On balanced sequences and their asymptotic critical exponent*

Francesco Dolce[1], Ľubomíra Dvořáková[1], and Edita Pelantová[1]

FNSPE, Czech Technical University in Prague, Czech Republic
{francesco.dolce, lubomira.dvorakova, edita.pelantova}@fjfi.cvut.cz

Abstract. We study aperiodic balanced sequences over finite alphabets. A sequence $\mathbf{v}$ of this type is fully characterised by a Sturmian sequence $\mathbf{u}$ and two constant gap sequences $\mathbf{y}$ and $\mathbf{y}'$. We study the language of $\mathbf{v}$, with focus on return words to its factors. We provide a uniform lower bound on the asymptotic critical exponent of all sequences $\mathbf{v}$ arising by $\mathbf{y}$ and $\mathbf{y}'$. It is a counterpart to the upper bound on the least critical exponent of $\mathbf{v}$ conjectured and partially proved recently in works of Baranwal, Rampersad, Shallit and Vandomme. We deduce a method computing the exact value of the asymptotic critical exponent of $\mathbf{v}$ provided the associated Sturmian sequence $\mathbf{u}$ has a quadratic slope. The method is used to compare the critical and the asymptotic critical exponent of balanced sequences over an alphabet of size $d \leq 10$ which are conjectured by Rampersad et al. to have the least critical exponent.

Keywords: balanced sequence · critical exponent · Sturmian sequence · return word · bispecial factors

1 Introduction

An infinite sequence over a finite alphabet is balanced if, for any two of its factors u and v of the same length, the number of occurrences of each letter in u and v differs by at most 1. Over a binary alphabet aperiodic balanced sequences coincide with Sturmian sequences, as shown by Hedlund and Morse [13]. Hubert [14] provided a construction of balanced sequences. It consists in colouring of entries of a Sturmian sequence $\mathbf{u}$ by two constant gap sequences $\mathbf{y}$ and $\mathbf{y}'$. In this paper we study combinatorial properties of balanced sequences. We first show that such sequences belong to the class of eventually dendric sequences introduced in [5]. We give formulæ for the factor complexity and the number of return words to each factor. The main goal of this paper is to develop a method computing the asymptotic critical exponent of a given balanced sequence. Our work can be understood as a continuation of research on balanced sequences with the least

* The research received funding from the project CZ.02.1.01/0.0/0.0/16_019/0000778. We would like to thank Daniela Opočenská for her careful and readily usable implementation of our program computing the asymptotic critical exponent.

critical exponent initiated by Rampersad, Shallit and Vandomme [19]. The relation between factor complexity and critical exponent for binary and ternary sequences was studied as well in [20]

Finding the least critical exponent of sequences is a classical problem. The answer is known as Dejean's conjecture [8], and the proof was provided step by step by many people. The least critical exponent was determined also for some particular classes of sequences: by Carpi and de Luca [6] for Sturmian sequences, and by Currie, Mol and Rampersad [7] for binary rich sequences. Recently, Rampersad, Shallit and Vandomme [19] found balanced sequences with the least critical exponent over alphabets of size 3 and 4 and also conjectured that the least critical exponent of balanced sequences over a d-letter alphabet with $d \geq 5$ is $\frac{d-2}{d-3}$. Their conjecture was confirmed for $d \leq 8$ [3,4].

Here we focus on the asymptotic critical exponent of balanced sequences. We show that the asymptotic critical exponent depends on the slope of the associated Sturmian sequence and, unlike the critical exponent, on the length of the minimal periods of $\mathbf{y}$ and $\mathbf{y}'$, but not on $\mathbf{y}$ and $\mathbf{y}'$ themselves. We also give a lower bound on the asymptotic critical exponent. We provide an algorithm computing the exact value of the asymptotic critical exponent for balanced sequences originated in Sturmian sequences with a quadratic slope (in this case the continued fraction of the slope is eventually periodic).

2 Preliminaries

An *alphabet* $\mathcal{A}$ is a finite set of symbols called *letters*. A *word* over $\mathcal{A}$ of *length* n is a string $u = u_0u_1 \cdots u_{n-1}$, where $u_i \in \mathcal{A}$ for all $i \in \{0, 1, \ldots, n-1\}$. The length of u is denoted by $|u|$. The set of all finite words over $\mathcal{A}$ together with the operation of concatenation forms a monoid, denoted $\mathcal{A}^*$. 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* $\boldsymbol{V}(u) \in \mathbb{N}^d$ defined as $(\boldsymbol{V}(u))_a = |u|_a$ 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 $i \in \mathbb{N}$. A sequence $\mathbf{u}$ is *eventually periodic* if $\mathbf{u} = vwww \cdots = v(w)^\omega$ for some $v \in \mathcal{A}^*$ and $w \in \mathcal{A}^+$. It is periodic if $\mathbf{u} = w^\omega$. If $\mathbf{u}$ is not eventually periodic, then it is *aperiodic*. A *factor* of $\mathbf{u} = u_0u_1u_2 \cdots$ is a word y such that $y = u_iu_{i+1}u_{i+2} \cdots u_{j-1}$ for some $i, j \in \mathbb{N}$, $i \leq j$. We usually denote $y = \mathbf{u}_{[i,j)}$. 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 a sequence $\mathbf{u}$ is the set of all its factors. We also define $\mathcal{L}(\mathbf{u})^+ = \mathcal{L}(\mathbf{u}) \setminus \{\varepsilon\}$. A factor w of $\mathbf{u}$ is *right special* if wa, wb are in $\mathcal{L}(\mathbf{u})$ 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. The *factor complexity* of a sequence $\mathbf{u}$ is the mapping $\mathcal{C}_{\mathbf{u}} : \mathbb{N} \to \mathbb{N}$ defined by $\mathcal{C}_{\mathbf{u}}(n) = \#\{w \in \mathcal{L}(\mathbf{u}) : |w| = n\}$. The first difference of the factor complexity is $s_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n)$. Aperiodic sequences with the lowest possible factor complexity, i.e., such that $\mathcal{C}_{\mathbf{u}}(n) = n + 1$ for all $n \in \mathbb{N}$, are called *Sturmian sequences* (for other equivalent definitions see [2]). Clearly, all Sturmian sequences are defined over a binary alphabet, e.g., $\{\mathtt{a}, \mathtt{b}\}$. If both sequences $\mathtt{a}\mathbf{u}$ and $\mathtt{b}\mathbf{u}$ 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. For other facts about Sturmian sequences see [18].

A sequence $\mathbf{u}$ over the alphabet $\mathcal{A}$ is *balanced* if for every letter $a \in \mathcal{A}$ and every pair of factors $u, v \in \mathcal{L}(\mathbf{u})$ with $|u| = |v|$, we have $||u|_a - |v|_a| \leq 1$. The class of Sturmian sequences and the class of aperiodic balanced sequences coincide over a binary alphabet (see [13]). Vuillon [22] provides a survey on some previous work on balanced sequences.

A *morphism* over $\mathcal{A}$ is a mapping $\psi : \mathcal{A}^* \to \mathcal{A}^*$ such that $\psi(uv) = \psi(u)\psi(v)$ for all $u, v \in \mathcal{A}^*$. The morphism ψ can be naturally extended to sequences by setting $\psi(u_0u_1u_2\cdots) = \psi(u_0)\psi(u_1)\psi(u_2)\cdots$.

