Available online at www.sciencedirect.com
ScienceDirect
Theoretical Computer Science 389 (2007) 12–25
Theoretical Computer Science
www.elsevier.com/locate/tcs

On a class of infinite words with affine factor complexity

Julien Bernat[a], Zuzana Masáková[b,*], Edita Pelantová[b]

[a] Institut de Mathématiques de Luminy - CNRS UMR 6206 163, avenue de Luminy, case 907, 13288 Marseille Cedex 09, France
[b] Doppler Institute for Mathematical Physics and Applied Mathematics and Department of Mathematics, FNSPE, Czech Technical University, Trojanova 13, 120 00 Praha 2, Czech Republic

Received 2 January 2007; received in revised form 11 June 2007; accepted 1 July 2007

Communicated by D. Perrin

Abstract

In this article, we consider the fixed point of a primitive substitution canonically defined by a β-numeration system. The problem of determination of the factor complexity of such an infinite word has been solved only partially. Here we provide a necessary and sufficient condition on the Rényi expansion of one for having an affine factor complexity map $\mathcal{C}(n)$, that is, such that $\mathcal{C}(n) = an + b$ for any positive integer n.

Keywords: Beta-expansions; Factor complexity of infinite words

1. Introduction

Factor complexity is one of the basic properties which is studied on infinite words $(u_n)_{n\in\mathbb{N}}$ over a finite alphabet $\mathcal{A}$. It is the function $\mathcal{C} : \mathbb{N} \to \mathbb{N}$, which counts the number of factors of a given length which occur in an infinite word.[1] In other words, factor complexity expresses the measure of irregularity in the word.

For eventually periodic words, the factor complexity is an eventually constant function. As shown by Morse and Hedlund [8], an infinite word $(u_n)_{n\in\mathbb{N}}$ which is not eventually periodic, i.e. is aperiodic, has factor complexity satisfying $\mathcal{C}(n) \geq n + 1$ for all $n \in \mathbb{N}$. Moreover, the language of the factors of an infinite word is factorial, hence one has $\mathcal{C}(n + m) \leq \mathcal{C}(n)\mathcal{C}(m)$ for all $n, m \in \mathbb{N}$. It is therefore obvious that not every function $\mathcal{C}$ can represent the factor complexity of an infinite word. For an overview of necessary conditions for a factor complexity function $\mathcal{C}$ see [4].

Aperiodic words with minimal complexity $\mathcal{C}(n) = n + 1$, for all $n \in \mathbb{N}$, are called Sturmian; their properties have been studied by many authors, see [6]. On the other hand, words having maximal complexity satisfy $\mathcal{C}(n) = m^n$, where m is the cardinality of the alphabet. Under the term infinite words of low factor complexity, one usually understands words for which $\mathcal{C}$ is a sublinear function, i.e. there exist constants a, b such that $\mathcal{C}(n) \leq an + b$ for all positive integers n. A special subclass is formed by infinite words with affine complexity, i.e. such that $\mathcal{C}(n) = an + b$ for

* Corresponding author. Tel.: +420 224358544.
E-mail addresses: bernat@iml.univ-mrs.fr (J. Bernat), masakova@km1.fjfi.cvut.cz (Z. Masáková), pelantova@km1.fjfi.cvut.cz (E. Pelantová).

[1] By $\mathbb{N}$ we denote the set of nonnegative integers $\mathbb{N} = \{0, 1, 2, \dots\}$.

0304-3975/$ - see front matter
doi:10.1016/j.tcs.2007.07.001

all $n \in \mathbb{N}$. Among the words with affine factor complexity, one finds Sturmian words, Arnoux–Rauzy words, words coding generic interval exchange transformation, and others.

As shown in [9], fixed points of a primitive substitution have low factor complexity. Let us mention that, relaxing the assumption of primitivity, the factor complexity is bounded by a quadratic function, see [10]. The determination of the factor complexity of a fixed point from the prescription of the substitution is not a simple task.

In this paper we consider canonical substitutions associated with simple Parry numbers β. These are numbers whose Rényi expansion of 1 is finite, i.e. is of the form $d_\beta(1) = t_1 \cdots t_m$. The canonical substitution corresponding to β is the substitution over the alphabet $\mathcal{A} = \{0, 1, \dots, m-1\}$ given by

$$\begin{aligned} \varphi(0) &= 0^{t_1}1 \\ \varphi(1) &= 0^{t_2}2 \\ &\vdots \\ \varphi(m-2) &= 0^{t_{m-1}}(m-1) \\ \varphi(m-1) &= 0^{t_m}. \end{aligned} \tag{1}$$

Since $t_1 \geq 1$ and $t_m \geq 1$, one easily checks that for any letter i, $\varphi^{2m}(i)$ contains at least one occurrence of each letter, hence the substitution is primitive. Moreover, the substitution φ admits a unique fixed point, which is the infinite word

$$u_\beta := \lim_{n\to\infty} \varphi^n(0).$$

In [5], the factor complexity of such fixed points is determined for substitutions satisfying the condition $t_1 > \max\{t_2, \dots, t_{m-1}\}$. In particular, it is shown that

$$(m-1)n + 1 \leq \mathcal{C}(n) \leq mn, \quad \text{for all } n \geq 1.$$

In the same paper it is shown that the word u_β is Arnoux–Rauzy, if and only if $t_m = 1$ and $t_1 = t_2 = \cdots = t_{m-1}$. In this case the factor complexity is equal to $(m-1)n + 1$, which is an affine function.

The aim of this article is the characterization of substitutions of the form (1), for which the fixed point u_β has affine factor complexity. We will show

Theorem 1.1. *Let β be a simple Parry number with the Rényi expansion of unity $d_\beta(1) = t_1 \cdots t_m$, and let u_β be the fixed point of the substitution (1). Then the factor complexity of u_β is an affine function if and only if the coefficients $t_1, \dots, t_m$ satisfy*

(1) $t_m = 1$
(2) *If there exists a nonempty word which is simultaneously a proper suffix and a proper prefix of $t_1 \cdots t_{m-1}$ then $t_1 \cdots t_{m-1} = w^k$ for some word w and $k \in \mathbb{N}$, $k \geq 2$.*

If both conditions are satisfied, then the complexity function equals $\mathcal{C}(n) = (m-1)n + 1$ for all $n \in \mathbb{N}$.

Note that infinite words u_β which are Arnoux–Rauzy (i.e. $t_m = 1$ and $t_1 = t_2 = \cdots = t_{m-1}$), satisfy Condition (2) of the above theorem with a one-letter word $w = \lfloor \beta \rfloor$. Condition (2) is satisfied also by other words u_β, which are not Arnoux–Rauzy, but have the same complexity $\mathcal{C}(n) = (m-1)n + 1$. These words illustrate the fact that Arnoux–Rauzy words of order $m \geq 3$ cannot be characterized by their complexity, as is the case for Arnoux–Rauzy words of order $m = 2$, i.e. Sturmian words.

Example 1.2. Consider fixed points u_β of substitution (1) over the alphabet $\{0, 1, \dots, 4\}$ with parameters $t_1, \dots, t_5$.

- If $t_1t_2t_3t_4t_5 = 32321$ then both conditions of Theorem 1.1 are satisfied, since $t_m = t_5 = 1$ and $t_1 \cdots t_4 = (32)^2$. The infinite word u_β has affine complexity $\mathcal{C}(n) = 4n + 1$, although it is not an Arnoux–Rauzy word.
- If $t_1t_2t_3t_4t_5 = 32221$ then both conditions of Theorem 1.1 are satisfied as well, since $t_m = t_5 = 1$ and there exists no nonempty word which is simultaneously a proper prefix and a proper suffix of $t_1 \cdots t_4 = 3222$. Again, the complexity of the corresponding u_β is affine and u_β is not an Arnoux–Rauzy word.
- If $t_1t_2t_3t_4t_5 = 32231$ then Condition (2) is not satisfied, since 3 is a prefix and a suffix of $t_1t_2t_3t_4 = 3223$ but $t_1t_2t_3t_4$ is not an integer power of any word.
- If $t_1t_2t_3t_4t_5 = 32322$ then complexity of u_β is not affine, since Condition (1) is not satisfied.

