Computation of critical exponent in balanced sequences*

Francesco Dolce, L'ubomíra Dvořáková, and Edita Pelantová

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

Abstract. We study balanced sequences over a d-letter alphabet. Each such sequence $\mathbf{v}$ is described by a Sturmian sequence and two constant gap sequences $\mathbf{y}$ and $\mathbf{y}'$. We provide an algorithm which for a given $\mathbf{y}$, $\mathbf{y}'$ and a quadratic slope of a Sturmian sequence computes the critical exponent of the balanced sequence $\mathbf{v}$.

Keywords: Critical exponent · Balanced sequences · Return words · Bispecial factors

1 Introduction

An infinite sequence $\mathbf{v}$ over a finite alphabet is called balanced if for each pair u, v of its factors having the same length and for each letter a of the alphabet, the number of occurrences of a in u and v differs at most by one. Balanced aperiodic sequences over a binary alphabet were introduced already in 1940 by Hedlund and Morse under the name Sturmian sequences (see [7]). Balanced sequences over a d-letter alphabet were characterized by Hubert in [8]; in particular he showed that each aperiodic balanced sequence over a d-letter alphabet can be mapped by a letter-to-letter projection π to a Sturmian sequence. In this paper we focus on the critical exponent of a balanced sequence $\mathbf{v}$. Roughly speaking, the critical exponent $E(\mathbf{v})$ describes the maximal repetition of factors in $\mathbf{v}$. For Sturmian sequences, the formula to evaluate the critical exponent was provided by Carpi and de Luca in [3] (see also [4]). Recently, Rampersad, Shallit and Vandomme in [9] and Baranwal and Shallit in [1] and [2] started looking for balanced sequences over a d-letter alphabet having the least critical exponent. They used the automated theorem prover Walnut to show that the smallest possible critical exponent of a balanced sequence over d letters is $\frac{d-2}{d-3}$ for $d = 5, \ldots, 8$. For $d = 9, 10$ they showed that the least critical exponent can not be smaller than $\frac{d-2}{d-3}$ and conjectured that this value is attained by the sequences $\mathbf{x}_9$ and $\mathbf{x}_{10}$ (see Example 5).

In [5], we gave a general method to compute the critical exponent $E(\mathbf{v})$ and the asymptotic critical exponent $E^*(\mathbf{v})$ of any uniformly recurrent sequence $\mathbf{v}$.

* The research received funding from the Ministry of Education, Youth and Sports of the Czech Republic through the project CZ.02.1.01/0.0/0.0/16_019/0000778.

Our method is based on looking for the shortest return words to bispecial factors in $\mathbf{v}$. The asymptotic critical exponent $E^*(\mathbf{v})$ reflects repetitions of factors of length growing to infinity. Since the letter-to-letter projection π maps every sufficiently long bispecial factor in a balanced sequence $\mathbf{v}$ to a bispecial factor in the underlying Sturmian sequence, we could apply our method to compute the asymptotic critical exponent of balanced sequences.

In this contribution we refine our approach to all bispecial factors, not only the long enough ones (Propositions 3, 5 and 6) and deduce an algorithm for computing the critical exponent of balanced sequences associated with Sturmian sequences with quadratic slopes (Section 6). In particular, we confirm the conjectured property of the sequences $\mathbf{x}_9$ and $\mathbf{x}_{10}$ (Example 7).

2 Preliminaries

An *alphabet* $\mathcal{A}$ is a finite set of symbols called *letters*. A (finite) *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 $i = 0, 1, \ldots, n-1$. 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 by $\mathcal{A}^*$. Its neutral element is the *empty word* ε and we denote $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$. If $u = xyz$ for some $x, y, z \in \mathcal{A}^*$, then x is a *prefix* of u, z is a *suffix* of u and y is a *factor* of u. To any word u over $\mathcal{A}$ with 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 $i \in \mathbb{N}$. In this paper we always denote sequences by bold letters. The shift of $\mathbf{u} = u_0u_1u_2 \cdots$ is the sequence $\sigma(\mathbf{u}) = u_1u_2u_3 \cdots$. 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 $v = \varepsilon$. If $\mathbf{u}$ is not eventually periodic, then it is *aperiodic*. A *factor* of $\mathbf{u} = u_0u_1u_2 \cdots$ is a word u such that $u = u_iu_{i+1}u_{i+2} \cdots u_{j-1}$ for some $i, j \in \mathbb{N}$, $i \leq j$. The number i is called an *occurrence* of the factor u in $\mathbf{u}$. 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 said to be *uniformly recurrent*.

The *language* $\mathcal{L}(\mathbf{u})$ of a sequence $\mathbf{u}$ is the set of all its factors. 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 central notion of our contribution is the *critical exponent* of an infinite sequence. Let $z \in \mathcal{A}^+$ be a prefix of a periodic sequence u^ω with $u \in \mathcal{A}^+$, and let us suppose that u is minimal in length with this property. We say that z has *fractional root* u and *exponent* $e = |z|/|u|$. We usually write $z = u^e$.

Definition 1. *Given a sequence* $\mathbf{u}$*, we define the* critical exponent *of* $\mathbf{u}$ *as*

$$E(\mathbf{u}) = \sup\{e \in \mathbb{Q} : \textit{there exist } x, y \in \mathcal{L}(\mathbf{u}),\ \textit{with } |x| > 0 \textit{ and } y = x^e\}.$$

If $E(\mathbf{u}) < +\infty$*, we define the* asymptotic critical exponent *of* $\mathbf{u}$ *as*

$$E^*(\mathbf{u}) = \lim_{n\to\infty} \sup\{e \in \mathbb{Q} : \text{there exist } x, y \in \mathcal{L}(\mathbf{u}), \text{ with } |x| > n \text{ and } y = x^e\}\,.$$

Otherwise $E^*(\mathbf{u}) = E(\mathbf{u}) = +\infty$.

In [5] we find a formula to compute $\mathrm{E}(\mathbf{u})$ and $\mathrm{E}^*(\mathbf{u})$ for a uniformly recurrent aperiodic sequence $\mathbf{u}$. This tool uses the notion of return words.

Let us 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, then the set $\mathcal{R}_{\mathbf{u}}(w)$ is finite for each prefix w. 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 $\mathbf{d_u}(w) = d_0d_1d_2\cdots$ over the alphabet of cardinality $\#\mathcal{R}_{\mathbf{u}}(w)$.

