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.

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 $\mathbf{u} = u_0 u_1 u_2 \cdots$ be a sequence and let $w = u_i u_{i+1} \cdots u_{i+n-1}$ be its factor. The index i is an occurrence of w. A return word to w is a word $u_i u_{i+1} \cdots u_{j-1}$ 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

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 $\mathbf{u} = u_0 u_1 u_2 \cdots$, 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 $y = u_i u_{i+1} u_{i+2} \cdots u_{j-1}$ for some $i, j \in \mathbb{N}$, $i \leq j$. 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 $\mathcal{C}_{\mathbf{u}}(n) \geq n + 1$ 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

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

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 $u_i u_{i+1} \cdots u_{j-1}$ 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 $\mathbf{d}_{\mathbf{u}}(w) = d_0 d_1 d_2 \cdots$ 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 $\mathbf{u} = u_0 u_1 u_2 \cdots \in \{0,1\}^\mathbb{N}$ 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 $0^k 1$ and $\mathbf{c}_\alpha$ can be uniquely written as a concatenation of the blocks $0^k 1$ and $0^{k+1} 1$;

– if $\alpha < \frac{1}{2}$, then $\mathbf{c}_\alpha$ has a prefix $1^k 0$ and $\mathbf{c}_\alpha$ can be uniquely written as a concatenation of the blocks $1^k 0$ 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

$$\varphi_b = F \circ E : \begin{cases} 0 \to 0 \\ 1 \to 01 \end{cases} \quad \text{and} \quad \varphi_\beta = E \circ F : \begin{cases} 0 \to 10 \\ 1 \to 1 \end{cases}.$$

Their corresponding matrices are:

$$M_b = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad M_\beta = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

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 $0^k 1$ 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 $0^k 1$. Therefore $\mathbf{u} = \varphi_b(\mathbf{u}')$. $\square$

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

$$\mathbf{u}^{(n)} \in \{0,1\}^{\mathbb{N}} \quad \text{is a standard Sturmian sequence} \quad \text{and} \quad \mathbf{u} = \varphi_{z_0 z_1 \ldots z_{n-1}}(\mathbf{u}^{(n)}) \quad \text{for every} \quad n \in \mathbb{N}.$$

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 $0^k 1$ 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 $1^k 0$ 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 $\mathbf{u} = u_0 u_1 \cdots$ and $\mathbf{v} = v_0 v_1 \cdots$ 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 $v = v_0 v_1 \cdots v_n \in \{0, 1\}^+$ of length at least 2 we put $\mathcal{S}(v_0 v_1 \cdots v_n) = u_0 u_1 \cdots u_{n-1}$, where

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

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. $\qquad\square$

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 $\mathbf{u} = u_0 u_1 \cdots$ which is fixed by the Sturmian morphism $\psi : 0 \to 010$, $1 \to 01001$, *i.e.*

$$\mathbf{u} = 01001001010010010010100100100101001001010 \cdots .$$

The associated Rote sequence $\mathbf{v} = v_0 v_1 \cdots$ (*i.e.* $\mathbf{u} = \mathcal{S}(\mathbf{v})$) starting with 0 is

$$\mathbf{v} = 0011100011000111000110001110001100011100011 \cdots .$$

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.* $v_0 v_1 = 00$. 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 $u_0 u_1 = 01$ (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.* $v_2 v_3 = 11$. 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 $v_3 v_4$ 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 $v_3 v_4 = 11$. 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 $v_5 v_6 = 00$, 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 $v_0 v_1 v_2 v_3 v_4 = 00111$.

Similarly we get $v_6 = v_5 + u_5 = 0 + 0 = 0$ and thus $v_6 v_7 = 00$, 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 $v_8 v_9 = 11$. So we have to wait until another factor 01 appears in $\mathbf{u}$. It happens immediately since $u_8 u_9 = 01$. Thus $v_{10} = v_8 + u_8 + u_9 = 0 \mod 2$ and $v_{10} v_{11} = 00$. Therefore the word $v_6 v_7 v_8 v_9 = 0011$ 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 $u = u_0 u_1 \cdots u_{n-1} \in \{0,1\}^*$ 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 $u = u_0 u_1 \cdots u_{m-1}$ 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 $u = u_0 u_1 \cdots u_{m-1}$ 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

$$\left\{ \binom{|u|_0}{|u|_1} \mod 2 : uw \text{ is a prefix of } \mathbf{u} \right\} = \left\{ \binom{0}{0}, \binom{0}{1}, \binom{1}{0}, \binom{1}{1} \right\}.$$

