. Thibon, *Noncommutative symmetric functions*, Adv. in Math. **112** (1995), 218–348.

[3] I. Gessel, *Multipartite P-partitions and inner product of skew Schur functions*, Contemp. Math. **34** (1984), 289–301.

[4] F. Hivert, J.-C. Novelli, J.-Y. Thibon, *Commutative Hopf algebras of permutations and trees*, preprint math.CO/0502456.

[5] D. Krob, B. Leclerc and J.-Y. Thibon, *Noncommutative symmetric functions II: Transformations of alphabets*, Internal J. Alg. Comput. **7** (1997), 181–264.

[6] J.-L. Loday and M. Ronco, *Hopf algebra of the planar binary trees*, Adv. in Math. **139** (1998), 293–309.

[7] I. G. Macdonald, *Symmetric functions and Hall polynomials*, 2nd ed., Oxford University Press, 1995.

[8] C. Malvenuto and C. Reutenauer, *Duality between quasi-symmetric functions and the Solomon descent algebra*, J. Algebra **177** (1995), 967–982.

[9] J.-C Novelli and J.-Y. Thibon, *A Hopf algebra of parking functions*, Proc. FPSAC/SFCA 2004, Vancouver (electronic).

[10] F. Patras, *L'algèbre des descentes d'une bigèbre graduée*, J. Algebra **170** (1994), 547–566.

[11] F. Patras and C. Reutenauer, *Lie representations and an algebra containing Solomon's*, J. Algebraic Combin. **16** (2002), 301–314.

[12] C. Reutenauer, *Free Lie algebras*, Oxford University Press, 1993.

[13] M. Schocker, *Lie idempotent algebras*, Adv. Math. **175** (2003), 243–270.

[14] L. Solomon, *A Mackey formula in the group ring of a Coxeter group*, J. Algebra **41** (1976), 255–268.

[15] J.-Y. Thibon and B.C.V. Ung, *Quantum quasi-symmetric functions and Hecke algebras*, J. Phys. A: Math. Gen. **29** (1996), 7337–7348.

Letter frequency in infinite repetition-free words[*]

Pascal Ochem[†]

Abstract

We estimate the extremal letter frequency in infinite words over a finite alphabet avoiding some repetitions. For ternary square-free word, we improve the bounds of Tarannikov on the minimal letter frequency, and prove that the maximal letter frequency is $\frac{255}{653}$. Kolpakov et al. have studied the function ρ such that $\rho(x)$ is the minimal letter frequency in an infinite binary x-free word. In particular, they have shown that ρ is discontinuous at $\frac{7}{3}$ and at every integer at least 3. We answer one their question by providing some other points of discontinuity for ρ. Finally, we propose stronger versions of Dejean's conjecture on repetition threshold in which unequal letter frequencies are required.

1 Introduction

In this paper, we study the extremal frequencies of a letter in factorial languages defined by an alphabet size and a set of forbidden repetitions. Given such a language, we denote by f^- (resp. f^+) the minimal (resp. maximal) letter frequency in an infinite word that belong to the language L. Letter frequencies have been mainly studied in [6, 12, 13]. Let Σ_i denote the i-letter alphabet $\{0, 1, \ldots, i-1\}$. We consider here the frequency of the letter 0. Let $n(w)$ denote the number of occurrences of 0 in the finite word w. So the letter frequency in w is $\frac{n(w)}{|w|}$. A negative result is either a lower bound on f^- or an upper bound on f^+. Notice that for binary words, we only need to consider f^- since $f^- + f^+ = 1$. We denote by $\rho(x)$ (resp. $\rho(x^+)$) the minimal letter frequency in an infinite x-free (resp. (x^+)-free) binary word. Our results are stated in Section 2. The proof technique for negative results is an improved version of the methods given in [7,13] to find lower bounds on the minimal frequency of square occurrences in an infinite binary word. It is detailled in Section 3. Positive results consist in uniform morphisms that can produce infinite words in L with a given letter frequency. The method used to find such morphisms is explained in Section 4. In Section 5, we make a conjecture related to Dejean's conjecture [3] involving unequal letter frequencies. The C++ sources of the programs and the

[*] Research supported in part by COMBSTRU.

[†] LaBRI, Université Bordeaux I, 351, cours de la Libération, 33405 Talence Cedex (France), ochem@labri.fr

morphisms used in this paper are available at: `http://dept-info.labri.fr/~ochem/morphisms/`.

2 Statement of main results

For ternary square-free words, Tarannikov [13] showed that $f^- \in \left[\frac{1780}{6481}, \frac{64}{233}\right] = [0.27464897\ldots, 0.27467811\ldots]$. According to [12], he also proved that $f^+ \leq \frac{469}{1201} = 0.39050791\ldots$. We obtain the following results:

Theorem 2.1 *For ternary square-free words, we have*

1. $f^- \in \left[\frac{1000}{3641}, \frac{883}{3215}\right] = [0.27464982\ldots, 0.27465007\ldots]$.
2. $f^+ = \frac{255}{653} = 0.39050535\ldots$.

A (β, n)-repetition is a repetition with prefix size n and exponent β. The notions of (β, n)-freeness and (β^+, n)-freeness are introduced in [4]. A word is said to be (β, n)-free (resp. (β^+, n)-free) if it contains no (β', n')-repetition such that $n' \geq n$ and $\beta' \geq \beta$ (resp. $\beta' > \beta$). we construct in [4] an infinite $\left(\frac{8}{5}^+, 3\right)$-free binary word.

Theorem 2.2 *For* $(\frac{5}{3}, 3)$*-free binary words, we have* $f^- = \frac{1}{2}$.

Theorem 2.2 implies that infinite $(\beta, 3)$-free binary words have equal letter frequency for $\beta \in \left[\frac{8}{5}^+, \frac{5}{3}\right]$. A similar result in [6] says that infinite $(\beta, 1)$-free binary words have equal letter frequency for $\beta \in \left[2^+, \frac{7}{3}\right]$, i.e. $\rho(2^+) = \rho\left(\frac{7}{3}\right) = \frac{1}{2}$. It is noticeable that these two cases of equal letter frequency have different kind of growth function. Karhumäki and Shallit have shown that the growth function of $\frac{7}{3}$-free binary words is polynomial [5], whereas the growth function of $(\frac{8}{5}^+, 3)$-free binary words is exponential. To see this, notice that the 992-uniform morphism $h : \Sigma_4^* \rightarrow \Sigma_2^*$ given in [4] produces a $\left(\frac{8}{5}^+, 3\right)$-free binary word $h(w)$ for every $\frac{7}{5}^+$-free word $w \in \Sigma_4^*$, and that an exponential lower bound on the number of 4-ary $\frac{7}{5}^+$-free words is shown in [10].

Kolpakov et al. [6] proved that the function ρ is discontinuous at $\frac{7}{3}$, more precisely they obtained that $\rho\left(\frac{7}{3}\right) = \frac{1}{2}$ and $\rho\left(\frac{7}{3}^+\right) \leq \frac{10}{21} = 0.47619047\ldots$.

The next result provides new points of discontinuity for ρ in the range $\left[\frac{7}{3}^+, 3\right]$, namely $\frac{17}{7}$, $\frac{5}{2}$, $\frac{23}{9}$, $\frac{16}{41}$, $\frac{18}{7}$, and $\frac{8}{3}$.

