ON THE IRRATIONALITY EXPONENT OF REAL NUMBERS WITH LOW COMPLEXITY EXPANSION

YANN BUGEAUD, HAJIME KANEKO, AND DONG HAN KIM

Abstract. Let ξ be a real number and $b \geq 2$ an integer. We study the relationship between the irrationality exponent of ξ and the subword complexity $p(n, \mathbf{x})$ of the b-ary expansion $\mathbf{x}$ of ξ, where $p(n, \mathbf{x})$ counts the number of distinct blocks of length n in $\mathbf{x}$, for $n \geq 1$. If the irrationality exponent of ξ is equal to 2, which is the case for almost all real numbers ξ, we show that the limit superior of the sequence $(p(n, \mathbf{x})/n)_{n \geq 1}$ is at least equal to $4/3$. The proof is based on a careful study of the evolution of the Rauzy graphs of infinite words of low complexity.

1. Introduction

Let ξ be an irrational real number. Its irrationality exponent $\mu(\xi)$ is the supremum of the real numbers μ for which the inequality

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

has infinitely many rational solutions p/q with $q \geq 1$. It follows from the theory of continued fractions that $\mu(\xi)$ is always at least equal to 2 and an easy covering argument shows that equality holds for almost all ξ with respect to the Lebesgue measure. Recently, Bugeaud and Kim [8] established a connection between the irrationality exponent of a real number and the complexity of its expansion in a given integer base. Before stating their result, let us introduce some notions from combinatorics on words.

Let $\mathcal{A}$ be a finite set called an alphabet and denote by $|\mathcal{A}|$ its cardinality. A word over $\mathcal{A}$ is a finite or infinite sequence of elements of $\mathcal{A}$. For a (finite or infinite) word $\mathbf{x} = x_1 x_2 \ldots$ written over $\mathcal{A}$, let $n \mapsto p(n, \mathbf{x})$ denote its subword complexity function which counts the number of different subwords of length n occurring in $\mathbf{x}$, that is,

$$p(n, \mathbf{x}) = \mathrm{Card}\{x_{j+1} x_{j+2} \ldots x_{j+n} : j \geq 0\}, \quad n \geq 1.$$

2020 *Mathematics Subject Classification.* 11A63, 11J82 (primary); 68R15 (secondary).

Key words and phrases. irrationality exponent, subword complexity, b-ary expansion.

Clearly, we have

$$1 \leq p(n, \mathbf{x}) \leq |\mathcal{A}|^n, \quad n \geq 1.$$

We set $p(0, \mathbf{x}) = 1$ for convenience. The Morse-Hedlund Theorem [15] states that if $\mathbf{x}$ is ultimately periodic, then there exists an integer C such that $p(n, \mathbf{x}) \leq C$ for $n \geq 0$; otherwise, we have

$$p(n+1, \mathbf{x}) \geq p(n, \mathbf{x}) + 1, \quad n \geq 0, \tag{1.1}$$

thus $p(n, \mathbf{x}) \geq n+1$ for $n \geq 0$. There exist uncountably many infinite words $\mathbf{s}$ over $\{0, 1\}$ such that $p(n, \mathbf{s}) = n+1$ for $n \geq 0$. These words are called Sturmian words. Classical references on combinatorics on words and on Sturmian sequences include [6,12,14]. The words $\mathbf{x}$ for which there exists an integer C such that $p(n, \mathbf{x}) \leq n+C$ for $n \geq 0$ are called quasi-Sturmian.

Let ξ be a real number. Let b be an integer greater than or equal to 2. There exists a unique infinite sequence $(x_j)_{j \geq 1}$ of integers from $\{0, 1, \ldots, b-1\}$, called the b-ary expansion of ξ, such that

$$\xi = \lfloor \xi \rfloor + \sum_{j \geq 1} \frac{x_j}{b^j} \tag{1.2}$$

and $x_j \neq b-1$ for infinitely many indices j. Here, $\lfloor \cdot \rfloor$ denotes the integer part function. The sequence $(x_j)_{j \geq 1}$ is ultimately periodic if, and only if, ξ is rational. A natural way to measure the complexity of the real number ξ written in base b as in (1.2) is to count the number of distinct blocks of given length in the infinite word $\mathbf{x} = x_1 x_2 \ldots$. We set

$$p(n, \xi, b) = p(n, \mathbf{x}), \quad n \geq 1,$$

and call the function $n \mapsto p(n, \xi, b)$ the complexity function of ξ in base b.

Adamczewski [3] has observed that the statements on the combinatorial structure of Sturmian words established in [7] and [4] almost immediately imply that the irrationality exponent of any real number, whose sequence of digits in some integer base is quasi-Sturmian, is greater than 2. The main result of [8] asserts that, if the complexity function of ξ in base b grows sufficiently slowly, then the irrationality exponent of ξ exceeds 2.

Theorem 1.1 ([8]). *Let $b \geq 2$ be an integer and ξ an irrational real number. If μ denotes the irrationality exponent of ξ, then*

$$\liminf_{n \to +\infty} \frac{p(n, \xi, b)}{n} \geq 1 + \frac{1 - 2\mu(\mu-1)(\mu-2)}{\mu^3(\mu-1)}$$

and

$$\limsup_{n\to+\infty} \frac{p(n,\xi,b)}{n} \geq 1 + \frac{1-2\mu(\mu-1)(\mu-2)}{3\mu^3-6\mu^2+4\mu-1}.$$

The lower bound for the limsup obtained in Theorem 1.1 has been refined in [10] as follows.

Theorem 1.2. *Let $b \geq 2$ be an integer and ξ an irrational real number. If μ denotes the irrationality exponent of ξ, then*

$$\liminf_{n\to+\infty} \frac{p(n,\xi,b)}{n} \geq 1 + \frac{-\mu^3+2\mu^2+\mu-1}{\mu^4-2\mu^3+3\mu^2-3\mu+1}$$

and

$$\limsup_{n\to+\infty} \frac{p(n,\xi,b)}{n} \geq \frac{\mu+\sqrt{4(\mu-1)^3+\mu^2}}{2\mu(\mu-1)}. \tag{1.3}$$

The purpose of the present paper is, by means of a totally different approach, to improve (1.3).

Theorem 1.3. *Let $b \geq 2$ be an integer and ξ an irrational real number. Let μ denote the irrationality exponent of ξ. If $\mu = 2$, then*

$$\limsup_{n\to+\infty} \frac{p(n,\xi,b)}{n} \geq \frac{4}{3}.$$

If $\mu > 2$, then

$$\limsup_{n\to+\infty} \frac{p(n,\xi,b)}{n} \geq 1 + \frac{\mu+1-\sqrt{81(\mu-2)^2-10\mu+29}}{8(\mu-2)}. \tag{1.4}$$

Note that the right-hand side of (1.4) tends to $4/3$ when μ tends to 2. Furthermore, it is greater than one (that is, it gives a non-trivial result) for $\mu < 2.2$. Observe also that Theorems 1.1 and 1.2 give respectively the values $8/7$ and $(1+\sqrt{2})/2 = 1.207\ldots$ when $\mu = 2$. Thus, Theorem 1.3 is noticeably stronger. As explained in [10], the method developed in [8] and slightly refined in [10] is unlikely to yield the result of the present paper.

Here, we obtain Theorem 1.3 after a very detailled study of Rauzy graphs, introduced by Gérard Rauzy in 1983 [16]. While this approach gives better estimates for the limsup, it gives unfortunately no estimate for the liminf, unlike the approach followed in [8, 10].

Theorem 1.3 is a restatement of Theorem 1.4 below, which involves the repetition function $n \mapsto r(n,\mathbf{x})$ of an infinite word $\mathbf{x}$, defined for any positive integer n by

$$\begin{aligned} r(n,\mathbf{x}) &= \max\{m \geq 1 \,|\, x_1\ldots x_n, x_2\ldots x_{n+1}, \ldots, x_m\ldots x_{n+m-1} \text{ are all distinct}\} \\ &= \min\{m \geq 1 \,|\, x_{m+1}\ldots x_{m+n} = x_{i+1}\ldots x_{i+n} \text{ for some } 0 \leq i < m\}. \end{aligned}$$

Observe that the present definition of $r(n, \mathbf{x})$ differs by $-n$ from that given in [8–10] and that we have $p(n, \mathbf{x}) \geq r(n, \mathbf{x})$ for $n \geq 1$.

