

Complementary symmetric Rote sequences: the critical exponent and the recurrence function*

L'ubom'ira Dvoř'akov'á Kateřina Medkov'á Edita Pelantov'á

FNSPE, Czech Technical University in Prague, Czech republic

received 17th Mar. 2020, revised 24th May 2020, accepted 24th May 2020.

We determine the critical exponent and the recurrence function of complementary symmetric Rote sequences. The formulae are expressed in terms of the continued fraction expansions associated with the S-adic representations of the corresponding standard Sturmian sequences. The results are based on a thorough study of return words to bispecial factors of Sturmian sequences. Using the formula for the critical exponent, we describe all complementary symmetric Rote sequences with the critical exponent less than or equal to 3, and we show that there are uncountably many complementary symmetric Rote sequences with the critical exponent less than the critical exponent of the Fibonacci sequence. Our study is motivated by a conjecture on sequences rich in palindromes formulated by Baranwal and Shallit. Its recent solution by Curie, Mol, and Rampersad uses two particular complementary symmetric Rote sequences.

Keywords: critical exponent, recurrence function, Rote sequence, Sturmian sequence, return word, bispecial factor

Introduction

We study the relation between the critical exponents of two binary sequences $\mathbf{v} = v_0v_1v_2 \cdots$ and $\mathbf{u} = u_0u_1u_2 \cdots$ over the alphabet $\{0, 1\}$, where $u_i = v_i + v_{i+1} \pmod 2$ for each $i \in \mathbb{N}$. We write $\mathbf{u} = \mathcal{S}(\mathbf{v})$. Our study is motivated by a conjecture formulated by Baranwal and Shallit in [3]. They searched for binary sequences rich in palindromes with a minimum critical exponent. They showed that the value of this critical exponent is greater than 2.707. Moreover, they found two sequences $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ having the critical exponent equal to $2 + \frac{1}{\sqrt{2}}$ and they conjectured that this is the minimum value. Both of these sequences belong to the class of complementary symmetric Rote sequences. Their conjecture has been recently proved by Curie, Mol, and Rampersad in [8].

A Rote sequence is a binary sequence \mathbf{v} containing $2n$ factors of length n for every $n \in \mathbb{N}, n \geq 1$. If the language of \mathbf{v} is invariant under the exchange of letters $0 \leftrightarrow 1$, the sequence \mathbf{v} is called a complementary

*The research received funding from the Ministry of Education, Youth and Sports of the Czech Republic through the project no. CZ.02.1.01/0.0/0.0/16_019/0000778 and from the Grant Agency of the Czech Technical University in Prague through the grant no. SGS20/183/OHK4/3T/14.

symmetric (CS) Rote sequence. Already in his original paper [24], Rote proved that these sequences are essentially connected with Sturmian sequences. He deduced that a binary sequence \mathbf{v} is a CS Rote sequence if and only if the sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$ is Sturmian. Both CS Rote sequences and Sturmian sequences are rich in palindromes, see [6, 10].

The formula for the critical exponent of Sturmian sequences was provided by Damanik and Lenz in [9]. The relation between the critical exponent of a CS Rote sequence \mathbf{v} and the associated Sturmian sequence $\mathcal{S}(\mathbf{v})$ is not straightforward: While the minimum exponent among all Sturmian sequences is reached by the Fibonacci sequence and it is $3 + \frac{2}{1+\sqrt{5}}$ (see [19]), the two CS Rote sequences $\mathbf{v}^{(1)}$, $\mathbf{v}^{(2)}$ whose critical exponent equals $2 + \frac{1}{\sqrt{2}}$, i.e., it is minimum among all binary rich sequences, are associated with the Sturmian sequences $\mathcal{S}(\mathbf{v}^{(1)})$ and $\mathcal{S}(\mathbf{v}^{(2)})$ whose critical exponent is $3 + \sqrt{2}$.

In this paper, we will first derive the relation between the critical exponents of the sequences \mathbf{v} and $\mathcal{S}(\mathbf{v})$, where \mathbf{v} is a uniformly recurrent binary sequence whose language is closed under the exchange of letters, see Theorem 14. Using this relation, we will determine the formula for the critical exponent of any CS Rote sequence, see Theorem 33.

One of the consequences of this theorem is for instance the fact that the languages of the sequences $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ are the only languages of CS Rote sequences with the critical exponent less than 3, see Proposition 34. In this context, let us mention that in [8] the authors showed that there are exactly two languages of rich binary sequences with the critical exponent less than $\frac{14}{5}$ and they are the languages of the sequences $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$. Furthermore, we show that there are uncountably many CS Rote sequences with the critical exponent strictly less than the critical exponent of the Fibonacci sequence, see Theorem 37.

Our main technical tool is the description of return words to bispecial factors of Sturmian sequences in terms of the continued fraction expansions related to the S-adic representations of Sturmian sequences. As a by-product, we obtain an explicit formula for the recurrence function of CS Rote sequences, see Theorem 54. When formulating our results, we use the convergents $(\frac{p_N}{q_N})$ of an irrational number $\theta = [0, a_1, a_2, a_3, \dots]$, where the coefficients a_i 's in the continued fraction expansion of θ correspond to the S-adic representation of the standard Sturmian sequence associated to a given CS Rote sequence.

There are many generalizations of Sturmian sequences to multiliteral alphabets, see [1]. The critical exponent and the recurrence function were studied for two of these generalizations. Justin and Pirillo described in [15] the critical exponent of substitutive Arnoux-Rauzy sequences. Recently, Rampersad, Shallit, and Vandomme in [22], and Baranwal and Shallit in [2] determined the minimal threshold for the critical exponent of balanced sequences over alphabets of cardinality 3, 4, and 5, respectively. The recurrence function of Sturmian sequences was found by Morse and Hedlund in [20], and their result was generalized by Cassaigne and Chekhova in [7] for Arnoux-Rauzy sequences.

The paper is organized as follows. We first introduce basic notions from combinatorics on words in Section 1. In Section 2, we recall how to simplify the formula for the critical exponent using return words to bispecial factors. The definitions of the already mentioned mapping \mathcal{S} and complementary symmetric Rote sequences and their basic properties are provided in Section 3. The relation between the critical exponents of the sequences \mathbf{v} and $\mathcal{S}(\mathbf{v})$ is described in Section 4. The main tool for further results – a thorough study of return words to bispecial factors of Sturmian sequences using the S-adic representation – is carried out in Section 5. An explicit formula for the critical exponent of CS Rote sequences is given in Section 6. CS Rote sequences with a small critical exponent are studied in Section 7. And finally, in Section 8, an explicit formula for the recurrence function of CS Rote sequences is derived.

1 Preliminaries

An *alphabet* \mathcal{A} is a finite set of symbols called *letters*. A *word* over \mathcal{A} of *length* n is a string $u = u_0u_1 \cdots u_{n-1}$, where $u_i \in \mathcal{A}$ for all $i \in \{0, 1, \dots, n-1\}$. The length of u is denoted by $|u|$. The set of all finite words over \mathcal{A} together with the operation of concatenation form a monoid \mathcal{A}^* . Its neutral element is the *empty word* ε and we denote $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$.

If $u = xyz$ for some $x, y, z \in \mathcal{A}^*$, then x is a *prefix* of u , z is a *suffix* of u and y is a *factor* of u . We sometimes use the notation $yz = x^{-1}u$.

To any word u over \mathcal{A} with cardinality $\#\mathcal{A} = d$, we assign its *Parikh vector* $\vec{V}(u) \in \mathbb{N}^d$ defined as $(\vec{V}(u))_a = |u|_a$ for all $a \in \mathcal{A}$, where $|u|_a$ is the number of letters a occurring in u .

A *sequence* over \mathcal{A} is an infinite string $\mathbf{u} = u_0u_1u_2 \cdots$, where $u_i \in \mathcal{A}$ for all $i \in \mathbb{N} = \{0, 1, 2, \dots\}$. We always denote sequences by bold letters. A sequence \mathbf{u} is *eventually periodic* if $\mathbf{u} = vwww \cdots = v(w)^\omega$ for some $v \in \mathcal{A}^*$ and $w \in \mathcal{A}^+$. Otherwise \mathbf{u} is *aperiodic*.

A *factor* of \mathbf{u} is a word y such that $y = u_iu_{i+1}u_{i+2} \cdots u_{j-1}$ for some $i, j \in \mathbb{N}$, $i \leq j$. The number i is called an *occurrence* of the factor y in \mathbf{u} . In particular, if $i = j$, the factor y is the empty word ε and any index i is its occurrence. If $i = 0$, the factor y is a *prefix* of \mathbf{u} . If each factor of \mathbf{u} has infinitely many occurrences in \mathbf{u} , the sequence \mathbf{u} is *recurrent*. Moreover, if for each factor the distances between its consecutive occurrences are bounded, \mathbf{u} is *uniformly recurrent*.

The *language* $\mathcal{L}(\mathbf{u})$ of the sequence \mathbf{u} is the set of all factors of \mathbf{u} . A factor w of \mathbf{u} is *right special* if both words wa and wb are factors of \mathbf{u} for at least two distinct letters $a, b \in \mathcal{A}$. Analogously we define a *left special* factor. A factor is *bispecial* if it is both left and right special. Note that the empty word ε is a bispecial factor if at least two distinct letters occur in \mathbf{u} .

The *factor complexity* of a sequence \mathbf{u} is a mapping $\mathcal{C}_{\mathbf{u}} : \mathbb{N} \rightarrow \mathbb{N}$ defined by

$$\mathcal{C}_{\mathbf{u}}(n) = \#\{w \in \mathcal{L}(\mathbf{u}) : |w| = n\}.$$

The aperiodic sequences with the lowest possible factor complexity are called *Sturmian sequences*. In other words, it means that a sequence \mathbf{u} is *Sturmian* if it has the factor complexity $\mathcal{C}_{\mathbf{u}}(n) = n + 1$ for all $n \in \mathbb{N}$. Clearly, all Sturmian sequences are defined over a binary alphabet, e.g., $\{0, 1\}$. There are many equivalent definitions of Sturmian sequences, see a survey in [1].

A *morphism* over \mathcal{A} is a mapping $\psi : \mathcal{A}^* \rightarrow \mathcal{A}^*$ such that $\psi(uv) = \psi(u)\psi(v)$ for all $u, v \in \mathcal{A}^*$. The morphism ψ can be naturally extended to sequences by

$$\psi(\mathbf{u}) = \psi(u_0u_1u_2 \cdots) = \psi(u_0)\psi(u_1)\psi(u_2) \cdots.$$

A *fixed point* of a morphism ψ is a sequence \mathbf{u} such that $\psi(\mathbf{u}) = \mathbf{u}$. The *matrix* of a morphism ψ over \mathcal{A} with the cardinality $\#\mathcal{A} = d$ is the matrix $M_\psi \in \mathbb{N}^{d \times d}$ defined as $(M_\psi)_{ab} = |\psi(a)|_b$ for all $a, b \in \mathcal{A}$. The Parikh vector of the ψ -image of a word $w \in \mathcal{A}^*$ can be obtained via multiplication by the matrix M_ψ , i.e.,

$$\vec{V}(\psi(w)) = M_\psi \vec{V}(w). \quad (1)$$

Consider a prefix w of a recurrent sequence \mathbf{u} . Let $i < j$ be two consecutive occurrences of w in \mathbf{u} . Then the word $u_iu_{i+1} \cdots u_{j-1}$ is a *return word* to w in \mathbf{u} . The set of all return words to w in \mathbf{u} is denoted by $\mathcal{R}_{\mathbf{u}}(w)$. If the sequence \mathbf{u} is uniformly recurrent, the set $\mathcal{R}_{\mathbf{u}}(w)$ is finite for each prefix w , i.e., $\mathcal{R}_{\mathbf{u}}(w) = \{r_0, r_1, \dots, r_{k-1}\}$. 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 $\#\mathcal{R}_{\mathbf{u}}(w) = k$. The concept of derived sequences was introduced by Durand in [12].

2 The critical exponent and its relation to return words

Let $z \in \mathcal{A}^+$ be a prefix of a periodic sequence u^ω with $u \in \mathcal{A}^+$. We say that z has the *fractional root* u and the *exponent* $e = |z|/|u|$. We usually write $z = u^e$. Let us emphasize that a word z can have multiple exponents and fractional roots. A word z is *primitive* if its only integer exponent is 1.

Let \mathbf{u} be a sequence and u its non-empty factor. The supremum of $e \in \mathbb{Q}$ such that u^e is a factor of \mathbf{u} is the *index* of u in \mathbf{u} :

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

If the sequence \mathbf{u} is clear from the context, we will write $\text{ind}(u)$ instead of $\text{ind}_{\mathbf{u}}(u)$.

Definition 1. The *critical exponent* of a sequence \mathbf{u} is

$$\begin{aligned} \text{cr}(\mathbf{u}) &= \sup\{e \in \mathbb{Q} : \text{there is a non-empty factor of } \mathbf{u} \text{ with the exponent } e\} \\ &= \sup\{\text{ind}_{\mathbf{u}}(u) : u \text{ is a non-empty factor of } \mathbf{u}\}. \end{aligned}$$

Remark 2. Let us comment on the above definition.

1. If a non-empty factor $u \in \mathcal{L}(\mathbf{u})$ is non-primitive, i.e., $u = x^k$ for some $x \in \mathcal{A}^+$ and $k \in \mathbb{N}, k \geq 2$, then $\text{ind}_{\mathbf{u}}(x) = k \text{ind}_{\mathbf{u}}(u) > \text{ind}_{\mathbf{u}}(u)$. Therefore, only primitive factors play a role for finding $\text{cr}(\mathbf{u})$.
2. If some non-empty factor occurs at least twice in \mathbf{u} , then $\text{ind}(x) > 1$ for some non-empty factor x and so $\text{cr}(\mathbf{u}) > 1$. Consequently, $\text{cr}(\mathbf{u}) > 1$ for each sequence \mathbf{u} .
3. We say that u is an *overlapping factor* in \mathbf{u} , if there exist $x, y \in \mathcal{A}^*$ such that $xu = uy \in \mathcal{L}(\mathbf{u})$ and $0 < |x| < |u|$. If \mathbf{u} has an overlapping factor, then $\text{cr}(\mathbf{u}) > 2$. Indeed, by [17] the equality $xu = uy$ implies that there exist $a, b \in \mathcal{A}^*$ and $k \in \mathbb{N}$ such that $u = (ab)^k a$, $x = ab$, and $y = ba$. If a is empty then the assumption $|u| > |x| > 0$ forces $k \geq 2$, otherwise $k \geq 1$. In both cases $\text{ind}_{\mathbf{u}}(ab) > 2$.
4. If \mathbf{u} is eventually periodic, then $\text{cr}(\mathbf{u})$ is infinite.
5. If \mathbf{u} is aperiodic and uniformly recurrent, then each factor of \mathbf{u} has a finite index. Nevertheless, $\text{cr}(\mathbf{u})$ may be infinite. A Sturmian sequence may serve as an example of such a sequence if the continued fraction expansion of its slope has unbounded partial quotients, see [9].
6. If \mathbf{u} is a binary sequence, then either 11, 00, or 0101 occur in \mathbf{u} . It means that the critical exponent of a binary sequence is at least 2. This value is attained by the famous Thue-Morse sequence, which is, of course, overlap-free, see [25] or [4]. The critical exponent of generalized Thue-Morse sequences over multi-letter alphabets were described in [5].
7. Recently, Ghareghani and Sharifani introduced a natural generalization of k -bonacci sequences to infinite alphabets. They also computed the critical exponent of these sequences, [13].

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

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

Proof: Let $u \in \mathcal{L}(\mathbf{u})$ be a non-empty factor with the index $\text{ind}(u) > 1$. Denote $|u| = n$. When searching for supremum, we may assume without loss of generality that u is a factor having the largest index among all factors of \mathbf{u} of length n , i.e., $\text{ind}(u) \geq \text{ind}(v)$ for all $v \in \mathcal{L}(\mathbf{u})$ of length n . Since \mathbf{u} is uniformly recurrent, $\text{ind}(u)$ is finite. We denote

$$z = u^{\text{ind}(u)} = u'u'' \cdots u'u''u' \quad \text{and} \quad b = u^{\text{ind}(u)-1},$$

where $u = u'u''$ and $u'' \neq \varepsilon$. Clearly, $z = bu''u' = u'u''b$.

Let us show that the word b is a bispecial factor of \mathbf{u} . The word z is a factor of \mathbf{u} and so z occurs in \mathbf{u} at some position j , i.e.,

$$z = u_j u_{j+1} \cdots u_{j+|z|-1}.$$

Then the letter $u_{j+|z|}$ which follows the word z is distinct from the first letter of u'' . Otherwise, we could prolong z to the right, which contradicts the definition of the index of $u = u_j u_{j+1} \cdots u_{j+n-1}$. Similarly, the letter u_{j-1} which precedes z is distinct from the last letter of u'' . Indeed, if those letters are the same, then the factor $u_{j-1} u_j \cdots u_{j+n-2}$ of length n has the index at least $\text{ind}(u) + \frac{1}{n}$, which contradicts the choice of u . We can conclude that the factor b is a bispecial factor of \mathbf{u} .

