Repetitions in beta-integers

Ľ.Balková (l.balkova@centrum.cz), K.Klouda (karel@kloudak.eu) and E.Pelantová (edita.pelantova@fjfi.cvut.cz)
Doppler Institute and Department of Mathematics, FNSPE, Czech Technical University in Prague

Abstract. Classical crystals are solid materials containing arbitrarily long periodic repetitions of a single motif. In this Letter, we study the maximal possible repetition of the same motif occurring in β-integers – one dimensional models of quasicrystals. We are interested in β-integers realizing only a finite number of distinct distances between neighboring elements. In such a case, the problem may be reformulated in terms of combinatorics on words as a study of the index of infinite words coding β-integers. We will solve a particular case for β being a quadratic non-simple Parry number.

Keywords: repetitions in words, powers of factors, index of infinite words, beta-integers, quasicrystals, aperiodic structures

MSC(2000): 11R06, 52C23, 11B05, 68R15

1. Introduction

This Letter takes up the study of β-integers initiated by the investigation of their asymptotic properties in [1]. Similarly as in the previous Letter, we restrict our consideration to β-integers realizing only a finite number of distinct distances between neighbors; β is then called a Parry number. For Parry numbers, the set of β-integers forms a discrete aperiodic Delone set with a self-similarity factor β and of finite local complexity. It follows herefrom that β-integers are suitable for modeling materials with aperiodic long range order, the so-called quasicrystals [2]. Classical crystals are solid materials containing arbitrarily long periodic repetitions of a single motif. Quasicrystals do not share this property.

In this Letter, we are interested in the maximal possible repetition of one motif occurring in β-integers. It turns out to be suitable to reformulate and study this problem in terms of combinatorics on words.

For Parry numbers, coding distinct distances between neighboring nonnegative β-integers with distinct letters, one obtains a right-sided infinite word u_β over a finite alphabet. The reformulation of our task in the language of combinatorics on words has the following reading: For a given factor w of the infinite word $u = u_\beta$, find the longest prefix v of the infinite periodic word $w^\omega = wwwwww\ldots$ such that v occurs as a factor in $u = u_\beta$. The ratio of the lengths of v and w is called the index of the factor w in $u = u_\beta$ and is denoted by $\mathrm{ind}(w)$. Let us note that $\mathrm{ind}(w)$ is not necessarily an integer. Denote by k the lower integer part $\lfloor \mathrm{ind}(w) \rfloor$ of the index of w, then the word w^k, i.e., the concatenation of k words w, is usually called the maximal integer power of w.

The index of any infinite word u can be naturally defined as

$$\mathrm{ind}(u) = \sup\{\mathrm{ind}(w) \mid w \text{ factor of } u\}.$$

Explicit values of the index are known only for few classes of infinite words. The index of Sturmian words has been studied in many papers [3, 4, 5, 6], the complete solution to the problem was provided independently by Carpi and de Luca [7] and by Damanik and Lenz [8]. Recently, the index of infinite words has reinforced its importance: Damanik in [9] considers discrete one-dimensional Schrödinger operators with aperiodic potentials generated by primitive morphisms and he establishes purely singular continuous spectrum with probability one provided that the potentials (infinite words) have the index greater than three. Let us stress that infinite words u_β associated with Parry numbers belong to the class of infinite words generated by primitive morphisms, too.

The study of spectra of discrete Schrödinger operators with aperiodic potentials generated by primitive substitutions has been in the center of attention of mathematical physicists since the publication of the paper of A. Hof, O. Knill, B. Simon from 1995 [10]. They have shown how the spectra are connected with one of the combinatorial characteristics of infinite words, namely the presence of arbitrarily long palindromes. Thus, the index of infinite words is another combinatorial characteristics for the investigation of spectra. It confirms that combinatorics on words has found new applications in mathematical physics.

Here we study the index of infinite words u_β for quadratic non-simple Parry numbers β. These words are determined by integer parameters p, q, where $p > q \geq 1$. We provide an explicit formula for $\mathrm{ind}(u_\beta)$. In the particular case of $p = q + 1$, the infinite word u_β is Sturmian and our result may be deduced also from the well-known formula for the index of Sturmian words. We have chosen the word u_β associated with quadratic non-simple Parry numbers β for our study of the index of non-Sturmian words since for such infinite words, we dispose of detailed knowledge on arithmetical properties of β-integers and combinatorial properties of the associated infinite words u_β [11, 12].

The Letter is organized in the following way. In Section 2, we introduce necessary notions from combinatorics on words and we cite a relevant result on the index of Sturmian words. In Section 3, we provide the background on infinite words u_β coding β-integers for β being a non-simple quadratic Parry number. In Section 4, we determine the maximal integer power occurring in u_β (Theorem 4.6). Section 5 is devoted to the index of u_β (Theorem 5.3) and to the comparison of our result with the formula for the index of Sturmian words.

2. Preliminaries

An *alphabet* $\mathcal{A}$ is a finite set of symbols, called *letters*. Throughout this paper, the binary alphabet $\mathcal{A} = \{0, 1\}$ is used. The string $w = w_1 w_2 \dots w_k$, where $w_i \in \mathcal{A}$ for each $i = 1, 2, \dots, k$, is called a *word of length* k on $\mathcal{A}$. The length of w is then k and it is denoted by $|w| = k$. The set of all finite words together with the operation of concatenation forms a monoid; its neutral element is the empty word ε. We denote this monoid $\mathcal{A}^*$. An infinite sequence $u = u_0 u_1 u_2 \dots$ of symbols from the alphabet $\mathcal{A}$ is called an *infinite word*. A finite word w is said to be a *factor* of the (finite or infinite) word v if there exists a finite word v' and a finite or infinite word v'' such that $v = v'wv''$. If v' is the empty word, then w is called a *prefix* of v, if v'' is the empty word, then w is a *suffix* of v. If $v = v'w$, then vw^{-1} denotes the word v' obtained from v by erasing its suffix w. The set of all factors of an infinite word u is said to be the *language* of u and is denoted $\mathcal{L}(u)$. An infinite word u is called *recurrent* if every of its factors occurs infinitely many times in u and u is called *uniformly recurrent* if for every of its factors w, the set of all factors in $\mathcal{L}(u)$ that do not contain w as their factor is finite. In other words, every sufficiently long element of $\mathcal{L}(u)$ contains w as its factor.

