A NEW COMPLEXITY FUNCTION, REPETITIONS IN STURMIAN WORDS, AND IRRATIONALITY EXPONENTS OF STURMIAN NUMBERS

YANN BUGEAUD AND DONG HAN KIM

Abstract. We introduce and study a new complexity function in combinatorics on words, which takes into account the smallest second occurrence time of a factor of an infinite word. We characterize the eventually periodic words and the Sturmian words by means of this function. Then, we establish a new result on repetitions in Sturmian words and show that it is best possible. Let $b \geq 2$ be an integer. We deduce a lower bound for the irrationality exponent of real numbers whose sequence of b-ary digits is a Sturmian sequence over $\{0, 1, \ldots, b-1\}$ and we prove that this lower bound is best possible. As an application, we derive some information on the b-ary expansion of $\log(1 + \frac{1}{a})$, for any integer $a \geq 34$.

1. Introduction

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,

$$p(n, \mathbf{x}) = \#\{x_k x_{k+1} \ldots x_{k+n-1} : k \geq 1\}, \quad n \geq 1.$$

Clearly, we have

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

A celebrated theorem by Morse and Hedlund [36] characterizes the eventually periodic words by means of the subword complexity function.

Theorem 1.1. *Let $\mathbf{x} = x_1 x_2 \ldots$ be an infinite word. The following statements are equivalent:*

2010 *Mathematics Subject Classification.* 68R15 (primary); 11A63, 11J82 (secondary).

Key words and phrases. Combinatorics on words, Sturmian word, complexity, b-ary expansion.

(i) $\mathbf{x}$ *is eventually periodic;*

(ii) There exists a positive integer n *with* $p(n, \mathbf{x}) \leq n$*;*

(iii) There exists M *such that* $p(n, \mathbf{x}) \leq M$ *for* $n \geq 1$*.*

Therefore, the least possible subword complexity for a non eventually periodic infinite word $\mathbf{x}$ is given by $p(n, \mathbf{x}) = n + 1$ for every $n \geq 1$.

Definition 1.2. A Sturmian word is an infinite word $\mathbf{x}$ such that $p(n, \mathbf{x}) = n + 1$ for every $n \geq 1$.

There are uncountably many Sturmian words. There are several ways for describing them, one of them is given at the beginning of Section 3.

In the present paper, we introduce and study a new complexity function, which takes into account the smallest second occurrence time of a factor of $\mathbf{x}$. For an infinite word $\mathbf{x} = x_1 x_2 \ldots$ set

$$r(n, \mathbf{x}) = \min\{m \geq 1 : x_i^{i+n-1} = x_{m-n+1}^m \text{ for some } i \text{ with } 1 \leq i \leq m - n\}.$$

Here and below, for integers i, j with $i \leq j$, we write x_i^j for the factor $x_i x_{i+1} \ldots x_j$ of $\mathbf{x}$.

Said differently, $r(n, \mathbf{x})$ denotes the length of the smallest prefix of $\mathbf{x}$ containing two (possibly overlapping) occurrences of some word of length n.

One of the purposes of the present work is to characterize the eventually periodic words and the Sturmian words by means of the function $n \mapsto r(n, \mathbf{x})$. This is the object of Theorems 2.3 and 2.4.

In Section 3, by means of a precise combinatorial study of Sturmian words, we establish that every Sturmian word $\mathbf{s}$ satisfies

$$\liminf_{n \to +\infty} \frac{r(n, \mathbf{s})}{n} \leq \sqrt{10} - \frac{3}{2}. \tag{1.1}$$

A similar result also follows from Theorem 2.1 of [22], but with $\sqrt{10} - \frac{3}{2}$ replaced by a larger value strictly less than 2. We prove that the inequality (1.1) is best possible by constructing explicitly a Sturmian word $\mathbf{s}$ for which we have equality in (1.1).

By Sturmian number, we mean a real number for which there exists an integer base $b \geq 2$ such that its b-ary expansion is a Sturmian sequence over $\{0, 1, \ldots, b-1\}$. We show in Section 4 how it easily follows from (1.1) that the irrationality exponent of any Sturmian number is at least equal to $\frac{5}{3} + \frac{4\sqrt{10}}{15}$. We establish that this lower bound is best possible and, more generally, that the irrationality exponent of any Sturmian number can be read on its b-ary expansion (which means that infinitely many of its very good rational approximants can be constructed by cutting its b-ary expansion and completing by periodicity; see below Theorem 4.3).

Combined with earlier results of Alladi and Robinson [7], our result implies that, for any integer $b \geq 2$, the tail of the b-ary expansion of $\log(1 + \frac{1}{a})$, viewed as an infinite word over $\{0, 1, \dots, b-1\}$, cannot be a Sturmian word when $a \geq 34$ is an integer.

The present paper illustrates the fruitful interplay between combinatorics on words and Diophantine approximation, which has already led recently to several progresses. It is organized as follows. Our new results are stated in Sections 2 to 4 and proved in Sections 5 to 8. We consider in Section 9 a recurrence function studied by Cassaigne in [24]. The link between the function $n \mapsto r(n, \mathbf{x})$ and other combinatorial exponents is discussed in Section 10.

2. A new characterization of periodic and Sturmian words

We begin this section by stating some immediate properties of the function $n \mapsto r(n, \mathbf{x})$.

Lemma 2.1. *For an arbitrary infinite word $\mathbf{x}$ written over a finite alphabet $\mathcal{A}$, we have:*

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

(ii) There exists a unique integer j such that $x_j^{j+n-1} = x_{r(n,\mathbf{x})-n+1}^{r(n,\mathbf{x})}$ *and* $1 \leq j \leq r(n, \mathbf{x}) - n$.

(iii) $r(n + 1, \mathbf{x}) \geq r(n, \mathbf{x}) + 1, \quad (n \geq 1)$.

Let $b \geq 2$ and $n \geq 1$ be integers. A de Bruijn word of order n over an alphabet of cardinality b is a word of length $b^n + n - 1$ in which every block of length n occurs exactly once. Every de Bruijn word of order n over an alphabet with at least three letters can be extended to a de Bruijn word of order $n + 1$ (see e.g. [26, 31, 13]). When $|\mathcal{A}| \geq 3$, this establishes the existence of infinite words $\mathbf{x}$ satisfying $r(n, \mathbf{x}) = |\mathcal{A}|^n + n$, for every $n \geq 1$. Thus, we can have equality in the right hand side of (i) for every $n \geq 1$.

The lemma below shows that $r(n, \mathbf{x})$ is bounded from above in terms of the subword complexity function of $\mathbf{x}$.

Lemma 2.2. *For any infinite word* $\mathbf{x}$*, we have*

$$r(n, \mathbf{x}) \le p(n, \mathbf{x}) + n, \quad n \ge 1.$$

Proof. By the definition of $r(n, \mathbf{x})$, all the $r(n, \mathbf{x}) - 1 - (n-1)$ factors of length n of $x_1^{r(n,\mathbf{x})-1}$ are distinct. Since $x_{r(n,\mathbf{x})-n+1}^{r(n,\mathbf{x})}$ is a factor of $x_1^{r(n,\mathbf{x})-1}$, we have

$$p(n, \mathbf{x}) \ge p(n, x_1^{r(n,\mathbf{x})-1}) = p(n, x_1^{r(n,\mathbf{x})}) = r(n, \mathbf{x}) - n. \qquad \square$$

We stress that there is no analogue upper bound for the subword complexity function of $\mathbf{x}$ in terms of $r(n, \mathbf{x})$. Indeed, any infinite word $\mathbf{x} = x_1 x_2 \ldots$ over a finite alphabet $\mathcal{A}$ and such that

$$x_1 \ldots x_{2^j} = x_{2^{j+1}+2^j+1} \ldots x_{2^{j+2}}, \text{ for } j \ge 1,$$

satisfies $r(2^j, \mathbf{x}) \le 2^{j+2}$ for $j \ge 1$, thus $r(n, \mathbf{x}) \le 8n$ for every $n \ge 1$. However, by a suitable choice of $x_{2^j+1}, \ldots, x_{2^{j+1}+2^j}$, we can guarantee that $p(n, \mathbf{x}) = |\mathcal{A}|^n$ for every $n \ge 1$.

Our first result is a characterization of eventually periodic words by means of the function $n \mapsto r(n, \mathbf{x})$. It is the analogue of Theorem 1.1.

Theorem 2.3. *Let* $\mathbf{x} = x_1 x_2 \ldots$ *be an infinite word. The following statements are equivalent:*

(i) $\mathbf{x}$ *is eventually periodic;*

(ii) $r(n, \mathbf{x}) \le 2n$ *for all sufficiently large integers* n*;*

(iii) There exists M *such that* $r(n, \mathbf{x}) - n \le M$ *for* $n \ge 1$.

Our second result is a characterization of Sturmian words by means of the function $n \mapsto r(n, \mathbf{x})$.

Theorem 2.4. *Let* $\mathbf{x} = x_1 x_2 \ldots$ *be an infinite word. The following statements are equivalent:*

(i) $\mathbf{x}$ *is a Sturmian word;*

(ii) $\mathbf{x}$ *satisfies* $r(n, \mathbf{x}) \le 2n + 1$ *for* $n \ge 1$*, with equality for infinitely many* n.

It is possible to precisely describe the sequence $(r(n, \mathbf{x}))_{n\geq 1}$ for some classical infinite words $\mathbf{x}$, including the Fibonacci word and the Thue-Morse word. The proofs of the next results can be obtained by induction.

Let $\mathbf{f}$ denote the Fibonacci word $\mathbf{f} = 01001010\ldots$ over $\{0,1\}$ and $(F_n)_{n\geq 0}$ the Fibonacci sequence given by $F_0 = 0$, $F_1 = 1$ and $F_{n+2} = F_{n+1} + F_n$ for $n \geq 0$. The Fibonacci word is a Sturmian word and it satisfies $r(m, \mathbf{f}) = F_n + m$ for $F_n - 2 < m \leq F_{n+1} - 2$ and $n \geq 3$.

Let $\mathbf{t} = 01101001\ldots$ denote the Thue–Morse word over $\{0,1\}$. Then, we have $r(1, \mathbf{t}) = 3$ and $r(2^n - m, \mathbf{t}) = 5 \cdot 2^{n-1} - m$, if $0 \leq m < 2^{n-1}$ and $n \geq 1$.

There are several ways to measure the complexity of an infinite word $\mathbf{x}$, beside the functions $n \mapsto p(n, \mathbf{x})$ and $n \mapsto r(n, \mathbf{x})$ already mentioned; see, for instance, [32]. One can also consider the return time function $n \mapsto R(n, \mathbf{x})$, which indicates the first return time of the prefix of length n of $\mathbf{x}$. The characterization of Sturmian words by means of the function $n \mapsto R(n, \mathbf{x})$ is studied in [33]. The main drawback is that $R(\cdot, \mathbf{x})$ is defined only when $\mathbf{x}$ is a recurrent word. Indeed, if $\mathbf{x}$ is an infinite word over a finite alphabet and a is a letter, then the fact that $R(n, \mathbf{x})$ is well defined does not imply that $R(n, a\mathbf{x})$ is also defined; however, we always have

$$r(n-1, \mathbf{x}) + 1 \leq r(n, a\mathbf{x}) \leq r(n, \mathbf{x}) + 1.$$

3. Combinatorial study of Sturmian and quasi-Sturmian words

We begin by a classical result on Sturmian words.

Theorem 3.1. *Let θ and ρ be real numbers with $0 < \theta < 1$ and θ irrational. For $n \geq 1$, set*

$$s_n := \lfloor (n+1)\theta + \rho \rfloor - \lfloor n\theta + \rho \rfloor, \quad s'_n := \lceil (n+1)\theta + \rho \rceil - \lceil n\theta + \rho \rceil,$$

and define the infinite words

$$\mathbf{s}_{\theta,\rho} := s_1 s_2 s_3 \ldots, \quad \mathbf{s}'_{\theta,\rho} := s'_1 s'_2 s'_3 \ldots$$

Then we have

$$p(n, \mathbf{s}_{\theta,\rho}) = p(n, \mathbf{s}'_{\theta,\rho}) = n + 1, \quad \textit{for } n \geq 1.$$

The infinite words $\mathbf{s}_{\theta,\rho}$ and $\mathbf{s}'_{\theta,\rho}$ are called the Sturmian words with slope θ and intercept ρ. *Conversely, for every infinite word $\mathbf{x}$ on $\{0,1\}$ such that $p(n, \mathbf{x}) = n+1$*

for $n \geq 1$, there exist real numbers $\theta_{\mathbf{x}}$ and $\rho_{\mathbf{x}}$ with $0 < \theta_{\mathbf{x}} < 1$ and $\theta_{\mathbf{x}}$ irrational, such that $\mathbf{x} = \mathbf{s}_{\theta_{\mathbf{x}},\rho_{\mathbf{x}}}$ or $\mathbf{s}'_{\theta_{\mathbf{x}},\rho_{\mathbf{x}}}$.

For θ and ρ as in Theorem 3.1 the words $\mathbf{s}_{\theta,\rho}$ and $\mathbf{s}'_{\theta,\rho}$ differ only by at most two letters. Classical references on Sturmian words include [30, Chapter 6], [35, Chapter 2], and [9, Chapter 9].

The function $n \mapsto r(n, \mathbf{x})$ motivates the introduction of the exponent of repetition of an infinite word. Although the term 'repetition' usually refers to consecutive copies of the same word, we have decided to use it in our context, where we allow overlaps.

Definition 3.2. The exponent of repetition of an infinite word $\mathbf{x}$, denoted by $\mathrm{rep}(\mathbf{x})$, is defined by

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

A combinatorial study of Sturmian words whose slope has an unbounded sequence of partial quotients in its continued fraction expansion has been made in Section 11 of [4].

Theorem 3.3. *Let $\mathbf{s}$ be a Sturmian word. If its slope has an unbounded sequence of partial quotients in its continued fraction expansion, then* $\mathrm{rep}(\mathbf{s}) = 1$.