Moreover, since $z = bu''u' = u'u''b$, the word $u = u'u''$ is a concatenation of the return words to the bispecial factor b . It suffices to prove that only the cases when u is a return word to the bispecial factor b have to be inspected. Let us assume that u is a concatenation of at least two return words to b . It means that

$$z = ub = sbt \quad \text{for some } s, t \text{ such that } s \text{ is a prefix of } u \text{ and } 0 < |s| < |u|. \quad (2)$$

We will find another factor of \mathbf{u} with the index strictly larger than $\text{ind}(u)$, which means that such a factor u can be omitted. We distinguish three cases:

- If $\text{ind}(u) \geq 2$, then $|b| \geq |u|$ and both words u and s are prefixes of b . Therefore, the relation (2) implies $us = su$ and we can easily conclude that there is a word x and an integer $k > 1$ such that $u = x^k$. As mentioned in Remark 2, $\text{ind}(x) > \text{ind}(u)$.
- If $1 < \text{ind}(u) < 2$ and $\text{cr}(\mathbf{u}) \leq 2$, then by Item 3 of Remark 2, \mathbf{u} has no overlapping factor. Clearly, $z = u'u''u'$ and $b = u'$ for $u', u'' \neq \varepsilon$. Then the relation (2) implies $u'v = su'$ for some v and $|u| > |s| \geq |u'|$. Indeed, if $0 < |s| < |u'|$, then u' is an overlapping factor, which is not possible. Therefore, u' is a prefix of s and we can easily deduce that

$$\text{ind}(s) \geq \frac{|s| + |u'|}{|s|} > \frac{|u| + |u'|}{|u|} = \text{ind}(u).$$

- If $1 < \text{ind}(u) < 2$ and $\text{cr}(\mathbf{u}) > 2$, then there is a factor $x \in \mathcal{L}(\mathbf{u})$ with $\text{ind}(x) > 2 > \text{ind}(u)$.

□

Remark 4. In fact, we proved that it suffices to consider the set

$$\mathcal{M}' = \{u : u \text{ is a return word to a bispecial factor of } \mathbf{u} \text{ with the fractional root } u\}$$

or, even more specifically, the set

$$\mathcal{M}'' = \{u : u \text{ is a return word to the bispecial factor } b = u^{\text{ind}(u)-1} \text{ of } \mathbf{u}\}$$

instead of \mathcal{M} . Clearly, $\mathcal{M}'' \subset \mathcal{M}' \subset \mathcal{M}$.

In Remark 2, we emphasize that only primitive factors are relevant for finding the critical exponent. Let us verify that all return words from the set \mathcal{M}' (and so \mathcal{M}'' , too) are primitive. We prove it by contradiction. Let us suppose that $u \in \mathcal{M}'$ is non-primitive, i.e., $u = x^k$ for some non-empty x and $k \in \mathbb{N}$, $k > 1$. Since b has the fractional root u , $b = u^\ell = x^{k\ell}$ for some $\ell \in \mathbb{Q}$. Therefore, $ub = x^{k(\ell+1)} = xx^{k\ell}x^{k-1} = xbx^{k-1}$, which contradicts that u is a return word to b in \mathbf{u} .

3 The mapping \mathcal{S} on binary words and complementary symmetric Rote sequences

In this section, we introduce a mapping \mathcal{S} which enables us to describe the properties of CS Rote sequences using Sturmian sequences. Nevertheless, this mapping \mathcal{S} can be applied to any binary sequence.

Definition 5. By \mathcal{S} we denote the mapping $\mathcal{S} : \{0, 1\}^+ \mapsto \{0, 1\}^*$ such that for every $v_0 \in \{0, 1\}$ we put $\mathcal{S}(v_0) = \varepsilon$ and for every $v = v_0v_1 \cdots v_n \in \{0, 1\}^+$ of length at least 2 we put $\mathcal{S}(v_0v_1 \cdots v_n) = u_0u_1 \cdots u_{n-1}$, where

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

Moreover, we extend the domain of \mathcal{S} naturally to $\{0, 1\}^{\mathbb{N}}$: for every $\mathbf{v} \in \{0, 1\}^{\mathbb{N}}$ we put $\mathcal{S}(\mathbf{v}) = \mathbf{u}$, where

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

By $E : \{0, 1\}^* \mapsto \{0, 1\}^*$ we denote the morphism which exchanges the letters, i.e., $E(0) = 1$, $E(1) = 0$.

Example 6. We have $E(001110) = 110001$ and $\mathcal{S}(001110) = \mathcal{S}(110001) = 01001$.

Clearly, the images of v and $E(v)$ under \mathcal{S} coincide for each $v \in \{0, 1\}^*$. Moreover, $\mathcal{S}(x) = \mathcal{S}(y)$ if and only if $x = y$ or $x = E(y)$. The following rule follows directly from the definition of \mathcal{S} :

$$\mathcal{S}(v_0v_1 \cdots v_n) = \mathcal{S}(v_0v_1 \cdots v_k)\mathcal{S}(v_kv_{k+1} \cdots v_n) \quad \text{for any } k = 0, \dots, n. \quad (3)$$

These observations hold also for infinite sequences.

Lemma 7. Let \mathbf{v} be a binary sequence whose language $\mathcal{L}(\mathbf{v})$ is closed under E . Then $w \neq \varepsilon$ is a right (left) special factor in \mathbf{v} if and only if $\mathcal{S}(w)$ is a right (left) special factor in $\mathcal{S}(\mathbf{v})$.

Proof: We will prove the statement for right special factors. The proof for left special factors is analogous. Let c be the last letter of w .

(\implies): Let $w0$ and $w1$ belong to $\mathcal{L}(\mathbf{v})$. Then $\mathcal{S}(w0)$ and $\mathcal{S}(w1)$ belong to $\mathcal{L}(\mathcal{S}(\mathbf{v}))$. By the rule (3), $\mathcal{S}(w0) = \mathcal{S}(w)\mathcal{S}(c0)$ and $\mathcal{S}(w1) = \mathcal{S}(w)\mathcal{S}(c1)$. As $\mathcal{S}(c0) \neq \mathcal{S}(c1)$, the factor $\mathcal{S}(w)$ is right special in $\mathcal{S}(\mathbf{v})$.

(\impliedby): Let $\mathcal{S}(w)0$ and $\mathcal{S}(w)1$ be in $\mathcal{L}(\mathcal{S}(\mathbf{v}))$. Then $\mathcal{S}(w)0 = \mathcal{S}(w)\mathcal{S}(cc) = \mathcal{S}(wc)$ and $\mathcal{S}(w)1 = \mathcal{S}(w)\mathcal{S}(cE(c)) = \mathcal{S}(wE(c))$. Since $\mathcal{L}(\mathbf{v})$ is closed under the exchange of letters, all factors wc , $E(wc)$, $wE(c)$ and $E(w)c$ are in $\mathcal{L}(\mathbf{v})$. It means that w and $E(w)$ are right special factors in \mathbf{v} . \square

A *Rote sequence* is a sequence \mathbf{v} with the factor complexity $\mathcal{C}_{\mathbf{v}}(n) = 2n$ for all $n \in \mathbb{N}, n \geq 1$. Clearly, all Rote sequences are defined over a binary alphabet, e.g., $\{0, 1\}$. If the language of a Rote sequence \mathbf{v} is closed under the exchange of letters, i.e., $E(v) \in \mathcal{L}(\mathbf{v})$ for each $v \in \mathcal{L}(\mathbf{v})$, the Rote sequence \mathbf{v} is called *complementary symmetric* (shortly CS).

Rote in [24] proved that these sequences are essentially connected with Sturmian sequences.

Proposition 8 ([24]). *Let \mathbf{u} and \mathbf{v} be two sequences over $\{0, 1\}$ such that $\mathbf{u} = \mathcal{S}(\mathbf{v})$. Then \mathbf{v} is a complementary symmetric Rote sequence if and only if \mathbf{u} is a Sturmian sequence.*

Let us emphasize that to a given CS Rote sequence \mathbf{v} there is the unique *associated* Sturmian sequence \mathbf{u} such that $\mathbf{u} = \mathcal{S}(\mathbf{v})$. On the other hand, for any Sturmian sequence \mathbf{u} there exist two *associated* CS Rote sequences \mathbf{v} and $E(\mathbf{v})$ such that $\mathbf{u} = \mathcal{S}(\mathbf{v}) = \mathcal{S}(E(\mathbf{v}))$. However, $\mathcal{L}(\mathbf{v}) = \mathcal{L}(E(\mathbf{v}))$.

Analogously, to a given factor $v \in \mathcal{L}(\mathbf{v})$ there is a unique associated word u such that $u = \mathcal{S}(v)$ and this word u is a factor of \mathbf{u} . In addition, to a given factor $u \in \mathcal{L}(\mathbf{u})$ there are exactly two associated words $v, E(v)$ such that $\mathcal{S}(v) = \mathcal{S}(E(v)) = u$ and both these words $v, E(v)$ are factors of \mathbf{v} .

Example 9. Let us underline that for Sturmian sequences \mathbf{u} and $E(\mathbf{u})$ the languages of their associated CS Rote sequences may essentially differ. Consider the Fibonacci sequence

$$\mathbf{f} = abaababaaba \dots,$$

which is the fixed point of the Fibonacci morphism $F : a \rightarrow ab, b \rightarrow a$.

- If $a = 0$ and $b = 1$, then the associated CS Rote sequence starting with 0 is $\mathbf{v} = 001110011100 \dots$. The prefix w of \mathbf{v} of length 7 is $w = 0011100 = (00111)^{\frac{7}{5}}$, i.e., w has the fractional root 00111 and the exponent $\frac{7}{5}$.
- If $a = 1$ and $b = 0$, then the associated CS Rote sequence starting with 0 is $\mathbf{v}' = 011011001001 \dots$. The prefix w' of \mathbf{v}' of length 7 is $w' = 0110110 = (011)^{\frac{7}{3}}$, i.e., w' has the fractional root 011 and the exponent $\frac{7}{3}$.

We will show later in Example 36 that even the critical exponent of CS Rote sequences associated with \mathbf{u} and $E(\mathbf{u})$ may be different.

In the next section, we will explain that the relation between the shortest fractional root of a factor v and the shortest fractional root of $\mathcal{S}(v)$ is influenced by the number of letters 1 occurring in the shortest fractional root of $\mathcal{S}(v)$. This is the reason for the following definition and lemma.

Definition 10. A word $u = u_0u_1 \dots u_{n-1} \in \{0, 1\}^*$ is called *stable* if $|u|_1 = 0 \pmod{2}$. Otherwise, u is *unstable*.

Lemma 11. *Let $\mathcal{S} : \{0, 1\}^+ \mapsto \{0, 1\}^*$.*

- (i) If 0 is a prefix of $v \in \{0, 1\}^*$, then $\mathcal{S}(v0)$ is stable.
- (ii) For every $u \in \{0, 1\}^*$ there exists a unique $w \in \{0, 1\}^+$ with a prefix 0 such that $u = \mathcal{S}(w)$. Moreover, w has a suffix 0 if and only if u is stable.
- (iii) If 0 is a prefix of w , then $\mathcal{S}(vw) = \mathcal{S}(v0)\mathcal{S}(w)$.
- (iv) Let 0 be a prefix of v and v' . Then $\mathcal{S}(v')$ is a prefix of $\mathcal{S}(v)$ if and only if v' is a prefix of v .

Proof:

- (i) Let $v = v_0v_1 \cdots v_{n-1}$, $n = |v|$, and $v_0 = 0$. Put $v_n = 0$. Then $\mathcal{S}(v0) = u_0u_1 \cdots u_{n-1}$, where $u_i = v_i + v_{i+1} \pmod 2$ for every $i \in \{0, 1, \dots, n-1\}$. It implies

$$|\mathcal{S}(v0)|_1 = \sum_{i=0}^{n-1} u_i = v_0 + v_n = 0 \pmod 2.$$

- (ii) Let $u = u_0u_1 \cdots u_{m-1}$ and $m = |u|$. We look for $w = w_0w_1 \cdots w_m$ such that $u_i = w_i + w_{i+1} \pmod 2$ for every $i \in \{0, 1, \dots, m-1\}$. Clearly, these equations can be equivalently rewritten as

$$u_i = w_{i+1} - w_i \pmod 2 \quad \text{for every } i \in \{0, 1, \dots, m-1\}. \quad (4)$$

Then starting with $w_0 = 0$ and summing up the equations (4) for $i = \{0, 1, \dots, j-1\}$, we determine the letter w_j of w as $w_j = \sum_{i=0}^{j-1} u_i \pmod 2$. In particular, $w_m = |u|_1 \pmod 2$.

- (iii) It is a particular case of the equation (3).
- (iv) It follows directly from the definition of \mathcal{S} .

□

4 The relation between the indices of factors in \mathbf{v} and $\mathcal{S}(\mathbf{v})$

In this section, we provide a tool for determining the critical exponent of a binary sequence \mathbf{v} whose language is closed under the exchange of letters. For any factor v of such a sequence, $\text{ind}_{\mathbf{v}}(v) = \text{ind}_{\mathbf{v}}(E(v))$ and we can consider only factors of \mathbf{v} starting with 0 without loss of generality.

Lemma 12. *Let \mathbf{v} be a binary aperiodic uniformly recurrent sequence whose language is closed under E . Denote $\mathbf{u} = \mathcal{S}(\mathbf{v})$. For a non-empty factor $v \in \mathcal{L}(\mathbf{v})$ with the prefix 0 and $\text{ind}_{\mathbf{v}}(v) > 1$, there exists a stable factor $u \in \mathcal{L}(\mathbf{u})$ such that*

$$\text{ind}_{\mathbf{u}}(u) + \frac{1}{|u|} = \text{ind}_{\mathbf{v}}(v) \quad \text{and} \quad u = \mathcal{S}(v0). \quad (5)$$

And vice versa, for a non-empty stable factor $u \in \mathcal{L}(\mathbf{u})$, there exists a factor $v \in \mathcal{L}(\mathbf{v})$ with the prefix 0 satisfying (5).

Proof: For a given $n \in \mathbb{N}$, $n \geq 1$, consider the set \mathcal{K}_n of factors $v \in \mathcal{L}(\mathbf{v})$ of length n with the prefix 0 and $\text{ind}_{\mathbf{v}}(v) > 1$. First, we show that the mapping $v \mapsto \mathcal{S}(v0)$ is a bijection between \mathcal{K}_n and the set of all stable factors of \mathbf{u} of length n .

Indeed, if $v \in \mathcal{K}_n$, then $v0 \in \mathcal{L}(\mathbf{v})$. The factor $u := \mathcal{S}(v0)$ belongs to $\mathcal{L}(\mathbf{u})$, $|u| = |v|$, and by Item (i) of Lemma 11, u is stable. On the other hand, if $u \in \mathcal{L}(\mathbf{u})$ is stable and of length n , then by Item (ii) of Lemma 11, there exists a unique w such that 0 is a prefix and a suffix of w and $\mathcal{S}(w) = u$. As $\mathcal{L}(\mathbf{v})$ is closed under E , necessarily $w \in \mathcal{L}(\mathbf{v})$ and $w = v0$ for some v with the prefix 0. In particular, $\text{ind}_{\mathbf{v}}(v) > 1$. As $u = \mathcal{S}(w) = \mathcal{S}(v0)$, the lengths of u and v coincide.

Now we show that any $v \in \mathcal{K}_n$ and its image $u = \mathcal{S}(v0)$ satisfy (5). Find $k \in \mathbb{N}$, $k \geq 1$, and $\theta \in (0, 1]$ such that $\text{ind}_{\mathbf{v}}(v) = k + \theta$. Denote $v' = v^\theta$. Obviously, $v' \neq \varepsilon$, v' is a prefix of v and 0 is a prefix of v' . Applying Item (iii) of Lemma 11, we get $\mathcal{S}(v^k v') = (\mathcal{S}(v0))^k \mathcal{S}(v')$. Clearly, $|\mathcal{S}(v')| = |v'| - 1$, $|u| = |v|$, and by Item (iv) of Lemma 11, $\mathcal{S}(v')$ is a prefix of $\mathcal{S}(v0)$. For $u = \mathcal{S}(v0)$ it means that

$$\text{ind}_{\mathbf{u}}(u) \geq k + \frac{|\mathcal{S}(v')|}{|u|} = k + \frac{|v'|}{|v|} - \frac{1}{|u|} = k + \theta - \frac{1}{|u|} = \text{ind}_{\mathbf{v}}(v) - \frac{1}{|u|}.$$

To show the opposite inequality, we find $\ell \in \mathbb{N}$ and $\eta \in [0, 1)$ such that $\text{ind}_{\mathbf{u}}(u) = \ell + \eta$. Denote $u' = u^\eta$. Using Item (ii) of Lemma 11, we find v' with the prefix 0 such that $u' = \mathcal{S}(v')$. By Item (iv), v' is a prefix of v , and by Item (iii), $u^\ell u' = (\mathcal{S}(v0))^\ell \mathcal{S}(v') = \mathcal{S}(v^\ell v')$. Therefore, $v^\ell v' \in \mathcal{L}(\mathbf{v})$ and

$$\text{ind}_{\mathbf{v}}(v) \geq \ell + \eta + \frac{1}{|u|} = \text{ind}_{\mathbf{u}}(u) + \frac{1}{|u|}.$$

□

As explained in Lemma 3, only return words to bispecial factors play a role for the determination of the critical exponent of a sequence. More specifically, we can restrict ourselves to factors from the set \mathcal{M}' (or \mathcal{M}'') introduced in Remark 4.

