European Journal of Combinatorics 118 (2024) 103942

Contents lists available at ScienceDirect

European Journal of Combinatorics

journal homepage: https://www.sciencedirect.com/journal/european-journal-of-combinatorics

Initial nonrepetitive complexity of regular episturmian words and their Diophantine exponents

Jarkko Peltomäki

The Turku Collegium for Science, Medicine and Technology TCSMT, University of Turku, Turku, Finland
Turku Centre for Computer Science TUCS, Turku, Finland
Department of Mathematics and Statistics, University of Turku, Turku, Finland

ARTICLE INFO

Article history:
Received 5 April 2022
Accepted 14 February 2024
Available online 29 February 2024

In memory of my beloved brother Sauli Peltomäki (1996–2021)

ABSTRACT

Regular episturmian words are episturmian words whose directive words have a regular and restricted form making them behave more like Sturmian words than general episturmian words. We present a method to evaluate the initial nonrepetitive complexity of regular episturmian words extending the work of Wojcik on Sturmian words. For this, we develop a theory of generalized Ostrowski numeration systems and show how to associate with each episturmian word a unique sequence of numbers written in this numeration system.

The description of the initial nonrepetitive complexity allows us to obtain novel results on the Diophantine exponents of regular episturmian words. We prove that the Diophantine exponent of a regular episturmian word is finite if and only if its directive word has bounded partial quotients. Moreover, we prove that the Diophantine exponent of a regular episturmian word is strictly greater than 2 if the sequence of partial quotients is eventually at least 3.

Given an infinite word x over an integer alphabet, we may consider a real number ξ_x having x as a fractional part. The Diophantine exponent of x is a lower bound for the irrationality exponent of ξ_x. Our results thus yield nontrivial lower bounds for the irrationality exponents of real numbers whose fractional parts are regular episturmian words. As a consequence, we identify a new uncountable class of transcendental numbers

E-mail address: r@turambar.org.

https://doi.org/10.1016/j.ejc.2024.103942

whose irrationality exponents are strictly greater than 2. This class contains an uncountable subclass of Liouville numbers.

1. Introduction

The fractional part of the expansion of a real number in some base can be interpreted as a right-infinite word. Major open problems in number theory concern the expansions of well-known numbers such as $\sqrt{2}$, π, or e. The inverse problem of inferring properties of a number $\xi_{\mathbf{x}}$ whose fractional part matches a prescribed infinite word $\mathbf{x}$ has attracted much attention especially in the last two decades. One of the most significant results is that of Adamczewski and Bugeaud [3] from 2007 stating that if the factor complexity function $p(\mathbf{x}, n)$ of an aperiodic infinite word $\mathbf{x}$ (a purely combinatorial notion) is sublinear, then $\xi_{\mathbf{x}}$ is transcendental. More recently, Bugeaud and Kim [13] introduced the notion of the exponent of repetition of an infinite word and studied it in relation to Sturmian words proving, among other results, that if $\lim_{n\to\infty}(p(\mathbf{x}, n) - n) < \infty$, then the irrationality exponent of $\xi_{\mathbf{x}}$ is at least $5/3 + 4\sqrt{10}/15$. The exponent of repetition is closely linked to the notion of the Diophantine exponent of an infinite word. The significance of this notion is that the Diophantine exponent of $\mathbf{x}$ is a lower bound to the irrationality exponent of $\xi_{\mathbf{x}}$. In this paper, we consider the class of regular episturmian words and prove results on their Diophantine exponents by characterizing their initial nonrepetitive complexity function. This provides novel results on the irrationality exponents of numbers whose fractional parts match a regular episturmian word.

1.1. Episturmian words

An infinite word is Sturmian if it has exactly $n + 1$ distinct factors (subwords) of length n for all n. Sturmian words have many equivalent definitions and they can be generalized in various ways depending on the definition being used. Generalizing the work of de Luca on iterated palindromic closure, Droubay, Justin, and Pirillo introduced in [16] episturmian words that further generalize the so-called Arnoux-Rauzy words [8]. This purely combinatorial generalization is defined as follows. For a finite word w, let $w^{(+)}$ be the shortest palindrome having w as a prefix. Let $\Delta = y_1 y_2 \cdots$ be an infinite word and define a sequence (u_k) of finite words as follows:

$$u_1 = \varepsilon,$$
$$u_{k+1} = (u_k y_k)^{(+)}.$$

The limit $\mathbf{c}_\Delta$ of the words (u_k) is the standard episturmian word with directive word Δ. An episturmian word with directive word Δ is an infinite word sharing the set of factors with $\mathbf{c}_\Delta$. Sturmian words correspond to directive words that are binary and not ultimately constant. The most famous standard episturmian word that is not Sturmian is the Tribonacci word

$$0102010010201010201001020102010010201010201001020 1\cdots$$

having directive word $012012012\cdots$. The main results of this paper are stated for regular episturmian words whose directive words are of special form. Write Δ in the form $x_1^{a_1} x_2^{a_2} \cdots$ with $x_k \neq x_{k+1}$ and $a_k > 0$ for all k. If the sequence $x_1 x_2 \cdots$ equals the periodic sequence with period $012 \cdots (d-1)$ for some d, then we say that the episturmian words with directive word Δ are *regular* with period d. This class of regular episturmian words contains Sturmian words (case $d = 2$) and d-bonacci words, generalizations of the Fibonacci and Tribonacci words.

Episturmian words enjoy many of the good properties of Sturmian words as described in the foundational papers [16,23,24] and the survey [19], but some properties such as interpretation as codings of irrational rotations are lost. Standard references for Sturmian words are [26, Ch. 2], [32, Ch. 6]. We refer the reader to [31, Ch. 4] for an introduction to Sturmian words as codings of irrational rotations.

1.2. Initial nonrepetitive complexity

We define the *initial nonrepetitive complexity function* $\operatorname{inrc}(\mathbf{x}, n)$ of an infinite word $\mathbf{x}$ by

$$\operatorname{inrc}(\mathbf{x}, n) = \max\{m : \mathbf{x}[i, i+n-1] \neq \mathbf{x}[j, j+n-1] \text{ for all } i, j \text{ with } 1 \leq i < j \leq m\}.$$

The number $\operatorname{inrc}(\mathbf{x}, n)$ is the maximum number of factors of length n seen when $\mathbf{x}$ is read from left to right prior to the first repeated factor of length n. In other words, the prefix of $\mathbf{x}$ of length $\operatorname{inrc}(\mathbf{x}, n) + n$ is the shortest prefix of $\mathbf{x}$ containing two occurrences of some factor of length n.

The notion of initial nonrepetitive complexity was introduced independently by Moothatu [28] and Bugeaud and Kim [13]. Nicholson and Rampersad [29] examine the general properties of this function and determine it explicitly for certain words such as the Thue–Morse word, the Fibonacci word, and the Tribonacci word. Their results were generalized for all standard Arnoux-Rauzy words in [27] by Medková et al.

In this paper, we set up a framework that allows us in principle to determine the initial nonrepetitive complexity of any episturmian word. In order to do this, we generalize the S-adic expansion of Sturmian words derived in [10] to all episturmian words and prove in Theorem 3.13 that an episturmian word $\mathbf{t}$ can be expressed in the form

$$\mathbf{t} = \lim_{k\to\infty} T^{\rho_k}(\mathbf{c})$$

where $\mathbf{c}$ is the corresponding standard episturmian word and (ρ_k) is an integer sequence expressed in a generalized Ostrowski numeration system. In other words, we show that each episturmian word is a limit of appropriate shifts of the corresponding standard word $\mathbf{c}$. This means that studying a property of episturmian words can be reduced to studying the property on shifts of standard episturmian words. Here we apply this principle and determine the initial nonrepetitive complexity of *regular* episturmian words generalizing a result of Wojcik [35] on Sturmian words. This results in Theorem 5.10 which is too complicated to be stated here. We leave the characterization of this function for all episturmian words open.

1.3. Diophantine exponents and main results

The Diophantine exponent is a combinatorial exponent of infinite words. It is introduced in [2], but is used implicitly in the earlier works by the same authors.

Definition 1.1. Let $\mathbf{x}$ be an infinite word. We let its *Diophantine exponent*, denoted by $\operatorname{dio}(\mathbf{x})$, to be the supremum of all real numbers ρ for which there exist arbitrarily long prefixes of $\mathbf{x}$ of the form UV^e, where U and V are finite words and e is a real number, such that

$$\frac{|UV^e|}{|UV|} \geq \rho.$$

The concept of Diophantine exponent has inherent interest to a word-combinatorist, and the concept has connections to other combinatorial exponents widely studied in combinatorics on words. What makes it special is the ingenious and simple result that $\operatorname{dio}(\mathbf{x})$ is a lower bound for the irrationality exponent $\mu(\xi_{\mathbf{x},b})$ of the real number $\xi_{\mathbf{x},b}$ having $\mathbf{x}$ as the fractional part of its base-b expansion.

The crucial result here is that of Bugeaud and Kim [13] stating that

$$\operatorname{dio}(\mathbf{x}) = 1 + \limsup_{n\to\infty} \frac{n}{\operatorname{inrc}(\mathbf{x}, n)}.$$

This, together with the characterization of the function $\operatorname{inrc}(\mathbf{x}, n)$ for regular episturmian words (Theorem 5.10), provides means to find lower bounds for $\mu(\xi_{\mathbf{x},b})$ when $\mathbf{x}$ is a regular episturmian word.

Our two main results are as follows. The results were previously proved for Sturmian words by Adamczewski (based on the results of [10]) and Komatsu [25] respectively.

Theorem 1.2. *Let* $\mathbf{t}$ *be a regular episturmian word of period* d *with directive word* $\Delta = x_1^{a_1}x_2^{a_2}\cdots$. *If* $d = 2$ *or* $\limsup_k a_k \geq 3$, *then* $\mu(\xi_{\mathbf{t},b}) > 2$.

Theorem 1.3. *Let* $\mathbf{t}$ *be a regular episturmian word with directive word* $\Delta = x_1^{a_1}x_2^{a_2}\cdots$. *Then* $\xi_{\mathbf{t},b}$ *is a Liouville number if and only if the sequence* (a_k) *is unbounded.*

We thus identify a new uncountable class of numbers with irrationality exponent strictly greater than 2 and a new uncountable class of Liouville numbers.

In addition, we show in Section 7 that it is possible that $\mu(\xi_{\mathbf{t},b}) > \operatorname{dio}(\mathbf{t})$ for an episturmian word over a 3-letter alphabet. For Sturmian words, we have the equality $\mu(\xi_{\mathbf{t},b}) = \operatorname{dio}(\mathbf{t})$ by a result of Bugeaud and Kim [13]. Additional results on the Diophantine exponents of d-bonacci words are provided in Section 6.

2. Preliminaries from combinatorics on words

We use standard notions and notations from combinatorics on words. For a general reference, see, e.g., [26].

A *word* is a finite sequence of symbols from some finite set of *letters* called an *alphabet*. If $w = w_1 \cdots w_n$ with $w_i \in \mathcal{A}$, then we say that w is a word of length n over $\mathcal{A}$, and we set $|w| = n$. The *empty word*, the unique word of length 0, is denoted by ε. The set of words over $\mathcal{A}$ is denoted by $\mathcal{A}^*$. If u and v are words such that $u = u_1 \cdots u_n$ and $v = v_1 \cdots v_m$ with $u_i, v_i \in \mathcal{A}$, then their *concatenation* uv is the word $u_1 \cdots u_n v_1 \cdots v_m$. If $w = uzv$, then u is a *prefix* of w, v is a *suffix* of w, and z is a *factor* of w. If $u \neq w$, then u is a *proper* prefix of w; proper suffix is defined analogously. If $w = uv$, then by $u^{-1}w$ and wv^{-1} we respectively refer to the words v and u. An *occurrence* of u in $w_1 \cdots w_n$ is an index i such that u is a prefix of $w_i w_{i+1} \cdots w_n$. By $w[i, j]$ we mean the factor $w_i \cdots w_j$ whenever the indices i and j make sense.

Let w be a word over $\mathcal{A}$. By w^n we refer to the concatenation $w \cdots w$ where w is repeated n times. This is an integer power, and by a fractional power w^e, $e \geq 1$, we mean the word $(uv)^n u$ with $uv = w$ and $e = n + |u|/|w|$. If $w = u^n$ only if $n = 1$, then we say that w is *primitive*. The word w is primitive if and only if w occurs exactly twice in w^2. If $w = uv$, then the word vu is a *conjugate* of w. If $w = w_1 \cdots w_n$, $w_i \in \mathcal{A}$, then the *reversal* of w is the word $w_n \cdots w_1$. If a word equals its reversal, then we say that it is a *palindrome* (we count the empty word as a palindrome). If $w_i = w_{i+p}$ for all i such that $0 \leq i < |w| - p$, then w has *period* p. A mapping $\tau\colon \mathcal{A}^* \to \mathcal{A}^*$ is a *morphism* if $\tau(uv) = \tau(u)\tau(v)$ for all $u, v \in \mathcal{A}^*$.

An infinite sequence of letters $\mathbf{x}$ over $\mathcal{A}$ is called an *infinite word*. The set of infinite words over $\mathcal{A}$ is denoted by $\mathcal{A}^\omega$. This set is naturally equipped with the product topology. If $\mathbf{x} = x_1x_2x_3 \cdots$ with $x_i \in \mathcal{A}$, then its *shift* $T(\mathbf{x})$ is the word $x_2x_3 \cdots$. The map T is continuous with respect to the product topology. The *language* $\mathcal{L}_{\mathbf{x}}$ of $\mathbf{x}$ is its set of factors (the preceding notions of prefix, suffix, factor, and occurrence directly generalize to infinite words), and by $\mathcal{L}_{\mathbf{x}}$ we refer to the set of factors of $\mathbf{x}$ of length n. We say that the language $\mathcal{L}_{\mathbf{x}}$ is *closed under reversal* if the reversal of each w in $\mathcal{L}_{\mathbf{x}}$ is also in $\mathcal{L}_{\mathbf{x}}$. If $wa, wb \in \mathcal{L}_{\mathbf{x}}$ for distinct letters a and b, then we say that the factor w of $\mathbf{x}$ is *right special*; *left special* factors are defined analogously. If a factor is both right and left special, we say it is *bispecial*. The set $\{\mathbf{w} \in \mathcal{A}^\omega : \mathcal{L}_{\mathbf{w}} \subseteq \mathcal{L}_{\mathbf{x}}\}$ is called the *subshift* generated by $\mathbf{x}$. The subshift is a T-invariant and closed set. An infinite word $\mathbf{x}$ is *ultimately periodic* if $\mathbf{x} = uv^\omega$ where u and v are finite words, $v \neq \varepsilon$, and $v^\omega = vvv \cdots$. If $\mathbf{x}$ is not ultimately periodic, then it is *aperiodic*. If $\mathbf{x} = x_1x_2 \cdots$ with $x_i \in \mathcal{A}$ and $\tau\colon A^* \to A^*$ is a morphism, then we define $\tau(\mathbf{x})$ to be the infinite word $\tau(x_1)\tau(x_2)\cdots$.

3. Episturmian words and generalized Ostrowski numeration systems

Episturmian words were introduced in [16] as generalizations of Sturmian words based on palindromic closure.

Let $\mathcal{A}$ be the integer alphabet $\{0, 1, \ldots, d-1\}$ of d letters. Let $w^{(+)}$ be the shortest palindrome having the word w as a prefix. Let $\Delta = y_1y_2 \cdots$ be an infinite word over $\mathcal{A}$ and define a sequence (u_k) of finite words as follows:

$$u_1 = \varepsilon,$$

$$u_{k+1} = (u_k y_k)^{(+)},$$

and set $\mathbf{c}_\Delta = \lim_{k\to\infty} u_k$. We say that $\mathbf{c}_\Delta$ is a *standard episturmian word* with directive word Δ. In what follows, we often use the name *epistandard* for a standard episturmian word. Each epistandard word has a unique directive word. The words u_k are called *central words* and they are exactly the palindromic prefixes of $\mathbf{c}_\Delta$. In fact, the words u_k are exactly the bispecial factors of $\mathbf{c}_\Delta$. This means in particular that every prefix of $\mathbf{c}_\Delta$ is left special.

An infinite word $\mathbf{t}$ is *episturmian* with directive word Δ if $\mathcal{L}_\mathbf{t} = \mathcal{L}_{\mathbf{c}_\Delta}$. Equivalently, a word $\mathbf{t}$ is episturmian if $\mathcal{L}_\mathbf{t}$ is closed under reversal and $\mathbf{t}$ has at most one right special factor of length n for all n [16, Thm. 5]. If $\mathbf{t}$ is binary and aperiodic, then we call $\mathbf{t}$ *Sturmian*. It is equivalent to require that Δ is binary and contains both 0 and 1 infinitely often.

An episturmian word is ultimately periodic if and only if the directive word Δ is eventually constant, that is, $y_n = a$ for some $a \in \mathcal{A}$ for all n large enough [16, Thm. 3]. In this paper, we consider only aperiodic episturmian words, so we assume that (y_n) is not eventually constant.

3.1. The intercept of an episturmian word

Our next aim is to desubstitute an episturmian word with episturmian morphisms in a certain way that gives rise to the concept of the intercept of an episturmian word. This generalizes the arguments of [10] for Sturmian words.

Let $\mathcal{A}$ be the integer alphabet $\{0, 1, \ldots, d-1\}$ as before. For each $y \in \mathcal{A}$, define the morphisms L_y as follows:

$$L_y(x) = \begin{cases} y, & \text{if } x = y, \\ yx, & \text{if } x \neq y. \end{cases}$$

These morphisms belong to the class of episturmian morphisms; see [19, Sect. 3]. Let $\mathbf{t}$ be an episturmian word over $\mathcal{A}$ with directive word $y_1 y_2 \cdots$. Then, depending on if the first letter of $\mathbf{t}$ is y_1, we have $\mathbf{t} = L_{y_1}(\mathbf{t}_1)$ or $\mathbf{t} = T(L_{y_1}(\mathbf{t}_1))$ for some unique infinite word $\mathbf{t}_1$. It is well-known that $\mathbf{t}_1$ is also an episturmian word over $\mathcal{A}$ (possibly over a strict subalphabet of $\mathcal{A}$) [23, Thm. 3.10]. Thus there exist an integer b_1 in $\{0, 1\}$ and a unique episturmian word $\mathbf{t}_1$ such that $\mathbf{t} = T^{b_1} \circ L_{y_1}(\mathbf{t}_1)$. By repeating this decoding, we see that there exists a unique integer sequence (b_k) and a unique sequence $(\mathbf{t}_k)$ of episturmian words such that

$$\mathbf{t} = (T^{b_1} \circ L_{y_1}) \circ \cdots \circ (T^{b_k} \circ L_{y_k})(\mathbf{t}_k) \tag{1}$$

for all $k \geq 1$. It is easy to see that if $y_k = y_{k+1}$ and $b_{k+1} = 0$, then $b_k = 0$. Indeed, if $b_k = 1$, then $\mathbf{t}_{k-1}$ does not begin with y_k so, by the form of the morphism $L_{y_{k+1}}$, neither does $\mathbf{t}_k$ begin with y_k. Since $y_k = y_{k+1}$, this implies that $b_{k+1} = 1$.

Let us write the directive word Δ more compactly as follows:

$$\Delta = y_1 y_2 \cdots = x_1^{a_1} x_2^{a_2} \cdots \tag{2}$$

with $a_k \geq 1$, and $x_k \neq x_{k+1}$ for all k. We call the sequence (a_k) the sequence of *partial quotients* of $\mathbf{t}$ (the choice of the name will become apparent below). Let $r_0 = 0$ and $r_k = a_1 + \cdots + a_k$ for $k \geq 1$. By the property given at the end of the previous paragraph, we see that $b_{r_k+1} \cdots b_{r_{k+1}}$ (viewed as a word over $\{0, 1\}$) is of the form 0^*1^*, so we may write

$$b_1 b_2 \cdots = 0^{a_1 - c_1} 1^{c_1} 0^{a_2 - c_2} 1^{c_2} \cdots$$

for some integers c_k such that $0 \leq c_k \leq a_k$ for all k. Therefore

$$\mathbf{t} = \left(L_{x_1}^{a_1 - c_1} \circ (T \circ L_{x_1})^{c_1}\right) \circ \cdots \circ \left(L_{x_k}^{a_k - c_k} \circ (T \circ L_{x_k})^{c_k}\right)(\mathbf{t}_{r_k})$$

for all k. It is easy to verify that $L_y \circ T \circ L_y = T \circ L_y \circ L_y$ for all letters y, so we find that

$$\mathbf{t} = T^{c_1} L_{x_1}^{a_1} \circ \cdots \circ T^{c_k} L_{x_k}^{a_k}(\mathbf{t}_{r_k}) \tag{3}$$

for all k. We call the sequence $c_1 c_2 \cdots$ the *intercept* of $\mathbf{t}$. The choice of the name will become apparent after we discuss Sturmian words below. Notice that the intercept of $\mathbf{t}$ is unique. Notice also the important fact that the above derivation guarantees that $T^{c_1}L_{x_1}^{a_1} \circ \cdots \circ T^{c_k}L_{x_k}^{a_k}(z)$ is nonempty when z is the first letter of $\mathbf{t}_{r_k}$.

Lemma 3.1. *Let Δ be a directive word as in (2). The intercept of the epistandard word $\mathbf{c}_\Delta$ is 0^ω.*

Proof. The word $\mathbf{c}_\Delta$ begins with x_1, so $b_1 = 0$. The claim follows by induction because $\mathbf{t}_1$ is epistandard by [16, Thm. 9]. □

Next we introduce several sequences of morphisms and words in order to define the important generalized standard words and show their connection to the central words u_k. See [23, Sect. 2] for a slightly more elaborate presentation.

Let Δ be a directive word as in (2). Set

$$\mu_k = L_{y_1} \circ \cdots \circ L_{y_k} \quad \text{and} \quad \tau_k = \mu_{r_k} = L_{x_1}^{a_1} \circ \cdots \circ L_{x_k}^{a_k}$$

with μ_0 and τ_0 being the identity map, and define the sequences $(h_k)_{k\geq 0}$, $(s_k)_{k\geq 0}$, and $(q_k)_{k\geq 0}$ by setting

$$h_k = \mu_k(y_{k+1}), \quad s_k = h_{r_k} = \tau_k(x_{k+1}), \quad \text{and} \quad q_k = |s_k|.$$

The words s_k are the (finite) *standard words* associated with the directive word Δ. By definition, the epistandard word $\mathbf{c}_\Delta$ is the limit of both (h_k) and (s_k). The words s_k are primitive [23, Prop. 2.8]. Notice that if i is such that $r_{k-1} \leq i < r_k$, then $h_k = s_{k-1}$.

Let $(\mathbf{c}_n)$ be the sequence of epistandard words such that $\mathbf{c}_\Delta = \mu_n(\mathbf{c}_n)$ as in (1) (see Lemma 3.1) and $(\mu_{n,k})$, $(\tau_{n,k})$, $(u_{n,k})$, $(h_{n,k})$, and $(s_{n,k})$ be the respective sequences for $\mathbf{c}_n$. A simple induction argument (see [23, Eq. 3]) shows that

$$u_k = \mu_p(u_{p,k-p})u_{p+1} \tag{4}$$

for all p such that $0 \leq p < k$. Replacing k by $k+1$ and p by $k-1$ in (4) yields

$$u_{k+1} = h_{k-1}u_k \tag{5}$$

for all $k \geq 1$. Using (5) repeatedly, we obtain

$$u_k = h_{k-2}h_{k-3}\cdots h_{p-1}u_p = h_{k-2}h_{k-3}\cdots h_1 h_0 \tag{6}$$

for $k \geq 2$. The Eq. (6) directly implies the following important formula for $k \geq 1$:

$$u_{r_k+1} = s_{k-1}^{a_k}s_{k-2}^{a_{k-1}}\cdots s_0^{a_1}. \tag{7}$$

Definition 3.2. Let Δ be as in (2), and let $P(k) = \max\{p < k : y_p = y_k\}$ if this integer exists, and leave $P(k)$ undefined otherwise. Define $j(k)$ as the largest j such that $j \leq k$ and $x_j = x_{k+1}$ when $P(r_k+1)$ exists and leave $j(k)$ undefined otherwise.

For the next lemma, we make the following convention which makes our formulas less "noisy". We often have formulas involving $s_k^{a_{k+1}}$ where the subscript $k+1$ in the superscript a_{k+1} is one greater than in the subscript k. This conveys no essential information, so we will write $s_k^{a_*}$ instead whenever there is no risk of confusion. We also take $s_k^{a_*-1}$ to mean $s_k^{a_{k+1}-1}$ and $s_k^{a_*-c_*}$ to mean $s_k^{a_{k+1}-c_{k+1}}$ etc.

Lemma 3.3. *Let Δ be a directive word as in (2) and $k \geq 0$. If $P(r_k+1)$ exists, then*

$$s_k = s_{k-1}^{a_*}\cdots s_{j(k)}^{a_*}s_{j(k)-1}. \tag{8}$$

If $P(r_k+1)$ does not exist, then

$$s_k = s_{k-1}^{a_*}\cdots s_0^{a_*}x_{k+1}. \tag{9}$$

Proof. Suppose first that $P(r_k+1)$ exists. Set $j=j(k)$. First of all, as $x_k \neq x_{k+1}$, we have $s_k = \tau_{k-1}(L_{x_k}^{a_k}(x_{k+1})) = \tau_{k-1}(x_k^{a_k} x_{k+1}) = s_{k-1}^{a_k}\tau_{k-1}(x_{k+1})$. As long as $k-t>j$, we have $x_{k-t}\neq x_{k+1}$ and $\tau_{k-t}(x_{k+1}) = s_{k-t-1}^{a_{k-t}}\tau_{k-t-1}(x_{k+1})$ by a similar computation. Hence we find that

$$s_k = s_{k-1}^{a_*}\cdots s_j^{a_*}\tau_j(x_{k+1}) = s_{k-1}^{a_*}\cdots s_j^{a_*}\tau_{j-1}(x_j) = s_{k-1}^{a_*}\cdots s_j^{a_*}s_{j-1}$$

because $x_j = x_{k+1}$. Say $P(r_k+1)$ does not exist. Then $x_{k-t}\neq x_{k+1}$ for all t such that $0\le t<k$ and the above arguments yield

$$s_k = s_{k-1}^{a_*}\tau_{k-1}(x_{k+1}) = \cdots = s_{k-1}^{a_*}\cdots s_0^{a_*}\tau_0(x_{k+1}).$$

The claim follows as $\tau_0(x_{k+1}) = x_{k+1}$. □

The Eqs. (8) and (9) imply that

$$q_k = a_k q_{k-1} + \cdots + a_{j(k)+1}q_{j(k)} + q_{j(k)-1} \tag{10}$$

when $P(r_k+1)$ exists and

$$q_k = a_k q_{k-1} + \cdots + a_1 q_0 + 1 \tag{11}$$

when $P(r_k+1)$ does not exist. If Δ is periodic with period d, then we see that (q_k) satisfies a linear recurrence of order d.

3.2. Regular episturmian words

In this section, we define regular episturmian words which have directive words of special form. This subclass of episturmian words has not had specific attention except for the paper of Glen [18] where the powers occurring in these words are studied.

Definition 3.4. Let Δ be a directive word as in (2). If there exists an integer d such that $d\ge 2$, the letters $x_1,\cdots,x_d$ are pairwise distinct, and $x_1x_2\cdots = (x_1\cdots x_d)^\omega$, then we say that the directive word Δ is *regular* with period d. An episturmian word is regular if its directive word is regular.

In what follows, we often assume that $x_1x_2\cdots = (012\cdots(d-1))^\omega$ for some integer d such that $d\ge 2$. Notice that regular episturmian words are exactly the Sturmian words when $d=2$. This class includes the d-bonacci words $\mathbf{f}_d$ which are the epistandard words having directive words $(012\cdots(d-1))^\omega$. The 2-bonacci word is called the *Fibonacci word*, and the 3-bonacci word is called the *Tribonacci word*.