The number of factors of the infinite word u gives us insight into its variability. The function $\mathcal{C}_u : \mathbb{N} \mapsto \mathbb{N}$ that to every n associates the number of distinct factors of length n occurring in u is called the *factor complexity* of the infinite word u. An infinite periodic word $u = www\dots$, where w is a finite word, is usually denoted w^ω. Its factor complexity $\mathcal{C}_u$ is bounded; it is readily seen that $\mathcal{C}_u(n) \leq |w|$. Similarly, the factor complexity of an eventually periodic word $u = w'w^\omega$, where w', w are finite words, is bounded. The necessary and sufficient condition for an infinite word to be aperiodic is the validity of the equation $\mathcal{C}_u(n) \geq n+1$ for all $n \in \mathbb{N}$ [13]. Infinite aperiodic words satisfying $\mathcal{C}_u(n) = n+1$ for all $n \in \mathbb{N}$ are called *Sturmian words*; these are thus infinite aperiodic words of the lowest possible factor complexity. Sturmian words are the best known aperiodic words; a survey on their properties may be consulted in [14]. In particular, any Sturmian word is uniformly recurrent.

For determination of the factor complexity of an infinite word u, an essential role is played by special factors. We recall that a factor w of an infinite word u over a binary alphabet $\{0,1\}$ is called *left special* if $0w$ and $1w$ are both factors of u, w is called *right special* if $w0$ and $w1$ are both factors of u, and w is said to be *bispecial* if w is both left special and right special.

The crucial notion of our study is the index of a factor w in a given infinite word u. Let us define first the integer power of w. For any $k \in \mathbb{N}$, the k-th *power* of w is the concatenation of k words w, usually denoted w^k. Analogously for $r \in \mathbb{Q}, r \geq 1$, we call a word v the r-th *power* of the word w if there exists a proper prefix w' of w such that

$$v = \underbrace{w \dots w}_{\lfloor r \rfloor - times} w' \quad \text{and} \quad r = \lfloor r \rfloor + \frac{|w'|}{|w|}.$$

For instance, $(011)^{2\frac{1}{3}} = 0110110$. The r-th power of w is denoted w^r.

Our aim is to find, for a given factor w of an infinite word u, the highest power of w occurring in u. We will be interested exclusively in aperiodic uniformly recurrent words. Such words contain for every factor w only r-th powers of w with r bounded by a constant depending on w, see [15]. Therefore, for aperiodic uniformly recurrent words u, it makes sense to define the index of w in u as

$$\operatorname{ind}(w) = \max\{r \in \mathbb{Q} \mid w^r \in \mathcal{L}(u)\}.$$

The word $w^{\operatorname{ind}(w)}$ is called the *maximal power* of w in u, the word $w^{\lfloor \operatorname{ind}(w) \rfloor}$ the *maximal integer power* of w in u. The index of the infinite word u is defined as

$$\operatorname{ind}(u) = \sup\{\operatorname{ind}(w) \mid w \in \mathcal{L}(u)\}.$$

Let us remark that an aperiodic uniformly recurrent word u can have an infinite index; even among Sturmian words, one can find words with an infinite index. The language of every Sturmian word is characterized by an irrational parameter $\alpha \in (0,1)$, called *slope*. If α is the slope of a Sturmian word u, then the word obtained by exchanging letters in u has the slope $1-\alpha$ and has evidently the same index as u. Consequently, we may assume without loss of generality that $\alpha > \frac{1}{2}$. In order to determine the index of a Sturmian word, we need to express α in the form of its continued fraction. Since $\frac{1}{2} < \alpha < 1$, the continued fraction of α equals $[0; 1, a_2, a_3, \dots]$. The results of [7] and [8] say that the index of a Sturmian word u with slope α is equal to

$$\operatorname{ind}(u) = 2 + \sup\Big\{a_{n+2} + \frac{q_n - 2}{q_{n+1}} \mid n \in \mathbb{N}\Big\}, \tag{1}$$

where q_n is the denominator of the n-th convergent of α.

Now, let us describe a large class of uniformly recurrent words: fixed points of primitive morphisms. This class includes infinite words u_β associated with Parry numbers β. The map $\varphi : \mathcal{A}^* \mapsto \mathcal{A}^*$ is called a *morphism* if $\varphi(wv) = \varphi(w)\varphi(v)$ for every $w, v \in \mathcal{A}^*$. One may associate with φ the *morphism matrix* M_φ satisfying

$$(M_\varphi)_{ab} = \text{number of letters } b \text{ occurring in } \varphi(a),$$

for any pair of letters $a, b \in \mathcal{A}$.

Knowing for any $a \in \mathcal{A}$ the number of letters a occurring in a factor w, we may obtain the same information for $\varphi(w)$ by a simple formula. We mention the formula only for the binary alphabet $\mathcal{A} = \{0, 1\}$ we are interested in. It follows from the definition of the morphism matrix that for every factor $w \in \mathcal{A}^*$

$$(|\varphi(w)|_0, |\varphi(w)|_1) = (|w|_0, |w|_1)M_\varphi. \tag{2}$$

where $|v|_a$ denotes the number of letters a occurring in a word v. Clearly, a similar formula holds for any finite alphabet.

A morphism is said to be *primitive* if a power of M_φ has all elements strictly positive. In other words, matrices of primitive morphisms fulfill the assumptions of the Perron-Frobenius theorem [16].

The action of the morphism φ may be naturally extended to an infinite word $u = u_0u_1u_2\ldots$ by

$$\varphi(u_0u_1u_2\ldots) = \varphi(u_0)\varphi(u_1)\varphi(u_2)\ldots$$

An infinite word u is called a *fixed point* of φ if $\varphi(u) = u$. It is known that any fixed point of a primitive morphism is uniformly recurrent and that any left eigenvector corresponding to the dominant eigenvalue of M_φ is proportional to the densities of letters in any fixed point of φ [17].