Lemma 13. *Let \mathbf{v} be a binary sequence whose language is closed under E . Assume that v with the prefix 0 is a return word in \mathbf{v} to a bispecial factor $b = v^{e-1}$, where $e > 2$. Denote $u = \mathcal{S}(v0)$. Then*

- either u is a stable return word in $\mathcal{S}(\mathbf{v})$ to a bispecial factor with the fractional root u ;
- or $u = x^2$, where x is an unstable return word in $\mathcal{S}(\mathbf{v})$ to a bispecial factor with the fractional root u .

Proof: The factor $\mathcal{S}(b)$ is bispecial in $\mathcal{S}(\mathbf{v})$ by Lemma 7. Moreover, by the rule (3), we can write $\mathcal{S}(b) = \mathcal{S}(v^{e-1}) = \mathcal{S}(v0)^f$ for $f = e - 1 - \frac{1}{|v|} \geq 1$. Thus $\mathcal{S}(b)$ has the fractional root $u = \mathcal{S}(v0)$.

The word vb is a complete return word to b in \mathbf{v} and thus $vb = bw$ for some w . Note that 0 is the first letter of b . We denote the last letter of b by z . By the rule (3), we get $\mathcal{S}(v0)\mathcal{S}(b) = \mathcal{S}(b)\mathcal{S}(zw)$. It means that $\mathcal{S}(b)$ is a prefix and a suffix of the word $\mathcal{S}(v0)\mathcal{S}(b)$. We discuss two cases:

– $\mathcal{S}(b)$ has exactly two occurrences in $\mathcal{S}(v0)\mathcal{S}(b)$, one as a prefix and one as a suffix. In this case $u = \mathcal{S}(v0)$ is a return word to $\mathcal{S}(b)$ in $\mathcal{S}(\mathbf{v})$ and by Item (i) of Lemma 11, u is stable.

– $\mathcal{S}(b)$ occurs in $\mathcal{S}(v0)\mathcal{S}(b)$ as an inner factor. In this case, there exists a return word $u' \neq \varepsilon$ to $\mathcal{S}(b)$ such that $|u'| < |u|$ and $u'\mathcal{S}(b)$ is a proper prefix of $\mathcal{S}(v0)\mathcal{S}(b)$. We take the word b' such that $b = vb'$. Clearly, b' has the prefix 0 and so $\mathcal{S}(b) = u\mathcal{S}(b')$. Then $u'\mathcal{S}(b) = u'u\mathcal{S}(b')$ is a proper prefix of $\mathcal{S}(v0)\mathcal{S}(b) = u\mathcal{S}(b) = uu\mathcal{S}(b')$. In other words, $u'uu'' = uu$ for some non-empty u'' and consequently, $u = u'u'' = u''u'$. This implies the existence of $x \in \{0, 1\}^+$ and $k', k'' \in \mathbb{N}$, $k', k'' \geq 1$ such that $u' = x^{k'}$ and $u'' = x^{k''}$. If we denote $k = k' + k'' \geq 2$, we can write $u = x^k$.

We show that x is unstable and $k = 2$. Indeed, assume x is stable, then by Item (ii) of Lemma 11, we find a unique y with the prefix 0 such that $x = \mathcal{S}(y0)$. Applying Item (iii), we obtain $\mathcal{S}(v0) = u = x^k = (\mathcal{S}(y0))^k = \mathcal{S}(y^k0)$ and thus $v = y^k$. Nevertheless, the factor v is primitive as explained in Remark 4. Thus this is a contradiction.

Since x is unstable and $u = x^k$ is stable, necessarily $k = 2p$ for some integer $p \geq 1$. Now we deduce that $k = 2$. Indeed, if $p \geq 2$, then u is a p -power of the stable factor x^2 which yields a contradiction with the primitivity of v as above. Finally, $k = 2$ implies $k' = 1$ and $u' = x$ is an unstable return word to the bispecial factor $\mathcal{S}(b)$ in $\mathcal{S}(\mathbf{v})$. \square

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

$$A_1 = \left\{ \text{ind}_{\mathbf{u}}(u) + \frac{1}{|u|} : u \text{ is a stable return word to a bispecial factor of } \mathbf{u} \right\} \text{ and}$$

$$A_2 = \left\{ \frac{1}{2}(\text{ind}_{\mathbf{u}}(u) + \frac{1}{|u|}) : u \text{ is an unstable return word to a bispecial factor of } \mathbf{u} \right\}.$$

Then

$$\text{cr}(\mathbf{v}) = \sup(A_1 \cup A_2).$$

Proof: First we show that

$$\text{ind}_{\mathbf{v}}(v) \leq \sup(A_1 \cup A_2) \quad \text{for any non-empty } v \in \mathcal{L}(\mathbf{v}). \quad (6)$$

If $\text{ind}_{\mathbf{v}}(v) \leq 2$, then the inequality (6) is trivially satisfied as A_1 contains the number $\text{ind}_{\mathbf{u}}(0) + \frac{1}{|0|} \geq 2$ (note that 0 is a stable return word in \mathbf{u} to the bispecial factor ε). Now we assume that $\text{ind}_{\mathbf{v}}(v) = e > 2$. By Lemma 3 and Remark 4, we may focus only on v which is a return word to the bispecial factor $b = v^{e-1}$ and v has the prefix 0. By Lemma 12, $\text{ind}_{\mathbf{v}}(v) = \text{ind}_{\mathbf{u}}(u) + \frac{1}{|u|}$, where $u = \mathcal{S}(v0)$. By Lemma 13, the factor u is either a stable return word to a bispecial factor in \mathbf{u} , or $u = x^2$, where x is an unstable return word to a bispecial factor in \mathbf{u} . In the first case we have $\text{ind}_{\mathbf{v}}(v) \leq \sup A_1$, while in the second case we have $\text{ind}_{\mathbf{v}}(v) = \text{ind}_{\mathbf{u}}(u) + \frac{1}{|u|} = \frac{1}{2}\text{ind}_{\mathbf{u}}(x) + \frac{1}{2|x|} \leq \sup A_2$. We may conclude that $\text{cr}(\mathbf{v}) \leq \sup(A_1 \cup A_2)$.

To prove the opposite inequality, we show

$$A_1 \cup A_2 \setminus [0, 1] \subset \{\text{ind}_{\mathbf{v}}(v) : v \in \mathcal{L}(\mathbf{v}), v \neq \varepsilon\}.$$

If $H \in A_1$, then there exists a stable factor u in \mathbf{u} such that $H = \text{ind}_{\mathbf{u}}(u) + \frac{1}{|u|}$, and by Lemma 12, we find v in \mathbf{v} such that $H = \text{ind}_{\mathbf{v}}(v)$. Analogously, if $H \in A_2$ and $H > 1$, then $\text{ind}_{\mathbf{u}}(u) = 2H - \frac{1}{|u|} \geq 2$ for some unstable factor $u \in \mathcal{L}(\mathbf{u})$. Thus the word $y = uu \in \mathcal{L}(\mathbf{u})$, it is a stable factor of \mathbf{u} and its index in \mathbf{u} is $\frac{1}{2}\text{ind}_{\mathbf{u}}(u)$. By Lemma 12, there exists v in \mathbf{v} such that $\text{ind}_{\mathbf{v}}(v) = \text{ind}_{\mathbf{u}}(y) + \frac{1}{|y|} = \frac{1}{2}\text{ind}_{\mathbf{u}}(u) + \frac{1}{2|u|} = H$. \square

Theorem 14 will be used in the next sections to determine the critical exponent of a complementary symmetric Rote sequence \mathbf{v} by exploiting the indices of factors in the Sturmian sequence $\mathcal{S}(\mathbf{v})$. The following example shows an opposite application of Theorem 14. But before that, let us state a simple auxiliary statement reflecting the behaviour of fractional roots under the application of a morphism.

Observation 15. Let $\phi : \mathcal{A}^* \mapsto \mathcal{A}^*$ be a morphism and let $w \in \mathcal{A}^*$ be a prefix of $\phi(a)$ for each $a \in \mathcal{A}$. If u is a fractional root of z , then $\phi(u)$ is a fractional root of $\phi(z)w$.

Example 16. Let us consider the Thue–Morse sequence $\mathbf{t} = 01101001 \dots$, which is fixed by the morphism $\psi : 0 \mapsto 01$ and $1 \mapsto 10$. As ψ is primitive, the sequence \mathbf{t} is uniformly recurrent. It is well-known that its language is closed under the exchange of letters and $\text{cr}(\mathbf{t}) = 2$. The corresponding sequence $\mathbf{u} = \mathcal{S}(\mathbf{t}) = 1011101 \dots$ is called the period doubling sequence and it is fixed by the morphism $\phi : 0 \mapsto 11$ and $1 \mapsto 10$, see [23].

We determine the critical exponent of \mathbf{u} . Theorem 14 implies $\text{cr}(\mathbf{u}) \leq 4$, as otherwise $\text{cr}(\mathbf{t}) > 2$, which is a contradiction. Now we show that the value 4 is attained.

By Observation 15 and the fact that both $\phi(0)$ and $\phi(1)$ have the prefix 1, the morphism ϕ has the following two properties:

1. If $w \in \mathcal{L}(\mathbf{u})$, then $\phi(w)1 \in \mathcal{L}(\mathbf{u})$.
2. If u is a fractional root of w , then $\phi(u)$ is a fractional root of $\phi(w)1$.

We will construct two sequences $(u^{(n)})$ and $(w^{(n)})$ of words belonging to $\mathcal{L}(\mathbf{u})$. We start with $u^{(0)} = 1$ and $w^{(0)} = 111 \in \mathcal{L}(\mathbf{u})$ and for each $n \in \mathbb{N}$ we define

$$w^{(n+1)} = \phi(w^{(n)})1 \quad \text{and} \quad u^{(n+1)} = \phi(u^{(n)}).$$

Note that $u^{(0)} = 1$ is a fractional root of $w^{(0)} = 111$. Because of the property (2), the word $u^{(n)}$ is a fractional root of $w^{(n)}$ for each $n \in \mathbb{N}$. Moreover, the specific form of the morphism ϕ implies $|u^{(n+1)}| = 2|u^{(n)}|$ and $|w^{(n+1)}| = 2|w^{(n)}| + 1$. It gives $|u^{(n)}| = 2^n$ and $|w^{(n)}| = 2^{n+2} - 1$. Therefore, $\text{ind}_{\mathbf{u}}(u^{(n)}) \geq \frac{2^{n+2}-1}{2^n} \rightarrow 4$. We may conclude that $\text{cr}(\mathbf{u}) = 4$.

5 Return words to bispecial factors of Sturmian sequences

The main goal of this article is to describe the critical exponent and the recurrence function of CS Rote sequences. Proposition 8 and Theorem 14 transform the first task to the computation of the indices of return words to bispecial factors in the associated Sturmian sequences.

This is a preparatory section for this computation. We introduce the directive sequence of a standard Sturmian sequence and recall some known results on bispecial factors, their return words, and derived sequences. It allows us to describe the longest factor of \mathbf{u} with the fractional root u , where u is any return word to a bispecial factor of a Sturmian sequence \mathbf{u} (Lemma 24). Furthermore, we explain how to express the lengths of these factors explicitly (Proposition 30), and eventually in Section 6, we determine the indices.

First, we recall that a binary sequence $\mathbf{u} \in \{0, 1\}^{\mathbb{N}}$ is Sturmian if it has the factor complexity $\mathcal{C}_{\mathbf{u}}(n) = n + 1$ for all $n \in \mathbb{N}$. If both sequences $0\mathbf{u}$ and $1\mathbf{u}$ are Sturmian, then \mathbf{u} is called a *standard Sturmian sequence*. It is well-known that for any Sturmian sequence there exists a unique standard Sturmian sequence with the same language. Since all properties which we are interested in (indices of factors, critical exponent, special factors, return words, recurrence function) depend only on the language of the sequence, we restrict ourselves to standard Sturmian sequences without loss of generality.

In the sequel, we use the characterization of standard Sturmian sequences by their directive sequences. To introduce them, we define two morphisms

$$G = \begin{cases} 0 \rightarrow 10 \\ 1 \rightarrow 1 \end{cases} \quad \text{and} \quad D = \begin{cases} 0 \rightarrow 0 \\ 1 \rightarrow 01 \end{cases}$$

with the corresponding matrices

$$M_G = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad M_D = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Let us note that $G = E \circ F$ and $D = F \circ E$, where E is the morphism which exchanges letters, i.e., $E : 0 \rightarrow 1, 1 \rightarrow 0$, and F is the Fibonacci morphism, i.e., $F : 0 \rightarrow 01, 1 \rightarrow 0$.

Proposition 17 ([15]). *For every standard Sturmian sequence \mathbf{u} there is a uniquely given sequence $\Delta = \Delta_0 \Delta_1 \Delta_2 \cdots \in \{G, D\}^{\mathbb{N}}$ of morphisms and a sequence $(\mathbf{u}^{(n)})_{n \geq 0}$ of standard Sturmian sequences such that*

$$\mathbf{u} = \Delta_0 \Delta_1 \cdots \Delta_{n-1}(\mathbf{u}^{(n)}) \text{ for every } n \in \mathbb{N}.$$

Moreover, the sequence Δ contains infinitely many letters G and infinitely many letters D , i.e.,

$$\Delta = G^{a_1} D^{a_2} G^{a_3} D^{a_4} \cdots \text{ or } \Delta = D^{a_1} G^{a_2} D^{a_3} G^{a_4} \cdots \text{ for some sequence } (a_i)_{i \geq 1} \text{ of positive integers.}$$

The sequence Δ is called the directive sequence of \mathbf{u} .

Remark 18. Let us note that \mathbf{u} has the directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \cdots$ if and only if $E(\mathbf{u})$ has the directive sequence $D^{a_1} G^{a_2} D^{a_3} G^{a_4} \cdots$. Obviously, both sequences \mathbf{u} and $E(\mathbf{u})$ have the same structure up to the exchange of letters $0 \leftrightarrow 1$. In particular, any Sturmian sequence with the directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \cdots$ can be written as a concatenation of the blocks $1^{a_1} 0$ and $1^{a_1+1} 0$, while any Sturmian sequence with the directive sequence $D^{a_1} G^{a_2} D^{a_3} G^{a_4} \cdots$ can be written as a concatenation of the blocks $0^{a_1} 1$ and $0^{a_1+1} 1$.

By Vuillon's result [26], every factor of any Sturmian sequence has exactly two return words. Thus for a given bispecial factor b of \mathbf{u} , we usually denote the more and the less frequent return word to b in \mathbf{u} by r and s . In this notation, the sequence \mathbf{u} can be decomposed into the blocks $r^k s$ and $r^{k+1} s$ for some $k \in \mathbb{N}, k \geq 1$.

We need to know how bispecial factors and their return words change under the application of morphisms G and D . The following description can be found in [18], where several partial statements from [16] are accumulated.