Consider a factor w of a recurrent sequence $\mathbf{u} = u_0u_1u_2\cdots$. Let $i < j$ be two consecutive occurrences of w in $\mathbf{u}$. Then the word $u_iu_{i+1}\cdots u_{j-1}$ is a *return word* to w in $\mathbf{u}$. The set of all return words to w in $\mathbf{u}$ is denoted by $\mathcal{R}_{\mathbf{u}}(w)$. If $\mathbf{u}$ is uniformly recurrent, the set $\mathcal{R}_{\mathbf{u}}(w)$ is finite for each prefix w. The opposite is true if $\mathbf{u}$ is recurrent. In this case $\mathbf{u}$ can be written as a concatenation $\mathbf{u} = r_{d_0}r_{d_1}r_{d_2}\cdots$ of return words to w. The *derived sequence* of $\mathbf{u}$ to w is the sequence $\mathbf{d}_{\mathbf{u}}(w) = d_0d_1d_2\cdots$ over the alphabet of cardinality $\#\mathcal{R}_{\mathbf{u}}(w)$. The concept of derived sequences was introduced by Durand [11].

Given a sequence $\mathbf{u}$ over an alphabet $\mathcal{A}$ and $w \in \mathcal{L}(\mathbf{u})$, we define the sets of left extensions, right extensions and bi-extensions of w in $\mathcal{L}(\mathbf{u})$ respectively as $L_{\mathbf{u}}(w) = \{a \in \mathcal{A} : aw \in \mathcal{L}(\mathbf{u})\}$, $R_{\mathbf{u}}(w) = \{b \in \mathcal{A} : wb \in \mathcal{L}(\mathbf{u})\}$ and $B_{\mathbf{u}}(w) = \{(a,b) \in \mathcal{A} \times \mathcal{A} : awb \in \mathcal{L}(\mathbf{u})\}$. The *extension graph* of w in $\mathcal{L}(\mathbf{u})$, denoted $\mathcal{E}_{\mathbf{u}}(w)$, is the undirected bipartite graph whose set of vertices is the disjoint union of $L_{\mathbf{u}}(w)$ and $R_{\mathbf{u}}(w)$ and with edges the elements of $B_{\mathbf{u}}(w)$. A sequence $\mathbf{u}$ (resp. a language $\mathcal{L}(\mathbf{u})$) is said to be *eventually dendric* with *threshold* $m \geq 0$ if $\mathcal{E}_{\mathbf{u}}(w)$ is a tree for every word $w \in \mathcal{L}(\mathbf{u})$ of length at least m. It is said to be *dendric* if we can choose $m = 0$. Dendric languages were introduced in [5] under the name of tree sets. It is known that Sturmian sequences are dendric.

Example 1. It is known that the sequence $\mathbf{u}_f = \mathtt{abaababaabaababaababaa}\cdots$, obtained as fixed point of the morphism $f : \mathtt{a} \mapsto \mathtt{ab}, \mathtt{b} \mapsto \mathtt{a}$, is Sturmian (see [18]).

3 Languages of balanced sequences

In 2000, Hubert [14] characterised balanced sequences over alphabets of higher cardinality. A suitable tool for their description is the notion of constant gap.

Definition 1. *A sequence* $\mathbf{y}$ *over an alphabet* $\mathcal{A}$ *is a* constant gap sequence *if for each letter* $a \in \mathcal{A}$ *appearing in* $\mathbf{y}$ *there is a positive integer* d *such that the distance between successive occurrences of* a *in* $\mathbf{y}$ *is always* d.

Obviously, any constant gap sequence is periodic. Given a constant gap sequence $\mathbf{y}$, we denote its minimal period length by $\mathrm{Per}(\mathbf{y})$.

Example 2. The sequence $\mathbf{y} = (\mathtt{0102})^\omega$ is a constant gap sequence because the distance between consecutive $\mathtt{0}$s is always 2, while the distance between consecutive $\mathtt{1}$s (resp. $\mathtt{2}$s) is always 4. Its minimal period is $\mathrm{Per}(\mathbf{y}) = 4$.

The sequence $(\mathtt{011})^\omega$ is periodic but it is not a constant gap sequence.

The i-th *shift* of a constant gap sequence $\mathbf{y} = (y_0 y_1 \cdots y_{k-1})^\omega$ with minimal period $k \geq 1$ (and $0 \leq i < k$) is the sequence $\sigma^i(\mathbf{y}) = (y_i \cdots y_{k-1} y_0 \cdots y_{i-1})^\omega$.

Example 3. Let $\mathbf{y}$ be the sequence seen in Example 2. Then we have $\sigma^0(\mathbf{y}) = \mathbf{y}$, $\sigma(\mathbf{y}) = (\mathtt{1020})^\omega$, $\sigma^2(\mathbf{y}) = (\mathtt{0201})^\omega$ and $\sigma^3(\mathbf{y}) = (\mathtt{2010})^\omega$.

Theorem 1 ([14]). *A recurrent aperiodic sequence* $\mathbf{v}$ *is balanced if and only if* $\mathbf{v}$ *is obtained from a Sturmian sequence* $\mathbf{u}$ *over* $\{\mathtt{a}, \mathtt{b}\}$ *by replacing the* $\mathtt{a}$*s in* $\mathbf{u}$ *by a constant gap sequence* $\mathbf{y}$ *over some alphabet* $\mathcal{A}$*, and replacing the* $\mathtt{b}$*s in* $\mathbf{u}$ *by a constant gap sequence* $\mathbf{y}'$ *over some alphabet* $\mathcal{B}$ *disjoint from* $\mathcal{A}$.

Definition 2. *Let* $\mathbf{u}$ *be a Sturmian sequence over the alphabet* $\{\mathtt{a}, \mathtt{b}\}$*, and* $\mathbf{y}, \mathbf{y}'$ *be two constant gap sequences over two disjoint alphabets* $\mathcal{A}$ *and* $\mathcal{B}$*. The* colouring *of* $\mathbf{u}$ *by* $\mathbf{y}$ *and* $\mathbf{y}'$*, denoted* $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$*, is the sequence over* $\mathcal{A} \cup \mathcal{B}$ *obtained by the procedure described in Theorem 1.*

For $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ we use the notation $\pi(\mathbf{v}) = \mathbf{u}$ and $\pi(v) = u$ for any $v \in \mathcal{L}(\mathbf{v})$ and the corresponding $u \in \mathcal{L}(\mathbf{u})$. Symmetrically, given a word $u \in \mathcal{L}(\mathbf{u})$, we denote by $\pi^{-1}(u) = \{v \in \mathcal{L}(\mathbf{v}) : \pi(v) = u\}$. We say that $\mathbf{u}$ (resp. u) is a *projection* of $\mathbf{v}$ (resp. v). The map $\pi : \mathcal{L}(\mathbf{v}) \to \mathcal{L}(\mathbf{u})$ is clearly a morphism.

