Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering

Doctoral Thesis

Substitutive sequences and their properties

Prague 2020 Kateřina Medková

I warmly thank my advisor Edita Pelantová for her wisdom, kind support and excellent advice. I thank my Master's thesis advisor Karel Klouda who patiently introduced me to scientific research during my undergraduate studies, and I thank the whole TIGR group for providing a pleasant and friendly study environment.

I also thank my family and my dear Víťa for everything, including their encouragements when finishing this thesis.

Bibliografický záznam

Autorka	Ing. Kateřina Medková České vysoké učení technické v Praze Fakulta jaderná a fyzikálně inženýrská Katedra matematiky
Název práce	Substitutivní posloupnosti a jejich vlastnosti
Studijní program	Aplikace přírodních věd
Studijní obor	Matematické inženýrství
Školitelka	prof. Ing. Edita Pelantová, CSc. České vysoké učení technické v Praze Fakulta jaderná a fyzikálně inženýrská Katedra matematiky
Školitel specialista	Ing. Karel Klouda, Ph.D. České vysoké učení technické v Praze Fakulta informačních technologií Katedra aplikované matematiky
Akademický rok	2019/2020
Počet stran	195
Klíčová slova	kombinatorika na slovech, derivované slovo, sturmovské slovo, Arnouxovo–Rauzyho slovo, Roteho slovo, kritický exponent, rekurentní funkce, komplexita bez opakování

Bibliographic entry

Author	Ing. Kateřina Medková Czech Technical University in Prague Faculty of Nuclear Sciences and Physical Engineering Department of Mathematics
Title of Dissertation	Substitutive sequences and their properties
Degree Programme	Application of Natural Sciences
Field of Study	Mathematical Engineering
Supervisor	prof. Ing. Edita Pelantová, CSc. Czech Technical University in Prague Faculty of Nuclear Sciences and Physical Engineering Department of Mathematics
Supervisor specialist	Ing. Karel Klouda, Ph.D. Czech Technical University in Prague Faculty of Information Technology Department of Applied Mathematics
Academic Year	2019/2020
Number of Pages	195
Keywords	combinatorics on words, derived sequence, Sturmian sequence, Arnoux–Rauzy sequence, Rote sequence, critical exponent, recurrence function, non-repetitive complexity

Abstrakt

Tato dizertační práce se věnuje studiu strukturních vlastností nekonečných slov s nízkou faktorovou komplexitou. Práce je koncipována jako soubor pěti autorčiných článků. Tři z nich již byly publikovány v odborných časopisech, zbývající dva články jsou v recenzním řízení.

Největší část práce se věnuje derivovaným posloupnostem. Připomeňme, že každé uniformně rekurentní slovo $\mathbf{u}$ můžeme zapsat jako zřetězení konečného počtu návratových slov ke zvolenému prefixu slova $\mathbf{u}$. Uspořádání návratových slov v tomto zřetězení je kódováno příslušnou derivovanou posloupností. Durand ukázal, že slovo $\mathbf{u}$ je primitivně substitutivní právě tehdy, když je množina $\mathrm{Der}(\mathbf{u})$ všech jeho derivovaných posloupností konečná.

Nejprve se věnujeme případu, kdy je sturmovské slovo $\mathbf{u}$ pevným bodem primitivního morfismu. V tomto případě přicházíme s algoritmem, který vrací všechny morfismy fixující derivované posloupnosti slova $\mathbf{u}$. Díky tomu můžeme zkonstruovat dobrý horní odhad na velikost množiny $\mathrm{Der}(\mathbf{u})$.

Poté zobecňujeme tyto výsledky a popisujeme množinu $\mathrm{Der}(\mathbf{u})$ pro libovolně zvolené Arnouxovo–Rauzyho slovo $\mathbf{u}$. Využíváme k tomu speciální S-adickou reprezentaci slova $\mathbf{u}$. Z tohoto popisu také přímo vyplývá, že derivovaná posloupnost k Arnouxovu–Rauzyho slovu $\mathbf{u}$ je vždy Arnouxovo–Rauzyho slovo.

Studujeme také derivované posloupnosti pro libovolné komplementárně symetrické (CS) Roteho slovo $\mathbf{v}$, které souvisí se standardním sturmovským slovem $\mathbf{u}$. Vysvětlíme, že libovolný neprázdný prefix takového slova $\mathbf{v}$ má právě tři návratová slova. Také ukážeme, že libovolná derivovaná posloupnost slova $\mathbf{v}$ je slovo kódující výměnu intervalů a najdeme parametry této transformace. Dále ukážeme, že slovo $\mathbf{v}$ je primitivně substitutivní právě tehdy, když je slovo $\mathbf{u}$ primitivně substitutivní.

Věnujeme se také jiným vlastnostem nekonečných slov. Pro CS Roteho slova najdeme vzorce pro výpočet kritického exponentu a rekurentní funkce. S využitím vztahu pro kritický exponent popíšeme všechna CS Roteho slova s kritickým exponentem menším nebo rovným třem. Dále ukážeme, že existuje nekonečně mnoho CS Roteho slov s kritickým exponentem menším, než je kritický exponent Fibonacciho slova.

Nakonec studujeme komplexitu bez opakování $\mathrm{nr}\mathcal{C}_{\mathbf{u}}$ a počáteční komplexitu bez opakování $\mathrm{inr}\mathcal{C}_{\mathbf{u}}$, což jsou funkce, které popisují strukturu slova $\mathbf{u}$ s ohledem na repetice jeho faktorů dané délky. Najdeme vyjádření funkce $\mathrm{nr}\mathcal{C}_{\mathbf{u}}$ pro všechna Arnouxova–Rauzyho slova a také vyjádření funkce $\mathrm{inr}\mathcal{C}_{\mathbf{u}}$ pro standardní Arnouxova–Rauzyho slova. Získané vzorce aplikujeme na d-bonacciho slovo.

Abstract

This thesis is devoted to the study of structural properties of sequences with low factor complexity. The presented work is a collection of five author's papers. Three of them have been already published in refereed journals, while the remaining two articles are currently being refereed.

The main part of this thesis deals with derived sequences. Any uniformly recurrent sequence $\mathbf{u}$ can be written as the concatenation of a finite number of return words to a chosen prefix of $\mathbf{u}$. Ordering of these return words in this concatenation is coded by the derived sequence. Durand proved that a sequence $\mathbf{u}$ is primitive substitutive if and only if the set $\mathrm{Der}(\mathbf{u})$ of all derived sequences to the prefixes of $\mathbf{u}$ is finite.

First we focus on a Sturmian sequence $\mathbf{u}$ fixed by a primitive morphism. In this case we provide an algorithm which lists the morphisms fixing the derived sequence of $\mathbf{u}$. This enables us to provide a sharp upper bound on the cardinality of the set $\mathrm{Der}(\mathbf{u})$.

More generally, we describe the set $\mathrm{Der}(\mathbf{u})$ for every Arnoux–Rauzy sequence $\mathbf{u}$ using a special S-adic representation of $\mathbf{u}$. As a corollary, we show that all derived sequences of $\mathbf{u}$ are Arnoux–Rauzy sequences.

We study derived sequences also for a complementary symmetric (CS) Rote sequence $\mathbf{v}$ which is related to a standard Sturmian sequence $\mathbf{u}$. We show that any non-empty prefix of $\mathbf{v}$ has exactly three return words. We prove that any derived sequence of $\mathbf{v}$ is the coding of three interval exchange transformation and we determine the parameters of this transformation. We also prove that $\mathbf{v}$ is primitive substitutive if and only if $\mathbf{u}$ is primitive substitutive.

In addition, for all CS Rote sequences we give the formulas for their critical exponent and recurrence function. Using the formula for the critical exponent, we describe all CS Rote sequences with the critical exponent less than or equal to three, and we show that there are uncountably many CS Rote sequences with the critical exponent less than the critical exponent of the Fibonacci sequence.

Finally, we study non-repetitive complexity $\mathrm{nr}\mathcal{C}_{\mathbf{u}}$ and initial non-repetitive complexity $\mathrm{inr}\mathcal{C}_{\mathbf{u}}$, which reflect the structure of a sequence $\mathbf{u}$ with respect to the repetitions of factors of a given length. We determine $\mathrm{nr}\mathcal{C}_{\mathbf{u}}$ for Arnoux–Rauzy sequences and $\mathrm{inr}\mathcal{C}_{\mathbf{u}}$ for standard Arnoux–Rauzy sequences. The obtained formulas are then used to evaluate the values of $\mathrm{nr}\mathcal{C}_{\mathbf{u}}$ and $\mathrm{inr}\mathcal{C}_{\mathbf{u}}$ for the d-bonacci sequence.

Contents

Chapter 1

Introduction

Combinatorics on words is dedicated to the study of properties of finite and infinite words (we call them sequences). Since words appear naturally in many contexts, it is closely related to many other fields, e.g., symbolic dynamical systems, some parts of number theory such as positional number systems or continued fraction expansions, or several fields of theoretical informatics such as theory of codes, languages, automata or L systems.

Among all sequences, the sequences which can be generated via morphisms (also called substitutions) are prominent, since these generating morphisms provide us with a useful tool for their study. They are called *substitutive sequences.*

For example, the sequence $\mathbf{u} = 0100101001001010010100100101001001\cdots$ can be obtained by the repeated application of the morphism $\psi : 0 \to 010, 1 \to 01$:

$$0 \to 010 \to 01001010 \to 010010100100101001010 \to \cdots \to \mathbf{u}\,.$$

In fact, to get the sequence $\mathbf{u}$ we have to apply the morphism ψ infinitely many times, and so we write $\mathbf{u} = \psi^{\omega}(0)$. More precisely we also say that $\mathbf{u}$ is a fixed point of the morphism ψ, i.e., $\psi(\mathbf{u}) = \mathbf{u}$.

Similarly, the sequence $\mathbf{v} = 0010001001000100010010001001000010001\cdots$ can be obtained as the image of $\mathbf{u}$ under the morphism $\tau : 0 \to 0, 1 \to 01$, i.e., $\mathbf{v} = \tau(\mathbf{u}) = \tau(\psi^{\omega}(0))$. Hence $\mathbf{v}$ can be easily generated using two morphisms τ and ψ.

More generally, we can consider sequences which are generated using a finite set of morphisms by applying them repeatedly in a given order. This leads us to the notion of S-adic representation, which is the central notion in our research.

In this thesis we study structural properties of well-known classes of sequences with low factor complexity, namely the Sturmian, Arnoux–Rauzy and Rote sequences. The main advantage of these classes is that they admit useful S-adic representations.

First of all, we focus on the derived sequences of these sequences, but we also study their other properties such as the critical exponent, the recurrence function and the non-repetitive complexity.

This thesis is organized as follows. We first give an overview of the studied properties and classes of sequences in Chapter 2. Then in Chapter 3 we describe the contents of the articles comprising the thesis. The articles themselves then follow:

[A] K. Klouda, K. Medková, E. Pelantová, and Š. Starosta, Fixed points of Sturmian morphisms and their derivated words. *Theoretical Computer Science* 743 (2018), 23–37.

[B] K. Medková, Derived sequences of Arnoux–Rauzy sequences. In: R. Mercas, D. Reidenbach (eds.), *WORDS 2019, Lecture notes in Computer Science* 11682, Springer (2019), 251–263.

[C] K. Medková, E. Pelantová, and L. Vuillon, Derived sequences of complementary symmetric Rote sequences. *RAIRO - Theoretical Informatics and Applications* 53 (2019), 125–151.

[D] K. Medková, E. Pelantová, and É. Vandomme, On non-repetitive complexity of Arnoux–Rauzy words. Submitted to *Discrete Applied Mathematics.*

[E] L. Dvořáková, K. Medková, and E. Pelantová, Complementary symmetric Rote sequences: the critical exponent and the recurrence function. Submitted to *Discrete Mathematics & Theoretical Computer Science.*

Chapter 2

Overview of the field

The aim of this chapter is to present several key properties and classes of sequences which are studied in this thesis. The basic notions of combinatorics on words are recalled in Section 2.1. In Section 2.2 we introduce well-known classes of sequences with low factor complexity such as Sturmian, Arnoux–Rauzy, Rote and dendric sequences. In Section 2.3 we summarize the known results about several key properties of sequences. We focus on the return words and the derived sequences, the critical exponent, the recurrence function and the non-repetitive complexity. Finally we present useful tools for their study in Section 2.4.

2.1 Preliminaries

In this section we recall well-known definitions from combinatorics on words and fix the basic notation that we use in the rest of the thesis. More details about the background can be found for example in the books [5, 23, 72–74].

2.1.1 Words, sequences, factors and languages

An *alphabet* $\mathcal{A}$ is a set of symbols called *letters*. In this thesis we always suppose that the alphabet is finite.

By a *(finite) word* of length n over $\mathcal{A}$ we mean a finite string $w = w_0w_1 \cdots w_{n-1}$ of n letters from the alphabet $\mathcal{A}$. We denote its length $|w| = n$. There is a unique word ε with the length $|\varepsilon| = 0$ which is called the *empty word*.

We denote $\mathcal{A}^*$ the set of all finite words over the alphabet $\mathcal{A}$. We can endow this set with a binary operation $\circ$ called the *concatenation* of words. The concatenation of the words $u = u_0u_1 \cdots u_{n-1}$ and $v = v_0v_1 \cdots v_{m-1}$ is the word $u \circ v = u_0u_1 \cdots u_{n-1}v_0v_1 \cdots v_{m-1}$. We typically abbreviate the notation and denote uv the concatenation of the words u and v. The algebraic structure $(\mathcal{A}^*, \circ)$ is the free monoid generated by $\mathcal{A}$ and its neutral element is the empty word ε. We also denote $\mathcal{A}^+$ the set of all non-empty finite words over $\mathcal{A}$.

A *sequence* $\mathbf{u} = u_0u_1u_2 \cdots$ over $\mathcal{A}$ is a right-infinite string of letters from $\mathcal{A}$ (it is also called an *infinite word*). Let us note that some authors work with the so-called bi-infinite sequences $\cdots u_{-2}u_{-1}u_0u_1u_2 \cdots$. While many properties remain the same, in some aspects right-infinite and bi-infinite sequences differ essentially. Thus we have to distinguish properly between these two concepts. Here we always work with right-infinite sequences and we denote them by bold letters. We also denote $\mathcal{A}^{\mathbb{N}}$ the set of all sequences over $\mathcal{A}$. Let us emphasize that by $\mathbb{N}$ we always mean the set of non-negative integers, i.e., $\mathbb{N} = \{0, 1, 2, \ldots\}$.

A sequence of the form $\mathbf{u} = vuuu\cdots = vu^{\omega}$, where $v, u \in \mathcal{A}^*$ and u is non-empty, is called *(eventually) periodic*. If, moreover, v is empty, then $\mathbf{u} = u^{\omega}$ is called *purely periodic*. Since the structure of eventually periodic sequences is clear, we usually consider these sequences as trivial and we focus especially on *aperiodic* sequences, i.e., the sequences which are not periodic.

Example 2.1. Champernowne (Barbier) sequence $\mathbf{h} = 011011100101\cdots$ obtained by the concatenation of the binary representations of natural numbers is an example of a binary aperiodic sequence.

A word w of length n is a *factor* of $\mathbf{u}$ if $w = u_i u_{i+1} \cdots u_{i+n-1}$ for some index $i \in \mathbb{N}$; this index i is called an *occurrence* of w in $\mathbf{u}$. If $i = 0$, then we say that w is a *prefix* of $\mathbf{u}$. Analogously, we define these terms for finite words. Let $w = pus$ for some $p, u, s \in \mathcal{A}^*$. Then p is a *prefix* of w, u is a *factor* of w and s is a *suffix* of w. We also use the notation $us = p^{-1}w$ and $pu = ws^{-1}$.

The set of all factors of a sequence $\mathbf{u}$ is called the *language* of $\mathbf{u}$ and is denoted $\mathcal{L}_{\mathbf{u}}$. Moreover, we denote $\mathcal{L}_{\mathbf{u}}(n)$ the set of all factors of $\mathbf{u}$ of length n. Typically, there are infinitely many sequences with the same language. However, these sequences have lots of common properties, since many properties depend only on the language and not on the sequences themselves.

The sequence $\mathbf{u}$ is called *recurrent* if each of its factors occurs in $\mathbf{u}$ infinitely many times. Moreover, $\mathbf{u}$ is *uniformly recurrent* if for each factor w of $\mathbf{u}$ the gaps between consecutive occurrences of w are bounded. In other words, $\mathbf{u}$ is uniformly recurrent if for every integer n there exists an integer m such that each factor of $\mathbf{u}$ of length n occurs in every factor of $\mathbf{u}$ of length m.

2.1.2 Morphisms, fixed points and substitutive sequences

Let $\mathcal{A}, \mathcal{B}$ be alphabets. A *morphism* from $\mathcal{A}$ to $\mathcal{B}$ is any homomorphism θ: $(\mathcal{A}^*, \circ) \to (\mathcal{B}^*, \circ)$, i.e., it is a mapping $\theta : \mathcal{A}^* \to \mathcal{B}^*$ which for all $u, v \in \mathcal{A}^*$ satisfies $\theta(uv) = \theta(u)\theta(v)$. Clearly, the morphism θ is uniquely determined by the images of letters from $\mathcal{A}$. Hence the domain of the morphism θ can be naturally extended to the set of all sequences over $\mathcal{A}$: for any sequence $\mathbf{u} = u_0 u_1 u_2 \cdots \in \mathcal{A}^{\mathbb{N}}$ we put

$$\theta(\mathbf{u}) = \theta(u_0)\theta(u_1)\theta(u_2)\cdots .$$

If $\mathcal{A} = \mathcal{B}$, we say that θ is a *morphism on* $\mathcal{A}$. A morphism θ on $\mathcal{A}$ is called *prolongable on* $a \in \mathcal{A}$ if

$$\theta(a) = aw \text{ for some non-empty word } w \in \mathcal{A}^* \quad \text{and} \quad \lim_{n \to +\infty} |\theta^n(a)| = +\infty .$$

Such a morphism is sometimes called a *substitution*. A morphism θ is *primitive* if there is an integer $k \geq 1$ such that for all pairs of letters $a, b \in \mathcal{A}$ the word $\theta^k(a)$ contains the letter b.

A sequence $\mathbf{u}$ such that $\theta(\mathbf{u}) = \mathbf{u}$ is called a *fixed point* of the morphism θ. Clearly, a morphism θ which is prolongable on some letter a has the fixed point starting with this letter a. Indeed, since for each $n \in \mathbb{N}$ the word $\theta^n(a)$ is a proper prefix of the word $\theta^{n+1}(a) = \theta^n(a)\theta^n(w)$, the limit of the sequence $(\theta^n(a))_{n\geq 0}$ exists and is the sequence

$$\theta^\omega(a) = \lim_{n\to\infty} \theta^n(a) = aw\theta(w)\theta^2(w)\cdots .$$

Clearly, the sequence $\theta^\omega(a)$ is a fixed point of θ and we usually say that $\theta^\omega(a)$ is a sequence generated by θ.

Example 2.2. The Fibonacci morphism φ is defined as $0 \to 01$, $1 \to 0$. It is primitive since for $k = 2$ we have $\varphi^2(0) = 010$ and $\varphi^2(1) = 01$. It is also prolongable on 0, hence it generates the sequence $\mathbf{f} = 0100101001001\cdots$ which is called the Fibonacci sequence. This sequence is the only fixed point of φ.

As a generalization of this concept, we can consider also sequences which arise as the images of fixed points under other morphisms. More precisely, a sequence $\mathbf{u}$ over $\mathcal{B}$ is called *substitutive* (or also *morphic*) if it is of the form $\mathbf{u} = \kappa(\theta^\omega(a))$, where θ is a morphism on $\mathcal{A}$ prolongable on $a \in \mathcal{A}$ and κ is a letter-to-letter morphism from $\mathcal{A}$ to $\mathcal{B}$ (i.e., $|\kappa(c)| = 1$ for all $c \in \mathcal{A}$).

In fact, every sequence of the form $\mathbf{u} = \tau(\psi^\omega(a))$, where ψ is a morphism prolongable on a and τ is any morphism, is substitutive, i.e., there is a letter-to-letter morphism κ and a morphism θ prolongable on a such that $\mathbf{u} = \kappa(\theta^\omega(a))$, see [5, Corollary 7.7.5].

A morphism θ can have more than one fixed point. However, if θ is primitive, then all its fixed points have the same language. Fixed points of primitive morphisms have other useful properties. First of all, they are always uniformly recurrent. Even more generally, every substitutive sequence $\kappa(\theta^\omega(a))$ with primitive θ is also uniformly recurrent. These sequences are sometimes called *primitive substitutive*.

Example 2.3. We take the Fibonacci morphism φ from Example 2.2 and consider the Thue–Morse morphism μ defined by $0 \to 01$, $1 \to 10$. Then the sequence

$$\mathbf{u} = \mu(\varphi^\omega(0)) = \mu(\mathbf{f}) = 011001011001100101100101 10\cdots$$

is primitive substitutive. Indeed, it can be also obtained as $\mathbf{u} = \kappa(\theta^\omega(A))$, where the morphisms θ and κ are defined by

$$\theta : A \to AB,\ B \to CD,\ C \to A,\ D \to B \quad \text{and} \quad \kappa : A \to 0,\ B \to 1,\ C \to 1,\ D \to 0\,.$$

One can read more about fixed points and substitutive sequence for example in [5, Section 7] or in the note [3] that describes a morphic taxonomy of sequences. In Section 2.4.2 we discuss the so-called S-adic representations of sequences which in some sense further generalize the concept of generating sequences via morphisms.

2.1.3 Parikh vectors

For any (finite) word w over $\mathcal{A}$ and any letter $a \in \mathcal{A}$ we denote $|w|_a$ the number of letters a occurring in w. Let $\mathcal{A}$ have size d. The *Parikh vector* of a word w is the vector $\vec{V}(w) \in \mathbb{N}^d$ defined as

$$\big(\vec{V}(w)\big)_a = |w|_a \text{ for all } \ a \in \mathcal{A}\,.$$

Similarly, the *matrix* of a morphism θ on $\mathcal{A}$ is the matrix $\mathbf{M}_\theta \in \mathbb{N}^{d\times d}$ defined as

$$\big(\mathbf{M}_\theta\big)_{ab} = |\theta(b)|_a \text{ for all } \ a, b \in \mathcal{A}\,.$$

Example 2.4. The word 0100 over the binary alphabet $\{0, 1\}$ has the Parikh vector $\vec{V}(0100) = \binom{3}{1}$. The Fibonacci morphism from Example 2.2 has the matrix

$$\mathbf{M}_\varphi = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} .$$

This notion is especially useful when dealing with the length of factors and morphisms since the relations

$$\vec{V}(\theta(w)) = \mathbf{M}_\theta \cdot \vec{V}(w) \quad \text{and} \quad |w| = (1, 1, \ldots, 1) \cdot \vec{V}(w)$$

convert it to the multiplication of vectors and matrices.

2.1.4 Special factors

Let $\mathbf{u}$ be a sequence over $\mathcal{A}$ and let w be its factor. The factor w is *left special* if there are at least two distinct letters a, b such that $aw, bw \in \mathcal{L}_\mathbf{u}$. Analogously, w is called *right special* if there are at least two distinct letters c, d such that $wc, wd \in \mathcal{L}_\mathbf{u}$. The factor is *bispecial* if it is both left and right special.

In addition, each letter a such that $aw \in \mathcal{L}_\mathbf{u}$ is called a *left extension* of the factor w in $\mathbf{u}$ and each letter b such that $wb \in \mathcal{L}_\mathbf{u}$ is called a *right extension* of w in $\mathbf{u}$. We denote $\mathrm{Lext}_\mathbf{u}(w)$ and $\mathrm{Rext}_\mathbf{u}(w)$ the set of all left and right extensions of w in $\mathbf{u}$, respectively. The *bilateral order* of w in $\mathbf{u}$ is defined as

$$m_\mathbf{u}(w) = \#\{(a, b) \in \mathcal{A}^2 : awb \in \mathcal{L}_\mathbf{u}\} - \#\mathrm{Lext}_\mathbf{u}(w) - \#\mathrm{Rext}_\mathbf{u}(w) + 1 .$$

Clearly, if w is not bispecial factor of $\mathbf{u}$, then $m_\mathbf{u}(w) = 0$. For bispecial factors the situation is more variable. We say that a bispecial factor w is *strong* if $m_\mathbf{u}(w) > 0$, it is *neutral (ordinary)* if $m_\mathbf{u}(w) = 0$ and it is *weak* if $m_\mathbf{u}(w) < 0$.

Example 2.5. The bispecial factor 0 of the Champernowne sequence $\mathbf{h}$ from Example 2.1 is strong, since $\mathcal{L}_\mathbf{h}(3) = \{000, 001, 010, 011, 100, 101, 110, 111\}$.

The bispecial factor 0 of the Fibonacci sequence $\mathbf{f}$ from Example 2.2 is neutral (ordinary), since $\mathcal{L}_\mathbf{f}(3) = \{001, 010, 100, 101\}$.

The bispecial factor 0 of the periodic sequence $\mathbf{u} = (001)^\omega$ is weak, since $\mathcal{L}_\mathbf{u}(3) = \{001, 010, 100\}$.

Understanding the structure of special factors is essential for the study of many properties of sequences since they often enable us to reduce the number of investigated cases substantially. We will see this in several places in this thesis, e.g., see Proposition 2.6, Remark 2.31 or Lemmas 3.35 and 3.41.

2.1.5 Factor complexity and other types of complexity

There are many ideas how to quantify the complexity of sequences. Overall, it depends on the context and the intended application which one is the best.

The simplest way of measuring the complexity of a sequence $\mathbf{u}$ is to count the number of its distinct factors of each length. More precisely, we can define the *factor complexity* of $\mathbf{u}$ as a function $\mathcal{C}_\mathbf{u} : \mathbb{N} \to \mathbb{N}$ which to each length n assigns the number of factors of $\mathbf{u}$ of length n, i.e., $\mathcal{C}_\mathbf{u}(n) = \#\mathcal{L}_\mathbf{u}(n)$.

Clearly, the factor complexity function is always non-decreasing. Nevertheless, not every non-decreasing function is the factor complexity of some sequence. Morse and Hedlund [82] showed that the factor complexity of $\mathbf{u}$ is bounded if and only if $\mathbf{u}$ is periodic and, moreover, for every aperiodic sequence $\mathbf{u}$ we have $\mathcal{C}_{\mathbf{u}}(n) \geq n+1$ for every $n \in \mathbb{N}$. There exist aperiodic sequences with this minimal factor complexity, i.e., $\mathcal{C}_{\mathbf{u}}(n) = n+1$ for every $n \in \mathbb{N}$. They are called Sturmian sequence and they are probably the most studied objects in combinatorics on words. We devote Section 2.2.1 to them.

Nevertheless, other classes of sequences with low factor complexity (i.e., $\mathcal{C}(n) = \mathcal{O}(n)$) are also studied. In this thesis we deal especially with Arnoux–Rauzy sequences (Section 2.2.2) and Rote sequences (Section 2.2.4), but in Section 2.2.3 we mention also other classes of sequences that generalize Sturmian sequences, such as sequences coding interval exchange transformations or dendric sequences.

Another important class of sequences with factor complexity $\mathcal{C}(n) = \mathcal{O}(n)$ are fixed points of primitive morphisms, as proved by Pansiot [86]. In fact, he proved that the factor complexities of fixed points generated by morphisms can only have five different asymptotic behaviours, see also [23, Section 4.7] or [5, Section 10.4] for more details.

On the other hand, the maximal factor complexity of a sequence over an alphabet $\mathcal{A}$ is $\mathcal{C}_{\mathbf{u}}(n) = (\#\mathcal{A})^n$, since this is the number of all words over $\mathcal{A}$ of length n. Such sequences also exist. For example the Champernowne sequence $\mathbf{h}$ from Example 2.1 contains every binary word and so has the complexity $\mathcal{C}_{\mathbf{h}}(n) = 2^n$.

The factor complexity of various classes of sequences has been studied (e.g., see [23, Chapter 4]). In fact, it can be easily computed using special factors.

Proposition 2.6 ([30, Propositions 3.4 and 3.5]). *Let $\mathbf{u}$ be a recurrent sequence over $\mathcal{A}$. Then the first difference of the factor complexity satisfies for every $n \in \mathbb{N}$*

$$\begin{aligned}\Delta\mathcal{C}_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n) &= \sum_{w \in \mathcal{L}_{\mathbf{u}}(n) \text{ is right special}} (\#\mathrm{Rext}_{\mathbf{u}}(w) - 1)\ , \\ &= \sum_{w \in \mathcal{L}_{\mathbf{u}}(n) \text{ is left special}} (\#\mathrm{Lext}_{\mathbf{u}}(w) - 1)\ ,\end{aligned}$$

and the second difference of the factor complexity satisfies for every $n \in \mathbb{N}$

$$\Delta^2\mathcal{C}_{\mathbf{u}}(n) = \Delta\mathcal{C}_{\mathbf{u}}(n+1) - \Delta\mathcal{C}_{\mathbf{u}}(n) = \sum_{w \in \mathcal{L}_{\mathbf{u}}(n) \text{ is bispecial}} m(w)\,,$$

where $m(w)$ is the bilateral order of w (see Section 2.1.4).

Several other variants of the complexity function have been introduced. For example, the notion of factor complexity can be modified using another equivalence relation on $\mathcal{A}^*$ instead of the usual equality of words.

Two words v and w over $\mathcal{A}$ are said to be *Abelian equivalent*, denoted $v \sim w$, if $|v|_a = |w|_a$ for all $a \in \mathcal{A}$. In other words, two words are Abelian equivalent if they are permutations of each other. It is easy to verify that $\sim$ is indeed an equivalence relation on $\mathcal{A}^*$.

Example 2.7. The words 00101 and 11000 are Abelian equivalent, while the pairs of words 00101 and 0100 or 00101 and 00111 are not Abelian equivalent.

The *Abelian complexity* of a sequence $\mathbf{u}$ is the function $\mathcal{AC}_{\mathbf{u}} : \mathbb{N} \to \mathbb{N}$ which to each $n \in \mathbb{N}$ assigns the number of pairwise Abelian non-equivalent factors of $\mathbf{u}$ of length n.

Example 2.8. Both the periodic sequence 01^{ω} and the Fibonacci sequence $\mathbf{f}$ (Example 2.2) has the Abelian complexity $\mathcal{AC}_{\mathbf{f}}(0) = 1$ and $\mathcal{AC}_{\mathbf{f}}(n) = 2$ for all $n \geq 1$.

There are several similarities between the usual factor complexity and Abelian complexity. For example, by Coven and Hedlund [39], Sturmian sequences can be characterized by means of Abelian complexity: an aperiodic binary sequence $\mathbf{u}$ is Sturmian if and only if $\mathcal{AC}_{\mathbf{u}}(n) = 2$ for all $n \geq 1$. On the other hand, the behaviour of these two complexities can differ essentially, too. Cassaigne, Ferenczi and Zamboni [37] constructed a sequence with factor complexity $\mathcal{C}(n) = 2n + 1$ and unbounded Abelian complexity. One can read more for instance in [97]. Further generalizations such as k-Abelian equivalence [68] or binomial equivalence [98] have been introduced, too.

Palindromic complexity is another type of complexity. If $w = w_0 w_1 \cdots w_{n-1}$, then the word $\tilde{w} = w_{n-1} w_{n-2} \cdots w_0$ is called the *reversal (mirror image)* of w. A word w which coincides with its reversal $\tilde{w}$ is a *palindrome*. The *palindromic complexity* of a sequence $\mathbf{u}$ is the function $\mathcal{PC}_{\mathbf{u}} : \mathbb{N} \to \mathbb{N}$ which for each $n \in \mathbb{N}$ counts the number of palindromes in $\mathbf{u}$ of length n.

Example 2.9. The Fibonacci sequence (Example 2.2) contains the palindromes

$$\varepsilon, 0, 1, 00, 010, 101, 1001, 00100, 01010, 010010, \ldots$$

Thus $\mathcal{PC}_{\mathbf{f}}(0) = 1$, $\mathcal{PC}_{\mathbf{f}}(1) = 2$, $\mathcal{PC}_{\mathbf{f}}(2) = 1$, $\mathcal{PC}_{\mathbf{f}}(3) = 2$, $\mathcal{PC}_{\mathbf{f}}(4) = 1$, etc.

An interesting relation between the palindromic and factor complexity has been revealed in [8] (see also [10, Section 3] for some notes). Let us recall that the language $\mathcal{L}_{\mathbf{u}}$ of a sequence $\mathbf{u}$ is *closed under reversal* if $\mathcal{L}_{\mathbf{u}}$ contains with every factor w also its reversal $\tilde{w}$.

Proposition 2.10 ([8, Theorem 1.2]). *Let $\mathbf{u}$ be a sequence whose language $\mathcal{L}_{\mathbf{u}}$ is closed under reversal. Then for every $n \in \mathbb{N}$ we have*

$$\mathcal{PC}_{\mathbf{u}}(n+1) + \mathcal{PC}_{\mathbf{u}}(n) \leq \Delta\mathcal{C}_{\mathbf{u}}(n) + 2\,. \tag{2.1}$$

Moreover, Bucci et al. [28] proved that among sequences with the language $\mathcal{L}_{\mathbf{u}}$ closed under reversal the sequences attaining the equality in (2.1) are exactly the so-called rich sequences. A sequence $\mathbf{u}$ is called *rich (in palindromes)* if every factor w of $\mathbf{u}$ contains $|w| + 1$ palindromes. Let us note that $|w| + 1$ is the maximal number of palindromes which w can contain.

For several classical sequences or classes of sequences the palindrome complexity is well-understood, e.g., see [1]. We just mention that a sequence $\mathbf{u}$ is Sturmian if and only if it has the palindrome complexity $\mathcal{PC}_{\mathbf{u}}(2k-1) = 2$ and $\mathcal{PC}_{\mathbf{u}}(2k) = 1$ for all $k \geq 1$, as proved by Droubay and Pirillo [48, Theorem 5].

Of course, there exist also other functions which reflect the complexity of sequences. We discuss the so-called non-repetitive complexity in Section 2.3.4.

2.2 Sequences with low factor complexity

This thesis deals with some classes of sequences with low factor complexity. We study especially Sturmian sequences, Arnoux–Rauzy sequences and Rote sequences. In this section we summarize the basic facts and needed results about these sequences.

Besides Arnoux–Rauzy and Rote sequences, we mention other generalizations of Sturmian sequences, too. In particular, we briefly present dendric and neutral sequences which have been recently introduced in [15]. They represent a natural general framework which includes many Sturmian generalizations.

2.2.1 Sturmian sequences

Sturmian sequences were first introduced by Morse and Hedlund [82] in 1940. Since then they have become the most studied objects in combinatorics on words and related fields as from many points of view they represent the simplest non-trivial case.

Let us recall that a sequence $\mathbf{u}$ is called *Sturmian* if it has $n+1$ factors of each length n, i.e., $\mathcal{C}_{\mathbf{u}}(n) = \#\mathcal{L}_{\mathbf{u}}(n) = n+1$. By the definition every Sturmian sequence is defined over a binary alphabet, e.g., $\mathcal{A} = \{0,1\}$. All Sturmian sequences are uniformly recurrent. The most famous Sturmian sequence is the Fibonacci sequence $\mathbf{f}$ introduced in Example 2.2.

There are many equivalent definitions of Sturmian sequences as well as huge amount of various results related to them. In this section we recall only several their classical characterizations and properties. More precisely, we briefly mention their special factors, palindromic properties and balancedness. We also explain their construction via coding of rotations on the unite circle and coding of two interval exchange transformations. In Section 2.3.1 we present the characterization of Sturmian sequences by their return words (Theorem 2.28). More details about these results together with many other facts about Sturmian sequences can be found for instance in [73, Chapter 2], [14] or [10].

Sturmian sequences can be describe by means of their special factors. By Proposition 2.6, a recurrent binary sequence $\mathbf{u}$ is Sturmian if and only if it has exactly one right special and one left special factor of each length. In fact, the language $\mathcal{L}_{\mathbf{u}}$ of each Sturmian sequence $\mathbf{u}$ is closed under reversal. Thus if a word w is a right special factor of $\mathbf{u}$, then its reversal $\tilde{w}$ is a left special factor of $\mathbf{u}$ and vice versa. It also means that all bispecial factor of a Sturmian sequence are palindromes.

Sturmian sequences have also other interesting palindromic properties. As we have mention at the end of Section 2.1.5, a sequence $\mathbf{u}$ is Sturmian if and only if $\mathbf{u}$ contains one palindrome of every even length and two palindromes of every odd length. Then Proposition 2.10 implies that every Sturmian sequence is rich in palindromes.

We say that a sequence $\mathbf{u}$ over $\mathcal{A}$ is *balanced* if for any pair of factors $v, w \in \mathcal{L}_{\mathbf{u}}$ of the same length the inequality $||v|_a - |w|_a| \leq 1$ holds for any letter $a \in \mathcal{A}$. Already Morse and Hedlund [82] proved that a binary sequence is Sturmian if and only if it is balanced and aperiodic. However, there exist balanced aperiodic sequences over larger alphabets, too. One can read for instance Vuillon's survey [105].

Similarly we can define slightly generalized balancedness. Let c be a positive integer. We say that a sequence $\mathbf{u}$ over $\mathcal{A}$ is *c-balanced* if for any pair of factors $v, w \in \mathcal{L}_{\mathbf{u}}$ of the same length the inequality $||v|_a - |w|_a| \leq c$ holds for any letter $a \in \mathcal{A}$. This generalized balancedness is closely related to the Abelian complexity defined in Section 2.1.5. For example, a sequence has bounded Abelian complexity if and only if it is c-balanced for some positive integer c. See for example [97] or [10, Section 5.6] for more details.

Sturmian sequences can be easily constructed as the so-called mechanical sequences. Given two real numbers α and ρ with $\alpha \in [0,1]$ and $\rho \in [0,1)$ (or $\rho \in (0,1]$), we define a *lower mechanical sequence* $\mathbf{s}_{\alpha,\rho} = s_0 s_1 s_2 \cdots$ and a *upper mechanical sequence* $\mathbf{t}_{\alpha,\rho} = t_0 t_1 t_2 \cdots$ by

$$\begin{aligned} s_n &= \lfloor \rho + \alpha(n+1) \rfloor - \lfloor \rho + \alpha n \rfloor \\ t_n &= \lceil \rho + \alpha(n+1) \rceil - \lceil \rho + \alpha n \rceil \end{aligned} \qquad \text{for all } n \in \mathbb{N}\,.$$

Clearly, mechanical sequences (except the trivial case 0^ω and 1^ω) are always binary. In addition, if α is rational, then the respective mechanical sequences are periodic, while if α is irrational, then the respective mechanical sequences are Sturmian. In fact, this characterizes Sturmian sequences.

Proposition 2.11 ([73, Theorem 2.1.13]). *A sequence is Sturmian if and only if it is (lower or upper) mechanical sequence with an irrational parameter α.*

For a Sturmian sequence $\mathbf{u} = \mathbf{s}_{\alpha,\rho}$ (or $\mathbf{u} = \mathbf{t}_{\alpha,\rho}$) the parameter α is called the *slope* of $\mathbf{u}$ and the parameter ρ is called the *intercept* of $\mathbf{u}$. In fact, the term slope can sometimes denote also the value $1 - \alpha$, but since $\mathbf{s}_{1-\alpha,\rho} = E(\mathbf{s}_{\alpha,\rho})$, where E is the morphism which exchanges the letters $0 \leftrightarrow 1$, it usually does not make a big difference.

This terminology follows from the graphic interpretation illustrated in Figure 2.1. We take the straight line $y = \alpha x + \rho$. Then the lower mechanical sequence $\mathbf{s}$ corresponds with the line which connects the points $P_n = (n, \lfloor \alpha n + \rho \rfloor)$, i.e., the points with integer coordinates which are situated just below this straight line. More precisely, the points P_n and P_{n+1} are joined by a horizontal line if $s_n = 0$ and by a diagonal line if $s_n = 1$. An analogous observation holds for the upper mechanical sequence and the points located just above the straight line, too.

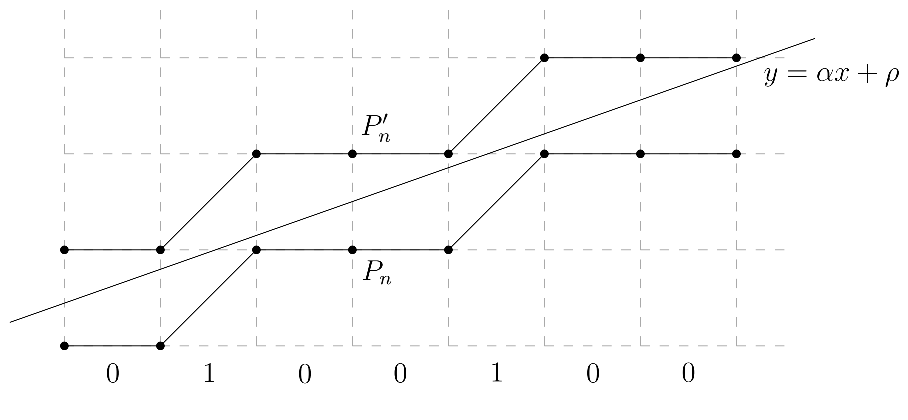

Figure 2.1: Graphic interpretation of mechanical sequences (according to [73, Fig. 2.2]).

Mechanical sequences (and so Sturmian sequences) can be also viewed as sequences coding two interval exchange transformations. It is captured in Figure 2.2. For a given parameter $\alpha \in (0, 1)$ we consider the partition of the interval

(a) $I = [0, 1)$ into $I_0 = [0, 1 - \alpha)$ and $I_1 = [1 - \alpha, 1)$;

(b) $I = (0, 1]$ into $I_0 = (0, 1 - \alpha]$ and $I_1 = (1 - \alpha, 1]$.

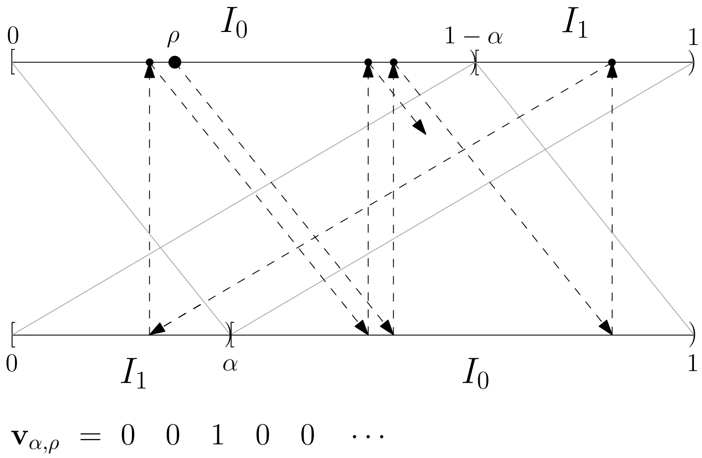

Figure 2.2: A sequence coding two interval exchange transformation.

Then the *two interval exchange transformation* is the mapping $T : I \to I$ defined by

$$T(x) = \begin{cases} x + \alpha & \text{if } x \in I_0 \,, \\ x + \alpha - 1 & \text{if } x \in I_1 \,. \end{cases}$$

If we take an initial point $\rho \in I$, the sequence $\mathbf{v}_{\alpha,\rho} = v_0 v_1 v_2 \cdots$ defined by

$$v_n = \begin{cases} 0 & \text{if } T^n(\rho) \in I_0 \,, \\ 1 & \text{if } T^n(\rho) \in I_1 \,. \end{cases}$$

is called a *2iet sequence* with the slope α and the intercept ρ.

Let us explain that 2iet sequences are just another interpretation of mechanical sequences. Indeed, if we consider the partition (a), we can rewrite the transformation T as

$$T(x) = \{x + \alpha\} \ , \text{ where } \{z\} = z - \lfloor z \rfloor \text{ denotes the fractional part of } z. \tag{2.2}$$

And since

$$\lfloor \rho + \alpha(n+1) \rfloor - \lfloor \rho + \alpha n \rfloor = 1 \quad \Longleftrightarrow \quad T^n(\rho) = \{\rho + n\alpha\} \in [1-\alpha, 1) = I_1 \,,$$

the lower mechanical sequence $\mathbf{s}_{\alpha,\rho}$ equals the 2iet sequence $\mathbf{v}_{\alpha,\rho}$. Similarly if we consider the partition (b), the respective 2iet sequences correspond with the upper mechanical sequences. Hence we may conclude that the class of Sturmian sequences is precisely the class of 2iet sequences with irrational slopes.

If we naturally identify the interval $[0, 1)$ with the unite circle, then the transformation T can be viewed as the rotation of an angle α on this circle (see Expression (2.2)). Hence if we take the partition (a) as the partition of the circle and iterate the rotation from the initial point ρ, we construct the sequence $\mathbf{v}_{\alpha,\rho}$. Clearly, we can proceed similarly with the partition (b), too. Thus the Sturmian sequences can be also constructed as sequences coding irrational rotations on the unite circle.

Let us point out that the language of a Sturmian sequence depends only on its slope α and not on its intercept ρ. It follows directly from the symbolic dynamics interpretations mentioned above. In addition, among all Sturmian sequences with a slope α, the sequence with the intercept $\rho = \alpha$ is exceptional and we call it *standard* Sturmian sequence. It can be defined also combinatorially. A Sturmian sequence $\mathbf{u}$ is standard if all its left special factors are prefixes of $\mathbf{u}$, i.e., both sequences $0\mathbf{u}$ and $1\mathbf{u}$ are Sturmian.

Example 2.12. The Fibonacci sequence $\mathbf{f}$ is standard Sturmian sequence with $\rho = \alpha = 2 - \phi$, where $\phi = (1 + \sqrt{5})/2$ denotes the golden ratio.

Standard Sturmian sequences have many comfortable properties. For example, a factor of a standard Sturmian sequence $\mathbf{u}$ is bispecial if and only if it is a palindromic prefix of $\mathbf{u}$. Hence if we study a property which depends only on the language of a sequence and not on the sequence itself, we usually work only with standard Sturmian sequences.

In this thesis we essentially use the characterization of Sturmian sequences via so-called Sturmian morphism which we introduce in Section 2.4.2.

Continued fraction expansion

Various properties of a Stumian sequence are very often expressed in terms of continued fraction expansion of its slope α. Hence we briefly recall this notion; one can find much more for instance in [60]. Let α be an irrational number. Then the *continued fraction expansion* of α is the expression of α in the form

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}\,.$$

We shortly write $\alpha = [a_0; a_1, a_2, a_3, \ldots]$.

Example 2.13. The golden ratio $\phi = (1 + \sqrt{5})/2$ has the continued fraction expansion $\phi = [1; 1, 1, \ldots] = [\bar{1}]$. Hence the slope $2 - \phi$ of the Fibonacci sequence $\mathbf{f}$ has the continued fraction expansion $2 - \phi = [0; 2, 1, 1, 1, \ldots] = [0; 2, \bar{1}]$.

By N^{th} *convergent* to the number α we mean the fraction $\frac{p_N}{q_N} = [a_0; a_1, a_2, \ldots, a_N]$. It is well known that the sequences (p_N) and (q_N) fulfil the following recurrence relations

$$\begin{aligned} p_{-1} &= 1\,,\ p_0 = a_0 && \text{and} && p_{N+1} = a_{N+1}p_N + p_{N-1} && \text{for all } N \in \mathbb{N}\,; \\ q_{-1} &= 0\,,\ q_0 = 1 && \text{and} && q_{N+1} = a_{N+1}q_N + q_{N-1} && \text{for all } N \in \mathbb{N}\,. \end{aligned}$$

Example 2.14. The convergents to the number $2 - \phi$ are

$$\frac{0}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{5}, \frac{3}{8}, \frac{5}{13}, \frac{8}{21}, \frac{13}{34}, \ldots$$

2.2.2 Arnoux–Rauzy sequences and episturmian sequences

The Arnoux–Rauzy sequences were introduced by Arnoux and Rauzy in [7] as a generalization of the Sturmian sequences to multi-letter alphabets which preserve the special factors properties.

A sequence $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is *Arnoux–Rauzy sequence* if its language $\mathcal{L}_{\mathbf{u}}$ is closed under reversal, $\mathbf{u}$ has exactly one right special factor of each length and every right special factor has $\#\mathcal{A}$ right extensions. Moreover, an Arnoux–Rauzy sequence $\mathbf{u}$ is *standard* if each of its prefixes is a left special factor.

Example 2.15. The *Tribonacci sequence* $\mathbf{t} = 0102010010201010201001\cdots$ which is the fixed point of the morphism $\varphi : 0 \to 01,\ 1 \to 02,\ 2 \to 0$ is a standard Arnoux–Rauzy sequence over $\mathcal{A} = \{0, 1, 2\}$.

More generally, we define for every $d \geq 2$ the *d-bonacci sequence* $\mathbf{t}$ as a sequence over the alphabet $\mathcal{A} = \{0, 1, \ldots, d-1\}$ which is the fixed point of the morphism

$$\varphi : a \to 0(a+1) \text{ for all } a = 0, 1, \ldots, d-2 \quad \text{and} \quad (d-1) \to 0\,.$$

The d-bonacci sequence is standard Arnoux–Rauzy sequence over $\mathcal{A}$.

If we slightly relax the requirements on special factors, we get the set of episturmian sequences. A sequence $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is *episturmian sequence* if its language $\mathcal{L}_{\mathbf{u}}$ is closed under reversal and $\mathbf{u}$ has at most one right special factor of each length. An episturmian sequence $\mathbf{u}$ is *standard* if all its left special factors are prefixes of $\mathbf{u}$.

Clearly, every Arnoux–Rauzy sequence is episturmian, but there are also aperiodic epistumian sequences which are not Arnoux–Rauzy. Hence Arnoux–Rauzy sequences over the alphabet $\mathcal{A}$ are sometimes called $\#\mathcal{A}$-strict episturmian sequences. More details about Arnoux–Rauzy sequences can be found for example in [94], while the knowledge about episturmian sequences is summarized in [58].

In the binary case, the set of Arnoux–Rauzy sequences equals the set of Sturmian sequences. Moreover, the Arnoux–Rauzy sequences (or even episturmian sequences) share many properties with the Sturmian sequences. Droubay, Justin and Pirillo [47] proved that any episturmian sequence is uniformly recurrent. In addition, the Arnoux–Rauzy sequences are always aperiodic since they have the factor complexity

$$\mathcal{C}_{\mathbf{u}}(n) = (\#\mathcal{A} - 1)n + 1 \quad \text{for every } n \in \mathbb{N}\,.$$

Indeed, it follows directly from Proposition 2.6. Nevertheless, there are also other sequences with the same factor complexity, e.g., the sequences coding interval exchange transformations or, more generally, the dendric sequences (see Section 2.2.3).

All bispecial factors of an episturmian sequence $\mathbf{u}$ are palindromes. Let us emphasize that for every $n \in \mathbb{N}$ there is at most one bispecial factor of length n. Hence we can order the bispecial factors of $\mathbf{u}$ by their lengths starting with the empty word ε.

Like in the Sturmian case, for every episturmian sequence there exists a unique standard episturmian sequence with the same language. Hence if we study properties which depend only on the language of a sequence, we can deal only with standard episturmian sequences without loss of generality.

On the other hand, there are also properties in which Sturmian and Arnoux–Rauzy sequences differ essentially. For example Cassaigne, Ferenczi and Zamboni [37] constructed an Arnoux–Rauzy sequence which is not c-balanced for any positive integer c.

Standard episturmian sequences can be constructed by the co-called palindromic closures. The *palindromic closure* $w^{(+)}$ of a finite word w is the shortest palindrome having w as a prefix. A sequence $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is standard episturmian if and only if there exist a directive sequence $\mathbf{d} = d_0 d_1 d_2 \cdots \in \mathcal{A}^{\mathbb{N}}$ and a sequence of palindromes $u^{(0)} = \varepsilon, u^{(1)}, u^{(2)}, \ldots$ such that $u^{(n+1)} = (u^{(n)} d_n)^{(+)}$ for all $n \in \mathbb{N}$ and $\mathbf{u} = \lim_{n\to\infty} u^{(n)}$.

Example 2.16. We consider the directive sequence $\mathbf{d} = (012)^\omega$ and we gradually construct the respective sequence of palindromes:

$$\begin{aligned}
u^{(0)} &= \varepsilon \\
u^{(1)} &= (0)^{(+)} = 0 \\
u^{(2)} &= (0\,1)^{(+)} = 010 \\
u^{(3)} &= (010\,2)^{(+)} = 0102010 \\
u^{(4)} &= (0102010\,0)^{(+)} = 01020100102010 \\
u^{(5)} &= (01020100102010\,1)^{(+)} = 0102010010201010 \\
u^{(6)} &= (0102010010201010\,2)^{(+)} = 01020100102010102010 0102010 \\
&\text{etc.}
\end{aligned}$$

The limit of this sequence of palindromes is the Tribonacci sequence (Example 2.15). Similarly the directive sequence of the d-bonacci sequence is $(012\cdots(d-1))^\omega$.

Let us emphasize that a standard episturmian sequence is completely determined by its directive sequence. The standard Arnoux–Rauzy sequences over $\mathcal{A}$ are exactly those episturmian sequences whose directive sequences contain every letter from $\mathcal{A}$ infinitely many times. In Section 2.4.2 we generalize this notion of directive sequences also for non-standard episturmian sequences using the so-called episturmian morphisms.

2.2.3 Dendric and neutral sequences

Dendric sequences represent a very general class of sequences which contains several generalizations of Sturmian sequences to multi-letter alphabets. They were introduced by Berthé et. al. in [15]. In fact, they define dendric (tree) sets since they work with languages rather than with sequences. However, for our purposes it does not make any difference.

Let $\mathbf{u}$ be a sequence and w be its factor. The *extension graph* of w in $\mathbf{u}$ is an undirected bipartite graph $\mathcal{E}_\mathbf{u}(w)$ defined as follows: its vertices are the letters from the set $\mathrm{Lext}_\mathbf{u}(w)$ and the letters from the set $\mathrm{Rext}_\mathbf{u}(w)$, and there is an edge between the vertices $a \in \mathrm{Lext}(w)$ and $b \in \mathrm{Rext}(w)$ if $awb \in \mathcal{L}_\mathbf{u}$.

Example 2.17. We consider the Tribonacci sequence $\mathbf{t} = 0102010010201010201001\cdots$ from Example 2.15. The extension graphs $\mathcal{E}_\mathbf{t}(\varepsilon)$, $\mathcal{E}_\mathbf{t}(2)$, $\mathcal{E}_\mathbf{t}(01)$ and $\mathcal{E}_\mathbf{t}(10)$ are displayed in Figure 2.3.

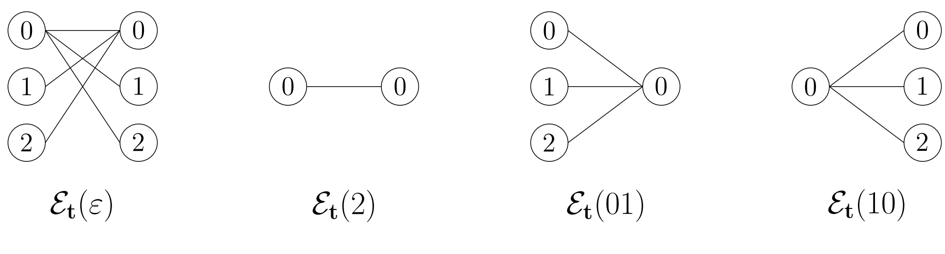

Figure 2.3: The extension graphs $\mathcal{E}_\mathbf{t}(\varepsilon)$, $\mathcal{E}_\mathbf{t}(2)$, $\mathcal{E}_\mathbf{t}(01)$ and $\mathcal{E}_\mathbf{t}(10)$ of the Tribonacci sequence $\mathbf{t}$.

A sequence $\mathbf{u}$ is *dendric* if for every $w \in \mathcal{L}_{\mathbf{u}}$ the extension graph $\mathcal{E}_{\mathbf{u}}(w)$ is a tree (i.e., it is both acyclic and connected).

Example 2.18. The Tribonacci sequence is dendric sequence since one can easily verify that each of its extensions graphs has one of the shapes displayed in Figure 2.3.

In fact, all Arnoux–Rauzy sequences (and so all Sturmian sequences) are dendric. Moreover, the sequences coding regular interval exchange transformations are also dendric, see [16]. These sequences were introduced in [85]. Since then they have been intensively studied as they represent another important generalization of Sturmian sequences (see 2iet sequences defined in Section 2.2.1). We skip their general definition (e.g., see [10, Section 5.1]), but we briefly indicate how the sequences coding three interval exchange transformations look like.

A *three interval exchange transformation* $T : [0, 1) \to [0, 1)$ is given by two parameters $\beta, \gamma \in (0, 1)$ such that $\beta + \gamma < 1$, and by a permutation π on the set $\{1, 2, 3\}$. The interval $[0, 1)$ is partitioned into three subintervals

$$I_A = [0, \beta)\,, \quad I_B = [\beta, \beta + \gamma) \quad \text{and} \quad I_C = [\beta + \gamma, 1)\,.$$

These subintervals are then rearranged by the transformation T according to the permutation π. For example, if we take the permutation $\pi = (3, 2, 1)$, then

$$T(x) = \begin{cases} x + 1 - \beta & \text{if } x \in I_A\,, \\ x + 1 - 2\beta - \gamma & \text{if } x \in I_B\,, \\ x - \beta - \gamma & \text{if } x \in I_C\,. \end{cases}$$

If we take another permutation $\pi = (2, 3, 1)$, then

$$T(x) = \begin{cases} x + 1 - \beta & \text{if } x \in I_A\,, \\ x - \beta & \text{if } x \in I_B\,, \\ x - \beta & \text{if } x \in I_C\,. \end{cases}$$

We can proceed similarly for other permutations. Let $\rho \in [0, 1)$. The sequence $\mathbf{u} = u_0u_1u_2 \cdots \in \{A, B, C\}^{\mathbb{N}}$ defined by

$$u_n = \begin{cases} A & \text{if } T^n(\rho) \in I_A\,, \\ B & \text{if } T^n(\rho) \in I_B\,, \\ C & \text{if } T^n(\rho) \in I_C \end{cases}$$

is called a *3iet sequence* coding the intercept ρ under the transformation T.

Dendric sequences can be further generalized, too. For example, we say that a sequence $\mathbf{u}$ is *neutral* if all its non-empty bispecial factors are neutral (see Section 2.1.4). The *characteristic* of a neutral sequence $\mathbf{u}$ is the integer $\chi(\mathbf{u}) = 1 - m_{\mathbf{u}}(\varepsilon)$, where $m_{\mathbf{u}}(\varepsilon)$ is the bilateral order of the empty word ε in $\mathbf{u}$. These sequences are studied by Dolce and Perrin in [45]. Clearly, every dendric sequence is a neutral sequence with the characteristic 1 since all its bispecial factors (including the empty word) are neutral. Other examples of neutral sequences are presented in Section 2.2.4.

Recently, dendric and neutral sequences have been intensively studied, one can read for example [15–18, 20, 21, 45, 46]. We just mention that all neutral sequences have low factor complexity.

Proposition 2.19 ([45, Proposition 2.4]). *The factor complexity of a neutral sequence* $\mathbf{u}$ *over* $\mathcal{A}$ *with the characteristic* $\chi(\mathbf{u})$ *is given by*

$$\mathcal{C}_{\mathbf{u}}(0) = 1 \quad \textit{and} \quad \mathcal{C}_{\mathbf{u}}(n) = n(\#\mathcal{A} - \chi(\mathbf{u})) + \chi(\mathbf{u}) \quad \textit{for every } n \geq 1\,.$$

2.2.4 Rote sequences

Rote sequences are another class of sequences with low factor complexity. A sequence $\mathbf{v}$ is called *Rote sequence* if it has the factor complexity $\mathcal{C}_{\mathbf{v}}(n) = 2n$ for every integer $n \geq 1$. By definition, all Rote sequences are defined over a binary alphabet, e.g., $\mathcal{A} = \{0, 1\}$.

They are named after Rote who proposed several construction of these sequences in [100]. He showed how one can in principal construct all Rote sequences via their Rauzy graphs (see [100, Section 2]) and for special subclasses he gave also alternative methods of construction. In particular, he proved that some of Rote sequences can be viewed as sequences coding irrational rotations on the unite circle. Let us recall that Sturmian sequences are also sequences coding irrational rotations (see Section 2.2.1).

Proposition 2.20 ([100, Theorem 2]). *Let ρ, α and β be real numbers with $0 < \beta < 1$, $0 < \alpha < \min\{\beta, 1-\beta\}$, α irrational and $n\alpha \neq \beta \mod 1$ for all $n \in \mathbb{N}$. Then the sequence $\mathbf{v} = v_0 v_1 v_2 \cdots$ defined by*

$$v_n = \begin{cases} 0 & \text{if } \{\rho + n\alpha\} \in [0, \beta) \\ 1 & \text{if } \{\rho + n\alpha\} \in [\beta, 1) \end{cases} \qquad \text{for all } n \in \mathbb{N}$$

is a Rote sequence.

However, there are also Rote sequences which cannot be generated by this way.

Example 2.21 ([100, Sections 3.1 and 3.2]). We consider the morphisms

$$\theta : \begin{cases} 0 \to 03 \\ 1 \to 12 \\ 2 \to 0312031 \\ 3 \to 1203120 \end{cases}, \qquad \kappa : \begin{cases} 0 \to 0 \\ 1 \to 1 \\ 2 \to 0 \\ 3 \to 1 \end{cases}, \qquad \psi : \begin{cases} 0 \to 03 \\ 1 \to 102 \\ 2 \to 10201 \\ 3 \to 1020 \end{cases} \quad \text{and} \quad \tau : \begin{cases} 0 \to 0 \\ 1 \to 1 \\ 2 \to 10110 \\ 3 \to 101 \end{cases}.$$

Both sequences $\mathbf{u} = \kappa(\theta^\omega(0))$ and $\mathbf{v} = \tau(\psi^\omega(0))$ are Rote sequences. Nevertheless, while the sequence $\mathbf{u}$ can be constructed by Proposition 2.20, the sequence $\mathbf{v}$ cannot be obtained by this process; see [100] for proofs and other details.

The given examples also illustrate that there exist primitive substitutive Rote sequences. Moreover, we can construct also a Rote sequence which is the fixed point of a non-identical morphism.

Example 2.22 ([100, Section 3.3]). The morphism $\theta : 0 \to 001, 1 \to 111$ fixes the Rote sequence $\mathbf{v} = 001001111001001111111111111001 \cdots$. However, let us notice that this morphism θ is not primitive and the sequence $\mathbf{v}$ is not uniformly recurrent.

Complementary symmetric Rote sequences

We are especially interested in the so-called complementary symmetric (CS) Rote sequences. They are Rote sequences whose languages are closed under the exchange of letters $0 \leftrightarrow 1$. More precisely, the binary sequence $\mathbf{u}$ is *complementary symmetric (CS)* if it contains with each factor w also the factor $E(w)$, where $E : 0 \to 1, 1 \to 0$ is the exchange morphism.

Rote [100] proved that these sequences are essentially connected with Sturmian sequences.

Theorem 2.23 ([100, Theorem 3]). *Let* $\mathbf{u} = u_0u_1\cdots$ *and* $\mathbf{v} = v_0v_1\cdots$ *be two sequences over* $\{0,1\}$ *such that* $u_i = v_i + v_{i+1} \mod 2$ *for all* $i \in \mathbb{N}$. *Then* $\mathbf{v}$ *is a CS Rote sequence if and only if* $\mathbf{u}$ *is a Sturmian sequence.*

This theorem indicates the usefulness of the following notation. By $\mathcal{S}$ we denote the mapping $\mathcal{S} : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ such that for every $\mathbf{v} \in \{0,1\}^{\mathbb{N}}$ we put $\mathcal{S}(\mathbf{v}) = \mathbf{u}$, where

$$u_i = v_i + v_{i+1} \mod 2 \quad \text{for all} \quad i \in \mathbb{N}\,. \tag{2.3}$$

In this notation we can reformulate Theorem 2.23:

A sequence $\mathbf{v}$ is a CS Rote sequence if and only if $\mathbf{u} = \mathcal{S}(\mathbf{v})$ is a Sturmian sequence.

We usually say that such a sequence $\mathbf{u}$ is *associated* with $\mathbf{v}$ and vice versa.

Example 2.24. If we start with the letter 0 and repeatedly use Equation (2.3) rewritten to the form $v_{i+1} = u_i + v_i \mod 2$, we can construct a CS Rote sequence $\mathbf{g}$ associated with the Fibonacci sequence $\mathbf{f}$, i.e., $\mathbf{f} = \mathcal{S}(\mathbf{g})$. We have

$$\begin{aligned} \mathbf{f} &= 0100101001001010010100100101001001\cdots \\ \mathbf{g} &= 0011100111000110001100011100111000 1\cdots \end{aligned}$$

In fact, the sequence $E(\mathbf{g}) = 110001100011100\cdots$ is associated with $\mathbf{f}$, too.

As in Example 2.24, any Sturmian sequence $\mathbf{u}$ has two associated CS Rote sequences $\mathbf{v}$ and $E(\mathbf{v})$ such that $\mathbf{u} = \mathcal{S}(\mathbf{v}) = \mathcal{S}(E(\mathbf{v}))$. However, since the exchange of letters does not affect the structure of the sequence, we usually consider only the CS Rote sequences starting with the letter 0 without loss of generality.

Similarly we can define this mapping $\mathcal{S}$ also for finite non-empty words. For every $v_0 \in \{0,1\}$ we put $\mathcal{S}(v_0) = \varepsilon$ and for every $v = v_0v_1\cdots v_n \in \{0,1\}^+$ of length at least 2 we put $\mathcal{S}(v_0v_1\cdots v_n) = u_0u_1\cdots u_{n-1}$, where

$$u_i = v_i + v_{i+1} \mod 2 \quad \text{for all} \quad i \in \{0,1,\ldots,n-1\}\,.$$

Clearly, the images of v and $E(v)$ under $\mathcal{S}$ coincide for each $v \in \{0,1\}^+$. In fact, $\mathcal{S}(x) = \mathcal{S}(y)$ if and only if $x = y$ or $x = E(y)$.

Example 2.25. We have $E(001110) = 110001$ and $\mathcal{S}(001110) = \mathcal{S}(110001) = 01001$.

Then we can comfortably write the relations between the factors of associated sequences $\mathbf{v}$ and $\mathbf{u} = \mathcal{S}(\mathbf{v})$.

Proposition 2.26. *Let* $\mathbf{v}$ *be a Rote sequence and* $\mathbf{u} = \mathcal{S}(\mathbf{v})$ *be its associated Sturmian sequence.*

(i) A word $v \neq \varepsilon$ *is a factor of* $\mathbf{u}$ *if and only if* $u = \mathcal{S}(v)$ *is a factor of* $\mathbf{u}$. *Moreover, for every* $m \in \mathbb{N}$, *the index* m *is an occurrence of* u *in* $\mathbf{u}$ *if and only if* m *is an occurrence of* v *in* $\mathbf{v}$ *or an occurrence of* $E(v)$ *in* $\mathbf{v}$.

(ii) A word $v \neq \varepsilon$ *is a left (right) special factor of* $\mathbf{v}$ *if and only if* $u = \mathcal{S}(v)$ *is a left (right) special factor of* $\mathbf{u}$.

(iii) *Each non-empty bispecial factor of* $\mathbf{v}$ *is neutral and the empty word is a strong bispecial factor of* $\mathbf{v}$.

Finally, we emphasize that CS Rote sequences can be defined in other ways, too. The class of CS Rote sequences is exactly the class of sequences constructed by Proposition 2.20 with $\beta = \frac{1}{2}$. In particular, by results of Blodin-Massé et al. [25] this means that CS Rote sequences are rich in palindromes. This observation also follows from the note stated after Proposition 2.10 since Allouche et al. [1] proved that every CS Rote sequence $\mathbf{u}$ has the palindromic complexity $\mathcal{PC}_{\mathbf{u}}(n) = 2$ for all $n \geq 1$. CS Rote sequences can be also constructed using the so-called pseudopalindromic closures, see [26, Section 6]. In more general setting, they represent an interesting example of neutral sequences with the characteristic 0 (see Item (iii) of Proposition 2.26). Hence they are not dendric.

2.3 Main subjects of our research

In this section we briefly recall several properties of sequences which we are interested in and we summarize the relevant known results about them. First we focus on the notion of return words and derived sequences since it creates a key part of this thesis. Then we introduce also the critical exponent, the recurrence function and the non-repetitive complexity.

2.3.1 Return words and derived sequences

Return words

Return words are well established notion in combinatorics on words. To this field they were first introduced by Durand [49], however, they can be seen as a kind of analogue to the first return map occurring in the theory of dynamical systems (e.g., see [24]).

First, let us recall the definition. We consider a sequence $\mathbf{u}$ and its factor w. Whenever $i < j$ are two consecutive occurrences of w in $\mathbf{u}$, then the string $u_i u_{i+1} \cdots u_{j-1}$ is a return word to w in $\mathbf{u}$. In other words, a *return word* to w in $\mathbf{u}$ is every factor r such that $rw \in \mathcal{L}_{\mathbf{u}}$ and w occurs exactly twice in the word rw: both as a prefix and a suffix. We denote $\mathcal{R}_{\mathbf{u}}(w)$ the set of all return words to w in $\mathbf{u}$.

Example 2.27. We consider the Fibonacci sequence $\mathbf{f} = 0100101001001010010100\cdots$ from Example 2.2. Its prefix 0 has two return words 01 and 0, hence $\mathcal{R}_{\mathbf{f}}(0) = \{01, 0\}$. Its prefix 0100 has two return words 01001 and 010, hence $\mathcal{R}_{\mathbf{f}}(0100) = \{01001, 010\}$.

In fact, the notion of return words make sense especially for recurrent sequences, otherwise there are factors with no return words. Moreover, we usually consider only uniformly recurrent sequences since each of their factors has the finite number of return words. That is, a recurrent sequence $\mathbf{u}$ is uniformly recurrent if and only if the set $\mathcal{R}_{\mathbf{u}}(w)$ is finite for each factor w of $\mathbf{u}$.

First let us observe that periodic sequences can be easily characterized by means of the return words. Indeed, a recurrent sequence $\mathbf{u}$ is periodic if and only if there exists a factor of $\mathbf{u}$ having only one return word (see [104, Proposition 3.1]).

Similarly, Vuillon [104] showed that Sturmian sequences can be also easily characterized by their return words.

Theorem 2.28 ([104, Main Theorem]). *A binary sequence* $\mathbf{u}$ *is Sturmian if and only if the set of return words to* w *has exactly two elements for every non-empty factor* w *of* $\mathbf{u}$*.*

In fact, if we consider also return words to the empty word ε, which are by definition all distinct letters from $\mathbf{u}$, we can exclude the binary assumption: A sequence $\mathbf{u}$ is Sturmian if and only if each of its factors has two return words.

This characterization inspired the following generalization of Sturmian words to m-letter alphabets described in [11, 106]. We say that the recurrent sequence has the property R_m if each of its factors has exactly m return words. The class of all sequences with R_m covers some other generalizations of Sturmian words. Justin and Vuillon [65] proved that Arnoux–Rauzy sequences over $\#\mathcal{A} = m$ satisfy R_m, Vuillon [106] proved this property for sequences coding regular m-interval exchange transformations.

It is worth to notice that all sequences from both these classes have the factor complexity $\mathcal{C}(n) = (m-1)n + 1$ for all $n \geq 0$. However, unlike the binary case, Vuillon [106] observed that this condition is not sufficient for sequences to satisfy R_m for $m \geq 3$. He found a sequence with the complexity $2n + 1$ which has some factors with more than three return words (see Section 3 in [106]). Balková, Pelantová, and Steiner [11] characterized the sequences with R_3.

Proposition 2.29 ([11, Theorem 5.7]). *A uniformly recurrent sequence* $\mathbf{u}$ *has three return words to each of its factors if and only if* $\mathcal{C}_{\mathbf{u}}(n) = 2n + 1$ *for all* $n \geq 0$ *and* $\mathbf{u}$ *has no weak bispecial factor.*

In addition, they showed that analogous conditions are sufficient, but not necessary if $m \geq 4$. More precisely, if a uniformly recurrent sequence $\mathbf{u}$ has the complexity $\mathcal{C}_{\mathbf{u}}(n) = (m-1)n + 1$ for all $n \geq 0$ and $\mathbf{u}$ has no weak bispecial factors, then $\mathbf{u}$ has m return words to each of its factors. Let us note that in the notion of [45] these sequences are neutral of characteristic 1 (see Section 2.2.3). On the other hand, they constructed a sequence which fulfils R_4 while it has different factor complexity and it contains weak bispecial factors [11, Proposition 6.1].

Return words were studied also for more general classes of sequences. Berthé et al. [20] studied return words in sequences coding linear involutions which generalize sequences coding interval exchange transformations. Even more generally, Dolce and Perrin [45] determined the number of return words for neutral sets of any characteristic. In particular, for a uniformly recurrent neutral sequence $\mathbf{u}$ their result implies that the number of return words to non-empty factors of $\mathbf{u}$ is constant:

Proposition 2.30 ([45, Corollary 5.4]). *Let* $\mathbf{u}$ *be a uniformly recurrent neutral sequence over* $\mathcal{A}$ *of characteristic* $\chi(\mathbf{u})$*. For any non-empty* $w \in \mathcal{L}_{\mathbf{u}}$*, the set* $\mathcal{R}_{\mathbf{u}}(w)$ *has* $\#\mathcal{A} - \chi(\mathbf{u}) + 1$ *elements.*

Nevertheless, they are known also some classes of sequences whose factors do not have constant number of return words. Justin and Vuillon [65] explicitly described return words in episturmian sequences which generalize the Arnoux–Rauzy sequences. It directly follows from [65, Theorem 4.4] that every episturmian sequence which is not k-strict for any k has factors with at least two different numbers of return words.

Balková, Pelantová and Steiner [11] studied a class of sequences associated to β-numeration systems. This class can be viewed as a generalization of Arnoux–Rauzy sequences, too. They identified the basis β for which the associated sequence has constant number of return words, see [11, Section 7].

Finally, we mention two results on particular sequences. As follows from Huang and Wen [64] the period doubling sequence, which is the fixed point of the primitive morphism $0 \to 01$, $1 \to 00$, has to each of its factors two or three return words. The Thue–Morse sequence, which is fixed by the primitive morphism $0 \to 01$, $1 \to 10$, has to each of its factors two (only to the empty word), three or four return words, see e.g. [35, Section 6].

Remark 2.31. Let us mention that to describe return words in aperiodic uniformly recurrent sequence $\mathbf{u}$, it basically suffices to investigate only return words to bispecial factors of $\mathbf{u}$. Indeed, in such a sequence $\mathbf{u}$ each factor w has the unique shortest bispecial factor containing w, we denote it $v = pws$, where $p, s \in \mathcal{A}^*$ (e.g., see [27, Proposition 5]). Then it is easy to realize that

$$\mathcal{R}_{\mathbf{u}}(w) = p^{-1}\mathcal{R}_{\mathbf{u}}(v)p = \{p^{-1}rp : r \in \mathcal{R}_{\mathbf{u}}(v)\}\,.$$

Other details can be found for example in [11].

Derived sequences

If $\mathbf{u}$ is a uniformly recurrent sequence and w is its factor, then the finite set $\mathcal{R}_{\mathbf{u}}(w) = \{r_0, r_1, \ldots, r_{k-1}\}$ creates a code. In particular, if w is a prefix of $\mathbf{u}$, then $\mathbf{u}$ can be written uniquely in the form

$$\mathbf{u} = r_{d_0} r_{d_1} r_{d_2} \cdots\,, \text{ where all } r_{d_j} \in \mathcal{R}_{\mathbf{u}}(w)\,, \text{ i.e., all } \ d_j \in \{0, 1, \ldots, k-1\}\,.$$

Similarly, if w is not a prefix of $\mathbf{u}$, i.e., the first occurrences of w in $\mathbf{u}$ is $i \geq 1$, then $\mathbf{u}$ can be written as

$$\mathbf{u} = u_0 u_1 \cdots u_{i-1} r_{d_0} r_{d_1} r_{d_2} \cdots\,, \text{ where all } r_{d_j} \in \mathcal{R}_{\mathbf{u}}(w)\,, \text{ i.e., all } \ d_j \in \{0, 1, \ldots, k-1\}\,,$$

in other words, $\mathbf{u}$ can be uniquely decoded to return words except for some finite prefix.

In both cases, the unique sequence $d_0 d_1 d_2 \cdots$ defined over the alphabet of cardinality $\#\mathcal{R}_{\mathbf{u}}(w)$ codes the order of return words in $\mathbf{u}$. It is called the *derived sequence* of $\mathbf{u}$ with respect to w and denoted $\mathbf{d}_{\mathbf{u}}(w)$.

Example 2.32 (Example 2.27 continued). Since the prefix 0 of the Fibonacci sequence $\mathbf{f}$ has the return words 01 and 0, the sequence $\mathbf{f}$ can be written as a concatenation of the words $r_0 = 01$ and $r_1 = 0$:

$$\mathbf{f} = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01\ 01\ 0\ 01\ 01\ 0\ \ = r_0 r_1 r_0 r_0 r_1 r_0 r_1 r_0 r_0 r_1 r_0 r_0 r_1 \cdots\,.$$

Hence the derived sequence of $\mathbf{f}$ with respect to 0 is

$$\mathbf{d}_{\mathbf{f}}(0) = 0100101001001 \cdots = \mathbf{f}\,.$$

In fact, any derived sequence of $\mathbf{f}$ with respect to its prefix is equal (up to exchange of letters $0 \leftrightarrow 1$ which make no difference) to the Fibonacci sequence $\mathbf{f}$. In other words, the Fibonacci sequence has only one derived sequence and it is the Fibonacci sequence itself.

Let us emphasize that, unlike the return words, derived sequence depends on the original sequence and not only on its language.

For w being a non-empty prefix, these sequences were introduced by Durand [49] and, independently, for a general factor w they were investigated by Holton and Zamboni [61]. In this thesis we follow the Durand's notion and we consider only derived sequences to prefixes.

When studying derived sequences, the usual aim is to describe the set $\mathrm{Der}(\mathbf{u})$ of all derived sequences of a given uniformly recurrent sequence $\mathbf{u}$:

$$\mathrm{Der}(\mathbf{u}) = \{\mathbf{d_u}(w) : w \text{ is a non-empty prefix of } \mathbf{u}\}\,.$$

In fact, it suffices to investigate only derived sequences to right special prefixes. The reason is the same as in Remark 2.31. By the definition, the return words to the empty word ε are all letters in $\mathbf{u}$ and so the corresponding derived sequence is $\mathbf{d_u}(\varepsilon) = \mathbf{u}$ for every sequence $\mathbf{u}$. Hence we omit the empty prefix.

It is also worth to realize that the sequence $\mathbf{u}$ is a morphic image of each of its derived sequences, i.e., for every prefix w there exists a morphism τ such that $\mathbf{u} = \tau(\mathbf{d_u}(w))$. In fact, τ is just the inverse morphism to the coding of return words which creates the derived sequence.

Let us mention one more Durand's simple observation.

Proposition 2.33 ([49, Proposition 2.6]). *Let $\mathbf{u}$ be a uniformly recurrent sequence, w be its non-empty prefix, $\mathbf{v} = \mathbf{d_u}(w)$ and v be a non-empty prefix of $\mathbf{v}$. Then the derived sequence $\mathbf{d_v}(v)$ is also the derived sequence of $\mathbf{u}$ with respect to some non-empty prefix u of $\mathbf{u}$, i.e., $\mathbf{d_v}(v) = \mathbf{d_u}(u)$.*

The derived sequences were so far studied especially for Sturmian sequences and their generalization. Araújo and Bruyère [6] precisely described derived sequences of standard Sturmian sequences in terms of continued fraction expansions of their slopes, see [6, Proposition 15]. Furthermore, the description of derived sequences of all standard episturmian sequences can be easily deduced from the work of Justin and Vuillon [65, Corollary 4.1 and Theorem 4.4]. They use the very comfortable notion of directive sequences and episturmian morphisms.

Huang and Wen studied the derived sequences to all factors in the Fibonacci sequence [62], Tribonacci sequence [63] and period doubling sequence [64]. In the case of the Fibonacci and Tribonacci sequences, they found out that each of its derived sequences is Fibonacci and Tribonacci sequence, respectively. In fact, this can be trivially deduced from much earlier work of Justin and Vuillon [65, Corollary 4.1 and Theorem 4.4].

Characterization of primitive substitutive sequences

Both Durand [49] and Holton and Zamboni [61] introduced derived sequences when studying substitutive sequences. Let us recall that a sequence $\mathbf{u}$ is called substitutive if $\mathbf{u} = \kappa(\theta^\omega(a))$ for some 1-uniform morphism κ and some substitution θ. Moreover, if θ is primitive, we call such a sequence $\mathbf{u}$ primitive substitutive. These sequences are always uniformly recurrent (e.g., see [5, Theorem 10.9.5]).

Let us emphasize that, in general, it can be hard to decide if a given sequence is substitutive or not. However, Durand [49] managed to characterize primitive substitutive sequences combinatorially using derived sequences.

Theorem 2.34 ([49, Theorem 2.5]). *A uniformly recurrent sequence* $\mathbf{u}$ *is primitive substitutive if and only if the number of its distinct derived sequences is finite.*

Moreover, all derived sequences of a primitive substitutive sequence are also primitive substitutive, and especially, all derived sequences of a fixed point of a primitive morphism are also fixed points of some primitive morphisms. The finiteness of $\mathrm{Der}(\mathbf{u})$ in this special case follows also from [61].

In fact, Theorem 2.34 works for every uniformly recurrent sequence of the form $\mathbf{u} = \tau(\psi^\omega(a))$, where τ and ψ are arbitrary morphisms. This assertion is a consequence of [52, Theorem 3], where Durand proved that every such a sequence is primitive substitutive.

Note on applications and generalizations of return words and derived sequences

Return words are now an integral part of combinatorics on words and it is almost impossible to mention all their usefulness and applications. And although the notion of derived sequences is probably not so well-known, they are also naturally present in many works. Hence we stated only several recent results which are somehow related to our objects. However, this list is still far from being complete.

Araújo and Bruyère [6] applied their description of Sturmian sequences to obtain a new proof of some characterization of Sturmian sequences. Blodin-Massé et al. [24] use the description of return words in sequences coding rotations to show that they are rich. Bucci and De Luca [27] use this notion to show that some kind of generalizations of Arnoux–Rauzy sequences which seem essential are just morphic images of Arnoux–Rauzy sequences. Berthé et al. [15] and Berthé et al. [21] exploited the properties of return words and derived sequences for studying various properties of dendric sequences and also for a characterization of substitutive dendric sequences. Among many other things, Durand [52] used this notion to prove that the uniform recurrence of substitutive sequences is decidable. In the same paper he also considers return words with respect to a set of factors of a sequence $\mathbf{u}$ (instead of to one single factor of $\mathbf{u}$).

Finally, let us mention that also Abelian variants of return words can be considered. We recall that two words v and w over $\mathcal{A}$ are Abelian equivalent if $|v|_a = |w|_a$ for all $a \in \mathcal{A}$ (see Section 2.1.5). All words which are Abelian equivalent to w create the *Abelian class* of w. For a recurrent sequence $\mathbf{u}$ and its factor w, let $n_0 < n_1 < n_2 < \cdots$ be all integers n_i such that the factor $u_{n_i} \cdots u_{n_i+|w|-1}$ is Abelian equivalent to w. Then each factor $w_{n_i} \cdots w_{n_{i+1}-1}$ is called a *semi-Abelian return* to the Abelian class of w. By an *Abelian return* to the Abelian class of w we mean an Abelian class of $w_{n_i} \cdots w_{n_{i+1}-1}$. Hence the number of Abelian returns is the number of distinct Abelian classes of semi-Abelian returns.

Example 2.35 ([89, Example 5]). We consider the Thue–Morse sequence $\mathbf{m}$ which is fixed by the morphism $0 \to 01$, $1 \to 10$.

The Abelian class of the word 01 is $\{01, 10\}$. If we mark the occurrences of words 01 and 10 in $\mathbf{m}$ by dots:

$$\mathbf{m} = \underset{\cdot}{0}1\underset{\cdot}{1}\underset{\cdot}{0}\underset{\cdot}{1}0\underset{\cdot}{0}1\underset{\cdot}{1}0\underset{\cdot}{0}\underset{\cdot}{1}\underset{\cdot}{0}1\underset{\cdot}{1}\underset{\cdot}{0}\underset{\cdot}{1}0\underset{\cdot}{0}\underset{\cdot}{1}\underset{\cdot}{0}1\underset{\cdot}{1}0\underset{\cdot}{0}1\underset{\cdot}{1}\underset{\cdot}{0}\underset{\cdot}{1}0\underset{\cdot}{0}1 \cdots ,$$

we can see that the Abelian class $\{01, 10\}$ has four semi-Abelian returns: 0, 01, 1 and 10. Since the words 01 and 10 are Abelian equivalent, the class $\{01, 10\}$ has three Abelian returns: $\{0\}$, $\{1\}$ and $\{01, 10\}$.

This notation was introduced by Puzynina and Zamboni in [89]. In this paper they study (semi-)Abelian return in Sturmian sequences. Among other things, they revealed two new characterizations of Sturmian sequences.

Proposition 2.36 ([89, Theorems 2 and 3]).

(i) A binary recurrent sequence $\mathbf{u}$ *is Sturmian if and only if the Abelian class of each factor* w *of* $\mathbf{u}$ *has two or three Abelian returns in* $\mathbf{u}$.

(ii) A binary recurrent sequence $\mathbf{u}$ *is Sturmian if and only if the Abelian class of each factor* w *of* $\mathbf{u}$ *has two or three semi-Abelian returns in* $\mathbf{u}$.

The Abelian returns of Sturmian sequences were examined also by Rigo, Salimov and Vandomme [99]. They also introduced the notion of Abelian derived sequences and they indicated that their role differs from the classical one essentially. More precisely, they showed that the Thue–Morse sequence (which is fixed by a primitive morphism), has infinitely many Abelian derived sequences [99, Proposition 38]. Other related results can be found in [76, 90].

2.3.2 Critical exponent

Roughly speaking, the critical exponent reflects the length of the longest repetition in a given sequence. By repetition of a non-empty factor w we mean every word of the form $z = ww \cdots ww'$, where w' is a proper prefix of w. Then we say that z has the *fractional root* w and the *exponent* $e = |z|/|w|$. We also write $z = w^e$ and z is called an *e-power* of w. Let us emphasize that a word z can have multiple exponents and fractional roots. The word z is *primitive* if the only integer exponent of z is 1.

Example 2.37. The word $z = 01101101$ has the fractional root 011 and the exponent $\frac{8}{3}$. However, z also has the fractional root 011011 and the exponent $\frac{8}{6}$ or the fractional root 01101101 and the exponent 1. Hence the word z is primitive.

