Recurrence and frequencies*

Valérie Berthé and Ahmed Mimouni

[1] IRIF, Université Paris Cité, 75013 Paris, France berthe@irif.fr
[2] LACL, Université Paris-Est-Créteil, 94000 Créteil, France amimouni@lacl.fr

Abstract. Given an infinite word with values in a finite alphabet that admit frequencies for all its factors, the smallest frequency function associates with a given positive integer the smallest frequency for factors of this length, and the recurrence function measures the gaps between successive occurrences of factors. We develop in this paper the relations between the growth orders of the smallest frequency and of the recurrence functions. As proved by M. Boshernitzan, linearly recurrent words are characterized in terms of the smallest frequency function. In this paper, we see how to extend this result beyond the case of linearly recurrent words. We first establish a general relation between the smallest frequency and the recurrence fuctions. Given a lower bound for the smallest frequency function, this relation provides an upper bound for the recurrence function which involves the primitive of this lower bound. We then see how to improve this relation in the case of Sturmian words where the product of the smallest frequency and the recurrence function is bounded for most of the lengths of factors. We also indicate how to construct Sturmian words having a prescribed behaviour for the smallest frequency function, and for the product of the smallest frequency and the recurrence function.

Keywords: Recurrence function · Frequencies · Smallest frequency function · Factor complexity · Sturmian words.

1 Introduction

There exist various combinatorial natural measures of disorder that occur simultaneously in word combinatorics, symbolic dynamics, and in the context of aperiodic order. Aperiodic order refers to the mathematical formalization of quasicrystals, i.e., solids with an aperiodic atom structure. This involves different mathematical objects, such as tilings, cut-and-project sets, Delone point sets, or infinite words over a finite alphabet, and various functions allow the quantification of order. For words, these notions involve factors, and for tilings and point sets, patches or patterns. In the case of infinite words (which is the case on which we focus here), it is natural to count how many different factors exist (factor complexity), whether or not they reappear (recurrence), and how often they occur (frequencies and shift invariant measures).

* This work was supported by the Agence Nationale de la Recherche through the project “Codys” (ANR-18-CE40-0007).

More precisely, given an infinite word x with values in a finite alphabet, the *factor complexity* p_x counts the number of factors of a given size of an infinite word. The *recurrence function* R_x measures the gaps between successive occurrences of factors, i.e., how often they come back. It is also called repetitivity in the tiling setting, such as introduced in the seminal paper [15], see also [16]. Lastly, the frequency of a factor w in the infinite word x is defined as the limit when n tends towards infinity, if it exists, of the number of occurrences of w in $u_0 u_1 \cdots u_{n-1}$ divided by n. If the infinite word x admits frequencies for all factors, we then consider the *smallest frequency function* e_x for factors of a given length.

The relevance in this context of the quantity $ne_x(n)$ has been highlighted by M. Boshernitzan. He introduced indeed in [6] the so-called ne_n condition; see also [8, Chap. 7] for more details and Theorem 1 below. In a nutshell, $\limsup ne_x(n) > 0$ implies that frequencies are uniform (in dynamical terms, one has unique ergodicity), and $\liminf ne_x(n) > 0$ is equivalent to linear recurrence, which means that there exists a constant $c > 0$ such that factors of length n in x occur in any factor of length cn in x. This is an important property in the context of aperiodc order as described in the survey [1]; see also [11].

In this paper, we see how to extend the strategy developed by M. Boshernitzan for his characterization of linear recurrence, by exploring the relations between these notions. This illustrates the fact that recurrence and frequencies are closely related. We then focus on the Sturmian case which allows to produce examples of infinite words with a prescribed lower bound on e_x.

Basic notions and general statements relating recurrence, frequencies and factor complexity are recalled in Section 2. Our main result relating frequencies and recurrence, namely Theorem 2, is stated in Section 3. We then focus on the case of Sturmian words in Section 4. We provide conditions on the sequence of partial quotients in order to get prescribed behaviour for the smallest frequency function (see Theorem 5 and 6). We also prove that the product $e_x R_x$ takes bounded values for most integer values (see Theorem 7), even in the case where the partial quotients in the continued fraction expansion of the angle of the Sturmian word takes non-bounded values.