Theorem 1.4 asserts that if $p(n, \mathbf{x})$ is sufficiently small for every large value of n, then there are some positive ε and arbitrarily large values of n such that the prefix of length $\lfloor (2-\varepsilon)n \rfloor$ of $\mathbf{x}$ has two (necessarily overlapping) occurrences of a same block of length n.

Theorem 1.4. *Let $b \geq 2$ be an integer. Let $\mathbf{x} = x_1 x_2 \ldots$ be an infinite word over $\{0, 1, \ldots, b-1\}$ which is not ultimately periodic and satisfies*

$$\rho = \limsup_{n \to +\infty} \frac{p(n, \mathbf{x})}{n} < \frac{4}{3}.$$

Then, we have

$$\liminf_{n \to \infty} \frac{r(n, \mathbf{x})}{n} \leq 1 - \frac{4 - 3\rho}{2(1 + 2\rho)(2 - \rho)}$$

and

$$\mu\Big(\sum_{j \geq 1} \frac{x_j}{b^j} \Big) \geq 2 + \frac{4 - 3\rho}{\rho(9 - 4\rho)}.$$

Let $\mathbf{x}$ denote the word given by the b-ary expansion of an irrational, real number ξ. In [8,10], we have shown that there are $\delta, \delta', \delta'' > 0$ such that, if $r(n, \mathbf{x}) \leq (1+\delta)n$ for every sufficiently large integer n, then there are arbitrarily large n such that $r(n, \mathbf{x}) \leq (1-\delta')n$, thus, by [8, Lemma 3.6], we have $\mu(\xi) \geq 2+\delta''$. Said differently, if $\mu(\xi) < 2 + \delta''$, then there are arbitrarily large n such that $p(n, \mathbf{x}) \geq r(n, \mathbf{x}) \geq (1 + \delta)n$.

In the present paper, our approach is different. Our assumption is on the function $n \mapsto p(n, \mathbf{x})$ and not on the function $n \mapsto r(n, \mathbf{x})$. We show that there are $\delta, \delta' > 0$ such that, if $p(n, \mathbf{x}) \leq (1 + \delta)n$ for every sufficiently large integer n, then, there are arbitrarily large n such that $r(n, \mathbf{x}) \leq (1 - \delta')n$.

To achieve this, we study very carefully the combinatorial structure of words of small complexity, focusing on the evolution of their Rauzy graphs defined in Subsection 2.1. Previous works on these words and their Rauzy graphs include [1,2], where Aberkane studies, among other, the infinite words $\mathbf{x}$ such that $p(n, \mathbf{x})/n$ tends to 1 as n tends to infinity. Also, Heinis [13] established that if the sequence $(p(n, \mathbf{x})/n)_{n \geq 1}$ has a limit, then this limit cannot be in the open interval $(1, 2)$. This interval is the best one can obtain, since there exist sequences $\mathbf{x}$ with $p(n, \mathbf{x}) = 2n+1$ for $n \geq 1$. Such sequences, under the additional assumption of minimality, have been studied by Arnoux and Rauzy [5], again by means of the evolution of their Rauzy graphs.

One important difficulty we had to overcome is that there is no additional assumption on $\mathbf{x}$ in Theorem 1.4. In particular, $\mathbf{x}$ is not supposed to be minimal, nor even recurrent: it may happen that arbitrarily long prefixes of $\mathbf{x}$ occur only once in $\mathbf{x}$.

2. Auxiliary lemmas

Throughout the paper, we assume that the words are written over a finite alphabet.

2.1. **Special words and the Rauzy graph.** Let n be a positive integer. An n-word is a word of length n. If $\mathbf{x}$ is an infinite word, an n-subword (or, n-factor) of $\mathbf{x}$ is a factor of $\mathbf{x}$ of length n.

Definition 2.1. *Let $\mathbf{x}$ be an infinite word and n a positive integer. We let $\mathcal{F}_n(\mathbf{x})$ denote the set of n-subwords of $\mathbf{x}$. A word w in $\mathcal{F}_n(\mathbf{x})$ is called right-special if there exist distinct letters x, y such that wx and wy are subwords of $\mathbf{x}$. A word w in $\mathcal{F}_n(\mathbf{x})$ is called left-special if there exist distinct letters x, y such that xw and yw are subwords of $\mathbf{x}$. Any word in $\mathcal{F}_n(\mathbf{x})$ which is right-special and left-special is called a bi-special word.*

The following lemma is an immediate consequence of Definition 2.1.

Lemma 2.2. *Let $\mathbf{x}$ be an infinite word and m, n integers with $n > m \geq 1$.*

(i) If w is a left-special n-word of $\mathbf{x}$, then the prefix of w of length m is a left special m-word of $\mathbf{x}$.

(ii) If w is a right-special n-word of $\mathbf{x}$, then the suffix of w of length m is a right special m-word of $\mathbf{x}$.

Lemma 2.3. *Suppose that $\mathbf{x}$ is a non-ultimately periodic word. Then there is at least one left-special word w with the property that there are two distinct letters x, y such that the words xw and yw occur infinitely many times in $\mathbf{x}$.*

Proof. If there is no such left-special word, then after deleting a suitable prefix of $\mathbf{x}$ the remaining word has no left-special words, which implies that $\mathbf{x}$ is ultimately periodic by (1.1). □

Let n be a positive integer. We define the Rauzy graph $\mathcal{G}_n = \mathcal{G}_n(\mathbf{x})$ of $\mathbf{x}$. The set of vertices of the oriented graph $\mathcal{G}_n$ is the set of subwords of $\mathbf{x}$ of length n.

There exists an edge from u to v in $\mathcal{G}_n$ if and only if there exist letters x, y such that $ux = yv$ is a subword of $\mathbf{x}$. In this case, we write

$$[uv] = ux = yv.$$

A finite sequence of edges $[u_1u_2], [u_2u_3], \ldots, [u_{k-1}u_k]$ is called a path and we write it as $[u_1u_2 \ldots u_{k-1}u_k]$. A subword of length $n + \ell$ of the word $\mathbf{x}$ induces a path of length ℓ in $\mathcal{G}_n$. Moreover, a path of length ℓ in $\mathcal{G}_{n+k}$ induces a path of length $\ell + k$ in $\mathcal{G}_n$. However, the converse does not hold in general.

Let U and V be two paths in $\mathcal{G}_n$. If the last vertex of U and the initial vertex of V are identical, we define the concatenated path UV in a natural way. We denote the length of a path U in $\mathcal{G}_n$ by $|U|_n$. For example, we have $|U|_n = k$ and $|U| = n + k$, where $|\cdot|$ denotes the length (that is, the number of letters) of a word.

In the Rauzy graph, any non-right special vertex has exactly one outgoing edge, and any non-left special vertex has exactly one incoming edge, unless it is a prefix of $\mathbf{x}$ having no other occurrence in $\mathbf{x}$. Any right special vertex has at least two outgoing edges. Any left special vertex has at least two incoming edges.