3. Infinite words associated with β-integers

Here, we provide the description of infinite words u_β associated with quadratic non-simple Parry numbers in terms of fixed points of primitive morphisms. We keep the notation from our precedent Letter [1], where the number theoretical background on β-integers and associated infinite words u_β is available. Nevertheless, to make this Letter self-contained, we will recall all notions needed for understanding of our results. In this section we also deduce some important properties of u_β, in particular, a transformation generating bispecial factors of u_β. Bispecial factors turn out to be essential for our main aim – determination of the maximal integer powers of factors and determination of the index of u_β.

A non-simple quadratic Parry number β is the larger root of $x^2 - (p+1)x + p - q$, where $p > q \geq 1$. Its Rényi expansion of unity is $d_\beta(1) = pq^\omega$. The set of β integers has two distances between neighbors: $\Delta_0 = 1$ and $\Delta_1 = \beta - p$. Consequently, the infinite word u_β coding distances between neighboring β-integers is binary.

As we have already said, the reader may find the notions of the Rényi expansion of unity, β-integers, distances between neighbors in $\mathbb{Z}_\beta$ etc. in our precedent Letter [1]. However, in order to follow the ideas in the sequel, it is sufficient to know that u_β is the unique fixed point of a morphism canonically associated with parameters p, q characterizing non-simple quadratic Parry numbers β. Therefore we will use the result of [18] for an equivalent definition of u_β.

Definition 3.1. *Let β is the larger root of $x^2 - (p+1)x + p - q$, where $p > q \geq 1$. The unique fixed point of the morphism*

$$\varphi(0) = 0^p1, \quad \varphi(1) = 0^q1 \tag{3}$$

will be denoted u_β.

The infinite word u_β starts as follows

$$u_\beta = \underbrace{\underbrace{0^p1\ \ldots\ 0^p1}_{p \text{ times}}\ 0^q1\ \underbrace{0^p1\ \ldots\ 0^p1}_{p \text{ times}}\ 0^q1\ \ldots\ \underbrace{0^p1\ \ldots\ 0^p1}_{p \text{ times}}\ 0^q1}_{p \text{ times}}\ \underbrace{0^p1\ \ldots\ 0^p1}_{q \text{ times}} 0^q1\ \ldots \tag{4}$$

The morphism matrix M_φ is $\left(\begin{smallmatrix} p & 1 \\ q & 1 \end{smallmatrix}\right)$ and φ is thus obviously primitive. Computing the left eigenvector of M_φ corresponding to the dominant eigenvalue β, we get the densities $1 - \frac{1}{\beta}$ and $\frac{1}{\beta}$ of letters 0 and 1, respectively.

Remark 3.2. *In the paper [11], it is shown that the factor complexity $\mathcal{C}$ of u_β satisfies*

$$\text{if } \ p > q+1, \quad \text{then } \ \{\mathcal{C}(n+1) - \mathcal{C}(n) \mid n \in \mathbb{N}\} = \{1, 2\},$$

$$\text{if } \ p = q+1, \quad \text{then } \ \{\mathcal{C}(n+1) - \mathcal{C}(n) \mid n \in \mathbb{N}\} = \{1\}.$$

Therefore, u_β is Sturmian if and only if $p = q + 1$.

First of all, some simple, but very important properties of the morphism φ are observed.

Observation 3.3. *Let 10^k1 be a factor of u_β, then $k = p$ or $k = q$.*

Observation 3.4. *Let v be any factor of u_β containing at least one 1. Then there exists k_1, $0 \le k_1 \le p$, such that $0^{k_1}1$ is a prefix of v and there exists k_2, $0 \le k_2 \le p$, such that 10^{k_2} is a suffix of v. The fact that $\varphi(0)$ and $\varphi(1)$ end in 1 and contain only one letter 1 implies that there exists a unique word w in $\{0,1\}^*$ satisfying $v = 0^{k_1}1\varphi(w)0^{k_2}$. Clearly, w is a factor of u_β.*

Of significant importance is the map $T : \{0,1\}^* \to \{0,1\}^*$ defined by

$$T(w) = 0^q1\varphi(w)0^q. \tag{5}$$

The map T helps to generate bispecial factors that play a crucial role in the determination of the index of factors. Therefore, the rest of this section is devoted to the description of properties of T.

Lemma 3.5. *Let T be the map defined in (5).*

1. *For every $w \in \mathcal{L}(u_\beta)$, it holds that $T(w) \in \mathcal{L}(u_\beta)$.*

2. *Let w be a factor of u_β and let $a, b \in \mathcal{A}$, then $awb \in \mathcal{L}(u_\beta)$ if and only if $aT(w)b \in \mathcal{L}(u_\beta)$.*

3. *Let v be a bispecial factor of u_β containing at least one letter 1, then there exists a unique factor w such that $v = T(w)$.*

4. *Let w, v be factors of u_β, then w is a prefix of v if and only if $T(w)$ is a prefix of $T(v)$.*

5. *Let w, v be factors of u_β, then w is a suffix of v if and only if $T(w)$ is a suffix of $T(v)$.*

Proof. 1. Take an arbitrary factor $w \in \mathcal{L}(u_\beta)$. Then w is extendable to the right, and, since u_β is recurrent, w is also extendable to the left. In other words, there exists $a, b \in \{0,1\}$ such that awb is also a factor of u_β. As u_β is a fixed point of φ, the image $\varphi(awb)$ belongs to $\mathcal{L}(u_\beta)$. Finally, $T(w)$ is a factor of u_β because $T(w)$ is a subword of $\varphi(awb)$.