Proposition 1 ([5]). *Let* $\mathbf{u}$ *be a uniformly recurrent aperiodic sequence. Let* $(w_n)_{n\in\mathbb{N}}$ *be a sequence of all bispecial factors of* $\mathbf{u}$ *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*

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

3 Balanced sequences

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 $\big||u|_a - |v|_a\big| \leq 1$. Aperiodic balanced sequences over binary alphabet, i.e., Sturmian sequences, can be characterized by many equivalent definitions. The definition we will need is based on return words. Vuillon in [10] shows that an infinite recurrent sequence $\mathbf{u}$ is Sturmian if and only if each of its factors has exactly two return words. Moreover, the derived sequence to a factor of a Sturmian sequence is Sturmian too. A Sturmian sequence $\mathbf{u}$ is called *standard* if each bispecial factor of $\mathbf{u}$ is a prefix of $\mathbf{u}$. To any Sturmian sequence $\mathbf{u}'$ there exists a standard Sturmian sequence $\mathbf{u}$ such that $\mathcal{L}(\mathbf{u}) = \mathcal{L}(\mathbf{u}')$. Balanced sequences over alphabets of higher cardinality can be constructed from Sturmian sequences. To describe the construction we need the following definition.

Definition 2. *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, denoted* $\mathrm{gap}_{\mathbf{y}}(a)$*, such that the distance between successive occurrences of* a *in* $\mathbf{y}$ *is always* $\mathrm{gap}_{\mathbf{y}}(a)$.

Any constant gap sequence is periodic. We denote by $\mathrm{Per}(\mathbf{y})$ the minimal period of $\mathbf{y}$. Note that $\mathrm{gap}_{\mathbf{y}}(a)$ divides $\mathrm{Per}(\mathbf{y})$ for each letter a appearing in $\mathbf{y}$. Given a constant gap sequence $\mathbf{y}$ and a word $y \in \mathcal{L}(\mathbf{y})$ we denote by $\mathrm{gap}_{\mathbf{y}}(y)$ the length of the gap between two successive occurrences of y in $\mathbf{y}$. Note that $\mathrm{gap}_{\mathbf{y}}(y) = \mathrm{lcm}\{\mathrm{gap}_{\mathbf{y}}(a) : a \in \mathcal{A} \text{ and } a \text{ occurs in } y\}$. Moreover $\mathrm{gap}_{\mathbf{y}}(y)$ divides $\mathrm{Per}(\mathbf{y})$ for every factor $y \in \mathcal{L}(\mathbf{y})$.

Example 1. In the sequel we will deal with the following constant gap sequences $\mathbf{y} = (\mathtt{01})^\omega$ and $\mathbf{y}' = (\mathtt{234567284365274863254768})^\omega$. The sequence $\mathbf{y}$ is evidently a constant gap sequence because $\mathrm{gap}_{\mathbf{y}}(\mathtt{0}) = \mathrm{gap}_{\mathbf{y}}(\mathtt{1}) = 2$. The sequence $\mathbf{y}'$ is also a constant gap sequence because $\mathrm{gap}_{\mathbf{y}'}(a) = 6$ for $a \in \{\mathtt{2},\mathtt{4},\mathtt{6}\}$ and $\mathrm{gap}_{\mathbf{y}'}(a) = 8$ for $a \in \{\mathtt{3},\mathtt{5},\mathtt{7},\mathtt{8}\}$. Moreover, for every $y \in \mathcal{L}(\mathbf{y}')$ with $|y| \geq 2$ we have $\mathrm{gap}_{\mathbf{y}'}(y) = 24$. The minimal periods are respectively $\mathrm{Per}(\mathbf{y}) = 2$ and $\mathrm{Per}(\mathbf{y}') = 24$.

Given a constant gap sequence $\mathbf{y}$ we define for every positive integer n the set $\mathrm{gap}(\mathbf{y}, n) = \{i \ : \ \exists y \in \mathcal{L}(\mathbf{y}), |y| = n, \ \mathrm{gap}_{\mathbf{y}}(y) = i\}$. It is clear that $\mathrm{gap}(\mathbf{y}, 0) = \{1\}$ for every constant gap sequence $\mathbf{y}$.

Example 2. Let $\mathbf{y}, \mathbf{y}'$ be the sequences in Example 1. One has $\mathrm{gap}(\mathbf{y}, n) = \{2\}$ for every $n \geq 1$, $\mathrm{gap}(\mathbf{y}', 1) = \{6, 8\}$ and $\mathrm{gap}(\mathbf{y}', n) = \{24\}$ for every $n \geq 2$.

Theorem 1 ([8]). *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* $\mathcal{A}$*.*

Let us recall that the frequencies of letters in any Sturmian sequence $\mathbf{u}$ are always well defined and irrational. We will assume here, without loss of generality, that $\rho_{\mathtt{a}} < \rho_{\mathtt{b}}$ and adopt the convention that the first component of the Parikh vector of a factor of $\mathbf{u}$ corresponds to the least frequent letter of $\mathbf{u}$ and the second component to the most frequent letter (even if we consider a Sturmian sequence over binary alphabets other than $\{\mathtt{a}, \mathtt{b}\}$).

Definition 3. *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).

Example 3. Let us consider the sequence $\mathbf{x}_9$ (see Example 5 later for a more precise definition) obtained as colouring by the constant gap sequences $\mathbf{y}$ and $\mathbf{y}'$ given in Example 1 of a Sturmian sequence $\mathbf{u}$ starting as follows:

$$\mathbf{u} = \mathtt{bbabbabbabbbabbabbabbbabbabbabbabbbabbabbabbbabbabb} \cdots .$$

Thus $\mathbf{x}_9$ starts as follows:

$$\mathbf{x}_9 = \mathtt{230451670284136052174806312504716820341560728143065} \cdots .$$

Such a sequence is balanced according to Theorem 1.

The language of balanced sequences has certain symmetries. In particular, the following result is proved in [5, Corollary 1].

Lemma 1 ([5]). *Let* $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ *and* $v \in \mathcal{L}(\mathbf{v})$. *For any* $i, j \in \mathbb{N}$ *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* $\mathcal{L}(\mathbf{v})$.

Note that $\mathcal{L}(\mathbf{v})$ does not depend on the sequence $\mathbf{u}$ itself but only on $\mathcal{L}(\mathbf{u})$. Having in mind the formula for computing the critical exponent given in Proposition 1, we focus on return words to factors of balanced sequences.