2.2. **Recurrent words.**

Definition 2.4. *A subword w of an infinite word $\mathbf{x}$ is* recurrent *if w occurs in $\mathbf{x}$ infinitely many times. An infinite word $\mathbf{x}$ is n-*recurrent *if every subword of $\mathbf{x}$ of length n is recurrent. An infinite word $\mathbf{x}$ is called* recurrent *if every subword of $\mathbf{x}$ is recurrent.*

Definition 2.5. *For any $n \geq 1$, let a_n denote the shortest prefix of $\mathbf{x}$ such that $\mathbf{x} = a_n\mathbf{z}_n$ and $\mathbf{z}_n$ is n-recurrent. We call a_n the n-*th non-recurrent prefix *of $\mathbf{x}$ and $\mathcal{G}'_n(\mathbf{x}) := \mathcal{G}_n(\mathbf{z}_{n+1})$ the n-*th reduced Rauzy graph *of $\mathbf{x}$, which is a subgraph of $\mathcal{G}_n(\mathbf{x})$ consisting of recurrent vertices (n-subwords) and recurrent edges ($n+1$-subwords). We let $s_n = |a_n|$ denote the length of a_n. When there is no ambiguity, we simply write a and $\mathbf{z}$ instead of a_n and $\mathbf{z}_n$.*

Lemma 2.6. *Suppose that $\mathbf{z}_n \neq \mathbf{z}_{n+1}$. Then there exists a bi-special $(n-1)$-word w in $\mathbf{z}_n$ such that xwy is an $(n+1)$-subword of $\mathbf{z}_n$ but xwy is not a subword of $\mathbf{z}_{n+1}$, where wy is the prefix of length n of $\mathbf{z}_{n+1}$ and x is the last letter of a_{n+1}. In particular, wy is a left-special word.*

Proof. We write $\mathbf{x} = x_1x_2\ldots$. Let $\mathbf{x} = [u_1u_2\ldots]$ with $u_i = x_ix_{i+1}\ldots x_{i+n-1}$. Let $s := s_{n+1} = |a_{n+1}|$. Then $\mathbf{z}_{n+1} = x_{s+1}x_{s+2}\cdots = [u_{s+1}u_{s+2}\ldots]$ and $[u_su_{s+1}]$ is not recurrent. Since both u_s and u_{s+1} are both recurrent, u_s is right special and u_{s+1} is left special. Let $u_s = xw$ and $u_{s+1} = wy$. Then w is a bi-special $(n-1)$-word by Lemma 2.2. Therefore, there are letters x' and y' such that $x' \neq x$, $y' \neq y$ and the n-words $x'w$ and wy' are factors of $\mathbf{z}_n$. However $xwy = [u_s u_{s+1}] = x_s \dots x_{s+n}$ is not recurrent, thus it is not a subword of $\mathbf{z}_{n+1}$. □

Lemma 2.7. *Let $n \geq 1$. Then we have*

$$p(n + s_n - 1, \mathbf{x}) = p(n + s_n - 1, \mathbf{z}_n) + s_n.$$

Proof. If $s_n = 0$, then $\mathbf{x} = \mathbf{z}_0$, thus the lemma holds. We assume that $s := s_n = |a_n| \geq 1$. We write $\mathbf{x} = x_1 x_2 \cdots = [u_1 u_2 \dots]$ with $u_i = x_i x_{i+1} \dots x_{i+n-1}$. Then u_s is not recurrent and u_i is recurrent for $i \geq s+1$.

For each i with $1 \leq i \leq s$, the subword $[u_i \dots u_{i+s-1}]$ contains u_s. However, for $i \geq s+1$, the subword $[u_i \dots u_{i+s-1}]$ is a factor of $\mathbf{z}_n = [u_{s+1} u_{s+2} \dots]$ and do not contain any non-recurrent n-subword.

Since $u_j \neq u_s$ for $j \geq s+1$, we deduce that the s words $[u_i \dots u_{i+s-1}]$ with $i = 1, \dots, s$ are all distinct. Therefore, we have $p(n+s-1, \mathbf{x}) = p(n+s-1, \mathbf{z}_n) + s$. □

The following lemma is [11, Theorem 4.5.4].

Lemma 2.8. *Let $\mathbf{x} = x_1 x_2 \dots$ be a non-ultimately periodic word. Let $n \geq 1$. Suppose that $p(n+1, \mathbf{x}) = p(n, \mathbf{x}) + 1$. Then there is one right-special word of length n. Moreover, if $x_1 \dots x_n$ occurs in $\mathbf{x}$ more than once, then there is a unique left-special word of length n in $\mathbf{x}$ and $\mathbf{x}$ is n-recurrent.*

Definition 2.9. *A right-special word w is called* essential *if w is a suffix of a right-special word of any length (or of arbitrarily large length, this is equivalent).*

From Lemma 2.2, we note that for any non-essential right-special word w there exists m such that w is not a suffix of a right special word of length bigger than m. Therefore, if $\mathbf{x}$ is not ultimately periodic, then for each n there exists an essential right-special word of length n.

Lemma 2.10. *Suppose that $\liminf_{n\to\infty} p(n, \mathbf{x})/n < 2$ and $\mathbf{x}$ is not ultimately periodic. Then, for each n, there exists a unique essential right-special word of length n.*

Proof. By the above remark, there exists an essential right-special word of length n. By assumption, there is an integer $m > n$ such that $p(m+1, \mathbf{x}) = p(m, \mathbf{x}) + 1$. By

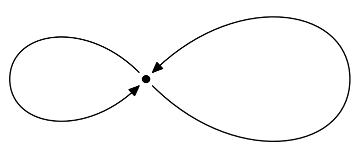
Lemma 2.8, there exists one right-special word of length m. Therefore, the number of essential right-special words of length n is at most one. $\square$

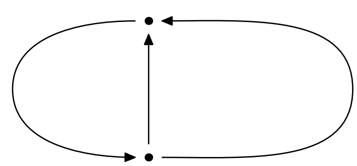
Lemma 2.11. *Suppose that* $\mathbf{x}$ *is* n*-recurrent and non-ultimately periodic. If* $p(n+1,\mathbf{x}) = p(n,\mathbf{x})+1$*, then* $\mathbf{x}$ *is* $(n+1)$*-recurrent.*

Proof. Write $\mathbf{x} = a\mathbf{z}$, where a is the $(n+1)$-th non-recurrent prefix of $\mathbf{x}$ and assume that a is non-empty. Let $\mathbf{z} = x_{s+1}x_{s+2}\dots$. Then $x_{s+1}\dots x_{s+n}$ is a left-special n-word of $\mathbf{x}$ by Lemma 2.6. Since there is only one left special n-word in $\mathbf{x}$ by Lemma 2.8 and the only left special word has two preceding letters, Lemma 2.3 implies that $x_s x_{s+1}\dots x_{s+n}$ occurs infinitely many times in $\mathbf{x}$, which contradicts the choice of a and $\mathbf{z}$. $\square$

2.3. **Rauzy graph of ∞-shape.** A Rauzy graph $\mathcal{G}_n$ (resp., a reduced Rauzy graph $\mathcal{G}'_n$) is called of ∞-*shape* if there exists exactly one special word w of length n in $\mathbf{x}$ (resp., in $\mathbf{z}_{n+1}$) and there are exactly two paths from w to itself. Note that w is a bi-special word.

Lemma 2.12. *Suppose that a non-ultimately periodic word* $\mathbf{x}$ *satisfies that* $p(n+1,\mathbf{x}) = p(n,\mathbf{x})+1$ *for some positive integer* n*. Then the reduced Rauzy graph* $\mathcal{G}'_n(\mathbf{x})$ *is isomorphic to one of the graphs*

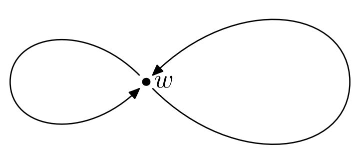




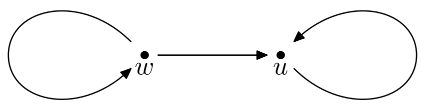
Moreover, in the latter case, there exists a positive integer d *such that* $\mathcal{G}'_{n+d}(\mathbf{x})$ *is of* ∞*-shape.*

Proof. Let $\mathbf{x} = a\mathbf{z}$ with the n-th non-recurrent prefix (possibly empty) a. Then, $p(n+1,\mathbf{z}) = p(n,\mathbf{z})+1$ and $\mathbf{z}$ has either one right special word and one left special word, or one bispecial word of length n.