2 Basic notions

Let $\mathcal{A}$ be an alphabet and let $\mathcal{A}^*$ stand for the set of all (finite) words over $\mathcal{A}$. Let x be an infinite word in $\mathcal{A}^{\mathbb{N}}$. Let $L_x(n)$ stand for the set of factors of length n of x, and L_x for the set of all its factors. The word x is said to be *uniformly recurrent* if every factor appears infinitely often and with bounded gaps (or, equivalently, if for every integer n, there exists an integer m such that every factor of x of length m contains every factor of length n). The recurrence functions R_x is then defined as

$$R_x(n) := \min\{m : \forall w \in L_x(m), L_x(n) \text{ is included in the set of factors of } w\}$$

and the infinite word x is then said to be *linearly recurrent* if there exists a positive integer $c > 0$ such that for every positive integer n, one has

$$R_x(n) \leq cn.$$

We then define $\mu_x(w)$ as the *frequency* of the word $w \in L_x$ in x, if it exists:

$$\mu_x(w) = \lim_{n\to\infty} \frac{|x_0 x_1 \cdots x_n|_w}{n+1}.$$

If every factor of x has a frequency, we will say that x has frequencies.

If x has frequencies, then x is said to have *uniform frequencies* if for every w, and for every k, $\frac{|x_k x_{k+1} \cdots x_{k+n}|_w}{n+1}$ tends with n to $\mu_x(w)$ uniformly in k.

The *smallest frequency* function $e_x(n)$ is then defined as

$$e_x(n) := \min_{w\in L_x(n)} \mu_x(w).$$

Note that every linearly recurrent word has (uniform) frequencies, by [9]. However there exist uniformly recurrent words that do not admit frequencies (see e.g. [14] for examples of codings of interval exchanges on four letters).

The following proposition is usually stated in terms of shift-invariant measures; see e.g. [8, Chapter 7].

Proposition 1. *Let x be an infinite word that has frequencies. Let $E \subseteq L_x$ be a set of words such that no element of E is prefix of another element of E. We then have*

$$\sum_{w\in E} \mu_x(w) \leq 1.$$

Proof. Suppose that E has at least two elements (otherwise the proof comes easily). Suppose that $\sum_{w\in E} \mu_x(w) > 1$. Then there exists a positive integer n such that $\sum_{w\in E} \frac{|x_0\cdots x_n|_w}{n+1} > 1$, i.e., $\sum_{w\in E} |x_0 \cdots x_n|_w > n+1$. Then the pigeonhole principle gives us the existence of a positive integer i, and of v, v' in E, such that v and v' appear at position i in x. That would mean that v is a prefix of v', or the other way around, contradicting the hypothesis.

The followng result relates in a natural way recurrence and factor complexity.

Proposition 2. *[18] Let x be an infinite word. One has*

$$R_x(n) \geq p_x(n) + n - 1 \text{ for all } n.$$

We now state a first bridge between recurrence and frequencies, by splitting words into smaller words in which we are sure to find some fixed factor.

Proposition 3. *Let x be an infinite word with frequencies. One has*

$$e_x(n) \geq \frac{1}{R_x(n)} \text{ for all } n. \tag{1}$$

Proof. Let us show that for every factor w of x of size m, $\mu_x(w) \geq \frac{1}{R_x(|w|)}$. Let n be given. We split the prefix $x_0 \cdots x_n$ of x into smaller factors of size $R_x(|w|)$, in which we must have at least one occurrence of w, by definition; hence

$$|x_0 \cdots x_n|_w \geq \lfloor (n+1)/R_x(|w|) \rfloor,$$

which gives us

$$\mu_x(w) = \lim_{n\to\infty} \frac{|x_0 \cdots x_n|_w}{n+1} \geq \lim_{n\to\infty} \frac{\lfloor (n+1)/R_x(|w|) \rfloor}{n+1} = \frac{1}{R_x(|w|)}.$$

Another easy-to-build bridge is the one giving a lower bound on the complexity function knowing the smallest frequency function e_x. It comes almost directly from Proposition 1.

Proposition 4. *Let x be an infinite word with frequencies, and n a positive integer such that $e_x(n) > 0$. We then have*

$$p_x(n) \leq \frac{1}{e_x(n)}.$$

Proof. By Proposition 1, one gets

$$1 \geq \sum_{v \in L_x(n)} \mu_x(v) \geq \sum_{v \in L_x(n)} e_x(n) \geq p_x(n) \cdot e_x(n).$$

Remark 1. Morse-Hedlund's theorem states that for any infinite word x, if there exists n such that $p_x(n) \leq n$, then, the word x is ultimately periodic. Hence, for any aperiodic infinite word x and for any integer n, one has $p_x(n) > n$. Proposition 4 then implies that $e_x(n) < \frac{1}{n}$ for all n.