Theorem 2.3

1. $\rho\left(\frac{7}{3}^+\right) \leq \frac{47}{101} = 0.46534653\ldots$.
2. $\rho\left(\frac{17}{7}\right) \geq \frac{467}{1004} = 0.46513944\ldots$.

3. $\rho\left(\frac{17}{7}^{+}\right) \leq \frac{81}{175} = 0.46285714\ldots.$

4. $\rho\left(\frac{5}{2}\right) \geq \frac{54286}{117293} = 0.46282386\ldots.$

5. $\rho\left(\frac{5}{2}^{+}\right) \leq \frac{23}{52} = 0.44230769\ldots.$

6. $\rho\left(\frac{23}{9}\right) > \frac{205}{464} = 0.44181034\ldots.$

7. $\rho\left(\frac{23}{9}^{+}\right) \leq \frac{91}{206} = 0.44174757\ldots.$

8. $\rho\left(\frac{16}{41}\right) > \frac{322}{729} = 0.44170096\ldots.$

9. $\rho\left(\frac{16}{41}^{+}\right) \leq \frac{143}{324} = 0.44135802\ldots.$

10. $\rho\left(\frac{18}{7}\right) \geq \frac{79}{179} = 0.44134078\ldots.$

11. $\rho\left(\frac{18}{7}^{+}\right) \leq \frac{41}{93} = 0.44086021\ldots.$

12. $\rho\left(\frac{8}{3}\right) > \frac{339}{769} = 0.44083224\ldots.$

13. $\rho\left(\frac{8}{3}^{+}\right) \leq \frac{24}{59} = 0.40677966\ldots.$

14. $\rho\left(3\right) > \frac{115}{283} = 0.40636042\ldots.$

3 Method for negative results

Let L be a factorial language. A word w is said to be k-biprolongable in L if there exists a word $lwr \in L$ such that $|l| = |r| = k$. A *suffix cover* of L is a set S of finite words in L such that every finite k-biprolongable word in L has a suffix that belongs to S, for some finite number k. Taking $k = 20$ is sufficient for every negative result in this paper. For a word $u \in S$, let

$$A_u(q) = \left\{ w \in L \mid uw \in L \text{ and for every prefix } w[1..k] \text{ of } w,\ \frac{n(w[1..k])}{k} < q \right\}.$$

Lemma 3.1 *Let L be a factorial language and S one of its suffix cover. Let $q \in \mathbb{Q}$. If $A_u(q)$ is finite for every word $u \in S$, then $f^- \geq q$.*

Proof Assume $A_u(q)$ is finite for every word $u \in S$. Then any infinite word $w \in L$ has a decomposition into finite factors $w_0w_1w_2\ldots$ such that $|w_0| = k + \max_{u \in S}|u|$ and $\frac{n(w_i)}{|w_i|} \geq q$ for every $i \geq 1$. □

Lemma 3.1 enables us to obtain bounds of the form $f^- \geq q$ by choosing an explicit suffix cover and checking by computer that every set $A_u(q)$ is finite. It is easy to see that Lemma 3.1 and the definition of $A_u(q)$ can be modified to provide bounds of the form $f^- > q$, $f^+ \leq q$, or $f^+ < q$. This method is a natural generalization of the one in [13], where the suffix cover consist in the empty word, and of the one in [7], where the suffix cover consist in all binary words of length three. Since we study here the frequency of the letter 0 in repetition-free words, every letter other than 0 play the same role. Let us say that two words u and u' in Σ_s are equivalent if and only if u can be obtained from u' by a permutation of the letters in $\Sigma_s \setminus \{0\}$. Notice that for two equivalent words u and u', $A_{u'}(q)$ is finite if and only if $A_u(q)$ is finite. We define the reduced suffix cover of a suffix cover S as the quotient of S by this equivalence relation. To prove the negative part of Theorem 2.1.1 we used the reduced suffix cover $\{1, 01210, 0210, 2010\}$, the computation took about 20 days on a XEON 2.2Gh. For Theorem 2.1.2 we used the reduced suffix cover $\{0, 01, 021, 0121\}$. For Theorem 2.2 we used the suffix cover $\{01, 111, 000, 1110, 1010, 0001111000010,\ 0111101000010,\ 1110101000010,\ 0111100010\}$. We omit the computer proof that this is indeed a suffix cover for $(\frac{5}{3}, 3)$-free binary words. The negative statements of Theorem 2.3 (even items) were obtained using the suffix cover $\{1, 10, 100\}$.

4 Method for positive results

Let L be a factorial language. To construct an infinite word $w \in L$ with a given letter frequency $q \in \mathbb{Q}$, we basically use the method described in [10]. We note $q = \frac{a}{b}$ with a coprime to b. For increasing values of k, we look for a $(k \times b)$-uniform morphism $h : \Sigma_e^* \to \Sigma_s^*$ producing (infinite) words in L such that $n(h(i)) = k \times a$ for every $i \in \Sigma_e$.

Consider the 8-uniform morphism $m : \Sigma_3^* \longrightarrow \Sigma_4^*$ defined by

$$m(0) = 01232103,$$
$$m(1) = 01230323,$$
$$m(2) = 01210321.$$

To get the bound $f^- \leq \frac{883}{3215}$ in Theorem 2.1, we found a square-free morphism $h^+ : \Sigma_3^* \longrightarrow \Sigma_3^*$ such that $h^+ = m \circ m^+$ where $m^+ : \Sigma_4^* \longrightarrow \Sigma_3^*$ is a 3215-uniform morphism. To get the bound $f^+ \geq \frac{255}{653}$ in Theorem 2.1, we found a square-free morphism $h^- : \Sigma_3^* \longrightarrow \Sigma_3^*$ such that $h^- = m \circ m^-$ where $m^- : \Sigma_4^* \longrightarrow \Sigma_3^*$ is a 9142-uniform morphism ($9142 = 14 \times 653$). We need a result of Crochemore [2] saying that a uniform morphism is square-free if the image of every square-free word of length 3 is square-free. The software **mreps** [9] written by Kucherov et al. can test if a word is square-free in linear time. We used it to prove that h^- and h^+ are square-free by checking that $h^-(w)$ and $h^+(w)$ are square-free, where $w = 010201210120212$ is square-free and contains every ternary square-free words of length 3 as factors. Checking the image of w is faster than checking the images of the 12 ternary square-free words of length 3 because **mreps** runs in linear time. Since the morphisms h^- (resp. h^+) is square-free, we obtain an exponential lower bound for ternary square-free words with letter frequency $\frac{883}{3215}$

(resp. $\frac{255}{653}$), which is interesting from the point of view of [12].

For each positive statement in Theorem 2.3 (odd items), we provide a uniform morphism $h : \Sigma_3^* \longrightarrow \Sigma_2^*$ such that for every $\left(\frac{7}{4}^+\right)$-free ternary word w, $h(w)$ has the corresponding properties of repetition-freeness and letter frequency.

5 Dejean's conjecture and letter frequencies

The *repetition threshold* is the least exponent $\alpha = \alpha(k)$ such that there exist an infinite (α^+)-free word over Σ_k. Dejean proved that $\alpha(3) = \frac{7}{4}$. She also conjectured that $\alpha(4) = \frac{7}{5}$ and $\alpha(k) = \frac{k}{k-1}$ for $k \geq 5$. In its full generality, this conjecture is still open, although Pansiot [11] proved that $\alpha(4) = \frac{7}{5}$ and Moulin-Ollagnier [8] proved that Dejean's conjecture holds for $5 \leq k \leq 11$. For more information, see [1]. Based on numerical evidences, we propose the following conjecture which implies Dejean's conjecture.