Theorem 3.3 follows from the proof of [4, Proposition 11.1]. For the sake of completeness, we provide an alternative (in our opinion, simpler) proof in Section 7.

A result of Berthé, Holton, and Zamboni [14] on the initial critical exponent (see Definition 10.1 below) of Sturmian words implies straightforwardly that, for every Sturmian word $\mathbf{s}$, there exists a positive real number $\delta(\mathbf{s})$ such that

$$\mathrm{rep}(\mathbf{s}) \leq 2 - \delta(\mathbf{s}).$$

However, the infimum of $\delta(\mathbf{s})$ taken over all the Sturmian words $\mathbf{s}$ is equal to 0. The purpose of the next result is to show that the exponents of repetition of Sturmian words are uniformly bounded from above by some constant strictly less than 2.

Theorem 3.4. *Every Sturmian word $\mathbf{s}$ satisfies*

$$\mathrm{rep}(\mathbf{s}) \leq \sqrt{10} - \frac{3}{2} = 1.6622776\ldots.$$

Moreover, if a Sturmian word $\mathbf{s}'$ *satisfies*

$$\mathrm{rep}(\mathbf{s}') = \sqrt{10} - \frac{3}{2}, \tag{3.1}$$

then the continued fraction expansion of the slope of $\mathbf{s}'$ *is eventually periodic and of the form* $[0; a_1, a_2, \ldots, a_K, \overline{2, 1, 1}]$ *for some integer* K.

It was tempting to conjecture that the upper bound $\sqrt{10} - \frac{3}{2}$ in Theorem 3.4 could be replaced by the Golden Ratio $\varphi := \frac{1+\sqrt{5}}{2}$ (note that the Fibonacci word $\mathbf{f}$ satisfies $\mathrm{rep}(\mathbf{f}) = \varphi$). However, we establish in Section 7 that Theorem 3.4 is best possible, by giving an explicit example of a Sturmian word whose exponent of repetition is equal to $\sqrt{10} - \frac{3}{2}$. For example, the Sturmian word $\mathbf{s}'$ of slope $\frac{\sqrt{10}-2}{3} = [0; \overline{2, 1, 1}]$ and intercept $\frac{1}{3}$ satisfies (3.1). A same kind of example has been already studied by Cassaigne [24]. We discuss Cassaigne's recurrence function $n \mapsto R'(n)$ in Section 9.

A more precise result is proved in Section 7. Namely, we establish a necessary and sufficient condition on a Sturmian word $\mathbf{s}'$ ensuring that $\mathrm{rep}(\mathbf{s}') = \sqrt{10} - \frac{3}{2}$ and give examples of such $\mathbf{s}'$. We also remark that $\sqrt{10} - \frac{3}{2}$ is an isolated point of the set of real numbers $\mathrm{rep}(\mathbf{s})$, where $\mathbf{s}$ runs over the Sturmian words.

Actually the conclusion of Theorem 3.4 remains true for a slightly larger class of words.

Definition 3.5. A quasi-Sturmian word $\mathbf{x}$ is an infinite word which satisfies

$$p(n, \mathbf{x}) = n + k, \quad \text{for } n \geq n_0.$$

A structure theorem of Cassaigne [25] allows us to deduce the next theorem from Theorem 3.4.

Theorem 3.6. *Every quasi-Sturmian word* $\mathbf{s}$ *satisfies* $\mathrm{rep}(\mathbf{s}) \leq \sqrt{10} - \frac{3}{2}$.

It can be deduced from Theorem 2.1 of [22] that every Sturmian or quasi-Sturmian word $\mathbf{s}$ satisfies $\mathrm{rep}(\mathbf{s}) \leq 1.83929\ldots$. The proof of Theorems 3.4 and 3.6 follows a completely different approach and yields a significant improvement.

We explain in the next section how Theorem 3.6 allows us to get new results on the b-ary expansion of real numbers whose irrationality exponent is slightly larger than 2.

4. Rational approximation of quasi-Sturmian numbers and applications

Ferenczi and Mauduit [28] studied the combinatorial properties of Sturmian words $\mathbf{s}$ and showed that, for some positive real number ε depending only on $\mathbf{s}$, they contain infinitely many $(2+\varepsilon)$-powers of blocks (that is, a block followed by itself and by its beginning of relative length at least ε) occurring not too far from the beginning. Then, by applying a theorem of Ridout [38] from transcendence theory, they deduce that, for any integer $b \geq 2$, the tail of the b-ary expansion of an irrational algebraic number, viewed as an infinite word over the alphabet $\{0, 1, \ldots, b-1\}$, cannot be a Sturmian word; see also [8].

Subsequently, Berthé, Holton and Zamboni [14] established that any Sturmian word $\mathbf{s}$, whose slope has a bounded continued fraction expansion, has infinitely many prefixes which are $(2+\varepsilon)$-powers of blocks, for some positive real number ε depending only on $\mathbf{s}$. This gives non-trivial information on the rational approximation to real numbers whose expansion in some integer base is a Sturmian word.

Definition 4.1. The irrationality exponent $\mu(\xi)$ of a real number ξ is the supremum of the real numbers μ such that the inequality

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

has infinitely many solutions in rational numbers $\frac{p}{q}$. If $\mu(\xi)$ is infinite, then ξ is called a Liouville number.

Recall that the irrationality exponent of an irrational number ξ is always at least equal to 2, with equality for almost all ξ, in the sense of the Lebesgue measure.

As observed in [1] (see also Section 8.5 of [17]), it follows from the results of [14] and [4] that, for any integer $b \geq 2$ and for any quasi-Sturmian word $\mathbf{s}$ over $\{0, 1, \ldots, b-1\}$, there exists a positive real number $\eta(\mathbf{s})$ such that the irrationality exponent of any real number whose b-ary expansion coincides with $\mathbf{s}$ is at least equal to $2 + \eta(\mathbf{s})$.

The reason for this is that, for an integer $b \geq 2$, there is a close connection between the exponent of repetition of an infinite word $\mathbf{x}$ written over $\{0, 1, \ldots, b-1\}$ and the irrationality exponent of the real number whose b-ary expansion is given by $\mathbf{x}$.

Theorem 4.2. *Let $b \geq 2$ be an integer and $\mathbf{x} = x_1 x_2 \dots$ an infinite word over $\{0, 1, \dots, b-1\}$, which is not eventually periodic. Then, the irrationality exponent of the irrational number $\xi_{\mathbf{x},b} := \sum_{k \geq 1} \frac{x_k}{b^k}$ satisfies*

$$\mu(\xi_{\mathbf{x},b}) \geq \frac{\mathrm{rep}(\mathbf{x})}{\mathrm{rep}(\mathbf{x}) - 1}, \tag{4.1}$$

where the right hand side is infinite if $\mathrm{rep}(\mathbf{x}) = 1$.

It immediately follows from Theorems 3.3 and 4.2 that any Sturmian number constructed from a Sturmian sequence whose slope has unbounded partial quotients is a Liouville number. This result was first established by Komatsu [34].

As mentioned in Section 3 for the related quantity $\delta(\mathbf{s})$, the infimum of $\eta(\mathbf{s})$ over all Sturmian words $\mathbf{s}$ is equal to 0 and one cannot deduce a non-trivial lower bound for the irrationality exponents of Sturmian numbers. We improve this as follows.

Theorem 4.3. *Let $b \geq 2$ be an integer. Let $\mathbf{s} = s_1 s_2 \dots$ be a Sturmian or a quasi-Sturmian word over $\{0, 1, \dots, b-1\}$. Then,*

$$\mu\Big(\sum_{j \geq 1} \frac{s_j}{b^j}\Big) \geq \frac{5}{3} + \frac{4\sqrt{10}}{15} = 2.5099\dots,$$

with equality when $\mathbf{s}$ is the Sturmian word $\mathbf{s}'$ defined in Theorem 3.4.

The first statement of Theorem 4.3 is an immediate consequence of Theorem 4.2 combined with Theorem 3.6. Its second statement directly follows from Theorem 4.5 below.

If there is equality in (4.1), we say that *the irrationality exponent of $\xi_{\mathbf{x},b}$ can be read on its b-ary expansion.* This is equivalent to say that, for every $\varepsilon > 0$, there exist positive integers r, s, with $r + s$ being arbitrarily large, such that

$$\Big|\xi_{\mathbf{x},b} - \frac{p_{r,s}}{b^r(b^s - 1)}\Big| \leq \frac{1}{b^{(r+s)(\mu(\xi_{\mathbf{x},b}) - \varepsilon)}},$$

where $p_{r,s}$ is the nearest integer to $b^r(b^s - 1)\xi_{\mathbf{x},b}$. Or, if one prefers, this is equivalent to say that, among the very good approximants to $\xi_{\mathbf{x},b}$, infinitely many of them can be constructed by cutting its b-ary expansion and completing by periodicity (this does not mean, however, that infinitely many convergents to $\xi_{\mathbf{x},b}$ have a denominator of the form $b^r(b^s - 1)$). Using the Diophantine exponent v'_b introduced in [10] (see also Section 7.1 of [17]), to say that *the irrationality exponent of $\xi_{\mathbf{x},b}$ can be read on its b-ary expansion.* simply means that $v'_b(\xi_{\mathbf{x},b}) = \mu(\xi_{\mathbf{x},b})$.

Let $b \geq 2$ be an integer. A covering argument shows that, for any positive real number ε, the set of real numbers ξ such that there are infinitely many integer triples $(r, s, p_{r,s})$ with $r \geq 0, s \geq 0$ and

$$\left| \xi - \frac{p_{r,s}}{b^r(b^s - 1)} \right| \leq \frac{1}{b^{(r+s)(1+\varepsilon)}},$$

has Lebesgue measure zero. Consequently, the b-ary expansion $\mathbf{x}_{\xi,b}$ of almost every real number ξ satisfies $\mathrm{rep}(\mathbf{x}_{\xi,b}) = +\infty$, thus the right-hand side of inequality (4.1) is equal to 1 almost always. This shows that, since the irrationality exponent of an irrational number is always at least equal to 2, it can only very rarely be read on its b-ary expansion. There are only few known examples for which this is the case; see [16, 23] and the following result of Adams and Davison [6] (additional references and a more detailed statement are given in Section 7.6 of [17]).

Theorem 4.4. *Let $b \geq 2$ be an integer and $\alpha = [a_1; a_2, a_3 \ldots]$ an irrational number greater than* 1*. The irrationality exponent of the real number*

$$\xi_{\alpha,b} = \sum_{j=1}^{+\infty} \frac{1}{b^{\lfloor j\alpha \rfloor}}$$

is given by

$$\mu(\xi_{\alpha,b}) = 1 + \limsup_{n \to +\infty} \, [a_n; a_{n-1}, \ldots, a_1].$$

Theorem 4.4 gives us the irrationality exponent of any real number whose expansion in some integer base is a *characteristic Sturmian word* (that is, a Sturmian word whose intercept is 0). It shows that equality holds in (4.1) when $\mathbf{x}$ is a characteristic Sturmian word. We extend this result in Section 8 by proving that the inequality in Theorem 4.2 is an equality for any Sturmian word $\mathbf{x}$ and any integer base $b \geq 2$.

Theorem 4.5. *Let $b \geq 2$ be an integer and $\mathbf{x} = x_1 x_2 \ldots$ a Sturmian word. Then, the irrationality exponent of the irrational number $\sum_{k \geq 1} \frac{x_k}{b^k}$ satisfies*

$$\mu\Big(\sum_{k \geq 1} \frac{x_k}{b^k}\Big) = \frac{\mathrm{rep}(\mathbf{x})}{\mathrm{rep}(\mathbf{x}) - 1},$$

where the right hand side is infinite if $\mathrm{rep}(\mathbf{x}) = 1$.

The proof of Theorem 4.5 rests on the theory of continued fractions combined with properties of the function $n \mapsto r(n, \mathbf{x})$ and of Sturmian words. Furthermore, a result obtained in the course of this proof implies that, given b and b' multiplicatively independent integers, an irrational real number cannot have simultaneously a Sturmian b-ary expansion and a Sturmian b'-ary expansion. This gives a partial answer to Problem 3 of [18]. We will return to this question in a subsequent work.

We display below a statement equivalent to Theorem 4.3, but we need first to introduce some notation. Let b denote an integer at least equal to 2. Any real number ξ has a unique b-ary expansion, that is, it can be uniquely written as

$$\xi = \lfloor \xi \rfloor + \sum_{\ell \geq 1} \frac{a_\ell}{b^\ell} = \lfloor \xi \rfloor + 0 \, . \, a_1 a_2 \ldots,$$

where $\lfloor \cdot \rfloor$ denotes the integer part function, the *digits* $a_1, a_2, \ldots$ are integers from the set $\{0, 1, \ldots, b-1\}$ and a_ℓ differs from $b-1$ for infinitely many indices ℓ. A natural way to measure the complexity of ξ is to count the number of distinct blocks of given length in the infinite word $\mathbf{a} = a_1 a_2 a_3 \ldots$ For $n \geq 1$, we set $p(n, \xi, b) = p(n, \mathbf{a})$ with $\mathbf{a}$ as above. Clearly, we have

$$p(n, \xi, b) = \#\{a_{\ell+1} a_{\ell+2} \ldots a_{\ell+n} : \ell \geq 0\}.$$

Theorem 4.6. *Every irrational real number ξ with $\mu(\xi) < \frac{5}{3} + \frac{4}{15}\sqrt{10}$ satisfies*

$$\lim_{n \to +\infty} \big(p(n, \xi, b) - n \big) = +\infty,$$

for every integer $b \geq 2$. Furthermore, for every integer $b \geq 2$, there exists an irrational real number ξ_b with $\mu(\xi_b) = \frac{5}{3} + \frac{4}{15}\sqrt{10}$ and $p(n, \xi_b, b) = n + 1$ for $n \geq 1$.