Lemma 19. *Let \mathbf{u}' , \mathbf{u} be standard Sturmian sequences such that $\mathbf{u} = G(\mathbf{u}')$.*

- (i) *For every bispecial factor b' of \mathbf{u}' , the factor $b = G(b')1$ is a bispecial factor of \mathbf{u} .*
- (ii) *Every bispecial factor b of \mathbf{u} which is not empty can be written as $b = G(b')1$ for a uniquely given bispecial factor $b' \in \mathcal{L}(\mathbf{u}')$.*
- (iii) *The words r', s' are return words to a bispecial prefix b' of \mathbf{u}' if and only if $r = G(r'), s = G(s')$ are return words to a bispecial prefix $b = G(b')1$ of \mathbf{u} . Moreover, the derived sequences satisfy $\mathbf{d}_{\mathbf{u}}(b) = \mathbf{d}_{\mathbf{u}'}(b')$.*

Lemma 20. *Let \mathbf{u}' , \mathbf{u} be standard Sturmian sequences such that $\mathbf{u} = D(\mathbf{u}')$.*

- (i) *For every bispecial factor b' of \mathbf{u}' , the factor $b = D(b')0$ is a bispecial factor of \mathbf{u} .*
- (ii) *Every bispecial factor b of \mathbf{u} which is not empty can be written as $b = D(b')0$ for a uniquely given bispecial factor $b' \in \mathcal{L}(\mathbf{u}')$.*
- (iii) *The words r' , s' are return words to a bispecial prefix b' of \mathbf{u}' if and only if $r = D(r')$, $s = D(s')$ are return words to a bispecial prefix $b = D(b')0$ of \mathbf{u} . Moreover, the derived sequences satisfy $\mathbf{d}_{\mathbf{u}}(b) = \mathbf{d}_{\mathbf{u}'}(b')$.*

Any prefix of a standard Sturmian sequence is a left special factor. Moreover, a factor of a standard Sturmian sequence \mathbf{u} is bispecial if and only if it is a palindromic prefix of \mathbf{u} . Therefore, we can order the bispecial factors of a given standard Sturmian sequence \mathbf{u} by their lengths: we start with the empty word ε , which is the 0^{th} bispecial factor, then the first letter of \mathbf{u} is the 1^{st} bispecial factor of \mathbf{u} etc.

Remark 21. If \mathbf{u} has the directive sequence $\Delta_0\Delta_1\Delta_2\cdots \in \{G, D\}^{\mathbb{N}}$, the derived sequence $\mathbf{d}_{\mathbf{u}}(b)$ to the n^{th} bispecial factor b of \mathbf{u} has the directive sequence $\Delta_n\Delta_{n+1}\Delta_{n+2}\cdots$. Indeed, we denote the sequence with the directive sequence $\Delta_n\Delta_{n+1}\Delta_{n+2}\cdots$ by \mathbf{u}' . It has the bispecial factor ε and by the definition $\mathbf{d}_{\mathbf{u}'}(\varepsilon) = \mathbf{u}'$. If we apply n times Lemmas 19 or 20, we get $\mathbf{d}_{\mathbf{u}}(b) = \mathbf{d}_{\mathbf{u}'}(\varepsilon) = \mathbf{u}'$.

Let us formulate a direct consequence of the relation (1) and Lemmas 19 and 20.

Corollary 22. *Let $k, h \in \mathbb{N}$. Let b' be the k^{th} bispecial factor of a standard Sturmian sequence \mathbf{u}' and u' be a return word to b' in \mathbf{u}' . Let $\Delta = \Delta_0\Delta_1\Delta_2\cdots \in \{G, D\}^{\mathbb{N}}$ be the directive sequence of \mathbf{u}' .*

1. *If $\mathbf{u} = G^h(\mathbf{u}')$, then the $(k+h)^{\text{th}}$ bispecial factor b of \mathbf{u} and a return word u to b satisfy*

$$\vec{V}(b) = \begin{pmatrix} 1 & 0 \\ h & 1 \end{pmatrix} \vec{V}(b') + h \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \vec{V}(u) = \begin{pmatrix} 1 & 0 \\ h & 1 \end{pmatrix} \vec{V}(u').$$

The directive sequence of \mathbf{u} is $G^h\Delta_0\Delta_1\Delta_2\cdots$.

2. *If $\mathbf{u} = D^h(\mathbf{u}')$, then the $(k+h)^{\text{th}}$ bispecial factor b of \mathbf{u} and a return word u to b satisfy*

$$\vec{V}(b) = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \vec{V}(b') + h \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \vec{V}(u) = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \vec{V}(u').$$

The directive sequence of \mathbf{u} is $D^h\Delta_0\Delta_1\Delta_2\cdots$.

As we have seen in Remark 4, when determining the critical exponent it suffices to take into account only bispecial factors whose fractional roots are equal to its return words. Lemma 24 says that all bispecial factors of a Sturmian sequence are of this type, and moreover, it enables one to determine the indices of their return words. The first auxiliary statement is a slightly strengthened variant of Observation 15 for the morphisms G and D .

Observation 23. *Let \mathbf{u} be a binary sequence and let $u \in \mathcal{L}(\mathbf{u})$. If z is the longest factor in $\mathcal{L}(\mathbf{u})$ with the fractional root u , then $G(z)1$ is the longest factor in $\mathcal{L}(G(\mathbf{u}))$ with the fractional root $G(u)$ and, analogously, $D(z)0$ is the longest factor in $\mathcal{L}(D(\mathbf{u}))$ with the fractional root $D(u)$.*

Lemma 24. *Let b be a bispecial factor of a standard Sturmian sequence \mathbf{u} . Let r and s be the return words to b in \mathbf{u} and let $k \in \mathbb{N}$, $k \geq 1$, be such that \mathbf{u} is concatenated from the blocks $r^k s$ and $r^{k+1} s$. Then $r^{k+1} b$ is the longest factor of \mathbf{u} with the fractional root r and sb is the longest factor of \mathbf{u} with the fractional root s .*

Proof: We proceed by induction on the length of b . Without loss of generality, we assume that \mathbf{u} has the directive sequence $\Delta = G^{a_1} D^{a_2} G^{a_3} D^{a_4} \dots$.

The bispecial factor $b = \varepsilon$ has the return words $r = 1, s = 0$, and by Remark 18, \mathbf{u} is concatenated from the blocks $1^{a_1} 0 = r^{a_1} s$ and $1^{a_1+1} 0 = r^{a_1+1} s$. Clearly, $r^{a_1+1} b = 1^{a_1+1}$ is the longest factor of \mathbf{u} with the fractional root 1. Similarly, $sb = 0$ is the longest factor of \mathbf{u} with the fractional root 0.

Let b be a bispecial factor of \mathbf{u} with $|b| \geq 1$ and let \mathbf{u} be concatenated from the blocks $r^k s$ and $r^{k+1} s$ for the return words r, s to b in \mathbf{u} and some $k \in \mathbb{N}$, $k \geq 1$. By Proposition 17, there is a unique standard Sturmian sequence \mathbf{u}' such that $\mathbf{u} = G(\mathbf{u}')$. By Lemmas 19 and 20, there is a unique bispecial factor b' of \mathbf{u}' with the return words r' and s' such that $b = G(b')1$, $r = G(r')$, and $s = G(s')$. Moreover, \mathbf{u}' is concatenated from the blocks $(r')^k s'$ and $(r')^{k+1} s'$. Clearly, $|b'| < |b|$ and so by the induction hypothesis, the words $(r')^{k+1} b'$ and $s' b'$ are the longest factors of \mathbf{u}' with the fractional root r' and s' , respectively. But then by Observation 23, the words $r^{k+1} b = G((r')^{k+1} b')1$ and $sb = G(s' b')1$ are the longest factors of \mathbf{u} with the fractional root $r = G(r')$ and $s = G(s')$, respectively. \square

Having in mind our goal to describe the critical exponent of any CS Rote sequence and Theorem 14, we need to determine the indices of return words to bispecial factors in standard Sturmian sequences, i.e., the lengths of factors from Lemma 24. We also want to distinguish, which of these return words are (un)stable. Both of these tasks can be solved using the Parikh vectors of the relevant bispecial factors and their return words. We deduce the explicit formulae for the needed Parikh vectors in Proposition 30. For this purpose, we adopt the following notation.

Notation 25. To a standard Sturmian sequence \mathbf{u} with the directive sequence $\Delta = G^{a_1} D^{a_2} G^{a_3} D^{a_4} \dots$ or $\Delta = D^{a_1} G^{a_2} D^{a_3} G^{a_4} \dots$ we assign an irrational number $\theta \in (0, 1)$ with the continued fraction expansion

$$\theta = [0, a_1, a_2, a_3, \dots].$$

For every $N \in \mathbb{N}$, we denote the N^{th} convergent to the number θ by $\frac{p_N}{q_N}$ and the N^{th} convergent to the number $\frac{\theta}{1+\theta}$ by $\frac{p'_N}{q'_N}$.

Remark 26. Let us recall some basic properties of convergents. They can be found in any number theory textbook, e.g., [14].

1. The sequences (p_N) , (q_N) , and (q'_N) fulfil the same recurrence relation for all $N \in \mathbb{N}$, $N \geq 1$, namely

$$X_N = a_N X_{N-1} + X_{N-2},$$

but they differ in their initial values: $p_{-1} = 1, p_0 = 0; q_{-1} = 0, q_0 = 1; q'_{-1} = q'_0 = 1$. It implies for all $N \in \mathbb{N}$

$$p_N + q_N = q'_N.$$

2. For all $N \in \mathbb{N}$, $N \geq 1$, we have

$$\begin{aligned} \begin{pmatrix} 1 & 0 \\ a_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & a_2 \\ 0 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ a_{2N-1} & 1 \end{pmatrix} \begin{pmatrix} 1 & a_{2N} \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} p_{2N-1} & p_{2N} \\ q_{2N-1} & q_{2N} \end{pmatrix}; \\ \begin{pmatrix} 1 & 0 \\ a_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & a_2 \\ 0 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & a_{2N-2} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ a_{2N-1} & 1 \end{pmatrix} &= \begin{pmatrix} p_{2N-1} & p_{2N-2} \\ q_{2N-1} & q_{2N-2} \end{pmatrix}. \end{aligned}$$

Remark 27. For the description of a standard Sturmian sequence \mathbf{u} we use the number θ . Usually, a standard Sturmian sequence is characterized by the so-called *slope*, which is equal to the density of the letter 1 in the sequence \mathbf{u} . In our notation, the slope of \mathbf{u} is $\frac{\theta}{1+\theta} = [0, 1 + a_1, a_2, a_3, \dots]$ if the directive sequence Δ starts with D , otherwise the slope is $\frac{1}{1+\theta} = [0, 1, a_1, a_2, a_3, \dots]$.

In the sequel, we will need two auxiliary statements on convergents $\frac{p_N}{q_N}$ to θ .

Lemma 28. For all $N \in \mathbb{N}$ we have

$$\begin{pmatrix} p_N \\ q_N \end{pmatrix} \not\equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix} \pmod{2} \quad \text{and} \quad \begin{pmatrix} p_N \\ q_N \end{pmatrix} \not\equiv \begin{pmatrix} p_{N-1} \\ q_{N-1} \end{pmatrix} \pmod{2}.$$

Proof: The first statement is a consequence of the fact that p_N and q_N are coprime. We show the second statement by contradiction. Assume that there exists $K \in \mathbb{N}$ such that $\begin{pmatrix} p_K \\ q_K \end{pmatrix} \equiv \begin{pmatrix} p_{K-1} \\ q_{K-1} \end{pmatrix} \pmod{2}$. Let K denote the smallest integer with this property. As $q_{-1} = 0$ and $q_0 = 1$, necessarily, $K > 0$. Using the recurrence relation satisfied by the convergents, we can write

$$\begin{pmatrix} p_K \\ q_K \end{pmatrix} = a_K \begin{pmatrix} p_{K-1} \\ q_{K-1} \end{pmatrix} + \begin{pmatrix} p_{K-2} \\ q_{K-2} \end{pmatrix} \equiv \begin{pmatrix} p_{K-1} \\ q_{K-1} \end{pmatrix} \pmod{2}.$$

If a_K is even, then the previous equation gives $\begin{pmatrix} p_{K-2} \\ q_{K-2} \end{pmatrix} \equiv \begin{pmatrix} p_{K-1} \\ q_{K-1} \end{pmatrix} \pmod{2}$, which is a contradiction with the minimality of K .

If a_K is odd, then the previous equation gives $\begin{pmatrix} p_{K-2} \\ q_{K-2} \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix} \pmod{2}$, which is a contradiction with the first statement. \square

Lemma 29. For all $N \in \mathbb{N}$, $N \geq 1$, we have

$$a_N \begin{pmatrix} p_{N-1} \\ q_{N-1} \end{pmatrix} + a_{N-1} \begin{pmatrix} p_{N-2} \\ q_{N-2} \end{pmatrix} + \cdots + a_2 \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} + a_1 \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} = \begin{pmatrix} p_N \\ q_N \end{pmatrix} + \begin{pmatrix} p_{N-1} \\ q_{N-1} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Proof: It can be easily proved by induction on N . \square

The Parikh vectors of the bispecial factors of \mathbf{u} and the corresponding return words can be easily expressed using the convergents $\frac{p_N}{q_N}$ to θ . We will use these expressions essentially in the next sections.

Proposition 30. *Let b be the n^{th} bispecial factor of \mathbf{u} with the directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \dots$. We denote the more and the less frequent return word to b in \mathbf{u} by r and s , respectively. Put $a_0 = 0$ and write n in the form $n = m + a_0 + a_1 + a_2 + \dots + a_N$ for a unique $N \in \mathbb{N}$ and $0 \leq m < a_{N+1}$. Then*

1. $\vec{V}(r) = \begin{pmatrix} p_N \\ q_N \end{pmatrix};$
2. $\vec{V}(s) = \begin{pmatrix} m p_N + p_{N-1} \\ m q_N + q_{N-1} \end{pmatrix};$
3. $\vec{V}(b) = (m+1) \begin{pmatrix} p_N \\ q_N \end{pmatrix} + \begin{pmatrix} p_{N-1} \\ q_{N-1} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$

Proof: First we suppose that N is even and we denote the standard Sturmian sequence with the directive sequence $G^{a_{N+1}} D^{a_{N+2}} G^{a_{N+3}} \dots$ by \mathbf{u}' . By Remark 18, \mathbf{u}' is concatenated from the blocks $1^{a_{N+1}}0$ and $1^{a_{N+1}+1}0$. Thus its m^{th} bispecial factor is $b' = 1^m$ and the return words to b' in \mathbf{u}' are $r' = 1$ and $s' = 1^m0$.

By Lemmas 19 and 20 and Remark 21,

$$r = G^{a_1} D^{a_2} \dots G^{a_{N-1}} D^{a_N}(r') \quad \text{and} \quad s = G^{a_1} D^{a_2} \dots G^{a_{N-1}} D^{a_N}(s').$$

By Corollary 22 and Lemma 29, the Parikh vectors of r and s satisfy

$$\begin{aligned} \vec{V}(r) &= \begin{pmatrix} 1 & 0 \\ a_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & a_2 \\ 0 & 1 \end{pmatrix} \dots \begin{pmatrix} 1 & 0 \\ a_{N-1} & 1 \end{pmatrix} \begin{pmatrix} 1 & a_N \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} p_N \\ q_N \end{pmatrix}; \\ \vec{V}(s) &= \begin{pmatrix} 1 & 0 \\ a_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & a_2 \\ 0 & 1 \end{pmatrix} \dots \begin{pmatrix} 1 & 0 \\ a_{N-1} & 1 \end{pmatrix} \begin{pmatrix} 1 & a_N \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ m \end{pmatrix} = \begin{pmatrix} m p_N + p_{N-1} \\ m q_N + q_{N-1} \end{pmatrix}. \end{aligned}$$

To find the Parikh vector of b we start with the bispecial factor $b' = 1^m$ and N times apply Corollary 22. Eventually, we rewrite the arising products of matrices by Lemma 29 and we get

$$\vec{V}(b) = m \begin{pmatrix} p_N \\ q_N \end{pmatrix} + a_N \begin{pmatrix} p_{N-1} \\ q_{N-1} \end{pmatrix} + a_{N-1} \begin{pmatrix} p_{N-2} \\ q_{N-2} \end{pmatrix} + \dots + a_2 \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} + a_1 \begin{pmatrix} p_0 \\ q_0 \end{pmatrix}.$$

This together with Lemma 29 implies the statement of Item 3 for N even. The proof for N odd is analogous. \square

Remark 31. If we assume in Proposition 30 that \mathbf{u} has the directive sequence $\mathbf{\Delta} = D^{a_1} G^{a_2} D^{a_3} G^{a_4} \dots$, then by Remark 18, the coordinates of the Parikh vectors will be exchanged, i.e.,

$$\vec{V}(r) = \begin{pmatrix} q_N \\ p_N \end{pmatrix}, \quad \vec{V}(s) = \begin{pmatrix} m q_N + q_{N-1} \\ m p_N + p_{N-1} \end{pmatrix}, \quad \text{and} \quad \vec{V}(b) = (m+1) \begin{pmatrix} q_N \\ p_N \end{pmatrix} + \begin{pmatrix} q_{N-1} \\ p_{N-1} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

6 The critical exponent of CS Rote sequences

We are going to give an explicit formula for the critical exponent of a CS Rote sequence \mathbf{v} . We will use Theorem 14 which requires the knowledge of the indices of return words to bispecial factors in the Sturmian sequence $\mathcal{S}(\mathbf{v})$. It is well-known that there is a unique standard Sturmian sequence \mathbf{u} such that both $\mathcal{S}(\mathbf{v})$ and \mathbf{u} have the same language. Since the critical exponent depends only on the language, we can work with the standard Sturmian sequence \mathbf{u} instead of $\mathcal{S}(\mathbf{v})$.

In the following proposition and theorem, we use Notation 25.