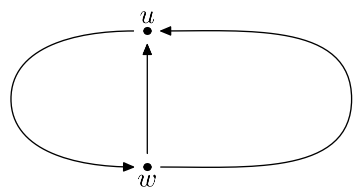
If there is a bi-special word w of length n in $\mathbf{z}$, then there are no other special vertices. Let $[wu_1\cdots u_{k-1}w]$ and $[wv_1\cdots v_{\ell-1}w]$ denote the two cycles containing w. When $k=1$, then $[wu_1\cdots u_{k-1}w]$ denotes $[ww]$. The Rauzy graph $\mathcal{G}_n(\mathbf{z})$ can be represented as follows:



If the right-special word and the left special word are distinct, then $\mathcal{G}_n(\mathbf{z})$ is either of the form



or



For the left-hand side graph, $\mathbf{z}$ is ultimately periodic, which is not allowed. Let u denote the left special word and w denote the right special word. In the Rauzy graph $\mathcal{G}_n(\mathbf{z})$, there exist two paths from w to u, namely, $[wu_1 \cdots u_{k-1}u]$ and $[wv_1 \cdots v_{\ell-1}u]$ and there exists one path from u to w, namely $[uw_1 \ldots w_{m-1}w]$. Then, the $(n+m)$-subword $[uw_1 \ldots w_{m-1}w]$ is bispecial since $[uw_1 \ldots w_{m-1}wu_1]$, $[uw_1 \ldots w_{m-1}wv_1]$ and $[u_{k-1}uw_1 \ldots w_{m-1}w]$, $[v_{\ell-1}uw_1 \ldots w_{m-1}w]$ are edges of the graph $\mathcal{G}_{n+m}(\mathbf{z})$. The two cycles containing the bispecial word are

$$[uw_1 \ldots w_{m-1}wu_1u_2 \cdots u_{k-1}uw_1 \ldots w_{m-1}w]$$

and

$$[uw_1 \ldots w_{m-1}wv_1v_2 \cdots v_{\ell-1}uw_1 \ldots w_{m-1}w].$$

Consequently, the graph $\mathcal{G}_{n+m}(\mathbf{z})$ is of ∞-shape.

Any right special word of $\mathbf{z}$ of length $n+d$ for $0 \le d \le m$ is a suffix of $[uw_1w_2 \ldots w_{m-1}w]$. Therefore, $p(n+d+1, \mathbf{z}) = p(n+d, \mathbf{z}) + 1$, for $0 \le d \le m$. By Lemma 2.11, the word $\mathbf{z}$ is $(n+m+1)$-recurrent. Hence, we have $\mathcal{G}'_{n+m}(\mathbf{x}) = \mathcal{G}_{n+m}(\mathbf{z})$. □

3. Evolution of the Rauzy graphs

In this section, we consider the evolution of the Rauzy graph of $\mathbf{x} = x_1x_2\ldots$ when $\mathbf{x}$ is not ultimately periodic and $p(n, \mathbf{x}) \le \frac{4}{3}n$ for every large integer n. Then Lemma 2.12 implies that there exist infinitely many n such that the reduced Rauzy graphs $\mathcal{G}'_n(\mathbf{x})$ are of ∞-shape. Let the Rauzy graph $\mathcal{G}_n(\mathbf{x})$ be of ∞-shape with the bispecial word w, and let $U = [wu_1 \ldots u_{k-1}w]$, $V = [wv_1 \ldots v_{\ell-1}w]$ denote the two cycles of $\mathcal{G}_n(\mathbf{x})$ from w to itself and put $u_0 = u_k = v_0 = v_\ell = w$. See Figure 1. Note that if $k = 1$, then $U = [ww]$. We write concatenated paths in the Rauzy graph $\mathcal{G}_n$ as $UU = [wu_1 \ldots u_{k-1}wu_1 \ldots u_{k-1}w]$, $UV = [wu_1 \ldots u_{k-1}wv_1 \ldots v_{\ell-1}w]$ and $V^b = V \cdots V$ is called the b-th concatenated path of V. We also define

$$U_i = [u_i \ldots u_{k-1}w], \quad i = 1, \ldots, k-1,$$

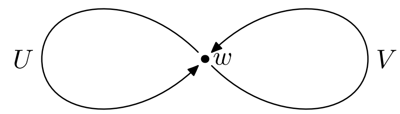


FIGURE 1. The Rauzy graph $\mathcal{G}_n(\mathbf{x})$ with a bispecial word w and two cycles U, V

and

$$V_i = [v_i \dots v_{\ell-1} w], \quad i = 1, \dots, \ell - 1.$$

We write $U_0 = U$, $V_0 = V$ and for convenience we let $U_k = V_\ell$ denote the empty path.

By Lemma 2.10, either $[u_{k-1}w]$ or $[v_{\ell-1}w]$ is essentially right-special. If $[u_{k-1}w]$ is essentially right-special, then so is U.

Lemma 3.1. *Suppose that $\mathbf{x}$ is not ultimately periodic and the Rauzy graph $\mathcal{G}_n(\mathbf{x})$ is of ∞-shape with the bi-special word w and two cycles $U = [wu_1 \dots u_{k-1}w]$, $V = [wv_1 \dots v_{\ell-1}w]$. We assume that, for any $m > n$, we have*

$$\frac{p(m, \mathbf{x})}{m} \leq \frac{4}{3}.$$

Assume further that U is essentially right-special.
(i) If UV^bU is a factor of $\mathbf{x}$ for some integer $b \geq 2$, then

$$k \geq (2b - 3)\ell + 4.$$

(ii) If UV^bU and $UV^{b'}U$ are factors of $\mathbf{x}$ for some integers $b, b' \geq 1$, then $b = b'$.

Proof. (i) If UV^bU is a factor of $\mathbf{x}$, then V^{b-1} is a right-special word. Let m be an integer with $n + 1 \leq m \leq n + (b - 1)\ell$. Then by Lemma 2.2, the suffix of V^{b-1} of length m is a right-special word. Since U is essentially right-special, there is another right-special word ending with $[u_{k-1}w]$ of length m. Therefore, $p(m + 1, \mathbf{x}) \geq p(m, \mathbf{x}) + 2$ for each $n + 1 \leq m \leq n + (b - 1)\ell$. Thus

$$p(n + (b - 1)\ell + 1, \mathbf{x}) \geq p(n + 1, \mathbf{x}) + 2(b - 1)\ell.$$

Since $p(n + 1, \mathbf{x}) = k + \ell \geq n + 2$, we obtain

$$\frac{4}{3} \geq \frac{p(n + (b - 1)\ell + 1, \mathbf{x})}{n + (b - 1)\ell + 1} \geq \frac{k + (2b - 1)\ell}{k + b\ell - 1} = 1 + \frac{(b - 1)\ell + 1}{k + b\ell - 1}.$$

Therefore, we get

$$k \geq (2b - 3)\ell + 4.$$

(ii) We may assume that $b > b'$. Note that V^{b-1} and $UV^{b'}$ are right-special and by Lemma 2.2 the suffixes of them are right-special. The number of suffixes of V^{b-1} ending by $[v_{\ell-1}w]$ is $(b-1)\ell$ and the number of suffixes of $UV^{b'}$ ending by $[v_{\ell-1}w]$ is $k + b'\ell$. Since $b'\ell$ of these suffixes coincide, the number of right-special subwords of $\mathbf{x}$ ending by $[v_{\ell-1}w]$ is at least

$$(b-1)\ell + k + b'\ell - b'\ell = k + (b-1)\ell,$$

and so

$$\begin{aligned} p(n+k+(b-1)\ell+1, \mathbf{x}) &\geq p(n+1, \mathbf{x}) + 2k + 2(b-1)\ell \\ &= 3k + (2b-1)\ell. \end{aligned}$$

Thus, since $b \geq 2$, we have

$$\frac{p(n+k+(b-1)\ell+1, \mathbf{x})}{n+k+(b-1)\ell+1} \geq \frac{3k+(2b-1)\ell}{2k+b\ell-1} > \frac{4}{3},$$

which gives a contradiction. $\square$

Definition 3.2. *Suppose that the Rauzy graph $\mathcal{G}_n(\mathbf{x})$ consists of one bispecial word w and two cycles U, V. We call U a special cycle if U is essentially right-special. A Rauzy graph $\mathcal{G}_n(\mathbf{x})$ with a bispecial word w, a special cycle U, and a non-special cycle V is called a Rauzy graph of configuration (w, U, V). We say that the Rauzy graph $\mathcal{G}_n(\mathbf{x})$ of configuration (w, U, V) has multiplicity $b \geq 1$ if the concatenated path UV^bU is a factor of $\mathbf{x}$.*