The conclusion of the first assertion of Theorem 4.6 was proved to be true for every irrational algebraic number ξ in [28], for every real number ξ whose irrationality exponent is equal to 2 in [1] (see also Section 8.5 of [17]; note that, by Roth's theorem [39], every irrational algebraic number satisfies $\mu(\xi) = 2$), and for every irrational real number ξ satisfying $\mu(\xi) < 2.19149\ldots$ in [22].

We can deduce from Theorem 4.6 some information on the b-ary expansion of several classes of real numbers, without knowing exactly their irrationality exponent. Recall that, for example, Alladi and Robinson [7] (who improved earlier results of A. Baker [12]) and Danilov [27] proved that, for any positive integer s, the irrationality exponents of $\log(1 + \frac{s}{t})$ and $\sqrt{t^2 - s^2} \arcsin \frac{s}{t}$ are bounded from above by functions of t which tend to 2 as the integer t tends to infinity. The next statement then follows at once from Corollary 1 of [7], which implies that the irrationality exponent of $\log(1+\frac{1}{a})$ is less than $\frac{5}{3}+\frac{4}{15}\sqrt{10}$ for every integer $a \geq 34$.

Corollary 4.7. *For every integer $b \geq 2$ and every integer $a \geq 34$, we have*

$$\lim_{n \to +\infty} \left(p\Big(n, \log\Big(1+\frac{1}{a}\Big), b\Big) - n \right) = +\infty,$$

For much larger values of a, a stronger result than the above corollary has been established in [22]. Namely, for any positive real number ε, there exists an integer a_0 such that, for every integer $b \geq 2$ and every integer $a \geq a_0$, we have

$$\liminf_{n \to +\infty} \frac{p\big(n, \log\big(1+\frac{1}{a}\big), b\big)}{n} \geq \frac{9}{8} - \varepsilon.$$

The approach followed in [22] gives a non-trivial result only when the integer a exceeds 23347.

5. Auxiliary combinatorial lemmas

The proofs of Theorems 2.3 and 2.4 rest on a series of combinatorial lemmas.

For a word $U = u_1 \ldots u_n$ composed of n letters, denote by $|U| = n$ its length and set

$$\Lambda(U) = \{1 \leq k < n : u_i = u_{i+k} \text{ for all } 1 \leq i \leq n-k\}.$$

An element of $\Lambda(U)$ is called a period of U. We stress that a period of a word of length n may not be a divisor of n. A finite word U is called *primitive* if there is no non-empty word V such that $U = V^n$ for some integer $n \geq 2$.

Lemma 5.1 (Fine and Wilf Theorem [29])**.** *Let $U = u_1 \ldots u_n$ and h, k be in $\Lambda(U)$. If $n \geq h + k - gcd(h,k)$, then U is periodic of period $gcd(h,k)$.*

Lemma 5.2. *Let $U = u_1 \ldots u_n$ be a finite word and λ in $\Lambda(U)$. Then $u_{n-\lambda+2}^n a$ with $a \neq u_{n-\lambda+1}$ is not a factor of U.*

Proof. Since λ is in $\Lambda(U)$, all the factors of length λ in U have the same number of a's. Since $u_{n-\lambda+1} \neq a$, the number of a's in $u_{n-\lambda+1}^n$ is one less than in $u_{n-\lambda+2}^n a$, thus the latter cannot be a factor of U. □

Lemma 5.3. *Let $\mathbf{x}$ be an infinite word and n an integer with $r(n,\mathbf{x}) \geq r(n-1,\mathbf{x}) + 2$. Then $r(n,\mathbf{x}) \geq 2n+1$.*

Proof. To shorten the notation, we simply write $r(\cdot)$ for $r(\cdot, \mathbf{x})$. Suppose that

$$r(n) \geq r(n-1) + 2 \quad \text{but} \quad r(n) \leq 2n.$$

Let s, ℓ be the nonnegative integers satisfying

$$x_{s+1}^{s+n-1} = x_{r(n-1)-n+2}^{r(n-1)}, \qquad x_{r(n)-n+1}^{r(n)} = x_{r(n)-n+1-\ell}^{r(n)-\ell} \tag{5.1}$$

with

$$0 \leq s \leq r(n-1) - n, \qquad 1 \leq \ell \leq r(n) - n \leq n. \tag{5.2}$$

Then, we have

$$x_{s+n} \neq x_{r(n-1)+1}, \tag{5.3}$$

for otherwise $r(n) = r(n-1) + 1$.

Since

$$r(n-1) - n - s + 1 \in \Lambda\left(x_{s+1}^{r(n-1)}\right),$$

by Lemma 5.2 and (5.3), the word $x_{n+s+1}^{r(n-1)+1}$ is not a factor of $x_{s+1}^{r(n-1)}$.

Our assumption implies $n+s+1 \geq r(n)-n+1$ and $r(n-1)+1 \leq r(n)$, thus by (5.1), we have $x_{n+s+1}^{r(n-1)+1} = x_{n+s+1-\ell}^{r(n-1)+1-\ell}$, which is not a factor of $x_{s+1}^{r(n-1)}$. Therefore, we have $n+s+1-\ell < s+1$, i.e., $n < \ell$, a contradiction to (5.2). □

Lemma 5.4. *Let* $\mathbf{x}$ *be an infinite word and* n *an integer such that* $r(n+1, \mathbf{x}) = r(n, \mathbf{x}) + 1$. *Let* j *be the integer satisfying* $1 \leq j < r(n, \mathbf{x}) - n + 1$ *and* $x_j^{j+n-1} = x_{r(n,\mathbf{x})-n+1}^{r(n,\mathbf{x})}$. *Then,* $x_{j+n} = x_{r(n,\mathbf{x})+1}$.

Proof. By assumption, there exists a unique integer h satisfying $1 \leq h < r(n+1, \mathbf{x}) - n$ and $x_h^{h+n} = x_{r(n+1,\mathbf{x})-n}^{r(n+1,\mathbf{x})}$. In particular, we have $x_h^{h+n-1} = x_{r(n+1,\mathbf{x})-n}^{r(n+1,\mathbf{x})-1}$, thus $h = j$ and $x_{j+n} = x_{r(n,\mathbf{x})+1}$. □

Lemma 5.5. *Let* $\mathbf{x}$ *be an infinite word satisfying* $r(i, \mathbf{x}) \leq 2i + 1$ *for all* $i \geq 1$. *Let* m, n *be positive integers such that* $r(n, \mathbf{x}) = 2n + 1$ *and* $m \geq 2n + 1$. *If* k *is the integer defined by* $r(k-1, \mathbf{x}) < m \leq r(k, \mathbf{x})$, *then* $k \geq n$ *and* $r(k, \mathbf{x}) - k \leq m - n$.

Proof. Write $r(\cdot)$ for $r(\cdot, \mathbf{x})$. Observe that $k \geq n$, since $r(n-1) < r(n) \leq m$. If $r(k) = m$, then we get $r(k) - k = m - k \leq m - n$, as required.

If $r(k-1) < m < r(k)$, then $r(k) \geq r(k-1)+2$ and we deduce from Lemma 5.3 that $r(k) = 2k+1$. Furthermore, we have $k \geq n+1$. Let $\ell = \min\{i \geq 1 : r(k-i) = 2(k-i)+1\}$. Since $r(n) = 2n+1$, the integer ℓ is well-defined and

$$k \geq n + \ell.$$

For $i = 1, \ldots, \ell-1$, we have $r(k-i) \leq 2(k-i)$ and it follows from Lemma 5.3 that $r(k-i) = r(k-i-1)+1$, thus,

$$r(k-1) - r(k-\ell) = \ell - 1.$$

Since $m \geq r(k-1)+1 = r(k-\ell)+\ell = r(k)-\ell$, we have

$$r(k) - k \leq (m+\ell) - (n+\ell) = m - n,$$

which completes the proof of the lemma. □

6. Proofs of Theorems 2.3 and 2.4

Proof of Theorem 2.3.

(iii) ⇒ (ii) : Immediate.

(ii) ⇒ (i) : It follows from Lemma 5.3 and Lemma 2.1 (iii) that there exists an integer n_0 such that $r(n+1, \mathbf{x}) = r(n, \mathbf{x}) + 1$ for every $n \geq n_0$. By Lemma 5.4, we deduce that there exists an integer j such that $x_{j+n} = x_{r(n_0,\mathbf{x})+n-n_0+1}$, for $n \geq n_0$. This shows that $\mathbf{x}$ is eventually periodic.

(i) ⇒ (iii) : Let r and s denote the length of the preperiod and that of the period of $\mathbf{x}$. Then, the infinite word starting at x_{r+1} is the same as the infinite word starting at x_{r+s+1}, thus we have $r(n, \mathbf{x}) \leq n + r + s$ for $n \geq 1$.

Proof of Theorem 2.4.

(i) ⇒ (ii) : The inequality is clear by Lemma 2.2 and Theorem 2.3 implies that there is equality for infinitely many n.

(ii) ⇒ (i) : Let n be an integer such that $r(n, \mathbf{x}) = 2n+1$. By the proof of Lemma 2.2 we have $p(n, x_1^{2n}) = n+1$.

Let m be an integer with $m \geq 2n+1$. Then, by Lemma 5.5, there exists an integer k such that $k \geq n$, $m \leq r(k, \mathbf{x})$ and $r(k, \mathbf{x}) - k \leq m - n$. By Lemma 2.1 (ii), we get that $x_{r(k,\mathbf{x})-k+1}^{r(k,\mathbf{x})} = x_{r(k,\mathbf{x})-k+1-j}^{r(k,\mathbf{x})-j}$ for some integer j with $1 \le j \le r(k,\mathbf{x}) - k$. Therefore, we have $x_{m-n+1}^{m} = x_{m-n+1-j}^{m-j}$, which implies that

$$p(n, x_1^m) = p(n, x_1^{m-1}).$$

Since this equality holds for every $m \ge 2n+1$ and $p(n, x_1^{2n}) = n+1$, we deduce that $p(n,\mathbf{x}) = n+1$. Thus, we have established the existence of arbitrary large integers n such that $p(n,\mathbf{x}) = n+1$. This shows that $\mathbf{x}$ is a Sturmian word.

7. Proof of Theorems 3.4 and 3.6

Through this section, we fix an infinite sequence $(a_k)_{k\ge1}$ of positive integers. We define inductively a sequence of words $(M_k)_{k\ge0}$ on the two letter-alphabet $\{0,1\}$ by the formulas

$$M_0 = 0, \quad M_1 = 0^{a_1-1}1 \text{ and } M_{k+1} = M_k^{a_{k+1}} M_{k-1} \qquad (k \ge 1). \tag{7.1}$$

It is easy to check that the last two letters of M_k are 10 (resp. 01) if k is even (resp. odd) and $|M_k| \ge 2$. This sequence converges to the infinite word

$$\mathbf{s}_{\theta,0} := \lim_{k\to+\infty} M_k = 0^{a_1-1}1\ldots,$$

which is usually called the characteristic Sturmian word of slope

$$\theta := [0; a_1, a_2, a_3, \ldots]$$

constructed over the alphabet $\{0,1\}$ (See e.g. [35]).

Let $\mathbf{x}$ be a Sturmian word of slope θ. We study the combinatorial properties of $\mathbf{x}$. An *admissible word* is a factor of $\mathbf{x}$ of finite length. Note that the set of factors of $\mathbf{x}$ is the same as that of $\mathbf{s}_{\theta,0}$ (see e.g. [35, Proposition 2.1.18]). Let $(\frac{p_\ell}{q_\ell})_{\ell\ge0}$ denote the sequence of convergents to the slope of $\mathbf{x}$. Then, for $k \ge 0$, we have $q_k = |M_k|$ and p_k is the number of digits 1 in M_k. It is known that only the last two letters of $M_{k+1}M_k$ and M_kM_{k+1} are different (see e.g. [35, Proposition 2.2.2]). For a non-empty finite word U, we write U^- for the word U deprived of its last letter. For $k \ge 1$, set

$$\tilde{M}_k = (M_kM_{k-1})^{--} = (M_{k-1}M_k)^{--}$$

and observe that $\tilde{M}_k$ is a prefix of M_{k+1}.

We will use the property that $M_{k+1}M_k$ and $M_{k+1}M_{k+1}M_k$ are primitive (see e.g. [35, Proposition 2.2.3]) in conjunction with the following lemma.

Lemma 7.1. *Let U be a primitive word. Then all the $|U|$ factors of length $|U|-1$ of UU^{--} are distinct.*

Proof. Assume that there are integers i, j with $0 \le i < j \le |U|-1$ and

$$(UU^{--})_{i+1}^{i+|U|-1} = (UU^{--})_{j+1}^{j+|U|-1}.$$

Then, $j-i$ and $|U|$ are periods of $(UU^{--})_{i+1}^{j+|U|-1}$ and

$$|U| + (j-i) - \gcd(|U|, j-i) \le |U| + j - i - 1.$$

Thus, we deduce from Lemma 5.1 that $(UU^{--})_{i+1}^{j+|U|-1}$ is periodic of period $\gcd(|U|, j-i)$. Since

$$\gcd(|U|, j-i) \le j-i \le |U|-1,$$

this contradicts the fact that U is primitive. □

The next lemma shows that repetitions occur near the beginning of any Sturmian word of slope θ.

Lemma 7.2. *Let $\mathbf{x}$ be a Sturmian word of slope θ. Then, for $k \ge 1$, there exists a unique word W_k satisfying*

(i) $\mathbf{x} = W_k M_k \tilde{M}_k \ldots$, where W_k is a non-empty suffix of M_k,

or

(ii) $\mathbf{x} = W_k M_{k-1} M_k \tilde{M}_k \ldots$, where W_k is a non-empty suffix of M_k,

or

(iii) $\mathbf{x} = W_k M_k \tilde{M}_k \ldots$, where W_k is a non-empty suffix of M_{k-1},

and all the $(2q_k + q_{k-1})$ cases are mutually exclusive.

Furthermore, if $\mathbf{x} = W_k M_{k-1} M_k \tilde{M}_k \ldots$ and W_k is a non-empty suffix of M_k, then $W_{k+1} = W_k M_{k-1}$. Moreover, if $\mathbf{x} = W_k M_k \tilde{M}_k \ldots$ and W_k is a non-empty suffix of M_{k-1}, then $W_{k+1} = W_k$.