Example 4. Let $\mathbf{u}_f$ be as in Example 1. Let us take the constant gap sequences $\mathbf{y} = (\mathtt{0102})^\omega$ and $\mathbf{y}' = (\mathtt{34})^\omega$ over the alphabets $\mathcal{A} = \{\mathtt{0}, \mathtt{1}, \mathtt{2}\}$ and $\mathcal{B} = \{\mathtt{3}, \mathtt{4}\}$ respectively. The sequence $\mathbf{v}_f = \mathrm{colour}(\mathbf{u}_\mathrm{f}, \mathbf{y}, \mathbf{y}') = \mathtt{0310423014023041032401} \cdots$ is balanced according to Theorem 1. One has $\pi(\mathbf{v}_f) = \mathbf{u}_f$. Moreover, $\pi(\mathtt{031}) = \pi(\mathtt{041}) = \mathtt{aba}$, and $\pi^{-1}(\mathtt{aba}) = \{\mathtt{031}, \mathtt{032}, \mathtt{041}, \mathtt{042}, \mathtt{130}, \mathtt{140}, \mathtt{230}, \mathtt{240}\}$.

Definition 3. *An aperiodic sequence* $\mathbf{u}$ *over* $\{\mathtt{a}, \mathtt{b}\}$ *has* well distributed occurrences*, or has the* WDO property*, if for every* $m \in \mathbb{N}$ *and for every* $w \in \mathcal{L}(\mathbf{u})$ *one has* $\{\boldsymbol{V}(p) \bmod m : pw \text{ is a prefix of } \mathbf{u}\} = \mathbb{Z}_m^2$.

It is known that Sturmian sequences have the WDO property (see [1]).

Example 5. Let $\mathbf{u}_f$ be as in Example 1 and let us consider $m = 2$ and $w = \mathtt{ab} \in \mathcal{L}(\mathbf{u}_f)$. Then it is easy to check that $\boldsymbol{V}(\varepsilon) \equiv \binom{0}{0} \mod 2$, $\boldsymbol{V}(\mathtt{aba}) \equiv \binom{0}{1} \mod 2$, $\boldsymbol{V}(\mathtt{abaab}) \equiv \binom{1}{0} \mod 2$ and $\boldsymbol{V}(\mathtt{abaababa}) \equiv \binom{1}{1} \mod 2$, where $w, \mathtt{aba}w, \mathtt{abaab}w$ and $\mathtt{abaababa}w$ are prefixes of $\mathbf{u}_f$.

Using the WDO property we can prove that, to study the language of aperiodic recurrent balanced sequences, it is enough to study standard Sturmian sequences.

Proposition 1. *Let* $\mathbf{u}, \mathbf{u}'$ *be two Sturmian sequences such that* $\mathcal{L}(\mathbf{u}) = \mathcal{L}(\mathbf{u}')$, $\mathbf{y}$ *and* $\mathbf{y}'$ *two constant gap sequences over disjoint alphabets and* $i, j \in \mathbb{N}$. *Let* $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$, $\mathbf{v}' = \mathrm{colour}(\mathbf{u}', \mathbf{y}, \mathbf{y}')$ *and* $\mathbf{v}'' = \mathrm{colour}(\mathbf{u}, \sigma^{\mathrm{i}}(\mathbf{y}), \sigma^{\mathrm{j}}(\mathbf{y}'))$. *Then* $\mathcal{L}(\mathbf{v}) = \mathcal{L}(\mathbf{v}') = \mathcal{L}(\mathbf{v}'')$.

Proof. Let $v \in \mathcal{L}(\mathbf{v})$ and w such that wv is a prefix of $\mathbf{v}$. Then $\pi(w)\pi(v)$ is a prefix of $\mathbf{u}$ and $|\pi(w)|$ is an occurrence of $\pi(v)$ in $\mathbf{u}$. Since $\pi(v) \in \mathcal{L}(\mathbf{u}')$, using the WDO property, we can find $p \in \mathcal{L}(\mathbf{u}')$ such that $p\pi(v)$ is a prefix of $\mathbf{u}'$ and $\boldsymbol{V}(\pi(w)) = \boldsymbol{V}(p) \bmod \mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')$. Thus v appears both in $\mathbf{v}$ at occurrence $|\pi(w)|$ and in $\mathbf{v}'$ at occurrence $|p|$. Hence $\mathcal{L}(\mathbf{v}) \subset \mathcal{L}(\mathbf{v}')$. Using the same argument we can prove the opposite inclusion.

Let p be a prefix of $\mathbf{u}$ such that $\boldsymbol{V}(p) = \binom{i}{j} \bmod \mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')$. Denote $\mathbf{u}'' = p^{-1}\mathbf{u}$. Then $\mathrm{colour}(\mathbf{u}'', \sigma^{\mathrm{i}}(\mathbf{y}), \sigma^{\mathrm{j}}(\mathbf{y}'))$ gives the same sequence as the one obtained by erasing the prefix of length $|p|$ from $\mathbf{v}$. Since $\mathcal{L}(\mathbf{u}) = \mathcal{L}(\mathbf{u}'')$, using the same argument as before we have $\mathcal{L}(\mathbf{v}'') = \mathcal{L}(\mathbf{v})$.

Corollary 1. *Let* $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ *and* $v \in \mathcal{L}(\mathbf{v})$. *For any* i, j *such that* $0 \le i < \mathrm{Per}(\mathbf{y})$ *and* $0 \le j < \mathrm{Per}(\mathbf{y}')$, *the word* v' *obtained from* $\pi(v)$ *by replacing the* $\mathtt{a}$*s by* $\sigma^i(\mathbf{y})$ *and the* $\mathtt{b}$*s by* $\sigma^j(\mathbf{y}')$ *is in* $\pi^{-1}(\pi(v))$, *and thus in* $\mathcal{L}(\mathbf{v})$.

Example 6. Let $\mathbf{u}_f, \mathbf{v}_f, \mathbf{y}$ and $\mathbf{y}'$ be as in Example 4. Let $v = \mathtt{03104} \in \mathcal{L}(\mathbf{v}_f)$ and let us denote $u = \pi(v) = \mathtt{abaab}$. One can easily check that the word $v' = \mathtt{24013}$ obtained from u by replacing the $\mathtt{a}$s by $\sigma^3(\mathbf{y})$ and the $\mathtt{b}$s by $\sigma(\mathbf{y}')$ is in $\mathcal{L}(\mathbf{v}_f)$.

Note that, if $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$, there exists an $m \in \mathbb{N}$ such that every factor $v \in \mathcal{L}(\mathbf{v})$ longer than m contains at least $\mathrm{Per}(\mathbf{y})$ letters in $\mathcal{A}$ and at least $\mathrm{Per}(\mathbf{y}')$ letters in $\mathcal{B}$. Indeed, it is enough to find m such that all factors of length m in $\mathcal{L}(\mathbf{u})$ contain at least $\mathrm{Per}(\mathbf{y})$ $\mathtt{a}$s and at least $\mathrm{Per}(\mathbf{y}')$ $\mathtt{b}$s.

Example 7. Let $\mathbf{u}_f, \mathbf{v}_f, \mathbf{y}$ and $\mathbf{y}'$ be as in Example 4. Then, it easy to check that all factors of length 7 in $\mathcal{L}(\mathbf{u}_f)$ contain at least 4 $\mathtt{a}$s and 2 $\mathtt{b}$s. Thus, all factors of length 7 in $\mathcal{L}(\mathbf{v}_f)$ contain at least four letters in $\mathcal{A}$ and at least two letters in $\mathcal{B}$. On the other hand, $\mathtt{babaab} \in \mathcal{L}(\mathbf{u}_f)$ has length 6 and contains only three $\mathtt{a}$s.