In the sequel we will use the following notation:

$$\begin{pmatrix} a \\ b \end{pmatrix} \mod \begin{pmatrix} n \\ n' \end{pmatrix} := \begin{pmatrix} a \mod n \\ b \mod n' \end{pmatrix}.$$

Proposition 2. *Let* $u, f \in \mathcal{L}(\mathbf{u})$ *such that* $fu \in \mathcal{L}(\mathbf{u})$ *and* u *is a prefix of* fu. *Then the two statements are equivalent:*

1. *there exist* w *and* v *such that* $vw \in \mathcal{L}(\mathbf{v})$, w *is a prefix of* vw, $|w| = |u|$ *and* $\pi(vw) = fu$;
2. $\vec{V}(f) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \mod \begin{pmatrix} n \\ n' \end{pmatrix}$ *for some* $n \in \mathrm{gap}(\mathbf{y}, |u|_{\mathtt{a}})$ *and* $n' \in \mathrm{gap}(\mathbf{y}', |u|_{\mathtt{b}})$.

Proof. Let v and w be as in Item 1. Then u is a prefix and a suffix of $\pi(vw)$ and $f = \pi(v)$. By Lemma 1, the factor w occurring as a prefix of vw is obtained from u by colouring the $\mathtt{a}$'s with $\sigma^s(\mathbf{y})$ and the $\mathtt{b}$'s with $\sigma^t(\mathbf{y}')$ for some $s, t \in \mathbb{N}$. Hence, the same factor w occurring as a suffix of vw is obtained from u by colouring the $\mathtt{a}$'s with $\sigma^S(\mathbf{y})$ and the $\mathtt{b}$'s with $\sigma^T(\mathbf{y}')$, where $S = s + |f|_{\mathtt{a}}$ and $T = t + |f|_{\mathtt{b}}$. Hence the prefixes of length $|u|_a$ of $\sigma^s(\mathbf{y})$ and $\sigma^S(\mathbf{y})$ coincide, and similarly the prefixes of length $|u|_b$ of $\sigma^t(\mathbf{y}')$ and $\sigma^T(\mathbf{y}')$ coincide. This implies that $|f|_{\mathtt{a}}$ is divisible by some $n \in \mathrm{gap}(\mathbf{y}, |u|_{\mathtt{a}})$ and that $|f|_{\mathtt{b}}$ is divisible by some $n' \in \mathrm{gap}(\mathbf{y}', |u|_{\mathtt{b}})$. In other words, $|f|_{\mathtt{a}} = 0 \mod n$ and $|f|_{\mathtt{b}} = 0 \mod n'$.

Let f, n and n' be as in Item 2. Let us consider $y \in \mathcal{L}(\mathbf{y})$ and $y' \in \mathcal{L}(\mathbf{y}')$ such that $\mathrm{gap}_{\mathbf{y}}(y) = n$ with $|y| = |u|_{\mathtt{a}}$ and $\mathrm{gap}_{\mathbf{y}}(y') = n'$ with $|y'| = |u|_{\mathtt{b}}$. Let $s, t \in \mathbb{N}$ be such that y is a prefix of $\sigma^s(\mathbf{y})$ and y' is a prefix of $\sigma^t(\mathbf{y}')$. Colouring the letters $\mathtt{a}$'s in fu with $\sigma^s(\mathbf{y})$ and the letters $\mathtt{b}$'s with $\sigma^t(\mathbf{y}')$, we get, by Lemma 1, a factor x of $\mathbf{v}$. Since $|f|_{\mathtt{a}}$ is a multiple of $\mathrm{gap}_{\mathbf{y}}(y)$ and $|f|_{\mathtt{b}}$ is a multiple of $\mathrm{gap}_{\mathbf{y}}(y')$, the prefix and the suffix of length $|u|$ of x coincide, i.e., $x = vw$, w is a prefix of vw, $|w| = |u|$ and $\pi(vw) = fu$.

As we have already mentioned, any factor of a Sturmian sequence has exactly two return words and thus any piece of $\mathbf{u}$ between occurrences of u is a concatenation of these two return words. This implies the following observation.

Observation 1 *Let* r *and* s *be respectively the most and the least frequent return words to* u *in* $\mathbf{u}$. *If* $fu \in \mathcal{L}(\mathbf{u})$ *and* u *is a prefix of* fu, *then* $\vec{V}(f) = k\vec{V}(r) + \ell\vec{V}(s)$, *where* $\left(\begin{smallmatrix} \ell \\ k \end{smallmatrix}\right)$ *is the Parikh vector of a factor of the derived sequence* $\mathbf{d_u}(u)$.

4 Shortest return words to factors in balanced sequences

The length of the return words to factors of a Sturmian sequence $\mathbf{u}$ is well-known. Our aim in this section is to find a formula for the length of the shortest return words to factors of a colouring of $\mathbf{u}$. As occurrences of a factor u in a Sturmian sequence $\mathbf{u}$ and occurrences of factors from $\pi^{-1}(u)$ in any colouring of $\mathbf{u}$ coincide, we can therefore be able to give a formula based on the knowledge of the length of return words in $\mathbf{u}$. Proposition 2 and Observation 1 justify the following definition.

Definition 4. *Let $u \in \mathcal{L}(\mathbf{u})$ and let r and s be respectively the most and the least frequent return words to u in $\mathbf{u}$. We denote $\mathcal{S}(u) = \mathcal{S}_1(u) \cap \mathcal{S}_2(u) \cap \mathcal{S}_3$, where*

$$\mathcal{S}_1(u) = \left\{ \begin{pmatrix} \ell \\ k \end{pmatrix} : \begin{pmatrix} \ell \\ k \end{pmatrix} \textit{ is the Parikh vector of a factor of } \mathbf{d_u}(u) \right\};$$

$$\mathcal{S}_2(u) = \bigcup_{n \in \mathrm{gap}(\mathbf{y}, |u|_{\mathtt{a}})} \ \bigcup_{n' \in \mathrm{gap}(\mathbf{y}', |u|_{\mathtt{b}})} \left\{ \begin{pmatrix} \ell \\ k \end{pmatrix} : \ k\vec{V}(r) + \ell\vec{V}(s) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \mod \begin{pmatrix} n \\ n' \end{pmatrix} \right\};$$

$$\mathcal{S}_3 = \left\{ \begin{pmatrix} \ell \\ k \end{pmatrix} : 1 \leq k + \ell \leq \mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}') \right\}.$$

Using the formula provided in Proposition 1, we can treat all bispecial factors of the same length simultaneously.