Let x be an infinite word that admits frequencies. The condition

$$\limsup n e_x(n) > 0$$

implies that the frequencies are uniform, as shown by M. Boshernitzan [6]. Following the approach developed by M. Boshernitzan, we investigate in the next section the condition $\liminf n e_x(n) > 0$ and see how it allows one to relate recurrence and frequencies.

3 Relating recurrence and frequencies

The following result, due to M. Boshernitzan, provides a characterization of linear recurrece in terms of frequencies.

Theorem 1. *[8, Ex. 174] An infinite word x is linearly recurrent if and only if $\liminf n e_x(n) > 0$.*

We now see how to extend Boshernitzan's result, namely Theorem 1, as a generalisation of [8, Ex. 174].

Theorem 2. *Let $f : \mathbb{R}_+^* \to \mathbb{R}_+^*$ be a continuous non-increasing function and F one of its primitives. Let x be a uniformly recurrent infinite word that has frequencies. We assume that for n large enough, we have*

$$e_x(n) \geq f(n).$$

We then have, for n large enough,

$$F(n) + 1 \geq F(R_x(n)).$$

Proof. The statement is true if x is constant. We now assume that x is a non-constant word. Let u be a factor of x. Let

$$r_x(u) := \min\{n : \forall w \in L_x(n),\ u \text{ is a factor of } w\}.$$

Let n be an integer, and $u = u_1 u_2 \cdots u_n$ a factor of x of size n such that $r_x(u) = R_x(n)$ (such u word exists because we work on a finite alphabet). The uniform recurrence of x gives us the existence of a factor $v = v_1 v_2 \cdots v_m$ of x of maximal length such that u appears only once in vu. Let us show that $|v| + |u| = m + n = r_x(u) = R_x(n)$.

Indeed, by definition of $r_x(u)$, and because $r_x(u) > 1$ (x is not constant), there exists a factor $w = w_1 \cdots w_{r_x(u)-1}$ of x such that u does not occur in w, but u occurs in wa, with a being a letter of $\mathcal{A}$, i.e., u appears only once in wa. Then, we can write wa as $w'u$, with $w' = w_1 \cdots w_{r_x(u)-n}$. Hence, since v is of maximal length, $|v| = m \geq r_x(u) - n$, i.e., $m + n \geq r_x(u)$.
Moreover, we suppose by contradiction that $m + n > r_x(u)$. Taking away the last letter of vu gives us the factor $vu_1 \cdots u_{n-1}$, of size $m+n-1 \geq r_x(u)$, which thus contains u. However, it cannot contain u, since u occured only once in vu, hence the desired contradiction.

Now, suppose that we have $i > j \geq 0$, such that $v_i \cdots v_m u$ is a prefix of $w := v_j \cdots v_m u$. This gives

$$v_j \cdots v_m u = w_1 \cdots w_{n+(m-j)} = v_i \cdots v_m u w_{n+(m-i)+1} \cdots w_{n+(m-j)}.$$

This shows that u appears twice in $w = v_j \cdots v_m u$, i.e., u appears twice in vu, contradicting the hypothesis: there do not exist $i \neq j$ such that $v_i \cdots v_m u$ is a prefix of $v_j \cdots v_m u$. Proposition 1 then gives us

$$1 \geq \mu_x(u) + \sum_{i=1}^{m} \mu_x(v_i \cdots v_m u) \geq \sum_{i=0}^{m} e_x(m-i+n) \geq \sum_{i=0}^{m} f(m-i+n) = \sum_{i=n}^{m+n} f(i),$$

and hence, since f is non-increasing and positive,

$$1 \geq \int_n^{m+n+1} f(t)\ \mathrm{d}t;$$

consequently, since f is positive and thus F increasing, one gets

$$F(n) + 1 \geq F(R_x(n) + 1) \geq F(R_x(n)). \tag{2}$$

The essence of the proof lies in the sequence of factors of increasing length ranging from n to $R_x(n)$, on which we can use Proposition 1. This creates the inequality in which appear $f(n)$ and $f(R_x(n))$, which then, when integrated, relates $F(n)$ and $F(R_x(n)$.

As a first application we recover Theorem 1, with a function F growing slowly towards infinity, i.e., $f(n) := \frac{c}{n}$ and $F = \int f = c \cdot \log$.