Proposition 3.3. *Suppose that the reduced Rauzy graph $\mathcal{G}'_n(\mathbf{x})$ is of configuration (w, U, V) with multiplicity $b \geq 1$ and let $|U|_n = k$ and $|V|_n = \ell$. Assume that*

$$\frac{p(m, \mathbf{x})}{m} \leq \frac{4}{3} \quad \text{for } m \geq n.$$

Then the reduced Rauzy graph $\mathcal{G}'_{n+k}(\mathbf{x})$ is of configuration either (U, UU, UV^bU) or (U, UV^bU, UU) with the bispecial word U and the two cycles of concatenated paths UU and UV^bU (see Figure 2). Furthermore, we have

$$p(n+(b-1)\ell+1, \mathbf{z}_{n+k+1}) = k + (2b-1)\ell \tag{3.1}$$

and

$$s_{n+(b-1)\ell+2} = s_{n+k+1}. \tag{3.2}$$

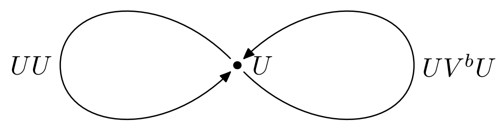


FIGURE 2. The reduced Rauzy graph $\mathcal{G}'_{n+k}(\mathbf{x})$

Moreover, if $k < \ell$ and $s_{n+1} < s_{n+k+1}$ with $b = 1$, then there exists i, $1 \le i \le \ell - 1$ such that

$$\mathbf{z}_{n+1} = V_i V U \dots, \ \mathbf{z}_{n+k+1} = VU \dots, \quad \text{and} \ \ s_{n+k+1} - s_{n+1} = |V_i|_n, \tag{3.3}$$

Proof. Since $\mathbf{x}$ is not ultimately periodic, both U and V are recurrent. By Lemma 3.1 (ii), UV^bU is recurrent as well. Since U is an essential right-special word, UU is also recurrent. By Lemma 3.1, we note that for any $b \ge 1$ we have

$$k \ge (b-1)\ell + 1.$$

All the recurrent words of length $n + (b-1)\ell + 1$ are prefixes of

$$\begin{cases} U_i, & \text{for} \ \ 0 \le i \le k - (b-1)\ell - 1, \\ U_i U, U_i V^b U, & \text{for} \ \ k - (b-1)\ell \le i \le k-1, \\ V_i U, V_i V U, \dots, V_i V^{b-1} U, & \text{for} \ \ 0 \le i \le \ell - 1. \end{cases}$$

Using Lemma 3.1 (ii), we deduce that there is no other recurrent word of length $n + (b-1)\ell + 1$. Consequently,

$$\begin{aligned} p(n + (b-1)\ell + 1, \mathbf{z}_{n+(b-1)\ell+1}) &= p(n + (b-1)\ell + 1, \mathbf{z}_{n+k+1}) \\ &= k + (2b-1)\ell. \end{aligned} \tag{3.4}$$

Suppose $\mathbf{z}_{n+(b-1)\ell+2} \ne \mathbf{z}_{n+k+1}$. Then by Lemma 2.6, there exists a recurrent bispecial word W of length m with $n+(b-1)\ell+1 \le m \le n+k-1$. By Lemma 2.2, the bispecial word W should be a path in $\mathcal{G}_n$ that starts and ends with U or V. However, since $|U| = n+k$, the word W cannot contain U. Thus $W = V^c$ for some $c > b$. Since W is recurrent, $UV^{c'}U$ for $c' \ge c$ is a factor of $\mathbf{z}_n$, which is not allowed by Lemma 3.1 (ii). Hence, we have $s_{n+(b-1)\ell+2} = s_{n+k+1}$.

We assume that $b = 1$ and $s_{n+1} < s_{n+2} = s_{n+k+1}$. Since U is the special cycle, UU, UV are recurrent factors. We also check that VU is recurrent, otherwise $\mathbf{x}$ is ultimately periodic. Therefore, by Lemma 2.6, $[v_{\ell-1} w v_1]$ is not a recurrent factor of $\mathbf{x}$ and

$$x_s x_{s+1} \cdots = [v_{\ell-1} w v_1 \ldots], \qquad s := s_{n+2}. \tag{3.5}$$

Suppose that $k < \ell$. We apply Lemma 2.7 for $n+2$. Then we have

$$p(n + s_{n+2} + 1, \mathbf{x}) = p(n + s_{n+2} + 1, \mathbf{z}_{n+2}) + s_{n+2} \geq n + 2s_{n+2} + 2.$$

Combined with $p(n+s_{n+2}+1, \mathbf{x}) \leq \frac{4}{3}(n+s_{n+2}+1)$ and $p(n, \mathbf{z}_{n+1}) = k+\ell-1 \geq n+1$, we have

$$2s_{n+2} + 2 \leq n \leq k + \ell - 2 < 2\ell - 2.$$

Therefore, $s_{n+2} < \ell - 2$ and (3.5) yields that

$$\mathbf{z}_{n+1} = [v_i \ldots v_{\ell-1} w v_1 \ldots], \qquad \text{for some } i \text{ with } \ 2 < i \leq \ell - 1.$$

Hence

$$\mathbf{z}_{n+1} = V_i V U \ldots, \quad \mathbf{z}_{n+2} = V U \ldots, \quad \text{and} \quad s_{n+2} - s_{n+1} = |V_i|_n. \qquad \square$$

Definition 3.4. *Suppose that $\mathbf{x}$ is not ultimately periodic and $p(n, \mathbf{x}) \leq \frac{4n}{3}$ for every large integer n. By Proposition 3.3, we have an increasing sequence of integers $(n_i)_{i=0}^{\infty}$ such that the reduced Rauzy graph $\mathcal{G}'_{n_i}$ is of ∞-shape. If $\mathcal{G}'_{n_i}$ is of configuration (w, U, V) with multiplicity $b \geq 1$, then the reduced Rauzy graph $\mathcal{G}'_{n_{i+1}}$ is of configuration (U, UU, UV^bU) or (U, UV^bU, UU).*

Definition 3.5. *We say that $\mathbf{x}$ has* bounded cycles *if there exists k such that all the reduced Rauzy graphs $\mathcal{G}'_n(\mathbf{x})$ have a cycle whose length is bounded by k.*

Proposition 3.6. *Suppose that $\mathbf{x}$ is not ultimately periodic and $p(n, \mathbf{x}) \leq \frac{4n}{3}$ for every large integer n.*

(i) If $\mathbf{x}$ has bounded cycles, then there exists k such that for large n_i the ∞-shape reduced Rauzy graph $\mathcal{G}'_{n_i}(\mathbf{x})$ is of configuration $(w_{(i)}, U_{(i)}, V_{(i)})$ with $|U_{(i)}|_{n_i} = k$ and multiplicity 1.

(ii) If $\mathbf{x}$ does not have bounded cycles, then there are infinitely many n_i's such that the ∞-shape reduced Rauzy graph $\mathcal{G}'_{n_i}(\mathbf{x})$ is of configuration $(w_{(i)}, U_{(i)}, V_{(i)})$ with $|U_{(i)}|_{n_i} > |V_{(i)}|_{n_i}$.

In both cases, we have $\lim_{i\to\infty} |V_{(i)}|_{n_i} = \infty$.

Proof. We use the sequence $(n_i)_{i=0}^{\infty}$ introduced in Definition 3.4. Let $(w_{(i)}, U_{(i)}, V_{(i)})$ be the configuration of $\mathcal{G}'_{n_i}$, b_i be its multiplicity and $k_i = |U_{(i)}|_{n_i}$, $\ell_i = |V_{(i)}|_{n_i}$. By Proposition 3.3, we have $n_{i+1} = n_i + k_i$ and the reduced Rauzy graph $\mathcal{G}'_{n_{i+1}}$ is of configuration $(U_{(i)}, U_{(i)}U_{(i)}, U_{(i)}(V_{(i)})^{b_i}U_{(i)})$ or $(U_{(i)}, U_{(i)}(V_{(i)})^{b_i}U_{(i)}, U_{(i)}U_{(i)})$. If $U_{(i+1)} = U_{(i)}U_{(i)}$, $V_{(i+1)} = U_{(i)}V_{(i)}^{b_i}U_{(i)}$ then

$$k_{i+1} = |U_{(i)}U_{(i)}|_{n_{i+1}} = k_i, \quad \ell_{i+1} = |U_{(i)}V_{(i)}^{b_i}U_{(i)}|_{n_{i+1}} = k_i + b_i\ell_i.$$

If $U_{(i+1)} = U_{(i)}V_{(i)}^{b_i}U_{(i)}$, $V_{(i+1)} = U_{(i)}$, then

$$k_{i+1} = |U_{(i)}V_{(i)}^{b_i}U_{(i)}|_{n_{i+1}} = k_i + b_i\ell_i, \quad \ell_{i+1} = |U_{(i)}U_{(i)}|_{n_{i+1}} = k_i.$$

If $k_{i+1} = k_i$ for all large i's, then k_i is bounded and ℓ_i goes to infinity. By Lemma 3.1, $b_i = 1$ for all large i's.