Proposition 3. *Let $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ and $u \in \mathcal{L}(\mathbf{u})$. The shortest words in the set $\{v : v \in \mathcal{R}_{\mathbf{v}}(w)$ and $\pi(w) = u\}$ have length*

$$|v| = \min\{k|r| + \ell|s| : \begin{pmatrix} \ell \\ k \end{pmatrix} \in \mathcal{S}(u)\}.$$

Proof. First, let us show that the length of any return word in $\mathbf{v}$ to a factor from $\pi^{-1}(u)$ is contained in the set $\left\{k|r| + \ell|s| : \binom{\ell}{k} \in \mathcal{S}_1(u) \cap \mathcal{S}_2(u)\right\}$. By Proposition 2 and Observation 1, a vector $\binom{\ell}{k}$ belongs to $\mathcal{S}_1(u) \cap \mathcal{S}_2(u)$ if and only if $k\vec{V}(r) + \ell\vec{V}(s)$ is the Parikh vector of $\pi(v)$, where v is a factor between two (possibly not consecutive) occurrences of a factor $w \in \pi^{-1}(u)$ in $\mathbf{v}$. Obviously, the length of v is $k|r| + \ell|s|$. It is evident that if we consider above $|v| = \min\{k|r| + \ell|s|\}$, where $\binom{\ell}{k} \in \mathcal{S}_1(u) \cap \mathcal{S}_2(u)$, then v is a return word to a factor $w \in \pi^{-1}(u)$.

To finish the proof, we have to show that the minimum value of $|v|$ is attained for k and ℓ satisfying $1 \leq k + \ell \leq \mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')$. Let $\binom{\ell}{k} \in \mathcal{S}_1(u) \cap \mathcal{S}_2(u)$ and $k + \ell > \mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')$. Thus $\vec{V}(d) = \binom{\ell}{k}$ for some $d = d_1 d_2 d_3 \cdots d_{k+\ell} \in \mathcal{L}(\mathbf{d_u}(u))$. For every $i = 1, 2, \ldots, k + \ell$, we denote $\binom{\ell_i}{k_i} = \vec{V}(d_1 d_2 \cdots d_i)$. We assign to each i the vector $X_i = k_i\vec{V}(r) + \ell_i\vec{V}(s)$. Since the number of equivalence classes $\mod \binom{n}{n'}$ is $nn' \leq \mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')$, there exist i, j with $1 \leq i < j \leq k + \ell$ such that $X_i = X_j \mod \binom{n}{n'}$. Denote $\binom{\ell'}{k'}$ the Parikh vector of $d_{i+1} d_{i+2} \cdots d_j$. Obviously, $\binom{\ell'}{k'} \in \mathcal{S}_1(u)$, $1 \leq j - i = k' + \ell' < k + \ell$ and $k' \leq k$ and $\ell' \leq \ell$. Hence

$k'|r| + \ell'|s| < k|r| + \ell|s|$. Since $k'\vec{V}(r) + \ell'\vec{V}(s) = X_j - X_i = \binom{0}{0} \bmod \binom{n}{n'}$, the vector $\binom{\ell'}{k'} \in \mathcal{S}_2(u)$. Therefore, the minimum length can not be achieved for $k + \ell > \mathrm{Per}(\mathbf{y})\mathrm{Per}(\mathbf{y}')$.

Since a constant gap sequence is periodic, it is clear that any long enough factor in the sequence is neither right special nor left special. Let us define, for a given constant gap sequence $\mathbf{y}$, the number

$$\beta(\mathbf{y}) = \max\{|u| : u \text{ is a bispecial factor of } \mathbf{y}\}.$$

It immediately follows that for $n > \beta(\mathbf{y})$, we have $\mathrm{gap}(\mathbf{y}, n) = \{\mathrm{Per}(\mathbf{y})\}$.

Example 4. Let us consider the sequences $\mathbf{y}$ and $\mathbf{y}'$ from Example 1. One can easily check that $\beta(\mathbf{y}) = 0$ and $\beta(\mathbf{y}') = 1$.

The following result is analogous to [5, Lemma 3].

Lemma 2. *Let* $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ *and* $w \in \mathcal{L}(\mathbf{v})$*.*

1. *If* $\pi(w)$ *is bispecial in* $\mathbf{u}$*, then* w *is bispecial in* $\mathbf{v}$*.*
2. *If* w *is bispecial in* $\mathbf{v}$*,* $|\pi(w)|_{\mathtt{a}} > \beta(\mathbf{y})$ *and* $|\pi(w)|_{\mathtt{b}} > \beta(\mathbf{y}')$*, then* $\pi(w)$ *is bispecial in* $\mathbf{u}$*. Moreover, in this case* $\pi(\mathcal{R}_{\mathbf{v}}(w)) = \pi(\mathcal{R}_{\mathbf{v}}(w'))$ *for each* $w' \in \mathcal{L}(\mathbf{v})$ *with* $\pi(w') = \pi(w)$*.*

If a projection of a bispecial factor w in $\mathbf{v}$ is bispecial in $\mathcal{L}(\mathbf{u})$, we can deduce an explicit formula for $1 + \frac{|w|}{|v|}$, where $|v|$ is the length of a shortest return word to w in $\mathbf{v}$. These values are crucial for the computation of $E(\mathbf{v})$ and $E^*(\mathbf{v})$.

First, we list some important facts on Sturmian sequences. They are partially taken from [6]. Recall our convention for the frequencies of letters $\rho_{\mathtt{a}} < \rho_{\mathtt{b}}$. The language of the Sturmian sequence $\mathbf{u}$ is fully described by the coefficients of the continued fraction of the number θ associated with $\mathbf{u}$, that is

$$\theta = \theta(\mathbf{u}) := \frac{\rho_{\mathtt{a}}}{\rho_{\mathtt{b}}} = [0, a_1, a_2, a_3, \ldots].$$

The relation to the slope α of $\mathbf{u}$ is $\alpha = \frac{1}{1+\theta}$. 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.

Proposition 4 ([6]). *Let* $\theta = [0, a_1, a_2, a_3, \ldots]$ *be the irrational number associated with a standard Sturmian sequence* $\mathbf{u}$ *and* b *a bispecial factor of* $\mathbf{u}$*. Then*

1. *there exists a unique pair* $(N, m) \in \mathbb{N}^2$ *with* $0 \le m < a_{N+1}$ *such that 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*

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

2. the derived sequence $\mathbf{d_u}(b)$ *to* b *in* $\mathbf{u}$ *is Sturmian and the irrational number associated with* $\mathbf{d_u}(b)$ *is* $\theta' = [0, a_{N+1} - m, a_{N+2}, a_{N+3}, \ldots]$.