Proof. We first claim that, for each $k \geq 1$, the word $M_k M_k M_{k-1} M_k \tilde{M}_k$ is admissible. This follows from the fact that $M_{k+3} M_{k+2}$ is admissible and

$$M_{k+3} = \cdots M_{k+2} M_{k+1} = \cdots M_k M_{k+1} = \cdots M_k M_k M_{k-1},$$
$$M_{k+2} = M_{k+1} M_k \cdots = M_k \tilde{M}_k \cdots .$$

Since $M_k M_k M_{k-1}$ is primitive, Lemma 7.1 implies that any admissible word of length $2q_k + q_{k-1} - 1$ is a factor of $M_k M_k M_{k-1} M_k \tilde{M}_k$. These admissible words are prefixes of $W M_k \tilde{M}_k$ or $W M_{k-1} M_k \tilde{M}_k$ for some non-empty W which is a suffix of M_k, and prefixes of $W M_k \tilde{M}_k$ for some non-empty W which is a suffix of M_{k-1}. Consequently, $\mathbf{x} = W M_k^- \ldots$ or $W M_{k-1} M_k^- \ldots$ with W which is a suffix of M_k or $\mathbf{x} = W M_k^- \ldots$ with W which is a suffix of or M_{k-1} .

Since there are two admissible words of length $2q_k + q_{k-1} - 1$ starting with M_k^-, namely $M_k M_{k-1} M_k^-$ and $M_k M_k M_{k-1}^-$, it follows that if $\mathbf{x} = U M_k^- \ldots$ for some U, then $\mathbf{x} = U M_k \tilde{M}_k \ldots$. Hence we conclude that $\mathbf{x} = W M_k \tilde{M}_k \ldots$ or $W M_{k-1} M_k \tilde{M}_k \ldots$ with W which is a suffix of M_k or $\mathbf{x} = W M_k \tilde{M}_k \ldots$ with W which is a suffix of or M_{k-1}. Putting $W_k = W$, we see that W_k satisfies one of the cases (i), (ii), (iii), which are mutually exclusive by Lemma 7.1.

By the first assertion of the lemma, $\mathbf{x}$ starts with $W_{k+1} \tilde{M}_{k+1}$, where W_{k+1} is a non-empty suffix of M_{k+1} or M_k. If W_{k+1} is a suffix of M_k, then put $W' = W_{k+1}$, thus

$$\mathbf{x} = W_{k+1} \tilde{M}_{k+1} \cdots = W' M_k \tilde{M}_k \ldots .$$

If W_{k+1} is a suffix of $M_{k+1} = M_k^{a_{k+1}} M_{k-1}$, then $W_{k+1} = W' M_k^t M_{k-1}$ for some integer $t \geq 0$ and a non-empty suffix W' of M_k or W_{k+1} is a non-empty suffix of M_{k-1}. If $W_{k+1} = W' M_k^t M_{k-1}$, with W' a suffix of M_k, then

$$\mathbf{x} = W_{k+1} \tilde{M}_{k+1} \cdots = \begin{cases} W' M_k \tilde{M}_k \ldots, & \text{if } t \geq 1, \\ W' M_{k-1} M_k \tilde{M}_k \ldots, & \text{if } t = 0. \end{cases}$$

If W_{k+1} is a suffix of M_{k-1}, then put $W' = W_{k+1}$, thus

$$\mathbf{x} = W_{k+1} \tilde{M}_{k+1} \cdots = W' M_k \tilde{M}_k \ldots .$$

By the first assertion of the lemma, we conclude that $W' = W_k$. If $\mathbf{x} = W' M_{k-1} M_k \tilde{M}_k \ldots$, then $W_{k+1} = W' M_{k-1}$ and if W' is a suffix of M_{k-1}, then $W_{k+1} = W'$. $\square$

We are now in position to establish Theorems 3.3 and 3.4.

Proof of Theorem 3.3.

Let k and t be large integers such that $M_k = (M_{k-1})^t M_{k-2}$. Let ℓ be the integer part of $\sqrt{t}$. We distinguish two cases. If $|W_k| > (\ell+1)|M_{k-1}|$, then

$$r((\ell-1)|M_{k-1}|, \mathbf{x}) \le \ell |M_{k-1}|$$

and, otherwise, we check that

$$r((t-1)|M_{k-1}|, \mathbf{x}) \le |W_k| + t|M_{k-1}| \le (t+\ell+1)|M_{k-1}|.$$

As k and t can be taken arbitrarily large, we deduce that $\mathrm{rep}(\mathbf{x}) = 1$.

Further auxiliary results for the proof of Theorem 3.4.

Lemma 7.3. *If $\mathbf{x} = UV\ldots$ where V is a factor of $M_k \tilde{M}_{k+1}$ such that $|V| > q_k$, then we have*

$$r(|V| - q_k, \mathbf{x}) \le |UV|.$$

Proof. Let $V = v_1 \ldots v_n$ be a factor of $M_k \tilde{M}_{k+1}$ such that $|V| = n > q_k$. Since $M_k \tilde{M}_{k+1} = M_k \ldots M_k M_{k-1}^{--}$ and M_{k-1} is a prefix of M_k, we get $v_1^{n-q_k} = v_{1+q_k}^n$. Thus we have $r(n - q_k, \mathbf{x}) \le |UV|$. □

We establish two further lemmas on the combinatorial structure of Sturmian words. For $k \ge 1$, we set

$$\eta_k := \frac{q_{k-1}}{q_k}, \qquad t_k := \frac{|W_k|}{q_k}, \qquad \varepsilon_k := \frac{2}{q_k}.$$

Recall that φ denotes the Golden Ratio $\frac{1+\sqrt{5}}{2}$.

In the rest of the proof of the theorem, we assume that k is large enough to ensure that $q_{k-2} \ge 6$, thus, $\varepsilon_k < \eta_k$, $\varepsilon_k < \frac{1}{6}$ and $\varepsilon_k < \frac{1-\eta_k}{2}$.

Lemma 7.4. *(i) If $\mathbf{x} = W_k M_k \tilde{M}_k \ldots$, where W_k is a suffix of M_k, then $\frac{r(n,\mathbf{x})}{n} < \varphi + 2\varepsilon_k$ for some n with $q_k - 2 \le n \le |W_k| + q_k + q_{k-1} - 2$.*

(ii) If $\mathbf{x} = W_k M_k \tilde{M}_k \ldots$, where W_k is a suffix of M_{k-1}, then $\frac{r(n,\mathbf{x})}{n} < \varphi + 2\varepsilon_k$ for some n with $|W_k| + q_k - 2 \le n \le |W_k| + q_{k+1} + q_k - 2$.

Proof. (i) Since $W_k M_k \tilde{M}_k$ is a factor of $M_k \tilde{M}_{k+1} = M_k M_k M_{k+1}^{--}$, we have by Lemma 7.3 $r(|W_k M_k \tilde{M}_k| - q_k, \mathbf{x}) \le |W_k M_k \tilde{M}_k|$, which yields that

$$\begin{aligned} \frac{r(|W_k| + q_k + q_{k-1} - 2, \mathbf{x})}{|W_k| + q_k + q_{k-1} - 2} &\le \frac{|W_k| + 2q_k + q_{k-1} - 2}{|W_k| + q_k + q_{k-1} - 2} \\ &= 1 + \frac{1}{t_k + 1 + \eta_k - \varepsilon_k} < 1 + \frac{1}{t_k + 1 + \eta_k} + \varepsilon_k. \end{aligned} \tag{7.2}$$

Furthermore, $\mathbf{x} = W_k M_k \tilde{M}_k \cdots = W_k \tilde{M}_k \ldots$, thus we have by Lemma 7.3 $r(|\tilde{M}_k| - q_{k-1}, \mathbf{x}) \le |W_k \tilde{M}_k|$, which yields that

$$\frac{r(q_k - 2, \mathbf{x})}{q_k - 2} \le \frac{|W_k| + q_k + q_{k-1} - 2}{q_k - 2} = 1 + \frac{t_k + \eta_k}{1 - \varepsilon_k} < 1 + t_k + \eta_k + 2\varepsilon_k. \tag{7.3}$$

Since for every positive real number x we have $\min(x, \frac{1}{1+x}) \le \frac{1}{\varphi}$, we derive from (7.2) and (7.3) that

$$\frac{r(n, \mathbf{x})}{n} < \varphi + 2\varepsilon_k \text{ for some } n \text{ with } q_k - 2 \le n \le |W_k| + q_k + q_{k-1} - 2.$$

(ii) Since $\mathbf{x} = W_k M_k \tilde{M}_k \cdots = W_k \tilde{M}_k \ldots$ and $W_k \tilde{M}_k$ is a factor of $M_{k-1} \tilde{M}_k$, by Lemma 7.3 we have $r(|W_k \tilde{M}_k| - q_{k-1}, \mathbf{x}) \le |W_k \tilde{M}_k|$, which yields that

$$\begin{aligned} \frac{r(|W_k| + q_k - 2, \mathbf{x})}{|W_k| + q_k - 2} &\le \frac{|W_k| + q_k + q_{k-1} - 2}{|W_k| + q_k - 2} \\ &= 1 + \frac{\eta_k}{t_k + 1 - \varepsilon_k} < 1 + \frac{\eta_k}{t_k + 1} + \varepsilon_k. \end{aligned} \tag{7.4}$$

Since W_k is a suffix of M_{k-1} which is a suffix of M_{k+1}, we deduce from Lemma 7.2 that $\mathbf{x}$ starts with either $W_{k+1} M_{k+1} \tilde{M}_{k+1}$ or $W_{k+1} M_k M_{k+1} \tilde{M}_{k+1}$, where $W_{k+1} = W_k$. If $\mathbf{x} = W_{k+1} M_{k+1} \tilde{M}_{k+1} \ldots$, then the proof is completed by (i) since $q_{k+1} \ge |W_k| + q_k$ and $|W_{k+1}| = |W_k|$. If $\mathbf{x} = W_{k+1} M_k M_{k+1} \tilde{M}_{k+1} \cdots = W_{k+1} M_k \tilde{M}_{k+1} \ldots$, then by Lemma 7.3 we obtain

$$r(|M_k \tilde{M}_{k+1}| - q_k, \mathbf{x}) \le |W_{k+1} M_k \tilde{M}_{k+1}|,$$

thus,

$$\begin{aligned} \frac{r(q_{k+1} + q_k - 2, \mathbf{x})}{q_{k+1} + q_k - 2} &\le \frac{|W_{k+1}| + q_{k+1} + 2q_k - 2}{q_{k+1} + q_k - 2} = 1 + \frac{|W_{k+1}| + q_k}{q_{k+1} + q_k - 2} \\ &\le 1 + \frac{|W_k| + q_k}{2q_k + q_{k-1} - 2} = 1 + \frac{t_k + 1}{2 + \eta_k - \varepsilon_k} < 1 + \frac{t_k + 1}{2 + \eta_k} + \varepsilon_k. \end{aligned}$$

Combined with (7.4), we deduce that there exists an integer n with $|W_k| + q_k - 2 \le n \le q_{k+1} + q_k - 2$ and

$$\frac{r(n, \mathbf{x})}{n} \le 1 + \sqrt{\frac{\eta_k}{2 + \eta_k}} + \varepsilon_k < 1 + \frac{1}{\sqrt{3}} + \varepsilon_k = 1.57735\ldots + \varepsilon_k.$$

This completes the proof of the lemma. $\square$

Lemma 7.5. *Assume that* $\mathbf{x} = W_k M_{k-1} M_k \tilde{M}_k \dots$, *where* W_k *is a suffix of* M_k *and* $a_k \geq 3$. *If* k *is sufficiently large, then, for some integer* n *with* $\frac{q_k}{2} - 2 \leq n \leq q_k + q_{k-1} - 2$, *we have*

$$\frac{r(n, \mathbf{x})}{n} < \frac{\sqrt{17}+9}{8} + 2\varepsilon_k = 1.640\ldots + 2\varepsilon_k.$$

Proof. By the assumption $a_k \geq 3$, we get $\eta_k = \frac{q_{k-1}}{q_k} < \frac{1}{3}$.