In order to prove that Conditions (1) and (2) of Theorem 1.1 are sufficient for affine factor complexity, we use purely the tools of combinatorics on words. For the opposite implication, we use the geometrical representation of the factors of the word u_β as coding of patterns occurring in the set of β-integers, see Section 2.

2. Preliminaries

2.1. β-numeration

In [12] the author introduces and studies the properties of the positional number system with base $\beta \in \mathbb{R}$, $\beta > 1$. For arbitrary real $x > 0$, the β-expansion of x can be found by the greedy algorithm, as follows. There exists a unique $k \in \mathbb{Z}$ such that $\beta^k \le x < \beta^{k+1}$. Set $x_k := \lfloor x/\beta^k \rfloor$ and $r_k := x - x_k\beta^k$. For each $i < k$, set $x_i := \lfloor \beta r_{i+1} \rfloor$ and $r_i := \beta r_{i+1} - x_i$. Obviously,

$$x = x_k\beta^k + x_{k-1}\beta^{k-1} + x_{k-2}\beta^{k-2} + \cdots \tag{2}$$

and $x_i \in \{0, 1, \ldots, \lceil\beta\rceil - 1\}$. Note that the elements of the sequence $(x_i)_{i\le k}$ satisfy the relation $x_i = \left\lfloor \beta T_\beta^{k-i}(x\beta^{-(k+1)}) \right\rfloor$, where the map T_β is defined by

$$T_\beta : [0, 1] \to [0, 1), \qquad T_\beta(x) = \beta x \mod 1. \tag{3}$$

For the expression of x in the form of its *β-expansion* (2) we use the notation $x = x_k \cdots x_0 \bullet x_{-1}x_{-2}\cdots$, if $k \ge 0$, or $x = 0 \bullet \underbrace{000\cdots0}_{-k-1 \text{ times}} x_k x_{k-1} \cdots$ if $k < 0$. If the β-expansion ends in infinitely many 0's, we omit them. These expansions are said to be finite.

Numbers x with vanishing β-fractional part, i.e. such that $x_i = 0$ for $i < 0$ are called nonnegative *β-integers* and we denote them $x = x_k \cdots x_1x_0\bullet$. The set of nonnegative β-integers is denoted by $\mathbb{Z}_\beta^+$, and the set of β-integers is defined as $\mathbb{Z}_\beta = \mathbb{Z}_\beta^+ \cup (-\mathbb{Z}_\beta^+)$.

Unlike the situation with integer base β, in case that $\beta \notin \mathbb{N}$, there exist sequences $(x_i)_{i\le k}$, $x \in \{0, 1, \cdots, \lceil\beta\rceil - 1\}$ that are not the β-expansion of some $x > 0$. For the description of admissible sequences of digits, one needs the so-called Rényi expansion of one. For $\beta \in \mathbb{R}$, $\beta > 1$, put $t_1 := \lfloor\beta\rfloor$ and let $0 \bullet t_2t_3t_4\cdots$ be the β-expansion of the number $\beta - \lfloor\beta\rfloor$. Then the sequence $d_\beta(1) = t_1t_2t_3\cdots$ is called the *Rényi expansion of* 1. We have obviously,

$$1 = \sum_{i=1}^{\infty} \frac{t_i}{\beta^i} \quad \text{and} \quad t_i \in \{0, 1, \ldots, \lceil\beta\rceil - 1\}.$$

In order that a sequence $t_1t_2t_3\cdots$ of integers be the Rényi expansion of 1 for some base β, the so-called Parry condition must be satisfied [11],

$$t_it_{i+1}t_{i+2}\cdots \prec t_1t_2t_3\cdots \quad \text{for all } i \in \mathbb{N},\ i \ge 2, \tag{4}$$

where the symbol $\prec$ stands for 'strictly lexicographically smaller'. In the same paper [11] it is shown that a finite sequence of digits $x_kx_{k-1}\cdots x_1x_0$ over the alphabet $\mathcal{A} = \{0, 1, \ldots, \lceil\beta\rceil - 1\}$ is the β-expansion of a β-integer if and only if

$$x_ix_{i-1}\cdots x_0 \prec d_\beta(1) \quad \text{for all } i \in \mathbb{N},\ i \le k. \tag{5}$$

Using the Rényi expansion of 1, one can even describe the distances between consecutive β-integers on the real line. If $\beta \in \mathbb{N}$, the β-integers are precisely the rational integers, therefore the distance between consecutive β-integers is always 1. The situation is very different if the base β is not an integer. The distances between consecutive β-integers are the elements of $\left\{T_\beta^i(1) \mid i \in \mathbb{N}\right\}$, see [13]. Note that, since $\mathbb{Z}_\beta$ is a discrete set for any $\beta > 1$, one may define the successor and predecessor maps, respectively as

$$\mathrm{pred}(x) = \max\{y \in \mathbb{Z}_\beta \mid y < x\} \quad \text{and} \quad \mathrm{succ}(x) = \min\{y \in \mathbb{Z}_\beta \mid y > x\}.$$

Example 2.1. Consider the base $\beta = \frac{1}{2}(1 + \sqrt{5})$, i.e., β is the golden ratio. The number β is a root of the equation $x^2 = x+1$, and its Rényi expansion is equal to $d_\beta(1) = 11$. Condition (5) implies in this case that $x_k \cdots x_0 \in \{0, 1\}^{k+1}$ is the β-expansion of a β-integer if and only if $x_ix_{i-1} \ne 11$ for all $i = 1, \ldots, k$. The set of β-integers thus starts with the numbers (written in their β-expansion)

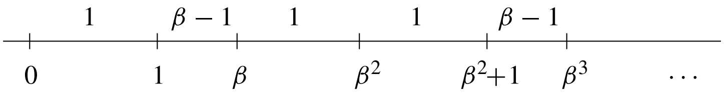

Fig. 1. The set of β-integers for $\beta = \frac{1}{2}(1+\sqrt{5})$ drawn on the real line.

$$0\bullet, \quad 1\bullet, \quad 10\bullet, \quad 100\bullet, \quad 101\bullet, \quad 1000\bullet, \quad \text{etc.}$$

The distances between consecutive β-integers take only two values, namely $T_\beta^0(1) = 1$ and $T_\beta(1) = \beta - 1$, see Fig. 1.

Numbers β, for which the Rényi expansion $d_\beta(1)$ is eventually periodic, are called *Parry numbers*. In this case the number of values for distances between consecutive β-integers is finite. In other words, Parry numbers are numbers for which the elements of the sequence $(T_\beta^i(1))_{i\in\mathbb{N}}$ take finitely many distinct values. If moreover $d_\beta(1) = t_1t_2\cdots t_m$, $t_m \neq 0$, then β is called a *simple* Parry number; one has $T_\beta^i(1) = 0\bullet t_{i+1}\cdots t_m$ for any $i \in \{1,\ldots,m-1\}$. Associating with these distances letters in the alphabet $\mathcal{A} = \{0,1,\ldots,m-1\}$ in a natural way, $T_\beta^i(1) \mapsto i$, the sequence of β-integers can be coded by an infinite word over $\mathcal{A}$. This infinite word is the fixed point of the substitution φ defined by (1).

Example 2.2. The infinite word u_β for $\beta = \frac{1}{2}(1+\sqrt{5})$ starts with

$$u_\beta = 01001\cdots$$

In [3] it is shown that the infinite word u_β is a fixed point of a canonical substitution (1) associated with β. Note that a canonical substitution can be associated also with a nonsimple Parry number β, see [3]. For more details about the properties of β-numeration we refer to [6].