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* **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* **v** *and* $w = \mathcal{S}(x)$. *Then the prefix* x *of* **v** *has three return words* $A, B, C \in \{0,1\}^+$ *satisfying* (r, s *and* k *are the same as in Def.* 3.8):

$\quad$ (i) *if* $\mathcal{T}_w = SU(k)$, *then* $\quad \mathcal{S}(A0) = r$, $\quad \mathcal{S}(B0) = sr^{k+1}s$ $\quad$ *and* $\quad \mathcal{S}(C0) = sr^k s$;
$\quad$ (ii) *if* $\mathcal{T}_w = US(k)$, *then* $\quad \mathcal{S}(A0) = rr$, $\quad \mathcal{S}(B0) = rsr$ $\qquad$ *and* $\quad \mathcal{S}(C0) = s$;
$\quad$ (iii) *if* $\mathcal{T}_w = UU(k)$, *then* $\quad \mathcal{S}(A0) = rr$, $\quad \mathcal{S}(B0) = rs$ $\qquad\;$ *and* $\quad \mathcal{S}(C0) = sr$.

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 **v**. Then by Lemma 3.7 the index n is an occurrence of w in **u** and the prefix $u = u_0 u_1 \cdots u_{n-1}$ is stable. Since **u** is a concatenation of the blocks $r^{k+1}s$ and $r^k s$, the sequence **u** has one of the prefixes ur, $usr^{k+1}s$ or $usr^k s$.

$\quad$ – If ur is a prefix of **u**, then $n + |r|$ is an occurrence of w in **u**. Moreover, the prefix of **u** of length $n + |r|$ is stable. It means that $m := n + |r|$ is the subsequent occurrence of x in **v** and $A := v_n v_{n+1} \cdots v_{m-1}$ is a return word to x in **v**. Let us recall our convention that 0 is a prefix of **v** and thus any return word to the prefix x begins with 0, in particular $v_m = 0$. Therefore, $r = u_n u_{n+1} \cdots u_{m-1} = \mathcal{S}(A0)$.

$\quad$ – If $usr^{k+1}s$ is a prefix of **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 **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 **v**. The next occurrence of w in **u** is $m := n + |s| + (k+1)|r| + |s|$. The prefix of **u** of length m is stable and thus m is the smallest occurrence of x in **v** grater than n. Therefore $B := v_n \cdots v_{m-1}$ is a return word to x in **v** and obviously $sr^{k+1}s = \mathcal{S}(B0)$.

$\quad$ The reasoning in all remaining cases is analogous and so we omit it. $\qquad\square$

Example 3.11 (Example 3.3 continued). Recall that the prefix 00 of **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 **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 $\mathcal{S}(C0) = \mathcal{S}(001110) = 01001 = rsr$.

$\quad$ It remains to explain how to determine the type of a given prefix w of a standard Sturmian sequence **u**. This question will be solved in Section 5.

4. Derived sequences of complementary symmetric Rote sequences

$\quad$ As we have proved in Theorem 3.1, any derived sequence of a Rote sequence **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 **v** is associated with a standard Sturmian sequence. First we mention an important direct consequence of Theorem 3.10.

Corollary 4.1. *Let* **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* **v**. *Then the derived sequence* $\mathbf{d_v}(x)$ *is uniquely determined by the derived sequence* $\mathbf{d_u}(w)$ *of* **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 **u** and let **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 **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. $\qquad\square$

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 $\mathbf{u} = u_0 u_1 u_2 \cdots \in \{A, B, C\}^{\mathbb{N}}$ 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.*,

$$T(I_A) = \left[1 - \alpha, 1\right), \quad T(I_B) = \left[(k + 1)(1 - \alpha) - \alpha, 1 - \alpha\right), \quad T(I_C) = \left[0, (k + 1)(1 - \alpha) - \alpha\right).$$

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. $\qquad\square$

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 $\mathbf{d_u}(w^{(1)}) = \mathbf{d_{u^{(1)}}}(\varepsilon) = \mathbf{u}^{(1)}$.