We now consider other lower bounds on $e_x(n)$ with Corollary 1.

Corollary 1. *Let x be a uniformly recurrent infinite word that has frequencies.*
If there exist $\beta \in (0,1]$ and $c > 0$ such that for n large enough

$$e_x(n) \geq \frac{c}{n(\log n)^\beta} \ ,$$

then there exists $c' > 1$ such that

$$R_x(n) = O(n^{c'}) \quad \textit{and} \quad p_x(n) = O(n(\log n)^\beta).$$

Proof. We first consider the case of $\beta = 1$. Suppose that there exists $c > 0$ such that, for n large enough, we have $n \log(n) e_x(n) \geq c$. Then, with $f(n) := c \cdot \frac{1/n}{\log(n)}$, and thus $F(n) := c \cdot \log(\log(n))$, Theorem 2 gives us

$$\log(\log(n)) + \frac{1}{c} \geq \log(\log(R_x(n))),$$

which yields

$$n^{\exp(\frac{1}{c})} \geq R_x(n),$$

and finally

$$R_x(n) = O(n^{\exp(\frac{1}{c})}).$$

The result for the complexity comes from Proposition 4.

We now consider the case of $\beta \in (0,1)$. Suppose that there exists $c > 0$ such that for n large enough, we have $n(\log n)^\beta e_x(n) \geq c$. Theorem 2 with $F(n) := \frac{c}{1-\beta} \cdot (\log n)^{1-\beta}$ gives us

$$\frac{c}{1-\beta} \cdot (\log(n))^{1-\beta} + 1 \geq \frac{c}{1-\beta} \cdot \log(R_x(n))^{1-\beta},$$

which yields, for n large enough and for some $\epsilon > 0$,

$$(1+\epsilon)(\log(n))^{1-\beta} \geq (\log(n))^{1-\beta} + \frac{1-\beta}{c} \geq \log(R_x(n))^{1-\beta},$$

and finally

$$R_x(n) \leq n^{(1+\epsilon)^{\frac{1}{1-\beta}}},$$

or, with $c' := (1+\epsilon)^{\frac{1}{1-\beta}} > 1$,

$$R_x(n) = O(n^{c'}).$$

Remark 2. We may want a priori to consider other values for β in the second statement of Corollay 1. Proposition 4 yields that for n large enough, $ne_x(n) \leq 1$ implies $p_x(n) \leq n$, which thus implies that x is ultimaltey periodic, by Morse-Hedlund's theorem. Hence, if $\beta < 0$, then for n large enough, $e_x(n) \geq \frac{c}{n(\log n)^\beta} \geq \frac{1}{n}$, which would thus imply x to be ultimately periodic. On the other hand, if $\beta > 1$, then F is a negative function that converges towards 0, making us unable to reach any conclusion using (2). This is also why we do not consider functions of the form $\frac{1}{n^\beta}$ for $\beta > 1$.

4 Frequencies and reucrence for Sturmian words

We now focus on the family of Sturmian words where the bounds provided by Theorem 1 can be improved, by investigating in more details the case where $\liminf ne_x(n) = 0$.

4.1 First properties

Sturmian words are defined as the (one-sided) words having exactly $n+1$ factors of length n, for every positive integer n. They are equivalently defined as symbolic codings (with respect to two-interval partitions) of the irrational translations R_α of the unit circle (that is, the one-dimensional torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$), where $R_\alpha \colon \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z},\ x \mapsto x + \alpha \bmod 1$. More precisely, the infinite word $x = (x_n)_{n\in\mathbb{N}} \in \{0,1\}^\mathbb{N}$ is a Sturmian word if there exist $\alpha \in (0,1)$, $\alpha \notin \mathbb{Q}$ (called its *angle*), $s \in \mathbb{R}$ such that

$$\forall n \in \mathbb{N},\ x_n = i \Longleftrightarrow R_\alpha^n(s) = n\alpha + s \in I_i \pmod 1,$$

with either $I_0 = [0, 1-\alpha),\ I_1 = [1-\alpha, 1)$, or $I_0 = (0, 1-\alpha],\ I_1 = (1-\alpha, 1]$. For more on Sturmian words, see the corresponding chapters in [17,20] and the references therein.