Conjecture 5.1

1. *For every $k \geq 5$, there exists an infinite $\left(\frac{k}{k-1}^+\right)$-free word over Σ_n with letter frequency $\frac{1}{k+1}$.*

2. *For every $k \geq 6$, there exists an infinite $\left(\frac{k}{k-1}^+\right)$-free word over Σ_n with letter frequency $\frac{1}{k-1}$.*

It is easy to see that the values $\frac{1}{k+1}$ and $\frac{1}{k-1}$ in Conjecture 5.1 would be best possible. For $\left(\frac{5}{4}^+\right)$-free words over Σ_5, we obtain $f^+ < \frac{103}{440} = 0.23409090\cdots < \frac{1}{4}$ using the reduced suffix cover $\{0, 01, 012, 0123, 012341, 401234, 4301234\}$. That is why Conjecture 5.1.2 is stated with $k \geq 6$.

Acknowledgements

I thank Julien Cassaigne for a fruitful discussion about the case of equal letter frequency for $\left(\frac{5}{3}, 3\right)$-free binary words.

References

[1] C. Choffrut and J. Karhumäki, Combinatorics of words. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, Vol. 1, pp. 329–438, Springer-Verlag, 1997.

[2] M. Crochemore, Sharp characterizations of squarefree morphisms, *Theoret. Comput. Sci.* **18** (1982), 221–226.

[3] F. Dejean, Sur un théorème de Thue, *J. Combin. Theory. Ser. A* **13** (1972), 90–99.

[4] L. Ilie, P. Ochem, and J.O. Shallit, A generalization of Repetition Threshold, *Theoret. Comput. Sci.* To appear.

[5] J. Karhumäki and J.O. Shallit, Polynomial versus exponential growth in repetition-free binary words, *J. Combin. Theory. Ser. A* **105(2)** (2004), 335–347.

[6] R. Kolpakov, G. Kucherov, and Y. Tarannikov, On repetition-free binary words of minimal density, *Theoret. Comput. Sci.* **218** (1999), 161–175.

[7] G. Kucherov, P. Ochem, and M. Rao How many square occurrences must a binary sequence contain ? *Electron. J. Comb.* **10(1)** (2003), #R12.

[8] J. Moulin-Ollagnier, Proof of Dejean's conjecture for alphabets with $5, 6, 7, 8, 9, 10$ and 11 letters, *Theoret. Comput. Sci.* **95** (1992), 187–205.

[9] `http://mreps.loria.fr/`

[10] P. Ochem, A generator of morphisms for infinite words. In *Proceedings of the Workshop on Word Avoidability, Complexity, and Morphisms*, Turku, Finland, July 17 2004, LaRIA Technical Report 2004-07, pp. 9–14.

[11] J.-J. Pansiot, A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, *Disc. Appl. Math.* **7** (1984), 297–311.

[12] C. Richard and U. Grimm, On the entropy and letter frequencies of ternary square-free words, *Electron. J. Comb.* **11** (2004), #R14

[13] Y. Tarannikov, The minimal density of a letter in an infinite ternary square-free word is 0.2746..., *J. Integer Sequences* 5(2):Article 02.2.2 (2002).

Conjugacy of morphisms and Lyndon decomposition of standard Sturmian words

Gwénäel Richomme[*]

Abstract

Using the notions of conjugacy of morphisms, we answer a question of G. Melançon concerning the decomposition in Lyndon words of standard Sturmian words. We show some connections with morphisms preserving Lyndon words

1 Introduction

Finite (or infinite) Lyndon words can be encountered in many studies (see for instance [8–10]). They are the nonempty words which are smaller in lexicographic order than all their proper suffixes. The Lyndon factorization theorem [4] states that any finite word can be decomposed uniquely in a product of nonincreasing (in lexicographic order) Lyndon words. This result was extended to infinite words [19] (In such a case, the decomposition can end with an infinite Lyndon word). Thus some works concern the decomposition in Lyndon words of some infinite words (see for instance [3, 5, 11, 12, 18] for such results).

In [12], G. Melançon gives a decomposition in Lyndon words of standard Sturmian words. He asks the following question: in which cases, the sequence of nonincreasing Lyndon words appearing in the decomposition of a standard Sturmian word can be written $(g^n(l_0))_{n\geq 0}$ with l_0 a Lyndon word and g a morphism. In Section 5, we answer this question.

For this, we use results about morphisms preserving Lyndon words [14] and about conjugacy of morphisms [13]. In particular, we show that when a positive answer exists to the previous question, g preserves Lyndon words and is the conjugate of a morphism f that generates the decomposed standard Sturmian word.

In Section 2, we recall notions on Sturmian words and morphisms. Section 3 recalls both the decomposition in Lyndon words of standard Sturmian words obtained by G. Melançon, and his question. This section also recall notions on morphisms preserving Lyndon words. Section 4 presents notions on conjugacy of morphisms and introduces a new particular case, namely the strong conjugacy. Using it in conjunction with morphisms preserving Lyndon words, we give a new

[*] LaRIA, Université de Picardie Jules Verne, 33, Rue Saint Leu, F-80039 Amiens cedex 1 (France), gwenael.richomme@u-picardie.fr

proof that for any standard Sturmian words w over $\{a < b\}$, aw is an infinite Lyndon word [3]. Finally, in Section 5, we answer G. Melançon. Note that at a first step, we express the decomposition of a standard Sturmian word using only morphisms.

2 Sturmian words and morphisms

We recall here notions on words (see for instance [8, 9] for more details).

An *alphabet* A is a set of symbols called *letters*. Here we consider only finite alphabets. A *word over* A is a sequence of letters from A. The *empty word* ε is the empty sequence. Equipped with the concatenation operation, the set A^* of finite words over A is a free monoid with neutral element ε and set of generators A. We denote by A^ω the set of infinite words over A. As usually, for a finite word u and an integer n, the n^{th} power of u, denoted u^n, is the word ε if $n = 0$ and the word $u^{n-1}u$ otherwise. If u is not the empty word, u^ω denotes the infinite word obtained by infinitely repeating u. A finite word w is said *primitive* if for any word u, the equality $w = u^n$ (with n an integer) implies $n = 1$. Any word is the power of a unique primitive word called the *primitive root* of w.

Given a nonempty word $u = u_1 \dots u_n$ with $u_i \in A$, the *length* $|u|$ of u is the integer n. One has $|\varepsilon| = 0$. If for some words u, v, p, s (possibly empty), $u = pvs$, then v is a *factor* of u, p is a *prefix* of u and s is a *suffix* of u. When $p \neq u$ (resp. $s \neq u$), we say that p is a *proper prefix* (resp. s is a *proper suffix*) of u. By $|u|_a$ we denote the number of occurrences of the letter a in the word u.

Sturmian words may be defined in many equivalent ways (see [1] for instance). They are infinite binary words. Here we will consider them as the infinite balanced non ultimately periodic words. We recall that a (finite or infinite) word w over $\{a, b\}$ is *balanced* if for any factors u and v of same length $||u|_a - |v|_a| \leq 1$, and that an infinite word w is *ultimately periodic* if $w = uv^\omega$ for some finite words u and v.