If $k_{i+1} > k_i$ for infinitely many i's, then $k_{i+1} = k_i + b_i\ell_i > k_i = \ell_{i+1}$ for infinitely many i's. Note that $(k_i + \ell_i)_{i=0}^{\infty}$ is strictly increasing. We see that k_i goes to infinity and so does ℓ_i. □

Example 3.7. Let $\mathbf{x} = U^{a_1}VU^{a_2}V \cdots VU^{a_n}V \cdots$ with $a_1 \le a_2 \le \dots$ for some cycles $U = [wu_1 \dots u_{k-1}w]$ and $V = [wv_1 \dots v_{\ell-1}w]$ with n-subwords w, u_j $(1 \le j \le k-1)$ and v_j $(1 \le j \le \ell - 1)$. Let n_i $(n_i > n)$ be the sequence such that the reduced Rauzy graph $\mathcal{G}'_{n_i}(\mathbf{x})$ is of ∞-shape. Then $n_i = n + ik$ and $\mathcal{G}'_{n_i}(\mathbf{x})$ has cycles U^{i+1} and U^iVU^i. Therefore, $\mathbf{x}$ has bounded cycles.

4. Recurrence time of low complexity sequences

Recall that s_n denotes the length of n-th non-recurrent prefix of $\mathbf{x}$. Moreover, we use the same notation for $U = [wu_1 \dots u_{k-1}w]$, $V = [wv_1 \dots v_{\ell-1}w]$ as in the previous section. Throughout this section, ρ is a positive real number less than $4/3$ and we assume that $\mathbf{x}$ is not ultimately periodic and that $p(m, \mathbf{x})/m < \rho$ for sufficiently large m.

Lemma 4.1. *Let n be given and suppose that*

$$\frac{p(m,\mathbf{x})}{m} < \rho$$

holds for any $m \ge n$. Suppose that $\mathcal{G}'_n(\mathbf{x})$ is of ∞-shape with two cycles U and V. Let b be the multiplicity of $\mathcal{G}'_n(\mathbf{x})$. Let $|U|_n = k$ and $|V|_n = \ell$. Then we have

$$k + \frac{2b+1}{3}\ell < \rho(n+1) \tag{4.1}$$

and

$$2s_{n+k+1} + k + (2b-1)\,\ell < \frac{\rho n}{2-\rho}. \tag{4.2}$$

Proof. By Proposition 3.3, we have $s := s_{n+(b-1)\ell+2} = s_{n+k+1}$. We apply Lemma 2.7 with n replaced by $n+(b-1)\ell+2$. Then, combined with (3.1), we have

$$\begin{aligned} p(n+(b-1)\ell+s+1,\mathbf{x}) &= p(n+(b-1)\ell+s+1,\mathbf{z}_{n+(b-1)\ell+2})+s \\ &\geq p(n+(b-1)\ell+1,\mathbf{z}_{n+(b-1)\ell+2})+2s \\ &= k+(2b-1)\ell+2s. \end{aligned}$$

By using $p(n+(b-1)\ell+s+1,\mathbf{x}) < \rho(n+(b-1)\ell+s+1)$, we deduce that

$$(2-\rho)s+k+(2b-1-\rho(b-1))\ell < \rho(n+1). \tag{4.3}$$

Therefore, we have

$$k+\left(\frac{2b}{3}+\frac{1}{3}\right)\ell \leq k+(2b-1-\rho(b-1))\,\ell < \rho(n+1),$$

which is (4.1).

We note that, since $p(n,\mathbf{z}_{n+1}) = k+\ell-1 \geq n+1$, we have

$$n \leq k+\ell-2. \tag{4.4}$$

Combined with (4.3), this gives

$$(2-\rho)s+k+\ell+(2-\rho)(b-1)\ell < \frac{\rho}{2}n+\frac{\rho}{2}(k+\ell-2)+\rho,$$

that is,

$$(2-\rho)s+\left(1-\frac{\rho}{2}\right)(k+\ell)+(2-\rho)(b-1)\ell < \frac{\rho}{2}n.$$

By dividing by $1-\frac{\rho}{2}$, we obtain (4.2). □

In the rest of the paper, we put

$$\delta := \frac{4-3\rho}{2(1+2\rho)(2-\rho)}.$$

Lemma 4.2. *We use the same notation as in Lemma 4.1. Let n be large enough. Then we have*

$$s_{n+k+1}+2k+b\ell < (1-\delta)(n+2k+b\ell). \tag{4.5}$$

For $k'=0,\ldots,k-1$, we have either

$$s_{n+k+1}+2k+b\ell < (1-\delta)(n+k+b\ell+k') \tag{4.6}$$

or

$$s_{n+k+1}+2k+b\ell+k' < (1-\delta)(n+2k). \tag{4.7}$$

Moreover, if $\ell < k$, then

$$s_{n+k+1} + 2k + b\ell < (1-\delta)(n+2k). \tag{4.8}$$

Proof. By (4.1) and the last statement of Proposition 3.6, for large n such that $\ell \geq 3 > 2\rho$, we have

$$2k + b\ell < 2\rho n. \tag{4.9}$$

Then we have by (4.2)

$$s_{n+k+1} + \delta(n+2k+b\ell) < \frac{\rho n}{2(2-\rho)} + (1+2\rho)\delta n = \frac{\rho n}{2(2-\rho)} + \frac{4-3\rho}{2(2-\rho)} n = n.$$

Therefore, we deduce (4.5).

From (4.2), for $k' = 0, \ldots, k-1$, we have either

$$s_{n+k+1} + k < \frac{\rho n}{2(2-\rho)} + k' \tag{4.10}$$

or

$$s_{n+k+1} + b\ell + k' < \frac{\rho n}{2(2-\rho)}. \tag{4.11}$$

If (4.10) holds, then by (4.9)

$$s_{n+k+1} + k + \delta(n+k+b\ell+k') < \frac{\rho n}{2(2-\rho)} + k' + (1+2\rho)\delta n = n + k',$$

which implies (4.6). If (4.11) holds, then by (4.9)

$$s_{n+k+1} + b\ell + k' + \delta(n+2k) < \frac{\rho n}{2(2-\rho)} + (1+2\rho)\delta n = n.$$

which implies (4.7).

If $\ell < k$, then by (4.2)

$$s_{n+k+1} + b\ell < \frac{\rho n}{2(2-\rho)}.$$

Using (4.9), we have

$$s_{n+k+1} + b\ell + \delta(n+2k) < \frac{\rho n}{2(2-\rho)} + (1+2\rho)\delta n = n.$$