For every non-empty factor u of $\mathbf{u}$ we call the *index* of u in $\mathbf{u}$ the supremum of $e \in \mathbb{Q}$ such that u^e is a factor of $\mathbf{u}$:

$$\mathrm{ind}_{\mathbf{u}}(u) = \sup\{e \in \mathbb{Q} : u^e \in \mathcal{L}(\mathbf{u})\} .$$

The *critical exponent* of a sequence $\mathbf{u}$ is

$$\mathrm{cr}(\mathbf{u}) = \sup\{\mathrm{ind}_{\mathbf{u}}(u) : u \text{ is a non-empty factor of } \mathbf{u}\}$$

or, equivalently,

$$\mathrm{cr}(\mathbf{u}) = \sup\{e \in \mathbb{Q} : e \text{ is an exponent of a non-empty factor of } \mathbf{u}\} .$$

In other words, each factors of $\mathbf{u}$ has the exponent $e \leq \mathrm{cr}(\mathbf{u})$ and, moreover, for every $\delta > 0$ there is a factor of $\mathbf{u}$ with the exponent $e > \mathrm{cr}(\mathbf{u}) - \delta$.

Clearly, $1 < \mathrm{cr}(\mathbf{u}) \leq +\infty$ for every sequence $\mathbf{u}$. The critical exponent can be also infinite, since there exist sequences with unbounded repetitions, i.e., for every $e \in \mathbb{Q}$

there is an e-power in $\mathbf{u}$. Such a sequence is for example Champernowne sequence or any periodic sequence.

In addition, the critical exponent can be both rational and irrational. For example, the Thue–Morse sequence $\mathbf{m}$ has $\mathrm{cr}(\mathbf{m}) = 2$ (as follows already from the work of Thue [102] published in 1912), while the Fibonacci sequence $\mathbf{f}$ has $\mathrm{cr}(\mathbf{f}) = 2 + \frac{1+\sqrt{5}}{2}$ (as shown by Mignosi and Pirillo [78]). In fact, Krieger and Shallit [71] proved that every real number grater than 1 is a critical exponent of some sequence.

There is a huge amount of results related to critical exponent, repetitions or power avoidance and we do not mention them. For survey of some of them see for example [70, Chapter 3] or [5, Sections 1.6–1.8]. Here we just recall the critical exponents of Sturmian sequences and we mention some recent results which are relevant for this thesis.

Critical exponent of Sturmian sequences

Preceded by several partial results of various authors, finally Damanik and Lenz [42] and Capri and de Luca [38] independently gave a general formula for the critical exponent of Sturmian sequences in terms of continued fraction expansion of their slopes. We use the notation from Section 2.2.1.

Theorem 2.38 ([42, Theorem 1], [38, Theorem 4]). *Let $\alpha = [0; a_1, a_2, a_3, \ldots]$ and let $\mathbf{u}$ be a Sturmian sequence with the slope α. Then the critical exponent of $\mathbf{u}$ is given by*

$$\mathrm{cr}(\mathbf{u}) = 2 + \sup_{N \geq 0} \left\{ a_{N+1} + \frac{q_{N-1} - 2}{q_N} \right\} .$$

From this formula can be easily deduced some interesting previous results. For example, Sturmian sequence has infinite critical exponent if and only if its slope has unbounded coefficients in its continued fraction expansion (which was first proved by Mignosi [77, Theorem 2.25]).

In addition, the Fibonacci sequences with the value $\mathrm{cr}(\mathbf{f}) = 2 + \phi$, where $\phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio, has the smallest critical exponent among all Sturmian sequences. However, the Fibonacci sequence is not the only Sturmian sequence with this value. Obviously, all Sturmian sequences with the same slope $2 - \phi$ has the same critical exponent, since it depends only on the language of the sequence. Also the sequence $E(\mathbf{f})$ with the slope $\phi - 1$ has the same critical exponent. Nevertheless, Capri and de Luca [38] showed that this value is achieved also by Sturmian sequences with different slopes.

Proposition 2.39 ([38, Proposition 15]). *The minimal critical exponent for a Sturmian sequence is $2 + \phi$, and moreover, a Sturmian sequence $\mathbf{u}$ has $\mathrm{cr}(\mathbf{u}) = 2 + \phi$ if and only if the slope of $\mathbf{u}$ is one of the numbers*

$$\begin{aligned} 2 - \phi &= [0; 2, \bar{1}] , & \tfrac{3-\phi}{5} &= [0; 3, \bar{1}] , & \tfrac{\phi+3}{11} &= [0; 2, 2, \bar{1}] , \\ \phi - 1 &= [0; \bar{1}] , & \tfrac{\phi+2}{5} &= [0; 1, 2, \bar{1}] , & \tfrac{8-\phi}{11} &= [0; 1, 1, 2, \bar{1}] . \end{aligned}$$

In addition, by [35, Section 2.1] Sturmian sequences of slopes $\frac{\phi+4}{19}$ and $\frac{15-\phi}{19}$ have critical exponent $\frac{11}{3} > \phi + 2$ and all other Sturmian sequences have critical exponent at least 4.

In the episturmian case Justin and Pirillo [66, Theorem 5.2] gave a formula for critical exponent of standard episturmian sequences which are fixed by a primitive morphism. This formula is stated in terms of directive sequences.

Repetition threshold

We have already mentioned that every real number grater than 1 is a critical exponent of some sequence. Nevertheless, when the values of critical exponent approach 1, the needed alphabets grow in size. To capture this growth, the notion of repetition threshold was introduced. The *repetition threshold* is a mapping $\mathrm{RT} : \mathbb{N}_{\geq 1} \to \mathbb{R}_{>1}$ defined for every positive integer n by

$$\mathrm{RT}(n) = \inf \{\gamma \in \mathbb{R} : \exists\, \mathbf{u} \text{ over } n\text{-letter alphabet with } \mathrm{cr}(\mathbf{u}) = \gamma\} .$$

It is easy to find out that every binary sequence has the critical exponent at least 2 and the value 2 is achieved for example by the Thue–Morse sequence (as shown by Thue [102]). Hence we get $\mathrm{RT}(2) = 2$. Dejean [43] found $\mathrm{RT}(3)$ and also stated a famous conjecture about the value of this threshold for every n. After many partial results, a proof of this conjecture was completed independently by Currie and Rampersad [41] and by Rao [92]. Thus the repetition threshold is as follows:

$$\mathrm{RT}(n) = \begin{cases} 2 & \text{for } n = 2; \\ \frac{7}{4} & \text{for } n = 3; \\ \frac{7}{5} & \text{for } n = 4; \\ \frac{n}{n-1} & \text{for } n \geq 5. \end{cases}$$

We are especially interested in some recent results about repetition thresholds in special classes of sequences. Rampersad, Shallit and Vandomme [91] and later also Baranwal and Shallit [12] studied this threshold for balanced sequences. Especially the second paper employs a computational approach using the automatic theorem-proving software Walnut, which has been recently used to variety of problems in combinatorics on words. One can read more about Walnut in [83].

In addition, Baranwal and Shallit [13] also studied the repetition threshold for rich sequences over small alphabets. Using Walnut they gave lower bounds on its value for binary and ternary sequences. In the binary case they also constructed a sequence with the critical exponent $2 + \sqrt{2}/2$ and they conjectured that this critical exponent is minimal among all binary rich sequences. This conjecture was proven by Curie, Mol and Rampersad [40].

Theorem 2.40 ([40, Theorem 2]). *Each binary rich sequence* $\mathbf{u}$ *has*

$$\mathrm{cr}(\mathbf{u}) \geq 2 + \tfrac{\sqrt{2}}{2}$$

and this bound is attained by sequences $\mathbf{v} = \tau(\psi^\omega(0))$ *and* $\mathbf{v}' = \sigma(\psi^\omega(0))$*, where*

$$\psi : \begin{cases} 0 \to 01 \\ 1 \to 02 \\ 2 \to 022 \end{cases}, \quad \tau : \begin{cases} 0 \to 0 \\ 1 \to 01 \\ 2 \to 011 \end{cases} \quad \textit{and} \quad \sigma : \begin{cases} 0 \to 00101 \\ 1 \to 00101101 \\ 2 \to 0010110101101 \end{cases}.$$

It is not difficult to find out that both sequences $\mathbf{v}$ and $\mathbf{v}'$ are complementary symmetric Rote sequences (Section 2.2.4).

Notions related to the critical exponent

There exist several quantities similar to the critical exponent. Let us mention at least some of them.

The *asymptotic critical exponent* of $\mathbf{u}$ is defined by

$$\mathrm{cr}^*(\mathbf{u}) = \lim_{n\to\infty} \sup\{e \in \mathbb{Q} : e \text{ is an exponent of a factor } w \text{ of } \mathbf{u} \text{ with } |w| \geq n\}\,.$$

Thus it takes into account only repetitions of factors with growing length. Clearly, $1 \leq \mathrm{cr}^*(\mathbf{u}) \leq \mathrm{cr}(\mathbf{u})$. Let us emphasize that while for some sequences these two quantities are the same (e.g., if one of them is infinite), it does not hold in general.

In the case of Sturmian sequences the exact formula for $\mathrm{cr}^*(\mathbf{u})$ is known. Vandeth [103] proved it for Sturmian sequences which are fixed points of some morphisms, but Cassaigne in [35, Section 2.2] noticed that it remains valid for all Sturmian sequences.

Theorem 2.41 ([103, Section 5]). *Let $\alpha = [0; a_1, a_2, a_3, \ldots]$ and let $\mathbf{u}$ be a Sturmian sequence with the slope α. Then the asymptotic critical exponent of $\mathbf{u}$ is given by*

$$\mathrm{cr}^*(\mathbf{u}) = 2 + \limsup_{N\to\infty}\, [a_N; a_{N-1}, \ldots, a_1]\,.$$

It directly implies that the Fibonacci sequence $\mathbf{f}$ with $\mathrm{cr}^*(\mathbf{f}) = \mathrm{cr}(\mathbf{f}) = \phi + 2$ is again optimal among Sturmian sequences (as well as the sequence $\varphi(\mathbf{f})$ for any Sturmian morphism φ).

However, the Thue–Morse sequence $\mathbf{t}$ with $\mathrm{cr}^*(\mathbf{t}) = \mathrm{cr}(\mathbf{t}) = 2$ is no more optimal among all binary sequences, since Cassaigne [35, Theorem 2.4] constructed a binary sequence with $\mathrm{cr}^*(\mathbf{u}) = 1$.

The *initial critical exponent* of $\mathbf{u}$ is defined by

$$\mathrm{icr}(\mathbf{u}) = \sup\{e \in \mathbb{Q} : e \text{ is an exponent of a non-empty prefix of } \mathbf{u}\}\,.$$

and the *asymptotic initial critical exponent* of $\mathbf{u}$ is

$$\mathrm{icr}^*(\mathbf{u}) = \lim_{n\to\infty} \sup\{e \in \mathbb{Q} : e \text{ is an exponent of a prefix } w \text{ of } \mathbf{u} \text{ with } |w| \geq n\}\,.$$

Similarly as in the case of non-initial critical exponent we have $1 \leq \mathrm{icr}^*(\mathbf{u}) \leq \mathrm{icr}(\mathbf{u})$.

Berthé, Holton, and Zamboni [22] studied these quantities in the case of Sturmian sequences. They gave a formula for the (asymptotic) initial critical exponent of Sturmian sequences in terms of some S-adic representations (see [22, Proposition 3.3 and Corollary 3.5]) as well as they deduced from it some other interesting results.

The definitions directly imply that $\mathrm{icr}(\mathbf{u}) \leq \mathrm{cr}(\mathbf{u})$ and $\mathrm{icr}^*(\mathbf{u}) \leq \mathrm{cr}^*(\mathbf{u})$. Berthé et al. [22, Theorem 1.2] proved that for each standard Sturmian sequence $\mathbf{u}$ one has

$$\mathrm{cr}^*(\mathbf{u}) = 1 + \mathrm{icr}^*(\mathbf{u})\,. \tag{2.4}$$

The Fibonacci sequence has $\mathrm{icr}^*(\mathbf{f}) = \mathrm{icr}(\mathbf{f}) = 1 + \phi$. In addition, Allouche et al. [4] showed that every Sturmian sequence $\mathbf{u}$ has $\mathrm{icr}^*(\mathbf{u}) \geq 2$. Berthé et al. characterized Sturmian sequence with $\mathrm{icr}^*(\mathbf{u}) = 2$ (see [22, Proposition 1.1]), obviously, the Fibonacci sequence is not one of them.

Finally, let us mention that while the (asymptotic) critical exponent depends only on the language of $\mathbf{u}$ (hence in the case of Sturmian sequence it depends only on its slope and not on its intercept), the (asymptotic) initial critical exponent depends on the sequence $\mathbf{u}$ itself. Hence it is natural to ask how the initial critical exponent varies among the sequences with the same slope. Surprisingly, it seems that it can differ essentially.

In particular, for each irrational slope α with unbounded coefficients in its continued fraction expansion the respective standard Sturmian sequence $\mathbf{u}$ has $\mathrm{icr}^*(\mathbf{u}) = \mathrm{cr}^*(\mathbf{u}) = +\infty$ (see Equation (2.4)). On the other hand, Berthé et al. [22, Proposition 4.1] found out that for every irrational α there is a Sturmian sequence $\mathbf{v}$ with the slope α which has $\mathrm{icr}^*(\mathbf{v}) \leq 1 + \phi$. In fact, Mignosi, Restivo and Salemi [79] proved that this is true for every infinite minimal subshift, see [79] and also [22] for more details.

Recently, the Abelian variant of the critical exponent has been considered, too. For details and references see, e.g., [56, 88].

2.3.3 Recurrence function

Like the return words, also the notion of recurrence origins in the theory of dynamical systems. The combinatorial point of view was initiated in 1938 by Morse and Hedlund [81]. Many details about recurrence and other related notions can be found for example in [5, Sections 10.8–10.10] or in the surveys [34] of [35].

Let us recall that a sequence $\mathbf{u}$ is *recurrent* if each of its factors occurs infinitely many times in $\mathbf{u}$. However, the occurrences of a given factor of a recurrent sequence can occur with arbitrary large gaps. If we suppose that these gaps are bounded, we get uniformly recurrent sequence. More precisely, a sequence $\mathbf{u}$ is *uniformly recurrent* if for every $n \in \mathbb{N}$ there exists $m \in \mathbb{N}$ such that each factor of length m contains every factor of length n.

The recurrence function quantifies the speed of recurrence in the uniformly recurrent sequence: it assigns to each length n the respective length m. More formally, the *recurrence function* of a uniformly recurrent sequences $\mathbf{u}$ is the function $R_{\mathbf{u}} : \mathbb{N} \to \mathbb{N}$ defined by

$$\mathrm{R}_{\mathbf{u}}(n) = \min\left\{m \in \mathbb{N} : \text{each factor from } \mathcal{L}_{\mathbf{u}}(m) \text{ contains all factors from } \mathcal{L}_{\mathbf{u}}(n)\right\} .$$

If we denote $f_n(i)$ the factor of length n occurring in $\mathbf{u}$ at the position i, we can express the recurrence function also as follows:

$$\mathrm{R}_{\mathbf{u}}(n) = \min\left\{m \in \mathbb{N} : \forall i \in \mathbb{N}\ \{f_n(i), f_n(i+1), \ldots, f_n(i+m-n)\} = \mathcal{L}_{\mathbf{u}}(n)\right\} . \tag{2.5}$$

From this expression it it easy to deduce the following inequality:

$$\mathrm{R}_{\mathbf{u}}(n) \geq \mathcal{C}_{\mathbf{u}}(n) + n - 1 \quad \text{for each } n \in \mathbb{N} . \tag{2.6}$$

In fact, Morse and Hedlund [81] showed that for aperiodic sequences this inequality can be improved to $\mathrm{R}_{\mathbf{u}}(n) \geq \mathcal{C}_{\mathbf{u}}(n) + n$ for each $n \in \mathbb{N}$.

On the other hand, the recurrence function cannot be bounded from above by the complexity function since, for example, it is possible to find Sturmian sequences whose recurrence functions grow fast. It follows from the formula for the recurrence function of Sturmian sequences which was constructed by Morse and Hedlund in [82]. We use the notation from Section 2.2.1.

Theorem 2.42 ([82]). *Let $\alpha = [0; a_1, a_2, a_3, \ldots]$ and let* $\mathbf{u}$ *be a Sturmian sequence with the slope α. Then the recurrence function of* $\mathbf{u}$ *is given by*

$$R_{\mathbf{u}}(n) = q_{N+1} + q_N + n - 1 \qquad \text{for every } n \in [q_N, q_{N+1})\ .$$

This result was further generalized by Cassaigne and Chekhova [36, Proposition 2.4] who gave the formula for Arnoux–Rauzy sequences. Among other things, they use the following general relation between the recurrence function and the lengths of the return words.

Proposition 2.43 ([34, Proposition 2]). *For any uniformly recurrent sequence* $\mathbf{u}$ *and for any $n \in \mathbb{N}$ we have*

$$R_{\mathbf{u}}(n) = \max\{|r| : r \text{ is a return word to a factor of } \mathbf{u} \text{ of length } n\} + n - 1\,.$$

In fact, to find the recurrence function of $\mathbf{u}$ it suffices to determine the return words to so-called *(essential) singular factors* of $\mathbf{u}$ (e.g., see [33]) which can be derived from bispecial factors. The method is precisely described in [34, Section 5] and it can be use for example to calculated the recurrence function of the Thue–Morse sequence (see also [35, Proposition 6.1]). This approach was also used in Balková [9] to obtain the recurrence function of some class of sequences associated with β-integers.

Recurrence function of Sturmian sequences was recently studied also from probabilistic point of view, see [101].

Recurrence quotient

When the recurrence function of $\mathbf{u}$ grows slowly, it means that all factors of $\mathbf{u}$ have to occur quite often and so $\mathbf{u}$ has to be highly structured. By Inequality (2.6) such a sequence $\mathbf{u}$ has also small factor complexity. Hence we define the *recurrence quotient* of $\mathbf{u}$, denoted $\rho(\mathbf{u})$, as

$$\rho(\mathbf{u}) = \limsup_{n\to\infty} \frac{\mathrm{R}_{\mathbf{u}}(n)}{n}\,.$$

Clearly, the recurrence quotient is always at least 1. Let us emphasize that it can be also infinite. For example, this is the case of Sturmian sequences whose slopes have unbounded coefficients in their continued fraction expansions.

If $\rho(\mathbf{u})$ is finite, then the sequence $\mathbf{u}$ is called linearly recurrent. These sequences have some interesting properties, e.g., see [53]. In particular, Durand showed in [50, 51] that they have a characterization in terms of S-adic representations.

Clearly, if $\mathbf{u}$ is periodic, then $\rho(\mathbf{u}) = 1$. For aperiodic sequences Morse and Hedlund [81] proposed an open problem to find the best lower bound for $\rho(\mathbf{u})$. Cassaigne [34] proved that $\rho(\mathbf{u}) \geq 3$, but the Rauzy's conjecture from [93] which states that $\rho(\mathbf{u}) \geq 2 + \phi$ (where ϕ is the golden ratio) is, as far as we know, still open.

Nevertheless, for Sturmian sequences this estimate is true and the bound is attained by the Fibonacci sequence (as well as by the sequence $\varphi(\mathbf{f})$ for any Sturmian morphism φ). In fact, Theorem 2.42 directly implies the following formula for the recurrence quotients of Sturmian sequences.

Theorem 2.44 ([82]). *Let $\alpha = [0; a_1, a_2, a_3, \ldots]$ and let* $\mathbf{u}$ *be a Sturmian sequence with the slope α. Then the recurrence quotient of* $\mathbf{u}$ *is*

$$\rho(\mathbf{u}) = 2 + \limsup_{N\to\infty} \frac{q_N}{q_{N-1}} = 2 + \limsup_{N\to\infty} [a_N; a_{N-1}, \ldots, a_1]\ .$$

Cassaigne in [33] studies the structure of the set $\{\rho(\mathbf{u}) : \mathbf{u} \text{ is Sturmian sequence}\}$ in detail, but also in this special case some questions remain open.

Modifications of the recurrence function

Some other functions related to the recurrence function have been studied, too. First of all, it is quite natural to consider the prefix variant of the recurrence function. We denote $\mathrm{R}'_{\mathbf{u}}(n)$ the length of the smallest prefix of $\mathbf{u}$ which contains all factors of $\mathbf{u}$ of length n, i.e.,

$$\mathrm{R}'_{\mathbf{u}}(n) = \min\left\{m \in \mathbb{N} : \{f_n(0), f_n(1), \ldots, f_n(m-n)\} = \mathcal{L}_{\mathbf{u}}(n)\right\} ,$$

and we consider also the respective quotient

$$\rho'(\mathbf{u}) = \limsup_{n\to\infty} \frac{\mathrm{R}'_{\mathbf{u}}(n)}{n} .$$

This function was defined by Allouche and Bousquet-Mélou [2] to reformulate a conjecture on automaticity function, see [2] or [31] for details. In fact, they proved that this conjecture on automaticity function is equivalent to another conjecture which says that every aperiodic sequence $\mathbf{u}$ has $\rho'(\mathbf{u}) \geq 1 + \phi$. It is well-known that the Fibonacci sequence satisfies $\rho'(\mathbf{f}) = 1 + \phi$.

However, Cassaigne [31, Theorem 1] disproved this conjecture and found the correct optimal lower bound. More precisely, he proved that every aperiodic sequence $\mathbf{u}$ has

$$\rho'(\mathbf{u}) \geq \frac{29 - 2\sqrt{10}}{9}$$

and this value is attained by the Sturmian sequence which is the fixed point of the morphism $0 \to 01001010$, $1 \to 010$. Thus the Fibonacci sequence is optimal only among standard Sturmian sequences and not in the general case.

Let us mention that one can find also closely related function $\alpha_{\mathbf{u}}(n) = \mathrm{R}'_{\mathbf{u}}(n) - n$, which is called *appearance*. The value $\mathrm{R}'_{\mathbf{u}}(n) - n + 1 = \alpha_{\mathbf{u}}(n) + 1$ expresses the maximal position where a factor of length n occurs in $\mathbf{u}$ for the first time. Some details can be found in [5, Sections 10.10 and 15.3].

Cassaigne in [32] studied the function $\mathrm{R}''_{\mathbf{u}}(n)$ which assigns to each n the length of the smallest factor of $\mathbf{u}$ which contains all factors of length n.

Clearly, the functions $\mathrm{R}_{\mathbf{u}}$ and $\mathrm{R}''_{\mathbf{u}}$ depend only on the language of $\mathbf{u}$, while the function $\mathrm{R}'_{\mathbf{u}}$ depends on the sequence $\mathbf{u}$ itself. The mentioned functions also fulfil the following inequalities:

$$\mathcal{C}_{\mathbf{u}}(n) + n - 1 \leq \mathrm{R}''_{\mathbf{u}}(n) \leq \mathrm{R}'_{\mathbf{u}}(n) \leq \mathrm{R}_{\mathbf{u}}(n) \quad \text{for every sequence } \mathbf{u} \text{ and } n \in \mathbb{N},$$

and so each aperiodic sequence $\mathbf{u}$ has $\mathrm{R}''_{\mathbf{u}}(n) \geq 2n$ for every $n \in \mathbb{N}$.

Cassaigne [32, Theorem 1 and Corollary 1] showed that Sturmian sequences can be described by means of this function: a sequence $\mathbf{u}$ is Sturmian if and only if $\mathrm{R}''_{\mathbf{u}}(n) = 2n$ for every $n \in \mathbb{N}$. In particular, it means that all $n+1$ distinct factors from $\mathcal{L}_{\mathbf{u}}(n)$ occur in some factor $w \in \mathcal{L}_{\mathbf{u}}(2n)$. This is only possible if each factor from $\mathcal{L}_{\mathbf{u}}(n)$ occurs once in w. In that case, Cassaigne says that $\mathbf{u}$ has *grouped factors*. In general, a sequence $\mathbf{u}$

has grouped factors if and only if $\mathrm{R}''_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n) + n - 1$ for every $n \in \mathbb{N}$. Among other things, he showed that Sturmian sequences are not the only sequence with grouped factors, but the precise characterization of these sequences and many other related questions remain open.

By studying the relation between the properties of sequences and its morphic images, Frid [57] computed the functions $\mathrm{R}_{\mathbf{u}}$, $\mathrm{R}'_{\mathbf{u}}$ and $\mathrm{R}''_{\mathbf{u}}$ (as well as some other quantities) for fixed points of a large class of morphisms.

Links between the critical exponent and the recurrence function

Clearly, the notions of critical exponent and recurrence function (as well as many other quantities describing sequences) are related. It is particularly evident in the case of Sturmian sequences, where, in fact, Carpi and de Luca [38] used the formula for the recurrence function (Theorem 2.42) to find the formula for the critical exponent (Theorem 2.38). Now we state some of these links more explicitly.

Cassaigne [35, Proposition 3.2] observed that the following bound on the asymptotic critical exponent can be derived from the recurrence quotient:

$$\mathrm{cr}^*(\mathbf{u}) \geq 1 + \frac{1}{\rho(\mathbf{u}) - 1} \quad \text{for any sequence } \mathbf{u}\,.$$

In particular, when $\mathrm{cr}^*(\mathbf{u}) = 1$, then $\rho(\mathbf{u})$ is infinite. He also explains that, apart from this case and the periodic case where $\mathrm{cr}^*(\mathbf{u}) = +\infty$ and $\rho(\mathbf{u}) = 1$, the equality cannot hold.

Hence it may seem that the quantities $\mathrm{cr}^*(\mathbf{u})$ and $\rho(\mathbf{u})$ vary in opposite directions. However, this is not true at least for Sturmian sequences, since by comparing Theorems 2.41 and 2.44, we get that each Sturmian sequence $\mathbf{u}$ has $\mathrm{cr}^*(\mathbf{u}) = \rho(\mathbf{u})$. Nevertheless, it is not clear if this property characterizes Sturmian sequences.

Masáková and Pelantová [75, Theorem 1] found another relation between the recurrent function and the indices of factors which holds exactly for Sturmian sequences. More precisely, they proved that a uniformly recurrent sequence $\mathbf{u}$ is Sturmian if and only if there exist infinitely many factors w of $\mathbf{u}$ such that $R_{\mathbf{u}}(|w|) = |w| \cdot \mathrm{ind}(w) + 1$. In addition, they can used it to present an alternative proof of Theorem 2.38, since their approach relies on the Vuillon's description of Sturmian sequences by return words (see Theorem 2.28) instead of the manipulation with the continued fractions of the slopes.

2.3.4 Non-repetitive complexity

Non-repetitive complexity is another type of complexity which can be viewed as a dual function to the recurrence function.

The *non-repetitive complexity* of a sequence $\mathbf{u}$ is a function $\mathrm{nr}\mathcal{C}_{\mathbf{u}} : \mathbb{N} \to \mathbb{N}$ which to each length n assigns the maximal integer m such that for some position $i \in \mathbb{N}$ any factor of $\mathbf{u}$ of length n occurs at most ones in $u_i u_{i+1} \cdots u_{i+m+n-2}$. In other words, $\mathrm{nr}\mathcal{C}_{\mathbf{u}}(n)$ expresses the maximal number of distinct factors of length n which can be seen one after another somewhere in $\mathbf{u}$ until some factor is repeated. If we denote $f_n(i)$ the factor of length n occurring in $\mathbf{u}$ at the position i, we can write

$$\begin{aligned}\mathrm{nr}\mathcal{C}_{\mathbf{u}}(n) = \max\{m \in \mathbb{N} : \exists\, i \in \mathbb{N} \text{ such that}& \\ f_n(i), f_n(i+1), \ldots, f_n(i+m-1) \text{ are pairwise distinct}\}&\,.\end{aligned}$$

Similarly we define also the prefix variant of this function. The *initial non-repetitive complexity* of a sequence $\mathbf{u}$ is a function $\mathrm{inr}\mathcal{C}_{\mathbf{u}} : \mathbb{N} \to \mathbb{N}$ which to each length n assigns the maximal integer m such that any factor of $\mathbf{u}$ of length n occurs at most ones in $u_0 u_1 \cdots u_{m+n-2}$, i.e.,

$$\mathrm{inr}\mathcal{C}_{\mathbf{u}}(n) = \max\{m \in \mathbb{N} : f_n(0), f_n(1), \ldots, f_n(m-1) \text{ are pairwise distinct}\}\,.$$

Let us emphasize that while the function $\mathrm{nr}\mathcal{C}_{\mathbf{u}}$ depends only on the language of $\mathbf{u}$ and not on the sequence $\mathbf{u}$ itself, the function $\mathrm{inr}\mathcal{C}_{\mathbf{u}}$ depends on the precise structure of the sequence $\mathbf{u}$. In fact, the situation is exactly the same also in the case of (initial) critical exponent or (initial) recurrence function.

What we meant by non-repetitive complexity was introduced by Nicholson and Rampersad in [84], where they study the initial non-repetitive complexity. Nevertheless, the original idea and also the name come from Moothathu [80]. He proposed a new type of entropy of dynamical systems, so-called Eulerian entropy, which can be in the setting of symbolics dynamics formulated as a combinatorial property named non-repetitive complexity, see [80, Section 3] for details. However, Moothathu used this term for the quantity

$$\limsup_{n\to\infty} \tfrac{\log \mathrm{inr}\mathcal{C}_{\mathbf{u}}(n)}{n}\,.$$

The initial non-repetitive complexity was independently defined also by Bugeaud and Kim [29], whose motivation comes from the interplay between combinatorics on words and Diophantine approximation of real numbers. For example, they use the initial non-repetitive complexity of $\mathbf{u}$ to study the irrational exponent of a number x, whose expansion in a given base corresponds with $\mathbf{u}$ (see [29, Section 4]).

In fact, they defined the function $\mathrm{r}_{\mathbf{u}} : \mathbb{N} \to \mathbb{N}$ which to each n assigns the length of the smallest prefix of $\mathbf{u}$ containing two (possibly overlapping) occurrences of some factor of length n, i.e.,

$$\mathrm{r}_{\mathbf{u}}(n) = \min\{m \in \mathbb{N} : f_n(i) = f_n(m-n+1) \text{ for some } i \text{ with } 0 \le i \le m-n\}\,.$$

However, since

$$\mathrm{r}_{\mathbf{u}}(n) = \mathrm{inr}\mathcal{C}_{\mathbf{u}}(n) + n \text{ for every } n \in \mathbb{N}\,,$$

we can easily reformulate their results into our notion of $\mathrm{inr}\mathcal{C}_{\mathbf{u}}$.

Both Nichoson and Rampersad [84] and Bugeaud and Kim [29] examined general properties of this function in comparison to the classical complexity function. Clearly, the following inequalities hold:

$$\mathrm{inr}\mathcal{C}_{\mathbf{u}}(n) \le \mathrm{nr}\mathcal{C}_{\mathbf{u}}(n) \le \mathcal{C}_{\mathbf{u}}(n) \quad \text{for each } n \in \mathbb{N}\,.$$

They showed that these functions can both be equal or differ essentially. More precisely, they proved that for every $d \ge 3$, there exists a sequence $\mathbf{u}$ over d-letter alphabet with $\mathrm{inr}\mathcal{C}_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n) = d^n$ for all $n \in \mathbb{N}$. On the other hand, for every $d > 1$ they constructed a sequence $\mathbf{v}$ with the factor complexity $\mathcal{C}_{\mathbf{v}}(n) = d^n$ and the non-repetitive complexity $\mathrm{inr}\mathcal{C}_{\mathbf{v}}(n) \le 4n$ for all $n \ge 1$. See [84, Propositions 2 and 3] and [29, Section 2].

In addition, they characterize the periodic sequences by means of the initial non-repetitive complexity. More precisely, a sequence $\mathbf{u}$ is eventually periodic if the function $\mathrm{inr}\mathcal{C}_{\mathbf{u}}(n)$ is bounded, see [84, Theorem 1] and [29, Theorem 2.3].

Moreover, Bugeaud and Kim also provide a new characterization of Sturmian sequences.

Theorem 2.45 ([29, Theorem 2.4]). *A sequence* $\mathbf{u}$ *is Sturmian if and only if* $\mathbf{u}$ *has* $\mathrm{inr}\mathcal{C}_{\mathbf{u}}(n) \leq n+1$ *for every* $n \geq 1$ *with the equality for infinitely many* n*.*

They also defined the *exponent of repetition* of $\mathbf{u}$, denoted $\mathrm{rep}(\mathbf{u})$, as

$$\mathrm{rep}(\mathbf{u}) = \liminf_{n\to\infty} \frac{r_{\mathbf{u}}(n)}{n}.$$

They proved that $1 \leq \mathrm{rep}(\mathbf{u}) \leq \sqrt{10} - 3/2$ when $\mathbf{u}$ runs over the Sturmian sequences as well as that both extremal values are attained, see [29, Section 3].

Similarly, Nicholson and Rampersad use the limes superior of $r_{\mathbf{u}}(n)/n$ to give a new criterion for aperiodicity. They showed that $\mathbf{u}$ is eventually periodic if and only if

$$\limsup_{n\to\infty} \frac{\mathrm{inr}\mathcal{C}_{\mathbf{u}}(n)}{n} < \frac{1}{1+\phi^2}, \qquad \text{where } \phi \text{ is the golden ratio}.$$

Nevertheless, it remains open if the constant $\frac{1}{1+\phi^2}$ is the best possible or if it can be replaced for example by 1.

Not surprisingly, (initial) non-repetitive complexity and related quantities have some links with other combinatorial properties such as (initial) recurrence function, (initial) critical exponent or Diophantine exponent, some details can be found in [29, Sections 9 and 10].

Although Moothathu introduced the concept of non-repetitive complexity, he did not explicitly compute this function for any particular sequence. Nicholson and Rampersad [84] obtained formulas for the Thue–Morse sequence, the Fibonacci sequence and the Tribonacci sequence. They also constructed some square-free sequences with slowly growing initial non-repetitive complexity. The formulas for the Thue–Morse and Fibonacci sequences are also mentioned in [29].

Proposition 2.46 ([84, Theorems 6, 10 and 16]). *The initial non-repetitive complexities of the Thue–Morse sequence* $\mathbf{m}$*, the Fibonacci sequence* $\mathbf{f}$ *and the Tribonacci sequence* $\mathbf{t}$ *are as follows:*

(i) If $2^{k-1} < n \leq 2^k$ *for some integer* $k \geq 1$*, then* $\mathrm{inr}\mathcal{C}_{\mathbf{m}}(n) = 3 \cdot 2^{k-1}$*.*

(ii) If $F_k - 2 < n \leq F_{k+1} - 2$ *for some integer* $k \geq 1$*, then* $\mathrm{inr}\mathcal{C}_{\mathbf{f}}(n) = F_k$*, where* F_k *is* k^{th} *Fibonacci number.*

(iii) If $\frac{T_k+T_{k-2}-3}{2} < n \leq \frac{T_{k+1}+T_{k-1}-3}{2}$ *for* $k \geq 1$*, then* $\mathrm{inr}\mathcal{C}_{\mathbf{t}}(n) = T_k$*, where* T_k *is* k^{th} *Tribonacci number.*

2.4 Our tools for studying sequences

The aim of this section is to briefly describe the essential tools which we used to obtain our results. In particular, we focus on Rauzy graphs and S-adic representation of sequences. Although both this concepts are now well established in combinatorics on words, we would like to highlight them since it seems that they can be very useful also for other studies.

Let us mention that we take advantage also from other well-known concepts of combinatorics on words such as special factors (Section 2.1.4) or Parikh vectors (Section 2.1.3). In the article [E] we significantly use continued fraction expansions related to Sturmian sequences (Section 2.2.1), too. In papers [D] and [E] we also utilize the results about return words and derived sequences obtained in the articles [A, B, C].

2.4.1 Rauzy graphs

Rauzy graphs can be useful when studying the properties of a sequence $\mathbf{u}$ since they visualize the factor structure of $\mathbf{u}$. They were introduced by Rauzy [93].

First let us recall some definitions. For each non-negative integer n, the Rauzy graph of order n, denoted $\Gamma_{\mathbf{u}}(n)$, is the oriented graph (V, E), where the set V of vertices is the set $\mathcal{L}_{\mathbf{u}}(n)$ of all factors of $\mathbf{u}$ of length n, the set E of oriented edges is the set $\mathcal{L}_{\mathbf{u}}(n+1)$ of all factors of $\mathbf{u}$ of length $n+1$ and there is an edge e from v to w if there exist two letters a and b such that $e = va = bw \in \mathcal{L}_{\mathbf{u}}(n+1)$. Let us emphasize that we label the vertices by the factors of $\mathbf{u}$ of length n and the edges by the factors of $\mathbf{u}$ of length $n+1$. However, in the literature there are also different ways of labelling the edge e, e.g., by the first or the last letter of e. Several first Rauzy graphs of the Fibonacci sequence $\mathbf{f}$ are displayed in Figure 2.4.

By a path P of length m in the Rauzy graph $\Gamma_{\mathbf{u}}(n)$ we mean a sequence of $m+1$ consecutive vertices from $\Gamma_{\mathbf{u}}(n)$

$$v_0 \xrightarrow{v_0 a_1} v_1 \xrightarrow{v_1 a_2} \cdots \xrightarrow{v_{m-1} a_m} v_m\,, \qquad v_0, \ldots, v_m \in \mathcal{L}_{\mathbf{u}}(n)\,,\ a_1, \ldots, a_m \in \mathcal{A},$$

and we label the path P by the word $p = v_0 a_1 a_2 \cdots a_m$ of length $n+m$.

Clearly, every factor of $\mathbf{u}$ of length ℓ is a label of some path in $\Gamma_{\mathbf{u}}(n)$ for each $n \leq \ell$. On the other hand, not all paths in $\Gamma_{\mathbf{u}}(n)$ belong to the language of $\mathbf{u}$. Indeed, for example the word 000 is the label of the paths

$$0 \xrightarrow{00} 0 \xrightarrow{00} 0$$

in the Rauzy graph $\Gamma_{\mathbf{f}}(1)$, but, obviously, it is not the factor of the Fibonacci sequence $\mathbf{f}$. The reason is that the Rauzy graph $\Gamma_{\mathbf{u}}(n)$ captures all possible one-letter prolongations

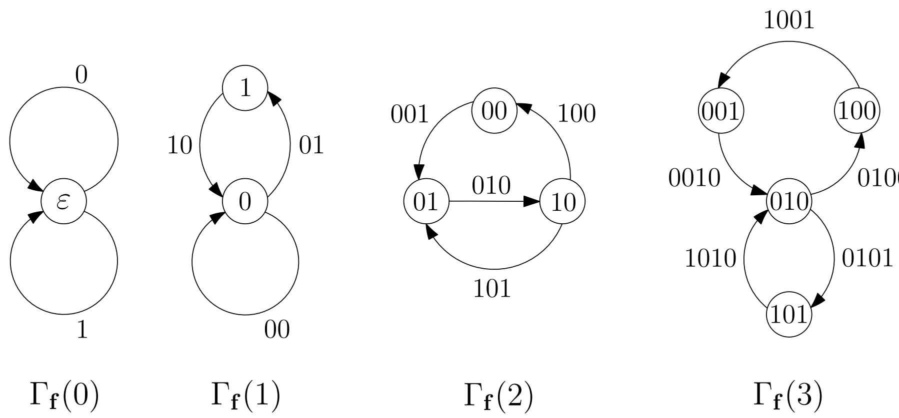

Figure 2.4: Rauzy graphs of the Fibonacci sequence.

of factors of length n, but it does not take into account the longer factors (or, in other words, the history). For example, the graph $\Gamma_{\mathbf{f}}(1)$ expresses that somewhere in the word $\mathbf{f}$ the factor 0 is followed by 0 and, elsewhere, it is followed by 1. However, it does not reflect the fact that while the factor 10 can be followed by both letters 0 and 1, the factor 00 cannot be followed by 0 since 000 is not a factor of $\mathbf{f}$. Nevertheless, if each of inner vertices of a path $v_0 v_1 \cdots v_m$ has indegree and outdegree 1, then the label of this path corresponds to a factor of $\mathbf{u}$.

Many properties of sequences can be reformulated in the notion of Rauzy graphs. For example, a sequence $\mathbf{u}$ is recurrent if and only if all its Rauzy graphs are strongly connected, i.e., every two vertices are connected by an oriented path.

It is also easy to realize which vertices in $\Gamma_{\mathbf{u}}(n)$ correspond with special factors of $\mathbf{u}$ of length n. A vertex w is left special if and only if it has indegree at least 2 and, analogously, w is right special if and only if it has outdegree at least 2. Moreover, the edges incoming to w and outcoming from w correspond with left and right extensions of w.

Evolution of Rauzy graphs

In fact, for our purposes it is important to understand how the Rauzy graphs evolve, i.e., what is the connection between $\Gamma_{\mathbf{u}}(n)$ and $\Gamma_{\mathbf{u}}(n+1)$. For simplicity, we suppose that the sequence $\mathbf{u}$ is recurrent. This, in particular, means that for every vertex w of $\Gamma_{\mathbf{u}}(n)$ there is at least one edge aw incoming to w and there is at least one edge wb outcoming from w.

We create the Rauzy graph $\Gamma_{\mathbf{u}}(n+1)$ from $\Gamma_{\mathbf{u}}(n)$. The vertices of $\Gamma_{\mathbf{u}}(n+1)$ are all edges of $\Gamma_{\mathbf{u}}(n)$ and the edges of $\Gamma_{\mathbf{u}}(n+1)$ are defined as follows:

(i) for each non-special vertex w of $\Gamma_{\mathbf{u}}(n)$ with the unique left extension a and right extension b there is one edge $aw \xrightarrow{awb} bw$;

(ii) for each left special vertex w of $\Gamma_{\mathbf{u}}(n)$ with k left extensions $a_1, a_2, \ldots, a_k$ and the unique right extension b (i.e., w is not bispecial) there are k edges

$$a_1 w \xrightarrow{a_1 wb} wb\,,\ a_2 w \xrightarrow{a_2 wb} wb\,,\ \ldots\,,\ a_k w \xrightarrow{a_k wb} wb\,;$$

(iii) for each right special vertex w of $\Gamma_{\mathbf{u}}(n)$ with ℓ right extensions $b_1, b_2, \ldots, b_\ell$ and the unique left extension a (i.e., w is not bispecial) there are ℓ edges

$$aw \xrightarrow{awb_1} wb_1\,,\ aw \xrightarrow{awb_2} wb_2\,,\ \ldots\,,\ aw \xrightarrow{awb_\ell} wb_\ell\,;$$

(iv) for each bispecial vertex w of $\Gamma_{\mathbf{u}}(n)$ with the left extensions $a_1, \ldots, a_k$ and the right extensions $b_1, \ldots, b_\ell$ there are edges of the form

$$a_i w \xrightarrow{a_i w b_j} wb_j\,,\ \text{where } i \in \{1, \ldots, k\}\,,\ j \in \{1, \ldots, \ell\} \ \text{ and } \ a_i w b_j \in \mathcal{L}_{\mathbf{u}}(n+2)\,.$$

Let us emphasize that in the bispecial case the last condition is essential. Usually, not all the edges of the form $a_i w b_j$ are included in $\Gamma_{\mathbf{u}}(n+1)$ since not all the words $a_i w b_j$ are factors of $\mathbf{u}$.

Example 2.47. We derive the graph $\Gamma_{\mathbf{f}}(2)$ from the graph $\Gamma_{\mathbf{f}}(1)$. The edges 00, 01 and 10 are the vertices of $\Gamma_{\mathbf{u}}(2)$. The non-special vertex 1 induces the edge $01 \xrightarrow{010} 10$ and the bispecial vertex 0 induces three edges $10 \xrightarrow{100} 00$, $10 \xrightarrow{101} 01$ and $00 \xrightarrow{001} 01$. The edge $00 \xrightarrow{000} 00$ is not included in $\Gamma_{\mathbf{f}}(2)$ as $000 \notin \mathcal{L}_{\mathbf{f}}(3)$.

We can summarize that if there is no bispecial factor in $\Gamma_{\mathbf{u}}(n)$, then the Rauzy graph $\Gamma_{\mathbf{u}}(n+1)$ is completely determined by $\Gamma_{\mathbf{u}}(n)$.

Rauzy graphs of Sturmian sequences

Rauzy graphs are especially useful for the sequences with relatively small factor complexity, since these sequences do not have to many special factors and so their graphs have simple structure.

Let us recall that every Sturmian sequence $\mathbf{u}$ is recurrent and it has exactly one left and one right special factor of each length n, we denote them x and y. Hence the Rauzy graph $\Gamma_{\mathbf{u}}(n)$ has one of two following shapes (see also Figure 2.4):

(I) If $x \neq y$ (i.e., there is no bispecial factor of length n), then $\Gamma_{\mathbf{u}}(n)$ consists of three paths with the only common vertices x and y: P_A is the minimal path that links x to y, P_B and P_C are paths that links y to x and do not contain P_A.

(II) If $x = y$ (i.e., x is a bispecial factor of length n), then $\Gamma_{\mathbf{u}}(n)$ consists of two cycles P_B, P_C with the only common vertex x.

Let us emphasize that the shape II can be understood as a special case of the shape I when $P_A = x$. Moreover, the labels of paths P_A and P_B, P_C are palindromes. For other details see for example [23, Section 4.5.6].

The Rauzy graphs of Arnoux–Rauzy sequences over d-letter alphabet with $d > 2$ are nearly the same. The only difference is that there are always d (instead of two) distinct paths from the right special factor to the left special factor.

Finally, let us mention that return words and derived sequences can be also naturally interpreted in terms of the Rauzy graphs. We utilize this in the article [D], where Rauzy graphs of Sturmian and Arnoux–Rauzy sequences play a role. The Rauzy graphs of Sturmian sequences are used in the article [D], too.

2.4.2 S-adic representation of sequences

In Section 2.1.2 we define substitutive sequences which are generated via two morphisms. More precisely, each substitutive sequence $\mathbf{u}$ can be written in the form $\mathbf{u} = \tau(\psi^\omega(a))$, where τ, ψ are morphisms and ψ is prolongable on a. We can further generalize this notion by considering an infinite sequence of generating morphisms. This is the idea of S-adic representation of sequences.

Let $\mathcal{A}$ be an alphabet and let S be a finite set of (non-erasing) morphisms on $\mathcal{A}$. Let $\mathbf{Z} = (\zeta_n)_{n\in\mathbb{N}}$ be a sequence of morphisms from S and let $(a_n)_{n\in\mathbb{N}}$ be a sequence of letters from $\mathcal{A}$. We say that the sequence $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ *admits* $((\zeta_n, a_n))_{n\in\mathbb{N}}$ *as an S-adic representation* if

$$\lim_{n\to\infty} |\zeta_0\zeta_1\cdots\zeta_{n-1}(a_n)| = +\infty \quad \text{and} \quad \mathbf{u} = \lim_{n\to\infty} \zeta_0\zeta_1\cdots\zeta_{n-1}(a_n).$$

The sequence $\mathbf{Z}$ is called a *directive sequence* of $\mathbf{u}$. The sequence of letters $(a_n)_{n\in\mathbb{N}}$ plays a minor role compared to the directive sequence. Let us remark that also infinite sets S are sometimes considered.

The notion of S-adicity was precisely introduced by Ferenczi [54] and one can read more about it for example in the interesting survey [19]. Clearly, substitutive sequences admit S-adic representations with periodic directive sequences.

A sequence can admit many different S-adic representations. But some S-adic representations might be more useful to get information about the sequence than the others. Handy S-adic representations are known especially for Sturmian sequence (found in [22]) and episturmian sequences (found in [66]). However, S-adic representation are (partially) known also for other classes of sequences, e.g., sequences coding interval exchange transformations (see [55]), sequences coding rotations (see [44]) or dendric sequences (see [21]).

In the sequel we introduce one very useful S-adic representation of episturmian sequences (and so Sturmian and Arnoux–Rauzy sequences, too). We use this notion in all our articles [A, B, C, D, E]. But first we recall needed facts about episturmian morphism.

Sturmian and episturmian morphisms

For every $a \in \mathcal{A}$ we define *elementary (fundamental) episturmian morphisms*:

$$L_a : \begin{cases} a \to a \\ b \to ab \quad \text{for all } b \neq a \end{cases} \qquad \text{and} \qquad R_a : \begin{cases} a \to a \\ b \to ba \quad \text{for all } b \neq a\,. \end{cases}$$

These $2\#\mathcal{A}$ morphisms generate the monoid $\mathcal{M}_{\mathcal{A}} = \langle L_a, R_a : a \in \mathcal{A}\rangle$ of *pure episturmian morphisms*. Let us remark that *episturmian morphisms* are all morphisms which can be obtained by composition of pure episturmian morphisms and permutations on $\mathcal{A}$. Similarly, *standard episturmian morphisms* are compositions of the morphisms from $\langle L_a : a \in \mathcal{A}\rangle$ and permutations. See [58, 66] for details.

Over a binary alphabet, (standard) episturmian morphisms are called *(standard) Sturmian morphisms*. The monoid $\mathcal{M} = \mathcal{M}_{\{0,1\}}$ of pure Sturmian morphisms is also called *special Sturmian monoid*. It is not difficult to realize that for any Sturmian morphism ψ either $\psi \in \mathcal{M}$ or $\psi^2 \in \mathcal{M}$.

Example 2.48. Fibonacci sequence $\mathbf{f}$ is the fixed point of the morphism $\varphi = 0 \to 01$, $1 \to 0$. Clearly, φ does not belong to $\mathcal{M}$. However, $\varphi = L_0 \circ E$ and so $\varphi^2 = L_0 L_1 \in \mathcal{M}$. Thus the Fibonacci sequence is fixed also by the pure Sturmian morphism $L_0 L_1$.

To express the morphisms from $\mathcal{M}_{\mathcal{A}}$, we use the following notation adopted from [59]. For a given alphabet $\mathcal{A}$ we define a new alphabet $\bar{\mathcal{A}} = \{\bar{a} : a \in \mathcal{A}\}$. We put $\varphi_a = L_a$ and $\varphi_{\bar{a}} = R_a$ for every letter $a \in \mathcal{A}$. Then for every word $z = z_0 z_1 \cdots z_{n-1} \in (\mathcal{A} \cup \bar{\mathcal{A}})^*$ we write

$$\varphi_z = \varphi_{z_0}\varphi_{z_1} \cdots \varphi_{z_{n-1}} \in \mathcal{M}_{\mathcal{A}}$$

and we say that z is a *directive word* of the morphism φ_z. A word is *L-spinned* (*R-spinned*, respectively) if all its letters are from $\mathcal{A}$ ($\bar{\mathcal{A}}$, respectively). The *opposite* word of z is obtained from z by switching spins of all its letters.

Example 2.49. The word 012 is L-spinned, while the word $\bar{0}\bar{1}\bar{2}$ is R-spinned. These two words are opposite words of each other.

The word $\bar{0}0\bar{1}\bar{2}0$ directs the morphism

$$\psi = \varphi_{\bar{0}0\bar{1}\bar{2}0} = R_0 L_0 R_1 R_2 L_0 \,.$$

However, the morphism ψ is also directed by the word $0012\bar{0}$, since one can easily verify that $\varphi_{\bar{0}0\bar{1}\bar{2}0} = \varphi_{0012\bar{0}}$.

We usually work with primitive morphisms. The primitivity of a pure episturmian morphism can be easily recognized from its directive word: the morphism $\varphi_z \in \mathcal{M}_\mathcal{A}$ is primitive if and only if its directive word z contains a or $\bar{a}$ for every letter $a \in \mathcal{A}$.

Example 2.50. The Sturmian morphism $\varphi_{\bar{0}0\bar{1}} : 0 \to 0010, 1 \to 010$ is primitive, while the Sturmian morphism $\varphi_{\bar{0}0\bar{0}} : 0 \to 0, 1 \to 0100$ is not primitive.

As indicated in Example 2.49, a pure episturmian morphism can have more than one directive word, i.e., the monoid $\mathcal{M}_\mathcal{A}$ is not free. Nevertheless, the presentation of the monoid $\mathcal{M}_\mathcal{A}$ is known. In fact, Richomme [95, Theorem 7.1] or [96, Proposition 6.5] described the presentation of the monoid of all episturmian morphisms, too.

Proposition 2.51 ([96, Proposition 6.5], [67, Theorem 2.2]). *The monoid of pure episturmian morphisms $\mathcal{M}_\mathcal{A}$ with generators $\{L_a : a \in \mathcal{A}\} \cup \{R_a : a \in \mathcal{A}\}$ has the following presentation:*

$$R_{a_1} R_{a_2} \cdots R_{a_k} L_{a_1} = L_{a_1} L_{a_2} \cdots L_{a_k} R_{a_1} \,, \tag{2.7}$$

where $k \in \mathbb{N}, k \geq 1$ and $a_1, a_2, \ldots, a_k \in \mathcal{A}$ with $a_1 \neq a_i$ for all i, $2 \leq i \leq k$.

This immediately implies that the monoid of pure standard episturmian morphisms is free.

In the notion of directive words, Relations (2.7) can be restated using the so-called block transformations introduced by Justin and Pirillo [67]. A *block-transformation* in the word $z \in (\mathcal{A} \cup \bar{\mathcal{A}})^*$ is the replacement of the factor $av\bar{a}$ of z, where $a \in \mathcal{A}$ and $v \in (\mathcal{A} \setminus \{a\})^*$, by the opposite word $\bar{a}\bar{v}a$ or vice-versa.

Proposition 2.52 ([96, Proposition 6.5], [67, Theorem 2.2]). *Let z, z' be two words over $\mathcal{A} \cup \bar{\mathcal{A}}$. Then $\varphi_z = \varphi_{z'}$ if and only if we can pass from z to z' by a chain of block-transformations.*

Example 2.53 (Example 2.49 continued). Using block-transformations we can rewrite:

$$\bar{0}0\bar{1}\bar{2}0 \longleftrightarrow 0\bar{0}\bar{1}\bar{2}0 \longleftrightarrow 0012\bar{0} \,.$$

Hence by the previous proposition all these words direct the same morphism, i.e., $\varphi_{\bar{0}0\bar{1}\bar{2}0} = \varphi_{0\bar{0}\bar{1}\bar{2}0} = \varphi_{0012\bar{0}}$. Equivalently, by Proposition 2.51 we rewrite:

$$R_a L_a R_b R_c R_a = L_a R_a R_b R_c L_a = L_a L_a L_b L_c R_a \,.$$

A directive word $z \in (\mathcal{A} \cup \bar{\mathcal{A}})^*$ is a *normalized directive word* if z has no factor from the set $\{\bar{a}\bar{\mathcal{A}}^* a : a \in \mathcal{A}\}$.

Example 2.54. (Example 2.53 continued) The directive words $\bar{0}0\bar{1}\bar{2}0$ and $0\bar{0}\bar{1}\bar{2}0$ of ψ are not normalized, while the directive word $0012\bar{0}$ of ψ is normalized.

Proposition 2.55 ([59, Lemma 5.3]). *Any pure episturmian morphism has the unique normalized directive word.*

Directive sequences of episturmian sequences

The following theorem ensures the existence of directive sequences of episturmian sequences.

Theorem 2.56 ([66, Theorem 3.10]). *A sequence* $\mathbf{u}$ *is episturmian if and only if there exist a sequence* $\mathbf{z} = z_0 z_1 z_2 \cdots \in (\mathcal{A} \cup \bar{\mathcal{A}})^{\mathbb{N}}$ *and an infinite sequence* $(\mathbf{u}^{(n)})_{n \geq 0}$ *of recurrent sequences such that* $\mathbf{u}^{(0)} = \mathbf{u}$ *and*

$$\mathbf{u}^{(n)} = \varphi_{z_n}(\mathbf{u}^{(n+1)}) \quad \text{for every } n \in \mathbb{N}\,. \tag{2.8}$$

This sequence $\mathbf{z}$ *is called a* directive sequence *of* $\mathbf{u}$.

In fact, Relation (2.8) can be restated as follows:

$$\mathbf{u} = \varphi_{z_0}\varphi_{z_1} \cdots \varphi_{z_n}(\mathbf{u}^{(n+1)})\,.$$

Hence we can understood the sequence $\mathbf{u}$ as an S-adic sequence with the set of morphisms $S = \{L_a : a \in \mathcal{A}\} \cup \{R_a : a \in \mathcal{A}\}$ and with the directive sequence $\mathbf{Z} = \varphi_{z_0}\varphi_{z_1}\varphi_{z_2}\cdots$.

Let us notice that in the case of a standard episturmian sequence $\mathbf{u}$ the directive sequence of $\mathbf{u}$ defined in Theorem 2.56 equals the directive sequence from the palindromic closure construction of $\mathbf{u}$ mentioned in Section 2.2.2 (e.g., see [58, Section 3]).

Clearly, an episturmian sequence over $\mathcal{A}$ is periodic if and only if its directive sequence $\mathbf{z}$ is of the form $\mathbf{z} = w\mathbf{a}$, where $w \in (\mathcal{A} \cup \bar{\mathcal{A}})^*$ and $\mathbf{a} \in \{a, \bar{a}\}^{\mathbb{N}}$ for some letter $a \in \mathcal{A}$. Similarly, the Arnoux–Rauzy (Sturmian) sequences can be easily recognised by their directive sequences, too (e.g., see [58, Section 2.3]).

Proposition 2.57. *An episturmian sequence* $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ *with the directive sequence* $\mathbf{z}$ *is an Arnoux–Rauzy sequence over* $\mathcal{A}$ *if and only if for every* $a \in \mathcal{A}$ *the letter* a *or* $\bar{a}$ *occurs infinitely many times in* $\mathbf{z}$.

In particular, fixed points of primitive episturmian morphisms are Arnoux–Rauzy sequences.

Remark 2.58. Theorem 2.56 and Proposition 2.57 immediately imply that for an Arnoux–Rauzy (Sturmian) sequence $\mathbf{u}$ each sequence $\mathbf{u}^{(i)}$ from Theorem 2.56 is an Arnoux–Rauzy (Sturmian) sequence with a directive sequence $z_i z_{i+1} z_{i+2} \cdots$.

Proposition 2.59 ([66, Proposition 3.11]).

(i) A sequence $\mathbf{z} \in (\mathcal{A} \cup \bar{\mathcal{A}})^{\mathbb{N}}$ *which has infinitely many L-spinned letters directs the unique episturmian sequence* $\mathbf{u}$.

(ii) A sequence $\mathbf{z} \in (\mathcal{A} \cup \bar{\mathcal{A}})^{\mathbb{N}}$ *which contains finitely many L-spinned letters directs one episturmian sequence for each* $\bar{a} \in \bar{\mathcal{A}}$ *which occurs in* $\mathbf{z}$ *infinitely many times.*

In particular, for a primitive episturmian morphism $\varphi_z \in \mathcal{M}_{\mathcal{A}}$ it means that

(i) if z contains at least one L-spinned letter, then φ_z has the unique fixed point;

(ii) otherwise, φ_z has $\#\mathcal{A}$ different fixed points.

In addition, an episturmian sequence can have more than one directive sequence. However, Glen, Levé, and Richomme [59] described all directive sequences which direct the same episturmian sequence.

Theorem 2.60 ([59, Theorem 4.1]). *Two sequences $\mathbf{z}^{(1)}, \mathbf{z}^{(2)} \in (\mathcal{A} \cup \bar{\mathcal{A}})^{\mathbb{N}}$ direct the same episturmian sequence if and only if one of the following cases holds for some i, j such that $\{i, j\} = \{1, 2\}$:*

(i) $\mathbf{z}^{(i)} = \prod_{n\geq 1} u^{(n)}$, $\mathbf{z}^{(j)} = \prod_{n\geq 1} v^{(n)}$, *where* $u^{(n)}$, $v^{(n)}$ *are words such that* $\varphi_{u^{(n)}} = \varphi_{v^{(n)}}$ *for all* $n \geq 1$*;*

(ii) $\mathbf{z}^{(i)} = wa\prod_{n\geq 1} u^{(n)}x^{(n)}$, $\mathbf{z}^{(j)} = w'\bar{a}\prod_{n\geq 1} \bar{u}^{(n)}y^{(n)}$, *where* w, w' *are words such that* $\varphi_w = \varphi_{w'}$, $a \in \mathcal{A}$ *and for all* $n \geq 1$, $u^{(n)}$ *is a non-empty a-free L-spinned word,* $\bar{u}^{(n)}$ *is the opposite word of* $u^{(n)}$ *and* $x^{(n)}$, $y^{(n)}$ *are non-empty words over* $\{a, \bar{a}\}$ *such that* $|x^{(n)}| = |y^{(n)}|$ *and* $|x^{(n)}|_a = |y^{(n)}|_a$.

(iii) $\mathbf{z}^{(i)} = w\mathbf{a}$ *and* $\mathbf{z}^{(j)} = w'\mathbf{b}$, *where* $a, b \in \mathcal{A}$, $\mathbf{a} \in \{a, \bar{a}\}^{\mathbb{N}}$, $\mathbf{b} \in \{b, \bar{b}\}^{\mathbb{N}}$ *and* w, w' *are words such that* $\varphi_w(a) = \varphi_{w'}(b)$.

Items (i) and (ii) are especially important for us since we focus on aperiodic episturmian sequences. One can notice that Item (i) is based on block-transformations of the directive words of episturmian morphisms, while Item (ii) brings new relations.

Example 2.61. We verify that the sequences $\mathbf{y} = (0\bar{1}\bar{2}\bar{0})^\omega$ and $\mathbf{z} = 0\bar{1}\bar{2}00(120\bar{0})^\omega$ direct the same Arnoux–Rauzy sequence. Indeed, we start with $\mathbf{y}$ and we set $u^{(1)} = 0\bar{1}\bar{2}$, $u^{(2k)} = \bar{0}0$ and $u^{(2k+1)} = \bar{1}\bar{2}$ for all $k > 0$. We make the block-transformations:

$$\underbrace{0\bar{1}\bar{2}}_{u^{(1)}}\underbrace{\bar{0}0}_{u^{(2)}}\underbrace{\bar{1}\bar{2}}_{u^{(3)}}\underbrace{\bar{0}0}_{u^{(4)}}\underbrace{\bar{1}\bar{2}}_{u^{(5)}}\underbrace{\bar{0}0}_{u^{(6)}}\cdots \quad\longleftrightarrow\quad \underbrace{0\bar{1}\bar{2}}_{u^{(1)}}\underbrace{0\bar{0}}_{u^{(2)}}\underbrace{\bar{1}\bar{2}}_{u^{(3)}}\underbrace{0\bar{0}}_{u^{(4)}}\underbrace{\bar{1}\bar{2}}_{u^{(5)}}\underbrace{0\bar{0}}_{u^{(6)}}\cdots$$

and we get the sequence $0\bar{1}\bar{2}(0\bar{0}\bar{1}\bar{2})^\omega$. Then we set $u^{(1)} = 0\bar{1}\bar{2}0$ and $u^{(k)} = \bar{0}\bar{1}\bar{2}0$ for all $k > 1$. After the block-transformations:

$$\underbrace{0\bar{1}\bar{2}0}_{u^{(1)}}\underbrace{\bar{0}\bar{1}\bar{2}0}_{u^{(2)}}\underbrace{\bar{0}\bar{1}\bar{2}0}_{u^{(3)}}\underbrace{\bar{0}\bar{1}\bar{2}0}_{u^{(4)}}\underbrace{\bar{0}\bar{1}\bar{2}0}_{u^{(5)}}\cdots \quad\longleftrightarrow\quad \underbrace{0\bar{1}\bar{2}0}_{u^{(1)}}\underbrace{012\bar{0}}_{u^{(2)}}\underbrace{012\bar{0}}_{u^{(3)}}\underbrace{012\bar{0}}_{u^{(4)}}\underbrace{012\bar{0}}_{u^{(5)}}\cdots$$

we get the sequence $0\bar{1}\bar{2}0(012\bar{0})^\omega$. Finally we set $u^{(1)} = 0\bar{1}\bar{2}00$, $u^{(2k)} = 12$ and $u^{(2k+1)} = \bar{0}0$ for all $k > 0$, and the block-transformations

$$\underbrace{0\bar{1}\bar{2}00}_{u^{(1)}}\underbrace{12}_{u^{(2)}}\underbrace{\bar{0}0}_{u^{(3)}}\underbrace{12}_{u^{(4)}}\underbrace{\bar{0}0}_{u^{(5)}}\underbrace{12}_{u^{(6)}}\cdots \quad\longleftrightarrow\quad \underbrace{0\bar{1}\bar{2}00}_{u^{(1)}}\underbrace{12}_{u^{(2)}}\underbrace{0\bar{0}}_{u^{(3)}}\underbrace{12}_{u^{(4)}}\underbrace{0\bar{0}}_{u^{(5)}}\underbrace{12}_{u^{(6)}}\cdots$$

lead us to the sequence $\mathbf{z} = 0\bar{1}\bar{2}00(120\bar{0})^\omega$.

Also the sequences $\mathbf{y} = (\bar{0}\bar{1}\bar{2})^\omega$ and $\mathbf{z} = \bar{0}1(20\bar{1})^\omega$ direct the same episturmian sequence, since in the notation of Item (ii) of Theorem 2.60 we have

$$\mathbf{y} = \underbrace{\bar{0}}_{w'}\underbrace{\bar{1}}_{\bar{a}}\underbrace{\bar{2}\bar{0}}_{\bar{u}^{(1)}}\underbrace{\bar{1}}_{y^{(1)}}\underbrace{\bar{2}\bar{0}}_{\bar{u}^{(2)}}\underbrace{\bar{1}}_{y^{(2)}}\cdots \quad\text{and}\quad \mathbf{z} = \underbrace{\bar{0}}_{w}\underbrace{1}_{a}\underbrace{20}_{u^{(1)}}\underbrace{\bar{1}}_{y^{(1)}}\underbrace{20}_{u^{(2)}}\underbrace{\bar{1}}_{y^{(2)}}\cdots.$$

A directive sequence $\mathbf{z} \in (\mathcal{A} \cup \bar{\mathcal{A}})^{\mathbb{N}}$ is *normalized* if it contains infinitely many L-spinned letters, but no factor from the set $\{\bar{a}\bar{\mathcal{A}}^*a : a \in \mathcal{A}\}$. By Proposition 2.59, every normalized directive sequence directs exactly one episturmian sequence.

Example 2.62 (Example 2.61 continued). The sequences $(0\bar{1}\bar{2}\bar{0})^\omega$ and $(\bar{0}\bar{1}\bar{2})^\omega$ are not normalized, while the sequences $0\bar{1}\bar{2}00(120\bar{0})^\omega$ and $\bar{0}1(20\bar{1})^\omega$ are normalized directive sequences.

Proposition 2.63 ([59, Theorem 5.2]). *Any aperiodic episturmian sequence* $\mathbf{u}$ *has a unique normalized directive sequence.*

We remark that this unambiguity need not hold for periodic episturmian sequences, see [59].

Moreover, the normalized directive sequences can be constructed using Theorem 2.60. If a directive sequence does not contain infinitely many L-spinned letters, then we use Item (ii) to find another one with infinitely many L-spinned letters. If a directive sequence contains infinitely many L-spinned letters, then it can be normalized by repeated applications of Item (i). See [59, Section 5] for more details.

Example 2.64. By Proposition 2.59, the sequence $\mathbf{z} = (\bar{0}\bar{1}\bar{2})^\omega$ directs three Arnoux–Rauzy sequences starting with the letters $0, 1, 2$, respectively. All these sequences have the same language as the Tribonacci sequence $\mathbf{t}$ with the directive sequence $(012)^\omega$. Their normalized directive sequences are $0(12\bar{0})^\omega$, $\bar{0}1(20\bar{1})^\omega$ and $\bar{0}\bar{1}2(01\bar{2})^\omega$, respectively (see Example 2.61).

Finally, let us point out that a sequence which is fixed by a primitive episturmian morphism has to have a purely periodic directive sequence. However, its normalized directive sequence need not be purely periodic (see Example 2.62). Nevertheless, the normalized directive sequence of a substitutive sequence is always periodic.

Chapter 3

Aims and results of the thesis

This chapter is dedicated to a brief summary of the main results of this thesis. First of all, we focus on the derived sequences of Sturmian, Arnoux–Rauzy and complementary symmetric (CS) Rote sequences (Section 3.1). We also study the non-repetitive complexity of Arnoux–Rauzy sequences (Section 3.2) and derive formulas for the critical exponent (Section 3.3) and the recurrence function (Section 3.4) of CS Rote sequences. To make the summary clearer we partially unified the notation, although the original papers [A, B, C, D, E] differ slightly in some aspects.

3.1 Derived sequences

A substantial part of this thesis is devoted to the investigation of derived sequences in the case of sequences with low factor complexity. The notions of return words and derived sequences were described in detail in Section 2.3.1. We just recall that the derived sequence $\mathbf{d_u}(w)$ of $\mathbf{u}$ with respect to a non-empty prefix w of $\mathbf{u}$ expresses the order of return words to w in the sequence $\mathbf{u}$.

Although Durand [49] proved that a uniformly recurrent sequence is primitive substitutive if and only if it has finite number of derived sequences (see Theorem 2.34), many related questions remain open.

Our main aim is to describe the set $\mathrm{Der}(\mathbf{u})$ of all derived sequences of $\mathbf{u}$ with respect to its non-empty prefixes in the case when $\mathbf{u}$ is Sturmian (Section 3.1.1), Arnoux–Rauzy (Section 3.1.2) or CS Rote sequence (Section 3.1.3). Let us recall that two derived sequences which differ only by a permutation of letters are identified with one another and counted as one derived sequence.

3.1.1 Derived sequences of Sturmian sequences

The aim of the article [A] is to study derived sequences of Sturmian sequences. In particular, we precisely describe the set $\mathrm{Der}(\mathbf{u})$ for Sturmian sequences which are the fixed points of primitive Sturmian morphisms. Sturmian sequences are discussed in Section 2.2.1; let us now just recall that Araújo and Bruyère [6] described derived sequences of standard Sturmian sequences using continued fraction expansions of their slopes. Nevertheless, we consider also the more general case of non-standard sequences.

Vuillon [104] proved that every Sturmian sequence has two return words to each of its factors (see Theorem 2.28). Thus the derived sequences of Sturmian sequences are binary and, moreover, they are Sturmian by Theorem 2.28 and Proposition 2.33.

Proposition 3.1. *If* $\mathbf{u}$ *is a Sturmian sequence and* w *is a prefix of* $\mathbf{u}$*, then the derived sequence* $\mathbf{d_u}(w)$ *is Sturmian as well.*

We focus especially on the fixed points of primitive Sturmian morphisms. Hence we can restrict ourselves on primitive Sturmian morphisms from the special Sturmian monoid $\mathcal{M}$ without loss of generality. Indeed, by Section 2.4.2 for every Sturmian morphism ψ either $\psi \in \mathcal{M}$ or $\psi^2 \in \mathcal{M}$ and the fixed points of ψ are fixed also by ψ^2.

Our key tool is the decomposition of Sturmian morphisms from $\mathcal{M}$ into elementary Sturmian morphisms, which was described in Section 2.4.2. We recall that for $\psi \in \mathcal{M}$ we write

$$\psi = \varphi_w = \varphi_{w_0}\varphi_{w_1}\cdots\varphi_{w_{n-1}}\,,$$

where $w \in (\mathcal{A} \cup \bar{\mathcal{A}})^* = \{0, 1, \bar{0}, \bar{1}\}^*$ is the directive word and all φ_{w_i} are elementary Sturmian morphisms. Since this decomposition need not be unique, we use the so-called normalized directive word: a directive word z is normalized if it does not contain a factor of the form $\bar{a}\bar{\mathcal{A}}^*a$ for any $a \in \mathcal{A}$. Normalized directive words are unique and can be easily found by Proposition 2.52.

In fact, this is closely related to S-adic representations of Sturmian sequences explained in Section 2.4.2. Indeed, if a sequence $\mathbf{u}$ is a fixed point of a morphism φ_w, then the directive sequence of $\mathbf{u}$ is $\mathbf{z} = w^\omega$ and the sequence $\mathbf{u}$ can be expressed in the form

$$\begin{aligned}\mathbf{u} = \varphi_w(\mathbf{u}) &= \varphi_{w_0}(\mathbf{u}^{(1)})\,, \text{ where } \mathbf{u}^{(1)} \text{ is fixed by } \varphi_{w_1\cdots w_{n-1}w_0}\,;\\ &= \varphi_{w_0}\varphi_{w_1}(\mathbf{u}^{(2)})\,, \text{ where } \mathbf{u}^{(2)} \text{ is fixed by } \varphi_{w_2\cdots w_{n-1}w_0w_1}\,;\\ &\text{etc.}\end{aligned}$$

Remark 3.2. Let us briefly comment on the notation. While the article [A] denotes the elementary Sturmian morphisms by φ_b, φ_β, $\varphi_a, \varphi_\alpha$, in this summary we prefer the notation L_0, L_1, R_0, R_1, or $\varphi_0, \varphi_1, \varphi_{\bar{0}}, \varphi_{\bar{1}}$ from Section 2.4.2. Hence we would like to emphasize their connections:

$$\varphi_b = L_0 = \varphi_0\,, \qquad \varphi_\beta = L_1 = \varphi_1\,, \qquad \varphi_a = R_0 = \varphi_{\bar{0}} \quad \text{and} \qquad \varphi_\alpha = R_1 = \varphi_{\bar{1}}\,.$$

Thus in [A] the directive word (called *name*) of ψ is a word over the alphabet $\{a, \alpha, b, \beta\}$ instead of $\mathcal{A} \cup \bar{\mathcal{A}} = \{0, 1, \bar{0}, \bar{1}\}$.

The structure of any elementary Sturmian morphism φ is simple enough to enable us to precisely describe the relation between the derived sequences of a sequence $\mathbf{u}$ and its preimage $\mathbf{u}'$, where $\mathbf{u} = \varphi(\mathbf{u}')$. In fact, special factors and return words of $\mathbf{u}$ and $\mathbf{u}'$ are closely linked, too (see [A, Section 3]).

Proposition 3.3. *Let* $\mathbf{u}, \mathbf{u}'$ *be Sturmian sequences and* $c \in \{0, 1\}$*.*

(i) If $\mathbf{u} = L_c(\mathbf{u}')$*, then* $\mathrm{Der}(\mathbf{u}) = \mathrm{Der}(\mathbf{u}') \cup \{\mathbf{u}'\}$.

(ii) If $\mathbf{u} = R_c(\mathbf{u}')$ *and* $\mathbf{u}$ *starts with a letter* $d \in \{0, 1\}$*,* $d \neq c$*, then* $\mathrm{Der}(\mathbf{u}) = \mathrm{Der}(\mathbf{u}')$.

Roughly speaking, we can gradually desubstitute the sequence $\mathbf{u}$ onto its preimages $\mathbf{u}^{(1)}$, $\mathbf{u}^{(2)}$, etc. and determine the derived sequences of $\mathbf{u}$ from this process using the proposition above. In fact, we have to do it more properly to ensure that we do not omit any derived sequence.

Although we use these observations only for Sturmian sequences which are the fixed points of primitive Sturmian morphisms, they can be similarly applied to general Sturmian sequences (i.e., to eventually periodic or even aperiodic directive sequences). We deal with these cases in the article [B].

Durand [49] shoved that all derived sequences of a sequence fixed by a primitive morphism are also fixed points of primitive morphisms, see Section 2.3.1 for further details. We provide an algorithm which for a given Sturmian morphism ψ lists the morphisms fixing the derived sequences of the fixed point of ψ.

First we focus on the case of a primitive Sturmian morphism $\psi = \varphi_w \in \mathcal{M}$ having the unique fixed point. By Proposition 2.59 its directive word w contains at least one letter from $\mathcal{A}$, i.e., $w \in (\mathcal{A} \cup \bar{\mathcal{A}})^* \setminus \bar{\mathcal{A}}^*$. Moreover, its normalized directive word z is of the form $z = \bar{a}^k b z'$ for some $a, b \in \mathcal{A}, a \neq b$, $k \in \mathbb{N}$ and $z' \in (\mathcal{A} \cup \bar{\mathcal{A}})^*$. Then we define the transformation Δ by

$$\Delta(z) = N(z'\bar{a}^k b) \quad \text{and} \quad \Delta(\psi) = \varphi_{\Delta(z)}\,, \tag{3.1}$$

where $N(v)$ is the normalization of the word v (i.e., it is a normalized directive word of the morphism φ_v). We have to normalize the obtained word since we want to apply the transformation Δ repeatedly and Δ acts only on normalized words.

Example 3.4. We consider the primitive morphism $\psi = \varphi_z$ with the normalized directive word $z = 1\bar{0}\bar{1}$. We apply repeatedly the transformation Δ on ψ:

$$\begin{aligned}
\Delta(1\bar{0}\bar{1}) &= N(\bar{0}\bar{1}1) = \bar{0}1\bar{1} & \Delta(\psi) &= \varphi_{\bar{0}1\bar{1}}\\
\Delta(\bar{0}1\bar{1}) &= N(\bar{1}\bar{0}1) = 10\bar{1} & \Delta^2(\psi) &= \varphi_{10\bar{1}}\\
\Delta(10\bar{1}) &= N(0\bar{1}1) = 01\bar{1} & \Delta^3(\psi) &= \varphi_{01\bar{1}}\\
\Delta(01\bar{1}) &= N(1\bar{1}0) = 1\bar{1}0 & \Delta^4(\psi) &= \varphi_{1\bar{1}0}\\
\Delta(1\bar{1}0) &= N(\bar{1}01) = \bar{1}01 & \Delta^5(\psi) &= \varphi_{\bar{1}01}\\
\Delta(\bar{1}01) &= N(1\bar{1}0) = 1\bar{1}0 & \Delta^6(\psi) &= \Delta^4(\psi)
\end{aligned}$$

Theorem 3.5 ([A, Theorem 25]). *Let $\psi = \varphi_z \in \mathcal{M}$ be a primitive Sturmian morphism with the normalized directive word $z \in (\mathcal{A} \cup \bar{\mathcal{A}})^* \setminus \bar{\mathcal{A}}^*$, i.e., ψ has the unique fixed point $\mathbf{u}$. Then $\mathbf{x}$ is a derived sequence of $\mathbf{u}$ with respect to one of its non-empty prefixes if and only if $\mathbf{x}$ is the fixed point of the morphism $\Delta^m(\psi)$ for some $m \geq 1$.*

Example 3.6. The unique fixed point $\mathbf{u}$ of the morphism $\psi = \varphi_{1\bar{0}\bar{1}}$ considered in Example 3.4 has five distinct derived sequences. They are fixed by morphisms $\Delta(\psi)$, $\Delta^2(\psi)$, $\Delta^3(\psi)$, $\Delta^4(\psi)$ and $\Delta^5(\psi)$, respectively. Nevertheless, only the fixed points of $\Delta^4(\psi)$ and $\Delta^5(\psi)$ represent the derived sequences of $\mathbf{u}$ to infinitely many prefixes of $\mathbf{u}$.

The remaining case of Sturmian sequences with two fixed points can be easily transformed to Theorem 3.5 by the following proposition.

Proposition 3.7 ([A, Proposition 29]). *Let $\psi = \varphi_z \in \mathcal{M}$ be a primitive Sturmian morphism which has two fixed points, i.e., $z = z_0 z_1 \cdots z_{n-1} \in \bar{\mathcal{A}}^*$. We denote c the letter from $\mathcal{A}$ such that $z_0 = \bar{c}$ and d the other letter from $\mathcal{A}$.*

(i) The fixed point $\mathbf{u}$ of ψ starting with the letter c has $\mathrm{Der}(\mathbf{u}) = \mathrm{Der}(\mathbf{v}) \cup \{\mathbf{v}\}$, where $\mathbf{v}$ is the unique fixed point of the morphism φ_v with $v = c^{-1}N(zc) \in \{\bar{c}, d\}^*$.

(ii) The fixed point $\mathbf{x}$ of ψ starting with the letter d has $\mathrm{Der}(\mathbf{x}) = \mathrm{Der}(\mathbf{y})$, where $\mathbf{y}$ is the fixed point starting with the letter d of the morphism φ_y with $y = z_1 z_2 \cdots z_{n-1} z_0$.

We give a sharp bound on the cardinality of $\mathrm{Der}(\mathbf{u})$, too.

Proposition 3.8 ([A, Corollary 35 and Proposition 37]). *Let $\psi = \varphi_w \in \mathcal{M}$ be a primitive Sturmian morphism. Then for its fixed point $\mathbf{u}$ we have*

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

Moreover, both bounds are attained for infinitely many morphisms which are not powers of any other morphisms.

Example 3.9. For every $n \geq 2$ the fixed point of the primitive Sturmian morphism $\psi = \varphi_z$, where $z = (\bar{a})^{n-1} b$ for $a, b \in \mathcal{A}, a \neq b$, has one derived sequence. This derived sequence is also fixed by the morphism ψ.

For every $n \geq 2$ the fixed point of the primitive Sturmian morphism $\psi = \varphi_z$, where $z = a^{n-2}\bar{b}\bar{a}$ for $a, b \in \mathcal{A}, a \neq b$, has $3n-4$ distinct derived sequences. See also Example 3.6 or [A, Example 33].

In addition, for the fixed points of two special classes of Sturmian morphisms we determine the precise numbers of their distinct derived sequences.

Proposition 3.10 ([A, Propositions 36 and 37]). *Let $\psi = \varphi_z \in \mathcal{M}$ be a primitive Sturmian morphism which is not a power of any other morphism.*

(i) If ψ is standard, i.e., $z \in \mathcal{A}^*$, then its unique fixed point has $|z|$ derived sequences. Moreover, if $z = z_0 z_1 \cdots z_{n-1}$, then their fixing morphisms have the directive words $z_1 z_2 \cdots z_{n-1} z_0$, $z_2 z_3 \cdots z_{n-1} z_0 z_1$, $\ldots$, $z_{n-1} z_0 \cdots z_{n-2}$ and $z_0 z_1 \cdots z_{n-1}$, respectively.

(ii) If ψ has two fixed points, i.e., $z \in \bar{\mathcal{A}}^*$, then its fixed point starting with the letter a has $1 + |z|_{\bar{b}}$ derived sequences, where $a, b \in \mathcal{A}, a \neq b$.

Proposition 3.8 and 3.10 can be stated analogously also for primitive Sturmian morphisms which are not included in the special Sturmian monoid $\mathcal{M}$, i.e., for the morphisms of the form $\psi = \varphi_w \circ E$, where $\varphi_w \in \mathcal{M}$ and E is the morphism which exchanges the letters $0 \leftrightarrow 1$.

To give the exact number of derived sequences, one needs to describe when the normalized directive word z corresponds to some power of a Sturmian morphism. This is not trivial, since z may be a normalized directive word of a power of a morphism without z being a power of some other word. For example, if $z = \bar{1}0\bar{0}\bar{1}\bar{1}$, then the normalized directive word of $(\varphi_z)^3$ is the primitive word $N(z^3) = \bar{1}001110\bar{0}111\bar{0}\bar{0}\bar{1}\bar{1}$.

3.1.2 Derived sequences of Arnoux–Rauzy sequences

In the article [B] we generalize the results of [A] to the case of Arnoux–Rauzy sequences (see Section 2.2.2). We use a similar technique based on the representation of Arnoux–Rauzy sequence by episturmian morphisms. However, we slightly modify the notation, which enables us to comfortably work also with primitive substitutive sequences (and not only with fixed points of primitive morphisms). More precisely, we use the normalized directive sequences (explained in Section 2.4.2) of Arnoux–Rauzy sequences which were introduced by Glen, Levé and Richomme [59].

Let us recall that by Theorem 2.56 and Proposition 2.57 each Arnoux–Rauzy sequence $\mathbf{u}$ has a directive sequence $\mathbf{z} \in (\mathcal{A} \cup \bar{\mathcal{A}})^{\mathbb{N}}$ such that, for every $a \in \mathcal{A}$, a or $\bar{a}$ occurs infinitely many times in $\mathbf{z}$, and a sequence of its recurrent preimages $(\mathbf{u}^{(n)})_{n\geq 0}$ such that

$$\mathbf{u} = \varphi_{z_0 z_1 \cdots z_{n-1}}(\mathbf{u}^{(n)}) \quad \text{for every } n \in \mathbb{N}\,. \tag{3.2}$$

These directive sequences do not have to be unique, but every Arnoux–Rauzy sequence has a unique normalized directive sequence (see Proposition 2.63) which can be constructed using Theorem 2.60. The directive sequence is normalized if it contains infinitely many letters from $\mathcal{A}$, but no factor from the set $\{\bar{a}\bar{\mathcal{A}}^* a : a \in \mathcal{A}\}$.

Justin and Vuillon [65] showed that every Arnoux–Rauzy sequence over $\mathcal{A}$ has $\#\mathcal{A}$ return words to each of its factors. Hence the corresponding derived sequences can be considered over the same alphabet $\mathcal{A}$. Nevertheless, it is not clear if these sequences are also Arnoux–Rauzy. In particular, we cannot use the same argument as in the case of Sturmian sequences since the number of return words does not characterize Arnoux–Rauzy sequences (see Section 2.3.1 for more details).

We first deduce the following proposition which is completely analogous to Proposition 3.3 from the previous section.

Proposition 3.11 ([B, Corollary 21]). *Let $\mathbf{u}, \mathbf{u}'$ be Arnoux–Rauzy sequences over $\mathcal{A}$ and $a \in \mathcal{A}$.*

(i) If $\mathbf{u} = L_a(\mathbf{u}')$, then $\mathrm{Der}(\mathbf{u}) = \mathrm{Der}(\mathbf{u}') \cup \{\mathbf{u}'\}$.

(ii) If $\mathbf{u} = R_a(\mathbf{u}')$ and $\mathbf{u}$ starts with a letter $b \in \mathcal{A}, b \neq a$, then $\mathrm{Der}(\mathbf{u}) = \mathrm{Der}(\mathbf{u}')$.

Hence, we can again determine the derived sequences of a given Arnoux–Rauzy sequence by analyzing the process of desubstitution of $\mathbf{u}$ onto its preimages $\mathbf{u}^{(i)}$ which is controlled by its directive sequence $\mathbf{z}$ (see Relation (3.2)). This proposition also indicates that only the letters of $\mathbf{z}$ which are from $\mathcal{A}$ (in [B] they are called L-spinned letters) are important. This leads us to the following definition of the transformation Δ (compare with (3.1)): Let $\mathbf{z} = z_0 z_1 z_2 \cdots \in (\mathcal{A} \cup \bar{\mathcal{A}})^{\mathbb{N}}$ be a normalized directive sequence. Then Δ applied to $\mathbf{z}$ removes the smallest prefix of $\mathbf{z}$ ending with a letter from $\mathcal{A}$ (L-spinned letter), i.e.,

$$\Delta(\mathbf{z}) = z_{k+1} z_{k+2} z_{k+3} \cdots\,, \text{ where } k \text{ is the smallest index such that } z_k \in \mathcal{A}.$$

One can notice that the sequence $\Delta(\mathbf{z})$ is still normalized. Hence we can apply this transformation repeatedly.

Example 3.12. We consider the normalized directive sequence $\mathbf{z} = \bar{2}10(\bar{2}1\bar{0}1)^\omega$ and we apply repeatedly the transformation Δ on $\mathbf{z}$:

$$\begin{aligned}
\Delta(\mathbf{z}) &= 0(\bar{2}1\bar{0}1)^\omega\,;\\
\Delta^2(\mathbf{z}) &= (\bar{2}1\bar{0}1)^\omega\,;\\
\Delta^3(\mathbf{z}) &= \bar{0}1(\bar{2}1\bar{0}1)^\omega = (\bar{0}1\bar{2}1)^\omega\,;\\
\Delta^4(\mathbf{z}) &= (\bar{2}1\bar{0}1)^\omega = \Delta^2(\mathbf{z})\,.
\end{aligned}$$

Using this Δ notation we can easily describe all the derived sequences of a given Arnoux–Rauzy sequence.