2.2. *Combinatorics on words*

Let $\mathcal{A} = \{0,1,\ldots,m-1\}$ be a finite alphabet. A finite concatenation $w = w_0w_1\cdots w_{n-1}$ of the letters is called a *word*, its length n is denoted by $|w|$. The set of finite words over an alphabet $\mathcal{A}$ together with the empty word ε and the concatenation operation forms a free monoid, denoted by $\mathcal{A}^*$.

The sequence $u = (u_i)_{i\in\mathbb{N}}$ of letters in the alphabet $\mathcal{A}$ is called an infinite word. A word w is a *factor* of a word u (finite or infinite) if there exist words $w^{(1)}$ and $w^{(2)}$ such that $u = w^{(1)}ww^{(2)}$. If $w^{(1)}$ is an empty word, then w is a prefix of u. If $w^{(2)} = \varepsilon$, then w is a suffix of u. The set of all factors of an infinite word u is called the *language* of u and denoted by $\mathcal{L}(u)$. The set of all factors of u of length n is denoted by $\mathcal{L}_n(u)$. Obviously $\mathcal{L}(u) = \bigcup_{n\in\mathbb{N}}\mathcal{L}_n(u)$. The cardinality of the set $\mathcal{L}_n(u)$ is the *factor complexity*, i.e., the function $\mathcal{C} : \mathbb{N} \to \mathbb{N}$, given by

$$\mathcal{C}(n) := \#\mathcal{L}_n(u).$$

Note that any language which is the set of factors of an infinite word is extendable, that is, every factor $w_0\cdots w_{n-1}$ of length n can be extended in at least one way to a factor $w_0\cdots w_{n-1}w_n$ of length $n+1$. Hence the factor complexity is a nondecreasing function. The set of letters by which it is possible to extend a factor w to the right is called the *right extension* of w,

$$\mathrm{Rext}(w) = \{a \in \mathcal{A} \mid wa \in \mathcal{L}(u)\}.$$

The increment of complexity can be calculated using the number of right extensions of all factors of length n,

$$\Delta\mathcal{C}(n) = \mathcal{C}(n+1) - \mathcal{C}(n) = \sum_{w\in\mathcal{L}_n(u)} \big(\#\mathrm{Rext}(w) - 1\big).$$

A factor w, for which $\#\mathrm{Rext}(w) \geq 2$ is called a *right special factor*. Only such factors are important for the determination of the first difference of factor complexity.

In this paper we study *recurrent words*. These are infinite words, in which every factor appears at least twice. Factors of a recurrent word can be extended in at least one way to the left, and so all the above considerations can be analogously stated. In particular, we have

$$\Delta\mathcal{C}(n) = \sum_{w\in\mathcal{L}_n(u)} \big(\#\mathrm{Lext}(w) - 1\big), \tag{6}$$

where $\mathrm{Lext}(w) = \{a \in \mathcal{A} \mid aw \in \mathcal{L}(u)\}$. Factors with $\#\mathrm{Lext}(w) \geq 2$ are called *left special*. Factors which are both right special and left special are called *bispecial*.

A *substitution* is a morphism on the free monoid $\mathcal{A}^*$, i.e. a mapping $\varphi : \mathcal{A}^* \to \mathcal{A}^*$ satisfying $\varphi(wv) = \varphi(w)\varphi(v)$ for every pair $w, v \in \mathcal{A}^*$. It is obvious that a substitution is uniquely determined by images of all letters $a \in \mathcal{A}$. The action of a substitution can be naturally extended to infinite words $(u_n)_{n\in\mathbb{N}}$ as

$$\varphi(u) = \varphi(u_0u_1u_2\cdots) := \varphi(u_0)\varphi(u_1)\varphi(u_2)\cdots.$$

An infinite word u satisfying $u = \varphi(u)$ is a fixed point of the substitution φ. We are concerned with substitutions satisfying $\varphi(a) \neq \varepsilon$ for all $a \in \mathcal{A}$, and such that there exists a_0 which is a proper prefix of $\varphi(a_0)$. Such substitution is called *nonerasing prolongable* substitution and has at least one fixed point, namely $\lim_{n\to\infty}\varphi^n(a_0)$. A substitution is called *primitive*, if there exists $k \in \mathbb{N}$ such that, for every pair of letters $a, b \in \mathcal{A}$, the letter a appears in the word $\varphi^k(b)$. It is known [2] that a fixed point of a primitive substitution is a linearly recurrent word, which implies that the distances between consecutive occurrences of a given factor are bounded.

3. Affine factor complexity of infinite words u_β

Our aim is to describe the substitutions of the form (1) whose fixed points

$$u_\beta = \underbrace{0^{t_1}1\,0^{t_1}1\,\cdots\,0^{t_1}1}_{t_1 \text{ times}}\,0^{t_2}2\cdots$$

have affine factor complexity, i.e. the first difference $\Delta\mathcal{C}(n)$ is constant. For the determination of $\Delta\mathcal{C}(n)$ we use the left special factors of u_β. In [5] it is shown that every prefix w of the infinite word u_β is a left special factor and its left extension is $\mathrm{Lext}(w) = \mathcal{A} = \{0, 1, \ldots, m-1\}$. Therefore using (6) we have $\Delta\mathcal{C}(n) \geq m - 1$ for every $n \in \mathbb{N}$.

Infinite words whose every prefix is a left special factor are called left special branches [1]. As we have mentioned, u_β is a left special branch of itself. In [5] it is, moreover, shown that u_β has no other left special branch.

For the description of left special factors of another type (i.e. which are not prefixes of a left special branch) we use a lemma from [5].

Definition 3.1. Let $d_\beta(1) = t_1t_2\cdots t_m$. For $2 \leq k \leq m$ we denote

$$j_k := \min\{i \in \mathbb{N} \mid 1 \leq i \leq k-1,\ t_{k-i} \neq 0\}.$$

Note that the index j_k always exists, because $t_1 > 0$.

Lemma 3.2. *All factors of u_β of the form $X0^rY$, where X, Y are non-zero letters and $r \in \mathbb{N}$, are the following,*

$$j_k0^{t_k}k, \qquad \textit{for } k = 2, 3, \ldots, m-1,$$

$$k\,0^{t_1}1, \qquad \textit{for } k = 1, 2, \ldots, m-1,$$

$$j_m0^{t_1+t_m}1.$$

This lemma is exactly Lemma 4.5 in [5].

Remark 3.3. (i) Since $j_k \leq k-1$, the only factors of the form $(m-1)0^rY$ are, according to Lemma 3.2, the factors $(m-1)0^{t_1}1$ and possibly $j_m0^{t_1+t_m}1$. In any case, the letter $(m-1)$ is always followed by the letter 0.

(ii) Recall that for parameters $t_1, \ldots, t_m$ of the substitution it holds that $t_m \geq 1$, and from the Parry condition $t_1 \geq t_i$ for all $i = 2, \ldots, m$.