Many studies of Sturmian words use Sturmian morphisms. Let A, B be two alphabets. A *morphism* (*endomorphism* if $A = B$) f from A^* to B^* is a mapping from A^* to B^* such that for all words u, v over A, $f(uv) = f(u)f(v)$. We also say that f is a morphism on A or that f is defined on A (without any other precision when B has no importance). A morphism on A is entirely known by the images of the letters of A. A morphism extends naturally on infinite words. We denote just by juxtaposition the composition of morphisms. Given an endomorphism f, if $\lim_{n\to\infty} f^n(a)$ exists, then this limit is denoted $f^\omega(a)$ and is a fixed point of f: the word $f^\omega(a)$ is said generated by f.

Sturmian morphisms are the morphisms in $\{E, L_a, L_b, R_a, R_b\}^*$ where E, L_a, L_b, R_a, R_b are the endomorphisms defined on $\{a, b\}$ by $E(a) = b$, $E(b) = a$, $L_a(a) = a$, $L_a(b) = ab$, $L_b(a) = ba$, $L_b(b) = b$, $R_a(a) = a$, $R_a(b) = ba$, $R_b(a) = ab$, $R_b(b) = b$. Many relations exists between Sturmian words and Sturmian morphisms. For instance, it is known [2, 6] that any Sturmian word can be defined as an infinite product of Sturmian morphisms.

A particular case of Sturmian words is the standard (or characteristic) one. For any standard Sturmian words, there exists a sequence $(a_n)_{n\geq 0}$ of integers, called *the directive sequence* verifying $a_1 \geq 0$ and $a_k \geq 1$ for all $k \geq 2$, such that

$$w = \lim_{n\to\infty} s_n$$

where the sequence $(s_n)_{n\geq -1}$ of words is defined by : $s_{-1} = b$, $s_0 = a$ and $s_n = s_{n-1}^{a_n} s_{n-2}$ for $n \geq 1$. Let us observe that for every $n \geq 0$, s_{2n} ends with a. Moreover [1],

$$\begin{aligned}
s_{2n} &= L_a^{a_1} L_b^{a_2} \dots L_a^{a_{2n-1}} L_b^{a_{2n}}(a)\\
&= L_a^{a_1} L_b^{a_2} \dots L_a^{a_{2n-1}} L_b^{a_{2n}} L_a^{a_{2n+1}}(a)\\
s_{2n+1} &= L_a^{a_1} L_b^{a_2} \dots L_a^{a_{2n-1}} L_b^{a_{2n}} L_a^{a_{2n+1}}(b)\\
&= L_a^{a_1} L_b^{a_2} \dots L_a^{a_{2n-1}} L_b^{a_{2n}} L_a^{a_{2n+1}} L_b^{a_{2n+2}}(b)
\end{aligned}$$

3 Lyndon words and morphisms

From now on we consider ordered alphabets. We denote $\{\alpha_1 < \dots < \alpha_n\}$ the n-letter alphabet $\{\alpha_1, \dots, \alpha_n\}$ with order $\alpha_1 < \dots < \alpha_n$. Given an ordered alphabet A, we denote by $\preceq$ the lexicographic order whenever used on A^* or on A^ω. Let recall that for two different (finite or infinite) words u and v, $u \prec v$ if and only if $u = xay$, $v = xbz$ with $a, b \in A$, $a < b$, $x \in A^*$, $y, z \in A^* \cup A^\omega$, or if (when u is finite) u is a prefix of v.

A nonempty finite word w is a *Lyndon word* if for all nonempty words u and v, $w = uv$ implies $w \prec vu$. Equivalently [4, 8], a nonempty word w is a Lyndon word if all its nonempty proper suffixes are greater than it for the lexicographic order. For instance, on the one-letter alphabet $\{a\}$, only a is a Lyndon word. On $\{a < b\}$ the Lyndon words of length at most 5 are a, b, ab, aab, abb, $aaab$, $aabb$, $abbb$, $aaaab$, $aaabb$, $aabab$, $aabbb$, $abbbb$. Lyndon words are primitive.

The second definition of Lyndon words extends to infinite words: An infinite word is an *infinite Lyndon word* if all its proper suffixes are greater than it for the lexicographic order. A useful result of G. Melançon [12] states that an infinite word is a Lyndon word if and only if it has an infinity of prefixes that are Lyndon words. See for instance [7] for a recent example of infinite Lyndon word.

Any nonempty finite or infinite Lyndon words can be decomposed as a nonincreasing product of Lyndon words. First, R. C. Lyndon proved (see [8] for instance):

> *Any word $w \in A^+$ may be written uniquely as a nonincreasing product of Lyndon words: $w = l_1 l_2 \dots l_n$ where for each i, l_i is a Lyndon word and $l_n \preceq l_{n-1} \preceq \dots l_1$.*

This result was generalized to infinite words [19]:

> *Any right infinite word w may be uniquely expressed as a nonincreasing product of Lyndon words, finite or infinite, in one of the two following forms: either there exists an infinite nonincreasing sequence of finite Lyndon words $(l_k)_{k\geq 0}$ such that*
>
> $$w = \prod_{n\geq 0} l_n = l_0 l_1 \ldots$$
>
> *or there exist finite Lyndon words $l_0, \ldots, l_{m-1}$ $(m \geq 0)$ and an infinite word l_m such that $l_m \prec l_{m-1} \preceq l_{m-2} \preceq \ldots l_0$ and*
>
> $$w = l_0 \ldots l_{m-1} l_m.$$

As already said in the introduction, many works concern the decomposition in Lyndon words of some infinite words. In [12], G. Melançon obtains the decomposition of standard Sturmian words. We consider these words here on the alphabet $\{a < b\}$. For any word w ending with the letter a, let us denote $\overline{w}$ the word such that $w = \overline{w}a$.

Theorem 3.1 [12] *Let s be a standard Sturmian word with directive sequence $(a_n)_{n\geq 1}$. Let $l_n = as_{2n}^{a_{2n+1}-1} s_{2n-1}\overline{s_{2n}}$. (if $a_1 = 0$ then $l_0 = b$).*

The words $(l_n)_{n\geq 0}$ form a strictly decreasing sequence of Lyndon words and the unique factorization of s as a nonincreasing product of Lyndon words is

$$s = \prod_{n\geq 0} l_n^{a_{2n+1}}.$$

G. Melançon wrote [12, Remark 3.7] :

> *When is the sequence $(l_n)_{n\geq 0}$ morphic ? More precisely, is it possible to give a morphism φ : $\{a, b\}^* \to \{a, b\}^*$ and a Lyndon word l_0 such that $l_{n+1} = \varphi(l_n)$? This question has a positive answer in the case where the directive sequence is constant. For instance, if $a_n = 2$ for all $n \geq 0$, then we may set $l_0 = aab$ and use the morphism mapping $a \mapsto aaabaab$ and $b \mapsto aab$.*
>
> *A characteristic Sturmian word may be itself morphic. That is, may be the limit $\lim_n \varphi^n(a)$ of a (nonerasing) morphism (satisfying $\varphi(a) \in aA^*$). It is known that this is essentially equivalent to the fact that its directive sequence is periodic. Unfortunately, even when a characteristic Sturmian word s has a periodic directive sequence, it seems that the sequence $(l_n)_{n\geq 0}$ is not always morphic, although it is possible to describe patterns in the factorization.*

The aim of this paper is to answer this question. The main ideas of our proof are generalizations of the following remarks: the morphism $a \mapsto aaabaab$ and $b \mapsto aab$ is the Sturmian morphism $L_a^2 R_b^2$ and preserves Lyndon words.