Since $\mathbf{x} = W_k M_{k-1} M_k \tilde{M}_k \ldots = W_k M_{k-1} \tilde{M}_k \ldots$, it follows from Lemma 7.3 that $r(|M_{k-1}\tilde{M}_k| - q_{k-1}, \mathbf{x}) \leq |W_k M_{k-1} \tilde{M}_k|$, which yields that

$$\begin{aligned} \frac{r(q_k + q_{k-1} - 2, \mathbf{x})}{q_k + q_{k-1} - 2} &\leq \frac{|W_k| + q_k + 2q_{k-1} - 2}{q_k + q_{k-1} - 2} \\ &= 1 + \frac{t_k + \eta_k}{1 + \eta_k - \varepsilon_k} < 1 + \frac{t_k + \eta_k}{1 + \eta_k} + \varepsilon_k. \end{aligned} \tag{7.5}$$

We also have that $\mathbf{x} = W_k M_{k-1} \tilde{M}_k \ldots = W_k M_{k-1}^{--} \ldots$. Assume that $|W_k| \geq \frac{q_k}{2} \geq 3$. Since $W_k M_{k-1}^{--}$ is a suffix of $M_k M_{k-1}^{--} = \tilde{M}_k$, by Lemma 7.3, $r(|W_k M_{k-1}^{--}| - q_{k-1}, \mathbf{x}) \leq |W_k M_{k-1}^{--}|$, thus

$$\begin{aligned} \frac{r(|W_k| - 2, \mathbf{x})}{|W_k| - 2} &\leq \frac{|W_k| + q_{k-1} - 2}{|W_k| - 2} = 1 + \frac{\eta_k}{t_k - \varepsilon_k} = 1 + \frac{\eta_k}{t_k} + \frac{\eta_k \varepsilon_k}{t_k(t_k - \varepsilon_k)} \\ &< 1 + \frac{\eta_k}{t_k} + \frac{4\varepsilon_k}{3(1 - 2\varepsilon_k)} < 1 + \frac{\eta_k}{t_k} + 2\varepsilon_k. \end{aligned} \tag{7.6}$$

By (7.5) and (7.6), we get

$$\min_{\frac{q_k}{2} - 2 \leq n \leq q_k + q_{k-1} - 2} \frac{r(n, \mathbf{x})}{n} < \begin{cases} 1 + \min\left\{ \frac{t_k + \eta_k}{1 + \eta_k}, \frac{\eta_k}{t_k} \right\} + 2\varepsilon_k, & \text{if } |W_k| \geq \frac{q_k}{2}. \\ 1 + \frac{1/2 + \eta_k}{1 + \eta_k} + \varepsilon_k, & \text{if } |W_k| < \frac{q_k}{2}. \end{cases}$$

Since $\min\left\{ \frac{t_k + \eta_k}{1 + \eta_k}, \frac{\eta_k}{t_k} \right\} \leq \frac{\eta_k + \sqrt{5\eta_k^2 + 4\eta_k}}{2(1 + \eta_k)}$, we get

$$\min_{\frac{q_k}{2} - 2 \leq n \leq q_k + q_{k-1} - 2} \frac{r(n, \mathbf{x})}{n} < 1 + \frac{\max\left\{ \eta_k + \sqrt{5\eta_k^2 + 4\eta_k}, 1 + 2\eta_k \right\}}{2(1 + \eta_k)} + 2\varepsilon_k.$$

Thus, using $\eta_k < \frac{1}{3}$, for some integer n with $\frac{q_k}{2} - 2 \leq n \leq q_k + q_{k-1} - 2$ we have

$$\frac{r(n, \mathbf{x})}{n} < 1 + \frac{\frac{1}{3} + \sqrt{\frac{5}{9} + \frac{4}{3}}}{2(1 + \frac{1}{3})} + 2\varepsilon_k = \frac{\sqrt{17}+9}{8} + 2\varepsilon_k. \qquad \square$$

Completion of the proof of Theorem 3.4.

Suppose that $\liminf_{n\to+\infty} \frac{r(n,\mathbf{x})}{n} > 1.65$. By Lemmas 7.2, 7.4 and 7.5, for all large k we have $a_k \in \{1,2\}$ and

$$\mathbf{x} = W_k M_{k-1} M_k \tilde{M}_k \ldots,$$

where W_k is a suffix of M_k. Thus, for all large k we have $W_{k+1} = W_k M_{k-1}$ from Lemma 7.2.

We gather two auxiliary statements in a lemma.

Lemma 7.6. *Assume that* $\mathbf{x} = W_k M_{k-1} M_k \tilde{M}_k \ldots$, *where* W_k *is suffix of* M_k. *If* k *is sufficiently large, then we have*

(i) $$\frac{r(|W_k| + q_k + q_{k-1} - 2, \mathbf{x})}{|W_k| + q_k + q_{k-1} - 2} < 1 + \frac{1+\eta_k}{t_k + 1 + \eta_k} + \varepsilon_k,$$

(ii) $$\frac{r(q_k + q_{k-1} - 2, \mathbf{x})}{q_k + q_{k-1} - 2} < 1 + \frac{t_k + \eta_k}{1+\eta_k} + \varepsilon_k.$$

Proof. Since $\mathbf{x}_1^{|W_k|+q_k+q_{k-1}-2} = W_k \tilde{M}_k = \mathbf{x}_{q_k+q_{k-1}+1}^{|W_k|+2q_k+2q_{k-1}-2}$, we get

$$r(|W_k| + q_k + q_{k-1} - 2, \mathbf{x}) \le |W_k| + 2q_k + 2q_{k-1} - 2.$$

Also by Lemma 7.3, from the fact $\mathbf{x} = W_k M_{k-1} M_k \tilde{M}_k \ldots = W_k M_{k-1} \tilde{M}_k \ldots$ we get

$$r(q_k + q_{k-1} - 2, \mathbf{x}) \le |W_k| + q_k + 2q_{k-1} - 2.$$ □

• If $a_k = 1$ for all large k then η_k tends to $\frac{1}{\varphi}$ as k tends to infinity and we deduce from

$$1 - t_{k+1} = 1 - \frac{t_k + \eta_k}{1+\eta_k} = \frac{1-t_k}{1+\eta_k}$$

that $\lim_{k\to+\infty} t_k = 1$. By Lemma 7.6 (i), we then get

$$\frac{r(|W_k| + q_k + q_{k-1} - 2, \mathbf{x})}{|W_k| + q_k + q_{k-1} - 2} < 1 + \frac{1+\eta_k}{t_k + 1 + \eta_k} + \varepsilon_k,$$

where the right hand side tends to $1 + \frac{1+1/\varphi}{2+1/\varphi} = \varphi$ as k tends to infinity.

Consequently, there are arbitrarily large integers k such that $a_k = 2$.

• If $a_{k+1} = 2$, $a_{k+2} = 2$, then $q_{k+2} = 5q_k + 2q_{k-1}$, $q_{k+1} = 2q_k + q_{k-1}$, thus

$$\eta_{k+2} = \frac{2q_k + q_{k-1}}{5q_k + 2q_{k-1}} = \frac{2+\eta_k}{5+2\eta_k},$$

$$t_{k+2} = \frac{|W_k M_{k-1} M_k|}{q_{k+2}} = \frac{|W_k| + q_k + q_{k-1}}{5q_k + 2q_{k-1}} = \frac{t_k + 1 + \eta_k}{5 + 2\eta_k}.$$

By Lemma 7.6 (ii), we get

$$\frac{r(q_{k+2} + q_{k+1} - 2, \mathbf{x})}{q_{k+2} + q_{k+1} - 2} < 1 + \frac{t_{k+2} + \eta_{k+2}}{1 + \eta_{k+2}} + \varepsilon_k = 1 + \frac{t_k + 3 + 2\eta_k}{7 + 3\eta_k} + \varepsilon_k < \varphi.$$

• If $a_k = 1$, $a_{k+1} = 2$, $a_{k+2} = 1$, $a_{k+3} = 2$, then we have

$$q_{k+3} = 11q_{k-1} + 8q_{k-2}, \quad q_{k+2} = 4q_{k-1} + 3q_{k-2},$$

thus

$$\eta_{k+3} = \frac{4q_{k-1} + 3q_{k-2}}{11q_{k-1} + 8q_{k-2}} = \frac{4 + 3\eta_{k-1}}{11 + 8\eta_{k-1}},$$

$$t_{k+3} = \frac{|W_{k-1}| + q_{k-2} + q_{k-1} + q_k + q_{k+1}}{q_{k+3}} = \frac{t_{k-1} + 5 + 4\eta_{k-1}}{11 + 8\eta_{k-1}}.$$

By Lemma 7.6 (i) we may assume that

$$\frac{1 + \eta_{k-1}}{t_{k-1} + 1 + \eta_{k-1}} > \varphi - 1, \quad \text{that is,} \qquad t_{k-1} < (\varphi - 1)(1 + \eta_{k-1}).$$

Using Lemma 7.6 (ii), we get

$$\frac{r(q_{k+3} + q_{k+2} - 2, \mathbf{x})}{q_{k+3} + q_{k+2} - 2} < 1 + \frac{t_{k+3} + \eta_{k+3}}{1 + \eta_{k+3}} + \varepsilon_k = 1 + \frac{t_{k-1} + 9 + 7\eta_{k-1}}{15 + 11\eta_{k-1}} + \varepsilon_k.$$

For $\eta_{k-1} \le \varphi - 1$, we obtain

$$\begin{aligned}\frac{t_{k-1} + 9 + 7\eta_{k-1}}{15 + 11\eta_{k-1}} &\le \frac{\varphi + 8 + (\varphi + 6)\eta_{k-1}}{15 + 11\eta_{k-1}} \le \frac{\varphi + 8 + (\varphi + 6)(\varphi - 1)}{15 + 11(\varphi - 1)} \\ &= \frac{7\varphi + 3}{11\varphi + 4} = \frac{5\sqrt{5} + 69}{122} = 0.6572\ldots\end{aligned}$$

For $\eta_{k-1} > \varphi - 1$, we get

$$\frac{t_{k-1} + 9 + 7\eta_{k-1}}{15 + 11\eta_{k-1}} \le \frac{10 + 7\eta_{k-1}}{15 + 11\eta_{k-1}} < \frac{7\varphi + 3}{11\varphi + 4} = \frac{5\sqrt{5} + 69}{122} = 0.6572\ldots < \sqrt{10} - \frac{5}{2}.$$

• If $a_{k+1} = 2$, $a_{k+2} = 1$, $a_{k+3} = 1$, $a_{k+4} = 1$, then by Lemma 7.6 (ii) we may assume that

$$\frac{t_{k+1} + \eta_{k+1}}{1 + \eta_{k+1}} = \frac{|W_{k+1}| + q_k}{q_{k+1} + q_k} = \frac{|W_k| + q_k + q_{k-1}}{3q_k + q_{k-1}} = \frac{t_k + 1 + \eta_k}{3 + \eta_k} > \varphi - 1. \tag{7.7}$$

We have $q_{k+4} = 8q_k + 3q_{k-1}$, $q_{k+3} = 5q_k + 2q_{k-1}$, thus

$$\eta_{k+4} = \frac{5q_k + 2q_{k-1}}{8q_k + 3q_{k-1}} = \frac{5 + 2\eta_k}{8 + 3\eta_k},$$

$$t_{k+4} = \frac{|W_k| + q_{k-1} + q_k + q_{k+1} + q_{k+2}}{q_{k+4}} = \frac{t_k + 6 + 3\eta_k}{8 + 3\eta_k}.$$

By Lemma 7.6 (i), we get

$$\begin{aligned}\frac{r(|W_{k+4}|+q_{k+4}+q_{k+3}-2,\mathbf{x})}{|W_{k+4}|+q_{k+4}+q_{k+3}-2} &< 1+\frac{1+\eta_{k+4}}{t_{k+4}+1+\eta_{k+4}}+\varepsilon_k \\ &= 1+\frac{13+5\eta_k}{t_k+19+8\eta_k}+\varepsilon_k \\ &< 1+\frac{13+5\eta_k}{(\varphi-1)(3+\eta_k)+18+7\eta_k}+\varepsilon_k \\ &< 1+\frac{18}{21+4\varphi}+\varepsilon_k \\ &= 1.6552\ldots+\varepsilon_k < \sqrt{10}-\frac{3}{2},\end{aligned}$$

where we used the inequality (7.7).

Suppose that $\liminf_{n\to+\infty}\frac{r(n,\mathbf{x})}{n}\geq\sqrt{10}-\frac{3}{2}$. We have established that there exists an integer K such that the slope of $\mathbf{x}$ is equal to $[0;a_1,a_2,\ldots,a_K,\overline{2,1,1}]$ and for all $k\geq K$

$$\mathbf{x}=W_{k+1}M_kM_{k+1}\tilde{M}_{k+1}\cdots=W_kM_{k-1}M_kM_k\tilde{M}_k\ldots.$$

We establish now that, under these assumptions, we have

$$\liminf_{n\to+\infty}\frac{r(n,\mathbf{x})}{n}=\sqrt{10}-\frac{3}{2}.$$

Let k be an integer with $k>K$. By Lemma 7.6 (i),

$$\frac{r(|W_{3k+K}|+q_{3k+K}+q_{3k-1+K}-2,\mathbf{x})}{|W_{3k+K}|+q_{3k+K}+q_{3k-1+K}-2}<1+\frac{1+\eta_{3k+K}}{t_{3k+K}+1+\eta_{3k+K}}+\varepsilon_{3k+K}.$$

Since

$$\eta_k=\frac{q_{k-1}}{q_k}=[0;a_k,a_{k-1},\ldots,a_1]$$

and

$$\begin{aligned}t_k&=\frac{q_{k-2}+q_{k-3}+\ldots+q_{K-1}+|W_K|}{q_k}\\ &=\eta_k\eta_{k-1}+\eta_k\eta_{k-1}\eta_{k-2}+\ldots+\eta_k\eta_{k-1}\cdots\eta_K+\frac{|W_K|}{q_k},\end{aligned}$$

we check that

$$\lim_{k\to+\infty}\eta_{3k+K}=\frac{\sqrt{10}}{2}-1\quad\text{and}\quad\lim_{k\to+\infty}t_{3k+K}=\frac{8-\sqrt{10}}{6},\tag{7.8}$$

giving that

$$\liminf_{n\to+\infty}\frac{r(n,\mathbf{x})}{n}\leq\sqrt{10}-\frac{3}{2}.$$

Let us now show that this inequality is indeed an equality.