Let us recall that the nominator p_N and the denominator q_N of the N^{th} *convergent* to θ satisfy for all $N \geq 1$ the recurrence relation $X_N = a_N X_{N-1} + X_{N-2}$, but that they differ in their initial values: $p_{-1} = 1, p_0 = 0; q_{-1} = 0, q_0 = 1$.

The following statement is a direct consequence of Propositions 3 and 4.

Proposition 5. *Let* $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ *and* $\left(\frac{p_N}{q_N}\right)_N$ *be the sequence of convergents to the irrational number* θ *associated with* $\mathbf{u}$. *Let* (N, m) *be the pair assigned in Proposition 4 to a bispecial factor* $b \in \mathcal{L}(\mathbf{u})$. *Then, a shortest return word* v *to a factor* $w \in \pi^{-1}(b)$ *satisfies*

$$I(N, m) := 1 + \frac{|w|}{|v|} = 1 + \max\left\{\frac{(1+m)Q_N + Q_{N-1} - 2}{(k + \ell m)Q_N + \ell Q_{N-1}} : \left(\begin{smallmatrix}\ell\\ k\end{smallmatrix}\right) \in \mathcal{S}(b)\right\}, \quad (1)$$

where $Q_N := p_N + q_N$ *and* $Q_{N-1} := p_{N-1} + q_{N-1}$.

The following lemma helps us to recognize which vector is the Parikh vector of a factor of a given Sturmian sequence. This is important to decide whether $\left(\begin{smallmatrix}\ell\\ k\end{smallmatrix}\right)$ belongs to $\mathcal{S}_1(b)$. The lemma can be shown using the facts that $\theta = \frac{\rho_{\mathtt{a}}}{\rho_{\mathtt{b}}}$ and that $\mathbf{u}$ is balanced.

Lemma 3. *Let* $\mathbf{u}$ *be a Sturmian sequence with associated irrational number* θ. *Then* $\mathbf{u}$ *contains a factor* u *such that* $|u|_{\mathtt{b}} = k$ *and* $|u|_{\mathtt{a}} = \ell$ *if and only if* $(k-1)\theta - 1 < \ell < (k+1)\theta + 1$ *and* $k, \ell \in \mathbb{N}$.

Example 5. In the sequel, we will illustrate our method for computing the critical exponent on the balanced sequences $\mathbf{x}_9$ and $\mathbf{x}_{10}$ introduced in [9] as candidates to be the balanced sequences having the minimal critical exponent over respectively a 9- and a 10-letter alphabet. Let us define $\mathbf{x}_9$ and $\mathbf{x}_{10}$.

- $\mathbf{x}_9 = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$, where $\mathbf{u}$ is the standard Sturmian sequence associated with $\theta = [0, 2, 3, 2^{\omega}]$, and $\mathbf{y}, \mathbf{y}'$ are the constant gap sequences introduced in Example 1. Prefixes of $\mathbf{u}$ and $\mathbf{x}_9$ are displayed in Example 3.
- $\mathbf{x}_{10} = \mathrm{colour}(\mathbf{u}', \mathbf{y}, \mathbf{y}'')$, where $\mathbf{u}'$ is the standard Sturmian sequence associated with $\theta = [0, 4, 2, 3^{\omega}]$, $\mathbf{y}$ is the constant gap sequence introduced in Example 1 and $\mathbf{y}'' = (\mathtt{234567284963254768294365274869})^{\omega}$.

5 Computation of the Asymptotic Critical Exponent

From now on we consider a standard Sturmian sequence $\mathbf{u}$ with associated irrational number θ having eventually periodic continued fraction expansion. The goal of this section is to compute the asymptotic critical exponent of a sequence $\mathbf{v}$ obtained by colouring of $\mathbf{u}$. By Proposition 1, to determine $E^*(\mathbf{v})$ we only need to consider long enough bispecial factors w.

For this purpose, we write the continued fraction expansion of θ as

$$\theta = [0, a_1, a_2, \ldots, a_h, (z_0, z_1, \ldots, z_{M-1})^\omega], \tag{2}$$

where the preperiod h is chosen so that each bispecial factor b associated with (N, m), $N \geq h$, satisfies $|b|_{\mathtt{a}} > \beta(\mathbf{y})$ and $|b|_{\mathtt{b}} > \beta(\mathbf{y}')$.

We then decompose the set $\mathcal{W}$ of all nonempty bispecial factors of $\mathbf{v} = \mathrm{colour}(\mathbf{u}, \mathbf{y}, \mathbf{y}')$ into two subsets:

$\mathcal{W}^{long} := \{w \in \mathcal{W} : \pi(w) \text{ bispecial in } \mathbf{u} \text{ assigned to } (N, m) \text{ with } N \geq h\}$.

$\mathcal{W}^{short} := \mathcal{W} \setminus \mathcal{W}^{long}$.

Using Proposition 5 and Lemma 2 in order to compute $E^*(\mathbf{v})$, we need to manipulate the numbers $I(N, m)$ defined in Equation (1).

Our approach consists in partitioning the set of all possible pairs (N, m), $N \geq h$, into a finite number of subsets such that $\mathcal{S}(b)$ is the same for each Sturmian bispecial factor b assigned to a pair in the given subset. A suitable partition uses the following equivalence relation on the first component of the pair.

Definition 5. *Let $N_1, N_2 \in \mathbb{N}$ and $N_1, N_2 \geq h$. We say that N_1 is equivalent to N_2, and write $N_1 \sim N_2$, if the following three conditions are satisfied:*

1. $N_1 = N_2 \bmod M$,
2. $\begin{pmatrix} p_{N_1-1} \\ q_{N_1-1} \end{pmatrix} = \begin{pmatrix} p_{N_2-1} \\ q_{N_2-1} \end{pmatrix} \mod \begin{pmatrix} \mathrm{Per}(\mathbf{y}) \\ \mathrm{Per}(\mathbf{y}') \end{pmatrix}$,
3. $\begin{pmatrix} p_{N_1} \\ q_{N_1} \end{pmatrix} = \begin{pmatrix} p_{N_2} \\ q_{N_2} \end{pmatrix} \mod \begin{pmatrix} \mathrm{Per}(\mathbf{y}) \\ \mathrm{Per}(\mathbf{y}') \end{pmatrix}$.