Moreover $L_a^2 L_b^2$ is a conjugate of $L_a^2 R_b^2$, $L_a^2 L_b^2(a) = (l_0)^2 a$ and $L_a^2 R_b^2(a) = a(l_0)^2$. Let us note that, in [14], similar remarks are made about the decomposition of the Fibonacci word (the standard Sturmian word of directive sequence $(1)_{n \geq 0}$). In Section 4, we recall notions on conjugacy of morphisms.

Let us now recall some results on morphisms preserving (finite) Lyndon words. These morphisms are studied in [14]. By definition, a morphism f *preserves Lyndon word* if for each Lyndon word w, $f(w)$ is a Lyndon word. Effective characterizations of such morphisms are given in [14]. Consequently Sturmian words preserving Lyndon words are known:

Proposition 3.2 [14] *A Sturmian morphism on* $\{a < b\}$ *is a Lyndon morphism if and only if it belongs to* $\{L_a, R_b\}^*$.

To end this section, let us observe that a study of morphisms preserving infinite Lyndon words is given in [15].

4 Strong Conjugacy

In this section, we recall the notion of conjugacy (see, e.g., [9, 13]). We also introduce the particular case of strong conjugacy which will be useful to answer G. Melançon.

Let A and B be two alphabets and let f and g be two morphisms from A^* to B^*. The morphism g is a (right) *conjugate* of f if there exists a word u such that for any word x over A, $f(x)u = ug(x)$. We will also say that f and g are u-*conjugated*, and we will denote $f \triangleleft_u g$. Moreover if $f(a) = ua$ and $g(a) = au$ for a letter a, f and g will be called *strongly* (on a) u-*conjugated*.

Let us recall that any morphism f has at least one conjugate: itself ($f \triangleleft_\varepsilon f$). The Fibonacci morphism $\varphi = L_a E$ defined by $\varphi(a) = ab$ and $\varphi(b) = a$ has exactly two conjugates, itself and the morphism $\tilde{\varphi} = R_a E$ ($\tilde{\varphi}(a) = ba$, $\tilde{\varphi}(b) = a$). A lot of relations between conjugacy of morphisms and Sturmian morphisms were given by P. Séébold [17] and generalized to a larger family of morphisms in [13].

Since $\varphi(a)$ does not end with the letter a, no morphism is strongly conjugate (on a) to the Fibonacci morphism. Nevertheless we can observe that φ^2 ($a \mapsto aba$, $b \mapsto ab$) is strongly ab-conjugated to $\varphi\tilde{\varphi}$ ($a \mapsto aab$, $b \mapsto ab$). More generally, for all integers x and y ($y \neq 0$), the morphism $L_a^x L_b^y$ is strongly conjugated to the morphism $L_a^x R_b^y$. This follows immediatly the formulas: $L_a^x L_b^y(a) = (a^x b)^y a$, $L_a^x L_b^y(b) = a^x b$, $L_a^x R_b^y(a) = a(a^x b)^y$, $L_a^x R_b^y(b) = a^x b$ ($L_a^x L_b^y \triangleleft_{(a^x b)^y} L_a^x R_b^y$).

A basic property of conjugacy is [9,13]: for morphisms f, f', g, g', and words u, u', if $f \triangleleft_u g$ and $f' \triangleleft_{u'} g'$ then $f f' \triangleleft_{f(u')u} gg'$ (of course $f(u')u = ug(u')$). This property extends to strong conjugacy:

Lemma 4.1 *Let f, f', g, g', (a a letter) and u, u' words such that f is strongly (on a) u-conjugated to g and f' is strongly (on a) u'- conjugated to g'. Then*

ff' is strongly (on a) $[f(u')u]$-conjugated to gg'.

Proof We already know $ff' \lhd_{f(u')u} gg'$. By hypothesis, $f(a) = ua$, $g(a) = au$, $f'(a) = u'a$ et $g'(a) = au'$. Thus $ff'(a) = f(u'a) = f(u')ua$ and $gg'(a) = g(au') = aug(u') = af(u')u$. So ff' is strongly $[f(u')u]$-conjugated to gg'. □

We end this section with a first use of strong conjugacy concerning Sturmian words. One particular property of any standard Sturmian word w over $\{a < b\}$ is that both aw and bw are Sturmian words [16]. Words aw (with w standard Sturmian) are also known as Christoffel words. In [3], it is shown, that Christoffel words are infinite Lyndon words:

Proposition 4.2 [3] *For any standard Sturmian word w over $\{a < b\}$, aw is an infinite Lyndon word.*

Proof Let w be a standard word with directive sequence $(a_n)_{n\geq 1}$. We have already said that a standard word can be viewed as $w = \lim_{n\to\infty} s_n$ for some words s_n defined in Section 2. In fact, we can verify that then $w = \lim_{n\to\infty} s_{2n}$. Let $n \geq 1$. We know that $s_{2n} = L_a^{a_1} L_b^{a_2} \dots L_a^{a_{2n-1}} L_b^{a_{2n}}(a)$. As a consequence of Lemma 4.1 and of the fact that for all integers x and y, the morphism $L_a^x L_b^y$ is strongly conjugated to the morphism $L_a^x R_b^y$, we can verify that $L_a^{a_1} L_b^{a_2} \dots L_a^{a_{2n-1}} L_b^{a_{2n}}$ is strongly conjugated to $L_a^{a_1} R_b^{a_2} \dots L_a^{a_{2n-1}} R_b^{a_{2n}}$.

In particular, $aL_a^{a_1} L_b^{a_2} \dots L_a^{a_{2n-1}} L_b^{a_{2n}}(a) = L_a^{a_1} R_b^{a_2} \dots L_a^{a_{2n-1}} R_b^{a_{2n}}(a)a$. By Proposition 3.2, the morphism $L_a^{a_1} R_b^{a_2} \dots L_a^{a_{2n-1}} R_b^{a_{2n}}$ preserves Lyndon words. Hence $L_a^{a_1} R_b^{a_2} \dots L_a^{a_{2n-1}} R_b^{a_{2n}}(a)$ is a Lyndon word. Consequently the word w has an infinity of Lyndon words as prefixes. It is a Lyndon word. □

Let us note that the previous proof technique can be used to state other results. For instance, we let the reader prove:

Proposition 4.3 *Let A be an alphabet and a a letter in A. Let f, g be two nonerasing endomorphisms on A and let u be a word over A such that f is u-strongly conjugate to g. Then $f^\omega(a)$ and $g^\omega(a)$ exist and $af^\omega(a) = g^\omega(a)$.*

Thus if g generates on a an infinite Lyndon word (which is the case if it preserves Lyndon words or if it preserves infinite Lyndon words (see [15])), $af^\omega(a)$ is an infinite Lyndon word.

The situation of Proposition 4.3 can be met for morphisms that are not Sturmian. For instance, this is the case with the morphisms:

$$f : \left\{ \begin{array}{l} a \mapsto aba \\ b \mapsto abb \end{array} \right. \qquad\qquad g : \left\{ \begin{array}{l} a \mapsto aab \\ b \mapsto bab \end{array} \right.$$

Moreover one can see that g preserves infinite Lyndon words and generates an infinite Lyndon word.