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 $r^{(n)} = \varphi_{z_0 z_1 \ldots z_{n-1}}(0)$ and $s^{(n)} = \varphi_{z_0 z_1 \ldots z_{n-1}}(1)$.
(ii) If $z_n = \beta$, then $r^{(n)} = \varphi_{z_0 z_1 \ldots z_{n-1}}(1)$ and $s^{(n)} = \varphi_{z_0 z_1 \ldots z_{n-1}}(0)$.

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 $\mathbf{z}^{(1)} = z_1 z_2 z_3 \cdots$. 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:

$$P_w = \left(\begin{array}{cc} |r|_0 & |s|_0 \\ |r|_1 & |s|_1 \end{array} \right) \quad \mod 2\,.$$

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 $P_w = \left(\begin{array}{cc} p & q \\ 0 & 1 \end{array} \right)$ for some numbers $p, q \in \{0, 1\}$;

(ii) US if $P_w = \left(\begin{array}{cc} p & q \\ 1 & 0 \end{array} \right)$ for some numbers $p, q \in \{0, 1\}$;

(iii) UU if $P_w = \left(\begin{array}{cc} p & q \\ 1 & 1 \end{array} \right)$ for some numbers $p, q \in \{0, 1\}$.

Example 5.7. Take the prefixes $\epsilon, 0$ and 010 of Sturmian sequence $\mathbf{u}$ from Example 3.3. Their matrices are

$$P_\varepsilon = \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right), \quad P_0 = \left(\begin{array}{cc} 1 & 1 \\ 1 & 0 \end{array} \right) \quad \text{and} \quad P_{010} = \left(\begin{array}{cc} 0 & 1 \\ 1 & 1 \end{array} \right)$$

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 $\varphi_{z_0 z_1 \ldots z_{n-1}}$. Thus the matrix $P^{(n)}$ can be obtained as a product of the matrix of the morphism $\varphi_{z_0 z_1 \ldots z_{n-1}}$ 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 $z_1 z_2 z_3 \cdots = \beta bb\beta \cdots$ 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
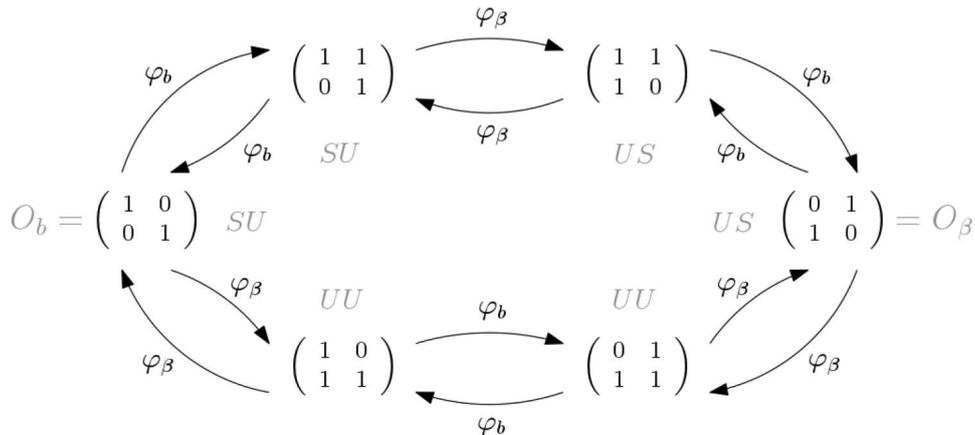
FIGURE 1. The diagram captures the multiplication mod 2 of matrices with the determinant 1 by morphism matrices M_b and M_β. A directed edge labelled by φ_b goes from matrix M to matrix M' if $M' = M_b M \mod 2$. Analogously for the label φ_β.

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 $P^{(n)}$:

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}. \qquad (5.2)$$

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 **u** with the directive sequence **z** has bispecial prefixes of at least two types. Indeed, by Proposition 5.9 (or by the graph in Fig. 1):
  – if **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 **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 **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 **g** associated with the Fibonacci sequence **f**.