The main advantage in studying regular episturmian words is that the function $j(k)$ is simple: $j(k) = k-(d-1)$ when $j(k)$ is defined, i.e., when $k\ge d$. This simplifies many properties. For example, from (8) and (9), we have

$$s_k = s_{k-1}^{a_*}\cdots s_0^{a_*}x_{k+1} \tag{12}$$

for $1\le k<d$, and

$$s_k = s_{k-1}^{a_*}\cdots s_{k-(d-1)}^{a_*}s_{k-d} \tag{13}$$

for $k\ge d$. Two consecutive applications of (13) show that

$$(a_{k+1}+1)q_k = q_{k+1} + (a_{k-(d-2)}-1)q_{k-(d-1)} + q_{k-d}$$

for $k\ge d$. In particular, we see that

$$q_{k+1} < (a_{k+1}+1)q_k \tag{14}$$

for $k\ge d$. In a similar fashion, combining (7) and (13) yields that

$$|u_{r_k}| \ge a_k q_{k-1} + q_{k-(d+2)} \tag{15}$$

for $k\ge d+2$.

We believe that most of the results of this paper can be carried out for general episturmian words. However, this leads to very complicated arguments; the arguments are already tedious and complicated in the regular case.

3.3. Generalized Ostrowski numeration systems

Let us now define a representation for a nonnegative integer n in terms of the shift $T^n(\mathbf{c}_\Delta)$ of the epistandard word $\mathbf{c}_\Delta$. First, we prove a generalization of a famous result of Brown [11, Thm. 2].

Proposition 3.5. *Let Δ be a directive word as in (2), and let n be a positive integer. Let $c_1c_2\cdots$ be the intercept of $T^n(\mathbf{c}_\Delta)$. Then there exists an integer k such that $c_k \neq 0$ and $c_i = 0$ for all $i > k$. Moreover the prefix of $\mathbf{c}_\Delta$ of length n equals $s_{k-1}^{c_k}s_{k-2}^{c_{k-1}}\cdots s_0^{c_1}$.*

Proof. We prove the claim by induction on n. From Lemma 3.1, it follows that $\mathbf{c}_\Delta = \tau_1(\mathbf{c}_{\Delta'})$ where $\Delta' = T^{a_1}(\Delta)$. Thus if $n \leq a_1$, then $n0^\omega$ is a valid intercept for $T^n(\mathbf{c}_\Delta)$, and the uniqueness of the intercept implies that $c_1 = n$. By definition, the word s_1 is a prefix of $\mathbf{c}_\Delta$ and $s_1 = \tau_1(x_2) = x_1^{a_1}x_2$, so $s_0^{c_1}$ is a prefix of $\mathbf{c}_\Delta$. This establishes the base case.

Suppose that $n > a_1$. Then it follows from (3) and the arguments preceding it that $T^{n-c_1q_0}(\mathbf{c}_\Delta)$ is a τ_1-image of an episturmian word $\mathbf{t}$ with intercept $c_2c_3\cdots$ (recall from above that $q_0 = 1$). In fact, as in the proof of Lemma 3.1, the word $\mathbf{t}$ is a suffix of the epistandard word $\mathbf{c}_{\Delta'}$. Let w be the prefix of $\mathbf{c}_{\Delta'}$ such that $|\tau_1(w)| = n - c_1q_0$, that is, say $\mathbf{t} = T^{|w|}(\mathbf{c}_{\Delta'})$. The word w must be nonempty as $n > a_1$ and $c_1 \leq a_1$. Now $0 < |w| < n$, so the induction hypothesis implies that there exists an integer k such that $c_k \neq 0$, $c_i = 0$ for all $i > k$, and

$$w = s_{\Delta',k-2}^{c_k}\cdots s_{\Delta',0}^{c_2}$$

where $s_{\Delta',j}$ is the jth standard word for the directive word Δ'. By definition, we have $\tau_1(s_{\Delta',j}) = s_{j+1}$ for all j, so

$$\tau_1(w) = s_{k-1}^{c_k}s_{k-2}^{c_{k-1}}\cdots s_1^{c_2}.$$

Since $T^{n-c_1q_0}(\mathbf{c}_\Delta) = \tau_1(\mathbf{t})$, the word $T^{n-c_1q_0}(\mathbf{c}_\Delta)$ must begin with $x_1^{a_1}$. It follows that $\tau_1(w)s_0^{c_1}$ is a prefix of $T^n(\mathbf{c}_\Delta)$. Since $|\tau_1(w)s_0^{c_1}| = n$, the claim follows. □

The connection to Brown's result becomes clearer as we study greedy expansions below. Thanks to Proposition 3.5, we can give the following definitions.

Definition 3.6. Let Δ be a directive word as in (2), and let n be a positive integer. We let the *representation*, or the Ostrowski expansion, $\mathrm{rep}_\Delta(n)$ of n to be the word $c_1\cdots c_k$ where $c_k \neq 0$ and $c_1\cdots c_k0^\omega$ is the intercept of the word $T^n(\mathbf{c}_\Delta)$. In addition, we set $\mathrm{rep}_\Delta(0) = \varepsilon$.

Definition 3.7. Let Δ be a directive word as in (2). If $c_1\cdots c_k$ is a sequence of nonnegative integers, then we set the *value* $\mathrm{val}_\Delta(c_1\cdots c_k)$ of $c_1\cdots c_k$ to be the number

$$\sum_{i=0}^{k} c_iq_{i-1}.$$

We often omit the subscript Δ in $\mathrm{rep}_\Delta(n)$ and $\mathrm{val}_\Delta(w)$ if the directive word Δ is clear from context.

It follows from Proposition 3.5 that if $\mathrm{rep}_\Delta(n) = c_1\cdots c_k$, then $\mathrm{val}_\Delta(c_1\cdots c_k) = n$. Therefore the Ostrowski expansion of an integer can be viewed as an expansion with respect to the numeration system associated with the sequence (q_k) (for a gentle introduction to numeration systems, see the book [33]). However, we emphasize that the Ostrowski expansion of a number n does not necessarily coincide with the greedy expansion of n with respect to (q_k) as indicated by the following example.

Example 3.8. Consider the nonregular directive word $\Delta = 010(201)^\omega$. Then $(q_k) = (1, 2, 3, 7, \ldots)$, so the greedy expansion of the number 6 is 002 since $6 = 0\cdot 1 + 0\cdot 2 + 2\cdot 3$. Now $\mathbf{c}_\Delta = 0100102\cdots$, so $T^6(\mathbf{c}_\Delta) = 2\cdots$ and $c_1 = 1$. As in the proof of Proposition 3.5, the prefix 01001 of $\mathbf{c}_\Delta$ is an L_0-image and $T^6(\mathbf{c}_\Delta) = T(L_0(T^3(\mathbf{c}_{\Delta'})))$ where $\Delta' = 10(201)^\omega$. Next $\mathbf{c}_{\Delta'} = 1012\ldots$, so $c_2 = 1$

and $T^3(\mathbf{c}_{\Delta'}) = T(L_1(T(\mathbf{c}_{\Delta''})))$ with $\Delta'' = 0(201)^\omega$. As the intercept of $\mathbf{c}_{\Delta''}$ is 0^ω, we find that $\mathrm{rep}_\Delta(6) = 111$, so the Ostrowski expansion of 6 is different from its greedy expansion.

Our next aim is to prove that the Ostrowski expansion of an integer coincides with the greedy expansion in the important special case of regular episturmian words. We leave the characterization open in the case of nonregular directive words. Whenever we discuss greedy expansions below, we assume that the greedy expansion is written the least significant digit first and without trailing zeros.

Definition 3.9. Given a directive word Δ as in (2), an infinite word $c_1 c_2 \cdots$ over the alphabet $\{0, 1, 2, \ldots\}$ satisfies the *Ostrowski conditions* if $0 \leq c_k \leq a_k$ for all k and for all $k \geq 1$ the following implication holds:

$$P(r_k + 1) \text{ exists and } c_i = a_i \text{ for all } i \text{ such that } j(k) < i \leq k \Longrightarrow c_{j(k)} = 0.$$

A finite word $c_1 \cdots c_k$ satisfies the Ostrowski conditions if it is a prefix of an infinite word satisfying the Ostrowski conditions.

If Δ is regular with period d, then the Ostrowski conditions state that $0 \leq c_k \leq a_k$ for all k and that if $c_i = a_i$ for all i such that $k - (d-1) < i \leq k$ and $k \geq d$, then $c_{k-(d-1)} = 0$.

Lemma 3.10. *The intercept of a regular episturmian word satisfies the Ostrowski conditions.*

Proof. Let $c_1 c_2 \cdots$ be the intercept of a regular episturmian word $\mathbf{t}$ with directive word Δ as in (2). The property that $0 \leq c_k \leq a_k$ for all k follows directly from the derivation of the intercept preceding Lemma 3.1. Suppose that $P(r_k + 1)$ exists, that is, say $k \geq d$, and assume that $c_i = a_i$ for all i such that $k - (d-1) < i \leq k$. Let z be the first letter of $\mathbf{t}_{r_k}$. From (3) and the discussion following it, we see that

$$T^{c_1} L_{x_1}^{a_1} \circ \cdots \circ T^{c_k} L_{x_k}^{a_k}(z) \tag{16}$$

is a nonempty prefix of $\mathbf{t}$. Suppose for a contradiction that there exists a largest ℓ such that $z = x_\ell$ with $k - (d-1) < \ell \leq k$. Thus $T^{c_i} L_{x_i}^{a_i}(z) = T^{a_i}(x_i^{a_i} z) = z$ when $i > \ell$. It follows that

$$T^{c_\ell} L_{x_\ell}^{a_\ell} \circ \cdots \circ T^{c_k} L_{x_k}^{a_k}(z) = T^{c_\ell} L_{x_\ell}^{a_\ell}(z) = \varepsilon$$

because $z = x_\ell$ and $c_\ell = a_\ell > 0$. This contradicts that (16) is nonempty. Since the directive word Δ contains d distinct letters, we deduce that $z = x_{k-(d-1)}$. Thus $z = x_{k+1}$ for $x_1 x_2 \cdots$ is d-periodic. Clearly $T^{c_i} L_{x_i}^{a_i}(x_{k+1}) = T^{a_i}(x_i^{a_i} x_{k+1}) = x_{k+1}$ for all i such that $k - (d-1) < i \leq k$. Since the prefix (16) must be nonempty, we deduce that $T^{c_{k-(d-1)}} L_{x_{k-(d-1)}}^{a_{k-(d-1)}}(x_{k+1}) \neq \varepsilon$. As $x_{k-(d-1)} = x_{k+1}$, the only option is that $c_{k-(d-1)} = 0$. Therefore $c_1 c_2 \cdots$ satisfies the Ostrowski conditions. □

The arguments presented so far generalize those of [10]. In the case of Sturmian words, the sequence of partial quotients (a_k) can be viewed as the continued fraction expansion $[0; a_1, a_2, \ldots]$ of a number α which equals the frequency of the letter x_2 in $\mathbf{t}$. A Sturmian word $\mathbf{t}$ can be seen as a coding of the rotation $x \mapsto x + \alpha$ of a point ρ in the torus $[-\alpha, 1 - \alpha)$. The point ρ, often called the intercept of $\mathbf{t}$ due to the connection to so-called mechanical words, can be expressed as a sum

$$\sum_{k=1}^{\infty} c_k (q_{k-1}\alpha - p_{k-1})$$

where p_{k-1}/q_{k-1} are the convergents of α and (c_k) is an integer sequence satisfying the conditions $0 \leq c_k \leq a_k$ for all k and $c_{k+1} = a_{k+1} \Longrightarrow c_k = 0$ for all k. The sequence (c_k) is exactly the intercept of $\mathbf{t}$ in the sense we defined above and the conditions match the Ostrowski conditions. Moreover, the denominators of the convergents match the lengths of the associated standard words. The sum representation of ρ often goes by the name of Ostrowski expansion of a real number. For the proofs of these facts and more details, see [10]. Thus the concepts we defined have nice number-theoretic interpretations in the context of Sturmian words. Unfortunately no such interpretations are known for general episturmian words. See [16, Sect. 5] for an intercept defined for episturmian words that are fixed points of morphisms.

Equalities like (3) make sense even when the word $c_1c_2\cdots$ does not satisfy the Ostrowski conditions. A whole theory of so-called spinned directive words has been developed for studying alternative representations of episturmian words; see [19, Sect. 4]. Our notion of intercept coincides with the normalized directive word of [20].

Lemma 3.11. *Let Δ be a directive word as in (2). If $c_1\cdots c_k$ satisfies the Ostrowski conditions, then* $\mathrm{val}_\Delta(c_1\cdots c_k) < q_k$.

Proof. Let us prove the claim by induction on k. The base case is established by observing that $\mathrm{val}(\varepsilon) = 0 < q_0 = 1$. Say $P(r_k+1)$ exists, and set $j = j(k)$. Assume first that there exists a largest i such that $j < i \leq k$ and $c_i < a_i$. Then

$$\begin{aligned}\mathrm{val}(c_1\cdots c_k) &= \mathrm{val}(c_1\cdots c_{i-1}) + c_iq_{i-1} + a_{i+1}q_i + \cdots + a_kq_{k-1} \\ &< q_{i-1} + (a_i-1)q_{i-1} + a_{i+1}q_i + \cdots + a_kq_{k-1} \\ &\leq a_{j+1}q_j + \cdots + a_kq_{k-1} \\ &< q_k\end{aligned}$$

where the first inequality follows from the induction hypothesis and the final inequality follows from (10). Suppose then that no i like above exists. The Ostrowski conditions now imply that $c_j = 0$, so

$$\begin{aligned}\mathrm{val}(c_1\cdots c_k) &= \mathrm{val}(c_1\cdots c_{j-1}) + a_{j+1}q_j + \cdots + a_kq_{k-1} \\ &< q_{j-1} + a_{j+1}q_j + \cdots + a_kq_{k-1} \\ &= q_k\end{aligned}$$

by the induction hypothesis and (10).

Suppose then that $P(r_k+1)$ does not exist. Then

$$\mathrm{val}(c_1\cdots c_k) \leq \mathrm{val}(a_1\cdots a_k) = a_1q_0 + \cdots a_kq_{k-1} < q_k$$

by (11). The claim follows. □

Proposition 3.12. *Let Δ be a regular directive word. Let $c_1\cdots c_k$ be the greedy expansion of a nonnegative integer n with respect to the numeration system associated with (q_k). Then $\mathrm{rep}_\Delta(n) = c_1\cdots c_k$.*

Proof. If $n = 0$, then the claim is clear as the greedy expansion of 0 is ε by convention. Suppose that $n > 0$. Let ℓ be the largest integer such that $q_{\ell-1} \leq n$. Then there exists a unique nonnegative integers b_ℓ and $r_{\ell-1}$ such that $n = b_\ell q_{\ell-1} + r_{\ell-1}$ with $r_{\ell-1} < q_{\ell-1}$. Writing similarly $r_i = b_iq_{i-1} + r_{i-1}$ for $i = \ell-1, \ldots, 1$ yields the greedy expansion $b_1\cdots b_\ell$ of n with respect to (q_k). It is evident that $\mathrm{val}_\Delta(b_1\cdots b_i) < q_i$ for $i = 1, \ldots, \ell$. Let then $\mathrm{rep}_\Delta(n) = c_1\cdots c_k$. It follows from Lemma 3.11 that $\mathrm{val}_\Delta(c_1\cdots c_i) < q_i$ for $i = 1, \ldots, k$. For the claim it suffices to prove that $b_1\cdots b_\ell = c_1\cdots c_k$.

Without loss of generality, we may assume that $k \leq \ell$. If $k < \ell$, then $\mathrm{val}(b_1\cdots b_\ell) \geq \mathrm{val}(0^{\ell-1}1) = q_{\ell-1}$. Then $\mathrm{val}(c_1\cdots c_k) < q_k \leq q_{\ell-1}$ which contradicts that $\mathrm{val}_\Delta(b_1\cdots b_\ell) = \mathrm{val}_\Delta(c_1\cdots c_k) = n$. Therefore $k = \ell$. Suppose by symmetry that $c_k \leq b_\ell$. If $c_k < b_\ell$, then $b_1\cdots b_{\ell-1}(b_\ell - c_k)$ and $c_1\cdots c_{k-1}0$ represent the same number which is impossible by what we just argued. Thus $b_\ell = c_k$. By repeating the argument for the words $c_1\cdots c_{k-1}$ and $b_1\cdots b_{\ell-1}$, which represent the same number, we see that $c_1\cdots c_k = b_1\cdots b_\ell$. The claim follows. □

Proposition 3.12 allows easy computation of Ostrowski expansions in the case of regular directive words. For example, if $\Delta = (012)^\omega$, then $(q_k) = (1, 2, 4, 7, 13, 24, 44, \ldots)$ and $\mathrm{rep}_\Delta(7) = 0001$ and $\mathrm{rep}_\Delta(10) = 1101$ as $10 = 1 + 2 + 7$.

Observe how the proof of Proposition 3.12 fails in Example 3.8. We have $\mathrm{rep}(6) = 111$, but it is not true that $\mathrm{val}(11) = q_0 + q_1 = 1 + 2 = 3 < q_2 = 3$. Thus a result akin to Lemma 3.11 is not valid for general directive words.

3.4. Auxiliary results on generalized standard words

In this section, we prove further results on generalized standard words needed in this paper. Most of the presented results appear in some form in [23, Sect. 3.3]; our proofs follow a somewhat different philosophy being based purely on properties of the numeration system. See also [24] on additional results on the numeration system.

Proposition 3.5 states that if in an episturmian word has intercept $c_1 \cdots c_k 0^\omega$, then it equals $T^{\mathrm{val}(c_1 \cdots c_k)}(\mathbf{c}_\Delta)$. We generalize this to arbitrary intercepts as follows. This result is found as [16, Thm. 3.20].

Theorem 3.13. *If* $\mathbf{t}$ *is an episturmian word with directive word* Δ *as in (2) and intercept* $c_1 c_2 \cdots$*, then*

$$\mathbf{t} = \lim_{k \to \infty} T^{\mathrm{val}(c_1 \cdots c_k)}(\mathbf{c}_\Delta).$$

For the proof, we need the following lemma.

Lemma 3.14. *Let* Δ *be a directive word as in (2) and* $c_1 c_2 \cdots$ *be an intercept. Then there exists an integer* k *such that* $c_k < a_k$*.*

Proof. This proof is similar to that of Lemma 3.10. Let z_k be the first letter of $\mathbf{t}_{r_k}$ where $\mathbf{t}_{r_k}$ is as in (3). Since Δ contains finitely many distinct letters and is not eventually constant, there must exist integers j and k such that $j < k$, $x_j = z_k$, and $x_i \neq z_k$ when $j < i \leq k$. If $c_i < a_i$ for some i such that $j < i \leq k$, then the claim is clear, so assume that $c_i = a_i$ when $j < i \leq k$. It follows that $T^{c_i} L_{x_i}^{a_i}(z_k) = T^{a_i}(x_i^{a_i} z_k) = z_k$ for all i such that $j < i \leq k$. Therefore

$$T^{c_j} L_{x_j}^{a_j} \circ \cdots T^{c_k} L_{x_k}^{a_k}(z_k) = T^{c_j} L_{x_j}^{a_j}(z_k) = T^{c_j}(z_k)$$

since $x_j = z_k$. It follows from (3) that $T^{c_j}(z_k)$ must be nonempty, so we conclude that $c_j = 0$. The claim follows. □

Proof of Theorem 3.13. Since the suffix of an intercept is a valid intercept, Lemma 3.14 implies that there exists an increasing integer sequence (k_n) such that $c_{k_n} < a_{k_n}$ for all n. From (3), we have

$$\mathbf{t} = T^{c_1} L_{x_1}^{a_1} \cdots T^{c_{k_n - 1}} L_{x_{k_n - 1}}^{a_{k_n - 1}}(x_{k_n}^{a_{k_n} - c_{k_n}} z_{k_n} \cdots)$$

for a letter z_{k_n}. Thus $\mathbf{t}$ and the episturmian word with intercept $c_1 \cdots c_{k_n} 0^\omega$ have common prefix

$$T^{c_1} L_{x_1}^{a_1} \cdots T^{c_{k_n - 1}} L_{x_{k_n - 1}}^{a_{k_n - 1}}(x_{k_n}^{a_{k_n} - c_{k_n}} z_{k_n}).$$

By definition, we have $|T^{c_1} L_{x_1}^{a_1} \cdots T^{c_{k_n - 1}} L_{x_{k_n - 1}}^{a_{k_n - 1}}(x_{k_n})| \geq 1$, so this common prefix equals

$$T^{c_1} L_{x_1}^{a_1} \cdots T^{c_{k_n - 1}} L_{x_{k_n - 1}}^{a_{k_n - 1}}(x_{k_n}) \cdot \tau_{k_n - 1}(x_{k_n}^{a_{k_n} - c_{k_n} - 1} z_{k_n}).$$

The length of this common prefix is thus has at least $|\tau_{k_n - 1}(z_{k_n})|$, and this length tends to infinity as $n \to \infty$ provided that the directive sequence Δ is not ultimately constant (we always assume this). If we denote the episturmian word with intercept $c_1 \cdots c_k 0^\omega$ by $\mathbf{t}_k$, we have $\mathbf{t} = \lim_{k \to \infty} \mathbf{t}_k$. Since $\mathbf{t}_k = T^{\mathrm{val}(c_1 \cdots c_k)}(\mathbf{c}_\Delta)$ by Proposition 3.5, the claim follows. □

The significance of Theorem 3.13 is that, in principle, the properties of a general episturmian word reduce to those of shifts of an epistandard word. This result suggests to consider the longest common prefixes of the words in the sequence $(T^{\mathrm{val}(c_1 \cdots c_k)}(\mathbf{c}_\Delta))_k$. This is found in Lemma 3.20, but we need several auxiliary lemmas for the proof.

Lemma 3.15. *The word* u_{r_k} *is a proper prefix of* s_k *for all* $k \geq 1$*.*

Proof. If $k = 1$, then $u_{r_k} = s_0^{a_1 - 1}$ by (5) and (7) and $s_1 = s_0^{a_1} x_2$, so the claim holds. Assume that $k > 1$. Say $P(r_k + 1)$ exists. Since both u_{r_k} and s_k are prefixes of $\mathbf{c}_\Delta$, it suffices to show that

$|u_{r_k}| < |s_k|$. By (5) and (7), we have $|u_{r_k}| = (a_k - 1)q_{k-1} + a_{k-1}q_{k-2} + \cdots + a_1 q_0$. Moreover, we have $q_k = a_k q_{k-1} + \cdots + a_{j(k)+1}q_{j(k)} + q_{j(k)-1}$ by (8). Thus $|s_k| - |u_{r_k}| \geq q_{k-1} - |u_{r_{j(k)}}|$. Since $j(k) < k$, we have $|u_{r_{j(k)}}| \leq |u_{r_{k-1}}|$. Thus by the induction hypothesis, we have $|s_k| - |u_{r_k}| \geq q_{k-1} - |u_{r_{k-1}}| > 0$. When $P(r_k + 1)$ does not exist, the claim follows by similar arguments. □

Lemma 3.16. *We have the following implications for all $k \geq 0$.*

(i) If $P(r_{k+1} + 1)$ and $P(r_k + 1)$ exist, then $s_{k+1}s_k = s_k^{a_*+1} u_{r_{j(k)}} u_{r_{j(k+1)}}^{-1} s_k$.
(ii) If $P(r_{k+1} + 1)$ exists and $P(r_k + 1)$ does not exist, then $s_{k+1}s_k = s_k^{a_*+1} x_{k+1}^{-1} u_{r_{j(k+1)}}^{-1} s_k$.
(iii) If $P(r_{k+1} + 1)$ does not exist and $P(r_k + 1)$ exists, then $s_{k+1}s_k = s_k^{a_*+1} u_{r_{j(k)}} x_{k+2} s_k$.
(iv) If neither $P(r_{k+1} + 1)$ nor $P(r_k + 1)$ exists, then $s_{k+1}s_k = s_k^{a_*+1} x_{k+1}^{-1} x_{k+2} s_k$.

Proof. Suppose that $P(r_{k+1} + 1)$ exists. As $j(k+1) < k+1$, we see that $u_{r_{j(k+1)}}$ is a prefix of s_k. Applying (7) and (8), we have

$$(s_k^{a_*})^{-1} s_{k+1} s_k = s_{k-1}^{a_*} \cdots s_{j(k+1)}^{a_*} s_{j(k+1)-1} \cdot u_{r_{j(k+1)}} u_{r_{j(k+1)}}^{-1} s_k = s_{k-1}^{a_*} \cdots s_0^{a_*} \cdot u_{r_{j(k+1)}}^{-1} s_k.$$

When $P(r_k + 1)$ exists, then $s_{k-1}^{a_*} \cdots s_0^{a_*} = s_k u_{r_{j(k)}}$ and (i) holds. If $P(r_k + 1)$ does not exist, then we deduce from (9) that

$$(s_k^{a_*})^{-1} s_{k+1} s_k = s_k x_{k+1}^{-1} u_{r_{j(k+1)}}^{-1} s_k,$$

so (ii) holds.

Suppose then that $P(r_{k+1} + 1)$ does not exist. Then $s_{k+1} = s_k^{a_*} \cdots s_0^{a_*} x_{k+2}$ by (9). If $P(r_k + 1)$ exists, then $s_{k-1}^{a_*} \cdots s_0^{a_*} = s_k u_{r_{j(k)}}$ like above, and we have

$$(s_k^{a_*})^{-1} s_{k+1} s_k = s_k u_{r_{j(k)}} x_{k+2} s_k,$$

so (iii) holds. If $P(r_k + 1)$ does not exist, then

$$(s_k^{a_*})^{-1} s_{k+1} s_k = s_k x_{k+1}^{-1} x_{k+2} s_k,$$

and we see that (iv) holds. □

Lemma 3.17. *For all $k \geq 0$, the word $s_k^{a_*+1}$ is a prefix of $\mathbf{c}_\Delta$ if and only if $P(r_k + 1)$ exists. If $P(r_k + 1)$ does not exist, then $s_k^{a_*+1} x_{k+1}^{-1} x_{k+2}$ is a prefix of $\mathbf{c}_\Delta$. The word $s_k^{a_*+2}$ is not a prefix of $\mathbf{c}_\Delta$.*

Proof. By Lemma 3.15, the word $u_{r_{j(k+1)}}$ is a proper prefix of s_k. Say $P(r_k + 1)$ exists. Lemma 3.16 implies that the word $s_k^{a_*+1}$ is a prefix of $s_{k+1}s_k$. Since $s_{k+1}s_k$ is a prefix of $\mathbf{c}_\Delta$, we see that $s_k^{a_*+1}$ is a prefix of $\mathbf{c}_\Delta$. Suppose that $P(r_k + 1)$ does not exist. Then (ii) or (iv) of Lemma 3.16 holds. In the latter case, the word $s_k^{a_*+1}$ cannot be a prefix of $s_{k+1}s_k$ because $x_{k+1} \neq x_{k+2}$ by definition. In the former case, we have $s_{k+1}s_k = s_k^{a_*+1} x_{k+1}^{-1} u_{r_{j(k+1)}}^{-1} s_k$. Since s_k is a prefix of $\mathbf{c}_\Delta$, we see that the words s_k and $u_{r_{j(k+1)}+1}$ share the prefix $u_{r_{j(k+1)}} x_{j(k+1)}$. Therefore $s_{k+1}s_k$ has prefix $s_k^{a_*+1} x_{k+1}^{-1} x_{j(k+1)}$. By definition, we have $x_{j(k+1)} = x_{k+2}$, so we see that $s_k^{a_*+1} x_{k+1}^{-1} x_{k+2}$ is a prefix of $\mathbf{c}_\Delta$. Like above, the word $s_k^{a_*+1}$ is not a prefix of $\mathbf{c}_\Delta$. Thus we have proved the first and second claims.