The properties of the equivalence $\sim$ are summarized in the following lemma. They follow from the definition of convergents to θ and from the periodicity of the continued fraction expansion of θ.

Lemma 4. *Let $\sim$ be the equivalence on the set $\{N \in \mathbb{N} : N \geq h\}$ introduced in Definition 5 and let H denote the number of equivalence classes.*

1. *If $N_1 \sim N_2$, then $a_{N_1+1} = a_{N_2+1}$.*
2. *$N_1 \sim N_2$ if and only if $N_1 + 1 \sim N_2 + 1$.*
3. *$N_1 \sim N_2$ if and only if $N_2 = N_1 \mod H$.*
4. *$H = \min\{i \in \mathbb{N}, i > 0 : h + i \sim h\} \leq M\mathrm{Per}(\mathbf{y})^2\mathrm{Per}(\mathbf{y}')^2$.*
5. *H is divisible by M.*

Definitions 4 and 5 together with Lemma 4 ensure the following property.

Corollary 1. *Let $b^{(1)}$ and $b^{(2)}$ be bispecial factors of $\mathbf{u}$ and (N_1, m_1) and (N_2, m_2), with $N_1 \geq h$ and $N_2 \geq h$, be the pairs assigned to $b^{(1)}$ and $b^{(2)}$ respectively.*

If $N_1 \sim N_2$ and $m_1 = m_2$, then $\mathcal{S}(b^{(1)}) = \mathcal{S}(b^{(2)})$.

Let us define a partition of the set $\mathcal{W}^{long}$ into subsets $C(i, m)$, where $0 \leq i < H$ and $0 \leq m < z_{i \bmod M}$, as follows: if $(h + i + NH, m)$ is the pair assigned to a bispecial factor $b = \pi(w)$ in $\mathbf{u}$, then we put w into the subset $C(i,m)$. Using Propositions 1 and 5, we have

$$E^*(\mathbf{v}) = \max\left\{E^*(i,m) \ : \ 0 \le i < H, \ \ 0 \le m < z_{i \bmod M}\right\} , \tag{3}$$

where $E^*(i,m) := \limsup_{N\to\infty} I(h+i+NH, \ m)$.

To compute $E^*(i,m)$, we need, according to Equation (1), to determine $\lim_{N\to\infty} \frac{Q_{N-1}}{Q_N}$. A direct consequence of the Perron-Frobenius theorem serves this purpose.

Lemma 5. *Let $A \in \mathbb{N}^{2\times 2}$ be a primitive matrix with* $\det A = \pm 1$, *and $(S_N)_N$, $(T_N)_N$ be two sequences of integers given by the recurrent relation $(S_{N+1}, T_{N+1}) = (S_N, T_N)A$ for each $N \in \mathbb{N}$, with $S_0, T_0 \in \mathbb{N}$ such that $S_0 + T_0 > 0$. Denote by $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ an eigenvector of A to the non-dominant eigenvalue λ. Then*

1. $\lim_{N\to\infty} \frac{S_N}{T_N} = -\frac{y}{x}$, *and*
2. $S_N + \frac{y}{x}T_N = \lambda^N(S_0 + \frac{y}{x}T_0)$ *for each* $N \in \mathbb{N}$.

Proof. As A is a primitive matrix with non-negative entries, the components x and y of an eigenvector to the non-dominant eigenvalue have opposite signs. In particular $x, y \neq 0$. Obviously, $(S_N, T_N) = (S_0, T_0)A^N$ for each $N \in \mathbb{N}$. Multiplying both sides of the equation by the eigenvector $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$, we obtain $xS_N + yT_N = \lambda^N(xS_0 + yT_0)$, i.e., Item 2 is proven.

As $|\lambda| < 1$, Item 2 implies that $\lim_{N\to\infty} xT_N\left(\frac{S_N}{T_N} + \frac{y}{x}\right) = \lim_{N\to\infty}(xS_N + yT_N) = 0$. Since $\lim_{N\to\infty} T_N = +\infty$, necessarily $\lim_{N\to\infty}\left(\frac{S_N}{T_N} + \frac{y}{x}\right) = 0$. This proves Item 1.

Periodicity of the continued fraction expansion of θ and the previous lemma ensure that the sequences $S_N := Q_{MN+h+i-1}$ and $T_N := Q_{MN+h+i}$ satisfy the recurrent relation $(S_{N+1}, T_{N+1}) = (S_N, T_N)A^{(i)}$ with

$$A^{(i)} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & z_i \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 1 \\ 1 & z_{i+1} \end{smallmatrix}\right)\cdots\left(\begin{smallmatrix} 0 & 1 \\ 1 & z_{M-1} \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 1 \\ 1 & z_0 \end{smallmatrix}\right)\cdots\left(\begin{smallmatrix} 0 & 1 \\ 1 & z_{i-1} \end{smallmatrix}\right) , \tag{4}$$

and hence also the existence of the limit

$$L_i = \lim_{N\to\infty} \frac{S_N}{T_N} = \lim_{N\to\infty} \frac{Q_{HN+h+i-1}}{Q_{HN+h+i}} \quad \text{for} \ \ i = 0, 1, \ldots, H-1. \tag{5}$$

Moreover, the non-dominant eigenvalue λ of $A^{(i)}$ satisfies

$$S_N - L_iT_N = \lambda^N(S_0 - L_iT_0) \quad \text{for each } N \in \mathbb{N}\,. \tag{6}$$

By Corollary 1, for all bispecial factors w in $C(i,m)$ we obtain the same set $\mathcal{S}(\pi(w))$. Let us denote $\mathcal{S}(i,m) := \mathcal{S}(\pi(w))$. Formula (1) then immediately gives

$$E^*(i,m) = 1 + \max\left\{\frac{1+m+L_i}{k+\ell m+\ell L_i} \ : \ \left(\begin{smallmatrix} \ell \\ k \end{smallmatrix}\right) \in \mathcal{S}(i,m)\right\} . \tag{7}$$

Example 6. Let us evaluate $E^*(\mathbf{x}_9)$, where $\mathbf{x}_9$ is the sequence defined in Example 5. It is easy to find that $H = 8$ and there are 16 distinct subsets $C(i,m)$ for $i \in \{0,1,\dots,7\}$ and $m \in \{0,1\}$. As $\theta = [0,2,3,2^\omega]$ has period 1, the recurrence relation for $(Q_N)_N$ is $Q_{N+1} = 2Q_N + Q_{N-1}$ for $N \geq 2$. In particular $L_i = \lim\limits_{N\to\infty} \frac{Q_{N-1}}{Q_N} = \sqrt{2}-1$ for each i. Listing all elements of $\mathcal{S}(i,m)$ is more laborious (but possible to do by hand as well). Thanks to a program implemented by our student Daniela Opočenská we find that $E^*(\mathbf{x}_9) = E^*(2,1)$. Since $\mathcal{S}(2,1) = \{\left(\begin{smallmatrix}6\\10\end{smallmatrix}\right)\}$, we have $E^*(\mathbf{x}_9) = E^*(2,1) = 1 + \frac{2+L_0}{16+6L_0} = 1 + \frac{2\sqrt{2}-1}{14} \doteq 1,1306$.