2. Let $1w1$ be a factor of $\mathcal{L}(u_\beta)$, then $01w1$ is as well a factor of u_β. Applying φ, we learn that $\varphi(01w1) = \varphi(0)T(w)1$ is a factor of u_β, which proves that $1T(w)1$ belongs to $\mathcal{L}(u_\beta)$. The other cases $0w0, 0w1, 1w0$ are analogous.

Let $0T(w)1 \in \mathcal{L}(u_\beta)$, using Observation 3.3, the word $v = 10^p1\varphi(w)0^q1$ is also a factor of u_β. Applying Observation 3.4, $v = 1\varphi(0w1)$ and $0w1$ is an element of $\mathcal{L}(u_\beta)$. All the other cases $0T(w)0, 1T(w)0, 1T(w)1$ are similar.

3. Observation 3.3 implies that each bispecial factor v containing at least one letter 1 has the prefix 0^q1 and the suffix 10^q. According to Observation 3.4, there exists a unique w such that $v = T(w)$.

4. The implication $\Rightarrow$ is obvious noticing that 0^q is a prefix of $\varphi(a)$ for $a \in \{0,1\}$. The opposite implication $\Leftarrow$ follows taking into account that $\varphi(1)$ is not a prefix of $\varphi(0)$ and $\varphi(0)$ is not a prefix of $\varphi(1a)$ for any $a \in \{0,1\}$.

5. The implication $\Rightarrow$ is obvious noticing that 0^q1 is a suffix of $\varphi(a)$ for $a \in \{0,1\}$. The opposite implication $\Leftarrow$ follows taking into account that $1\varphi(1)$ is not a suffix of $\varphi(0)$ and $\varphi(0)$ is not a suffix of $\varphi(x1)$ for any $x \in \{0,1\}^*$. □

4. Integer powers in u_β

Even if we want to describe the maximal integer powers of factors of u_β, it turns out to be useful to study first the relation between bispecial factors and the maximal rational powers of factors.

Lemma 4.1. *Let u be an infinite uniformly recurrent word over an alphabet $\mathcal{A}$. Let $w^k w'$ be its factor for some proper prefix w' of w and some positive integer k. Let us denote by* **P1**, **P2**, **P3** *the following statements:*

P1 *The factor w has the maximal index in u among all factors of u with the same length $|w|$ and $w^k w'$ is the maximal power of w in u.*

P2 *There exist $a, b \in \{0,1\}$ such that*

$$aw^kw'b \in \mathcal{L}(u) \quad \text{and} \quad w'b \text{ is not a prefix of } w \quad \text{and} \quad a \text{ is not a suffix of } w.$$

P3 *All the following factors are bispecial:*

$$w', ww', www', \ldots, w^{k-1}w'.$$

Then **P1** *implies* **P2** *and* **P2** *implies* **P3**.

Proof. **P**1 $\Rightarrow$ **P**2 : As u is recurrent, there exists a such that aw^kw' is a factor of u. Since w^kw' is the maximal power of w, the letter a is not a suffix of w, otherwise the factor awa^{-1} (usually called a conjugate of w) would have a larger index than w. On the other hand, if $w^kw'b$ is a factor of w, then $w'b$ is not a prefix of w, otherwise it contradicts the fact that w^kw' is the maximal power of w in u.

P2 $\Rightarrow$ **P**3 : Since w' is a proper prefix of w, there exists $x \in \mathcal{A}$ such that $w'x$ is a prefix of w. Denote by y the last letter of w. Obviously, $x \neq b$ and $y \neq a$. As $aw^jw'x$ is a prefix of $aw^kw'b$ and $yw^jw'b$ is a suffix of $aw^kw'b$, both $aw^jw'x$ and $yw^jw'b$ are in $\mathcal{L}(u)$ for all $j,\ 0 \le j \le k-1$. It follows that all factors listed in **P3** are bispecial. □

In this section, our aim is to describe the maximal integer powers occurring in u_β. Since the letter 0 has the maximal index p, we may restrict our consideration to k-th powers of factors of u_β with $k \geq p$. Crucial for the determination of the index of u_β are Propositions 4.2 and 4.5.

Proposition 4.2. *Let p, q be integers, $p > q \geq 1$, and u_β be the fixed point of the morphism (3). Assume $p > 3$. Let w be a factor of u_β containing at least two 1s and w' be a proper prefix of w. Denote $v = w^k w'$ for some $k \in \mathbb{N}$, $k \geq p$. If there exist $a, b \in \{0, 1\}$ so that*

$$avb \in \mathcal{L}(u_\beta) \quad \text{and} \quad w'b \text{ is not a prefix of } w \quad \text{and} \quad a \text{ is not a suffix of } w,$$

then there exist a unique $\tilde{w}$ of length ≥ 2 and a proper prefix $\tilde{w}'$ of $\tilde{w}$ such that

$$w = 0^q 1 \; \varphi(\tilde{w})(0^q 1)^{-1} \quad \text{and} \quad v = T(\tilde{v}) = T(\tilde{w}^k \tilde{w}'); \tag{6}$$

moreover,

$$a\tilde{v}b \in \mathcal{L}(u_\beta) \quad \text{and} \quad \tilde{w}'b \text{ is not a prefix of } \tilde{w} \quad \text{and} \quad a \text{ is not a suffix of } \tilde{w}.$$

In order to prove Proposition 4.2, we will use the following lemma.

Lemma 4.3. *Let p, q be integers, $p > q \geq 1$, and u_β be the fixed point of the morphism (3). The following statements hold:*

1. *If $0(x1)^\ell x0 \in \mathcal{L}(u_\beta)$ for some integer $\ell \geq 2$, then $\ell = 2$ and $x = 0^q$.*
2. *If $1(x0)^\ell x1 \in \mathcal{L}(u_\beta)$ for some integer $\ell \geq p-1$ and $p \leq 2q$, then x is the empty word ε.*

Proof. 1. At first, we exclude the case when x contains a non-zero letter. Suppose that the letter 1 occurs in x. Since the factors $0x1$ and $1x0$ belong to $\mathcal{L}(u_\beta)$, it follows that x is bispecial. By Lemma 3.5 Item *3.*, x starts in 0^q1 and ends in 10^q. Therefore $10^q10^q1 \in \mathcal{L}(u_\beta)$. As $10^q10^q1 = 1\varphi(11)$, we have according to Observation 3.4 that $11 \in \mathcal{L}(u_\beta)$ – a contradiction.