Therefore, we obtain (4.8). $\square$

Proposition 4.3. *If $\mathbf{x}$ has bounded cycles, then for infinitely many integers n we have*

$$r(n, \mathbf{x}) < (1-\delta)n.$$

Proof. We use the sequence $(n_i)_{i=0}^{\infty}$ introduced in Definition 3.4. Let $(w_{(i)}, U_{(i)}, V_{(i)})$ be the configuration of $\mathcal{G}'_{n_i}$. By Proposition 3.6, we may assume that the multiplicity of $\mathcal{G}'_{n_i}$ is 1 for any $i \geq 0$ and that $|U_{(i)}|_{n_i} = k$ is a positive constant. Set $(w, U, V) := (w_{(0)}, U_{(0)}, V_{(0)})$. Then we have $w_{(i)} = U^i$, $U_{(i)} = U^{i+1}$ and $V_{(i)} = U^i V U^i$ for each $i \geq 1$. Putting $\ell := |V|_{n_0}$, we note that $n_i = |w_{(i)}| = n_0 + ki$, $|U_{(i)}| = n_0 + (i+1)k$, $|V_{(i)}| = n + \ell + 2ki$ and that $|U_{(i)}|_{n_i} = k$, $|V_{(i)}|_{n_i} = \ell + ki$. Since $\mathbf{x}$ is not ultimately periodic, $s_{n_i+1} < s_{n_{i+1}+1}$ for infinitely many i's. Choose i large enough to ensure that

$$(\ell + ki)\delta > 2k. \tag{4.12}$$

By Proposition 3.3, we see that $s_{n_i+2} = s_{n_{i+1}+1} = s_{n_i+k+1}$ and

$$\mathbf{z}_{n_i+2} = V_{(i)} U_{(i)} \ldots, \tag{4.13}$$

where the concatenation rule follows from the word of length n_i. Let

$$h = h(i) := \max\{j \geq 1 \mid s_{n_{i+1}+1} = s_{n_{i+j}+1}\}.$$

Then $s_{n_{i+h}+1} < s_{n_{i+h+1}+1}$. We, again, apply Proposition 3.3. Then we have

$$\mathbf{z}_{n_{i+h}+1} = \widetilde{V_{(i+h)}} V_{(i+h)} U_{(i+h)} \ldots, \qquad \mathbf{z}_{n_{i+h}+2} = V_{(i+h)} U_{(i+h)} \ldots \tag{4.14}$$

and

$$s_{n_{i+h}+2} - s_{n_{i+h}+1} = |\widetilde{V_{(i+h)}}|_{n_{i+h}},$$

where the concatenation rule follows from the word of length n_{i+h} and $\widetilde{V_{(i+h)}}$ is a suffix of $V_{(i+h)} = U_{(i)}^h V_{(i)} U_{(i)}^h$. Since $\mathbf{z}_{n_i+2} = \mathbf{z}_{n_{i+h}+1}$, by (4.13) and (4.14), we have $\widetilde{V_{(i+h)}} = V_{(i)} U_{(i)}^h$. Therefore, we have

$$s_{n_{i+h}+k+1} - s_{n_i+2} = s_{n_{i+h}+2} - s_{n_{i+h}+1} = |V_{(i)} U_{(i)}^h|_{n_{i+h}} = |V_{(i)}|_{n_i} = \ell + ki \tag{4.15}$$

and

$$\mathbf{z}_{n_i+2} = \mathbf{z}_{n_{i+h}+1} = V_{(i)} U_{(i)}^h V_{(i)} U_{(i)}^{h+1} \ldots, \qquad \mathbf{z}_{n_{i+h}+2} = U_{(i)}^h V_{(i)} U_{(i)}^{i+h+1} \ldots.$$

Therefore

$$r(n_i + (h-1)k, \mathbf{x}) \leq s_{n_i+2} + (i+1)k + \ell. \tag{4.16}$$

Note that the lengths of the two cycles of $\mathcal{G}'_{n_{i+h}}$ are

$$|U_{(i+h)}|_{n_{i+h}} = k, \qquad |V_{(i+h)}|_{n_{i+h}} = \ell + (i+h)k.$$

Thus, we get by (4.5) that

$$s_{n_{i+h}+k+1} + 2k + \ell + (i+h)k < (1-\delta)(n_{i+h} + 2k + \ell + (i+h)k).$$

Therefore, combined with (4.15), we have

$$s_{n_i+2} + (h+2i+2)k + 2\ell < (1-\delta)(n_i + (2h+i+2)k + \ell).$$

Combined with (4.16) and (4.12), we have

$$\begin{aligned} r(n_i + (h-1)k, \mathbf{x}) &\le s_{n_i+2} + (i+1)k + \ell \\ &< (1-\delta)(n_i + (2h+i+2)k + \ell) - (i+h+1)k - \ell \\ &= (1-\delta)(n_i + 2(h+1)k) - (h+1)k - \delta(ki + \ell) \\ &< (1-\delta)(n_i + 2(h+1)k) - (h+3)k \\ &< (1-\delta)(n_i + (h-1)k). \end{aligned}$$

Since there are infinitely many such n_i's, the proof is complete. □

Lemma 4.4. *Suppose that $\mathcal{G}'_n(\mathbf{x})$ is of configuration (w, U, V) with $|U|_n = k$, $|V|_n = \ell$ and multiplicity $b \ge 1$. Then $\mathbf{z}_{n+k+1}$ belongs to one of the following cases, where U' (resp. V') denotes a suffix of U (resp. V) with $0 \le |U'|_n =: k' < k$ (resp. $0 \le |V'|_n =: \ell' < \ell$)*

(1) *If $\mathbf{z}_{n+k+1} = U'UU\ldots$, then $r(n+k+k', \mathbf{x}) \le s_{n+k+1} + k$.*
(2) *If $\mathbf{z}_{n+k+1} = U'UV^bUUU\ldots$, then $r(n+k', \mathbf{x}) \le s_{n+k+1} + k$ and $r(n+2k, \mathbf{x}) \le s_{n+k+1} + 2k + b\ell + k'$.*
(3) *If $\mathbf{z}_{n+k+1} = U'UV^bUUV^bU\ldots$, then $r(n+2k+b\ell+k', \mathbf{x}) \le s_{n+k+1}+2k+b\ell$.*
(4) *If $\mathbf{z}_{n+k+1} = U'UV^bUV^bU\ldots$, then $r(n+2k+b\ell, \mathbf{x}) \le s_{n+k+1}+k+b\ell+k'$.*
(5) *If $\mathbf{z}_{n+k+1} = U'V^bUUU\ldots$, then $r(n+2k, \mathbf{x}) \le s_{n+k+1} + k + b\ell + k'$.*
(6) *If $\mathbf{z}_{n+k+1} = U'V^bUUV^bU\ldots$, then $r(n+(b-1)\ell, \mathbf{x}) \le s_{n+k+1} + \ell + k'$ and $r(n+k+b\ell+k', \mathbf{x}) \le s_{n+k+1} + 2k + b\ell$.*
(7) *If $\mathbf{z}_{n+k+1} = U'V^bUV^bU\ldots$, then $r(n+k+b\ell+k', \mathbf{x}) \le s_{n+k+1} + k + b\ell$.*
(8) *If $\mathbf{z}_{n+k+1} = V'V^cUU\ldots$ for some $0 \le c < b$, then $r(n+k, \mathbf{x}) \le s_{n+k+1} + k + c\ell + \ell'$.*
(9) *If $\mathbf{z}_{n+k+1} = V'V^cUV^bU\ldots$ for some $0 \le c < b$, then $r(n+k+c\ell+\ell', \mathbf{x}) \le s_{n+k+1} + k + b\ell$.*

Proof. By Proposition 3.3, every subword of $\mathbf{z}_{n+k+1}$ starting with U is a concatenation of the two words UU and UV^bU. Therefore, if $\mathbf{z}_{n+k+1}$ starts with U', then the prefix of $\mathbf{z}_{n+k+1}$ is either $U'UU$, $U'UV^bU$ or $U'V^bU$. The second and the third cases are divided into the three subcases respectively: $U'UV^bUUU$, $U'UV^bUUV^bU$, $U'UV^bUV^bU$, and $U'V^bUUU$, $U'V^bUUV^bU$, $U'V^bUV^bU$.