Let us then prove the final claim. If $P(r_k + 1)$ does not exist, then the second claim shows that $s_k^{a_*+1}$ is not a prefix of $\mathbf{c}_\Delta$, so assume that $P(r_k + 1)$ exists. Suppose first that $P(r_{k+1} + 1)$ also exists. The first claim shows that s_{k+1}^2 is a prefix of $\mathbf{c}_\Delta$. From Lemma 3.16, we see that

$$s_{k+1}^2 = s_k^{a_*+1} u_{r_{j(k)}} u_{r_{j(k+1)}}^{-1} s_k^{a_*+1} u_{r_{j(k)}} u_{r_{j(k+1)}}^{-1}.$$

Because s_k is a prefix of $\mathbf{c}_\Delta$ and $u_{r_{j(k+1)}}$ is a proper prefix of s_k by Lemma 3.15, we see that the prefix $s_k^{a_*+1}$ of s_{k+1}^2 is followed by $u_{r_{j(k)}} x_{k+2}$. Similarly, the word s_k has the word $u_{r_{j(k)}} x_{k+1}$ as a prefix. Since $x_{k+1} \neq x_{k+2}$, we conclude that $s_k^{a_{k+1}+2}$ is not a prefix of $\mathbf{c}_\Delta$.

Assume then that $P(r_{k+1} + 1)$ does not exist. Then $s_{k+1} = s_k^{a_*+1} u_{r_{j(k)}} x_{k+2}$. Since s_{k+1}^2 is a prefix of $\mathbf{c}_\Delta$ and $u_{r_{j(k)}}$ is a proper prefix of s_k, we see that if $s_k^{a_*+2}$ is a prefix of $\mathbf{c}_\Delta$, then $x_{k+2} = x_{j(k)}$. Because $x_{j(k)} = x_{k+1}$ and $x_{k+1} \neq x_{k+2}$, we conclude that $s_k^{a_*+2}$ is not a prefix of $\mathbf{c}_\Delta$. □

Based on Lemma 3.17, we give the following definition.

Definition 3.18. If $P(r_k+1)$ exists, then we let t_k to be the longest word such that $s_k^{a_*+1}t_k$ is a prefix of $\mathbf{c}_\Delta$ with period q_k. If $P(r_k+1)$ does not exist, then we set $t_k = x_{k+1}^{-1}$.

Notice that the word t_k is a proper prefix of s_k when $P(r_k+1)$ exists, but t_k can be empty.

Lemma 3.19. *For all $k \geq 0$, we have $s_k^i t_k = u_{r_k+i}$ for all i such that $1 \leq i \leq a_{k+1}$.*

Proof. Suppose that $P(r_k+1)$ exists. By Lemma 3.17, the word s_k^2 is a prefix of $\mathbf{c}_\Delta$, so $s_k t_k$ is a right special prefix of $\mathbf{c}_\Delta$. Therefore $s_k t_k = u_{\ell+2}$ for some integer ℓ (recall that the central words u_k are as bispecial factors exactly the right special prefixes of $\mathbf{c}_\Delta$). It follows from [23, Eq. 2] that $u_{\ell+2} = L_{x_1}(u_{\ell+1}^{(1)})x_1$ where $u_{\ell+1}^{(1)}$ is the $(\ell+1)$th central word associated with the directive word $T(\Delta)$. The word $u_{\ell+2}$ can be repeatedly decoded like this for a total of $\ell+1$ times to obtain the empty word $u_1^{(\ell+1)}$. On the other hand, we have $s_k t_k = \mu_{r_k}(x_{k+1})t_k$, so it must be possible to decode $s_k t_k$ at least r_k times before obtaining an empty word. Therefore $r_k + 1 \leq \ell + 2$.

Assume for a contradiction that $r_k + 1 < \ell + 2$, that is, $r_k \leq \ell$. By (8), we may write

$$s_k t_k = s_{k-1}^{a_*} \cdots s_{j(k)}^{a_*} s_{j(k)-1} t_k = h_{r_k-1} \cdots h_{r_{j(k)}-1} t_k.$$

On the other hand, by (6), we have $s_k t_k = h_\ell \cdots h_0$. As $r_k \leq \ell$, the word $h_\ell \cdots h_0$ contains the product $h_{r_k-1} \cdots h_{r_{j(k)}-1}$, and we see that $|t_k| \geq |h_\ell| \geq |h_{r_k}| = q_k$. This contradicts the definition of t_k. We have thus proved that $s_k t_k = u_{r_k+1}$. The rest of the claim follows by observing from (5) that the differences $|u_{r_k+i+1}| - |u_{r_k+i}|$ equal q_k for $i = 1, \ldots, a_{k+1}-1$. These correspond to the differences $|s_k^{i+1}t_k| - |s_k^i t_k|$ for $i = 1, \ldots, a_{k+1}-1$.

Let us then assume that $P(r_k+1)$ does not exist. Lemma 3.17 again implies that $s_k t_k$ is a central word. By (9), we have $s_k = s_{k-1}^{a_*} \cdots s_0^{a_*} x_{k+1}$ so $s_k t_k = s_{k-1}^{a_*} \cdots s_0^{a_*} = h_{r_k-1} \cdots h_0 = u_{r_k+1}$ because $t_k = x_{k+1}^{-1}$. The rest of the claim follows as above. □

The following lemma was proved for Sturmian words in [35, Prop. 5.2..2].

Lemma 3.20. *Let Δ be a regular directive word and $c_1c_2\cdots$ be an intercept. Let $k \geq 1$, and assume that there exists a least positive integer n such that $c_{k+n} \neq 0$. The length of the longest common prefix of $T^{\mathrm{val}(c_1\cdots c_k)}(\mathbf{c}_\Delta)$ and $T^{\mathrm{val}(c_1\cdots c_{k+n})}(\mathbf{c}_\Delta)$ equals $|s_{k+n-1}^{a_*-c_*} s_{k+n-2}^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1\cdots c_k)$.*

Proof. As Δ is regular, it follows from Lemmas 3.10 and 3.11 that $q_{k+n-1} - \mathrm{val}(c_1\cdots c_k) \geq 0$. Thus we can let v to be the suffix of s_{k+n-1} of length $q_{k+n-1} - \mathrm{val}(c_1\cdots c_k)$. Suppose that $P(r_{k+n-1}+1)$ exists. Then Lemma 3.17 implies that $s_{k+n-1}^{a_*+1}t_{k+n-1}$ is the longest prefix of $\mathbf{c}_\Delta$ with period q_{k+n-1}. Therefore the word $T^{\mathrm{val}(c_1\cdots c_k)}(\mathbf{c}_\Delta)$ has prefix $v s_{k+n-1}^{a_{k+n}} t_{k+n-1}$. Since $c_{k+1} = \cdots = c_{k+n-1} = 0$, the word $T^{\mathrm{val}(c_1\cdots c_{k+n})}(\mathbf{c}_\Delta)$ has prefix $v s_{k+n-1}^{a_*-c_*} t_{k+n-1}$. Since $c_{k+n} > 0$, we have by the definition of the word t_{k+n-1} that the longest common prefix of the words $T^{\mathrm{val}(c_1\cdots c_k)}(\mathbf{c}_\Delta)$ and $T^{\mathrm{val}(c_1\cdots c_{k+n})}(\mathbf{c}_\Delta)$ equals $v s_{k+n-1}^{a_*-c_*} t_{k+n-1}$. The claim follows by a short computation applying Lemma 3.19.

The same arguments apply when $P(r_{k+n-1}+1)$ does not exist as then the longest prefix of $\mathbf{c}_\Delta$ with period q_{k+n-1} equals $s_{k+n-1}^{a_*+1}t_{k+n-1}$ and $t_{k+n-1} = x_{k+n}^{-1}$. Notice that $|x_{k+n}^{-1}| = -1$. □

Let us then find out when the length of the longest common prefix increases. Let $k \geq 1$ and n_1 and n_2 be the two least positive integers such that $n_1 < n_2$, $c_{k+n_1} \neq 0$, and $c_{k+n_2} \neq 0$. Let v_1 be the longest common prefix of $T^{\mathrm{val}(c_1\cdots c_k)}(\mathbf{c}_\Delta)$ and $T^{\mathrm{val}(c_1\cdots c_{k+n_1})}(\mathbf{c}_\Delta)$ and v_2 be the longest common prefix of $T^{\mathrm{val}(c_1\cdots c_{k+n_1})}(\mathbf{c}_\Delta)$ and $T^{\mathrm{val}(c_1\cdots c_{k+n_2})}(\mathbf{c}_\Delta)$. By Lemma 3.20, we have

$$|v_2| - |v_1| = |s_{k+n_2-1}^{a_*-c_*} s_{k+n_2-2}^{a_*} \cdots s_0^{a_*}| - |s_{k+n_1-1}^{a_*} s_{k+n_1-2}^{a_*} \cdots s_0^{a_*}|,$$

so $|v_2| - |v_1| > 0$ unless $n_2 = n_1 + 1$ and $a_{k+n_2} = c_{k+n_2}$.

Lemma 3.21. *Let Δ be a regular word and $c_1c_2\cdots$ be an intercept. Define*

$$\eta_k = |s_k^{a_*-c_*} s_{k-1}^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1\cdots c_k)$$

for all $k \geq 0$. The sequence (η_k) is nondecreasing and $\lim_{k\to\infty} \eta_k = \infty$.

Proof. Let $k \geq 0$. Then

$$\begin{aligned}\eta_{k+1} &= |s_{k+1}^{a_{k+2}-c_{k+2}} s_k^{a_{k+1}} \cdots s_0^{a_1}| - \mathrm{val}(c_1 \cdots c_{k+1}) \\ &\geq |s_k^{a_{k+1}} \cdots s_0^{a_1}| - \mathrm{val}(c_1 \cdots c_{k+1}) \\ &= |s_k^{a_{k+1}-c_{k+1}} s_{k-1}^{a_k} \cdots s_0^{a_1}| - \mathrm{val}(c_1 \cdots c_k) \\ &= \eta_k,\end{aligned}$$

so (η_k) is nondecreasing. From the discussion preceding this lemma, we see that $\eta_{k+1} = \eta_k$ if and only if $a_{k+2} = c_{k+2}$. By Lemma 3.14, there exist infinitely many k such that $c_{k+2} < a_{k+2}$. Since (η_k) is nondecreasing, this shows that $\lim_{k\to\infty} \eta_k = \infty$. □

Many additional properties of the words s_k could be easily derived, but we do not need them in this paper, so we will stop after the following required result.

Lemma 3.22 ([23, Thm. 3.17])**.** *Let Δ be a directive word as in (2) and y be a letter occurring infinitely many times in Δ. Define an intercept $c_1 c_2 \cdots$ as follows: $c_k = 0$ if $x_k = y$ and $c_k = a_k$ if $x_k \neq y$. Then $c_1 c_2 \cdots$ is the intercept of the word $y\mathbf{c}_\Delta$.*

Proof. Since y occurs infinitely many times in Δ, arbitrarily long central words u_k are followed by the letter y. As the language of $\mathbf{c}_\Delta$ is closed under reversal and u_k are palindromes, we see that yu_k is a factor of $\mathbf{c}_\Delta$ for infinitely many k. Since $\mathbf{c}_\Delta$ is the limit of the sequence (u_k), it follows that $y\mathbf{c}_\Delta$ is an episturmian word with directive word Δ.

Let $d_1 d_2 \cdots$ be the intercept of $y\mathbf{c}_\Delta$. Say $y \neq x_1$. Then $y\mathbf{c}_\Delta = yx_1^{a_1} x_2 \cdots$, so $y\mathbf{c}_\Delta = T \circ L_{x_1}(y\mathbf{c}_{\Delta'})$ where $\Delta' = T(\Delta)$. It follows that $d_1 = a_1$. Suppose then that $y = x_1$. Now it is plain that $y\mathbf{c}_\Delta = L_{x_1}(y\mathbf{c}_{\Delta'})$, so we have $d_1 = 0$. It follows from induction that $y\mathbf{c}_\Delta$ has the claimed intercept. □

4. Rauzy graphs of episturmian words

Let $\mathbf{x}$ be an infinite word. The Rauzy graph $\Gamma(n)$ of order n associated with the language of $\mathbf{x}$ is a directed graph with vertices $\mathcal{L}_\mathbf{x}$ and edges $\mathcal{L}_\mathbf{x}$. There is an edge e from vertex u to vertex v if and only if e has prefix u and suffix v. Each word with the language $\mathcal{L}_\mathbf{x}$ corresponds to an infinite path in the graph $\Gamma(n)$ starting from its prefix of length n. The initial nonrepetitive complexity $\mathrm{inrc}(\mathbf{x}, n)$ can be determined from $\Gamma(n)$: start from the vertex corresponding to the prefix of $\mathbf{x}$ of length n and follow the path dictated by $\mathbf{x}$ until a vertex is repeated for the first time. In general, this is of no help as the graph $\Gamma(n)$ can be very complicated. However, when $\mathbf{x}$ has low factor complexity, there are only few right special factors of length n and the analysis is more likely to succeed. This is indeed the case with episturmian words whose Rauzy graphs have especially nice form.

An episturmian word $\mathbf{t}$ with directive word Δ can be equivalently defined as an infinite word such that its language is closed under reversal and it has exactly one right special factor of each length [16, Thm. 5]. The reversal of the right special factor must thus be left special, so there is exactly one left special factor and exactly one right special factor of each length. A moment's thought shows that this means that $\Gamma(n)$ is composed of cycles sharing a common part, called the *central path*, like in Fig. 1. The number of cycles depends on the number of letters that eventually appear in the directive word. If $\mathbf{t}$ is regular with period d, then there are exactly d cycles. Indeed, each central word u_k has as a suffix all shorter central words, so each central word is followed by each letter 0, 1, . . ., $d - 1$ in $\mathbf{c}_\Delta$. The suffixes of the central words yield right special factors for each length, and the claim follows. Notice that we just argued that the central words are right special. This means that they are left special because they are palindromes. Therefore the central path of $\Gamma(|u_k|)$ reduces to a single vertex. Notice that the graph $\Gamma(n)$ "evolves" to $\Gamma(n+1)$ in a deterministic fashion whenever n does not equal the length of a central word: the central path is shortened by one edge and all cycles maintain their number of edges. When n equals some $|u_k|$, the evolution or "bursting of the bispecial factor" is determined by Δ. In the case of episturmian words, determining $\mathrm{inrc}(\mathbf{t}, n)$ is thus rather straightforward: find out the location of the vertex v corresponding to the prefix of $\mathbf{t}$ of length n and determine the length L of the next cycle taken. If v is on the central path, then $\mathrm{inrc}(\mathbf{t}, n)$ equals L. Otherwise we need to add to L the number of edges that need to be traversed from v to the vertex of the left special factor.

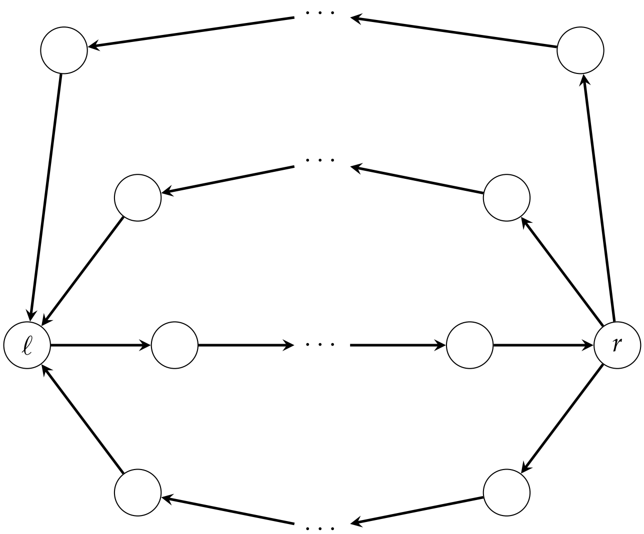

Fig. 1. The Rauzy graph of an episturmian word. The left special factor corresponds to the vertex ℓ and the right special to the vertex r. The directed path from ℓ to r is the central path.

5. Initial nonrepetitive complexity of regular episturmian words

In this section, we derive a complete description of the initial nonrepetitive complexity of regular episturmian words; see Theorem 5.10. We specialize the most significant propositions to the case of Sturmian words. Our proof method generalizes that of Wojcik who determined the initial nonrepetitive complexity of Sturmian words [35, Sect. 5.3].

The majority of the results presented need the assumption that the directive word Δ is regular. We make the convention that this is implicitly assumed in the following discussion, but we make the assumption explicit in the statements of lemmas, propositions, etc.

It is natural to partition the positive integers according to the sequence (q_k), but it is in fact better to do it using the central words u_k. We set

$$\mathcal{I}_k = \{|u_{r_k}| + 1, |u_{r_k}| + 2, \ldots, |u_{r_{k+1}}|\}$$

for $k \geq 0$ with the convention $|u_{r_0}| = -1$. Clearly $\mathbb{N}$ is a disjoint union of these intervals. We further subdivide each interval $\mathcal{I}_k$ into a_{k+1} subintervals by setting

$$\mathcal{I}_{k,\ell} = \{|u_{r_k+\ell}| + 1, \ldots, |u_{r_k+\ell+1}|\}.$$

for $\ell = 0, \ldots, a_{k+1} - 1$. Notice the following peculiarity: the first subinterval $\mathcal{I}_{k,0}$ has q_{k-1} elements while the remaining intervals have q_k elements (when $k > 0$). Indeed by (5), we have $u_{r_k+1} = h_{r_k-1}u_{r_k}$, so $|u_{r_k+1}| - |u_{r_k}| = |h_{r_k-1}| = q_{k-1}$. Similarly $|u_{r_k+t}| - |u_{r_k+t-1}| = q_k$ for t such that $2 \leq t \leq a_{k+1}$. The chief reason for defining the intervals like this is that when the directive word Δ is regular with period d, we are guaranteed that $q_{k+d-1} > |u_{r_{k+1}}|$, that is, q_{k+d-1} exceeds the right endpoint of $\mathcal{I}_k$. If we had defined the right endpoint of $\mathcal{I}_k$ to be $|u_{r_{k+1}+1}|$, which is a priori more natural, this is not true when $d = 2$ as indicated by the proof of the following lemma.

Lemma 5.1. *Let Δ be a regular directive word with period d. We have $q_{k+d-1} - 1 > |u_{r_{k+1}}|$ when $k \geq 0$ and $d \geq 2$. If $d > 2$, then $q_{k+d-1} - 1 > |u_{r_{k+1}+1}|$ for $k \geq 0$.*

Proof. Let us first prove the latter claim. By (7) and (13), we have $u_{r_{k+1}+1} = s_k^{a_*} \cdots s_0^{a_*}$ and $s_{k+d-1} = s_{k+d-2}^{a_*} \cdots s_k^{a_*} s_{k-1}$. When $d > 2$, the word s_{k+d-1} has the factors $s_{k-1+d-1}$ and $s_k^{a_*} s_{k-1}$. We

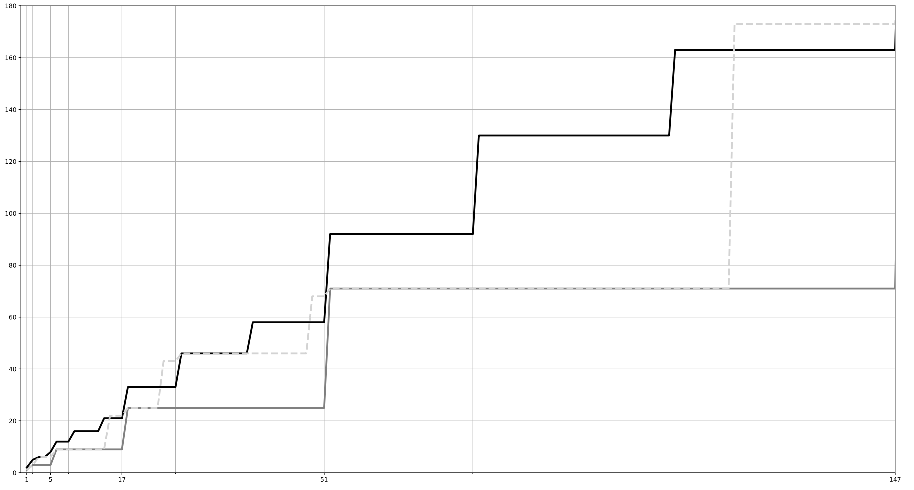

Fig. 2. Plots of the initial nonrepetitive complexity of the episturmian words with directive word $(001122)^\omega$ having intercepts 0^ω (dark gray), $(01)^\omega$ (dashed light gray), and 1^ω (black). The major ticks 1, 5, 17, 51, 147 on the x-axis are the endpoints of the intervals $\mathcal{I}_k$, and the minor ticks are the endpoints of the intervals $\mathcal{I}_{k,0}$.

conclude that if $d > 2$ and $q_{k-1+d-1} > |u_{r_k+1}|$, then $q_{k+d-1} - 1 > |u_{r_{k+1}+1}|$. It then suffices to check that $q_{d-1} > |u_{r_1+1}|$, but this is trivially true as $u_{r_1+1} = s_0^{a_*}$ and $s_{d-1} = s_{d-2}^{a_*} \cdots s_0^{a_*} x_d$ by (12).

Say $d = 2$. Like above, we have $u_{r_{k+1}} = s_k^{a_*-1} s_{k-1}^{a_*} \cdots s_0^{a_*}$ and $s_{k+1} = s_k^{a_*} s_{k-1}$. Consequently $q_{k+1} - 1 > |u_{r_{k+1}}|$ if and only if $|s_k| - 1 > |s_{k-1}^{a_*-1} s_{k-2}^{a_*} \cdots s_0^{a_*}|$. By repeating the argument, we see that the claim is true if and only if $|s_1| - 1 > |s_0^{a_*-1}|$. Since $s_1 = s_0^{a_1} x_2$, the claim follows. □

We set out to figure out $\operatorname{inrc}(\mathbf{t}, n)$ for a regular episturmian word $\mathbf{t}$ when $n \in \mathcal{I}_k$ for $k \geq 0$. In view of Theorem 3.13, the aim is to reduce finding this number to the study of shifts of $\mathbf{c}_\Delta$. In fact, if $\mathbf{t}$ is regular with period d and intercept $c_1 c_2 \cdots$, then in most cases $\operatorname{inrc}(\mathbf{t}, n)$ is determined by the word $T^{\operatorname{val}(c_1 \cdots c_{k+d-1})}(\mathbf{c}_\Delta)$; see Proposition 5.7 for the complete details. See Fig. 2 for example plots of the function $\operatorname{inrc}(\mathbf{t}, n)$.

Let $n \in \mathcal{I}_{k,\ell}$ for k and ℓ such that $k \geq 0$ and $0 \leq \ell < a_{k+1}$, and let θ_n be the length of the central path of the Rauzy graph $\Gamma(n)$ of $\mathbf{c}_\Delta$ (the number of edges on the central path). The number $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n)$ is determined by the cycle sequence taken in the Rauzy graph $\Gamma(n)$ when the word $\mathbf{c}_\Delta$ is read. We denote by C_y the cycle of $\Gamma(|u_{r_k+\ell+1}|)$ containing the edge corresponding to the factor $u_{r_k+\ell+1} y$. We denote the length of C_y by $\|C_y\|$. Notice that the graph $\Gamma(n)$ has the same cycle lengths as the graph $\Gamma(|u_{r_k+1+\ell}|)$.

Let us next show that $\|C_y\| = |\mu_{r_k+\ell}(y)|$. Say we start at the vertex $u_{r_k+\ell+1}$ of $\Gamma(|u_{r_k+\ell+1}|)$, take the cycle C_y, and return to the vertex $u_{r_k+\ell+1}$. This sequence of vertices corresponds to a factor w of length $|u_{r_k+\ell-1}| + \|C_y\|$ such that w contains exactly two occurrences of $u_{r_k+\ell+1}$, one as a prefix and one as a suffix. If follows from [23, Eq. 2] that $u_{r_k+\ell+1} = L_{x_1}(v_1)x_1$ where v_1 is the $(r_k+\ell)$th central word associated with the directive word $T(\Delta)$. Hence the word obtained from w by removing its last letter decodes to a word w_1 such that w_1 has exactly two occurrences of v_1, one as a prefix and one as a suffix. Moreover, we deduce from the form of the morphism L_{x_1} that the prefix v_1 of w_1 is followed by y. This procedure may be repeated $r_k + \ell$ times to obtain $w_{r_k+\ell} = y$. The procedure removes the suffix $u_{r_k+\ell+1}$ completely, so it must be that $|\mu_{r_k+\ell}(y)| + |u_{r_k+\ell+1}| = |w|$, that is, $\|C_y\| = |\mu_{r_k+\ell}(y)|$.

Next we partition the interval $\{0, 1, \ldots, q_{k+d-1} - 1\}$ into intervals λ_i, and we further divide these intervals into subintervals $\lambda_{i,j}$ according to the cycle sequence as described below. Our aim is to show that the initial nonrepetitive complexity has simple description on each $\lambda_{i,j}$. See Proposition 5.3.

Let

$$\Delta' = T^{r_{k+1}}(\Delta) = x_{k+2}^{a_{k+2}} \cdots \quad \text{and} \quad \Delta'' = T^{r_k+\ell}(\Delta) = x_{k+1}^{a_{k+1}-\ell} x_{k+2}^{a_{k+2}} \cdots.$$

Preceding the first occurrence of $x_{k+1}^{a_{k+1}-\ell+1}$, the standard word $\mathbf{c}_{\Delta''}$ is formed of blocks $x_{k+1}^{a_{k+1}-\ell} y$ with $y \neq x_{k+1}$. We say that such a block is of type y. Notice that the block types are given by the letters of the standard word with directive word Δ'. It is straightforward to verify that the number of blocks including the first block of type x_{k+d} equals K_d, where

$$K_d = \prod_{i=2}^{d-1} (a_{k+i} + 1).$$

In particular, $K_2 = 1$.

Corresponding to the ith block, $1 \le i \le K_d$, we define an interval λ_i as follows. If $i = 1$, then we let $L_i = 0$, and otherwise we let $L_i - 1$ to be the largest element of λ_{i-1}. We define

$$\lambda_i = \{L_i, \ldots, L_i + |\tau_{k+1}(y_i)| - 1\}$$

where y_i is the type of the ith block. The number of elements of λ_i is simply the length of the $\mu_{r_k+\ell}$-image of the block since $\mu_{r_k+\ell}(x_{k+1}^{a_{k+1}-\ell} y_i) = \mu_{r_k+\ell} L_{x_{k+1}}^{a_{k+1}-\ell}(y_i) = \tau_{k+1}(y_i)$.