Theorem 3.13 ([B, Theorem 24]). *Let $\mathbf{u}$ be an Arnoux–Rauzy sequence over $\mathcal{A}$ with the normalized directive sequence $\mathbf{z}$. Then a sequence $\mathbf{x}$ is a derived sequence of $\mathbf{u}$ with respect to one of its non-empty prefixes if and only if $\mathbf{x}$ is an Arnoux–Rauzy sequence directed by $\Delta^m(\mathbf{z})$ for some $m \geq 1$, i.e.,*

$$\mathrm{Der}(\mathbf{u}) = \{\textit{sequence directed by } \Delta^m(\mathbf{z}) : m \geq 1\}\,.$$

Example 3.14. We consider the Arnoux–Rauzy sequence $\mathbf{u}$ directed by the normalized directive sequence $\mathbf{z} = \bar{2}10(\bar{2}1\bar{0}1)^\omega$ from Example 3.12. One can easily verify that the sequences directed by $(\bar{2}1\bar{0}1)^\omega$ and $(\bar{0}1\bar{2}1)^\omega$ are the same up to the exchange of letters $0 \leftrightarrow 2$. Hence we consider them as the same derived sequence. We may conclude that the sequence $\mathbf{u}$ has two derived sequences directed by $\Delta(\mathbf{z})$ and $\Delta^2(\mathbf{z})$. However, only the derived sequence directed by $\Delta^2(\mathbf{z})$ appears for infinitely many prefixes of $\mathbf{u}$.

Let us mention that while the description of derived sequences of standard Arnoux–Rauzy sequences can be easily deduced from the work of Justin and Vuillon [65], we cover also the more complicated case of non-standard sequences.

Since the sequence directed by $\Delta^m(\mathbf{z})$, where $m \geq 1$, is obviously Arnoux–Rauzy sequence, we immediately obtain the following corollary that confirms the natural conjecture that derived sequences of Arnoux–Rauzy sequences are also Arnoux–Rauzy sequences.

Corollary 3.15 ([B, Corollary 25]). *Each derived sequence with respect to a non-empty prefix of a given Arnoux–Rauzy sequence over $\mathcal{A}$ is an Arnoux–Rauzy sequence over $\mathcal{A}$ as well.*

Finally, we focus on Arnoux–Rauzy sequences with periodic directive sequences. If an Arnoux–Rauzy sequence $\mathbf{u}$ is the fixed point of a primitive episturmian morphism φ_w, i.e., its (not necessarily normalized) directive sequence is $\mathbf{w} = w^\omega$, then

$$1 \leq \mathrm{Der}(\mathbf{u}) \leq 3|w| - 2\#\mathcal{A}$$

and both bounds are attained for infinitely many sequences. Since the statement and its proof are completely analogous to the Sturmian case (Proposition 3.8), this is not included in the paper [B].

It is possible to similarly bound the cardinality of $\mathrm{Der}(\mathbf{u})$ also for a primitive substitutive Arnoux–Rauzy sequence $\mathbf{u} = \tau(\psi^\omega(a))$ in terms of the lengths of decompositions of τ and ψ into elementary episturmian morphisms. Moreover, if we know the normalized directive sequence of $\mathbf{u}$, we can determine the exact number of derived sequences of $\mathbf{u}$.

Proposition 3.16 ([B, Corollary 29]). *Let* $\mathbf{u}$ *be an Arnoux–Rauzy sequence over* $\mathcal{A}$ *with the eventually periodic normalized directive sequence* $\mathbf{z} = x\left(yP(y)\cdots P^{n-1}(y)\right)^\omega \in (\mathcal{A}\cup\bar{\mathcal{A}})^{\mathbb{N}}$, *where the words* x *and* y *are the shortest possible and* P *is a permutation on* $\mathcal{A}$ *of the order* n. *We denote* $|xy|_\mathcal{A}$ *the numbers of letters from* $\mathcal{A}$ *in the word* xy.

(i) If the last letters of both x, y *are L-spinned, then* $\#\mathrm{Der}(\mathbf{u}) = |xy|_\mathcal{A} - 1$.

(ii) If the last letter of x *or* y *is R-spinned or* $x = \varepsilon$, *then* $\#\mathrm{Der}(\mathbf{u}) = |xy|_\mathcal{A}$.

Example 3.17. The Tribonacci sequence $\mathbf{t}$ (see Example 2.15) with the normalized directive sequence $\mathbf{z} = (012)^\omega = (0P(0)P^2(0))^\omega$, where $P: 0 \to 1, 1 \to 2, 2 \to 0$, has only one derived sequence and it is equal to $\mathbf{t}$.

On the other hand, the Arnoux–Rauzy sequence with the normalized directive sequence $\mathbf{z} = 0(12\bar{0})^\omega$, which has the same language as $\mathbf{t}$ (see Example 2.64), has three derived sequences directed by $\Delta(\mathbf{z}) = (12\bar{0})^\omega$, $\Delta^2(\mathbf{z}) = (2\bar{0}1)^\omega$ and $\Delta^3(\mathbf{z}) = (\bar{0}12)^\omega$, respectively.

Let us mention one related open question. For any Sturmian sequence $\mathbf{u}$ we can decide if $\mathbf{u}$ is a fixed point of a primitive morphism by the well-known Yasutomi's condition [107]. However, we do not known any analogous result for Arnoux–Rauzy sequences over the alphabet of size grater than two.

In particular, it means that we are not able to (easily) recognize Arnoux–Rauzy sequences which are fixed points from their normalized directive sequences. Indeed, an Arnoux–Rauzy sequence which is a fixed point has to have a purely periodic directive sequence, but its normalized directive sequence can be eventually periodic with a non-empty pre-period.

Example 3.18. The fixed point of the morphism $\varphi_{1\bar{0}\bar{2}\bar{1}}$ has the directive sequence $(1\bar{0}\bar{2}\bar{1})^\omega$, but its normalized directive sequence is $\mathbf{z} = 1\bar{0}\bar{2}11(021\bar{1})^\omega$.

Hence some derived sequences may be falsely considered to not be fixed by a primitive morphism.

Example 3.19. The primitively substitutive sequence $\mathbf{u}$ with the normalized directive sequence $\mathbf{z} = 11\bar{0}\bar{2}11(021\bar{1})^\omega$ is not the fixed point of any morphism. Although, all its derived sequences with respect to its non-empty prefixes are fixed points of morphisms:

$$
\begin{aligned}
\Delta(\mathbf{z}) &= 1\bar{0}\bar{2}11(021\bar{1})^\omega = (1\bar{0}\bar{2}\bar{1})^\omega\,; & \Delta^5(\mathbf{z}) &= (21\bar{1}0)^\omega\,;\\
\Delta^2(\mathbf{z}) &= \bar{0}\bar{2}11(021\bar{1})^\omega = (\bar{0}\bar{2}1\bar{1})^\omega\,; & \Delta^6(\mathbf{z}) &= (1\bar{1}02)^\omega\,;\\
\Delta^3(\mathbf{z}) &= 1(021\bar{1})^\omega = (102\bar{1})^\omega\,; & \Delta^7(\mathbf{z}) &= (\bar{1}021)^\omega\,;\\
\Delta^4(\mathbf{z}) &= (021\bar{1})^\omega\,; & \Delta^8(\mathbf{z}) &= \Delta^5(\mathbf{z})\,.
\end{aligned}
$$

From this point of view, the method used in [A] is for fixed points slightly more informative, since it enable us to obtain directly the fixing morphisms of derived sequences.

3.1.3 Derived sequences of complementary symmetric Rote sequences

The article [C] focuses on the return words and derived sequences of complementary symmetric (CS) Rote sequences which were introduced in Section 2.2.4.

First, we easily deduce from some results of [11] that every non-empty prefix of a CS Rote sequence has exactly three return words. The same also follows directly from Proposition 2.30, since CS Rote sequences are neutral sequences of characteristic 0. Thus derived sequences of a CS Rote sequence with respect to its non-empty prefixes are ternary sequences.

The detailed study of return words and derived sequences is based on Theorem 2.23 which expresses the link between CS Rote sequences and Sturmian sequences: the sequence $\mathbf{v}$ is a CS Rote sequence if and only if $\mathbf{u} = \mathcal{S}(\mathbf{v})$ is a Sturmian sequence. These sequences $\mathbf{u}$ and $\mathbf{v}$ are called *associated*. Let us recall that the mapping $\mathcal{S}$ assigns to each sequence $\mathbf{v} \in \{0,1\}^{\mathbb{N}}$ the sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$ such that

$$u_i = v_i + v_{i+1} \mod 2 \quad \text{for all} \quad i \in \mathbb{N}\,.$$

The mapping $\mathcal{S}$ is analogously defined also for non-empty words. For every $v_0 \in \{0,1\}$ we put $\mathcal{S}(v_0) = \varepsilon$ and for every $v = v_0v_1\cdots v_n \in \{0,1\}^+$ of length at least 2 we put $\mathcal{S}(v_0v_1\cdots v_n) = u_0u_1\cdots u_{n-1}$, where

$$u_i = v_i + v_{i+1} \mod 2 \quad \text{for all} \quad i \in \{0,1,\ldots,n-1\}\,.$$

Proposition 2.26 shows that the factors of associated sequences $\mathbf{v}$ and $\mathbf{u}$ are closely related, too. Using Theorem 2.23 and Proposition 2.26 we transform our task to precise description of return words and derived sequences of Sturmian sequence studied in [A].

We would like to emphasize that we study only CS Rote sequences associated with standard Sturmian sequences (in [26] they are called *standard Rote sequences*). Let us recall that for a standard Sturmian sequence $\mathbf{u}$ it suffices to determine return words and derived sequences to its bispecial factors, since those factors coincides with the right special prefixes of $\mathbf{u}$. In addition, the structure of derived sequences of $\mathbf{u}$ is quite simple (see Item (i) of Proposition 3.10). Nevertheless, it seems that it is possible to do the same also for general CS Rote sequences.

The form of return words (and so derived sequences) of a CS Rote sequence $\mathbf{v}$ depends on the number of ones in the return words of the associated Sturmian sequence $\mathbf{u}$. Hence we define the following notion of stability. A word $u = u_0u_1\cdots u_{n-1} \in \{0,1\}^*$ is called *stable* (S) if $|u|_1 = 0 \mod 2$. Otherwise, it is called *unstable* (U).

Example 3.20. The word $u = 0110101$ is stable while the word $v = 011010$ is unstable.

Then we classify prefixes of a standard Sturmian sequence into three types according to the stability of their return words. Let w be a prefix of a standard Sturmian sequence $\mathbf{u}$ and let r, s be its return words and k be a positive integer such that $\mathbf{u}$ is a concatenation of the blocks r^ks and $r^{k+1}s$. Then the type $\mathcal{T}_w$ of w is

(i) $\mathcal{T}_w = SU(k)$, if r is stable and s is unstable;

(ii) $\mathcal{T}_w = US(k)$, if r is unstable and s is stable;

(iii) $\mathcal{T}_w = UU(k)$, if both r and s are unstable.

Example 3.21. We consider the Fibonacci sequence $\mathbf{f}$. Its empty prefix ε has return words $r = 0$ and $s = 1$ and since $\mathbf{f}$ is composed of the blocks 01 and 001, the respective parameter is $k = 1$. Hence the type of ε is $\mathcal{T}_\varepsilon = SU(1)$.

The prefix 0 of $\mathbf{f}$ has return words $r = 01$ and $s = 0$. Since $\mathbf{f}$ is composed of the blocks 010 and 01010, the parameter is $k = 1$. Hence the type of 0 is $\mathcal{T}_0 = US(1)$.

It is easy to verify that all these types appear in the case of prefixes of Sturmian sequences, while the fourth possible case, i.e., the type SS, cannot appear. In fact, these prefix types can be determine from the directive sequence of the standard Sturmian sequence, we explain details in [C, Section 5].

Proposition 3.22 ([C, Theorem 3.10]). *Let* $\mathbf{v}$ *be a CS Rote sequence associated with a standard Sturmian sequence* $\mathbf{u} = \mathcal{S}(\mathbf{v})$. *Let* x *be a non-empty prefix of* $\mathbf{v}$ *and* $w = \mathcal{S}(x)$. *Then the prefix* x *of* $\mathbf{v}$ *has three return words* $A, B, C \in \{0,1\}^*$ *satisfying*

(i) if $\mathcal{T}_w = SU(k)$, *then* $\mathcal{S}(A0) = r$, $\mathcal{S}(B0) = sr^{k+1}s$ *and* $\mathcal{S}(C0) = sr^k s$;

(ii) if $\mathcal{T}_w = US(k)$, *then* $\mathcal{S}(A0) = rr$, $\mathcal{S}(B0) = rsr$ *and* $\mathcal{S}(C0) = s$;

(iii) if $\mathcal{T}_w = UU(k)$, *then* $\mathcal{S}(A0) = rr$, $\mathcal{S}(B0) = rs$ *and* $\mathcal{S}(C0) = sr$.

Among other things, this proposition directly implies that the derived sequences of a CS Rote sequence depend only on the derived sequences of the associated standard Sturmian sequence and on the types of respective prefixes.

Proposition 3.23 ([C, Corollary 4.1]). *Let* $\mathbf{v}$ *be a CS Rote sequence associated with a standard Sturmian sequence* $\mathbf{u} = \mathcal{S}(\mathbf{v})$ *and let* x *be a non-empty prefix of* $\mathbf{v}$. *Then the derived sequence* $\mathbf{d_v}(x)$ *is uniquely determined by the derived sequence* $\mathbf{d_u}(w)$ *of* $\mathbf{u}$ *to the prefix* $w = \mathcal{S}(x)$ *and by the type* $\mathcal{T}_w$ *of the prefix* w.

Example 3.24. The CS Rote sequence $\mathbf{g}$ from Example 2.24 associated with the Fibonacci sequence $\mathbf{f}$ has three derived sequences. Indeed, $\mathbf{f}$ has only one derived sequence which is (up to a permutation of letters) equal to $\mathbf{f}$ and by [C, Example 5.13] the prefixes of $\mathbf{f}$ are of three distinct types $SU(1)$, $US(1)$ and $UU(1)$.

In the article [A] we explain that all derived sequences of a standard Sturmian sequence are standard Sturmian sequences as well. Hence they can be interpreted as 2iet sequences, i.e., sequences coding two interval exchange transformations (see section 2.2.1). Similarly, all derived sequences of the associated CS Rote sequence with respect to its non-empty prefixes are 3iet sequences, i.e., sequences coding three interval exchange transformations. Thus these sequences are dendric (see Section 2.2.3).

Proposition 3.25 ([C, Proposition 4.2]). *Let* $\mathbf{v}$ *be a CS Rote sequence associated with a standard Sturmian sequence* $\mathbf{u} = \mathcal{S}(\mathbf{v})$, *let* x *be a non-empty prefix of* $\mathbf{v}$ *and* $w = \mathcal{S}(x)$. *Let* $(1-\alpha) < \frac{1}{2}$ *be the slope of the Sturmian sequence* $\mathbf{d_u}(w)$. *Then the derived sequence* $\mathbf{d_v}(x)$ *is a 3iet sequence coding the intercept* $\rho = 1-\alpha$ *under the three interval exchange transformation* T *given by the following parameters* β, γ *and permutation* π*:*

(i) if $\mathcal{T}_w = SU(k)$, *then* $\beta = \alpha$, $\gamma = \alpha - k(1-\alpha)$, *and* $\pi = (3,2,1)$;

(ii) if $\mathcal{T}_w = US(k)$, *then* $\beta = 2\alpha - 1$, $\gamma = 1-\alpha$, *and* $\pi = (3,2,1)$;

(iii) if $\mathcal{T}_w = UU(k)$, *then* $\beta = 2\alpha - 1$, $\gamma = 1-\alpha$, *and* $\pi = (2,3,1)$.

Example 3.26 (Example 3.24 continued). The Fibonacci sequence $\mathbf{f}$ has the slope $1-\frac{1}{\phi} = 2-\phi$ and the intercept $2-\phi$, where $\phi = (1+\sqrt{5})/2$ denotes the golden ratio. Thus each of three derived sequences of $\mathbf{g}$ is a 3iet sequence coding the intercept $2-\tau$ under the three interval exchange transformation T with the parameters:

(i) $\beta = \frac{1}{\tau}$, $\gamma = \frac{2}{\tau} - 1$ and $\pi = (3, 2, 1)$,

(ii) $\beta = \frac{2}{\tau} - 1$, $\gamma = 2 - \tau$ and $\pi = (3, 2, 1)$ or

(iii) $\beta = \frac{2}{\tau} - 1$, $\gamma = 2 - \tau$ and $\pi = (2, 3, 1)$.

Finally, we discuss the case of primitive substitutive CS Rote sequences.

Theorem 3.27. *Let* $\mathbf{v}$ *be a CS Rote sequence associated with a standard Sturmian sequence* $\mathbf{u} = \mathcal{S}(\mathbf{v})$. *Then* $\mathbf{v}$ *is primitive substitutive if and only if* $\mathbf{u}$ *is primitive substitutive.*

For CS Rote sequences associated with fixed points of standard Sturmian morphisms we bound the number of their distinct derived sequences. Clearly, similar bound can be constructed also for primitive substitutive CS Rote sequences.

Proposition 3.28 ([C, Corollary 6.4]). *Let* $\mathbf{v}$ *be a CS Rote sequence associated with a standard Sturmian sequence* $\mathbf{u} = \mathcal{S}(\mathbf{v})$ *fixed by a primitive morphism* φ_z, *where* $z \in \mathcal{A}^*$. *Then* $\mathbf{v}$ *has at most* $3|z|$ *derived sequences with respect to its non-empty prefixes and each of them is fixed by a primitive morphism over a ternary alphabet.*

Using Durand's general construction (see [49, Proposition 5.1]) we can also find the primitive morphisms which fix these derived sequences. The process is summarized in [C, Algorithm 6.7].

Example 3.29 (Example 3.26 continued). The three derived sequences of $\mathbf{g}$ are fixed by the following morphisms:

$$\sigma_0 : \begin{cases} A \to AB \\ B \to ABAACAACA \\ C \to ABAACA \end{cases}, \quad \sigma_1 : \begin{cases} A \to BBCAC \\ B \to BBCACAC \\ C \to B \end{cases}, \quad \sigma_2 : \begin{cases} A \to BACCB \\ B \to BACC \\ C \to BACB \end{cases}.$$

On the other hand, every CS Rote sequence associated with a standard Sturmian sequence has at least two derived sequences. Moreover, this bound is attained by only one CS Rote sequence which is associated with the Sturmian sequence directed by $(1001)^\omega$. See [A, Remark 7.1].

If φ_z is not a power of any other morphism, then the Sturmian sequence $\mathbf{u}$ has exactly $|z|$ distinct derived sequences (see Proposition 3.10) and thus by Proposition 3.23 the CS Rote sequence $\mathbf{v}$ has at least $|z|$ derived sequences. In each of our examples the actual number of derived sequences was $|z|$, $2|z|$ or $3|z|$, but we do not know whether some other values can also appear.

3.2 Non-repetitive complexity of Arnoux–Rauzy sequences

The article [D] is dedicated to the non-repetitive complexity and the initial non-repetitive complexity of Arnoux–Rauzy sequences. These functions were discussed in Section 2.3.4. The study is motivated by recent results of Nicholson and Rampersad [84] on the initial non-repetitive complexity of the Fibonacci and Tribonacci sequences stated in Proposition 2.46.

Our aim is to express the values of these functions for an arbitrary Arnoux–Rauzy sequence $\mathbf{u}$. For this purpose, we first use the Rauzy graphs of $\mathbf{u}$ (see Section 2.4.1) to transform our task into the evaluation of the lengths of return words to the bispecial factors of $\mathbf{u}$ (see [D, Proposition 7]). Then we utilize the knowledge of the form of these return words and respective derived sequences obtained in [B]. It leads to the desired formulas in terms of the directive sequence of $\mathbf{u}$.

To state the relevant theorems we have to recall that $B_{\mathbf{u}}(k)$ denotes the k-th bispecial factor of $\mathbf{u}$ (see Section 2.2.2) and φ_a for each $a \in \mathcal{A}$ is the elementary episturmian morphism (see Section 2.4.2).

Theorem 3.30 ([D, Theorem 13]). *Let* $\mathbf{u}$ *be an Arnoux–Rauzy sequence over* $\mathcal{A}$ *and let* $\mathbf{z} = z_0 z_1 z_2 \cdots$ *be the directive sequence of the standard Arnoux–Rauzy sequence with the language* $\mathcal{L}_{\mathbf{u}}$. *Let* $n \in \mathbb{N}, n \geq 1$. *Find the unique* k *such that* $|B_{\mathbf{u}}(k-1)| < n \leq |B_{\mathbf{u}}(k)|$ *and for every* $b \in \mathcal{A}$ *define* $S_b(k) = \sup\{\ell : 0 \leq \ell < k, z_\ell = b\}$. *Then we have*

$$\mathrm{nr}\mathcal{C}_{\mathbf{u}}(n) = |\varphi_{z_0}\varphi_{z_1}\cdots\varphi_{z_{k-1}}\varphi_{z_k}(a)| - 1 - |B_{\mathbf{u}}(k)| + n\,,$$

where $a \in \mathcal{A} \setminus \{z_k\}$ *is the letter such that* $S_a(k) = \inf\{S_b(k) : b \in \mathcal{A} \setminus \{z_k\}\}$.

Let us notice that the case of Sturmian sequences is much easier. In fact, the simple form of their Rauzy graphs enables us to state immediately an explicit formula.

Theorem 3.31 ([D, Theorem 5]). *Let* $\mathbf{u}$ *be a Sturmian sequence. Then* $\mathrm{nr}\mathcal{C}_{\mathbf{u}}(n) = n+1$ *for every* $n \in \mathbb{N}$.

Nevertheless, Sturmian sequences are not the only sequences with the equality $\mathrm{nr}\mathcal{C}_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n)$ for every $n \in \mathbb{N}$, one can reads more in Section 2.3.4 or in the paper [84].

We determine the initial non-repetitive complexity only for standard Arnoux–Rauzy sequence. It seems that for non-standard Arnoux–Rauzy sequences the evaluating of the initial non-repetitive complexity is much more complicated, as, unlike the standard case, we do not have the control over the positions of the vertices corresponding to prefixes in the respective Rauzy graphs.

Theorem 3.32 ([D, Theorem 15]). *Let* $\mathbf{u}$ *be a standard Arnoux–Rauzy sequence with the directive sequence* $\mathbf{z} = z_0 z_1 z_2 \cdots$. *For every integer* $n \geq 1$ *we take the unique* k *such that* $|B_{\mathbf{u}}(k-1)| < n \leq |B_{\mathbf{u}}(k)|$. *Then we have*

$$\mathrm{inr}\mathcal{C}_{\mathbf{u}}(n) = |\varphi_{z_0}\varphi_{z_1}\cdots\varphi_{z_{k-1}}(z_k)|\,.$$

Moreover, for standard Sturmian sequence we deduce the following corollary. However, Bugeaud and Kim [29] showed the more general result, see Theorem 2.45.

Corollary 3.33 ([D, Corollary 16]). *Let* $\mathbf{u}$ *be a standard Sturmian sequence. Then* $\mathrm{inr}\mathcal{C}_{\mathbf{u}}(n) = n+1$ *for infinitely many* $n \in \mathbb{N}$.

Finally, we apply Theorems 3.30 and 3.32 to the d-bonacci sequence $\mathbf{t}$ (see Example 2.15). We get the formulas for $\mathrm{nr}\mathcal{C}_{\mathbf{t}}(n)$ and $\mathrm{inr}\mathcal{C}_{\mathbf{t}}(n)$ in terms of the so-called d-bonacci numbers which naturally generalized the famous Fibonacci numbers. The sequence of *d-bonacci numbers* $(D_k)_{k\geq 0}$ is defined by the linear recurrence:

$$D_k = \sum_{j=1}^{d} D_{k-j} \quad \text{for } k \geq d \qquad \text{and} \qquad D_k = 2^k \quad \text{for all } k = 0, 1, \ldots, d-1\,.$$

Proposition 3.34 ([D, Theorems 20 and 21]). *Let* $\mathbf{t}$ *be the* d*-bonacci sequence and let* n, k *be positive integers such that*

$$\frac{1}{d-1}\sum_{i=0}^{d-1}(d-i)D_{k-i-2} - \frac{d}{d-1} < n \leq \frac{1}{d-1}\sum_{i=0}^{d-1}(d-i)D_{k-i-1} - \frac{d}{d-1}\,.$$

Then

$$\begin{aligned}\mathrm{nr}\mathcal{C}_{\mathbf{t}}(n) &= D_{k+1} - 1 - \frac{1}{d-1}\sum_{i=0}^{d-1}(d-i)D_{k-i-1} + \frac{d}{d-1} + n \qquad \textit{and}\\ \mathrm{inr}\mathcal{C}_{\mathbf{t}}(n) &= D_k\,.\end{aligned}$$

This generalizes the results of Nicholson and Rampersad [84] stated in Item (ii) and (iii) of Proposition 2.46.

3.3 Critical exponent of complementary symmetric Rote sequences

In the article [E] we determine the critical exponent of complementary symmetric (CS) Rote sequences. The critical exponent was explained in Section 2.3.2. The motivation for this work comes from the result on the repetition threshold of binary rich sequences stated in Theorem 2.40. This statement was formulated as a conjecture by Baranwal and Shallit [13] and proved by Curie, Mol, and Rampersad [40]. For us it is especially important that the two sequences with the minimal critical exponent among all binary rich sequences are CS Rote sequences.

First of all, it is worth to realize that for finding the critical exponent of a sequence only some of its factors have to be considered.

Lemma 3.35 ([E, Lemma 3]). *Let* $\mathbf{u}$ *be a uniformly recurrent aperiodic sequence. Then* $\mathrm{cr}(\mathbf{u}) = \sup\{\mathrm{ind}_{\mathbf{u}}(u) : u \in \mathcal{M}\}$*, where*

$$\mathcal{M} = \{u : u \textit{ is a return word to a bispecial factor of } \mathbf{u}\}\,.$$

The relation between a CS Rote sequence $\mathbf{v}$ and the associated Sturmian sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$ stated in Proposition 2.23 directly implies the following method of computation of the critical exponent of $\mathbf{v}$. In fact, it holds for more general pairs of sequences satisfying $\mathbf{u} = \mathcal{S}(\mathbf{v})$, too.

Proposition 3.36 ([E, Theorem 14]). *Let* $\mathbf{v}$ *be a binary aperiodic uniformly recurrent sequence whose language is closed under* E*. Denote* $\mathbf{u} = \mathcal{S}(\mathbf{v})$*,*

$$\begin{aligned}A_1 &= \left\{\mathrm{ind}_{\mathbf{u}}(u) + \tfrac{1}{|u|} : u \textit{ is a stable return word to a bispecial factor of } \mathbf{u}\right\} \quad \textit{and}\\ A_2 &= \left\{\tfrac{1}{2}\big(\mathrm{ind}_{\mathbf{u}}(u) + \tfrac{1}{|u|}\big) : u \textit{ is an unstable return word to a bispecial factor of } \mathbf{u}\right\}.\end{aligned}$$

Then

$$\mathrm{cr}(\mathbf{v}) = \sup\big(A_1 \cup A_2\big)\,.$$

Hence it suffices to study the return words to bispecial factors of the Sturmian sequence $\mathbf{u}$. We continue in the results of [A] and using the directive sequences of

Sturmian sequences (described in Section 2.4.2) we determine both the stability ([E, Proposition 30]) and the indices ([E, Proposition 32]) of these return words.

Thus we get the formula for the critical exponent of $\mathbf{v}$ in terms of continued fraction expansions related to the directive sequence of $\mathbf{u}$. Let us emphasize that in this section the directive sequence is a sequence of morphisms $D = \varphi_0$ and $G = \varphi_1$ (instead of a sequence of letters 0 and 1). To a sequence $\mathbf{u}$ with the directive sequence $G^{a_1}D^{a_2}G^{a_3}\cdots$ or $D^{a_1}G^{a_2}D^{a_3}\cdots$ we assign an irrational number $\theta \in (0,1)$ with the continued fraction expansion $\theta = [0; a_1, a_2, a_3, \ldots]$. We denote $\frac{p_N}{q_N}$ the N^{th} convergent to the number θ and $\frac{p'_N}{q'_N}$ the N^{th} convergent to the number $\frac{\theta}{1+\theta}$. Other details about these continued fraction expansions are summarized in [E, Section 5].

Theorem 3.37 ([E, Theorem 33]). *Let $\mathbf{v}$ be a CS Rote sequence and let $\mathbf{u}$ be the standard Sturmian sequence such that $\mathcal{L}_{\mathcal{S}(\mathbf{v})} = \mathcal{L}_{\mathbf{u}}$.*

If $\mathbf{u}$ has the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$, then we have $\operatorname{cr}(\mathbf{v}) = \sup(M_1 \cup M_2 \cup M_3)$, where

$$M_1 = \left\{ a_{N+1} + 2 + \frac{q'_{N-1} - 1}{q'_N} \; : \; q_N \textit{ is even, } N \in \mathbb{N} \right\} ;$$
$$M_2 = \left\{ \frac{a_{N+1} + 2}{2} + \frac{q'_{N-1} - 1}{2q'_N} \; : \; q_N \textit{ is odd, } N \in \mathbb{N} \right\} ;$$
$$M_3 = \left\{ 2 + \frac{q'_N - 1}{q'_{N-1} + q'_N} \; : \; q_{N-1}, q_N \textit{ are odd and } a_{N+1} > 1, N \geq 1 \right\}$$

If $\mathbf{u}$ has the directive sequence $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$, then the formula is the same except for the replacement of q_N and q_{N-1} by p_N and p_{N-1}.

Using this formula we describe all CS Rote sequences with the critical exponent less than or equal to 3.

Proposition 3.38 ([E, Proposition 34]). *Let $\mathbf{v}$ be a CS Rote sequence associated with the standard Sturmian sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$. If $\operatorname{cr}(\mathbf{v}) \leq 3$, then the directive sequence of $\mathbf{u}$ is of one of the following forms:*

(i) $G^{a_1}(D^2G^2)^\omega$, *where $a_1 = 1$ or $a_1 = 3$; in this case* $\operatorname{cr}(\mathbf{v}) = 2 + \frac{1}{\sqrt{2}}$*;*

(ii) $G^{a_1}D^4(G^2D^2)^\omega$, *where $a_1 = 1$ or $a_1 = 3$; in this case* $\operatorname{cr}(\mathbf{v}) = 3$*;*

(iii) $G^{a_1}D^1G^{a_3}(D^2G^2)^\omega$, *where $a_1 = 2$ or $a_1 = 4$ and $a_3 = 1$ or $a_3 = 3$; in this case* $\operatorname{cr}(\mathbf{v}) = 3$*;*

(iv) $D^1G^{a_2}(D^2G^2)^\omega$, *where $a_2 = 1$ or $a_2 = 3$; in this case* $\operatorname{cr}(\mathbf{v}) = 3$.

Let us mention that the sequences from Item (i) are the sequences $\mathbf{v}$ and $\mathbf{v}'$ from Theorem 2.39.

In addition, we show that there are uncountably many CS Rote sequences with the critical exponent less than $\frac{7}{2}$ (see [E, Theorem 37]). By Proposition 2.39 all these sequences has smaller critical exponent that any Sturmian sequence. Nevertheless, the detailed structure of the set $\{\operatorname{cr}(\mathbf{v}) : \mathbf{v} \text{ is CS Rote sequence}\}$ remains unclear.

3.4 Recurrence function of complementary symmetric Rote sequences

The article [E] contains the formula for the recurrence function of CS Rote sequences, too. This function is described in Section 2.3.3. As in the case of the critical exponent, the values of the recurrence function are expressed by means of continued fraction expansions related to the directive sequences of associated Sturmian sequences (see Section 3.3). Also in this section the directive sequence is a sequence of morphisms $D = \varphi_0$ and $G = \varphi_1$.

Theorem 3.39 ([E, Theorem 54]). *Let* $\mathbf{v}$ *be a CS Rote sequence and let* $\mathbf{u}$ *be the standard Sturmian sequence such that* $\mathcal{L}_{S(\mathbf{v})} = \mathcal{L}_{\mathbf{u}}$.

If $\mathbf{u}$ *has the directive sequence* $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$, *then the recurrence function* $R_\mathbf{v}$ *for* $n \in [q'_N, q'_{N+1})$, $N \in \mathbb{N}$, *is given by*

Case q_N **even**
$$R_\mathbf{v}(n+1) = \begin{cases} 2q'_{N+1} + q'_N + n & \text{if } a_{N+2} > 1, \\ 2q'_{N+2} + n & \text{if } a_{N+2} = 1. \end{cases}$$

Case q_{N+1} **even**
$$R_\mathbf{v}(n+1) = \begin{cases} q'_{N+2} + 2q'_{N+1} + q'_N + n & \text{if } a_{N+2} > 1 \\ 2q'_{N+2} + q'_{N+1} + n & \text{if } a_{N+2} = 1. \end{cases}$$

Case q_N, q_{N+1} **odd**
$$R_\mathbf{v}(n+1) = \begin{cases} 3q'_{N+1} + q'_N + n & \text{if } a_{N+2} > 1 \\ q'_{N+3} + q'_{N+2} + q'_{N+1} + n & \text{if } a_{N+2} = 1. \end{cases}$$

If $\mathbf{u}$ *has the directive sequence* $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$, *then the admissible values of* $R_\mathbf{v}$ *for* $n \in [q'_N, q'_{N+1})$, $N \in \mathbb{N}$, *are the same, but the cases above have to be distinguished according to the parity of* p_N *and* p_{N+1} *instead of* q_N *and* q_{N+1}.

Example 3.40. We consider the CS Rote sequence $\mathbf{v}$ such that $S(\mathbf{v})$ has the directive sequence $G(D^2G^2)^\omega$. Since in this case all q_N are odd, we obtain for every $n \in [q'_N, q'_{N+1})$ the formula

$$R_\mathbf{v}(n+1) = 3q'_{N+1} + q'_N + n = n + \frac{1}{2\sqrt{2}}\Big((4+3\sqrt{2})(1+\sqrt{2})^{N+1} - (4-3\sqrt{2})(1-\sqrt{2})^{N+1}\Big).$$

We briefly indicate how we derive the statement of Theorem 3.39. First, we use the following reformulation of Proposition 2.43.

Lemma 3.41 ([E, Lemma 41]). *Let* $\mathbf{u}$ *be a uniformly recurrent aperiodic sequence. For* $n \in \mathbb{N}$, *we denote*

$$\mathcal{B}_\mathbf{u}(n) = \{b \in \mathcal{L}_\mathbf{u} : \text{there is a factor } w \in \mathcal{L}_\mathbf{u}(n) \text{ such that } b \text{ is the shortest bispecial factor containing } w\}.$$

Then

$$R_\mathbf{u}(n) = \max\{|r| : r \text{ is a return word to } b \in \mathcal{B}_\mathbf{u}(n)\} + n - 1.$$

From the close relation between a CS Rote sequence $\mathbf{v}$ and the Sturmian sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$ it is easy to realize that the sets $\mathcal{B}_{\mathbf{v}}(n+1)$ and $\mathcal{B}_{\mathbf{u}}(n)$ correspond naturally for every $n \in \mathbb{N}$. Thus it suffices to describe the bispecial factors from the set $\mathcal{B}_{\mathbf{u}}(n)$ (see [E, Theorem 48]). This can be done through the investigation of the Rauzy graphs and palindromic properties of Sturmian sequences.

Finally, we use the comfortable description of the lengths of return words to bispecial factors of $\mathbf{u}$ ([E, Proposition 32]) as well as the relation between the return words in the Sturmian sequence $\mathbf{u}$ and the CS Rote sequence $\mathbf{v}$ explained in [C] (see Proposition 3.22).

3.5 Future directions

We hope that some of our tools and ideas can be utilized also for other tasks. We can see at least four possible future directions:

- The crucial notion for our study of derived sequences of Sturmian and Arnoux–Rauzy sequences is their handy S-adic representation. S-adic representations are (partially) known also for other classes of sequences. It would be nice to utilize them for the description of derived sequences of sequences coding interval exchange transformation or even dendric sequences.
- To study properties of Rote sequences, we especially use their $\mathcal{S}$-relation to Sturmian sequences (see Section 2.2.4). This S-relation (or its generalizations) can be considered also for other classes of sequences. It seems that similar methods can lead to some interesting results also in this cases.
- Following Durand's example we study only derived sequences with respect to prefixes of sequences. However, it could be also interesting to understand the (more complicated) structure of derived sequences to non-prefixes. Some results in this direction can be found in [64, 69, 87].
- Like many other authors we use our results on return words and derived sequences to study other properties of sequences such as critical exponent, recurrence function or non-repetitive complexity. It seems that there are other similar possibilities how utilize the results.

3.6 Note on authorship

Besides this extensive Introduction, the thesis is a collection of articles, most of which are co-authored. My contribution to each of these articles corresponds to the number of authors: all of them contributed equally.

References

[A] K. Klouda, K. Medková, E. Pelantová, and Š. Starosta, Fixed points of Sturmian morphisms and their derivated words. *Theoretical Computer Science* 743 (2018), 23–37.

[B] K. Medková, Derived sequences of Arnoux–Rauzy sequences. In: R. Mercas, D. Reidenbach (eds.), *WORDS 2019, Lecture notes in Computer Science*, vol. 11682, Springer (2019), 251–263.

[C] K. Medková, E. Pelantová, and L. Vuillon, Derived sequences of complementary symmetric Rote sequences. *RAIRO - Theoretical Informatics and Applications* 53 (2019), 125–151.

[D] K. Medková, E. Pelantová, and É. Vandomme, On non-repetitive complexity of Arnoux–Rauzy words. Submitted to *Discrete Applied Mathematics*.

[E] L. Dvořáková, K. Medková, and E. Pelantová, Complementary symmetric Rote sequences: the critical exponent and the recurrence function. Submitted to *Discrete Mathematics & Theoretical Computer Science.*

[1] J.-P. Allouche, M. Baake, J. Cassaigne, and D. Damanik, Palindrome complexity. *Theoret. Comput. Sci.* 292 (2003), 9–31.

[2] J.-P. Allouche and M. Bousquet-Mélou, On the conjectures of Rauzy and Shallit for infinite words. *Comment. Math. Univ. Carolinae* 36 (1995), 705–711.

[3] J.-P. Allouche, J. Cassaigne, J. Shallit, and L. Q. Zamboni, A Taxonomy of Morphic Sequences. Preprint: `https://arxiv.org/abs/1711.10807` (2017).

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

[5] J.-P. Allouche and J. Shallit, *Automatic Sequences, Theory, Applications, Generalizations.* Cambridge University Press, Cambridge (2003).

[6] I. M. Araújo and V. Bruyère, Words derivated from Sturmian words. *Theoret. Comput. Sci.* 340 (2005), 204–219.

[7] P. Arnoux and G. Rauzy, Représentation géométrique de suites de complexité $2n+1$. *Bulletin de la Société Mathématique de France* 119 (1991), 199–215.

[8] P. Baláži, Z. Masáková, and E. Pelantová, Factor versus palindromic complexity of uniformly recurrent infinite words. *Theoret. Comput. Sci.* 380 (2007), 266–275.

[9] L. Balková, Return words and recurrence function of a class of infinite words. *Acta Polytechnica* 47 (2007), 15–19.

[10] L. Balková, E. Pelantová, and Š. Starosta, Sturmian jungle (or garden?) on multiliteral alphabets. *RAIRO Theor. Inf. Appl.* 44 (2010), 443–470.

[11] L. Balková, E. Pelantová, and W. Steiner, Sequences with constant number of return words. *Monatsh. Math.* 155 (2008), 251–263.

[12] A. R. Baranwal and J. Shallit, Critical expoment of infinite balanced words via the Pell number system. In: R. Mercas and D. Reidenbach (eds.), *Proceedings WORDS 2019, Lecture Notes in Computer Science*, vol. 11682, Springer (2019), 80–92.

[13] A. R. Baranwal and J. Shallit, Repetitions in infinite palindrome-rich words. In: R. Mercas and D. Reidenbach (eds.), *Proceedings WORDS 2019, Lecture Notes in Computer Science*, vol. 11682, Springer (2019), 93–105.

[14] J. Berstel, Sturmian and Episturmian words. In: S. Bozapalidis and G. Rahonis (eds.), *Algebraic Informatics*, Springer Berlin Heidelberg (2007), 23–47.

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

[16] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone, Bifix codes and interval exchanges. *J. Pure Appl. Algebra* 219 (2015), 2781–2798.

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

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

[19] V. Berthé and V. Delecroix, Beyond substitutive dynamical systems: S-adic expansions. In: S. Akiyama (ed.), *Numeration and Substitution 2012, RIMS Kôkyûroku Bessatsu*, vol. B46 (2014), 81–123.

[20] V. Berthé, V. Delecroix, F. Dolce, D. Perrin, Ch. Reutenauer, and G. Rindone, Return words of linear involutions and fundamental groups. *Ergodic Theory and Dynamical Systems* 37 (2017), 693–715.

[21] V. Berthé, F. Dolce, F. Durand, J. Leroy, and D. Perrin, Rigidity and substitutive dendric words. *IJFCS International Journal of Foundations of Computer Science* 29 (2018), 705–720.

[22] V. Berthé, C. Holton, and L. Q. Zamboni, Initial powers of Sturmian sequences. *Acta Arith.* 122 (2006), 315–347.

[23] V. Berthé and M. Rigo (eds.), *Combinatorics, Automata and Number Theory*, Encyclopedia of Mathematics and Its Applications, vol. 135. Cambridge University Press (2010).

[24] A. Blondin-Massé, S. Brlek, S. Labbé, and L. Vuillon, Codings of rotations on two intervals are full. *Electronic Notes in Discrete Mathematics* 34 (2009), 289–393.

[25] A. Blondin-Massé, S. Brlek, S. Labbé, and L. Vuillon, Palindromic complexity of codings of rotations. *Theoret. Comput. Sci.* 412 (2011), 6455–6463.

[26] A. Blondin-Massé, G. Paquin, H. Tremblay, and L. Vuillon, On Generalized Pseudostandard Words over Binary Alphabet. *Journal of Int. Sequences* 16 (2013), article 13.2.11.

[27] M. Bucci and A. De Luca, On a family of morphic images of Arnoux–Rauzy words. In: *LATA '09: Proceedings of the 3rd International Conference on Language and Automata Theory and Applications*, Springer-Verlag, (2009), 259–266.

[28] M. Bucci, A. De Luca, A. Glen, and L. Q. Zamboni, A connection between palindromic and factor complexity using return words. *Adv. in Appl. Math* 42 (2009), 60–74.

[29] Y. Bugeaud and D. H. Kim, A new complexity function, repetitions in Sturmian words, and irrationality exponents of Sturmian numbers. *Trans. Amer. Math. Soc.* 371 (2019), 3281–3308.

[30] J. Cassaigne, Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin* 4(1) (1997), 67–88.

[31] J. Cassaigne, On a conjecture of J. Shallit. In: *ICALP'97, Lect. Notes Comput. Sci.*, vol. 1256, Springer-Verlag (1997), 693–704.

[32] J. Cassaigne, Sequences with grouped factors. In: *Developments in Language Theory III*, Aristotle University of Thessaloniki (1998), 211–222.

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

[34] J. Cassaigne, Recurrence in Infinite Words. In: A. Ferreira and H. Reichel (eds.), *STACS 2001, Lecture Notes in Computer Science*, vol. 2010, Springer, Berlin, Heidelberg (2001), 1–11.

[35] J. Cassaigne, On extremal properties of the Fibonacci word. *RAIRO-Theor. Inf. Appl.* 42(4) (2008), 701–715.

[36] J. Cassaigne and N. Chekhova, Fonctions de récurrence des suites d'Arnoux–Rauzy et réponse à une question de Morse et Hedlund. *Ann. Inst. Fourier (Grenoble)* 56(7) (2006), 2249–2270.

[37] J. Cassaigne, S. Ferenczi, and L. Q. Zamboni, Imbalances in Arnoux–Rauzy sequences. *Ann. Inst. Fourier (Grenoble)* 50(4) (2000), 1265–1276.

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

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

[40] J. D. Currie, L. Mol, and N. Rampersad, The repetition threshold for binary rich words. *Discrete Math. Theoret. Comput. Sci.* 22(1) (2020), no. 6.

[41] J. D. Currie and N. Rampersad, A proof of Dejean's conjecture. *Mathematics of Computation* 80 (2011), 1063–1070.

[42] D. Damanik and D. Lenz, The index of Sturmian sequences. *European J. Combin.* 23 (2002), 23–29.

[43] F. Dejean, Sur un théorème de Thue. *J. Combin. Theory. Ser. A* 13 (1972), 90–99.

[44] G. Didier, Codages de rotations et fractions continues. *J. Num-ber Theory*, 71(2) (1998), 275–306.

[45] F. Dolce and D. Perrin, Neutral and tree sets of arbitrary characteristic. *Theoret. Comput. Sci.* 658 (2017), 159–174.

[46] F. Dolce and D. Perrin, Eventually dendric subshifts. Preprint: `https://arxiv.org/abs/1807.05124` (2018).

[47] X. Droubay, J. Justin, and G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.* 225(1–2) (2001), 539–553.

[48] X. Droubay and G. Pirillo, Palindromes and Sturmian words. *Theoret. Comput. Sci.* 223 (1999), 73–85.

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

[50] F. Durand, Linearly recurrent subshifts have a finite number of non-periodic subshift factors. *Ergodic Theory and Dynamical Systems* 20 (2000), 1061–1078.

[51] F. Durand, Erratum: Linearly recurrent subshifts have a finite number of non-periodic factors. *Ergodic Theory and Dynamical Systems* 23(2) (2003), 663–669.

[52] F. Durand, Decidability of uniform recurrence of morphic sequence. *International Journal of Foundations of Computer Science* 24(1) (2013), 123–146.

[53] F. Durand, B. Host, and C. Skau, Substitutions, Bratteli diagrams and dimension groups. *Ergod. Th. Dyn. Sys.* 19 (1999), 952—993.

[54] S. Ferenczi, Rank and symbolic complexity. *Ergodic Theory Dynam. Syst.* 16 (1996), 663–682.

[55] S. Ferenczi, C. Holton, and L. Q. Zamboni, Structure of three-interval exchange transformations: II. A combinatorial description of the trajectories. *J. Anal. Math.* 89 (2003), 239–276.

[56] G. Fici, A. Langiu, T. Lecroq, A. Lefebvre, F. Mignosi, J. Peltomäki, and É. Prieur-Gaston, Abelian powers and repetitions in Sturmian words. *Theoret. Comput. Sci.* 635 (2016), 16–34.

[57] A. Frid, Applying a uniform marked morphism to a word. *Discrete Math. and Theoret. Comput. Sci.* 3 (1999), 125–140.

[58] A. Glen and J. Justin, Episturmian words: a survey. *RAIRO-Theoret. Inf. Appl.* 43 (2009), 403–442.

[59] A. Glen, F. Levé, and G. Richomme, Directive words of episturmian words: equivalences and normalization. *RAIRO-Theoret. Inf. Appl.* 43 (2009), 299–319.

[60] D. Hensley, *Continued Fractions*, World Scientific Publishing (2006).

[61] C. Holton and L. Q. Zamboni, Descendants of primitive substitutions. *Theoret. Comput. Syst.* 32 (1999), 133–157.

[62] Y. Huang and Z. Wen, The sequence of return words of the Fibonacci sequence. *Theor. Comput. Sci.* 593 (2015), 106–116.

[63] Y. Huang and Z. Wen, Kernel words and gap sequence of the Tribonacci sequence. *Acta Mathematica Scientia* 36B(1) (2016), 173–194.

[64] Y. Huang and Z. Wen, Envelope words and the reflexivity of the return word sequences in the period-doubling sequence. Preprint: `https://arxiv.org/abs/1703.07157` (2017).

[65] J. Justin and L. Vuillon, Return words in Sturmian and episturmian words. *Theor. Inform. Appl.* 34 (2000), 343–356.

[66] J. Justin and G. Pirillo, Episturmian words and episturmian morphisms. *Theoret. Comput. Sci.* 276(1–2) (2002), 281–313.

[67] J. Justin and G. Pirillo, Episturmian words: shifts, morphisms and numeration systems. *Internat. J. Found. Comput. Sci.* 15(2) (2004), 329–348.

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

[69] V. Košík and Š. Starosta, On Substitutions Closed Under Derivation: Examples. In: R. Mercas and D. Reidenbach (eds.), *Proceedings WORDS 2019, Lecture Notes in Computer Science*, vol. 11682, Springer (2019), 226–237.

[70] D. Krieger, *Critical exponents and stabilizers of infinite words.* Ph.D. thesis, `http://hdl.handle.net/10012/3599` (2008).

[71] D. Krieger and J. Shallit, Every real number greater than 1 is a critical exponent. *Theoret. Comput. Sci.* 381 (2007), 177–182.

[72] M. Lothaire, *Combinatorics on Words*, Encyclopaedia of Mathematics and its Applications, vol. 17, Addison-Wesley (1983). Reprinted in the Cambridge Mathematical Library, Cambridge University Press (1997).

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

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

[75] Z. Masáková and E. Pelantová, Relation between powers of factors and the recurrence function characterizing Sturmian words. *Theoret. Comput. Sci.* 410 (2009), 589–3596.

[76] Z. Masáková and E. Pelantová, Enumerating Abelian Returns to Prefixes of Sturmian Words. In: J. Karhumäki, A. Lepistö, L. Zamboni (eds.), *Combinatorics on Words, Lecture Notes in Computer Science*, vol. 8079, Springer (2013), 193–204.

[77] F. Mignosi, Infinite words with linear subword complexity. *Theoret. Comput. Sci.* 65 (1989), 221–242.

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

[79] F. Mignosi, A. Restivo, and S. Salemi, Periodicity and the golden ratio. *Theoret. Comput. Sci.* 204 (1998), 153–167.

[80] T. K. S. Moothathu, Eulerian entropy and non-repetitive subword complexity. *Theoret. Comput. Sci.* 420 (2012), 80–88.

[81] M. Morse and G. A. Hedlund, Symbolic dynamics. *Amer. J. Math.* 60 (1938), 815–866.

[82] M. Morse and G. A. Hedlund, Symbolic dynamics II - Sturmian trajectories. *Amer. J. Math.* 62 (1940), 1–42.

[83] H. Mousavi, Automatic theorem proving in Walnut. Preprint: `https://arxiv.org/abs/1603.06017` (2016).

[84] J. Nicholson and N. Rampersad, Initial non-repetitive complexity of infinite words. *Discrete Appl. Math.* 208 (C) (2016), 114–122.

[85] V. I. Oseledec, The spectrum of ergodic automorphisms. *Doklady Akademii Nauk SSSR* 168 (1966), 1009–1011 (in Russian).

[86] J. J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés. In: J. Paredaens (ed.), *Proc. 11th Int. Conf. on Automata, Languages, and Programming (ICALP), Lecture Notes in Computer Science*, vol. 172, Springer-Verlag (1984), 380–389.

[87] E. Pelantová and Š. Starosta, On Sturmian substitutions closed under derivation. Preprint: `https://arxiv.org/abs/1908.11095` (2019).

[88] J. Peltomäki and M. A. Whiteland, Every Nonnegative Real Number Is an Abelian Critical Exponent. In: R. Mercas and D. Reidenbach (eds.), *Proceedings WORDS 2019, Lecture Notes in Computer Science*, vol. 11682, Springer (2019), 275–285.

[89] S. Puzynina and L. Zamboni, Abelian returns in Sturmian words. *J. Combin. Theory Ser. A* 120(2) (2013), 390–408.

[90] N. Rampersad, M. Rigo, and P. Salimov, A note on abelian returns in rotation words. *Theoret. Comput. Sci.* 528 (2014), 101–107.

[91] N. Rampersad, J. Shallit, and É. Vandomme, Critical exponents of infinite balanced words. *Theoret. Comput. Sci.* 777 (2019), 454–463.

[92] M. Rao, Last cases of Dejean's conjecture. *Theoret. Comput. Sci.* 412 (2011), 3010–3018.

[93] G. Rauzy, Suites à termes dans un alphabet fini. In: *Séminaire de Théorie des Nombres de Bordeaux (1982–1983)* 12 (1983), 1–16.

[94] R. N. Risley and L. Q. Zamboni, A generalization of Sturmian sequences: Combinatorial structure and transcendence. *Acta Arith.* 95 (2000), 167–184.

[95] G. Richomme, Conjugacy and episturmian morphisms, *Theoret. Comput. Sci.* 302 (2003), 1–34.

[96] G. Richomme, Lyndon morphisms. *Bull. Belg. Math. Soc. Simon Stevin* 10 (2003), 761–785.

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

[98] M. Rigo and P. Salimov: Another generalization of abelian equivalence: Binomial complexity of infinite words. *Theoret. Comput. Sci.* 601 (2015), 47–57.

[99] M. Rigo, P. Salimov and É. Vandomme, Some properties of abelian return words. *Journal of Int. Sequences* 16 (2013), article 13.2.5.

[100] G. Rote, Sequences with subword complexity $2n$. *J. Number Theory* 46 (1994), 196–213.

[101] P. Rotondo and B. Vallée, The recurrence function of a random Sturmian word. In: C. Martinez et al. (eds.), *Proceedings of 14th ANALCO17*, SIAM (2017), 100–114.

[102] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl.* 1 (1912), 1–67. Reprinted in T. Nagell et al. (eds.), *Selected Mathematical Papers of Axel Thue.* Universitetsforlaget, Oslo (1977), 413–477.

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

[104] L. Vuillon, A characterization of Sturmian words by return words. *Eur. J. Combin.* 22 (2001), 263–275.

[105] L. Vuillon, Balanced words. *Bull. Belg. Math. Soc. Simon Stevin* 10 (2003), 787–805.

[106] L. Vuillon, On the number of return words in infinite words constructed by interval exchange transformations. *Pure Mathematics and Applications* 18 (2007), issue no. 3–4.

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

Article A

Fixed points of Sturmian morphisms and their derivated words

Theoretical Computer Science 743 (2018) 23–37

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Fixed points of Sturmian morphisms and their derivated words

Karel Klouda [b], Kateřina Medková [a,*], Edita Pelantová [a], Štěpán Starosta [b]

[a] Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 7, 115 19, Prague 1, Czech Republic

[b] Department of Applied Mathematics, Faculty of Information Technology, Czech Technical University in Prague, Thákurova 9, 160 00, Prague 6, Czech Republic

ARTICLE INFO

Article history:
Received 28 January 2018
Received in revised form 8 May 2018
Accepted 21 June 2018
Available online 2 July 2018
Communicated by P. Spirakis

Keywords:
Derivated word
Return word
Sturmian morphism
Sturmian word

ABSTRACT

Any infinite uniformly recurrent word $\mathbf{u}$ can be written as concatenation of a finite number of return words to a chosen prefix w of $\mathbf{u}$. Ordering of the return words to w in this concatenation is coded by derivated word $d_{\mathbf{u}}(w)$. In 1998, Durand proved that a fixed point $\mathbf{u}$ of a primitive morphism has only finitely many derivated words $d_{\mathbf{u}}(w)$ and each derivated word $d_{\mathbf{u}}(w)$ is fixed by a primitive morphism as well. In our article we focus on Sturmian words fixed by a primitive morphism. We provide an algorithm which to a given Sturmian morphism ψ lists the morphisms fixing the derivated words of the Sturmian word $\mathbf{u} = \psi(\mathbf{u})$. We provide a sharp upper bound on length of the list.

1. Introduction

Sturmian words are probably the most studied object in combinatorics on words. They are aperiodic words over a binary alphabet having the least factor complexity possible. Many properties, characterizations and generalizations are known, see for instance [5,4,2].

One of their characterizations is in terms of return words to their factors. Let $\mathbf{u} = u_0u_1u_2\cdots$ be a binary infinite word with $u_i \in \{0,1\}$. Let $w = u_iu_{i+1}\cdots u_{i+n-1}$ be its factor. The integer i is called an *occurrence* of the factor w. A return word to a factor w is a word $u_iu_{i+1}\cdots u_{j-1}$ with i and j being two consecutive occurrences of w such that $i < j$. In [22], Vuillon showed that an infinite word $\mathbf{u}$ is Sturmian if and only if each nonempty factor w has exactly two distinct return words. A straightforward consequence of this characterization is that if w is a prefix of $\mathbf{u}$, we may write

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

with $s_i \in \{0,1\}$ and r_0 and r_1 being the two return words to w. The coding of these return words, the word $d_{\mathbf{u}}(w) = s_0s_1s_2\cdots$ is called the *derivated word of* $\mathbf{u}$ *with respect to* w, introduced in [10]. A simple corollary of the characterization by return words and a result of [10] is that the derivated word $d_{\mathbf{u}}(w)$ is also a Sturmian word (see Theorem 1). This simple corollary follows also from other results. For instance, it follows from [1], where the authors investigate the derivated word

* Corresponding author.
E-mail address: medkokat@fjfi.cvut.cz (K. Medková).

https://doi.org/10.1016/j.tcs.2018.06.037
0304-3975/

of a standard Sturmian word and give its precise description. It also follows from the investigation of a more general setting in [7], which may in fact be used to describe derivated words of any episturmian word — generalized Sturmian words [12].

By the main result of [10], if $\mathbf{u}$ is a fixed point of a primitive morphism, the set of all derivated words of $\mathbf{u}$ is finite (the result also follows from [13]). In this case, again by [10], a derivated word itself is a fixed point of a primitive morphism.

In this article we study derivated words of fixed points of primitive Sturmian morphisms. By the results of [18], any primitive Sturmian morphism may be decomposed using elementary Sturmian morphisms — generators of the Sturmian monoid. In Theorems 14 and 18, we describe the relation between the set of derivated words of a Sturmian sequence $\mathbf{u}$ and the set of derivated words of $\varphi(\mathbf{u})$, where φ is a generator of the Sturmian monoid.

The main result of our article is an exact description of the morphisms fixing the derivated words $d_{\mathbf{u}}(w)$ of $\mathbf{u}$, where $\mathbf{u}$ is fixed by a Sturmian morphism ψ and w is its prefix. For this purpose, we introduce an operation Δ acting on the set of Sturmian morphisms with unique fixed point, see Definition 22. Iterating this operation we create the desired list of the morphisms as stated in Theorem 25. The Sturmian morphisms with two fixed points are treated separately, see Proposition 29.

We continue our study by counting the number of derivated words, in particular by counting the distinct elements in the sequence $\left(\Delta^k(\psi)\right)_{k\geq 1}$. This number depends on the decomposition of ψ into the generators of the special Sturmian monoid, see below in Section 2.3.

Using this decomposition, Propositions 36 and 37 provide the exact number of derivated words for two specific classes of Sturmian morphisms.

For a general Sturmian morphism ψ, Corollary 35 gives a sharp upper bound on their number. The upper bound depends on the number of the elementary morphisms in the decomposition of ψ. In the last section, we give some comments and state open questions.

2. Preliminaries

An *alphabet* $\mathcal{A}$ is a finite set of symbols called *letters*. A *finite word* of length n over $\mathcal{A}$ is a string $u = u_0u_1\cdots u_{n-1}$, where $u_i \in \mathcal{A}$ for all $i = 0, 1, \ldots, n-1$. The *length* of u is denoted by $|u| = n$. By $|u|_a$ we denote the number of copies of the letter a used in u, i.e. $|u|_a = \#\{i \in \mathbb{N} \colon i < n, u_i = a\}$. The set of all finite words over $\mathcal{A}$ together with the operation of concatenation forms a monoid $\mathcal{A}^*$. Its neutral element is the *empty word* ε and $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$. On this monoid we work with two operations which preserve the length of words. The *mirror image* or *reversal* of a word $u = u_0u_1\cdots u_{n-1} \in \mathcal{A}^*$ is the word $\overline{u} = u_{n-1}u_{n-2}\cdots u_1u_0$. The *cyclic shift* of u is the word

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

An *infinite word* over $\mathcal{A}$ is a sequence $\mathbf{u} = u_0u_1u_2\cdots = (u_i)_{i\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ with $u_i \in \mathcal{A}$ for all $i \in \mathbb{N} = \{0, 1, 2, \ldots\}$. Bold letters are systematically used to denote infinite words throughout this article.

A finite word $p \in \mathcal{A}^*$ is a *prefix* of $u = u_0u_1\cdots u_{n-1}$ if $p = u_0u_1u_2\cdots u_{k-1}$ for some $k \leq n$, the word $u_ku_{k+1}\cdots u_{n-1}$ is denoted $p^{-1}u$. Similarly, $p \in \mathcal{A}^*$ is a prefix of $\mathbf{u} = u_0u_1u_2\cdots$ if $p = u_0u_1u_2\cdots u_{k-1}$ for some integer k. We usually abbreviate $u_0u_1u_2\cdots u_{k-1} = \mathbf{u}_{[0,k)}$.

A finite word w is a *factor* of $\mathbf{u} = u_0u_1u_2\cdots$ if there exists an index i such that w is a prefix of the infinite word $u_iu_{i+1}u_{i+2}\cdots$. The index i is called an *occurrence* of w in $\mathbf{u}$. If each factor of $\mathbf{u}$ has infinitely many occurrences in $\mathbf{u}$, the word $\mathbf{u}$ is *recurrent*.

The *language* $\mathcal{L}(\mathbf{u})$ *of an infinite word* $\mathbf{u}$ is the set of all its factors. The mapping $\mathcal{C}_{\mathbf{u}} : \mathbb{N} \mapsto \mathbb{N}$ defined by $\mathcal{C}_{\mathbf{u}}(n) = \#\{w \in \mathcal{L}(\mathbf{u}) : |w| = n\}$ is called the *factor complexity* of the word $\mathbf{u}$.

An infinite word $\mathbf{u}$ is *eventually periodic* if $\mathbf{u} = wvvvvv\cdots$ for some $v, w \in \mathcal{A}^*$. If w is the shortest such word possible, we say that $|w|$ is the *preperiod* of $\mathbf{u}$; if v is the shortest possible, we say that $|v|$ is the *period* of $\mathbf{u}$. If $\mathbf{u}$ is not eventually periodic, it is *aperiodic*. A factor w of $\mathbf{u}$ is a *right special* factor if there exist at least two letters $a, b \in \mathcal{A}$ such that wa, wb belong to the language $\mathcal{L}(\mathbf{u})$. A *left special* factor is defined analogously.

An infinite word $\mathbf{u}$ is eventually periodic if and only if $\mathcal{L}(\mathbf{u})$ contains only finitely many right special factors. Equivalently, $\mathbf{u}$ is eventually periodic if and only if its factor complexity $\mathcal{C}_{\mathbf{u}}$ is bounded. On the other hand, the factor complexity of any aperiodic word satisfies $\mathcal{C}_{\mathbf{u}}(n) \geq n + 1$ for every $n \in \mathbb{N}$.

An infinite word $\mathbf{u}$ with $\mathcal{C}_{\mathbf{u}}(n) = n + 1$ for each $n \in \mathbb{N}$ is called Sturmian. A Sturmian word is *standard* (or *characteristic*) if each of its prefixes is a left special factor.

2.1. Derivated words

Consider a prefix w of an infinite recurrent word $\mathbf{u}$. Let $i < j$ be two consecutive occurrences of w in $\mathbf{u}$. The string $u_iu_{i+1}\cdots u_{j-1}$ is a *return word* to w in $\mathbf{u}$. The set of all return words to w in $\mathbf{u}$ is denoted by $\mathcal{R}_{\mathbf{u}}(w)$. Let us suppose that the set of return words to w is finite, i.e. $\mathcal{R}_{\mathbf{u}}(w) = \{r_0, r_1, \ldots, r_{k-1}\}$. The word $\mathbf{u}$ can be written as unique concatenation of the return words $\mathbf{u} = r_{s_0}r_{s_1}r_{s_2}\cdots$. The *derivated* word of $\mathbf{u}$ with respect to the prefix w is the infinite word

$$d_{\mathbf{u}}(w) = s_0s_1s_2\cdots$$

over the alphabet of cardinality $\#\mathcal{R}_{\mathbf{u}}(w) = k$. In his original definition, Durand [10] fixed the alphabet of the derivated word to the set $\{0, 1, \ldots, k-1\}$. Moreover, Durand's definition requires that for $i < j$ the first occurrence of r_i in $\mathbf{u}$ is less than the first occurrence of r_j in $\mathbf{u}$. In particular, a derivated word always starts with the letter 0. In the article [1], where derivated words of standard Sturmian words are studied, the authors required that the starting letters of the original word and its derivated word coincide. For our purposes, we do not need to fix the alphabet of derivated words: two derivated words which differ only by a permutation of letters are identified one with another.

In the sequel, we work only with infinite words which are *uniformly recurrent*, i.e. each prefix w of $\mathbf{u}$ occurs in $\mathbf{u}$ infinitely many times and the set $\mathcal{R}_{\mathbf{u}}(w)$ is finite. Our aim is to describe the set

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

Clearly, if a prefix w is not right special, then there exists a unique letter x such that $wx \in \mathcal{L}(\mathbf{u})$. Thus the occurrences of w and wx coincide, $\mathcal{R}_{\mathbf{u}}(w) = \mathcal{R}_{\mathbf{u}}(wx)$ and $d_{\mathbf{u}}(w) = d_{\mathbf{u}}(wx)$. If $\mathbf{u}$ is not eventually periodic, then w is a prefix of a right special prefix of $\mathbf{u}$. Therefore for an aperiodic uniformly recurrent word $\mathbf{u}$ we have

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

2.2. Sturmian words

Any Sturmian word $\mathbf{u}$ can be identified with an upper or lower mechanical word. A mechanical word is described by two parameters: slope and intercept. The slope is an irrational number $\gamma \in (0, 1)$ and the intercept is a real number $\rho \in [0, 1)$. To define the lower mechanical word $\mathbf{s}(\gamma, \rho) = (s_n(\gamma, \rho))_{n \in \mathbb{N}}$ we put $I_0 = [0, 1-\gamma)$. The nth letter of $\mathbf{s}(\gamma, \rho)$ is as follows:

$$s_n(\gamma, \rho) = \begin{cases} 0 & \text{if the number } \gamma n + \rho \mod 1 \text{ belongs to } I_0, \\ 1 & \text{otherwise.} \end{cases}$$

The definition of the upper mechanical word $\mathbf{s}'(\gamma, \rho) = \big(s'_n(\gamma, \rho)\big)_{n \in \mathbb{N}}$ is analogous, it just uses the interval $I_0 = (0, 1-\gamma]$. Let us stress that $s_n(\gamma, \rho) \neq s'_n(\gamma, \rho)$ for at most two neighboring indices n and $n+1$. All upper and lower mechanical words with irrational slope are Sturmian and any Sturmian word equals to a lower or to an upper mechanical word. Let us stress that one-sided Sturmian words with irrational slope are always uniformly recurrent. The language of a Sturmian word depends only on γ. The number γ is in fact the density of the letter 1, i.e., $\gamma = \lim_{n \to \infty} \frac{1}{n} \# \big\{ i \in \mathbb{N} : i < n, s_i(\gamma, \rho) = 1 \big\}$. Consequently, $1 - \gamma$ is the density of the letter 0.

For any irrational $\gamma \in (0, 1)$ there exists a unique mechanical word $\mathbf{c}(\gamma)$ with slope γ such that both $0\mathbf{c}(\gamma)$ and $1\mathbf{c}(\gamma)$ are Sturmian. The word $\mathbf{c}(\gamma)$ is a standard Sturmian word and $\mathbf{c}(\gamma) = \mathbf{s}(\gamma, \gamma) = \mathbf{s}'(\gamma, \gamma)$. Many further properties of Sturmian words can be found in [16,5].

For our study of derivated words, the following result of Vuillon from [22] is important: a word $\mathbf{u}$ is Sturmian if and only if any prefix of $\mathbf{u}$ has exactly two return words. By combining this result with [10], we obtain an essential observation about derivated words of Sturmian words, which also follows from [1].

Theorem 1. *If $\mathbf{u}$ is a Sturmian word and w is a prefix of $\mathbf{u}$, then its derivated word $d_{\mathbf{u}}(w)$ is Sturmian as well.*

Proof. Set $\mathbf{v} = d_{\mathbf{u}}(w)$. Let p be a prefix of $\mathbf{v}$. Due to Proposition 2.6 in [10], there exists a prefix q of $\mathbf{u}$ such that $d_{\mathbf{v}}(p) = d_{\mathbf{u}}(q)$. By Vuillon's characterization of Sturmian words, the word $d_{\mathbf{u}}(q)$ is binary. It means that any prefix p of $\mathbf{v}$ has two return words in $\mathbf{v}$ and so $\mathbf{v}$ is Sturmian. □

Remark 2 *(Historical).* The Sturmian words (sequences) were originally defined by Hedlund and Morse in [19]. Their definition is more general as they consider also biinfinite words and (in terms of our definition above) rational slopes. Hence their Sturmian words may not be recurrent. For details on the history of definition of Sturmian words see [11], especially the historical remark at page 146. Interestingly enough, the term *derivated* sequence is also used in [19], however, its definition differs from our one (as taken from [10]): Using again our terminology, their derivated word is a derivated word with respect to a one-letter word in a biinfinite Sturmian word.

2.3. Sturmian morphisms

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

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

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

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

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

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

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

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

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

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

$$\alpha a^k \beta = \beta b^k \alpha \quad \text{and} \quad a \alpha^k b = b \beta^k a \qquad \text{for any } k \in \mathbb{N}\,. \tag{2}$$

In [21], the author reveals a presentation of the Sturmian monoid which includes the special Sturmian monoid $\mathcal{M} = \langle \varphi_a, \varphi_b, \varphi_\alpha, \varphi_\beta \rangle$. A presentation of the special Sturmian monoid follows from this result. It is also given explicitly in [15]:

Theorem 3. *Let $w, v \in \{a, b, \alpha, \beta\}^*$. The morphism φ_w is equal to φ_v if and only if the word v can be obtained from w by applying the rewriting rules (2).*

Note that the presentation of a generalization of the Sturmian monoid, the so-called *episturmian monoid*, is also known, see [20]. The next lemma summarizes several simple and well-known properties of Sturmian morphisms we exploit in the sequel.

Lemma 4. *Let $w \in \{a, b, \alpha, \beta\}^+$.*

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

For $w \in \{a, b, \alpha, \beta\}^*$ the rules (2) preserve positions in w occupied by Latin letters $\{a, b\}$ and positions occupied by Greek letters $\{\alpha, \beta\}$. We define that $a < b$ and $\alpha < \beta$ which allows the following definition.

Definition 5. Let $w \in \{a, b, \alpha, \beta\}^*$. The lexicographically greatest word in $\{a, b, \alpha, \beta\}^*$ which can be obtained from w by application of rewriting rules (2) is denoted $N(w)$. If $\psi = \varphi_w$, then the word $N(w)$ is the *normalized name* of the morphism ψ and it is also denoted by $N(\psi) = N(w)$.

The next lemma is a direct consequence of Theorem 3.

Lemma 6. *Let $w \in \{a, b, \alpha, \beta\}^*$. We have $w = N(w)$ if and only if w does not contain $\alpha a^k \beta$ or $a\alpha^k b$ as a factor for any $k \in \mathbb{N}$. In particular, if $w \in \{a, b, \alpha, \beta\}^* \setminus \{a, \alpha\}^*$, the normalized name $N(w)$ has prefix either $a^i \beta$ or $\alpha^i b$ for some $i \in \mathbb{N}$.*

Example 7. Since $\psi = \varphi_a \varphi_b \varphi_\alpha \varphi_b = \varphi_b \varphi_a \varphi_\alpha \varphi_b = \varphi_b \varphi_b \varphi_\beta \varphi_a$, the normalized name of ψ is $N(\psi) = bb\beta a$.

The morphism $E : 0 \to 1, 1 \to 0$ which exchanges letters in words over $\{0, 1\}$ cannot change the factor complexity of an infinite word. Thus, E is clearly a Sturmian morphism. But E does not belong to the monoid $\mathcal{M} = \langle \varphi_a, \varphi_b, \varphi_\alpha, \varphi_\beta \rangle$. In fact, E is the only missing morphism. More precisely, any Sturmian morphism ψ either belongs to $\mathcal{M}$ or $\psi = \eta \circ E$, where $\eta \in \mathcal{M}$ (see [18]). To generate the whole monoid of Sturmian morphisms St, one needs only three morphisms, say E, φ_a and φ_b (see [16]). We have

$$\varphi_\alpha = E \varphi_a E \quad \text{and} \quad \varphi_\beta = E \varphi_b E. \tag{3}$$

Our aim is to study derivated words of fixed points of Sturmian morphisms. If $\mathbf{u}$ is a fixed point of ψ, it is also a fixed point of ψ^2. Due to (3), the square ψ^2 always belongs to $\mathcal{M}$. To illustrate why this is true, assume, e.g., that $\psi \in St = \langle E, \varphi_a, \varphi_b \rangle$ equals $\psi = \varphi_a E \varphi_b \varphi_a$. Using (3) and the fact that E^2 is the identity morphism, we have

$$\psi = \varphi_a E \varphi_b E E \varphi_a E E = \varphi_a \varphi_\beta \varphi_\alpha E$$

and hence

$$\psi^2 = \varphi_a\varphi_\beta\varphi_\alpha E\varphi_a\varphi_\beta\varphi_\alpha E = \varphi_a\varphi_\beta\varphi_\alpha E\varphi_a EE\varphi_\beta EE\varphi_\alpha E = \varphi_a\varphi_\beta\varphi_\alpha\varphi_\alpha\varphi_b\varphi_a \in \mathcal{M}.$$

Therefore we may restrict ourselves to fixed points of morphisms from the special Sturmian monoid $\mathcal{M}$. Note that this would not be true if we consider only the morphisms from $\langle\varphi_a, \varphi_b\rangle$, see also Lemma 4.

Example 8. The Fibonacci word is the fixed point of the morphism $\tau : 0 \to 01, 1 \to 0$. The morphism τ is Sturmian, but $\tau \notin \mathcal{M}$. We see that $\tau = \varphi_b \circ E$ and by the relations (3) we have $\tau^2 = \varphi_b\varphi_\beta$.

Remark 9. Two infinite words $\mathbf{u}$ and $E(\mathbf{u})$ over the alphabet $\{0, 1\}$ coincide up to a permutation of the letters 0 and 1. If a word $\mathbf{u}$ is a fixed point of a morphism φ_w, then $E(\mathbf{u})$ is a fixed point of the morphism $E \circ \varphi_w \circ E = \varphi_v$ for some v. By (3), the word v is obtained from w by exchange of letters $a \leftrightarrow \alpha$ and $b \leftrightarrow \beta$. Therefore we introduce the following morphism $F : \{a, b, \alpha, \beta\}^* \mapsto \{a, b, \alpha, \beta\}^*$ by

$$F(a) = \alpha, \quad F(\alpha) = a, \quad F(b) = \beta, \quad F(\beta) = b. \tag{4}$$

This notation enables us to formulate two useful facts on composition of E with morphisms from $\mathcal{M}$. Namely,

$$E \circ \varphi_w \circ E = \varphi_{F(w)} \qquad \text{and} \qquad (\varphi_w \circ E)^2 = \varphi_{wF(w)}\,. \tag{5}$$

Later on we will need the following statement on the morphism F. First we recall two classical results on word equations:

Lemma 10 *([17]). Let $y \in \mathcal{A}^*$ and $x, z \in \mathcal{A}^+$. Then $xy = yz$ if and only if there are $u, v \in \mathcal{A}^*$ and $\ell \in \mathbb{N}$ such that $x = uv$, $z = vu$ and $y = (uv)^\ell u$.*

Lemma 11 *([17]). Let $x, y \in \mathcal{A}^+$. The following three conditions are equivalent:*

(i) $xy = yx$;
(ii) There exist integers $i, j > 0$ such that $x^i = y^j$;
(iii) There exist $z \in \mathcal{A}^+$ and integers $p, q > 0$ such that $x = z^p$ and $y = z^q$.

With these two lemmas we prove the following result on word equations involving the morphism F. Note that this result is within the general setting considered in [9], however we give an explicit solution of cases that we need later.

Lemma 12. *Let z and p be nonempty words from $\{a, b, \alpha, \beta\}^+$.*

(i) If $zp = F(p)F(z)$, then there is $x \in \{a, b, \alpha, \beta\}^+$ such that

$$z = x\big(F(x)x\big)^i \quad \text{and} \quad p = \big(F(x)x\big)^j F(x) \quad \text{for some } i, j \in \mathbb{N}.$$

(ii) If $zp = pF(z)$, then there is $x \in \{a, b, \alpha, \beta\}^+$ such that

$$z = \big(F(x)x\big)^i \quad \text{and} \quad p = \big(F(x)x\big)^j F(x) \quad \text{for some } i, j \in \mathbb{N}.$$

Proof. We prove Item (i) by induction on $|zp| \geq 2$. If $|z| = |p|$, then $z = F(p)$ and the statement is true for $x = z$ and $i = j = 0$.

Assume $|z| > |p|$ (the case of $|z| < |p|$ is analogous). There must be a nonempty word q such that $z = F(p)q$ and this yields $qp = F(z) = pF(q)$. By Lemma 10 there are words u and v and $\ell \in \mathbb{N}$ such that $q = uv$, $p = (uv)^\ell u$ and $F(q) = vu$. This implies that $vu = F(u)F(v)$ and we can apply the induction hypothesis as $|uv| < |pz|$. Therefore, there are x and $s, r \in \mathbb{N}$ such that $v = x\big(F(x)x\big)^s$ and $u = \big(F(x)x\big)^t F(x)$. Putting this altogether we obtain

$$\begin{aligned} q &= uv = \big(F(x)x\big)^t F(x)x\big(F(x)x\big)^s = \big(F(x)x\big)^{t+s+1}, \\ p &= (uv)^\ell u = \big(F(x)x\big)^j F(x), \quad \text{with } j = \ell(t+s+1) + t, \\ z &= F(p)q = x\big(F(x)x\big)^i, \quad \text{with } i = \ell(t+s+1) + 2t + s + 1. \end{aligned}$$

To prove Item (ii), we apply Lemma 10 on $zp = pF(z)$. We have $z = uv$, $F(z) = vu$ and $p = (uv)^\ell u$ for some words u and v and $\ell \in \mathbb{N}$. It follows that $vu = F(u)F(v)$ and so, by Item (i), there is x such that $v = x\big(F(x)x\big)^i$ and $u = \big(F(x)x\big)^j F(x)$ for some $i, j \in \mathbb{N}$. Using all these equations we finish the proof by stating that

$$z = uv = \big(F(x)x\big)^j F(x)x\big(F(x)x\big)^i = \big(F(x)x\big)^{j+i+1} \quad \text{and} \quad p = (uv)^\ell u = \big(F(x)x\big)^{\ell(j+i+1)+j} F(x). \qquad \square$$

3. Derivated words of Sturmian preimages

In this section we study relations between derivated words of a Sturmian word and derivated words of its preimage under one of the morphisms $\varphi_a, \varphi_b, \varphi_\alpha$ and φ_β. We prove that the set of all derivated words of these two infinite words coincide up to at most one derivated word, see Theorems 14 and 18. This will be crucial fact for proving the main results of this paper. Because of (3), the roles of φ_a and φ_α and, analogously, the roles of φ_b and φ_β are symmetric. Therefore we can restrict the statements and proofs in this section to the morphisms φ_a and φ_b with no loss of generality. Again we use results from [16], in particular this slightly modified Proposition 2.3.2:

Proposition 13 *([16]). Let* $\mathbf{x}$ *be an infinite word.*

(i) If $\varphi_b(\mathbf{x})$ *is Sturmian, then* $\mathbf{x}$ *is Sturmian.*
(ii) If $\varphi_a(\mathbf{x})$ *is Sturmian and* $\mathbf{x}$ *starts with the letter* 1*, then* $\mathbf{x}$ *is Sturmian.*

Theorem 14. *Let* $\mathbf{u}$ *and* $\mathbf{u}'$ *be Sturmian words such that* $\mathbf{u} = \varphi_b(\mathbf{u}')$*. Then the sets of their derivated words satisfy*

$$\mathrm{Der}(\mathbf{u}) = \mathrm{Der}(\mathbf{u}') \cup \{\mathbf{u}'\}\,.$$

The proof of the previous theorem is split into two parts: In Proposition 16, Item (i) says $\{\mathbf{u}'\} \subset \mathrm{Der}(\mathbf{u})$ and Item (ii) says $\mathrm{Der}(\mathbf{u}) \subset \mathrm{Der}(\mathbf{u}') \cup \{\mathbf{u}'\}$. Proposition 17 says $\mathrm{Der}(\mathbf{u}') \subset \mathrm{Der}(\mathbf{u})$. Proofs of these propositions use the following simple property of the injective morphism φ_b.

Lemma 15. *Let* $\mathbf{u} = \varphi_b(\mathbf{u}')$ *be a Sturmian word. If* $p0 \in \mathcal{L}(\mathbf{u})$ *and* 0 *is a prefix of* p*, then there exists a unique factor* $p' \in \mathcal{L}(\mathbf{u}')$ *such that* $p0 = \varphi_b(p')0$*.*

Proposition 16. *Let* $\mathbf{u}$ *and* $\mathbf{u}'$ *be Sturmian words such that* $\mathbf{u} = \varphi_b(\mathbf{u}')$ *and let* w *be a prefix of* $\mathbf{u}$*.*

(i) If $|w| = 1$*, then* $d_{\mathbf{u}}(w) = \mathbf{u}'$ *(up to a permutation of letters).*
(ii) If $|w| > 1$*, then there exists a prefix* w' *of* $\mathbf{u}'$ *such that* $|w'| < |w|$ *and* $d_{\mathbf{u}}(w) = d_{\mathbf{u}'}(w')$ *(up to a permutation of letters). Moreover, if* w *is right special,* w' *is right special as well.*

Proof. Since $\varphi_b(0) = 0$ and $\varphi_b(1) = 01$, the word $\mathbf{u} = \varphi_b(\mathbf{u}')$ has a prefix 0 and the letter 1 is in $\mathbf{u}$ separated by blocks 0^k with $k \geq 1$. Therefore, the two return words in $\mathbf{u}$ to the word $w = 0$ are $r_0 = 0$ and $r_1 = 01$. We may write $\mathbf{u} = r_{s_0} r_{s_1} r_{s_2} \cdots$, where $r_{s_j} \in \{r_0, r_1\}$ and thus $d_{\mathbf{u}}(w) = s_0 s_1 s_2 \cdots$. Since $r_0 = \varphi_b(0)$ and $r_1 = \varphi_b(1)$, we obtain also $\varphi_b(\mathbf{u}') = \mathbf{u} = \varphi_b(s_0)\varphi_b(s_1)\varphi_b(s_2)\cdots = \varphi_b(s_0 s_1 s_2 \cdots)$. The statement in (i) now follows from injectivity of φ_b.

Now suppose that the prefix w of $\mathbf{u}$ is of length > 1. As explained earlier, it suffices to consider right special prefixes. Since the letter 1 is always followed by 0, each right special factor must end in 0. So the first and the last letter of w is 0, hence by Lemma 15 there is a unique prefix w' of $\mathbf{u}'$ such that $\varphi_b(w')0 = w$. Let r_0 and r_1 be the two return words to w and let $\mathbf{u} = r_{s_0} r_{s_1} r_{s_2} \cdots$. Since the first letter of both r_0 and r_1 is 0, there are uniquely given r'_0 and r'_1 such that $r_0 = \varphi_b(r'_0)$ and $r_1 = \varphi_b(r'_1)$ and $\mathbf{u}' = r'_{s_0} r'_{s_1} r'_{s_2} \cdots$.

Clearly w' is a prefix of $r'_{s_j} r'_{s_{j+1}} r'_{s_{j+2}} \cdots$ for all $j \in \mathbb{N}$ and so the number $|r'_{s_0} r'_{s_1} \cdots r'_{s_k}|$ is an occurrence of w' in $\mathbf{u}'$ for all $k \in \mathbb{N}$. Let $i > 0$ be an occurrence of w' in $\mathbf{u}'$. It follows that $\varphi_b\big(\mathbf{u}'_{[0,i)}\big)w$ is a prefix of $\mathbf{u}$ and $|\varphi_b\big(\mathbf{u}'_{[0,i)}\big)|$ is an occurrence of w in $\mathbf{u}$. There must be $j \in \mathbb{N}$ such that $\varphi_b\big(\mathbf{u}'_{[0,i)}\big) = r_{s_0} r_{s_1} \cdots r_{s_j}$ and hence, by injectivity of φ_b, $\mathbf{u}'_{[0,i)} = r'_{s_0} r'_{s_1} \cdots r'_{s_j}$ and $i = |r'_{s_0} r'_{s_1} \cdots r'_{s_j}|$.

We have proved that the numbers 0 and $|r'_{s_0} r'_{s_1} \cdots r'_{s_j}|$, $j = 0, 1, \ldots$, are all occurrences of w' in $\mathbf{u}'$. It follows that r'_0 and r'_1 are the two return words to w' in $\mathbf{u}'$ and

$$d_{\mathbf{u}'}(w') = s_0 s_1 s_2 \cdots = d_{\mathbf{u}}(w).$$

Since $w = \varphi_b(w')0$ is a right special factor, we must have that both $\varphi_b(w')00$ and $\varphi_b(w')01$ are factors of $\mathbf{u}$. It follows that both $w'0$ and $w'1$ are factors of $\mathbf{u}'$ and w' is right special. $\square$

Proposition 17. *Let* $\mathbf{u}$ *and* $\mathbf{u}'$ *be Sturmian words such that* $\mathbf{u} = \varphi_b(\mathbf{u}')$ *and let* w' *be a nonempty right special prefix of* $\mathbf{u}'$*. Then* $d_{\mathbf{u}'}(w') = d_{\mathbf{u}}(w)$*, where* $w = \varphi_b(w')0$*.*

Proof. Let r'_0 and r'_1 be the two return words to w' in $\mathbf{u}'$ and $d_{\mathbf{u}'}(w') = s_0 s_1 s_2 \cdots$. Put $w = \varphi_b(w')0$, $r_0 = \varphi_b(r'_0)$ and $r_1 = \varphi_b(r'_1)$. We obtain

$$\mathbf{u} = \varphi_b(\mathbf{u}') = \varphi_b(r'_{s_0} r'_{s_1} r'_{s_2} \cdots) = r_{s_0} r_{s_1} r_{s_2} \cdots$$

Clearly, w is prefix of $r_{s_k} r_{s_{k+1}} r_{s_{k+2}} \cdots$ for all $k \in \mathbb{N}$ and $|r_{s_0} r_{s_1} \cdots r_{s_j}|$ is an occurrence of w in $\mathbf{u}$ for all $j \in \mathbb{N}$.

Assume now $i > 0$ is an occurrence of w in $\mathbf{u}$. This means that $\mathbf{u}_{[0,i)}w$ is a prefix of $\mathbf{u}$ and hence, by Lemma 15 (note that w begins with 0), there must be p' a prefix of $\mathbf{u}'$ such that $\varphi_b(p') = \mathbf{u}_{[0,i)}$ and $p'w'$ is a prefix of $\mathbf{u}'$. Since $|p'|$ is an occurrence of w' in $\mathbf{u}'$, there is $j \in \mathbb{N}$ such that $p' = r'_{s_0}r'_{s_1}\cdots r'_{s_j}$. It follows that

$$\mathbf{u}_{[0,i)} = \varphi_b(r'_{s_0}r'_{s_1}\cdots r'_{s_j}) = r_{s_0}r_{s_1}\cdots r_{s_j}$$

and $i = |r_{s_0}r_{s_1}\cdots r_{s_j}|$.