Let x be a Sturmian word with angle $\alpha \in (0,1) \setminus \mathbb{Q}$. Note that Sturmian words have frequencies. The continued fraction expansion of the angle α allows the expression of the smallest frequency function e_x (see Theorem 4) and of the recurrence function R_x (see Section 4.3). Before stating them, let us introduce some notation. Consider the continued fraction expansion of α, with

$$\alpha = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}$$

The positive integer digits a_n are called *partial quotients*. The rational numbers p_n/q_n, with p_n, q_n being coprime positive integers being defined as

$$\frac{p_n}{q_n} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{}{\ddots + \cfrac{1}{a_n}}}},$$

are called *convergents*. Let

$$\theta_n = (-1)^n (q_n\alpha - p_n) = |q_n\alpha - p_n|$$

for all nonnegative n. One has $q_{-1} = 0$, $p_{-1} = 1$, $q_0 = 1$, $p_0 = 0$, and for all n

$$q_{n+1} = a_{n+1}q_n + q_{n-1},\ p_{n+1} = a_{n+1}p_n + p_{n-1},\ \text{and } \theta_{n-1} = a_{n+1}\theta_n + \theta_{n+1}. \quad (3)$$

4.2 On the frequencies of Sturmian words

In this section, we indicate how to obtain Sturmian words with a prescribed behaviour for the smallest frequency function (see in particular Remark 4).

Theorem 3. *[2] Let x be a Sturmian word of angle α, with $\alpha \in (0,1)$, α irrational. Let $m \geq 1$.*

Assume that $kq_n + q_{n-1} < m < (k+1)q_n + q_{n-1}$, with $n \geq 1$ and $1 \leq k \leq a_{n+1}$. The frequencies of factors of length m belong to the set

$$\{\theta_n, \theta_{n-1} - k\theta_n, \theta_{n-1} - (k-1)\theta_n\} = \{\theta_n, \theta_{n+1} + (a_{n+1} - k)\theta_n, \theta_{n+1} + (a_{n+1} - k + 1)\theta_n\}.$$

Assume $m = kq_n + q_{n-1}$, with $n \geq 1$ and $1 \leq k \leq a_{n+1}$. The frequencies of factors of length m belong to the set

$$\{\theta_n, \theta_{n-1} - k\theta_n\} = \{\theta_n, \theta_{n+1} + (a_{n+1} - k)\theta_n\}.$$

We thus deduce the following, with the notation e_α standing for the smallest frequency function for the Sturmian words of angle α.

Theorem 4. *Let α be an irrational number, $m > 0$, and n be the unique integer such that $q_n < m \leq q_{n+1}$. Then, $e_\alpha(m) = \theta_n = |q_n\alpha - p_n|$.*

The following estimates hold for the quantity $\theta_n = |q_n\alpha - p_n|$.

Proposition 5. *For every integer n, we have*

$$\frac{1}{2q_{n+1}} \leq \frac{1}{q_n + q_{n+1}} \leq \theta_n \leq \frac{1}{q_{n+1}} \quad \text{and} \quad \prod_{k=1}^{n} a_k \leq q_n \leq \prod_{k=1}^{n} (a_k + 1).$$

Proof. The first statement is a classical property of continued fractions (see e.g. [10]) based on the fact that, for all n, there exists $\alpha_n \in (0,1)$ such that one has

$$\left|\alpha - \frac{p_n}{q_n}\right| = \left|\frac{p_n + \alpha_n p_{n-1}}{q_n + \alpha_n q_{n-1}} - \frac{p_n}{q_n}\right|.$$

The second statement comes from the recurrence relation $q_{n+1} = a_{n+1}q_n + q_{n-1}$ (see (3) together with the fact that the sequence $(q_n)_n$ is increasing.

Theorem 5. *Let x be a Sturmian word of angle α, with α being an irrational number in $(0,1)$. Let $f : \mathbb{R}_+^* \to \mathbb{R}_+^*$ be a non-increasing function. If*

$$a_{n+1} = O\left(\frac{1}{q_n f(q_n)}\right),$$

then there exists $c > 0$ such that, for all n

$$e_\alpha(n) \geq c \cdot f(n).$$

Proof. The assumptions on f imply that there exists $c > 0$ such that, for n large enough, one has

$$a_{n+1} + 2 \leq \frac{c}{q_n f(q_n)}.$$

Let m be large enough, and let n be the unique integer such that $q_n < m \leq q_{n+1}$. We have that $\frac{q_{n-1}}{q_n} \leq 1$, hence

$$a_{n+1} + 1 + \frac{q_{n-1}}{q_n} \leq \frac{c}{q_n f(q_n)}, \text{ i.e., } q_{n+1} + q_n \leq \frac{c}{f(q_n)}.$$