Proposition 32. *Let \mathbf{u} be a standard Sturmian sequence with the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\dots$ or $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\dots$ and let $n \in \mathbb{N}$ be given. We put $a_0 = 0$ and we denote the more and the less frequent return word to the n^{th} bispecial factor of \mathbf{u} by r and s , respectively.*

1. *If $n = m + a_0 + a_1 + a_2 + \dots + a_N$, where $0 \leq m < a_{N+1}$, then $\text{ind}(r) = a_{N+1} + 2 + \frac{q'_{N-1}-2}{q'_N}$.*
2. *If $n = a_0 + a_1 + a_2 + \dots + a_N$, then $\text{ind}(s) = a_N + 2 + \frac{q'_{N-2}-2}{q'_{N-1}}$.*
3. *If $n = m + a_0 + a_1 + a_2 + \dots + a_N$, where $0 < m < a_{N+1}$, then $\text{ind}(s) = 2 + \frac{q'_N-2}{mq'_N+q'_{N-1}}$.*

Let us comment on what is meant by q'_{-2} in the case $N = 0$ in Item 2: we define q'_{-2} to satisfy the recurrence relation $1 = q'_0 = a_0q'_{-1} + q'_{-2} = q'_{-2}$.

Proof: We assume that N is even (the case of N odd is analogous). Moreover, we assume that \mathbf{u} has the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\dots$.

Since b is the n^{th} bispecial factor in \mathbf{u} , the derived sequence $\mathbf{d}_{\mathbf{u}}(b)$ is standard Sturmian with the directive sequence $G^kD^{a_{N+2}}G^{a_{N+3}}\dots$, where $k = a_{N+1} - m$, see Corollary 22. By Remark 18, $\mathbf{d}_{\mathbf{u}}(b)$ is a concatenation of the blocks 1^k0 and $1^{k+1}0$. Therefore, the sequence \mathbf{u} is concatenated from the blocks $r^k s$ and $r^{k+1} s$, where r and s are the return words to b . By Lemma 24, the factor $r^{k+1}b$ is the longest factor of \mathbf{u} with the fractional root r and sb is the longest factor of \mathbf{u} with the fractional root s . In other words, $r^{\text{ind}(r)} = r^{k+1}b$ and $s^{\text{ind}(s)} = sb$. By Proposition 30 and Remark 26, we have

$$|r| = p_N + q_N = q'_N, \quad |s| = mq'_N + q'_{N-1}, \quad \text{and} \quad |b| = (m+1)q'_N + q'_{N-1} - 2.$$

As $|r^{k+1}b| = (k+1)|r| + |b| = (a_{N+1} - m + 1)|r| + |b| = (a_{N+1} + 2)q'_N + q'_{N-1} - 2$, we get

$$\text{ind}(r) = \frac{|r^{k+1}b|}{|r|} = a_{N+1} + 2 + \frac{q'_{N-1}-2}{q'_N}.$$

As $|sb| = (2m+1)q'_N + 2q'_{N-1} - 2$, we get for $m = 0$

$$\text{ind}(s) = \frac{|sb|}{|s|} = \frac{q'_N + 2q'_{N-1} - 2}{q'_{N-1}} = \frac{a_N q'_{N-1} + q'_{N-2} + 2q'_{N-1} - 2}{q'_{N-1}} = a_N + 2 + \frac{q'_{N-2}-2}{q'_{N-1}}$$

and for $m > 0$

$$\text{ind}(s) = \frac{|sb|}{|s|} = \frac{(2m+1)q'_N + 2q'_{N-1} - 2}{mq'_N + q'_{N-1}} = 2 + \frac{q'_N - 2}{mq'_N + q'_{N-1}}.$$

If the directive sequence equals $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\dots$, only the coordinates of the Parikh vectors of r , s , b are exchanged (see Remark 31). \square

Theorem 33. Let \mathbf{v} be a CS Rote sequence and let \mathbf{u} be the standard Sturmian sequence such that $\mathcal{L}(\mathcal{S}(\mathbf{v})) = \mathcal{L}(\mathbf{u})$. Then $\text{cr}(\mathbf{v}) = \sup(M_1 \cup M_2 \cup M_3)$, where

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

if the directive sequence of \mathbf{u} is $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \dots$, and

$$\begin{aligned} M_1 &= \left\{ a_{N+1} + 2 + \frac{q'_{N-1} - 1}{q'_N} : p_N \text{ is even, } N \in \mathbb{N} \right\}; \\ M_2 &= \left\{ \frac{a_{N+1} + 2}{2} + \frac{q'_{N-1} - 1}{2q'_N} : p_N \text{ is odd, } N \in \mathbb{N} \right\}; \\ M_3 &= \left\{ 2 + \frac{q'_N - 1}{q'_{N-1} + q'_N} : p_{N-1}, p_N \text{ are odd and } a_{N+1} > 1, N \geq 1 \right\} \end{aligned}$$

if the directive sequence of \mathbf{u} is $D^{a_1} G^{a_2} D^{a_3} G^{a_4} \dots$.

Proof: Let us recall that every CS Rote sequence is uniformly recurrent and aperiodic. In addition, to a CS Rote sequence \mathbf{v} we can always find a unique standard Sturmian sequence \mathbf{u} such that \mathbf{u} has the same language as $\mathcal{S}(\mathbf{v})$. It is important to realize that Theorem 14 holds for the pair \mathbf{v} and \mathbf{u} , too.

First, we assume that \mathbf{u} has the directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \dots$. We compute the suprema of the sets A_1 and A_2 defined in Theorem 14, since by this theorem, $\text{cr}(\mathbf{v}) = \sup(A_1 \cup A_2)$.

Let us decompose A_1 into $A_1 = \bigcup_{N=0}^{\infty} A_1^{(N)}$, where

$$A_1^{(N)} = \left\{ \text{ind}_{\mathbf{u}}(u) + \frac{1}{|u|} : u \text{ is a stable return word to the } n^{\text{th}} \text{ bispecial f. of } \mathbf{u}, \sum_{k=0}^N a_k \leq n < \sum_{k=0}^{N+1} a_k \right\}.$$

By definition, a word u is stable if the number of ones occurring in u is even, i.e., the second component of its Parikh vector $\vec{V}(u)$ is even. Combining Propositions 30 and 32, we obtain that $A_1^{(N)}$ contains

- $a_{N+1} + 2 + \frac{q'_{N-1} - 1}{q'_N}$ if q_N is even,
- $a_N + 2 + \frac{q'_{N-2} - 1}{q'_{N-1}}$ if q_{N-1} is even,
- the subset $B_1^{(N)} = \left\{ 2 + \frac{q'_N - 1}{m q'_N + q'_{N-1}} : m q_N + q_{N-1} \text{ even, } 0 < m < a_{N+1} \right\}$.

First we look at $A_1^{(0)}$. Since $q_0 = 1$ is odd, $q_{-1} = 0$ is even, $a_0 = 0$, $q'_0 = q'_{-1} = q'_{-2} = 1$, we get $A_1^{(0)} = \{2\}$. Since we know that $\text{cr}(\mathbf{v}) > 2$, we can consider only $N \geq 1$. Let us note that all elements

in $B_1^{(N)}$ are strictly less than 3. If q_N or q_{N-1} is even, the set $A_1^{(N)}$ contains an element ≥ 3 and the set $B_1^{(N)}$ does not play any role for $\sup A_1$. If both q_N and q_{N-1} are odd, there is an element in $B_1^{(N)}$ only for odd $m < a_{N+1}$, and obviously, $\sup B_1^{(N)}$ is attained for $m = 1$ (if $a_{N+1} = 1$ the set is empty). Together it gives $\sup A_1 = \sup(M_1 \cup M_3)$.

Analogously we define the sets $A_2^{(N)}$ for unstable return words. Then $A_2^{(N)}$ consists of

- $\frac{1}{2}(a_{N+1} + 2 + \frac{q'_{N-1}-1}{q'_N})$ if q_N is odd,
- $\frac{1}{2}(a_N + 2 + \frac{q'_{N-2}-1}{q'_{N-1}})$ if q_{N-1} is odd,
- the subset $B_2^{(N)} = \{\frac{1}{2}(2 + \frac{q'_N-1}{m q'_N + q'_{N-1}}) : m q_N + q_{N-1} \text{ odd}, 0 < m < a_{N+1}\}$.

We easily compute that $\sup A_2^{(0)} = \frac{1}{2}(a_1 + 2)$. All elements in $B_2^{(N)}$ are strictly less than $\frac{3}{2}$. Thus if q_N or q_{N-1} is odd, the set $B_2^{(N)}$ does not play any role for $\sup A_2$. If q_N and q_{N-1} are even, then the set $B_2^{(N)}$ is empty. It means that $\sup A_2 = \sup M_2$. We can conclude that $\text{cr}(\mathbf{v}) = \sup(A_1 \cup A_2) = \sup(M_1 \cup M_2 \cup M_3)$.

If the directive sequence equals $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\dots$, only the coordinates of the Parikh vectors of r and s are exchanged (see Remark 31). \square

7 CS Rote sequences with a small critical exponent

In this section, we present some corollaries of Theorem 33. As we have mentioned, Currie, Mol, and Rampersad proved in [8] that there are exactly two languages of rich binary sequences with the critical exponent less than $\frac{14}{5}$. Both of them are languages of CS Rote sequences.

Let us recall that the critical exponent depends only on the language of a sequence and not on the sequence itself. Therefore, there are infinitely many CS Rote sequences with the critical exponent less than $\frac{14}{5}$, but all of them have one of two languages. We show that among all languages of CS Rote sequences only these two languages have the critical exponent less than 3. We also describe all languages of CS Rote sequences with the critical exponent equal to 3.

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

1. $G^{a_1}(D^2G^2)^\omega$, where $a_1 = 1$ or $a_1 = 3$; in this case $\text{cr}(\mathbf{v}) = 2 + \frac{1}{\sqrt{2}}$;
2. $G^{a_1}D^4(G^2D^2)^\omega$, where $a_1 = 1$ or $a_1 = 3$; in this case $\text{cr}(\mathbf{v}) = 3$;
3. $G^{a_1}D^1G^{a_3}(D^2G^2)^\omega$, where $a_1 = 2$ or $a_1 = 4$ and $a_3 = 1$ or $a_3 = 3$; in this case $\text{cr}(\mathbf{v}) = 3$;
4. $D^1G^{a_2}(D^2G^2)^\omega$, where $a_2 = 1$ or $a_2 = 3$; in this case $\text{cr}(\mathbf{v}) = 3$.

Proof: For each $N \in \mathbb{N}$ we denote by $\beta_N = a_{N+1} + 2 + \frac{q'_{N-1}-1}{q'_N}$ the number which is a candidate to join the set M_1 . We can easily compute that $\beta_0 = a_1 + 2$, $\beta_1 = a_2 + 2$ and $\beta_N > 3$ for every $N \geq 2$. Indeed, it

suffices to realize that $q'_{-1} = q'_0 = 1$, $q'_1 = a_1 + 1 > 1$ and $(q'_N)_{N \geq 1}$ is an increasing sequence of integers, so $\frac{q'_{N-1}-1}{q'_N} \in (0, 1)$ for all $N \geq 2$. Since we look for a sequence \mathbf{v} with $\text{cr}(\mathbf{v}) \leq 3$, we have to ensure that $\beta_N \notin M_1$ for all $N \geq 2$ by the parity conditions. It is also important to notice that $\sup M_3 \leq 3$. Indeed, since $\frac{q'_N-1}{q'_N+q'_{N-1}} \in [0, 1)$ for all $N \in \mathbb{N}$, all elements of M_3 are less than 3.

First we assume that the directive sequence of \mathbf{u} is $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\dots$. By Theorem 33, if q_N is even, then $\beta_N \in M_1$, otherwise $\frac{1}{2}\beta_N \in M_2$. To ensure $\sup M_1 \leq 3$, q_N has to be odd for all $N \geq 2$. Moreover, to ensure $\sup M_2 \leq 3$, $\beta_N \leq 6$ and so $a_{N+1} \leq 3$ for all $N \geq 2$. Since $q_0 = 1$ is odd, $\frac{1}{2}\beta_0 = \frac{a_1}{2} + 1 \in M_2$ and so $a_1 \leq 4$. We distinguish two cases.

(i) If $q_1 = a_1$ is odd, then M_1 is empty and $\frac{1}{2}\beta_1 = \frac{a_2}{2} + 1 \in M_2$. Thus $a_2 \leq 4$. The recurrence relation $q_N = a_N q_{N-1} + q_{N-2}$ with the odd initial conditions q_0 and q_1 produces q_N odd for all $N \geq 2$ if and only if a_N is even for all $N \geq 2$. We can summarize that $a_1 \in \{1, 3\}$, $a_2 \in \{2, 4\}$ and $a_N = 2$ for all $N \geq 3$.

Let us observe that if $a_2 = 4$, then $\sup M_2 = 3$. Since $\sup M_3 \leq 3$ and M_1 is empty, we conclude that $\text{cr}(\mathbf{v}) = \sup M_2 = 3$. It gives us Item 2 of our proposition.

If $a_2 = 2$, then it is easy to check that all elements of M_2 are smaller than $\frac{5}{2}$ and thus $\sup M_2 \leq \frac{5}{2}$. Since M_1 is empty, to prove Item 1, it remains to deduce

$$\sup M_3 = \sup \left\{ 2 + \frac{q'_N - 1}{q'_{N-1} + q'_N} : N \in \mathbb{N} \right\} = 2 + \frac{1}{\sqrt{2}} > \frac{5}{2}.$$

The sequence $(q'_N)_{N \geq 2}$ fulfils the recurrence relation $q'_N = 2q'_{N-1} + q'_{N-2}$ with the initial conditions $q'_0 = 1$ and $q'_1 = a_1 + 1$. This linear recurrence has the solution

$$q'_N = \frac{1}{2\sqrt{2}} \left((a_1 + \sqrt{2})(1 + \sqrt{2})^N - (a_1 - \sqrt{2})(1 - \sqrt{2})^N \right). \quad (7)$$

Using the solution one can easily check that $q'_{N-1} + q'_N = \sqrt{2}q'_N + c_N$, where $c_N = (a_1 - \sqrt{2})(1 - \sqrt{2})^N$. As $\lim_{N \rightarrow \infty} c_N = 0$ and $\lim_{N \rightarrow \infty} q'_N = +\infty$, we have

$$\lim_{n \rightarrow \infty} \left(2 + \frac{q'_N - 1}{q'_{N-1} + q'_N} \right) = \lim_{n \rightarrow \infty} \left(2 + \frac{q'_N - 1}{\sqrt{2}q'_N + c_N} \right) = 2 + \frac{1}{\sqrt{2}}.$$

Moreover, since $c_N > -1$, all elements of M_3 are smaller than the limit $2 + \frac{1}{\sqrt{2}}$. Indeed,

$$2 + \frac{q'_N - 1}{\sqrt{2}q'_N + c_N} < 2 + \frac{q'_N - 1}{\sqrt{2}q'_N - 1} < 2 + \frac{1}{\sqrt{2}}.$$

(ii) If $q_1 = a_1$ is even, then $M_1 = \{\beta_1 = a_2 + 2\}$ since all the other q_N are odd. Thus $a_2 = 1$ and $\text{cr}(\mathbf{v}) \geq 3$. The recurrence relation $q_N = a_N q_{N-1} + q_{N-2}$ with the initial conditions $q_0 = 1$, $q_1 = a_1 \in \{2, 4\}$ produces odd q_N for all $N \geq 2$ if and only if a_3 is odd and a_N is even for all $N \geq 4$. Moreover, to ensure $\sup M_2 \leq 3$, we have to take $a_1 \leq 4$ and $a_N \leq 3$ for every $N \geq 3$. Together with the parity conditions we get $a_1 \in \{2, 4\}$, $a_2 = 1$, $a_3 \in \{1, 3\}$, and $a_N = 2$ for all $N \geq 4$. Since $\sup M_3 \leq 3$, the set M_3 can be omitted. Together it means that in this case $\text{cr}(\mathbf{v}) = 3$ and it corresponds to Item 3 of our statement.

Now we assume that the directive sequence of \mathbf{u} is $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\dots$. By Theorem 33, if p_N is even, then $\beta_N \in M_1$, otherwise $\frac{\beta_N}{2} \in M_2$. Since $p_0 = 0$ is even, $\beta_0 = a_1 + 2 \in M_1$. Therefore, $a_1 = 1$ and $\text{cr}(\mathbf{v}) \geq 3$. Since $p_1 = a_1 p_0 + p_{-1} = 1$ is odd, $\frac{\beta_1}{2} = \frac{a_2}{2} + 1 \in M_2$ and so $a_2 \leq 4$. To guarantee $\text{cr}(\mathbf{v}) \leq 3$, p_N has to be odd and $a_{N+1} \leq 3$ for all $N \geq 2$. The recurrence relation $p_N = a_N p_{N-1} + p_{N-2}$ with the initial conditions $p_0 = 0$ and $p_1 = 1$ produces p_N odd for all $N \geq 2$ if and only if a_2 is odd and a_N is even for all $N \geq 2$. Clearly, the set M_3 can be omitted. We may conclude that $\text{cr}(\mathbf{v}) = 3$ and $a_1 = 1$, $a_2 \in \{1, 3\}$, and $a_N = 2$ for all $N \geq 3$, which corresponds to Item 4. \square

Remark 35. Let us emphasize that all standard Sturmian sequences from Proposition 34, i.e., which are associated with CS Rote sequences with the critical exponent ≤ 3 , are morphic images of the fixed point of the morphism D^2G^2 . It follows directly from the fact that their directive sequences have the periodic suffix $(D^2G^2)^\omega$.