Now consider $x = 0^s$ for some $s \in \mathbb{N}$. Then $0x1, 1x1 \in \mathcal{L}(u_\beta)$, which implies by Observation 3.3 that $s = q$. If ℓ was at least 3, then $1x1x1 = 10^q10^q1 \in \mathcal{L}(u_\beta)$, which leads to the same contradiction as before.

2. Again we start with the case of x containing the letter 1. Since factors $1x0$ and $0x1$ belong to the language $\mathcal{L}(u_\beta)$, it follows that x is bispecial. Hence x starts in 0^q1 and ends in 10^q. Since $10^q00^q1 \in \mathcal{L}(u_\beta)$, by Observation 3.3, we have $p = 2q + 1$. This contradicts the assumption $p \leq 2q$.

Suppose now that $x = 0^s$ for some $s \in \mathbb{N}$. Since $1(x0)^\ell x1 = 1(0^{s+1})^\ell 0^s 1$ is a factor of u_β, Observation 3.3 gives that $(s+1)\ell + s \leq p$, which is impossible if $s \geq 1$ and $\ell \geq p - 1$. Therefore $s = 0$ and x is the empty word.

□

Proof of Proposition 4.2. The factor w contains at least two 1s. Since w and $v = w^k w'$ satisfy Item **P2** of Lemma 4.1, both ww' and www' are bispecial, and therefore start in 0^q1 and end in 10^q. Consequently, their form is $ww' = T(x)$ and $www' = T(y)$, where $x, y \neq \varepsilon$. According to Lemma 3.5 Item *5.*, x is a suffix of y, i.e., $y = zx$ for some $z \neq \varepsilon$. Observing $ww' = 0^q1\varphi(x)0^q$ and $www' = 0^q1\varphi(z)\varphi(x)0^q$, it follows directly that $w = 0^q1\varphi(z)(0^q1)^{-1}$.

Let us, at first, show that

1. either z is a prefix of x,
2. or $z = x1$,
3. or $z = x0$.

Assume $z = tdz'$ and $x = t(1-d)x'$ for a word $t \in \{0,1\}^*$ and for a letter d. Then $w = 0^q1\varphi(t)\varphi(d)\varphi(z')(0^q1)^{-1}$ is a prefix of $ww' = 0^q1\varphi(t)\varphi(1-d)\varphi(x')0^q$. If $z' \neq \varepsilon$, we have a contradiction immediately. If $z' = \varepsilon$, then $t \neq \varepsilon$ knowing that z contains at least two letters (w contains at least two 1s).

- If $d = 1$, then $w = 0^q1\varphi(t)$ and $ww' = 0^q1\varphi(t)0^p1\varphi(x')0^q$, thus w' starts in 0^p1, which is not a prefix of w – a contradiction.
- If $d = 0$, then $w = 0^q1\varphi(t)0^{p-q}$ and $ww' = 0^q1\varphi(t)0^{p-q}0^{2q-p}1\varphi(x')0^q$, thus w' starts in $0^{2q-p}1$, which is not a prefix of w because $p \neq q$ – a contradiction again.

The situation $z = xz'$ for some z' of length ≥ 2 cannot occur because it implies $|w| > |ww'|$. Consequently, one of the situations $1., 2.,$ or 3. occurs.

1. If z is a prefix of x, i.e., $x = zx''$, then $ww' = 0^q1\varphi(z)\varphi(x'')0^q$ and $w = 0^q1\varphi(z)(0^q1)^{-1}$, thus $w' = 0^q1\varphi(x'')0^q$. Then $v = w^kw' = 0^q1\varphi(z^kx'')0^q$, therefore $\widetilde{w} = z$, $\widetilde{w}' = x''$. As w' is a proper prefix of w, it follows by Lemma 3.5 Item *4* that $\widetilde{w}'$ is a proper prefix of $\widetilde{w}$.
2. Assume $z = x1$, then $w = 0^q1\varphi(x)$ and $ww' = 0^q1\varphi(x)0^q$, thus $w' = 0^q$. Then $v = 0^q1\varphi((x1)^{k-1}x)0^q$. Since $w'0$ is not a prefix of w and 0 is not a suffix of w, the assumptions imply that $0v0 \in \mathcal{L}(u_\beta)$. By Observation 3.3, we have $1\varphi(0(x1)^{k-1}x0) \in \mathcal{L}(u_\beta)$. Since $k-1 \geq 3$, we deduce by Lemma 4.3 that $0(x1)^{k-1}x0$ is not a factor of u_β. It contradicts Observation 3.4. Hence, the case $z = x1$ does not occur.
3. If $z = x0$, then $w = 0^q1\varphi(x)0^{p-q}$ and $ww' = 0^q1\varphi(x)0^q$, thus $w' = 0^{2q-p}$, which can happen only for $p \leq 2q$. Then $v = 0^q1\varphi((x0)^{k-1}x)0^q$. Since $w'1$ is not a prefix of w and 1 is not a suffix of w, the assumptions imply that $1v1 \in \mathcal{L}(u_\beta)$. Hence, $1\varphi(1(x0)^{k-1}x1) \in \mathcal{L}(u_\beta)$. However, by Lemma 4.3, it follows that $x = \varepsilon$. Then $z = 0$, which contradicts the condition $|z| \geq 2$. Thus, the case $z = x0$ does not occur.