By Theorem 4, $e_\alpha(m) = \theta_n$, and by Proposition 5,

$$e_\alpha(m) = \theta_n \geq \frac{1}{q_{n+1} + q_n} \geq \frac{f(q_n)}{c}.$$

Finally, since f is non-increasing,

$$e_\alpha(m) \geq \frac{f(q_n)}{c} \geq \frac{f(m)}{c}.$$

Remark 3. Theorem 5 (see also Theorem 6 below) relies on assumptions invoving relations between both sequences $(a_n)_{n\in\mathbb{N}}$ and $(q_n)_{n\in\mathbb{N}}$. However, the following inequality allows the statement of hypothesis to involve only the sequence $(a_n)_{n\in\mathbb{N}}$. Indeed, let $f : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-increasing function. By Proposition 5, one has, for all n,

$$\frac{1}{(\prod_{k=1}^n a_k) f(\prod_{k=1}^n a_k)} \leq \frac{1}{q_n f(q_n)}.$$

The next statement is the counterpart of Theorem 5 in terms of the existence of an upper bound for infinitely many integers for e_x.

Theorem 6. *Let x be a Sturmian word of angle α, with α being an irrational number in $(0,1)$. Let $f : \mathbb{R}_+^* \to \mathbb{R}_+^*$. If $a_{n+1} \geq \frac{1}{q_n f(q_n+1)}$ for infinitely many n, then $e_\alpha(n) \leq f(n)$, for infinitely many n.*

Proof. Let n be such that $a_{n+1} \geq \frac{1}{q_n f(q_n+1)}$. One has

$$\frac{1}{q_n f(q_n+1)} \leq a_{n+1} \leq a_{n+1} + \frac{q_{n-1}}{q_n}.$$

Multiplying by q_n together with (3) yields

$$\frac{1}{f(q_n+1)} \leq q_{n+1}.$$

We take $m = q_n + 1$. By Theorem 4, $e_\alpha(m) = \theta_n$, and by Proposition 5, one has

$$e_\alpha(q_n+1) = \theta_n \leq \frac{1}{q_{n+1}} \leq f(q_n+1).$$

Remark 4. Let us illustrate Theorem 5 and 6 with the function $f : x \mapsto \frac{1}{x \log x}$ from Corollary 1. Let x be a Sturmian word of angle α, with α being an irrational number. Theorems 5 and 6 give the following. If $a_{n+1} \leq \log q_n$ for n large enough, then there exists $c > 0$ such that $e_\alpha(n) \geq \frac{c}{n \log n}$, for all n. If $a_{n+1} \geq \frac{q_n+1}{q_n} \log(q_n+1)$ for infinitely many n, then $e_\alpha(n) \leq \frac{1}{n \log n}$, for infinitely many n.

4.3 On the recurrence of Sturmian words

Let x be a Sturmian word with angle α ($\alpha \notin \mathbb{Q}$, $\alpha \in (0,1)$). Let $(q_k)_{k \in \mathbb{N}}$ denote the sequence of denominators of the convergents of the continued fraction expansion of α. One has, by [19], the following:

$$R_x(m) = m - 1 + q_n + q_{n+1}, \text{ for } q_n \leq m < q_{n+1}. \quad (4)$$

In particular, a Sturmian word is linearly recurrent if and only if its angle has bounded partial quotients a_k. For more on the recurrence function of Sturmian words, see e.g. [7,4,21].

Next theorem shows that, beyond the case of linear recurrence (that is, even when $\liminf n e_x(n) = 0$), the smallest frequency and the recurrence functions are still related in the Sturmian case: the functions e_x and $1/R_x$ have the same growth order for most of the integers. The statement about almost every α in the next theorem refers to a set of full Lebesgue measure. We also revisit the condition from Theorems 5 and 6 for the functions $e_x R_x$, R_x, and e_x.

Theorem 7. *Let x be a Sturmian word of angle α, with $\alpha \in (0,1)$, α irrational. Let $(a_n)_{n \in \mathbb{N}}$ be the sequence of partial quotients and let $(q_n)_{n \in \mathbb{N}}$ be the sequence* *of denominators of the convergents in the continued fraction expansion of α. One has, for all positive integers m that are not in the sequence $(q_n)_{n\in\mathbb{N}}$,*

$$1 \le e_x(m)R_x(m) \le 3.$$

Moreover, for almost all $\alpha \in (0,1)$,

$$\limsup_n e_x(q_n)R_x(q_n) = +\infty.$$

More precisely, let $f : \mathbb{R}_+^* \to \mathbb{R}_+^*$ be a non-increasing function.