As we saw in Example 4, the set $\pi^{-1}(u)$, for a word $u \in \mathcal{L}(\mathbf{u})$, is not in general a singleton. However, it is not difficult to prove that any long enough factor in $\mathbf{v}$ is uniquely determined, between the words having the same projection in $\mathbf{u}$, by the first $\mathrm{Per}(\mathbf{y})$ letters in $\mathcal{A}$ and the first $\mathrm{Per}(\mathbf{y}')$ letters in $\mathcal{B}$ (the number of needed letters can be reduced by studying the bispecial factors in $\mathbf{y}$ and $\mathbf{y}'$).

Lemma 1. *Let* $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ *and* $u \in \mathcal{L}(\mathbf{u})$ *such that* $|u|_{\mathtt{a}} \ge \mathrm{Per}(\mathbf{y})$ *and* $|u|_{\mathtt{b}} \ge \mathrm{Per}(\mathbf{y}')$. *Let* $a_0, a_1, \ldots, a_{\mathrm{Per}(\mathbf{y})-1} \in \mathcal{A}$, *and* $b_0, b_1, \ldots, b_{\mathrm{Per}(\mathbf{y}')-1} \in \mathcal{B}$. *There exists at most one word in* $\pi^{-1}(u)$ *having* $a_0, a_1, \ldots, a_{\mathrm{Per}(\mathbf{y})} - 1$ *(in this order) as first letters in* $\mathcal{A}$ *and* $b_0, b_1, \ldots, b_{\mathrm{Per}(\mathbf{y}')}$ *(in this order) as first letters in* $\mathcal{B}$.

Example 8. Let $\mathbf{u}_f, \mathbf{v}_f, \mathbf{y}$ and $\mathbf{y}'$ be as in Example 4 and $u = \mathtt{abaabaaba} \in \mathcal{L}(\mathbf{u}_f)$. One has $|u|_\mathtt{a} = 6 > \mathrm{Per}(\mathbf{y})$ and $|u|_\mathtt{b} = 3 > \mathrm{Per}(\mathbf{y}')$. One can check that the only word in $\pi^{-1}(u)$ having $\mathtt{0}, \mathtt{2}, \mathtt{0}, \mathtt{1}$ as first letters in $\mathcal{A}$ and $\mathtt{4}, \mathtt{3}$ as first letters in $\mathcal{B}$ is $\mathtt{042031042}$, that is the word obtained from u by $\sigma^2(\mathbf{y})$ and $\sigma(\mathbf{y}')$. On the other hand, no word in $\mathcal{L}(\mathbf{v})$ can have $\mathtt{0}, \mathtt{0}, \mathtt{1}, \mathtt{2}$ (in this order) as first letters in $\mathcal{A}$ or $\mathtt{3}, \mathtt{3}$ as first letters in $\mathcal{B}$.

Putting together Corollary 1 and Lemma 1, we obtain the following result.

Lemma 2. *Let* $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ *and* $u \in \mathcal{L}(\mathbf{u})$ *be such that* $|u|_\mathtt{a} \geq \mathrm{Per}(\mathbf{y})$ *and* $|u|_\mathtt{b} \geq \mathrm{Per}(\mathbf{y}')$. *Then* $\#(\pi^{-1}(u)) = \mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')$.

Example 9. Let $\mathbf{u}_f, \mathbf{v}_f, \mathbf{y}, \mathbf{y}'$ be as in Example 4 and $u = \mathtt{abaabaab} \in \mathcal{L}(\mathbf{u}_f)$. The set $\pi^{-1}(u) = \{$ 03104203, 03204103, 04103204, 04203104, 13024013, 14023014, 23014023, 24013024$\}$ has exactly 8 elements, according to Lemma 2.

The following result easily follows from Lemma 1 and the WDO property.

Lemma 3. *Let* $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ *and* $u \in \mathcal{L}(\mathbf{u})$ *be such that* $|u|_\mathtt{a} \geq \mathrm{Per}(\mathbf{y})$ *and* $|u|_\mathtt{b} \geq \mathrm{Per}(\mathbf{y}')$. *Let* $v \in \pi^{-1}(u)$. *Then v is right special (resp. left special) if and only if u is right special (resp. left special). Moreover, in this case the unique two right (resp. left) extensions of v belong to different alphabets $\mathcal{A}$ and $\mathcal{B}$.*

Proposition 2. *The language* $\mathcal{L}(\mathbf{v})$ *is eventually dendric.*

Proof. Let m be a positive integer such that for every word $w \in \mathcal{L}(\mathbf{u})$ of length at least m one has $|w|_\mathtt{a} \geq \mathrm{Per}(\mathbf{y})$ and $|w|_\mathtt{b} \geq \mathrm{Per}(\mathbf{y}')$. Let $v \in \mathcal{L}(\mathbf{v})$ and $u = \pi(v)$, and suppose that $|v| \geq m$. It easily follows from Lemmata 1 and 3 that $\mathcal{E}_\mathbf{v}(u)$ is isomorphic to $\mathcal{E}_\mathbf{u}(u)$ via the projection π. Since $\mathbf{u}$ is Sturmian, then $\mathcal{L}(\mathbf{u})$ is dendric. Thus $\mathcal{E}_\mathbf{v}(v)$ is a tree. Hence $\mathcal{L}(\mathbf{v})$ is eventually dendric of threshold m.

The following result easily follows from Lemma 2.

Proposition 3. *Let* $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ *and m be a positive integer such that every word in* $\mathcal{L}(\mathbf{u})$ *of length m has at least* $\mathrm{Per}(\mathbf{y})$ $\mathtt{a}$*s and at least* $\mathrm{Per}(\mathbf{y}')$ $\mathtt{b}$*s. Then for any* $n \geq m$ *one has* $\mathcal{C}_\mathbf{v}(n) = \mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')(n+1)$.

Example 10. Let $\mathbf{u}_f, \mathbf{v}_f, \mathbf{y}$ and $\mathbf{y}'$ be as in Example 4. The language $\mathcal{L}(\mathbf{v}_f)$ is eventually dendric with threshold 7. The factor complexity of $\mathbf{v}_f$ is defined by $\mathcal{C}_{\mathbf{v}_f}(n) = 8(n+1)$ for every $n \geq 7$, according to Proposition 3.

Proposition 4. *Let* $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ *and* $v \in \mathcal{L}(\mathbf{v})$ *such that* $|\pi(v)|_\mathtt{a} \geq \mathrm{Per}(\mathbf{y})$ *and* $|\pi(v)|_\mathtt{b} \geq \mathrm{Per}(\mathbf{y}')$. *Then* $\#(\mathcal{R}_\mathbf{v}(v)) = 1 + \mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')$.

Proof. From Proposition 3 we have $s_\mathbf{v}(n) = \mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')$ for every n large enough. The result thus follows from Proposition 2 and [10, Theorem 7.3].

Corollary 2. *A recurrent aperiodic balanced sequence is uniformly recurrent.*

Proof. A recurrent language is uniformly recurrent if and only if the number of return words to a given word in the language is finite. The result then just follows from Proposition 4 and [10, Theorem 7.3].