Corollary 3.4. *Every left special factor w with $|w| \leq t_1$ is a prefix of u_β.*

Proof. We prove the statement by contradiction. Let w be a left special factor satisfying $|w| \le t_1$, and suppose that w is not a prefix of u_β. Since u_β has a prefix 0^{t_1}, necessarily w is of the form $0^r Y$ for some $Y \neq 0$, $r \in \mathbb{N}$, $r < t_1$. If $Y = 1$, then from Lemma 3.2 we know that w has a unique left extension, namely 0, and thus cannot be a left special factor. If $Y > 1$, then again, w has a unique left extension, namely 0, if $r < t_Y$, or j_k, if $r = t_Y$. □

Proposition 3.5. *Let β be a simple Parry number. The infinite word u_β has affine factor complexity if and only if every left special factor is a prefix of u_β.*

Proof. Since every prefix w of u_β satisfies $\#\mathrm{Lext}(w) = m$, Corollary 3.4 implies that $\Delta\mathcal{C}(n) = m - 1$ for all $n \le t_1$. If u_β has affine factor complexity, then $\Delta\mathcal{C}(n) = m - 1$ for all $n \in \mathbb{N}$, and so no left special factors other than prefixes of u_β can exist. The opposite implication is obvious. □

Corollary 3.6. *If u_β has affine factor complexity, then $t_m = 1$.*

Proof. Suppose that $t_m \ge 2$. Then according to Lemma 3.2, the word $0^{t_1+t_m-1}$ is a left special factor, because it has two distinct left extensions, namely zero and j_m. In the same time, $0^{t_1+t_m-1}$ is not a prefix of u_β. Proposition 3.5 implies that the factor complexity of u_β is not an affine function. □

In [5] it is shown that under the conditions

$$\text{(a)}\quad t_m = 1 \qquad\qquad \text{(b)}\quad t_1 = t_2 = \cdots = t_{m-1} \quad \text{or} \quad t_1 > \max\{t_2, \dots, t_{m-1}\},$$

the factor complexity of u_β is affine. Note that the condition (b) is a very special case of Condition (2) of Theorem 1.1, whose proof is the aim of this paper.

Definition 3.7. A left special factor w of an infinite word u is called maximal if for any letter $a \in \mathcal{A}$ the word wa is not a left special factor of u.

If $t_m \ge 2$, then $0^{t_1+t_m-1}$ is maximal, since extending it to the right using Lemma 3.2, we do not obtain a left special factor. Let us mention that if w is a maximal left special factor, then it is a bispecial factor: since w is left special, there exist $X_1, X_2 \in \mathcal{A}$ such that $X_1w, X_2w \in \mathcal{L}(u_\beta)$. Every factor of u_β can be extended in at least one way to the right, and thus we can find $Y_1, Y_2 \in \mathcal{A}$ so that X_1wY_1 and X_2wY_2 belong to $\mathcal{L}(u_\beta)$. Since w is a maximal left special factor, we have $Y_1 \neq Y_2$. This, however, means that w is a right special factor.

Every left special factor w is either maximal or it can be extended by a letter $a \in \mathcal{A}$ such that wa is again a left special factor. Since the only infinite left special branch of u_β is u_β itself, every left special factor which is not a prefix of u_β is a prefix of a maximal left special factor. Proposition 3.5 therefore implies the following corollary:

Corollary 3.8. *The infinite word u_β has affine factor complexity if and only if u_β has no maximal left special factor.*

3.1. Sufficient condition for affine factor complexity of u_β

In the previous part we have derived that u_β can have affine factor complexity only if $t_m = 1$. Therefore we shall consider only simple Parry numbers with Rényi expansion

$$d_\beta(1) = t_1 t_2 \cdots t_{m-1} 1$$

and study the substitution

$$\begin{aligned} \varphi(0) &= 0^{t_1}1 \\ \varphi(1) &= 0^{t_2}2 \\ &\vdots \\ \varphi(m-2) &= 0^{t_{m-1}}(m-1) \\ \varphi(m-1) &= 0. \end{aligned} \tag{7}$$

In agreement with Corollary 3.8, the study of conditions under which the factor complexity is an affine function, resumes into the study of existence of maximal left special factors in the language of u_β. Lemma 3.2 under the condition $t_m = 1$ states that the longest factor containing only zero letters is 0^{t_1+1}, and this factor has a unique extension to the left and to the right. Therefore a left special factor of the form 0^r satisfies $r \le t_1$, and hence it is a prefix of the infinite left special branch u_β.

We have thus shown the following simple observations:

Lemma 3.9. *Any maximal left special factor contains at least one nonzero letter.*

From the form of the substitution (7) one can deduce the structure of left special factors.

Lemma 3.10. *If $w \in \mathcal{L}(u_\beta)$ is a left special factor (not necessarily maximal) then*

$$w = \begin{cases} 0^r, & \text{for some } r \in \mathbb{N},\ r \le t_1; \\ \varphi(v)0^s, & \text{for some left special factor } v \text{ and } s \in \mathbb{N}. \end{cases}$$

Lemma 3.11. *Let $w \in \mathcal{L}(u_\beta)$.*

(1) *If w is a left special factor then $\varphi(w)$ is a left special factor with the same number of left extensions;*
(2) *If w is a maximal left special factor then there exists $q \in \mathbb{N}, q \le t_1$ such that $\varphi(w)0^q$ is a maximal left special factor.*

Statement (2) of Lemma 3.11 says that if there exists one maximal left special factor, then there exist infinitely many such factors.

Definition 3.12. A maximal left special factor w is called noninitial if there exists a maximal left special factor v and an integer $q \in \mathbb{N}$ such that $w = \varphi(v)0^q$. A maximal left special factor which is not non-initial is called initial maximal left special factor.

If $\mathcal{L}(u_\beta)$ contains a maximal left special factor, then it contains an initial maximal left special factor as well. In order to describe initial maximal left special factors, we introduce the notion of trident.

Definition 3.13. A factor $w \in \mathcal{L}(u_\beta)$ is called a trident if there exist letters $X, Y, Z \in \mathcal{A}$ such that

(1) wX is a left special factor;
(2) wY and wZ are not left special factors;
(3) the unique left extensions of wY and wZ are distinct.

The letter X is called the rooted tooth, the letters Y and Z are called nonrooted teeth of the trident w.

Clearly, the teeth X, Y, Z are different.

Remark 3.14. If 0^r is a trident, then the rooted tooth $X = 0$ or 1. This fact follows from Lemma 3.2, since 0^rX is a left special factor only if $X \le 1$.

Lemma 3.15. *Let w be a trident containing a nonzero letter with rooted tooth X and nonrooted teeth Y, Z.*

(i) *If $X = 0$ then $t_Y = t_Z$.*
(ii) *If $X \ne 0$ then there exist an integer $s \in \mathbb{N}$ and a trident $\hat{w}$ with rooted tooth $\hat{X} \ne m - 1$ and nonrooted teeth $\hat{Y}$, $\hat{Z}$, such that*
   (a) $w = \varphi(\hat{w})0^s$,
   (b) $A = \hat{A} + 1$ *for every nonzero tooth A of the trident w,*
   (c) $s = t_A$ *for every nonzero tooth A of the trident w,*
   (d) *if $A = 0$ is a nonrooted tooth of w, then $\hat{A} = m - 1$ or $t_{\hat{A}+1} > t_X = t_{\hat{X}+1}$.*

Proof. From the definition of a trident, it follows that w is a left special factor. According to Lemma 3.10, there exist a left special factor $\hat{w}$ and $s \in \mathbb{N}$ such that $w = \varphi(\hat{w})0^s$.

(i) Let $X = 0$. Since $wY = \varphi(\hat{w})0^sY$ and $wZ = \varphi(\hat{w})0^sZ$ are factors of u_β, and $Y, Z \ne X = 0$, it follows that $s = t_Y = t_Z$.

(ii) Let $X \ne 0$. Since $wX = \varphi(\hat{w})0^sX = \varphi\big(\hat{w}(X-1)\big)$ is a left special factor, also $\hat{w}(X-1)$ is a left special factor and $s = t_X$. As teeth Y, Z are distinct, at least one of them is nonzero, say $Y \ne 0$. Since

$wY = \varphi(\hat{w})0^s Y = \varphi\big(\hat{w}(Y-1)\big)$ is not a left special factor, due to Lemma 3.11, $\hat{w}(Y-1)$ is also not a left special factor and $t_X = t_Y$.