If $a_{n+1} = O\left(\frac{1}{q_n f(q_n)}\right)$, then there exist $c, c' > 0$ such that

$$e_x(q_n)R_x(q_n) \le \frac{c}{q_n f(q_n)}, \quad R_x(n) \le \frac{c}{f(n)} \text{ and } e_x(n) \ge c' f(n) \text{ for all } n.$$

If $a_{n+1} \ge \frac{1}{q_n f(q_n+1)}$, then the following properties hold for infinitely many n:

$$e_\alpha(q_n)R_x(q_n) \ge \frac{1}{2q_n f(q_n)}, \quad R_x(n) \ge \frac{1}{f(n)} \quad \text{and} \quad e_x(n) \le f(n).$$

Proof. Let x be a Sturmian word of angle α and let m be a positive integer. The lower bound for $e_x(m)R_x(m)$ comes from Proposition 3.

Now let n be such that $q_n \le m < q_{n+1}$. Let us first assume $m \ne q_n$. By Theorem 4, $e_x(m) = \theta_n$ and moreover, $R_x(m) = m - 1 + q_n + q_{n+1}$. Regarding the upper bound, one thus gets, by Proposition 5, that

$$e_x(m)R_x(m) = \theta_n(m - 1 + q_{n+1} + q_n) \le \frac{1}{q_{n+1}}(m - 1 + q_{n+1} + q_n) \le 3,$$

since $m < q_{n+1}$. We thus have proved that $1 \le e_x(m)R_x(m) \le 3$ for $m \ne q_n$.

We now assume $m = q_n$. Then, by Proposition 5, one has $e_x(m) = \theta_{n-1}$ and

$$e_x(q_n)R_x(q_n) = \theta_{n-1}(q_n - 1 + q_{n+1} + q_n) \le \frac{1}{q_n}(q_n - 1 + q_{n+1} + q_n),$$

hence

$$1 \le e_x(q_n)R_x(q_n) \le a_{n+1} + 3. \tag{5}$$

Moreover, again by by Proposition 5, one has

$$e_x(m)R_x(m) = \theta_{n-1}(q_n - 1 + q_{n+1} + q_n) \ge \frac{1}{q_n + q_{n-1}}(q_n - 1 + q_{n+1} + q_n).$$

Hence

$$e_x(m)R_x(m) \ge \frac{(a_{n+1}+2)q_n + q_{n-1} - 1}{q_n + q_{n-1}} \ge \frac{a_{n+1}}{2}. \tag{6}$$

We now consider $f : \mathbb{R}^+ \to \mathbb{R}^+$ to be a non-increasing function. We assume $a_{n+1} = O\left(\frac{1}{q_n f(q_n)}\right)$. There exists $c > 0$ such that, for n large enough, $a_{n+1}+3 \le \frac{c}{q_n f(q_n)}$. Hence, by (5), one has

$$1 \le e_x(q_n)R_x(q_n) \le \frac{c}{q_n f(q_n)}.$$

Moreover, let m be given and let n be such that $q_n \leq m < q_{n+1}$. Then, by (4), there exists $c > 0$ such that

$$R_x(m) \leq (2a_{n+1} + 3)q_n \leq \frac{2c}{f(q_n)} \leq \frac{2c}{f(m)}$$

since f is non-increasing. The lower bound on e_x comes from Proposition 3.

We now assume $a_{n+1} \geq \frac{1}{q_n f(q_n+1)}$ for infinitely many n. Then, by (6), one gets, for infinitely many n, $e_x(q_n)R_x(q_n) \geq \frac{1}{2q_n f(q_n)}$. Moreover, let $m = q_n + 1$. Then, by (4), $R_x(q_n + 1) \geq q_{n+1} \geq a_{n+1}q_n \geq \frac{1}{f(q_n+1)}$. The upper bound on e_x comes from the proof of Theorem 6.

It remains to prove the statement on the generic behaviour of $e_x(q_n)R_x(q_n)$. According to [5], one has $\limsup a_n = +\infty$, for Lebesgue almost all $\alpha \in (0, 1)$. Consequently, by (6), $\limsup_n e_x(q_n)R_x(q_n) = +\infty$ for almost all α.