Given a vector $b = \binom{b_1}{b_2} \in \mathbb{N}^2$ and two periodic sequences $\mathbf{y}, \mathbf{y}'$, we use the notation $b \bmod \boldsymbol{Per}(\mathbf{y}, \mathbf{y}') := \begin{pmatrix} b_1 \bmod \mathrm{Per}(\mathbf{y}) \\ b_2 \bmod \mathrm{Per}(\mathbf{y}') \end{pmatrix}$.

Lemma 4. *Let $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$, $u \in \mathcal{L}(\mathbf{u})$ with $|u|_{\mathtt{a}} \geq \mathrm{Per}(\mathbf{y})$, $|u|_{\mathtt{b}} \geq \mathrm{Per}(\mathbf{y}')$ and $v, w \in \mathcal{L}(\mathbf{v})$ such that $\pi(v) = \pi(w) = u$. Let i, j be occurrences of v and w in $\mathbf{v}$ respectively and let us assume that $i < j$. Then $v = w$ if and only if $\boldsymbol{V}(\mathbf{u}_{[i,j)}) = \binom{0}{0} \mod \boldsymbol{Per}(\mathbf{y}, \mathbf{y}')$.*

Proof. By Lemma 1, $v = w$ if and only if there exist $0 \leq s < \mathrm{Per}(\mathbf{y})$ and $0 \leq t < \mathrm{Per}(\mathbf{y}')$ such that both v and w are obtained from u by replacing the $\mathtt{a}$s by $\sigma^s(\mathbf{y})$ and the $\mathtt{b}$s by $\sigma^t(\mathbf{y}')$. Furthermore, in this case we have $\boldsymbol{V}(\mathbf{u}_{[0,i)}) = \boldsymbol{V}(\mathbf{u}_{[0,j)}) \mod \boldsymbol{Per}(\mathbf{y}, \mathbf{y}')$, that is $\boldsymbol{V}(\mathbf{u}_{[i,j)}) = \binom{0}{0} \mod \boldsymbol{Per}(\mathbf{y}, \mathbf{y}')$.

Lemma 5. *Let $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$, $u \in \mathcal{L}(\mathbf{u})$ with $|u|_{\mathtt{a}} \geq \mathrm{Per}(\mathbf{y})$, $|u|_{\mathtt{b}} \geq \mathrm{Per}(\mathbf{y}')$ and $v, w \in \mathcal{L}(\mathbf{v})$ with $\pi(v) = \pi(w) = u$. Then $\pi(\mathcal{R}_{\mathbf{v}}(v)) = \pi(\mathcal{R}_{\mathbf{v}}(w))$.*

Proof. Let $r \in \mathcal{R}_{\mathbf{v}}(v)$. Then u is both a prefix and a suffix of $\pi(rv)$. By Lemma 1 there exist a unique $0 \leq s < \mathrm{Per}(\mathbf{y})$ and a unique $0 \leq t < \mathrm{Per}(\mathbf{y}')$ such that w is obtained from u by replacing the $\mathtt{a}$s by $\sigma^s(\mathbf{y})$ and the $\mathtt{b}$s by $\sigma^t(\mathbf{y}')$. By Corollary 1 the word obtained from $\pi(rv)$ by replacing the $\mathtt{a}$s by $\sigma^s(\mathbf{y})$ and the $\mathtt{b}$s by $\sigma^t(\mathbf{y}')$ is in $\mathcal{L}(\mathbf{v})$ and has w as a prefix. This factor is equal to $r'w$ and it contains only two occurrences of w. Indeed, it follows from Lemma 4 that $\pi(r') = \pi(r)$ is the unique non-empty prefix of $\pi(rv)$ satisfying $\boldsymbol{V}(\pi(r)) = \binom{0}{0} \mod \boldsymbol{Per}(\mathbf{y}, \mathbf{y}')$. Thus, r' is a return word to w with $\pi(r') = \pi(r)$, which implies $\pi(\mathcal{R}_{\mathbf{v}}(v)) \subset \pi(\mathcal{R}_{\mathbf{v}}(w))$. The opposite inclusion can be proved symmetrically.

4 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 *fractional root* u and the *exponent* $e = |z|/|u|$. We usually write $z = u^e$. Let us emphasise that a word z can have multiple exponents and fractional roots.

Definition 4. *Given a sequence $\mathbf{u}$ and $u \in \mathcal{L}(\mathbf{u})^+$, we define the* index *of u in $\mathbf{u}$ as $ind_{\mathbf{u}}(u) = \sup\{e \in \mathbb{Q} : u^e \in \mathcal{L}(\mathbf{u})\}$. The* critical exponent *of a sequence $\mathbf{u}$ is defined as $E(\mathbf{u}) = \sup\{ind_{\mathbf{u}}(u) : u \in \mathcal{L}(\mathbf{u})^+\}$. Its* asymptotic critical exponent *is defined as $E^*(\mathbf{u}) = \lim_{n\to\infty} (\sup\{ind_{\mathbf{u}}(u) : u \in \mathcal{L}(\mathbf{u}),\ |u| \geq n\})$.*

Clearly, $\mathrm{E}(\mathbf{u}) \geq \mathrm{E}^*(\mathbf{u})$. If $\mathbf{u}$ is eventually periodic, then both $\mathrm{E}(\mathbf{u})$ and $\mathrm{E}^*(\mathbf{u})$ are infinite. If $\mathbf{u}$ is aperiodic and uniformly recurrent, then each factor of $\mathbf{u}$ has finite index. Nevertheless, $\mathrm{E}^*(\mathbf{u})$ may be infinite. An example of such a sequence is given by Sturmian sequences whose continued fraction expansions of their slope have unbounded partial quotients (see [9]).

Lemma 6. *Let u, w be non-empty factors of a recurrent sequence* $\mathbf{u}$. *If $u \in \mathcal{R}_{\mathbf{u}}(w)$, then $w = u^e$ for some $e \in \mathbb{Q}$. Moreover, if* $\mathbf{u}$ *is aperiodic and uniformly recurrent, then u is a return word to a finite number of factors in* $\mathbf{u}$.

Proof. Since $u \in \mathcal{R}_{\mathbf{u}}(w)$, w is a prefix of uw. Hence there exists $z \in \mathcal{L}(\mathbf{u})$ such that $uw = wz$. A known result from equations on words implies that there exist $x, y \in \mathcal{L}(\mathbf{u})$ and a non-negative integer i such that $u = xy$, $z = yx$ and $w = (xy)^i x$. Thus, w is a prefix of $u^\omega = (xy)^\omega$.

Let us now suppose that u is a return word to infinitely many factors. By the previous argument, u is a fractional root of all those factors. This implies that $u^n \in \mathcal{L}(\mathbf{u})$ for all $n \in \mathbb{N}$. Thus, $\mathbf{u}$ is either periodic or not uniformly recurrent.