So, again as in the previous proof, we have shown that the numbers 0 and $|r_{s_0}r_{s_1}\cdots r_{s_j}|$, $j = 0, 1, \ldots$, are all occurrences of w in $\mathbf{u}$. It follows that r_0 and r_1 are the two return words to w in $\mathbf{u}$ and

$$d_{\mathbf{u}}(w) = s_0s_1s_2\cdots = d_{\mathbf{u}'}(w'). \qquad \square$$

Theorem 18. *Let $\mathbf{u}$ and $\mathbf{u}'$ be Sturmian words such that $\mathbf{u}$ starts with the letter 1 and $\mathbf{u} = \varphi_a(\mathbf{u}')$. Then $\mathbf{u}'$ starts with 1 and the sets of their derivated words coincide, i.e.,*

$$\mathrm{Der}(\mathbf{u}) = \mathrm{Der}(\mathbf{u}')\,.$$

In particular, for any prefix w of $\mathbf{u}$ there exists a prefix w' of $\mathbf{u}'$ such that $|w'| \le |w|$ and $d_{\mathbf{u}}(w) = d_{\mathbf{u}'}(w')$ (up to a permutation of letters). Moreover, if w is right special, w' is right special as well.

Proof. The morphisms φ_a and φ_b are conjugate, that is, $0\varphi_a(x) = \varphi_b(x)0$ for each word x. This means that for any prefix $u'_0u'_1u'_2\cdots u'_k$ of $\mathbf{u}'$ we have $0\varphi_a(u'_0u'_1u'_2\cdots u'_k) = \varphi_b(u'_0u'_1u'_2\cdots u'_k)0$. As this holds true for each k, we obtain $0\mathbf{u} = 0\varphi_a(\mathbf{u}') = \varphi_b(\mathbf{u}')$.

Denote $\mathbf{v} = v_0v_1v_2\cdots = 0u_0u_1u_2\cdots$. We have $v_i = u_{i-1}$ for each $i \ge 1$. Let w be a nonempty prefix of $\mathbf{u}$ and (i_n) be the increasing sequence of its occurrences in $\mathbf{u}$. Note that w starts with the letter 1. This letter is in $\mathbf{u}$ surrounded by 0's. Thus the sequence (i_n) is also the sequence of occurrences of $0w$ in $\mathbf{v}$ and thus $d_{\mathbf{v}}(0w) = d_{\mathbf{u}}(w)$. It follows that

$$\mathrm{Der}(\mathbf{u}) = \{d_{\mathbf{v}}(v) \colon v \text{ is a prefix of } \mathbf{v} \text{ and } |v| > 1\}\,.$$

We finish the proof by applying Theorem 14 and Proposition 16 to the word $\mathbf{v} = \varphi_b(\mathbf{u}')$. $\square$

The only case which is not treated by Theorems 14 and 18, namely the case when $\mathbf{u} = \varphi_a(\mathbf{u}')$ and $\mathbf{u}$ begins with 0, can be translated into one of the previous cases.

Lemma 19. *Let $\mathbf{u}$ be a Sturmian word such that $\mathbf{u}$ starts with the letter 0 and $\mathbf{u} = \varphi_a(\mathbf{u}')$ for some word $\mathbf{u}'$. Then there exists a Sturmian word $\mathbf{v}$ such that $\mathbf{u}' = 0\mathbf{v}$ and $\mathbf{u} = \varphi_b(\mathbf{v})$.*

Proof. Since $\mathbf{u}$ starts with 0, the form of φ_a implies that $\mathbf{u}' = 0\mathbf{v}$ for some Sturmian word $\mathbf{v}$. As $0\varphi_a(x) = \varphi_b(x)0$ for each word x, we have

$$\mathbf{u} = \varphi_a(\mathbf{u}') = \varphi_a(0\mathbf{v}) = 0\varphi_a(\mathbf{v}) = \varphi_b(\mathbf{v}). \qquad \square$$

To sum up the results of this section, let us assume we have a sequence of Sturmian words $\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \ldots$ such that $\mathbf{u} = \mathbf{u}_0$ and for every $i \in \mathbb{N}$ one of the following is true:

(i) $\mathbf{u}_i = \varphi_b(\mathbf{u}_{i+1})$ or $\mathbf{u}_i = \varphi_\beta(\mathbf{u}_{i+1})$,
(ii) $\mathbf{u}_i$ begins with 1 and $\mathbf{u}_i = \varphi_a(\mathbf{u}_{i+1})$,
(iii) $\mathbf{u}_i$ begins with 0 and $\mathbf{u}_i = \varphi_\alpha(\mathbf{u}_{i+1})$.

If (i) holds for $\mathbf{u}_i$, then by Theorem 14

$$\mathrm{Der}(\mathbf{u}_i) = \mathrm{Der}(\mathbf{u}_{i+1}) \cup \{\mathbf{u}_{i+1}\},$$

moreover, $\mathbf{u}_{i+1}$ is the derivated word of the first letter of $\mathbf{u}_i$. This first letter is also the shortest right special prefix. If (ii) or (iii) holds for $\mathbf{u}_i$, then by Theorem 18

$$\mathrm{Der}(\mathbf{u}_i) = \mathrm{Der}(\mathbf{u}_{i+1}).$$

The crucial assumption, namely the existence of the above described sequence $(\mathbf{u}_k)_{k\ge 0}$, is guaranteed by the well-known fact on the desubstitution of Sturmian words (see, e.g., [14] and [19] and also Lemma 19). Here we formulate this fact as the following theorem:

Theorem 20 *([14], [19]). An infinite binary word $\mathbf{u}$ is Sturmian if and only if there exists an infinite word $\mathbf{w} = w_0 w_1 w_2 \cdots$ over the alphabet $\{a, b, \alpha, \beta\}$ and an infinite sequence $(\mathbf{u}_i)_{i\geq 0}$, such that $\mathbf{u} = \mathbf{u}_0$ and $\mathbf{u}_i = \varphi_{w_i}(\mathbf{u}_{i+1})$ for all $i \in \mathbb{N}$.*

In the following section we work only with the sequence $(\mathbf{u}_i)_{i\geq 0}$ corresponding to a fixed point $\mathbf{u}$ of a Sturmian morphism ψ. The next lemma provides us a simple technical tool for a description of the elements $\mathbf{u}_i$ as fixed points of some Sturmian morphisms.

Lemma 21. *Let ξ and η be Sturmian morphisms and $\mathbf{u} = \big(\xi \circ \eta\big)(\mathbf{u})$. If $\mathbf{u} = \xi(\mathbf{u}')$ for some $\mathbf{u}'$, then $\mathbf{u}'$ is the fixed point of the morphism $\eta \circ \xi$, i.e. $\mathbf{u}' = \big(\eta \circ \xi\big)(\mathbf{u}')$.*

Proof. For any Sturmian morphism ξ, the equation $\xi(\mathbf{x}) = \xi(\mathbf{y})$ implies that $\mathbf{x} = \mathbf{y}$. We deduce that

$$\xi(\mathbf{u}') = \mathbf{u} = \big(\xi \circ \eta\big)(\mathbf{u}) = \big(\xi \circ \eta\big)\big(\xi(\mathbf{u}')\big) = \big(\xi \circ \eta \circ \xi\big)(\mathbf{u}'),$$

and so $\mathbf{u}' = \big(\eta \circ \xi\big)(\mathbf{u}')$. $\square$

4. Derivated words of fixed points of Sturmian morphisms

Let $\mathbf{u}$ be an fixed point of a primitive Sturmian morphism (note that if the morphism is primitive, all its fixed points are aperiodic). It is known due to Durand [10] that the set $\mathrm{Der}(\mathbf{u})$ is finite (as the morphism is primitive). Put

$$\mathrm{Der}(\mathbf{u}) = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\ell\}.$$

Our main result is an algorithm that returns a list of Sturmian morphisms $\psi_1, \psi_2, \ldots, \psi_\ell$ such that $\mathbf{x}_i$ is a fixed point of ψ_i (up to a permutation of letters) for all i such that $1 \leq i \leq \ell$.

As we have noticed before, we can restrict ourselves to the morphisms belonging to the monoid $\mathcal{M} = \langle \varphi_a, \varphi_b, \varphi_\alpha, \varphi_\beta \rangle$. Let us recall (see Lemma 4) that a morphism from $\langle \varphi_a, \varphi_b \rangle$ or from $\langle \varphi_\alpha, \varphi_\beta \rangle$ is not primitive and has no aperiodic fixed point. Thus we consider only morphisms φ_w whose normalized name w contains at least one Latin and one Greek letter.

We will treat two cases separately. The first one is the case when the morphism φ_w has only one fixed point. Lemma 4 says that in such a case $w \notin \{a, \alpha\}^*$. In the second case, when $w \in \{a, \alpha\}^*$, the morphism φ_w has two fixed points.

4.1. Morphisms with unique fixed point

Let $\psi \in \langle \varphi_a, \varphi_b, \varphi_\alpha, \varphi_\beta \rangle$ and $N(\psi) = w \in \{a, b, \alpha, \beta\}^* \setminus \{a, \alpha\}^*$ be the normalized name of the morphism ψ. By Lemma 6 the word w has a prefix $a^k\beta$ or $\alpha^k b$ for some $k \in \mathbb{N}$. This property enables us to define a transformation on the set of morphisms from $\mathcal{M} \setminus \langle \varphi_a, \varphi_\alpha \rangle$. As we will demonstrate later, this transformation is in fact the desired algorithm returning the morphisms $\psi_1, \psi_2, \ldots, \psi_\ell$ mentioned above.

Definition 22. Let $w \in \{a, b, \alpha, \beta\}^* \setminus \{a, \alpha\}^*$ be the normalized name of a morphism ψ, i.e., $\psi = \varphi_w$. We put

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

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

Example 23. Consider the morphism $\psi = \varphi_w$, where $w = \beta\alpha aa\alpha$, and apply repeatedly the transformation Δ on ψ.

$$\begin{aligned}
\psi &= \varphi_{\beta\alpha aa\alpha} && \text{and} \quad N(\psi) = w = \beta\alpha aa\alpha \\
\Delta(\psi) &= \varphi_{\alpha aa\alpha\beta} && \text{and} \quad N\big(\Delta(\psi)\big) = \beta bb\alpha\alpha \\
\Delta^2(\psi) &= \varphi_{bb\alpha\alpha\beta} && \text{and} \quad N\big(\Delta^2(\psi)\big) = bb\beta\alpha\alpha \\
\Delta^3(\psi) &= \varphi_{b\beta\alpha\alpha b} && \text{and} \quad N\big(\Delta^3(\psi)\big) = b\beta\alpha\alpha b \\
\Delta^4(\psi) &= \varphi_{\beta\alpha\alpha bb} && \text{and} \quad N\big(\Delta^4(\psi)\big) = \beta\alpha\alpha bb \\
\Delta^5(\psi) &= \varphi_{\alpha\alpha bb\beta} && \text{and} \quad N\big(\Delta^5(\psi)\big) = \alpha\alpha bb\beta \\
\Delta^6(\psi) &= \Delta^3(\psi)
\end{aligned}$$

In what follows we prove that the five fixed points of morphisms $\Delta(\psi), \Delta^2(\psi), \Delta^3(\psi), \Delta^4(\psi), \Delta^5(\psi)$ are exactly the five derivated words of the fixed point of ψ.

Lemma 24. *Let* $\mathbf{u}$ *be a fixed point of a morphism* ψ *and* $N(\psi) = w \in \{a, b, \alpha, \beta\}^*$ *be the normalized name of the morphism* ψ*. If one of the following condition is satisfied*

(i) $\mathbf{u}$ *starts with* 0 *and* w *starts with* a*,*
(ii) $\mathbf{u}$ *starts with* 1 *and* w *starts with* α*,*

then $w \in \{a, \alpha\}^*$.

Proof. We consider only the case (i), the case (ii) is analogous. Let us assume $w \notin \{a, \alpha\}^*$. According to Lemma 6, the word w has a prefix $a^k\beta$, for some $k \geq 1$. Consequently, the morphism ψ equals $\varphi_a^k \circ \varphi_\beta \circ \eta$ for some morphism η. Any morphism of this form maps 0 to $1w_1$ and 1 to $1w_2$ for some words w_1 and w_2. Therefore, the fixed point starts with the letter 1, which is a contradiction. □

The following theorem along with Definition 22 provides the algorithm which to a given Sturmian morphism ψ lists the morphisms fixing the derivated words of the Sturmian word $\mathbf{u} = \psi(\mathbf{u})$.

Theorem 25. *Let* $\psi \in \langle \varphi_a, \varphi_b, \varphi_\alpha, \varphi_\beta \rangle$ *be a primitive morphism and* $N(\psi) = w \in \{a, b, \alpha, \beta\}^* \setminus \{a, \alpha\}^*$ *be its normalized name. Denote* $\mathbf{u}$ *the fixed point of* ψ*. Then* $\mathbf{x}$ *is (up to a permutation of letters) a derivated word of* $\mathbf{u}$ *with respect to one of its prefixes if and only if* $\mathbf{x}$ *is the fixed point of the morphism* $\Delta^j(\psi)$ *for some* $j \geq 1$*.*

Proof. Denote $\mathbf{x}_j$ the fixed point of $\Delta^j(\psi)$, $j = 1, 2, \dots$ and assume that v is a right special prefix of $\mathbf{u}$. We will prove that if $|v| = 1$, then $\mathrm{d}_{\mathbf{u}}(v) = \mathbf{x}_1$, and if $|v| > 1$, then there is a right special prefix v' of $\mathbf{x}_1$ such that $|v'| < |v|$ and $\mathrm{d}_{\mathbf{u}}(v) = \mathrm{d}_{\mathbf{x}_1}(v')$. We can repeat this proof for the prefix v' of $\mathbf{x}_1$ and eventually prove that $\mathrm{d}_{\mathbf{u}}(v) = \mathbf{x}_j$ for some j and that for any j there is a right special prefix v of $\mathbf{u}$ so that $\mathrm{d}_{\mathbf{u}}(v) = \mathbf{x}_j$.

Without loss of generality we assume that the normalized name of ψ is $w = a^k\beta z$. This means that $\Delta(\psi) = \varphi_z \circ \varphi_{a^k\beta}$.

First we assume $|v| = 1$. If $k > 0$, then the first letter of $\mathbf{u}$ is 1 which is not a right special factor. This implies that $k = 0$. Hence we have that $\mathbf{u} = \varphi_\beta(\mathbf{u}')$, where $\mathbf{u}' = \varphi_z(\mathbf{u})$. By Item (i) of Proposition 16 we obtain $\mathrm{d}_{\mathbf{u}}(v) = \mathbf{u}'$. Lemma 21 says the word $\mathbf{u}'$ is fixed by the morphism $\varphi_z \circ \varphi_\beta = \Delta(\psi)$, which implies $\mathbf{u}' = \mathbf{x}_1$.

Now assume $|v| > 1$. If $k = 0$, then by Item (ii) of Proposition 16 there is a right special prefix v' of $\mathbf{u}' = \varphi_z(\mathbf{u})$ such that $|v'| < |v|$ and $\mathrm{d}_{\mathbf{u}}(v) = \mathrm{d}_{\mathbf{u}'}(v')$. Again by Lemma 21 we obtain $\mathbf{u}' = \mathbf{x}_1$.

Let $k > 0$. For $i = 0, 1, \dots, k$ we define $\mathbf{u}^{(i)} = \varphi_{a^{k-i}\beta z}(\mathbf{u})$. By Lemma 24, the words $\mathbf{u}^{(i)}$ all start with the letter 1. Obviously, $\mathbf{u}^{(0)} = \mathbf{u}$ and $\mathbf{u}^{(i)} = \varphi_a\big(\mathbf{u}^{(i+1)}\big)$ for $i = 0, 1, \dots, k-1$. By Theorem 18, there are factors $v^{(i)}$ with $i = 0, 1, \dots, k$ such that $v^{(i)}$ is a right special prefix of $\mathbf{u}^{(i)}$,

$$|v| = |v^{(0)}| \geq |v^{(1)}| \geq |v^{(2)}| \geq \cdots \geq |v^{(k)}|$$

and

$$\mathrm{d}_{\mathbf{u}}(v) = \mathrm{d}_{\mathbf{u}^{(1)}}(v^{(1)}) = \mathrm{d}_{\mathbf{u}^{(2)}}(v^{(2)}) \cdots = \mathrm{d}_{\mathbf{u}^{(k)}}(v^{(k)})\,.$$

Define $\mathbf{u}' = \varphi_z(\mathbf{u})$. Then $\mathbf{u}^{(k)} = \varphi_{\beta z}(\mathbf{u}) = \varphi_\beta(\mathbf{u}')$ and by Item (ii) of Proposition 16 there is a right special prefix v' of $\mathbf{u}' = \varphi_z(\mathbf{u})$ such that $|v'| < |v^{(k)}|$ and $\mathrm{d}_{\mathbf{u}^{(k)}}(v^{(k)}) = \mathrm{d}_{\mathbf{u}'}(v')$. According to Lemma 21, the word $\mathbf{u}'$ is fixed by the morphism $\varphi_z \circ \varphi_{a^k\beta} = \Delta(\psi)$. Thus, we have again proved that there is a prefix v' of $\mathbf{u}' = \mathbf{x}_1$ such that $|v'| < |v|$ and $\mathrm{d}_{\mathbf{u}}(v) = \mathrm{d}_{\mathbf{u}'}(v')$. □

Remark 26. In Example 23 we considered the morphism $\psi = \varphi_w$, where $w = \beta\alpha a a \alpha$. We have found only five different morphisms $\Delta^i(\psi)$ for $i = 1, \dots, 5$. The sixth morphism $\Delta^6(\psi)$ already coincides with $\Delta^3(\psi)$. As it follows from the proofs of Theorems 14 and 25, the fixed points of $\Delta^3(\psi)$, $\Delta^4(\psi)$ and $\Delta^5(\psi)$ represent the derivated words of $\mathbf{u}$ to infinitely many prefixes of $\mathbf{u}$. Whereas the fixed point of $\Delta(\psi)$ or $\Delta^2(\psi)$ is a derivated word of $\mathbf{u}$ to only one prefix of $\mathbf{u}$.

Example 27. As explained in Example 8, to find the derivated words of the Fibonacci word we consider the morphism $\psi = \tau^2 = \varphi_b\varphi_\beta$. We have $\Delta(\psi) = \varphi_\beta\varphi_b$ and $\Delta^2(\psi) = \psi$. But these two morphisms are equal up to a permutation of letters, as $E\psi E = \Delta(\psi)$. This means that all derivated words of the Fibonacci word are the same and coincide with the Fibonacci word itself.

4.2. Morphisms with two fixed points

Let us now consider a Sturmian morphism ψ which has two fixed points. Let us denote $\mathbf{u}^{(0)}$ and $\mathbf{u}^{(1)}$ the fixed points of ψ starting with 0 and 1, respectively. Clearly, $\psi(0)$ starts with 0 and $\psi(1)$ with 1. Since the morphism ψ has to belong to the monoid $\langle \varphi_a, \varphi_\alpha \rangle$, the transformation Δ cannot be applied on it. However, we will show that there is a morphism from $\langle \varphi_a, \varphi_\beta \rangle$ (or $\langle \varphi_b, \varphi_\alpha \rangle$) with a unique fixed point $\mathbf{v}$ such that the set of derivated words of $\mathbf{u}^{(0)}$ (or $\mathbf{u}^{(1)}$) equals to

$\{\mathbf{v}\} \cup \mathrm{Der}(\mathbf{v})$. And since $\mathbf{v}$ is a fixed point of some morphism from $\langle \varphi_a, \varphi_b, \varphi_\beta, \varphi_\alpha \rangle \setminus \langle \varphi_a, \varphi_\alpha \rangle$, the set $\mathrm{Der}(\mathbf{v})$ can be described using Theorem 25.

Here we give results only for the case when the normalized name $w \in \{a, \alpha\}^*$ of the morphism begins with a. The case when the first letter is α is completely analogous. It suffices to exchange $a \leftrightarrow b$ and $\alpha \leftrightarrow \beta$ in the statements and proofs.

Lemma 28. *Let $w \in \{a, \alpha\}^*$ be the normalized name of a morphism starting with the letter a. Then the normalized name $N(wb)$ has a prefix b and a suffix a, the word $v = b^{-1}N(wb)$ belongs to $\{a, \beta\}^*$, and $|v|_\beta = |w|_\alpha$.*

Proof. First, we consider the special case when $w = a^k\alpha^\ell$, with $k \geq 1$ and $\ell \geq 0$. By the relation (2), $N(wb) = ba^{k-1}\beta^\ell a$ and the statement is true.

Let $w \in \{a, \alpha\}^*$ be arbitrary. It can be decomposed to several blocks of the form $a^k\alpha^\ell$ with $k \geq 1$, $\ell \geq 0$. Now the proof can be easily finished by induction on the number of these blocks. □

Proposition 29. *Let $w \in \{a, \alpha\}^*$ be the normalized name of a primitive morphism ψ and let a be its first letter.*

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

Proof. Let us start with proving (i). Let $\mathbf{v}$ be the infinite word given by Lemma 19. Then

$$\varphi_b(\mathbf{v}) = \mathbf{u} = \psi(\mathbf{u}) = \varphi_w(\mathbf{u}) = \big(\varphi_w \circ \varphi_b\big)(\mathbf{v}) = \varphi_{wb}(\mathbf{v}) = \varphi_{N(wb)}(\mathbf{v})\,.$$

By definition of v we have $N(wb) = bv$ and thus

$$\varphi_b(\mathbf{v}) = \varphi_{bv}(\mathbf{v}) = \varphi_b\big(\varphi_v(\mathbf{v})\big).$$

This implies that $\mathbf{v} = \varphi_v(\mathbf{v})$. Since $v \notin \{a, \alpha\}^*$, the morphism φ_v has a unique fixed point, namely the word $\mathbf{v}$. By Theorem 14, $\mathrm{Der}(\mathbf{u}) = \{\mathbf{v}\} \cup \mathrm{Der}(\mathbf{v})$ as stated in (i).

Statement (ii) is a direct consequence of Theorem 18 and Lemma 21. □

5. Bounds on the number of derivated words

In this section we study the relation between the normalized name w of a primitive morphism $\psi = \varphi_w$ and the number of distinct return words to its fixed point. We restrict ourselves to the case when $w \notin \{a, \alpha\}^*$, as the case $w \in \{a, \alpha\}^*$ is treated in the next section.

Theorem 25 says that the number of derivated words of $\mathbf{u}$ cannot exceed the upper bound:

$$\text{number of distinct words in the sequence } \Big(\Delta^k(w)\Big)_{k\geq 1}\,.$$

Since the words $\Delta^k(w) \in \{a, b, \alpha, \beta\}^*$ are all of the same length and $\Delta^{k+1}(w)$ is completely determined by $\Delta^k(w)$, the sequence $\big(\Delta^k(w)\big)_{k\geq 1}$ is eventually periodic.

The number of distinct elements in $\big(\Delta^k(w)\big)_{k\geq 1}$ is only an upper bound on the number of derivated words of $\mathbf{u}$. As we have already mentioned in Remark 9, fixed points of morphisms corresponding to the names v and $F(v)$ coincide up to exchange of letters 0 and 1 and hence define the same derivated word. On the other hand, if v and v' are normalized names with $|v| = |v'|$ and fixed points of φ_v and $\varphi_{v'}$ coincide (up to exchange of letters), then either $v' = v$ or $v' = F(v)$.

First we look at two examples that illustrate some special cases of the general Proposition 32 on the period and preperiod of the sequence $\big(\Delta^k(w)\big)_{k\geq 1}$.

Example 30. Consider a word w of length n in the form $w = b^{n-2}\beta a$. The sequence of $\big(\Delta^k(w)\big)_{k\geq 1}$ is eventually periodic. Its preperiod equals $n-2$ and is given by the words $b^{n-k}\beta b^{k-2}a$, for $k = 3, 4, \ldots, n$. The period equals $n-1$ and is given by the words $b^{n-k}a\beta b^{k-2}$, for $k = 2, 3, \ldots, n$.

Let us stress that for any $v \in \{a, b, \alpha, \beta\}^*$ the equation $v' = F(v)$ implies $|v|_a = |v'|_\alpha$ and $|v|_b = |v'|_\beta$. Since all words $\Delta^k(w)$ we listed above contain one letter a and no letter α, we can conclude that the morphism φ_w has $2n-3$ distinct derivated words.

Example 31. Consider a normalized name w in which the letter b is missing and w contains all the three remaining letters. Necessarily w has the form

$$\beta^{\ell_1}a^{k_1}\beta^{\ell_2}a^{k_2}\cdots\beta^{\ell_s}a^{k_s}\alpha^j,$$

where $s \geq 1$, $\ell_i \geq 1$ for all $i = 2, \ldots, s$ and $k_i \geq 1$ for all $i = 1, 2, \ldots, s-1$ and $j \geq 1$. It is easy to see that the normalized names of words obtained by repeated application of the mapping Δ are

$$\Delta^{\ell_1}(w) = a^{k_1}\beta^{\ell_2}a^{k_2}\cdots\beta^{\ell_s}a^{k_s}\beta^{\ell_1}\alpha^j \qquad \text{and} \qquad \Delta^{\ell_1+1}(w) = \beta^{\ell_2-1}a^{k_2}\cdots\beta^{\ell_s}a^{k_s}\beta^{\ell_1+1}\alpha^{j-1}b^{k_1}\alpha$$

We see that the $(\ell_1+1)^{st}$ iteration already contains all four letters.

Proposition 32. *Let $w \in \{a, b, \alpha, \beta\}^* \setminus \{a, \alpha\}^*$ be the normalized name of a primitive Sturmian morphism $\psi = \varphi_w$. Then the sequence $\left(\Delta^k(w)\right)_{k\geq 1}$ is eventually periodic and:*

(i) If it is purely periodic, then its period is at most $|w|$, otherwise, its period is at most $|w| - 1$.
(ii) If both b and β occur in w, then the preperiod is at most $|w| - 2$, otherwise the preperiod is at most $2|w| - 3$.

Proof. By Lemma 6, the word w (and all the elements of the sequence $\left(\Delta^k(w)\right)_{k\geq 1}$) has the form $w = a^i\beta w'$ or $w = \alpha^i b w'$ for some $i \geq 0$. In this proof we distinguish three cases such that exactly one of them is valid for all $\Delta^k(w), k = 1, 2, \ldots$ The first two cases correspond to the "periodic" part of the sequence $\left(\Delta^k(w)\right)_{k\geq 1}$.

Case 1: If w has a suffix β or b, then the word $\Delta(w)$ equals to $w'a^i\beta$ or $w'\alpha^i b$ and thus has again a suffix β or b. Indeed, since $N(w) = w$, the words $\alpha a^j\beta$ and $a\alpha^j b$ are not factors of w and so they are not even factors of w'. As the last letter of w' is b or β, neither $\alpha a^j\beta$ nor $a\alpha^j b$ is a factor of $w'a^i\beta$ and hence $\Delta(w) = N(w'a^i\beta) = w'a^i\beta$. This means that for any k the word $\Delta^k(w)$ is just a cyclic shift of w (see (1)). Therefore, $\left(\Delta^k(w)\right)_{k\geq 1}$ is purely periodic and its period is given by the number of letters β and b in w which is clearly at most $|w|$. Moreover, the word w belongs to the sequence $\left(\Delta^k(w)\right)_{k\geq 1}$ and the fixed point $\mathbf{u}$ of ψ itself is a derivated word of $\mathbf{u}$.

Without loss of generality we assume that $w = a^i\beta w'$; the case of $w = \alpha^i b w'$ can be treated in the same way, it suffices to exchange letters $a \leftrightarrow b$ and $\alpha \leftrightarrow \beta$. Denote p the longest suffix of w such that $p \in \{a, \alpha\}^*$. It remains to consider only the case of nonempty p.

Case 2: If $p = a^j$ for some $j \geq 1$, then w' has a suffix ba^j or βa^j. No rewriting rule from (2) can be applied to $w'a^i\beta$, hence, $\Delta(w) = w'a^i\beta$ has a suffix β. So, we can apply the reasoning from Case 1 on the word $\Delta(w)$ and hence the sequence $\left(\Delta^k(w)\right)_{k\geq 1}$ is purely periodic. As w contains at least one letter a as a suffix, the period is shorter than $|w|$ and w itself does not occur in $\left(\Delta^k(w)\right)_{k\geq 1}$.

Case 3: Now assume that the letter α occurs in p. We split this case into three subcases and show that if one of these subcases is valid for a word $\Delta^k(w)$, then this word belongs to the "preperiodic" part of $\left(\Delta^k(w)\right)_{k\geq 1}$. These three subcases (for word w) read:
(i) w begins with the letter a, i.e., $i \geq 1$;
(ii) w has a prefix β and p has a factor αa;
(iii) w has a prefix β and $p = a^j\alpha^s$ for $j \geq 0$ and $s \geq 1$.

(i) Since we assume that α occurs in p, a suffix of p has a form αa^t for some $t \geq 0$. It follows that $w'a^i\beta$, has a suffix $\alpha a^{t+i}\beta$. After applying the rewriting rules (2) to $w'a^i\beta$ we obtain the normalized name $\Delta(w)$ which has a suffix $b\alpha$.
(ii) A suffix of w can be expressed in the form $\alpha a^r\alpha^s a^t$, where $r \geq 1$ and $s, t \geq 0$. Therefore $w'\beta$ has a suffix $\alpha a^r\alpha^s a^t\beta$. After normalization we get that $\Delta(w)$ has a suffix in the form of $b\alpha^\ell$ for some $\ell \geq 1$.
(iii) As $w = \beta w'$ has a suffix $\beta a^j\alpha^s$ or $ba^j\alpha^s$, the word $N(w'\beta)$ has a suffix $\beta\alpha^s$.

All the three discussed subcases share the following property: The longest suffix $p' \in \{a, \alpha\}^*$ of the normalized name $v = \Delta(w)$ is of the form $p' = \alpha^m$, for some $m \geq 1$. It means that Case 3 (ii) is not applicable in the second iteration of Δ. By Lemma 6, the word v has a prefix $a^n\beta$ or $\alpha^n b, n \geq 0$.
If the prefix of v is of the form $\alpha^n b$, then the word $\Delta(v) = \Delta^2(w)$ belongs to Case 2. This means that v is the last member of the preperiodic part.
If the prefix of v is of the form $a^n\beta$, then we must apply either Case 3 (i) or 3 (iii) which means that α is again a suffix of the word obtained in the next iteration of Δ.
Let us give a bound on the number of times that we have to use Case 3 (i) or 3 (iii) before we reach Case 2.
If w contains both β and b, then the number of times of using Case 3 (i) or 3 (iii) is at most the number of letters β occurring in w before the first occurrence of b. Thus there are at most $|w| - 2$ such letters since w contains β, b and α.
If w does not contain b, then w must contain besides the letters β and α also the letter a; otherwise the morphism φ_w would be not primitive (see Lemma 4). The word w has a form described in Example 31 and thus $\Delta^{\ell_1+1}(w)$ contains both letter b and β (for the meaning of ℓ_1 see Example 31). For this word we can apply the reasoning from the previous paragraph, meaning that after $\ell_1 + 1$ iterations we need at most $|w| - 2$ further iterations before reaching the periodic part of $\left(\Delta^k(w)\right)_{k\geq 1}$. Since $\ell_1 \leq |w| - 2$, we get that the preperiod is at most $2|w| - 3$. $\square$

Example 30 illustrates that in the case that w contains the letters b and β the upper bounds on preperiod and period provided by the previous proposition are attained. The following example proves that the bound from Proposition 32 for w which does not contain both letters b and β is attained as well.

Example 33. Let us consider the normalized name $w = \beta^{n-2}a\alpha$. It is easy to evaluate iterations of the operator Δ:

$$
\begin{aligned}
\Delta^{n-2}(w) &= a\beta^{n-2}\alpha \\
\Delta^{n-1}(w) &= \beta^{n-2}b\alpha \\
\Delta^{2n-3}(w) &= b\beta^{n-2}\alpha \\
\Delta^{2n-2}(w) &= \beta^{n-2}\alpha b \quad \text{— the first member of the periodic part of } \big(\Delta^k(w)\big)_{k\geq 1} \\
\Delta^{3n-4}(w) &= \alpha b\beta^{n-2} \quad \text{— the last member of the periodic part of } \big(\Delta^k(w)\big)_{k\geq 1} \\
\Delta^{3n-3}(w) &= \Delta^{2n-2}(w).
\end{aligned}
$$

In Example 27 we showed that for the Fibonacci word the derivated words to all prefixes coincide. There are infinitely many words with this property:

Example 34. Consider $w = a^{n-1}\beta$ and the morphism $\psi = \varphi_w$. Then $\Delta(\psi) = \psi$ and thus the fixed point $\mathbf{u}$ of ψ is the derivated word to any prefix of $\mathbf{u}$.

Combining Proposition 32 and the last two examples we can give an upper and lower bound on the number of distinct derivated words.

Corollary 35. *Let $w \in \{a, b, \alpha, \beta\}^* \setminus \{a, \alpha\}^*$ be normalized name of a primitive Sturmian morphism $\psi = \varphi_w$ and $\mathbf{u}$ be a fixed point of ψ. Then*

$$1 \leq \#\mathrm{Der}(\mathbf{u}) \leq 3|w| - 4\,. \tag{6}$$

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

(i) $\varphi_{w'}$ and $\varphi_{w''}$ are not powers of other Sturmian morphisms,
(ii) for the fixed points $\mathbf{u}'$ and $\mathbf{u}''$ of the morphism $\varphi_{w'}$ and $\varphi_{w''}$, the lower resp. the upper bound in (6) is attained.

6. Standard Sturmian morphisms and their reversals

In this section we provide precise numbers of distinct derivated words for these three types of morphisms:

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

First we explain the title of this section and the fact that the fourth type of Sturmian morphism, namely a Sturmian morphism from $\langle \varphi_a, \varphi_\alpha \rangle \circ E$, is not considered at all.

A standard Sturmian morphism is a morphism fixing some standard Sturmian word. A reversal morphism $\overline{\psi}$ to a morphism ψ is defined by $\overline{\psi}(0) = \overline{\psi(0)}$ and $\overline{\psi}(1) = \overline{\psi(1)}$. Since $\varphi_a = \overline{\varphi_b}$ and $\varphi_\alpha = \overline{\varphi_\beta}$, any morphism in $\langle \varphi_a, \varphi_\alpha \rangle$ is just a reversal of a morphism in $\langle \varphi_b, \varphi_\beta \rangle$.

Due to the form of the morphisms φ_a and φ_α, any morphism $\eta \in \langle \varphi_a, \varphi_\alpha \rangle$ satisfies that the letter 0 is a prefix of $\eta(0)$ and the letter 1 is a prefix of $\eta(1)$. As any morphism $\xi \in \langle \varphi_a, \varphi_\alpha \rangle \circ E$ can be written in the form $\xi(0) = \eta(1)$ and $\xi(1) = \eta(0)$ for some $\eta \in \langle \varphi_a, \varphi_\alpha \rangle$, the morphism ξ cannot have any fixed point.

The normalized name w of a standard morphism from $\mathcal{M}$ is composed of the letters b and β only. Thus $\Delta(w) = \mathrm{cyc}(w)$ (see (1)).

To describe all standard morphisms we have to take into account also the morphisms of the form $\psi = \varphi_w \circ E$. In this case $\psi^2 \in \langle \varphi_b, \varphi_\beta \rangle$, in particular $\psi^2 = \varphi_{wF(w)}$. To describe the derivated words of fixed points of these standard morphisms, we need the notation

$$\mathrm{cyc_F}(w_1 w_2 w_3 \cdots w_n) = w_2 w_3 \cdots w_n F(w_1)\,.$$

Proposition 36. *Let $\mathbf{u}$ be a fixed point of a standard Sturmian morphism ψ which is not a power of any other Sturmian morphism.*

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

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

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

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

Proof. (i) Since $\psi = \varphi_w$ is a standard morphism, its normalized name w belongs to $\{b, \beta\}^*$ and $\Delta(w) = \mathrm{cyc}(w)$. By Theorem 25, all derivated words of **u** are fixed by one of the morphisms listed in (i). We only need to show that fixed points of the listed morphisms differ. More precisely, we need to show that $v_s \neq v_t$ and $v_s \neq F(v_t)$ for all $0 \leq t < s \leq |w| - 1$. Here the assumption that ψ is not a power of any other Sturmian morphism is crucial.

Let us recall simple facts about powers of morphisms: For any $\ell = 1, 2, \dots$ and $u \in \{b, \beta\}^+$ we have

$$(\varphi_u)^\ell = \varphi_{u^\ell}, \quad (\varphi_u \circ E)^{2\ell} = \varphi_{(uF(u))^\ell} \quad \text{and} \quad (\varphi_u \circ E)^{2\ell+1} = \varphi_{(uF(u))^\ell u} \circ E.$$

If $\psi = \varphi_w$ is not a power of any Sturmian morphism, we have

$$w \neq u^\ell \quad \text{and} \quad w \neq \big(uF(u)\big)^k \quad \text{for any } u \in \{b, \beta\}^+ \text{ and any } \ell, k \in \mathbb{N}, \ell \geq 2, k \geq 1. \tag{7}$$

Lemma 11 implies that equation $\mathrm{cyc}^s(w) = \mathrm{cyc}^t(w)$ has no solution if $w \neq u^\ell$ and $0 \leq t < s \leq |w| - 1$. Therefore all the normalized names $v_0, v_1, \dots, v_{|w|-1}$ are distinct.

Now assume that $v_s = \mathrm{cyc}^s(w) = F\big(\mathrm{cyc}^t(w)\big) = F(v_t)$, where $0 \leq t < s \leq |w| - 1$.

Let z and p be the words such that $\mathrm{cyc}^s(w) = zp$, where $|z| = s - t$. We have $zp = F(p)F(z)$ and by Lemma 12 there is x such that $\mathrm{cyc}^s(w) = zp = x(F(x)x)^i(F(x)x)^j F(x) = (xF(x))^{i+j+1}$ for some non-negative integers i, j. This implies that there is a factor y of $xF(x)$ such that $|y| = |x|$ and $w = (yF(y))^{i+j+1}$ which is a contradiction with (7).

(ii) If we apply Theorem 25 to the morphism $\big(\varphi_w \circ E\big)^2 = \varphi_{wF(w)}$, we obtain the list of $2|w|$ normalized names $\mathrm{cyc}^s(wF(w))$, with $s = 0, 1, \dots, 2|w| - 1$. As $\mathrm{cyc}^{|w|+i}(wF(w)) = \mathrm{cyc}^i(F(w)w)$, all the derivated words are given by the fixed points of morphisms

$$\varphi_{v_0F(v_0)}, \varphi_{v_1F(v_1)}, \varphi_{v_2F(v_3)}, \dots, \varphi_{v_{|w|-1}F(v_{|w|-1})}$$

that are just squares of morphisms listed in Item (ii) of the proposition. To finish the proof, we need to show that the fixed points of the listed morphisms do not coincide nor coincide after exchange of the letters $0 \leftrightarrow 1$. In other words we need to show $v_sF(v_s) \neq v_tF(v_t)$ and $v_sF(v_s) \neq F(v_t)v_t$.

Assume the contrary. Then $v_s = v_t$ or $v_s = F(v_t)$ for some $t < s$. If we put $k = s - t$, then $v_s = \mathrm{cyc}_F^k(v_t)$. Let $v_t = zp$, where $|z| = k$, then $v_s = pF(z)$. Since the morphism $\psi = \varphi_w \circ E$ is not a power of other morphism we know that

$$w \neq \big(uF(u)\big)^\ell u \quad \text{for any } u \in \{b, \beta\}^+ \text{ and any } \ell \in \mathbb{N}, \ell \geq 1. \tag{8}$$

Two cases $v_s = v_t$ and $v_s = F(v_t)$ will be discussed separately.

- If $v_s = v_t$, then $zp = pF(z)$ and Lemma 12 says there is x so that $v_t = zp = (F(x)x)^{i+j}F(x)$, which contradicts (8).
- If $v_s = F(v_t)$, then $zp = F(p)z$ and by Lemma 12 there is x so that $v_t = zp = (F(x)x)^{i+j}F(x)$ which is again a contradiction with (8). □

Proposition 37. *Let $w \in \{\alpha, a\}^*$ be the normalized name of a primitive morphism ψ such that the letter a is a prefix of w. Moreover, assume that ψ is not a power of any other Sturmian morphism.*

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

Proof. We prove only Item (*i*), the proof of (*ii*) is analogous. Let **u** denote the fixed point starting with 0.

Proposition 29 says that we have to count elements in the set $\{\mathbf{v}\} \cup \mathrm{Der}(\mathbf{v})$, where $\mathbf{v}$ is a fixed point of φ_v with the normalized name $v = b^{-1}N(wb)$. By Lemma 28, the word $v \in \{a, \beta\}^*$. This property of v implies that $\Delta^k(v)$ is equal to some cyclic shift $\mathrm{cyc}^j(v)$ having a suffix β. There are $|v|_\beta$ cyclic shifts of v with this property and hence this number is an upper bound for the period of the sequence $\big(\Delta^k(v)\big)_{k\geq 1}$. By Lemma 28, the normalized name v has a suffix a and thus the word v itself does not appear in $\big(\Delta^k(v)\big)_{k\geq 1}$. We can conclude that **u** has at most $1 + |w|_\alpha$ derivated words.

For each k the iteration $\Delta^k(v)$ belongs to $\{a, \beta\}^*$ and consequently $F\big(\Delta^i(v)\big)$ belongs to $\{a, b\}^*$. Therefore, $\Delta^j(v) \neq F\big(\Delta^i(v)\big)$ for any pair of positive integers i, j.

As ψ is not a power of any other morphism, we can use the same technique as in the proof of Proposition 36 to show that $\mathrm{cyc}^i(v) \neq \mathrm{cyc}^j(v)$ for $i, j = 1, \ldots, |v|$, $i \neq j$. This means that the period of the sequence $\left(\Delta^k(v)\right)_{k\geq 1}$ is indeed equal to $|w|_\alpha$ and its preperiod is zero. $\square$

7. Comments and conclusions

1. In [1], the authors studied derivated words only for standard Sturmian words $\mathbf{c}(\gamma)$.
   However, they did not restrict their study to words fixed by a primitive morphism.
   Let us show an alternative proof of their result.
   The proof is a direct corollary of our Theorem 14 and the following result of [5]:

   Lemma 38 *([5, Lemma 2.2.18]). For any irrational $\gamma \in (0, 1)$ we have*

   $$\varphi_b(\mathbf{c}(\gamma)) = \mathbf{c}\left(\frac{\gamma}{1+\gamma}\right).$$

   As we have already mentioned, the authors of [1] required that any derivated word $\mathrm{d}_{\mathbf{u}}(v)$ to a prefix v of a Sturmian word $\mathbf{u}$ starts with the same letter as the word $\mathbf{u}$.
   By interchanging letters $0 \leftrightarrow 1$ in a characteristic word $\mathbf{c}(\gamma)$, we obtain the characteristic word $\mathbf{c}(1-\gamma)$. If $\gamma < \frac{1}{2}$, then the continued fraction of γ is of the form $[0, c_1+1, c_2, c_3, \ldots]$ with $c_1 > 0$ and the continued fraction of $1-\gamma$ equals $[0, 1, c_1, c_2, c_3, \ldots]$. Clearly, $\mathrm{Der}(\mathbf{c}(\gamma))$ and $\mathrm{Der}(\mathbf{c}(1-\gamma))$ coincide up to a permutation of letters. Without loss of generality we state the next theorem for the slope $\gamma < \frac{1}{2}$ only.

   Theorem 39 *([1]). Let $\mathbf{c}(\gamma)$ be a standard Sturmian word and $\gamma = [0, c_1+1, c_2, c_3, \ldots]$ with $c_1 > 0$. Then*

   $$\mathrm{Der}(\mathbf{c}(\gamma)) = \left\{\mathbf{c}(\delta) : \delta = [0, c_k + 1 - i, c_{k+1}, c_{k+2}, \ldots] \text{ with } 0 \leq i \leq c_k - 1 \text{ and } (k, i) \neq (1, 0)\right\}.$$

   Proof. Let $\delta = [0, d_1+1, d_2, d_3, \ldots]$ with $d_1 > 0$. Set $\delta' = \frac{\delta}{1-\delta}$. It is easy to see that $\delta' = [0, d_1, d_2, d_3, \ldots]$. Since $\delta' \in (0, 1)$ and $\delta = \frac{\delta'}{1+\delta'}$, Lemma 38 implies that $\mathbf{c}(\delta) = \varphi_b(\mathbf{c}(\delta'))$. Applying Theorem 14 we obtain that $\mathrm{Der}(\mathbf{c}(\delta)) = \{\mathbf{c}(\delta')\} \cup \mathrm{Der}(\mathbf{c}(\delta'))$. We have transformed the original task to the task to determine the set of derivated words of the standard sequence $\mathbf{c}(\delta')$. If $\delta' < \frac{1}{2}$, i.e., $d_1 > 1$, we repeat this procedure with δ'. If $d_1 = 1$, i.e., $\delta' > \frac{1}{2}$, we use the fact that $\mathrm{Der}(\mathbf{c}(\delta))$ and $\mathrm{Der}(\mathbf{c}(1-\delta))$ coincide, and replace δ' by $1-\delta'$ and repeat the procedure with its continued fraction $[0, d_2+1, d_3, d_4, \ldots]$.
   In the terms of corresponding continued fractions, one step of the described procedure can be represented as

   $$[0, d_1+1, d_2, d_3, \ldots] \mapsto \begin{cases} [0, d_1, d_2, d_3, \ldots] & \text{if } d_1 > 1, \\ [0, d_2+1, d_3, d_4, \ldots] & \text{if } d_1 = 1. \end{cases}$$

   We conclude that the set $\mathrm{Der}(\mathbf{c}(\gamma))$ is in the form given in the theorem. $\square$

2. In case that $\mathbf{u}$ is a fixed point of a standard Sturmian morphisms, we have determined the exact number of distinct derivated words of $\mathbf{u}$, see Proposition 36. Let us mention that this result can be inferred from [1]. We also have provided the exact number of derivated words when $\mathbf{u}$ is a fixed point of a Sturmian morphism which has two fixed points, see Proposition 37.
   For fixed points of other Sturmian morphisms we only gave an upper bound on the number of their distinct derivated words, see Corollary 35. To give an exact number, one needs to describe when the normalized name $w \in \{a, b, \alpha, \beta\}^*$ corresponds to some power of a Sturmian morphism. Clearly, w may be a normalized name of a power of a Sturmian morphism without w being a power of some other word from $\{a, b, \alpha, \beta\}^*$. For example, if $v = \alpha ba\alpha\alpha = N(v)$, then the normalized name of v^3 is the primitive word $N(v^3) = \alpha bb\beta\beta\beta ba\beta\beta\beta aa\alpha\alpha$.
3. The key tool we used to determine the set $\mathrm{Der}(\mathbf{u})$ is provided by Theorems 14 and 18. We believe that an analogue of these theorems can be found also for Arnoux–Rauzy words over multiliteral alphabet. For definition and properties of these words see [4,12].
   In [8], the authors described a new class of ternary sequences with complexity $2n+1$. These sequences are constructed from infinite products of two morphisms. The structure of their bispecial factors suggests that due to result of [3], any derivated word of such a word is over a ternary alphabet. Probably, even for these words an analogue of Theorems 14 and 18 can be proved. Other candidates for generalization of Theorems 14 and 18 seem to be the infinite words whose language forms tree sets as defined in [6].

Acknowledgements

We acknowledge financial support by the Ministry of Education, Youth and Sports of the Czech Republic project no. CZ.02.1.01/0.0/0.0/16_019/0000778. K.M., E.P. and Š.S. also acknowledge financial support by the Czech Science Foundation grant GAČR 13-03538S and K.M. acknowledges financial support by the Czech Technical University in Prague grant SGS17/193/OHK4/3T/14.

References

[1] I.M. Araújo, V. Bruyère, Words derivated from Sturmian words, Theoret. Comput. Sci. 340 (2005) 204–219.
[2] L. Balková, E. Pelantová, Š. Starosta, Sturmian jungle (or garden?) on multiliteral alphabets, RAIRO Theor. Inform. Appl. 44 (2010) 443–470.
[3] L. Balková, E. Pelantová, W. Steiner, Sequences with constant number of return words, Monatsh. Math. 155 (2008) 251–263.
[4] J. Berstel, Sturmian and episturmian words, in: S. Bozapalidis, G. Rahonis (Eds.), Algebraic Informatics, in: Lecture Notes in Computer Science, vol. 4728, Springer, Berlin Heidelberg, 2007, pp. 23–47.
[5] J. Berstel, P. Séébold, Sturmian words, in: M. Lothaire (Ed.), Algebraic Combinatorics on Words, Cambridge University Press, 2002, pp. 45–110.
[6] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, G. Rindone, Acyclic, connected and tree sets, Monatsh. Math. 176 (2015) 521–550.
[7] M. Bucci, A. De Luca, On a family of morphic images of Arnoux–Rauzy words, in: LATA '09: Proceedings of the 3rd International Conference on Language and Automata Theory and Applications, Springer-Verlag, 2009, pp. 259–266.
[8] J. Cassaigne, S. Labbé, J. Leroy, A set of sequences of complexity $2n + 1$, in: S. Brlek, F. Dolce, C. Reutenauer, É. Vandomme (Eds.), Combinatorics on Words, WORDS 2017, in: Lecture Notes in Computer Science, vol. 10432, Springer, Cham, 2017, pp. 144–156.
[9] J.D. Day, P. Fleischmann, F. Manea, D. Nowotka, Equations enforcing repetitions under permutations, in: S. Brlek, F. Dolce, C. Reutenauer, É. Vandomme (Eds.), Combinatorics on Words: 11th International Conference, WORDS 2017, Montréal, QC, Canada, September 11–15, 2017, Proceedings, Springer International Publishing, Cham, 2017, pp. 72–84.
[10] F. Durand, A characterization of substitutive sequences using return words, Discrete Math. 179 (1998) 89–101.
[11] N.P. Fogg, Substitutions in Arithmetics, Dynamics and Combinatorics, 1st ed., Lecture Notes in Mathematics, vol. 1794, Springer, 2002.
[12] A. Glen, J. Justin, Episturmian words: a survey, Theor. Inform. Appl. 43 (2009) 403–442.
[13] C. Holton, L.Q. Zamboni, Descendants of primitive substitutions, Theoret. Comput. Syst. 32 (1999) 133–157.
[14] J. Justin, G. Pirillo, Episturmian words and episturmian morphisms, Theoret. Comput. Sci. 276 (2002) 281–313.
[15] C. Kassel, C. Reutenauer, Sturmian morphisms, the braid group B_4, Christoffel words and bases of F_2, Ann. Mat. Pura Appl. 186 (2007) 317–339.
[16] M. Lothaire, Combinatorics on Words, Encyclopaedia of Mathematics and Its Applications, vol. 17, Addison-Wesley, Reading, Mass, 1983. Reprinted in the Cambridge Mathematical Library Cambridge University Press, Cambridge UK, 1997.
[17] R.C. Lyndon, M.P. Schützenberger, The equation $a^m = b^n c^p$ in free group, Michigan Math. J. 9 (1962) 289–298.
[18] F. Mignosi, P. Séébold, Morphismes sturmiens et règles de Rauzy, J. Théor. Nombres Bordeaux 5 (1993) 221–233.
[19] M. Morse, G.A. Hedlund, Symbolic dynamics II – Sturmian trajectories, Amer. J. Math. 62 (1940) 1–42.
[20] G. Richomme, Conjugacy and episturmian morphisms, Theoret. Comput. Sci. 302 (2003) 1–34.
[21] P. Séébold, Fibonacci morphisms and Sturmian words, Theoret. Comput. Sci. 88 (1991) 365–384.
[22] L. Vuillon, A characterization of Sturmian words by return words, European J. Combin. 22 (2001) 263–275.

Article B

Derived sequences of Arnoux–Rauzy sequences

Derived Sequences of Arnoux–Rauzy Sequences

Kateřina Medková(✉)

Department of Mathematics, Faculty of Nuclear Science and Physical Engineering, Czech Technical University in Prague, Prague, Czech Republic
medkokat@fjfi.cvut.cz

Abstract. For an Arnoux–Rauzy sequence $\mathbf{u}$ we describe the set $\mathrm{Der}(\mathbf{u})$ of derived sequences corresponding to all nonempty prefixes of $\mathbf{u}$ using the normalized directive sequence of $\mathbf{u}$. As a corollary, we show that all derived sequences of $\mathbf{u}$ are also Arnoux–Rauzy sequences. Moreover, if $\mathbf{u}$ is primitive substitutive, we precisely determine the cardinality of the set $\mathrm{Der}(\mathbf{u})$.

Keywords: Arnoux–Rauzy sequence · Derived sequence · Return word

1 Introduction

Derived sequences were introduced by Durand [4] to characterize the primitive substitutive sequences, i.e., the sequences which are morphic images of fixed points of primitive morphisms.

Let $\mathbf{u} = u_0u_1u_2\cdots$ be a recurrent sequence. An *occurrence* of the factor w in $\mathbf{u}$ is the index i such that w is a prefix of the sequence $u_iu_{i+1}u_{i+2}\cdots$. Let $i < j$ be two consecutive occurrences of w in $\mathbf{u}$. Then the word $u_iu_{i+1}\cdots u_{j-1}$ is a *return word* to w in $\mathbf{u}$. We take into consideration only the sequence $\mathbf{u}$ for which each factor w has finitely many return words, and we denote these return words by $r_0, r_1, \ldots, r_{k-1}$. Such a sequence is called *uniformly recurrent*. In addition, if w is a prefix of $\mathbf{u}$, then the sequence $\mathbf{u}$ can be written as the unique concatenation of the return words to w: $\mathbf{u} = r_{d_0}r_{d_1}r_{d_2}\cdots$ with all $d_i \in \{0, 1, \ldots, k-1\}$. The ordering of the return words in this concatenation is coded by the sequence $\mathbf{d_u}(w) = d_0d_1d_2\cdots$ which is called the *derived sequence* of $\mathbf{u}$ with respect to w.

Return words and derived sequences were especially studied in the case of *Sturmian sequences*, which are the aperiodic binary sequences having the least factor complexity possible. Every Sturmian sequence $\mathbf{u}$ has exactly one left and one right special factor per length. The factor w is left (right, respectively) special if the words aw, bw (wa, wb, respectively) are factors of $\mathbf{u}$ for two different letters

This work was supported by the project CZ.02.1.01/0.0/0.0/16_019/0000778 from European Regional Development Fund and by the grant No. SGS17/193/OHK4/3T/14 from the Grant Agency of the CTU in Prague.

R. Mercaş and D. Reidenbach (Eds.): WORDS 2019, LNCS 11682, pp. 251–263, 2019.
https://doi.org/10.1007/978-3-030-28796-2_20

a, b. Moreover, a Sturmian sequence is *standard* if all its prefixes are left special factors.

Vuillon [14] showed that a binary sequence is Sturmian if and only if each of its factors has exactly two return words. This property implies that the derived sequence with respect to each prefix of a Sturmian sequence is Sturmian as well. The derived sequences of standard Sturmian sequences were precisely described in [1], where the one-to-one correspondence between standard Sturmian sequences and continued fractions of irrational numbers from the interval $(0, 1)$ is used. Clearly, this approach does not work in the non-standard case, but using a special representation of Sturmian sequences by Sturmian morphisms, we can deal with it, too. This technique is basically used in [12] to study the derived sequences of fixed points of primitive Sturmian morphisms.

As is well known, Sturmian sequences have various generalizations for multi-letter alphabets. The first one was introduced by Arnoux and Rauzy [2]: a uniformly recurrent sequence $\mathbf{u}$ over $\mathcal{A}$ is called *Arnoux–Rauzy* if it has exactly one left and one right special factor per length and all left (right, respectively) special factors appear in $\mathbf{u}$ immediately preceded (followed, respectively) by all letters from $\mathcal{A}$.

Many properties of the Arnoux–Rauzy sequences are known (see for example the survey [8]). For our considerations the work [9] is especially important since its authors showed that each factor of an Arnoux–Rauzy sequence over $\mathcal{A}$ has exactly $\#\mathcal{A}$ return words. It means that the derived sequences of Arnoux–Rauzy sequences over $\mathcal{A}$ can be considered over the same alphabet $\mathcal{A}$. Nevertheless, such a property does not characterize Arnoux–Rauzy sequences if $\#\mathcal{A} > 2$. For example, by [6] the sequences coding interval exchange transformations can have this property, too. More generally, the sequences over $\mathcal{A}$ each of whose factors has exactly $\#\mathcal{A}$ return words were studied in [3].

The aim of this paper is to study the derived sequences of Arnoux–Rauzy sequences. Let us emphasize that the description of derived sequences of standard Arnoux–Rauzy sequences can be easily deduced from the work of Justin and Vuillon [9], while here we cover also the more complicated case of non-standard Arnoux–Rauzy sequences. As in [12], our main tool is a special representation of Arnoux–Rauzy sequences, namely the directive sequences containing pure episturmian morphisms (see Sect. 2.3). Since these directive sequences need not be unique, in [7] the authors introduce so-called *normalized directive sequences* and show that these representations are unique. Moreover, they have also other useful properties which allow us to use them for a construction of derived sequences (see Sects. 2.3 and 3).

For every Arnoux–Rauzy sequence $\mathbf{u}$ we describe the set $\mathrm{Der}(\mathbf{u})$ of derived sequences with respect to all nonempty prefixes of $\mathbf{u}$ (see Theorem 24). As a corollary, we show that every derived sequence of an Arnoux–Rauzy sequence is an Arnoux–Rauzy sequence as well. By Durand's fundamental result [4] the sequence $\mathbf{u}$ is primitive substitutive if and only if the set $\mathrm{Der}(\mathbf{u})$ is finite. Here we precisely determine the cardinality of $\mathrm{Der}(\mathbf{u})$ for all primitive substitutive Arnoux–Rauzy sequences (see Corollary 29). It generalizes the results from [12], where the cardinality of $\mathrm{Der}(\mathbf{u})$ is bounded for the fixed points of primitive Sturmian morphisms.

2 Preliminaries

2.1 Words, Sequences and Morphisms

An *alphabet* $\mathcal{A}$ is a finite set of symbols called *letters*. A *word* of length n over $\mathcal{A}$ is a string $u = u_0 u_1 \cdots u_{n-1}$, where all $u_i \in \mathcal{A}$. The *length* of u is denoted by $|u| = n$. The unique word ε of length 0 is called the *empty word*. The symbol $\mathcal{A}^*$ denotes the set of all finite words over $\mathcal{A}$ and $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$. By $|u|_a$ we denote the number of copies of the letter a used in u. The *reversal* of a word $u = u_0 u_1 \cdots u_{n-1}$ is the word $u_{n-1} \cdots u_1 u_0$.

A *sequence* over $\mathcal{A}$ is a right infinite string $\mathbf{u} = u_0 u_1 u_2 \cdots \in \mathcal{A}^{\mathbb{N}}$ with letters $u_i \in \mathcal{A}$ for all $i \in \mathbb{N} = \{0, 1, 2, \ldots\}$. A sequence $\mathbf{u}$ is *eventually periodic* if $\mathbf{u} = wvvv \cdots = wv^\omega$ for some $v, w \in \mathcal{A}^*$; otherwise it is *aperiodic*.

A word w of length n is a *factor* of $\mathbf{u} = u_0 u_1 u_2 \cdots$ if there is an index i such that $w = u_i u_{i+1} u_{i+2} \cdots u_{i+n-1}$. The index i is called an *occurrence* of w in $\mathbf{u}$. Further, if $i = 0$, then w is a *prefix* of $\mathbf{u}$. We will also use the abbreviated notation $u_i u_{i+1} \cdots u_{j-1} = \mathbf{u}_{[i,j)}$ and $u_i u_{i+1} \cdots = \mathbf{u}_{[i,\infty)}$ for all integers $0 \leq i < j$.

The *language* $\mathcal{F}(\mathbf{u})$ of a sequence $\mathbf{u}$ is the set of all its factors. A factor w of $\mathbf{u}$ is *right special* (*left special*, resp.) if there exist at least two letters $a, b \in \mathcal{A}$ such that $wa, wb \in \mathcal{F}(\mathbf{u})$ ($aw, bw \in \mathcal{F}(\mathbf{u})$, resp.).

If each factor w of $\mathbf{u}$ has infinitely many occurrences in $\mathbf{u}$, the sequence $\mathbf{u}$ is *recurrent*. Moreover, if the distances between two consecutive occurrences of w are bounded, then $\mathbf{u}$ is *uniformly recurrent*.

A *morphism* over $\mathcal{A}^*$ is a mapping $\psi : \mathcal{A}^* \mapsto \mathcal{A}^*$ such that $\psi(vw) = \psi(v)\psi(w)$ for all $v, w \in \mathcal{A}^*$. We consider only non-erasing morphisms for which $\psi(a) \neq \varepsilon$ for every $a \in \mathcal{A}$. Then the domain of the morphism ψ can be naturally extended to $\mathcal{A}^{\mathbb{N}}$ by $\psi(u_0 u_1 \cdots) = \psi(u_0)\psi(u_1) \cdots$. A morphism ψ is *primitive* if there is $k \in \mathbb{N}$ such that for every $a, b \in \mathcal{A}$ the letter a occurs in $\psi^k(b)$.

A *fixed point* of a morphism ψ is a sequence $\mathbf{u}$ such that $\psi(\mathbf{u}) = \mathbf{u}$. A sequence $\mathbf{v}$ is *primitive substitutive* if $\mathbf{v} = \theta(\mathbf{u})$, where θ is a morphism and $\mathbf{u}$ is a fixed point of a primitive morphism.

A *permutation* P on $\mathcal{A}$ is a morphism over $\mathcal{A}^*$ such that $\{P(a) : a \in \mathcal{A}\} = \mathcal{A}$. The *order* of the permutation P is the smallest integer $n > 0$ such that $P^n = \mathrm{Id}$.

2.2 Return Words and Derived Sequences

Let $i < j$ be two consecutive occurrences of a factor w in a recurrent sequence $\mathbf{u}$. Then the word $\mathbf{u}_{[i,j)}$ is a *return word* to w in $\mathbf{u}$. The set of all return words to w in $\mathbf{u}$ is denoted $\mathcal{R}_{\mathbf{u}}(w)$. If the sequence $\mathbf{u}$ is uniformly recurrent, then every factor w of $\mathbf{u}$ has a finite number of return words, we denote them $\mathcal{R}_{\mathbf{u}}(w) = \{r_0, r_1, \ldots, r_{k-1}\}$. In addition, if w is a prefix of $\mathbf{u}$, the sequence $\mathbf{u}$ can be written as the unique concatenation of these return words: $\mathbf{u} = r_{d_0} r_{d_1} r_{d_2} \cdots$ and the *derived sequence* of $\mathbf{u}$ to the prefix w is the sequence $\mathbf{d}_{\mathbf{u}}(w) = d_0 d_1 d_2 \cdots$ over the alphabet of cardinality $\#\mathcal{R}_{\mathbf{u}}(w) = k$. Originally, Durand [4] fixed this alphabet to the set $\{0, 1, \ldots, k-1\}$ and required that for $i < j$ the first occurrence of r_i in $\mathbf{u}$ is less than the first occurrence of r_j in $\mathbf{u}$. In particular, this means that his derived sequences always start with the letter 0. In this article, we do not need to fix the alphabet of derived sequences: two derived sequences which differ only by a permutation of letters are identified with one another.

We consider only aperiodic and uniformly recurrent sequences $\mathbf{u}$. Our aim is to describe the set

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

Let us emphasize that we study only derived sequences with respect to nonempty prefixes since the derived sequence with respect to the empty word is trivial.

Clearly, if a nonempty prefix w of $\mathbf{u}$ is not right special, then there exists a unique letter a such that $wa \in \mathcal{F}(\mathbf{u})$. Thus the occurrences of w and wa coincide, and so $\mathcal{R}_{\mathbf{u}}(w) = \mathcal{R}_{\mathbf{u}}(wa)$ and $\mathbf{d}_{\mathbf{u}}(w) = \mathbf{d}_{\mathbf{u}}(wa)$. Since $\mathbf{u}$ is aperiodic, w is a prefix of some right special prefix of $\mathbf{u}$. Therefore, it suffices to take into consideration only right special prefixes of $\mathbf{u}$, i.e.,

$$\mathrm{Der}(\mathbf{u}) = \{\mathbf{d}_{\mathbf{u}}(w) \colon w \text{ is a nonempty right special prefix of } \mathbf{u}\}.$$

2.3 Episturmian and Arnoux–Rauzy Sequences

Definition 1. *A sequence* $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ *is* episturmian *if its language is closed under reversal and* $\mathbf{u}$ *has at most one right special factor of each length.*

An episturmian sequence $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ *is an* Arnoux–Rauzy sequence *if* $\mathbf{u}$ *has exactly one right special factor of each length and* $wa \in \mathcal{F}(\mathbf{u})$ *for every right special factor* w *of* $\mathbf{u}$ *and every letter* $a \in \mathcal{A}$*. An Arnoux–Rauzy sequence* $\mathbf{u}$ *is* standard *if each of its prefixes is a left special factor of* $\mathbf{u}$.

The Arnoux–Rauzy sequences over $\mathcal{A}$ are sometimes called $\#\mathcal{A}$-strict episturmian sequences, since there are also epistumian sequences which are not Arnoux–Rauzy (e.g., see [8]). In the binary case, the set of all Arnoux–Rauzy sequences coincides with the set of all Sturmian sequences. Clearly, all Arnoux–Rauzy sequences are aperiodic and by [5] they are also uniformly recurrent.

Example 2. The Tribonacci sequence $\mathbf{u}_\tau = abacabaabacababacabaa\cdots$ which is the fixed point of the morphism $\tau : a \to ab$, $b \to ac$, $c \to a$ is a standard Arnoux–Rauzy sequence over $\{a, b, c\}$.

In the sequel, we will use the description of episturmian sequences in terms of sequences of pure episturmian morphisms. We follow the notation from [7].

Definition 3. *For every* $a \in \mathcal{A}$ *we define* elementary episturmian morphisms*:*

$$L_a : \begin{cases} a \to a \\ b \to ab \quad \text{for all } b \neq a \end{cases} \qquad \text{and} \qquad R_a : \begin{cases} a \to a \\ b \to ba \quad \text{for all } b \neq a. \end{cases}$$

These $2\#\mathcal{A}$ *morphisms generate the monoid* $\mathcal{M}_{\mathcal{A}} = \langle L_a, R_a : a \in \mathcal{A}\rangle$ *of* pure episturmian morphisms*.*

Let us remark that *episturmian morphisms* are the morphisms obtained by composition of pure episturmian morphisms and permutations (e.g., see [8,10]). All episturmian morphisms are injective.

Definition 4. *For a given alphabet $\mathcal{A}$ we define a new alphabet $\bar{\mathcal{A}} = \{\bar{a} : a \in \mathcal{A}\}$ and we consider words and sequences over the alphabet $\mathcal{A} \cup \bar{\mathcal{A}}$ called* spinned. *We put $\varphi_a = L_a$ and $\varphi_{\bar{a}} = R_a$ for every letter $a \in \mathcal{A}$. Then for every spinned word $z = z_0 z_1 \cdots z_{n-1} \in (\mathcal{A} \cup \bar{\mathcal{A}})^*$ we write*

$$\varphi_z = \varphi_{z_0}\varphi_{z_1} \cdots \varphi_{z_{n-1}} \in \mathcal{M}_{\mathcal{A}}$$

and we say that z is a directive word *of the morphism φ_z. A spinned word is* L-spinned *(*R-spinned*, respectively) if all its letters are from $\mathcal{A}$ ($\bar{\mathcal{A}}$, respectively). The* opposite *word $\bar{z}$ of a spinned word z is obtained from z by switching spins of all its letters.*

Example 5. The words $\bar{a}a\bar{b}\bar{c}a$, abc, $\bar{b}\bar{b}$ are spinned words over $\{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$. The word $z = \bar{a}a\bar{b}\bar{c}a$ directs the morphism $\psi = \varphi_z = \varphi_{\bar{a}a\bar{b}\bar{c}a} = R_a L_a R_b R_c L_a$. The word abc is L-spinned, while $\bar{b}\bar{b}$ is R-spinned. The opposite word of z is $a\bar{a}bc\bar{a}$.

Pure episturmian morphisms can have more than one directive word, i.e., the monoid $\mathcal{M}_{\mathcal{A}}$ is not free. Nevertheless, the presentation of the monoid $\mathcal{M}_{\mathcal{A}}$ is known. Here we state it in the notion of directive words using the so-called *block-transformation* from [11], but it also follows from more general presentation of the whole episturmian monoid as stated in [13].

Definition 6. *A* block-transformation *in the word z is the replacement of the factor $av\bar{a}$ of z, where $a \in \mathcal{A}$ and $v \in (\mathcal{A} \setminus \{a\})^*$, by the opposite word $\bar{a}\bar{v}a$ or vice-versa.*

Proposition 7 ([11])**.** *Let z, z' be two spinned words over $\mathcal{A} \cup \bar{\mathcal{A}}$. Then $\varphi_z = \varphi_{z'}$ if and only if we can pass from z to z' by a chain of block-transformations.*

Example 8. Using the block-transformations from Definition 6 we may rewrite $\bar{a}a\bar{b}\bar{c}a \longleftrightarrow a\bar{a}\bar{b}\bar{c}a \longleftrightarrow aabc\bar{a}$, and so by Proposition 7 all these words direct the same morphism, i.e., $\varphi_{\bar{a}a\bar{b}\bar{c}a} = \varphi_{a\bar{a}\bar{b}\bar{c}a} = \varphi_{aabc\bar{a}}$.

The following theorem extends the notion of directive words to infinite episturmian sequences.

Theorem 9 ([10])**.** *A sequence $\mathbf{u}$ is episturmian if and only if there exists a spinned sequence $\mathbf{z} = z_0 z_1 z_2 \cdots \in (\mathcal{A} \cup \bar{\mathcal{A}})^{\mathbb{N}}$ and an infinite sequence $(\mathbf{u}^{(i)})_{i \geq 0}$ of recurrent sequences such that $\mathbf{u}^{(0)} = \mathbf{u}$ and*

$$\mathbf{u}^{(i)} = \varphi_{z_i}(\mathbf{u}^{(i+1)}).$$

This sequence $\mathbf{z}$ is called a directive sequence *of $\mathbf{u}$.*

Let us notice that the directive sequence from Theorem 9 is the same object as the directive sequence from the construction of episturmian sequences using *palindromic closures* (e.g., see Sect. 3 in [8]).

Proposition 10 ([10]).

(i) *A spinned sequence* $\mathbf{z} \in (\mathcal{A} \cup \bar{\mathcal{A}})^{\mathbb{N}}$ *which has infinitely many L-spinned letters directs the unique episturmian sequence* $\mathbf{u}$. *Moreover, the sequence* $\mathbf{u}$ *starts with the left-most L-spinned letter in* $\mathbf{z}$.

(ii) *A spinned sequence* $\mathbf{z} \in (\mathcal{A} \cup \bar{\mathcal{A}})^{\mathbb{N}}$ *which contains finitely many L-spinned letters directs one episturmian sequence for each* $\bar{a} \in \bar{\mathcal{A}}$ *which occurs in* $\mathbf{z}$ *infinitely many times.*

Proposition 10 implies that some directive sequences direct more than one episturmian sequence. In addition, an episturmian sequence can have more than one directive sequence. However, in [7] the authors describe all directive sequences which direct the same episturmian sequence. Here we state this result only for the case of aperiodic episturmian sequences.

Theorem 11 (Theorem 4.1 in [7]). *Two spinned sequences* $\mathbf{z}^{(1)}$ *and* $\mathbf{z}^{(2)}$ *direct the same aperiodic episturmian sequence if and only if one of the following cases holds for some* i, j *such that* $\{i, j\} = \{1, 2\}$*:*

(i) $\mathbf{z}^{(i)} = \prod_{n\geq 1} u^{(n)}$, $\mathbf{z}^{(j)} = \prod_{n\geq 1} v^{(n)}$, *where* $u^{(n)}$, $v^{(n)}$ *are spinned words such that* $\varphi_{u^{(n)}} = \varphi_{v^{(n)}}$ *for all* $n \geq 1$*;*

(ii) $\mathbf{z}^{(i)} = wa \prod_{n\geq 1} u^{(n)} x^{(n)}$, $\mathbf{z}^{(j)} = w'\bar{a} \prod_{n\geq 1} \bar{u}^{(n)} y^{(n)}$, *where* w, w' *are spinned words such that* $\varphi_w = \varphi_{w'}$, a *is an L-spinned letter and for all* $n \geq 1$, $u^{(n)}$ *is a nonempty* a*-free L-spinned word,* $\bar{u}^{(n)}$ *is the opposite word of* $u^{(n)}$ *and* $x^{(n)}$, $y^{(n)}$ *are nonempty spinned words over* $\{a, \bar{a}\}$ *such that* $|x^{(n)}| = |y^{(n)}|$ *and* $|x^{(n)}|_a = |y^{(n)}|_a$.

Item (i) is based on block-transformations of the directive words of episturmian morphisms, while Item (ii) brings new relations. Now we define the normalized directive sequences which are unique for all aperiodic episturmian sequences.

Definition 12. *A spinned sequence* $\mathbf{z} \in (\mathcal{A} \cup \bar{\mathcal{A}})^{\mathbb{N}}$ *is* normalized *if it contains infinitely many L-spinned letters, but no factor from the set* $\{\bar{a}\bar{\mathcal{A}}^* a : a \in \mathcal{A}\}$.

Theorem 13 (Theorem 5.2 in [7]). *Any aperiodic episturmian sequence* $\mathbf{u}$ *has a unique normalized directive sequence.*

Every normalized spinned sequence directs exactly one episturmian sequence, see Proposition 10. Moreover, the normalized directive sequences can be constructed using Theorem 13. If a directive sequence does not contain infinitely many L-spinned letters, then we use Item (ii) to find another one with infinitely many L-spinned letters. If a directive sequence contains infinitely many L-spinned letters, then it can be normalized from left to right by repeated applications of Item (i) (see [7] for more details).

The Arnoux–Rauzy sequences can be easily recognised by their directive sequences (e.g., see Sect. 2.3 in [8]).

Proposition 14. *An episturmian sequence* $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ *with the directive sequence* $\mathbf{z}$ *is an Arnoux–Rauzy sequence over* $\mathcal{A}$ *if and only if for every* $a \in \mathcal{A}$ *the letter* a *or* $\bar{a}$ *occurs infinitely many times in* $\mathbf{z}$.

Remark 15. Theorem 9 and Proposition 14 immediately imply that for an Arnoux–Rauzy sequence $\mathbf{u}$ each sequence $\mathbf{u}^{(i)}$ from Theorem 9 is an Arnoux–Rauzy sequence with a directive sequence $\mathbf{z}_{[i,\infty)} = z_i z_{i+1} \cdots$.

Example 16. By Propositions 10 and 14, the spinned sequence $\mathbf{y} = a(a\bar{b}\bar{c}\bar{a})^{\omega}$ directs the unique Arnoux–Rauzy sequence $\mathbf{u}$ over $\{a, b, c\}$. Obviously, $\mathbf{y}$ is not normalized. We can normalize it using Item (i) of Theorem 11. First we set $u^{(1)} = aa\bar{b}\bar{c}$, $u^{(2k)} = \bar{a}a$ and $u^{(2k+1)} = \bar{b}\bar{c}$ for all $k > 0$ and make the block-transformations in all even blocks. We get $\mathbf{y}' = aa\bar{b}\bar{c}(a\bar{a}\bar{b}\bar{c})^{\omega}$. Then we set $u^{(1)} = aa\bar{b}\bar{c}a$ and $u^{(k)} = \bar{a}\bar{b}\bar{c}a$ for all $k > 1$. After the relevant block-transformations we get $\mathbf{y}'' = aa\bar{b}\bar{c}a(abc\bar{a})^{\omega}$. Finally we set $u^{(1)} = aa\bar{b}\bar{c}aa$, $u^{(2k)} = bc$ and $u^{(2k+1)} = \bar{a}a$ for all $k > 0$, which leads us to the normalized sequence $\mathbf{y}''' = aa\bar{b}\bar{c}aa(bca\bar{a})^{\omega}$.

By Proposition 10, the spinned sequence $\mathbf{z} = (\bar{a}\bar{b}\bar{c})^{\omega}$ directs three Arnoux–Rauzy sequences $\mathbf{u}^{(a)}$, $\mathbf{u}^{(b)}$, $\mathbf{u}^{(c)}$ starting with the letters a, b, c, respectively. Using Item (ii) of Theorem 11 we find their normalized directive sequences $\mathbf{z}^{(a)} = a(bc\bar{a})^{\omega}$, $\mathbf{z}^{(b)} = \bar{a}b(ca\bar{b})^{\omega}$ and $\mathbf{z}^{(c)} = \bar{a}\bar{b}c(ab\bar{c})^{\omega}$, respectively.

Justin and Vuillon [9] completely describe the return words to any factor of an episturmian sequence. In particular, an Arnoux–Rauzy sequence has the same number of return words to each of its factors.

Proposition 17 ([9])**.** *Let* $\mathbf{u}$ *be an Arnoux–Rauzy sequence over* $\mathcal{A}$. *Then every factor* w *of* $\mathbf{u}$ *has exactly* $\#\mathcal{A}$ *different return words.*

3 Derived Sequences of Episturmian Preimages

In this section we study the relations between the derived sequences of a given Arnoux–Rauzy sequence and the derived sequences of its preimage under the morphisms L_a or R_a. In the binary case, these relations are completely analogous to those described in Section 3 of [12]. Proposition 19 can be also deduced from the results in [9].

For simplicity, we now define the return words and the derived sequence with respect to the empty prefix ε of a sequence $\mathbf{u}$ over $\mathcal{A}$ as $\mathcal{R}_{\mathbf{u}}(\varepsilon) = \mathcal{A}$ and $\mathbf{d}_{\mathbf{u}}(\varepsilon) = \mathbf{u}$. We start with an auxiliary lemma which follows directly from the form of the morphism L_a.

Lemma 18. *Let* $\mathbf{u}$, $\mathbf{u}'$ *be Arnoux–Rauzy sequences over* $\mathcal{A}$ *such that* $\mathbf{u} = L_a(\mathbf{u}')$ *for some* $a \in \mathcal{A}$. *For each factor* $pa \in \mathcal{F}(\mathbf{u})$ *with the prefix* a *there is exactly one word* $p' \in \mathcal{F}(\mathbf{u}')$ *such that* $pa = L_a(p')a$.

Proposition 19. *Let* $\mathbf{u}$ *and* $\mathbf{u}'$ *be Arnoux–Rauzy sequences over* $\mathcal{A}$ *such that* $\mathbf{u} = L_a(\mathbf{u}')$ *for some* $a \in \mathcal{A}$.

(i) *If w is a nonempty right special prefix of $\mathbf{u}$, then there exists a right special prefix w' of $\mathbf{u}'$ such that $w = L_a(w')a$ and $\mathbf{d_u}(w) = \mathbf{d_{u'}}(w')$.*
(ii) *If w' is a right special prefix of $\mathbf{u}'$, then $w = L_a(w')a$ is a right special prefix of $\mathbf{u}$ and $\mathbf{d_{u'}}(w') = \mathbf{d_u}(w)$.*

Proof. We start with Item (i). For a nonempty right special prefix w of $\mathbf{u}$ we denote its return words $\mathcal{R}_\mathbf{u}(w) = \{r_c : c \in \mathcal{A}\}$ and its derived sequence $\mathbf{d_u}(w) = d_0d_1\cdots$. Thus $\mathbf{u} = r_{d_0}r_{d_1}\cdots$. By the form of the morphism L_a, the sequence $\mathbf{u}$ starts with the letter a and a is also separating in $\mathbf{u}$, i.e., every factor of $\mathbf{u}$ of length two contains the letter a. Since w is nonempty right special prefix, it both starts and ends with the letter a and by Lemma 18 there is a unique prefix w' of $\mathbf{u}'$ such that $w = L_a(w')a$. Since w is a right special factor of the Arnoux–Rauzy sequence $\mathbf{u}$, the word $wc = L_a(w')ac \in \mathcal{F}(\mathbf{u})$ for every $c \in \mathcal{A}$. Thus $w'c \in \mathcal{F}(\mathbf{u}')$ for every $c \in \mathcal{A}$ and so w' is a right special factor of $\mathbf{u}'$. In addition, all return words r_c start with the letter a and so by Lemma 18 there are uniquely given words r'_c such that $r_c = L_a(r'_c)$ for all $c \in \mathcal{A}$. Since L_a is injective, we have $\mathbf{u}' = r'_{d_0}r'_{d_1}\cdots$.

Now it suffices to prove that the set $\{|r'_{d_0}\cdots r'_{d_j}| : j \in \mathbb{N}\} \cup \{0\}$ is the set of all occurrences of w' in $\mathbf{u}'$. Then the words r'_c, $c \in \mathcal{A}$, are return words to w' in $\mathbf{u}'$ and $\mathbf{d_u}(w) = \mathbf{d_{u'}}(w')$. Let $i > 0$ be an occurrence of w' in $\mathbf{u}'$. It means that $\mathbf{u}'_{[0,i)}w'c$ is a prefix of $\mathbf{u}'$ for some $c \in \mathcal{A}$. Then $L_a(\mathbf{u}'_{[0,i)}w'c)$ is a prefix of $\mathbf{u}$, the word $L_a(w'c)$ has a prefix $L_a(w')a = w$ and $|L_a(\mathbf{u}'_{[0,i)})|$ is an occurrence of w in $\mathbf{u}$. Thus $L_a(\mathbf{u}'_{[0,i)}) = r_{d_0}\cdots r_{d_j}$ for some $j \in \mathbb{N}$ and by injectivity of L_a, it follows that $\mathbf{u}'_{[0,i)} = r'_{d_0}\cdots r'_{d_j}$ and so $i = |r'_{d_0}\cdots r'_{d_j}|$ for some $j \in \mathbb{N}$.

Conversely, we suppose that $i = |r'_{d_0}\cdots r'_{d_j}|$ for some $j \in \mathbb{N}$. If we denote $p = r_{d_0}\cdots r_{d_j}$, then pw is a prefix of $\mathbf{u}$ and by Lemma 18 there is a unique prefix p' of $\mathbf{u}'$ such that $p = L_a(p')$. Clearly, $p'w'$ is also a prefix of $\mathbf{u}'$ and by injectivity of L_a we can conclude that $p' = r'_{d_0}\cdots r'_{d_j}$. Thus i is an occurrence of w' in $\mathbf{u}'$.

To prove Item (ii) we suppose that w' is a right special prefix of $\mathbf{u}'$. We denote its return words $\mathcal{R}_{\mathbf{u}'}(w') = \{r'_c : c \in \mathcal{A}\}$ and its derived sequence $\mathbf{d_{u'}}(w') = d_0d_1\cdots$. Thus $\mathbf{u}' = r'_{d_0}r'_{d_1}\cdots$. If we set $w = L_a(w')a$ and $r_c = L_a(r'_c)$ for all $c \in \mathcal{A}$, we get $\mathbf{u} = r_{d_0}r_{d_1}\cdots$. Now it remains to prove that w is a right special prefix of $\mathbf{u}$ and the set $\{|r_{d_0}\cdots r_{d_j}| : j \in \mathbb{N}\} \cup \{0\}$ is the set of all occurrences of w in $\mathbf{u}$. We skip these proofs since the arguments are completely analogous to those used in the proof of Item (i).

Proposition 20. *Let $\mathbf{u}$ and $\mathbf{u}'$ be Arnoux–Rauzy sequences over $\mathcal{A}$ such that $\mathbf{u} = R_a(\mathbf{u}')$ for some $a \in \mathcal{A}$ and let $\mathbf{u}$ start with the letter $b \in \mathcal{A}, b \neq a$.*

(i) *If w is a nonempty right special prefix of $\mathbf{u}$, then there exists a nonempty right special prefix w' of $\mathbf{u}'$ such that $w = R_a(w')$ and $\mathbf{d_u}(w) = \mathbf{d_{u'}}(w')$.*
(ii) *If w' is a nonempty right special prefix of $\mathbf{u}'$, then $w = R_a(w')$ is a nonempty right special prefix of $\mathbf{u}$ and $\mathbf{d_{u'}}(w') = \mathbf{d_u}(w)$.*

Proof. The morphisms L_a and R_a are conjugate, i.e., $aR_a(x) = L_a(x)a$ for every word $x \in \mathcal{A}^*$. Thus for the Arnoux–Rauzy sequence $\mathbf{v} = a\mathbf{u}$ we get $\mathbf{v} = aR_a(\mathbf{u}') = L_a(\mathbf{u}')$, since the conjugacy holds for every prefix of $\mathbf{u}'$.

Let w be a nonempty right special prefix of $\mathbf{u}$ and let (i_n) be the increasing sequence of the occurrences of w in $\mathbf{u}$. By the form of the morphism R_a, each letter $b \neq a$ (excluding the first letter of $\mathbf{u}$) is preceded by the letter a. Thus the sequence (i_n) is also the sequence of the occurrences of the word aw in $\mathbf{v}$ and $\mathbf{d}_{\mathbf{u}}(w) = \mathbf{d}_{\mathbf{v}}(aw)$. Moreover, aw is a right special prefix of $\mathbf{v}$ and so we can apply Proposition 19 and find the right special prefix w' of $\mathbf{u}'$ such that $aw = L_a(w')a = aR_a(w')$ and $\mathbf{d}_{\mathbf{u}}(w) = \mathbf{d}_{\mathbf{v}}(aw) = \mathbf{d}_{\mathbf{u}'}(w')$. The proof of Item (ii) is similar and so we skip it.

Propositions 19 and 20 can be also restated as follows.

Corollary 21. *Let $\mathbf{u}, \mathbf{u}'$ be Arnoux–Rauzy sequences over $\mathcal{A}$ and $a \in \mathcal{A}$.*

(i) *If $\mathbf{u} = L_a(\mathbf{u}')$, then $\operatorname{Der}(\mathbf{u}) = \operatorname{Der}(\mathbf{u}') \cup \{\mathbf{u}'\}$.*
(ii) *If $\mathbf{u} = R_a(\mathbf{u}')$ and $\mathbf{u}$ starts with a letter $b \in \mathcal{A}, b \neq a$, then $\operatorname{Der}(\mathbf{u}) = \operatorname{Der}(\mathbf{u}')$.*

4 Derived Sequences of Arnoux–Rauzy Sequences

First, we introduce a transformation Δ on the set of normalized directive sequences. Subsequently, we use this transformation to describe the set $\operatorname{Der}(\mathbf{u})$ of derived sequences of an Arnoux–Rauzy sequence $\mathbf{u}$.

Definition 22. *Let $\mathbf{z} = z_0 z_1 z_2 \cdots$ be a normalized spinned sequence and let k be the unique index such that z_k is an L-spinned letter and $z_0 z_1 \cdots z_{k-1}$ is an R-spinned word (or is empty). Then $\Delta(\mathbf{z}) = \mathbf{z}_{[k+1,\infty)} = z_{k+1} z_{k+2} z_{k+3} \cdots$.*

Clearly, if $\mathbf{z}$ is the normalized directive sequence of an Arnoux–Rauzy sequence $\mathbf{u}$, then $\Delta(\mathbf{z})$ is the normalized directive sequence of an Arnoux–Rauzy sequence as well. For every integer $m \geq 1$ we let $\mathbf{d}_m$ denote the Arnoux–Rauzy sequence directed by $\Delta^m(\mathbf{z})$ and we also set $\mathbf{d}_0 = \mathbf{u}$.