Next we subdivide the interval λ_i into four adjacent intervals that respectively have sizes $(a_{k+1} - \ell - 1)\|C_{x_{k+1}}\| + \theta_n + 1$, $\|C_{x_{k+1}}\| - (\theta_n + 1)$, $\theta_n + 1$, and $\|C_{y_i}\| - (\theta_n + 1)$. More formally, we define

$$\begin{aligned}
\lambda_{i,1} &= \{L_i, \ldots, L_i + (a_{k+1} - \ell - 1)\|C_{x_{k+1}}\| + \theta_n\}, \\
\lambda_{i,2} &= \{L_i + (a_{k+1} - \ell - 1)\|C_{x_{k+1}}\| + \theta_n + 1, \ldots, L_i + (a_{k+1} - \ell)\|C_{x_{k+1}}\| - 1\}, \\
\lambda_{i,3} &= \{L_i + (a_{k+1} - \ell)\|C_{x_{k+1}}\|, \ldots, L_i + (a_{k+1} - \ell)\|C_{x_{k+1}}\| + \theta_n\}, \text{ and} \\
\lambda_{i,4} &= \{L_i + (a_{k+1} - \ell)\|C_{x_{k+1}}\| + \theta_n + 1, \ldots, L_i + (a_{k+1} - \ell)\|C_{x_{k+1}}\| + \|C_{y_i}\| - 1\}.
\end{aligned}$$

Let us find out the size of the union of the intervals λ_i for $i = 1, \ldots, K_d$. The intervals are clearly disjoint and adjacent, so the size equals $|\tau_{k+1}(y)|$ summed over the block types y. Thus the size of the union equals the length of the τ_{k+1}-image of the prefix of $\mathbf{c}_{\Delta'}$ having the first occurrence of x_{k+d} as a suffix. By (12), this prefix equals v_{d-2} where v_{d-2} is the $(d-2)$th standard word for the directive word Δ'. Then the τ_{k+1}-image of v_{d-2} equals s_{k+d-1}. Indeed, by definition, we have $v_{d-2} = L_{x_{k+2}}^{a_{k+2}} \circ \cdots \circ L_{x_{k+d-1}}^{a_{k+d-1}}(x_{k+d})$, so $\tau_{k+1}(v_{d-2}) = \tau_{k+d-1}(x_{k+d}) = s_{k+d-1}$. It follows that

$$\bigcup_{i=1}^{K_d} \lambda_i = \{0, 1, \ldots, q_{k+d-1} - 1\}.$$

Example 5.2 (*Sturmian Case*)**.** When Δ is binary and $d = 2$, we have $K_d = 1$, so there is only one block. Now $\|C_{x_{k+1}}\| = |\tau_k(x_{k+1})| = q_k$. The type y of the block is clearly x_{k+2}, so

$$\|C_y\| = |\mu_{r_k+\ell}(x_{k+2})| = |\mu_{r_k}(x_{k+1}^{\ell} x_{k+2})| = \ell q_k + |\mu_{r_k}(x_k)| = \ell q_k + q_{k-1}.$$

By recalling that $L_1 = 0$, we find that the intervals $\lambda_{1,j}$ are as follows:

$$\begin{aligned}
\lambda_{1,1} &= \{0, \ldots, (a_{k+1} - \ell - 1)q_k + \theta_n\}, \\
\lambda_{1,2} &= \{(a_{k+1} - \ell - 1)q_k + \theta_n + 1, \ldots, (a_{k+1} - \ell)q_k - 1\}, \\
\lambda_{1,3} &= \{(a_{k+1} - \ell)q_k, \ldots, (a_{k+1} - \ell)q_k + \theta_n\}, \text{ and} \\
\lambda_{1,4} &= \{(a_{k+1} - \ell)q_k + \theta_n + 1, \ldots, q_{k+1} - 1\}.
\end{aligned}$$

The union of the intervals equals $\{0, 1, \ldots, q_{k+1} - 1\}$.

Proposition 5.3. *Let Δ be a regular directive word. Let n and i be integers such that $n \in \mathcal{I}_{k,\ell}$ with $k \ge 0$ and $0 \le \ell < a_{k+1}$ and $1 \le i \le K_d$. Suppose that the ith block has type y_i.*

(i) *If* $m \in \lambda_{i,1}$, *then* $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k$.
(ii) *If* $m \in \lambda_{i,2}$, *then* $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = |\tau_{k+1}(y_i)| + L_i - m$.
(iii) *If* $m \in \lambda_{i,3}$, *then* $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = \ell q_k + |\tau_k(y_i)|$.
(iv) *If* $m \in \lambda_{i,4}$, *then* $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k + L_{i+1} - m$.

Proof. We are concerned with the cycles taken in the graph $\Gamma(n)$, which evolves to $\Gamma(u_{r_k+1+\ell})$. Recall that the cycle sequence taken in $\Gamma(u_{r_k+1+\ell})$ is determined by the letters of the standard word $\mathbf{c}_{\Delta''}$ with $\Delta'' = T^{r_k+\ell}(\Delta) = x_{k+1}^{a_{k+1}-\ell} x_{k+2}^{a_{k+2}} \cdots$ and that the cycle lengths are given by the lengths of the $\mu_{r_k+\ell}$-images of these letters. Let v be the prefix of $T^m(\mathbf{c}_\Delta)$ of length n.

Case A. Suppose that $m \in \lambda_{i,1}$. By the discussion preceding Example 5.2, the number L_i equals the length of the $\mu_{r_k+\ell}$-image of the first $i-1$ blocks. Then, from the remark at the beginning of this proof, we see that reading off L_i letters from the beginning of $\mathbf{c}_\Delta$ amounts to traveling complete cycles in $\Gamma(u_{r_k+1+\ell})$. Since all prefixes of $\mathbf{c}_\Delta$ are left special, we see that the prefix of $T^{L_i}(\mathbf{c}_\Delta)$ of length n corresponds to the left special vertex of $\Gamma(n)$. Assume that $L_i \leq m < L_i + (a_{k+1} - \ell - 1)\|C_{x_{k+1}}\|$. Since the cycle C_1 of $\Gamma(n)$ having length $\|C_{x_{k+1}}\|$ is initially taken $a_{k+1} - \ell$ times (as the word $x_{k+1}^{a_{k+1}-\ell}$ is a prefix of $\mathbf{c}''_\Delta$), it follows that v lies on C_1 and C_1 is traversed at least once more after the prefix v of $T^m(\mathbf{c}_\Delta)$. It follows that

$$\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = \|C_1\| = \|C_{x_{k+1}}\| = |\mu_{r_k+\ell}(x_{k+1})| = q_k.$$

Assume then that $L_i + (a_{k+1} - \ell - 1)\|C_{x_{k+1}}\| \leq m \leq L_i + (a_{k+1} - \ell - 1)\|C_{x_{k+1}}\| + \theta_n$. Then v lies on the central path of $\Gamma(n)$ during the $(a_{k+1} - \ell)$th traversal of C_1. Hence $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k$ in this case as well.

Case B. Suppose that $m \in \lambda_{i,2}$. By the arguments in the latter case of the previous paragraph, we see that v lies on the cycle C_1 during the $(a_{k+1} - \ell)$th traversal of C_1. Moreover, the vertex v is not on the central path of $\Gamma(n)$ and exactly $\|C_1\| - (m - (L_i + (a_{k+1} - \ell - 1)\|C_1\|))$ edges need to be traversed to return to the left special vertex of $\Gamma(n)$. The ith block equals $x_{k+1}^{a_{k+1}-\ell} y_i$, so the cycle C_1 is followed by a cycle C_y having length $|\mu_{r_k+\ell}(y_i)|$. Therefore the initial nonrepetitive complexity is determined by the return to the left special vertex of $\Gamma(n)$ when traversing the cycle C_y, that is, we have

$$\begin{aligned}
\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) &= \|C_1\| + \|C_y\| - (m - (L_i + (a_{k+1} - \ell - 1)\|C_1\|)) \\
&= |\mu_{r_k+\ell}(y_i)| - (m - (L_i + (a_{k+1} - \ell)q_k)) \\
&= |\tau_k(x_{k+1}^{\ell} y_i)| - (m - (L_i + (a_{k+1} - \ell)q_k)) \\
&= \ell q_k + |\tau_k(y_i)| - (m - (L_i + (a_{k+1} - \ell)q_k)) \\
&= a_{k+1} q_k + |\tau_k(y_i)| + L_i - m \\
&= |\tau_{k+1}(y_i)| + L_i - m.
\end{aligned}$$

Case C. Suppose that $m \in \lambda_{i,3}$. Now v lies on the central path of $\Gamma(n)$ and the next cycle to be traversed is C_y. Since v is on the central path, the initial nonrepetitive complexity is simply $\|C_y\|$. The claim follows from the preceding computations.

Case D. Suppose that $m \in \lambda_{i,4}$. In this case v lies on C_y but not on the central path. From the form of the word $\mathbf{c}_{\Delta''}$, we deduce that the next cycle taken is C_1. Exactly $L_{i+1} - m$ edges need to be traversed to arrive at the left special factor of $\Gamma(n)$. Therefore

$$\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = L_{i+1} - m + \|C_1\| = q_k + L_{i+1} - m. \quad \square$$

In the Sturmian case, Proposition 5.3 simplifies as follows. This result was obtained in [35, Prop. 5.3.0.1].

Proposition 5.4 (*Sturmian Case*)**.** *Assume that* Δ *is a binary directive word. Let* $n \in \mathcal{I}_{k,\ell}$ *for* $k \geq 0$ *and* $0 \leq \ell < a_{k+1}$.

(i) *If* $0 \leq m \leq (a_{k+1} - \ell - 1)q_k + \theta_n$, *then* $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k$.
(ii) *If* $(a_{k+1} - \ell - 1)q_k + \theta_n < m < (a_{k+1} - \ell)q_k$, *then* $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_{k+1} - m$.

(iii) *If* $(a_{k+1} - \ell)q_k \leq m \leq (a_{k+1} - \ell)q_k + \theta_n$*, then* $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = \ell q_k + q_{k-1}$.
(iv) *If* $(a_{k+1} - \ell)q_k + \theta_n < m < q_{k+1}$*, then* $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_{k+1} + q_k - m$.

Proof. The claim follows by short computations using the information provided in Example 5.2. □

The following proposition gives the nonrepetitive initial complexity of certain shifts of the regular standard episturmian words for the lengths in the interval $\mathcal{I}_k$. The statement is quite complicated. We advise the reader to read the proof and study the implications (18), (19), (20), and (21) rather than spending much time on the statement itself.

Proposition 5.5. *Let* Δ *be a regular directive word and* $\Delta' = T^{r_{k+1}}(\Delta)$ *for* $k \geq 0$*. Suppose that* $n \in \mathcal{I}_k$*, and let* m *be an integer such that* $0 \leq m < q_{k+d-1}$ *and* $\operatorname{rep}_\Delta(m) = c_1 \cdots c_{k+d-1}$ *(possibly with trailing zeros). Let* y_i *be the ith letter of* $\mathbf{c}_{\Delta'}$ *when* $i = \operatorname{val}_{\Delta'}(c_{k+2} \cdots c_{k+d-1}) + 1$.

(i) *If* $0 < c_{k+1} < a_{k+1} - 1$ *or* $c_{k+1} = a_{k+1} - 1 > 0$ *and* $c_k = 0$*, then we have the following implications:*

- $|u_{r_k}| < n \leq |u_{r_{k+1}}| - \operatorname{val}_\Delta(c_1 \cdots c_{k+1})$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k$,
- $|u_{r_{k+1}}| - \operatorname{val}_\Delta(c_1 \cdots c_{k+1}) < n \leq |u_{r_{k+1}}| - c_{k+1}q_k$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = |\tau_{k+1}(y_i)| - \operatorname{val}_\Delta(c_1 \cdots c_{k+1})$,
- $|u_{r_{k+1}}| - c_{k+1}q_k < n \leq |u_{r_{k+1}}| + q_k - \operatorname{val}_\Delta(c_1 \cdots c_{k+1})$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = (a_{k+1} - c_{k+1})q_k + |\tau_k(y_i)|$,
- $|u_{r_{k+1}}| + q_k - \operatorname{val}_\Delta(c_1 \cdots c_{k+1}) < n \leq |u_{r_{k+1}}|$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k + |\tau_{k+1}(y_i)| - \operatorname{val}_\Delta(c_1 \cdots c_{k+1})$.

(ii) *If* $c_{k+1} = a_{k+1}$*, then we have the following implications:*

- $|u_{r_k}| < n \leq |u_{r_{k+1}}| + q_k - \operatorname{val}_\Delta(c_1 \cdots c_k)$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = |\tau_k(y_i)|$,
- $|u_{r_{k+1}}| + q_k - \operatorname{val}_\Delta(c_1 \cdots c_k) < n \leq |u_{r_{k+1}}|$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k + |\tau_k(y_i)| - \operatorname{val}_\Delta(c_1 \cdots c_k)$.

(iii) *If* $c_{k+1} = a_{k+1} - 1 > 0$ *and* $c_k \neq 0$*, then we have the following implications:*

- $|u_{r_k}| < n \leq |u_{r_k+1}|$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k + |\tau_k(y_i)| - \operatorname{val}_\Delta(c_1 \cdots c_k)$,
- $|u_{r_k+1}| < n \leq |u_{r_k+1}| + q_k - \operatorname{val}_\Delta(c_1 \cdots c_k)$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k + |\tau_k(y_i)|$,
- $|u_{r_k+1}| + q_k - \operatorname{val}_\Delta(c_1 \cdots c_k) < n \leq |u_{r_{k+1}}|$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = 2q_k + |\tau_k(y_i)| - \operatorname{val}_\Delta(c_1 \cdots c_k)$.

(iv) *If* $c_{k+1} = 0$ *and* $c_k = 0$*, then we have the following implications:*

- $|u_{r_k}| < n \leq |u_{r_{k+1}}| - \operatorname{val}_\Delta(c_1 \cdots c_{k-1})$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k$,
- $|u_{r_{k+1}}| - \operatorname{val}_\Delta(c_1 \cdots c_{k-1}) < n \leq |u_{r_{k+1}}|$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = |\tau_{k+1}(y_i)| - \operatorname{val}_\Delta(c_1 \cdots c_{k-1})$.

(v) *If* $c_{k+1} = a_{k+1} - 1 = 0$ *and* $c_k \neq 0$*, then we have the following implications:*

- $|u_{r_k}| < n \leq |u_{r_{k+1}}|$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = |\tau_{k+1}(y_i)| - \operatorname{val}_\Delta(c_1 \cdots c_k)$.

(vi) *If* $a_{k+1} - 1 > c_{k+1} = 0$ *and* $c_k \neq 0$*, then we have the following implications:*

- $|u_{r_k}| < n \leq |u_{r_{k+1}}| - \operatorname{val}_\Delta(c_1 \cdots c_k)$
  $\implies \operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k$.

- $|u_{r_{k+1}}| - \mathrm{val}_\Delta(c_1 \cdots c_k) < n \leq |u_{r_{k+1}}|$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = |\tau_{k+1}(y_i)| - \mathrm{val}_\Delta(c_1 \cdots c_k)$.

Proof. Since $0 \leq m < q_{k+d-1}$, we see that $m \in \lambda_i$ for some i. The discussion preceding Example 5.2 tells us that the left endpoint L_i of λ_i equals the τ_{k+1}-image of the prefix of $\mathbf{c}_{\Delta'}$ of length $i-1$. Let $v_1 = \mathrm{rep}_{\Delta'}(i-1)$ and $v_2 = \mathrm{rep}_{\Delta'}(i)$. It follows that $0^{k+1}v_1 \leq_{\mathrm{lex}} \mathrm{rep}_\Delta(m) <_{\mathrm{lex}} 0^{k+1}v_2$ (here $<_{\mathrm{lex}}$ is the lexicographic order on $\mathbb{N}$). Since v_1 and v_2 are the representations of two consecutive integers, we see that $\mathrm{rep}_\Delta(m)$ must end with v_1. In other words, we have showed that $i = \mathrm{val}_{\Delta'}(c_{k+2} \cdots c_{k+d-1}) + 1$. Thus the type of the ith block is y_i where y_i is the ith letter of $\mathbf{c}_{\Delta'}$.

Assume that $n \in \mathcal{I}_{k,l}$ for some ℓ such that $0 \leq \ell < a_{k+1}$, and write

$$n = |u_{r_k+\ell+1}| - \theta_n = |u_{r_k+1}| + \ell q_k - \theta_n. \tag{17}$$

Recall that θ_n is the length of the central path of $\Gamma(n)$ and that the intervals $\mathcal{I}_{k,\ell}$ have size q_k except in the case $\ell = 0$ when the size is q_{k-1}.

Case A. Let us first consider the case where $m \in \lambda_{i,1}$. Now $L_i = \mathrm{val}_\Delta(0^{k+1}c_{k+2} \cdots c_{k+d-1})$, so $m \in \lambda_{i,1}$ if and only if

$$0 \leq \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) \leq (a_{k+1} - \ell - 1)q_k + \theta_n.$$

Substituting θ_n from (17) to the right inequality yields

$$\begin{aligned} n &\leq (a_{k+1} - \ell - 1)q_k + |u_{r_k+1}| + \ell q_k - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) \\ &= (a_{k+1} - 1)q_k + |u_{r_k+1}| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) \\ &= |u_{r_{k+1}}| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) \end{aligned}$$

The left inequality $\mathrm{val}_\Delta(c_1 \cdots c_{k+1}) \geq 0$ is trivially true, so we conclude using Proposition 5.3 that

$$|u_{r_k}| < n \leq |u_{r_{k+1}}| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) \implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k. \tag{18}$$

Now $|u_{r_{k+1}}| = |u_{r_k}| + q_{k-1} + (a_{k+1} - 1)q_k$, so the antecedent is false only if $\mathrm{val}_\Delta(c_1 \cdots c_{k+1}) \geq (a_{k+1} - 1)q_k + q_{k-1}$. Using Lemma 3.11, we see that this happens exactly when $c_{k+1} = a_{k+1}$ or $c_{k+1} = a_{k+1} - 1$ and $c_k \neq 0$.

Case B. Assume that $m \in \lambda_{i,2}$. Like above, this holds if and only if

$$(a_{k+1} - \ell - 1)q_k + \theta_n < \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) < (a_{k+1} - \ell)q_k.$$

The only option is that $c_{k+1} = a_{k+1} - \ell - 1$, which cannot happen if $a_{k+1} = c_{k+1}$. Substituting θ_n from (17) to the left inequality implies that

$$n > (a_{k+1} - 1)q_k + |u_{r_k+1}| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) = |u_{r_{k+1}}| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}).$$

The right inequality $\mathrm{val}_\Delta(c_1 \cdots c_{k+1}) < (a_{k+1} - \ell)q_k$ is trivially true. Since $|u_{r_k+1}| + \ell q_k = |u_{r_k+1}| + (a_{k+1} - c_{k+1} - 1)q_k = |u_{r_{k+1}}| - c_{k+1}q_k$, we conclude from Proposition 5.3 that

$$\begin{aligned} &|u_{r_{k+1}}| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) < n \leq |u_{r_{k+1}}| - c_{k+1}q_k \\ &\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = |\tau_{k+1}(y_i)| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}). \end{aligned} \tag{19}$$

Notice indeed that here $L_i - m = \mathrm{val}_\Delta(0^{k+1}c_{k+2} \cdots c_{k+d-1}) - m = -\mathrm{val}_\Delta(c_1 \cdots c_{k+1})$.

Case C. Suppose that $m \in \lambda_{i,3}$. This holds if and only if

$$(a_{k+1} - \ell)q_k \leq \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) \leq (a_{k+1} - \ell)q_k + \theta_n.$$

Since $\theta_n < q_k$, it must be that $c_{k+1} = a_{k+1} - \ell$, and this cannot happen if $c_{k+1} = 0$ as $0 \leq \ell < a_{k+1}$. The left inequality is trivial. Utilizing again (17), the right inequality transforms into

$$n \leq a_{k+1}q_k + |u_{r_k+1}| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) = |u_{r_{k+1}}| + q_k - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}).$$

As $|u_{r_k+\ell}| = |u_{r_k+1}| + (\ell - 1)q_k = |u_{r_k+1}| + (a_{k+1} - c_{k+1} - 1)q_k = |u_{r_{k+1}}| - c_{k+1}q_k$, we have the following:

$$|u_{r_k+1}| - c_{k+1}q_k < n \leq |u_{r_{k+1}}| + q_k - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) \implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = (a_{k+1} - c_{k+1})q_k + |\tau_k(y_i)|. \tag{20}$$

Case D. Assume finally that $m \in \lambda_{i,4}$. This is true only if

$$(a_{k+1} - \ell)q_k + \theta_n < \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) < (a_{k+1} - \ell)q_k + |\tau_{k+1}(y_i)|.$$

From the left inequality, we obtain the following inequality:

$$n > a_{k+1}q_k + |u_{r_k+1}| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) = |u_{r_{k+1}}| + q_k - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}).$$

Since $n \in \mathcal{I}_{k,\ell}$, we see that this is possible only when $q_k < \mathrm{val}_\Delta(c_1 \cdots c_{k+1})$, that is, when $c_{k+1} \neq 0$. The right inequality holds trivially, so we have

$$|u_{r_{k+1}}| + q_k - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) < n \leq |u_{r_{k+1}}| \implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k + |\tau_{k+1}(y_i)| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}). \tag{21}$$

Here the facts $L_{i+1} = L_i + |\tau_{k+1}(y_i)|$ and $L_i - m = -\mathrm{val}_\Delta(c_1 \cdots c_{k+1})$ were used.

Let us then put the above results together. If c_{k+1} satisfies $0 < c_{k+1} < a_{k+1} - 1$, then the antecedents of (18), (19), (20), and (21) are all satisfied. Clearly the interval $\{|u_{r_k}| + 1, \ldots, |u_{r_{k+1}}|\}$ is partitioned by these four cases and the initial nonrepetitive complexity is determined on each partition. Exactly the same happens if $c_{k+1} = a_{k+1} - 1$ and $c_k = 0$. This gives (i). Suppose then that $c_{k+1} = a_{k+1}$. Then, as we saw above, the antecedents of (18) and (19) are not satisfied, so these cases are omitted. The left inequality is trivial in the Case C and $n \in \mathcal{I}_{k,0}$, so we may now deduce that $\mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = |\tau_k(y_i)|$ provided that $|u_{r_k}| < n \leq |u_{r_{k+1}}| + q_k - \mathrm{val}_\Delta(c_1 \cdots c_{k+1})$. The Case D directly applies, and we have (ii). The remaining cases are similar. □

The Sturmian case was determined in [35, Cor. 5.3.0.2].

Proposition 5.6 (*Sturmian Case*)**.** *Let Δ be a binary directive word. Suppose that $n \in \mathcal{I}_k$ for $k \geq 0$, and let m be an integer such that $0 \leq m < q_{k+1}$ and $\mathrm{rep}(m) = c_1 \cdots c_{k+1}$ (possibly with trailing zeros).*

(i) If $0 < c_{k+1} < a_{k+1} - 1$ or $c_{k+1} = a_{k+1} - 1 > 0$ and $c_k = 0$, then we have the following implications:

- $q_k - 2 < n \leq q_{k+1} - 2 - \mathrm{val}(c_1 \cdots c_{k+1})$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k$,
- $q_{k+1} - 2 - \mathrm{val}(c_1 \cdots c_{k+1}) < n \leq q_{k+1} - 2 - c_{k+1}q_k$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_{k+1} - \mathrm{val}(c_1 \cdots c_{k+1})$,
- $q_{k+1} - 2 - c_{k+1}q_k < n \leq q_{k+1} + q_k - 2 - \mathrm{val}(c_1 \cdots c_{k+1})$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_{k+1} - c_{k+1}q_k$,
- $q_{k+1} + q_k - 2 - \mathrm{val}(c_1 \cdots c_{k+1}) < n \leq q_{k+1} - 2$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_{k+1} + q_k - \mathrm{val}(c_1 \cdots c_{k+1})$.

(ii) If $c_{k+1} = a_{k+1}$, then we have the following implications:

- $q_k - 2 < n \leq q_{k+1} - 2 - \mathrm{val}(c_1 \cdots c_k)$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_{k-1}$,
- $q_{k+1} - 2 - \mathrm{val}(c_1 \cdots c_k) < n \leq q_{k+1} - 2$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k + q_{k-1} - \mathrm{val}(c_1 \cdots c_k)$.

(iii) If $c_{k+1} = a_{k+1} - 1 > 0$ and $c_k \neq 0$, then we have the following implications:

- $q_k - 2 < n \leq q_k + q_{k-1} - 2$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k + q_{k-1} - \mathrm{val}(c_1 \cdots c_k)$,
- $q_k + q_{k-1} - 2 < n \leq 2q_k + q_{k-1} - 2 - \mathrm{val}(c_1 \cdots c_k)$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k + q_{k-1}$,

- $2q_k + q_{k-1} - 2 - \mathrm{val}(c_1 \cdots c_k) < n \le q_{k+1} - 2$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = 2q_k + q_{k-1} - \mathrm{val}(c_1 \cdots c_k)$.

(iv) If $c_{k+1} = 0$ and $c_k = 0$, then we have the following implications:

- $q_k - 2 < n \le q_{k+1} - 2 - \mathrm{val}(c_1 \cdots c_{k-1})$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k$,
- $q_{k+1} - 2 - \mathrm{val}(c_1 \cdots c_{k-1}) < n \le q_{k+1} - 2$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_{k+1} - \mathrm{val}(c_1 \cdots c_{k-1})$.

(v) If $c_{k+1} = a_{k+1} - 1 = 0$ and $c_k \neq 0$, then we have the following implications:

- $q_k - 2 < n \le q_{k+1} - 2$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_{k+1} - \mathrm{val}(c_1 \cdots c_k)$.

(vi) If $a_{k+1} - 1 > c_{k+1} = 0$ and $c_k \neq 0$, then we have the following implications:

- $q_k - 2 < n \le q_{k+1} - 2 - \mathrm{val}(c_1 \cdots c_k)$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k$.
- $q_{k+1} - 2 - \mathrm{val}(c_1 \cdots c_k) < n \le q_{k+1} - 2$
  $\implies \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) = q_{k+1} - \mathrm{val}(c_1 \cdots c_k)$.

Proof. An easy induction argument shows that $|u_{r_k}| = q_k - 2$ for all $k \ge 1$ in the Sturmian case. Utilize again the computations of Example 5.2. □

Keeping in mind Theorem 3.13, we now determine which shifts of $\mathbf{c}_\Delta$ need to be considered in order to determine the initial nonrepetitive complexity. Notice that if precise information is not required, the first case of the following proposition can be omitted since the longest common prefix of $\mathbf{t}$ and $T^{\mathrm{val}(c_1 \cdots c_{k+d})}(\mathbf{c}_\Delta)$ is at least as long as that of $\mathbf{t}$ and $T^{\mathrm{val}(c_1 \cdots c_{k+d-1})}(\mathbf{c}_\Delta)$ according to Lemma 3.21. Notice also that when the case (i) of the proposition applies, we can compute the initial nonrepetitive complexity using Proposition 5.5. This result was proved for Sturmian words in [35, Prop. 5.3.0.3].

Proposition 5.7. *Let $\mathbf{t}$ be a regular episturmian word with directive word as in (2) and intercept $c_1 c_2 \cdots$. Let $n \in \mathcal{I}_k$ for $k \ge 0$.*

(i) If $c_i < a_i$ for some i such that $k+2 \le i \le k+d$, then $\mathrm{inrc}(\mathbf{t}, n) = \mathrm{inrc}(T^{\mathrm{val}(c_1 \cdots c_{k+d-1})}(\mathbf{c}_\Delta), n)$.
(ii) If $c_i = a_i$ for all i such that $k+2 \le i \le k+d$, then $\mathrm{inrc}(\mathbf{t}, n) = \mathrm{inrc}(T^{\mathrm{val}(c_1 \cdots c_{k+d})}(\mathbf{c}_\Delta), n)$.

For the proof, we need some auxiliary lemmas. They include more information than we need, but the complete statements could be useful in some other contexts.

Lemma 5.8. *Let $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ be a regular directive word with period d. Let k and i be such that $1 \le k \le d$ and $0 \le i < k$. Then $\tau_k(x_{k-i}) = s_{k-1}^{a_*} \cdots s_{k-i}^{a_*} s_{k-(i+1)}$.*

Proof. Let $k = 1$ and $i = 0$. Now $\tau_k(x_{k-i}) = \tau_1(x_1) = x_1 = s_0$, so the base case is established. Let then $k > 1$ and i be such that $1 \le i < k$. If $i > 0$, then $x_k \neq x_{k-i}$ because Δ is regular, so $\tau_k(x_{k-i}) = s_{k-1}^{a_*} \tau_{k-1}(x_{k-i})$, and the claim follows from the induction hypothesis. If $i = 0$, then $\tau_{k-1}(x_{k-i}) = s_{k-1}$, and the claim follows. □

Lemma 5.9. *Let $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ be a regular directive word with period d. For k and ℓ such that $0 \le k < d$ and $1 \le \ell \le a_{k+1}$, we have*

(i) $\tau_k L_{x_{k+1}}^{\ell}(x_{k+1}) = s_k$,
(ii) $\tau_k L_{x_{k+1}}^{\ell}(x_i) = s_k^{\ell} s_{k-1}^{a_*} \cdots s_i^{a_*} s_{i-1}$ when $1 \le i \le k$, and
(iii) $\tau_k L_{x_{k+1}}^{\ell}(x_i) = s_k^{\ell} s_{k-1}^{a_*} \cdots s_0^{a_*} x_i$ when $i > k+1$.