Lemma 7 ([12]). *Let* $\mathbf{u}$ *be a uniformly recurrent aperiodic sequence and $f \in \mathcal{L}(\mathbf{u})^+$ such that $ind_{\mathbf{u}}(f) > 1$. Then there exist a factor $u \in \mathcal{L}(\mathbf{u})$ and a bispecial factor w in* $\mathbf{u}$ *such that $|f| = |u|$, $ind_{\mathbf{u}}(f) \le ind_{\mathbf{u}}(u) = 1 + \frac{|w|}{|u|}$ and $u \in \mathcal{R}_{\mathbf{u}}(w)^+$.*

Proposition 5. *Let* $\mathbf{u}$ *be a uniformly recurrent aperiodic sequence. Let $(w_n)_{n\in\mathbb{N}}$ be a sequence of all bispecial factors ordered by their length. For every $n \in \mathbb{N}$, let v_n be a shortest return word to w_n in* $\mathbf{u}$. *Then*

$$E(\mathbf{u}) = 1 + \sup_{n\in\mathbb{N}} \left\{ \frac{|w_n|}{|v_n|} \right\} \qquad \textit{and} \qquad E^*(\mathbf{u}) = 1 + \limsup_{n\to\infty} \frac{|w_n|}{|v_n|}.$$

Proof. By Lemma 6, $v_n w_n = v_n^{e_n}$ for some exponent $e_n \in \mathbb{Q}$ and thus $\mathrm{ind}_{\mathbf{u}}(v_n) \ge e_n = \frac{|v_n w_n|}{|v_n|} = 1 + \frac{|w_n|}{|v_n|}$. Hence $\mathrm{E}(\mathbf{u}) \ge 1 + \sup\{\frac{|w_n|}{|v_n|}\} > 1$. By the second statement of the same lemma, $\lim\limits_{n\to\infty} |v_n| = \infty$. Therefore, $\mathrm{E}^*(\mathbf{u}) \ge 1 + \limsup \frac{|w_n|}{|v_n|} \ge 1$.

To show the opposite inequality, let $\delta > 0$ be such that $\mathrm{E}(\mathbf{u}) - \delta > 1$. Thus there exists $f \in \mathcal{L}(\mathbf{u})$ satisfying $\mathrm{E}(\mathbf{u}) - \delta < \mathrm{ind}_{\mathbf{u}}(f)$. Using Lemma 7, we find $u \in \mathcal{L}(\mathbf{u})$ and a bispecial factor w such that $\mathrm{ind}_{\mathbf{u}}(f) \le \mathrm{ind}_{\mathbf{u}}(u) = 1 + \frac{|w|}{|u|}$, where $u \in \mathcal{R}_{\mathbf{u}}(w)^+$. Therefore, for some index $m \in \mathbb{N}$, one has $w = w_m$ and $|u| \ge |v_m|$. Altogether, for arbitrarily positive δ we have

$$\mathrm{E}(\mathbf{u}) - \delta < \mathrm{ind}_{\mathbf{u}}(f) \le \mathrm{ind}_{\mathbf{u}}(u) = 1 + \frac{|w|}{|u|} \le 1 + \frac{|w_m|}{|v_m|} \le 1 + \sup \left\{ \frac{|w_n|}{|v_n|} \right\}.$$

Consequently, $\mathrm{E}(\mathbf{u}) \le 1 + \sup\left\{\frac{|w_n|}{|v_n|}\right\}$.

If $\mathrm{E}^*(\mathbf{u}) = 1$, then the above proven inequality $\mathrm{E}^*(\mathbf{u}) \ge 1 + \limsup \frac{|w_n|}{|v_n|} \ge 1$ implies the second statement of the proposition. If $\mathrm{E}^*(\mathbf{u}) > 1$, then there exists a sequence of factors $f^{(n)} \in \mathcal{L}(\mathbf{u})$ with $\mathrm{ind}_{\mathbf{u}}(f^{(n)}) > 1$ such that $|f^{(n)}| \to \infty$ and $\mathrm{ind}_{\mathbf{u}}(f^{(n)}) \to \mathrm{E}^*(\mathbf{u})$. For each n, we find the factor $u^{(n)}$ and the bispecial factor $w^{(n)}$ with the properties given in Lemma 7 and we proceed analogously as before.

5 Asymptotic critical exponent of balanced sequences

To describe the asymptotic critical exponent of a balanced sequence, we first list important facts on Sturmian sequences. They are partially taken from [12], where they are used to compute the critical exponent of complementary symmetric Rote sequences.

In the sequel, we use the characterisation of standard Sturmian sequences by their directive sequences. To introduce them, we define two morphisms $G : \mathtt{a} \to \mathtt{ba}$, $\mathtt{b} \to \mathtt{b}$ and $D : \mathtt{a} \to \mathtt{a}$, $\mathtt{b} \to \mathtt{ab}$.

Proposition 6 ([15]). *For every standard Sturmian sequence* $\mathbf{u}$ *there exists a unique sequence* $\mathbf{\Delta} = \Delta_0\Delta_1\Delta_2\cdots \in \{G,D\}^{\mathbb{N}}$ *of morphisms and a sequence* $(\mathbf{u}^{(n)})_{n\geq 0}$ *of standard Sturmian sequences such that* $\mathbf{u} = \Delta_0\Delta_1\cdots\Delta_{n-1}(\mathbf{u}^{(n)})$ *for every* $n \in \mathbb{N}$*. Moreover, the sequence* $\mathbf{\Delta}$ *contains infinitely many letters* G *and infinitely many letters* D*, i.e., for some sequence* $(a_i)_{i\geq 1}$ *of positive integers we can write* $\mathbf{\Delta} = G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ *or* $\mathbf{\Delta} = D^{a_1}G^{a_2}D^{a_3}G^{a_4}$*.*

We call the sequence $\mathbf{\Delta}$ in Proposition 6 the *directive sequence* of $\mathbf{u}$.

Let us fix the notation by adopting the following **convention**: To a standard Sturmian sequence $\mathbf{u}$ with directive sequence $\mathbf{\Delta} = G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$, we assign an irrational number $\theta \in (0,1)$ having the continued fraction expansion $\theta = [a_0, a_1, a_2, a_3, \ldots]$ with $a_0 = 0$. The frequencies of the letters in the Sturmian sequence $\mathbf{u}$ are $\frac{\theta}{1+\theta}$ (for the least frequent letter) and $\frac{1}{1+\theta}$ (for the most frequent letter). For every $N \in \mathbb{N}$, we define the N^{th} *convergent* to θ as $\frac{p_N}{q_N}$ and the N^{th} convergent to $\frac{\theta}{1+\theta}$ as $\frac{P_N}{Q_N}$, where p_N, q_n, Q_N satisfy the following recurrence relation for all $N \geq 1$: $X_N = a_N X_{N-1} + X_{N-2}$, but they differ in their initial values: $p_{-1} = 1, p_0 = 0$; $q_{-1} = 0, q_0 = 1$; $Q_{-1} = Q_0 = 1$. This implies $p_N + q_N = Q_N$ for all $N \in \mathbb{N}$. Note that $\mathbf{u}$ has directive sequence $G^{a_1}D^{a_2}G^{a_3}D^{a_4}\cdots$ if and only if $\mathbf{u}$ after exchange of letters $\mathtt{a} \leftrightarrow \mathtt{b}$ has directive sequence $D^{a_1}G^{a_2}D^{a_3}G^{a_4}\cdots$.

By Vuillon's result [21], every factor of any Sturmian sequence has exactly two return words and its derived sequence is Sturmian as well. The Parikh vectors of the bispecial factors in $\mathbf{u}$ and the corresponding return words can be easily expressed using the convergents $\frac{p_N}{q_N}$ to θ. In the following proposition we order the bispecial factors in the Sturmian sequence by their length.