If moreover $Z \neq 0$, we have analogously $t_X = t_Z$ and $\hat{w}(Z-1)$ is not a left special factor. As factors $\varphi\big(\hat{w}(Y-1)\big)$ and $\varphi\big(\hat{w}(Z-1)\big)$ have distinct left extensions, also factors $\hat{w}(Y-1)$ and $\hat{w}(Z-1)$ have distinct left extensions, and therefore $\hat{w}$ is a trident with teeth $X-1, Y-1, Z-1$.

Suppose now that $Z = 0$. Since $\varphi\big(\hat{w}(X-1)\big)$, $\varphi\big(\hat{w}(Y-1)\big)$ and $\varphi(\hat{w})0^{t_X}0$ are factors of equal length, there must exist a letter $\hat{Z} \neq X-1, Y-1$ such that $\hat{w}\hat{Z} \in \mathcal{L}(u_\beta)$ and $\hat{w}\hat{Z}$ has a unique left extension. If $\hat{Z} \neq m-1$, then $w0 = \varphi(\hat{w})0^{t_X}0$ is a proper prefix of $\varphi(\hat{w}\hat{Z}) = \varphi(\hat{w})(\hat{Z}+1)$, and hence $t_{\hat{Z}+1} \geq t_X + 1$. □

Corollary 3.16. *If w is a trident with rooted tooth $X = 1$ and $Y \neq 0$ is a nonrooted tooth, then $t_Y = t_1$.*

Tridents play an important role for the existence of maximal left special factors.

Proposition 3.17. *Let v be an initial maximal left special factor. Then there exists a trident w with rooted tooth X and an integer $s \in \mathbb{N}$ such that*

$$\begin{aligned} &v = \varphi(w)0^s, \quad X \neq 0, m-1, \quad \textit{and} \\ &t_{X+1} < s = \min\{t_{A+1} \mid A \textit{ is a nonrooted tooth of } w,\ A \neq m-1\}. \end{aligned} \tag{8}$$

Proof. Let v be an initial maximal left special factor, and let $y', z' \in \mathcal{A}$ be its distinct left extensions. Denote by Y' the right extension of $y'v$ and by Z' the right extension of $z'v$. Since v is a maximal left special factor, necessarily $Y' \neq Z'$. According to Lemmas 3.9 and 3.11, we have $v = \varphi(w)0^s$ for some left special factor w and some $s \in \mathbb{N}$, $s \leq t_1$. Since $\varphi(w)0^s Y'$ and $\varphi(w)0^s Z'$ belong to the language $\mathcal{L}(u_\beta)$, there exist distinct letters Y, Z such that wY, $wZ \in \mathcal{L}(u_\beta)$, and wY, wZ have unique left extensions.

Since $v = \varphi(w)0^s$ is an initial maximal left special factor, the left special factor w is not maximal, and thus there exists a letter X such that wX is a left special factor. This shows that the factor w is a trident with rooted tooth X and nonrooted teeth Y, Z.

Let us now show that $X \neq 0, m-1$. Suppose that $X = 0$. Then using Lemma 3.11, the factor $\varphi(wX) = \varphi(w)0^{t_1}1$ is left special. Since $v = \varphi(w)0^s$ and $s \leq t_1$, this implies that v is a prefix of a left special factor $\varphi(w)0^{t_1}1$, which contradicts the maximality of v.

Suppose now that $X = m-1$. Then using (i) of Remark 3.3, the factor $w(m-1)0$ is left special, and thus $\varphi(w)0^{t_1+1}$ is also a left special factor. Again, we obtain a contradiction with the maximality of v, since v is then a proper prefix of another left special factor.

The same reason leads us to the fact that $s > t_{X+1}$, because otherwise $v = \varphi(w)0^s$ is a proper prefix of the left special factor $\varphi(w)\varphi(X) = \varphi(X)0^{t_{X+1}}(X+1)$, where we use that $X \neq m-1$.

It remains to determine the value of s. Since at least one of the letters Y', Z' is nonzero, say $Y' \neq 0$, we have $vY' = \varphi(w)0^s Y' = \varphi(wY)$, and thus $Y' = Y+1$, $s = t_{Y+1} \leq t_1$ and $Y \neq m-1$. If, moreover, $Z' \neq 0$, we have by the same arguments that $s = t_{Z+1} = t_{Y+1}$, and $Z \neq m-1$. If $Z' = 0$, then either $Z = m-1$ or $Z \neq m-1$ and $t_{Z+1} > s = t_{Y+1}$. □

We are now in position to prove that condition (2) of Theorem 1.1 is sufficient for u_β to have affine factor complexity.

Proposition 3.18. *Let u_β be the infinite word associated with the Parry number β with $d_\beta(1) = t_1 \cdots t_{m-1}1$. If u_β does not have affine factor complexity, then*

(1) *there exists a nonempty word which is both a proper prefix and a proper suffix of the word $t_1 \cdots t_{m-1}$;*
(2) *for every $k \in \mathbb{N}$, $k \geq 2$, and every word w it holds that $w^k \neq t_1 \cdots t_{m-1}$.*

Proof. If the factor complexity of u_β is not an affine function, then there exists an initial maximal left special factor v. According to Proposition 3.17, there exist an integer s and a trident w with rooted tooth X and nonrooted teeth Y, Z satisfying conditions (8). Denote $l = X$. Relations (8) imply that $1 \leq l < m-1$. We want to construct l tridents $w^{(1)}, w^{(2)}, \ldots, w^{(l)}$ with triples of teeth $(1, Y_1, Z_1), (2, Y_2, Z_2), \ldots, (l, Y_l, Z_l)$, and integers $s_1, s_2, \ldots, s_l$ such that $w^{(i)}, w^{(i+1)}$ and s_{i+1} have properties of tridents $\hat{w}$, w and the integer s from Lemma 3.15 for all $i = 1, \ldots, l-1$, and $Y_l = Y$ and $Z_l = Z$. If $l = 1$, this role is played obviously by the trident w, its triple of teeth $(1, Y, Z)$ and the integer

s. If $l \geq 2$, then according to Remark 3.14, the trident w contains a non-zero letter and satisfies the assumptions of Lemma 3.15, which implies the existence of the sequence of tridents $w^{(1)}, w^{(2)}, \dots, w^{(l)}$ with triples of teeth $(1, Y_1, Z_1), (2, Y_2, Z_2), \dots, (l, Y_l, Z_l)$, and integers $s_1, s_2, \dots, s_l$ with required properties.

According to Corollary 3.16, we have

$$s_1 = t_1. \tag{9}$$

Since the rooted teeth $X_1 = 1, X_2 = 2, \dots, X_l = l$ are non-zero, (c) of Lemma 3.15 implies

$$s_2 = t_2, \quad s_3 = t_3, \dots, \quad s_l = t_l. \tag{10}$$

Using Proposition 3.17 we obtain

$$t_{l+1} < s := \min\{t_{A+1} \mid A \text{ is a nonrooted tooth of } w^{(l)},\ A \neq m-1\}. \tag{11}$$

Lemma 3.15 implies that the sequence $Y_1, Y_2, \dots, Y_l$ is formed by consecutive integers separated by blocks of 0's. More precisely, for any $i = 1, \dots, l-1$, we have

$$Y_{i+1} = \begin{cases} Y_i + 1 & \text{if } Y_i < m-1 \text{ and } t_{Y_i+1} = s_{i+1}, \\ 0 & \text{if } Y_i = m-1 \text{ or } t_{Y_i+1} > s_{i+1}. \end{cases} \tag{12}$$

The same rule is valid for the sequence $Z_1, \dots, Z_l$.

Since nonrooted teeth Y_1, Z_1 are distinct, we can without loss of generality assume that $Y_1 \geq 2$.