5 An answer to G. Melançon

In this section, we consider a standard Sturmian word w over the ordered alphabet $\{a < b\}$ with directive sequence $(a_n)_{n\geq 1}$ (Let recall that $a_1 \geq 0$ and $a_n \geq 1$ for all $n \geq 2$). The sequence of words $(s_n)_{n\geq 0}$ and $(l_n)_{n\geq 0}$ are those defined respectively at the end of Section 2 and in Theorem 3.1. In particular, $w = \lim_{n\to\infty} \prod_{n\geq 0} l_n^{a_{2n}}$ is the decomposition in Lyndon words of w (for each $n \geq 0$, l_n is a Lyndon word and $l_{n+1} \preceq l_n$). Our result is:

Theorem 5.1 *With the hypotheses of this section, there exists a morphism g such that for all $n \geq 0$, $l_{n+1} = g(l_n)$ if and only if one of the two following cases hold:*

- *$1 \leq a_1 \leq a_3$, and for all $n \geq 1$, $a_{2n} = a_2$ and $a_{2n+1} = a_3$. In this case, $l_0 = a^{a_1}b$ and $g = L_a^{a_1} R_b^{a_2} L_a^{a_3-a_1}$.*
- *$a_1 = 0$, $1 \leq a_2 \leq a_4$, and for all $n \geq 1$, $a_{2n+2} = a_4$ and $a_{2n-1} = a_3$. In this case, $l_0 = b$ and $g = R_b^{a_2} L_a^{a_3} R_b^{a_4-a_2}$.*

We observe that in each case, the morphism g is a Sturmian morphism that preserves Lyndon words (see Proposition 3.2). Moreover the word w is generated by a Sturmian morphism ($L_a^{a_1} L_b^{a_2} L_a^{a_3-a_1}$ or $L_b^{a_2} L_a^{a_3} L_b^{a_4-a_2}$).

In order to prove the previous theorem, using the strong conjugacy, we first express each Lyndon word l_n with morphisms. For $n \geq 0$, we denote:

$$f_n = (L_a^{a_1} L_b^{a_2}) \dots (L_a^{a_{2n-1}} L_b^{a_{2n}})$$

$$g_n = (L_a^{a_1} R_b^{a_2}) \dots (L_a^{a_{2n-1}} R_b^{a_{2n}})$$

The interest of the morphisms f_n is immediate since we have already seen relations between them and the words s_n ($s_{2n} = f_n(a)$, $s_{2n+1} = f_{n+1}(b)$). We also observe that each g_n is a morphism that preserves Lyndon words. As a consequence of Lemma 4.1 and of the fact that for all integers x and y, the morphism $L_a^x L_b^y$ is strongly conjugated to the morphism $L_a^x R_b^y$, we have:

Lemma 5.2 *For all $n \geq 1$, f_n is strongly (on a) conjugated to g_n.*

Now we give a new formula for the words $(l_n)_{n\geq 0}$:

Lemma 5.3 *For all $n \geq 0$, $l_n = g_n L_a^{a_{2n+1}}(b)$*

Proof

$$\begin{aligned} l_n a &= a s_{2n}^{a_{2n+1}-1} s_{2n-1} \overline{s_{2n}} a \\ &= a s_{2n}^{a_{2n+1}-1} s_{2n-1} s_{2n} \\ &= a f_n(a^{a_{2n+1}-1} ba). \end{aligned}$$

If $n = 0$, $l_n a = a^{a_1} ba = L_a^{a_1}(b)a = g_0 L_a^{a_1}(b)a$.

When $n \geq 1$, let u_n be the word such that $f \triangleleft_{u_n} g_n$. By Lemma 5.2, $f_n(a) = u_n a$, $g_n(a) = au_n$. Thus

$$\begin{aligned} l_n a &= af_n(a^{a_{2n+1}-1}b)u_n a \\ &= au_n g_n(a^{a_{2n+1}-1}b)a \\ &= g_n(a^{a_{2n+1}}b)a \\ &= g_n L_a^{a_{2n+1}}(b)a \end{aligned}$$

Consequently for all $n \geq 0$, $l_n = g_n L_a^{a_{2n+1}}(b)$. □

Let us observe that Lemma 5.2 allows to give a new proof of the fact that the words $(l_n)_{n\geq 0}$ form a strictly decreasing sequence of Lyndon words. Indeed, by Proposition 3.2, each morphism $g_n L_a^{a_{2n+1}}$ is a Lyndon morphism, hence $g_n L_a^{a_{2n+1}}$ is a Lyndon word. Moreover $R_b^{a_{2n}} L_a^{a_{2n+1}}(b)$ for each $n \geq 1$, then $R_b^{a_{2n}} L_a^{a_{2n+1}}(b) \prec b$ which implies $l_n = g_n L_a^{a_{2n+1}}(b) \prec g_{n-1} L_a^{a_{2n-1}}(b) = l_{n-1}$ (since any morphism preserving Lyndon words also strictly preserves the lexicographic order on finite words [14]).

Proof Theorem 5.1 Note that the "if" part of the theorem is immediate. Assume the sequence $(l_n)_{n\geq 0}$ is morphic. Let g be the morphism such that, for all $n \geq 0$, $g(l_n) = l_{n+1}$. Observe that the morphism g cannot be erasing since otherwise this contradicts the fact that l_2 is a primitive word (as a Lyndon word).

We first consider the case $a_1 \geq 2$. Observe $l_0 = a^{a_1}b$ and

$$g(a^{a_1}b) = l_1 = [a(a^{a_1}b)^{a_2}]^{a_3} a^{a_1} b.$$

Assume $g(a) = a$, and so $g(b) = ab(a^{a_1}b)^{a_2-1}[a(a^{a_1}b)^{a_2}]^{a_3-1}a^{a_1}b$. The word $l_2 = g(l_1)$ has $g(a^{a_1+1}b)$ as prefix. Thus the words a^{a_1+2} and $ba^{a_1}b$ are factors of l_2. This contradicts the fact that l_2, as a factor of a Sturmian word, is balanced. Hence $g(a) \neq a$.

Since $a_1 \geq 2$ and $g(a^{a_1}b)$ starts with $a^{a_1+1}b$, the word $a^{a_1+1}b$ is a prefix of $g(a)$. More precisely, $a(a^{a_1}b)^{a_2}$ must be a prefix of $g(a)$. Finally, we can verify that $g(a) = (a(a^{a_1}b)^{a_2})^k$ for an integer $k \geq 1$. It follows $g(b) = (a(a^{a_1}b)^{a_2})^{a_3-ka_1}a^{a_1}b$ which implies $a_3 \geq ka_1$.

Assume $k \geq 2$. The word $l_2 = g(l_1)$ contains $g(ba^{a_1}b)$ and $g(a^{a_1+1}b)$ as factors. The word $g(ba^{a_1}b)$ ends with bub where $u = (a^{a_1}b)^{a_2}[a(a^{a_1}b)^{a_2}]^{a_3}a^{a_1}$. Furthermore the word $g(a^{a_1+1}b) = [a(a^{a_1}b)^{a_2}]^{a_3+k}a^{a_1}b$ starts with aua. This contradicts the fact that l_2 is balanced.

Hence $k = 1$, $a_3 \geq a_1$, $g(a) = a(a^{a_1}b)^{a_2}$, $g(b) = [a(a^{a_1}b)^{a_2}]^{a_3-a_1}a^{a_1}b$. We observe that $g = L_a^{a_1} R_b^{a_2} L_a^{a_3-a_1}$ and that it is an injective morphism.