Example 23. For the normalized spinned sequence $\mathbf{z} = \bar{c}ba(\bar{c}b\bar{a}b)^\omega$ we get $\Delta(\mathbf{z}) = a(\bar{c}b\bar{a}b)^\omega$, $\Delta^2(\mathbf{z}) = (\bar{c}b\bar{a}b)^\omega$, $\Delta^3(\mathbf{z}) = (\bar{a}b\bar{c}b)^\omega$ and $\Delta^4(\mathbf{z}) = (\bar{c}b\bar{a}b)^\omega = \Delta^2(\mathbf{z})$.

Theorem 24. *Let $\mathbf{u}$ be an Arnoux–Rauzy sequence over $\mathcal{A}$ with the normalized directive sequence $\mathbf{z}$. Then $\mathbf{d}$ is the derived sequence with respect to a nonempty prefix of $\mathbf{u}$ if and only if $\mathbf{d} = \mathbf{d}_m$ for some $m \geq 1$, i.e., $\mathbf{d}$ is an Arnoux–Rauzy sequence directed by $\Delta^m(\mathbf{z})$ for some $m \geq 1$.*

Proof. ($\Rightarrow$) We consider a nonempty right special prefix w of $\mathbf{u}$ and prove that $\mathbf{d}_{\mathbf{u}}(w) = \mathbf{d}_m$ for some $m \geq 1$. In fact, we prove that for every $i \in \mathbb{N}$ and a right special prefix v of $\mathbf{d}_i$ there is a right special prefix v' of $\mathbf{d}_{i+1}$ such that $|v'| < |v|$ and $\mathbf{d}_{\mathbf{d}_i}(v) = \mathbf{d}_{\mathbf{d}_{i+1}}(v')$. Then starting with a nonempty right special prefix w of $\mathbf{u}$ we eventually find the index $m \geq 1$ and the prefix w'' of $\mathbf{d}_m$ such that $w'' = \varepsilon$ and so $\mathbf{d}_{\mathbf{u}}(w) = \mathbf{d}_{\mathbf{d}_m}(\varepsilon) = \mathbf{d}_m$.

Since $\mathbf{z}$ is normalized, $\Delta^i(\mathbf{z}) = \mathbf{y}$ is also normalized and so it has a prefix $\bar{x}a$ for some $a \in \mathcal{A}$ and $\bar{x} \in (\bar{\mathcal{A}} \setminus \{\bar{a}\})^*$. If $\bar{x} = \varepsilon$, then $\Delta^{i+1}(\mathbf{z}) = \mathbf{y}_{[1,\infty)}$ and so

$\mathbf{d}_i = L_a(\mathbf{d}_{i+1})$. By Proposition 19 there is a right special prefix v' of $\mathbf{d}_{i+1}$ such that $v = L_a(v')a$ and $\mathbf{d}_{\mathbf{d}_i}(v) = \mathbf{d}_{\mathbf{d}_{i+1}}(v')$. If $\bar{x}$ is nonempty, we denote $|\bar{x}| = n$. Then $\Delta^{i+1}(\mathbf{z}) = \mathbf{y}_{[n+1,\infty)}$. Let us denote $\mathbf{u}^{(\ell)}$ the sequence directed by $\mathbf{y}_{[\ell,\infty)}$ for all $\ell \in \mathbb{N}$. In particular, $\mathbf{u}^{(0)} = \mathbf{d}_i$, $\mathbf{u}^{(n+1)} = \mathbf{d}_{i+1}$ and $\mathbf{u}^{(\ell)} = \varphi_{y_\ell}(\mathbf{u}^{(\ell+1)})$ for all $\ell \in \mathbb{N}$. By Proposition 10 all sequences $\mathbf{u}^{(0)}, \dots, \mathbf{u}^{(n)}$ starts with the letter a and so by Proposition 20 there are nonempty right special prefixes $v^{(\ell)}$ of $\mathbf{u}^{(\ell)}$ for all $\ell = 0, \dots, n$ such that $v^{(0)} = v$, $v^{(\ell)} = \varphi_{y_\ell}(v^{(\ell+1)})$ for all $\ell = 0, \dots, n-1$ and

$$\mathbf{d}_{\mathbf{d}_i}(v) = \mathbf{d}_{\mathbf{u}^{(1)}}(v^{(1)}) = \cdots = \mathbf{d}_{\mathbf{u}^{(n)}}(v^{(n)}).$$

By Proposition 19 there is a right special prefix v' of $\mathbf{u}^{(n+1)} = \mathbf{d}_{i+1}$ such that $v^{(n)} = L_a(v')a$ and $\mathbf{d}_{\mathbf{u}^{(n)}}(v^{(n)}) = \mathbf{d}_{\mathbf{d}_{i+1}}(v')$. Since we also have

$$|v| > |v^{(1)}| > |v^{(2)}| > \cdots > |v^{(n)}| > |v'|,$$

v' is the desired right special prefix of $\mathbf{d}_{i+1}$.

($\Leftarrow$) For arbitrary $m \geq 1$ we find a nonempty right special prefix w of $\mathbf{u}$ such that $\mathbf{d}_{\mathbf{u}}(w) = \mathbf{d}_m$. We set $\mathbf{z} = z_0 z_1 \cdots z_i \Delta^m(\mathbf{z})$ for some $i \in \mathbb{N}$ and we let $\mathbf{u}^{(\ell)}$ denote the sequence directed by $\mathbf{z}_{[\ell,\infty)}$ for all $\ell \in \mathbb{N}$. In particular, $\mathbf{u}^{(0)} = \mathbf{u}$ and $\mathbf{u}^{(i+1)} = \mathbf{d}_m$. Now we take the right special prefix ε of $\mathbf{d}_m$ and using Propositions 19 and 20 we successively find right special prefixes $w^{(\ell)}$ of $\mathbf{u}^{(\ell)}$ for all $\ell = i, \dots, 0$. Since z_i is L-spinned, the inequalities $0 < |w^{(i)}| < |w^{(i-1)}| < \cdots < |w^{(0)}|$ hold and

$$\mathbf{d}_m = \mathbf{d}_{\mathbf{d}_m}(\varepsilon) = \mathbf{d}_{\mathbf{u}^{(i)}}(w^{(i)}) = \cdots = \mathbf{d}_{\mathbf{u}^{(1)}}(w^{(1)}) = \mathbf{d}_{\mathbf{u}}(w^{(0)}).$$

Then $w^{(0)}$ is the desired prefix w of $\mathbf{u}$.

Corollary 25. *All derived sequences with respect to nonempty prefixes of a given Arnoux–Rauzy sequence over $\mathcal{A}$ are Arnoux–Rauzy sequences over $\mathcal{A}$ as well.*

Proof. This follows directly from Theorems 24 and 9.

By Durand's result [4] the set $\mathrm{Der}(\mathbf{u})$ is finite if and only if $\mathbf{u}$ is a primitive substitutive sequence. An Arnoux–Rauzy sequence is primitive substitutive if and only if its normalized directive sequence is eventually periodic. Indeed, a pure episturmian morphism is primitive if and only if its directive word contains at least one letter a or $\bar{a}$ for every $a \in \mathcal{A}$ and the normalization of an eventually periodic directive sequence always produces an eventually periodic normalized directive sequence (see [7] for more details).

Now we specify the cardinality of $\mathrm{Der}(\mathbf{u})$ according to the normalized directive sequence of an Arnoux–Rauzy sequence $\mathbf{u}$. Let us recall that two derived sequences $\mathbf{d}^{(1)}, \mathbf{d}^{(2)}$ such that $\mathbf{d}^{(1)} = P(\mathbf{d}^{(2)})$ for some permutation P are considered as equal since their structure is the same. Let us emphasize that a permutation P on $\mathcal{A}$ can be naturally extended to the alphabet $\mathcal{A} \cup \bar{\mathcal{A}}$: P acts on the letters from $\mathcal{A}$ without any changes and for every letter $\bar{a} \in \bar{\mathcal{A}}$ we put $P(\bar{a}) = \bar{b}$ if $P(a) = b$.

Observation 26. *Let $\mathbf{u}^{(1)}$, $\mathbf{u}^{(2)}$ be Arnoux–Rauzy sequences with the normalized directive sequences $\mathbf{z}^{(1)}$, $\mathbf{z}^{(2)}$, respectively, and let P be a permutation. Then $\mathbf{u}^{(1)} = P(\mathbf{u}^{(2)})$ if and only if $\mathbf{z}^{(1)} = P(\mathbf{z}^{(2)})$.*

Lemma 27. *Let $\mathbf{z}$ be the normalized directive sequence and let $k < \ell$ be the minimal indices such that there is a permutation P satisfying $\mathbf{z}_{[\ell,\infty)} = P(\mathbf{z}_{[k,\infty)})$. We denote $x = \mathbf{z}_{[0,k)}$, $y = \mathbf{z}_{[k,\ell)}$ and n the order of the permutation P. Then $\mathbf{z}$ is eventually periodic: $\mathbf{z} = x\left(yP(y)\cdots P^{n-1}(y)\right)^{\omega}$. Moreover, every sequence $\mathbf{z}_{[i,\infty)}$ with $i \geq \ell$ is equal (up to permutation of letters) to the sequence $\mathbf{z}_{[j,\infty)}$ for some $j \in \{k,\ldots,\ell-1\}$ and if $\mathbf{z}_{[i,\infty)} = Q(\mathbf{z}_{[j,\infty)})$ for some $j < i$ and a permutation Q, then $i \geq \ell$.*

Proof. In the notation from the statement we can write

$$\begin{aligned}\mathbf{z} = xy\mathbf{z}_{[\ell,\infty)} &= xyP(\mathbf{z}_{[k,\infty)}) = xyP(y\mathbf{z}_{[\ell,\infty)}) = xyP(y)P^2(\mathbf{z}_{[k,\infty)}) = \cdots \\ &= xyP(y)\cdots P^{n-1}(y)P^n(y)P^{n+1}(y)\cdots = x\left(yP(y)\cdots P^{n-1}(y)\right)^{\omega}.\end{aligned}$$

Moreover, for every $i \geq l$ we can write $\mathbf{z}_{[i,\infty)} = P(\mathbf{z}_{[i-\ell+k,\infty)})$. Thus eventually we get $\mathbf{z}_{[i,\infty)} = P^m(\mathbf{z}_{[i',\infty)})$ for some positive integer m and an index i' such that $k \leq i' < \ell$. The last part of the statement clearly holds since otherwise it leads us to the contrary with the minimality of the indices k, ℓ.

Corollary 28. *Let $\mathbf{u}$ be an Arnoux–Rauzy sequence over $\mathcal{A}$ with the aperiodic normalized directive sequence and let v, w be two distinct nonempty right special prefixes of $\mathbf{u}$. Then the derived sequences with respect to v and w are distinct, i.e., $\mathbf{d}_{\mathbf{u}}(v) \neq P(\mathbf{d}_{\mathbf{u}}(w))$ for any permutation P.*

Proof. We argue by contradiction. By Theorem 24 all derived sequences with respect to nonempty prefixes of $\mathbf{u}$ are the elements of the sequence $(\mathbf{d}_m)_{m\geq 1}$. Thus we can suppose that $\mathbf{d}_m = P(\mathbf{d}_\ell)$ for some positive integers m, ℓ and a permutation P. Since v, w are distinct right special prefixes, we get $m \neq \ell$. By Observation 26, it means that $\Delta^m(\mathbf{z}) = P(\Delta^\ell(\mathbf{z}))$ and so by Lemma 27 $\mathbf{z}$ is eventually periodic, which is the contradiction.

Corollary 29. *Let $\mathbf{u}$ be an Arnoux–Rauzy sequence over $\mathcal{A}$ with the eventually periodic normalized directive sequence $\mathbf{z} = x\left(yP(y)\cdots P^{n-1}(y)\right)^{\omega}$, where the words $x \in (\mathcal{A}\cup\bar{\mathcal{A}})^*$, $y \in (\mathcal{A}\cup\bar{\mathcal{A}})^+$ are the shortest possible and P is a permutation with the order n. We denote $|x|_L$, $|xy|_L$ the numbers of L-spinned letters in the words x, xy, respectively.*

- *(i) If the last letters of both x, y are L-spinned, then $\#\mathrm{Der}(\mathbf{u}) = |xy|_L - 1$. More precisely, there are $|x|_L - 1$ derived sequences belonging to exactly one nonempty right special prefix of $\mathbf{u}$ and $|y|_L$ derived sequences belonging to infinitely many right special prefixes of $\mathbf{u}$.*
- *(ii) If the last letter of x or y is R-spinned or $x = \varepsilon$, then $\#\mathrm{Der}(\mathbf{u}) = |xy|_L$. More precisely, there are $|x|_L$ derived sequences belonging to exactly one nonempty right special prefix of $\mathbf{u}$ and $|y|_L$ derived sequences belonging to infinitely many right special prefixes of $\mathbf{u}$.*

Proof. By Theorem 24 all elements of $\mathrm{Der}(\mathbf{u})$ are the elements of the sequence $(\mathbf{d}_m)_{m\geq 1}$. To prove Item (i), we have to show that the sequence $(\mathbf{d}_m)_{m\geq 1}$ has pre-period $|x|_L - 1$ and period $|y|_L$ (up to permutation of letters). However, the sequence $(\Delta^m(\mathbf{z}))_{m\geq 1}$ has the same pre-period and period, see Observation 26. Now it suffices to apply Lemma 27 with $k = |x|$ and $\ell = |xy|$. Since both x and y end with L-spinned letters, the sequences that occur once in $(\Delta^m(\mathbf{z}))_{m\geq 1}$ are exactly the elements of the set $\{\mathbf{z}_{[i,\infty)} : 0 < i < |x|\}$ for which z_{i-1} is L-spinned. Thus they are the sequences $\Delta(\mathbf{z}), \ldots, \Delta^{|x|_L - 1}(\mathbf{z})$. Similarly, the sequences that occur (up to permutation of letters) infinitely many times in $(\Delta^m(\mathbf{z}))_{m\geq 1}$ are exactly the elements of the set $\{\mathbf{z}_{[i,\infty)} : |x| \leq i < |xy|\}$ for which z_{i-1} is L-spinned, so they are the sequences $\Delta^{|x|_L}(\mathbf{z}), \ldots, \Delta^{|xy|_L - 1}(\mathbf{z})$.

We prove Item (ii) analogously. It suffices to realize that if x or y ends with an R-spinned letter or x is the empty word, then the periodic part of $(\Delta^m(\mathbf{z}))_{m\geq 1}$ starts with the element $\Delta^{|x|_L + 1}(\mathbf{z})$. Thus the sequence $(\Delta^m(\mathbf{z}))_{m\geq 1}$ has pre-period $|x|_L$ and period $|y|_L$ (up to permutation of letters).

Example 30. The Arnoux–Rauzy sequence $\mathbf{u}$ is directed by the normalized directive sequence $\mathbf{z} = \bar{c}ba(\bar{c}b\bar{a}b)^\omega = \bar{c}ba(\bar{c}bP(\bar{c}b))^\omega$ for the permutation $P : a \to c, b \to b, c \to a$ with the order 2. By Item (i) of Corollary 29, the sequence $\mathbf{u}$ has two derived sequences: $\mathbf{d}_1$ directed by $\Delta(\mathbf{z}) = a(\bar{c}b\bar{a}b)^\omega$ belonging to the shortest nonempty right special prefix of $\mathbf{u}$ and $\mathbf{d}_2$ directed by $\Delta^2(\mathbf{z}) = (\bar{c}b\bar{a}b)^\omega$ belonging to all the others right special prefixes of $\mathbf{u}$.

The Tribonacci sequence $\mathbf{u}_\tau$ from Example 2 is directed by the normalized directive sequence $\mathbf{z} = (abc)^\omega = (aP(a)P^2(a))^\omega$ for the permutation $P : a \to b, b \to c, c \to a$ with the order 3. Then by Item (ii) of Corollary 29, $\mathbf{u}_\tau$ has one derived sequence $\mathbf{d}$ directed by $(abc)^\omega$. In other words, the derived sequence with respect to any prefix of $\mathbf{u}_\tau$ is the sequence $\mathbf{u}_\tau$ itself.

The Arnoux–Rauzy sequence $\mathbf{u}^{(a)}$ directed by the normalized directive sequence $\mathbf{z}^{(a)} = a(bc\bar{a})^\omega$ (see Example 16) has by Item (ii) of Corollary 29 three derived sequences $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ directed by $\Delta(\mathbf{z}^{(a)}) = (bc\bar{a})^\omega$, $\Delta^2(\mathbf{z}^{(a)}) = (c\bar{a}b)^\omega$, $\Delta^3(\mathbf{z}^{(a)}) = (\bar{a}bc)^\omega$, respectively. The sequence $\mathbf{d}_1$ is the derived sequence with respect to the shortest nonempty right special prefix of $\mathbf{u}^{(a)}$, while both $\mathbf{d}_2, \mathbf{d}_3$ belong to infinitely many right special prefixes of $\mathbf{u}^{(a)}$.

References

1. Araújo, I.M., Bruyère, V.: Words derivated from Sturmian words. Theoret. Comput. Sci. **340**, 204–219 (2005)
2. Arnoux, P., Rauzy, G.: Représentation géométrique de suites de complexité $2n+1$. Bull. de la Société Mathématique de France **119**, 199–215 (1991)
3. Balková, L., Pelantová, E., Steiner, W.: Sequences with constant number of return words. Monatsh. Math. **155**, 251–263 (2008)
4. Durand, F.: A characterization of substitutive sequences using return words. Discrete Math. **179**, 89–101 (1998)
5. Droubay, X., Justin, J., Pirillo, G.: Episturmian words and some constructions of de Luca and Rauzy. Theoret. Comput. Sci. **225**(1–2), 539–553 (2001)

6. Ferenczi, S., Holton, C., Zamboni, L.Q.: Structure of three interval exchange transformations I. An arithmetic study. Ann. Inst. Fourier (Grenobles) **51**, 861–901 (2001)
7. Glen, A., Levé, F., Richomme, G.: Directive words of episturmian words: equivalences and normalization. RAIRO-Theoret. Inf. Appl. **43**, 299–319 (2009)
8. Glen, A., Justin, J.: Episturmian words: a survey. RAIRO-Theoret. Inf. Appl. **43**, 403–442 (2009)
9. Justin, J., Vuillon, L.: Return words in Sturmian and episturmian words. RAIRO-Theoret. Inf. Appl. **34**, 343–356 (2000)
10. Justin, J., Pirillo, G.: Episturmian words and episturmian morphisms. Theoret. Comput. Sci. **276**(1–2), 281–313 (2002)
11. Justin, J., Pirillo, G.: Episturmian words: shifts, morphisms and numeration systems. Int. J. Found. Comput. Sci. **15**(2), 329–348 (2004)
12. Klouda, K., Medková, K., Pelantová, E., Starosta, Š.: Fixed points of Sturmian morphisms and their derivated words. Theoret. Comput. Sci. **743**, 23–37 (2018)
13. Richomme, G.: Lyndon morphisms. Bull. Belg. Math. Soc. Simon Stevin **10**, 761–785 (2003)
14. Vuillon, L.: A characterization of Sturmian words by return words. Eur. J. Combin. **22**, 263–275 (2001)

Article C

Derived sequences of complementary symmetric Rote Sequences

RAIRO-Theor. Inf. Appl. 53 (2019) 125–151
https://doi.org/10.1051/ita/2019004

RAIRO - Theoretical Informatics and Applications
www.rairo-ita.org

DERIVED SEQUENCES OF COMPLEMENTARY SYMMETRIC ROTE SEQUENCES

Kateřina Medková[1,*], Edita Pelantová[1] and Laurent Vuillon[2]

Abstract. Complementary symmetric Rote sequences are binary sequences which have factor complexity $\mathcal{C}(n) = 2n$ for all integers $n \geq 1$ and whose languages are closed under the exchange of letters. These sequences are intimately linked to Sturmian sequences. Using this connection we investigate the return words and the derived sequences to the prefixes of any complementary symmetric Rote sequence $\mathbf{v}$ which is associated with a standard Sturmian sequence $\mathbf{u}$. We show that any non-empty prefix of $\mathbf{v}$ has three return words. We prove that any derived sequence of $\mathbf{v}$ is coding of three interval exchange transformation and we determine the parameters of this transformation. We also prove that $\mathbf{v}$ is primitive substitutive if and only if $\mathbf{u}$ is primitive substitutive. Moreover, if the sequence $\mathbf{u}$ is a fixed point of a primitive morphism, then all derived sequences of $\mathbf{v}$ are also fixed by primitive morphisms. In that case we provide an algorithm for finding these fixing morphisms.

Mathematics Subject Classification. 68R15.

Received December 10, 2018. Accepted October 7, 2019.

1. Introduction

The notion of return words and derived sequences has been introduced by Durand in [16] and seems to be a powerful tool for studying the structure of aperiodic infinite sequences, and so also of the corresponding dynamical systems.

A return word can be considered as a symbolical analogy of return time occurring in the theory of dynamical systems. Let $\mathbf{u} = u_0u_1u_2\cdots$ be a sequence and let $w = u_iu_{i+1}\cdots u_{i+n-1}$ be its factor. The index i is an occurrence of w. A return word to w is a word $u_iu_{i+1}\cdots u_{j-1}$ with $i < j$ being two consecutive occurrences of w.

Return words are well understood in the case of Sturmian sequences, *i.e.* aperiodic sequences with the lowest possible factor complexity $C(n) = n + 1$ for all $n \in \mathbb{N}$. They can be also seen as the coding of rotation with an irrational angle α on the unit circle with the partition in two intervals of lengths α and $1 - \alpha$, respectively.

The third author characterizes Sturmian sequences as sequences with two return words to each their factor in [31]. Similarly the paper [4] is dedicated to investigation of sequences with a fixed number of return words to any factors, in particular, Arnoux-Rauzy sequences and sequences coding interval exchange transformations are

Keywords and phrases: Derived sequence, return word, Rote sequence, Sturmian sequence.

[1] Czech Technical Univ. in Prague, 120 00 Prague 2, Trojanova 13, Czech Republic.
[2] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LAMA, 73000 Chambéry, France.

[*] Corresponding author: katerinamedkova@gmail.com

of this type, see [32]. Besides, return words in episturmian sequences were described in [21] while the description of return words in the coding of rotations was used to show their fullness in [10].

A derived sequence expresses the order of return words in the sequence $\mathbf{u}$. More precisely, if w is a prefix of $\mathbf{u}$ with k return words $r_1, r_2, \ldots, r_k$, then $\mathbf{u}$ can be written as the concatenation of these return words: $\mathbf{u} = r_{d_0} r_{d_1} r_{d_2} \cdots$. Then the derived sequence of $\mathbf{u}$ to the prefix w is the sequence $\mathbf{d} = d_0 d_1 d_2 \cdots$ over an alphabet of size k.

Durand's result from [16] states that a sequence is primitive substitutive if and only if its number of derived sequences is finite. Now the goal is to understand the structure of the derived sequences. Derived sequences of standard Sturmian sequences were investigated in [1] and the derived sequences of fixed points of primitive Sturmian morphisms were described in [23].

Recently, new developments are done to understand the structure of more complicated objects, *e.g.* acyclic, neutral and tree sequences introduced in [7]. Return words in sequences coding linear involutions were studied in [8] and the number of return words for more general neutral sequences was determined in [15]. In [9] the properties of return words and derived sequences were exploited for the characterization of substitutive tree sequences.

In this paper, we study complementary symmetric Rote sequences, *i.e.* the binary sequences with factor complexity $C(n) = 2n$ for all $n \geq 1$ whose languages are closed under the exchange of letters. These sequences are not tree, but they represent an interesting example of neutral sequences with characteristics 1. They are named after Rote, who proposed several constructions of these sequences in [30]. For example, he constructed them as projections of fixed points over a four letter alphabet (see Sect. 7 of our paper) or as the coding of irrational rotations on a unit circle with the partition on two intervals of length $1/2$. Later on, they were also constructed using palindromic and pseudopalindromic closures, see [11]. This construction was proposed and firstly applied to the Thue-Morse sequence in [13] and later extended in [27] to a broader class of sequences.

Our techniques are based on the close relation between complementary symmetric Rote sequences and Sturmian sequences shown in [30]: a sequence $\mathbf{v} = v_0 v_1 v_2$ is a complementary symmetric Rote sequence if and only if its difference sequence $\mathbf{u}$, which is defined by $u_i = v_{i+1} - v_i \mod 2$, is a Sturmian sequence. In fact, we investigate the consequences of this relation, see Sections 2.5, 3 and 5. We also use the description of derived sequences of Sturmian sequences as studied in detail in [23]. Here we focus on complementary symmetric Rote sequences which are associated with standard Sturmian sequences.

First we recall needed terminology and notations in Section 2. Section 3 is dedicated to return words: in Theorem 3.1 we show that every non-empty prefix of any complementary symmetric Rote sequence $\mathbf{v}$ has three return words. In other words, all derived sequences of $\mathbf{v}$ are over a ternary alphabet. Then we proceed with the study of derived sequences. In Proposition 4.2 we prove that any derived sequence of $\mathbf{v}$ is the coding of a three interval exchange transformation and we determine the parameters of this transformation. Then in Theorem 6.3 and Lemma 6.1 we concentrate on the question of substitutivity of Rote sequences. In the case when the associated standard Sturmian sequence $\mathbf{u}$ is fixed by a primitive morphism, Corollary 6.4 estimates the number of distinct derived sequences of $\mathbf{v}$ from above and Algorithm 6.7 provides a list of all derived sequences of $\mathbf{v}$. Section 7 compares our techniques with the original Rote's construction of substitutive Rote sequences and the last section collects related open questions.

2. Preliminaries

2.1. Sequences and morphisms

An *alphabet* $\mathcal{A}$ is a finite set of symbols called *letters*. A *word* over $\mathcal{A}$ of *length* n is a string $u = u_0 u_1 \cdots u_{n-1}$, where $u_i \in \mathcal{A}$ for all $i \in \{0, 1, \ldots, n-1\}$. The length of u is denoted by $|u|$. The set of all finite words over $\mathcal{A}$ together with the operation of concatenation form a monoid $\mathcal{A}^*$. Its neutral element is the *empty word* ε and we denote $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$.

If $u = xyz$ for some $x, y, z \in \mathcal{A}^*$, then x is a *prefix* of u, z is a *suffix* of u and y is a *factor* of u.

To any word u over $\mathcal{A}$ with the cardinality $\#\mathcal{A} = d$ we assign the vector $V_u \in \mathbb{N}^d$ defined as $(V_u)_a = |u|_a$ for all $a \in \mathcal{A}$, where $|u|_a$ is the number of letters a occurring in u. The vector V_u is usually called the *Parikh vector* of u.

A *sequence* over $\mathcal{A}$ is an infinite string $\mathbf{u} = u_0u_1u_2\cdots$, where $u_i \in \mathcal{A}$ for all $i \in \mathbb{N} = \{0, 1, 2, \ldots\}$. We always denote sequences by bold letters. The set of all sequences over $\mathcal{A}$ is denoted $\mathcal{A}^{\mathbb{N}}$. A sequence $\mathbf{u}$ is *eventually periodic* if $\mathbf{u} = vwww\cdots = v(w)^\infty$ for some $v \in \mathcal{A}^*$ and $w \in \mathcal{A}^+$, moreover, $\mathbf{u}$ is *purely periodic* if $\mathbf{u} = www\cdots = w^\infty$. Otherwise $\mathbf{u}$ is *aperiodic*.

A *factor* of $\mathbf{u}$ is a word y such that $y = u_iu_{i+1}u_{i+2}\cdots u_{j-1}$ for some $i, j \in \mathbb{N}$, $i \le j$. The index i is called an *occurrence* of the factor y in $\mathbf{u}$. In particular, if $i = j$, the factor y is the empty word ε and any index i is its occurrence. If $i = 0$, the factor y is a *prefix* of $\mathbf{u}$.

If each factor of $\mathbf{u}$ has infinitely many occurrences in $\mathbf{u}$, the sequence $\mathbf{u}$ is *recurrent*. Moreover, if the distances between two consecutive occurrences are bounded, $\mathbf{u}$ is *uniformly recurrent*.

The *language* $\mathcal{L}(\mathbf{u})$ of the sequence $\mathbf{u}$ is the set of all factors of $\mathbf{u}$. A factor w of $\mathbf{u}$ is *right special* if both words wa and wb are factors of $\mathbf{u}$ for at least two distinct letters $a, b \in \mathcal{A}$. Analogously we define the *left special* factor. The factor is *bispecial* if it is both left and right special. Note that the empty word ε is the bispecial factor if at least two distinct letters occur in $\mathbf{u}$.

The *factor complexity* of a sequence $\mathbf{u}$ is the mapping $\mathcal{C}_{\mathbf{u}} : \mathbb{N} \to \mathbb{N}$ defined by

$$\mathcal{C}_{\mathbf{u}}(n) = \#\{w \in \mathcal{L}(\mathbf{u}) : |w| = n\}\,.$$

A classical result of Hedlund and Morse [25] says that a sequence is eventually periodic if and only if its factor complexity is bounded. The factor complexity of any aperiodic sequence $\mathbf{u}$ satisfies $\mathcal{C}_{\mathbf{u}}(n) \ge n + 1$ for every $n \in \mathbb{N}$.

A *morphism* from a monoid $\mathcal{A}^*$ to a monoid $\mathcal{B}^*$ is a mapping $\psi : \mathcal{A}^* \to \mathcal{B}^*$ such that $\psi(uv) = \psi(u)\psi(v)$ for all $u, v \in \mathcal{A}^*$. In particular, if $\mathcal{A} = \mathcal{B}$, ψ is a morphism over $\mathcal{A}$. The domain of a morphism ψ can be naturally extended to $\mathcal{A}^{\mathbb{N}}$ by

$$\psi(\mathbf{u}) = \psi(u_0u_1u_2\cdots) = \psi(u_0)\psi(u_1)\psi(u_2)\cdots\,.$$

The *matrix* of a morphism ψ over $\mathcal{A}$ with the cardinality $\#\mathcal{A} = d$ is the matrix $M_\psi \in \mathbb{N}^{d\times d}$ defined as $(M_\psi)_{ab} = |\psi(a)|_b$ for all $a, b \in \mathcal{A}$. The Parikh vector of the ψ-image of a word $w \in \mathcal{A}^*$ can be obtained via multiplication by the matrix M_ψ, *i.e.* $V_{\psi(w)} = M_\psi V_w$.

The morphism is *primitive* if there is a positive integer k such that all elements of the matrix M_ψ^k are positive. A *fixed point* of a morphism ψ is a sequence $\mathbf{u}$ such that $\psi(\mathbf{u}) = \mathbf{u}$. It is well known that all fixed points of a primitive morphism have the same language. The sequence $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is *primitive substitutive* if $\mathbf{u} = \sigma(\mathbf{v})$ for a morphism $\sigma : \mathcal{B}^* \to \mathcal{A}^*$ and a sequence $\mathbf{v} \in \mathcal{B}^{\mathbb{N}}$ which is a fixed point of a primitive morphism over $\mathcal{B}$.

2.2. Derived sequences

Consider a prefix w of a recurrent sequence $\mathbf{u}$. Let $i < j$ be two consecutive occurrences of w in $\mathbf{u}$. Then the word $u_iu_{i+1}\cdots u_{j-1}$ is a *return word* to w in $\mathbf{u}$. The set of all return words to w in $\mathbf{u}$ is denoted $\mathcal{R}_{\mathbf{u}}(w)$.

If the sequence $\mathbf{u}$ is uniformly recurrent, the set $\mathcal{R}_{\mathbf{u}}(w)$ is finite for each prefix w, *i.e.* $\mathcal{R}_{\mathbf{u}}(w) = \{r_0, r_1, \ldots, r_{k-1}\}$. Then the sequence $\mathbf{u}$ can be written as a concatenation of these return words:

$$\mathbf{u} = r_{d_0}r_{d_1}r_{d_2}\cdots$$

and the *derived sequence* of $\mathbf{u}$ to the prefix w is the sequence $\mathbf{d_u}(w) = d_0d_1d_2\cdots$ over the alphabet of cardinality $\#\mathcal{R}_{\mathbf{u}}(w) = k$. For simplicity, we do not fix this alphabet and we consider two derived sequences which differ only in a permutation of letters as identical. The set of all derived sequences to the prefixes of $\mathbf{u}$ is

$$\mathrm{Der}(\mathbf{u}) = \{\mathbf{d_u}(w) : w \text{ is a prefix of } \mathbf{u}\}\,.$$

If the prefix w is not right special, there is a unique letter a such that wa is a factor of $\mathbf{u}$. It means that the occurrences of factors w and wa in $\mathbf{u}$ coincides, thus $\mathcal{R}_\mathbf{u}(w) = \mathcal{R}_\mathbf{u}(wa)$ and $\mathbf{d_u}(w) = \mathbf{d_u}(wa)$. If $\mathbf{u}$ is aperiodic, then any prefix of $\mathbf{u}$ is a prefix of some right special prefix of $\mathbf{u}$. Therefore, for an aperiodic uniformly recurrent sequence $\mathbf{u}$ we can take into consideration only right special prefixes since

$$\mathrm{Der}(\mathbf{u}) = \{\mathbf{d_u}(w) : w \text{ is a right special prefix of } \mathbf{u}\}\,. \tag{2.1}$$

In the sequel we will essentially use the following Durand's result.

Theorem 2.1 (Durand [16])**.** *A sequence* $\mathbf{u}$ *is substitutive primitive if and only if the set* $\mathrm{Der}(\mathbf{u})$ *is finite.*

2.3. Sturmian sequences

Sturmian sequences are aperiodic sequences with the lowest possible factor complexity. In other words, a sequence $\mathbf{u}$ is *Sturmian* if it has its factor complexity $\mathcal{C}_\mathbf{u}(n) = n + 1$ for all $n \in \mathbb{N}$. Clearly, all Sturmian sequences are defined over a binary alphabet.

There are many equivalent definitions of Sturmian sequences, see for example [3, 5, 6]. One of the most important characterizations of Sturmian sequences comes from the symbolic dynamics: any Sturmian sequence can be obtained by a coding of two interval exchange transformation. Here we recall only the basic facts about this transformation, a detailed explanation can be found in [24].

For a given parameter $\alpha \in (0, 1)$, consider the partition of the interval $I = [0, 1)$ into $I_0 = [0, \alpha)$ and $I_1 = [\alpha, 1)$ or the partition of $I = (0, 1]$ into $I_0 = (0, \alpha]$ and $I_1 = (\alpha, 1]$. Then the *two interval exchange transformation* $T : I \to I$ is defined by

$$T(y) = \begin{cases} y + 1 - \alpha & \text{if } y \in I_0\,, \\ y - \alpha & \text{if } y \in I_1\,. \end{cases}$$

If we take an initial point $\rho \in I$, the sequence $\mathbf{u} = u_0u_1u_2\cdots \in \{0, 1\}^\mathbb{N}$ defined by

$$u_n = \begin{cases} 0 & \text{if } T^n(\rho) \in I_0\,, \\ 1 & \text{if } T^n(\rho) \in \ I_1 \end{cases}$$

is a *2iet sequence* with the *slope* α and the *intercept* ρ. It is well known that the set of all 2iet sequences with irrational slopes coincides with the set of all Sturmian sequences.

Any Sturmian sequence is uniformly recurrent. The language of a Sturmian sequence is independent of its intercept ρ, *i.e.* it depends only on its slope α. The frequencies of the letters 0 and 1 in a Sturmian sequence with the slope α are α and $1 - \alpha$, respectively. In the case that $\alpha > \frac{1}{2}$, the form of the transformation T implies that two consecutive occurrences of the letter 1 are separated by the block 0^k or 0^{k+1}, where $k = \lfloor \frac{\alpha}{1-\alpha} \rfloor$. Similarly, if $\alpha < \frac{1}{2}$, two 0's are separated by the block 1^k or 1^{k+1}, where $k = \lfloor \frac{1-\alpha}{\alpha} \rfloor$.

Among all Sturmian sequences with a given slope α, the sequence with the intercept $\rho = 1 - \alpha$ plays a special role. Such a sequence is called a *standard Sturmian sequence* and it is usually denoted by $\mathbf{c}_\alpha$. Any prefix of $\mathbf{c}_\alpha$ is a left special factor. In other words, a Sturmian sequence $\mathbf{u} \in \{0, 1\}^\mathbb{N}$ is standard if both sequences $0\mathbf{u}$, $1\mathbf{u}$ are Sturmian. In particular, it means that

– if $\alpha > \frac{1}{2}$, then $\mathbf{c}_\alpha$ has a prefix 0^k1 and $\mathbf{c}_\alpha$ can be uniquely written as a concatenation of the blocks 0^k1 and $0^{k+1}1$;

– if $\alpha < \frac{1}{2}$, then $\mathbf{c}_\alpha$ has a prefix 1^k0 and $\mathbf{c}_\alpha$ can be uniquely written as a concatenation of the blocks 1^k0 and $1^{k+1}0$.

Moreover, a factor of $\mathbf{c}_\alpha$ is bispecial if and only if it is a palindromic prefix of $\mathbf{c}_\alpha$.

In the context of derived sequences the most important characterization of Sturmian sequences is provided by the third author in [31]: a given sequence $\mathbf{u}$ is Sturmian if and only if any prefix of $\mathbf{u}$ has exactly two return words.

Let us denote the return words to a prefix w of a Sturmian sequence $\mathbf{u}$ as $\mathcal{R}_\mathbf{u}(w) = \{r, s\}$. Then the derived sequence $\mathbf{d}_\mathbf{u}(w)$ of $\mathbf{u}$ to the prefix w can be considered over the alphabet $\{r, s\}$, *i.e.* $\mathbf{d}_\mathbf{u}(w) \in \{r, s\}^\mathbb{N}$. As follows from Vuillon's [31] and Durand's result [16], the derived sequence $\mathbf{d}_\mathbf{u}(w)$ of a Sturmian sequence $\mathbf{u}$ is also a Sturmian sequence. Moreover, if $\mathbf{u}$ is standard, then both sequences $0\mathbf{u}$ and $1\mathbf{u}$ are Sturmian. It implies that $r\mathbf{d}_\mathbf{u}(w)$ and $s\mathbf{d}_\mathbf{u}(w)$ are Sturmian as well. We can conclude that the derived sequence to any prefix of a standard Sturmian sequence is a standard Sturmian sequence. It also means that $\mathbf{d}_\mathbf{u}(w) \in \{r, s\}^\mathbb{N}$ can be decomposed into blocks $r^k s$ and $r^{k+1}s$, where k is a positive integer and r is the most frequent return word. We will strictly use this notation through the whole paper.

In [1], Araújo and Bruyère described derived sequences of any standard Sturmian sequence $\mathbf{u}$. Their description uses the continued fraction of the slope α of $\mathbf{u}$. Derived sequences of all Sturmian sequences are studied in [23]. In the sequel, we will work only with standard Sturmian sequences since especially in this case the elements of the set $\mathrm{Der}(\mathbf{u})$ are easily expressible. In accordance with a wording provided in [23], the S-adic representation of $\mathbf{u}$ by a sequence of Sturmian morphisms will be used for expression of the set $\mathrm{Der}(\mathbf{u})$.

2.4. Sturmian morphisms

A morphism $\psi : \{0,1\}^* \to \{0,1\}^*$ is *Sturmian* if $\psi(\mathbf{u})$ is a Sturmian sequence for any Sturmian sequence $\mathbf{u}$. The set of all Sturmian morphisms together with the operation of composition form the so-called Sturmian monoid St. This monoid is generated by two morphisms E and F, where E is the morphism which exchanges letters, *i.e.* $E : 0 \to 1,\ 1 \to 0$, and F is the Fibonacci morphism, *i.e.* $F : 0 \to 01,\ 1 \to 0$. In the sequel, we work with the submonoid of St which is generated by two elementary morphisms φ_b and φ_β defined by

$$\varphi_b = F \circ E : \quad \begin{cases} 0 \to 0 \\ 1 \to 01 \end{cases} \qquad \text{and} \qquad \varphi_\beta = E \circ F : \quad \begin{cases} 0 \to 10 \\ 1 \to 1 \end{cases}.$$

Their corresponding matrices are:

$$M_b = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad M_\beta = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

The image of a standard Sturmian sequence under φ_b or φ_β is a standard Sturmian sequence as well. Therefore, any element of the submonoid $\langle \varphi_b, \varphi_\beta \rangle$ preserves the set of standard Sturmian sequences. For some $z = z_0 z_1 \cdots z_{n-1} \in \{b, \beta\}^+$, the composition of the morphisms $\varphi_{z_0}, \varphi_{z_1}, \varphi_{z_2}, \ldots, \varphi_{z_{n-1}}$ will be denoted by $\varphi_z = \varphi_{z_0}\varphi_{z_1} \cdots \varphi_{z_{n-1}}$. Let us stress that the morphism φ_z is primitive if and only if z contains both letters b and β. By φ_ε we denote the identity morphism.

Lemma 2.2. *For every standard Sturmian sequence* $\mathbf{u}$ *there is a uniquely given standard Sturmian sequence* $\mathbf{u}'$ *such that* $\mathbf{u} = \varphi_b(\mathbf{u}')$ *or* $\mathbf{u} = \varphi_\beta(\mathbf{u}')$.

Proof. Let us suppose that the letter 0 is more frequent in $\mathbf{u}$ (the second case can be proved analogously). Since $\mathbf{u}$ is a standard Sturmian sequence, it can be written as a concatenation of blocks 0^k1 and $0^{k+1}1$ for some integer $k \geq 1$. Thus $\mathbf{u}$ can be uniquely desubstituted by $0 \to 0$ and $01 \to 1$ to the standard Sturmian sequence $\mathbf{u}'$ which is a concatenation of blocks $0^{k-1}1$ and 0^k1. Therefore $\mathbf{u} = \varphi_b(\mathbf{u}')$. □

By the previous lemma, to a given standard Sturmian sequence $\mathbf{u}$ we can uniquely assign the pair: the *directive sequence* $\mathbf{z} = z_0 z_1 \cdots \in \{b, \beta\}^{\mathbb{N}}$ and the sequence $(\mathbf{u}^{(n)})_{n\geq 0}$, such that

$$\mathbf{u}^{(n)} \in \{0,1\}^{\mathbb{N}} \quad \text{is a standard Sturmian sequence and} \quad \mathbf{u} = \varphi_{z_0 z_1 \ldots z_{n-1}}(\mathbf{u}^{(n)}) \ \text{ for every } \ n \in \mathbb{N}\,.$$

In fact, a sequence $\mathbf{z} \in \{b, \beta\}^{\mathbb{N}}$ containing infinitely many occurrences of both letters already determines a unique standard Sturmian sequence $\mathbf{u}$, as

$$\mathbf{u} = \lim_{n\to\infty} \varphi_{z_0 z_1 \ldots z_{n-1}}(0) = \lim_{n\to\infty} \varphi_{z_0 z_1 \ldots z_{n-1}}(1)\,.$$

Now we can formulate several simple consequences of Lemma 2.2.

Observation 2.3. *Let $\mathbf{u}$ be a standard Sturmian sequence with the directive sequence $\mathbf{z} \in \{b, \beta\}^{\mathbb{N}}$.*

(i) The sequence $\mathbf{z}$ contains infinitely many letters b and infinitely many letters β.
(ii) If $\mathbf{z}$ has a prefix $b^k\beta$ for some positive integer k, then the letter 0 is more frequent in $\mathbf{u}$ and $\mathbf{u}$ can be written as a concatenation of blocks 0^k1 and $0^{k+1}1$.
(iii) If $\mathbf{z}$ has a prefix $\beta^k b$ for some positive integer k, then the letter 1 is more frequent in $\mathbf{u}$ and $\mathbf{u}$ can be written as a concatenation of blocks 1^k0 and $1^{k+1}0$.
(iv) The directive sequence $\mathbf{z}$ is eventually periodic if and only if the sequence $\mathbf{u}$ is substitutive. Moreover, $\mathbf{z}$ is purely periodic, i.e. $\mathbf{z} = z^\infty$ for some $z \in \{b,\beta\}^+$, if and only if $\mathbf{u}$ is a fixed point of the morphism φ_z.

2.5. Complementary symmetric Rote sequences

A *Rote sequence* is a sequence $\mathbf{v}$ with the factor complexity $\mathcal{C}_{\mathbf{v}}(n) = 2n$ for all integer $n \geq 1$. Clearly, all Rote sequences are defined over a binary alphabet, *e.g.* $\{0,1\}$. If the language of a Rote sequence $\mathbf{v}$ is closed under the exchange of letters, *i.e.* $E(v) \in \mathcal{L}(\mathbf{v})$ for each $v \in \mathcal{L}(\mathbf{v})$, the Rote sequence $\mathbf{v}$ is called *complementary symmetric*. Rote in [30] proved that these sequences are essentially connected with Sturmian sequences:

Proposition 2.4 (Rote [30])**.** *Let $\mathbf{u} = u_0u_1\cdots$ and $\mathbf{v} = v_0v_1\cdots$ be two sequences over $\{0,1\}$ such that $u_i = v_i + v_{i+1} \mod 2$ for all $i \in \mathbb{N}$. Then $\mathbf{v}$ is a complementary symmetric Rote sequence if and only if $\mathbf{u}$ is a Sturmian sequence.*

Convention. In this paper, we work only with complementary symmetric Rote sequences and for simplicity we usually call them shortly Rote sequences.

As indicated by Proposition 2.4, it will be useful to introduce the following notation.

Definition 2.5. By $\mathcal{S}$ we denote the mapping $\mathcal{S} : \{0,1\}^+ \to \{0,1\}^*$ such that for every $v_0 \in \{0,1\}$ we put $\mathcal{S}(v_0) = \varepsilon$ and for every $v = v_0v_1\cdots v_n \in \{0,1\}^+$ of length at least 2 we put $\mathcal{S}(v_0v_1\cdots v_n) = u_0u_1\cdots u_{n-1}$, where

$$u_i = v_i + v_{i+1} \mod 2 \quad \text{for all} \ \ i \in \{0,1,\ldots,n-1\}\,.$$

Example 2.6. Let $v = 001110$. Then $\mathcal{S}(v) = \mathcal{S}(E(v)) = 01001$. Clearly, the images of v and $E(v)$ under $\mathcal{S}$ coincide for each $v \in \{0,1\}^+$. Moreover, $\mathcal{S}(x) = \mathcal{S}(y)$ if and only if $x = y$ or $x = E(y)$.

If we extend the domain of $\mathcal{S}$ naturally to $\{0,1\}^{\mathbb{N}}$, Proposition 2.4 says: $\mathbf{v}$ is a Rote sequence if and only if $\mathcal{S}(\mathbf{v})$ is a Sturmian sequence. Moreover, for any Sturmian sequence $\mathbf{u}$ there exist two Rote sequences $\mathbf{v}$ and $E(\mathbf{v})$ such that $\mathbf{u} = \mathcal{S}(\mathbf{v}) = \mathcal{S}(E(\mathbf{v}))$. Since a permutation of letters in the sequence does not influence its derived sequences, we will work only with Rote sequences starting with the letter 0 without lose of generality. We will also use the bar notation $\bar{\mathbf{v}} = E(\mathbf{v})$ or $\bar{v} = E(v)$ to express the sequence or the word with exchanged letters $0 \leftrightarrow 1$.

Convention. We consider only Rote sequences $\mathbf{v} \in \{0,1\}^{\mathbb{N}}$ with the prefix 0. If a Sturmian sequence $\mathbf{u} \in \{0,1\}^{\mathbb{N}}$ satisfies $\mathbf{u} = \mathcal{S}(\mathbf{v})$, we say that $\mathbf{v}$ is *associated* with $\mathbf{u}$ or equivalently $\mathbf{u}$ is *associated* with $\mathbf{v}$.

To a given word $u \in \{0,1\}^*$ there are exactly two words $v, \bar{v}$ such that $S(v) = S(\bar{v}) = u$. Moreover, if the first letter of v is given, then the rest of the word $v = v_0 \cdots v_n$ is completely determined by $u = u_0 \cdots u_{n-1}$:

$$v_{i+1} = v_0 + u_0 + u_1 + \cdots + u_i \mod 2 \quad \text{for all } i \in \{0,1,\ldots,n-1\}\,. \tag{2.2}$$

Lemma 2.7. *Let $\mathbf{u}$ be a Sturmian sequence associated with a Rote sequence $\mathbf{v}$. A word u is a factor of $\mathbf{u}$ if and only if both words $v, \bar{v}$ such that $u = \mathcal{S}(v) = \mathcal{S}(\bar{v})$ are factors of $\mathbf{v}$. Moreover, for every $m \in \mathbb{N}$, the index m is an occurrence of u in $\mathbf{u}$ if and only if m is an occurrence of v in $\mathbf{v}$ or an occurrence of $\bar{v}$ in $\mathbf{v}$.*

Bispecial factors of a sequence $\mathbf{u}$ play a crucial role in finding its derived sequences. We use the terminology introduced by Cassaigne [12] to distinguish three types of bispecial factors. Let w be a bispecial factor of $\mathbf{u}$. Then the *bilateral order* of w is the number

$$B(w) = \#\{(a,b) \in \mathcal{A} \times \mathcal{A} : awb \in \mathcal{L}(\mathbf{u})\} - \#\{a \in \mathcal{A} : aw \in \mathcal{L}(\mathbf{u})\} - \#\{b \in \mathcal{A} : wb \in \mathcal{L}(\mathbf{u})\} + 1\,.$$

The bispecial factor w is *weak* if $B(w) < 0$, it is *ordinary* if $B(w) = 0$ and it is *strong* if $B(w) > 0$.

Corollary 2.8. *Let $\mathbf{u}$ be a Sturmian sequence associated with a Rote sequence $\mathbf{v}$, let $\ell \in \mathbb{N}$. If w is a bispecial factor of length ℓ in $\mathbf{u}$, then there are two bispecial factors $x, \bar{x}$ of length $\ell+1$ in $\mathbf{v}$ such that $w = \mathcal{S}(x) = \mathcal{S}(\bar{x})$. Conversely, if x is a bispecial factor of length $\ell+1$ in $\mathbf{v}$, then $\mathcal{S}(x)$ of length ℓ is a bispecial factor in $\mathbf{u}$. Moreover, each non-empty bispecial factor of $\mathbf{v}$ is ordinary and the empty word is a strong bispecial factor of $\mathbf{v}$.*

Proof. Let w be a bispecial factor of $\mathbf{u}$. By the well known balance properties of Sturmian sequences, the bispecial factor w is ordinary. Indeed, the words $0w1, 1w0$ are always factors of $\mathbf{u}$, in addition, just one word from $\{1w1, 0w0\}$ is a factor of $\mathbf{u}$. Without lose of generality let us suppose that $1w1 \in \mathcal{L}(\mathbf{u})$. The associated factors of the Rote sequence $\mathbf{v}$ are $0x\bar{a}$, $1\bar{x}a$, $0\bar{x}\bar{a}$, $1xa$, $0\bar{x}a$ and $1x\bar{a}$, where $w = \mathcal{S}(x)$ and x starts with 0 and ends with a. Combining with Lemma 2.7 we get that both words $x, \bar{x}$ are ordinary bispecial factors of $\mathbf{v}$.

Conversely, let us suppose that x is a non-empty bispecial factor of $\mathbf{v}$. It means that the words $0x$, $1x$, $x0$, $x1$ are factors of $\mathbf{v}$. Then $\mathcal{S}(0x) = aw$, $\mathcal{S}(1x) = \bar{a}w$, $\mathcal{S}(x0) = wb$, $\mathcal{S}(x1) = w\bar{b}$, where $w = \mathcal{S}(x)$, a is the first letter of x and b is the last letter of x. Thus w is a bispecial factor of $\mathbf{u}$.

Since $00, 11, 01, 10 \in \mathcal{L}(\mathbf{v})$, the bilateral order of ε is 1, *i.e.* ε is strong. □

3. Return words to prefixes of complementary symmetric Rote sequences

Complementary symmetric Rote sequences form a special subclass of binary sequences coding the rotations. The return words in the sequences coding the rotations were studied in [10] in particular for palindromic factors. To compute the exact number of return words to a factor of a given Rote sequence, we use the following results from [4] (Lems. 4.2 and 4.4):

(i) If $\mathbf{v}$ is uniformly recurrent sequence with no weak bispecial factor, then $\#\mathcal{R}_{\mathbf{v}}(x) \geq 1 + \Delta\mathcal{C}_{\mathbf{v}}(|x|)$ for every factor $x \in \mathcal{L}(\mathbf{v})$.

(ii) If $\mathbf{v}$ has no weak bispecial factor and $\Delta\mathcal{C}_{\mathbf{v}}(n) < m$ for all $n \geq 0$, then $\#\mathcal{R}_{\mathbf{v}}(w) \leq m$ for every factor $w \in \mathcal{L}(\mathbf{v})$.

Recall that $\Delta\mathcal{C}_{\mathbf{v}}$ denotes the first difference of the factor complexity $\mathcal{C}_{\mathbf{v}}$, *i.e.* $\Delta\mathcal{C}_{\mathbf{v}}(n) = \mathcal{C}_{\mathbf{v}}(n+1) - \mathcal{C}_{\mathbf{v}}(n)$ for each $n \in \mathbb{N}$.

Theorem 3.1. *Let $\mathbf{v}$ be a Rote sequence. Then every non-empty prefix x of $\mathbf{v}$ has exactly three distinct return words.*

Proof. By Corollary 2.8, no bispecial factor of a Rote sequence $\mathbf{v}$ is weak. Every Rote sequence is uniformly recurrent and for all $n \geq 1$ it holds true $\Delta\mathcal{C}_{\mathbf{v}}(n) = 2$. Thus by Lemma 4.2 from [4], we have $\#\mathcal{R}_{\mathbf{v}}(x) \geq 3$ for every non-empty prefix x of $\mathbf{v}$.

On the other hand, since $\Delta\mathcal{C}_{\mathbf{v}}(n) < 3$ for all $n \geq 0$, by Lemma 4.4 from [4] we have $\#\mathcal{R}_{\mathbf{v}}(x) \leq 3$ for every prefix x of $\mathbf{v}$. Therefore, $\#\mathcal{R}_{\mathbf{v}}(x) = 3$ for every non-empty prefix x of $\mathbf{v}$. □

Remark 3.2. The previous theorem also follows from a more general result obtained by Dolce and Perrin in [15]. They studied the so-called neutral sets. By our Corollary 2.8, the language $\mathcal{L}$ of a Rote sequence is a neutral set with the characteristic $\chi(\mathcal{L}) = 0$. As the language $\mathcal{L}$ is uniformly recurrent, we can apply Corollary 5.4 of [15] to deduce that any non-empty factor of a Rote sequence has exactly three return words.

A direct consequence of Theorem 3.1 is that all derived sequences of a Rote sequence to its non-empty prefixes are over a ternary alphabet. However, to study derived sequences we need to know also the structure of return words, not only their number.

For this purpose we now describe the crucial relation between return words of Sturmian and Rote sequences. Suppose that $\mathbf{v}$ is a Rote sequence with a prefix x. Then by Proposition 2.4 and Lemma 2.7, $\mathbf{u} = \mathcal{S}(\mathbf{v})$ is a Sturmian sequence, $w = \mathcal{S}(x)$ is a prefix of $\mathbf{u}$ and the occurrences of w in $\mathbf{u}$ coincide with the occurrences of x and $\bar{x}$ in $\mathbf{v}$. Let r, s be two return words to w in $\mathbf{u}$, r is the most frequent one. Our aim is to find three return words to x in $\mathbf{v}$. We start with an example.

Example 3.3. Consider the Sturmian sequence $\mathbf{u} = u_0u_1\cdots$ which is fixed by the Sturmian morphism $\psi : 0 \to 010,\ 1 \to 01001$, *i.e.*

$$\mathbf{u} = 010010010100100100101001001001010010010 10\cdots .$$

The associated Rote sequence $\mathbf{v} = v_0v_1\cdots$ (*i.e.* $\mathbf{u} = \mathcal{S}(\mathbf{v})$) starting with 0 is

$$\mathbf{v} = 0011100011000111000110001110001100011100 11\cdots .$$

Take the prefix $w = 0$ of $\mathbf{u}$. It has two return words $r = 01$, $s = 0$ and the occurrences of w in $\mathbf{u}$ are $0, 2, 3, 5, 6, 8, 10, 11, \ldots$ The associated prefix of $\mathbf{v}$ is $x = 00$ since $0 = \mathcal{S}(00)$. As we know from Lemma 2.7, the occurrences of $w = 0$ in $\mathbf{u}$ correspond to the occurrences of $x = 00$ and $\bar{x} = 11$ in $\mathbf{v}$. To find the return words to $x = 00$ we have to determine precisely when the words 00 and 11 occur in $\mathbf{v}$.

Clearly, there is the factor 00 at the position 0, *i.e.* $v_0v_1 = 00$. Which word from $\{00, 11\}$ starts at the position 2 depends only on the letter v_2, see equation (2.2). This letter is completely determined by the prefix of $\mathbf{u}$ of length 2, which is $u_0u_1 = 01$ (this is also the first return word to 0 in $\mathbf{u}$). Indeed, $v_2 = v_0 + u_0 + u_1 \mod 2$. Since $v_2 = 0 + 0 + 1 = 1$, there is the factor 11 starting at position 2, *i.e.* $v_2v_3 = 11$. In other words, the return word 01 causes the alternation of the factors x and $\bar{x}$, since it has an odd number of 1's.

To determine the factor v_3v_4 starting at position 3 we have to compute the letter $v_3 = v_0 + u_0 + u_1 + u_2 = v_2 + u_2 \mod 2$. Since $v_2 = 0$, we get $v_3 = 1$ and $v_3v_4 = 11$. Notice that the word u_2 is the second return word to 0 in $\mathbf{u}$. Since u_2 has an even number of 1's, it leaves the factors x, $\bar{x}$ unchanged. In the next step we get $v_5v_6 = 00$, since $v_5 = v_3 + u_3 + u_4 = 1 + 0 + 1 = 0 \mod 2$. So we find the first return word to 00 in $\mathbf{v}$, it is the word $v_0v_1v_2v_3v_4 = 00111$.

Similarly we get $v_6 = v_5 + u_5 = 0 + 0 = 0$ and thus $v_6v_7 = 00$, so the next return word to 00 in $\mathbf{v}$ is the word $v_5 = 0$.

As $v_8 = v_6 + u_6 + u_7 = 0 + 0 + 1 = 1$, it holds true $v_8v_9 = 11$. So we have to wait until another factor 01 appears in $\mathbf{u}$. It happens immediately since $u_8u_9 = 01$. Thus $v_{10} = v_8 + u_8 + u_9 = 0 \mod 2$ and $v_{10}v_{11} = 00$. Therefore the word $v_6v_7v_8v_9 = 0011$ is the last return word to 00 in $\mathbf{v}$.

In total, the prefix $x = 00$ of $\mathbf{v}$ has three return words 0, 0011 and 00111.

As we have seen in Example 3.3, to describe the return words to x, we have to distinguish if a given return word to w causes the alternation of the factors x, $\bar{x}$ or not. This is the meaning of the following definition.

Definition 3.4. A word $u = u_0u_1 \cdots u_{n-1} \in \{0,1\}^*$ is called *stable* (S) if $|u|_1 = 0 \mod 2$. Otherwise, u is *unstable* (U).

Example 3.5. The word $u = 0110101$ is stable while the word $v = 011010$ is unstable.

Remark 3.6. In the notion of Parikh vectors, the factor u is stable if its Parikh vector $V_u = \binom{p}{0} \mod 2$ and it is unstable if $V_u = \binom{p}{1} \mod 2$ for some number $p \in \{0,1\}$.

Lemma 3.7. *Let* $\mathbf{v}$ *be a Rote sequence and let* x *be its prefix. Denote* $\mathbf{u} = \mathcal{S}(\mathbf{v})$ *and* $w = \mathcal{S}(x)$*. An index* m *is an occurrence of* x *in* $\mathbf{v}$ *if and only if* m *is an occurrence of* w *in* $\mathbf{u}$ *and the prefix* $u = u_0u_1 \cdots u_{m-1}$ *of* $\mathbf{u}$ *is stable.*

Proof. Recall that $u_i = v_{i+1} + v_i \mod 2$ holds true for all $i \in \mathbb{N}$. By summing up mod 2 we get for the prefix $u = u_0u_1 \cdots u_{m-1}$ of $\mathbf{u}$:

$$|u|_1 = \sum_{i=0}^{m-1} u_i = \sum_{i=0}^{m-1} (v_{i+1} + v_i) = v_m + v_0 \mod 2. \tag{3.1}$$

By Lemma 2.7, m is an occurrence of the prefix x in $\mathbf{v}$ if and only if m is an occurrence of w in $\mathbf{u}$ and the letter v_m coincides with v_0, which is the first letter of x. The equation (3.1) says that the letters v_0 and v_m coincide if and only if the prefix of $\mathbf{u}$ of length m is stable. □

We have seen that the form of return words in a Rote sequence depends on the stability of the return words in the associated Sturmian sequence. The following definition sorts the prefixes of standard Sturmian sequences according to the stability of their return words.

Definition 3.8. Let w be a prefix of a standard Sturmian sequence $\mathbf{u}$ with return words $\mathcal{R}_{\mathbf{u}}(w) = \{r, s\}$, where r is the most frequent return word. Let k be a positive integer such that $\mathbf{u}$ is a concatenation of blocks $r^k s$ and $r^{k+1}s$. We distinguish three cases:

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

The type of the prefix w is denoted $\mathcal{T}_w$ (or $\mathcal{T}$ if the respective factor w is clear). If the number k is not essential, we write only SU, US and UU.

Remark 3.9. It is easy to verify that all these types appear in the case of prefixes of Sturmian sequences. On the other hand, the fourth possible case, *i.e.* the type SS, cannot appear. We can prove this using the results from [2]. It also follows from the proof of Theorem 4 in [30].

First we recall the WELLDOC property. A sequence $\mathbf{u} \in \{0,1\}^{\mathbb{N}}$ has *well distributed occurrences modulo 2* (shortly WELLDOC(2) property) if for every factor $w \in \mathcal{L}(\mathbf{u})$ we have

$$\left\{ \begin{pmatrix} |u|_0 \\ |u|_1 \end{pmatrix} \mod 2 : uw \text{ is a prefix of } \mathbf{u} \right\} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}.$$

As shown in [2], all Sturmian sequences have the WELLDOC(2) property.

Let us suppose that w is a prefix of $\mathbf{u}$ with two stable return words, *i.e.* the numbers of 1's occurring in r and s are even. Since any word u such that uw is a prefix of $\mathbf{u}$ is a concatenation of words r and s, u contains an even number of 1's. It contradicts the WELLDOC(2) property of $\mathbf{u}$.

We use these prefix types to describe the return words to corresponding Rote prefixes.

Theorem 3.10. *Let* $\mathbf{v}$ *be a Rote sequence associated with a standard Sturmian sequence* $\mathbf{u} = \mathcal{S}(\mathbf{v})$*. Let* x *be a non-empty prefix of* $\mathbf{v}$ *and* $w = \mathcal{S}(x)$*. Then the prefix* x *of* $\mathbf{v}$ *has three return words* $A, B, C \in \{0,1\}^+$ *satisfying (*r*,* s *and* k *are the same as in Def. 3.8):*

(i) if $\mathcal{T}_w = SU(k)$*, then* $\mathcal{S}(A0) = r$*,* $\mathcal{S}(B0) = sr^{k+1}s$ *and* $\mathcal{S}(C0) = sr^k s$*;*
(ii) if $\mathcal{T}_w = US(k)$*, then* $\mathcal{S}(A0) = rr$*,* $\mathcal{S}(B0) = rsr$ *and* $\mathcal{S}(C0) = s$*;*
(iii) if $\mathcal{T}_w = UU(k)$*, then* $\mathcal{S}(A0) = rr$*,* $\mathcal{S}(B0) = rs$ *and* $\mathcal{S}(C0) = sr$*.*

Proof. Let us suppose that $\mathcal{T}_w = SU(k)$, *i.e.* $|r|_1 = 0 \mod 2$ and $|s|_1 = 1 \mod 2$. Let n be an occurrence of x in $\mathbf{v}$. Then by Lemma 3.7 the index n is an occurrence of w in $\mathbf{u}$ and the prefix $u = u_0u_1 \cdots u_{n-1}$ is stable. Since $\mathbf{u}$ is a concatenation of the blocks $r^{k+1}s$ and $r^k s$, the sequence $\mathbf{u}$ has one of the prefixes ur, $usr^{k+1}s$ or $usr^k s$.

– If ur is a prefix of $\mathbf{u}$, then $n + |r|$ is an occurrence of w in $\mathbf{u}$. Moreover, the prefix of $\mathbf{u}$ of length $n + |r|$ is stable. It means that $m := n + |r|$ is the subsequent occurrence of x in $\mathbf{v}$ and $A := v_n v_{n+1} \cdots v_{m-1}$ is a return word to x in $\mathbf{v}$. Let us recall our convention that 0 is a prefix of $\mathbf{v}$ and thus any return word to the prefix x begins with 0, in particular $v_m = 0$. Therefore, $r = u_n u_{n+1} \cdots u_{m-1} = \mathcal{S}(A0)$.

– If $usr^{k+1}s$ is a prefix of $\mathbf{u}$, then any index $\ell \in \{n+|s|, n+|s|+|r|, n+|s|+2|r|, \cdots, n+|s|+(k+1)|r|\}$ is an occurrence of w in $\mathbf{u}$. Since r is stable and s is unstable, prefixes of these lengths ℓ are unstable and by Lemma 3.7, such a index ℓ is not an occurrence of x in $\mathbf{v}$. The next occurrence of w in $\mathbf{u}$ is $m := n + |s| + (k+1)|r| + |s|$. The prefix of $\mathbf{u}$ of length m is stable and thus m is the smallest occurrence of x in $\mathbf{v}$ grater than n. Therefore $B := v_n \cdots v_{m-1}$ is a return word to x in $\mathbf{v}$ and obviously $sr^{k+1}s = \mathcal{S}(B0)$.

The reasoning in all remaining cases is analogous and so we omit it. □

Example 3.11 (Example 3.3 continued)**.** Recall that the prefix 00 of $\mathbf{v}$ has three return words $A = 0$, $B = 0011$ and $C = 00111$. The associated Sturmian prefix $S(00) = 0$ has the return words $r = 01$, $s = 0$ and $\mathbf{u}$ is a concatenation of blocks $rs = 010$ and $rrs = 01010$. Thus the type of 0 is $\mathcal{T}_0 = US(1)$. It holds true $\mathcal{S}(A0) = \mathcal{S}(00) = 0 = s$, $\mathcal{S}(B0) = \mathcal{S}(00110) = 0101 = rr$ and $\mathcal{S}(C0) = \mathcal{S}(001110) = 01001 = rsr$.

It remains to explain how to determine the type of a given prefix w of a standard Sturmian sequence $\mathbf{u}$. This question will be solved in Section 5.

4. Derived sequences of complementary symmetric Rote sequences

As we have proved in Theorem 3.1, any derived sequence of a Rote sequence $\mathbf{v}$ is over a ternary alphabet (we use the alphabet $\{A, B, C\}$). In this section we study the structure of these ternary sequences in the case that $\mathbf{v}$ is associated with a standard Sturmian sequence. First we mention an important direct consequence of Theorem 3.10.

Corollary 4.1. *Let* $\mathbf{v}$ *be a Rote sequence associated with a standard Sturmian sequence* $\mathbf{u} = \mathcal{S}(\mathbf{v})$ *and let* x *be a non-empty prefix of* $\mathbf{v}$*. Then the derived sequence* $\mathbf{d_v}(x)$ *is uniquely determined by the derived sequence* $\mathbf{d_u}(w)$ *of* $\mathbf{u}$ *to the prefix* $w = \mathcal{S}(x)$ *and by the type* $\mathcal{T}_w$ *of the prefix* w*.*

Proof. Let r, s be the return words to w in $\mathbf{u}$ and let $\mathbf{u}$ be a concatenation of blocks $r^k s$ and $r^{k+1}s$ for some positive integer k. We decompose the sequence $\mathbf{d_u}(w) \in \{r, s\}^{\mathbb{N}}$ from the left to the right into three types of blocks $\mathcal{S}(A0), \mathcal{S}(B0)$ and $\mathcal{S}(C0)$ according to the type $\mathcal{T}_w$ (the relevant blocks are listed in Thm. 3.10). Then the order of letters A, B, C in this decomposition is the desired derived sequence $\mathbf{d_v}(x)$ of $\mathbf{v}$ to x. It remains to explain that this decomposition is unique. In the case i) we decompose $\mathbf{d_u}(w)$ into the minimal blocks with an even number of letter s, similarly in the case ii) we decompose $\mathbf{d_u}(w)$ into the minimal blocks with an even number of letter r. In the case iii) we decompose $\mathbf{d_u}(w)$ into the pairs of letters. □

The main goal of this section is to show that any derived sequence $\mathbf{d_v}(x)$ of $\mathbf{v}$ is in fact coding of a three interval exchange transformation. The sequences coding the interval exchange transformation were introduced in [26] and they are intensively studied as they represent an important generalization of Sturmian sequences to the multi-literal alphabets, see [29]. Here we define only those interval exchange transformations which appear in our description of derived sequences $\mathbf{d_v}(x)$.

A three interval exchange transformation $T : [0,1) \to [0,1)$ is given by two parameters $\beta, \gamma \in (0,1)$, $\beta+\gamma < 1$, and by a permutation π on the set $\{1,2,3\}$. The interval $[0,1)$ is partitioned into three subintervals

$$I_A = [0,\beta)\,, \quad I_B = [\beta, \beta+\gamma) \quad \text{and} \quad I_C = [\beta+\gamma, 1)$$

of lengths β, γ and $1-\beta-\gamma$ respectively. These intervals are then rearranged by the transformation T according to the permutation π. More specifically:

– If the permutation $\pi = (3,2,1)$, then

$$T(y) = \begin{cases} y+1-\beta & \text{if } y \in I_A\,, \\ y+1-2\beta-\gamma & \text{if } y \in I_B\,, \\ y-\beta-\gamma & \text{if } y \in I_C\,. \end{cases}$$

– If the permutation $\pi = (2,3,1)$, then

$$T(y) = \begin{cases} y+1-\beta & \text{if } y \in I_A\,, \\ y-\beta & \text{if } y \in I_B\,, \\ y-\beta & \text{if } y \in I_C\,. \end{cases}$$

Let $\rho \in [0,1)$. The sequence $\mathbf{u} = u_0u_1u_2\cdots \in \{A,B,C\}^{\mathbb{N}}$ defined by

$$u_n = \begin{cases} A & \text{if } T^n(\rho) \in I_A\,, \\ B & \text{if } T^n(\rho) \in I_B\,, \\ C & \text{if } T^n(\rho) \in I_C \end{cases}$$

is called a *3iet sequence* coding the intercept ρ under the transformation T.

Take a standard Sturmian sequence $\mathbf{u}$. As we have mentioned before, every derived sequence $\mathbf{d_u}(w)$ of $\mathbf{u}$ to a given prefix w is also a standard Sturmian sequence. Thus $\mathbf{d_u}(w)$ is expressible as a 2iet sequence with the slope α and the intercept $\rho = 1-\alpha$.

Proposition 4.2. *Let $\mathbf{v}$ be a Rote sequence associated with a standard Sturmian sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$, let x be a non-empty prefix of $\mathbf{v}$ and $w = \mathcal{S}(x)$. Let $\alpha > \frac{1}{2}$ be the slope of the Sturmian sequence $\mathbf{d_u}(w)$. Then the derived sequence $\mathbf{d_v}(x)$ is a 3iet sequence coding the intercept $\rho = 1-\alpha$ under the three interval exchange transformation T, where T is given by the following parameters β, γ and permutation π:*

(i) if $\mathcal{T}_w = SU(k)$, then $\beta = \alpha$, $\gamma = \alpha - k(1-\alpha)$, and $\pi = (3,2,1)$;
(ii) if $\mathcal{T}_w = US(k)$, then $\beta = 2\alpha - 1$, $\gamma = 1-\alpha$, and $\pi = (3,2,1)$;
(iii) if $\mathcal{T}_w = UU(k)$, then $\beta = 2\alpha - 1$, $\gamma = 1-\alpha$, and $\pi = (2,3,1)$.

Proof. Since any derived sequence of a standard Sturmian sequence is standard as well, $\mathbf{d_u}(w)$ is coding of the intercept $1-\alpha$ under the transformation $G : [0,1) \to [0,1)$ defined by

$$G(y) = y+1-\alpha, \ \text{ if } \ y \in I_r = [0,\alpha) \quad \text{and} \quad G(y) = y-\alpha, \ \text{ if } \ y \in I_s = [\alpha, 1)\,.$$

Let us start with the simplest case iii): By Theorem 3.10, the derived sequence $\mathbf{d_v}(x)$ of $\mathbf{v}$ to the prefix x is determined by the decomposition of $\mathbf{d_u}(w)$ into blocks of length 2. The order of blocks rr, rs and sr in the decomposition of $\mathbf{d_u}(w)$ is given by the transformation G^2 under which the point $\rho = 1-\alpha$ is coded. A simple computation gives:

$$G^2(y) = \begin{cases} y+2-2\alpha & \text{if} \quad y \in [0, 2\alpha-1)\,, \\ y+1-2\alpha & \text{if} \quad y \in [2\alpha-1, \alpha)\,, \\ y+1-2\alpha & \text{if} \quad y \in [\alpha, 1)\,. \end{cases}$$

It means that G^2 exchanges three intervals under the permutation $(2,3,1)$ with the parameters β, γ as claimed in point iii) of the statement.

Let w be of type $SU(k)$ as assumed in i). Let us denote the intervals

$$I_A = [0,\alpha), \quad I_B = [\alpha, 2\alpha - k(1-\alpha)), \quad I_C = [2\alpha - k(1-\alpha), 1)$$

and define the transformation

$$T(y) = \begin{cases} G(y) & \text{if} \quad y \in I_A\,, \\ G^{k+3}(y) & \text{if} \quad y \in I_B\,, \\ G^{k+2}(y) & \text{if} \quad y \in I_C\,. \end{cases}$$

Recall that the parameter k in the type of w means that $\mathbf{d_u}(w)$ is a concatenation of blocks $r^k s$ and $r^{k+1}s$. By Theorem 3.10, the derived sequence $\mathbf{d_v}(x)$ of $\mathbf{v}$ to the prefix x is determined by the unique decomposition of $\mathbf{d_u}(w)$ into blocks r, $sr^{k+1}s$ and $sr^k s$. As mentioned in Section 2.3, $k = \lfloor \frac{\alpha}{1-\alpha} \rfloor$, *i.e.* $\alpha > k(1-\alpha)$ and $\alpha < (k+1)(1-\alpha)$. Therefore the intervals I_A, I_B, and I_C are well defined.

To prove i), one has to check

(1) $I_A \subset I_r$;
(2) $I_B \subset I_s$, $G^j(I_B) \subset I_r$ for all $j = 1, 2, \ldots, k+1$, $G^{k+2}(I_B) \subset I_s$;
(3) $I_C \subset I_s$, $G^j(I_C) \subset I_r$ for all $j = 1, 2, \ldots, k$, $G^{k+1}(I_C) \subset I_s$;
(4) T is an interval exchange transformation under the permutation (3,2,1), *i.e.*,

$$T(I_A) = \big[1-\alpha, 1\big), \quad T(I_B) = \big[(k+1)(1-\alpha) - \alpha, 1-\alpha\big), \quad T(I_C) = \big[0, (k+1)(1-\alpha) - \alpha\big).$$

Validity of (1)–(4) follows directly from the definition of G.

Proof of point ii) is analogous. □

Remark 4.3. It can be shown that all three transformations T from Proposition 4.2 satisfy the so called *i.d.o.c. property* [22]. For a three interval exchange transformation with the discontinuity points β and $\beta+\gamma$ it means that $T^n(\beta) \neq \beta+\gamma$ for all $n \in \mathbb{Z}$. Property i.d.o.c. implies that the factor complexity of any derived sequence $\mathbf{d_v}(x)$ is $\mathcal{C}(n) = 2n+1$.

Corollary 4.4. *Let $\mathbf{v}$ be a Rote sequence associated with a standard Sturmian sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$ and let x, x' be two non-empty prefixes of $\mathbf{v}$. Denote $w = \mathcal{S}(x)$ and $w' = \mathcal{S}(x')$. The derived sequence of $\mathbf{v}$ to the prefix x coincides with the derived sequence of $\mathbf{v}$ to the prefix x' if and only if the types of w and w' are the same and the derived sequence of $\mathbf{u}$ to the prefix w coincides with the derived sequence of $\mathbf{u}$ to the prefix w'. In other words,*

$$\mathbf{d_v}(x) = \mathbf{d_v}(x') \qquad \textit{iff} \qquad \mathcal{T}_w = \mathcal{T}_{w'} \quad \textit{and} \quad \mathbf{d_u}(w) = \mathbf{d_u}(w')\,.$$

Proof. Let us assume that $\mathbf{d_v}(x) = \mathbf{d_v}(x')$. We use two well known properties of 3iet sequences, see for example [18, 19]:

– the frequencies of letters in a 3iet sequence correspond to the lengths of the intervals I_A, I_B and I_C;
– the language of a 3iet sequence is closed under reversal if and only if the permutation is $(3,2,1)$.

By Proposition 4.2, the language of $\mathbf{d_v}(x)$ is not closed under reversal if and only if $w = \mathcal{S}(x)$ is of type UU. Moreover, if w is of type SU, the frequencies of letters are: α, $\alpha - k(1-\alpha)$ and $(k+1)(1-\alpha) - \alpha$. Since α is irrational, these three lengths are pairwise distinct. If w is of type US or UU, the letters B and C have the same frequency $1-\alpha$. Therefore, the assumption $\mathbf{d_v}(x) = \mathbf{d_v}(x')$ implies that the type of w and the type of w' are the same. Moreover, the lengths of the intervals I_A, I_B, I_C, *i.e.* the frequencies of the letters, must be the same. It implies that the slopes of $\mathbf{d_u}(w)$ and $\mathbf{d_u}(w')$ are equal. Since $\mathbf{d_u}(w)$ and $\mathbf{d_u}(w')$ are both standard Sturmian sequences with the same slope, obviously $\mathbf{d_u}(w) = \mathbf{d_u}(w')$.

The opposite implication follows from Corollary 4.1. □

The proof of Corollary 4.1 gives us the instructions how to construct the derived sequences of a Rote sequence: we need to know both the derived sequences of the associated Sturmian sequence $\mathbf{u}$ and the types of prefixes of $\mathbf{u}$. Remind that we work only with Rote sequences associated with standard Sturmian sequences, thus $\mathbf{u}$ is always standard and any prefix of $\mathbf{u}$ is left special. Due to (2.1) and Corollary 2.8, we can focus only on the bispecial prefixes of standard Sturmian sequences.

5. Types of bispecial prefixes of Sturmian sequences

Consider a standard Sturmian sequence $\mathbf{u}$ with the directive sequence $\mathbf{z} \in \{b, \beta\}^{\mathbb{N}}$. It means that there is a sequence $(\mathbf{u}^{(n)})_{n\geq 0}$ of standard Sturmian sequences such that for every $n \in \mathbb{N}$

$$\mathbf{u} = \varphi_{z_0 z_1 \ldots z_{n-1}}(\mathbf{u}^{(n)})\,. \tag{5.1}$$

Convention. We order the bispecial prefixes of $\mathbf{u}$ by their length and we denote the nth bispecial prefix of $\mathbf{u}$ by $w^{(n)}$. In particular, $w^{(0)} = \varepsilon$, $w^{(1)} = 0$ if $z_0 = b$ and $w^{(1)} = 1$ if $z_0 = \beta$.

Our aim is to find for each $n \in \mathbb{N}$ the derived sequence of $\mathbf{u}$ to the prefix $w^{(n)}$ and to determine the type of $w^{(n)}$. First we need to know how bispecial factors and their return words change under the application of morphisms φ_b and φ_β. It is shown in [23].

Lemma 5.1. *Let $\mathbf{u}', \mathbf{u}$ be Sturmian sequences such that $\mathbf{u} = \varphi_b(\mathbf{u}')$.*

(i) For every bispecial factor w' of $\mathbf{u}'$, the factor $w = \varphi_b(w')0$ is a bispecial factor of $\mathbf{u}$.
(ii) Every bispecial factor w of $\mathbf{u}$ which is not empty can be written as $w = \varphi_b(w')0$ for a uniquely given bispecial factor w' of $\mathbf{u}'$.
(iii) The words r', s' are return words to a bispecial prefix w' of $\mathbf{u}'$ if and only if $r = \varphi_b(r'), s = \varphi_b(s')$ are return words to a bispecial prefix $w = \varphi_b(w')0$ of $\mathbf{u}$. Moreover, $\mathbf{d_u}(w) = \mathbf{d_{u'}}(w')$.

Lemma 5.2. *Let $\mathbf{u}', \mathbf{u}$ be Sturmian sequences such that $\mathbf{u} = \varphi_\beta(\mathbf{u}')$.*

(i) For every bispecial factor w' of $\mathbf{u}'$, the factor $w = \varphi_\beta(w')1$ is a bispecial factor of $\mathbf{u}$.
(ii) Every bispecial factor w of $\mathbf{u}$ which is not empty can be written as $w = \varphi_\beta(w')1$ for a uniquely given bispecial factor w' of $\mathbf{u}'$.
(iii) The words r', s' are return words to a bispecial prefix w' of $\mathbf{u}'$ if and only if $r = \varphi_\beta(r'), s = \varphi_\beta(s')$ are return words to a bispecial prefix $w = \varphi_\beta(w')1$ of $\mathbf{u}$. Moreover, $\mathbf{d_u}(w) = \mathbf{d_{u'}}(w')$.

Example 5.3. The sequence from Example 3.3 is the fixed point of the morphism $\varphi_{b\beta b}$. So it is a standard Sturmian sequence with the directive sequence $\mathbf{z} = b\beta bb\beta bb\beta b\cdots$.

The 0th bispecial prefix of $\mathbf{u}$ is the empty word $w^{(0)} = \varepsilon$. Its return words are 0 and 1 and clearly $\mathbf{d_u}(w^{(0)}) = \mathbf{u}$.

By Lemma 5.1, the bispecial prefix $w^{(1)}$ can be obtained from ε using the morphism φ_b: $w^{(1)} = \varphi_b(\varepsilon)0 = 0$. It means that $w^{(1)} = 0$ originates in the sequence $\mathbf{u}^{(1)}$ which has the directive sequence $\beta bb\beta bb\cdots$. The return words to $w^{(1)}$ are $\varphi_b(0) = 0$ and $\varphi_b(1) = 01$ and its derived sequence is $\mathbf{d_u}(w^{(1)}) = \mathbf{d}_{\mathbf{u}^{(1)}}(\varepsilon) = \mathbf{u}^{(1)}$.

Similarly, the prefix $w^{(2)}$ arises from ε by application of $\varphi_b\varphi_\beta$, *i.e.*

$$w^{(2)} = \varphi_b(\varphi_\beta(\varepsilon)1)0 = 010\,.$$

Thus $w^{(2)}$ originates in $\mathbf{u}^{(2)}$ with the directive sequence $bb\beta bb\beta\cdots$. The return words to $w^{(2)}$ are $\varphi_b\varphi_\beta(0) = 010$ and $\varphi_b\varphi_\beta(1) = 01$ and its derived sequence is $\mathbf{d_u}(w^{(2)}) = \mathbf{d}_{\mathbf{u}^{(2)}}(\varepsilon) = \mathbf{u}^{(2)}$.

The method explained in Example 5.3 can be easily generalized. Let us formalize this procedure.

Corollary 5.4. *Let* $\mathbf{z} \in \{b,\beta\}^{\mathbb{N}}$ *be a directive sequence of a standard Sturmian sequence* $\mathbf{u}$ *and let* $n \in \mathbb{N}$. *Denote by* $r^{(n)}$ *the most frequent and by* $s^{(n)}$ *the less frequent return words to the* nth *bispecial prefix* $w^{(n)}$ *of* $\mathbf{u}$. *Then the derived sequence* $\mathbf{d_u}(w^{(n)})$ *is a standard Sturmian sequence with the directive sequence* $\mathbf{z}^{(n)} = z_n z_{n+1} z_{n+2}\cdots$. *Moreover*

(i) If $z_n = b$, *then* $r^{(n)} = \varphi_{z_0z_1\cdots z_{n-1}}(0)$ *and* $s^{(n)} = \varphi_{z_0z_1\ldots z_{n-1}}(1)$.
(ii) If $z_n = \beta$, *then* $r^{(n)} = \varphi_{z_0z_1\cdots z_{n-1}}(1)$ *and* $s^{(n)} = \varphi_{z_0z_1\ldots z_{n-1}}(0)$.

Proof. We proceed with induction on $n \in \mathbb{N}$.

Clearly, the return words to the bispecial prefix $w^{(0)} = \varepsilon$ are letters 0 and 1 and thus the derived sequence of $\mathbf{u}$ to $w^{(0)}$ is the sequence $\mathbf{u}$ itself. Using the notation of (5.1), we have $\mathbf{u} = \varphi_{z_0}(\mathbf{u}^{(1)})$. If $z_0 = b$ then 0 is the most frequent letter in $\mathbf{u}$, if $z_0 = \beta$ then 1 is the most frequent letter of $\mathbf{u}$.

The directive sequence of $\mathbf{u}^{(1)}$ is $\mathbf{z}^{(1)} = z_1z_2z_3\cdots$. Denote $\mathbf{u}' = \mathbf{u}^{(1)}$. Let w' be the nth bispecial prefix of $\mathbf{u}'$ and r', s' be its return words. By Lemmas 5.1 and 5.2, the words $\varphi_{z_0}(r')$ and $\varphi_{z_0}(s')$ are the return words to the $(n+1)$st bispecial prefix of $\mathbf{u}$ (we add 1 for the bispecial prefix ε of $\mathbf{u}$).

If we now apply the induction hypothesis on $\mathbf{u}'$ and take into consideration that the application of φ_{z_0} to the return words does not change their frequencies, the statement is proved. □

It remains to determine the types of bispecial prefixes of standard Sturmian sequences. We will use the following matrix formalism.

Definition 5.5. Let w be a prefix of a standard Sturmian sequence $\mathbf{u}$. Let r, s be the return words to w in $\mathbf{u}$, where r is the most frequent return word. Then the matrix P_w is defined as:

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

Remark 5.6. The type $\mathcal{T}_w$ of the prefix w depends on the bottom row of the matrix P_w. $\mathcal{T}_w$ is

(i) SU if $P_w = \begin{pmatrix} p & q \\ 0 & 1 \end{pmatrix}$ for some numbers $p, q \in \{0,1\}$;
(ii) US if $P_w = \begin{pmatrix} p & q \\ 1 & 0 \end{pmatrix}$ for some numbers $p, q \in \{0,1\}$;
(iii) UU if $P_w = \begin{pmatrix} p & q \\ 1 & 1 \end{pmatrix}$ for some numbers $p, q \in \{0,1\}$.

Example 5.7. Take the prefixes ϵ, 0 and 010 of Sturmian sequence $\mathbf{u}$ from Example 3.3. Their matrices are

$$P_\varepsilon = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad P_0 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad P_{010} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

and their types are $\mathcal{T}_\varepsilon = SU(1)$, $\mathcal{T}_0 = US(1)$ and $\mathcal{T}_{010} = UU(2)$ respectively.