Let k, ℓ, and i be such that $k \ge d$, $1 \le \ell \le a_{k+1}$, and $1 \le i \le d$. Then *(iv)* $\tau_k L^{\ell}_{x_{k+1}}(x_i) = s_k$ *if* $i \equiv k+1 \pmod{d}$ *and*

(v) $\tau_k L^{\ell}_{x_{k+1}}(x_i) = s_k^{\ell} s_{k-1}^{a_*} \cdots s_{k-j}^{a_*} s_{k-j-1}$ *when* $i \not\equiv k+1 \pmod{d}$ *and* j *is the smallest integer such that* $k - j \equiv i \pmod{d}$.

Proof. Assume that k and ℓ are such that $0 \le k < d$ and $1 \le \ell \le a_{k+1}$. Notice that since Δ is regular, we have $x_{k+1} \neq x_i$ for $i = 1, \ldots, k$ and $i = k+2, \ldots, k+d$. Clearly $\tau_k L^{\ell}_{x_{k+1}}(x_{k+1}) = \tau_k(x_{k+1}) = s_k$. It is straightforward to show that $\tau_k(x_i) = s_{k-1}^{a_*} \cdots s_0^{a_*} x_i$ for i such that $k+1 < i < d$, so we obtain $\tau_k L^{\ell}_{x_{k+1}}(x_i) = \tau_k(x^{\ell}_{k+1} x_i) = s_k^{\ell} \tau_k(x_i) = s_k^{\ell} s_{k-1}^{a_*} \cdots s_0^{a_*} x_i$ for these i. Let then i be such that $0 \le i < k$. Since $x_{k-i} \neq x_{k+1}$, we infer from Lemma 5.8 that

$$\tau_k L^{\ell}_{x_{k+1}}(x_{k-i}) = s_k^{\ell} \tau_k(x_{k-i}) = s_k^{\ell} s_{k-1}^{a_*} \cdots s_{k-i}^{a_*} s_{k-(i+1)}.$$

This proves the first part of the claim.

Suppose then that $k \ge d$, $1 \le \ell \le a_{k+1}$, and $1 \le i \le d$. Again we have $\tau_k L^{\ell}_{x_{k+1}}(x_{k+1}) = s_k$ which gives (iv). Assume that $i \not\equiv k+1 \pmod{d}$. Let j be the smallest number such that $k - j \equiv i \pmod{d}$. Applying computations as above, we obtain that

$$\tau_k L^{\ell}_{x_{k+1}}(x_i) = s_k^{\ell} s_{k-1}^{a_*} \tau_{k-1}(x_i) = \cdots = s_k^{\ell} s_{k-1}^{a_*} \cdots s_{k-j}^{a_*} \tau_{k-j}(x_i) = s_k^{\ell} s_{k-1}^{a_*} \cdots s_{k-j}^{a_*} s_{k-j-1}.$$

This proves (v). □

The point of Lemma 5.9 is that the words $\tau_k L^{\ell}_{x_{k+1}}(x_i)$ are ordered by length in a predictable pattern. Let us consider the words $\tau_k L^{\ell}_{x_{k+1}}(x_i)$ ordered by the index i in the natural order of $\{1, 2, \ldots, d\}$. Then the first part of Lemma 5.9 states that the lengths $|\tau_k L^{\ell}_{x_{k+1}}(x_i)|$ strictly decrease when i increases from 1 to $k+1$. The remaining words are of equal length that is strictly greater than the preceding values. In other words, the shortest word $\tau_k L^{\ell}_{x_{k+1}}(x_{k+1})$ is pushed to the right in a cyclical fashion. The second part of the lemma states that this cyclical pattern continues modulo d.

Proof of Proposition 5.7. Suppose that $n \in \mathcal{I}_{k,\ell}$ for some ℓ such that $0 \le \ell < a_{k+1}$. Let $m = \mathrm{val}(c_1 \cdots c_{k+d-1})$. Now m belongs to the ith block where $i = \mathrm{val}_{\Delta'}(c_{k+2} \cdots c_{k+d-1}) + 1$ and $\Delta' = T^{r_{k+1}}(\Delta)$ (see the first paragraph of the proof of Proposition 5.5). Say the block has type y_i. Write $n = |u_{r_k+1}| + \ell q_k - \theta_n$.

We assume first that there exists a largest i such that $k+2 \le i \le k+d$ and $c_i < a_i$. In order to show that $\mathrm{inrc}(\mathbf{t}, n) = \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n)$, it suffices to demonstrate that the longest common prefix of these words has length at least $\mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) + n$. Below we do this depending on which interval $\lambda_{i,j}$ the number m belongs to. If $\mathbf{t} = T^m(\mathbf{c}_\Delta)$, there is nothing to prove, so we assume that there exists a least positive integer j such that $c_{k+d-1+j} \neq 0$. By Lemma 3.20, the longest common prefix of $T^m(\mathbf{c}_\Delta)$ and $T^{\mathrm{val}(c_1 \cdots c_{k+d-1+j})}(\mathbf{c}_\Delta)$ has length P where

$$P = |s_{k+d-1+j-1}^{a_* - c_*} s_{k+d-1+j-2}^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1 \cdots c_{k+d-1}).$$

It follows from Lemma 3.21 that $\mathbf{t}$ and $T^{\mathrm{val}(c_1 \cdots c_{k+d-1+j})}(\mathbf{c}_\Delta)$ also have a common prefix of length P. It thus suffices to show that $P \ge \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) + n$ to establish (i).

Case A. Assume that $m \in \lambda_{i,1}$. As in the proof of Proposition 5.5, we see that this means that $0 \le \mathrm{val}(c_1 \cdots c_{k+1}) \le (a_{k+1} - \ell - 1)q_k + \theta_n$, so $-\theta_n \le (a_{k+1} - \ell - 1)q_k - \mathrm{val}(c_1 \cdots c_{k+1})$. Hence we obtain from Proposition 5.3 that

$$\begin{aligned} \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) + n &= q_k + n \\ &= q_k + |u_{r_{k+1}}| + \ell q_k - \theta_n \\ &\le |u_{r_k+1}| + a_{k+1} q_k - \mathrm{val}(c_1 \cdots c_{k+1}) \\ &= |u_{r_{k+1}+1}| - \mathrm{val}(c_1 \cdots c_{k+1}). \end{aligned}$$

By applying Lemma 3.21 and the preceding inequality, we obtain

$$\begin{aligned} P &\geq |s_{k+1}^{a_*-c_*} s_k^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1 \cdots c_{k+1}) \\ &\geq |s_k^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1 \cdots c_{k+1}) \\ &= |u_{r_{k+1}+1}| - \mathrm{val}(c_1 \cdots c_{k+1}) \\ &\geq \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) + n, \end{aligned}$$

so we conclude that $\mathrm{inrc}(\mathbf{t}, n) = \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n)$.

Case B. Assume that $m \in \lambda_{i,2}$. As in the proof of Proposition 5.5, we have

$$(a_{k+1} - \ell - 1)q_k + \theta_n < \mathrm{val}(c_1 \cdots c_{k+1}) < (a_{k+1} - \ell)q_k$$

and $c_{k+1} = a_{k+1} - \ell - 1$. Thus Proposition 5.3 gives

$$\begin{aligned} \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) + n &= |\tau_{k+1}(y_i)| - \mathrm{val}(c_1 \cdots c_{k+1}) + |u_{r_k+1}| + \ell q_k + \theta_n \\ &= |\tau_{k+1}(y_i)| - \mathrm{val}(c_1 \cdots c_{k+1}) + |u_{r_k+1}| + (a_{k+1} - c_{k+1} - 1)q_k + \theta_n \\ &= |\tau_{k+1}(y_i)| - \mathrm{val}(c_1 \cdots c_{k+1}) + |u_{r_{k+1}+1}| - (c_{k+1} + 1)q_k + \theta_n. \end{aligned}$$

Recall that there exists a largest i such that $k + 2 \leq i \leq k + d$ and $c_i < a_i$. An application of Lemma 3.21 yields

$$\begin{aligned} P &\geq |s_{i-1}^{a_*-c_*} s_{i-2}^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1 \cdots c_{i-1}) \\ &\geq q_{i-1} + |s_{i-2}^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1 \cdots c_{i-1}) \\ &\geq q_{k+1} + |s_k^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1 \cdots c_{k+1}) \\ &= q_{k+1} + |u_{r_{k+1}+1}| - \mathrm{val}(c_1 \cdots c_{k+1}). \end{aligned}$$

It thus suffices to prove that $|\tau_{k+1}(y_i)| \leq q_{k+1}$ in order to conclude that $\mathrm{inrc}(\mathbf{t}, n) = \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n)$. Since $y_i \neq x_{k+1}$, the j in (v) of Lemma 5.9 is at most $d - 2$ which implies that $\tau_{k+1}(y_i)$ is a prefix of $s_k^{a_*} \cdots s_{k-(d-2)}^{a_*} s_{k-(d-1)}$. Since $s_k^{a_*} \cdots s_{k+1-d+1}^{a_*} s_{k+1-d} = s_{k+1}$ by (12) and (13), it follows that $|\tau_{k+1}(y_i)| \leq q_{k+1}$.

Case C. Suppose that $m \in \lambda_{i,3}$. Then $(a_{k+1} - \ell)q_k \leq \mathrm{val}(c_1 \cdots c_{k+1}) \leq (a_{k+1} - \ell)q_k + \theta_n$, $c_{k+1} = a_{k+1} - \ell$, and $c_{k+1} \neq 0$. Proposition 5.3 implies that

$$\begin{aligned} \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) + n &= |\tau_k(y_i)| + \ell q_k + |u_{r_k+1}| + \ell q_k - \theta_n \\ &= |\tau_{k+1}(y_i)| - (a_{k+1} - \ell)q_k + |u_{r_k+1}| + \ell q_k - \theta_n \\ &= |\tau_{k+1}(y_i)| - a_{k+1}q_k + 2\ell q_k + |u_{r_k+1}| - \theta_n \\ &= |\tau_{k+1}(y_i)| - a_{k+1}q_k + 2(a_{k+1} - c_{k+1})q_k + |u_{r_k+1}| - \theta_n \\ &= |\tau_{k+1}(y_i)| + |u_{r_{k+1}+1}| - 2c_{k+1}q_k - \theta_n. \end{aligned}$$

Like in the previous case, we have $P \geq q_{k+1} + |u_{r_{k+1}+1}| - \mathrm{val}(c_1 \cdots c_{k+1})$, so we just need to show that $\mathrm{val}(c_1 \cdots c_{k+1}) \leq 2c_{k+1}q_k + \theta_n$. This is equivalent with $\mathrm{val}(c_1 \cdots c_k) \leq c_{k+1}q_k + \theta_n$ and this in turn is true when $c_{k+1} \neq 0$ because $\mathrm{val}(c_1 \cdots c_k) < q_k$ by Lemma 3.11. We conclude that $\mathrm{inrc}(\mathbf{t}, n) = \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n)$.

Case D. Suppose that $m \in \lambda_{i,4}$. Now

$$(a_{k+1} - \ell)q_k + \theta_n < \mathrm{val}(c_1 \cdots c_{k+1}) < (a_{k+1} - \ell)q_k + |\tau_{k+1}(y_i)|$$

and $c_{k+1} \neq 0$. From Proposition 5.3, we obtain that

$$\begin{aligned} \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n) + n &= q_k + |\tau_{k+1}(y_i)| - \mathrm{val}(c_1 \cdots c_{k+1}) + |u_{r_k+1}| + \ell q_k - \theta_n \\ &\leq q_k + q_{k+1} + |u_{r_{k+1}+1}| - \mathrm{val}(c_1 \cdots c_{k+1}) - a_{k+1}q_k + \ell q_k - \theta_n \\ &\leq P + q_k - a_{k+1}q_k + \ell q_k - \theta_n \\ &\leq P - (a_{k+1} - \ell - 1)q_k \\ &\leq P \end{aligned}$$

because $P \geq q_{k+1} + |u_{r_{k+1}+1}| - \mathrm{val}(c_1 \cdots c_{k+1})$, $\theta_n \geq 0$, and $\ell < a_{k+1}$. Therefore $\mathrm{inrc}(\mathbf{t}, n) = \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n)$ also in this case.

We have now finished the first part of the proof. Let $m' = \mathrm{val}(c_1 \cdots c_{k+d})$, and assume that $c_i = a_i$ for all i such that $k+2 \leq i \leq k+d$. Since the intercept $c_1c_2\cdots$ satisfies the Ostrowski conditions, it follows that $c_{k+1} = 0$ and $c_{k+d+1} < a_{k+d+1}$. Thus we immediately see that the preceding Cases C and D do not occur as then we had $c_{k+1} \neq 0$. The arguments given in the Case A still work, and we see that $\mathrm{inrc}(\mathbf{t}, n) = \mathrm{inrc}(T^m(\mathbf{c}_\Delta), n)$ when $m \in \lambda_{i,1}$. Since the longest common prefix of $\mathbf{t}$ and $T^{m'}(\mathbf{c}_\Delta)$ is at least as long as that of $\mathbf{t}$ and $T^m(\mathbf{c}_\Delta)$ by Lemma 3.21, we deduce that $\mathrm{inrc}(\mathbf{t}, n) = \mathrm{inrc}(T^{m'}(\mathbf{c}_\Delta), n)$ if $m \in \lambda_{i,1}$. We are thus left with the case $m \in \lambda_{i,2}$. Earlier in Case B we deduced that $\ell = a_{k+1} - c_{k+1} - 1$, that is, $\ell = a_{k+1} - 1$. This means that $n = |u_{r_{k+1}}| - \theta_n$.

If $T^{m'}(\mathbf{c}_\Delta) = \mathbf{t}$, then the claim is clear, so we assume that there exists a least positive integer j' such that $c_{k+d+j'} \neq 0$ (notice that since $c_{k+d} = a_{k+d}$, we have $j = 1$). By Lemma 3.20, the longest common prefix of $\mathbf{t}$ and $T^{m'}(\mathbf{c}_\Delta)$ has length P' where

$$
\begin{aligned}
P' &= |s_{k+d+j'-1}^{a_*-c_*} s_{k+d+j'-2}^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1 \cdots c_{k+d}) \\
&\geq |s_{k+d}^{a_*-c_*} s_{k+d-1}^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1 \cdots c_{k+d}) \\
&\geq q_{k+d} + |s_{k+d-1}^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1 \cdots c_{k+d}) \\
&= q_{k+d} + |s_k^{a_*} \cdots s_0^{a_*}| - \mathrm{val}(c_1 \cdots c_{k+1}) \\
&= q_{k+d} + |u_{r_{k+1}+1}| - \mathrm{val}(c_1 \cdots c_k).
\end{aligned}
$$

We used above the facts that $c_{k+d+1} < a_{k+d+1}$, $c_{k+1} = 0$, and $c_i = a_i$ for all i such that $k+2 \leq i \leq k+d$. Apply Proposition 5.5 (ii) to obtain that

$$\mathrm{inrc}(T^{m'}(\mathbf{c}_\Delta), n') = |\tau_{k+1}(y)|$$

for a letter y when $|u_{r_{k+1}}| < n' \leq |u_{r_{k+1}}| + q_k - \mathrm{val}(c_1 \cdots c_k)$. Since the initial nonrepetitive complexity is nondecreasing and $|\tau_{k+1}(y)| \leq q_{k+1}$ (see Case B), we find that

$$\mathrm{inrc}(T^{m'}(\mathbf{c}_\Delta), n) + n \leq q_{k+1} + n = q_{k+1} + |u_{r_{k+1}}| - \theta_n.$$

Therefore

$$
\begin{aligned}
P' &\geq q_{k+d} + |u_{r_{k+1}+1}| - \mathrm{val}(c_1 \cdots c_k) \\
&= q_{k+d} + |u_{r_{k+1}}| + q_k - \mathrm{val}(c_1 \cdots c_k) \\
&> q_{k+1} + |u_{r_{k+1}}| \\
&\geq \mathrm{inrc}(T^{m'}(\mathbf{c}_\Delta), n) + n.
\end{aligned}
$$

It follows that $\mathrm{inrc}(\mathbf{t}, n) = \mathrm{inrc}(T^{m'}(\mathbf{c}_\Delta), n)$. This establishes (ii) and ends the proof. □

We are finally in the position to give a complete description of the initial nonrepetitive complexity of a regular episturmian word. This mostly amounts to putting together Propositions 5.5 and 5.7. The statement is again complicated, but all listed cases are different. The initial nonrepetitive complexity of a general epistandard word is determined in [27, Thm. 16].

Theorem 5.10. *Let $\mathbf{t}$ be a regular episturmian word with a directive word Δ of period d and intercept $c_1c_2\cdots$. Suppose that $n \in \mathcal{I}_k$ for $k \geq 0$, and let $\Delta' = T^{r_{k+1}}(\Delta)$ and y_i be the ith letter of $\mathbf{c}_{\Delta'}$ when $i = \mathrm{val}_{\Delta'}(c_{k+2} \cdots c_{k+d-1}) + 1$.*

(i) If $0 < c_{k+1} < a_{k+1} - 1$ or $c_{k+1} = a_{k+1} - 1 > 0$ and $c_k = 0$, then we have the following implications:

- $|u_{r_k}| < n \leq |u_{r_{k+1}}| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1})$
  $\Longrightarrow \mathrm{inrc}(\mathbf{t}, n) = q_k$,
- $|u_{r_{k+1}}| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1}) < n \leq |u_{r_{k+1}}| - c_{k+1}q_k$
  $\Longrightarrow \mathrm{inrc}(\mathbf{t}, n) = |\tau_{k+1}(y_i)| - \mathrm{val}_\Delta(c_1 \cdots c_{k+1})$,

- $|u_{r_{k+1}}| - c_{k+1}q_k < n \leq |u_{r_{k+1}}| + q_k - \text{val}_\Delta(c_1 \cdots c_{k+1})$
$\implies \text{inrc}(\mathbf{t}, n) = (a_{k+1} - c_{k+1})q_k + |\tau_k(y_i)|$,
- $|u_{r_{k+1}}| + q_k - \text{val}_\Delta(c_1 \cdots c_{k+1}) < n \leq |u_{r_{k+1}}|$
$\implies \text{inrc}(\mathbf{t}, n) = q_k + |\tau_{k+1}(y_i)| - \text{val}_\Delta(c_1 \cdots c_{k+1})$.

(ii) If $c_{k+1} = a_{k+1}$, then we have the following implications:

- $|u_{r_k}| < n \leq |u_{r_k+1}| - \text{val}_\Delta(c_1 \cdots c_k)$
$\implies \text{inrc}(\mathbf{t}, n) = |\tau_k(y_i)|$,
- $|u_{r_k+1}| - \text{val}_\Delta(c_1 \cdots c_k) < n \leq |u_{r_{k+1}}|$
$\implies \text{inrc}(\mathbf{t}, n) = q_k + |\tau_k(y_i)| - \text{val}_\Delta(c_1 \cdots c_k)$.

(iii) If $c_{k+1} = a_{k+1} - 1 > 0$ and $c_k \neq 0$, then we have the following implications:

- $|u_{r_k}| < n \leq |u_{r_k+1}|$
$\implies \text{inrc}(\mathbf{t}, n) = q_k + |\tau_k(y_i)| - \text{val}_\Delta(c_1 \cdots c_k)$,
- $|u_{r_k+1}| < n \leq |u_{r_k+1}| + q_k - \text{val}_\Delta(c_1 \cdots c_k)$
$\implies \text{inrc}(\mathbf{t}, n) = q_k + |\tau_k(y_i)|$,
- $|u_{r_k+1}| + q_k - \text{val}_\Delta(c_1 \cdots c_k) < n \leq |u_{r_{k+1}}|$
$\implies \text{inrc}(\mathbf{t}, n) = 2q_k + |\tau_k(y_i)| - \text{val}_\Delta(c_1 \cdots c_k)$.

(iv) Suppose that $c_{k+1} = 0$ and there exists i such that $k + 2 \leq i \leq k + d$ and $c_i < a_i$.

(iv.a) If $c_k = 0$, then we have the following implications:

- $|u_{r_k}| < n \leq |u_{r_{k+1}}| - \text{val}_\Delta(c_1 \cdots c_{k-1})$
$\implies \text{inrc}(\mathbf{t}, n) = q_k$,
- $|u_{r_{k+1}}| - \text{val}_\Delta(c_1 \cdots c_{k-1}) < n \leq |u_{r_{k+1}}|$
$\implies \text{inrc}(\mathbf{t}, n) = |\tau_{k+1}(y_i)| - \text{val}_\Delta(c_1 \cdots c_{k-1})$.

(iv.b) If $a_{k+1} = 1$ and $c_k \neq 0$, then we have the following implication:

- $|u_{r_k}| < n \leq |u_{r_{k+1}}|$
$\implies \text{inrc}(\mathbf{t}, n) = |\tau_{k+1}(y_i)| - \text{val}_\Delta(c_1 \cdots c_k)$.

(iv.c) If $a_{k+1} > 1$ and $c_k \neq 0$, then we have the following implications:

- $|u_{r_k}| < n \leq |u_{r_{k+1}}| - \text{val}_\Delta(c_1 \cdots c_k)$
$\implies \text{inrc}(\mathbf{t}, n) = q_k$.
- $|u_{r_{k+1}}| - \text{val}_\Delta(c_1 \cdots c_k) < n \leq |u_{r_{k+1}}|$
$\implies \text{inrc}(\mathbf{t}, n) = |\tau_{k+1}(y_i)| - \text{val}_\Delta(c_1 \cdots c_k)$.

(v) If $c_{k+1} = 0$ and $a_i = c_i$ for all i such that $k + 2 \leq i \leq k + d$, then we have the following implication:

- $|u_{r_k}| < n \leq |u_{r_{k+1}}|$
$\implies \text{inrc}(\mathbf{t}, n) = q_k$.

Proof. Let $m = \text{val}(c_1 \cdots c_{k+d-1})$ and $m' = \text{val}(c_1 \cdots c_{k+d})$.

Suppose that $0 < c_{k+1} < a_{k+1} - 1$ or $c_{k+1} = a_{k+1} - 1 > 0$ and $c_k = 0$. Then the Ostrowski conditions guarantee that there exists i such that $k + 2 \leq i \leq k + d$ and $c_i < a_i$. Therefore Proposition 5.7 implies that $\text{inrc}(\mathbf{t}, n) = \text{inrc}(T^m(\mathbf{c}_\Delta), n)$. Thus (i) follows directly from Proposition 5.5(i). We apply Proposition 5.5 similarly when $c_{k+1} = a_{k+1}$ or $c_{k+1} = a_{k+1} - 1 > 0$ and $c_k \neq 0$ to obtain (ii) and (iii). When there exists i such that $k + 2 \leq i \leq k + d$ and $c_i < a_i$, then we obtain (iv) directly from Proposition 5.5.

Consider the final case (v), that is, assume that $c_i = a_i$ for all i such that $k + 2 \leq i \leq k + d$. Then the Ostrowski conditions imply that $c_{k+1} = 0$. In the latter part of the proof of Proposition 5.7, we saw that our assumptions implied that $m \notin \lambda_{i,3}, \lambda_{i,4}$. Moreover, we proved that if $m \in \lambda_{i,1}$, then $\text{inrc}(\mathbf{t}, n) = \text{inrc}(T^m(\mathbf{c}_\Delta), n)$. If $m \in \lambda_{i,1}$, then the Case A of the proof of Proposition 5.5 applies, and we have $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n) = q_k$ when $|u_{r_k}| < n \leq |u_{r_k+1}| - \operatorname{val}_\Delta(c_1 \cdots c_k)$. Assume then that $m \in \lambda_{i,2}$. As we saw in the Case B of the proof of Proposition 5.7, this means that $n \in \mathcal{I}_{k,\ell}$ with $\ell = a_{k+1} - 1$. Now $\operatorname{inrc}(\mathbf{t}, n) = \operatorname{inrc}(T^{m'}(\mathbf{c}_\Delta), n)$ by Proposition 5.7, so it suffices to prove that $\operatorname{inrc}(T^{m'}(\mathbf{c}_\Delta), n) = q_k$ to obtain (v).

Recall that the number $\operatorname{inrc}(T^m(\mathbf{c}_\Delta), n)$ is determined by the cycle sequence taken in the graph $\Gamma(n)$ and that the cycle sequence is determined by the letters of the standard word $\mathbf{c}_{\Delta''}$ with $\Delta'' = T^{r_k+\ell}(\Delta) = x_{k+1}x_{k+2}^{a_{k+2}} \cdots$. Recall also that the prefix of $T^{L_i}(\mathbf{c}_\Delta)$ of length n corresponds to the left special vertex of $\Gamma(n)$ and that L_i equals the length of the $\mu_{r_k+\ell}$-images of the first $i-1$ blocks. Phrased alternatively, L_i is the length of the τ_{k+1}-image of the prefix of Δ' of length $i-1$ where $\Delta' = T^{r_{k+1}}(\Delta) = T(\Delta'')$. Now $m' - m = c_{k+d}q_{k+d-1}$, so reading off the first $L_i + m' - m$ letters of $\mathbf{c}_\Delta$ corresponds exactly to taking the cycles determined by the $L_{x_{k+1}}^{a_{k+1}-\ell}$-image of $\mathbf{c}_{\Delta'}$ of length $\operatorname{val}_{\Delta'}(c_{k+2} \cdots c_{k+d})$.

Let us then find out the letter y of $\mathbf{c}_{\Delta'}$ at position $\operatorname{val}_{\Delta'}(c_{k+2} \cdots c_{k+d}) + 1$. By (12), we have $v_{d-1} = v_{d-2}^{a_{k+d}} \cdots v_0^{a_{k+2}} x_{k+1}$ for standard words v_t with the directive word Δ'. The length of the prefix of v_{d-1} obtained by removing the suffix x_{k+1} equals $\operatorname{val}_{\Delta'}(c_{k+2} \cdots c_{k+d})$, so $y = x_{k+1}$. The next cycles taken after reading off the first $L_i + m' - m$ letters of $\mathbf{c}_\Delta$ are thus determined by the $L_{x_{k+1}}^{a_{k+1}-\ell}$-image of the letter y and the letter immediately after it. Since the $L_{x_{k+1}}^{a_{k+1}-\ell}$-image of any letter begins with x_{k+1}, we see that $T^{L_i+m'-m}(\mathbf{c}_\Delta)$ initially takes the cycle $C_{x_{k+1}}$ twice.

Since $m \in \lambda_{i,2}$, we see as in the Case B of the proof of Proposition 5.3 that the prefix of $T^m(\mathbf{c}_\Delta)$ of length n lies on the cycle $C_{x_{k+1}}$ but not on the central path. In other words, when $m - L_i$ edges are traversed on the cycle $C_{x_{k+1}}$, starting from the left special vertex, the whole cycle is not traversed. Consider now the word $T^{L_i+m'-m}(\mathbf{c}_\Delta)$ that takes the cycle $C_{x_{k+1}}$ twice. When $m - L_i$ more letters are read, the current vertex is vertex v on the cycle $C_{x_{k+1}}$. When $\|C_{x_{k+1}}\|$ more letters are read, the vertex v is encountered again. Since $L_i + m' - m + m - L_i = m'$, we see that $\operatorname{inrc}(T^{m'}(\mathbf{c}_\Delta), n) = \|C_{x_{k+1}}\|$. Now $\|C_{x_{k+1}}\| = q_k$, and the claim is proved. □

The Sturmian case was determined in [35, Théor. 5.3.0.4].

Theorem 5.11 (*Sturmian Case*). *Let* $\mathbf{t}$ *be a Sturmian word with directive word* Δ *and intercept* $c_1c_2\cdots$. *Suppose that* $n \in \mathcal{I}_k$ *for* $k \geq 0$.