Let us finally note that the very last statement on the extensions of $\widetilde{v}$ follows from Lemma 3.5 Items *2.*, *4.*, and *5.* ☐

Remark 4.4. *Proposition 4.2 does not take into account u_β given by parameters $p = 2, q = 1$, $p = 3, q = 2$, and $p = 3, q = 1$. In the two first cases, u_β is a Sturmian word. Therefore, exclusion of the first two cases does not mean any loss. For the case of $p = 3, q = 1$, in the proof of Proposition 4.2, we cannot exclude the situation 2.; in this case, Lemma 4.3 Item 1. implies either the validity of (6) or of*

$$w = 01\varphi(01)(01)^{-1} \quad v = T(01010).$$

Proposition 4.2 thus claims that for every factor $w \in \mathcal{L}(u_\beta)$ containing at least two 1s such that its k-th power w^k is a factor of u_β with $k \geq p$, there exists a shorter factor $\tilde{w}$ such that its k-th power $\tilde{w}$ is also in the language of u_β. As a consequence, in order to determine the maximal integer power present in u_β, it is sufficient to study the index of factors w containing only one letter 1.

Proposition 4.5. *Let p, q be integers, $p > q \geq 1$, and u_β be the fixed point of the morphism (3). Let w be a factor of u_β containing one letter 1 and of the maximal index $ind(w)$ among all factors of length $|w|$ and such that $ind(w) \geq p \geq 3$. Denote $k := \lfloor ind(w) \rfloor$ and $v = w^k w'$ the maximal power of w. Then*

$$w = 0^q 1\ \varphi(0)(0^q1)^{-1} \quad \textit{and} \quad v = T(0^p)$$

and

$$ind(w) = p + \frac{2q+1}{p+1}.$$

Proof. According to Lemma 4.1, ww' is a bispecial factor. Lemma 3.5 Item *3.* claims that ww' starts in 0^q1. Therefore $w = 0^q10^s$ with $s \in \{0, p-q\}$ (Observation 3.3). The case $s = 0$ does not occur since $www = 0^q10^s0^q10^s0^q10^s \in \mathcal{L}(u_\beta)$ and $0^q10^q10^q1 = 0^q1\varphi(11)$, but $11 \notin \mathcal{L}(u_\beta)$ – a contradiction to Observation 3.4. Hence $w = 0^q10^{p-q} = 0^q1\varphi(0)(0^q1)^{-1}$. It remains to determine the form of v. Again, since ww' is bispecial, ww' ends in 10^q. As w' is a prefix of w, at most one 1 occurs in w'.

- Suppose w' contains 1, then w' starts in 0^q1 and ends in 10^q, thus $w' = 0^q10^q$. This is possible only in case when $p-q-1 \geq q$, i.e., $p \geq 2q+1$. Then, $v = (0^q10^{p-q})^k0^q10^q$. Consequently, $v = 0^q1\varphi(0^k)0^q = T(0^k)$. On one hand, Observations 3.3 and 3.4 imply that $k \leq p$. On the other hand, since v is the maximal power of w, it follows that $k \geq p$.

- Assume w' does not contain 1. Since ww' ends in 10^q and $w = 0^q10^{p-q}$, the only possibility for w' is $w' = 0^{2q-p}$. This comes in question only for $p \leq 2q$. Then $v = (0^q10^{p-q})^k0^{2q-p} = 0^q1\varphi(0^{k-1})0^q = T(0^{k-1})$. The same arguments as in the previous case imply that $k - 1 = p$.

Clearly, $\mathrm{ind}(w) = \frac{|v|}{|w|} = p + \frac{2q+1}{p+1}$. □

Let us state the main result of this section.

Theorem 4.6. *Let p, q be integers, $p > q \geq 1$, and u_β be the fixed point of the morphism (3). Assume $p \geq 3$.*

- *If $p \leq 2q$, then there exists a factor $w \neq \varepsilon$ satisfying $w^{p+1} \in \mathcal{L}(u_\beta)$ and no $(p+2)$-nd power of any factor belongs to the language $\mathcal{L}(u_\beta)$.*

- *If $p > 2q$, then there exists a factor $w \neq \varepsilon$ satisfying $w^p \in \mathcal{L}(u_\beta)$ and no $(p+1)$-st power of any factor belongs to the language $\mathcal{L}(u_\beta)$.*

Proof. Proposition 4.2 implies that in order to determine the maximal integer power present in u_β, we can restrict our consideration to powers of factors containing only one letter 1. When we compute the integer part $\lfloor \mathrm{ind}(w) \rfloor$ of such factors w in Proposition 4.5, we find that the maximum is $p+1$ if $p \leq 2q$ and p otherwise. □

5. Index of u_β

The task of this section is to compute the index of u_β, i.e.,

$$\mathrm{ind}(u_\beta) = \sup\{\mathrm{ind}(w) \mid w \in \mathcal{L}(u_\beta)\}.$$

We already know that $\mathrm{ind}(u_\beta) \geq p$. Using Lemma 4.1, it suffices to study rational powers $v = w^k w'$ of factors w with the property **P2**. As a direct consequence of Propositions 4.2 and 4.5, we have the following corollary.

Corollary 5.1. *Let p, q be integers, $p > q \geq 1$, and u_β be the fixed point of the morphism (3). Assume $p > 3$. The index of u_β is given by the following formula*

$$ind(u_\beta) = \sup\{ind(w^{(n)}) \mid n \in \mathbb{N}\},$$

where

$$w^{(0)} = 0, \qquad w^{(n+1)} = 0^q 1 \varphi(w^{(n)})(0^q 1)^{-1}. \tag{7}$$

Moreover, the maximal power of $w^{(n)}$ is $v^{(n)}$, where

$$v^{(0)} = 0^p, \qquad v^{(n+1)} = T(v^{(n)}). \tag{8}$$

In the sequel, let us determine the index of $w^{(n)}$ for every $n \in \mathbb{N}$.

Lemma 5.2. *The number of 0s and 1s in the words $w^{(n)}$ and $v^{(n)}$ satisfy*

$$(|w^{(n)}|_0, |w^{(n)}|_1) = (1, 0)M_\varphi^n,$$

$$(|v^{(n)}|_0, |v^{(n)}|_1) = (p+1, \tfrac{2q+1-p}{q})M_\varphi^n - (1, \tfrac{2q+1-p}{q}),$$

where $M_\varphi = \left(\begin{smallmatrix} p & 1 \\ q & 1 \end{smallmatrix}\right)$ is the morphism matrix.