Convention. To simplify the notation, for the nth bispecial prefix $w^{(n)}$ of a standard Sturmian sequence we will denote its type $\mathcal{T}^{(n)}$ instead of $\mathcal{T}_{w^{(n)}}$ and its matrix $P^{(n)}$ instead of $P_{w^{(n)}}$.

Observation 5.8. *Let* $\mathbf{u}$ *be a standard Sturmian sequence with the directive sequence* $\mathbf{z} \in \{b, \beta\}^{\mathbb{N}}$.

(i) If $\mathbf{u}$ *has a prefix* $b^k\beta$ *for some positive integer* k*, then by Observation 2.3,* $\mathbf{u}$ *is a concatenation of the blocks* 0^k1 *and* $0^{k+1}1$*. Therefore, the bispecial prefix* $w^{(0)} = \varepsilon$ *has the stable return word* $r^{(0)} = 0$ *and the unstable return word* $s^{(0)} = 1$*,* $w^{(0)}$ *is of type* $\mathcal{T}^{(0)} = SU(k)$ *and its matrix* $P^{(0)}$ *is*

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} =: O_b \,.$$

(ii) If $\mathbf{u}$ *has a prefix* $\beta^k b$ *for some positive integer* k*, then* $w^{(0)} = \varepsilon$ *has the unstable return word* $r^{(0)} = 1$ *and the stable return word* $s^{(0)} = 0$*, its type is* $\mathcal{T}^{(0)} = US(k)$ *and its matrix* $P^{(0)}$ *is*

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} =: O_\beta \,.$$

Let us recall that the Parikh vector $V_{\psi(u)}$ can be computed by multiplication $V_{\psi(u)} = M_\psi V_u$ for all $u \in \mathcal{A}^*$. By Corollary 5.4, the Parikh vectors of the return words $r^{(n)}$ and $s^{(n)}$ can be computed using the matrix of the morphism $\varphi_{z_0 z_1 \dots z_{n-1}}$. Thus the matrix $P^{(n)}$ can be obtained as a product of the matrix of the morphism $\varphi_{z_0 z_1 \dots z_{n-1}}$ and the matrix O_b or O_β. The following proposition summarizes these observations.

Proposition 5.9. *Let* $\mathbf{u}$ *be a standard Sturmian sequence with the directive sequence* $\mathbf{z} \in \{b, \beta\}^{\mathbb{N}}$ *and let* $n \in \mathbb{N}$.

(i) If the sequence $z_n z_{n+1} z_{n+2} \cdots$ *has a prefix* $b^k\beta$*, then*

$$P^{(n)} = M_{z_0} M_{z_1} \cdots M_{z_{n-1}} O_b \mod 2 \,.$$

(ii) If the sequence $z_n z_{n+1} z_{n+2} \cdots$ *has a prefix* $\beta^k b$*, then*

$$P^{(n)} = M_{z_0} M_{z_1} \cdots M_{z_{n-1}} O_\beta \mod 2 \,.$$

The type $\mathcal{T}^{(n)}$ *is given by the bottom row of matrix* $P^{(n)}$ *and the number* k.

Example 5.10 (Example 5.3 continued)**.** The 0th bispecial prefix of $\mathbf{u}$ is $w^{(0)} = \varepsilon$. Since $\mathbf{z}$ has the prefix $b\beta$, its matrix is O_b and its type is $SU(1)$.

For the bispecial prefix $w^{(1)} = 0$, we have $z_0 = b$ and $z_1 z_2 z_3 \cdots = \beta b b \beta \cdots$ has the prefix βb. Thus the corresponding matrix is

$$P^{(1)} = M_b O_\beta = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \mod 2$$

and the 1st bispecial prefix $w^{(1)}$ is of type $\mathcal{T}^{(1)} = US(1)$.

For the bispecial prefix $w^{(2)}$ we have $z_0 z_1 = b\beta$ and $z_2 z_3 z_4 \cdots = bb\beta b \cdots$ has the prefix $b^2\beta$. Therefore its matrix is

$$P^{(2)} = M_b M_\beta O_b = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \mod 2$$

and its type is $\mathcal{T}^{(2)} = UU(2)$.

Finally, we study what kind of matrices can appear among the matrices $P^{(n)}$ of Sturmian bispecial prefixes. Clearly, all matrices $M_b, M_\beta, O_b, O_\beta$ have their determinants equal to 1. By Proposition 5.9, the matrix $P^{(n)}$ is

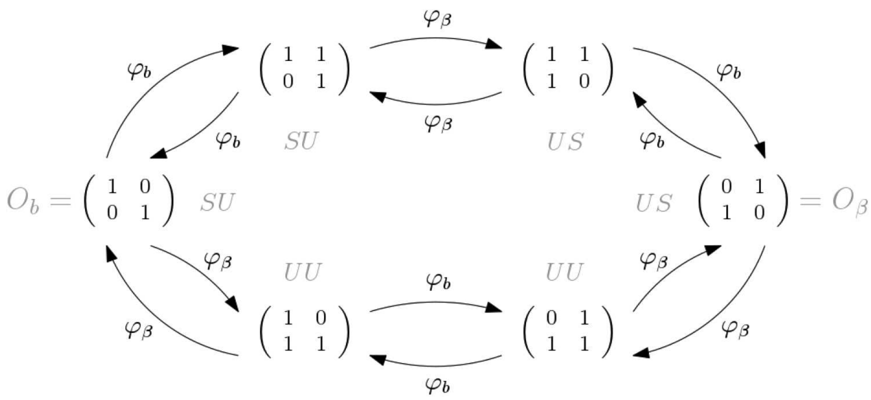

FIGURE 1. The diagram captures the multiplication mod 2 of matrices with the determinant 1 by morphism matrices M_b and M_β. A directed edge labelled by φ_b goes from matrix M to matrix M' if $M' = M_b M \mod 2$. Analogously for the label φ_β.

a product of these matrices modulo 2. So the determinant of $P^{(n)}$ has to be equal to 1. Therefore there are only six candidates for $P^{(n)}$:

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}. \tag{5.2}$$

The relations between these matrices are captured in Figure 1. We can go through this graph instead of calculating the respective products mod 2.

Example 5.11. The result of the product $P^{(2)} = M_b M_\beta O_b$ can be obtained as follows. We start in the vertex $O_b = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Then we move along the edge labelled by φ_β to the vertex $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. After that we move along the edge labelled by φ_b to the vertex $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. This vertex is the desired matrix $P^{(2)}$.

Remark 5.12. Any standard Sturmian sequence $\mathbf{u}$ with the directive sequence $\mathbf{z}$ has bispecial prefixes of at least two types. Indeed, by Proposition 5.9 (or by the graph in Fig. 1):
- if $\mathbf{z}$ has a prefix $b^\ell \beta$, $\ell \geq 1$, then the types of $w^{(0)}$ and $w^{(\ell)}$ are SU and US, respectively;
- if $\mathbf{z}$ has a prefix $\beta^{2\ell} b$, $\ell \geq 1$, then the types of $w^{(0)}$, $w^{(1)}$ and $w^{(2\ell)}$ are US, UU and SU, respectively;
- if $\mathbf{z}$ has a prefix $\beta^{2\ell-1} b$, $\ell \geq 1$, then the types of $w^{(0)}$ and $w^{(2\ell-1)}$ are US and UU, respectively.

It may happen that these bispecial prefixes are only of two types. For example, if the directive sequence is $\mathbf{z} = (\beta b b \beta)^\infty$, then for each $n \in \mathbb{N}$, the bispecial factor $w^{(2n)}$ is of the type $US(1)$ and the bispecial factor $w^{(2n+1)}$ is of the type $UU(2)$.

We illustrate our results on the Rote sequence $\mathbf{g}$ associated with the Fibonacci sequence $\mathbf{f}$.

Example 5.13. The Fibonacci sequence $\mathbf{f}$ is fixed by the Fibonacci morphism $F: 0 \to 01$, $1 \to 0$, *i.e.*

$$\mathbf{f} = 0100101001001010010100100101001001\cdots.$$

Clearly, the sequence $\mathbf{f}$ is a fixed point of the morphism F^2, too. Since $F^2 = \varphi_{b\beta}$, the Fibonacci sequence $\mathbf{f}$ has the directive sequence $\mathbf{z} = (b\beta)^\infty$, see Observation 2.3. By Corollary 5.4, the derived sequence $\mathbf{d_f}(w^{(2n)})$ has the directive sequence $(b\beta)^\infty$ and the derived sequence $\mathbf{d_f}(w^{(2n+1)})$ has the directive sequence $(\beta b)^\infty$. It means that $\mathbf{d_f}(w^{(2n)})$ is the Fibonacci sequence itself and $\mathbf{d_f}(w^{(2n+1)})$ can be obtained from the Fibonacci sequence by exchange of letters $0 \leftrightarrow 1$. If we rewrite the derived sequences into the alphabet $\{r, s\}$, where the most frequent letter is denoted by r and the less frequent letter by s (as required in Def. 3.8), we obtain only one derived sequence $\mathbf{d}$ (*i.e.* the Fibonacci sequence over the new alphabet):

$$\mathbf{d} = rsrrsrsrrsrrsrsrrsrsrrsrrsrsrrsrrsrs\cdots.$$

Therefore, the derived sequence to any prefix of the Fibonacci sequence is the Fibonacci sequence itself.

The types $\mathcal{T}^{(n)}$ of the bispecial prefixes of the Fibonacci sequence $\mathbf{f}$ can be determined by Proposition 5.9, where the matrix products can be computed using Figure 1:

- $P^{(0)} = O_b = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ mod 2 and $z_0z_1z_2\cdots = b\beta b\beta\cdots$, thus $w^{(0)}$ has the type $\mathcal{T}^{(0)} = SU(1)$;
- $P^{(1)} = M_bO_\beta = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ mod 2 and $z_1z_2z_3\cdots = \beta b\beta b\cdots$, thus $\mathcal{T}^{(1)} = US(1)$;
- $P^{(2)} = M_bM_\beta O_b = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ mod 2 and $z_2z_3z_4\cdots = b\beta b\beta\cdots$, thus $\mathcal{T}^{(2)} = UU(1)$;
- $P^{(3)} = M_bM_\beta M_bO_\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ mod 2 and $z_3z_4z_5\cdots = \beta b\beta b\cdots$, thus $\mathcal{T}^{(3)} = SU(1)$;
- $P^{(4)} = M_bM_\beta M_bM_\beta O_b = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ mod 2 and $z_4z_5z_6\cdots = b\beta b\beta\cdots$, thus $\mathcal{T}^{(4)} = US(1)$;
- $P^{(5)} = M_bM_\beta M_bM_\beta M_bO_\beta = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ mod 2 and $z_5z_6z_7\cdots = \beta b\beta b\cdots$, thus $\mathcal{T}^{(5)} = UU(1)$.

By simple computations we get

$$M_bM_\beta M_bM_\beta M_bM_\beta = I \mod 2\,,$$

where I is the identity matrix. Then we have

$$P^{(6)} = M_bM_\beta M_bM_\beta M_bM_\beta O_b = O_b \mod 2\,.$$

We have also $z_6z_7z_8\cdots = b\beta b\beta\cdots = z_0z_1z_2\cdots$. So we can conclude that $w^{(6)}$ has the same type as $w^{(0)}$, which is $SU(1)$. Similarly, $w^{(7)}$ has the same type as $w^{(1)}$ etc.

Now we use Corollary 4.4 to describe the derived sequences of the Rote sequence

$$\mathbf{g} = 001110011100011000110001110011100011\cdots$$

associated with the Fibonacci sequence $\mathbf{f}$. Since all derived sequences of $\mathbf{f}$ are the same and $\mathbf{f}$ has three distinct types of bispecial prefixes, there are exactly three distinct derived sequences of $\mathbf{g}$: $\mathbf{d_g}(x^{(0)})$, $\mathbf{d_g}(x^{(1)})$ and $\mathbf{d_g}(x^{(2)})$.

Finally, we show how to construct these derived sequences $\mathbf{d_g}(x^{(0)})$, $\mathbf{d_g}(x^{(1)})$ and $\mathbf{d_g}(x^{(2)})$. Since the type of $x^{(0)}$ is $SU(1)$, the return words A, B, C to the prefix $x^{(0)}$ correspond to the Sturmian factors r, sr^2s and srs, respectively, where r, s are the return words to $w^{(0)}$, see Theorem 3.10. Thus we have to decompose the sequence $\mathbf{d} \in \{r, s\}^{\mathbb{N}}$

$$\mathbf{d} = rsrrsrsrrsrrsrsrrsrsrrsrrsrsrrsrrsrs\cdots$$

onto blocks r, sr^2s and srs. The order of these blocks gives us the derived sequence of $\mathbf{g}$ to $x^{(0)}$

$$\mathbf{d_g}(x^{(0)}) = ABABAACAACAABABAAC\cdots.$$

The type of $w^{(1)}$ is $US(1)$, so the return words A, B, C to $x^{(1)}$ correspond to the Sturmian factors rr, rsr, s, respectively. So we decompose $\mathbf{d}$ onto blocks rr, rsr, s and we get

$$\mathbf{d_g}(x^{(1)}) = BBCACACBBCACACBBBC\cdots.$$

The type of $w^{(2)}$ is $UU(1)$, so the return words A, B, C to $x^{(2)}$ correspond to the Sturmian factors rr, rs, sr, respectively. So we decompose $\mathbf{d}$ onto blocks rr, rs, sr and we get

$$\mathbf{d_g}(x^{(2)}) = BACCBACCBBACBBACBB\cdots.$$

As explained in Section 4, the derived sequences $\mathbf{d_g}(x^{(0)})$, $\mathbf{d_g}(x^{(1)})$ and $\mathbf{d_g}(x^{(2)})$ are 3iet sequences. We can find the parameters of their interval exchange transformations using Proposition 4.2. The Fibonacci sequence has the slope $\alpha = \frac{1}{\tau}$ and the intercept $\rho = 1 - \frac{1}{\tau} = 2 - \tau$, where τ denotes the golden ratio $(1+\sqrt{5})/2$. Thus these derived sequences are 3iet sequences coding the intercept $2-\tau$ under the three interval exchange transformation T with the parameters β, γ and the permutation π as follows:

- for $\mathbf{d_g}(x^{(0)})$ the parameters are $\beta = \frac{1}{\tau}$, $\gamma = \frac{2}{\tau} - 1$ and $\pi = (3,2,1)$;
- for $\mathbf{d_g}(x^{(1)})$ the parameters are $\beta = \frac{2}{\tau} - 1$, $\gamma = 2 - \tau$ and $\pi = (3,2,1)$;
- for $\mathbf{d_g}(x^{(2)})$ the parameters are $\beta = \frac{2}{\tau} - 1$, $\gamma = 2 - \tau$ and $\pi = (2,3,1)$.

6. Derived sequences of substitutive complementary symmetric Rote sequences

The aim of this section is to decide when a Rote sequence associated with a standard Sturmian sequence is primitive substitutive, *i.e.* it is a morphic image of a fixed point of a primitive morphism. First we explain why a Rote sequence $\mathbf{v}$ cannot be purely primitive substitutive, *i.e.* cannot be fixed by a primitive morphism.

Lemma 6.1. *Let $\mathbf{v}$ be a Rote sequence associated with a standard Sturmian sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$. Then $\mathbf{v}$ is not a fixed point of a primitive morphism.*

Proof. Let us assume that $\mathbf{v}$ is fixed by a primitive morphism φ. Then the vector of letter frequencies $(\rho_0, \rho_1)^\top$ is an eigenvector to the dominant eigenvalue Λ of the matrix $M_\varphi \in \mathbb{N}^{2\times 2}$. As the language of $\mathbf{v}$ is closed under the exchange of letters $0 \leftrightarrow 1$, the vector of frequencies is $(\rho_0, \rho_1)^\top = (\frac{1}{2}, \frac{1}{2})^\top$, see [28]. Since all entries of the primitive matrix M_φ are integer, an eigenvalue to a rational eigenvector is an integer number. Moreover, by the Perron-Frobenius theorem the dominant eigenvalue of any primitive matrix with entries in $\mathbb{N}$ is bigger then 1, *i.e.* $\Lambda > 1$. The second eigenvalue Λ' (*i.e.* the other zero of the quadratic characteristic polynomial of M_φ) is integer, too.

Let x be a prefix of $\mathbf{v}$ and $\mathbf{d_v}(x)$ be the derived sequence of $\mathbf{v}$ to the prefix x. Let us assume that $\mathbf{d_v}(x)$ is fixed by a primitive morphism ψ. By Proposition 4.2, the derived sequence $\mathbf{d_v}(x)$ is a ternary sequence coding a three interval exchange transformation. The letter frequencies in any 3iet sequence are given by the lengths of the corresponding subintervals. The lengths of three subintervals described in Proposition 4.2 are irrational as the slope α of a Sturmian sequence is irrational. Therefore, the vector of frequencies and consequently the dominant eigenvalue of the matrix M_ψ is irrational as well.

For a sequence fixed by a primitive morphism η, Durand in [17] proved that any its derived sequence is fixed by some morphism, say ξ, and each eigenvalue λ of M_ξ either belongs to the spectrum of M_η or its modulus $|\lambda|$ belongs to $\{0, 1\}$.

Applying this result to the morphisms φ fixing the Rote sequence $\mathbf{v}$ and the morphism ψ fixing its derived sequence $\mathbf{d_v}(x)$ we get that the spectrum of M_ψ is a subset of $\{\Lambda, \Lambda', 0\} \cup \{y \in \mathbb{C} : |y| = 1\}$. Thus the dominant eigenvalue of the matrix M_ψ (which is by the Perron-Frobenius theorem bigger than 1) cannot be irrational. This is a contradiction. $\square$

Despite the previous lemma, we will show in Theorem 6.3 that a Rote sequence is primitive substitutive whenever the associated Sturmian sequence is primitive substitutive. For the proof we need to study the periodicity of the types of bispecial prefixes in Sturmian sequences.

Proposition 6.2. *Let $\mathbf{u}$ be a standard Sturmian sequence with an eventually periodic directive sequence $\mathbf{z}$ with a period Q. Then there exists $q \in \{1,2,3\}$ such that the sequence $\left(P^{(n)}\right)_{n\in\mathbb{N}}$ is eventually periodic with a period qQ.*

Proof. Let p be a preperiod of the directive sequence $\mathbf{z} = z_0z_1z_2\cdots$. We denote

$$H = M_{z_p}M_{z_{p+1}}\cdots M_{z_{p+Q-1}} \mod 2\,.$$

The matrix H belongs to the set of matrices displayed in (5.2). One can easily verify that

(i) if $H \in \left\{\begin{pmatrix}1&1\\0&1\end{pmatrix}, \begin{pmatrix}1&0\\1&1\end{pmatrix}, \begin{pmatrix}0&1\\1&0\end{pmatrix}\right\}$, then $H^2 = I \mod 2$;

(ii) if $H \in \left\{\begin{pmatrix}0&1\\1&1\end{pmatrix}, \begin{pmatrix}1&1\\1&0\end{pmatrix}\right\}$, then $H^3 = I \mod 2$.

Let q be the smallest positive integer such that $H^q = I \mod 2$, obviously $q \in \{1,2,3\}$. To conclude the proof we show that the sequence $\left(P^{(n)}\right)_{n\in\mathbb{N}}$ has a preperiod p and a period qQ. Let $n \geq p$ and $m = n + qQ$. By Proposition 5.9 and the fact that $z_i = z_{i+qQ}$ for any $i \geq p$ we can write

$$P^{(n)} = M_{z_0}\cdots M_{z_{n-1}}O_{z_n} = \Big(M_{z_0}\cdots M_{z_{p-1}}\Big)\Big(M_{z_p}\cdots M_{z_{n-1}}\Big)O_{z_n}\,,$$
$$P^{(n+qQ)} = P^{(m)} = M_{z_0}\cdots M_{z_{m-1}}O_{z_m} = \Big(M_{z_0}\cdots M_{z_{p-1}}\Big)\, H^q\, \Big(M_{z_p}\cdots M_{z_{n-1}}\Big)O_{z_n} = P^{(n)}\,.$$

□

Theorem 6.3. *Let $\mathbf{v}$ be a Rote sequence associated with a standard Sturmian sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$. Then the Rote sequence $\mathbf{v}$ is primitive substitutive if and only if the Sturmian sequence $\mathbf{u}$ is primitive substitutive.*

Proof. The proof is based on two results:

– A standard Sturmian sequence is primitive substitutive if and only if its directive sequence $\mathbf{z} \in \{b,\beta\}^{\mathbb{N}}$ is eventually periodic, see Observation 2.3.

– A sequence is primitive substitutive if and only if it has finitely many derived sequences, see Theorem 2.1.

Let us assume that $\mathbf{u}$ is primitive substitutive and Q is a period of its directive sequence $\mathbf{z} = z_0z_1z_2\cdots$. By Corollary 5.4, the sequence $\left(\mathbf{d_u}(w^{(n)})\right)_{n\in\mathbb{N}}$ is eventually periodic with the same period Q.

Remind that the type $\mathcal{T}^{(n)}$ of a bispecial prefix $w^{(n)}$ is determined by the bottom row of the matrix $P^{(n)}$ and by the length of the maximal monochromatic prefix of the sequence $z_nz_{n+1}z_{n+2}\cdots$. By Proposition 6.2, the sequence $\left(\mathcal{T}^{(n)}\right)_{n\in\mathbb{N}}$ is eventually periodic with the period qQ.

It implies that the sequence of pairs $\left(\mathcal{T}^{(n)}, \mathbf{d_u}(w^{(n)})\right)_{n\in\mathbb{N}}$ is eventually periodic with the period qQ, too. Then by Corollary 4.4, the Rote sequence $\mathbf{v}$ has only finitely many derived sequences and thus the sequence $\mathbf{v}$ is primitive substitutive.

On the other hand, if $\mathbf{v}$ is primitive substitutive, $\mathbf{v}$ has only finitely many derived sequences. Then by Corollary 4.4 the associated Sturmian sequence $\mathbf{u}$ has only finitely many derived sequences and thus $\mathbf{u}$ is primitive substitutive. □

Corollary 6.4. *Let $\mathbf{v}$ be a Rote sequence associated with a standard Sturmian sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$ fixed by a primitive morphism φ_z, where $z \in \{b,\beta\}^+$. Then $\mathbf{v}$ has at most $3|z|$ distinct derived sequences to its non-empty prefixes and each of them is fixed by a primitive morphism over a ternary alphabet.*

Proof. Using the notation from the proofs of Proposition 6.2 and Theorem 6.3, the sequence of pairs $\left(\mathcal{T}^{(n)}, \mathbf{d_u}(w^{(n)})\right)_{n\in\mathbb{N}}$ has a preperiod p and a period qQ, where $q \in \{1,2,3\}$. Since now the directive sequence $\mathbf{z} = z^\infty$ is purely periodic, we can choose $p = 0$ and $Q = |z|$. Each pair $\left(\mathcal{T}^{(n)}, \mathbf{d_u}(w^{(n)})\right)$ uniquely determines a derived sequence in $\mathbf{v}$, and thus there are at most $qQ \le 3|z|$ distinct derived sequences to non-empty prefixes of $\mathbf{v}$.

For the second part of the statement, it suffices to apply Durand's result from [16]: if a sequence $\mathbf{d}$ is the derived sequence to two distinct prefixes, then $\mathbf{d}$ is a fixed point of some morphism. □

Let us stress that the previous corollary does not speak about the derived sequence to the prefix ε. In this case the derived sequence $\mathbf{d_v}(\varepsilon)$ is the sequence $\mathbf{v}$ itself and by Lemma 6.1 it is not a fixed point of any primitive morphism.

Remark 6.5. The number of derived sequences of a Rote sequence $\mathbf{v}$ may be smaller than the value $3|z|$ announced in Corollary 6.4 and also smaller than the value qQ found in the proof. There are two reasons which may diminish the number:

(i) The period Q of the sequence $\left(\mathbf{d_u}(w^{(n)})\right)_{n\in\mathbb{N}}$ comes from the period Q of the directive sequence $\mathbf{z} = z^\infty$, where the word $z \in \{b,\beta\}^+$ describes the Sturmian morphism φ_z. If the word z is not primitive, *i.e.* $z = y^m$ for some $y \in \{b,\beta\}^+$ and $m \in \mathbb{N}, m \ge 2$, we can replace $|z|$ by the smaller number $|y|$. But even if z is primitive, the minimal period of $\left(\mathbf{d_u}(w^{(n)})\right)_{n\in\mathbb{N}}$ may be smaller. It happens for example in the Fibonacci case, where we consider the morphism $\varphi_{b\beta}$, *i.e.* $z = b\beta$, see Example 5.13.

(ii) The sequence of matrices $\left(P^{(n)}\right)_{n\in\mathbb{N}}$ has the guaranteed period qQ. But since the type $\mathcal{T}^{(n)}$ is determined only by the bottom row of the matrix $P^{(n)}$, it may also happen that $\mathcal{T}^{(m)} = \mathcal{T}^{(n)}$ for a pair $n, m \in \mathbb{N}$, $n < m < n + qQ$.

By the proof of Corollary 6.4, if two distinct prefixes of $\mathbf{v}$ has the same derived sequence, then this common derived sequence is fixed by some morphism. Durand in [16] provided a construction of this fixing morphism.

Durand's construction of fixing morphisms

Here we remind the construction only for the case when each non-empty prefix x of the sequence $\mathbf{v}$ has exactly three return words in $\mathbf{v}$. We assume:

- x and x' are prefixes of a sequence $\mathbf{v} \in \mathcal{A}^\mathbb{N}$ such that $|x| < |x'|$;
- $A, B, C \in \mathcal{A}^+$ are the return words to x and $A', B', C' \in \mathcal{A}^+$ are the return words to x';
- the derived sequences $\mathbf{d_v}(x)$ over $\{A,B,C\}$ and $\mathbf{d_v}(x')$ over $\{A',B',C'\}$ satisfy

$$\mathbf{d_v}(x') = \pi\big(\mathbf{d_v}(x)\big), \quad \text{where } \pi \text{ is the projection } A \to A', B \to B', C \to C'.$$

Since x is a prefix of x', the words A', B', C' are concatenations of the words A, B, C and we can write $A', B', C' \in \{A,B,C\}^+$. Thus one can find the words $w_A, w_B, w_C \in \{A,B,C\}^+$ such that $A' = w_A$, $B' = w_B$ and $C' = w_C$. Then the derived sequence $\mathbf{d_v}(x)$ is fixed by the morphism $\sigma : \{A,B,C\}^* \to \{A,B,C\}^*$ defined by

$$\sigma(A) = w_A, \quad \sigma(B) = w_B, \quad \sigma(C) = w_C\,.$$

If $\mathbf{v}$ is a Rote sequence associated with a standard Sturmian sequence $\mathbf{u}$, the Durand's construction can be transformed into the manipulation with the factors of the Sturmian sequence $\mathbf{u}$ instead of factors of the Rote sequence $\mathbf{v}$.

Let x, x' be two non-empty bispecial prefixes of $\mathbf{v}$ with the same derived sequence, *i.e.* $\mathbf{d_v}(x) = \mathbf{d_v}(x')$, and let $|x| < |x'|$. By Corollary 2.8, $w := \mathcal{S}(x)$ and $w' := \mathcal{S}(x')$ are bispecial prefixes of $\mathbf{u}$. By Corollary 4.4, w and w' have the same type and the same derived sequence of $\mathbf{u}$. Denote by r, s the most frequent and the less frequent return word to w in $\mathbf{u}$ and analogously denote the return words r', s' to the prefix w'.

Let A, B, C and A', B', C' be the return words to x and x' in $\mathbf{v}$, respectively. Put

$$a := \mathcal{S}(A0)\,,\quad b := \mathcal{S}(B0)\,,\quad c := \mathcal{S}(C0)\qquad \text{and}\qquad a' := \mathcal{S}(A'0)\,,\quad b' := \mathcal{S}(B'0)\,,\quad c' := \mathcal{S}(C'0)\,.$$

Theorem 3.10 implies $a, b, c \in \{r, s\}^+$ and $a', b', c' \in \{r', s'\}^+$. Recall that $A', B', C' \in \{A, B, C\}^+$ and as the types of bispecial prefixes w and w' are the same, necessarily $a', b', c' \in \{a, b, c\}^+$ as well. Therefore we can find the words $w_a, w_b, w_c \in \{a, b, c\}^+$ such that $a' = w_a, b' = w_b, c' = w_c$. Then the desired morphism σ fixing the derived sequence $\mathbf{d_v}(x)$ is defined by $A \to \pi(w_a), B \to \pi(w_b), C \to \pi(w_c)$, where π is the projection $a \to A, b \to B, c \to C$.

Moreover, if $\mathbf{u}$ is a fixed point of a primitive Sturmian morphism, then all its derived sequences are fixed by some primitive Sturmian morphisms as well, see Corollary 5.4. Thus the search for the morphism σ can be simplified as $r', s' \in \{r, s\}^*$ are images of r, s under a Sturmian morphism over the alphabet $\{r, s\}$.

Example 6.6. In Example 5.13 we have showed that the Rote sequence $\mathbf{g}$ associated with the Fibonacci sequence $\mathbf{f}$ has three derived sequences $\mathbf{d_g}(x^{(0)})$, $\mathbf{d_g}(x^{(1)})$ and $\mathbf{d_g}(x^{(2)})$. Now we find their fixing morphisms σ_0, σ_1, and σ_2, respectively.

Let us start with $\mathbf{d_g}(x^{(0)})$. We have $\mathbf{d_g}(x^{(0)}) = \mathbf{d_g}(x^{(3)})$ and so $\mathbf{d_f}(w^{(0)}) = \mathbf{d_f}(w^{(3)})$. The return words to $w^{(0)}$ are $r = 0$, $s = 1$ and the return words to $w^{(3)}$ are $r' = 01001$, $s' = 010$. Both $w^{(0)}$ and $w^{(3)}$ have the same type $SU(1)$. Thus the return words to $x^{(0)}$ correspond to blocks

$$a := \mathcal{S}(A0) = r = 0\,,\quad b := \mathcal{S}(B0) = srrs = 1001\,,\quad c := \mathcal{S}(C0) = srs = 101$$

and the return words to $x^{(3)}$ correspond to blocks

$$a' = r' = 01001\,,\qquad b' = s'r'r's' = 0100100101001010\,,\qquad c' = s'r's' = 01001001010\,,$$

where we denote $a' := \mathcal{S}(A'0)$, $b' := \mathcal{S}(B'0)$ and $c' := \mathcal{S}(C'0)$.

If we decompose $a' = 01001$ into $a = 0$, $b = 1001$, $c = 101$, we get $a' = ab$. Similarly $b' = abaacaaca$ and $c' = abaaca$. So the fixing morphism σ_0 is defined as follows:

$$\sigma_0 : \begin{cases} A \to AB \\ B \to ABAACAACA \\ C \to ABAACA \end{cases}.$$

Equivalently, we can also find σ_0 without knowledge of the return words since the Fibonacci sequence $\mathbf{f}$ is the fixed point of the Fibonacci morphism. As shown in Example 5.13, all derived sequences of $\mathbf{f}$ are equal to the Fibonacci sequence over the alphabet $\{r, s\}$ and so they are fixed by the morphism $\psi : r \to rs, s \to r$. Since the sequence of types $(\mathcal{T}^{(n)})_{n\in\mathbb{N}}$ has a period 3, the return words r' and s' satisfy $r' = \psi^3(r)$ and $s' = \psi^3(s)$. Therefore, it is enough to factorize $a' = \psi^3(r)$, $b' = \psi^3(srrs)$ and $c' = \psi^3(srs)$ into the blocks $a = r$, $b = srrs$ and $c = srs$. We get

$$\begin{array}{lllllllll} a' & = & r' & = & \psi^3(r) & = & rsrrs & = & ab, \\ b' & = & s'r'r's' & = & \psi^3(srrs) & = & rsrrsrrsrsrrsrsr & = & abaacaaca, \\ c' & = & s'r's' & = & \psi^3(srs) & = & rsrrsrrsrsr & = & abaaca\,. \end{array}$$

It exactly corresponds to the morphism σ_0.

Now we find a morphism σ_1 fixing the derived sequence of $\mathbf{g}$ to the prefix $x^{(1)}$. As the bispecial prefix $w^{(1)}$ has the type $US(1)$, we work with the blocks $a = rr$, $b = rsr$ and $c = s$. We apply the same method and we get

$$\begin{array}{lllllllll} a' & = & r'r' & = & \psi^3(rr) & = & rsrrsrsrrs & = & bbcac, \\ b' & = & r's'r' & = & \psi^3(rsr) & = & rsrrsrsrrsrrs & = & bbcacac, \\ c' & = & s' & = & \psi^3(s) & = & rsr & = & b\,. \end{array}$$

Thus the derived sequence $\mathbf{d_g}(x^{(1)})$ is fixed by the morphism

$$\sigma_1 : \begin{cases} A \to BBCAC \\ B \to BBCACAC \\ C \to B \end{cases}.$$

Finally, as $w^{(2)}$ has the type $UU(1)$, we work with the blocks $a = rr$, $b = rs$ and $c = sr$ and we get

$$\begin{array}{lllllllll} a' & = & r'r' & = & \psi^3(rr) & = & rsrrsrsrrs & = & baccb, \\ b' & = & r's' & = & \psi^3(rs) & = & rsrrsrsr & = & bacc, \\ c' & = & s'r' & = & \psi^3(sr) & = & rsrrsrrs & = & bacb\,. \end{array}$$

Therefore, the morphism fixing the derived sequence $\mathbf{d_g}(x^{(2)})$ is

$$\sigma_2 : \begin{cases} A \to BACCB \\ B \to BACC \\ C \to BACB \end{cases}.$$

We finish this section by an algorithm for finding the morphisms fixing the derived sequences of Rote sequences. To simplify the notation we use the cyclic shift operation $\mathrm{cyc} : \{b,\beta\}^+ \to \{b,\beta\}^+$ defined by

$$\mathrm{cyc}(z_0 z_1 \cdots z_{Q-1}) = z_1 z_2 \cdots z_{Q-1} z_0\,.$$

By Corollary 5.4, if $\mathbf{z} = z^\infty$, where $z = z_0 z_1 \cdots z_{Q-1} \in \{b,\beta\}^+$, is a directive sequence of a Sturmian sequence $\mathbf{u}$, then the derived sequence $\mathbf{d_u}(w^{(n)})$ to the nth bispecial prefix of $\mathbf{u}$ has the directive sequence $\mathbf{z}^{(n)} = \left(\mathrm{cyc}^n(z)\right)^\infty$, and thus $\mathbf{d_u}(w^{(n)})$ is fixed by the morphism $\varphi_{\mathrm{cyc}^n(z)}$.

Algorithm 6.7.
Input: $z \in \{b,\beta\}^*$ such that both b and β occur in z.

Output: the list of morphisms over $\{A,B,C\}$ fixing the derived sequences of the Rote sequence $\mathbf{v}$ associated with the fixed point of the morphism φ_z.

1. Denote $Q = |z|$, $\mathbf{z} = z_0 z_1 z_2 \cdots = z^\infty$ and $H = M_z$.
2. Find the minimal $q \in \{1,2,3\}$ such that $H^q = I \mod 2$.
3. For $i = 0,1,2,\ldots,qQ-1$ do:

- Compute $P^{(i)} = M_{z_0} M_{z_1} \cdots M_{z_{i-1}} O_{z_i} \mod 2$ and determine the type $\mathcal{T}^{(i)}$.
- Rewrite $\varphi_{\mathrm{cyc}^i(z)}$ into the alphabet $\{r,s\}$ by the rule $0 \to r, 1 \to s$ if the first letter of $\mathrm{cyc}^i(z)$ is b and by the rule $0 \to s, 1 \to r$ otherwise Denote this morphism ψ.
- Define $a,b,c \in \{r,s\}^+$ according to the type $\mathcal{T}^{(i)}$.
- Compute $\psi^q(a)$, $\psi^q(b)$, $\psi^q(c)$.
- Decompose the words $\psi^q(a)$, $\psi^q(b)$, $\psi^q(c)$ into the blocks a,b,c, *i.e.*, find $w_a, w_b, w_c \in \{a,b,c\}^+$ such that $\psi^q(a) = w_a,\ \psi^q(b) = w_b,\ \psi^q(c) = w_c$.

– Put into the list the morphism $\sigma_i : A \to \pi(w_a)$, $B \to \pi(w_b)$, $C \to \pi(w_c)$, where the projection π rewrites $a \to A$, $b \to B$, $c \to C$.

Example 6.8 (Example 5.10 continued)**.** Our aim is to describe the derived sequences of the Rote sequence associated with the fixed point of $\varphi_{b\beta b}$. Each its derived sequence is fixed by a primitive morphism which we find with Algorithm 6.7 to the input $z = b\beta b$.

1. $Q = |b\beta b| = 3$, $\mathbf{z} = (b\beta b)^\infty$ and $H = M_b M_\beta M_b = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.
2. $q = 2$ as $H^2 = I \mod 2$.
3. For $i = 0, 1, ..., 5$ (we illustrate the step only for $i = 2$) do:

– $P^{(2)} = M_b M_\beta O_b = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \mod 2$ and $z_2 z_3 z_4 \cdots = bb\beta \cdots$, thus $\mathcal{T}^{(2)} = UU(2)$.

– $\varphi_{\mathrm{cyc}^2(b\beta b)} = \varphi_{bb\beta} : \begin{cases} 0 \to 0010 \\ 1 \to 001 \end{cases}$ and thus $\psi : \begin{cases} r \to rrsr \\ s \to rrs \end{cases}$.

– $a = rr$, $b = rs$, $c = sr$.

– Since $\psi^2(r) = rrsrrrsrrrsrrsr$ and $\psi^2(s) = rrsrrrsrrrs$, we have
$\psi^2(a) = \psi^2(rr) = rrsrrrsrrrsrrsrrrsrrrsrrrsrrsr$;
$\psi^2(b) = \psi^2(rs) = rrsrrrsrrrsrrsrrrsrrrsrrrs$;
$\psi^2(c) = \psi^2(sr) = rrsrrrsrrrsrrsrrrsrrrsrrsr$.

– $\psi^2(a) = \underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{sr}_{c}$,
thus $w_a = acacacbabababac$;
$\psi^2(b) = \underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}$,
thus $w_b = acacacbababab$;
$\psi^2(c) = \underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{sr}_{c}$,
thus $w_c = acacacbababac$.

– We add to the list the morphism

$$\sigma_2 : \begin{cases} A \to ACACACBABABABAC \\ B \to ACACACBABABAB \\ C \to ACACACBABABAC \end{cases}.$$

7. Original Rote's construction and morphisms on four versus three letter alphabet

In the original Rote's paper [30], the author also construct Rote sequences as projections of fixed points on a four letter alphabet. Let us define the morphism ξ and the projection π as follows:

$$\xi : \begin{cases} 1 \to 13 \\ 2 \to 24 \\ 3 \to 241 \\ 4 \to 132 \end{cases} \quad \text{and} \quad \pi : \begin{cases} 1 \to 0 \\ 2 \to 1 \\ 3 \to 0 \\ 4 \to 1 \end{cases}.$$

The morphism ξ has two fixed points. We take its fixed point $\mathbf{q}$ starting with the letter 1, *i.e.*

$$\mathbf{q} = 13241241321324132132412413241241321324124132412413\cdots,$$

and we apply the projection π to construct the complementary symmetric Rote sequence:

$$\mathbf{r} = \pi(\mathbf{q}) = 00110110010011001001101100110110010011011001101100\cdots.$$

Using the operation $\mathcal{S}$ (see Def. 2.5), we obtain the associated Sturmian sequence

$$\mathbf{s} = \mathcal{S}\big(\mathbf{r}\big) = 0101101011010101101011010101101011010110101011010\cdots.$$

This example was in fact the beginning of this work. We notice that if we add the first letter 1 to the associated Sturmian sequence $\mathbf{s}$, we get the Sturmian sequence

$$\mathbf{u} = 1\mathbf{s} = 10101101011010101101011010101101011010110101011010 1\cdots$$

which is also fixed by a morphism, namely by the morphism:

$$\psi : \begin{cases} 0 \to 101 \\ 1 \to 10 \end{cases}.$$

The question is how to link this Rote's example to our construction.

In fact, the morphism ψ is a standard Sturmian morphism with the decomposition $\psi = \varphi_\beta \varphi_b E$. Thus we can use our techniques to find the return words and the derived sequence to the prefix $x = 01$ of the Rote sequence

$$\mathbf{v} = 01100100110110011011001001100100110110010011001001 10\cdots.$$

associated with the sequence $\mathbf{u}$. We obtain the return words $A = 0110$, $B = 010$ and $C = 011$. The derived sequence $\mathbf{d_v}(x)$ of $\mathbf{v}$ to the prefix $x = 01$ is fixed by the morphism

$$\sigma : \begin{cases} A \to ABC \\ B \to AC \\ C \to AB \end{cases}.$$

The Rote sequence $\mathbf{v}$ is clearly the image of the fixed point of σ, *i.e.* the derived sequence $\mathbf{d_v}(x)$, under the projection ρ defined as:

$$\rho(A) = 0110\,, \quad \rho(B) = 010\,, \quad \rho(C) = 011\,.$$

As $\mathbf{u} = 1\mathbf{s}$, their associated Rote sequences $\mathbf{v}$ and $\mathbf{r}$ are tied by

$$\mathbf{v} = 0110010011011001101100100 11\cdots = \overline{100110110010011001001101100}\cdots = \overline{1\mathbf{r}}\,.$$

Moreover, all return words to 01 in $\mathbf{v}$ obviously start with 0. Thus the original Rote sequence $\mathbf{r}$ is the image of the fixed point of σ under the projection ρ' defined as

$$\rho'(A) = \overline{1100} = 0011, \quad \rho'(B) = \overline{100} = 011, \quad \rho'(C) = \overline{110} = 001.$$

In other words, $\mathbf{r} = \pi(\mathbf{q}) = \rho'(\mathbf{d_v}(x))$.

The morphism ξ on a four letter alphabet can be recovered as follows. We write the images of the projection ρ' as suitable projections by π, more precisely

$$\rho'(A) = 0011 = \pi(1324), \quad \rho'(B) = 011 = \pi(124), \quad \rho'(C) = 001 = \pi(132).$$

Then the projections of the images of the morphism σ can be expressed as

$$\begin{aligned}
\rho'(\sigma(A)) &= \rho'(ABC) = \pi(1324124132) = \pi(\xi(1)\xi(3)\xi(2)\xi(4)),\\
\rho'(\sigma(B)) &= \rho'(AC) = \pi(1324132) = \pi(\xi(1)\xi(2)\xi(4)),\\
\rho'(\sigma(C)) &= \rho'(AB) = \pi(1324124) = \pi(\xi(1)\xi(3)\xi(2)),
\end{aligned}$$

and by recoding to the suffix code $\{13, 24, 241, 132\}$ we get the morphism ξ as above.

Finally, let us mention that we are able to generalize the original example as follows. For $n \in \mathbb{N}$ consider the morphism ξ_n defined as

$$\xi_n : \begin{cases} 1 \to 13 \\ 2 \to 24 \\ 3 \to (2413)^n 241 \\ 4 \to (1324)^n 132 \end{cases}.$$

Then the π projection of its fixed point $\mathbf{r}$ starting with 1 can be seen as the ρ' projection of the fixed point of the morphism σ_n, where σ_n and ρ' are the following:

$$\sigma_n : \begin{cases} A \to A^n ABA^n C \\ B \to A^n AC \\ C \to A^n AB \end{cases} \qquad \text{and} \qquad \rho' : \begin{cases} A \to 0011 \\ B \to 011 \\ C \to 001 \end{cases}.$$

Nevertheless, our technique with morphisms on a three letter alphabet is more natural and is based on the nice properties of return words and derived sequences. This is why we have chosen to write the paper to develop the whole theory of substitutive Rote sequences.

Remark 7.1. The Rote sequence $\mathbf{r}$ from the beginning of this section is connected to the fixed point $\mathbf{u}$ of the Sturmian morphism $\psi = \varphi_\beta \varphi_b E$. Of course, $\mathbf{u}$ is fixed also by the morphisms $\psi^2 = \varphi_{\beta bb\beta}$ which we have studied in Remark 5.12. It can be shown that the Rote sequence $\mathbf{v}$ associated with this $\mathbf{u}$ is exceptional among all Rote sequences associated with standard Sturmian sequences since it has only two distinct derived sequences to its prefixes. The other Rote sequences have at least three derived sequences.

8. Comments

In this paper, we have studied only the Rote sequences whose associated Sturmian sequences are standard. By definition, the intercept of a standard Sturmian sequence $\mathbf{u}$ is equal to $1 - \alpha$, where α is the density of the letter 0 in $\mathbf{u}$. For such a sequence we have used its S-adic representation $\mathbf{z}$ consisting of the morphisms $\varphi_b : 0 \to 0, 1 \to 01$ and $\varphi_\beta : 0 \to 10, 1 \to 1$. In particular, we have used the result from [23] which says that each suffix of $\mathbf{z}$ represents a derived sequence of $\mathbf{u}$.

In [14] M. Dekking studied properties of some submonoids of the Sturmian monoid. In particular, he considered the submonoid (we kept his notation) $\mathcal{M}_{3,8}$ generated by two morphisms $\psi_3 : 0 \to 0, 1 \to 01$ and $\psi_8 : 0 \to 01, 1 \to 1$. Theorem 3 from [14] says that any fixed point $\mathbf{u}$ of a primitive morphism from $\mathcal{M}_{3,8}$ is a Sturmian sequence with the intercept 0. Obviously, this $\mathbf{u}$ has an S-adic representation $\mathbf{z}$ consisting of the morphisms ψ_3 and ψ_8. But unlike the case of standard Sturmian sequences, in this case only some of suffixes of $\mathbf{z}$ represent derived sequences to prefixes of $\mathbf{u}$, see [23]. It would be interesting to know how this fact influences the set of derived sequences of a Rote sequence associated with a Sturmian sequence with the intercept 0.

The definition of derived sequences of $\mathbf{u}$ as introduced in [16] takes into account only the prefixes of $\mathbf{u}$. Recently, Yu-Ke Huang and Zhi-Ying Wen in [20] have considered also the derived sequences of $\mathbf{u}$ to non-prefix factors of $\mathbf{u}$. Recall that if $\mathbf{u}$ is a fixed point of a primitive morphism, then by Durand's result from [16], any derived sequence to a prefix of $\mathbf{u}$ is fixed by a primitive morphism as well. However, a derived sequence to a non-prefix factor of $\mathbf{u}$ need not to be fixed by a non-identical morphism at all.

Huang and Wen study the period-doubling sequence $\mathbf{p}$, *i.e.* the sequence fixed by the morphism $0 \mapsto 11, 1 \mapsto 10$. They show that there exist two sequences Θ_1 and Θ_2 such that any derived sequence $\mathbf{d}$ to a factor of $\mathbf{u}$ is equal to Θ_1 or Θ_2. Moreover, any derived sequence $\mathbf{d}'$ to a factor of $\mathbf{d}$ is equal to Θ_1 or Θ_2 and any derived sequence $\mathbf{d}''$ to a factor of $\mathbf{d}'$ is equal to Θ_1 or Θ_2, etc. They called this property Reflexivity. It may be interesting to look for sequences with Reflexivity among Rote or Sturmian sequences.

Let us note that the period-doubling sequence $\mathbf{p}$ and the Thue-Morse sequence $\mathbf{t}$, *i.e.* the sequence fixed by the morphism $0 \mapsto 01, 1 \mapsto 10$, are linked with the same mapping $\mathcal{S}$ which associates the Rote and Sturmian sequences (see Def. 2.5): $\mathcal{S}(\mathbf{t}) = \mathbf{p}$.

By Corollary 6.4, a Rote sequence $\mathbf{v}$ associated with a fixed point $\mathbf{u}$ of a primitive standard Sturmian morphism φ_z has at most $3|z|$ distinct derived sequences. On the other hand, if φ_z is not a power of any other morphism, then the Sturmian sequence $\mathbf{u}$ has exactly $|z|$ distinct derived sequences (see [23]) and thus by Corollary 4.1, the Rote sequence $\mathbf{v}$ has at least $|z|$ distinct derived sequences. In all examples for which we have listed the derived sequences of the fixed point of such a φ_z, the actual number of derived sequences was $|z|$, $2|z|$ or $3|z|$. We do not know whether some other values can also appear.

Acknowledgements. The research received funding from the Ministry of Education, Youth and Sports of the Czech Republic through the project no. CZ.02.1.01/0.0/0.0/16_019/0000778, from the Czech Technical University in Prague through the project SGS17/193/OHK4/3T/14, and from the French Agence Nationale de la Recherche (ANR) through the GlobNets (ANR-16-CE02-0009) project.

References

[1] I.M. Araújo and V. Bruyùre, Words derivated from Sturmian words. *Theoret. Comput. Sci.* **340** (2005) 204–219.
[2] L. Balková, M. Bucci, A. De Luca, J. Hladký and S. Puzynina, Aperiodic pseudorandom number generators based on infinite words. *Theoret. Comput. Sci.* **647** (2016) 85–100.
[3] L. Balková, E. Pelantová and Š. Starosta, Sturmian jungle (or garden?) on multiliteral alphabets. *RAIRO: ITA* **44** (2010) 443–470.
[4] L. Balková, E. Pelantová and W. Steiner, Sequences with constant number of return words. *Monatsh. Math.* **155** (2008) 251–263.
[5] J. Berstel, Sturmian and Episturmian words, in Algebraic Informatics, edited by S. Bozapalidis and G. Rahonis. Springer, Berlin, Heidelberg (2007) 23–47.
[6] J. Berstel and P. Séébold, Sturmian words, in Algebraic Combinatorics on Words, edited by M. Lothaire. Cambridge University Press (2002) 45–110.
[7] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer and G. Rindone, Acyclic, connected and tree sets. *Monatsh. Math.* **176** (2015) 521–550.
[8] V. Berthé, V. Delecroix, F. Dolce, D. Perrin, Ch. Reutenauer and G. Rindone, Return words of linear involutions and fundamental groups. *Ergodic Theory Dyn. Syst.* **37** (2017) 693–715.
[9] V. Berthé, F. Dolce, F. Durand, J. Leroy and D. Perrin, Rigidity and substitutive dendric words. *IJFCS Int. J. Found. Comput. Sci.* **29** (2018) 705–720.
[10] A. Blondin-Massé, S. Brlek, S. Labbé and L. Vuillon, Codings of rotations on two intervals are full. *Electr. Notes Discrete Math.* **34** (2009) 289–393.
[11] A. Blondin-Massé, G. Paquin, H. Tremblay and L. Vuillon, On generalized pseudostandard words over binary alphabet. *J. Int. Sequen.* **16** (2013) 13.2.11.
[12] J. Cassaigne, Complexité et facteurs spéciaux. Journées Montoises (Mons, 1994). *Bull. Belg. Math. Soc. Simon Stevin* **4** (1997) 67–88.
[13] A. de Luca and A. De Luca, Pseudopalindromic closure operators in free monoids. *Theoret. Comput. Sci.* **362** (2006) 282–300.
[14] M. Dekking, Substitution invariant words and binary trees. *Electr. J. Int.* **18A** (2018) #A7.

[15] F. Dolce and D. Perrin, Neutral and tree sets of arbitrary characteristic. *Theoret. Comput. Sci.* **658** (2017) 159–174.

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

[17] F. Durand, A generalization of Cobham's Theorem. *Theory Comput. Syst.* **31** (1998) 169–185.

[18] S. Ferenczi and T. Monteil, Infinite Words with Uniform frequencies and invariant measures, in Combinatorics, Automata and Number Theory, edited by V. Berthé and M. Rigo. Cambridge University Press (2010) 373–409.

[19] S. Ferenczi, C. Holton and L. Zamboni, The structure of three-interval exchange transformations II: a combinatorial description of the trajectories. *J. Anal. Math.* **89** (2003) 239–276.

[20] Y.K. Huang and Z.Y. Wen, Envelope words and the reflexivity of the return word sequences in the period-doubling sequence. Preprint arXiv:1703.07157v1 (2017).

[21] J. Justin and L. Vuillon, Return words in Sturmian and episturmian words. *RAIRO: ITA* **34** (2000) 343–356.

[22] M.S. Keane, Interval exchange transformations. *Math. Zeitsch.* **141** (1975) 25–31.

[23] K. Klouda, K. Medková, E. Pelantová and S. Starosta, Fixed points of Sturmian morphisms and their derivated words. *Theoret. Comput. Sci.* **743** (2018) 23–37.

[24] M. Lothaire, Combinatorics on Words, Vol. 17 of *Encyclopaedia of Mathematics and its Applications.* Addison-Wesley, Reading, Mass (1983). Reprinted in the Cambridge Mathematical Library, Cambridge University Press, Cambridge, UK (1997).

[25] M. Morse and G.A. Hedlund, Symbolic dynamics II – Sturmian trajectories. *Am. J. Math.* **62** (1940) 1–42.

[26] V.I. Oseledec, The spectrum of ergodic automorphisms. *Doklady Akademii Nauk SSSR* **168** (1966) 1009–1011.

[27] E. Pelantová and Š. Starosta, Constructions of words rich in palindromes and pseudopalindromes. *Discr. Math. Theor. Comput. Sci.* **18** (2016).

[28] M. Queffélec, Substitution dynamical systems – spectral analysis. Vol. 1294 of *Lecture Notes in Math.* Springer, Berlin (1987).

[29] G. Rauzy, Echanges d'intervalles et transformations induites. *Acta Arith.* **34** (1979) 315–328.

[30] G. Rote, Sequences with subword complexity $2n$. *J. Number Theory* **46** (1994) 196–213.

[31] L. Vuillon, A characterization of Sturmian words by return words. *Eur. J. Combin.* **22** (2001) 263–275.

[32] L. Vuillon, On the number of return words in infinite words with complexity 2n + 1. *Pure Math. Appl.* **18** (2007) 345–355.

Article D

On non-repetitive complexity of Arnoux–Rauzy words

On non-repetitive complexity of Arnoux–Rauzy words*

Kateřina Medková[a,*], Edita Pelantová[a], Élise Vandomme[b]

[a] *Department of Mathematics, FNSPE, Czech Technical University in Prague, Trojanova 13, 120 00 Praha 2, Czech Republic*

[b] *Department of Mathematics, University of Liege, Allée de la découverte 12, 4000 Liege, Belgium*

Abstract

The non-repetitive complexity $nr\mathcal{C}_{\mathbf{u}}$ and the initial non-repetitive complexity $inr\mathcal{C}_{\mathbf{u}}$ are functions which reflect the structure of the infinite word $\mathbf{u}$ with respect to the repetitions of factors of a given length. We determine $nr\mathcal{C}_{\mathbf{u}}$ for the Arnoux–Rauzy words and $inr\mathcal{C}_{\mathbf{u}}$ for the standard Arnoux–Rauzy words. Our main tools are S-adic representation of Arnoux–Rauzy words and description of return words to their factors. The formulas we obtain are then used to evaluate $nr\mathcal{C}_{\mathbf{u}}$ and $inr\mathcal{C}_{\mathbf{u}}$ for the d-bonacci word.

Keywords: Arnoux–Rauzy word, directive sequence, factor complexity, non-repetitivity
2000 MSC: 68R15

---

1. Introduction

Variability of an infinite word $\mathbf{u} = u_0u_1u_2\cdots$ over a finite alphabet can be judged from distinct points of view depending on applications or combinatorial properties one is interested in. The factor complexity of $\mathbf{u}$, here denoted $\mathcal{C}_{\mathbf{u}}$, is a function which to any $n \in \mathbb{N}$ assigns the number of distinct factors of length n occurring in $\mathbf{u}$. More formally, $\mathcal{C}_{\mathbf{u}}(n) = \#\{u_iu_{i+1}\cdots u_{i+n-1} \ : \ i \in \mathbb{N}\}$.

For the simplest infinite words, namely the eventually periodic words, the factor complexity is bounded from above by a constant. In [13], Morse and Hedlund showed that the factor complexity of an infinite word which is not eventually periodic satisfies $\mathcal{C}_{\mathbf{u}}(n) \geq n+1$ for each $n \in \mathbb{N}$. If the equality takes place for each n, the word $\mathbf{u}$ is called Sturmian. Sturmian words represent the most intensively studied class of infinite words. To measure the regularity of an infinite word, Morse and Hedlund introduced the recurrence function $R_{\mathbf{u}}$. The value $R_{\mathbf{u}}(n)$ is defined to be the minimal integer m such that any factor of $\mathbf{u}$ of length n occurs at least once in $u_iu_{i+1}u_{i+2}\cdots u_{i+m-1}$ for every $i \in \mathbb{N}$. In the same paper [13], the authors evaluated $R_{\mathbf{u}}(n)$ for any Sturmian word.

A dual function to $R_{\mathbf{u}}$ was recently introduced by Moothathu [12] under the name non-repetitive complexity function $nr\mathcal{C}_{\mathbf{u}}$. The value $nr\mathcal{C}_{\mathbf{u}}(n)$ is defined as the maximal m such that for some $i \in \mathbb{N}$ any factor of $\mathbf{u}$ of length n occurs at most once in $u_iu_{i+1}u_{i+2}\cdots u_{i+m+n-2}$. He also considered a "prefix variant" of this function called the initial non-repetitive complexity function $inr\mathcal{C}_{\mathbf{u}}$. By

---

*The research received funding from the Ministry of Education, Youth and Sports of the Czech Republic through the project no. CZ.02.1.01/0.0/0.0/16 019/0000778 and from the Czech Technical University in Prague through the project SGS17/193/OHK4/3T/14. The last author acknowledges partial funding via a Welcome Grant of the University of Liege.

*Corresponding author
Email address: katerinamedkova@gmail.com (Kateřina Medková)

definition, $inr\mathcal{C}_{\mathbf{u}}(n)$ is the maximal length m of a prefix of $\mathbf{u}$ such that each factor of $\mathbf{u}$ of length n occurs in $u_0u_1\cdots u_{m+n-2}$ at most once. Obviously,

$$inr\mathcal{C}_{\mathbf{u}}(n) \;\leq\; nr\mathcal{C}_{\mathbf{u}}(n) \;\leq\; \mathcal{C}_{\mathbf{u}}(n) \;\leq\; R_{\mathbf{u}}(n) - n + 1 \quad \text{for each } n \in \mathbb{N}.$$

Moothathu's concept of the initial non-repetitive complexity function was developed in [14] by Nicholson and Rampersad. They described some general properties of $inr\mathcal{C}_{\mathbf{u}}$ and evaluated $inr\mathcal{C}_{\mathbf{u}}$ for the Fibonacci, Tribonacci and Thue–Morse words. Note that the Fibonacci word and the Tribonacci word belong to the class of standard binary and ternary, respectively, Arnoux–Rauzy words. The Arnoux–Rauzy words represent one of the generalizations of Sturmian words to multi-letter alphabets. The recurrence function $R_{\mathbf{u}}$ for Arnoux–Rauzy words was determined in [7]. The initial non-repetitive complexity function for Sturmian sequences was recently studied by Bugeaud and Kim [5]. Their motivation for this study comes from the connection between the irrational exponent of a number x and $inr\mathcal{C}_{\mathbf{u}}$, where $\mathbf{u}$ corresponds to the expansion of x in a given base.

In the present article we focus on the non-repetitive complexity of Arnoux–Rauzy words. Using the S-adic representation of a given Arnoux–Rauzy word $\mathbf{u}$, we provide in Theorem 13 a formula for computing $nr\mathcal{C}_{\mathbf{u}}(n)$ for each $n \in \mathbb{N}$. In particular, we show (Theorem 5) that any Sturmian word (i.e., binary Arnoux–Rauzy word) $\mathbf{u}$ satisfies $nr\mathcal{C}_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n)$ for each $n \in \mathbb{N}$. It is interesting that this phenomenon can be observed also among the words with the maximal factor complexity. In [14], the authors constructed a word over q letter alphabet such that $q^n = \mathcal{C}_{\mathbf{u}}(n) = nr\mathcal{C}_{\mathbf{u}}(n)$.

For standard Arnoux–Rauzy words we determine in Theorem 21 also $inr\mathcal{C}_{\mathbf{u}}$ and thus we generalize Nicholson and Rampersad's result on the Fibonacci and the Tribonacci words.

2. Preliminaries

An *alphabet* $\mathcal{A}$ is a finite set of symbols called *letters*. Here we fix the alphabet $\mathcal{A} = \{0, 1, \ldots, d-1\}$, where d is a positive integer. A *word* $w = w_0 \cdots w_{n-1}$ over $\mathcal{A}$ is a finite sequence of letters from $\mathcal{A}$. The number of its letters is called the *length* of w and it is denoted by $|w| = n$. The notation $|w|_a$ is used for the number of occurrences of the letter a in w. The *empty word*, i.e., the unique word of length zero, is denoted by ε. The *concatenation* of words $v = v_0 \cdots v_k$ and $w = w_0 \cdots w_\ell$ is the word $vw = v_0 \cdots v_k w_0 \cdots w_\ell$. The set of all finite words over $\mathcal{A}$ equipped with the operation concatenation of words is a free monoid and it is denoted $\mathcal{A}^*$. The *Parikh vector* of a word $w \in \mathcal{A}^*$ is the vector $\vec{V}(w) = (|w|_0, |w|_1, \ldots, |w|_{d-1})^\top$. Obviously, $|w| = (1, 1, \cdots, 1) \cdot \vec{V}(w)$.

An infinite sequence of letters $\mathbf{u} = (u_i)_{i\geq 0}$ in $\mathcal{A}$ is called *infinite word*. The set of all infinite words over $\mathcal{A}$ is denoted $\mathcal{A}^{\mathbb{N}}$. The word $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is said to be *eventually periodic* if it is of the form $\mathbf{u} = vz^\omega$, where $v, z \in \mathcal{A}^*$, $z \neq \varepsilon$ and $z^\omega = zzz\cdots$. Otherwise, $\mathbf{u}$ is *aperiodic*.

A *factor* of a (finite of infinite) word w is a finite word v such that $w = svt$ for some words $s, t \in \mathcal{A}^*$. Moreover, if $s = \varepsilon$, then v is called a *prefix* of w and if $t = \varepsilon$, then v is called a *suffix* of w. The set of all factors of an infinite word $\mathbf{u}$ is called the *language* of $\mathbf{u}$ and denoted by $\mathcal{L}_{\mathbf{u}}$. By $\mathcal{L}_{\mathbf{u}}(n)$ we denote the set of factors of $\mathbf{u}$ of length n, i.e., $\mathcal{L}_{\mathbf{u}}(n) = \mathcal{L}_{\mathbf{u}} \cap \mathcal{A}^n$. Using this notation, the factor complexity of $\mathbf{u}$ can be expressed as $\mathcal{C}_{\mathbf{u}}(n) = \#\mathcal{L}_{\mathbf{u}}(n)$ for every $n \in \mathbb{N}$. In this paper, we focus on the (initial) non-repetitive complexity.

Definition 1. *The* non-repetitive complexity $nr\mathcal{C}_{\mathbf{u}}$ *and the* initial non-repetitive complexity $inr\mathcal{C}_{\mathbf{u}}$ *of an infinite word* $\mathbf{u}$ *are functions defined for each* $n \in \mathbb{N}$ *as follows*

$$nr\mathcal{C}_{\mathbf{u}}(n) := \max\{m \in \mathbb{N} : \exists k \in \mathbb{N} \ s.t.\ u_i \cdots u_{i+n-1} \neq u_j \cdots u_{j+n-1} \ \forall i, j \ \textit{with} \ k \leq i < j \leq k+m-1\},$$

$$inr\mathcal{C}_{\mathbf{u}}(n) := \max\{m \in \mathbb{N} : u_i \cdots u_{i+n-1} \neq u_j \cdots u_{j+n-1} \ \forall i, j \ \textit{with} \ 0 \leq i < j \leq m-1\}\,.$$

A factor w of $\mathbf{u}$ is *right special* if there exist two distinct letters $a, b \in \mathcal{A}$ such that wa and wb belong to $\mathcal{L}_{\mathbf{u}}$. Analogously, w is *left special* if aw and bw belong to $\mathcal{L}_{\mathbf{u}}$ for two distinct letters $a, b \in \mathcal{A}$. A factor which is both left and right special is called *bispecial.* If $\mathbf{u}$ is aperiodic, then for any length n at least one factor $w \in \mathcal{L}_{\mathbf{u}}(n)$ is left special and at least one factor $v \in \mathcal{L}_{\mathbf{u}}(n)$ is right special.

Factors of an infinite word $\mathbf{u}$ can be visualized by the so-called *Rauzy graphs* $\Gamma_{\mathbf{u}}(n)$, $n \in \mathbb{N}$. The set of vertices of $\Gamma_{\mathbf{u}}(n)$ is $\mathcal{L}_{\mathbf{u}}(n)$ and the set of its edges is $\mathcal{L}_{\mathbf{u}}(n+1)$. An oriented edge $e \in \mathcal{L}_{\mathbf{u}}(n+1)$ starts in $u \in \mathcal{L}_{\mathbf{u}}(n)$ and ends in $v \in \mathcal{L}_{\mathbf{u}}(n)$ if u is a prefix of e and v is a suffix of e. If $w \in \mathcal{L}_{\mathbf{u}}(n)$, we denote

$$\begin{aligned} N_+(w) &= \{v \in \mathcal{L}_{\mathbf{u}}(n) \,:\, w \text{ is a prefix and } v \text{ is a suffix of an edge } e \in \mathcal{L}_{\mathbf{u}}(n+1)\}\,, \\ N_-(w) &= \{v \in \mathcal{L}_{\mathbf{u}}(n) \,:\, v \text{ is a prefix and } w \text{ is a suffix of an edge } e \in \mathcal{L}_{\mathbf{u}}(n+1)\}\,. \end{aligned}$$

Any factor $v \in \mathcal{L}_{\mathbf{u}}(n+m)$ with a prefix $u \in \mathcal{L}_{\mathbf{u}}(n)$ corresponds to an oriented path of length m in $\Gamma_{\mathbf{u}}(n)$ starting with the vertex u.

The *occurrence* of the word w in $\mathbf{u} = u_0u_1u_2\cdots$ is every index $i \in \mathbb{N}$ such that w is a prefix of the word $u_iu_{i+1}u_{i+2}\cdots$. The factor of length n which occurs at the position i is denoted by $f_n(i)$. Hence, $f_n(i) = w$ if $w \in \mathcal{L}_{\mathbf{u}}(n)$ and w is a prefix of $u_iu_{i+1}u_{i+2}\cdots$. An infinite word $\mathbf{u}$ is said to be *recurrent* if each of its factors has at least two occurrences in $\mathbf{u}$. If $i < j$ are two consecutive occurrences of w in $\mathbf{u}$, then the word $u_iu_{i+1}\cdots u_{j-1}$ is called the *return word* to w in $\mathbf{u}$. If the set of all return words to w in $\mathbf{u}$ is finite for each factor w of $\mathbf{u}$, the word $\mathbf{u}$ is called *uniformly recurrent.*

A *morphism* of the free monoid $\mathcal{A}^*$ is a map $\psi : \mathcal{A}^* \to \mathcal{A}^*$ such that $\psi(vw) = \psi(v)\psi(w)$ for all $v, w \in \mathcal{A}^*$. The *incidence matrix* of ψ is $d \times d$ matrix $\boldsymbol{M}_\psi$ given by $[\boldsymbol{M}_\psi]_{ab} = |\psi(b)|_a$. The incidence matrix of ψ can be used to compute the Parikh vector of the image of a word w under ψ:

$$\vec{V}(\psi(w)) = \boldsymbol{M}_\psi \cdot \vec{V}(w)\,. \tag{1}$$

The domain of a morphism ψ of $\mathcal{A}^*$ can be naturally extended to $\mathcal{A}^{\mathbb{N}}$ by putting $\psi(\mathbf{u}) = \psi(u_0u_1u_2\cdots) = \psi(u_0)\psi(u_1)\psi(u_2)\cdots$. An infinite word $\mathbf{u}$ is called a *fixed point* of the morphism ψ if $\mathbf{u} = \psi(\mathbf{u})$.

3. Arnoux–Rauzy words

The Sturmian words can be described by many equivalent properties, for their list (which is far from being complete) see for example [2]. These properties offer several possibilities for generalization. One of them was used by Arnoux and Rauzy in [1] to introduce the words today known under their names.

Definition 2. *A recurrent infinite word* $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ *is a* d-ary Arnoux–Rauzy word *if for all* n *it has* $(d-1)n+1$ *factors of length* n *with exactly one left and one right special factor of length* n.

Over the binary alphabet the Arnoux–Rauzy words coincide with the Sturmian words. The Arnoux–Rauzy words belong to a broader family of episturmian words (e.g., see [9]). They are also embedded in the very general concept of tree sets introduced in [4] which comprises several generalizations of Sturmian words to multi-letter alphabet. The Arnoux–Rauzy words share many properties with the Sturmian words (e.g., see [15, 10, 8]). Here we recall some of them. If $\mathbf{u}$ is a d-ary Arnoux–Rauzy word, then

- there exists a dominant letter $a \in \mathcal{A}$ such that a occurs in each factor from $\mathcal{L}_{\mathbf{u}}(2)$;
- $\mathcal{L}_{\mathbf{u}}$ is closed under reversal, i.e., $w = w_0 w_1 \cdots w_{n-1} \in \mathcal{L}_{\mathbf{u}}$ implies $\bar{w} = w_{n-1} \cdots w_1 w_0 \in \mathcal{L}_{\mathbf{u}}$;
- each bispecial factor w of $\mathbf{u}$ is a palindrome, i.e., $w = \bar{w}$;
- $\mathbf{u}$ is uniformly recurrent;
- any factor of $\mathbf{u}$ has exactly d return words in $\mathbf{u}$.

On the other hand, some properties of Sturmian words are not present in d-ary Arnoux–Rauzy words when $d \geq 3$. An example of such a property is the so-called balancedness. Already Hedlund and Morse [13] proved that a binary aperiodic word $\mathbf{u}$ is Sturmian if and only if for any pair $v, w \in \mathcal{L}_{\mathbf{u}}$ of factors of the same length the inequality $|v|_a - |w|_a \leq c = 1$ holds for any letter $a \in \mathcal{A}$. This property is not preserved in Arnoux–Rauzy words, even if the constant $c = 1$ is allowed to depend on d. For a detailed study of this problem, see [3].

If each prefix of an Arnoux–Rauzy word $\mathbf{u}$ is left special, then $\mathbf{u}$ is called *standard*. For each Arnoux–Rauzy word $\mathbf{v}$, there exists a unique standard Arnoux–Rauzy word $\mathbf{u}$ such that $\mathcal{L}_{\mathbf{u}} = \mathcal{L}_{\mathbf{v}}$. We will work with the S-adic representation of the Arnoux–Rauzy words as described in [8]. Therefore we define the set S of elementary morphisms over the alphabet $\mathcal{A} = \{0, 1, \ldots, d-1\}$.

$$\text{For } i = 0, 1, \ldots, d-1 \quad \text{we put} \quad \varphi_i \ : \ \begin{cases} i \to i\,; \\ j \to ij \quad \text{for } j \neq i\,. \end{cases} \tag{2}$$

Any standard Arnoux–Rauzy word $\mathbf{u}$ is an image of a standard Arnoux–Rauzy word $\mathbf{u}'$ under a morphism φ_i, where the letter i coincides with the dominant letter of $\mathbf{u}$. This property enables us to assign to any standard Arnoux–Rauzy word a sequence $(i_n)_{n \geq 0}$ of indices and a sequence $\big(\mathbf{u}^{(n)}\big)_{n \geq 0}$ of standard Arnoux–Rauzy words such that

$$\mathbf{u} = \mathbf{u}^{(0)} \ \text{ and } \ \mathbf{u}^{(n)} = \varphi_{i_n}\big(\mathbf{u}^{(n+1)}\big) \ \text{ for each } n \in \mathbb{N}. \tag{3}$$

The sequence $(i_n)_{n \geq 0}$ is called the *directive sequence* of $\mathbf{u}$.

For any standard Arnoux–Rauzy word $\mathbf{u}$, both sequences $(i_n)_{n \geq 0}$ and $\big(\mathbf{u}^{(n)}\big)_{n \geq 0}$ are uniquely given. Moreover, every letter $i \in \mathcal{A}$ occurs in $(i_n)_{n \geq 0}$ infinitely many times. On the other hand, a sequence $(i_n)_{n \geq 0}$ which contains each letter of $\mathcal{A}$ infinitely many times determines a unique Arnoux–Rauzy word and thus the unique sequence $\big(\mathbf{u}^{(n)}\big)_{n \geq 0}$, cf. [15].

Example 3. *The most famous Sturmian word is the* Fibonacci word *which is the fixed point of so-called* Fibonacci morphism *defined as* $\tau : 0 \to 01,\ 1 \to 0$. *Analogously, for every integer* $d \geq 2$ *we define the* d-bonacci word $\mathbf{t}$ *as the fixed point of the* d-bonacci morphism

$$\tau \ : \ \begin{cases} \quad a \quad \to \ 0(a+1) \quad \textit{for } a = 0, \ldots, d-2\,, \\ (d-1) \ \to \ 0\,. \end{cases}$$

It is a d-ary standard Arnoux–Rauzy word. By simple computations we get $\tau^d = \varphi_0 \varphi_1 \cdots \varphi_{d-1}$ *and so its directive sequence* $(i_n)_{n \geq 0}$ *is* $(0\,1\,2 \cdots d-1)^\omega$, *i.e., its* n^{th} *element* $i_n \in \mathcal{A}$ *satisfies* $i_n \equiv n \bmod d$ *for any* $n \in \mathbb{N}$. *Over a ternary alphabet the word and the corresponding morphism is usually called* Tribonacci *word and morphism, respectively.*

4. Special factors and non-repetitive complexity

First we show the role that special factors play in the evaluation of non-repetitive complexity. Let us recall that for a given infinite word $\mathbf{u}$ we denoted by $f_n(i)$ the factor of length n occurring in $\mathbf{u}$ at the position i.

Lemma 4. *Let $\mathbf{u} = u_0u_1u_2\cdots$ be a recurrent aperiodic infinite word, $n \in \mathbb{N}$ and $m = nr\mathcal{C}_{\mathbf{u}}(n)$. Then there exists $h \in \mathbb{N}$ such that*

- *the set $L = \{f_n(h), f_n(h+1), \dots, f_n(h+m-1)\}$ contains m distinct factors of $\mathcal{L}_{\mathbf{u}}(n)$;*
- *the factor $f_n(h-1)$ is right special and belongs to L;*
- *the factor $f_n(h+m)$ is left special and belongs to L.*

Proof. Let k be an integer such that the factors from $L' = \{f_n(k), f_n(k+1), \dots, f_n(k+m-1)\}$ are pairwise distinct. As $\mathbf{u}$ is recurrent, we can assume $k \geq 1$. Since m is the maximal number of distinct consecutive factors, there exist integers i and j such that

$$k \leq i, j \leq k+m-1, \quad f_n(k-1) = f_n(i) \quad \text{and} \quad f_n(k+m) = f_n(j)\,.$$

We discuss two cases.

Case I: Assume $k < j$ and $i < k+m-1$. As $f_n(k+m) = f_n(j)$, the factors $f_{n+1}(k+m-1)$ and $f_{n+1}(j-1)$ of length $n+1$ have a common suffix of length n. It follows that $u_{k+m-1} \neq u_{j-1}$. Otherwise $f_n(k+m-1)$ and $f_n(j-1)$ would coincide, which is a contradiction with our choice of k. It means that $u_{k+m-1}f_n(j)$ and $u_{j-1}f_n(j)$ both belong to the language $\mathcal{L}_{\mathbf{u}}$. Thus $f_n(k+m) = f_n(j)$ is a left special factor. Analogously one can show that $f_n(k-1)$ is a right special factor. Thus we can choose $h = k$ and $L = L'$.

Case II: Assume $k = j$ or $i = k+m-1$. Without loss of generality we may assume $k = j$, i.e., $f_n(k) = f_n(k+m)$. Aperiodicity of $\mathbf{u}$ guarantees that there exists $\ell \in \mathbb{N}$ such that

$$f_n(k+q) = f_n(k+m+q) \ \text{ for each } q = 0, 1, \dots, \ell \quad \text{and} \quad f_n(k+\ell+1) \neq f_n(k+m+\ell+1).$$

Therefore $\{f_n(k+\ell+1), f_n(k+\ell+2), \dots, f_n(k+\ell+m)\} = L'$. We set $h = k+\ell+1$ and we show that the factor $f_n(h-1) = f_n(k+\ell)$ is right special and the factor $f_n(h+m) = f_n(k+\ell+m+1)$ is left special.

Since $f_n(k+\ell) = f_n(k+\ell+m)$ and $f_n(k+\ell+1) \neq f_n(k+\ell+m+1)$, the letters $u_{k+\ell+n}$ and $u_{k+\ell+m+n}$ differ. Hence, the factor $f_n(k+\ell)$ is right special. Since m is the maximal number of distinct consecutive factors, $f_n(k+\ell+m+1) \in L'$. By definition of ℓ, $f_n(k+\ell+m+1) \neq f_n(k+\ell+1)$, and so $f_n(k+\ell+m+1) = f_n(k+\ell+p)$ for some $1 < p \leq m$. We conclude that $f_n(k+\ell+m+1)$ is left special using the same arguments as in Case I. □

Theorem 5. *Let $\mathbf{u}$ be a Sturmian word. Then $nr\mathcal{C}_{\mathbf{u}}(n) = n+1$ for every $n \in \mathbb{N}$.*

Proof. Let $n \in \mathbb{N}$. Any Sturmian word $\mathbf{u}$ has exactly one left and one right special factor of length n. Let us denote them α and β, respectively. Therefore, in the Rauzy graph $\Gamma_{\mathbf{u}}(n)$ the vertex α has indegree 2 and all other vertices have indegree 1 and the vertex β has outdegree 2 and all others vertices have outdegree 1. Thus $\Gamma_{\mathbf{u}}(n)$ is a union of two cycles C_0 and C_1 which have a

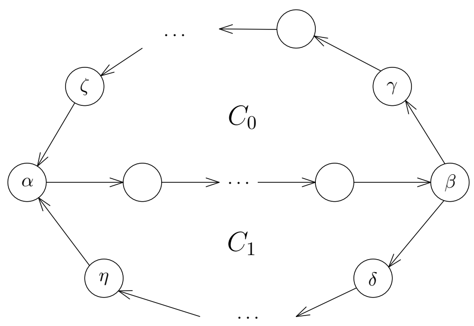

Figure 1: The Rauzy graph $\Gamma_{\mathbf{u}}(n)$ of the Sturmian word $\mathbf{u}$.

common part, namely the path from α to β (see Figure 1). Denote by $\gamma, \delta, \zeta, \eta$ the vertices such that $(\beta, \gamma), (\beta, \delta), (\zeta, \alpha)$ and (η, α) are edges in $\Gamma_{\mathbf{u}}(n)$.

By Lemma 4, let $h, m \in \mathbb{N}$ be such that $m = nr\mathcal{C}_{\mathbf{u}}(n)$, $L = \{f_n(h), f_n(h+1), \ldots, f_n(h+m-1)\}$, $\#L = nr\mathcal{C}_{\mathbf{u}}(n)$, $\alpha = f_n(h+m) \in L$ and $\beta = f(h-1) \in L$. Hence in the Rauzy graph $\Gamma_{\mathbf{u}}(n)$, there exists a path starting in a vertex of $N_+(\beta)$, passing through α and β, and ending in a vertex of $N_-(\alpha)$. Moreover, this path cannot pass twice through the same vertex. The only possible paths are the

$$\gamma \to \cdots \to \alpha \to \cdots \to \beta \to \delta \to \cdots \to \eta \text{ and } \delta \to \cdots \to \alpha \to \cdots \to \beta \to \gamma \to \cdots \to \zeta.$$

Both paths are hamiltonian, i.e., they are passing through all vertices of $\Gamma_{\mathbf{u}}(n)$ exactly once. It follows that $nr\mathcal{C}_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n) = n + 1$. □

The previous theorem states that the factor complexity and the non-repetitive complexity coincide for Sturmian words. In the next section we prove that this property is not preserved in d-ary Arnoux–Rauzy words with $d \geq 3$. Nevertheless, the equality $nr\mathcal{C}_{\mathbf{u}} = \mathcal{C}_{\mathbf{u}}$ we observed in binary aperiodic words with the smallest factor complexity can take place also in a word with the maximal factor complexity, as shown in [14]. The next corollary of Lemma 4 illustrates that the equality $nr\mathcal{C}_{\mathbf{u}} = \mathcal{C}_{\mathbf{u}}$ forces the Rauzy graphs of a word $\mathbf{u}$ to have a very special form.

Corollary 6. *Let $\mathbf{u}$ be a recurrent aperiodic word, $n \in \mathbb{N}$, $w \in \mathcal{L}_{\mathbf{u}}(n)$ and $m = nr\mathcal{C}_{\mathbf{u}}(n)$. Let $h \in \mathbb{N}$ be such that $f_n(h-1)$ and $f_n(h+m)$ are respectively the right and left special factors from Lemma 4. Assume $nr\mathcal{C}_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n)$.*

1. *If $w \neq f_n(h-1)$, then $N_+(w)$ contains at least $\#N_+(w) - 1$ left special factors.*
2. *If $w \neq f_n(h+m)$, then $N_-(w)$ contains at least $\#N_-(w) - 1$ right special factors.*

Proof. If the factor w is not right special, then the set $N_+(w)$ consists of one element and the statement is trivial. Let $w \neq f_n(h-1)$ be a right special factor. We write it in the form $w = as$, where $a \in \mathcal{A}$ and $s \in \mathcal{L}_{\mathbf{u}}(n-1)$. We denote $q = \#N_+(w)$ and find distinct letters $b_1, b_2, \ldots, b_q$ such that $N_+(w) = \{sb_1, sb_2, \ldots, sb_q\}$. Obviously, $asb_k \in \mathcal{L}_{\mathbf{u}}(n+1)$ for each $k = 1, 2, \ldots, q$. The assumption $nr\mathcal{C}_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n)$ implies that sb_k occurs in the set L described in Lemma 4. It means that $f_n(h+j_k) = sb_k$ for some index j_k, $0 \leq j_k \leq m-1$. Moreover, there exists an index p, $0 \leq p \leq m-1$ such that $f_n(h+p) = w = as$.

Let us look at the letter which precedes $f_n(h+j_k) = sb_k$, i.e., at the letter u_{h+j_k-1}:

– if $h = h + j_k$, then $u_{h+j_k-1} \neq a$ as $w = as \neq f_n(h-1)$;

– if $h \neq h+j_k \neq h+p+1$, then $u_{h+j_k-1} \neq a$, otherwise the factor $as = f_n(h+p) = f_n(h+j_k-1)$ occurs twice in L, which is a contradiction.

We showed that for all $k = 1, 2, \ldots, q$ (up to one possible exception when $h + j_k = h + p + 1$), the factors asb_k and $u_{h+j_k-1}sb_k$ belong to the language of $\mathbf{u}$ and $u_{h+j_k-1} \neq a$. It means that sb_k is a left special factor.

The proof of the second part of the statement is analogous. $\square$

5. Non-repetitive complexity of Arnoux–Rauzy words

For every Arnoux–Rauzy word $\mathbf{u}$, there exists at most one bispecial factor of $\mathbf{u}$ of length n. Thus we can order the bispecial factors by their lengths: for $k \in \mathbb{N}$ we denote $B_{\mathbf{u}}(k)$ the k^{th} bispecial factor of $\mathbf{u}$. In particular, $B_{\mathbf{u}}(0) = \varepsilon$, $B_{\mathbf{u}}(1) = u_0$ (the first letter of $\mathbf{u}$), etc.

Now we can formulate the link between the lengths of the return words to the bispecial factors and the values of non-repetitive complexity. Let us recall that any factor of a d-ary Arnoux–Rauzy word $\mathbf{u}$ has exactly d return words, cf. [10].

Proposition 7. *Let $\mathbf{u}$ be a d-ary Arnoux–Rauzy word and let $n, k \in \mathbb{N}$ be such that $B_{\mathbf{u}}(k-1) < n \leq B_{\mathbf{u}}(k)$. Denote by $r_0, r_1, \ldots, r_{d-1}$ the return words to $B_{\mathbf{u}}(k)$ in $\mathbf{u}$.*

1. *If $n = |B_{\mathbf{u}}(k)|$, then*

$$nr\mathcal{C}_{\mathbf{u}}(n) = \max\{|r_i r_j| \,:\, r_i r_j \in \mathcal{L}_{\mathbf{u}},\ 0 \leq i, j \leq d-1, i \neq j\} - 1\,.$$

2. *If $|B_{\mathbf{u}}(k-1)| < n < |B_{\mathbf{u}}(k)|$, then*

$$nr\mathcal{C}_{\mathbf{u}}(n) = nr\mathcal{C}_{\mathbf{u}}\big(|B_{\mathbf{u}}(k)|\big) - |B_{\mathbf{u}}(k)| + n\,.$$

Proof. Let $\mathbf{u}$ be a d-ary Arnoux–Rauzy word. Its n^{th} Rauzy graph $\Gamma_{\mathbf{u}}(n)$ contains exactly one vertex α with the indegree d and all other vertices have indegree 1. It also contains exactly one vertex β with the outdegree d, all other vertices have outdegree 1. It means that $\Gamma_{\mathbf{u}}(n)$ is composed of d cycles $C_0, C_1, \ldots, C_{d-1}$ which only have in common the path from α to β (see Figure 2). For every $i \in \{0, \ldots, d-1\}$, we denote by ℓ_i the number of vertices in the cycle C_i and by γ_i, ζ_i the vertices from the cycle C_i such that $(\beta, \gamma_i), (\zeta_i, \alpha)$ are edges in $\Gamma_{\mathbf{u}}(n)$. Let p be the number of vertices on the minimal path from α to β.

Let $h \in \mathbb{N}$ be such that $L = \{f_n(h), f_n(h+1), \ldots, f_n(h+m-1)\}$ is the set from Lemma 4 with $m = \#L = nr\mathcal{C}_{\mathbf{u}}(n)$. Then $f_n(h-1) = \beta$, $f_n(h+m) = \alpha$ and $\alpha, \beta \in L$. Hence the path in $\Gamma_{\mathbf{u}}(n)$ corresponding to L is of the form:

$$\gamma_i \to \cdots \to \zeta_i \to \alpha \to \cdots \to \beta \to \gamma_j \to \cdots \to \zeta_j$$

for some $i, j \in \mathcal{A}, i \neq j$, and it contains $nr\mathcal{C}_{\mathbf{u}}(n) = \ell_i + \ell_j - p$ vertices. So it suffices to compute the numbers ℓ_i, ℓ_j and p.