In order to show Statement (1) of the proposition, denote by $k \leq l$ the maximal index such that the sequence $Y_1, \dots, Y_k$ is formed by consecutive non-zero integers, i.e.

$$Y_1\ Y_2\ \cdots\ Y_k \ = \ j\ (j+1)\ \cdots\ (j+k-1) \quad \text{for some } \ j \in \mathbb{N},\ j \geq 2.$$

This, however, means, using (12), (10) and Corollary 3.16 that

$$t_j = t_1, \quad t_{j+1} = t_2, \dots, \quad t_{j+k-1} = t_k. \tag{13}$$

We now show that the nonrooted tooth $Y_k = (j+k-1)$ is equal to $(m-1)$, which together with (13) results in Statement (1) of the proposition. For the contradiction, assume that $Y_k = (j+k-1) < m-1$. Let us distinguish two cases according to whether $k < l$ or $k = l$. If $k < l$, then from the definition of k it follows that $Y_{k+1} = 0$, which, due to (12), can happen only if

$$t_{Y_k+1} = t_{j+k} \ > \ s_{k+1} = t_{k+1}. \tag{14}$$

If $k = l$, then (11) implies

$$t_{k+1} < s \leq t_{Y_k+1} = t_{j+k}. \tag{15}$$

In any case, (13) together with (14), or (15) gives

$$t_j t_{j+1} \cdots t_{j+k} \succ t_1 t_2 \cdots t_{k+1},$$

which contradicts the Parry condition (4).

Besides the validity of Statement (1) of the proposition, we have thus proved that the sequence $Y_1, \dots, Y_l$ contains at least one letter $m-1$.

In order to show Statement (2) of proposition, denote by p the shortest nonempty word which is both a proper prefix and a proper suffix of the word $t_1 \cdots t_{m-1}$. It is obvious that p is not a power of a shorter word.

We show Statement (2) by contradiction. Assume that there exists a word w such that $w^k = t_1 \cdots t_{m-1}$ for some $k \geq 2$, $k \in \mathbb{N}$. First we claim that such an assumption implies that $t_1 \cdots t_{m-1} = p^n$ for some $n \in \mathbb{N}$, $n \geq 2$. Since w is a prefix and a suffix of $t_1 \cdots t_{m-1}$, we must have $|w| \geq p$. If $|w| = |p|$, the claim is valid. If $|w| > |p|$, then p is a proper prefix and a proper suffix of w. Moreover, the prefix p and the suffix p do not overlap in the word w, since otherwise the overlap would be a proper prefix and a proper suffix of $t_1 \cdots t_{m-1}$ shorter than p, which contradicts the minimality of p. The condition $|w| > |p|$ thus implies that $w = pw'p$ for some (possibly empty) word w'. If $w' = \varepsilon$, the claim is valid. In the opposite case, the word $t_1 t_2 \cdots t_{m-1}$ has the prefix $pw'ppw'p$. The Parry condition for $d_\beta(1)$ implies that $w'p \preceq pw'$, and $ppw' \preceq pw'p$ which then implies $pw' \preceq w'p$, and therefore $pw' = w'p$. It is known that if two words commute, then they are powers of the same word. For proof of this assertion, see for example Chapter 1 of [7]. Since p itself is not a power, we must have $w' = p^j$ for some $j \in \mathbb{N}$, as we wanted to show.

Let now $t_1t_2\cdots t_{m-1} = p^n$ for some $n \in \mathbb{N}$, $n \geq 2$. Denote $s = |p|$. Obviously $m-1 = ns$. If $s = 1$, then $d_\beta(1) = t_1t_1\cdots t_1 1$, and in that case u_β is an Arnoux–Rauzy word, for which directly from the definition follows that the factor complexity is an affine function. Thus $s \geq 2$.

Let us come back to the sequence of tridents and the triples of their teeth, $(1, Y_1, Z_1), (2, Y_2, Z_2), \ldots, (l, Y_l, Z_l)$. We already know that one of the letters $Y_1, \ldots, Y_l$ is equal to $m-1$. Denote by q the maximal index, such that Y_q or Z_q is equal to $m-1 = ns$. Since the role of Y_q and Z_q is symmetric, without loss of generality we can assume that the last $m-1$ occurred was $Y_q = m-1$. We will show that both the corresponding rooted tooth q and the other nonrooted tooth Z_q are multiples of s.

For a contradiction, suppose that $q = as + b$, where $1 \leq b < s$. According to Lemma 3.15, we have

$$t_q = t_{Y_q} = t_{m-1}, \quad t_{q-1} = t_{m-2}, \ldots, \quad t_{q-b+1} = t_{m-1-b+1}.$$

Since the word p of the length s is the period of $t_1\cdots t_{m-1}$, we have

$$t_q = t_{as+b} = t_b = t_{m-1}, \quad t_{b-1} = t_{m-2}, \ldots, \quad t_{q-b+1} = t_1 = t_{m-b},$$

and therefore $t_1\cdots t_b$ is both a prefix and a suffix of $t_1\cdots t_{m-1}$, shorter than p, which contradicts the choice of p. In the same way, one can show that the nonrooted tooth Z_q is a multiple of s, say $Z_q = cs$ for some $c \in \mathbb{N}$.

Since for the sequence of letters $Z_1, \ldots, Z_l$ one can derive a rule analogous to (12), we obtain from the periodicity of $t_1\cdots t_{m-1}$ and the assumption $Z_i \neq m-1$ for $i \geq q$, given by the definition of the index q, that $t_{Z_i} = t_i = t_{i \bmod s}$, and therefore $Z_{i+1} = Z_i + 1$ for all i, $q \leq i \leq l$. The periodicity of $t_1\cdots t_{m-1}$ also implies $t_{l+1} = t_{Z_l+1} = t_Z + 1$, which contradicts (11). □

3.2. Necessary condition for affine factor complexity of u_β

We now show that if there exists a word p which is both a proper prefix and a proper suffix of $t_1\cdots t_{m-1}$, and $t_1\cdots t_{m-1}$ is not an integer power of p, then the factor complexity of u_β is not an affine function. According to Proposition 3.5 it suffices to find a left special factor which is not a prefix of u_β.

For that, we use the fact that the words of $\mathcal{L}(u_\beta)$ code the patterns of $\mathbb{Z}_\beta^+$. Indeed, the word $w = w_0w_1\cdots w_n \in \{0, 1, \ldots, m-1\}^*$ is a coding of the set $[x, y] \cap \mathbb{Z}_\beta$, if the distances between points of $[x, y] \cap \mathbb{Z}_\beta$ are consecutively $T_\beta^{w_0}(1), \ldots, T_\beta^{w_n}(1)$. With this, we can reformulate the main problem of this section in the language of $\mathbb{Z}_\beta^+$. Construction of a left special factor of u_β which is not a prefix of u_β is equivalent to the construction of β-integers z, x_1, x_2 such that

(i) the codings of the sets $[x_1, x_1 + z] \cap \mathbb{Z}_\beta$, $[x_2, x_2 + z] \cap \mathbb{Z}_\beta$ and $[0, z] \cap \mathbb{Z}_\beta$ are equal to the same word w; more formally,

$$\big([x_1, x_1 + z] \cap \mathbb{Z}_\beta\big) - x_1 = \big([x_2, x_2 + z] \cap \mathbb{Z}_\beta\big) - x_2 = [0, z] \cap \mathbb{Z}_\beta\,;$$

(ii) $x_1 - \operatorname{pred}(x_1) \neq x_2 - \operatorname{pred}(x_2)$;

(iii) $1 = \operatorname{succ}(x_1 + z) - (x_1 + z) = \operatorname{succ}(x_2 + z) - (x_2 + z)$;

(iv) $1 \neq \operatorname{succ}(z) - z$.

Note that as the distance $1 = T_\beta^0(1)$ is coded by the letter 0, Conditions (i)–(iv) ensure that the word $w0 \in \mathcal{L}(u_\beta)$ is a left special factor of u_β which is not a prefix of u_β.

The construction of suitable β-integers z,x_1,x_2 with the above properties, is the contents of this section, we shall, however, need some preparation.