Since M_kM_{k-1} is primitive, Lemma 7.1 implies that all of the first $(q_k + q_{k-1})$ factors of length $(q_k + q_{k-1} - 1)$ of the word $\mathbf{x} = W_k M_{k-1} M_k M_k \tilde{M}_k \dots$ are distinct, thus we have

$$r(q_k + q_{k-1} - 1, \mathbf{x}) \ge 2q_k + 2q_{k-1} - 1.$$

The next $|W_k|$ factors of $\mathbf{x}$ of length $(q_k + q_{k-1} - 1)$ are identical with its first $|W_k|$ factors since, for $1 \le i \le |W_k|$, we have

$$x_i^{i+q_k+q_{k-1}-2} = x_{i+q_k+q_{k-1}}^{i+2q_k+2q_{k-1}-2} = (W_k)_i^{|W_k|} (\tilde{M}_k)_1^{i+q_k+q_{k-1}-|W_k|-2}.$$

By the fact that the last two letters of M_kM_{k-1} and $M_{k-1}M_k$ are different, we get

$$x_{|W_k|+q_k+q_{k-1}-1} \ne x_{|W_k|+2q_k+2q_{k-1}-1}.$$

It follows that, for $1 \le i \le |W_k|$, we have

$$x_i^{i+|W_k|+q_k+q_{k-1}-2} \ne x_{i+q_k+q_{k-1}}^{i+|W_k|+2q_k+2q_{k-1}-2}.$$

Therefore, we get

$$r(|W_k| + q_k + q_{k-1} - 1, \mathbf{x}) \ge 2|W_k| + 2q_k + 2q_{k-1} - 1.$$

It then follows from Lemma 2.1 (iii) that

$$r(n, \mathbf{x}) \ge \begin{cases} n + q_k + q_{k-1}, & q_k + q_{k-1} - 1 \le n \le |W_k| + q_k + q_{k-1} - 2, \\ n + |W_k| + q_k + q_{k-1}, & |W_k| + q_k + q_{k-1} - 1 \le n \le q_{k+1} + q_k - 2. \end{cases}$$

We also check that

$$\lim_{k\to+\infty} \eta_{3k+1+K} = \frac{\sqrt{10}-2}{3}, \qquad \lim_{k\to+\infty} \eta_{3k+2+K} = \frac{\sqrt{10}-1}{3}.$$

$$\lim_{k\to+\infty} t_{3k+1+K} = \frac{8-\sqrt{10}}{9}, \qquad \lim_{k\to+\infty} t_{3k+2+K} = \frac{2}{3}.$$

Combined with (7.8) we get

$$\liminf_{k\to+\infty} \frac{|W_k| + 2q_k + 2q_{k-1} - 2}{|W_k| + q_k + q_{k-1} - 2} = 1 + \liminf_{k\to+\infty} \frac{1+\eta_k}{t_k + 1 + \eta_k} = \sqrt{10} - \frac{3}{2}$$

and

$$\liminf_{k\to+\infty} \frac{|W_k| + q_{k+1} + 2q_k + q_{k-1} - 2}{q_{k+1} + q_k - 2} = 1 + \liminf_{k\to+\infty} \frac{t_{k+1} + \eta_{k+1}}{1 + \eta_{k+1}} = \frac{5}{3}.$$

Therefore, we conclude that

$$\mathrm{rep}(\mathbf{x}) = \liminf_{n\to+\infty} \frac{r(n, \mathbf{x})}{n} = \sqrt{10} - \frac{3}{2}.$$

This completes the proof of Theorem 3.4.

We remark that, in the course of the proof of Theorem 3.4, we have established that if $\mathrm{rep}(\mathbf{x}) < \sqrt{10} - \frac{3}{2}$ for a Sturmian word $\mathbf{x}$, then $\mathrm{rep}(\mathbf{x}) \le \frac{5\sqrt{5}+191}{122} = 1.6572\ldots$. Consequently, $\sqrt{10} - \frac{3}{2}$ is an isolated point of the set of real numbers $\mathrm{rep}(\mathbf{s})$, where $\mathbf{s}$ runs over the Sturmian words.

Examples of Sturmian words $\mathbf{x}$ *such that* $\mathrm{rep}(\mathbf{x}) = \sqrt{10} - \frac{3}{2}$.

In the proof of Theorem 3.4 we have established that a Sturmian word $\mathbf{s}'$ satisfies

$$\mathrm{rep}(\mathbf{s}') = \sqrt{10} - \frac{3}{2}$$

if and only if the continued fraction expansion of the slope of $\mathbf{s}'$ is eventually periodic and of the form $[0; a_1, a_2, \ldots, a_K, \overline{2,1,1}]$ for some integer K and $\mathbf{s}' = W_k M_{k-1} M_k \tilde{M}_k \ldots$ for all sufficiently large k.

Set $\theta = [0; a_1, a_2, \ldots] = [0; \overline{2,1,1}] = \frac{\sqrt{10}-2}{3}$. With M_k defined as before, for $k \ge 2$, the word $W_k = 1 M_0 M_1 \ldots M_{k-2}$ is a suffix of M_k. Define

$$\mathbf{s}' = \lim_{k\to+\infty} W_k = \lim_{k\to+\infty} \bigl(1 M_0 M_1 \ldots M_{k-2}\bigr) = 100101001001\ldots.$$

By applying Theorem 1 and Proposition 1 of [11] with $e_n = 1$ for $n \ge 1$, we see that the intercept of $\mathbf{s}'$ is equal to

$$(1-\theta)\left(1 + \sum_{n=0}^{\infty} (-1)^{n+1} \theta_1 \cdots \theta_{n+1} e_{n+1}\right) = 1 - \theta - \sum_{k=0}^{\infty} (q_k \theta - p_k) = \frac{1}{3},$$

where $\theta_1 = [0; a_1 - 1, a_2, \ldots]$ and $\theta_k = [0; a_k, a_{k+1}, \ldots]$.

The example of Cassaigne [24] for the minimal value of $\limsup_{n\to+\infty} \frac{R'(n)}{n}$ is given by the fixed point of the substitution σ defined by

$$\sigma(0) = 01001010, \qquad \sigma(1) = 010.$$

Set

$$\mathbf{c} := \lim_{k\to+\infty} \sigma^k(0) = 0100101001001001010010010100100100101001001\ldots$$

The word $\mathbf{c}$ is a Sturmian word of slope $[0; 2, \overline{1,2,1}]$. Let $(M_k^{\mathbf{c}})_{k\ge 0}$ be the corresponding sequence of words given by (7.1). Then it is easy to check by induction that $010\sigma(M_k^{\mathbf{c}}) = M_{k+3}^{\mathbf{c}}010$ for $k \geq 0$. Therefore, we have

$$\begin{aligned}\sigma(01M_0^{\mathbf{c}}M_1^{\mathbf{c}}M_2^{\mathbf{c}}\dots) &= 01001010\ 010\ \sigma(M_0^{\mathbf{c}})\ \sigma(M_1^{\mathbf{c}})\ \sigma(M_2^{\mathbf{c}})\dots\\ &= 01M_0^{\mathbf{c}}M_1^{\mathbf{c}}M_2^{\mathbf{c}}\ 010\sigma(M_0^{\mathbf{c}})\ \sigma(M_1^{\mathbf{c}})\sigma(M_2^{\mathbf{c}})\dots\\ &= 01M_0^{\mathbf{c}}M_1^{\mathbf{c}}M_2^{\mathbf{c}}M_3^{\mathbf{c}}M_4^{\mathbf{c}}\dots,\end{aligned}$$

and it follows that $\mathbf{c} = 01M_0^{\mathbf{c}}M_1^{\mathbf{c}}M_2^{\mathbf{c}}M_3^{\mathbf{c}}M_4^{\mathbf{c}}\dots$, thus $\mathrm{rep}(\mathbf{c}) = \sqrt{10} - \frac{3}{2}$.

Let τ be the substitution given by $\tau(0) = 10$ and $\tau(1) = 0$. We check by induction that $0\tau(M_k^{\mathbf{c}}) = M_{k+1}0$ holds for all $k \geq 0$. We conclude that $\mathbf{c}$ and $\mathbf{s}'$ are related by

$$\begin{aligned}\tau(\mathbf{c}) &= \tau(01M_0^{\mathbf{c}}M_1^{\mathbf{c}}M_2^{\mathbf{c}}M_3^{\mathbf{c}}M_4^{\mathbf{c}}\dots)\\ &= 10\ 0\tau(M_0^{\mathbf{c}})\tau(M_1^{\mathbf{c}})\tau(M_2^{\mathbf{c}})\tau(M_3^{\mathbf{c}})\tau(M_4^{\mathbf{c}})\dots\\ &= 10M_1M_2M_3M_4M_5\dots = \mathbf{s}'.\end{aligned}$$

Proof of Theorem 3.6.

Let $\mathbf{y}$ be an infinite word defined over a finite alphabet $\mathcal{A}$ such that the sequence $(p(n,\mathbf{y}) - n)_{n\geq 1}$ is bounded and $\mathbf{y}$ is not ultimately periodic. It follows from Theorem 1.1 that the sequence $(p(n,\mathbf{y}) - n)_{n\geq 1}$ of positive integers is nondecreasing and bounded. Thus, it is eventually constant. There exist positive integers k and n_0 such that

$$p(n,\mathbf{y}) = n + k, \qquad \text{for } n \geq n_0.$$

It then follows from a result of Cassaigne [25] that there are a finite word W, a Sturmian word $\mathbf{s}$ defined over $\{0,1\}$ and a morphism ϕ from $\{0,1\}^*$ into $\mathcal{A}^*$ such that $\phi(01) \neq \phi(10)$ and

$$\mathbf{y} = W\phi(\mathbf{s}).$$

Write $\mathbf{s} = s_1s_2\dots$ Let n be a large positive integer. The word $V_n := s_{r(n,\mathbf{s})-n+1}^{r(n,\mathbf{s})}$ of length n has two occurrences in $s_1^{r(n,\mathbf{s})}$. Consequently, the word $\phi(V_n)$ has two occurrences in the prefix of $\mathbf{y}$ of length $|W| + |\phi(s_1^{r(n,\mathbf{s})})|$, thus

$$r(|\phi(V_n)|,\mathbf{y}) \leq |W| + |\phi(s_1^{r(n,\mathbf{s})})|.$$

A classical property of Sturmian words asserts that 0 and 1 have a frequency in $\mathbf{s}$. Consequently, by arguing as in [1], there exists a real number δ such that

$$|\phi(s_1 s_2 \ldots s_n)| = \delta n + o(n), \quad \text{for every } n \geq 1.$$

Let ε be a positive real number. For n large enough there exist real numbers η_n and μ_n with $|\eta_n|, |\mu_n| \leq \varepsilon n$ and

$$r(\delta n + \eta_n, \mathbf{y}) \leq |W| + \delta r(n, \mathbf{s}) + \mu_n.$$

As n can be taken arbitrarily large, this implies that

$$\operatorname{rep}(\mathbf{y}) = \liminf_{n \to +\infty} \frac{r(n, \mathbf{y})}{n} \leq \frac{\delta}{\delta - \varepsilon} \liminf_{n \to +\infty} \frac{r(n, \mathbf{s})}{n} + \frac{\varepsilon}{\delta - \varepsilon}.$$

Since ε can be chosen arbitrarily small, we deduce that

$$\operatorname{rep}(\mathbf{y}) \leq \liminf_{n \to +\infty} \frac{r(n, \mathbf{s})}{n} = \operatorname{rep}(\mathbf{s}).$$

In view of Theorem 3.4, this proves Theorem 3.6.

8. Rational approximation

In this section and in the next one, for a finite word W and a real number $w \geq 1$, we write W^w for the concatenation of $\lfloor w \rfloor$ copies of W and the prefix of length $\lceil (w - \lfloor w \rfloor)|W| \rceil$ of W.

Proof of Theorem 4.2.

Since the irrationality exponent of an irrational real number is at least equal to 2, we can assume that $\operatorname{rep}(\mathbf{x}) < 2$. Let n be a positive integer such that $r(n, \mathbf{x}) < 2n$. By the theorem of Lyndon and Schützenberger (Theorem 1.5.2 in [9]), this implies that there are finite words W, U, V (we do not indicate the dependence on n) and a positive integer t such that $|(UV)^t U| = n$ and $W(UV)^{t+1}U$ is the prefix of $\mathbf{x}$ of length $r(n, \mathbf{x})$. Observe that

$$|WUV| = |W(UV)^{t+1}U| - |(UV)^t U| = r(n, \mathbf{x}) - n.$$

Setting $\xi = \sum_{k \geq 1} \frac{x_k}{b^k}$, there exists an integer s such that ξ and the rational number $\frac{s}{b^{|W|}(b^{|UV|}-1)}$ have the same $r(n, \mathbf{x})$ first digits in their b-ary expansions, thus

$$\left| \xi - \frac{s}{b^{|W|}(b^{|UV|}-1)} \right| \leq \frac{1}{b^{|W(UV)^{t+1}U|}} = \frac{1}{b^{|WUV|+|(UV)^t U|}}$$
$$= \frac{1}{b^{|WUV|} b^{n|WUV|/(r(n,\mathbf{x})-n)}}.$$

We derive that

$$\mu(\xi) \geq 1 + \limsup_{n \to +\infty} \frac{n}{r(n, \mathbf{x}) - n},$$

thus, $\mu(\xi)$ is infinite if $\mathrm{rep}(\mathbf{x}) = 1$ and

$$\mu(\xi) \geq 1 + \frac{1}{\mathrm{rep}(\mathbf{x}) - 1},$$

otherwise. This proves the theorem.

Proof of Theorem 4.5.

We assume that the reader is familiar with the theory of continued fractions (see e.g. Section 1.2 of [15]).

Set $\xi := \sum_{k \geq 1} \frac{x_k}{b^k}$. Write $\xi = [0; d_1, d_2, \ldots]$ and let $(\frac{p_j}{q_j})_{j \geq 1}$ denote the sequence of its convergents.

Let $\mathcal{N} := (n_k)_{k \geq 1}$ be the increasing sequence of all the integers n such that $r(n+1, \mathbf{x}) \geq r(n, \mathbf{x}) + 2$. Let k be a positive integer. By Lemma 5.3 we have $r(n_k + 1, \mathbf{x}) = 2n_k + 3$.

We deduce from the definition of the sequence $\mathcal{N}$ that

$$r(n_k + \ell, \mathbf{x}) = 2n_k + 2 + \ell, \quad 1 \leq \ell \leq n_{k+1} - n_k. \tag{8.1}$$

Set $\alpha_k = \frac{r(n_k, \mathbf{x})}{n_k}$. Observe that $\alpha_k \leq 2 + \frac{1}{n_k}$ and

$$\mathrm{rep}(\mathbf{x}) = \liminf_{k \to +\infty} \alpha_k. \tag{8.2}$$

Let k be an integer for which $\alpha_k < 2$ (infinitely many such k do exist since $\mathrm{rep}(\mathbf{x}) < 2$). Let W_k, U_k, V_k be the words associated with n_k as in the previous proof and w_k, u_k, v_k their lengths, which satisfy $w_k + u_k + v_k = (\alpha_k - 1)n_k$. There exists an integer s_k such that the $\alpha_k n_k$ first digits of $\mathbf{x}$ and those of the b-ary expansion of the rational number $\frac{s_k}{b^{w_k}(b^{u_k+v_k}-1)}$ coincide. Consequently, we get

$$\left| \xi - \frac{s_k}{b^{w_k}(b^{u_k+v_k} - 1)} \right| \leq \frac{1}{b^{\alpha_k n_k}}. \tag{8.3}$$

A classical theorem of Legendre (see e.g. Theorem 1.8 of [15]) asserts that, if the irrational real number ζ and the rational number $\frac{p}{q}$ with $q \geq 1$ satisfy $|\zeta - \frac{p}{q}| < \frac{1}{2q^2}$, then $\frac{p}{q}$ is a convergent of the continued fraction expansion of ζ.