Using the same program we also find that $E^*(\mathbf{x}_{10}) = 1 + \frac{\sqrt{13}}{26} \doteq 1,1387$.

6 Computation of the critical exponent

In order to evaluate the critical exponent of $\mathbf{v} = \mathrm{colour}(\mathbf{u},\mathbf{y},\mathbf{y}')$, we have to determine, by Proposition 1,

$$E(\mathbf{v}) = 1 + \sup\left\{\frac{|w|}{|v|} : w \in \mathcal{L}(\mathbf{v}), w \text{ bispecial and } v \in \mathcal{R}_{\mathbf{v}}(w)\right\}.$$

To find the maximum value of $\frac{|w|}{|v|}$ among $w \in \mathcal{W}^{short}$ and $v \in \mathcal{R}_{\mathbf{v}}(w)$ we use Propositions 3 and 5. To determine $\sup\left\{\frac{|w|}{|v|} : w \in \mathcal{W}^{long} \text{ and } v \in \mathcal{R}_{\mathbf{v}}(w)\right\}$ we use the partition of $\mathcal{W}^{long}$ into subsets $C(i,m)$ which have been introduced in the previous section to count the asymptotic critical exponent. For each $C(i,m)$ we have to find

$$E(i,m) := \sup\left\{I(h+i+NH,\ m) : N \in \mathbb{N}\right\} \geq E^*(i,m)$$

and then to determine the maximal value among $E(i,m)$. We show that $I(h+i+NH,m)$ may exceed $E^*(i,m)$ only for a finite number of indices $N \in \mathbb{N}$.

Proposition 6. *Let λ be the non-dominant eigenvalue of the matrix $A^{(i)}$ defined in Equation* (4) *and L_i be the limit given in Equation* (5)*. Denote $\mu = |\lambda|^{H/M} < 1$. If $N_0 \in \mathbb{N}$ satisfies $\mu^{N_0}\left|Q_{h+i-1} - L_i Q_{h+i}\right| \leq 2L_i$, then $I(h+i+NH,m) \leq E^*(i,m)$ for all $N \geq N_0$ and $0 \leq m < z_{i \bmod M}$.*

Proof. Equation (6) gives $\left|Q_{h+i+NH-1} - L_i Q_{h+i+NH}\right| = \mu^N \left|Q_{h+i-1} - L_i Q_{h+i}\right|$. Thus, it is enough to show the implication:

If $I(h+i+NH,m) > E^*(i,m)$, then $\left|Q_{h+i+NH-1} - L_i Q_{h+i+NH}\right| > 2L_i$.

For this sake, we abbreviate notation by putting $S = Q_{h+i+NH-1}$, $T = Q_{h+i+NH}$ and $L = L_i$. Recall that $0 < L_i < 1$. Let $\binom{\ell}{k} \in \mathcal{S}(i,m)$ such that

$$I(h+i+NH,m) = 1 + \frac{(1+m)T + S - 2}{(k+\ell m)T + \ell S} > E^*(i,m) \geq 1 + \frac{1+m+L}{k+\ell m+\ell L}.$$

Thus we have $(k-\ell)(S-LT) > 2(k+\ell m+\ell L) \geq 2L|k-\ell|$, hence $|S-LT| > 2L$.

Example 7. Let us show that $E(\mathbf{x}_9) = \frac{7}{6}$. To do that we have to consider the sets of short and long bispecial factors.

$\mathcal{W}^{short}$: It is easy to check that $\pi\left(\mathcal{W}^{short}\right) = \left\{\mathtt{a}, \mathtt{b}, \mathtt{ab}, \mathtt{ba}, \mathtt{b}^2, \mathtt{b}^3, \mathtt{b}^2\mathtt{ab}^2, \mathtt{b}^2\mathtt{ab}^2\mathtt{ab}^2\right\}$. For each element w in $\mathcal{W}^{short}$ we have to prove that $1 + \frac{|w|}{|v|} \leq 1 + \frac{1}{6}$, i.e., that $\frac{|w|}{|v|} \leq \frac{1}{6}$, where v is a shortest return word to w.

Let $\pi(w) = \mathtt{a}$, which is not a bispecial factor in $\mathbf{u}$. We use Proposition 3. Looking into the prefix of $\mathbf{u}$ (as in Example 3) we see that the return words to $\mathtt{a}$ are $r = \mathtt{ab}^2$ and $s = \mathtt{ab}^3$. By Definition 4, each vector $\left(\begin{smallmatrix}\ell\\k\end{smallmatrix}\right) \in \mathcal{S}(\mathtt{a})$ satisfies $k + \ell \geq 1$ and $k\left(\begin{smallmatrix}1\\2\end{smallmatrix}\right) + \ell\left(\begin{smallmatrix}1\\3\end{smallmatrix}\right) = \left(\begin{smallmatrix}0\\0\end{smallmatrix}\right) \mod \left(\begin{smallmatrix}2\\1\end{smallmatrix}\right)$ since $\mathrm{gap}(\mathbf{y}, 1) = \{2\}$ and $\mathrm{gap}(\mathbf{y}', 0) = \{1\}$. This implies for each solution that $k + \ell \geq 2$. Since $3k + 4\ell \geq 3k + 3\ell \geq 6$, we have $\frac{|w|}{|v|} = \max\left\{\frac{|\mathtt{a}|}{|\mathtt{ab}^2|k+|\mathtt{ab}^3|\ell} : \left(\begin{smallmatrix}\ell\\k\end{smallmatrix}\right) \in \mathcal{S}(\mathtt{a})\right\} \leq \max\left\{\frac{1}{3k+4\ell} : k + \ell \geq 2\right\} \leq \frac{1}{6}$.
On the other hand, $\frac{|w|}{|v|} \geq \frac{1}{3\cdot 2+4\cdot 0} = \frac{1}{6}$ as the solution $\left(\begin{smallmatrix}\ell\\k\end{smallmatrix}\right) = \left(\begin{smallmatrix}0\\2\end{smallmatrix}\right)$ is the Parikh vector of a factor of any Sturmian sequence, in particular of $\mathbf{d_u}(\mathtt{a})$. Note that the value $1/6$ is attained for any w with $\pi(w) = \mathtt{a}$. For instance we can consider $w = \mathtt{0}$ and $v = \mathtt{045167}$.