Let $p = p_1\cdots p_s$, be a proper prefix and a proper suffix of the word $t_1\cdots t_{m-1}$ of the minimal nonzero length. From the Parry condition and the fact that $t_1\cdots t_{m-1} \neq p^k$ for $k \geq 2$ one can easily deduce that there exist words p', q, and a positive integer r such that

$$d_\beta(1) = p^r p' q p 1, \tag{16}$$

where p' is a prefix of p and $|p| > |p'| := j$, and q is a nonempty word starting with the letter $q_1 < p_{j+1}$. Let us mention that the words p, p', q are words over the alphabet $\{t_1, t_2, \ldots, t_{m-1}\}$. Since $t_1 \cdots t_{m-1}$ is not an integer power of p, we must have

$$pp'q \neq p'qp.$$

As $|pp'q| = |p'qp|$, we can find a word $c \in \{t_1, t_2, \ldots, t_{m-1}\}^*$ and digits $h_1, h_2 \in \{t_1, \ldots, t_{m-1}\}$ such that $h_1 \neq h_2$, h_1c is a suffix of $pp'q$, and h_2c is a suffix of $p'qp$. Note that since $q_1 < p_{j+1}$, c as a common suffix of $pp'q$ and $p'qp$ must satisfy

$$|c| \leq |p| + |q| - 1. \tag{17}$$

Denote $h := \min(h_1, h_2)$ and $A = |p^r p' q_1| = rs + j + 1$. Then we can define β-integers x_1, x_2, z using their β-expansion as

$$\begin{aligned} z &:= \quad p^r p' q_1 \bullet + hc0^A \bullet \\ x_1 &:= \quad p^r p' q 0^A \bullet - hc0^A \bullet \\ x_2 &:= p^r p' q p 0^A \bullet - hc0^A \bullet . \end{aligned} \tag{18}$$

Directly from the definition of A, h and c, it follows that the word $hcp^r p' q_1$ satisfies the Parry condition, and thus $z = hcp^r p' q_1 \bullet$ is a β-integer. In the same time, from the definition of h and c it is obvious that subtraction in the prescription for x_1 and x_2 can be performed digitwise and hence also $x_1, x_2 \in \mathbb{Z}_\beta^+$.

We now prove that the above defined z, x_1, x_2 satisfy Conditions (i)–(iv).

ad(i) In order to prove that the word w coding the distances between consecutive β-integers in the segment $[0, z]$, codes also the segments $x_1, x_1 + z$ and $x_2, x_2 + z$, we use the following lemma, which is a consequence of the fact that the infinite word u_β codes the distances between consecutive β-integers.

Lemma 3.19. *Let $x, z \in \mathbb{Z}_\beta^+$ such that*

$$\textit{for every } z' \in \mathbb{Z}_\beta^+,\ z' \leq z \textit{ we have } x + z' \in \mathbb{Z}_\beta^+. \tag{19}$$

Then codings of $[0, z] \cap \mathbb{Z}_\beta$ and $[x, x + z] \cap \mathbb{Z}_\beta$ coincide.

For $x = x_1$ we divide the verification of condition (19) into three cases.

- If $z' \in \mathbb{Z}_\beta$, $0 \leq z' < hc0^A \bullet + p10^{A-s-1}\bullet$, then the summation $x_1 + z'$ can be performed digitwise. The result is again a string satisfying the Parry condition, and therefore $x_1 + z' \in \mathbb{Z}_\beta$.
- If $z' = hc0^A \bullet + p10^{A-s-1}\bullet$, then we have $x_1 + z' = p^r p' q 0^A \bullet + p10^{A-s-1}\bullet = 10^m 0^{A-s-1}\bullet$.
- If $z' \in \mathbb{Z}_\beta$, $hc0^A \bullet + p10^{A-s-1}\bullet < z' \leq z$, then $x_1 + z' = 10^m 0^{A-s-1} \bullet + z''$, where $z'' \in \mathbb{Z}_\beta$, $0 < z'' \leq p^r p' q_1 \bullet - p10^{A-s-1}\bullet$, and again by digitwise summation we obtain an admissible β-expansion of $x_1 + z'$.

In order to prove condition (19) for $x = x_2$, we again separate $z' \in \mathbb{Z}_\beta$, $z' \leq z$ into three cases. Now the separating point is $z' = hc0^A \bullet + 10^{A-1}\bullet$. For such z' we have $x_2 + z' = 10^{m+A-1}\bullet$. The remaining cases $x_2 + z'$ can be solved by digitwise summation, similarly as for $x = x_1$.

ad(ii) For the proof of property (ii) we use another statement, which allows one to determine the distance of an element of $\mathbb{Z}_\beta^+$ to its predecessor.

Lemma 3.20. *Let the β-expansion of a β-integer y be $y_n y_{n-1} \cdots y_k 0^k \bullet$, where $y_k \neq 0$, $k \in \mathbb{N}$. Then $y - \operatorname{pred}(y) = T_\beta^{k'}(y)$, where $k' \in \{0, 1, \ldots, m-1\}$ is such that $k' = k \mod m$.*

Proof. If $k = 0$, the statement is obvious. Assume that $k \geq 1$. Denote $d_\beta^*(1) = \left(t_1 t_2 \cdots t_{m-1} 0\right)^\omega$. This is the lexicographically greatest word which is lexicographically strictly smaller than $d_\beta(1) = t_1 \cdots t_{m-1} 1$. Therefore the predecessor of y has the β-expansion of the form $y_n \cdots y_{k+1}(y_k - 1)v\bullet$, where v is a prefix of $d_\beta^*(1)$ of length $|v| = k$. We thus have

$$y - \operatorname{pred}(y) = 10^k \bullet - v\bullet = 0 \bullet t_{k'+1} t_{k'+2} \cdots t_m = T_\beta^{k'}(1),$$

where $k' \in \{0, 1, \ldots, m-1\}$ is such that $k' \equiv k \mod m$. The latter follows from the m-periodicity of $d^*_\beta(1)$. □

The definition of h implies that the number of zeros at the end of β-expansions of x_1, x_2 differ modulo m. Therefore property (ii) is valid.

ad(iii) For verifying the property (iii) we have to show that both $x_1 + z + 1$ and $x_2 + z + 1$ belong to $\mathbb{Z}_\beta$. Let l be the length of the Rényi expansion of 1. It is easy to compute that $x_1 + z = 10^l(p_1 - 1)p_2 \cdots p_s p^{r-2} p' q_1 \bullet$ if $r \geq 2$ and $d_\beta(x_1 + z) = 10^l(p'_1 - 1)p'_2 \cdots p'_j q_1 \bullet$ otherwise. In both cases, since $q_1 < p_{j+1}$, the digit 1 can be added at the last position of $x_1 + z$ without altering the validity of the Parry condition. The same argument holds for $x_2 + z = 10^l(p_1 - 1)p_2 \cdots p_s p^{r-1} p' q_1$.

ad(iv) In order to prove that $\mathrm{succ}(z) \neq z + 1$ we use the statement which is a simple consequence of the proof of Lemma 3.20.

Lemma 3.21. *Let the β-expansion of a β-integer y be $y_n y_{n-1} \cdots y_0 \bullet$. Denote by k the maximal index such that $y_{k-1} y_{k-2} \cdots y_0$ is a prefix of $d^*_\beta(1) = \left(t_1 t_2 \cdots t_{m-1} 0\right)^\omega$. Then $\mathrm{succ}(y) = y + T^{k'}_\beta(1)$, where $k' \in \{0, 1, \ldots, m-1\}$ is such that $k' \equiv k \mod m$.*

For $y = z = hcp^r p' q_1 \bullet$, the index k satisfies

$$0 < |p^r p' q_1| \leq k < |hcp^r p' q_1| \leq m,$$

where we have used inequality (17). Therefore $T^k_\beta(1) \neq T^0_\beta(1)$.