(i) If $0 < c_{k+1} < a_{k+1} - 1$ *or* $c_{k+1} = a_{k+1} - 1 > 0$ *and* $c_k = 0$, *then we have the following implications:*

- $q_k - 2 < n \leq q_{k+1} - 2 - \operatorname{val}(c_1 \cdots c_{k+1})$
  $\implies \operatorname{inrc}(\mathbf{t}, n) = q_k$,
- $q_{k+1} - 2 - \operatorname{val}(c_1 \cdots c_{k+1}) < n \leq q_{k+1} - 2 - c_{k+1}q_k$
  $\implies \operatorname{inrc}(\mathbf{t}, n) = q_{k+1} - \operatorname{val}(c_1 \cdots c_{k+1})$,
- $q_{k+1} - 2 - c_{k+1}q_k < n \leq q_{k+1} + q_k - 2 - \operatorname{val}(c_1 \cdots c_{k+1})$
  $\implies \operatorname{inrc}(\mathbf{t}, n) = q_{k+1} - c_{k+1}q_k$,
- $q_{k+1} + q_k - 2 - \operatorname{val}(c_1 \cdots c_{k+1}) < n \leq q_{k+1} - 2$
  $\implies \operatorname{inrc}(\mathbf{t}, n) = q_{k+1} + q_k - \operatorname{val}(c_1 \cdots c_{k+1})$.

(ii) If $c_{k+1} = a_{k+1}$, *then we have the following implications:*

- $q_k - 2 < n \leq q_{k+1} - 2 - \operatorname{val}(c_1 \cdots c_k)$
  $\implies \operatorname{inrc}(\mathbf{t}, n) = q_{k-1}$,
- $q_{k+1} - 2 - \operatorname{val}(c_1 \cdots c_k) < n \leq q_{k+1} - 2$
  $\implies \operatorname{inrc}(\mathbf{t}, n) = q_k + q_{k-1} - \operatorname{val}(c_1 \cdots c_k)$.

(iii) If $c_{k+1} = a_{k+1} - 1 > 0$ *and* $c_k \neq 0$, *then we have the following implications:*

- $q_k - 2 < n \leq q_k + q_{k-1} - 2$
  $\implies \operatorname{inrc}(\mathbf{t}, n) = q_k + q_{k-1} - \operatorname{val}(c_1 \cdots c_k)$,
- $q_k + q_{k-1} - 2 < n \leq 2q_k + q_{k-1} - 2 - \operatorname{val}(c_1 \cdots c_k)$
  $\implies \operatorname{inrc}(\mathbf{t}, n) = q_k + q_{k-1}$,

- $2q_k + q_{k-1} - 2 - \text{val}(c_1 \cdots c_k) < n \le q_{k+1} - 2$
  $\Longrightarrow \text{inrc}(\mathbf{t}, n) = 2q_k + q_{k-1} - \text{val}(c_1 \cdots c_k)$.

(iv) Suppose that $c_{k+1} = 0$ and $c_{k+2} < a_{k+2}$.

(iv.a) If $c_k = 0$, then we have the following implications:

- $q_k - 2 < n \le q_{k+1} - 2 - \text{val}(c_1 \cdots c_{k-1})$
  $\Longrightarrow \text{inrc}(\mathbf{t}, n) = q_k$,
- $q_{k+1} - 2 - \text{val}(c_1 \cdots c_{k-1}) < n \le q_{k+1} - 2$
  $\Longrightarrow \text{inrc}(\mathbf{t}, n) = q_{k+1} - \text{val}(c_1 \cdots c_{k-1})$.

(iv.b) If $a_{k+1} - 1 = c_{k+1}$ and $c_k \neq 0$, then we have the following implication:

- $q_k - 2 < n \le q_{k+1} - 2$
  $\Longrightarrow \text{inrc}(\mathbf{t}, n) = q_{k+1} - \text{val}(c_1 \cdots c_k)$.

(iv.c) If $a_{k+1} - 1 > c_{k+1}$ and $c_k \neq 0$, then we have the following implications:

- $q_k - 2 < n \le q_{k+1} - 2 - \text{val}(c_1 \cdots c_k)$
  $\Longrightarrow \text{inrc}(\mathbf{t}, n) = q_k$.
- $q_{k+1} - 2 - \text{val}(c_1 \cdots c_k) < n \le q_{k+1} - 2$
  $\Longrightarrow \text{inrc}(\mathbf{t}, n) = q_{k+1} - \text{val}(c_1 \cdots c_k)$.

(v) If $c_{k+1} = 0$ and $a_{k+2} = c_{k+2}$, then we have the following implication:

- $q_k - 2 < n \le q_{k+1} - 2$
  $\Longrightarrow \text{inrc}(\mathbf{t}, n) = q_k$.

Proof. Utilize again the computations of Example 5.2. □

6. Diophantine exponents

Recall the definition of the Diophantine exponent $\text{dio}(\mathbf{x})$ of an infinite word $\mathbf{x}$ from Section 1.3. It is clear from the definition that $\text{dio}(\mathbf{x}) \ge 1$ for all infinite words $\mathbf{x}$. It is possible that $\text{dio}(\mathbf{x}) = \infty$ (as we shall see). It is proved in [17] that for each real number θ such that $\theta \ge 1$, there exists an infinite binary word $\mathbf{x}$ such that $\text{dio}(\mathbf{x}) = \theta$. Propositions 4.3 and 6.2 of the recent paper [6] provide lower and upper bounds for Diophantine exponents of infinite words generated by morphisms.

Diophantine exponents relate to our work through the following result that allows us to compute the Diophantine exponent of a regular episturmian word with the help of Theorem 5.10.

Proposition 6.1 *([13, Lemma 10.3]). If $\mathbf{x}$ is an infinite word, then*

$$\text{dio}(\mathbf{x}) = 1 + \limsup_{n\to\infty} \frac{n}{\text{inrc}(\mathbf{x}, n)}.$$

Proof. There are some notational differences. The claim follows directly from [13, Lemma 10.3] with the observation that $\text{inrc}(\mathbf{x}, n) = r(n, \mathbf{x}) - n$ where r is defined as on p. 3282 of [13]. □

By Proposition 6.1 and Theorem 5.10, finding $\text{dio}(\mathbf{t})$ for a regular episturmian word $\mathbf{t}$ amounts to determining the ratio $n/\text{inrc}(\mathbf{t}, n)$ on the right endpoint of each of the subintervals of $\mathcal{I}_k$ described by Theorem 5.10.

Let us then define the closely-related notions of initial critical exponent and index.

Definition 6.2. Let $\mathbf{x}$ be an infinite word. The *initial critical exponent* of $\mathbf{x}$, denoted by $\text{ice}(\mathbf{x})$, is the supremum of the rational numbers ρ for which there exist arbitrarily long prefixes V of $\mathbf{x}$ such that V^ρ is a prefix of $\mathbf{x}$.

A formula for the initial critical exponent of a Sturmian word is given in [10, Cor. 3.5]. According to our knowledge, no attempt to study the initial critical exponents of general episturmian words has been attempted. The methods used to prove Theorem 5.10 could be used for such a study. It is worth mentioning that a slight modification of the proof of [13, Lemma 10.3] yields

$$\operatorname{ice}(\mathbf{x}) = 1 + \limsup_{n\to\infty} \frac{n}{\operatorname{pnrc}(\mathbf{x}, n)}$$

for an infinite word $\mathbf{x}$. Here $\operatorname{pnrc}(\mathbf{x}, n)$ is the *prefix nonrepetitive complexity function* of $\mathbf{x}$ defined by setting

$$\operatorname{pnrc}(\mathbf{x}, n) = \max\{m : \mathbf{x}[i, i+n-1] \neq \mathbf{x}[1, n-1] \text{ for all } 2 \leq i < m\}.$$

Thus the prefix of $\mathbf{x}$ of length $\operatorname{pnrc}(\mathbf{x}, n) + n$ is the shortest prefix of $\mathbf{x}$ containing two occurrences of the prefix of $\mathbf{x}$ of length n. We have the obvious relation $\operatorname{inrc}(\mathbf{x}, n) \leq \operatorname{pnrc}(\mathbf{x}, n)$ for all n.

Definition 6.3. Let $\mathbf{x}$ be an infinite word. The *index* (or critical exponent) of $\mathbf{x}$, denoted by $\operatorname{ind}(\mathbf{x})$, is the number

$$\sup\{e \in \mathbb{Q} \cap [1, \infty) : \mathbf{x} \text{ has an } e\text{-power as a factor}\}.$$

The indices have been determined for various classes of infinite words. It suffices to say here that a formula for the index of Sturmian words was derived independently in [14,15,22] (see also [30]). The index of a regular episturmian word can be found by applying [18, Lemma 6.5,Thm. 6.19]. It seems that the discussion in Sections 4.1 and 5.5 of [23] is all that has been said of the indices of general episturmian words.

From the definitions, we infer the following inequalities:

$$\operatorname{ice}(\mathbf{x}) \leq \operatorname{dio}(\mathbf{x}) \leq \operatorname{ind}(\mathbf{x}).$$

It is possible that $\operatorname{ice}(\mathbf{x}) < \operatorname{dio}(\mathbf{x})$ (see [10, p. 18]) or $\operatorname{dio}(\mathbf{x}) < \operatorname{ind}(\mathbf{x})$ (see Proposition 6.5).

Let us then show that the Diophantine exponent is shift-invariant. The same is not true for the initial critical exponent; see [10, Prop. 2.1, Sect. 4.2].

Proposition 6.4. *Let $\mathbf{x}$ be an infinite word. Then $\operatorname{dio}(T^m(\mathbf{x})) = \operatorname{dio}(\mathbf{x})$ for all $m \geq 0$.*

Proof. Let (U_n), (V_n), and (e_n) be sequences such that $|U_n V_n^{e_n}|/|U_n V_n|$ converges to $\operatorname{dio}(\mathbf{x})$ and $(|U_n V_n|)$ is increasing. Since either $|U_n| \to \infty$ or $|V_n| \to \infty$, we may factorize the prefix of $T^m(\mathbf{x})$ of length $|U_n V_n^{e_n}| - m$ as $U_n' V_n'^{e_n'}$ for a conjugate V_n' of V_n for n large enough. For n large enough, we have

$$\frac{|U_n' V_n'^{e_n'}|}{|U_n' V_n'|} = \frac{|U_n V_n^{e_n}| - m}{|U_n V_n| - m} = \frac{|U_n V_n^{e_n}|/|U_n V_n| - m/|U_n V_n|}{1 - m/|U_n V_n|} \to \lim_{n\to\infty} \frac{|U_n V_n^{e_n}|}{|U_n V_n|} = \operatorname{dio}(\mathbf{x}).$$

Therefore $\operatorname{dio}(T^m(\mathbf{x})) \geq \operatorname{dio}(\mathbf{x})$. A similar argument establishes that $\operatorname{ice}(T^m(\mathbf{x})) \geq \operatorname{ice}(\mathbf{x})$.

A symmetric argument where we prepend a factor of length m to words $U_n V_n^{e_n}$ such that $(|U_n V_n|)$ is increasing and $|U_n V_n^{e_n}|/|U_n V_n|$ converges to $\operatorname{dio}(T^m(\mathbf{x}))$ establishes that $\operatorname{dio}(\mathbf{x}) \geq \operatorname{dio}(T^m(\mathbf{x}))$. This argument does not apply to the initial critical exponent unless the prepended factor is compatible with the prefix powers of $T^m(\mathbf{x})$. □

6.1. Diophantine exponents of regular episturmian words

As the first application of Theorem 5.10, we find the Diophantine exponent of a regular epistandard word.

Proposition 6.5. *Let Δ be a regular directive word as in (2). Then*

$$\operatorname{dio}(\mathbf{c}_\Delta) = 1 + \limsup_{k\to\infty}(a_{k+1} + |u_{r_{k-(d-1)}}|/q_k) = \operatorname{ind}(\mathbf{c}_\Delta) - 1.$$

Proof. The intercept of $\mathbf{c}_\Delta$ is 0^ω. Thus we find from the first implication of Theorem 5.10 (iv.a) that

$$\max_{n\in\mathcal{I}_k} \frac{n}{\operatorname{inrc}(\mathbf{c}_\Delta, n)} = \frac{|u_{r_{k+1}}|}{q_k}$$

for $k \geq 0$. Since $u_{r_{k+1}} = s_k^{a_*-1}s_{k-1}^{a_*}\cdots s_0^{a_*} = s_k^{a_*}u_{r_{k-d+1}}$ by (7) and (13), it follows from Proposition 6.1 that $\operatorname{dio}(\mathbf{c}_\Delta) = 1 + \limsup_{k\to\infty}(a_{k+1} + |u_{r_{k-(d-1)}}|/q_k)$. If $\mathbf{c}_\Delta$ has unbounded partial quotients, that is, the sequence (a_k) is unbounded, then $\operatorname{dio}(\mathbf{c}_\Delta) = \infty$. Since $\operatorname{dio}(\mathbf{c}_\Delta) \leq \operatorname{ind}(\mathbf{c}_\Delta)$, the claim follows. If $\mathbf{c}_\Delta$ has bounded partial quotients, then it follows from [18, Lemma 6.5,Thm. 6.19] that

$$\operatorname{ind}(\mathbf{c}_\Delta) = 2 + \limsup_{k\to\infty}(a_{k+1} + |u_{r_{k-(d-1)}}|/q_k) \quad \square$$

We saw in the preceding proof that $\operatorname{dio}(\mathbf{c}_\Delta) = \infty$ if and only if $\mathbf{c}_\Delta$ has unbounded partial quotients. This is in fact true for all words in the subshift generated by $\mathbf{c}_\Delta$ as the next theorem states. This theorem was originally proved for Sturmian words in [4, Prop. 11.1] and an alternative proof was given in [13, Thm. 3.3]. Our more general proof uses yet another approach.

Theorem 6.6. *Let $\mathbf{t}$ be a regular episturmian word. Then $\operatorname{dio}(\mathbf{t}) < \infty$ if and only if $\mathbf{t}$ has bounded partial quotients.*

Before the proof, let us derive some helpful inequalities.

Lemma 6.7. *Let $\mathbf{t}$ be a regular episturmian word of period d with directive word Δ and intercept $c_1c_2\cdots$. Let y be a letter occurring in Δ and Q a nonnegative integer. Then*

$$\frac{|u_{r_{k+1}}|}{q_k + |\tau_{k+1}(y)| - \operatorname{val}(c_1\cdots c_{k+1})} \geq \frac{a_{k+1} + q_{k-(d+1)}/q_k}{a_{k+1} - c_{k+1} + 2}, \tag{22}$$

$$\frac{|u_{r_{k+1}}| - \operatorname{val}(c_1\cdots c_{k+1})}{q_k} \geq a_{k+1} - c_{k+1} - 1 + q_{k-(d+1)}/q_k, \tag{23}$$

$$\frac{|u_{r_{k+1}}|}{Qq_k + |\tau_k(y)| - \operatorname{val}(c_1\cdots c_k)} \geq \frac{1}{Q+1}\left(a_{k+1} + q_{k-(d+1)}/q_k\right) \tag{24}$$

for all k large enough.

Proof. From Lemma 3.11, (14), Lemma 5.8, and (15), we find that

$$\begin{aligned} 0 &\leq \operatorname{val}(c_1\cdots c_k) < q_k, \\ q_{k+1} &< (a_{k+1}+1)q_k, \\ |\tau_k(y)| &\leq q_k, \text{ and} \\ a_kq_{k-1} + q_{k-(d+2)} &\leq |u_{r_k}| \end{aligned}$$

for all k large enough. Therefore

$$\begin{aligned} \frac{|u_{r_{k+1}}|}{q_k + |\tau_{k+1}(y)| - \operatorname{val}(c_1\cdots c_{k+1})} &\geq \frac{a_{k+1}q_k + q_{k-(d+1)}}{q_k + q_{k+1} - \operatorname{val}(c_1\cdots c_k) - c_{k+1}q_k} \\ &\geq \frac{a_{k+1}q_k + q_{k-(d+1)}}{q_k + (a_{k+1}+1)q_k - c_{k+1}q_k} \\ &= \frac{a_{k+1} + q_{k-(d+1)}/q_k}{a_{k+1} - c_{k+1} + 2} \end{aligned}$$

for all k large enough. This proves the first inequality. The second inequality is derived similarly:

$$\begin{aligned} \frac{|u_{r_{k+1}}| - \operatorname{val}(c_1\cdots c_{k+1})}{q_k} &\geq \frac{a_{k+1}q_k + q_{k-(d+1)} - \operatorname{val}(c_1\cdots c_{k+1})}{q_k} \\ &\geq \frac{a_{k+1}q_k + q_{k-(d+1)} - c_{k+1}q_k - q_k}{q_k} \\ &= a_{k+1} - c_{k+1} - 1 + q_{k-(d+1)}/q_k \end{aligned}$$

for all k large enough. The third inequality is derived analogously by using the inequality $|u_{r_{k+1}}| \geq a_{k+1}q_k + q_{k-(d+1)}$ in the numerator and the inequalities $|\tau_k(y)| \leq q_k$ and $\mathrm{val}(c_1 \cdots c_k) \geq 0$ in the numerator. □

Proof of Theorem 6.6. Say $\mathbf{t}$ has directive word as in (2) and intercept $c_1c_2\cdots$. If $\mathbf{t}$ has bounded partial quotients, then $\mathrm{ind}(\mathbf{t}) < \infty$ [23, Sect. 4.1], so $\mathrm{dio}(\mathbf{t}) \leq \mathrm{ind}(\mathbf{t}) < \infty$. We may thus assume that $\mathbf{t}$ has unbounded partial quotients. Thus there exists a sequence (k_j) such that $a_{k_j+1} \to \infty$ as $j \to \infty$. Below we impose restrictions on (k_j) and show that under each restriction $\mathrm{dio}(\mathbf{t}) = \infty$. It is straightforward to see that (by taking an appropriate subsequence) some (k_j) must satisfy at least one of the restrictions or $c_k = 0$ for k large enough. When $c_k = 0$ for all k large enough, the word $\mathbf{t}$ is a shift of $\mathbf{c}_\Delta$, and Propositions 6.4 and 6.5 imply that $\mathrm{dio}(\mathbf{t}) = \infty$.

Case A. Assume that $0 < c_{k_j+1} < a_{k_j+1} - 1$ for all j. Suppose additionally that there exists a nonnegative constant M such that $a_{k_j+1} - c_{k_j+1} \leq M$ for all j. By the fourth implication of Theorem 5.10(i), it suffices to show that

$$\frac{|u_{r_{k_j+1}}|}{q_{k_j} + |\tau_{k_j+1}(y_i)| - \mathrm{val}(c_1 \cdots c_{k_j+1})} \xrightarrow{j\to\infty} \infty$$

in order to conclude that $\mathrm{dio}(\mathbf{t}) = \infty$ (here y_i is as in Theorem 5.10). By (22), we have

$$\frac{|u_{r_{k_j+1}}|}{q_{k_j} + |\tau_{k_j+1}(y_i)| - \mathrm{val}(c_1 \cdots c_{k_j+1})} \geq \frac{a_{k_j+1} + q_{k_j-(d+1)}/q_{k_j}}{a_{k_j+1} - c_{k_j+1} + 2} \geq \frac{1}{M+2} a_{k_j+1}$$

for j large enough. As $a_{k_j+1} \to \infty$ as $j \to \infty$, it follows that $\mathrm{dio}(\mathbf{t}) = \infty$. Assume then that the difference $a_{k_j+1} - c_{k_j+1}$ is unbounded. This time around, we consider the first implication of Theorem 5.10(i). Notice that we need to ensure that the interval $\{|u_{r_{k_j}}|, \ldots, |u_{r_{k_j+1}}| - \mathrm{val}(c_1 \cdots c_{k_j+1})\}$ is nonempty, but this is straightforward to do. From (23), we have

$$\frac{|u_{r_{k_j+1}}| - \mathrm{val}(c_1 \cdots c_{k_j+1})}{q_{k_j}} \geq a_{k_j+1} - c_{k_j+1} - 1 + q_{k_j-(d+1)}/q_{k_j} \xrightarrow{j\to\infty} \infty$$

because the difference $a_{k_j+1} - c_{k_j+1}$ is unbounded. Hence $\mathrm{dio}(\mathbf{t}) = \infty$ in this case as well.

Case B. If $c_{k_j+1} = a_{k_j+1} - 1 > 0$ and $c_{k_j} = 0$ for all j, then the difference $a_{k_j+1} - c_{k_j+1}$ is bounded. By repeating the arguments of Case A, we see that $\mathrm{dio}(\mathbf{t}) = \infty$.

Case C. Suppose that $c_{k_j+1} = a_{k_j+1}$ for all j. From (24), we see that

$$\frac{|u_{r_{k_j+1}}|}{q_{k_j} + |\tau_{k_j}(y_i)| - \mathrm{val}(c_1 \cdots c_k)} \geq \frac{1}{2}(a_{k_j+1} + q_{k-(d+1)}/q_k)$$

so $\mathrm{dio}(\mathbf{t}) = \infty$ by the last implication of Theorem 5.10 (ii).

Case D. Assume that $c_{k_j+1} = a_{k_j+1} - 1 > 0$ and $c_{k_j} \neq 0$ for all j. We apply the last implication of Theorem 5.11 (iii) and obtain similar to the Case C that

$$\frac{|u_{r_{k_j+1}}|}{2q_{k_j} + |\tau_{k_j}(y_i)| - \mathrm{val}(c_1 \cdots c_{k_j})} \geq \frac{1}{3}(a_{k_j+1} + q_{k-(d+1)}/q_k).$$

Hence again $\mathrm{dio}(\mathbf{t}) = \infty$. □

Interestingly a statement analogous to Theorem 6.6 is not true for the initial critical exponent. It is shown in [10, Prop. 4.1] that every Sturmian subshift contains a word $\mathbf{t}$ such that $\mathrm{ice}(\mathbf{t}) \leq 1+\varphi \approx 2.6180$ where φ is the Golden ratio. Even more interestingly it is possible that $\mathrm{ice}(\mathbf{t}) = 2$ for certain Sturmian words $\mathbf{t}$ with unbounded partial quotients [10, Thm 1.1]. We suspect that it is equally possible that the initial critical exponent is finite while the Diophantine exponent is infinite when $d > 2$.

Next we show that $\mathrm{dio}(\mathbf{t}) > 2$ for essentially all regular episturmian words. For Sturmian words, this result can be inferred from the results of [10] as indicated in the proof of [1, Prop. 4].

Theorem 6.8. *Let* $\mathbf{t}$ *be a regular episturmian word of period* d *with directive word as in (2). If* $d = 2$ *or* $\limsup_k a_k \geq 3$*, then* $\mathrm{dio}(\mathbf{t}) > 2$*.*

Proof. The claim follows from Theorem 6.6 if $\mathbf{t}$ has unbounded partial quotients. Suppose that $\mathbf{t}$ has bounded partial quotients and intercept $c_1 c_2 \cdots$. Let $C = M + 1$ where $M = \limsup_k a_k$. It follows from (14) that $q_{k+1}/q_k \leq C$ for all k large enough.

Case A. Assume first that there exists infinitely many k such that $0 < c_{k+1} < a_{k+1} - 1$. By the first implication of Theorem 5.10(i), we have

$$\mathrm{dio}(\mathbf{t}) - 1 \geq \limsup_{k\to\infty} \frac{|u_{r_{k+1}}| - \mathrm{val}(c_1 \ldots c_{k+1})}{q_k}$$

where we consider the limit superior over an appropriate subsequence like in the proof of Theorem 6.6. From (23), we find that

$$\frac{|u_{r_{k+1}}| - \mathrm{val}(c_1 \ldots c_{k+1})}{q_k} \geq a_{k+1} - c_{k+1} - 1 + q_{k-(d+1)}/q_k \geq 1 + C^{-(d+1)}.$$

The claim follows.

Case B. Let us assume that $c_{k+1} = a_{k+1} - 1 > 0$ and $c_k = 0$ for infinitely many k. Suppose in addition that $a_{k+1} \geq 3$. By the final implication of Theorem 5.10(i), we have

$$\mathrm{dio}(\mathbf{t}) - 1 \geq \limsup_{k\to\infty} \frac{|u_{r_{k+1}}|}{q_k + |\tau_{k+1}(y_i)| - \mathrm{val}(c_1 \cdots c_{k+1})}.$$

By applying (22), we see that

$$\begin{aligned}\frac{|u_{r_{k+1}}|}{q_k + |\tau_{k+1}(y)| - \mathrm{val}(c_1 \cdots c_{k+1})} &\geq \frac{a_{k+1} + q_{k-(d+1)}/q_k}{a_{k+1} - c_{k+1} + 2} \\ &= \frac{a_{k+1} + q_{k-(d+1)}/q_k}{3} \\ &\geq 1 + C^{-(d+1)}/3,\end{aligned}$$

so the claim follows.

Case C. Assume that $c_{k+1} = a_{k+1} - 1 > 0$ and $c_k \neq 0$ for infinitely many k. Assume moreover that $a_{k+1} \geq 3$. The last implication of Theorem 5.10 (iii) gives

$$\mathrm{dio}(\mathbf{t}) - 1 \geq \limsup_{k\to\infty} \frac{|u_{r_{k+1}}|}{2q_k + |\tau_k(y_i)| - \mathrm{val}(c_1 \cdots c_k)}.$$

From (24), we obtain that

$$\frac{|u_{r_{k+1}}|}{2q_k + |\tau_k(y_i)| - \mathrm{val}(c_1 \cdots c_k)} \geq \frac{1}{3}\left(a_{k+1} + q_{k-(d+1)}/q_k\right) \geq 1 + C^{-(d+1)}/3,$$

so the claim holds in this case as well.

Case D. Suppose that $c_{k+1} = a_{k+1}$ for infinitely many k. Assume moreover that $a_{k+1} \geq 2$. By the second implication of Theorem 5.10 (ii), we have

$$\mathrm{dio}(\mathbf{t}) - 1 \geq \limsup_{k\to\infty} \frac{|u_{r_{k+1}}|}{q_k + |\tau_k(y_i)| - \mathrm{val}(c_1 \cdots c_k)}.$$

Again, from (24), we see that

$$\frac{|u_{r_{k+1}}|}{q_k + |\tau_k(y_i)| - \mathrm{val}(c_1 \cdots c_k)} \geq \frac{1}{2}\left(a_{k+1} + q_{k-(d+1)}/q_k\right) \geq 1 + C^{-(d+1)}/2.$$

Case E. Suppose that $c_{k+1} = c_k = 0$ for infinitely many k. Suppose additionally that $a_{k+1} \geq 2$. Then the first implication of Theorem 5.10 (iv.a) gives

$$\mathrm{dio}(\mathbf{t}) - 1 \geq \limsup_{k\to\infty} \frac{|u_{r_{k+1}}| - \mathrm{val}(c_1 \cdots c_{k-1})}{q_k}.$$

Now from (23), we find that

$$\frac{|u_{r_{k+1}}| - \mathrm{val}(c_1 \cdots c_{k-1})}{q_k} \geq a_{k+1} - 1 + C^{-(d+1)},$$

so the claim follows.

Case F. Assume finally that $c_{k+1} = 0$ and $c_k \neq 0$ for infinitely many k. Suppose in addition that $a_{k+1} \geq 2$. We deduce from the first implication of Theorem 5.10 (iv.c) that

$$\mathrm{dio}(\mathbf{t}) - 1 \geq \limsup_{k \to \infty} \frac{|u_{r_{k+1}}| - \mathrm{val}(c_1 \cdots c_k)}{q_k}.$$

Thus the claim follows by an application of (23) as in the Case E.