Let $\pi(w) = \mathtt{ab}$. Again, $\mathcal{R}_\mathbf{u}(\mathtt{ab}) = \{\mathtt{ab}^2, \mathtt{ab}^3\}$. We have $\mathrm{gap}(\mathbf{y}, 1) = \{2\}$ and $\mathrm{gap}(\mathbf{y}', 1) = \{6, 8\}$. Thus $\left(\begin{smallmatrix}\ell\\k\end{smallmatrix}\right) \in \mathcal{S}(\mathtt{ab})$ satisfies

$$k\left(\begin{smallmatrix}1\\2\end{smallmatrix}\right) + \ell\left(\begin{smallmatrix}1\\3\end{smallmatrix}\right) = \left(\begin{smallmatrix}0\\0\end{smallmatrix}\right) \mod \left(\begin{smallmatrix}2\\6 \text{ or } 8\end{smallmatrix}\right). \tag{8}$$

If $k + \ell \geq 4$, then $\frac{|\mathtt{ab}|}{|\mathtt{ab}^2|k+|\mathtt{ab}^3|\ell} = \frac{2}{3k+4\ell} \leq \frac{2}{3k+3\ell} \leq \frac{2}{3\cdot 4} = \frac{1}{6}$.
When $1 \leq k + \ell \leq 3$, the only vector $\left(\begin{smallmatrix}\ell\\k\end{smallmatrix}\right)$ satisfying Equation (8) is $\left(\begin{smallmatrix}2\\0\end{smallmatrix}\right)$. However, this is never the Parikh vector of a Sturmian factor (cf. Lemma 3).

Let $\pi(w) = \mathtt{b}^2\mathtt{ab}^2$. Then $b = \mathtt{b}^2\mathtt{ab}^2$ is a bispecial factor of $\mathbf{u}$ associated with $(N, m) = (1, 1)$. We have $\mathrm{gap}(\mathbf{y}, 1) = \{2\}$ and $\mathrm{gap}(\mathbf{y}', 4) = \{24\}$. By Proposition 4 we know the Parikh vectors of r and s, thus $\left(\begin{smallmatrix}\ell\\k\end{smallmatrix}\right) \in \mathcal{S}(b)$ satisfies

$$k\left(\begin{smallmatrix}1\\2\end{smallmatrix}\right) + \ell\left(\begin{smallmatrix}1\\3\end{smallmatrix}\right) = \left(\begin{smallmatrix}0\\0\end{smallmatrix}\right) \mod \left(\begin{smallmatrix}2\\24\end{smallmatrix}\right). \tag{9}$$

It is not difficult to see that $\frac{5}{3k+4\ell} \leq \frac{1}{6}$.

Similar computations show that $\frac{|w|}{|v|} \leq \frac{1}{6}$ for each $w \in \mathcal{W}^{short}$ and $v \in \mathcal{R}_\mathbf{v}(w)$.

$\mathcal{W}^{long}$: From Example 6 it follows that $E^*(\mathbf{x}_9) \doteq 1,1306 < \frac{7}{6}$. Apply Proposition 6. We have $\mu = (\sqrt{2}-1)^8$. Since $\mu|Q_1 - L_0Q_2| = \mu|3 - (\sqrt{2}-1)10| \leq 2(\sqrt{2}-1)$ and $|Q_2 - L_0Q_3| = |10 - (\sqrt{2}-1)23| \leq 2(\sqrt{2}-1)$, we have $I(2+i+NH, m) \leq E^*(i, m) \leq E^*(\mathbf{x}_9)$ for all i, m and N besides $i = 0, N = 0$, i.e., we have to consider separately the bispecial factors associated with the pairs $(2, 0)$ and $(2, 1)$. Again, both $I(2, 0)$ and $I(2, 1)$ are smaller than $\frac{7}{6}$.

A similar computation can be done for the sequence $\mathbf{x}_{10}$. In this case we can show that $E(\mathbf{x}_{10}) = 1 + \frac{1}{7}$ and that the value $\frac{8}{7}$ is attained for instance for $w = \mathtt{2}$ and $v = \mathtt{2345067}$.

References

1. A. R. Baranwal, *Decision Algorithms for Ostrowski-Automatic Sequences*, master thesis, University of Waterloo, http://hdl.handle.net/10012/15845 (2020).
2. A. R. Baranwal and J. Shallit, *Critical exponent of infinite balanced words via the Pell number system*, Proceedings WORDS 2019, LNCS vol. 11682, 80–92 (2019).
3. A. Carpi and A. de Luca, *Special factors, periodicity, and an application to Sturmian words*, Acta Informatica 37, 986–1006 (2000).
4. D. Damanik and D. Lenz, *The index of Sturmian sequences*, European Journal of Combinatorics, 23, 23–29 (2002).
5. F. Dolce, L'. Dvořáková and E. Pelantová, *On balanced sequences and their asymptotic critical exponent*, Proceedings LATA 2021, LNCS vol. 12638, 293–304 (2021).
6. L'. Dvořáková, K. Medková and E. Pelantová, *Complementary symmetric Rote sequences: the critical exponent and the recurrence function*, Discrete Mathematics & Theoretical Computer Science, 20 (1) (2020).
7. G. A. Hedlund and M. Morse, *Symbolic dynamics II - Sturmian trajectories*, American Journal of Mathematics, 62, 1–42 (1940).
8. P. Hubert, *Suites équilibrées*, Theoretical Computer Science, 242, 91–108 (2000).
9. N. Rampersad, J. Shallit and É. Vandomme, *Critical exponents of infinite balanced words*, Theoretical Computer Science, 777, 454–463 (2019).
10. L. Vuillon, *A characterization of Sturmian words by return words*, European Journal of Combinatorics, 22, 263–275 (2001).