Proof. As $w^{(n)}$ is a conjugate of $\varphi(w^{(n-1)})$, the first formula holds by (2). Let us show the second one by induction on n.
For $n = 0$,

$$(|v^{(0)}|_0, |v^{(0)}|_1) = (p, 0) = (p+1, \tfrac{2q+1-p}{q}) - (1, \tfrac{2q+1-p}{q}).$$

For $n > 0$,

$$(|v^{(n)}|_0, |v^{(n)}|_1) = (|v^{(n-1)}|_0, |v^{(n-1)}|_1)M_\varphi + (2q, 1) =$$

by the induction assumption,

$$= \left[(p+1, \tfrac{2q+1-p}{q})M_\varphi^{n-1} - (1, \tfrac{2q+1-p}{q})\right] M_\varphi + (2q, 1) = (p+1, \tfrac{2q+1-p}{q})M_\varphi^n - (1, \tfrac{2q+1-p}{q}).$$

$\square$

Since the eigenvalues β and β' of M_φ are roots of the Parry polynomial $x^2 - (p+1)x + (p-q)$, it is straightforward to show that $\vec{x_1} = (\beta - 1, 1)$ is a left eigenvector of M_φ corresponding to β and $\vec{x_2} = (\beta' - 1, 1)$ is a left eigenvector of M_φ corresponding to β'. The index of $w^{(n)}$ may be expressed as follows

$$\mathrm{ind}(w^{(n)}) = \frac{|v^{(n)}|}{|w^{(n)}|} = \frac{(p+1, 0)M_\varphi^n \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right) + (0, \frac{2q+1-p}{q})M_\varphi^n \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right) - (1, \frac{2q+1-p}{q}) \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)}{(1, 0)M_\varphi^n \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)},$$

$$\mathrm{ind}(w^{(n)}) = p + 1 + \frac{\frac{2q+1-p}{q}(\alpha_1 \vec{x_1} + \alpha_2 \vec{x_2})M_\varphi^n \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right) - \frac{3q+1-p}{q}}{(\gamma_1 \vec{x_1} + \gamma_2 \vec{x_2})M_\varphi^n \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)},$$

where $\alpha_1 \vec{x_1} + \alpha_2 \vec{x_2} = (0, 1)$ and $\gamma_1 \vec{x_1} + \gamma_2 \vec{x_2} = (1, 0)$. Using the fact that $\vec{x_1}$ and $\vec{x_2}$ are eigenvectors of M_φ, we have

$$\mathrm{ind}(w^{(n)}) = p + 1 + \frac{\frac{2q+1-p}{q}\left(\alpha_1 \beta^n \vec{x_1} \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right) + \alpha_2 \beta'^n \vec{x_2} \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)\right) - \frac{3q+1-p}{q}}{\gamma_1 \beta^n \vec{x_1} \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right) + \gamma_2 \beta'^n \vec{x_2} \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)}.$$

It is easy to calculate that $\alpha_1 = \frac{1-\beta'}{\beta-\beta'}$, $\alpha_2 = \frac{\beta-1}{\beta-\beta'}$, $\gamma_1 = \frac{1}{\beta-\beta'}$, $\gamma_2 = \frac{-1}{\beta-\beta'}$, and $\vec{x_1}\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right) = \beta$, $\vec{x_2}\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right) = \beta'$. The final formula for the index of $w^{(n)}$ has the form

$$\begin{aligned}\mathrm{ind}(w^{(n)}) &= p+1+\frac{\frac{2q+1-p}{q}\left((1-\beta')\beta^{n+1}-(1-\beta)\beta'^{n+1}\right)-\frac{3q+1-p}{q}(\beta-\beta')}{\beta^{n+1}-\beta'^{n+1}} = \\ &= p+1+\frac{2q+1-p}{q}(1-\beta')+\underbrace{\frac{\beta-\beta'}{q(\beta^{n+1}-\beta'^{n+1})}}_{>0}\underbrace{\Big((2q+1-p)\beta'^{n+1}-(3q+1-p)\Big)}_{\mathrm{A}(n)}.\end{aligned}$$

Using the fact $0 < \beta' < 1 < \beta$, we determine the limit

$$\lim_{n\to\infty}\mathrm{ind}(w^{(n)}) = p+1+\frac{2q+1-p}{q}(1-\beta') = p+1+\frac{2q+1-p}{\beta-1}.$$

This limit is the supreme of $\{\mathrm{ind}(w^{(n)}) \mid n \geq 0\}$ if and only if $\mathrm{A}(n) < 0$ for all $n \in \mathbb{N}$. It is an easy exercise to show that $\mathrm{A}(n) = (2q+1-p)(\beta'^n - 1) - q < 0$ for all $n \in \mathbb{N}$ if and only if $p \leq 3q+1$, otherwise $\mathrm{A}(n) > 0$ for all sufficiently large n.

Let us sum up the results in a theorem.

Theorem 5.3. *Let p, q be integers, $p > q \geq 1$, and u_β be the fixed point of the morphism (3). Assume $p > 3$. Then the index of u_β satisfies*
for $p \leq 3q+1$

$$ind(u_\beta) = p+1+\frac{2q+1-p}{\beta-1},$$

otherwise there exists $n_0 \in \mathbb{N}$ such that

$$ind(u_\beta) = ind(w^{(n_0)}) > p+1+\frac{2q+1-p}{\beta-1}.$$

Remark 5.4. *Similarly as in the previous section, we have to treat the case of $p = 3$ and $q = 1$ separately. According to Remark 4.4, we have to determine the index of $(w^{(n)})$ defined in (7), but moreover the index of $(\hat{w}^{(n)})$ defined recursively by*

$$\hat{w}^{(0)} = 01\varphi(01)(01)^{-1}, \qquad \hat{w}^{(n+1)} = 0^q 1\varphi(\hat{w}^{(n)})(0^q 1)^{-1}.$$

Using the same technique as before, we obtain

$$\sup\{ind(\hat{w}^{(n)}) \mid n \in \mathbb{N}\} \quad = \quad \beta \quad < \quad 4 \quad = \quad \sup\{ind(w^{(n)}) \mid n \in \mathbb{N}\}.$$

Hence, Theorem 5.3 holds in fact also in this case.