Example 36. In Proposition 34, we have shown that the CS Rote sequence \mathbf{v} such that $\mathcal{S}(\mathbf{v})$ has the directive sequence $G(D^2G^2)^\omega$ has the critical exponent $\text{cr}(\mathbf{v}) = 2 + \frac{1}{\sqrt{2}}$. Let us determine the critical exponent $\text{cr}(\mathbf{v}')$ of the CS Rote sequence \mathbf{v}' associated to the standard Sturmian sequence $\mathcal{S}(\mathbf{v}')$ obtained by the exchange of letters from $\mathcal{S}(\mathbf{v})$, i.e., $\mathcal{S}(\mathbf{v}') = E(\mathcal{S}(\mathbf{v}))$.

By Remark 18, the directive sequence of $\mathcal{S}(\mathbf{v}')$ equals $D(G^2D^2)^\omega$. Thus we have $\theta = [0, 1, 2, 2, 2, \dots]$ and it is readily seen that p_N is even if and only if N is even. Let us calculate $\text{cr}(\mathbf{v}')$ by Theorem 33. We have

$$M_1 = \{a_{2N+1} + 2 + \frac{q'_{2N-1} - 1}{q'_{2N}} : N \in \mathbb{N}\} = \{3\} \cup \{4 + \frac{q'_{2N-1} - 1}{q'_{2N}} : N \in \mathbb{N}, N \geq 1\}.$$

Using equation (7), we can check that the sequence $(\frac{q'_{2N-1}-1}{q'_{2N}})$ is increasing and has the limit $\frac{1}{1+\sqrt{2}}$, and therefore $\sup M_1 = 4 + \frac{1}{1+\sqrt{2}}$. Since the elements of M_2 and M_3 are ≤ 3 , we can conclude that $\text{cr}(\mathbf{v}') = 4 + \frac{1}{1+\sqrt{2}}$.

It is well-known that among Sturmian sequences the sequence with the lowest possible critical exponent is the Fibonacci sequence \mathbf{f} , which has $\text{cr}(\mathbf{f}) = 3 + \frac{2}{1+\sqrt{5}} \sim 3.602$. The following theorem implies that there are uncountably many CS Rote sequences with the critical exponent smaller than $\text{cr}(\mathbf{f})$.

Theorem 37. Let $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\dots$ be the directive sequence of a standard Sturmian sequence \mathbf{u} and let \mathbf{v} be the CS Rote sequence associated with \mathbf{u} . Then $\text{cr}(\mathbf{v}) < \frac{7}{2}$ if and only if the sequence $a_1a_2a_3\dots$ is a concatenation of the blocks from the following list:

- L_0 : 111;
- L_1 : $s1$, where $s \in \{2, 4\}$;
- L_2 : $cs31$, where $c \in \{1, 3\}$ and $s \in \{2, 4\}^*$;
- L_3 : cs , where $c \in \{1, 3\}$ and $s \in \{2, 4\}^{\mathbb{N}}$,

and if the block L_0 appears in $a_1a_2a_3\dots$, then it is a prefix of $a_1a_2a_3\dots$.

Proof: As in the proof of Proposition 34, we again use the sets M_1 and M_2 from Theorem 33. The set M_3 can be omitted since $\sup M_3 \leq 3$.

Let us recall that $q_0 = 1$, $q_1 = a_1$, and if q_N is even, then $\beta_N = a_{N+1} + 2 + \frac{q'_{N-1}-1}{q'_N} \in M_1$, otherwise $\frac{1}{2}\beta_N \in M_2$. First we suppose that $\text{cr}(\mathbf{v}) < \frac{7}{2}$ and we deduce several auxiliary observations for each $N \in \mathbb{N}$:

1. If q_N even, then $a_{N+1} = 1$.
Proof: It follows from the inequality $\beta_N < \frac{7}{2}$.
2. If q_N odd, then $a_{N+1} \in \{1, 2, 3, 4\}$.
Proof: It follows from the inequality $\frac{1}{2}\beta_N < \frac{7}{2}$.
3. If q_{N-1} odd and q_N even, then q_{N+1} odd.
Proof: It follows from Item 1 and the relation $q_{N+1} = a_{N+1}q_N + q_{N-1} = q_N + q_{N-1}$.
4. If q_N even, q_{N+1} odd, and q_{N+2} even, then $a_{N+2} \in \{2, 4\}$.
Proof: It follows from Item 2 and the relation $q_{N+2} = a_{N+2}q_{N+1} + q_N$.
5. If q_N even, q_{N+1} odd, and q_{N+2} odd, then $a_{N+2} \in \{1, 3\}$.
Proof: It follows from Item 2 and the relation $q_{N+2} = a_{N+2}q_{N+1} + q_N$.
6. If q_N odd, q_{N+1} odd, and q_{N+2} odd, then $a_{N+2} \in \{2, 4\}$.
Proof: It follows from Item 2 and the relation $q_{N+2} = a_{N+2}q_{N+1} + q_N$.
7. If q_N odd, q_{N+1} odd, and q_{N+2} even, then $a_{N+2} \in \{1, 3\}$. Moreover, if $a_{N+2} = 1$, then $N = 0$ and $a_1 = 1$.
Proof: Item 2 and the relation $q_{N+2} = a_{N+2}q_{N+1} + q_N$ imply $a_{N+2} \in \{1, 3\}$. Assume that $a_{N+2} = 1$. By Item 1, $a_{N+3} = 1$. As q_{N+2} is even, $\beta_{N+2} \in M_1$ and so $\beta_{N+2} = 3 + \frac{q'_{N+1}-1}{q'_{N+2}} < \frac{7}{2}$. By some simple rearrangements and applications of the recurrence relation, we can rewrite this inequality equivalently as $(a_{N+1} - 1)q'_N + q'_{N-1} < 2$. It is easy to verify that this inequality holds only for $N = 0$ and $a_1 = 1$ or $N = 1$ and $a_2 = 1$. Nevertheless, the second case leads to a contradiction with the assumption that both $q_1, q_2 = a_2q_1 + 1$ are odd.
8. Let $M > N + 2$. If q_N even, q_M even, and q_K odd for all $K, N < K < M$, then $a_{N+2} \in \{1, 3\}$, $a_M = 3$ and $a_K \in \{2, 4\}$ for all $K, N + 2 < K < M$.
Proof: Item 6 implies $a_K \in \{2, 4\}$ for all $K, N + 2 < K < M$. By Item 5, $a_{N+2} \in \{1, 3\}$ and by Item 7, $a_M = 3$.

Using the previous claims we show that for each J for which q_J is even, at the position $J + 1$ ends one of the blocks L_0, L_1 , or L_2 . Moreover, the block L_0 can only occur as a prefix of the sequence $a_1a_2a_3 \cdots$, while each of the blocks L_1 and L_2 is either a prefix of $a_1a_2a_3 \cdots$ or it starts at the position $I + 2$, where I is the greatest integer smaller than J for which q_I is even. We discuss three cases:

- Let q_1 be even. Then Item 1 implies $a_2 = 1$. And since $a_1 = q_1$ is even and $q_0 = 1$ is odd, by Item 2, we get $a_1 \in \{2, 4\}$. Thus the prefix a_1a_2 of the sequence $a_1a_2a_3 \cdots$ is of the form L_1 from our list.

- Let $J > 1$ be the first index such that q_J is even. As $q_1 = a_1$, a_1 is odd, and by Item 2, we get $a_1 \in \{1, 3\}$. Item 6 implies $a_K \in \{2, 4\}$ for all $K, 1 < K < J$. By Item 7, $a_J = 3$ or $a_J = 1$. But if $a_J = 1$, then $J = 2$ and $a_1 = 1$. Finally, Item 1 implies $a_{J+1} = 1$. Thus the prefix $a_1 a_2 \cdots a_J a_{J+1}$ of the sequence $a_1 a_2 a_3 \cdots$ is of the form L_0 or L_2 from our list.
- Let I be an index such that q_I is even and let J be the smallest index greater than I for which q_J is even. By Item 1, $a_{I+1} = 1$. The word $a_{I+2} \cdots a_J a_{J+1}$ is by Item 4 or Item 8 either of the form L_1 or L_2 .

If there are infinitely many even denominators q_N , then we have shown that the sequence $a_1 a_2 a_3 \cdots$ is concatenated from the blocks L_0, L_1 , and L_2 (L_0 can only be a prefix). It remains to consider the case when only finitely many denominators q_N are even.

- Let q_N be odd for every $N \in \mathbb{N}$. Then $q_1 = a_1$ is odd. Especially, since both q_0 and $q_1 = a_1$ are odd, Item 2 implies $a_1 \in \{1, 3\}$ and it follows from Item 6 that $a_2 a_3 a_4 \cdots \in \{2, 4\}^{\mathbb{N}}$. Therefore, the sequence $a_1 a_2 a_3 \cdots$ is equal to the block L_3 from our list.
- Let $L \geq 1$ be the last index such that q_L is even. In particular, it means that q_N is even only for a finite number of indices. Item 1 implies that $a_{L+1} = 1$. By Items 5 and 6, the suffix $a_{L+2} a_{L+3} a_{L+4} \cdots$ of the sequence $a_1 a_2 a_3 \cdots$ equals to L_3 .

Now we have to show that any directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \cdots$ such that $a_1 a_2 a_3 \cdots$ is concatenated of the blocks from the list gives a standard Sturmian sequence \mathbf{u} such that the CS Rote sequence \mathbf{v} associated to \mathbf{u} has the critical exponent less than $\frac{7}{2}$.

If q_N is odd, then $\frac{1}{2}\beta_N = \frac{1}{2}a_{N+1} + 1 + \frac{q'_{N-1}-1}{2q'_N} < 3 + \frac{q'_{N-1}}{2q'_{N-1}+q'_{N-2}} < \frac{7}{2}$, as each a_{N+1} is ≤ 4 .

If q_N is even, it is easy to prove by induction on N that there is a block of the form L_0, L_1 , or L_2 ending at the position $N + 1$ (and L_0 only for $N = 2$). In particular, it means that $a_{N+1} = 1$ and $a_N \geq 2$ or $a_1 = a_2 = a_3 = 1$. In the first case, we get $\beta_N \leq 3 + \frac{q'_{N-1}}{2q'_{N-1}+q'_{N-2}} < \frac{7}{2}$, while in the second case, we get $\beta_2 = 3 + \frac{a_1+1-1}{a_2(a_1+1)+1} = 3 + \frac{1}{3} < \frac{7}{2}$.

To show that $\sup(M_1 \cup M_2) < \frac{7}{2}$, we need to show that $\sup \frac{q'_{N-1}}{2q'_{N-1}+q'_{N-2}} < \frac{1}{2}$. As $a_N \leq 4$ for all N , we can estimate $\frac{q'_{N-2}}{q'_{N-1}} \geq \frac{q'_{N-2}}{4q'_{N-2}+q'_{N-2}} = \frac{1}{5}$ and thus $\frac{q'_{N-1}}{2q'_{N-1}+q'_{N-2}} = \frac{1}{2 + \frac{q'_{N-2}}{q'_{N-1}}} \leq \frac{5}{11}$. \square

Remark 38. It would be interesting to reveal some topological properties of the set

$$\text{cr}_{\text{Rote}} := \{\text{cr}(\mathbf{v}) : \mathbf{v} \text{ is a CS Rote sequence}\},$$

for instance, to find its accumulation points in the interval $(3, \frac{7}{2})$. The proof of the previous theorem implies that there is no CS Rote sequence with the critical exponent between $3 + \frac{5}{11}$ and $3 + \frac{1}{2}$. We even believe that for any CS Rote sequence \mathbf{v} , the following implication holds: If $\text{cr}(\mathbf{v}) < 3 + \frac{1}{2}$, then $\text{cr}(\mathbf{v}) < 3 + \frac{1}{1+\sqrt{3}}$.

8 The recurrence function of CS Rote sequences

The main result of this section is Theorem 54, where we describe the recurrence function of any CS Rote sequence in terms of the convergents related to the associated Sturmian sequence. To obtain this result, we proceed similarly as in the previous parts concerning the critical exponent, i.e., we transform our task of finding the recurrence function of a CS Rote sequence into studying some properties of its associated Sturmian sequence. Let us emphasize that we may still restrict our consideration to CS Rote sequences associated with standard Sturmian sequences without loss of generality.

Definition 39. Let \mathbf{u} be a uniformly recurrent sequence. The mapping $R_{\mathbf{u}} : \mathbb{N} \rightarrow \mathbb{N}$ defined by

$$R_{\mathbf{u}}(n) = \min\{N \in \mathbb{N} : \text{each factor of } \mathbf{u} \text{ of length } N \text{ contains all factors of } \mathbf{u} \text{ of length } n\}$$

is called the *recurrence function* of \mathbf{u} .

The definition of the recurrence function may be reformulated in terms of return words [7].

Observation 40. Let \mathbf{u} be a uniformly recurrent sequence. Then

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

Moreover, to determine $R_{\mathbf{u}}(n)$ we can restrict our consideration to return words to bispecial factors of \mathbf{u} .

Lemma 41. Let \mathbf{u} be a uniformly recurrent aperiodic sequence. For $n \in \mathbb{N}$, we denote

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

Then

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

Proof: For evaluation of $R_{\mathbf{u}}(n)$ we use Observation 40. Let $w \in \mathcal{L}(\mathbf{u})$ and $|w| = n$.

If w is not right special, then there exists a unique letter x such that $wx \in \mathcal{L}(\mathbf{u})$. Obviously, the occurrences of w and wx in \mathbf{u} coincide. Therefore, return words to w and wx coincide as well.

If w is not left special, then there is a unique letter y such that $yw \in \mathcal{L}(\mathbf{u})$. If r is a return word to w , then the word $yr y^{-1}$ is a return word to yw and the return words r and $yr y^{-1}$ are of the same length.

These two facts imply that the lengths of return words to w equal the lengths of return words to the shortest bispecial factor containing w . \square

The following lemma shows that for a CS Rote sequence \mathbf{v} associated with the Sturmian sequence \mathbf{u} the sets $\mathcal{B}_{\mathbf{v}}(n+1)$ and $\mathcal{B}_{\mathbf{u}}(n)$ correspond naturally for every $n \in \mathbb{N}$. Thus to determine the set $\mathcal{B}_{\mathbf{v}}(n+1)$, we first describe the set $\mathcal{B}_{\mathbf{u}}(n)$ for a standard Sturmian sequence \mathbf{u} .

Lemma 42. Let w be a factor of length $n+1$ in a CS Rote sequence \mathbf{v} and let v be the shortest bispecial factor of \mathbf{v} containing w . Then the factor $\mathcal{S}(v)$ is the shortest bispecial factor of the associated Sturmian sequence $\mathcal{S}(\mathbf{v})$ such that $\mathcal{S}(v)$ contains $\mathcal{S}(w)$.

Proof: The statement is a consequence of the simple fact that $\mathcal{S}(v)$ is a bispecial factor of $\mathcal{S}(\mathbf{v})$ if and only if v and $E(v)$ are bispecial factors of \mathbf{v} . (See Lemma 7.) \square

In the sequel, we will essentially use a characterization of Sturmian sequences by palindromes from [11]. Let us first recall some basic notions. Consider an alphabet \mathcal{A} . The assignment $w = w_0w_1 \cdots w_{n-1} \rightarrow \bar{w} = w_{n-1}w_{n-2} \cdots w_0$ is called a *mirror mapping*, and the word \bar{w} is called the *reversal* or the *mirror image* of $w \in \mathcal{A}^*$. A word w which coincides with its mirror image \bar{w} is a *palindrome*. If p is a palindrome of odd length, then the *center* of p is a letter a such that $p = sa\bar{s}$ for some $s \in \mathcal{A}^*$. The center of a palindrome p of even length is the empty word ε .

Theorem 43 ([11]). *A sequence \mathbf{u} is Sturmian if and only if \mathbf{u} contains one palindrome of every even length and two palindromes of every odd length.*

Moreover, when studying in detail the proof of this theorem presented by Droubay and Pirillo [11], we deduce that any two palindromes of the same odd length have distinct centers, one has the center 0 and the other one has the center 1. In fact, we get the following corollary.