Example 5.13. The Fibonacci sequence **f** is fixed by the Fibonacci morphism $F : 0 \to 01$, $1 \to 0$, *i.e.*

$$\mathbf{f} = 01001010010010100101001001010100100101 \cdots .$$

Clearly, the sequence **f** is a fixed point of the morphism F^2, too. Since $F^2 = \varphi_{b\beta}$, the Fibonacci sequence **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 $z_0 z_1 z_2 \cdots = b\beta b\beta \cdots$, thus $w^{(0)}$ has the type $\mathcal{T}^{(0)} = SU(1)$;

- $P^{(1)} = M_b O_\beta = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ mod 2 and $z_1 z_2 z_3 \cdots = \beta b\beta b \cdots$, thus $\mathcal{T}^{(1)} = US(1)$;

- $P^{(2)} = M_b M_\beta O_b = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ mod 2 and $z_2 z_3 z_4 \cdots = b\beta b\beta \cdots$, thus $\mathcal{T}^{(2)} = UU(1)$;

- $P^{(3)} = M_b M_\beta M_b O_\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ mod 2 and $z_3 z_4 z_5 \cdots = \beta b\beta b \cdots$, thus $\mathcal{T}^{(3)} = SU(1)$;

- $P^{(4)} = M_b M_\beta M_b M_\beta O_b = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ mod 2 and $z_4 z_5 z_6 \cdots = b\beta b\beta \cdots$, thus $\mathcal{T}^{(4)} = US(1)$;

- $P^{(5)} = M_b M_\beta M_b M_\beta M_b O_\beta = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ mod 2 and $z_5 z_6 z_7 \cdots = \beta b\beta b \cdots$, thus $\mathcal{T}^{(5)} = UU(1)$.

By simple computations we get

$$M_b M_\beta M_b M_\beta M_b M_\beta = I \mod 2 ,$$

where I is the identity matrix. Then we have

$$P^{(6)} = M_b M_\beta M_b M_\beta M_b M_\beta O_b = O_b \mod 2 .$$

We have also $z_6 z_7 z_8 \cdots = b\beta b\beta \cdots = z_0 z_1 z_2 \cdots$. 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, $sr^2 s$ 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, $sr^2 s$ and srs. The order of these blocks gives us the derived sequence of $\mathbf{g}$ to $x^{(0)}$

$$\mathbf{d_g}(x^{(0)}) = ABABAACAACAABABABAAC \cdots .$$

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 $\mathbf{z} = z_0 z_1 z_2 \cdots$. 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

$$P^{(n)} = M_{z_0} \cdots M_{z_{n-1}} O_{z_n} = \left(M_{z_0} \cdots M_{z_{p-1}} \right) \left(M_{z_p} \cdots M_{z_{n-1}} \right) O_{z_n},$$

$$P^{(n+qQ)} = P^{(m)} = M_{z_0} \cdots M_{z_{m-1}} O_{z_m} = \left(M_{z_0} \cdots M_{z_{p-1}} \right) H^q \left(M_{z_p} \cdots M_{z_{n-1}} \right) O_{z_n} = P^{(n)}.$$

$\square$

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 $\mathbf{z} = z_0 z_1 z_2 \cdots$. By Corollary 5.4, the sequence $\left(\mathbf{d}_{\mathbf{u}}(w^{(n)})\right)_{n \in \mathbb{N}}$ 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 $z_n z_{n+1} z_{n+2} \cdots$. 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 $\left(\mathcal{T}^{(n)}, \mathbf{d}_{\mathbf{u}}(w^{(n)})\right)_{n \in \mathbb{N}}$ 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. $\square$

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. $\square$

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

$$a' = r' = 01001, \quad b' = s'r'r's' = 0100100101001010, \quad c' = s'r's' = 01001001010,$$

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

$$\begin{array}{ccccccc} a' & = & r' & = & \psi^3(r) & = & rsrrs & = & ab, \\ b' & = & s'r'r's' & = & \psi^3(srrs) & = & rsrrsrrsrsrrsrsr & = & abaacaaca, \\ c' & = & s'r's' & = & \psi^3(srs) & = & rsrrsrrsrsr & = & abaaca. \end{array}$$

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

$$
\begin{array}{ccccccccc}
a' & = & r'r' & = & \psi^3(rr) & = & rsrrsrsrrs & = & bbcac, \\
b' & = & r's'r' & = & \psi^3(rsr) & = & rsrrsrsrrsrrs & = & bbcacac, \\
c' & = & s' & = & \psi^3(s) & = & rsr & = & b.
\end{array}
$$

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

$$
\begin{array}{ccccccccc}
a' & = & r'r' & = & \psi^3(rr) & = & rsrrsrsrrs & = & baccb, \\
b' & = & r's' & = & \psi^3(rs) & = & rsrrsrsr & = & bacc, \\
c' & = & s'r' & = & \psi^3(sr) & = & rsrrsrrs & = & bacb.
\end{array}
$$

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 $\mathcal{T}^{(2)} = UU(2)$.