(1): If $n = |B_{\mathbf{u}}(k)|$, then $\alpha = \beta = B_{\mathbf{u}}(k)$, $p = 1$ and the Rauzy graph $\Gamma_{\mathbf{u}}(n)$ contains d cycles $C_0, C_1, \ldots, C_{d-1}$ with only the vertex $B_{\mathbf{u}}(k)$ in common. Clearly, these cycles correspond with the return words to $B_{\mathbf{u}}(k)$: if we start in $B_{\mathbf{u}}(k)$ and concatenate the first letters of all vertices

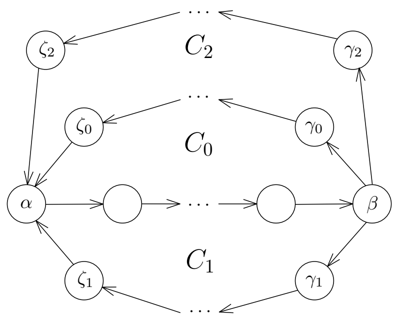

Figure 2: The Rauzy graph $\Gamma_{\mathbf{u}}(n)$ of a ternary Arnoux–Rauzy word $\mathbf{u}$.

of C_i, we get r_i for all $i \in \mathcal{A}$. Thus the number ℓ_i of vertices in C_i is equal to $|r_i|$. Hence $nr\mathcal{C}_{\mathbf{u}}(n) = \ell_i + \ell_j - 1 = |r_i| + |r_j| - 1$ for some $i \neq j$. By definition of $nr\mathcal{C}_{\mathbf{u}}(n)$, we have to choose $i \neq j$ such that the word $r_i r_j$ is a factor of $\mathbf{u}$ and its length is maximal possible.

(2): If $|B_{\mathbf{u}}(k-1)| < n < |B_{\mathbf{u}}(k)|$, then $\alpha \neq \beta$ and $p > 1$. Observe that if $p > 1$, then $\Gamma_{\mathbf{u}}(n+1)$ has also cycles of lengths ℓ_i for all $i \in \mathcal{A}$ and the minimal path from the left special factor to the right special factor contains $p-1$ vertices. It follows that $\Gamma_{\mathbf{u}}(n+p-1)$ contains the bispecial factor $B_{\mathbf{u}}(k)$ and so $|B_{\mathbf{u}}(k)| = n + p - 1$. By the Rauzy graph $\Gamma_{\mathbf{u}}(|B_{\mathbf{u}}(k)|)$ we have $nr\mathcal{C}_{\mathbf{u}}(|B_{\mathbf{u}}(k)|) = \ell_i + \ell_j - 1$, as the lengths of the cycles are preserved. So

$$nr\mathcal{C}_{\mathbf{u}}(n) = \ell_i + \ell_j - p = nr\mathcal{C}_{\mathbf{u}}(|B_{\mathbf{u}}(k)|) + 1 - (|B_{\mathbf{u}}(k)| - n + 1) = nr\mathcal{C}_{\mathbf{u}}(|B_{\mathbf{u}}(k)|) - |B_{\mathbf{u}}(k)| + n\,.$$

□

In the introduction we stated the inequality between the recurrence function $R_{\mathbf{u}}$ and the non-repetitive complexity. It is worth mentioning that $R_{\mathbf{u}}$ is also linked to return words, as stated by Cassaigne in [6].

Proposition 8 ([6])**.** *Let* $\mathbf{u}$ *be a recurrent infinite word. Then for each* $n \in \mathbb{N}$*,*

$$R_{\mathbf{u}}(n) - n + 1 = \max\{|r| \,:\, r \text{ is a return word to } w \in \mathcal{L}_{\mathbf{u}}(n)\}.$$

To transform Proposition 7 into an explicit formula for $nr\mathcal{C}_{\mathbf{u}}$, we have to compute the lengths of the return words to the bispecial factors in $\mathbf{u}$ and also decide which return words are neighbouring in $\mathbf{u}$. For this purpose we will essentially use the directive sequence $(i_n)_{n\geq 0}$ of a standard Arnoux-Rauzy word $\mathbf{u}$ introduced in Section 3. Let us emphasize that the non-repetitive complexity of $\mathbf{u}$ depends only on the language $\mathcal{L}_{\mathbf{u}}$ and not on the word $\mathbf{u}$ itself. Since for every Arnoux-Rauzy word $\mathbf{u}$ there exists a unique standard Arnoux-Rauzy word $\mathbf{v}$ such that $\mathcal{L}_{\mathbf{u}} = \mathcal{L}_{\mathbf{v}}$, we can restrict our considerations only to standard Arnoux–Rauzy words. Note that if $\mathbf{u}$ is standard Arnoux–Rauzy word, all its bispecial factors are prefixes of $\mathbf{u}$.

The following notion of *derived word* which codes the order of the return words in $\mathbf{u}$ will be also useful.

Definition 9. *Let* w *be a prefix of a uniformly recurrent word* $\mathbf{u}$ *and let* $r_0, r_1, \ldots, r_{\ell-1}$ *be the return words to* w *in* $\mathbf{u}$*. If we write* $\mathbf{u}$ *as a concatenation* $\mathbf{u} = r_{j_0} r_{j_1} r_{j_2} \cdots$*, then the word* $j_0 j_1 j_2 \cdots$ *is called the* derived word *to* w *in* $\mathbf{u}$ *and is denoted* $\mathrm{d}_{\mathbf{u}}(w)$*.*

We do not specify the order of the return words and thus the derived word is determined uniquely up to a permutation of letters. Clearly, the derived word to the empty word ε in $\mathbf{u}$ is the word $\mathbf{u}$ itself. The simple form of the morphisms φ_i defined by (2) gives immediately the following claim, which can be also deduced from the results in [10] or [11].

Claim 10. *Let $\mathbf{u}$ and $\mathbf{v}$ be standard d-ary Arnoux–Rauzy words such that $\mathbf{v} = \varphi_i(\mathbf{u})$ with $i \in \mathcal{A}$. Then for any $k \in \mathbb{N}$ it holds:*

- $B_{\mathbf{v}}(k+1) = \varphi_i\big(B_{\mathbf{u}}(k)\big)i$;
- *if $r_0, r_1, \ldots, r_{d-1}$ are the return words to $B_{\mathbf{u}}(k)$ in $\mathbf{u}$, then $\varphi_i(r_0), \varphi_i(r_1), \ldots, \varphi_i(r_{d-1})$ are the return words to $B_{\mathbf{v}}(k+1)$ in $\mathbf{v}$;*
- $\mathrm{d}_{\mathbf{u}}\big(B_{\mathbf{u}}(k)\big) = \mathrm{d}_{\mathbf{v}}\big(B_{\mathbf{v}}(k+1)\big)$ *up to permutation of letters.*

Corollary 11. *Let $\mathbf{u}$ be a standard Arnoux–Rauzy word with the directive sequence $(i_n)_{n\geq 0}$ and $\big(\mathbf{u}^{(n)}\big)_{n\geq 0}$ be the sequence satisfying (3). Then the derived word to $B_{\mathbf{u}}(k)$ in $\mathbf{u}$ is (up to permutation of letters) the word $\mathbf{u}^{(k)}$ and the corresponding return words are $\psi(0)$, $\psi(1), \ldots, \psi(d-1)$, where $\psi = \varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}$.*

Proof. Obviously, the bispecial factor ε has in the word $\mathbf{u}^{(k)}$ the return words $0, 1, \ldots, d-1$ and the derived word (up to permutation of letters) to $B_{\mathbf{u}^{(k)}}(0) = \varepsilon$ in $\mathbf{u}^{(k)}$ is $\mathbf{u}^{(k)}$. By repeated application of Claim 10 we get

$$\mathrm{d}_{\mathbf{u}}(B_{\mathbf{u}}(k)) = \mathrm{d}_{\mathbf{u}^{(1)}}(B_{\mathbf{u}^{(1)}}(k-1)) = \cdots = \mathrm{d}_{\mathbf{u}^{(k)}}(B_{\mathbf{u}^{(k)}}(0)) = \mathbf{u}^{(k)}$$

and

$$\{r_0, \ldots, r_{d-1}\} = \{\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(0), \ldots, \varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(d-1)\}\,.$$

□

Corollary 11 enable us to express the return words to the k^{th} bispecial factor $B_{\mathbf{u}}(k)$. However, we also need to know which return words are neighbouring, i.e., for which $i \neq j$ the word $r_i r_j$ is a factor of $\mathbf{u}$. Corollary 11 transforms this question to the description of neighbouring letters in the Arnoux–Rauzy word $\mathbf{u}^{(k)}$ with the directive sequence $(i_{n+k})_{n\geq 0}$, which is trivial.

Claim 12. *Let $\mathbf{u}$ be a standard Arnoux–Rauzy word with the directive sequence $(i_n)_{n\geq 0}$. Then i_0 is the dominant letter in $\mathbf{u}$ and the factors of length 2 in $\mathbf{u}$ are the words $i_0 a, a i_0$ for all $a \in \mathcal{A}$.*

For every $k \in \mathbb{N}$ and every letter $a \in \mathcal{A}$ we define $S_a(k) = \sup\{\ell : 0 \leq \ell < k, i_\ell = a\}$. As usual, if the set is empty, i.e., $i_\ell \neq a$ for all $\ell < k$, then $S_a(k) = -\infty$. Let us emphasize that $S_a(k) = S_b(k)$ for two distinct letters a and b if and only if $S_a(k) = S_b(k) = -\infty$.

Theorem 13. *Let $\mathbf{u}$ be a d-ary Arnoux–Rauzy word. For every integer $n \geq 1$ we take the unique k such that $|B_{\mathbf{u}}(k-1)| < n \leq |B_{\mathbf{u}}(k)|$. Then we have*

$$nr\mathcal{C}_{\mathbf{u}}(n) = |\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}\varphi_{i_k}(a)| - 1 - |B_{\mathbf{u}}(k)| + n\,,$$

where $(i_n)_{n\geq 0}$ is the directive sequence of the standard Arnoux-Rauzy word with the language $\mathcal{L}_{\mathbf{u}}$ and $a \in \mathcal{A}$ is any letter different from i_k such that $S_a(k) = \inf\{S_b(k) : b \in \mathcal{A}, b \neq i_k\}$.

Proof. Since the function $nr\mathcal{C}_{\mathbf{u}}$ depends only on the language $\mathcal{L}_{\mathbf{u}}$ and not on the word $\mathbf{u}$ itself, we can work with the standard Arnoux-Rauzy word $\mathbf{v}$ such that $\mathcal{L}_{\mathbf{v}} = \mathcal{L}_{\mathbf{u}}$ instead of $\mathbf{u}$. We denote $(i_n)_{n\geq 0}$ the directive sequence of $\mathbf{v}$ and to simplify the notation we also denote $\psi = \varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}$.

We start from Proposition 7 and using the previous claims we express $nr\mathcal{C}_{\mathbf{v}}(|B_{\mathbf{v}}(k)|)$ more explicitly. By Corollary 11 and Claim 12 the admissible pairs of return words to $B_{\mathbf{v}}(k)$ are

$$\{|r_ir_j| \,:\, r_ir_j \in \mathcal{L}_{\mathbf{v}},\, 0 \leq i,j \leq d-1, i \neq j\} = \{|\psi(i_k a)| : a \in \mathcal{A}, a \neq i_k\}\,.$$

It suffices to determine for which letter $a \neq i_k$ the image $|\psi(a)|$ is the longest possible. Let us emphasize that for every $i \in \{0,\dots,d-1\}$ and every words $x, y \in \mathcal{A}^*$ we have

(i) $\varphi_i(xa) = \varphi_i(x)ia$ if $i \neq a$ and $\varphi_i(xa) = \varphi_i(x)i$ if $i = a$;

(ii) if x is a proper prefix of y, i.e., $y = xz$ for some non-empty word z, then $\varphi_i(x)$ is a proper prefix of $\varphi_i(y) = \varphi_i(x)\varphi_i(z)$.

For two distinct letters $a, b \in \mathcal{A}$ we discuss two cases.

– If $S_a(k) = S_b(k)$, then the morphisms φ_a, φ_b are not included in the decomposition of ψ. Thus by application of Item (i) we get $\psi(a) = x'a$ and $\psi(b) = x'b$ for some non-empty word $x' \in \mathcal{A}^*$ and so $|\psi(a)| = |\psi(b)|$.

– If $S_a(k) < S_b(k)$, then we split $\psi = \sigma\varphi_b\theta$ such that the decomposition of the morphism θ contains neither φ_a nor φ_b. Then by Item (i) we have $\theta(a) = x'a$ and $\theta(b) = x'b$ for some non-empty word $x' \in \mathcal{A}^*$ and since $\varphi_b(x'a) = \varphi_b(x')ba$ and $\varphi_b(x'b) = \varphi_b(x')b$, the word $\varphi_b(\theta(b))$ is a proper prefix of $\varphi_b(\theta(a))$. By Item (ii) it means that also $\sigma(\varphi_b(\theta(b))) = \psi(b)$ is a proper prefix of $\sigma(\varphi_b(\theta(a))) = \psi(a)$ and so $|\psi(b)| < |\psi(a)|$.

We may conclude that

$$nr\mathcal{C}_{\mathbf{v}}(|B_{\mathbf{v}}(k)|) = \max\{|r_ir_j| \,:\, r_ir_j \in \mathcal{L}_{\mathbf{v}},\, 0 \leq i,j \leq d-1, i \neq j\} - 1 = |\psi(i_k a)| - 1 = |\psi\varphi_{i_k}(a)| - 1\,,$$

where a is any letter different from i_k such that $S_a(k) = \inf\{S_b(k) : b \neq i_k\}$. By Proposition 7 it concludes the proof, since for all $n, k \in \mathbb{N}$ we clearly have $B_{\mathbf{u}}(k) = B_{\mathbf{v}}(k)$ and $nr\mathcal{C}_{\mathbf{u}}(n) = nr\mathcal{C}_{\mathbf{v}}(n)$. □

6. Initial non-repetitive complexity of standard Arnoux–Rauzy words

The following lemma uses again the notation $f_n(i)$ for the factor of length n occurring in $\mathbf{u}$ at the position i.

Lemma 14. *Let $\mathbf{u} = u_0u_1u_2\cdots$ be a recurrent infinite word, $n \in \mathbb{N}$ and $m = inr\mathcal{C}_{\mathbf{u}}(n)$. Then the set $L = \{f_n(0), f_n(1), \dots, f_n(m-1)\}$ contains m distinct factors of $\mathcal{L}_{\mathbf{u}}(n)$ and the factor $f_n(m)$ is either left special and $f_n(m) = f_n(i)$ for some $i, 0 < i < m$, or $f_n(m) = f_n(0)$.*

Proof. The proof of Case I of Lemma 4 immediately gives this statement. □

Theorem 15. *Let $\mathbf{u}$ be a standard d-ary Arnoux–Rauzy word with the directive sequence $(i_n)_{n\geq 0}$. For every integer $n \geq 1$ we take the unique k such that $|B_{\mathbf{u}}(k-1)| < n \leq |B_{\mathbf{u}}(k)|$. Then we have*

$$inr\mathcal{C}_{\mathbf{u}}(n) = |\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k)|\,.$$

Proof. Let $\mathbf{u}$ be a standard d-ary Arnoux–Rauzy word and $n \in \mathbb{N}$. We denote $m = inr\mathcal{C}_{\mathbf{u}}(n)$ and $L = \{f_n(0), \ldots, f_n(m-1)\}$ the set from Lemma 14. Then $f_n(m) = f_n(0)$, since the word $f_n(0)$ is the only left special factor of $\mathbf{u}$ of length n. It means that m is equal to the length of the first return word to $f_n(0)$. We now determine its length.

If $n = |B_{\mathbf{u}}(k)|$ for some $k \in \mathbb{N}$, it means that $f_n(0) = B_{\mathbf{u}}(k)$ is bispecial factor. Then by Corollary 11 the first return word to $f_n(0)$ is equal to the word $\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k)$, since the word $\mathbf{u}^{(k)}$ is standard and so it starts with its dominant letter, which is by Claim 12 the letter i_k. Thus $m = |\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k)|$.

If $|B_{\mathbf{u}}(k-1)| < n < |B_{\mathbf{u}}(k)|$, then $B_{\mathbf{u}}(k) = f_n(0)w$ for some non-empty word $w \in \mathcal{A}^*$ since all prefixes of $\mathbf{u}$ are left special factors. Moreover, the word $f_n(0)$ is in $\mathbf{u}$ always followed by the word w. Indeed, since $f_n(0)$ is not right special, there is a unique letter $a \in \mathcal{A}$ such that $f_n(0)a \in \mathcal{L}_{\mathbf{u}}$ and we can repeat the same process until we reach $B_{\mathbf{u}}(k)$. But it means that the words $f_n(0)$ and $B_{\mathbf{u}}(k)$ have the same return words and derived words and so the first return word to $f_n(0)$ is equal to the word $\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k)$. Thus $m = |\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k)|$. $\square$

Let us emphasize that for non-standard Arnoux–Rauzy words the evaluating of the initial non-repetitive complexity is much more complicated, as, unlike the standard case, we do not have the control over the positions of the vertices corresponding to prefixes in the respective Rauzy graphs.

Corollary 16. *Let $\mathbf{u}$ be a standard Sturmian word. Then $inr\mathcal{C}_{\mathbf{u}}(n) = n+1$ for infinitely many $n \in \mathbb{N}$.*

Proof. Let $(i_\ell)_{\ell\geq 0}$ denote the directive sequence of $\mathbf{u}$. We will prove that $inr\mathcal{C}_{\mathbf{u}}(n) = n+1$ for every n such that $n = |B_{\mathbf{u}}(k)| + 1$ for some $k \in \mathbb{N}$ and $i_k \neq i_{k+1}$. Since the directive sequence $(i_\ell)_{\ell\geq 0}$ contains both letters 0 and 1 infinitely many times, it implies the statement of the corollary.

We take $n = |B_{\mathbf{u}}(k)| + 1$ such that $i_k \neq i_{k+1}$ and denote r_0 the more frequent return word to $B_{\mathbf{u}}(k)$ and r_1 the other return word. By Corollary 11 and Claim 12 we have $r_0 = \varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k)$ and $r_1 = \varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_{k+1})$. It also implies that r_1 is always followed by r_0, while r_0 can be followed both by r_0 and r_1.

As explained before, the Rauzy graph $\Gamma_{\mathbf{u}}(n-1)$ is composed of two cycles C_0 and C_1 with only the vertex $B_{\mathbf{u}}(k)$ in common (see Figure 3). Moreover, these cycles correspond with the return words r_0 and r_1: if we start in $B_{\mathbf{u}}(k)$ and concatenate the first letters of vertices from C_0, we get the return word r_0. Thus the number of vertices of C_0 is equal to $|r_0|$. It is analogous for C_1 and r_1. This connection also means that the cycle C_1 is always followed by C_0, while C_0 can be followed by both C_0 and C_1.

We denote α the edge from the cycle C_0 outcoming from the vertex $B_{\mathbf{u}}(k)$ and β the edge from C_0 incoming to $B_{\mathbf{u}}(k)$ (see Figure 3). Then α is the left special factor of $\mathbf{u}$ of length n and β is the right special factor of $\mathbf{u}$ of length n. It means that the Rauzy graph $\Gamma_{\mathbf{u}}(n)$ is composed of the cycle with $|r_0| + |r_1|$ vertices and one extra edge going from the vertex β to the vertex α (see Figure 3). It follows that $|r_0| + |r_1| = n+1$. Moreover, the return words to the factor α are r_0 and r_0r_1 and $|r_0r_1| = |r_0| + |r_1| = n+1$. Finally, it suffices to apply Theorem 15 for $B_{\mathbf{u}}(k) < n < B_{\mathbf{u}}(k+1)$ such that $i_k \neq i_{k+1}$:

$$inr\mathcal{C}_{\mathbf{u}}(n) = |\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_k}(i_{k+1})| = |\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k i_{k+1})| = |r_0r_1| = n+1.$$

$\square$

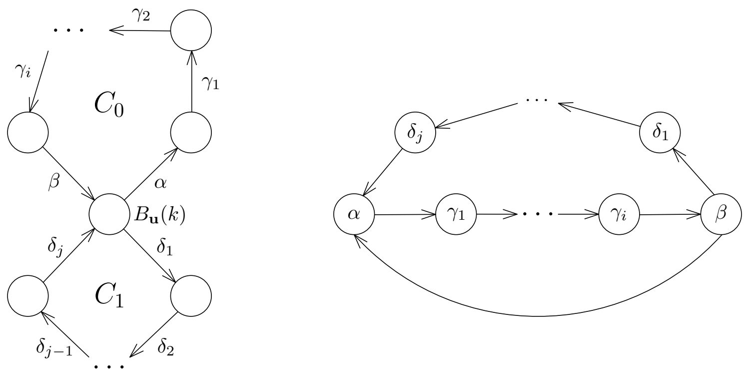

Figure 3: The Rauzy graphs $\Gamma_{\mathbf{u}}(n-1)$ (left) and $\Gamma_{\mathbf{u}}(n)$ (right) of the standard Sturmian word $\mathbf{u}$ for $n = |B_{\mathbf{u}}(k)|+1$.

Recently, Bugeaud and Kim [5] proved the new characterization of Sturmian words using the initial non-repetitive function: an infinite word $\mathbf{u}$ is Sturmian if and only if $inr\mathcal{C}_{\mathbf{u}}(n) \leq n+1$ for all $n \in \mathbb{N}$ with the equality for infinitely many n. So their result is more general than the previous corollary.

7. Enumeration of non-repetitive complexity for d-bonacci word

In this section we demonstrate the usefulness of Theorems 13 and 15 on the d-bonacci words. Let us recall that the d-bonacci word $\mathbf{t}$ (see Example 3) is the fixed point of the morphism

$$\tau \, : \left\{ \begin{array}{ccl} a & \to & 0(a+1) \ \text{ for } a = 0, \ldots, d-2 \\ (d-1) & \to & 0 \end{array} \right. \quad \text{with the matrix } \boldsymbol{M} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

In the sequel, we will use so-called *d-bonacci numbers*, which are the natural generalizations of the famous Fibonacci numbers. The sequence of the d-bonacci numbers $(D_k)_{k\geq 0}$ is defined by the linear recurrence:

$$D_k = \sum_{j=1}^{d} D_{k-j} \ \text{ for } k \geq d \quad \text{ and } \quad D_k = 2^k \ \text{ for all } k = 0, 1, \ldots, d-1\,.$$

Equivalently, the d-bonacci numbers can be expressed using the matrix recurrence. To simplify the notation we put $D_{-1} = 1$ and $D_{-k} = 0$ for all $k = 2, \ldots, d$. We also denote the vector $\vec{D}(n) = (D_n, D_{n-1}, \ldots, D_{n-d+1})^\top$ for all $n \geq -1$. Then the recurrence relation for the d-bonacci numbers can be rewritten in the following vector form:

$$\vec{D}(n) = \boldsymbol{M}\vec{D}(n-1) \quad \text{ for all } n \in \mathbb{N} \quad \text{ and } \quad \vec{D}(-1) = (1, 0, \ldots, 0)^\top =: \vec{e},$$

where $\boldsymbol{M}$ is the matrix of the d-bonacci morphism τ. Obviously, we can write

$$\vec{D}(n) = \boldsymbol{M}^{n+1}\vec{e}. \tag{4}$$

The simple form of the morphism τ gives us immediately the relation between the consecutive bispecial factors in the d-bonacci word $\mathbf{t}$ (compare with Claim 10), which allows us to express the lengths of the bispecial factors of $\mathbf{t}$.

Claim 17. *For every $k \geq 1$ the bispecial factors of the d-bonacci word $\mathbf{t}$ fulfil the equation*

$$B_{\mathbf{t}}(k) = \tau(B_{\mathbf{t}}(k-1))0\,.$$

Lemma 18. *For every $k \in \mathbb{N}$ the k^{th} bispecial factor $B_{\mathbf{t}}(k)$ of the d-bonacci word $\mathbf{t}$ has the length*

$$|B_{\mathbf{t}}(k)| = \frac{1}{d-1}\sum_{i=0}^{d-1}(d-i)D_{k-i-1} - \frac{d}{d-1}\,, \qquad \textit{where } D_j \textit{ is the } j^{th} \textit{ d-bonacci number.}$$

Proof. We denote the Parikh vector of the k^{th} bispecial factor $\vec{V}(k)$. Then using Claim 17 and Relation (1) we may write:

$$\vec{V}(k) = \boldsymbol{M}\vec{V}(k-1) + \vec{e} \quad \text{and so} \quad \vec{V}(k) = \boldsymbol{M}^k\vec{V}(0) + (\boldsymbol{M}^{k-1} + \boldsymbol{M}^{k-2} + \cdots + \boldsymbol{M}^0)\vec{e}.$$

Since $B_{\mathbf{t}}(0) = \varepsilon$, it is $\vec{V}(0) = (0, \ldots, 0)^\top$ and

$$\vec{V}(k) = (\boldsymbol{M}^{k-1} + \boldsymbol{M}^{k-2} + \cdots + \boldsymbol{M}^0)\vec{e}.$$

If we multiply this equality by the matrix $(\boldsymbol{M} - \boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix, we get:

$$(\boldsymbol{M} - \boldsymbol{I})\vec{V}(k) = (\boldsymbol{M}^k + \boldsymbol{M}^{k-1} + \cdots + \boldsymbol{M}^1 - \boldsymbol{M}^{k-1} - \cdots - \boldsymbol{M}^0)\vec{e} = \boldsymbol{M}^k\vec{e} - \vec{e}\,.$$

Finally, the application of Equation (4) gives us:

$$\vec{V}(k) = (\boldsymbol{M} - \boldsymbol{I})^{-1}\left(\boldsymbol{M}^k\vec{e} - \vec{e}\right) = (\boldsymbol{M} - \boldsymbol{I})^{-1}\left(\vec{D}(k-1) - \vec{e}\right)\,.$$

Now we can express the length of the k^{th} bispecial factor as:

$$|B_{\mathbf{t}}(k)| = (1, \ldots, 1) \cdot \vec{V}(k) = (1, \ldots, 1) \cdot (\boldsymbol{M} - \boldsymbol{I})^{-1}\left(\vec{D}(k-1) - \vec{e}\right)\,.$$

It suffices to compute the inverse matrix $(\boldsymbol{M} - \boldsymbol{I})^{-1}$. One can verify that it is

$$(\boldsymbol{M} - \boldsymbol{I})^{-1} = \frac{1}{d-1}\begin{pmatrix} 1 & d-1 & d-2 & \cdots & 2 & 1 \\ 1 & 0 & d-2 & \cdots & 2 & 1 \\ 1 & 0 & -1 & \cdots & 2 & 1 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 1 & 0 & -1 & \cdots & -d+3 & 1 \\ 1 & 0 & -1 & \cdots & -d+3 & -d+2 \end{pmatrix}$$

and thus $(1, \ldots, 1) \cdot (\boldsymbol{M} - \boldsymbol{I})^{-1} = \frac{1}{d-1}(d, d-1, d-2, \ldots, 1)$. Consequently,

$$\begin{aligned} |B_{\mathbf{t}}(k)| &= \frac{1}{d-1}(d, d-1, d-2, \ldots, 1)\vec{D}(k-1) - \frac{d}{d-1} \\ &= \frac{1}{d-1}\sum_{i=0}^{d-1}(d-i)D_{k-i-1} - \frac{d}{d-1}\,. \end{aligned}$$

$\square$

To find the simple expression for the lengths of the return words to $B_{\mathbf{t}}(k)$, we state one more auxiliary lemma. Let us remind that the d-bonacci word has the directive sequence $(0\,1\,2\cdots d-1)^\omega$, as explained in Example 3.

Lemma 19. *For the d-bonacci word with the directive sequence $(i_n)_{n\in\mathbb{N}} = (0\,1\,2\cdots d-1)^\omega$ and for every integer $k \geq 1$ we have*

$$|\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k)| = |\tau^k(0)| = D_k\,, \qquad \text{where } D_k \text{ is the } k^{th} \text{ } d\text{-bonacci number.}$$

Proof. One can simply verify that $\tau = \varphi_0 \circ P$, where P is a permutation such that $P(a) \equiv a+1 \mod d$ for all $a \in \{0,1,\ldots,d-1\}$. It is also easy to realize that $P \circ \varphi_a = \varphi_b \circ P$ for every $a, b \in \{0,1,\ldots,d-1\}$ such that $b \equiv a+1 \mod d$. These two facts give us

$$\tau^k = (\varphi_0 \circ P)^k = \varphi_0 \circ (P\circ\varphi_0)^{k-1} \circ P = \varphi_0 \circ (\varphi_1 \circ P)^{k-1} \circ P = \varphi_0\varphi_1 \circ (P \circ \varphi_1)^{k-2} \circ P^2 = \cdots$$
$$= \varphi_{j_0}\varphi_{j_1}\cdots\varphi_{j_{k-1}}P^k\,,$$

where $j_n \in \mathcal{A}$ and $j_n \equiv n \mod d$. But since the sequence $(j_n)_{n\in\mathbb{N}}$ is exactly the directive sequence of the d-bonacci word, i.e., $j_n = i_n$ for every $n \in \mathbb{N}$, we may conclude that

$$\tau^k(0) = \varphi_{j_0}\varphi_{j_1}\cdots\varphi_{j_{k-1}}P^k(0) = \varphi_{j_0}\varphi_{j_1}\cdots\varphi_{j_{k-1}}(j_k) = \varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k)\,.$$

It remains to prove that $|\tau^k(0)| = D_k$. We will prove that both sequences $(|\tau^n(0)|)_{n\in\mathbb{N}}$ and $(D_n)_{n\in\mathbb{N}}$ fulfil the same linear recurrence with the same initial conditions. In fact, we will show that for every $a \in \mathcal{A}$ the following equalities hold:

$$|\tau^k(a)| = \sum_{j=1}^{d-a}|\tau^{k-j}(0)| \ \text{ for all } k \geq d-a \quad \text{ and } \quad |\tau^k(a)| = 2^k \ \text{ for all } k = 0,\ldots,d-a-1\,. \tag{5}$$

We will proceed by induction on k. Simple computations verify the initial conditions. Now we suppose that the equality is true for $k-1$ and every letter $a \in \mathcal{A}$ and we prove that it is true also for k. If $a = d-1$, it is clear since $\tau^k(d-1) = \tau^{k-1}(0)$. If $a \neq d-1$, we rewrite as follows:

$$|\tau^k(a)| = |\tau^{k-1}(0)| + |\tau^{k-1}(a+1)| = |\tau^{k-1}(0)| + \sum_{j=1}^{d-a-1}|\tau^{k-1-j}(0)| = \sum_{j=1}^{d-a}|\tau^{k-j}(0)|\,.$$

If we consider the relations (5) for the letter $a = 0$, we get exactly the same recurrence as in the case of d-bonacci numbers. Thus these two sequences are the same and $|\tau^k(0)| = D_k$. $\square$

Theorem 20. *Let $\mathbf{t}$ be the d-bonacci word and let n, k be positive integers such that*

$$\frac{1}{d-1}\sum_{i=0}^{d-1}(d-i)D_{k-i-2} - \frac{d}{d-1} < n \leq \frac{1}{d-1}\sum_{i=0}^{d-1}(d-i)D_{k-i-1} - \frac{d}{d-1}\,.$$

Then

$$nr\mathcal{C}_{\mathbf{t}}(n) = D_{k+1} - 1 - \frac{1}{d-1}\sum_{i=0}^{d-1}(d-i)D_{k-i-1} + \frac{d}{d-1} + n\,.$$

Proof. It follows directly from Theorem 13. It suffices to replace the lengths of the bispecial factors by the expressions from Lemma 18 and determine the value of

$$|\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k a)|\,,$$

where $a \in \mathcal{A}$ is any letter such that $S_a(k) = \inf\{S_b(k) : b \in \mathcal{A}, b \neq i_k\}$. Since $\mathbf{t}$ has the directive sequence $(i_n)_{n\geq 0}$ given by $i_n \equiv n \mod d$, it is easy to realize that the desired letter a is the letter i_{k+1} (note that for $k < d-2$ there are also other possible choices of a). Then using Lemma 19 we get

$$|\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k a)| = |\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k i_{k+1})| = |\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_k}(i_{k+1})| = |\tau^{k+1}(0)| = D_{k+1}\,.$$

□

Theorem 21. *Let* $\mathbf{t}$ *be the d-bonacci word and let* n, k *be positive integers such that*

$$\frac{1}{d-1}\sum_{i=0}^{d-1}(d-i)D_{k-i-2} - \frac{d}{d-1} < n \leq \frac{1}{d-1}\sum_{i=0}^{d-1}(d-i)D_{k-i-1} - \frac{d}{d-1}\,.$$

Then $inr\mathcal{C}_{\mathbf{t}}(n) = D_k\,.$

Proof. It follows directly from Theorem 15. It suffices to realize that by Lemma 18 we know the lengths of the bispecial factors of $\mathbf{t}$ and by Lemma 19 we have $|\varphi_{i_0}\varphi_{i_1}\cdots\varphi_{i_{k-1}}(i_k)| = |\tau^k(0)| = D_k$. □

Note that for $d = 2$ and $d = 3$ the previous theorem gives the results stated in [14] as Theorems 10 and 16.

Corollary 22. *Let* $\mathbf{f}$ *and* $\mathbf{t}$ *be the Fibonacci and the Tribonacci word, respectively.*

- *Let* n, k *be positive integers such that* $F_k - 2 < n \leq F_{k+1} - 2$. *Then* $inr\mathcal{C}_{\mathbf{f}}(n) = F_k$, *where* F_k *is the* k^{th} *Fibonacci number.*
- *Let* n, k *be positive integers such that* $\frac{T_k+T_{k-2}-3}{2} < n \leq \frac{T_{k+1}+T_{k-1}-3}{2}$. *Then* $inr\mathcal{C}_{\mathbf{t}}(n) = T_k$, *where* T_k *is the* k^{th} *Tribonacci number.*

References

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

[2] L'. Balková, E. Pelantová, Š. Starosta: Sturmian Jungle (or Garden?) on Multiliteral Alphabets. RAIRO - Theoret. Inf. Appl. 44 (2010), 443–470.

[3] V. Berthé, J. Cassaigne, W. Steiner: Balance properties of Arnoux–Rauzy words. Int. J. of Algebra and Computation, Vol. 23, No. 04, (2013), 689–703.

[4] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, Ch. Reutenauer, G. Rindone: Acyclic, connected and tree sets. Monatshefte für Mathematik 176 (4) (2015).

[5] Y. Bugeaud, D. H. Kim: A new complexity function, repetitions in Sturmian words, and irrationality exponents of Sturmian numbers. Trans. Amer. Math. Soc. 371 (2019), 3281–3308.

[6] J.Cassaigne: Limit values of the recurrence quotient of Sturmian sequences. Theoretical Computer Science 218 (1999), 3–12.

[7] J. Cassaigne, N. Chekhova: Fonctions de récurrence des suites d' Arnoux–Rauzy et réponsé a une question de Morse et Hedlund. Annales de l'Institut Fourier 56 (2006) (7), 2249–2270.

[8] N. Pytheas Fogg: Substitutions in dynamics, arithmetics and combinatorics, Lecture Notes in Math. 1794 (2002), Springer-Verlag.

[9] A. Glen, J. Justin: Episturmian words: a survey. Theor. Inform. Appl. 43 (3) (2009), 403–442.

[10] J. Justin, L. Vuillon: Return words in Sturmian and episturmian words. RAIRO-Theoret. Inf. Apppl. 34 (2000), 343–356.

[11] K. Medková: Derived sequences of Arnoux–Rauzy sequences. In: R. Mercas and D. Reidenbach (eds.), Proceedings WORDS 2019, Lecture Notes in Computer Science 11682 (2019), Springer, pp. 251–263.

[12] T. K. S. Moothathu: Eulerian entropy and non-repetitivity subword complexity. Theoretical Computer Science 420 (2012), 80–88.

[13] M. Morse, G. A. Hedlund: Symbolic dynamics II: Sturmian trajectories. Amer. J. Math. 61 (1940), 1–42.

[14] J. Nicholson, N.Rampersad: Initial non-repetitivity complexity of infinite words. Discrete Appl. Math. 208 (C) (2016), 114–122.

[15] R. N. Risley and L. Q. Zamboni: A generalization of Sturmian sequences: Combinatorial structure and transcendence. Acta Arith. 95 (2000), 167–184.

Article E

Complementary symmetric Rote sequences: the critical exponent and the recurrence function

Discrete Mathematics and Theoretical Computer Science DMTCS vol. **VOL**:ISS, 2020, #NUM

*Complementary symmetric Rote sequences: the critical exponent and the recurrence function**

Ľubomíra Dvořáková Kateřina Medková Edita Pelantová

FNSPE, Czech Technical University in Prague, Czech republic

received xxxx-xx-xx, revised xxxx-xx-xx, accepted xxxx-xx-xx.

We determine the critical exponent and the recurrence function of complementary symmetric Rote sequences. The formulae are expressed in terms of the continued fraction expansions associated with the S-adic representations of the corresponding standard Sturmian sequences. The results are based on a thorough study of return words to bispecial factors of Sturmian sequences. Using the formula for the critical exponent, we describe all complementary symmetric Rote sequences with the critical exponent less than or equal to 3, and we show that there are uncountably many complementary symmetric Rote sequences with the critical exponent less than the critical exponent of the Fibonacci sequence. Our study is motivated by a conjecture on sequences rich in palindromes formulated by Baranwal and Shallit. Its recent solution by Curie, Mol, and Rampersad uses two particular complementary symmetric Rote sequences.

Keywords: critical exponent, recurrence function, Rote sequence, Sturmian sequence, return word, bispecial factor

1 Introduction

We study the relation between the critical exponents of two binary sequences $\mathbf{v} = v_0v_1v_2\cdots$ and $\mathbf{u} = u_0u_1u_2\cdots$ over the alphabet $\{0,1\}$, where $u_i = v_i + v_{i+1} \mod 2$ for each $i \in \mathbb{N}$. We write $\mathbf{u} = \mathcal{S}(\mathbf{v})$. Our study is motivated by a conjecture formulated by Baranwal and Shallit in [3]. They searched for binary sequences rich in palindromes with a minimum critical exponent. They showed that the value of this critical exponent is greater than 2.707. Moreover, they found two sequences $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ having the critical exponent equal to $2 + \frac{1}{\sqrt{2}}$ and they conjectured that this is the minimum value. Both of these sequences belong to the class of complementary symmetric Rote sequences. Their conjecture has been recently proved by Curie, Mol, and Rampersad in [7].

A Rote sequence is a binary sequence $\mathbf{v}$ containing $2n$ factors of length n for every $n \in \mathbb{N}, n \geq 1$. If the language of $\mathbf{v}$ is invariant under the exchange of letters $0 \leftrightarrow 1$, the sequence $\mathbf{v}$ is called a complementary

*The research received funding from the Ministry of Education, Youth and Sports of the Czech Republic through the project no. CZ.02.1.01/0.0/0.0/16_019/0000778 and from the Grant Agency of the Czech Technical University in Prague through the grant no. SGS20/183/OHK4/3T/14.

symmetric (CS) Rote sequence. Already in his original paper [22], Rote proved that these sequences are essentially connected with Sturmian sequences. He deduced that a binary sequence $\mathbf{v}$ is CS Rote sequence if and only if the sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$ is Sturmian. Both CS Rote sequences and Sturmian sequences are rich in palindromes, see [5, 9].

The formula for the critical exponent of Sturmian sequences was provided by Damanik and Lenz in [8]. The relation between the critical exponent of a CS Rote sequence $\mathbf{v}$ and the associated Sturmian sequence $\mathcal{S}(\mathbf{v})$ is not straightforward: While the minimum exponent among all Sturmian sequences is reached by the Fibonacci sequence and it is $3 + \frac{2}{1+\sqrt{5}}$ (see [17]), the two CS Rote sequences $\mathbf{v}^{(1)}$, $\mathbf{v}^{(2)}$ whose critical exponent equals $2 + \frac{1}{\sqrt{2}}$, i.e., it is minimum among all binary rich sequences, are associated with the Sturmian sequences $\mathcal{S}(\mathbf{v}^{(1)})$ and $\mathcal{S}(\mathbf{v}^{(2)})$ whose critical exponent is $3 + \sqrt{2}$.

In this paper, we will first derive the relation between the critical exponents of the sequences $\mathbf{v}$ and $\mathcal{S}(\mathbf{v})$, where $\mathbf{v}$ is a uniformly recurrent binary sequence whose language is closed under the exchange of letters, see Theorem 14. Using this relation, we will determine the formula for the critical exponent of any CS Rote sequence, see Theorem 33.

One of the consequences of this theorem is for instance the fact that the languages of the sequences $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ are the only languages of CS Rote sequences with the critical exponent less than 3, see Proposition 34. In this context, let us mention that in [7] the authors showed that there are exactly two languages of rich binary sequences with the critical exponent less than $\frac{14}{5}$ and they are the languages of the sequences $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$. Furthermore, we show that there are uncountably many CS Rote sequences with the critical exponent strictly less than the critical exponent of the Fibonacci sequence, see Theorem 37.

Our main technical tool is the description of return words to bispecial factors of Sturmian sequences in terms of the continued fraction expansions related to the S-adic representations of Sturmian sequences. As a by-product, we obtain an explicit formula for the recurrence function of CS Rote sequences, see Theorem 54. When formulating our results, we use the convergents $\left(\frac{p_N}{q_N}\right)$ of an irrational number $\theta = [0, a_1, a_2, a_3, \ldots]$, where the coefficients a_i's in the continued fraction expansion of θ correspond to the S-adic representation of the standard Sturmian sequence associated to a given CS Rote sequence.

There are many generalizations of Sturmian sequences to multiliteral alphabets, see [1]. The critical exponent and the recurrence function were studied for two of these generalizations. Justin and Pirillo described in [13] the critical exponent of substitutive Arnoux-Rauzy sequences. Recently, Rampersad, Shallit, and Vandomme in [20], and Baranwal and Shallit in [2] determined the minimal threshold for the critical exponent of balanced sequences over alphabets of cardinality $3, 4$, and 5, respectively. The recurrence function of Sturmian sequences was found by Morse and Hedlund in [18], and their result was generalized by Cassaigne and Chekhova in [6] for Arnoux-Rauzy sequences.

The paper is organized as follows. We first introduce basic notions from combinatorics on words in Section 2. In Section 3, we recall how to simplify the formula for the critical exponent using return words to bispecial factors. The definitions of the already mentioned mapping $\mathcal{S}$ and complementary symmetric Rote sequences and their basic properties are provided in Section 4. The relation between the critical exponents of the sequences $\mathbf{v}$ and $\mathcal{S}(\mathbf{v})$ is described in Section 5. The main tool for further results – a thorough study of return words to bispecial factors of Sturmian sequences using the S-adic representation – is carried out in Section 6. An explicit formula for the critical exponent of CS Rote sequences is given in Section 7. CS Rote sequences with a small critical exponent are studied in Section 8. And finally, in Section 9, an explicit formula for the recurrence function of CS Rote sequences is derived.

2 Preliminaries

An *alphabet* $\mathcal{A}$ is a finite set of symbols called *letters*. A *word* over $\mathcal{A}$ of *length* n is a string $u = u_0u_1\cdots u_{n-1}$, where $u_i \in \mathcal{A}$ for all $i \in \{0,1,\ldots,n-1\}$. The length of u is denoted by $|u|$. The set of all finite words over $\mathcal{A}$ together with the operation of concatenation form a monoid $\mathcal{A}^*$. Its neutral element is the *empty word* ε and we denote $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$.

If $u = xyz$ for some $x, y, z \in \mathcal{A}^*$, then x is a *prefix* of u, z is a *suffix* of u and y is a *factor* of u. We sometimes use the notation $yz = x^{-1}u$.

To any word u over $\mathcal{A}$ with the cardinality $\#\mathcal{A} = d$, we assign its *Parikh vector* $\vec{V}(u) \in \mathbb{N}^d$ defined as $(\vec{V}(u))_a = |u|_a$ for all $a \in \mathcal{A}$, where $|u|_a$ is the number of letters a occurring in u.

A *sequence* over $\mathcal{A}$ is an infinite string $\mathbf{u} = u_0u_1u_2\cdots$, where $u_i \in \mathcal{A}$ for all $i \in \mathbb{N} = \{0,1,2,\ldots\}$. We always denote sequences by bold letters. A sequence $\mathbf{u}$ is *eventually periodic* if $\mathbf{u} = vwww\cdots = v(w)^\omega$ for some $v \in \mathcal{A}^*$ and $w \in \mathcal{A}^+$. Otherwise $\mathbf{u}$ is *aperiodic*.

A *factor* of $\mathbf{u}$ is a word y such that $y = u_iu_{i+1}u_{i+2}\cdots u_{j-1}$ for some $i, j \in \mathbb{N}$, $i \le j$. The number i is called an *occurrence* of the factor y in $\mathbf{u}$. In particular, if $i = j$, the factor y is the empty word ε and any index i is its occurrence. If $i = 0$, the factor y is a *prefix* of $\mathbf{u}$. If each factor of $\mathbf{u}$ has infinitely many occurrences in $\mathbf{u}$, the sequence $\mathbf{u}$ is *recurrent*. Moreover, if for each factor the distances between its consecutive occurrences are bounded, $\mathbf{u}$ is *uniformly recurrent*.

The *language* $\mathcal{L}(\mathbf{u})$ of the sequence $\mathbf{u}$ is the set of all factors of $\mathbf{u}$. A factor w of $\mathbf{u}$ is *right special* if both words wa and wb are factors of $\mathbf{u}$ for at least two distinct letters $a, b \in \mathcal{A}$. Analogously we define a *left special* factor. A factor is *bispecial* if it is both left and right special. Note that the empty word ε is a bispecial factor if at least two distinct letters occur in $\mathbf{u}$.

The *factor complexity* of a sequence $\mathbf{u}$ is a mapping $\mathcal{C}_{\mathbf{u}} : \mathbb{N} \to \mathbb{N}$ defined by

$$\mathcal{C}_{\mathbf{u}}(n) = \#\{w \in \mathcal{L}(\mathbf{u}) : |w| = n\}\,.$$

The aperiodic sequences with the lowest possible factor complexity are called *Sturmian sequences*. In other words, it means that a sequence $\mathbf{u}$ is *Sturmian* if it has the factor complexity $\mathcal{C}_{\mathbf{u}}(n) = n + 1$ for all $n \in \mathbb{N}$. Clearly, all Sturmian sequences are defined over a binary alphabet, e.g., $\{0,1\}$. There are many equivalent definitions of Sturmian sequences, see a survey in [1].

A *morphism* over $\mathcal{A}$ is a mapping $\psi : \mathcal{A}^* \to \mathcal{A}^*$ such that $\psi(uv) = \psi(u)\psi(v)$ for all $u, v \in \mathcal{A}^*$. The morphism ψ can be naturally extended to sequences by

$$\psi(\mathbf{u}) = \psi(u_0u_1u_2\cdots) = \psi(u_0)\psi(u_1)\psi(u_2)\cdots\,.$$

A *fixed point* of a morphism ψ is a sequence $\mathbf{u}$ such that $\psi(\mathbf{u}) = \mathbf{u}$. The *matrix* of a morphism ψ over $\mathcal{A}$ with the cardinality $\#\mathcal{A} = d$ is the matrix $M_\psi \in \mathbb{N}^{d\times d}$ defined as $(M_\psi)_{ab} = |\psi(a)|_b$ for all $a, b \in \mathcal{A}$. The Parikh vector of the ψ-image of a word $w \in \mathcal{A}^*$ can be obtained via multiplication by the matrix M_ψ, i.e.,

$$\vec{V}(\psi(w)) = M_\psi \vec{V}(w)\,. \tag{1}$$

Consider a prefix w of a recurrent sequence $\mathbf{u}$. Let $i < j$ be two consecutive occurrences of w in $\mathbf{u}$. Then the word $u_iu_{i+1}\cdots u_{j-1}$ is a *return word* to w in $\mathbf{u}$. The set of all return words to w in $\mathbf{u}$ is denoted $\mathcal{R}_{\mathbf{u}}(w)$. If the sequence $\mathbf{u}$ is uniformly recurrent, the set $\mathcal{R}_{\mathbf{u}}(w)$ is finite for each prefix w, i.e., $\mathcal{R}_{\mathbf{u}}(w) = \{r_0, r_1, \ldots, r_{k-1}\}$. Then the sequence $\mathbf{u}$ can be written as a concatenation of these return words:

$$\mathbf{u} = r_{d_0}r_{d_1}r_{d_2}\cdots$$

and the *derived sequence* of $\mathbf{u}$ to the prefix w is the sequence $\mathbf{d_u}(w) = d_0 d_1 d_2 \cdots$ over the alphabet of cardinality $\#\mathcal{R}_\mathbf{u}(w) = k$. The concept of derived sequences was introduced by Durand in [11].

3 The critical exponent and its relation to return words

Let $z \in \mathcal{A}^+$ be a prefix of a periodic sequence u^ω with $u \in \mathcal{A}^+$. We say that z has the *fractional root* u and the *exponent* $e = |z|/|u|$. We usually write $z = u^e$. Let us emphasize that a word z can have multiple exponents and fractional roots. A word z is *primitive* if its only integer exponent is 1.

Let $\mathbf{u}$ be a sequence and u its non-empty factor. The supremum of $e \in \mathbb{Q}$ such that u^e is a factor of $\mathbf{u}$ is the *index* of u in $\mathbf{u}$:

$$\mathrm{ind}_\mathbf{u}(u) = \sup\{e \in \mathbb{Q} : u^e \in \mathcal{L}(\mathbf{u})\}\,.$$

If the sequence $\mathbf{u}$ is clear from the context, we will write $\mathrm{ind}(u)$ instead of $\mathrm{ind}_\mathbf{u}(u)$.

Definition 1. The *critical exponent* of a sequence $\mathbf{u}$ is

$$\begin{aligned}\mathrm{cr}(\mathbf{u}) &= \sup\{e \in \mathbb{Q} : \text{ there is a non-empty factor of } \mathbf{u} \text{ with the exponent } e\}\\ &= \sup\{\mathrm{ind}_\mathbf{u}(u) : u \text{ is a non-empty factor of } \mathbf{u}\}\,.\end{aligned}$$

Remark 2. Let us comment the above definition.

1. If a non-empty factor $u \in \mathcal{L}(\mathbf{u})$ is non-primitive, i.e., $u = x^k$ for some $x \in \mathcal{A}^+$ and $k \in \mathbb{N}, k \geq 2$, then $\mathrm{ind}_\mathbf{u}(x) = k\,\mathrm{ind}_\mathbf{u}(u) > \mathrm{ind}_\mathbf{u}(u)$. Therefore, only primitive factors play a role for finding $\mathrm{cr}(\mathbf{u})$.
2. If some non-empty factor occurs at least twice in $\mathbf{u}$, then $\mathrm{ind}(x) > 1$ for some non-empty factor x and so $\mathrm{cr}(\mathbf{u}) > 1$. Consequently, $\mathrm{cr}(\mathbf{u}) > 1$ for each sequence $\mathbf{u}$.
3. We say that u is an *overlapping* factor in $\mathbf{u}$, if there exist $x, y \in \mathcal{A}^*$ such that $xu = uy \in \mathcal{L}(\mathbf{u})$ and $0 < |x| < |u|$. If $\mathbf{u}$ has an overlapping factor, then $\mathrm{cr}(\mathbf{u}) > 2$. Indeed, by [15] the equality $xu = uy$ implies that there exist $a, b \in \mathcal{A}^*$ and $k \in \mathbb{N}$ such that $u = (ab)^k a$, $x = ab$, and $y = ba$. If a is empty then the assumption $|u| > |x| > 0$ forces $k \geq 2$, otherwise $k \geq 1$. In both cases $\mathrm{ind}_\mathbf{u}(ab) > 2$.
4. If $\mathbf{u}$ is eventually periodic, then $\mathrm{cr}(\mathbf{u})$ is infinite.
5. If $\mathbf{u}$ is aperiodic and uniformly recurrent, then each factor of $\mathbf{u}$ has a finite index. Nevertheless, $\mathrm{cr}(\mathbf{u})$ may be infinite. As an example of such a sequence may serve a Sturmian sequence, for which the coefficients in the continued fraction expansion of its slope are not bounded, see [8].
6. If $\mathbf{u}$ is a binary sequence, then either 11, 00, or 0101 occur in $\mathbf{u}$. It means that the critical exponent of a binary sequence is at least 2. This value is attained by the famous Thue-Morse sequence, which is, of course, overlap-free, see [23] or [4].

Lemma 3. *Let* $\mathbf{u}$ *be a uniformly recurrent aperiodic sequence. Then* $\mathrm{cr}(\mathbf{u}) = \sup\{\mathrm{ind}_\mathbf{u}(u) : u \in \mathcal{M}\}$, *where*

$$\mathcal{M} = \{u : u \textit{ is a return word to a bispecial factor of } \mathbf{u}\}\,.$$

Proof: Let $u \in \mathcal{L}(\mathbf{u})$ be a non-empty factor with the index $\mathrm{ind}(u) > 1$. Denote $|u| = n$. When searching for supremum, we may assume without loss of generality that u is a factor having the largest index among all factors of $\mathbf{u}$ of length n, i.e., $\mathrm{ind}(u) \geq \mathrm{ind}(v)$ for all $v \in \mathcal{L}(\mathbf{u})$ of length n. Since $\mathbf{u}$ is uniformly recurrent, $\mathrm{ind}(u)$ is finite. We denote

$$z = u^{\mathrm{ind}(u)} = u'u'' \cdots u'u''u' \quad \text{and} \quad b = u^{\mathrm{ind}(u)-1},$$

where $u = u'u''$ and $u'' \neq \varepsilon$. Clearly, $z = bu''u' = u'u''b$.

Let us show that the word b is a bispecial factor of $\mathbf{u}$. The word z is a factor of $\mathbf{u}$ and so z occurs in $\mathbf{u}$ at some position j, i.e.,

$$z = u_j u_{j+1} \cdots u_{j+|z|-1}.$$

Then the letter $u_{j+|z|}$ which follows the word z is distinct from the first letter of u''. Otherwise, we could prolong z to the right, which contradicts the definition of the index of $u = u_j u_{j+1} \cdots u_{j+n-1}$. Similarly, the letter u_{j-1} which precedes z is distinct from the last letter of u''. Indeed, if those letters are the same, then the factor $u_{j-1}u_j \cdots u_{j+n-2}$ of length n has the index at least $\mathrm{ind}(u) + \frac{1}{n}$, which contradicts the choice of u. We can conclude that the factor b is a bispecial factor of $\mathbf{u}$.

Moreover, since $z = bu''u' = u'u''b$, the word $u = u'u''$ is a concatenation of the return words to the bispecial factor b. It suffices to prove that only the cases when u is a return word to the bispecial factor b have to be inspected. Let us assume that u is a concatenation of at least two return words to b. It means that

$$z = ub = sbt \quad \text{for some } s, t \text{ such that } s \text{ is a prefix of } u \text{ and } 0 < |s| < |u|. \tag{2}$$

We will find another factor of $\mathbf{u}$ with the index strictly larger than $\mathrm{ind}(u)$, which means that such a factor u can be omitted. We distinguish three cases:

- If $\mathrm{ind}(u) \geq 2$, then $|b| \geq |u|$ and both words u and s are prefixes of b. Therefore, the relation (2) implies $us = su$ and we can easily conclude that there is a word x and an integer $k > 1$ such that $u = x^k$. As mentioned in Remark 2, $\mathrm{ind}(x) > \mathrm{ind}(u)$.
- If $1 < \mathrm{ind}(u) < 2$ and $\mathrm{cr}(\mathbf{u}) \leq 2$, then by Item (3) of Remark 2, $\mathbf{u}$ has no overlapping factor. Clearly, $z = u'u''u'$ and $b = u'$ for $u', u'' \neq \varepsilon$. Then the relation (2) implies $u'v = su'$ for some v and $|u| > |s| \geq |u'|$. Indeed, if $0 < |s| < |u'|$, then u' is an overlapping factor, which is not possible. Therefore, u' is a prefix of s and we can easily deduce that

$$\mathrm{ind}(s) \geq \frac{|s| + |u'|}{|s|} > \frac{|u| + |u'|}{|u|} = \mathrm{ind}(u).$$

- If $1 < \mathrm{ind}(u) < 2$ and $\mathrm{cr}(\mathbf{u}) > 2$, then there is a factor $x \in \mathcal{L}(\mathbf{u})$ with $\mathrm{ind}(x) > 2 > \mathrm{ind}(u)$.

□

Remark 4. In fact, we proved that it suffices to consider the set

$$\mathcal{M}' = \{u : u \text{ is a return word to a bispecial factor of } \mathbf{u} \text{ with the fractional root } u\}$$

or, even more specifically, the set

$$\mathcal{M}'' = \{u : u \text{ is a return word to the bispecial factor } b = u^{\mathrm{ind}(u)-1} \text{ of } \mathbf{u}\}$$

instead of $\mathcal{M}$. Clearly, $\mathcal{M}'' \subset \mathcal{M}' \subset \mathcal{M}$.

In Remark 2, we emphasize that only primitive factors are relevant for finding the critical exponent. Let us verify that all return words from the set $\mathcal{M}'$ (and so $\mathcal{M}''$, too) are primitive. We prove it by contradiction. Let us suppose that $u \in \mathcal{M}'$ is non-primitive, i.e., $u = x^k$ for some non-empty x and $k \in \mathbb{N}$, $k > 1$. Since b has the fractional root u, $b = u^\ell = x^{k\ell}$ for some $\ell \in \mathbb{Q}$. Therefore, $ub = x^{k(\ell+1)} = xx^{k\ell}x^{k-1} = xbx^{k-1}$, which contradicts that u is a return word to b in $\mathbf{u}$.

4 The mapping $\mathcal{S}$ on binary words and complementary symmetric Rote sequences

In this section, we introduce a mapping S which enables us to describe the properties of CS Rote sequences using Sturmian sequences. Nevertheless, this mapping S can be applied to any binary sequence.

Definition 5. By $\mathcal{S}$ we denote the mapping $\mathcal{S} : \{0,1\}^+ \mapsto \{0,1\}^*$ such that for every $v_0 \in \{0,1\}$ we put $\mathcal{S}(v_0) = \varepsilon$ and for every $v = v_0v_1 \cdots v_n \in \{0,1\}^+$ of length at least 2 we put $\mathcal{S}(v_0v_1 \cdots v_n) = u_0u_1 \cdots u_{n-1}$, where

$$u_i = v_i + v_{i+1} \mod 2 \quad \text{for all } \; i \in \{0,1,\ldots,n-1\}\,.$$

Moreover, we extend the domain of $\mathcal{S}$ naturally to $\{0,1\}^{\mathbb{N}}$: for every $\mathbf{v} \in \{0,1\}^{\mathbb{N}}$ we put $\mathcal{S}(\mathbf{v}) = \mathbf{u}$, where

$$u_i = v_i + v_{i+1} \mod 2 \quad \text{for all } \; i \in \mathbb{N}\,.$$

By $E : \{0,1\}^* \mapsto \{0,1\}^*$ we denote the morphism which exchanges the letters, i.e., $E(0) = 1$, $E(1) = 0$.

Example 6. We have $E(001110) = 110001$ and $\mathcal{S}(001110) = \mathcal{S}(110001) = 01001$.

Clearly, the images of v and $E(v)$ under $\mathcal{S}$ coincide for each $v \in \{0,1\}^*$. Moreover, $\mathcal{S}(x) = \mathcal{S}(y)$ if and only if $x = y$ or $x = E(y)$. The following rule follows directly from the definition of $\mathcal{S}$:

$$\mathcal{S}(v_0v_1 \cdots v_n) = \mathcal{S}(v_0v_1 \cdots v_k)\mathcal{S}(v_kv_{k+1} \cdots v_n) \quad \text{for any } k = 0,\ldots,n. \tag{3}$$

These observations hold also for infinite sequences.

Lemma 7. *Let $\mathbf{v}$ be a binary sequence whose language $\mathcal{L}(\mathbf{v})$ is closed under E. Then $w \neq \varepsilon$ is a right (left) special factor in $\mathbf{v}$ if and only if $\mathcal{S}(w)$ is a right (left) special factor in $\mathcal{S}(\mathbf{v})$.*

Proof: We will prove the statement for right special factors. The proof for left special factors is analogous. Let c be the last letter of w.
($\Longrightarrow$): Let $w0$ and $w1$ belong to $\mathcal{L}(\mathbf{v})$. Then $\mathcal{S}(w0)$ and $\mathcal{S}(w1)$ belong to $\mathcal{L}(\mathcal{S}(\mathbf{v}))$. By the rule (3), $\mathcal{S}(w0) = \mathcal{S}(w)\mathcal{S}(c0)$ and $\mathcal{S}(w1) = \mathcal{S}(w)\mathcal{S}(c1)$. As $\mathcal{S}(c0) \neq \mathcal{S}(c1)$, the factor $\mathcal{S}(w)$ is right special in $\mathcal{S}(\mathbf{v})$.
($\Longleftarrow$): Let $\mathcal{S}(w)0$ and $\mathcal{S}(w)1$ be in $\mathcal{L}(\mathcal{S}(\mathbf{v}))$. Then $\mathcal{S}(w)0 = \mathcal{S}(w)\mathcal{S}(cc) = \mathcal{S}(wc)$ and $\mathcal{S}(w)1 = \mathcal{S}(w)\mathcal{S}(cE(c)) = \mathcal{S}(wE(c))$. Since $\mathcal{L}(\mathbf{v})$ is closed under the exchange of letters, all factors wc, $E(wc)$, $wE(c)$ and $E(w)c$ are in $\mathcal{L}(\mathbf{v})$. It means that w and $E(w)$ are right special factors in $\mathbf{v}$.

$\square$

A *Rote sequence* is a sequence $\mathbf{v}$ with the factor complexity $\mathcal{C}_{\mathbf{v}}(n) = 2n$ for all $n \in \mathbb{N}, n \geq 1$. Clearly, all Rote sequences are defined over a binary alphabet, e.g., $\{0, 1\}$. If the language of a Rote sequence $\mathbf{v}$ is closed under the exchange of letters, i.e., $E(v) \in \mathcal{L}(\mathbf{v})$ for each $v \in \mathcal{L}(\mathbf{v})$, the Rote sequence $\mathbf{v}$ is called *complementary symmetric* (shortly CS).

Rote in [22] proved that these sequences are essentially connected with Sturmian sequences.

Proposition 8 ([22]). *Let $\mathbf{u}$ and $\mathbf{v}$ be two sequences over $\{0, 1\}$ such that $\mathbf{u} = \mathcal{S}(\mathbf{v})$. Then $\mathbf{v}$ is a complementary symmetric Rote sequence if and only if $\mathbf{u}$ is a Sturmian sequence.*

Let us emphasize that to a given CS Rote sequence $\mathbf{v}$ there is the unique *associated* Sturmian sequence $\mathbf{u}$ such that $\mathbf{u} = \mathcal{S}(\mathbf{v})$. On the other hand, for any Sturmian sequence $\mathbf{u}$ there exist two *associated* CS Rote sequences $\mathbf{v}$ and $E(\mathbf{v})$ such that $\mathbf{u} = \mathcal{S}(\mathbf{v}) = \mathcal{S}(E(\mathbf{v}))$. However, $\mathcal{L}(\mathbf{v}) = \mathcal{L}(E(\mathbf{v}))$.

Analogously, to a given factor $v \in \mathcal{L}(\mathbf{v})$ there is a unique associated word u such that $u = \mathcal{S}(v)$ and this word u is a factor of $\mathbf{u}$. In addition, to a given factor $u \in \mathcal{L}(\mathbf{u})$ there are exactly two associated words $v, E(v)$ such that $S(v) = \mathcal{S}(E(v)) = u$ and both these words $v, E(v)$ are factors of $\mathbf{v}$.

Example 9. Let us underline that for Sturmian sequences $\mathbf{u}$ and $E(\mathbf{u})$ the languages of their associated CS Rote sequences may essentially differ. Consider the Fibonacci sequence

$$\mathbf{f} = abaababaaba \cdots ,$$

which is the fixed point of the Fibonacci morphism $F : a \to ab$, $b \to a$.

- If $a = 0$ and $b = 1$, then the associated CS Rote sequence starting with 0 is $\mathbf{v} = 001110011100 \cdots$. The prefix w of $\mathbf{v}$ of length 7 is $w = 0011100 = (00111)^{\frac{7}{5}}$, i.e., w has the fractional root 00111 and the exponent $\frac{7}{5}$.
- If $a = 1$ and $b = 0$, then the associated CS Rote sequence starting with 0 is $\mathbf{v}' = 011011001001 \cdots$. The prefix w' of $\mathbf{v}'$ of length 7 is $w' = 0110110 = (011)^{\frac{7}{3}}$, i.e., w' has the fractional root 011 and the exponent $\frac{7}{3}$.

We will show later in Example 36 that even the critical exponent of CS Rote sequences associated with $\mathbf{u}$ and $E(\mathbf{u})$ may be different.

In the next section, we will explain that the relation between the shortest fractional root of a factor v and the shortest fractional root of $\mathcal{S}(v)$ is influenced by the number of letters 1 occurring in the shortest fractional root of $\mathcal{S}(v)$. This is the reason for the following definition and lemma.

Definition 10. A word $u = u_0 u_1 \cdots u_{n-1} \in \{0, 1\}^*$ is called *stable* if $|u|_1 = 0 \mod 2$. Otherwise, u is *unstable*.

Lemma 11. *Let $\mathcal{S} : \{0, 1\}^+ \mapsto \{0, 1\}^*$.*

(i) If 0 is a prefix of $v \in \{0, 1\}^*$, then $\mathcal{S}(v0)$ is stable.

(ii) For every $u \in \{0, 1\}^*$ there exists a unique $w \in \{0, 1\}^+$ with a prefix 0 such that $u = \mathcal{S}(w)$. Moreover, w has a suffix 0 if and only if u is stable.

(iii) If 0 is a prefix of w, then $\mathcal{S}(vw) = \mathcal{S}(v0)\mathcal{S}(w)$.

(iv) Let 0 be a prefix of v and v'. Then $\mathcal{S}(v')$ is a prefix of $\mathcal{S}(v)$ if and only if v' is a prefix of v.

Proof:

(i) Let $v = v_0 v_1 \cdots v_{n-1}$, $n = |v|$, and $v_0 = 0$. Put $v_n = 0$. Then $\mathcal{S}(v0) = u_0 u_1 \cdots u_{n-1}$, where $u_i = v_i + v_{i+1} \mod 2$ for every $i \in \{0, 1, \ldots, n-1\}$. It implies

$$|\mathcal{S}(v0)|_1 = \sum_{i=0}^{n-1} u_i = v_0 + v_n = 0 \mod 2\,.$$

(ii) Let $u = u_0 u_1 \cdots u_{m-1}$ and $m = |u|$. We look for $w = w_0 w_1 \cdots w_m$ such that $u_i = w_i + w_{i+1} \mod 2$ for every $i \in \{0, 1, \ldots, m-1\}$. Clearly, these equations can be equivalently rewritten as

$$u_i = w_{i+1} - w_i \mod 2 \quad \text{for every } i \in \{0, 1, \ldots, m-1\}\,. \tag{4}$$

Then starting with $w_0 = 0$ and summing up the equations (4) for $i = \{0, 1, \ldots, j-1\}$, we determine the letter w_j of w as $w_j = \sum_{i=0}^{j-1} u_i \mod 2$. In particular, $w_m = |u|_1 \mod 2$.

(iii) It is a particular case of the equation (3).

(iv) It follows directly from the definition of $\mathcal{S}$.

□

5 The relation between the indices of factors in $\mathbf{v}$ and $\mathcal{S}(\mathbf{v})$

In this section, we provide a tool for determining the critical exponent of a binary sequence $\mathbf{v}$ whose language is closed under the exchange of letters. For any factor v of such a sequence, $\text{ind}_{\mathbf{v}}(v) = \text{ind}_{\mathbf{v}}(E(v))$ and we can consider only factors of $\mathbf{v}$ starting with 0 without loss of generality.

Lemma 12. *Let $\mathbf{v}$ be a binary aperiodic uniformly recurrent sequence whose language is closed under E. Denote $\mathbf{u} = \mathcal{S}(\mathbf{v})$. For a non-empty factor $v \in \mathcal{L}(\mathbf{v})$ with the prefix 0 and $\text{ind}_{\mathbf{v}}(v) > 1$, there exists a stable factor $u \in \mathcal{L}(\mathbf{u})$ such that*

$$\text{ind}_{\mathbf{u}}(u) + \tfrac{1}{|u|} = \text{ind}_{\mathbf{v}}(v) \quad \textit{and} \quad u = \mathcal{S}(v0)\,. \tag{5}$$

And vice versa, for a non-empty stable factor $u \in \mathcal{L}(\mathbf{u})$, there exists a factor $v \in \mathcal{L}(\mathbf{v})$ with the prefix 0 satisfying (5).

Proof: For a given $n \in \mathbb{N}, n \geq 1$, consider the set $\mathcal{K}_n$ of factors $v \in \mathcal{L}(\mathbf{v})$ of length n with the prefix 0 and $\text{ind}_{\mathbf{v}}(v) > 1$. First, we show that the mapping $v \mapsto \mathcal{S}(v0)$ is a bijection between $\mathcal{K}_n$ and the set of all stable factors of $\mathbf{u}$ of length n.

Indeed, if $v \in \mathcal{K}_n$, then $v0 \in \mathcal{L}(\mathbf{v})$. The factor $u := \mathcal{S}(v0)$ belongs to $\mathcal{L}(\mathbf{u})$, $|u| = |v|$, and by Item (i) of Lemma 11, u is stable. On the other hand, if $u \in \mathcal{L}(\mathbf{u})$ is stable and of length n, then by Item (ii) of Lemma 11, there exists a unique w such that 0 is a prefix and a suffix of w and $\mathcal{S}(w) = u$. As $\mathcal{L}(\mathbf{v})$ is closed under E, necessarily $w \in \mathcal{L}(\mathbf{v})$ and $w = v0$ for some v with the prefix 0. In particular, $\text{ind}_{\mathbf{v}}(v) > 1$. As $u = \mathcal{S}(w) = \mathcal{S}(v0)$, the lengths of u and v coincide.

Now we show that any $v \in \mathcal{K}_n$ and its image $u = \mathcal{S}(v0)$ satisfy (5). Find $k \in \mathbb{N}, k \geq 1$, and $\theta \in (0,1]$ such that $\mathrm{ind}_{\mathbf{v}}(v) = k + \theta$. Denote $v' = v^\theta$. Obviously, $v' \neq \varepsilon$, v' is a prefix of v and 0 is a prefix of v'. Applying Item (iii) of Lemma 11, we get $\mathcal{S}(v^k v') = (\mathcal{S}(v0))^k \mathcal{S}(v')$. Clearly, $|\mathcal{S}(v')| = |v'| - 1$, $|u| = |v|$, and by Item (iv) of Lemma 11, $\mathcal{S}(v')$ is a prefix of $\mathcal{S}(v0)$. For $u = \mathcal{S}(v0)$ it means that

$$\mathrm{ind}_{\mathbf{u}}(u) \geq k + \tfrac{|S(v')|}{|u|} = k + \tfrac{|v'|}{|v|} - \tfrac{1}{|u|} = k + \theta - \tfrac{1}{|u|} = \mathrm{ind}_{\mathbf{v}}(v) - \tfrac{1}{|u|}\,.$$

To show the opposite inequality, we find $\ell \in \mathbb{N}$ and $\eta \in [0,1)$ such that $\mathrm{ind}_{\mathbf{u}}(u) = \ell + \eta$. Denote $u' = u^\eta$. Using Item (ii) of Lemma 11, we find v' with the prefix 0 such that $u' = \mathcal{S}(v')$. By Item (iv), v' is a prefix of v, and by Item (iii), $u^\ell u' = (\mathcal{S}(v0))^\ell \mathcal{S}(v') = \mathcal{S}(v^\ell v')$. Therefore, $v^\ell v' \in \mathcal{L}(\mathbf{v})$ and

$$\mathrm{ind}_{\mathbf{v}}(v) \geq \ell + \eta + \tfrac{1}{|u|} = \mathrm{ind}_{\mathbf{u}}(u) + \tfrac{1}{|u|}\,.$$

□

As explained in Lemma 3, only return words to bispecial factors play a role for the determination of the critical exponent of a sequence. More specifically, we can restrict ourselves to factors from the set $\mathcal{M}'$ (or $\mathcal{M}''$) introduced in Remark 4.

Lemma 13. *Let $\mathbf{v}$ be a binary sequence whose language is closed under E. Assume that v with the prefix 0 is a return word in $\mathbf{v}$ to a bispecial factor $b = v^{e-1}$, where $e > 2$. Denote $u = \mathcal{S}(v0)$. Then*
– either u is a stable return word in $\mathcal{S}(\mathbf{v})$ to a bispecial factor with the fractional root u;
– or $u = x^2$, where x is an unstable return word in $\mathcal{S}(\mathbf{v})$ to a bispecial factor with the fractional root u.

Proof: The factor $\mathcal{S}(b)$ is bispecial in $\mathcal{S}(\mathbf{v})$ by Lemma 7. Moreover, by the rule (3), we can write $\mathcal{S}(b) = \mathcal{S}(v^{e-1}) = \mathcal{S}(v0)^f$ for $f = e - 1 - \frac{1}{|v|} \geq 1$. Thus $\mathcal{S}(b)$ has the fractional root $u = \mathcal{S}(v0)$.

The word vb is a complete return word to b in $\mathbf{v}$ and thus $vb = bw$ for some w. Note that 0 is the first letter of b and denote z the last letter of b. By the rule (3), we get $\mathcal{S}(v0)\mathcal{S}(b) = \mathcal{S}(b)\mathcal{S}(zw)$. It means that $\mathcal{S}(b)$ is a prefix and a suffix of the word $\mathcal{S}(v0)\mathcal{S}(b)$. We discuss two cases:

– $\mathcal{S}(b)$ has exactly two occurrences in $\mathcal{S}(v0)\mathcal{S}(b)$, one as a prefix and one as a suffix. In this case $u = \mathcal{S}(v0)$ is a return word to $\mathcal{S}(b)$ in $S(\mathbf{v})$ and by Item (i) of Lemma 11, u is stable.

– $\mathcal{S}(b)$ occurs in $\mathcal{S}(v0)\mathcal{S}(b)$ as an inner factor. In this case, there exists a return word $u' \neq \varepsilon$ to $\mathcal{S}(b)$ such that $|u'| < |u|$ and $u'\mathcal{S}(b)$ is a proper prefix of $\mathcal{S}(v0)\mathcal{S}(b)$. We take the word b' such that $b = vb'$. Clearly, b' has the prefix 0 and so $\mathcal{S}(b) = u\mathcal{S}(b')$. Then $u'\mathcal{S}(b) = u'u\mathcal{S}(b')$ is a proper prefix of $\mathcal{S}(v0)\mathcal{S}(b) = u\mathcal{S}(b) = uu\mathcal{S}(b')$. In other words, $u'uu'' = uu$ for some non-empty u'' and consequently, $u = u'u'' = u''u'$. This implies the existence of $x \in \{0,1\}^+$ and $k', k'' \in \mathbb{N}, k', k'' \geq 1$ such that $u' = x^{k'}$ and $u'' = x^{k''}$. If we denote $k = k' + k'' \geq 2$, we can write $u = x^k$.

We show that x is unstable and $k = 2$. Indeed, assume x is stable, then by Item (ii) of Lemma 11, we find a unique y with the prefix 0 such that $x = \mathcal{S}(y0)$. Applying Item (iii), we obtain $\mathcal{S}(v0) = u = x^k = (\mathcal{S}(y0))^k = \mathcal{S}(y^k 0)$ and thus $v = y^k$. Nevertheless, the factor v is primitive as explained in Remark 4. Thus this is a contradiction.

Since x is unstable and $u = x^k$ is stable, necessarily $k = 2p$ for some integer $p \geq 1$. Now we deduce that $k = 2$. Indeed, if $p \geq 2$, then u is a p-power of the stable factor x^2 which yields a contradiction with the primitivity of v as above. Finally, $k = 2$ implies $k' = 1$ and $u' = x$ is an unstable return word to the bispecial factor $\mathcal{S}(b)$ in $S(\mathbf{v})$. □

Theorem 14. *Let* $\mathbf{v}$ *be a binary aperiodic uniformly recurrent sequence whose language is closed under* E*. Denote* $\mathbf{u} = \mathcal{S}(\mathbf{v})$*,*

$A_1 = \Big\{\mathrm{ind}_{\mathbf{u}}(u) + \frac{1}{|u|} : u \text{ is a stable return word to a bispecial factor of } \mathbf{u}\Big\}$ *and*

$A_2 = \Big\{\frac{1}{2}\big(\mathrm{ind}_{\mathbf{u}}(u) + \frac{1}{|u|}\big) : u \text{ is an unstable return word to a bispecial factor of } \mathbf{u}\Big\}$ *.*

Then

$$\mathrm{cr}(\mathbf{v}) = \sup\big(A_1 \cup A_2\big)\,.$$

Proof: First we show that

$$\mathrm{ind}_{\mathbf{v}}(v) \leq \sup\big(A_1 \cup A_2\big) \quad \text{for any non-empty } v \in \mathcal{L}(\mathbf{v})\,. \tag{6}$$

If $\mathrm{ind}_{\mathbf{v}}(v) \leq 2$, then the inequality (6) is trivially satisfied as A_1 contains the number $\mathrm{ind}_{\mathbf{u}}(0) + \frac{1}{|0|} \geq 2$ (note that 0 is a stable return word in $\mathbf{u}$ to the bispecial factor ε). Now we assume that $\mathrm{ind}_{\mathbf{v}}(v) = e > 2$. By Lemma 3 and Remark 4, we may focus only on v which is a return word to the bispecial factor $b = v^{e-1}$ and v has the prefix 0. By Lemma 12, $\mathrm{ind}_{\mathbf{v}}(v) = \mathrm{ind}_{\mathbf{u}}(u) + \frac{1}{|u|}$, where $u = \mathcal{S}(v0)$. By Lemma 13, the factor u is either a stable return word to a bispecial factor in $\mathbf{u}$, or $u = x^2$, where x is an unstable return word to a bispecial factor in $\mathbf{u}$. In the first case we have $\mathrm{ind}_{\mathbf{v}}(v) \leq \sup A_1$, while in the second case we have $\mathrm{ind}_{\mathbf{v}}(v) = \mathrm{ind}_{\mathbf{u}}(u) + \frac{1}{|u|} = \frac{1}{2}\mathrm{ind}_{\mathbf{u}}(x) + \frac{1}{2|x|} \leq \sup A_2$. We may conclude that $\mathrm{cr}(\mathbf{v}) \leq \sup\big(A_1 \cup A_2\big)$.

To prove the opposite inequality, we show

$$A_1 \cup A_2 \setminus [0,1] \subset \{\mathrm{ind}_{\mathbf{v}}(v) : v \in \mathcal{L}(\mathbf{v}), v \neq \varepsilon\}\,.$$

If $H \in A_1$, then there exists a stable factor u in $\mathbf{u}$ such that $H = \mathrm{ind}_{\mathbf{u}}(u) + \frac{1}{|u|}$, and by Lemma 12, we find v in $\mathbf{v}$ such that $H = \mathrm{ind}_{\mathbf{v}}(v)$. Analogously, if $H \in A_2$ and $H > 1$, then $\mathrm{ind}_{\mathbf{u}}(u) = 2H - \frac{1}{|u|} \geq 2$ for some unstable factor $u \in \mathcal{L}(\mathbf{u})$. Thus the word $y = uu \in \mathcal{L}(\mathbf{u})$, it is a stable factor of $\mathbf{u}$ and its index in $\mathbf{u}$ is $\frac{1}{2}\mathrm{ind}_{\mathbf{u}}(u)$. By Lemma 12, there is v in $\mathbf{v}$ such that $\mathrm{ind}_{\mathbf{v}}(v) = \mathrm{ind}_{\mathbf{u}}(y) + \frac{1}{|y|} = \frac{1}{2}\mathrm{ind}_{\mathbf{u}}(u) + \frac{1}{2|u|} = H$. □

Theorem 14 will be used in the next sections to determine the critical exponent of a complementary symmetric Rote sequence $\mathbf{v}$ by exploiting the indices of factors in the Sturmian sequence $\mathcal{S}(\mathbf{v})$. The following example shows an opposite application of Theorem 14. But before that, let us state a simple auxiliary statement reflecting the behaviour of fractional roots under the application of a morphism.

Observation 15. *Let* $\phi : \mathcal{A}^* \mapsto \mathcal{A}^*$ *be a morphism and let* $w \in \mathcal{A}^*$ *be a prefix of* $\phi(a)$ *for each* $a \in \mathcal{A}$*. If* u *is a fractional root of* z*, then* $\phi(u)$ *is a fractional root of* $\phi(z)w$*.*

Example 16. Let us consider the Thue–Morse sequence $\mathbf{t} = 01101001\cdots$, which is fixed by the morphism $\psi : 0 \mapsto 01$ and $1 \mapsto 10$. It is well-known that $\mathbf{t}$ is uniformly recurrent, its language is closed under the exchange of letters and $\mathrm{cr}(\mathbf{t}) = 2$. The corresponding sequence $\mathbf{u} = \mathcal{S}(\mathbf{t}) = 1011101\cdots$ is called the period doubling sequence and it is fixed by the morphism $\phi : 0 \mapsto 11$ and $1 \mapsto 10$, see [21].

We determine the critical exponent of $\mathbf{u}$. Theorem 14 implies $\mathrm{cr}(\mathbf{u}) \leq 4$, as otherwise $\mathrm{cr}(\mathbf{t}) > 2$, which is a contradiction. Now we show that the value 4 is attained.

By Observation 15 and the fact that both $\phi(0)$ and $\phi(1)$ have the prefix 1, the morphism ϕ has the following two properties:

1. If $w \in \mathcal{L}(\mathbf{u})$, then $\phi(w)1 \in \mathcal{L}(\mathbf{u})$.
2. If u is a fractional root of w, then $\phi(u)$ is a fractional root of $\phi(w)1$.

We will construct two sequences $\left(u^{(n)}\right)$ and $\left(w^{(n)}\right)$ of words belonging to $\mathcal{L}(\mathbf{u})$. We start with $u^{(0)} = 1$ and $w^{(0)} = 111 \in \mathcal{L}(\mathbf{u})$ and for each $n \in \mathbb{N}$ we define

$$w^{(n+1)} = \phi(w^{(n)})1 \qquad \text{and} \qquad u^{(n+1)} = \phi(u^{(n)})\,.$$

Note that $u^{(0)} = 1$ is a fractional root of $w^{(0)} = 111$. Because of the property (2), the word $u^{(n)}$ is a fractional root of $w^{(n)}$ for each $n \in \mathbb{N}$. Moreover, the specific form of the morphism ϕ implies $|u^{(n+1)}| = 2|u^{(n)}|$ and $|w^{(n+1)}| = 2|w^{(n)}| + 1$. It gives $|u^{(n)}| = 2^n$ and $|w^{(n)}| = 2^{n+2} - 1$. Therefore, $\mathrm{ind}_{\mathbf{u}}(u^{(n)}) \geq \frac{2^{n+2}-1}{2^n} \to 4$. We may conclude that $\mathrm{cr}(\mathbf{u}) = 4$.

6 Return words to bispecial factors of Sturmian sequences

The main goal of this article is to describe the critical exponent and the recurrence function of CS Rote sequences. Proposition 8 and Theorem 14 transform the first task to the computation of the indices of return words to bispecial factors in the associated Sturmian sequences.

This is a preparatory section for this computation. We introduce the directive sequence of a standard Sturmian sequence and recall some known results on bispecial factors, their return words, and derived sequences. It allows us to describe the longest factor of $\mathbf{u}$ with the fractional root u, where u is any return word to a bispecial factor of a Sturmian sequence $\mathbf{u}$ (Lemma 24). Further on, we explain how to express the lengths of these factors explicitly (Proposition 30), and eventually in Section 7, we determine the indices.

First, we recall that a binary sequence $\mathbf{u} \in \{0,1\}^{\mathbb{N}}$ is Sturmian if it has the factor complexity $\mathcal{C}_{\mathbf{u}}(n) = n + 1$ for all $n \in \mathbb{N}$. If both sequences $0\mathbf{u}$ and $1\mathbf{u}$ are Sturmian, then $\mathbf{u}$ is called a *standard Sturmian sequence*. It is well-known that for any Sturmian sequence there exists a unique standard Sturmian sequence with the same language. Since all properties which we are interested in (indices of factors, critical exponent, special factors, return words, recurrence function) depend only on the language of the sequence, we restrict ourselves to standard Sturmian sequences without loss of generality.

In the sequel, we use the characterization of standard Sturmian sequences by their directive sequences. To introduce them, we define two morphisms

$$G = \begin{cases} 0 \to 10 \\ 1 \to 1 \end{cases} \qquad \text{and} \qquad D = \begin{cases} 0 \to 0 \\ 1 \to 01 \end{cases}$$

with the corresponding matrices

$$M_G = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \qquad \text{and} \qquad M_D = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Let us note that $G = E \circ F$ and $D = F \circ E$, where E is the morphism which exchanges letters, i.e., $E : 0 \to 1, 1 \to 0$, and F is the Fibonacci morphism, i.e., $F : 0 \to 01, 1 \to 0$.

Proposition 17 ([13]). *For every standard Sturmian sequence* $\mathbf{u}$ *there is a uniquely given sequence* $\mathbf{\Delta} = \Delta_0\Delta_1\Delta_2\cdots \in \{G,D\}^{\mathbb{N}}$ *of morphisms and a sequence* $(\mathbf{u}^{(n)})_{n\geq 0}$ *of standard Sturmian sequences such that*

$$\mathbf{u} = \Delta_0\Delta_1\ldots\Delta_{n-1}(\mathbf{u}^{(n)}) \text{ for every } n \in \mathbb{N}\,.$$

Moreover, the sequence $\mathbf{\Delta}$ *contains infinitely many letters* G *and infinitely many letters* D*, i.e.,*

$\mathbf{\Delta} = G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ *or* $\mathbf{\Delta} = D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$ *for some sequence* $(a_i)_{i\geq 1}$ *of positive integers.*

The sequence $\mathbf{\Delta}$ *is called the* directive sequence *of* $\mathbf{u}$.

Remark 18. Let us note that $\mathbf{u}$ has the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ if and only if $E(\mathbf{u})$ has the directive sequence $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$. Obviously, both sequences $\mathbf{u}$ and $E(\mathbf{u})$ have the same structure up to the exchange of letters $0 \leftrightarrow 1$. In particular, any Sturmian sequence with the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ can be written as a concatenation of the blocks $1^{a_1}0$ and $1^{a_1+1}0$, while any Sturmian sequence with the directive sequence $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$ can be written as a concatenation of the blocks $0^{a_1}1$ and $0^{a_1+1}1$.

By Vuillon's result [24], every factor of any Sturmian sequence has exactly two return words. Thus for a given bispecial factor b of $\mathbf{u}$, we usually denote by r the more frequent and by s the less frequent return word to b in $\mathbf{u}$. In this notation, the sequence $\mathbf{u}$ can be decomposed into the blocks $r^k s$ and $r^{k+1}s$ for some $k \in \mathbb{N}, k \geq 1$.

We need to know how bispecial factors and their return words change under the application of morphisms G and D. The following description can be found in [16], where several partial statements from [14] are accumulated.

Lemma 19. *Let* $\mathbf{u}', \mathbf{u}$ *be standard Sturmian sequences such that* $\mathbf{u} = G(\mathbf{u}')$.

(i) For every bispecial factor b' *of* $\mathbf{u}'$*, the factor* $b = G(b')1$ *is a bispecial factor of* $\mathbf{u}$.

(ii) Every bispecial factor b *of* $\mathbf{u}$ *which is not empty can be written as* $b = G(b')1$ *for a uniquely given bispecial factor* $b' \in \mathcal{L}(\mathbf{u}')$.