Corollary 44. *A binary sequence \mathbf{u} is Sturmian if and only if every palindrome in $\mathcal{L}(\mathbf{u})$ has a unique palindromic extension, i.e., for any palindrome $p \in \mathcal{L}(\mathbf{u})$ there exists a unique letter $a \in \{0, 1\}$ such that $apa \in \mathcal{L}(\mathbf{u})$.*

We believe that Theorem 46 is already known. However, since we have not found it in the literature, we add its proof. For this purpose, we need an auxiliary lemma. Let us recall that the language $\mathcal{L}(\mathbf{u})$ of a Sturmian sequence \mathbf{u} is *closed under reversal*, i.e., $\mathcal{L}(\mathbf{u})$ contains with every factor w also its reversal \bar{w} , and all bispecial factors of \mathbf{u} are palindromes.

Lemma 45. *Let \mathbf{u} be a Sturmian sequence. Let $p \in \mathcal{L}(\mathbf{u})$ be a palindrome and let v be the shortest bispecial factor containing p .*

1. *Then p is a central factor of v , i.e., $v = sp\bar{s}$ for some word s .*
2. *If v' is the shortest bispecial factor with the same center as p and $|v'| \geq |p|$, then $v' = v$.*

Proof:

1. Let $v = sp$ be the shortest left special factor containing p , in particular $0sp$ and $1sp$ belong to $\mathcal{L}(\mathbf{u})$. Since the language $\mathcal{L}(\mathbf{u})$ is closed under reversal, $p\bar{s}$ is right special, i.e., $p\bar{s}0$ and $p\bar{s}1$ belong to $\mathcal{L}(\mathbf{u})$. As s is the only possible extension of p to the left by a factor of length $|s|$, both $sp\bar{s}0$ and $sp\bar{s}1$ belong to $\mathcal{L}(\mathbf{u})$. By the same argument, \bar{s} is the only possible extension of p to the right by a factor of length $|s|$. Therefore, $0sp\bar{s}$ and $1sp\bar{s}$ belong to $\mathcal{L}(\mathbf{u})$. Thus $sp\bar{s}$ is the shortest bispecial factor containing p .
2. Assume for contradiction that $v \neq v'$. Since v and v' are palindromes with the same centers, there exists a palindrome q such that $v' = s'q\bar{s}'$ and $v = sq\bar{s}$. Let q be the longest palindrome with this property. If $|q| = |v'|$, then necessarily $v' = v$. If $|v'| > |q|$, then the last letters of s' and s are distinct and q is a palindrome with two distinct palindromic extensions. This contradicts Corollary 44.

□

Theorem 46. *Let \mathbf{u} be a Sturmian sequence and $n \in \mathbb{N}, n \geq 1$. Find the shortest bispecial factors P_ε , resp. P_0 , resp. P_1 of length greater than or equal to n with the center ε , resp. 0, resp. 1. Then the following statements hold:*

1. Let w be a factor of \mathbf{u} of length n and let v be the shortest bispecial factor containing w . Then $v \in \{P_\varepsilon, P_0, P_1\}$.
2. If \mathbf{u} contains no bispecial factor of length $n - 1$, then for each $i \in \{\varepsilon, 0, 1\}$ there exists a factor w of \mathbf{u} of length n such that the shortest bispecial factor containing w is P_i .
3. If there exists a bispecial factor v of \mathbf{u} of length $n - 1$ and let $i \in \{\varepsilon, 0, 1\}$ be the center of the palindrome v , then for each $j \in \{\varepsilon, 0, 1\}$, $j \neq i$, there exists a factor w of \mathbf{u} of length n such that the shortest bispecial factor containing w is P_j , while P_i is not the shortest bispecial factor containing w for any factor w of \mathbf{u} of length n .

Proof: Consider the Rauzy graph of \mathbf{u} of order $n - 1$, i.e., a directed graph Γ_{n-1} whose vertices are factors of \mathbf{u} of length $n - 1$ and edges are factors of \mathbf{u} of length n . An edge e starts in the vertex x and ends in the vertex y if x is a prefix and y is a suffix of e . Denote ℓ , resp. r the vertex corresponding to the unique left special, resp. right special factor of length $n - 1$. Furthermore, we denote the shortest path from ℓ to r by p_A , and the shortest paths of non-zero length starting in r and ending in ℓ by p_B and p_C . If \mathbf{u} has no bispecial factor of length $n - 1$, then p_A has a positive length, see Figure 1(a). If \mathbf{u} has a bispecial factor of length $n - 1$, then the path p_A consists of a unique vertex – the bispecial factor b , see Figure 1(b).

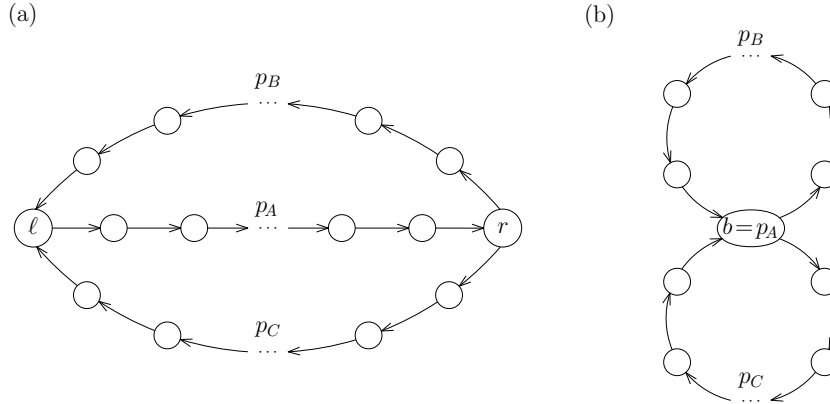


Fig. 1: The Rauzy graph of a Sturmian word (a) without a bispecial vertex, (b) with a bispecial vertex.

Observing these Rauzy graphs, it is obvious that for each edge e from the path p_x , where $x \in \{A, B, C\}$, the shortest bispecial factor containing e is the same as the shortest bispecial factor containing p_x . Clearly, if p_A has length 0, then $x \in \{B, C\}$.

Since the language $\mathcal{L}(\mathbf{u})$ is closed under reversal, the mirror mapping restricted to the factors of length $n - 1$ and n is an automorphism of the graph Γ_{n-1} . Let us suppose that a palindrome q of length $n - 1$ or n is contained in p_x as an edge or an inner vertex. As any palindrome is mapped onto itself, $\bar{r} = \ell$, and $\bar{\ell} = r$, this path p_x is mapped onto itself, i.e., $\overline{p_x} = p_x$, and the palindrome q is a central factor of p_x . On one hand, it means that p_x is a palindrome with the same center as q , on the other hand, it also means that p_x cannot contain any other palindrome of length $n - 1$ and n as its edge or an inner vertex.

By Theorem 43 and the comment after it, there are exactly three palindromes among all vertices and edges of Γ_{n-1} (all factors of length $n-1$ or n), and moreover, they have distinct centers. We may conclude that the paths p_A, p_B, p_C are palindromes and their centers are distinct. The rest of the proof follows from Lemma 45. \square

Observation 47. *Let P be a palindrome. Then its Parikh vector satisfies:*

1. $\vec{V}(P) = \binom{0}{0} \pmod 2$ if and only if P has the center ε ;
2. $\vec{V}(P) = \binom{1}{0} \pmod 2$ if and only if P has the center 0;
3. $\vec{V}(P) = \binom{0}{1} \pmod 2$ if and only if P has the center 1.

Let us recall that a factor of a standard Sturmian sequence \mathbf{u} is bispecial if and only if it is a palindromic prefix of \mathbf{u} . Therefore we can order the bispecial factors of a given standard Sturmian sequence \mathbf{u} according to their lengths. We denote the k -th bispecial factor of \mathbf{u} by $BS(k)$. Thus $BS(0) = \varepsilon$, $BS(1) = a$, where a is the first (and the more frequent) letter of \mathbf{u} etc.

The sequences (p_N) , (q_N) , and (q'_N) we use in the remaining part of the paper were introduced in Notation 25, the notation $\mathcal{B}_{\mathbf{u}}(n)$ comes from Lemma 41.

Theorem 48. *Let \mathbf{u} be a standard Sturmian sequence with the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\dots$ or $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\dots$, and $n \in [q'_N, q'_{N+1})$ for some $N \in \mathbb{N}$. Put $M = a_0 + a_1 + a_2 + \dots + a_N$, where $a_0 = 0$.*

- *If $n \in [q'_N, q'_{N+1} - 1)$ and $n - 1$ is not the length of a bispecial factor, then*
 $\mathcal{B}_{\mathbf{u}}(n) = \{BS(M+m), BS(M+m+1), BS(M+a_{N+1}+1)\}$ for some $m \in \{0, \dots, a_{N+1} - 1\}$.
- *If $n \in [q'_N, q'_{N+1} - 1]$ and $n - 1$ is the length of a bispecial factor, then*
 $\mathcal{B}_{\mathbf{u}}(n) = \{BS(M+m), BS(M+a_{N+1}+1)\}$ for some $m \in \{0, \dots, a_{N+1}\}$.

Proof: Assume that \mathbf{u} has the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\dots$. By Proposition 30, the Parikh vectors of bispecial factors satisfy

$$\vec{V}(BS(M+i+1)) = \vec{V}(BS(M+i)) + \binom{p_N}{q_N} \quad \text{for } i = 0, 1, \dots, a_{N+1} - 1 \quad (8)$$

$$\text{and} \quad \vec{V}(BS(M+a_{N+1}+1)) = \vec{V}(BS(M+a_{N+1})) + \binom{p_{N+1}}{q_{N+1}}. \quad (9)$$

Using Observation 47 and the relation $\binom{p_N}{q_N} \neq \binom{0}{0} \pmod 2$ from Lemma 28, we deduce that the centers of palindromes $BS(M+i)$, where $i = 0, 1, \dots, a_{N+1}$, alternate between two distinct elements of $\{\varepsilon, 0, 1\}$. The third element of $\{\varepsilon, 0, 1\}$ is the center of the palindrome $BS(M+a_{N+1}+1)$, as $\binom{p_N}{q_N} \neq \binom{p_{N+1}}{q_{N+1}} \pmod 2$, see Lemma 28. By Proposition 30, the length of $BS(M-1)$ equals $q'_N - 2$ and the length of $BS(M+a_{N+1}-1)$ equals $q'_{N+1} - 2$.

- Let us discuss the case $n = q'_{N+1} - 1$. The palindromes $BS(M + a_{N+1} - 1)$, $BS(M + a_{N+1})$ and $BS(M + a_{N+1} + 1)$ have distinct centers, and $n - 1$ is the length of the palindrome $BS(M + a_{N+1} - 1)$. Item 3 of Theorem 46 implies that all factors of length n occur in $BS(M + a_{N+1})$ and $BS(M + a_{N+1} + 1)$. Therefore, the set $\mathcal{B}_{\mathbf{u}}(n)$ consists of these two bispecial palindromes.
- Now we assume that $n \in [q'_N, q'_{N+1} - 2]$. Clearly, the length of $BS(M - 1)$ is strictly smaller than $n - 1$ and n does not exceed the length of $BS(M + a_{N+1} - 1)$. We choose the smallest $m \in \{0, 1, \dots, a_{N+1} - 1\}$ such that $n \leq |BS(M + m)|$. The bispecial factors $BS(M + m)$, $BS(M + m + 1)$, and $BS(M + a_{N+1} + 1)$ have distinct centers.

If $n - 1$ is not the length of any bispecial factor, then Item 2 of Theorem 46 implies that $\mathcal{B}_{\mathbf{u}}(n) = \{BS(M + m), BS(M + m + 1), BS(M + a_{N+1} + 1)\}$.

If $n - 1$ is the length of a bispecial factor, then Item 3 of Theorem 46 together with the fact that the centers of $BS(M + i)$ alternate for $i = 0, 1, \dots, a_{N+1} - 1$ implies that $\mathcal{B}_{\mathbf{u}}(n) = \{BS(M + m), BS(M + a_{N+1} + 1)\}$.

If \mathbf{u} has the directive sequence $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\dots$, then the proof will be analogous, only the coordinates of the Parikh vectors will be exchanged, see Remark 31. \square

Remark 49. The recurrence function of a standard Sturmian sequence \mathbf{u} with the directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\dots$ or $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\dots$ is known to satisfy $R_{\mathbf{u}}(n) = q'_{N+1} + q'_N + n - 1$ for every $n \in [q'_N, q'_{N+1})$. Let us show that this formula is a consequence of the previous statements. Indeed, by Lemma 41, we have to find the longest return word to the bispecial factor from the set $\mathcal{B}_{\mathbf{u}}(n)$ described in Theorem 48. Using Proposition 30, we find that the longest one is the return word s corresponding to the bispecial factor $BS(M + a_{N+1} + 1)$. Its length is $|s| = q'_{N+1} + q'_N$. And thus Lemma 41 implies the above mentioned formula, which was obtained by Hedlund and Morse already in 1940, see [20].

We have prepared everything we need to derive the formula for the recurrence function of CS Rote sequences. For this purpose, we recall Theorem 3.10 from [18]:

Theorem 50. *Let \mathbf{v} be a CS Rote sequence associated with the standard Sturmian sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$. Let v be a non-empty prefix of \mathbf{v} and $u = \mathcal{S}(v)$. Let r , resp. s be the more frequent, resp. the less frequent return word to u in \mathbf{u} and let ℓ be a positive integer such that \mathbf{u} is a concatenation of the blocks $r^\ell s$ and $r^{\ell+1}s$. Then the prefix v of \mathbf{v} has three return words A, B, C satisfying:*

1. *If r is stable and s unstable, then $\mathcal{S}(A0) = r$, $\mathcal{S}(B0) = sr^\ell s$, $\mathcal{S}(C0) = sr^{\ell+1}s$.*
2. *If r is unstable and s stable, then $\mathcal{S}(A0) = s$, $\mathcal{S}(B0) = rr$, $\mathcal{S}(C0) = rsr$.*
3. *If both r and s are unstable, then $\mathcal{S}(A0) = rr$, $\mathcal{S}(B0) = rs$, $\mathcal{S}(C0) = sr$.*

We will use the previous theorem for the determination of return words to bispecial factors of CS Rote sequences (which are by Lemma 42 associated to bispecial factors of Sturmian sequences). In particular, we focus on v such that $\mathcal{S}(v)$ is a bispecial factor of the Sturmian sequence $\mathbf{u} = \mathcal{S}(\mathbf{v})$ and $\mathcal{S}(v)$ belongs to the set $\mathcal{B}_{\mathbf{u}}(n)$ described in Theorem 48.

In fact, the following lemmas explain that only the bispecial factor $\mathcal{S}(v) = BS(M + a_{N+1} + 1)$, where $N \in \mathbb{N}$, $M = a_0 + a_1 + \dots + a_N$, is important. By Proposition 30, its return words r and s have either the lengths $|r| = q'_{N+1}$ and $|s| = q'_{N+1} + q'_N$ if $a_{N+2} > 1$, or $|r| = q'_{N+1} + q'_N$ and $|s| = q'_{N+1}$ if $a_{N+2} = 1$.

Lemma 51. *Let \mathbf{v} be a CS Rote sequence and $\mathbf{u} = \mathcal{S}(\mathbf{v})$ be the associated standard Sturmian sequence with the directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \dots$ or $D^{a_1} G^{a_2} D^{a_3} G^{a_4} \dots$. Put $M = a_0 + a_1 + a_2 + \dots + a_N$, where $a_0 = 0$. Let x and y be the bispecial factors in \mathbf{v} such that $\mathcal{S}(x) = BS(M + a_{N+1} + 1)$ and $\mathcal{S}(y) = BS(M + m)$, where $m \in \{0, 1, \dots, a_{N+1} - 1\}$. Then at least one return word to x in \mathbf{v} is longer than every return word to y in \mathbf{v} .*

Proof: On one hand, by Lemmas 19, 20 and Remark 21, the derived sequence $\mathbf{d}_{\mathbf{u}}(\mathcal{S}(y))$ is a standard Sturmian sequence with the directive sequence $G^{a_{N+1}-m} D^{a_{N+2}} G^{a_{N+3}} \dots$ or $D^{a_{N+1}-m} G^{a_{N+2}} D^{a_{N+3}} \dots$. It implies that \mathbf{u} is a concatenation of the blocks $r'^{\ell} s'$ and $r'^{\ell+1} s'$, where $\ell = a_{N+1} - m$. The return words to $\mathcal{S}(y)$ are by Proposition 30 of length $|r'| = p_N + q_N = q'_N$ and $|s'| = m q'_N + q'_{N-1}$. Regardless of (un)stability of the return words to $\mathcal{S}(y)$, the longest return word to y in \mathbf{v} is by Theorem 50 of length at most

$$(\ell + 1)|r'| + 2|s'| = (a_{N+1} - m + 1)q'_N + 2(m q'_N + q'_{N-1}) = q'_{N+1} + (m + 1)q'_N + q'_{N-1} \leq 2q'_{N+1}.$$

On the other hand, the return words r, s to $\mathcal{S}(x)$ in \mathbf{u} have by Proposition 30 either lengths $|r| = q'_{N+1}$ and $|s| = q'_{N+1} + q'_N$ (if $a_{N+2} > 1$), or lengths $|r| = q'_{N+1} + q'_N$ and $|s| = q'_{N+1}$ (if $a_{N+2} = 1$). Regardless of (un)stability of the return words to $\mathcal{S}(x)$, one of the return words to x in \mathbf{v} is of length at least

$$|r| + |s| = 2q'_{N+1} + q'_N.$$

□

Lemma 52. *Let \mathbf{v} be a CS Rote sequence and $\mathbf{u} = \mathcal{S}(\mathbf{v})$ be the associated standard Sturmian sequence with the directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \dots$ or $D^{a_1} G^{a_2} D^{a_3} G^{a_4} \dots$. Put $M = a_0 + a_1 + a_2 + \dots + a_N$, where $a_0 = 0$. Let x and y be the bispecial factors in \mathbf{v} such that $\mathcal{S}(x) = BS(M + a_{N+1} + 1)$ and $\mathcal{S}(y) = BS(M + a_{N+1})$. Then at least one return word to x in \mathbf{v} is at least as long as every return word to y in \mathbf{v} .*

Proof: Let us denote the return words to $\mathcal{S}(y)$ by r' and s' , and the return words to $\mathcal{S}(x)$ by r and s . By Proposition 30, $|r'| = q'_{N+1}$ and $|s'| = q'_N$.