Now we can prove that, for all $n \geq 1$, $a_{2n} = a_2$ and $a_{2n+1} = a_3$. We act by induction on n. There is nothing to do for $n = 1$. Let $n \geq 1$. Assume that we have already proved $a_{2p} = a_2$ and $a_{2p+1} = a_3$ for all integers p with $1 \leq p \leq n$. We have

$$\begin{aligned} l_{n+1} = g_{n+1}L_a^{a_{2n+3}}(b) &= L_a^{a_1}(R_b^{a_2}L_a^{a_3})^n R_b^{a_{2n+2}}L_a^{a_{2n+3}}(b) \\ &= (L_a^{a_1}R_b^{a_2}L_a^{a_3-a_1})^n L_a^{a_1}R_b^{a_{2n+2}}L_b^{a_{2n+3}}(b) \\ &= g^n(L_a^{a_1}R_b^{a_{2n+2}}L_b^{a_{2n+3}}(b))\,. \end{aligned}$$

Moreover $l_{n+1} = g^n(l_1)$. Since g is injective, $l_1 = L_a^{a_1}R_b^{a_{2n+2}}L_b^{a_{2n+3}}(b)$. This implies $a_{2n+2} = a_2$ and $a_{2n+3} = a_3$.

Now we consider the case $a_1 = 1$. We have $l_0 = ab$ and $l_1 = [a(ab)^{a_2}]^{a_3}ab$. As in case $a_1 \geq 2$, we cannot have $g(a) = a$. Hence $g(a)$ starts with aa. We observe that $g(a)$ cannot ends with a, since otherwise the balanced word $l_2 = g(l_1)$ contains aaa and bab. We observe also that $g(a) \neq [a(ab)^{a_2}]^i a(ab)^k$ for any integer k, i such that $1 \leq k < a_2$ and $i \geq 0$. Indeed otherwise the word l_2 containing both $g(aa)$ and $g(ab)$ should contains the factors $a(ab)^k aa$ and $b(ab)^k ab$ (since $(ab)^{a_2+1}$ ends $g(ab)$): this contradicts the fact that l_2 is balanced. It follows that $g(a) = [a(ab)^{a_2}]^k$ with $1 \leq k \leq a_3$ and $g(b) = [a(a^{a_1}b)^{a_2}]^{a_3-k}a^{a_1}b$. Exactly as in case $a_1 \geq 2$, we can then prove that $k = 1$, $g = L_aR_b^{a_2}L_a^{a_3-1}$ and for all integers $n \geq 1$, $a_{2n} = a_2$ and $a_{2n+1} = a_3$.

From now on, we consider the case $a_1 = 0$. we have $l_0 = b$ and so $g(b) = l_1 = (ab^{a_2})^{a_3}b$. Moreover $l_2 = R_b^{a_2}L_a^{a_3}R_b^{a_4}L_a^{a_5}(b)$, that is

$$l_2 = [ab^{a_2}[(ab^{a_2})^{a_3}b]^{a_4}]^{a_5}(ab^{a_2})^{a_3}b.$$

Furthermore $l_2 = g^2(b) = g((ab^{a_2})^{a_3}b)$. It follows that

$$g((ab^{a_2})^{a_3}) = [ab^{a_2}g(b)^{a_4}]^{a_5}$$

Since the word $ab^{a_2}g(b)^{a_4} = ab^{a_2}[(ab^{a_2})^{a_3}b]^{a_4}$ is a primitive word, $g(ab^{a_2}) = [ab^{a_2}g(b)^{a_4}]^x$ and $xa_3 = a_5$ for an integer $x \geq 1$. Since ab^{a_2} is not a suffix of $g(b)$, $a_2 \leq a_4$.

Let us prove that $x = 1$, that is, $a_3 = a_5$. Assume by contradiction that $x \geq 2$. The word l_2 has $(ab^{a_2})^{a_3+1}$ as a prefix and $[(ab^{a_2})^{a_3}b]^2$ as a suffix. Let $u = ab^{a_2}g(b)^{a_4}$: $g(ab^{a_2}) = u^x$. The word $l_3 = g(l_2)$ contains the factor $g((ab^{a_2})^{a_3+1}) = u^{(a_3+1)x} = uu^{a_5}uu^{x-2}$ which contains the factor $ab^{a_2}g(b)^{a_4}u^{a_5}ab^{a_2}(ab^{a_2})^{a_3}b$ which starts with $ab^{a_2}g(b)^{a_4}u^{a_5}(ab^{a_2})^{a_3}a$. Observe now that $g((ab^{a_2})^{a_3}b) = [ab^{a_2}g(b)^{a_4}]^{a_5}g(b)$ ends with $b^{a_2+1}g(b)^{a_4}$. Consequently the word l_3 also contains the factor

$$b^{a_2+1}g(b)^{a_4}g(((ab)^{a_2})^{a_3}b) = bb^{a_2}g(b)^{a_4}u^{a_5}(ab^{a_2})^{a_3}b\,.$$

We have a contradiction with the fact that l_3 is a balanced word.

From what precedes, $g(ab^{a_2}) = ab^{a_2}g(b)^{a_4}$ and so $g(a) = ab^{a_2}g(b)^{a_4-a_2} = ab^{a_2}((ab^{a_2})^{a_3}b)^{a_4-a_2}$. Moreover $g(b) = (ab^{a_2})^{a_3}b$. We observe $g = R_b^{a_2}L_a^{a_3}R_b^{a_4-a_2}$. As in case $a_1 \geq 2$, we can state that, for all integers $n \geq 2$, $a_{2n} = a_4$ and $a_{2n-1} = a_3$. □

6 Conclusion

This paper shows the interest of conjugacy of morphisms and of morphisms preserving Lyndon words as tools to tackle problems concerning Sturmian words and/or Lyndon words. We are now working to find other situations where these tools can be useful. In particular, we are looking for the decomposition in Lyndon words of any Sturmian words.

Acknowledgement

Many thanks to P. Séébold for useful discussions on strong conjugacy.

References

[1] J. Berstel and P. Séébold, Sturmian words, chapter 2 in [9].

[2] V. Berthé, C. Holton, and L. Q. Zamboni, Initial powers of sturmian sequences. To appear in *Acta Informatica*, see www.lirmm.fr/∼berthe.

[3] J. P. Borel and F. Laubie, Quelques mots sur la droite projective réelle, *Journal de Théorie des Nombres de Bordeaux*, 5:23–51, 1993.

[4] K. T. Chen, R. H. Fox, and R C. Lyndon, Free differential calculus IV – the quotient groups of the lower central series, *Ann. Math. 68*, 68:81–95, 1958.

[5] A. Ido and G. Melançon, Lyndon factorization of the Thue-Morse word and its relatives, *Discret. Math. and Theoret. Comput. Sci.*, 1:43–52, 1997.

[6] J. Justin and G. Pirillo, Episturmian words and episturmian morphisms, *Theoretical Computer Science*, 276(1-2):281–313, 2002.

[7] J. Justin and G. Pirillo, On a characteristic property of Arnoux-Rauzy sequences, *RAIRO Theoretical Informatics and Applications*, 36:385–388, 2002.

[8] M. Lothaire, *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics*, Addison-Wesley, 1983. Reprinted in 1997 by Cambridge University Press in the Cambridge Mathematical Library, Cambridge, UK, 1997.