(iii) The words r', s' *are return words to a bispecial prefix* b' *of* $\mathbf{u}'$ *if and only if* $r = G(r'), s = G(s')$ *are return words to a bispecial prefix* $b = G(b')1$ *of* $\mathbf{u}$*. Moreover, the derived sequences satisfy* $\mathbf{d_u}(b) = \mathbf{d_{u'}}(b')$.

Lemma 20. *Let* $\mathbf{u}', \mathbf{u}$ *be standard Sturmian sequences such that* $\mathbf{u} = D(\mathbf{u}')$.

(i) For every bispecial factor b' *of* $\mathbf{u}'$*, the factor* $b = D(b')0$ *is a bispecial factor of* $\mathbf{u}$.

(ii) Every bispecial factor b *of* $\mathbf{u}$ *which is not empty can be written as* $b = D(b')0$ *for a uniquely given bispecial factor* $b' \in \mathcal{L}(\mathbf{u}')$.

(iii) The words r', s' *are return words to a bispecial prefix* b' *of* $\mathbf{u}'$ *if and only if* $r = D(r'), s = D(s')$ *are return words to a bispecial prefix* $b = D(b')0$ *of* $\mathbf{u}$*. Moreover, the derived sequences satisfy* $\mathbf{d_u}(b) = \mathbf{d_{u'}}(b')$.

Any prefix of a standard Sturmian sequence is a left special factor. Moreover, a factor of a standard Sturmian sequence $\mathbf{u}$ is bispecial if and only if it is a palindromic prefix of $\mathbf{u}$. Therefore, we can order the bispecial factors of a given standard Sturmian sequence $\mathbf{u}$ by their lengths: we start with the empty word ε, which is the 0^{th} bispecial factor, then the first letter of $\mathbf{u}$ is the 1^{st} bispecial factor of $\mathbf{u}$ etc.

Remark 21. If $\mathbf{u}$ has the directive sequence $\Delta_0\Delta_1\Delta_2\cdots \in \{G, D\}^{\mathbb{N}}$, the derived sequence $\mathbf{d}_{\mathbf{u}}(b)$ to the n^{th} bispecial factor b of $\mathbf{u}$ has the directive sequence $\Delta_n\Delta_{n+1}\Delta_{n+2}\cdots$. Indeed, we denote $\mathbf{u}'$ the sequence with the directive sequence $\Delta_n\Delta_{n+1}\Delta_{n+2}\cdots$. It has the bispecial factor ε and by the definition $\mathbf{d}_{\mathbf{u}'}(\varepsilon) = \mathbf{u}'$. If we apply n times Lemmas 19 or 20, we get $\mathbf{d}_{\mathbf{u}}(b) = \mathbf{d}_{\mathbf{u}'}(\varepsilon) = \mathbf{u}'$.

Let us formulate a direct consequence of the relation (1) and Lemmas 19 and 20.

Corollary 22. *Let $k, h \in \mathbb{N}$. Let b' be the k^{th} bispecial factor of a standard Sturmian sequence $\mathbf{u}'$ and u' be a return word to b' in $\mathbf{u}'$. Let $\mathbf{\Delta} = \Delta_0\Delta_1\Delta_2\cdots \in \{G, D\}^{\mathbb{N}}$ be the directive sequence of $\mathbf{u}'$.*

1. *If $\mathbf{u} = G^h(\mathbf{u}')$, then the $(k+h)^{th}$ bispecial factor b of $\mathbf{u}$ and a return word u to b satisfy*

$$\vec{V}(b) = \begin{pmatrix} 1 & 0 \\ h & 1 \end{pmatrix} \vec{V}(b') + h\begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \textit{and} \quad \vec{V}(u) = \begin{pmatrix} 1 & 0 \\ h & 1 \end{pmatrix} \vec{V}(u')\,.$$

   The directive sequence of $\mathbf{u}$ is $G^h\Delta_0\Delta_1\Delta_2\cdots$.

2. *If $\mathbf{u} = D^h(\mathbf{u}')$, then the $(k+h)^{th}$ bispecial factor b of $\mathbf{u}$ and a return word u to b satisfy*

$$\vec{V}(b) = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \vec{V}(b') + h\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \textit{and} \quad \vec{V}(u) = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \vec{V}(u')\,.$$

   The directive sequence of $\mathbf{u}$ is $D^h\Delta_0\Delta_1\Delta_2\cdots$.

As we have seen in Remark 4, when determining the critical exponent it suffices to take into account only bispecial factors whose fractional roots are equal to its return words. Lemma 24 says that all bispecial factors of a Sturmian sequence are of this type, and moreover, it enables to determine the indices of their return words. The first auxiliary statement is a slightly strengthened variant of Observation 15 for the morphisms G and D.

Observation 23. *Let $\mathbf{u}$ be a binary sequence and let $u \in \mathcal{L}(\mathbf{u})$. If z is the longest factor in $\mathcal{L}(\mathbf{u})$ with the fractional root u, then $G(z)1$ is the longest factor in $\mathcal{L}(G(\mathbf{u}))$ with the fractional root $G(u)$ and, analogously, $D(z)0$ is the longest factor in $\mathcal{L}(D(\mathbf{u}))$ with the fractional root $D(u)$.*

Lemma 24. *Let b be a bispecial factor of a standard Sturmian sequence $\mathbf{u}$. Let r and s be the return words to b in $\mathbf{u}$ and let $k \in \mathbb{N}$, $k \geq 1$, be such that $\mathbf{u}$ is concatenated from the blocks $r^k s$ and $r^{k+1}s$. Then $r^{k+1}b$ is the longest factor of $\mathbf{u}$ with the fractional root r and sb is the longest factor of $\mathbf{u}$ with the fractional root s.*

Proof: We proceed by induction on the length of b. Without loss of generality, we assume that $\mathbf{u}$ has the directive sequence $\mathbf{\Delta} = G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$.

The bispecial factor $b = \varepsilon$ has the return words $r = 1, s = 0$, and by Remark 18, $\mathbf{u}$ is concatenated from the blocks $1^{a_1}0 = r^{a_1}s$ and $1^{a_1+1}0 = r^{a_1+1}s$. Clearly, $r^{a_1+1}b = 1^{a_1+1}$ is the longest factor of $\mathbf{u}$ with the fractional root 1. Similarly, $sb = 0$ is the longest factor of $\mathbf{u}$ with the fractional root 0.

Let b be a bispecial factor of $\mathbf{u}$ with $|b| \geq 1$ and let $\mathbf{u}$ be concatenated from the blocks $r^k s$ and $r^{k+1}s$ for the return words r, s to b in $\mathbf{u}$ and some $k \in \mathbb{N}$, $k \geq 1$. By Proposition 17, there is a unique standard Sturmian sequence $\mathbf{u}'$ such that $\mathbf{u} = G(\mathbf{u}')$. By Lemmas 19 and 20, there is a unique bispecial factor b' of $\mathbf{u}'$ with the return words r' and s' such that $b = G(b')1$, $r = G(r')$, and $s = G(s')$. Moreover, $\mathbf{u}'$ is concatenated from the blocks $(r')^k s'$ and $(r')^{k+1}s'$. Clearly, $|b'| < |b|$ and so by the induction hypothesis, the words $(r')^{k+1}b'$ and $s'b'$ are the longest factors of $\mathbf{u}'$ with the fractional root r' and s', respectively. But then by Observation 23, the words $r^{k+1}b = G((r')^{k+1}b')1$ and $sb = G(s'b')1$ are the longest factors of $\mathbf{u}$ with the fractional root $r = G(r')$ and $s = G(s')$, respectively. $\square$

Having in mind our goal to describe the critical exponent of any CS Rote sequence and Theorem 14, we need to determine the indices of return words to bispecial factors in standard Sturmian sequences, i.e., the lengths of factors from Lemma 24. We also want to distinguish, which of these return words are (un)stable. Both of these tasks can be solved using the Parikh vectors of the relevant bispecial factors and their return words. We deduce the explicit formulae for the needed Parikh vectors in Proposition 30. For this purpose, we adopt the following notation.

Notation 25. To a standard Sturmian sequence $\mathbf{u}$ with the directive sequence $\mathbf{\Delta} = G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ or $\mathbf{\Delta} = D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$ we assign an irrational number $\theta \in (0,1)$ with the continued fraction expansion

$$\theta = [0, a_1, a_2, a_3, \ldots]\,.$$

For every $N \in \mathbb{N}$, we denote by $\frac{p_N}{q_N}$ the N^{th} *convergent* to the number θ and by $\frac{p'_N}{q'_N}$ the N^{th} convergent to the number $\frac{\theta}{1+\theta}$.

Remark 26. Let us recall some basic properties of convergents. They can be found in any number theory textbook, e.g., [12].

1. The sequences (p_N), (q_N), and (q'_N) fulfil the same recurrence relation for all $N \in \mathbb{N}, N \geq 1$, namely
$$X_N = a_N X_{N-1} + X_{N-2}\,,$$
but they differ in their initial values: $p_{-1} = 1, p_0 = 0$; $q_{-1} = 0, q_0 = 1$; $q'_{-1} = q'_0 = 1$. It implies for all $N \in \mathbb{N}$
$$p_N + q_N = q'_N\,.$$

2. For all $N \in \mathbb{N}, N \geq 1$, we have
$$\begin{pmatrix} 1 & 0 \\ a_1 & 1 \end{pmatrix}\begin{pmatrix} 1 & a_2 \\ 0 & 1 \end{pmatrix}\cdots\begin{pmatrix} 1 & 0 \\ a_{2N-1} & 1 \end{pmatrix}\begin{pmatrix} 1 & a_{2N} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} p_{2N-1} & p_{2N} \\ q_{2N-1} & q_{2N} \end{pmatrix};$$
$$\begin{pmatrix} 1 & 0 \\ a_1 & 1 \end{pmatrix}\begin{pmatrix} 1 & a_2 \\ 0 & 1 \end{pmatrix}\cdots\begin{pmatrix} 1 & a_{2N-2} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ a_{2N-1} & 1 \end{pmatrix} = \begin{pmatrix} p_{2N-1} & p_{2N-2} \\ q_{2N-1} & q_{2N-2} \end{pmatrix}.$$

Remark 27. For the description of a standard Sturmian sequence $\mathbf{u}$ we use the number θ. Usually, a standard Sturmian sequence is characterized by the so-called *slope*, which is equal to the density of the letter 1 in the sequence $\mathbf{u}$. In our notation, the slope of $\mathbf{u}$ is $\frac{\theta}{1+\theta} = [0, 1 + a_1, a_2, a_3, \ldots]$ if the directive sequence $\mathbf{\Delta}$ starts with D, otherwise the slope is $\frac{1}{1+\theta} = [0, 1, a_1, a_2, a_3, \ldots]$.

In the sequel, we will need two auxiliary statements on convergents $\frac{p_N}{q_N}$ to θ.

Lemma 28. *For all $N \in \mathbb{N}$ we have*

$$\begin{pmatrix} p_N \\ q_N \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \mod 2 \qquad \textit{and} \qquad \begin{pmatrix} p_N \\ q_N \end{pmatrix} \neq \begin{pmatrix} p_{N-1} \\ q_{N-1} \end{pmatrix} \mod 2\,.$$

Proof: The first statement is a consequence of the fact that p_N and q_N are coprime. We show the second statement by contradiction. Assume that there exists $K \in \mathbb{N}$ such that $\begin{pmatrix} p_K \\ q_K \end{pmatrix} = \begin{pmatrix} p_{K-1} \\ q_{K-1} \end{pmatrix} \mod 2$. Let K denote the smallest integer with this property. As $q_{-1} = 0$ and $q_0 = 1$, necessarily, $K > 0$. Using the recurrence relation satisfied by the convergents, we can write

$$\begin{pmatrix} p_K \\ q_K \end{pmatrix} = a_K \begin{pmatrix} p_{K-1} \\ q_{K-1} \end{pmatrix} + \begin{pmatrix} p_{K-2} \\ q_{K-2} \end{pmatrix} = \begin{pmatrix} p_{K-1} \\ q_{K-1} \end{pmatrix} \mod 2\,.$$

If a_K is even, then the previous equation gives $\begin{pmatrix} p_{K-2} \\ q_{K-2} \end{pmatrix} = \begin{pmatrix} p_{K-1} \\ q_{K-1} \end{pmatrix} \mod 2$, which is a contradiction with the minimality of K.

If a_K is odd, then the previous equation gives $\begin{pmatrix} p_{K-2} \\ q_{K-2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \mod 2$, which is a contradiction with the first statement. $\square$

Lemma 29. *For all $N \in \mathbb{N}$, $N \geq 1$, we have*

$$a_N \begin{pmatrix} p_{N-1} \\ q_{N-1} \end{pmatrix} + a_{N-1} \begin{pmatrix} p_{N-2} \\ q_{N-2} \end{pmatrix} + \cdots + a_2 \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} + a_1 \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} = \begin{pmatrix} p_N \\ q_N \end{pmatrix} + \begin{pmatrix} p_{N-1} \\ q_{N-1} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Proof: It can be easily proved by induction on N. $\square$

The Parikh vectors of the bispecial factors of $\mathbf{u}$ and the corresponding return words can be easily expressed using the convergents $\frac{p_N}{q_N}$ to θ. We will use these expressions essentially in the next sections.

Proposition 30. *Let b be the n^{th} bispecial factor of $\mathbf{u}$ with the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$. We denote r and s the more and the less frequent return word to b in $\mathbf{u}$, respectively. Put $a_0 = 0$ and write n in the form $n = m + a_0 + a_1 + a_2 + \cdots + a_N$ for a unique $N \in \mathbb{N}$ and $0 \leq m < a_{N+1}$. Then*

1. $\vec{V}(r) = \begin{pmatrix} p_N \\ q_N \end{pmatrix}$;

2. $\vec{V}(s) = \begin{pmatrix} m\,p_N + p_{N-1} \\ m\,q_N + q_{N-1} \end{pmatrix}$;

3. $\vec{V}(b) = (m+1) \begin{pmatrix} p_N \\ q_N \end{pmatrix} + \begin{pmatrix} p_{N-1} \\ q_{N-1} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Proof: First we suppose that N is even and denote $\mathbf{u}'$ the standard Sturmian sequence with the directive sequence $G^{a_{N+1}}D^{a_{N+2}}G^{a_{N+3}}\cdots$. By Remark 18, $\mathbf{u}'$ is concatenated from the blocks $1^{a_{N+1}}0$ and $1^{a_{N+1}+1}0$. Thus its m^{th} bispecial factor is $b' = 1^m$ and the return words to b' in $\mathbf{u}'$ are $r' = 1$ and $s' = 1^m 0$.

By Lemmas 19 and 20 and Remark 21,

$$r = G^{a_1}D^{a_2}\cdots G^{a_{N-1}}D^{a_N}(r') \qquad \text{and} \qquad s = G^{a_1}D^{a_2}\cdots G^{a_{N-1}}D^{a_N}(s')\,.$$

By Corollary 22 and Lemma 29, the Parikh vectors of r and s satisfy

$$\vec{V}(r) = \begin{pmatrix}1 & 0\\ a_1 & 1\end{pmatrix}\begin{pmatrix}1 & a_2\\ 0 & 1\end{pmatrix}\cdots\begin{pmatrix}1 & 0\\ a_{N-1} & 1\end{pmatrix}\begin{pmatrix}1 & a_N\\ 0 & 1\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix} = \begin{pmatrix}p_N\\ q_N\end{pmatrix};$$
$$\vec{V}(s) = \begin{pmatrix}1 & 0\\ a_1 & 1\end{pmatrix}\begin{pmatrix}1 & a_2\\ 0 & 1\end{pmatrix}\cdots\begin{pmatrix}1 & 0\\ a_{N-1} & 1\end{pmatrix}\begin{pmatrix}1 & a_N\\ 0 & 1\end{pmatrix}\begin{pmatrix}1\\ m\end{pmatrix} = \begin{pmatrix}m\,p_N + p_{N-1}\\ m\,q_N + q_{N-1}\end{pmatrix}.$$

To find the Parikh vector of b we start with the bispecial factor $b' = 1^m$ and N times apply Corollary 22. Eventually, we rewrite the arising products of matrices by Lemma 29 and we get

$$\vec{V}(b) = m\begin{pmatrix}p_N\\ q_N\end{pmatrix} + a_N\begin{pmatrix}p_{N-1}\\ q_{N-1}\end{pmatrix} + a_{N-1}\begin{pmatrix}p_{N-2}\\ q_{N-2}\end{pmatrix} + \cdots + a_2\begin{pmatrix}p_1\\ q_1\end{pmatrix} + a_1\begin{pmatrix}p_0\\ q_0\end{pmatrix}.$$

This together with Lemma 29 implies the statement of Item (3) for N even. The proof for N odd is analogous. $\square$

Remark 31. If we assume in Proposition 30 that $\mathbf{u}$ has the directive sequence $\mathbf{\Delta} = D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$, then by Remark 18, the coordinates of the Parikh vectors will be exchanged, i.e.,

$$\vec{V}(r) = \begin{pmatrix}q_N\\ p_N\end{pmatrix},\ \vec{V}(s) = \begin{pmatrix}m\,q_N + q_{N-1}\\ m\,p_N + p_{N-1}\end{pmatrix},\ \text{and}\ \vec{V}(b) = (m+1)\begin{pmatrix}q_N\\ p_N\end{pmatrix} + \begin{pmatrix}q_{N-1}\\ p_{N-1}\end{pmatrix} - \begin{pmatrix}1\\ 1\end{pmatrix}.$$

7 The critical exponent of CS Rote sequences

We are going to give an explicit formula for the critical exponent of a CS Rote sequence $\mathbf{v}$. We will use Theorem 14 which requires the knowledge of the indices of return words to bispecial factors in the Sturmian sequence $\mathcal{S}(\mathbf{v})$. It is well-known that there is a unique standard Sturmian sequence $\mathbf{u}$ such that both $\mathcal{S}(\mathbf{v})$ and $\mathbf{u}$ have the same language. Since the critical exponent depends only on the language, we can work with the standard Sturmian sequence $\mathbf{u}$ instead of $\mathcal{S}(\mathbf{v})$.

In the following proposition and theorem, we use Notation 25.

Proposition 32. *Let $\mathbf{u}$ be a standard Sturmian sequence with the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ or $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$ and let $n \in \mathbb{N}$ be given. We put $a_0 = 0$ and we denote r and s the more and the less frequent return words to the n^{th} bispecial factor of $\mathbf{u}$, respectively.*

1. *If $n = m + a_0 + a_1 + a_2 + \cdots + a_N$, where $0 \le m < a_{N+1}$, then* $\mathrm{ind}(r) = a_{N+1} + 2 + \frac{q'_{N-1}-2}{q'_N}$.

2. *If $n = a_0 + a_1 + a_2 + \cdots + a_N$, then* $\mathrm{ind}(s) = a_N + 2 + \frac{q'_{N-2}-2}{q'_{N-1}}$.

3. If $n = m + a_0 + a_1 + a_2 + \cdots + a_N$*, where* $0 < m < a_{N+1}$*, then* $\mathrm{ind}(s) = 2 + \frac{q'_N - 2}{m\, q'_N + q'_{N-1}}$.

Let us comment what is meant by q'_{-2} in the case $N = 0$ in Item (2): we define q'_{-2} to satisfy the recurrent relation $1 = q'_0 = a_0 q'_{-1} + q'_{-2} = q'_{-2}$.

Proof: We assume that N is even (the case of N odd is analogous). Moreover, we assume that $\mathbf{u}$ has the directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \cdots$.

Since b is the n^{th} bispecial factor in $\mathbf{u}$, the derived sequence $\mathbf{d_u}(b)$ is standard Sturmian with the directive sequence $G^k D^{a_{N+2}} G^{a_{N+3}} \cdots$, where $k = a_{N+1} - m$, see Corollary 22. By Remark 18, $\mathbf{d_u}(b)$ is a concatenation of the blocks $1^k 0$ and $1^{k+1}0$. Therefore, the sequence $\mathbf{u}$ is concatenated from the blocks $r^k s$ and $r^{k+1}s$, where r and s are the return words to b. By Lemma 24, the factor $r^{k+1}b$ is the longest factor of $\mathbf{u}$ with the fractional root r and sb is the longest factor of $\mathbf{u}$ with the fractional root s. In other words, $r^{\mathrm{ind}(r)} = r^{k+1}b$ and $s^{\mathrm{ind}(s)} = sb$. By Proposition 30 and Remark 26, we have

$$|r| = p_N + q_N = q'_N, \qquad |s| = m q'_N + q'_{N-1}, \quad \text{and} \quad |b| = (m+1)q'_N + q'_{N-1} - 2\,.$$

As $|r^{k+1}b| = (k+1)|r| + |b| = (a_{N+1} - m + 1)|r| + |b| = (a_{N+1} + 2)q'_N + q'_{N-1} - 2$, we get

$$\mathrm{ind}(r) = \frac{|r^{k+1}b|}{|r|} = a_{N+1} + 2 + \frac{q'_{N-1} - 2}{q'_N}\,.$$

As $|sb| = (2m+1)q'_N + 2q'_{N-1} - 2$, we get for $m = 0$

$$\mathrm{ind}(s) = \frac{|sb|}{|s|} = \frac{q'_N + 2q'_{N-1} - 2}{q'_{N-1}} = \frac{a_N q'_{N-1} + q'_{N-2} + 2q'_{N-1} - 2}{q'_{N-1}} = a_N + 2 + \frac{q'_{N-2} - 2}{q'_{N-1}}$$

and for $m > 0$

$$\mathrm{ind}(s) = \frac{|sb|}{|s|} = \frac{(2m+1)q'_N + 2q'_{N-1} - 2}{m\, q'_N + q'_{N-1}} = 2 + \frac{q'_N - 2}{m\, q'_N + q'_{N-1}}\,.$$

If the directive sequence equals $D^{a_1} G^{a_2} D^{a_3} G^{a_4} \cdots$, only the coordinates of the Parikh vectors of r, s, b are exchanged (see Remark 31). $\square$

Theorem 33. *Let* $\mathbf{v}$ *be a CS Rote sequence and let* $\mathbf{u}$ *be the standard Sturmian sequence such that* $\mathcal{L}(\mathcal{S}(\mathbf{v})) = \mathcal{L}(\mathbf{u})$*. Then* $\mathrm{cr}(\mathbf{v}) = \sup(M_1 \cup M_2 \cup M_3)$*, where*

$$M_1 = \left\{ a_{N+1} + 2 + \frac{q'_{N-1} - 1}{q'_N} \,:\, q_N \text{ is even, } N \in \mathbb{N} \right\};$$

$$M_2 = \left\{ \frac{a_{N+1} + 2}{2} + \frac{q'_{N-1} - 1}{2q'_N} \,:\, q_N \text{ is odd, } N \in \mathbb{N} \right\};$$

$$M_3 = \left\{ 2 + \frac{q'_N - 1}{q'_{N-1} + q'_N} \,:\, q_{N-1}, q_N \text{ are odd and } a_{N+1} > 1, N \geq 1 \right\}$$

if the directive sequence of $\mathbf{u}$ *is* $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \cdots$*, and*

$$M_1 = \left\{ a_{N+1} + 2 + \frac{q'_{N-1} - 1}{q'_N} \,:\, p_N \text{ is even, } N \in \mathbb{N} \right\};$$

$$M_2 = \left\{ \frac{a_{N+1} + 2}{2} + \frac{q'_{N-1} - 1}{2q'_N} \,:\, p_N \text{ is odd, } N \in \mathbb{N} \right\};$$

$$M_3 = \left\{ 2 + \frac{q'_N - 1}{q'_{N-1} + q'_N} \,:\, p_{N-1}, p_N \text{ are odd and } a_{N+1} > 1, N \geq 1 \right\}$$

if the directive sequence of $\mathbf{u}$ *is* $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$.

Proof: Let us recall that every CS Rote sequence is uniformly recurrent and aperiodic. In addition, to a CS Rote sequence $\mathbf{v}$ we can always find a unique standard Sturmian sequence $\mathbf{u}$ such that $\mathbf{u}$ has the same language as $\mathcal{S}(\mathbf{v})$. It is important to realize that Theorem 14 holds for the pair $\mathbf{v}$ and $\mathbf{u}$, too.

First, we assume that $\mathbf{u}$ has the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$. We compute the suprema of the sets A_1 and A_2 defined in Theorem 14, since by this theorem, $\mathrm{cr}(\mathbf{v}) = \sup(A_1 \cup A_2)$.

Let us decompose A_1 into $A_1 = \bigcup_{N=0}^{\infty} A_1^{(N)}$, where

$$A_1^{(N)} = \Big\{\mathrm{ind}_{\mathbf{u}}(u) + \tfrac{1}{|u|} : \ u \text{ is a stable return word to the } n^{th} \text{ bispecial f. of } \mathbf{u}, \ \sum_{k=0}^{N} a_k \le n < \sum_{k=0}^{N+1} a_k\Big\}.$$

By definition, a word u is stable if the number of ones occurring in u is even, i.e., the second component of its Parikh vector $\vec{V}(u)$ is even. Combining Propositions 30 and 32, we obtain that $A_1^{(N)}$ contains

- $a_{N+1} + 2 + \frac{q'_{N-1}-1}{q'_N}$ if q_N is even,
- $a_N + 2 + \frac{q'_{N-2}-1}{q'_{N-1}}$ if q_{N-1} is even,
- the subset $B_1^{(N)} = \{2 + \frac{q'_N - 1}{m\, q'_N + q'_{N-1}} : m\, q_N + q_{N-1} \text{ even}, \ 0 < m < a_{N+1}\}$.

First we look at $A_1^{(0)}$. Since $q_0 = 1$ is odd, $q_{-1} = 0$ is even, $a_0 = 0$, $q'_0 = q'_{-1} = q'_{-2} = 1$, we get $A_1^{(0)} = \{2\}$. Since we know that $\mathrm{cr}(\mathbf{v}) > 2$, we can consider only $N \ge 1$. Let us note that all elements in $B_1^{(N)}$ are strictly less than 3. If q_N or q_{N-1} is even, the set $A_1^{(N)}$ contains an element ≥ 3 and the set $B_1^{(N)}$ does not play any role for $\sup A_1$. If both q_N and q_{N-1} are odd, there is an element in $B_1^{(N)}$ only for odd $m < a_{N+1}$, and obviously, $\sup B_1^{(N)}$ is attained for $m = 1$ (if $a_{N+1} = 1$ the set is empty). Together it gives $\sup A_1 = \sup(M_1 \cup M_3)$.

Analogously we define the sets $A_2^{(N)}$ for unstable return words. Then $A_2^{(N)}$ consists of

- $\frac{1}{2}\big(a_{N+1} + 2 + \frac{q'_{N-1}-1}{q'_N}\big)$ if q_N is odd,
- $\frac{1}{2}\big(a_N + 2 + \frac{q'_{N-2}-1}{q'_{N-1}}\big)$ if q_{N-1} is odd,
- the subset $B_2^{(N)} = \{\frac{1}{2}\big(2 + \frac{q'_N-1}{m\, q'_N + q'_{N-1}}\big) : m\, q_N + q_{N-1} \text{ odd}, \ 0 < m < a_{N+1}\}$.

We easily compute that $\sup A_2^{(0)} = \frac{1}{2}(a_1 + 2)$. All elements in $B_2^{(N)}$ are strictly less than $\frac{3}{2}$. Thus if q_N or q_{N-1} is odd, the set $B_2^{(N)}$ does not play any role for $\sup A_2$. If q_N and q_{N-1} are even, then the set $B_2^{(N)}$ is empty. It means that $\sup A_2 = \sup M_2$. We can conclude that $\mathrm{cr}(\mathbf{v}) = \sup(A_1 \cup A_2) = \sup(M_1 \cup M_2 \cup M_3)$.

If the directive sequence equals $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$, only the coordinates of the Parikh vectors of r and s are exchanged (see Remark 31). $\square$

8 CS Rote sequences with a small critical exponent

In this section, we present some corollaries of Theorem 33. As we have mentioned, Currie, Mol, and Rampersad proved in [7] that there are exactly two languages of rich binary sequences with the critical exponent less than $\frac{14}{5}$. Both of them are languages of CS Rote sequences.

Let us remind that the critical exponent depends only on the language of a sequence and not on the sequence itself. Therefore, there are infinitely many CS Rote sequences with the critical exponent less than $\frac{14}{5}$, but all of them have one of two languages. We show that among all languages of CS Rote sequences only these two languages have the critical exponent less than 3. We also describe all languages of CS Rote sequences with the critical exponent equal to 3.

Proposition 34. *Let* $\mathbf{v}$ *be a CS Rote sequence associated with the standard Sturmian sequence* $\mathbf{u} = \mathcal{S}(\mathbf{v})$. *If* $\mathrm{cr}(\mathbf{v}) \leq 3$, *then the directive sequence of* $\mathbf{u}$ *is of one of the following forms:*

1. $G^{a_1}(D^2G^2)^\omega$, *where* $a_1 = 1$ *or* $a_1 = 3$; *in this case* $\mathrm{cr}(\mathbf{v}) = 2 + \frac{1}{\sqrt{2}}$;
2. $G^{a_1}D^4(G^2D^2)^\omega$, *where* $a_1 = 1$ *or* $a_1 = 3$; *in this case* $\mathrm{cr}(\mathbf{v}) = 3$;
3. $G^{a_1}D^1G^{a_3}(D^2G^2)^\omega$, *where* $a_1 = 2$ *or* $a_1 = 4$ *and* $a_3 = 1$ *or* $a_3 = 3$; *in this case* $\mathrm{cr}(\mathbf{v}) = 3$;
4. $D^1G^{a_2}(D^2G^2)^\omega$, *where* $a_2 = 1$ *or* $a_2 = 3$; *in this case* $\mathrm{cr}(\mathbf{v}) = 3$.

Proof: For each $N \in \mathbb{N}$ we denote $\beta_N = a_{N+1} + 2 + \frac{q'_{N-1}-1}{q'_N}$ the number which is a candidate to join the set M_1. We can easily compute that $\beta_0 = a_1 + 2$, $\beta_1 = a_2 + 2$ and $\beta_N > 3$ for every $N \geq 2$. Indeed, it suffices to realize that $q'_{-1} = q'_0 = 1$, $q'_1 = a_1 + 1 > 1$ and $(q'_N)_{N \geq 1}$ is an increasing sequence of integers, so $\frac{q'_{N-1}-1}{q'_N} \in (0,1)$ for all $N \geq 2$. Since we look for a sequence $\mathbf{v}$ with $\mathrm{cr}(\mathbf{v}) \leq 3$, we have to ensure that $\beta_N \notin M_1$ for all $N \geq 2$ by the parity conditions. It is also important to notice that $\sup M_3 \leq 3$. Indeed, since $\frac{q'_N-1}{q'_N+q'_{N-1}} \in [0,1)$ for all $N \in \mathbb{N}$, all elements of M_3 are less than 3.

First we assume that the directive sequence of $\mathbf{u}$ is $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$. By Theorem 33, if q_N is even, then $\beta_N \in M_1$, otherwise $\frac{1}{2}\beta_N \in M_2$. To ensure $\sup M_1 \leq 3$, q_N has to be odd for all $N \geq 2$. Moreover, to ensure $\sup M_2 \leq 3$, $\beta_N \leq 6$ and so $a_{N+1} \leq 3$ for all $N \geq 2$. Since $q_0 = 1$ is odd, $\frac{1}{2}\beta_0 = \frac{a_1}{2} + 1 \in M_2$ and so $a_1 \leq 4$. We distinguish two cases.

(i) If $q_1 = a_1$ is odd, then M_1 is empty and $\frac{1}{2}\beta_1 = \frac{a_2}{2} + 1 \in M_2$. Thus $a_2 \leq 4$. The recurrent relation $q_N = a_N q_{N-1} + q_{N-2}$ with the odd initial conditions q_0 and q_1 produces q_N odd for all $N \geq 2$ if and only if a_N is even for all $N \geq 2$. We can summarize that $a_1 \in \{1,3\}$, $a_2 \in \{2,4\}$ and $a_N = 2$ for all $N \geq 3$.

Let us observe that if $a_2 = 4$, then $\sup M_2 = 3$. Since $\sup M_3 \leq 3$ and M_1 is empty, we conclude that $\mathrm{cr}(\mathbf{v}) = \sup M_2 = 3$. It gives us Item (2) of our proposition.

If $a_2 = 2$, then it is easy to check that all elements of M_2 are smaller than $\frac{5}{2}$ and thus $\sup M_2 \leq \frac{5}{2}$. Since M_1 is empty, to prove Item (1), it remains to compute

$$\sup M_3 = \sup \left\{ 2 + \frac{q'_N - 1}{q'_{N-1} + q'_N} : N \in \mathbb{N} \right\} = 2 + \frac{1}{\sqrt{2}} > \frac{5}{2}\,.$$

The sequence $(q'_N)_{N\geq 2}$ fulfils the recurrence relation $q'_N = 2q'_{N-1} + q'_{N-2}$ with the initial conditions $q'_0 = 1$ and $q'_1 = a_1 + 1$. This linear recurrence has the solution

$$q'_N = \tfrac{1}{2\sqrt{2}}\Big((a_1+\sqrt{2})(1+\sqrt{2})^N - (a_1-\sqrt{2})(1-\sqrt{2})^N\Big)\,. \tag{7}$$

Now it suffices to use the expression (7) to verify that

$$\lim_{N\to\infty} \frac{q'_N - 1}{q'_{N-1}+q'_N} = \frac{1}{\sqrt{2}} \quad \text{and} \quad \frac{q'_N - 1}{q'_{N-1}+q'_N} \leq \frac{1}{\sqrt{2}} \ \text{ for all } N \in \mathbb{N}\,.$$

(ii) If $q_1 = a_1$ is even, then $M_1 = \{\beta_1 = a_2 + 2\}$ since all the others q_N are odd. Thus $a_2 = 1$ and $\mathrm{cr}(\mathbf{v}) \geq 3$. The recurrent relation $q_N = a_N q_{N-1} + q_{N-2}$ with the initial conditions $q_0 = 1$, $q_1 = a_1 \in \{2,4\}$ produces odd q_N for all $N \geq 2$ if and only if a_3 is odd and a_N is even for all $N \geq 4$. Moreover, to ensure $\sup M_2 \leq 3$, we have to take $a_1 \leq 4$ and $a_N \leq 3$ for every $N \geq 3$. Together with the parity conditions we get $a_1 \in \{2,4\}$, $a_2 = 1$, $a_3 \in \{1,3\}$, and $a_N = 2$ for all $N \geq 4$. Since $\sup M_3 \leq 3$, the set M_3 can be omitted. Together it means that in this case $\mathrm{cr}(\mathbf{v}) = 3$ and it corresponds to Item (3) of our statement.

Now we assume that the directive sequence of $\mathbf{u}$ is $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$. By Theorem 33, if p_N is even, then $\beta_N \in M_1$, otherwise $\frac{\beta_N}{2} \in M_2$. Since $p_0 = 0$ is even, $\beta_0 = a_1 + 2 \in M_1$. Therefore, $a_1 = 1$ and $\mathrm{cr}(\mathbf{v}) \geq 3$. Since $p_1 = a_1 p_0 + p_{-1} = 1$ is odd, $\frac{\beta_1}{2} = \frac{a_2}{2} + 1 \in M_2$ and so $a_2 \leq 4$. To guarantee $\mathrm{cr}(\mathbf{v}) \leq 3$, p_N has to be odd and $a_{N+1} \leq 3$ for all $N \geq 2$. The recurrence relation $p_N = a_N p_{N-1} + p_{N-2}$ with the initial conditions $p_0 = 0$ and $p_1 = 1$ produces p_N odd for all $N \geq 2$ if and only if a_2 is odd and a_N is even for all $N \geq 2$. Clearly, the set M_3 can be omitted. We may conclude that $\mathrm{cr}(\mathbf{v}) = 3$ and $a_1 = 1$, $a_2 \in \{1,3\}$, and $a_N = 2$ for all $N \geq 3$, which corresponds to Item (4).

□

Remark 35. Let us emphasize that all standard Sturmian sequences from Proposition 34, i.e., which are associated with CS Rote sequences with the critical exponent ≤ 3, are morphic images of the fixed point of the morphism D^2G^2. If follows directly from the fact that their directive sequences have the periodic suffix $(D^2G^2)^\omega$.

Example 36. In Proposition 34, we have shown that the CS Rote sequence $\mathbf{v}$ such that $\mathcal{S}(\mathbf{v})$ has the directive sequence $G(D^2G^2)^\omega$ has the critical exponent $\mathrm{cr}(\mathbf{v}) = 2 + \frac{1}{\sqrt{2}}$. Let us determine the critical exponent $\mathrm{cr}(\mathbf{v}')$ of the CS Rote sequence $\mathbf{v}'$ associated to the standard Sturmian sequence $\mathcal{S}(\mathbf{v}')$ obtained by the exchange of letters from $\mathcal{S}(\mathbf{v})$, i.e., $\mathcal{S}(\mathbf{v}') = E(\mathcal{S}(\mathbf{v}))$.

By Remark 18, the directive sequence of $\mathcal{S}(\mathbf{v}')$ equals $D(G^2D^2)^\omega$. Thus we have $\theta = [0,1,2,2,2,\ldots]$ and it is readily seen that p_N is even if and only if N is even. Let us calculate $\mathrm{cr}(\mathbf{v}')$ by Theorem 33. We have

$$M_1 = \{a_{2N+1} + 2 + \frac{q'_{2N-1}-1}{q'_{2N}} : N \in \mathbb{N}\} = \{3\} \cup \{4 + \frac{q'_{2N-1}-1}{q'_{2N}} : N \in \mathbb{N}, N \geq 1\}\,.$$

Using equation (7), we can check that the sequence $\big(\frac{q'_{2N-1}-1}{q'_{2N}}\big)$ is increasing and has the limit $\frac{1}{1+\sqrt{2}}$, and therefore $\sup M_1 = 4 + \frac{1}{1+\sqrt{2}}$. Since the elements of M_2 and M_3 are ≤ 3, we can conclude that $\mathrm{cr}(\mathbf{v}') = 4 + \frac{1}{1+\sqrt{2}}$.

It is well-known that among Sturmian sequences the sequence with the lowest possible critical exponent is the Fibonacci sequence $\mathbf{f}$, which has $\mathrm{cr}(\mathbf{f}) = 3 + \frac{2}{1+\sqrt{5}} \sim 3.602$. The following theorem implies that there are uncountably many CS Rote sequences with the critical exponent smaller than $\mathrm{cr}(\mathbf{f})$.

Theorem 37. *Let $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ be the directive sequence of a standard Sturmian sequence $\mathbf{u}$ and let $\mathbf{v}$ be the CS Rote sequence associated with $\mathbf{u}$. Then $\mathrm{cr}(\mathbf{v}) < \frac{7}{2}$ if and only if the sequence $a_1a_2a_3\cdots$ is a concatenation of the blocks from the following list:*
L_0: 111;
L_1: $s1$, where $s \in \{2,4\}$;
L_2: $cs31$, where $c \in \{1,3\}$ and $s \in \{2,4\}^*$;
L_3: $c\mathbf{s}$, where $c \in \{1,3\}$ and $\mathbf{s} \in \{2,4\}^{\mathbb{N}}$,
and if the block L_0 appears in $a_1a_2a_3\cdots$, then it is a prefix of $a_1a_2a_3\cdots$.

Proof: As in the proof of Proposition 34, we again use the sets M_1 and M_2 from Theorem 33. The set M_3 can be omitted since $\sup M_3 \leq 3$.

Let us recall that $q_0 = 1$, $q_1 = a_1$, and if q_N is even, then $\beta_N = a_{N+1} + 2 + \frac{q'_{N-1}-1}{q'_N} \in M_1$, otherwise $\frac{1}{2}\beta_N \in M_2$. First we suppose that $\mathrm{cr}(\mathbf{v}) < \frac{7}{2}$ and we deduce several auxiliary observations for each $N \in \mathbb{N}$:

1. If q_N even, then $a_{N+1} = 1$.
   Proof: It follows from the inequality $\beta_N < \frac{7}{2}$.
2. If q_N odd, then $a_{N+1} \in \{1,2,3,4\}$.
   Proof: It follows from the inequality $\frac{1}{2}\beta_N < \frac{7}{2}$.
3. If q_{N-1} odd and q_N even, then q_{N+1} odd.
   Proof: It follows from Item (1) and the relation $q_{N+1} = a_{N+1}q_N + q_{N-1} = q_N + q_{N-1}$.
4. If q_N even, q_{N+1} odd, and q_{N+2} even, then $a_{N+2} \in \{2,4\}$.
   Proof: It follows from Item (2) and the relation $q_{N+2} = a_{N+2}q_{N+1} + q_N$.
5. If q_N even, q_{N+1} odd, and q_{N+2} odd, then $a_{N+2} \in \{1,3\}$.
   Proof: It follows from Item (2) and the relation $q_{N+2} = a_{N+2}q_{N+1} + q_N$.
6. If q_N odd, q_{N+1} odd, and q_{N+2} odd, then $a_{N+2} \in \{2,4\}$.
   Proof: It follows from Item (2) and the relation $q_{N+2} = a_{N+2}q_{N+1} + q_N$.
7. If q_N odd, q_{N+1} odd, and q_{N+2} even, then $a_{N+2} = \{1,3\}$. Moreover, if $a_{N+2} = 1$, then $N = 0$ and $a_1 = 1$.
   Proof: Item (2) and the relation $q_{N+2} = a_{N+2}q_{N+1} + q_N$ imply $a_{N+2} \in \{1,3\}$. Assume that $a_{N+2} = 1$. By Item (1), $a_{N+3} = 1$. As q_{N+2} is even, $\beta_{N+2} \in M_1$ and so $\beta_{N+2} = 3 + \frac{q'_{N+1}-1}{q'_{N+2}} < \frac{7}{2}$. By some simple rearrangements and applications of the recurrence relation, we can rewrite this inequality equivalently as $(a_{N+1}-1)q'_N + q'_{N-1} < 2$. It is easy to verify that this inequality holds only for $N = 0$ and $a_1 = 1$ or $N = 1$ and $a_2 = 1$. Nevertheless, the second case leads to a contradiction with the assumption that both $q_1, q_2 = a_2q_1 + 1$ are odd.

8. Let $M > N+2$. If q_N even, q_M even, and q_K odd for all $K, N < K < M$, then $a_{N+2} \in \{1,3\}$, $a_M = 3$ and $a_K \in \{2,4\}$ for all $K, N+2 < K < M$.

   Proof: Item (6) implies $a_K \in \{2,4\}$ for all K, $N+2 < K < M$. By Item (5), $a_{N+2} \in \{1,3\}$ and by Item (7), $a_M = 3$.

Using the previous claims we show that for each J for which q_J is even, at the position $J+1$ ends one of the blocks L_0, L_1, or L_2. Moreover, the block L_0 can only occur as a prefix of the sequence $a_1a_2a_3\cdots$, while each of the blocks L_1 and L_2 is either a prefix of $a_1a_2a_3\cdots$ or it starts at the position $I+2$, where I is the greatest integer smaller than J for which q_I is even. We discuss three cases:

- Let q_1 be even. Then Item (1) implies $a_2 = 1$. And since $a_1 = q_1$ is even and $q_0 = 1$ is odd, by Item (2), we get $a_1 \in \{2,4\}$. Thus the prefix a_1a_2 of the sequence $a_1a_2a_3\cdots$ is of the form L_1 from our list.
- Let $J > 1$ be the first index such that q_J is even. As $q_1 = a_1$, a_1 is odd, and by Item (2), we get $a_1 \in \{1,3\}$. Item (6) implies $a_K \in \{2,4\}$ for all $K, 1 < K < J$. By Item (7), $a_J = 3$ or $a_J = 1$. But if $a_J = 1$, then $J = 2$ and $a_1 = 1$. Finally, Item (1) implies $a_{J+1} = 1$. Thus the prefix $a_1a_2\cdots a_Ja_{J+1}$ of the sequence $a_1a_2a_3\cdots$ is of the form L_0 or L_2 from our list.
- Let I be an index such that q_I is even and let J be the smallest index greater than I for which q_J is even. By Item (1), $a_{I+1} = 1$. The word $a_{I+2}\cdots a_Ja_{J+1}$ is by Item (4) or Item (8) either of the form L_1 or L_2.

If there are infinitely many even denominators q_N, then we have shown that the sequence $a_1a_2a_3\cdots$ is concatenated from the blocks L_0, L_1, and L_2 (L_0 can only be a prefix). It remains to consider the case when only finitely many denominators q_N are even.

- Let q_N be odd for every $N \in \mathbb{N}$. Then $q_1 = a_1$ is odd. Especially, since both q_0 and $q_1 = a_1$ are odd, Item (2) implies $a_1 \in \{1,3\}$ and it follows from Item (6) that $a_2a_3a_4\cdots \in \{2,4\}^{\mathbb{N}}$. Therefore, the sequence $a_1a_2a_3\cdots$ is equal to the block L_3 from our list.
- Let $L \geq 1$ be the last index such that q_L is even. In particular, it means that q_N is even only for a finite number of indices. Item (1) implies that $a_{L+1} = 1$. By Items (5) and (6), the suffix $a_{L+2}a_{L+3}a_{L+4}\cdots$ of the sequence $a_1a_2a_3\cdots$ equals to L_3.

Now we have to show that any directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ such that $a_1a_2a_3\cdots$ is concatenated of the blocks from the list gives a standard Sturmian sequence $\mathbf{u}$ such that the CS Rote sequence $\mathbf{v}$ associated to $\mathbf{u}$ has the critical exponent less than $\frac{7}{2}$.

If q_N is odd, then $\frac{1}{2}\beta_N = \frac{1}{2}a_{N+1} + 1 + \frac{q'_{N-1}-1}{2q'_N} < 3 + \frac{q'_{N-1}}{2q'_{N-1}+q'_{N-2}} < \frac{7}{2}$, as each a_{N+1} is ≤ 4.

If q_N is even, it is easy to prove by induction on N that there is a block of the form L_0, L_1, or L_2 ending at the position $N+1$ (and L_0 only for $N = 2$). In particular, it means that $a_{N+1} = 1$ and $a_N \geq 2$ or $a_1 = a_2 = a_3 = 1$. In the first case, we get $\beta_N \leq 3 + \frac{q'_{N-1}}{2q'_{N-1}+q'_{N-2}} < \frac{7}{2}$, while in the second case, we get $\beta_2 = 3 + \frac{a_1+1-1}{a_2(a_1+1)+1} = 3 + \frac{1}{3} < \frac{7}{2}$.

To show that $\sup(M_1 \cup M_2) < \frac{7}{2}$, we need to show that $\sup \frac{q'_{N-1}}{2q'_{N-1}+q'_{N-2}} < \frac{1}{2}$. As $a_N \leq 4$ for all N, we can estimate $\frac{q'_{N-2}}{q'_{N-1}} \geq \frac{q'_{N-2}}{4q'_{N-2}+q'_{N-2}} = \frac{1}{5}$ and thus $\frac{q'_{N-1}}{2q'_{N-1}+q'_{N-2}} = \frac{1}{2+\frac{q'_{N-2}}{q'_{N-1}}} \leq \frac{5}{11}$. □

Remark 38. It would be interesting to reveal some topological properties of the set

$$\mathrm{cr}_{Rote} := \{\mathrm{cr}(\mathbf{v}) : \mathbf{v} \text{ is a CS Rote sequence}\},$$

for instance, to find its accumulation points in the interval $(3, \frac{7}{2})$. The proof of the previous theorem implies that there is no CS Rote sequence with the critical exponent between $3 + \frac{5}{11}$ and $3 + \frac{1}{2}$. We even believe that for any CS Rote sequence $\mathbf{v}$, the following implication holds: If $\mathrm{cr}(\mathbf{v}) < 3 + \frac{1}{2}$, then $\mathrm{cr}(\mathbf{v}) < 3 + \frac{1}{1+\sqrt{3}}$.

9 The recurrence function of CS Rote sequences

The main result of this section is Theorem 54, where we describe the recurrence function of any CS Rote sequence in terms of the convergents related to the associated Sturmian sequence. To obtain this result, we proceed similarly as in the previous parts concerning the critical exponent, i.e., we transform our task of finding the recurrence function of a CS Rote sequence into studying some properties of its associated Sturmian sequence. Let us emphasize that we may still restrict our consideration to CS Rote sequences associated with standard Sturmian sequences without loss of generality.

Definition 39. Let $\mathbf{u}$ be a uniformly recurrent sequence. The mapping $R_{\mathbf{u}} : \mathbb{N} \to \mathbb{N}$ defined by

$$R_{\mathbf{u}}(n) = \min\{N \in \mathbb{N} : \text{each factor of } \mathbf{u} \text{ of length } N \text{ contains all factors of } \mathbf{u} \text{ of length } n\}$$

is called the *recurrence function* of $\mathbf{u}$.

The definition of the recurrence function may be reformulated in terms of return words [6].

Observation 40. *Let $\mathbf{u}$ be a uniformly recurrent sequence. Then*

$$R_{\mathbf{u}}(n) = \max\{|r| \in \mathbb{N} : r \textit{ is a return word to a factor of } \mathbf{u} \textit{ of length } n\} + n - 1.$$

Moreover, to determine $R_{\mathbf{u}}(n)$ we can restrict our consideration to return words to bispecial factors of $\mathbf{u}$.

Lemma 41. *Let $\mathbf{u}$ be a uniformly recurrent aperiodic sequence. For $n \in \mathbb{N}$, we denote*

$$\mathcal{B}_{\mathbf{u}}(n) = \{b \in \mathcal{L}(\mathbf{u}) : \exists w \in \mathcal{L}(\mathbf{u}), |w| = n, \textit{ such that } b \textit{ is the shortest bispecial factor containing } w\}.$$

Then

$$R_{\mathbf{u}}(n) = \max\{|r| : r \textit{ is a return word to } b \in \mathcal{B}_{\mathbf{u}}(n)\} + n - 1.$$

Proof: For evaluation of $R_{\mathbf{u}}(n)$ we use Observation 40. Let $w \in \mathcal{L}(\mathbf{u})$ and $|w| = n$.

If w is not right special, then there exists a unique letter x such that $wx \in \mathcal{L}(\mathbf{u})$. Obviously, the occurrences of w and wx in $\mathbf{u}$ coincide. Therefore, return words to w and wx coincide as well.

If y is not left special, then there is a unique letter y such that $yw \in \mathcal{L}(\mathbf{u})$. If r is a return word to w, then the word yry^{-1} is a return word to yw and the return words r and yry^{-1} are of the same length.

These two facts imply that the lengths of return words to w equal the lengths of return words to the shortest bispecial factor containing w. □

The following lemma shows that for a CS Rote sequence $\mathbf{v}$ associated with the Sturmian sequence $\mathbf{u}$ the sets $\mathcal{B}_{\mathbf{v}}(n+1)$ and $\mathcal{B}_{\mathbf{u}}(n)$ correspond naturally for every $n \in \mathbb{N}$. Thus to determine the set $\mathcal{B}_{\mathbf{v}}(n+1)$, we first describe the set $\mathcal{B}_{\mathbf{u}}(n)$ for a standard Sturmian sequence $\mathbf{u}$.

Lemma 42. *Let w be a factor of length $n+1$ in a CS Rote sequence $\mathbf{v}$ and let v be the shortest bispecial factor of $\mathbf{v}$ containing w. Then the factor $\mathcal{S}(v)$ is the shortest bispecial factor of the associated Sturmian sequence $\mathcal{S}(\mathbf{v})$ such that $\mathcal{S}(v)$ contains $\mathcal{S}(w)$.*

Proof: The statement is a consequence of the simple fact that $\mathcal{S}(v)$ is a bispecial factor of $\mathcal{S}(\mathbf{v})$ if and only if v and $E(v)$ are bispecial factors of $\mathbf{v}$. (See Lemma 7.) □

In the sequel, we will essentially use a characterization of Sturmian sequences by palindromes from [10]. Let us first remind some basic notions. Consider an alphabet $\mathcal{A}$. The assignment $w = w_0w_1\cdots w_{n-1} \to \overline{w} = w_{n-1}w_{n-2}\cdots w_0$ is called a *mirror mapping*, and the word $\overline{w}$ is called the *reversal* or the *mirror image* of $w \in \mathcal{A}^*$. A word w which coincides with its mirror image $\overline{w}$ is a *palindrome*. If p is a palindrome of odd length, then the *center* of p is a letter a such that $p = sa\overline{s}$ for some $s \in \mathcal{A}^*$. The center of a palindrome p of even length is the empty word ε.

Theorem 43 ([10])**.** *A sequence $\mathbf{u}$ is Sturmian if and only if $\mathbf{u}$ contains one palindrome of every even length and two palindromes of every odd length.*

Moreover, when studying in detail the proof of Droubay and Pirillo 43, we deduce that any two palindromes of the same odd length have distinct centers, one has the center 0 and the other one has the center 1. In fact, we get the following corollary.

Corollary 44. *A binary sequence $\mathbf{u}$ is Sturmian if and only if every palindrome in $\mathcal{L}(\mathbf{u})$ has a unique palindromic extension, i.e., for any palindrome $p \in \mathcal{L}(\mathbf{u})$ there exists a unique letter $a \in \{0,1\}$ such that $apa \in \mathcal{L}(\mathbf{u})$.*

We believe that Theorem 46 is already known. However, since we have not found it in the literature, we add its proof. For this purpose, we need an auxiliary lemma. Let us remind that the language $\mathcal{L}(\mathbf{u})$ of a Sturmian sequence $\mathbf{u}$ is *closed under reversal*, i.e., $\mathcal{L}(\mathbf{u})$ contains with every factor w also its reversal $\overline{w}$, and all bispecial factors of $\mathbf{u}$ are palindromes.

Lemma 45. *Let $\mathbf{u}$ be a Sturmian sequence. Let $p \in \mathcal{L}(\mathbf{u})$ be a palindrome and let v be the shortest bispecial factor containing p.*

1. *Then p is a central factor of v, i.e., $v = sp\overline{s}$ for some word s.*
2. *If v' is the shortest bispecial factor with the same center as p and $|v'| \geq |p|$, then $v' = v$.*

Proof:

1. Let $v = sp$ be the shortest left special factor containing p, in particular $0sp$ and $1sp$ belong to $\mathcal{L}(\mathbf{u})$. Since the language $\mathcal{L}(\mathbf{u})$ is closed under reversal, $p\overline{s}$ is right special, i.e., $p\overline{s}0$ and $p\overline{s}1$ belong to

$\mathcal{L}(\mathbf{u})$. As s is the only possible extension of p to the left by a factor of length $|s|$, both $sp\overline{s}0$ and $sp\overline{s}1$ belong to $\mathcal{L}(\mathbf{u})$. By the same argument, $\overline{s}$ is the only possible extension of p to the right by a factor of length $|s|$. Therefore, $0sp\overline{s}$ and $1sp\overline{s}$ belong to $\mathcal{L}(\mathbf{u})$. Thus $sp\overline{s}$ is the shortest bispecial factor containing p.

2. Assume for contradiction that $v \neq v'$. Since v and v' are palindromes with the same centers, there exists a palindrome q such that $v' = s'q\overline{s'}$ and $v = sq\overline{s}$. Let q be the longest palindrome with this property. If $|q| = |v'|$, then necessarily $v' = v$. If $|v'| > |q|$, then the last letters of s' and s are distinct and q is a palindrome with two distinct palindromic extensions. This contradicts Corollary 44.

□

Theorem 46. *Let $\mathbf{u}$ be a Sturmian sequence and $n \in \mathbb{N}, n \geq 1$. Find the shortest bispecial factors P_ε, resp. P_0, resp. P_1 of length greater than or equal to n with the center ε, resp. 0, resp. 1. Then the following statements hold:*

1. *Let w be a factor of $\mathbf{u}$ of length n and let v be the shortest bispecial factor containing w. Then $v \in \{P_\varepsilon, P_0, P_1\}$.*

2. *If $\mathbf{u}$ contains no bispecial factor of length $n-1$, then for each $i \in \{\varepsilon, 0, 1\}$ there exists a factor w of $\mathbf{u}$ of length n such that the shortest bispecial factor containing w is P_i.*

3. *If there exists a bispecial factor v of $\mathbf{u}$ of length $n-1$ and let $i \in \{\varepsilon, 0, 1\}$ be the center of the palindrome v, then for each $j \in \{\varepsilon, 0, 1\}$, $j \neq i$, there exists a factor w of $\mathbf{u}$ of length n such that the shortest bispecial factor containing w is P_j, while P_i is not the shortest bispecial factor containing w for any factor w of $\mathbf{u}$ of length n.*

Proof: Consider the *Rauzy graph* of $\mathbf{u}$ of order $n-1$, i.e., a directed graph Γ_{n-1} whose vertices are factors of $\mathbf{u}$ of length $n-1$ and edges are factors of $\mathbf{u}$ of length n. An edge e starts in the vertex x and ends in the vertex y if x is a prefix and y is a suffix of e. Denote ℓ, resp. r the vertex corresponding to the unique left special, resp. right special factor of length $n-1$. Further on, denote p_A the shortest path from ℓ to r, and p_B and p_C the shortest paths of non-zero length starting in r and ending in ℓ. If $\mathbf{u}$ has no bispecial factor of length $n-1$, then p_A has a positive length, see Figure 1(a). If $\mathbf{u}$ has a bispecial factor of length $n-1$, then the path p_A consists of a unique vertex – the bispecial factor b, see Figure 1(b).

Observing these Rauzy graphs, it is obvious that for each edge e from the path p_x, where $x \in \{A, B, C\}$, the shortest bispecial factor containing e is the same as the shortest bispecial factor containing p_x.

Since the language $\mathcal{L}(\mathbf{u})$ is closed under reversal, the mirror mapping restricted to the factors of length $n-1$ and n is an automorphism of the graph Γ_{n-1}. Let us suppose that p_x contains a palindrome q of length $n-1$ or n. As any palindrome is mapped onto itself, $\overline{r} = \ell$, and $\overline{\ell} = r$, this path p_x is mapped onto itself, i.e., $\overline{p_x} = p_x$, and the palindrome q is a central factor of p_x. On one hand, it means that p_x is a palindrome with the same center as q, on the other hand, it also means that p_x cannot contain any other palindrome of length $n-1$ and n.

By Theorem 43 and the comment after it, there are exactly three palindromes among all vertices and edges of Γ_{n-1} (all factors of length $n-1$ or n), and moreover, they have distinct centers. We may

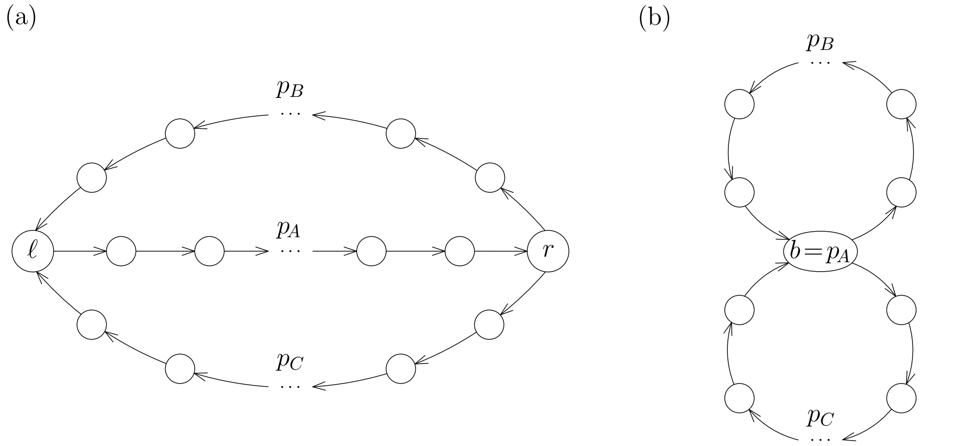

Fig. 1: The Rauzy graph of a Sturmian word (a) without a bispecial vertex, (b) with a bispecial vertex.

conclude that each path p_A, p_B, p_C contains exactly one palindrome of length $n-1$ or n. Therefore, all these paths are palindromes and their centers are distinct. The rest of the proof follows from Lemma 45. $\square$

Observation 47. *Let P be a palindrome. Then its Parikh vector satisfies:*

1. $\vec{V}(P) = \binom{0}{0} \mod 2$ *if and only if P has the center ε;*
2. $\vec{V}(P) = \binom{1}{0} \mod 2$ *if and only if P has the center* 0*;*
3. $\vec{V}(P) = \binom{0}{1} \mod 2$ *if and only if P has the center* 1.

Let us recall that a factor of a standard Sturmian sequence $\mathbf{u}$ is bispecial if and only if it is a palindromic prefix of $\mathbf{u}$. Therefore we can order the bispecial factors of a given standard Sturmian sequence $\mathbf{u}$ according to their lengths and denote $BS(k)$ the k-th bispecial factor of $\mathbf{u}$. Thus $BS(0) = \varepsilon$, $BS(1) = a$, where a is the first (and the more frequent) letter of $\mathbf{u}$ etc.

The sequences (p_N), (q_N), and (q'_N) we use in the remaining part of the paper were introduced in Notation 25, the notation $\mathcal{B}_{\mathbf{u}}(n)$ comes from Lemma 41.

Theorem 48. *Let $\mathbf{u}$ be a standard Sturmian sequence with the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ or $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$, and $n \in [q'_N, q'_{N+1})$ for some $N \in \mathbb{N}$. Put $M = a_0 + a_1 + a_2 + \cdots + a_N$, where $a_0 = 0$.*

- *If $n \in [q'_N, q'_{N+1} - 1)$ and $n-1$ is not the length of a bispecial factor, then*
$$\mathcal{B}_{\mathbf{u}}(n) = \{BS(M{+}m), BS(M{+}m{+}1), BS(M{+}a_{N+1}{+}1)\} \textit{ for some } m \in \{0, 1, \ldots, a_{N+1} - 1\}.$$
- *If $n \in [q'_N, q'_{N+1} - 1]$ and $n-1$ is the length of a bispecial factor, then*
$$\mathcal{B}_{\mathbf{u}}(n) = \{BS(M+m), BS(M+a_{N+1}+1)\} \textit{ for some } m \in \{0, 1, \ldots, a_{N+1}\}.$$

Proof: Assume that $\mathbf{u}$ has the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$. By Proposition 30, the Parikh vectors of bispecial factors satisfy

$$\vec{V}\big(BS(M+i+1)\big) = \vec{V}\big(BS(M+i)\big) + \begin{pmatrix} p_N \\ q_N \end{pmatrix} \quad \text{for } i = 0,1,\ldots,a_{N+1}-1 \tag{8}$$

$$\text{and} \qquad \vec{V}\big(BS(M+a_{N+1}+1)\big) = \vec{V}\big(BS(M+a_{N+1})\big) + \begin{pmatrix} p_{N+1} \\ q_{N+1} \end{pmatrix}. \tag{9}$$

Using Observation 47 and the relation $\begin{pmatrix} p_N \\ q_N \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \mod 2$ from Lemma 28, we deduce that the centers of palindromes $BS(M+i)$, where $i = 0,1,\ldots,a_{N+1}$, alternate between two distinct elements of $\{\varepsilon, 0, 1\}$. The third element of $\{\varepsilon, 0, 1\}$ is the center of the palindrome $BS(M+a_{N+1}+1)$, as $\begin{pmatrix} p_N \\ q_N \end{pmatrix} \neq \begin{pmatrix} p_{N+1} \\ q_{N+1} \end{pmatrix} \mod 2$, see Lemma 28. By Proposition 30, the length of $BS(M-1)$ equals $q'_N - 2$ and the length of $BS(M+a_{N+1}-1)$ equals $q'_{N+1} - 2$.

- Let us discuss the case $n = q'_{N+1} - 1$. The palindromes $BS(M+a_{N+1}-1)$, $BS(M+a_{N+1})$ and $BS(M+a_{N+1}+1)$ have distinct centers, and $n-1$ is the length of the palindrome $BS(M+a_{N+1}-1)$. Item (3) of Theorem 46 implies that all factors of length n occurs in $BS(M+a_{N+1})$ and $BS(M+a_{N+1}+1)$. Therefore, the set $\mathcal{B}_{\mathbf{u}}(n)$ consists of these two bispecial palindromes.

- Now we assume that $n \in [q'_N, q'_{N+1}-2]$. Clearly, the length of $BS(M-1)$ is strictly smaller then $n-1$ and n does not exceed the length of $BS(M+a_{N+1}-1)$. We choose the smallest $m \in \{0,1,\ldots,a_{N+1}-1\}$ such that $n \leq |BS(M+m)|$. The bispecial factors $BS(M+m)$, $BS(M+m+1)$, and $BS(M+a_{N+1}+1)$ have distinct centers.

  If $n-1$ is not the length of any bispecial factor, then Item (2) of Theorem 46 implies that $\mathcal{B}_{\mathbf{u}}(n) = \{BS(M+m), BS(M+m+1), BS(M+a_{N+1}+1)\}$.

  If $n-1$ is the length of a bispecial factor, then Item (3) of Theorem 46 together with the fact that the centers of $BS(M+i)$ alternate for $i = 0,1,\ldots,a_{N+1}-1$ implies that $\mathcal{B}_{\mathbf{u}}(n) = \{BS(M+m), BS(M+a_{N+1}+1)\}$.

If $\mathbf{u}$ has the directive sequence $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$, then the proof will be analogous, only the coordinates of the Parikh vectors will be exchanged, see Remark 31. □

Remark 49. The recurrence function of a standard Sturmian sequence $\mathbf{u}$ with the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ or $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$ is known to satisfy $R_{\mathbf{u}}(n) = q'_{N+1} + q'_N + n - 1$ for every $n \in [q'_N, q'_{N+1})$. Let us show that this formula is a consequence of the previous statements. Indeed, by Lemma 41, we have to find the longest return word to the bispecial factor from the set $\mathcal{B}_{\mathbf{u}}(n)$ described in Theorem 48. Using Proposition 30, we find that the longest one is the return word s corresponding to the bispecial factor $BS(M+a_{N+1}+1)$. Its length is $|s| = q'_{N+1} + q'_N$. And thus Lemma 41 implies the above mentioned formula, which was obtained by Hedlund and Morse already in 1940, see [18].

We have prepared everything we need to derive the formula for the recurrence function of CS Rote sequences. For this purpose, we recall Theorem 3.10 from [16]:

Theorem 50. *Let $\mathbf{v}$ be a CS Rote sequence associated with the standard Sturmian sequence $\mathbf{u} = S(\mathbf{v})$. Let v be a non-empty prefix of $\mathbf{v}$ and $u = S(v)$. Let r, resp. s be the more frequent, resp. the less frequent return word to u in $\mathbf{u}$ and let ℓ be a positive integer such that $\mathbf{u}$ is a concatenation of the blocks $r^\ell s$ and $r^{\ell+1}s$. Then the prefix v of $\mathbf{v}$ has three return words A, B, C satisfying:*

1. *If r is stable and s unstable, then $S(A0) = r, \quad S(B0) = sr^\ell s, \quad S(C0) = sr^{\ell+1}s.$*
2. *If r is unstable and s stable, then $S(A0) = s, \quad S(B0) = rr, \quad S(C0) = rsr.$*
3. *If both r and s are unstable, then $S(A0) = rr, \quad S(B0) = rs, \quad S(C0) = sr.$*

We will use the previous theorem for the determination of return words to bispecial factors of CS Rote sequences (which are by Lemma 42 associated to bispecial factors of Sturmian sequences). In particular, we focus on v such that $\mathcal{S}(v)$ is a bispecial factor of the Sturmian sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$ and $\mathcal{S}(v)$ belongs to the set $\mathcal{B}_{\mathbf{u}}(n)$ described in Theorem 48.

Lemma 51. *Let $\mathbf{v}$ be a CS Rote sequence and $\mathbf{u} = \mathcal{S}(\mathbf{v})$ be the associated standard Sturmian sequence with the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ or $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$. Put $M = a_0+a_1+a_2+\cdots+a_N$, where $a_0 = 0$. Let x and y be the bispecial factors in $\mathbf{v}$ such that $\mathcal{S}(x) = BS(M + a_{N+1} + 1)$ and $\mathcal{S}(y) = BS(M+m)$, where $m \in \{0, 1, \ldots, a_{N+1} - 1\}$. Then at least one return word to x in $\mathbf{v}$ is longer than every return word to y in $\mathbf{v}$.*

Proof: On one hand, by Lemmas 19, 20 and Remark 21, the derived sequence $\mathbf{d}_{\mathbf{u}}(\mathcal{S}(y))$ is a standard Sturmian sequence with the directive sequence $G^{a_{N+1}-m}D^{a_{N+2}}G^{a_{N+3}}\cdots$ or $D^{a_{N+1}-m}G^{a_{N+2}}D^{a_{N+3}}\cdots$. It implies that $\mathbf{u}$ is a concatenation of the blocks ${r'}^\ell s'$ and ${r'}^{\ell+1}s'$, where $\ell = a_{N+1} - m$. The return words to $\mathcal{S}(y)$ are by Proposition 30 of length $|r'| = p_N + q_N = q'_N$ and $|s'| = m\,q'_N + q'_{N-1}$. Regardless of (un)stability of the return words to $\mathcal{S}(y)$, the longest return word to y in $\mathbf{v}$ is by Theorem 50 of length at most

$$(\ell+1)|r'| + 2|s'| = (a_{N+1} - m + 1)q'_N + 2(m\,q'_N + q'_{N-1}) = q'_{N+1} + (m+1)q'_N + q'_{N-1} \le 2q'_{N+1}.$$

On the other hand, the return words r, s to $\mathcal{S}(x)$ in $\mathbf{u}$ have by Proposition 30 either lengths $|r| = q'_{N+1}$ and $|s| = q'_{N+1} + q'_N$ (if $a_{N+2} > 1$), or lengths $|r| = q'_{N+1} + q'_N$ and $|s| = q'_{N+1}$ (if $a_{N+2} = 1$). Regardless of (un)stability of the return words to $\mathcal{S}(x)$, one of the return words to x in $\mathbf{v}$ is of length at least

$$|r| + |s| = 2q'_{N+1} + q'_N\,.$$

□

Lemma 52. *Let $\mathbf{v}$ be a CS Rote sequence and $\mathbf{u} = \mathcal{S}(\mathbf{v})$ be the associated standard Sturmian sequence with the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ or $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$. Put $M = a_0+a_1+a_2+\cdots+a_N$, where $a_0 = 0$. Let x and y be the bispecial factors in $\mathbf{v}$ such that $\mathcal{S}(x) = BS(M + a_{N+1} + 1)$ and $\mathcal{S}(y) = BS(M + a_{N+1})$. Then at least one return word to x in $\mathbf{v}$ is longer than every return word to y in $\mathbf{v}$.*

Proof: Let us denote the return words to $\mathcal{S}(y)$ by r' and s', and the return words to $\mathcal{S}(x)$ by r and s. By Proposition 30, $|r'| = q'_{N+1}$ and $|s'| = q'_N$.

First, we assume that r' is unstable. Then by Theorem 50, the return words to y in $\mathbf{v}$ are of length at most $2|r'|+|s'| = 2q'_{N+1}+q'_N$. Regardless of (un)stability of the return words to $\mathcal{S}(x)$, one of the return words to x in $\mathbf{v}$ is of length at least $|r|+|s| = 2q'_{N+1}+q'_N$.

It remains to discuss the case when r' is stable. Let us assume that $\mathbf{u}$ has the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$. By Proposition 30, it means that $|r'|_1 = q_{N+1}$ is even. We use Theorem 50 to find the longest return word to y. Similarly as in the proof of Lemma 51, we determine that $\ell = a_{N+2}$ and thus the longest return word to y in $\mathbf{v}$ is of length

$$L' = 2|s'| + (\ell+1)|r'| = 2\,q'_N + (a_{N+2}+1)q'_{N+1} = q'_{N+2} + q'_{N+1} + q'_N\,.$$

Let us compare L' with the length of the return words to x in $\mathbf{v}$. If $a_{N+2} > 1$, then by Proposition 30, $|r|_1 = |r'|_1 = q_{N+1}$ and r is stable as well. The longest return word to x is by Theorem 50 the return word $sr^{\ell+1}s$, where $\ell = a_{N+2}-1$. Its length is

$$L = 2|s| + a_{N+2}|r| = 2(q'_{N+1}+q'_N) + a_{N+2}q'_{N+1} = q'_{N+2} + 2q'_{N+1} + q'_N > L'\,.$$

If $a_{N+2} = 1$, then by Proposition 30, $|r|_1 = q_{N+2} = q_{N+1}+q_N$ and $|s|_1 = q_{N+1}$. Since q_{N+1} is even, necessarily q_N is odd (as follows from the well-known relation $p_Nq_{N+1} - p_{N+1}q_N = (-1)^{N+1}$ for all N). It means that r is unstable and s is stable. Thus the longest return word to x in $\mathbf{v}$ has the length

$$L = 2|r| + |s| = 2(q'_{N+1}+q'_N) + q'_{N+1} = 2q'_{N+2} + q'_{N+1} > L'.$$

If $\mathbf{u}$ has the directive sequence $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$, the proof is analogous, we only have to take into account that the coordinates of the Parikh vectors are exchanged (see Remark 31). In particular, instead of considering q_N when determining the number of ones, we consider p_N. □

Proposition 53. *Let* $\mathbf{v}$ *be a CS Rote sequence. Let* $\mathbf{u}$ *be the standard Sturmian sequence such that* $\mathcal{L}(\mathcal{S}(\mathbf{v})) = \mathcal{L}(\mathbf{u})$ *and let* $\mathbf{u}$ *have the directive sequence* $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ *or* $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$*. Put* $M = a_0+a_1+a_2+\cdots+a_N$*, where* $a_0 = 0$*. Let* L *be the length of the longest return word to the bispecial factor* v *of* $\mathbf{v}$ *such that* $\mathcal{S}(v)$ *is the bispecial factor* $BS(M+a_{N+1}+1)$ *in* $\mathbf{u}$*. Then the recurrence function of* $\mathbf{v}$ *satisfies* $R_{\mathbf{v}}(n+1) = L+n$ *for any* $n \in [q'_N, q'_{N+1})$*, where* $N \in \mathbb{N}$*.*

Proof: Let $\mathbf{v}'$ be a CS Rote sequence associated with the standard Sturmian sequence $\mathbf{u}$ such that $\mathcal{L}(\mathcal{S}(\mathbf{v})) = \mathcal{L}(\mathbf{u})$. Clearly, $\mathcal{L}(\mathbf{v}) = \mathcal{L}(\mathbf{v}')$. Since the recurrence function depends only on the language of the sequence and not on the sequence itself, we can work with the CS Rote sequence $\mathbf{v}'$ instead of $\mathbf{v}$.

It follows from Lemma 41 that $R_{\mathbf{v}}(n+1) = R_{\mathbf{v}'}(n+1) = L+n$, where L is the length of the longest return word to a bispecial factor from the set $\mathcal{B}_{\mathbf{v}'}(n+1)$. Lemma 42 shows the correspondence between the bispecial factors from the set $\mathcal{B}_{\mathbf{v}'}(n+1)$ and the bispecial factors from the set $\mathcal{B}_{\mathbf{u}}(n)$. In particular, if $v \in \mathcal{B}_{\mathbf{v}'}(n+1)$, then $\mathcal{S}(v) \in \mathcal{B}_{\mathbf{u}}(n)$. For every $n \in [q'_N, q'_{N+1})$, where $N \in \mathbb{N}$, the set $\mathcal{B}_{\mathbf{u}}(n)$ is described in Theorem 48. Together with Lemmas 51 and 52, it implies that the bispecial factor v such that $S(v)$ equals $BS(M+a_{N+1}+1)$ has the longest return word among all bispecial factors from the set $\mathcal{B}_{\mathbf{v}'}(n+1)$. □

Theorem 54. *Let* $\mathbf{v}$ *be a CS Rote sequence and let* $\mathbf{u}$ *be the standard Sturmian sequence such that* $\mathcal{L}(\mathcal{S}(\mathbf{v})) = \mathcal{L}(\mathbf{u})$*. If* $\mathbf{u}$ *has the directive sequence* $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$*, then the value of the recurrence function* $R_{\mathbf{v}}$ *for* $n \in [q'_N, q'_{N+1})$*,* $N \in \mathbb{N}$*, is given by*

Case q_N even: $R_{\mathbf{v}}(n+1) = \begin{cases} 2q'_{N+1} + q'_N + n & \textit{if} \ \ a_{N+2} > 1, \\ 2q'_{N+2} + n & \textit{if} \ \ a_{N+2} = 1. \end{cases}$

Case q_{N+1} even: $R_{\mathbf{v}}(n+1) = \begin{cases} q'_{N+2} + 2q'_{N+1} + q'_N + n & \textit{if} \ \ a_{N+2} > 1 \\ 2q'_{N+2} + q'_{N+1} + n & \textit{if} \ \ a_{N+2} = 1. \end{cases}$

Case q_N, q_{N+1} odd: $R_{\mathbf{v}}(n+1) = \begin{cases} 3q'_{N+1} + q'_N + n & \textit{if} \ \ a_{N+2} > 1 \\ q'_{N+3} + q'_{N+2} + q'_{N+1} + n & \textit{if} \ \ a_{N+2} = 1. \end{cases}$

If $\mathbf{u}$ *has the directive sequence* $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$, *then the value of the recurrence function* $R_{\mathbf{v}}$ *for* $n \in [q'_N, q'_{N+1})$, $N \in \mathbb{N}$, *is given by*

Case p_N even: $R_{\mathbf{v}}(n+1) = \begin{cases} 2q'_{N+1} + q'_N + n & \textit{if} \ \ a_{N+2} > 1, \\ 2q'_{N+2} + n & \textit{if} \ \ a_{N+2} = 1. \end{cases}$

Case p_{N+1} even: $R_{\mathbf{v}}(n+1) = \begin{cases} q'_{N+2} + 2q'_{N+1} + q'_N + n & \textit{if} \ \ a_{N+2} > 1 \\ 2q'_{N+2} + q'_{N+1} + n & \textit{if} \ \ a_{N+2} = 1. \end{cases}$

Case p_N, p_{N+1} odd: $R_{\mathbf{v}}(n+1) = \begin{cases} 3q'_{N+1} + q'_N + n & \textit{if} \ \ a_{N+2} > 1 \\ q'_{N+3} + q'_{N+2} + q'_{N+1} + n & \textit{if} \ \ a_{N+2} = 1. \end{cases}$

Proof: By Proposition 53, $R_{\mathbf{v}}(n+1) = L + n$, where L is the length of the longest return word to the bispecial factor v in $\mathbf{v}$ such that $\mathcal{S}(v) = BS(M + a_{N+1} + 1)$. Consider first that $\mathbf{u}$ has the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$. By Proposition 30, the Parikh vectors of the return words r and s to the bispecial factor $BS(M + a_{N+1} + 1)$ are

1. $\vec{V}(r) = \begin{pmatrix} p_{N+1} \\ q_{N+1} \end{pmatrix}$, $\vec{V}(s) = \begin{pmatrix} p_{N+1} + p_N \\ q_{N+1} + q_N \end{pmatrix}$ if $a_{N+2} > 1$;
2. $\vec{V}(r) = \begin{pmatrix} p_{N+1} + p_N \\ q_{N+1} + q_N \end{pmatrix}$, $\vec{V}(s) = \begin{pmatrix} p_{N+1} \\ q_{N+1} \end{pmatrix}$ if $a_{N+2} = 1$.

Let us emphasize that at most one of the numbers q_N and q_{N+1} is even. It follows from the well-known relation $p_N q_{N+1} - p_{N+1} q_N = (-1)^{N+1}$ for all N. Moreover, let us recall that $p_N + q_N = q'_N$.

First we discuss the case $a_{N+2} > 1$.

- If q_N is even, then q_{N+1} is odd, i.e., r and s are unstable. Since $|r| < |s|$, Item (3) of Theorem 50 gives $L = |rs| = 2q'_{N+1} + q'_N$.
- If q_{N+1} is even, then q_N is odd, i.e., r is stable and s is unstable. We use Item (1) of Theorem 50 with $\ell = a_{N+2} - 1$. Clearly, $L = 2|s| + (\ell+1)|r| = 2(q'_{N+1} + q'_N) + a_{N+2}q'_{N+1} = q'_{N+2} + 2q'_{N+1} + q'_N$.
- If both q_N, q_{N+1} are odd, then r is unstable and s is stable. Item (2) of Theorem 50 implies $L = |rsr| = 3q'_{N+1} + q'_N$.

It remains to discuss the case $a_{N+2} = 1$.

- If q_N is even, then q_{N+1} is odd, i.e., r and s are unstable. Since $|r| > |s|$, Item (3) of Theorem 50 gives $L = |rr| = 2q'_{N+1} + 2q'_N = 2q'_{N+2}$.
- If q_{N+1} is even, then q_N is odd, i.e., r is unstable and s is stable. Item (2) of Theorem 50 implies $L = |rsr| = 3q'_{N+1} + 2q'_N = 2q'_{N+2} + q'_{N+1}$.
- If both q_N, q_{N+1} are odd, then r is stable and s is unstable. We use Item (1) of Theorem 50 with $\ell = a_{N+3}$. Thus $L = 2|s| + (\ell+1)|r| = 2q'_{N+1} + (a_{N+3}+1)(q'_{N+1}+q'_N) = q'_{N+3}+q'_{N+2}+q'_{N+1}$.

If $\mathbf{u}$ has the directive sequence $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$, then the statement of Theorem 54 will stay the same, only q_N and q_{N+1} will be replaced by p_N and p_{N+1} because the Parikh vectors of r and s have the coordinates exchanged, see Remark 31. □

Example 55. By Proposition 34, the critical exponent of the CS Rote sequence $\mathbf{v}$ such that $S(\mathbf{v})$ has the directive sequence $G(D^2G^2)^\omega$ is $\mathrm{cr}(\mathbf{v}) = 2 + \frac{1}{\sqrt{2}}$. In Example 36, we have shown that the CS Rote sequence $\mathbf{v}'$ associated to the Sturmian sequence $S(\mathbf{v}') = E(S(\mathbf{v}))$ has the critical exponent $\mathrm{cr}(\mathbf{v}') = 4 + \frac{1}{1+\sqrt{2}}$.

Let us find an explicit formula for the recurrence function $R_{\mathbf{v}}$, resp. $R_{\mathbf{v}'}$ of the CS Rote sequence $\mathbf{v}$, resp. $\mathbf{v}'$. We will see that these recurrence functions differ essentially, too. In the proof of Proposition 34, we have shown that all q_N are odd and we have found an explicit formula for q'_N, see (7). Applying Theorem 54, we obtain for every $n \in [q'_N, q'_{N+1})$

$$R_{\mathbf{v}}(n+1) = 3q'_{N+1} + q'_N + n = n + \frac{1}{2\sqrt{2}}\Big((4+3\sqrt{2})(1+\sqrt{2})^{N+1} - (4-3\sqrt{2})(1-\sqrt{2})^{N+1}\Big).$$

Further on, p_N is even if and only if N is even. Therefore, we obtain for every $n \in [q'_{2N}, q'_{2N+1})$

$$R_{\mathbf{v}'}(n+1) = 2q'_{2N+1} + q'_{2N} + n = n + \frac{1}{2\sqrt{2}}\Big((1+\sqrt{2})^{2N+3} - (1-\sqrt{2})^{2N+3}\Big);$$

and for every $n \in [q'_{2N-1}, q'_{2N})$

$$R_{\mathbf{v}'}(n+1) = q'_{2N+1} + 2q'_{2N} + q'_{2N-1} + n = n + \frac{1}{\sqrt{2}}\Big((1+\sqrt{2})^{2N+2} - (1-\sqrt{2})^{2N+2}\Big).$$

References

[1] L. Balková, E. Pelantová, and Š. Starosta, *Sturmian jungle (or garden?) on multiliteral alphabets*, RAIRO Theor. Inf. Appl. 44 (2010), 443–470.

[2] A. R. Baranwal and J. Shallit, *Critical exponent of infinite balanced words via the Pell number system*, in: R. Mercas and D. Reidenbach (eds.), Proceedings WORDS 2019, Lecture Notes in Computer Science, vol. 11682, Springer (2019), 80–92.

[3] A. R. Baranwal and J. Shallit, *Repetitions in infinite palindrome-rich words*, in: R. Mercas and D. Reidenbach (eds.), Proceedings WORDS 2019, Lecture Notes in Computer Science, vol. 11682, Springer (2019), 93–105.

[4] J. Berstel, *Axel Thue's papers on repetitions in words: a translation,* Publications du LaCIM 20, Université du Québec à Montréal (1995).

[5] A. Blondin Massé, S. Brlek, S. Labbé, and L. Vuillon, *Palindromic complexity of codings of rotations*, Theoret. Comput. Sci. 412(46) (2011), 6455–6463.

[6] J. Cassaigne and N. Chekhova, *Fonctions de récurrence des suites d'Arnoux-Rauzy et réponse à une question de Morse et Hedlund*, Ann. Inst. Fourier (Grenoble) 56(7) (2006), 2249–2270.

[7] J. D. Currie, L. Mol, and N. Rampersad, *The repetition threshold for binary rich words*, Discrete Math. Theoret. Comput. Sci. 22(1) (2020), no. 6.

[8] D. Damanik and D. Lenz, *The index of Sturmian sequences*, European J. Combin. 23 (2002), 23–29.

[9] X. Droubay, J. Justin, and G. Pirillo, *Episturmian words and some construction of de Luca and Rauzy*, Theor. Comput. Sci. 255 (2001), 539–553.

[10] X. Droubay and G. Pirillo, *Palindromes and Sturmian words*, Theoret. Comput. Sci. 223 (1999), 73–85.

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

[12] D. Hensley, *Continued Fractions*, World Scientific Publishing (2006).

[13] J. Justin, and G. Pirillo, *Episturmian words and episturmian morphisms*, Theoret. Comput. Sci. 276 (2002), 281–313.

[14] K. Klouda, K. Medková, E. Pelantová, and Š. Starosta, *Fixed points of Sturmian morphisms and their derivated words*, Theoret. Comput. Sci. 743 (2018), 23–37.

[15] M. Lothaire, *Combinatorics on Words*, Encyclopaedia of Mathematics and its Applications, vol. 17, Addison-Wesley, Reading, Mass. (1983). Reprinted in the Cambridge Mathematical Library, Cambridge University Press (1997).

[16] K. Medková, E. Pelantová, and L. Vuillon, *Derived sequences of complementary symmetric Rote sequences*, RAIRO-Theor. Inf. Appl. 53 (2019), 125–151.

[17] F. Mignosi and G. Pirillo, *Repetitions in the Fibonacci words*, RAIRO-Theor. Inf. Appl. 26 (1992), 199–204.

[18] M. Morse and G. A. Hedlund, *Symbolic dynamics II. Sturmian trajectories*, Amer. J. Math. 62 (1940), 1–42.

[19] E. Pelantová and Š. Starosta, *Languages invariant under more symmetries: overlapping factors versus palindromic richness*, Discrete Math. 313 (2013), 2432–2445.

[20] N. Rampersad, J. Shallit, and É. Vandomme, *Critical exponents of infinite balanced words*, Theor. Comp. Sci. 777 (2019), 454–463.

[21] M. Rigo, *Formal Languages, Automata and Numeration Systems*, 2 vols., Wiley (2014).

[22] G. Rote, *Sequences with subword complexity* $2n$, J. Number Theory 46 (1994), 196–213.

[23] A. Thue, *Probleme über Veränderungen von Zeichenreihen nach gegebene Regeln*, Christiana Videnskabs-Selskabs Skrifter, I. Math.-naturv. Klasse 10 (1914).

[24] L. Vuillon, *A characterization of Sturmian words by return words*, Eur. J. Combin. 22 (2001), 263–275.