First, we assume that r' is unstable. Then by Theorem 50, the return words to y in \mathbf{v} are of length at most $2|r'| + |s'| = 2q'_{N+1} + q'_N$. Regardless of (un)stability of the return words to $\mathcal{S}(x)$, one of the return words to x in \mathbf{v} is of length at least $|r| + |s| = 2q'_{N+1} + q'_N$.

It remains to discuss the case when r' is stable. Let us assume that \mathbf{u} has the directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \dots$. By Proposition 30, it means that $|r'|_1 = q_{N+1}$ is even. We use Theorem 50 to find the longest return word to y . Similarly as in the proof of Lemma 51, we determine that $\ell = a_{N+2}$ and thus the longest return word to y in \mathbf{v} is of length

$$L' = 2|s'| + (\ell + 1)|r'| = 2q'_N + (a_{N+2} + 1)q'_{N+1} = q'_{N+2} + q'_{N+1} + q'_N.$$

Let us compare L' with the length of the return words to x in \mathbf{v} . If $a_{N+2} > 1$, then by Proposition 30, $|r|_1 = |r'|_1 = q_{N+1}$ and r is stable as well. The longest return word to x is by Theorem 50 the return word $sr^{\ell+1}s$, where $\ell = a_{N+2} - 1$. Its length is

$$L = 2|s| + a_{N+2}|r| = 2(q'_{N+1} + q'_N) + a_{N+2}q'_{N+1} = q'_{N+2} + 2q'_{N+1} + q'_N > L'.$$

If $a_{N+2} = 1$, then by Proposition 30, $|r|_1 = q_{N+2} = q_{N+1} + q_N$ and $|s|_1 = q_{N+1}$. Since q_{N+1} is even, necessarily q_N is odd (as follows from the well-known relation $p_N q_{N+1} - p_{N+1} q_N = (-1)^{N+1}$ for all N). It means that r is unstable and s is stable. Thus the longest return word to x in \mathbf{v} has the length

$$L = 2|r| + |s| = 2(q'_{N+1} + q'_N) + q'_{N+1} = 2q'_{N+2} + q'_{N+1} > L'.$$

If \mathbf{u} has the directive sequence $D^{a_1} G^{a_2} D^{a_3} G^{a_4} \dots$, the proof is analogous, we only have to take into account that the coordinates of the Parikh vectors are exchanged (see Remark 31). In particular, instead of considering q_N when determining the number of ones, we consider p_N . \square

Proposition 53. *Let \mathbf{v} be a CS Rote sequence. Let \mathbf{u} be the standard Sturmian sequence such that $\mathcal{L}(\mathcal{S}(\mathbf{v})) = \mathcal{L}(\mathbf{u})$ and let \mathbf{u} have the directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \dots$ or $D^{a_1} G^{a_2} D^{a_3} G^{a_4} \dots$. Put $M = a_0 + a_1 + a_2 + \dots + a_N$, where $a_0 = 0$. Let L be the length of the longest return word to the bispecial factor v of \mathbf{v} such that $\mathcal{S}(v)$ is the bispecial factor $BS(M + a_{N+1} + 1)$ in \mathbf{u} . Then the recurrence function of \mathbf{v} satisfies $R_{\mathbf{v}}(n + 1) = L + n$ for any $n \in [q'_N, q'_{N+1})$, where $N \in \mathbb{N}$.*

Proof: Let \mathbf{v}' be a CS Rote sequence associated with the standard Sturmian sequence \mathbf{u} such that $\mathcal{L}(\mathcal{S}(\mathbf{v}')) = \mathcal{L}(\mathbf{u})$. Clearly, $\mathcal{L}(\mathbf{v}) = \mathcal{L}(\mathbf{v}')$. Since the recurrence function depends only on the language of the sequence and not on the sequence itself, we can work with the CS Rote sequence \mathbf{v}' instead of \mathbf{v} .

It follows from Lemma 41 that $R_{\mathbf{v}}(n + 1) = R_{\mathbf{v}'}(n + 1) = L + n$, where L is the length of the longest return word to a bispecial factor from the set $\mathcal{B}_{\mathbf{v}'}(n + 1)$. Lemma 42 shows the correspondence between the bispecial factors from the set $\mathcal{B}_{\mathbf{v}'}(n + 1)$ and the bispecial factors from the set $\mathcal{B}_{\mathbf{u}}(n)$. In particular, if $v \in \mathcal{B}_{\mathbf{v}'}(n + 1)$, then $\mathcal{S}(v) \in \mathcal{B}_{\mathbf{u}}(n)$. For every $n \in [q'_N, q'_{N+1})$, where $N \in \mathbb{N}$, the set $\mathcal{B}_{\mathbf{u}}(n)$ is described in Theorem 48. Together with Lemmas 51 and 52, it implies that the bispecial factor v such that $\mathcal{S}(v)$ equals $BS(M + a_{N+1} + 1)$ has the longest return word among all bispecial factors from the set $\mathcal{B}_{\mathbf{v}'}(n + 1)$. \square

Theorem 54. *Let \mathbf{v} be a CS Rote sequence and let \mathbf{u} be the standard Sturmian sequence such that $\mathcal{L}(\mathcal{S}(\mathbf{v})) = \mathcal{L}(\mathbf{u})$. If \mathbf{u} has the directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \dots$, then the value of the recurrence function $R_{\mathbf{v}}$ for $n \in [q'_N, q'_{N+1})$, $N \in \mathbb{N}$, is given by*

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

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

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

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

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

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

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

Proof: By Proposition 53, $R_{\mathbf{v}}(n+1) = L + n$, where L is the length of the longest return word to the bispecial factor v in \mathbf{v} such that $\mathcal{S}(v) = BS(M + a_{N+1} + 1)$. Consider first that \mathbf{u} has the directive sequence $G^{a_1} D^{a_2} G^{a_3} D^{a_4} \dots$. By Proposition 30, the Parikh vectors of the return words r and s to the bispecial factor $BS(M + a_{N+1} + 1)$ are

1. $\vec{V}(r) = \begin{pmatrix} p_{N+1} \\ q_{N+1} \end{pmatrix}$, $\vec{V}(s) = \begin{pmatrix} p_{N+1} + p_N \\ q_{N+1} + q_N \end{pmatrix}$ if $a_{N+2} > 1$;
2. $\vec{V}(r) = \begin{pmatrix} p_{N+1} + p_N \\ q_{N+1} + q_N \end{pmatrix}$, $\vec{V}(s) = \begin{pmatrix} p_{N+1} \\ q_{N+1} \end{pmatrix}$ if $a_{N+2} = 1$.

Let us emphasize that at most one of the numbers q_N and q_{N+1} is even. It follows from the well-known relation $p_N q_{N+1} - p_{N+1} q_N = (-1)^{N+1}$ for all N . Moreover, let us recall that $p_N + q_N = q'_N$.

First we discuss the case $a_{N+2} > 1$.

- If q_N is even, then q_{N+1} is odd, i.e., r and s are unstable. Since $|r| < |s|$, Item 3 of Theorem 50 gives $L = |rs| = 2q'_{N+1} + q'_N$.
- If q_{N+1} is even, then q_N is odd, i.e., r is stable and s is unstable. We use Item 1 of Theorem 50 with $\ell = a_{N+2} - 1$. Clearly, $L = 2|s| + (\ell + 1)|r| = 2(q'_{N+1} + q'_N) + a_{N+2}q'_{N+1} = q'_{N+2} + 2q'_{N+1} + q'_N$.
- If both q_N, q_{N+1} are odd, then r is unstable and s is stable. Item 2 of Theorem 50 implies $L = |rsr| = 3q'_{N+1} + q'_N$.

It remains to discuss the case $a_{N+2} = 1$.

- If q_N is even, then q_{N+1} is odd, i.e., r and s are unstable. Since $|r| > |s|$, Item 3 of Theorem 50 gives $L = |rr| = 2q'_{N+1} + 2q'_N = 2q'_{N+2}$.
- If q_{N+1} is even, then q_N is odd, i.e., r is unstable and s is stable. Item 2 of Theorem 50 implies $L = |rsr| = 3q'_{N+1} + 2q'_N = 2q'_{N+2} + q'_{N+1}$.
- If both q_N, q_{N+1} are odd, then r is stable and s is unstable. We use Item 1 of Theorem 50 with $\ell = a_{N+3}$. Thus $L = 2|s| + (\ell + 1)|r| = 2q'_{N+1} + (a_{N+3} + 1)(q'_{N+1} + q'_N) = q'_{N+3} + q'_{N+2} + q'_{N+1}$.

If \mathbf{u} has the directive sequence $D^{a_1} G^{a_2} D^{a_3} G^{a_4} \dots$, then the statement of Theorem 54 will stay the same, only q_N and q_{N+1} will be replaced by p_N and p_{N+1} because the Parikh vectors of r and s have the coordinates exchanged, see Remark 31. \square

Example 55. By Proposition 34, the critical exponent of the CS Rote sequence \mathbf{v} such that $\mathcal{S}(\mathbf{v})$ has the directive sequence $G(D^2G^2)^\omega$ is $\text{cr}(\mathbf{v}) = 2 + \frac{1}{\sqrt{2}}$. In Example 36, we have shown that the CS Rote sequence \mathbf{v}' associated to the Sturmian sequence $\mathcal{S}(\mathbf{v}') = E(\mathcal{S}(\mathbf{v}))$ has the critical exponent $\text{cr}(\mathbf{v}') = 4 + \frac{1}{1+\sqrt{2}}$.

Let us find an explicit formula for the recurrence function $R_{\mathbf{v}}$, resp. $R_{\mathbf{v}'}$ of the CS Rote sequence \mathbf{v} , resp. \mathbf{v}' . We will see that these recurrence functions differ essentially, too. In the proof of Proposition 34, we have shown that all q_N are odd and we have found an explicit formula for q'_N , see (7). Applying Theorem 54, we obtain for every $n \in [q'_N, q'_{N+1})$

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

Furthermore, p_N is even if and only if N is even. Therefore, we obtain for every $n \in [q'_{2N}, q'_{2N+1})$

$$R_{\mathbf{v}'}(n+1) = 2q'_{2N+1} + q'_{2N} + n = n + \frac{1}{2\sqrt{2}} \left((1 + \sqrt{2})^{2N+3} - (1 - \sqrt{2})^{2N+3} \right);$$

and for every $n \in [q'_{2N-1}, q'_{2N})$

$$R_{\mathbf{v}'}(n+1) = q'_{2N+1} + 2q'_{2N} + q'_{2N-1} + n = n + \frac{1}{\sqrt{2}} \left((1 + \sqrt{2})^{2N+2} - (1 - \sqrt{2})^{2N+2} \right).$$

References

- [1] L. Balková, E. Pelantová, and Š. Starosta, *Sturmian jungle (or garden?) on multiliteral alphabets*, RAIRO-Theor. Inf. Appl. 44 (2010), 443–470.
- [2] A. R. Baranwal and J. Shallit, *Critical exponent of infinite balanced words via the Pell number system*, in: R. Mercas and D. Reidenbach (eds.), Proceedings WORDS 2019, Lecture Notes in Computer Science, vol. 11682, Springer (2019), 80–92.
- [3] A. R. Baranwal and J. Shallit, *Repetitions in infinite palindrome-rich words*, in: R. Mercas and D. Reidenbach (eds.), Proceedings WORDS 2019, Lecture Notes in Computer Science, vol. 11682, Springer (2019), 93–105.
- [4] J. Berstel, *Axel Thue's papers on repetitions in words: a translation*, Publications du LaCIM 20, Université du Québec à Montréal (1995).
- [5] A. Blondin Massé, S. Brlek, A. Glen, and S. Labbé, *On the Critical Exponent of Generalized Thue-Morse Words*, Discret. Math. Theor. Comput. Sci. 9(1) (2007), 293–304.
- [6] A. Blondin Massé, S. Brlek, S. Labbé, and L. Vuillon, *Palindromic complexity of codings of rotations*, Theoret. Comput. Sci. 412(46) (2011), 6455–6463.
- [7] J. Cassaigne and N. Chekhova, *Fonctions de récurrence des suites d'Arnoux-Rauzy et réponse à une question de Morse et Hedlund*, Ann. Inst. Fourier (Grenoble) 56(7) (2006), 2249–2270.
- [8] J. D. Currie, L. Mol, and N. Rampersad, *The repetition threshold for binary rich words*, Discrete Math. Theoret. Comput. Sci. 22(1) (2020), no. 6.

- [9] D. Damanik and D. Lenz, *The index of Sturmian sequences*, European J. Combin. 23 (2002), 23–29.
- [10] X. Droubay, J. Justin, and G. Pirillo, *Episturmian words and some construction of de Luca and Rauzy*, Theor. Comput. Sci. 255 (2001), 539–553.
- [11] X. Droubay and G. Pirillo, *Palindromes and Sturmian words*, Theoret. Comput. Sci. 223 (1999), 73–85.
- [12] F. Durand, *A characterization of substitutive sequences using return words*, Discrete Math. 179 (1998), 89–101.
- [13] N. Ghareghani and P. Sharifani, *On square factors and critical factors of k -bonacci words on infinite alphabet*, preprint: arXiv:1912.05253 (2019), 26 pp.
- [14] D. Hensley, *Continued Fractions*, World Scientific Publishing (2006).
- [15] J. Justin and G. Pirillo, *Episturmian words and episturmian morphisms*, Theoret. Comput. Sci. 276 (2002), 281–313.
- [16] K. Klouda, K. Medková, E. Pelantová, and Š. Starosta, *Fixed points of Sturmian morphisms and their derivated words*, Theoret. Comput. Sci. 743 (2018), 23–37.
- [17] M. Lothaire, *Combinatorics on Words*, Encyclopaedia of Mathematics and its Applications, vol. 17, Addison-Wesley, Reading, Mass. (1983). Reprinted in the Cambridge Mathematical Library, Cambridge University Press (1997).
- [18] K. Medková, E. Pelantová, and L. Vuillon, *Derived sequences of complementary symmetric Rote sequences*, RAIRO-Theor. Inf. Appl. 53 (2019), 125–151.
- [19] F. Mignosi and G. Pirillo, *Repetitions in the Fibonacci words*, RAIRO-Theor. Inf. Appl. 26 (1992), 199–204.
- [20] M. Morse and G. A. Hedlund, *Symbolic dynamics II. Sturmian trajectories*, Amer. J. Math. 62 (1940), 1–42.
- [21] E. Pelantová and Š. Starosta, *Languages invariant under more symmetries: overlapping factors versus palindromic richness*, Discrete Math. 313 (2013), 2432–2445.
- [22] N. Rampersad, J. Shallit, and É. Vandomme, *Critical exponents of infinite balanced words*, Theor. Comp. Sci. 777 (2019), 454–463.
- [23] M. Rigo, *Formal Languages, Automata and Numeration Systems*, 2 vols., Wiley (2014).
- [24] G. Rote, *Sequences with subword complexity $2n$* , J. Number Theory 46 (1994), 196–213.
- [25] A. Thue, *Probleme über Veränderungen von Zeichenreihen nach gegebene Regeln*, Christiana Videnskabs-Selskabs Skrifter, I. Math.-naturv. Klasse 10 (1914).
- [26] L. Vuillon, *A characterization of Sturmian words by return words*, Eur. J. Combin. 22 (2001), 263–275.