The Cases A–F prove the claim when $\limsup_k a_k \geq 3$, so we may consider the special case $d = 2$. Next we handle the Cases B–D unconditionally. Consider the Case D first. The Ostrowski conditions imply that $c_k = 0$. Thus the first implication of Theorem 5.11 (ii) gives

$$\begin{aligned}\frac{q_{k+1} - 2 - \mathrm{val}(c_1 \cdots c_k)}{q_{k-1}} &= \frac{a_{k+1}q_k + q_{k-1} - 2 - \mathrm{val}(c_1 \cdots c_{k-1})}{q_{k-1}} \\ &\geq \frac{a_{k+1}q_k - 2}{q_{k-1}} \\ &= a_{k+1}a_k + \frac{a_{k+1}q_{k-2} - 2}{q_{k-1}},\end{aligned}$$

so

$$\mathrm{dio}(\mathbf{t}) - 1 \geq \limsup_{k \to \infty} \left(a_{k+1}a_k + \frac{a_{k+1}q_{k-2} - 2}{q_{k-1}} \right) \geq 1 + C^{-1}.$$

Suppose then that $c_{k+1} = a_{k+1} - 1 > 0$ and $c_k = 0$. Taking into account what we have already proved, we see that we may assume that $c_{k-1} < a_{k-1}$. Now the first implication of Theorem 5.11(i) yields

$$\begin{aligned}\frac{q_{k+1} - 2 - \mathrm{val}(c_1 \cdots c_{k+1})}{q_k} &= \frac{q_k + q_{k-1} - 2 - \mathrm{val}(c_1 \cdots c_{k-1})}{q_k} \\ &= 1 + \frac{q_{k-1} - 2 - c_{k-1}q_{k-2} - \mathrm{val}(c_1 \cdots c_{k-2})}{q_k} \\ &\geq 1 + \frac{q_{k-1} - 2 - (c_{k-1} + 1)q_{k-2}}{q_k} \\ &\geq 1 + \frac{q_{k-3} - 2}{q_k},\end{aligned}$$

so $\mathrm{dio}(\mathbf{t}) - 1 \geq 1 + C^{-3}$. Suppose finally that $c_{k+1} = a_{k+1} - 1 > 0$ and $c_k \neq 0$. We may again assume that $c_k < a_k$. From the second implication of Theorem 5.11 (iii), we have

$$\begin{aligned}\frac{2q_k + q_{k-1} - 2 - \mathrm{val}(c_1 \cdots c_k)}{q_k + q_{k-1}} &= 1 + \frac{q_k - 2 - c_kq_{k-1} - \mathrm{val}(c_1 \cdots c_{k-1})}{q_k + q_{k-1}} \\ &\geq 1 + \frac{q_k - (c_k + 1)q_{k-1} - 2}{q_k + q_{k-1}} \\ &\geq 1 + \frac{q_{k-2} - 2}{q_k + q_{k-1}} \\ &\geq 1 + \frac{q_{k-2} - 2}{q_{k+1}},\end{aligned}$$

so $\mathrm{dio}(\mathbf{t}) - 1 \geq 1 + C^{-3}$ in this case as well. As the cases A–D are now handled when $d = 2$, the claim is true or $c_k = 0$ for all k large enough. In the latter case the word $\mathbf{t}$ is a shift of $\mathbf{c}_\Delta$, and the claim follows from Propositions 6.4 and 6.5. □

For a fixed d, the conclusion of Theorem 6.8 can be improved. For example, it is shown in [13, Thm. 4.3] that

$$\operatorname{dio}(\mathbf{t}) \geq \frac{5}{3} + \frac{4\sqrt{10}}{15} \approx 2.5099$$

for a Sturmian word $\mathbf{t}$. We have not attempted such a study for $d > 2$. However, if d is not fixed, it is unlikely that the lower bound 2 can be improved. The main indication to this is the fact that the Diophantine exponent of the standard d-bonacci tends to 2 as $d \to \infty$ (see the discussion after Proposition 6.15).

Theorem 6.8 leaves open if $\operatorname{dio}(\mathbf{t}) > 2$ when $\limsup_k a_k \leq 2$ and $d > 2$. We show next that this is not necessarily true by providing explicit counterexamples. Before this, we prove the following lemmas that narrow down the exceptional directive words and intercepts even further.

Lemma 6.9. *Let $\mathbf{t}$ be an episturmian word of period d with directive word as in (2) and intercept $c_1c_2\cdots$. If one of the conditions*

(i) 0^d occurs infinitely many times in $c_1c_2\cdots$,
(ii) there exist infinitely many k such that $c_i = a_i$ for all i such that $k \leq i \leq k+d-2$

is satisfied, then $\operatorname{dio}(\mathbf{t}) > 2$.

Proof. We may assume that $\mathbf{t}$ has bounded partial quotients and let $C = 1 + \limsup_k a_k$. Suppose that 0^d occurs infinitely many times in $c_1c_2\cdots$. Let k be such that $c_i = 0$ for all i such that $k-(d-1) \leq i \leq k$. Suppose that $c_{k+1} = 0$. Then we apply the first implication of Theorem 5.10 (iv.a), (7), Lemma 3.11, and (14) as follows:

$$\begin{aligned}
\operatorname{dio}(\mathbf{t}) - 1 &\geq \limsup_{k\to\infty} \frac{|u_{r_{k+1}}| - \operatorname{val}(c_1\cdots c_{k-d})}{q_k} \\
&\geq \limsup_{k\to\infty} \frac{q_k + |s_{k-d}^{a_*-1}\cdots s_0^{a_*}| - q_{k-d}}{q_k} \\
&\geq 1 + \limsup_{k\to\infty} \frac{q_{k-d} + |s_{k-2d}^{a_*-1}\cdots s_0^{a_*}| - q_{k-d}}{q_k} \\
&\geq 1 + C^{-(2d+1)}.
\end{aligned}$$

Consider then the case $c_{k+1} \neq 0$. By inspecting the first implications of Theorem 5.10(i) and (ii), we find that

$$\operatorname{dio}(\mathbf{t}) - 1 \geq \limsup_{k\to\infty} \frac{|u_{r_k+1}| - \operatorname{val}(c_1\cdots c_k)}{q_k},$$

so $\operatorname{dio}(\mathbf{t}) - 1 \geq 1 + C^{-(2d+1)}$ like above.

Assume then that there exist infinitely many k such that $c_i = a_i$ for all i such that $k+2 \leq i \leq k+d$. We must have $c_{k+1} = 0$ by the Ostrowski conditions. We infer from Theorem 5.10 (v) that

$$\operatorname{dio}(\mathbf{t}) - 1 \geq \limsup_{k\to\infty} \frac{|u_{r_{k+1}}|}{q_k} \geq \limsup_{k\to\infty} \left(a_{k+1} + \frac{|s_{k-d}^{a_*-1}\cdots s_0^{a_*}|}{q_k} \right) \geq 1 + C^{-(d+1)} \quad \square$$

Lemma 6.10. *Let $\mathbf{t}$ be a regular episturmian word with directive word as in (2) with $d = 3$. If $\limsup_k a_k = 1$, then $\operatorname{dio}(\mathbf{t}) > 2$.*

Proof. Say $\limsup_k a_k = 1$ and that $\mathbf{t}$ has intercept $c_1c_2\cdots$. By Lemma 6.9, the claim follows if the intercept contains 0^3 infinitely many times or 11 infinitely many times. Therefore we may assume that $c_1c_2\cdots$ is eventually a product of the words 01 and 001.

Assume that 001 occurs infinitely often in the intercept. Let k be large enough such that we are guaranteed that $a_{k+1} = a_k = a_{k-1} = 1$, $c_{k+1} = 1$, and $c_k = c_{k-1} = 0$. By the first implication of

Theorem 5.10 (ii), we have

$$\operatorname{dio}(\mathbf{t}) - 1 \geq \limsup_{k\to\infty} \frac{|u_{r_k+1}| - \operatorname{val}(c_1 \cdots c_k)}{|\tau_k(y_i)|}.$$

By Lemma 5.8, we have $|\tau_k(y_i)| \leq q_{k-1} + q_{k-2}$. Therefore

$$\begin{aligned}
\frac{|u_{r_k+1}| - \operatorname{val}(c_1 \cdots c_k)}{|\tau_k(y_i)|} &\geq \frac{|s_{k-1}s_{k-2}s_{k-3}^{a_*} \cdots s_0^{a_*}| - \operatorname{val}(c_1 \cdots c_{k-2})}{q_{k-1} + q_{k-2}} \\
&\geq \frac{q_{k-1} + |s_{k-3}^{a_*} \cdots s_0^{a_*}|}{q_{k-1} + q_{k-2}} \\
&= \frac{q_{k-1} + q_{k-2} + |s_{k-5}^{a_*-1} s_{k-6}^{a_*} \cdots s_0^{a_*}|}{q_{k-1} + q_{k-2}} \\
&\geq 1 + \frac{q_{k-6}}{q_k} \\
&\geq 1 + C^{-6}
\end{aligned}$$

where $C = 1 + \limsup_k a_k = 2$. Therefore $\operatorname{dio}(\mathbf{t}) > 2$.

The only case left is the case where $c_1c_2 \cdots$ has suffix $(01)^\omega$. Say k is large enough and $c_{k+1} = 0$ and $c_{k+2} = c_k = 1$. Since $c_{k+2} = 1$, we have $y_i \neq x_{k+2}$, and so $|\tau_{k+1}(y_i)| = q_k + q_{k-1}$. From Theorem 5.10 (iv.b), we have

$$\begin{aligned}
\operatorname{dio}(\mathbf{t}) - 1 &\geq \limsup_{k\to\infty} \frac{|u_{r_{k+1}}|}{|\tau_{k+1}(y_i)| - \operatorname{val}(c_1 \cdots c_k)} \\
&\geq \limsup_{k\to\infty} \frac{|s_{k-1}^{a_*} \cdots s_0^{a_*}|}{q_k + q_{k-1} - q_{k-1}} \\
&\geq 1 + \limsup_{k\to\infty} \frac{|s_{k-4}^{a_*} \cdots s_0^{a_*}|}{q_k} \\
&\geq 1 + C^{-4}. \quad \square
\end{aligned}$$

By Lemma 6.10, the conclusion of Theorem 6.8 can fail only if $\limsup_k a_k = 2$ when $d = 3$. The next proposition shows that this may happen. For the proof, recall that the Stolz-Cesàro Theorem states that

$$\lim_{k\to\infty} \frac{a_k}{b_k} = \lim_{k\to\infty} \frac{a_k - a_{k-1}}{b_k - b_{k-1}}$$

whenever the right side limit exists and the sequence (b_k) is strictly monotone.

Proposition 6.11. *Let* $\mathbf{t}$ *be the epistürmian word with directive word* $(001122)^\omega$ *having intercept* 1^ω. *Then*

$$\operatorname{dio}(\mathbf{t}) = 1 + \frac{1}{2}(\beta - 1) \approx 1.9156$$

where β*, approximately 2.8312, is the real root of the polynomial* $x^3 - 2x^2 - 2x - 1$.

Proof. Suppose that k is such that $k \geq 4$. It follows from Theorem 5.10 (iii) that $\operatorname{dio}(\mathbf{t}) - 1$ equals the largest of the limits of the ratios

$$\frac{|u_{r_k+1}|}{q_k + |\tau_k(y_i)| - \operatorname{val}(c_1 \cdots c_k)}, \frac{|u_{r_k+1}| + q_k - \operatorname{val}(c_1 \cdots c_k)}{q_k + |\tau_k(y_i)|}, \frac{|u_{r_k+1}|}{2q_k + |\tau_k(y_i)| - \operatorname{val}(c_1 \cdots c_k)}$$

as $k \to \infty$. Here $\Delta' = T^{r_{k+1}}(\Delta) = T^{2(k+1)}(\Delta)$, so the words Δ and Δ' are isomorphic. Therefore the Ostrowski numeration systems associated with Δ and Δ' are the same. It follows that $\operatorname{val}_{\Delta'}(c_{k+2} \cdots c_{k+d-1}) = \operatorname{val}_\Delta(1) = 1$ meaning that y_i is the second letter of the word $\mathbf{c}_{\Delta'}$, that is,

$y_i = x_{k+2}$. Thus $\tau_k(y_i) = s_{k-1}^2 s_{k-2}$. The numbers q_i satisfy the linear recurrence $q_i = 2q_{i-1} + 2q_{i-2} + q_{i-3}$, so it follows from the theory of linear recurrences that $q_{k+\ell}/q_k \to \beta^\ell$ as $k \to \infty$.

Let us compute the last limit. From (5), we find that $|u_{r_{k+1}}| - |u_{r_k}| = q_k + q_{k-1}$. In addition, we have $2q_k + |\tau_k(y_i)| = q_{k+1}$ and $\mathrm{val}(1^k) - \mathrm{val}(1^{k-1}) = q_{k-1}$. Therefore the Stolz-Cesàro Theorem implies that

$$\begin{aligned}\lim_{k\to\infty} \frac{|u_{r_{k+1}}|}{2q_k + |\tau_k(y_i)| - \mathrm{val}(c_1\cdots c_k)} &= \lim_{k\to\infty} \frac{q_k + q_{k-1}}{q_{k+1} - q_k - q_{k-1}} \\ &= \lim_{k\to\infty} \frac{q_k + q_{k-1}}{q_k + q_{k-1} + q_{k-2}} \\ &= \frac{\beta^2 + \beta}{\beta^2 + \beta + 1} \\ &= \frac{1}{2}(\beta - 1) \\ &\approx 0.9156.\end{aligned}$$

Therefore $\mathrm{dio}(\mathbf{t})$ is at least as large as claimed. It thus suffices to show that the other two limits do not exceed this value. For the first ratio, we find like above that

$$\lim_{k\to\infty} \frac{|u_{r_k+1}|}{q_k + |\tau_k(y_i)| - \mathrm{val}(c_1\cdots c_k)} = \lim_{k\to\infty} \frac{2q_{k-1}}{2q_{k-1} + q_{k-2}} = \frac{2\beta}{2\beta+1} \approx 0.8499$$

and, for the second ratio, we have

$$\lim_{k\to\infty} \frac{|u_{r_k+1}| + q_k - \mathrm{val}(c_1\cdots c_k)}{q_k + |\tau_k(y_i)|} = \lim_{k\to\infty} \frac{q_k}{3q_{k-1} + q_{k-2}} = \frac{\beta}{3\beta+1} \approx 0.8443. \quad \square$$

Notice that Proposition 6.11 shows that the Diophantine exponent can be less than that of the corresponding standard word. The next proposition demonstrates that Lemma 6.10 does not generalize to $d > 3$. Therefore the assumptions of Theorem 6.8 are necessary.

Proposition 6.12. *Let* $\mathbf{t}$ *be the episturmian word with directive word* $(0123)^\omega$ *having intercept* $(001)^\omega$, $(010)^\omega$, *or* $(100)^\omega$. *Then*

$$\mathrm{dio}(\mathbf{t}) = 1 + \frac{1}{27}(-7\zeta_4^3 + 15\zeta_4^2 + 13\zeta_4 - 4) \approx 1.9873$$

where ζ_4*, approximately 1.9276, is the positive real root of the polynomial* $x^4 - x^3 - x^2 - x - 1$.

Proof. Assume that the intercept $c_1c_2\cdots$ equals $(001)^\omega$. Suppose that k is such that $k \geq 3$ and $c_{k+1} = 1$. It follows from Theorem 5.10 (ii) that $\mathrm{dio}(\mathbf{t}) - 1$ is it least as large as the limits of

$$\frac{|u_{r_{k+1}}| - \mathrm{val}(c_1\cdots c_k)}{|\tau_k(y_i)|} \quad \text{and} \quad \frac{|u_{r_{k+1}}|}{q_k + |\tau_k(y_i)| - \mathrm{val}(c_1\cdots c_k)}$$

as $k \to \infty$ (along appropriate subsequences). Now $c_{k+2}\cdots c_{k+d-2} = 00$, so y_i is the first letter of the epistandard word with intercept $x_{k+2}x_{k+3}\cdots$, that is, $y_i = x_{k+2}$ (notice that the word Δ' of Theorem 5.10 is isomorphic to the directive word $(0123)^\omega$, so the numeration systems associated to both directive words are the same). It follows that $\tau_k(y_i) = s_{k-1}s_{k-2}s_{k-3}$. The numbers q_i satisfy the linear recurrence $q_i = q_{i-1} + q_{i-2} + q_{i-3} + q_{i-4}$, so it follows from the theory of linear recurrences that $q_{k+\ell}/q_k \to \zeta_4^\ell$ as $k \to \infty$.

From (7), we find that $|u_{r_{k+1}}| - |u_{r_{k-2}}| = q_k - q_{k-4}$. In addition, $\mathrm{val}(c_1\cdots c_k) - \mathrm{val}(c_1\cdots c_{k-3}) = q_{k-3}$. Using the Stolz-Cesàro Theorem, the limit of the first ratio is found as follows:

$$\begin{aligned}\lim_{k\to\infty} \frac{|u_{r_{k+1}}| - \mathrm{val}(c_1\cdots c_k)}{|\tau_k(y_i)|} &= \lim_{k\to\infty} \frac{|u_{r_{k+1}}| - |u_{r_{k-2}}| - (\mathrm{val}(c_1\cdots c_k) - \mathrm{val}(c_1\cdots c_{k-3}))}{|\tau_k(y_i)| - |\tau_{k-3}(y_i)|} \\ &= \lim_{k\to\infty} \frac{q_{k-1} + q_{k-2}}{q_{k-1} + q_{k-2} + q_{k-7}}\end{aligned}$$

$$\begin{aligned}&= \frac{\zeta_4^6 + \zeta_4^5}{\zeta_4^6 + \zeta_4^5 + 1} \\ &= \frac{1}{27}(-7\zeta_4^3 + 15\zeta_4^2 + 13\zeta_4 - 4) \\ &\approx 0.9873.\end{aligned}$$

Therefore $\mathrm{dio}(\mathbf{t})$ is at least as large as claimed. It thus suffices to show that the analogous limits do not exceed this value in the other cases. For the latter ratio, we find that

$$\lim_{k\to\infty} \frac{|u_{r_{k+1}}|}{q_k + |\tau_k(y_i)| - \mathrm{val}(c_1 \cdots c_k)} = \lim_{k\to\infty} \frac{q_k - q_{k-4}}{q_k + q_{k-1} + q_{k-2} - 2q_{k-3} + q_{k-7}} \approx 0.6107.$$

Assume then that $c_{k+1} = 0$ and $c_k = 1$. By Theorem 5.10 (iv.b), we need to find the limit of

$$\frac{|u_{r_{k+1}}|}{|\tau_{k+1}(y_i)| - \mathrm{val}(c_1 \cdots c_k)}.$$

Now $c_{k+2} \cdots c_{k+d-1} = 01$ and $\mathrm{val}(c_{k+2} \cdots c_{k+d-1}) = 2$, so again $y_i = x_{k+2}$ and $|\tau_{k+1}(y_i)| = q_{k+1}$. Moreover, we have $\mathrm{val}(c_1 \cdots c_k) - \mathrm{val}(c_1 \cdots c_{k-2}) = q_{k-1}$. Hence

$$\lim_{k\to\infty} \frac{|u_{r_{k+1}}|}{|\tau_{k+1}(y_i)| - \mathrm{val}(c_1 \cdots c_k)} = \lim_{k\to\infty} \frac{q_k - q_{k-4}}{q_{k+1} - q_{k-2} - q_{k-1}} \approx 0.8139.$$

Suppose finally that $c_{k+1} = c_k = 0$. By Theorem 5.10 (iv.a), we need to consider the ratios

$$\frac{|u_{r_{k+1}}| - \mathrm{val}(c_1 \cdots c_{k-1})}{q_k} \quad \text{and} \quad \frac{|u_{r_{k+1}}|}{|\tau_{k+1}(y_i)| - \mathrm{val}(c_1 \cdots c_{k-1})}.$$

This time around $c_{k+2} \cdots c_{k+d-1} = 10$, $\mathrm{val}(c_{k+2} \cdots c_{k+d-1}) = 1$, $y_i = x_{k+3}$, $\tau_{k+1}(y_i) = s_k s_{k-1} s_{k-2}$, and $\mathrm{val}(c_1 \cdots c_k) - \mathrm{val}(c_1 \cdots c_{k-2}) = q_{k-2}$. Proceeding as above, it is straightforward to show that the limits are approximately 0.7653 and 0.7309. This proves the claim. It is straightforward to check that the intercepts $(010)^\omega$ and $(100)^\omega$ lead to exactly the same result. □

The intercepts $(001)^\omega$, $(010)^\omega$, and $(100)^\omega$ are not the only interesting ones for words in the 4-bonacci subshift. It can be computed (like in the proof of Proposition 6.12) that the words with intercepts $(01)^\omega$ or $(10)^\omega$ have Diophantine exponent 2. The word with intercept $(011)^\omega$ and the words with conjugate intercepts $(110)^\omega$ and $(101)^\omega$ have Diophantine exponent $1+\zeta_4^2-\zeta_4 \approx 2.7879$. The intercepts $(0001)^\omega$ and $(0011)^\omega$ and their conjugate intercepts give Diophantine exponent $\mathrm{dio}(\mathbf{c}_{(0123)^\omega}) \approx 2.0781$. We have not attempted to compute Diophantine exponents for aperiodic intercepts. Computer experiments suggest that the Diophantine exponent of the word in the 5-bonacci subshift with intercept $(001)^\omega$ is approximately 1.9148 and approximately 1.8535 for the intercept $(01)^\omega$. We believe that similar results are obtained for d-bonacci words for $d > 4$ and for words that satisfy $\limsup_k a_k = 1$.

Since $\mathrm{ice}(\mathbf{x}) \leq \mathrm{dio}(\mathbf{x})$ for any infinite word $\mathbf{x}$, Proposition 6.11 has the following remarkable consequence.

Corollary 6.13. *There exists an episturmian word over a 3-letter alphabet having only finitely many square prefixes.*

This is indeed unexpected since every Sturmian word and every regular epistandard word has arbitrarily long square prefixes [18, Lemma 6.5]. We expect that a regular episturmian word has infinitely many square prefixes when $\limsup_{k\to\infty} a_k \geq 3$, but we have not attempted to prove this. We also expect that every infinite word in the Tribonacci subshift has arbitrarily long square prefixes.

6.2. Diophantine exponents of d-bonacci words

Here we prove more explicit results on the Diophantine exponents of d-bonacci words.

Definition 6.14. The *d-bonacci constant* ζ_d is the positive real root of the polynomial $x^d - x^{d-1} - \cdots - x - 1$.

Proposition 6.15. *For the d-bonacci word $\mathbf{f}_d$ with directive word $(01\cdots(d-1))^\omega$ and intercept 0^ω, we have* $\mathrm{dio}(\mathbf{f}_d) = \mathrm{ice}(\mathbf{f}_d) = \mathrm{ind}(\mathbf{f}_d) - 1 = 1 + 1/(\zeta_d - 1)$.

Proof. We deduce from Theorem 5.10 (iv.a) and the Stolz-Cesàro Theorem that

$$\mathrm{dio}(\mathbf{f}_d) - 1 = \lim_{k\to\infty} \frac{|u_{r_k+1}|}{q_k} = \lim_{k\to\infty} \frac{q_k}{q_k - q_{k-1}} = \frac{1}{\zeta_d - 1}. \quad \square$$

In particular, we have $\mathrm{dio}(\mathbf{f}_2) \approx 2.6180$, $\mathrm{dio}(\mathbf{f}_3) \approx 2.1915$, and $\mathrm{dio}(\mathbf{f}_4) \approx 2.0781$. Moreover, we have $\mathrm{dio}(\mathbf{f}_d) \to 2$ as $d \to \infty$ since it is well-known that $\zeta_d \to 2$ as $d \to \infty$ (see, e.g., [21]).

Lemma 6.16. *Let $\mathbf{t}$ be a word in the d-bonacci subshift with intercept $c_1c_2\cdots$ that is not in the shift orbit of $\mathbf{f}_d$. If 0^d occurs infinitely many times in $c_1c_2\cdots$, then*

$$\mathrm{dio}(\mathbf{t}) \geq 2 + \frac{1}{\zeta_d^d - \zeta_d} > \mathrm{dio}(\mathbf{f}_d).$$

If 0^ℓ occurs in $c_1c_2\cdots$ for arbitrarily large ℓ, then

$$\mathrm{dio}(\mathbf{t}) \geq 1 + \frac{\zeta_d^d}{(\zeta_d - 1)(\zeta_d^d - 1)}.$$

Proof. Since $\mathbf{t}$ is not in the shift orbit of $\mathbf{f}_d$, the intercept does not end with 0^ω. Therefore it contains $0^\ell 1$ infinitely many times for some ℓ. Let k be such that $c_{k+1} = 1$ and $c_k = \cdots = c_{k-(\ell-1)} = 0$. Using Lemma 5.9, we see that

$$\begin{aligned}\frac{|u_{r_k+1}| - \mathrm{val}(c_1\cdots c_k)}{|\tau_k(y_i)|} &\geq \frac{|s_{k-1}\cdots s_0| - \mathrm{val}(c_1\cdots c_{k-\ell})}{|s_{k-1}\cdots s_{k-(d-1)}|} \\ &= \frac{|s_{k-1}\cdots s_0| - \mathrm{val}(c_1\cdots c_{k-\ell})}{q_k - q_{k-d}}. \end{aligned} \tag{25}$$

Now $-\mathrm{val}(c_1\cdots c_{k-\ell})/(q_k - q_{k-d}) \geq -q_{k-\ell}/(q_k - q_{k-d}) \xrightarrow{k\to\infty} -\zeta_d^{-\ell}/(1 - \zeta_d^{-d})$, so if ℓ can be taken arbitrarily large, we have from Theorem 5.10 (ii) that

$$\mathrm{dio}(\mathbf{t}) - 1 \geq \lim_{k\to\infty} \frac{|s_{k-1}\cdots s_0|}{q_k - q_{k-d}} = \lim_{k\to\infty} \frac{q_k + |u_{r_{k-d+1}}|}{q_k - q_{k-d}} = \frac{\mathrm{dio}(\mathbf{f}_d) - 1}{1 - \zeta_d^{-d}}$$

where the last equality follows from the proof of Proposition 6.5. The latter claim now follows from Proposition 6.15 and simplification. If $\ell \geq d$, then we deduce from (25) that

$$\frac{|u_{r_k+1}| - \mathrm{val}(c_1\cdots c_k)}{|\tau_k(y_i)|} \geq \frac{|s_{k-1}\cdots s_{k-(d-1)}| + |s_{k-(d+1)}\cdots s_0|}{q_k - q_{k-d}} = 1 + \frac{|u_{r_{k-d+1}}|}{q_k - q_{k-d}},$$

so, like above,

$$\mathrm{dio}(\mathbf{t}) - 1 \geq 1 + \frac{\mathrm{dio}(\mathbf{f}_d) - 2}{1 - \zeta_d^{-d}} = 1 + \frac{1}{\zeta_d^d - \zeta_d}.$$

It is easy to check that this lower bound is larger than $\mathrm{dio}(\mathbf{f}_d)$. Thus the first claim is proved. $\square$

We believe that there always exists a word in the subshift of the d-bonacci word such that $\mathrm{dio}(\mathbf{t}) = 2 + 1/(\zeta_d^d - \zeta_d)$, but we have not attempted to prove this rigorously.

Corollary 6.17. *Let* $\mathbf{t}$ *be a word in the Fibonacci subshift with intercept* $c_1c_2\cdots$. *If* $\mathbf{t}$ *is in the shift orbit of the standard word* $\mathbf{c}_\Delta$, *then* $\operatorname{dio}(\mathbf{t}) = \operatorname{ice}(\mathbf{t}) = 1 + \zeta_2$. *Otherwise* $\operatorname{dio}(\mathbf{t}), \operatorname{ice}(\mathbf{t}) \in [3, 2+\zeta_2]$. *Moreover, if* $c_1c_2\cdots$ *contains* $10^\ell 1$ *for arbitrarily large* ℓ, *then* $\operatorname{dio}(\mathbf{t}) = \operatorname{ice}(\mathbf{t}) = 2 + \zeta_2 = \operatorname{ind}(\mathbf{t})$.