[9] M. Lothaire, *Algebraic Combinatorics on words*, volume 90 of *Encyclopedia of Mathematics*, Cambridge University Press, Cambridge, UK, 2002.

[10] M. Lothaire, *Applied Combinatorics on Words.* To appear.
(see www-igm.univ-mlv.fr/∼berstel).

[11] G. Melançon, Lyndon factorization of infinite words. In *STACS'96*, volume 1046 of *Lect. Notes in Comp. Sci.*, pages 147–154, 1996.

[12] G. Melançon, Lyndon factorization of Sturmian words, *Discrete Mathematics*, 210:137–149, 2000.

[13] G. Richomme, Conjugacy and episturmian morphisms, *Theoretical Computer Science*, 302:1–34, 2003.

[14] G. Richomme, Lyndon morphisms, *Bulletin of the Belgian Mathematical Society*, 10:761–785, 2003.

[15] G. Richomme, On morphisms preserving infinite lyndon words, *Journal of Automata, Languages and Combinatorics*, submitted. An extended abstract appears in the abstract of the conference "Journées Montoises d'Informatique Théorique", Liège, Belgium, 2004 (Prépublication 04.006, Institut de Mathématique, Université de Liège).

[16] P. Séébold, Fibonacci morphisms and Sturmian words, *Theoretical Computer Science*, 88:365–384, 1991.

[17] P. Séébold, On the conjugation of standard morphisms, *Theoretical Computer Science*, 195:91–109, 1998.

[18] P. Séébold, Lyndon factorization of the Prouhet words, *Theoretical Computer Science*, 307:179–197, 2003.

[19] R. Siromoney, L. Mathew, V. R. Dare, and K. G. Subramanian, Infinite Lyndon words, *Information Processing Letters*, 50:101–104, 1994.

Publications du **La**boratoire de **C**ombinatoire et d'**I**nformatique **M**athématique

.... *suite de la page couverture 2*

18 *Hyperbinomiales: doubles suites satisfaisant à des équations aux différences partielles de dimension et d'ordre deux de la forme* $H(n,k) = p(n,k)H(n-1,k) + q(n,k)H(n-1,k-1)$, P. Théoret, **1994**

19 *Théorie des espèces et combinatoire des structures arborescentes*, F. Bergeron, G. Labelle, P. Leroux, **1994**

20 *Axel Thue's papers on repetitions in words: a translation*, J. Berstel, **1995**

21 *Une généralisation des séries indicatrices d'espèces de structures*, K. Pineau, **1996**

22 *Fonctions symétriques, ensembles de descentes et classes de conjugaison dans les produits en couronne*, S. Poirier, **1996**

23 *Combinatoire des systèmes d'équations différentielles et aux dérivées partielles*, B. Randrianirina, **1998**

24 *Espèces de structures et applications au dénombrement de cartes et de cactus planaires*, M. Bousquet, **1998**

25 *Contribution à l'étude des mots sturmiens*, B. Parvaix, **1998**

26 *Analyse d'équations combinatoires en théorie des espèces*, P. Auger, **1999**

27 *Colloque LaCIM 2000 — Combinatoire, Informatique et Applications*, P. Leroux (éd.), Septembre 2000, Actes, **2000**

28 *Fonctions de Bessel et combinatoire*, F. Gascon, **2001**

29 *The Saga of Reduced Factorizations of Elements of the Symmetric Group*, A. Garsia, **2002**

30 *Variétés de carquois et homologie d'intersection*, R. Schiffler, **2003**

31 *Algorithmes vectoriels et bioinformatique*, S. Hamel, **2003**

32 *Conjecture* $n!$ *et généralisations*, J.-C. Aval, **2004**

33 *Spécification de classes de structures arborescentes*, C. Lamathe, **2004**

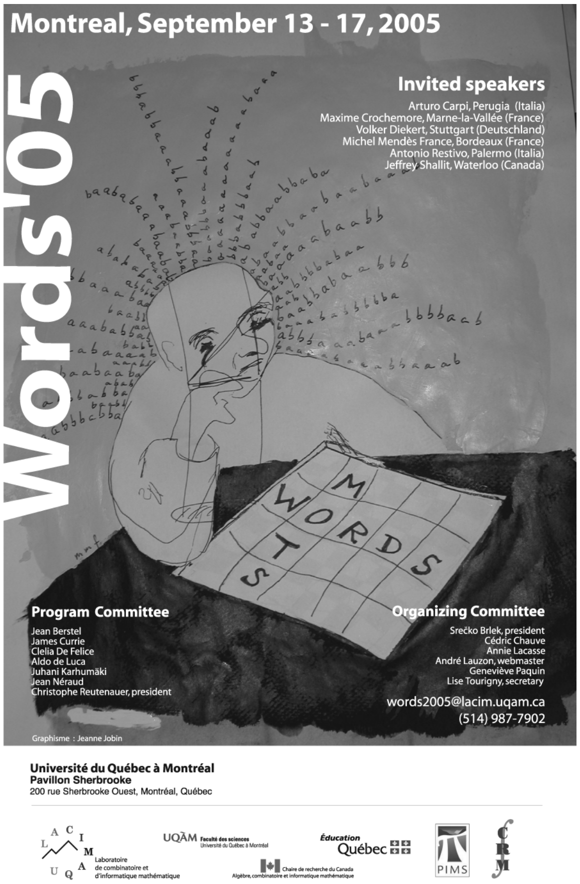
34 *A Holonomic Systems Approach to Algebraic Combinatorics*, M. Mishna, **2005**

35 *Une famille d'algorithmes à balayage circulaire pour le calcul de diagrammes de Voronoï de points ou de cercles pondérés*, A. Pavillet, **2005**

36 *Words 2005 — 5^{th} International Conference on Words*, Srečko Brlek, Christophe Reutenauer (eds.), **2005**

WORDS 2005
5th INTERNATIONAL CONFERENCE ON WORDS

La conférence WORDS 2005 était la cinquième édition d'une série qui débuta en 1997 à Rouen (France). Le sujet de la conférence est l'étude des mots avec une emphase sur le point de vue théorique. En particulier les aspects combinatoires algébriques et algorithmiques sont privilégiés bien que les motivations puissent provenir d'autres domaines tels que l'informatique théorique. La conférence consista en six conférences plénières et également 28 communications sélectionnées par le comité de programme.



Montreal, September 13 - 17, 2005

Words'05

Invited speakers

Arturo Carpi, Perugia (Italia)
Maxime Crochemore, Marne-la-Vallée (France)
Volker Diekert, Stuttgart (Deutschland)
Michel Mendès France, Bordeaux (France)
Antonio Restivo, Palermo (Italia)
Jeffrey Shallit, Waterloo (Canada)

Program Committee

Jean Berstel
James Currie
Clelia De Felice
Aldo de Luca
Juhani Karhumäki
Jean Néraud
Christophe Reutenauer, president

Organizing Committee

Srečko Brlek, president
Cédric Chauve
Annie Lacasse
André Lauzon, webmaster
Geneviève Paquin
Lise Tourigny, secretary

words2005@lacim.uqam.ca
(514) 987-7902

Graphisme : Jeanne Jobin

Université du Québec à Montréal
Pavillon Sherbrooke
200 rue Sherbrooke Ouest, Montréal, Québec

Laboratoire de combinatoire et d'informatique mathématique
Université du Québec à Montréal
C.P. 8888, Succ. Centre-Ville
Montréal (Québec) Canada
H3C 3P8

ISBN 2-89276-376-2