Example 3.22. Let us illustrate the use of β-integers for the construction of a left special factor not being a prefix of u_β. We consider the most simple example of a simple Parry number β not satisfying Condition (2) of Theorem 1.1, namely such that $d_\beta(1) = 2121$. Note that the word $t_1 \cdots t_{m-1} = t_1 t_2 t_3 = 212$ has a nonempty proper prefix and proper suffix 2, in the same time it is not an integer power of any word. Since $m = 4$, the infinite word u_β is over the alphabet $\{0, 1, 2, 3\}$. The β-integers are strings with digits $0, 1, 2 = \lfloor \beta \rfloor = t_1$. The distances between consecutive β-integers take the following four values:

$$\begin{aligned}
T^0_\beta(1) &= 1\bullet, \\
T^1_\beta(1) &= 0 \bullet 121 = 10 \bullet -2\bullet, \\
T^2_\beta(1) &= 0 \bullet 21 \;\;= 100 \bullet -21\bullet, \\
T^3_\beta(1) &= 0 \bullet 1 \;\;\;\;= 1000 \bullet -212 \bullet .
\end{aligned} \tag{20}$$

With these distances we associate letters 0, 1, 2, 3, respectively.

Let us find β-integers z, x_1, x_2 satisfying Conditions (i)–(iv) and explain how they correspond to a left special factor of u_β which is not a prefix of u_β. Using (18), we obtain $z = 121\bullet$, $x_1 = 2001\bullet$, $x_2 = 21100\bullet$.

Each of the intervals $[0, z]$, $[x_1, x_1 + z]$, $[x_2, x_2 + z]$ contains sixteen β-integers. These β-integers are listed in Table 1 in the form of their β-expansion and drawn on the real line in Fig. 2. Using (20) one can verify that the segments of β-integers in all the three intervals are coded by the same word $w = 001001020010010$. Note that the distance between $x_1 + z$ and its successor is the same as the distance between $x_2 + z$ and its successor, namely $T^0_\beta(1) = 1$. Therefore the word $w0$ belongs to the language of u_β. Since the distance between z and its successor is equal to $T^2_\beta(1)$, the word $w0$ is not a prefix of $u_\beta = w2\cdots$. In order to verify that $w0$ is a left special factor of u_β, it suffices to see that the distance between x_1 and its predecessor (namely $T^3_\beta(1)$) is different from the distance of x_2 and its predecessor (namely $T^2_\beta(1)$). In particular, both $3w0$ and $2w0$ belong to the language of u_β.

4. Conclusions

Among the words u_β which have affine factor complexity are words for which the Rényi expansion of unity in base β is of the form $d_\beta(1) = t_1 t_2 \cdots t_{m-1} 1 = p^k 1$, for some $k \geq 2$. If p is a word of length 1, such words are Arnoux–Rauzy, and thus have for each n exactly one left special and one right special factor of length n. If p is of length $|p| \geq 2$, then u_β has for every $n \in \mathbb{N}$ one left special and $|p|$ right special factors.

Table 1
The list of β-integers contained in intervals $[0, z]$, $[x_1, x_1 + z]$ and $[x_2, x_2 + z]$ where $z = 121\bullet$, $x_1 = 2001\bullet$, $x_2 = 21100\bullet$

	$\mathrm{pred}(x_1) = 1212$	$\mathrm{pred}(x_2) = 21021$
0	$x_1 = 2000$	$x_2 = 21100$
1	2001	21101
2	2002	21102
10	2010	21110
11	2011	21111
12	2012	21112
20	2020	21120
21	2021	21121
100	2100	21200
101	2101	21201
102	2102	21202
110	2110	100000
111	2111	100001
112	2112	100002
120	2120	100010
$z = 121$	$x_1 + z = 10000$	$x_2 + z = 100011$
$\mathrm{succ}(z) = 200$	$\mathrm{succ}(x_1 + z) = 10001$	$\mathrm{succ}(x_2 + z) = 100012$

These are strings of digits $\{0, 1, 2\}$ ordered by the radix order, which moreover satisfy the Parry condition.

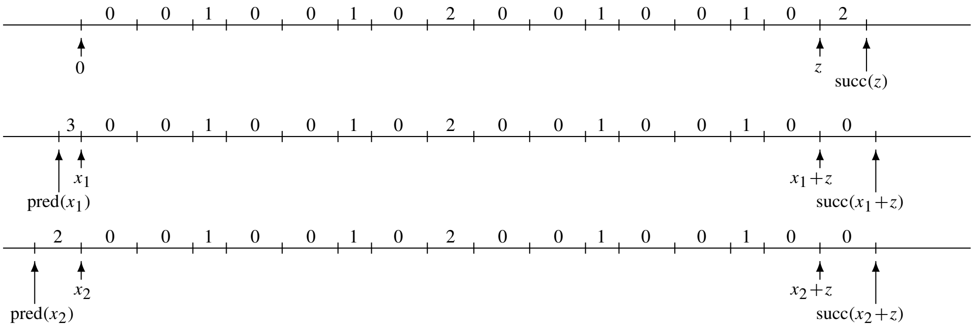

Fig. 2. The β-integers from Table 1 drawn on the real line.

As a continuation of this paper, it would be interesting to study the factor complexity of a fixed point of a substitution defined by a nonsimple Parry number. It would also be interesting to compute explicitly the factor complexity in the nonaffine case. In particular, is it possible that the factor complexity is ultimately affine, that is, $\mathcal{C}(n) = an + b$ for $n \geq n_0$? Due to Lemma 3.11, there cannot exist finitely many maximal left special factors in the nonaffine case, hence $a > m - 1$ in such a case.

Acknowledgement

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

References

[1] J. Cassaigne, Complexité et facteurs spéciaux, Bull. Belg. Math. Soc. Simon Stevin 4 (1997) 67–88.
[2] F. Durand, Linearly recurrent subshifts have a finite number of non-periodic subshift factors, Ergodic Theory Dynam. Syst. 20 (4) (2000) 1061–1078.
[3] S. Fabre, Substitutions et β-systèmes de numération, Theoret. Comput. Sci. 137 (1995) 219–236.
[4] S. Ferenczi, Complexity of sequences and dynamical systems, in: Combinatorics and Number Theory (Tiruchirappalli, 1996), Discrete Math. 206 (1–3) (1999) 145–154.

[5] Ch. Frougny, Z. Masáková, E. Pelantová, Complexity of infinite words associated with beta-expansions, RAIRO Theoret. Inform. Appl. 38 (2004) 163–185; Corrigendum, RAIRO Theoret. Inform. Appl. 38 (2004) 269–271.
[6] M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, 2002.
[7] M. Lothaire, Combinatorics on Words, Cambridge University Press, 1997.
[8] M. Morse, G.A. Hedlund, Symbolic dynamics II. Sturmian trajectories, Amer. J. Math. 62 (1940) 1–42.
[9] P. Michel, Stricte ergodicit d'ensembles minimaux de substitution. Thorie ergodique (Actes Journes Ergodiques, Rennes, 1973/1974), in: Lecture Notes in Mathematics, vol. 532, Springer, Berlin, 1976, pp. 189–201.
[10] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, in: Automata, Languages and Programming (Antwerp, 1984), in: Lecture Notes in Comput. Sci., vol. 172, Springer, 1984, pp. 380–389.
[11] W. Parry, On the β-expansions of real numbers, Acta Math. Acad. Sci. Hungar. 11 (1960) 401–416.
[12] A. Rényi, Representations for real numbers and their ergodic properties, Acta Math. Acad. Sci. Hungar. 8 (1957) 477–493.
[13] W.P. Thurston, Groups, tilings, and finite state automata, Geometry Supercomputer Project Research Report GCG1, University of Minnesota, 1989.