Proof. If 0^2 occurs infinitely many times in $c_1c_2\cdots$, then Lemma 6.16 implies that $\operatorname{dio}(\mathbf{t}) \geq 3$. If 0^2 does not occur in $c_1c_2\cdots$ infinitely many times, then the intercept ends with 1^ω or $(01)^\omega$. The former case is impossible since the intercept satisfies the Ostrowski conditions by Lemma 3.10. It follows from Lemma 3.22 that

$$\mathbf{t} = T^{c_1}L_{x_1}^{a_1} \circ \cdots \circ T^{c_k}L_{x_k}^{a_k}(y\mathbf{c}_{\Delta'}) = T^{c_1}L_{x_1}^{a_1} \circ \cdots \circ T^{c_k}L_{x_k}^{a_k}(y)\tau_k(\mathbf{c}_{\Delta'})$$

for some integer k, letter y, and $\Delta' = T^k(\Delta)$. Clearly $\tau_k(\mathbf{c}_{\Delta'})$ is an episturmian word with directive word $(01)^\omega$ and intercept 0^ω, so $\tau_k(\mathbf{c}_{\Delta'}) = \mathbf{f}_2$. Therefore $\mathbf{t} = w\mathbf{f}_2$ for some finite word w. It follows from Propositions 6.4 and 6.15 that $\operatorname{dio}(\mathbf{t}) = \operatorname{dio}(\mathbf{f}_2) = 1+\zeta_2$. The last claim follows from Lemma 6.16 because $1 + \zeta_2^2/((\zeta_2 - 1)(\zeta_2^2 - 1)) = \operatorname{ind}(\mathbf{f}_2) = 2 + \zeta_2$. The claim for $\operatorname{ice}(\mathbf{t})$ was established in the proof of [10, Prop. 4.3]. □

We can prove a result like the preceding corollary for the Tribonacci subshift, but this result does not anymore generalize to $d > 3$ as indicated by Proposition 6.12.

Proposition 6.18. *If* $\mathbf{t}$ *is a word in the Tribonacci subshift, then* $\operatorname{dio}(\mathbf{f}_3) \leq \operatorname{dio}(\mathbf{t}) \leq \operatorname{ind}(\mathbf{f}_3)$.

Proof. Say $\mathbf{t}$ has intercept $c_1c_2\cdots$. If the intercept has suffix 0^ω, then $\operatorname{dio}(\mathbf{t}) = \operatorname{dio}(\mathbf{f}_3)$, so we may assume that 1 occurs infinitely often in $c_1c_2\cdots$. By Lemma 6.16, the claim is clear if the intercept contains 0^3 infinitely many times.

Suppose that 0011 occurs infinitely often in $c_1c_2\cdots$. Then there exists infinitely many k such that $c_{k+2} = c_{k+1} = 1$ and $c_k = c_{k-1} = 0$. Then the first implication of Theorem 5.10 (ii) yields

$$\operatorname{dio}(\mathbf{t}) - 1 \geq \limsup_{k\to\infty} \frac{|u_{r_k+1}| - \operatorname{val}(c_1\cdots c_k)}{|\tau_k(y_i)|}.$$

Since $c_{k+2} = 1$, we have $y_i \neq x_{k+2}$, so $|\tau_k(y_i)| = q_{k-1}$ by Lemma 5.8. Therefore

$$\frac{|u_{r_k+1}| - \operatorname{val}(c_1\cdots c_k)}{|\tau_k(y_i)|} = \frac{|u_{r_k+1}| - \operatorname{val}(c_1\cdots c_{k-2})}{q_{k-1}} \geq 1 + \frac{q_{k-3}}{q_{k-1}} \xrightarrow{k\to\infty} 1 + \zeta_3^{-2}.$$

Hence $\operatorname{dio}(\mathbf{t}) \geq 2 + \zeta_3^{-2} > \operatorname{dio}(\mathbf{f}_3)$. We may thus assume that eventually all occurrences of 11 in $c_1c_2\cdots$ are preceded by 10. Assume that 1101 occurs infinitely many times in $c_1c_2\cdots$. Letting $k+1$ to correspond to the third letter of this pattern, we have from Theorem 5.10 (iv.b) that

$$\operatorname{dio}(\mathbf{t}) - 1 \geq \limsup_{k\to\infty} \frac{|u_{r_{k+1}}|}{|\tau_{k+1}(y_i)| - \operatorname{val}(c_1\cdots c_k)}.$$

Again $c_{k+2} = 1$, so $|\tau_{k+1}(y_i)| = q_k + q_{k-1}$. Using the fact that 1101 is preceded by 10, we have

$$\begin{aligned}
\frac{|u_{r_{k+1}}|}{|\tau_{k+1}(y_i)| - \operatorname{val}(c_1\cdots c_k)} &= \frac{|u_{r_{k+1}}|}{q_k + q_{k-1} - q_{k-1} - q_{k-2} - q_{k-4} - \operatorname{val}(c_1\cdots c_{k-4})} \\
&\geq \frac{|u_{r_{k+1}}|}{q_k - q_{k-2} - q_{k-4}} \\
&\geq \frac{q_k}{q_k - q_{k-2} - q_{k-4}} \\
&\xrightarrow{k\to\infty} \frac{1}{1 - \zeta_3^{-2} - \zeta_3^{-4}} \\
&\geq 1.6206.
\end{aligned}$$

Hence $\operatorname{dio}(\mathbf{t}) \geq 2.6206 > \operatorname{dio}(\mathbf{f}_3)$. We may thus assume that 11 is eventually always followed by 00. Suppose then that 100101 occurs infinitely many times in the intercept. Letting $k+1$ to correspond to the fifth letter of this pattern, we have from Theorem 5.10 (iv.b) that

$$\mathrm{dio}(\mathbf{t}) - 1 \geq \limsup_{k\to\infty} \frac{|u_{r_{k+1}}|}{|\tau_{k+1}(y_i)| - \mathrm{val}(c_1 \cdots c_k)}.$$

Now $|\tau_{k+1}(y_i)| = q_k + q_{k-1}$ and so

$$\frac{|u_{r_{k+1}}|}{|\tau_{k+1}(y_i)| - \mathrm{val}(c_1 \cdots c_k)} \geq \frac{|u_{r_{k+1}}|}{q_k + q_{k-1} - q_{k-1} - q_{k-4}} \geq \frac{q_k + q_{k-4}}{q_k - q_{k-4}} \xrightarrow{k\to\infty} \mathrm{dio}(\mathbf{f}_3) - 1,$$

meaning that $\mathrm{dio}(\mathbf{t}) \geq \mathrm{dio}(\mathbf{f}_3)$. We may thus assume that 100 is eventually always followed by 100. We may now prove that the intercept contains only finitely many occurrences of 11. Indeed, the preceding assumptions imply that late enough occurrences of 11 must be followed by 00 and each late enough occurrence of 100 must be followed by 100. Thus $c_1c_2\cdots$ has suffix $(100)^\omega$ and 11 occurs only finitely many times.

We have now argued that the claim holds or the intercept is eventually a product of the words 10 and 100. Moreover, the pattern 100101 does not eventually occur, so the intercept has suffix $(10)^\omega$ or $(100)^\omega$. Say it has suffix $(10)^\omega$. Let k be large enough such that $c_{k+1} = c_{k-1} = 0$ and $c_{k+2} = c_k = c_{k-2} = 1$. Again $|\tau_{k+1}(y_i)| = q_k + q_{k-1}$, and

$$\begin{aligned}\frac{|u_{r_{k+1}}|}{|\tau_{k+1}(y_i)| - \mathrm{val}(c_1 \cdots c_k)} &= \frac{|u_{r_{k+1}}|}{q_k + q_{k-1} - q_{k-1} - q_{k-3} - \mathrm{val}(c_1 \cdots c_{k-3})} \\ &\geq \frac{|s_{k-1} \cdots s_0|}{q_{k-1} + q_{k-2}} \\ &\geq \frac{q_{k-1} + q_{k-2} + q_{k-3}}{q_{k-1} + q_{k-2}} \\ &\xrightarrow{k\to\infty} \mathrm{dio}(\mathbf{f}_3) - 1,\end{aligned}$$

so $\mathrm{dio}(\mathbf{t}) \geq \mathrm{dio}(\mathbf{f}_3)$ by Theorem 5.10 (iv.b).

We can now assume that the intercept has suffix $(100)^\omega$. Since the Diophantine exponent is shift-invariant, we may focus on the intercept $(100)^\omega$ and its conjugate intercepts $(010)^\omega$, and $(001)^\omega$.

Let us first consider the intercept $(100)^\omega$. Let k be such that $c_{k+1} = 1$. Here $|\tau_k(y_i)| = q_{k-1} + q_{k-2}$, $|u_{r_k+1}| - |u_{r_{k-3}+1}| = q_k$, and $\mathrm{val}(c_1 \cdots c_k) - \mathrm{val}(c_1 \cdots c_{k-3}) = q_{k-3}$, so we find with the help of the Stolz-Cesàro Theorem that (along an appropriate subsequence)

$$\lim_{k\to\infty} \frac{|u_{r_k+1}| - \mathrm{val}(c_1 \cdots c_k)}{|\tau_k(y_i)|} = \lim_{k\to\infty} \frac{q_{k-1} + q_{k-2}}{q_{k-1} + q_{k-3}} = \frac{\zeta_3^2 + \zeta_3}{\zeta_3^2 + 1} = \frac{1}{\zeta_3 - 1},$$

so $\mathrm{dio}(\mathbf{t}) \geq \mathrm{dio}(\mathbf{f}_3)$ by Theorem 5.10 (ii). In fact, it can be showed that $\mathrm{dio}(\mathbf{t}) = \mathrm{dio}(\mathbf{f}_3)$. The remaining cases lead to the same result as is straightforward to show. □

7. Irrationality exponents

The results of the previous section on Diophantine exponents allows us to obtain novel results on irrationality exponents of numbers whose fractional parts are regular episturmian.

Definition 7.1. The *irrationality exponent* $\mu(\xi)$ of a real number ξ is the supremum of the real numbers ρ such that the inequality

$$\left|\xi - \frac{p}{q}\right| < \frac{1}{q^\rho}$$

has infinitely many rational solutions p/q. If $\mu(\xi) = \infty$, then we say that ξ is a *Liouville number*.

Recall that $\mu(\xi) \geq 2$ whenever ξ is irrational and that $\mu(\xi) = 2$ for almost all real numbers ξ with respect to the Lebesgue measure. Roth's theorem states that $\mu(\xi) = 2$ when ξ is algebraic.

Definition 7.2. Let b be an integer such that $b \geq 2$ and $\mathbf{x}$ be an infinite word over the alphabet $\{0, 1, \ldots, b-1\}$ such that $\mathbf{x} = x_1 x_2 \cdots$. Then $\xi_{\mathbf{x},b}$ is the real number $\sum_{k \geq 1} x_k / b^k$.

The following result is the key result linking combinatorial properties of $\mathbf{x}$ to the Diophantine properties of $\xi_{\mathbf{x},b}$. For its proof, see, e.g., [1, Sect. 3] and [4, Thm. 2.1]. The function p of the statement is the factor complexity function that counts the number of factors of length n, that is, $p(\mathbf{x}, n) = |\mathcal{L}_{\mathbf{x}}|$ for all n.

Proposition 7.3. *Let b be an integer such that $b \geq 2$ and $\mathbf{x}$ be an aperiodic infinite word. Then $\mu(\xi_{\mathbf{x},b}) \geq \mathrm{dio}(\mathbf{x})$. If, moreover, there exists an integer K such that $p(\mathbf{x}, n) \leq Kn$ for all n large enough, then $\mu(\xi_{\mathbf{x},b}) \leq (2K+1)^3(\mathrm{dio}(\mathbf{x}) + 1)$.*

Showing that $\mu(\xi_{\mathbf{x},b}) > 2$ implies that $\xi_{\mathbf{x},b}$ is transcendental by Roth's theorem, so computing Diophantine exponents of infinite words can be used to show transcendence results. While the following results allow us to conclude the transcendence of certain numbers, these facts are not new. One of the main results of [3] states that if $\mathbf{x}$ is aperiodic and has sublinear factor complexity, then $\xi_{\mathbf{x},b}$ is transcendental. Since episturmian words have sublinear factor complexity [16, Thm. 7], it follows that $\xi_{\mathbf{t},b}$ is transcendental for an *arbitrary* aperiodic episturmian word $\mathbf{t}$.

Theorem 6.8 together with Proposition 7.3 directly implies the following result. The result was obtained in [1] when $d = 2$ based on the results of [10].

Theorem 7.4. *Let $\mathbf{t}$ be a regular episturmian word of period d with directive word as in (2). If $d = 2$ or $\limsup_k a_k \geq 3$, then $\mu(\xi_{\mathbf{t},b}) > 2$.*

Hence we have identified a new class of transcendental numbers whose irrationality exponents are greater than 2. This class is uncountable because the set of directive words in the statement is uncountable and an episturmian word has a unique directive word.

The next theorem identifies a new uncountable class of Liouville numbers. The case $d = 2$ was first established by Komatsu [25].

Theorem 7.5. *Let $\mathbf{t}$ be a regular episturmian word. Then $\xi_{\mathbf{t},b}$ is a Liouville number if and only if $\mathbf{t}$ has unbounded partial quotients.*

Proof. If $\mathbf{t}$ has unbounded partial quotients, then $\mathrm{dio}(\mathbf{t}) = \infty$ by Theorem 6.6 and $\xi_{\mathbf{t},b}$ is a Liouville number by Proposition 7.3. If $\mathbf{t}$ has bounded partial quotients, then $\mathrm{dio}(\mathbf{t}) < \infty$ by Theorem 6.6, and Proposition 7.3 yields a finite upper bound for $\mu(\xi_{\mathbf{t},b})$ since a regular episturmian word over d letters has $(d-1)n + 1$ factors of length n for all n [16, Thm. 7]. □

Adamczewski and Cassaigne established in [5] that $\xi_{\mathbf{x},b}$ is not a Liouville number if $\mathbf{x}$ is a k-automatic word. Theorem 7.5 establishes the following weak analogue of this result: if $\mathbf{t}$ is a regular epistandard word (or its shift) and the sequence (q_k) satisfies a linear recurrence, then $\xi_{\mathbf{x},b}$ is not a Liouville number. Indeed, when the sequence (q_k) satisfies a linear recurrence, the corresponding Ostrowski numeration system can be viewed as a positional numeration system and the corresponding standard word is automatic with respect to this numeration system. For more on Ostrowski-automatic words, see the recent paper [9]. For a general reference on these topics, see, e.g., [33, Ch. 2].

In [13], Bugeaud and Kim prove the following remarkable result.

Proposition 7.6 *([13, Thm. 4.5]).* *If $\mathbf{t}$ is a Sturmian word, then $\mu(\xi_{\mathbf{t},b}) = \mathrm{dio}(\mathbf{t})$.*

This result states that the irrationality exponent of $\xi_{\mathbf{t},b}$ can be read off its base-b expansion for a Sturmian word $\mathbf{t}$. For most real numbers, this is not possible [13, p. 3287]. Since many properties of Sturmian words transfer to all episturmian words, it is natural to wonder if this result can be generalized. As a consequence to Proposition 6.12, we obtain a negative answer to this question.

Proposition 7.7. *There exists an episturmian word $\mathbf{t}$ over a 3-letter alphabet such that $\mu(\xi_{\mathbf{t},b}) > \mathrm{dio}(\mathbf{t})$.*

Proof. Let $\mathbf{t}$ be the episturmian word with directive word $(001122)^\omega$ and intercept 1^ω. Since $\mathbf{t}$ is aperiodic, the number $\xi_{\mathbf{t},b}$ is irrational, and hence $\mu(\xi_{\mathbf{t},b}) \geq 2$. On the other hand, we showed in Proposition 6.11 that $\text{dio}(\mathbf{t}) < 2$. The claim follows. □

This leaves open if the conclusion of Proposition 7.6 is true for regular episturmian words satisfying the assumptions of Theorem 7.4. We see no reason to believe this as the proof of Proposition 7.6 relies heavily on the theory of continued fractions and no such theory is available for general episturmian words.

Bugeaud and Kim discuss on p. 3288 of [13] the base-b expansions of the numbers of the form $\log(1 + 1/a)$. Their absolute lower bound for the Diophantine exponents of Sturmian words implies that the fractional part of the base-b expansion of $\log(1+1/a)$ cannot be a Sturmian word for $a \geq 34$. This result is a direct consequence of [7, Cor. 1] which provides an upper bound for the irrationality exponent of $\log(1 + 1/a)$ which tends to 2 as $a \to \infty$. Applying this corollary to Proposition 6.18 yields the following.

Corollary 7.8. *For every integer b such that $b \geq 2$ and every integer $a \geq 23347$, the fractional part of the base-b expansion of* $\log(1 + 1/a)$ *is not isomorphic to any word in the Tribonacci subshift.*

Results like this can be generated by finding results on the irrationality exponents of specific numbers. Since the irrationality exponent of e is 2 (see, e.g., the proof of [1, Cor. 2]), we obtain the following corollary. While this result is very minor, it seems that, besides the results of [1], very little is known about the base-b expansions of e.

Corollary 7.9. *Let* $\mathbf{t}$ *be a regular episturmian word of period d with directive word as in (2) such that $d = 2$ or $\limsup_k a_k \geq 3$. For every integer b such that $b \geq 2$, the fractional part of the base-b expansion of e is not isomorphic to* $\mathbf{t}$.

Obviously we conjecture that the conclusion is true when $d > 2$. We obtain the same conclusion if we replace the number e by a badly approximable number. This observation is rather interesting since there exist badly approximable numbers of sublinear complexity; see [12, Sect. 8.5].

8. Open problems

We have characterized the initial nonrepetitive complexity of regular episturmian words in Theorem 5.10. It is unclear how this result can be generalized to all episturmian words.

Problem 8.1. *Characterize the initial nonrepetitive complexity of a general episturmian word.*

Such a characterization would allow to determine the Diophantine exponent of a general episturmian word, but it is possible that this problem can be attacked in some other way.

Problem 8.2. *Determine the Diophantine exponent of an episturmian word. Determine conditions that ensure that the Diophantine exponent of an episturmian word is strictly greater than* 2.

The Rauzy graph approach should work for studying the prefix nonrepetitive complexity function and the initial critical exponent. However, we find this more difficult than characterizing initial nonrepetitive complexity because we need to detect returning to the initial vertex in the Rauzy graph, not a return to some vertex.

Problem 8.3. *Adjust the methods of Section 5 for the study of prefix nonrepetitive complexity function of regular episturmian words.*

We believe Theorem 6.6 generalizes to all aperiodic episturmian words.

Conjecture 8.4. *Let* $\mathbf{t}$ *be an aperiodic episturmian word. Then* $\text{dio}(\mathbf{t}) < \infty$ *if and only if* $\mathbf{t}$ *has bounded partial quotients.*

Let X_d be the set of regular episturmian words with period d. Define

$$\mathcal{B}(d) = \{\mathrm{dio}(\mathbf{t}) : \mathbf{t} \in X_d\} \quad \text{and}$$
$$\mathcal{K}(d) = \{\mathrm{dio}(\mathbf{t}) : \mathbf{t} \in X_d \text{ and } \limsup_k a_k \geq 3 \text{ or } d = 2\}.$$

It is proved in [13, Thm. 4.3] that $\mathcal{B}(2) \geq \frac{5}{3} + \frac{4\sqrt{10}}{15}$ and that this lower bound is optimal. This is a remarkable result, and we propose the following problems and questions.

Problem 8.5. *Reprove that $\mathcal{B}(2) \geq \frac{5}{3} + \frac{4\sqrt{10}}{15}$ using Theorem 5.10. Find and prove an optimal lower bound for $\mathcal{K}(d)$ when $d > 2$.*

Question 8.6. *What is the least element of $\mathcal{K}(d)$ and what is the least accumulation point of $\mathcal{K}(d)$? Is it true that $\inf \mathcal{B}(d) = 1$ when $d > 2$?*

Theorem 6.8 was not enough to conclude that $\mu(\xi_{\mathbf{t},b}) > 2$ for all regular episturmian words $\mathbf{t}$. It would be interesting to know if this is true.

Question 8.7. *Is it true that $\mu(\xi_{\mathbf{t},b}) > 2$ for all regular episturmian words $\mathbf{t}$? What about all aperiodic episturmian words?*

Our results provide lower bounds for $\mu(\xi_{\mathbf{t},b})$ when $\mathbf{t}$ is a regular episturmian word, and Proposition 7.7 suggests that the lower bounds are strict. Hence we propose the following problem.

Problem 8.8. *Find a better upper bound than that of Proposition 7.3 for the irrationality exponent of a regular episturmian word with bounded partial quotients.*

Finally we propose the following problem related to the discussion after Theorem 7.5. This is an analogue of a conjecture of Shallit [34] that was settled in [5].

Problem 8.9. *Develop rigorously the notion of an Ostrowski-automatic word in the setting $d > 2$ and settle the conjecture that if $\mathbf{t}$ is an Ostrowski-automatic word, then $\xi_{\mathbf{t},b}$ is not a Liouville number.*

Acknowledgments

We thank the reviewers for pointing out several problems in the original manuscript. Due to your help, these issues have been identified and fixed. We are especially thankful for Example 3.8 which indicates that not all Ostrowski expansions equal the corresponding greedy expansions.

References

[1] B. Adamczewski, On the expansion of some exponential periods in an integer base, Math. Ann. 346 (2010) 107–116, http://dx.doi.org/10.1007/s00208-009-0391-z.
[2] B. Adamczewski, Y. Bugeaud, Dynamics for β-shifts and diophantine approximation, Ergodic Theory Dynam. Systems 27 (6) (2007) 1695–1711, http://dx.doi.org/10.1017/S0143385707000223.
[3] B. Adamczewski, Y. Bugeaud, On the complexity of algebraic numbers I. Expansions in integer bases, Ann. of Math. 165 (2) (2007) 547–565, http://dx.doi.org/10.4007/annals.2007.165.547.
[4] B. Adamczewski, Y. Bugeaud, Nombres réels de complexité sous-linéaire: mesures d'irrationalité et de transcendance, J. Reine Ange. Math. 658 (2011) 65–98, http://dx.doi.org/10.1515/crelle.2011.061.
[5] B. Adamczewski, J. Cassaigne, Diophantine properties of real numbers generated by finite automata, Compos. Math. 142 (6) (2006) 1351–1372, http://dx.doi.org/10.1112/S0010437X06002247.
[6] B. Adamczewski, J. Cassaigne, M. Le Gonidec, On the computational complexity of algebraic numbers: the Hartmanis-Stearns problem revisited, Trans. Amer. Math. Soc. 373 (5) (2020) 3085–3115, http://dx.doi.org/10.1090/tran/7915.
[7] K. Alladi, M.L. Robinson, Legendre polynomials and irrationality, J. Reine Ange. Math. 1980 (318) (1980) 137–155, http://dx.doi.org/10.1515/crll.1980.318.137.
[8] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité $2n+1$, Bull. Soc. Math. France 119 (2) (1991) 199–215.
[9] A. Baranwal, L. Schaeffer, J. Shallit, Ostrowski-automatic sequences: Theory and applications, Theoret. Comput. Sci. 858 (2021) 122–142, http://dx.doi.org/10.1016/j.tcs.2021.01.018.

[10] V. Berthé, C. Holton, L.Q. Zamboni, Initial powers of Sturmian sequences, Acta Arith. 122 (4) (2006) 315–347, http://dx.doi.org/10.4064/aa122-4-1.
[11] T.C. Brown, Descriptions of the characteristic sequence of an irrational, Canad. Math. Bull. 36 (1) (1993) 15–21.
[12] Y. Bugeaud, Distribution Modulo One and Diophantine Approximation, Cambridge University Press, 2012, http://dx.doi.org/10.1017/CBO9781139017732.
[13] Y. Bugeaud, D.H. Kim, A new complexity function, repetitions in Sturmian words, and irrationality exponents of Sturmian numbers, Trans. Amer. Math. Soc. 371 (5) (2019) 3281–3308, http://dx.doi.org/10.1090/tran/7378.
[14] A. Carpi, A. de Luca, Special factors, periodicity, and an application to Sturmian words, Acta Inform. 36 (2000) 983–1006, http://dx.doi.org/10.1007/PL00013299.
[15] D. Damanik, D. Lenz, The index of Sturmian sequences, European J. Combin. 23 (2002) 23–29, http://dx.doi.org/10.1006/eujc.2000.0496.
[16] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, Theoret. Comput. Sci. 225 (2001) 539–553, http://dx.doi.org/10.1016/S0304-3975(99)00320-5.
[17] A. Dubickas, Binary words with a given diophantine exponent, Theoret. Comput. Sci. 410 (47–49) (2009) 5191–5195, http://dx.doi.org/10.1016/j.tcs.2009.08.013.
[18] A. Glen, Powers in a class of $\mathcal{A}$-strict standard episturmian words, Theoret. Comput. Sci. 380 (3) (2007) 330–354, http://dx.doi.org/10.1016/j.tcs.2007.03.023.
[19] A. Glen, J. Justin, Episturmian words: A survey, RAIRO - Theor. Inf. Appl. 43 (3) (2009) 403–442, http://dx.doi.org/10.1051/ita/2009003.
[20] A. Glen, F. Levé, G. Richomme, Quasiperiodic and Lyndon episturmian words, Theoret. Comput. Sci. 409 (2008) 578–600, http://dx.doi.org/10.1016/j.tcs.2008.09.056.
[21] K. Hare, H. Prodinger, J. Shallit, Three series for the generalized golden mean, Fibonacci Quart. 52 (4) (2014) 307–313.
[22] J. Justin, G. Pirillo, Fractional powers in Sturmian words, Theoret. Comput. Sci. 255 (2001) 363–376, http://dx.doi.org/10.1016/S0304-3975(99)90294-3.
[23] J. Justin, G. Pirillo, Episturmian words and episturmian morphisms, Theoret. Comput. Sci. 276 (2002) 281–313, http://dx.doi.org/10.1016/S0304-3975(01)00207-9.
[24] J. Justin, G. Pirillo, Episturmian words: Shifts, morphisms and numeration systems, Internat. J. Found Comput. Sci. 15 (2) (2004) 329–348, http://dx.doi.org/10.1142/S0129054104002455.
[25] T. Komatsu, A certain power series and the inhomogeneous continued fraction expansions, J. Number Theory 59 (2) (1996) 291–312, http://dx.doi.org/10.1006/jnth.1996.0099.
[26] M. Lothaire, Algebraic Combinatorics on Words, in: Encyclopedia of Mathematics and Its Applications, (90) Cambridge University Press, 2002.
[27] K. Medková, E. Pelantová, É. Vandomme, On non-repetitive complexity of Arnoux-Rauzy words, Discrete Appl. Math. 285 (2020) 423–433, http://dx.doi.org/10.1016/j.dam.2020.06.016.
[28] T.K.S. Moothathu, Eulerian entropy and non-repetitive subword complexity, Theoret. Comput. Sci. 420 (2012) 80–88, http://dx.doi.org/10.1016/j.tcs.2011.11.013.
[29] J. Nicholson, N. Rampersad, Initial non-repetitive complexity of infinite words, Discrete Appl. Math. 208 (2016) 114–122, http://dx.doi.org/10.1016/j.dam.2016.03.010.
[30] J. Peltomäki, Characterization of repetitions in Sturmian words: A new proof, Inform. Process. Lett. 115 (11) (2015) 886–891, http://dx.doi.org/10.1016/j.ipl.2015.05.011.
[31] J. Peltomäki, Privileged Words and Sturmian Words (Ph.D. thesis), (214) Turku Centre for Computer Science, University of Turku, Turku, Finland, 2016, URL http://urn.fi/URN:ISBN:978-952-12-3422-4.
[32] N. Pytheas Fogg, Substitutions in dynamics, arithmetics and combinatorics, in: Lecture Notes in Mathematics, (1794) Springer, 2002, http://dx.doi.org/10.1007/b13861.
[33] M. Rigo, Formal languages, automata and numeration systems 2, in: Applications to recognizability and decidability, ISTE, John Wiley & Sons, 2014, http://dx.doi.org/10.1002/9781119042853.
[34] J. Shallit, Number theory and formal languages, in: Emerging Applications of Number Theory, in: The IMA Volumes in Mathematics and its Applications, (109) Springer, New York, pp. 547–570, http://dx.doi.org/10.1007/978-1-4612-1544-8_22.
[35] C. Wojcik, Factorisations des mots de basse complexité (Ph.D. thesis), Université Claude Bernard Lyon 1, 2020.