At the conclusion, let us compare in case of Sturmian words u_β the formula for $\mathrm{ind}(w^{(n)})$ with the formula (1) for the index of general Sturmian words. As we have already stated, u_β is Sturmian if and only if $p = q+1$, i.e., β is the larger root of the polynomial $x^2 - (p+1)x + 1$. For such β, we have

$$\mathrm{ind}(u_\beta) = p+1+\frac{2p-1}{\beta-1} = \beta+1\,.$$

In order to apply the formula from (1), we need to determine the slope α of the Sturmian word u_β. Since $\alpha > \frac{1}{2}$ is the density of the more frequent letter, according to

Section 3, $\alpha = 1 - \frac{1}{\beta}$. Let us use some basic properties of continued fractions available at any book on Number Theory to determine the continued fraction of this value. Since

$$1 - \frac{1}{\beta} = \frac{1}{1 + \frac{1}{\beta - 1}} = \frac{1}{1 + \frac{1}{p-1+1-\frac{1}{\beta}}},$$

one obtains $\beta = [0; 1, (p-1), 1, (p-1), \ldots] = [0; \overline{1, (p-1)}]$. Denominators q_n of the convergents of β fulfill therefore the following recurrent relations

$$q_{2n+1} = (p-1)q_{2n} + q_{2n-1} \qquad \text{and} \qquad q_{2n} = q_{2n-1} + q_{2n-2}$$

with initial values $q_1 = 1,\ q_2 = p,\ q_3 = p+1$. By mathematical induction on n, it may be shown easily that

$$q_{2n-1} = \frac{1}{\beta - \beta'}\Big(\beta^n - \beta'^n\Big) \qquad \text{and} \qquad q_{2n} = \frac{1}{\beta - \beta'}\Big((1-\beta')\beta^{n+1} - (1-\beta)\beta'^{n+1}\Big).$$

As it holds for coefficients of the continued fraction of β that $a_{2n-1} = 1$ and $a_{2n} = p-1$, it suffices to consider even n in (1). We obtain then finally

$$a_{2n+2} + 2 + \frac{q_{2n} - 2}{q_{2n+1}} = p + 1 + \frac{\big((1-\beta')\beta^{n+1} - (1-\beta)\beta'^{n+1}\big) - 2(\beta - \beta')}{\beta^{n+1} - \beta'^{n+1}},$$

which is exactly $\operatorname{ind}(w^{(n)})$. This result holds for all parameters p, q satisfying $p = q+1$, even for $p \leq 3$. Consequently, Theorem 5.3 is in fact valid for all parameters p, q with $p > q \geq 1$.

Acknowledgements

The authors acknowledge financial support by Czech Science Foundation GA ČR 201/05/0169, by the grants MSM6840770039 and LC06002 of the Ministry of Education, Youth, and Sports of the Czech Republic.

References

1. Balková, L'., Gazeau, J.-P., Pelantová, E.: Asymptotic behavior of beta-integers, Letters in Mathematical Physics 84 179–198 (2008)
2. Frougny, Ch., Gazeau, J.-P., Krejcar, R.: Additive and multiplicative properties of point-sets based on beta-integers, Theor. Comp. Sci. 303 491–516 (2003)
3. Berstel, J.: On the index of Sturmian words, in: Jewels are forever, Springer 287–294 (1999)
4. Justin, J., Pirillo, G.: Fractional powers in Sturmian words, Theoret. Comput. Sci. 255 363-376 (2001)
5. Mignosi, F., Pirillo, G.: Repetitions in the Fibonacci infinite word, RAIRO Inform. Theor. Appl. 26 199-204 (1992)
6. Masáková, Z., Pelantová, E.: Relation between powers of factors and recurrence function characterizing Sturmian words, submitted, arXiv 0809.0603v2[math.CO]
7. Carpi, A., de Luca, A.: Special factors, periodicity, and an apllication to Sturmian words, Acta Inform. 36 983–1006 (2000)
8. Damanik, D., Lenz, D.: The index of Sturmian sequences, European J. Combin. 23 23–29 (2002)
9. Damanik, D.: Singular continuous spectrum for a class of substitution Hamiltonians II, Letters in Mathematical Physics 54 25–31 (2000)

10. Hof, A. and Knill, O. and Simon, B.: Singular continuous spectrum for palindromic Schrödinger operators, Comm. Math. Phys. 1 149–159 (1995)
11. Balková, L'., Masáková, Z.: Palindromic complexity of infinite words associated with non-simple Parry numbers, RAIRO - Theoret. Informatics Appl. 43 145–163 (2009)
12. Balková, L'., Pelantová, E., Turek, O.: Combinatorial and arithmetical properties of infinite words associated with quadratic non-simple Parry numbers, RAIRO - Theoret. Informatics Appl. 41 307–328 (2007)
13. Hedlund, G. A., Morse, M.: Symbolic dynamics, Amer. J. Math. 60 815–866(1938)
14. Berstel, J.: Sturmian and episturmian words (a survey of some recent results), in: S. Bozapalidis, G. Rahonis (eds), Conference on Algebraic Informatics, Thessaloniki, Lecture Notes Comput. Sci. 4728 23–47 (2007)
15. Lothaire, M.: Algebraic combinatorics on words, volume 90 of Encyclopedia of Mathematics and its Applications, Cambridge University Press (2002)
16. Minc, H.: Nonnegative matrices, John Wiley&Sons, New York (1988)
17. Queffélec, M.: Substitution dynamical systems—spectral analysis, Lecture Notes in Mathematics, Springer-Verlag (1987)
18. Fabre, S.: Substitutions et β-systèmes de numération, Theoret. Comput. Sci. 137 219–236 (1995)