– $\varphi_{\text{cyc}^2(b\beta b)} = \varphi_{bb\beta} : \begin{cases} 0 \to 0010 \\ 1 \to 001 \end{cases}$ 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

$\psi^2(a) = \psi^2(rr) = rrsrrrsrrrsrrsrrrsrrrsrrsr$;

$\psi^2(b) = \psi^2(rs) = rrsrrrsrrrsrrsrrrsrrrsrrs$;

$\psi^2(c) = \psi^2(sr) = rrsrrrsrrrsrrsrrrsrrrsrrsr$.

– $\psi^2(a) = \underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{sr}_{c}$,
thus $w_a = acacacbababababac$;

$\psi^2(b) = \underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rr}_{a}\underbrace{sr}_{c}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}\underbrace{rr}_{a}\underbrace{rs}_{b}$,
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.*

$$\mathbf{q} = 1324124132132413213241241324124132132412413 \cdots ,$$

and we apply the projection π to construct the complementary symmetric Rote sequence:

$$\mathbf{r} = \pi(\mathbf{q}) = 0011011001001100100110110011011001001011001101100 \cdots .$$

Using the operation $\mathcal{S}$ (see Def. 2.5), we obtain the associated Sturmian sequence

$$\mathbf{s} = \mathcal{S}(\mathbf{r}) = 0101101011010101101011010101101011010110101011010 \cdots .$$

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

$$\mathbf{u} = 1\mathbf{s} = 10101101011010101101011010101101011010110101011 0101 \cdots$$

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

$$\mathbf{v} = 0110010011011001101100100110010011011001001100100110 \cdots .$$

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

$$\mathbf{v} = 0110010011011001101100 10011 \cdots = \overline{1001101100100110010011 01100} \cdots = \overline{1\mathbf{r}} .$$

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

$$\rho'(\sigma(A)) = \rho'(ABC) = \pi(1324124132) = \pi(\xi(1)\xi(3)\xi(2)\xi(4)),$$
$$\rho'(\sigma(B)) = \rho'(AC) = \pi(1324132) = \pi(\xi(1)\xi(2)\xi(4)),$$
$$\rho'(\sigma(C)) = \rho'(AB) = \pi(1324124) = \pi(\xi(1)\xi(3)\xi(2)),$$

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:

$$\sigma_n : \begin{cases} A \to A^n ABA^n C \\ B \to A^n AC \\ C \to A^n AB \end{cases} \quad \text{and} \quad \rho' : \begin{cases} A \to 0011 \\ B \to 011 \\ C \to 001 \end{cases}.$$

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 $\psi^2 = \varphi_{\beta b b \beta}$ 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 **u** has an S-adic representation **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 **z** represent derived sequences to prefixes of **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 **u** as introduced in [16] takes into account only the prefixes of **u**. Recently, Yu-Ke Huang and Zhi-Ying Wen in [20] have considered also the derived sequences of **u** to non-prefix factors of **u**. Recall that if **u** is a fixed point of a primitive morphism, then by Durand's result from [16], any derived sequence to a prefix of **u** is fixed by a primitive morphism as well. However, a derived sequence to a non-prefix factor of **u** need not to be fixed by a non-identical morphism at all.

Huang and Wen study the period-doubling sequence **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 **d** to a factor of **u** is equal to Θ_1 or Θ_2. Moreover, any derived sequence **d**' to a factor of **d** is equal to Θ_1 or Θ_2 and any derived sequence **d**'' to a factor of **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 **p** and the Thue-Morse sequence **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 **v** associated with a fixed point **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 **u** has exactly $|z|$ distinct derived sequences (see [23]) and thus by Corollary 4.1, the Rote sequence **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.

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.