Proposition 7 ([12]). *Let* $\theta = [a_0, a_1, a_2, a_3, \ldots]$ *be the irrational number associated with a Sturmian sequence* $\mathbf{u}$ *and let us suppose that* $\mathtt{b}$ *is the most frequent letter. Let* b *be the* n^{th} *bispecial factor of* $\mathbf{u}$*. Then there exists a unique pair* $(N,m) \in \mathbb{N}^2$ *with* $0 \leq m < a_{N+1}$ *such that* $n = m + a_0 + a_1 + a_2 + \cdots + a_N$*. The Parikh vectors of the most frequent return word* r *to* b*, of the least frequent return word* s *to* b *and of* b *itself are* $\boldsymbol{V}(r) = \binom{p_N}{q_N}$*,* $\boldsymbol{V}(s) = \binom{m\,p_N + p_{N-1}}{m\,q_N + q_{N-1}}$ *and* $\boldsymbol{V}(b) = \boldsymbol{V}(r) + \boldsymbol{V}(s) - \binom{1}{1}$*. The irrational number associated with the derived sequence* $\mathbf{d_u}(b)$ *to* b *in* $\mathbf{u}$ *is* $\theta' = [0, a_{N+1} - m, a_{N+2}, a_{N+3}, \ldots]$*.*

We will describe how to compute the asymptotic critical exponent of the balanced sequence $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ associated with a standard Sturmian sequence $\mathbf{u}$ with θ having an eventually periodic continued fraction expansion. Our main tool for computing $\mathrm{E}^*(\mathbf{v})$ is Proposition 5. Thus we need to find for any bispecial factor of length $|b|$ in $\mathbf{v}$ the length $|v|$ of its shortest return word.

As stated in Lemma 2, if w is a bispecial factor of $\mathbf{v}$ and w is long enough, then there exist $\mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')$ bispecial factors of the same length in $\mathbf{v}$, all of them having the same projection in $\mathbf{u}$, and this projection is bispecial in $\mathbf{u}$. By Lemma 5, the shortest return words to these bispecial factors of $\mathbf{v}$ have the same length. Therefore, we can consider only one representative. In the sequel, we denote by w_n a bispecial factor of $\mathbf{v}$ such that $b_n = \pi(w_n)$ is the n^{th} bispecial factor in the Sturmian sequence $\mathbf{u}$, when these are ordered by length. We want to compute $\mathrm{E}^*(\mathbf{v}) = 1 + \limsup \frac{|w_n|}{|v_n|}$, where v_n is a shortest return word to w_n in $\mathbf{v}$. The fact that the continued fraction expansion of θ is eventually periodic enables us to split the sequence $(|w_n|/|v_n|)$ into a finite number of subsequences such that each of them has a finite limit. The largest limit of these subsequences is the searched $\mathrm{E}^*(\mathbf{v})$. To find a suitable partition of the index set $\mathbb{N}$ into a finite number of subsets of indices describing subsequences, we define an equivalence on $\mathbb{N}$. First we fix our notation: $\theta = [a_0, a_1, a_2, a_3, \dots] = [0, a_1 a_2 \dots a_h (z_0 z_1 \dots z_{M-1})^\omega]$. In particular, $a_i = z_j$, if $i > h$ and $i - 1 - h = j \mod M$.

Definition 5. *To any $n \in \mathbb{N}$ we assign a unique pair $(N, m) \in \mathbb{N}^2$ as in Proposition 7. Let (N_1, m_1) and (N_2, m_2) be assigned to the integers n_1 and n_2 respectively. We say that n_1 is equivalent to n_2 and write $n_1 \sim n_2$ if*

$$\begin{aligned} m_1 &= m_2, & N_1 &= N_2 \mod M, \\ \binom{p_{N_1-1}}{q_{N_1-1}} &= \binom{p_{N_2-1}}{q_{N_2-1}} \mod \boldsymbol{Per}(\mathbf{y}, \mathbf{y}'), & \binom{p_{N_1}}{q_{N_1}} &= \binom{p_{N_2}}{q_{N_2}} \mod \boldsymbol{Per}(\mathbf{y}, \mathbf{y}'). \end{aligned}$$

Obviously, the above defined relation on $\mathbb{N}$ is an equivalence and there are only finitely many equivalence classes, say $C_1, C_2, \dots, C_T$. Now we can define subsequences of the sequence $(|w_n|/|v_n|)$: if $\#C_t = \infty$, then we insert $|w_n|/|v_n|$ into the t^{th} subsequence for each $n \in C_t$. For each $n \in \mathbb{N}$, up to a finite number of exceptions, $|w_n|/|v_n|$ belongs to a subsequence. The number of subsequences is at most $Z\ \mathrm{Per}(\mathbf{y})^2\ \mathrm{Per}(\mathbf{y}')^2$, where $Z = z_0 + z_1 + \dots + z_{M-1}$. We obtain thus the following algorithm computing the asymptotic critical exponent.

Algorithm for determining $\mathrm{E}^*(\mathbf{v})$, where $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$:

Input: $\theta = [0, a_1 a_2 \cdots a_h (z_0 z_1 \cdots z_{M-1})^\omega]$, $\mathrm{Per}(\mathbf{y})$ and $\mathrm{Per}(\mathbf{y}')$.

Step 1. Find all infinite equivalence classes C_t introduced in Definition 5.
Step 2. For each class C_t

- insert $|w_n|/|v_n|$ into the t^{th} subsequence for each $n \in C_t$;
- find the limit e_t of the t^{th} subsequence.

Output: $\mathrm{E}^*(\mathbf{v}) = 1+$ the maximum value among all limits e_t.

Proposition 5 and a thorough study of short return words provide a lower bound on the asymptotic critical exponent.

Theorem 2. *Let $\mathbf{u}$ be a Sturmian sequence, $\mathbf{y}, \mathbf{y}'$ two constant gap sequences and $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$. Then $E(\mathbf{v}) \geq E^*(\mathbf{v}) \geq 1 + \frac{1}{\mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')}$. Moreover, $E^*(\mathbf{v})$ depends only on $\mathrm{Per}(\mathbf{y})$ and $\mathrm{Per}(\mathbf{y}')$ (not on the structure of $\mathbf{y}$ and $\mathbf{y}'$).*

On one hand, the asymptotic critical exponent depends only on the length of the periods of $\mathbf{y}$ and $\mathbf{y}'$ and it does not depend on their structure, in contrast to the critical exponent. On the other hand, the asymptotic critical exponent depends on the preperiod of the continued fraction of θ, in contrast to the asymptotic critical exponent of the associated Sturmian sequence (see [16]).

Example 11. Let $\mathbf{v}$ be the balanced sequence given by the parameters $\theta = [0, \overline{2}]$, $\mathrm{Per}(\mathbf{y}) = 1$ and $\mathrm{Per}(\mathbf{y}') = 2$. One can check that $\mathrm{E}^*(\mathbf{v}) = 3 + \sqrt{2} \doteq 4.41$.

For the balanced sequence $\mathbf{v}'$ given by the parameters $\theta = [0, 1, \overline{2}]$, $\mathrm{Per}(\mathbf{y}) = 1$ and $\mathrm{Per}(\mathbf{y}') = 2$, one has $\mathrm{E}^*(\mathbf{v}') = 2 + \frac{\sqrt{2}}{2} \doteq 2.7$.