Since $\alpha_k < 2$, we get $\alpha_k \leq 2 - \frac{1}{n_k}$. As

$$2\big(b^{w_k}(b^{u_k+v_k} - 1)\big)^2 < 2b^{2(\alpha_k-1)n_k} \leq b^{\alpha_k n_k}$$

holds if $\alpha_k n_k \leq 2n_k - 1$, Legendre's theorem and the assumption $\alpha_k < 2$ imply that the rational number $\frac{s_k}{b^{w_k}(b^{u_k+v_k}-1)}$, which may not be written under its reduced form, is a convergent, say $\frac{p_h}{q_h}$, of the continued fraction expansion of ξ.

Let ℓ be the smallest positive integer such that $\alpha_{k+\ell} < 2$.

We first establish that $\ell \leq 2$ if n_k is sufficiently large.

Assume that $r(n_{k+1}, \mathbf{x}) = 2n_{k+1} + \varepsilon_{k+1}$ and $r(n_{k+2}, \mathbf{x}) = 2n_{k+2} + \varepsilon_{k+2}$, with $\varepsilon_{k+1}, \varepsilon_{k+2} \in \{0, 1\}$. Put $\eta_k := r(n_{k+2}, \mathbf{x}) - r(n_{k+1}, \mathbf{x})$. Since

$$\begin{aligned} \alpha_{k+2} n_{k+2} &= r(n_{k+2}, \mathbf{x}) = r(n_{k+1} + (n_{k+2} - n_{k+1}), \mathbf{x}) \\ &= 2n_{k+1} + 2 + n_{k+2} - n_{k+1} = n_{k+2} + n_{k+1} + 2, \end{aligned} \tag{8.4}$$

we get $n_{k+2} = n_{k+1} + 2 - \varepsilon_{k+2}$, thus

$$\eta_k = 2(n_{k+1} + 2 - \varepsilon_{k+2}) + \varepsilon_{k+2} - 2n_{k+1} - \varepsilon_{k+1} = 4 - \varepsilon_{k+1} - \varepsilon_{k+2}.$$

This shows that $\eta_k \in \{2, 3, 4\}$.

By a well-known property of Sturmian sequences (see [35] on page 46), for any $n \geq 1$, there exists a unique factor Z_n (called a right special factor) of $\mathbf{x}$ of length n such that $Z_n 0$ and $Z_n 1$ are both factors of $\mathbf{x}$.

It follows from our assumption $r(n_{k+1} + 1, \mathbf{x}) > r(n_{k+1}, \mathbf{x}) + 1$ that $Z_{n_{k+1}} = x_{r(n_{k+1},\mathbf{x})-n_{k+1}+1}^{r(n_{k+1},\mathbf{x})}$. Likewise, we get $Z_{n_{k+2}} = x_{r(n_{k+2},\mathbf{x})-n_{k+2}+1}^{r(n_{k+2},\mathbf{x})}$, thus

$$Z_{n_{k+1}} = x_{r(n_{k+1},\mathbf{x})-n_{k+1}+1}^{r(n_{k+1},\mathbf{x})} = x_{r(n_{k+2},\mathbf{x})-n_{k+1}+1}^{r(n_{k+2},\mathbf{x})} = x_{r(n_{k+1},\mathbf{x})+\eta_k-n_{k+1}+1}^{r(n_{k+1},\mathbf{x})+\eta_k}.$$

It then follows from the theorem of Lyndon and Schützenberger (Theorem 1.5.2 in [9]) that there exists an integer t_k, a word T_k of length η_k and a prefix T_k' of T_k such that

$$Z_{n_{k+1}} = (T_k)^{t_k} T_k'.$$

We deduce that

$$t_k \geq \frac{n_{k+1} - 3}{4}.$$

Since $|T_k| \leq 4$ and a Sturmian word cannot contain unbounded powers of a fixed word (see [9, Corollary 10.6.6]), there exists an integer t such that no factor of $\mathbf{x}$ is a t-th power.

Consequently, if k is large enough, then we cannot have simultaneously $r(n_{k+1}, \mathbf{x}) \geq 2n_{k+1}$ and $r(n_{k+2}, \mathbf{x}) \geq 2n_{k+2}$. This implies that $\ell = 1$ or $\ell = 2$.

Since $\alpha_{k+\ell} < 2$, it follows from Legendre's theorem that the rational number $\frac{s_{k+\ell}}{b^{w_{k+\ell}}(b^{u_{k+\ell}+v_{k+\ell}}-1)}$, which may not be written under its reduced form, is a convergent, say $\frac{p_j}{q_j}$, of the continued fraction expansion of ξ. The $(\alpha_k n_k + 1)$-th digit of the b-ary expansion of $\frac{p_j}{q_j}$ is equal to the $(\alpha_k n_k + 1)$-th digit of $\mathbf{x}$ and differs from the $(\alpha_k n_k + 1)$-th digit of the b-ary expansion of $\frac{p_h}{q_h}$. Consequently, the rational numbers $\frac{p_h}{q_h}$ and $\frac{p_j}{q_j}$ are distinct.

Here, the indices h and j depend on k. We have

$$q_h \leq b^{w_k}(b^{u_k+v_k} - 1) \leq b^{(\alpha_k - 1)n_k} \tag{8.5}$$

and

$$q_j \leq b^{w_{k+\ell}}(b^{u_{k+\ell}+v_{k+\ell}} - 1) \leq b^{(\alpha_{k+\ell} - 1)n_{k+\ell}}.$$

Note that it follows from (8.4) that

$$(\alpha_{k+2} - 1)n_{k+2} = n_{k+1} + 2.$$

and, likewise,

$$(\alpha_{k+1} - 1)n_{k+1} = n_k + 2,$$

Note that $n_{k+1} \leq n_k + 2$ if $\alpha_{k+1} \geq 2$.

The properties of continued fractions give that

$$\frac{1}{2q_h q_{h+1}} \leq \left|\xi - \frac{p_h}{q_h}\right| \leq \frac{1}{q_h q_{h+1}} \tag{8.6}$$

and

$$\frac{1}{2q_j q_{j+1}} \leq \left|\xi - \frac{p_j}{q_j}\right| \leq \frac{1}{q_j q_{j+1}}.$$

This implies that

$$q_{j+1} \geq \frac{b^{\alpha_{k+\ell} n_{k+\ell}}}{2q_j} \geq \frac{b^{n_{k+\ell}}}{2}.$$

Since $\alpha_k < 2$, we get

$$q_h \leq b^{(\alpha_k - 1)n_k} \leq b^{n_k - 1} < \frac{b^{n_{k+\ell}}}{2} \leq q_{j+1}.$$

Combined with $p_h/q_h \neq p_j/q_j$, this gives

$$q_h < q_{h+1} \leq q_j < q_{j+1}.$$

It follows from

$$q_h \geq \frac{b^{\alpha_k n_k}}{2q_{h+1}}$$

and

$$q_{h+1} \leq q_j \leq b^{(\alpha_{k+\ell}-1)n_{k+\ell}} \leq b^{n_k+4}, \tag{8.7}$$

that

$$q_h \geq \frac{b^{\alpha_k n_k}}{2b^{n_k+4}}. \tag{8.8}$$

Since $q_h \leq b^{w_k}(b^{u_k+v_k}-1) \leq b^{(\alpha_k-1)n_k}$, this shows that the rational number $\frac{s_k}{b^{w_k}(b^{u_k+v_k}-1)}$ is not far from being reduced, in the sense that the greatest common divisor of its numerator and denominator is at most equal to $2b^4$. Furthermore, we deduce from (8.3), (8.5), (8.6), (8.7), and (8.8) that

$$\frac{1}{(2b^4 q_h)^{\alpha_k/(\alpha_k-1)}} \leq \Big|\xi - \frac{p_h}{q_h}\Big| \leq \frac{1}{q_h^{\alpha_k/(\alpha_k-1)}}. \tag{8.9}$$

Moreover, it follows from

$$q_{h+1} \geq \frac{b^{\alpha_k n_k}}{2q_h} \geq \frac{b^{n_k}}{2}$$

that

$$1 \leq \frac{q_j}{q_{h+1}} \leq 2b^4.$$

Consequently, all the partial quotients $d_{h+2}, \ldots, d_j$ are less than $2b^4$ and we get

$$\Big|\xi - \frac{p_\ell}{q_\ell}\Big| > \frac{1}{(q_\ell + q_{\ell+1})q_\ell} > \frac{1}{(d_{\ell+1}+2)q_\ell^2} \geq \frac{1}{2(b^4+1)q_\ell^2},$$

for $\ell = h+1, \ldots, j-1$.

Now, we are armed to conclude the proof. We consider the increasing sequence $\mathcal{K}$ of integers k such that $\alpha_k < 2$. Let k be an element of $\mathcal{K}$ and assume that k is sufficiently large. We have established that there exist integers $h(k)$ and $j(k)$ such that all the partial quotients $d_{h(k)+2}, \ldots, d_{j(k)}$ are less than $2b^4$. Furthermore, (8.9) provides us with a precise estimate of $d_{h(k)+1}$. The definitions of h and j show that if k' is the next element after k in the sequence $\mathcal{K}$, then $h(k') = j(k)$. Consequently, we have a precise estimate of all but finitely many partial quotients of ξ and we deduce from (8.2) and (8.9) that

$$\mu(\xi) = \limsup_{k\to+\infty} \frac{\alpha_k}{\alpha_k - 1} = \frac{\mathrm{rep}(\mathbf{x})}{\mathrm{rep}(\mathbf{x}) - 1}.$$

This completes the proof of the theorem.

9. On the recurrence function of an infinite word

Cassaigne [24] studied the recurrence function $n \mapsto R'(n)$ of an infinite word $\mathbf{x} = x_1 x_2 \dots$, which is defined as the length of the shorted prefix of $\mathbf{x}$ containing an occurrence of every factor of $\mathbf{x}$ of length n. Then it is not difficult to check that $R'(n) \ge p(n, \mathbf{x}) + n - 1$ and the equality holds if and only if $r(n, \mathbf{x}) = p(n, \mathbf{x}) + n$. Moreover, for a Sturmian word $\mathbf{x}$, we have the following relation between $r(n, \mathbf{x})$ and $R'(n)$.

Proposition 9.1. *For any Sturmian word* $\mathbf{x}$*, we have*

$$\limsup_{n\to+\infty} \frac{R'(n)}{n} = \frac{\mathrm{rep}(\mathbf{x})}{\mathrm{rep}(\mathbf{x}) - 1}.$$

Therefore, it follows from Theorem 3.4 that

$$\limsup_{n\to+\infty} \frac{R'(n)}{n} \ge \frac{5}{3} + \frac{4\sqrt{10}}{15} = 2.5099\dots,$$

and this value is optimal.

Proof. Let $\mathbf{x} = x_1 x_2 x_3 \dots$ be a Sturmian word. Let n be a positive integer such that $R'(n) \ge 2n + 1$. Since $p(n, \mathbf{x}) = n + 1$, there exist integers i, j such that

$$0 \le i < j \le R'(n) - n \quad \text{and} \quad x_{i+1}^{i+n} = x_{j+1}^{j+n}.$$

It follows from the definition of $R'(n)$ that $x_{R'(n)-n+1}^{R'(n)}$ is not a factor of $x_1^{R'(n)-1}$. Thus, there exists $m \ge 0$ such that

$$x_{i+1}^{i+n+m} = x_{j+1}^{j+n+m}, \quad x_{i+n+m+1} \ne x_{j+n+m+1}, \quad \text{and} \quad j + n + m + 1 \le R'(n).$$

Therefore, $x_{i+m+1}^{i+n+m+1}$ and $x_{j+m+1}^{j+n+m+1}$ are the two factors of $\mathbf{x}$ of length $n+1$ extending the right special factor x_{i+m+1}^{i+m+n}, and $x_1^{R'(n)+1}$ contains all the factors of $\mathbf{x}$ of length $n + 1$. This shows that $R'(n + 1) = R'(n) + 1$ whenever $R'(n) \ge 2n + 1$.

Let $(n_k)_{k\geq 1}$ be the increasing sequence of all the integers n such that $r(n+1, \mathbf{x}) \geq r(n, \mathbf{x}) + 2$. It then follows from (8.1) that

$$\mathrm{rep}(\mathbf{x}) = \liminf_{k\to+\infty} \frac{r(n_k, \mathbf{x})}{n_k} = \liminf_{k\to+\infty} \frac{n_k + n_{k-1} + 2}{n_k} = 1 + \liminf_{k\to+\infty} \frac{n_{k-1}}{n_k}.$$

For every positive integer n, we have $R'(n) = 2n$ if, and only if, $r(n, \mathbf{x}) = 2n + 1$. This shows that $R'(n_k + 1) = 2(n_k + 1)$ holds for every positive integer k. Furthermore, we have established above that $R'(n + 1) = R'(n) + 1$ if n is not an element of the sequence $(n_k + 1)_{k\geq 1}$. Consequently, we have

$$\begin{aligned}\limsup_{n\to+\infty} \frac{R'(n)}{n} &= \limsup_{k\to+\infty} \frac{R'(n_k + 2)}{n_k + 2} = \limsup_{k\to+\infty} \frac{n_{k+1} + n_k + 3}{n_k + 2} \\ &= 1 + \limsup_{k\to+\infty} \frac{n_{k+1}}{n_k} = 1 + \frac{1}{\mathrm{rep}(\mathbf{x}) - 1}.\end{aligned}$$

This proves the proposition. □

10. Links with other combinatorial exponents

There are various combinatorial exponents associated with infinite words. One of them, the initial critical exponent, was introduced in 2006 by Berthé, Holton, and Zamboni [14].