If $\mathbf{z}_{n+k+1}$ starts with V', a suffix of V, then the prefix of $\mathbf{z}_{n+k+1}$ is either $V'V^cUU$ or $V'V^cUV^bU$ for some $c \leq b-1$, thus giving the last two cases. □

Proposition 4.5. *Let $1 \leq \rho < 4/3$ be defined as in Lemma 4.1. Moreover, let n be large enough that $\ell \geq 3$. Let $\mathcal{G}_n'(\mathbf{x})$ be of configuration (w, U, V) with multiplicity $b \geq 1$. Suppose that $|U|_n = k > \ell = |V|_n$. Then there exists m with $n \leq m < n + 3k + b\ell$ satisfying that*

$$r(m, \mathbf{x}) < (1-\delta)m.$$

Proof. By Lemma 4.4, we have

(1) If $\mathbf{z}_{n+k+1} = U'UU\ldots$, then by (4.5) we have

$$r(n+k+k', \mathbf{x}) \leq s_{n+k+1} + k < (1-\delta)(n+k) \leq (1-\delta)(n+k+k').$$

(2) If $\mathbf{z}_{n+k+1} = U'UV^bUUU\ldots$, then by (4.6) or (4.7), we have either

$$r(n+k', \mathbf{x}) \leq s_{n+k+1} + k < (1-\delta)(n+k')$$

or

$$r(n+2k, \mathbf{x}) \leq s_{n+k+1} + 2k + b\ell + k' < (1-\delta)(n+2k).$$

(3) If $\mathbf{z}_{n+k+1} = U'UV^bUUV^bU\ldots$, then by (4.5)

$$\begin{aligned} r(n+2k+b\ell+k', \mathbf{x}) &\leq s_{n+k+1} + 2k + b\ell \\ &< (1-\delta)(n+2k+b\ell) \leq (1-\delta)(n+2k+b\ell+k'). \end{aligned}$$

(4) If $\mathbf{z}_{n+k+1} = U'UV^bUV^bU\ldots$, then by (4.5)

$$\begin{aligned} r(n+2k+b\ell, \mathbf{x}) &\leq s_{n+k+1} + k + b\ell + k' \\ &\leq s_{n+k+1} + 2k + b\ell < (1-\delta)(n+2k+b\ell). \end{aligned}$$

(5) If $\mathbf{z}_{n+k+1} = U'V^bUUU\ldots$, then by (4.8)

$$r(n+2k, \mathbf{x}) \leq s_{n+k+1} + k + b\ell + k' < s_{n+k+1} + 2k + b\ell < (1-\delta)(n+2k).$$

(6) If $\mathbf{z}_{n+k+1} = U'V^bUUV^bU\ldots$, then by (4.7) or (4.6), we have either

$$r(n+(b-1)\ell, \mathbf{x}) \leq s_{n+k+1} + \ell + k' \leq (1-\delta)n \leq (1-\delta)(n+(b-1)\ell)$$

or

$$r(n+k+b\ell+k', \mathbf{x}) \leq s_{n+k+1} + 2k + b\ell < (1-\delta)(n+k+b\ell+k').$$

(7) If $\mathbf{z}_{n+k+1} = U'V^bUV^bU\dots$, then by (4.5)

$$\begin{aligned} r(n+k+b\ell+k', \mathbf{x}) &\le s_{n+k+1} + k + b\ell \\ &< (1-\delta)(n+k+b\ell) \le (1-\delta)(n+k+b\ell+k'). \end{aligned}$$

(8) If $\mathbf{z}_{n+k+1} = V'V^cUU\dots$, then by (4.8)

$$r(n+k, \mathbf{x}) \le s_{n+k+1} + k + c\ell + \ell' \le s_{n+k+1} + k + b\ell < (1-\delta)(n+k).$$

(9) If $\mathbf{z}_{n+k+1} = V'V^cUV^bU\dots$, then by (4.8)

$$r(n+k+c\ell+\ell', \mathbf{x}) \le s_{n+k+1} + k + b\ell < (1-\delta)(n+k) \le (1-\delta)(n+k+c\ell+\ell').$$

□

Proof of Theorem 1.4. By Proposition 3.6, $\mathbf{x}$ is either of bounded cycle with multiplicity 1 for large n_i or unbounded cycle with infinitely many n_i's with $|U_{(i)}| > |V_{(i)}|$. By Propositions 4.3 and 4.5, we get the first assertion. Set

$$\operatorname{rep}(\mathbf{x}) = \liminf_{n\to\infty} \frac{r(n,\mathbf{x})}{n}.$$

By [8, Lemma 3.6] (recall that the definition of rep is different there), we have

$$\mu\Big(\sum_{j\ge 1} \frac{x_j}{b^j}\Big) \ge 1 + \frac{1}{\operatorname{rep}(\mathbf{x})}.$$

This implies the second assertion. □

Acknowledgment

The authors are thankful to Julien Cassaigne for introducing them to Aberkane's papers. The second author's work was supported by the JSPS KAKENHI Grant Number 24K06641. The third author was supported by the National Research Foundation of Korea (NRF-2018R1A2B6001624, RS-2023-00245719).

References

[1] Ali Aberkane, *Exemples de suites de complexité inférieure à 2n*, Bull. Belg. Math. Soc. Simon Stevin **8** (2001), no. 2, 161–180 (French, with French summary). Journées Montoises d'Informatique Théorique (Marne-la-Vallée, 2000). MR1838940

[2] ______, *Words whose complexity satisfies* $\lim \frac{p(n)}{n} = 1$, Theoret. Comput. Sci. **307** (2003), no. 1, 31–46, DOI 10.1016/S0304-3975(03)00091-4. Words. MR2014729

[3] Boris Adamczewski, *On the expansion of some exponential periods in an integer base*, Math. Ann. **346** (2010), no. 1, 107–116, DOI 10.1007/s00208-009-0391-z. MR2558889

[4] Boris Adamczewski and Yann Bugeaud, *Nombres réels de complexité sous-linéaire: mesures d'irrationalité et de transcendance*, J. Reine Angew. Math. **658** (2011), 65–98, DOI 10.1515/CRELLE.2011.061 (French, with English summary). MR2831513

[5] Pierre Arnoux and Gérard Rauzy, *Représentation géométrique de suites de complexité* $2n+1$, Bull. Soc. Math. France **119** (1991), no. 2, 199–215 (French, with English summary). MR1116845

[6] Jean-Paul Allouche and Jeffrey Shallit, *Automatic sequences*, Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations. MR1997038

[7] Valérie Berthé, Charles Holton, and Luca Q. Zamboni, *Initial powers of Sturmian sequences*, Acta Arith. **122** (2006), no. 4, 315–347, DOI 10.4064/aa122-4-1. MR2234421

[8] Yann Bugeaud and Dong Han Kim, *On the b-ary expansions of* $\log(1+\frac{1}{a})$ *and* e, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **17** (2017), no. 3, 931–947. MR3726831

[9] ______, *A new complexity function, repetitions in Sturmian words, and irrationality exponents of Sturmian numbers*, Trans. Amer. Math. Soc. **371** (2019), no. 5, 3281–3308, DOI 10.1090/tran/7378. MR3896112

[10] ______, *On the b-ary expansion of a real number whose irrationality exponent is close to* 2, available at `arXiv:2510.02059[math.NT]`. preprint.

[11] Julien Cassaigne and François Nicolas, *Factor complexity*, Combinatorics, automata and number theory, Encyclopedia Math. Appl., vol. 135, Cambridge Univ. Press, Cambridge, 2010, pp. 163–247. MR2759107

[12] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics* (V. Berthé, S. Ferenczi, C. Mauduit, and A. Siegel, eds.), Lecture Notes in Mathematics, vol. 1794, Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel. MR1970385

[13] Alex Heinis, *The* $P(n)/n$*-function for bi-infinite words*, Theoret. Comput. Sci. **273** (2002), no. 1-2, 35–46, DOI 10.1016/S0304-3975(00)00432-1. WORDS (Rouen, 1999). MR1872441

[14] M. Lothaire, *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, 2002. MR1905123

[15] Marston Morse and Gustav A. Hedlund, *Symbolic dynamics II. Sturmian trajectories*, Amer. J. Math. **62** (1940), 1–42, DOI 10.2307/2371431. MR0000745

[16] Gérard Rauzy, *Suites à termes dans un alphabet fini*, Seminar on number theory, 1982–1983 (Talence, 1982/1983), Univ. Bordeaux I, Talence, 1983, pp. Exp. No. 25, 16 (French). MR750326

I.R.M.A., UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg Cedex, France

Institut universitaire de France

Email address: bugeaud@math.unistra.fr

Institute of Mathematics, University of Tsukuba, 1-1-1 Tenodai, Tsukuba, Ibaraki, 305-8571, Japan

Email address: kanekoha@math.tsukuba.ac.jp

Department of Mathematics Education, Dongguk University, 30 Pildong-ro 1-gil, Jung-gu, Seoul 04620, Korea.

Email address: kim2010@dgu.ac.kr