We used a program implemented by our student Daniela Opočenská computing the asymptotic critical exponent of balanced sequences $\mathbf{x}_d$ defined in [19] for $d \in \{3, 4, \ldots, 10\}$. The authors of [19] conjectured that the least critical exponent over an alphabet of cardinality d equals $\frac{d-2}{d-3}$ and this is achieved on the sequences $\mathbf{x}_d$. This conjecture was proved for $d = 3$ and $d = 4$ in [19]. [1] Later, in [4,3] it is shown that $\mathbf{x}_d$ are indeed the sequences with the least critical exponent over alphabets of size 5 to 8. The balanced sequences $\mathbf{x}_d$ and their critical exponent are listed in Table 1.

The table is taken from [19] (instead of the slope α of a Sturmian sequence, used in the original table, we use the parameter θ corresponding to the directive sequence). We also added to the table a column containing the asymptotic critical exponent. We see that $\mathrm{E}^*(\mathbf{x}_d) = \mathrm{E}(\mathbf{x}_d)$ for $d = 3, 4, 5, 6, 7$. However $\mathrm{E}^*(\mathbf{x}_8) < \mathrm{E}(\mathbf{x}_8)$. Moreover, using the table we can deduce that there exists a balanced sequence $\mathbf{x}$ over an 8-letter alphabet with $\mathrm{E}^*(\mathbf{x}) < \mathrm{E}^*(\mathbf{x}_8)$. The sequence $\mathbf{x}$ uses the same pair $\mathbf{y}$ and $\mathbf{y}'$ as $\mathbf{x}_8$. The parameter θ corresponding to $\mathbf{x}$ is $\theta = [0, 2, 3, \overline{2}]$. Since $\mathbf{x}_8$ and $\mathbf{x}_9$ have same θ and same lengths of constant gap sequences, we have $\mathrm{E}^*(\mathbf{x}) = \mathrm{E}^*(\mathbf{x}_9) < \mathrm{E}^*(\mathbf{x}_8)$. The method used for finding the candidates with the least critical exponent cannot be applied to find a suitable $\mathbf{x}_d$ for a general d. The same is true for the least asymptotic critical exponent. Indeed, even a proof that the candidates should be given by θ with an eventually periodic continued fraction expansion is still missing.

References

1. L. Balková, M. Bucci, A. De Luca, J. Hladký, and S. Puzynina, *Aperiodic pseudorandom number generators based on infinite words*, Theoret. Comput. Sci. 647 (2016), 85–100.
2. L. Balková, E. Pelantová, and Š. Starosta, *Sturmian jungle (or garden?) on multiliteral alphabets*, RAIRO-Theor. Inf. Appl. 44 (2010), 443–470.
3. A. R. Baranwal, *Decision Algorithms for Ostrowski-Automatic Sequences*, the master thesis, University of Waterloo, http://hdl.handle.net/10012/15845 (2020).
4. 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.

[1] More precisely, the minimality in the case $d = 4$ was proved by Peltomäki in a private communication to Rampersad.

d	θ	$\mathbf{y}$	$\mathbf{y}'$	$\mathrm{E}(\mathbf{v})$	$\mathrm{E}^*(\mathbf{v})$
3	$[0,1,\overline{2}]$	0^ω	$(12)^\omega$	$2+\frac{1}{\sqrt{2}}$	$2+\frac{1}{\sqrt{2}}$
4	$[0,\overline{1}]$	$(01)^\omega$	$(23)^\omega$	$1+\frac{1+\sqrt{5}}{4}$	$1+\frac{1+\sqrt{5}}{4}$
5	$[0,1,\overline{2}]$	$(01)^\omega$	$(2324)^\omega$	$\frac{3}{2}$	$\frac{3}{2}$
6	$[0,2,1,1,\overline{1,1,1,2}]$	0^ω	$(123415321435)^\omega$	$\frac{4}{3}$	$\frac{4}{3}$
7	$[0,1,3,\overline{1,2,1}]$	$(01)^\omega$	$(234526432546)^\omega$	$\frac{5}{4}$	$\frac{5}{4}$
8	$[0,3,1,\overline{2}]$	$(01)^\omega$	$(23452673254623752643257 6)^\omega$	$\frac{6}{5}=1.2$	$\frac{12+3\sqrt{2}}{14}\doteq 1.16$
9	$[0,2,3,\overline{2}]$	$(01)^\omega$	$(2345672843652748632547 68)^\omega$	$?\frac{7}{6}\doteq 1.167$	$1+\frac{2\sqrt{2}-1}{14}\doteq 1.13$
10	$[0,4,2,\overline{3}]$	$(01)^\omega$	$(23456728496325476829436527486 9)^\omega$	$?\frac{8}{7}\doteq 1.14$	$1+\frac{\sqrt{13}}{26}\doteq 1.139$

Table 1. The balanced sequences with the least critical exponent over alphabets of size d. We denote by question mark the conjectures that are not yet proved.

5. V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone, *Acyclic, connected and tree sets,* Monatshefte für Mathematik, vol. 176(4) (2015), 521–550.
6. A. Carpi and A. de Luca, *Special factors, periodicity, and an application to Sturmian words*, Acta Inform. 36 (2000), 983–1006.
7. J. D. Currie, L. Mol, and N. Rampersad, *The repetition threshold for binary rich words*, Discrete Math. Theoret. Comput. Sci. 22(1) (2020), no. 6.
8. F. Dejean, *Sur un théorème de Thue*, J. Combin. Theory Ser. A 13 (1972), 90–99.
9. D. Damanik and D. Lenz, *The index of Sturmian sequences*, European J. Combin. 23 (2002), 23–29.
10. F. Dolce and D. Perrin, *Eventually dendric shift spaces*, Ergodic Theory and Dynamical Systems (2020), 1–26.
11. F. Durand, *A characterization of substitutive sequences using return words*, Discrete Math. 179 (1998), 89–101.
12. L. Dvořáková, K. Medková, and E. Pelantová, *Complementary symmetric Rote sequences: the critical exponent and the recurrence function*, Discrete Math. Theoret. Comput. Sci. 20 (1) (2020) #20
13. G. A. Hedlund and M. Morse, *Symbolic dynamics II - Sturmian trajectories*, Amer. J. Math. 62 (1940), 1–42
14. P. Hubert, *Suites équilibrées*, Theoret. Comput. Sci. 242 (2000), 91–108.
15. J. Justin and G. Pirillo, *Episturmian words and episturmian morphisms*, Theoret. Comput. Sci. 276 (2002), 281–313.
16. J. Justin and G. Pirillo, *Fractional powers in Sturmian words*, Theoret. Comput. Sci. 223 (2001), 363—376.
17. H. W. Lenstra and J. O. Shallit, *Continued fractions and linear recurrences*, Mathematics of Computation 61, (1993) 351-–354.
18. N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Lecture Notes in Mathematics, vol. 313. Springer-Verlag (2002). Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.
19. N. Rampersad, J. Shallit, and É. Vandomme, *Critical exponents of infinite balanced words*, Theoret. Comput. Sci. 777 (2019), 454–463.
20. J. Shallit and A. Shur, *Subword complexity and power avoidance*, Theoret. Comput. Sci. 792 (2019), 96–116.
21. L. Vuillon, *A characterization of Sturmian words by return words*, Eur. J. Combin. 22 (2001), 263–275.
22. L. Vuillon, *Balanced words*, Bull. Belgian Math. Soc. 10 (2003), 787–805.