Definition 10.1. The initial critical exponent of an infinite word $\mathbf{x}$, denoted by $\mathrm{ice}(\mathbf{x})$, is the supremum of the real numbers ρ for which there exist arbitrary long prefixes V of $\mathbf{x}$ such that V^ρ is a prefix of $\mathbf{x}$.

The definition of the Diophantine exponent of an infinite word appeared in [2], but this notion was implicitly used in earlier works of the same authors.

Definition 10.2. The Diophantine exponent of an infinite word $\mathbf{x}$, denoted by $\mathrm{dio}(\mathbf{x})$, is the supremum of the real numbers ρ for which there exist arbitrary long prefixes of $\mathbf{x}$ that can be factorized as UV^w, with U and V finite words and w a real number such that

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

It follows from Definitions 9.1 and 9.2 that every infinite word $\mathbf{x}$ satisfies

$$1 \leq \mathrm{ice}(\mathbf{x}) \leq \mathrm{dio}(\mathbf{x}) \leq +\infty. \tag{9.1}$$

Furthermore, there are words $\mathbf{x}$ such that $\mathrm{ice}(\mathbf{x}) < \mathrm{dio}(\mathbf{x})$.

The following lemma shows that the Diophantine exponent and the exponent of repetition are closely related.

Lemma 10.3. *Let* $\mathbf{x}$ *be an infinite word written over a finite alphabet. We have* $\mathrm{rep}(\mathbf{x}) = 1$ *(resp.* $= +\infty$*) if and only if* $\mathrm{dio}(\mathbf{x}) = +\infty$ *(resp.* $= 1$*). Furthermore, if* $1 < \mathrm{dio}(\mathbf{x}) < +\infty$*, then we have*

$$\mathrm{rep}(\mathbf{x}) = \frac{\mathrm{dio}(\mathbf{x})}{\mathrm{dio}(\mathbf{x}) - 1} \le \frac{\mathrm{ice}(\mathbf{x})}{\mathrm{ice}(\mathbf{x}) - 1}.$$

Proof. In view of (9.1), it only remains for us to prove the first equality. To see that $\mathrm{rep}(\mathbf{x}) \le \frac{\mathrm{dio}(\mathbf{x})}{\mathrm{dio}(\mathbf{x})-1}$ it suffices to note that if UV^w is a prefix of $\mathbf{x}$, where $w > 1$ is chosen such that $|UV^w| = |U| + w|V|$, then

$$\frac{r(|V|^{w-1}, \mathbf{x})}{|V|^{w-1}} \le \frac{|UV^w|}{|V|^{w-1}} \le \frac{|UV^w|/|UV|}{(|UV^w|/|UV|) - 1}.$$

Conversely, if $r(n, \mathbf{x}) = Cn$ for some rational number C and some integer n, then the prefix of $\mathbf{x}$ of length Cn can be written under the form UV^w, where $w > 1$ and $|UV| = (C-1)n$. This implies that $\mathrm{dio}(\mathbf{x}) \ge \frac{C}{C-1}$. Letting C tend to $\mathrm{rep}(\mathbf{x})$, we get $\mathrm{dio}(\mathbf{x})(\mathrm{rep}(\mathbf{x}) - 1) \ge \mathrm{rep}(\mathbf{x})$, that is, $\mathrm{rep}(\mathbf{x}) \ge \frac{\mathrm{dio}(\mathbf{x})}{\mathrm{dio}(\mathbf{x})-1}$. □

One motivation for considering the function $n \mapsto r(n, \mathbf{x})$ comes from Diophantine approximation. Indeed, the following transcendence criteria have been recently established in [5, 3, 21, 19], although they were not highlighted in these papers, in which the subword complexity function $n \mapsto p(n, \mathbf{x})$ occurs in place of $n \mapsto r(n, \mathbf{x})$.

Theorem 10.4. *Let* $\mathcal{A}$ *be a finite set of integers. Let* $\mathbf{x} = x_1 x_2 \ldots$ *be an infinite word over* $\mathcal{A}$*, which is not eventually periodic. If*

$$\liminf_{n \to +\infty} \frac{r(n, \mathbf{x})}{n} < +\infty,$$

or if there exists a real number η *with* $\eta < 1/11$ *and*

$$\limsup_{n \to +\infty} \frac{r(n, \mathbf{x})}{n (\log n)^{\eta}} < +\infty,$$

then, for every integer $b \ge 2$*, the real number* $\sum_{k \ge 1} \frac{x_k}{b^k}$ *is transcendental.*

Recall that a real number is algebraic of degree two if and only if its continued fraction expansion is eventually periodic.

Theorem 10.5. *Let $\mathcal{A}$ be a finite set of positive integers. Let $\mathbf{x} = x_1 x_2 \dots$ be an infinite word over $\mathcal{A}$. If $\mathbf{x}$ is not eventually periodic and*

$$\liminf_{n \to +\infty} \frac{r(n, \mathbf{x})}{n} < +\infty,$$

then the real number $[0; x_1, x_2, \dots]$ is transcendental.

The interested reader is referred to the survey [20], where the combinatorial assumption made on the infinite word $\mathbf{x}$ is precisely the following (the same assumption is made in [3, 21, 19]): we suppose that $\mathbf{x}$ is not eventually periodic and that there exist three sequences of finite words $(U_n)_{n \geq 1}$, $(V_n)_{n \geq 1}$ and $(W_n)_{n \geq 1}$ such that:

(i) For every $n \geq 1$, the word $W_n U_n V_n U_n$ is a prefix of the word $\mathbf{x}$;
(ii) The sequence $(|V_n|/|U_n|)_{n \geq 1}$ is bounded from above;
(iii) The sequence $(|W_n|/|U_n|)_{n \geq 1}$ is bounded from above;
(iv) The sequence $(|U_n|)_{n \geq 1}$ is increasing.

One sees that this assumption exactly means that $\mathrm{dio}(\mathbf{x})$ exceeds 1 and is, by Lemma 10.3, equivalent to the one made in the above transcendence criteria. Using Lemma 2.2, we deduce immediately that $r(n, \mathbf{x})$ can be replaced by $p(n, \mathbf{x})$ in Theorems 10.4 and 10.5. Consequently, Lemma 8.1 of [20] (which was also used in [3, 19]) is not needed to deduce Theorems 3.1 and 3.2 of [20] from the combinatorial transcendence criteria stated in Section 4 of that paper. This shows that considering the function $n \mapsto r(n, \mathbf{x})$ is indeed the right point of view.

We end this section with a theorem established in [4]. It is stated in that paper with the subword complexity function $n \mapsto p(n, \mathbf{x})$, but, in that paper as well, the proofs actually work if this function is replaced by $n \mapsto r(n, \mathbf{x})$. For the definition of Mahler's classification, the reader is directed to Chapter 3 of [15].

Theorem 10.6. *Let ξ be a real number such that its expansion $\mathbf{x}$ in some integer base $b \geq 2$ satisfies*

$$\limsup_{n \to +\infty} \frac{r(n, \mathbf{x})}{n} < +\infty.$$

If $\mathrm{rep}(\mathbf{x}) = 1$, then ξ is a Liouville number. Otherwise, ξ is either an S-number or a T-number in Mahler's classification.

Acknowledgement

Dong Han Kim was supported by the National Research Foundation of Korea (NRF-2015R1A2A2A01007090). The authors wish to warmly thank the anonymous referee for very careful reading and many valuable comments.

References

[1] B. Adamczewski, *On the expansion of some exponential periods in an integer base*, Math. Ann. 346 (2010), 107–116.

[2] B. Adamczewski and Y. Bugeaud, *Dynamics for β-shifts and Diophantine approximation*, Ergod. Th. & Dynam. Sys. 27 (2007), 1695–1710.

[3] B. Adamczewski and Y. Bugeaud, *On the complexity of algebraic numbers I. Expansions in integer bases*, Ann. of Math. 165 (2007), 547–565.

[4] B. Adamczewski and Y. Bugeaud, *Nombres réels de complexité sous-linéaire : mesures d'irrationalité et de transcendance*, J. Reine Angew. Math. 658 (2011), 65–98.

[5] B. Adamczewski, Y. Bugeaud et F. Luca, *Sur la complexité des nombres algébriques*, C. R. Acad. Sci. Paris 339 (2004), 11–14.

[6] W. W. Adams and J. L. Davison, *A remarkable class of continued fractions*, Proc. Amer. Math. Soc. 65 (1977), 194–198.

[7] K. Alladi and M. L. Robinson, *Legendre polynomials and irrationality*, J. Reine Angew. Math. 318 (1980), 137–155.

[8] J.-P. Allouche, *Nouveaux résultats de transcendance de réels à développements non aléatoire*, Gaz. Math. 84 (2000), 19–34.

[9] J.-P. Allouche and J. Shallit, Automatic Sequences: Theory, Applications, Generalizations, Cambridge University Press, 2003.

[10] M. Amou and Y. Bugeaud, *Expansions in integer bases and exponents of Diophantine approximation*, J. London Math. Soc. 81 (2010), 297–316.

[11] P. Arnoux, S. Ferenczi and P. Hubert, *Trajectories of rotations*, Acta Arith. 87 (1999), 209–217.

[12] A. Baker, *Approximations to the logarithms of certain rational numbers*, Acta Arith. 10 (1964), 315–323.

[13] V. Becher and P. A. Heiber, *On extending de Bruijn sequences*, Inform. Process Lett. 111 (2011), 930–932.

[14] V. Berthé, C. Holton, and L. Q. Zamboni, *Initial powers of Sturmian sequences*, Acta Arith. 122 (2006), 315–347.

[15] Y. Bugeaud, Approximation by algebraic numbers, Cambridge Tracts in Mathematics 160, Cambridge, 2004.

[16] Y. Bugeaud, *Diophantine approximation and Cantor sets*, Math. Ann. 341 (2008), 677–684.

[17] Y. Bugeaud, Distribution modulo one and Diophantine approximation. Cambridge Tracts in Mathematics 193, Cambridge, 2012.

[18] Y. Bugeaud, *On the expansions of a real number to several integer bases*, Rev. Mat. Iberoam. 28 (2012), 931–946.

[19] Y. Bugeaud, *Automatic continued fractions are transcendental or quadratic*, Ann. Sci. École Norm. Sup. 46 (2013), 1005–1022.

[20] Y. Bugeaud, *Expansions of algebraic numbers.* In: Four Faces of Number Theory, EMS Series of Lectures in Mathematics, 2015.

[21] Y. Bugeaud and J.-H. Evertse, *On two notions of complexity of algebraic numbers*, Acta Arith. 133 (2008), 221–250.

[22] Y. Bugeaud and D.H. Kim, *On the b-ary expansions of* $\log(1+\frac{1}{a})$ *and* e. Ann. Sc. Norm. Super. Pisa Cl. Sci., to appear.

[23] Y. Bugeaud, D. Krieger, and J. Shallit, *Morphic and Automatic Words: Maximal Blocks and Diophantine Approximation*, Acta Arith. 149 (2011), 181–199.

[24] J. Cassaigne, *On a conjecture of J. Shallit*, In: ICALP 1997, Lecture Notes in Computer Science 1256, Springer, Berlin, 1997, pp. 693–704.

[25] J. Cassaigne, *Sequences with grouped factors.* In: DLT'97, Developments in Language Theory III, Thessaloniki, Aristotle University of Thessaloniki, 1998, pp. 211–222.

[26] L. J. Cummings and D. Wiedemann, *Embedded de Bruijn sequences*, Proc. 7th Southeastern international conference on combinatorics, graph theory, and computing, Congr. Numer. 53 (1986), 155–160.

[27] L. V. Danilov, *Rational approximations of some functions at rational points*, Mat. Zametki 24 (1978), 449–458, 589. English translation: Math. Notes 24 (1978), 741–746.

[28] S. Ferenczi and C. Mauduit, *Transcendence of numbers with a low complexity expansion*, J. Number Theory 67 (1997), 146–161.

[29] N. J. Fine and H. S. Wilf, *Uniqueness theorems for periodic functions*, Proc. Amer. Math. Soc. 16 (1965), 109–114.

[30] N. P. Fogg, Substitutions in dynamics, arithmetics and combinatorics. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel. Lecture Notes in Mathematics 1794. Springer-Verlag, Berlin, 2002.

[31] A. Iványi, *On the d-complexity of words*, Ann. Univ. Sci. Budapest Sect. Comput. 8 (1987), 69–90.

[32] T. Kamae and D.H. Kim, *A characterization of eventual periodicity*, Theoret. Comput. Sci., 581 (2015), 1–8.

[33] D.H. Kim, *Return time complexity of Sturmian sequences*, Theoret. Comput. Sci. 412 (2011), 3413–3417.

[34] T. Komatsu, *A certain power series and the inhomogeneous continued fraction expansions*, J. Number Theory 59 (1996), 291–312.

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

[36] M. Morse and G.A. Hedlund, *Symbolic dynamics II: Sturmian sequences*, Amer. J. Math. 62 (1940), 1–42.

[37] C Pomerance, J.M. Robson and J. Shallit, *Automaticity. II. Descriptional complexity in the unary case*, Theoret. Comput. Sci. 180 (1997), 181–201.

[38] D. Ridout, *Rational approximations to algebraic numbers*, Mathematika 4 (1957), 125–131.

[39] K. F. Roth, *Rational approximations to algebraic numbers*, Mathematika 2 (1955), 1–20; corrigendum, 168.

Université de Strasbourg, CNRS, IRMA, UMR 7501, 7 rue René Descartes, 67084 Strasbourg, France

E-mail address: bugeaud@math.unistra.fr

Department of Mathematics Education, Dongguk University – Seoul, Seoul 04620, Korea.

E-mail address: kim2010@dongguk.edu