Remark 5. Let us continue with the example of the function $f(x) = \frac{1}{x \log x}$, such as considered in Remark 4. We assume $a_{n+1} = O(\log q_n)$. Then, there exists $c > 0$ such that $e_x(n) \geq \frac{c}{n \log n}$, for all n, and $R_x = O(n \log n)$, which improves the bound of Corollary 1.

5 Concluding remarks

We have seen with Theorem 2 that the approach developed by M. Boshernitzan for his characterization of linearly recuurent words in terms of the smallest frequency function (see Theorem 1) can be extended beyond the case of linearly recurrent words. Corollary 1 shows for instance that a lower bound of order $\frac{1}{n \log}$ for the smallest frequency function induces a polynomial upper bound for the recurrence function. A strategy in order to provide concrete examples of words having a prescribed behaviour for the growth of the smallest frequency function is to work with explicit S-adic expansions, such as the expansions provided by continued fractions for Sturmian words. The precise arithmetic descriptions of the smallest frequency and of the recurrence functions then allow to improve in Theorems 5 and 7 the results obtained in Theorem 2, as stressed in Remark 4 and 5, when comparing with Corollary 1.

The study of multidimensional words is natural in the context of aperiodic order. We plan to investigate the case of multidimensional Sturmian words (see [3]) in order to see how to extend the relation from Theorem 7, i.e., the quantity $e_x R_x$. Note that the study of the frequencies for factors is related to the three-length theorem. See [13,12] for higher-dimensional generalizations.

References

1. Aliste-Prieto, J., Coronel, D., Cortez, M. I. and Durand, F. and Petite, S.: Linearly repetitive Delone sets, In: Mathematics of aperiodic order. Progr. Math., **309**, 195–222 (2015)

2. Berthé, V.: Fréquences des facteurs des suites sturmiennes, Theoret. Comput. Sci., **165**(2), 295–309 (1996)
3. Berthé, V., Vuillon, L.: Tilings and rotations on the torus: a two-dimensional generalization of Sturmian sequences. Discrete Math. **223** (2000), 27–53 (2000)
4. Berthé, V., Cesaratto, E., Rotondo, P., Vallée, B., Viola, A.: Recurrence function on Sturmian words: a probabilistic study. In: Mathematical foundations of computer science 2015. Part I, LNCS, vol. 9234, pp. 116–128. Springer, Heidelberg (2015)
5. Borel, E.: Les probabilités dénombrables et leurs applications arithmétiques, Rend. Circ. Mat. Palermo **27**, 247–271 (1909)
6. Boshernitzan, M: A condition for unique ergodicity of minimal symbolic flows. Ergod. Th. & Dynam. Sys. **12**(3), 425–428 (1992)
7. Cassaigne, J.: Limit values of the recurrence quotient of Sturmian sequences. Theoret. Comput. Sci. **218**(1), 3–12 (1999)
8. Combinatorics, automata and number theory. Encyclopedia of Mathematics and its Applications **135**. Eds. Berthé, V. and Rigo, M., Cambridge University Press, Cambridge, (2010)
9. Durand, F., Host, B., Skau, C.: Substitutions, Bratteli diagrams and dimension groups. Ergod. Th. Dyn. Sys. **19**, 952–993 (1999).
10. Hardy, G. H., Wright, E. M.: An Introduction to the Theory of Numbers. Oxford Science Publications, Published by Oxford University Press (1980)
11. Haynes, A., Koivusalo, H., Walton, J.: A characterization of linearly repetitive cut and project sets. Nonlinearity, **31**(2), 515–539 (2018)
12. Haynes, A., Marklof, J.: Higher dimensional Steinhaus and Slater problems via homogeneous dynamics. Ann. Sci. Éc. Norm. Supér. (4) **53**, 537–557 (2020)
13. Haynes, A., Marklof, J.: A five distance theorem for Kronecker sequences. Int. Math. Res. Not. IMRN **24**, 19747–19789 (2022)
14. Keane, M.: Non-ergodic interval exchange transformations. Israel J. Math. **26**, 188–196 (1977)
15. Lagarias, J. C., Pleasants, P. A. B.: Local complexity of Delone sets and crystallinity. Canad. Math. Bull., **45**(4), 634–652 (2002)
16. Lagarias, J. C., Pleasants, P. A. B: Repetitive Delone sets and quasicrystals. Ergodic Theory Dynam. Systems **23** , 831–867 (2003)
17. Lothaire, M.: Algebraic Combinatorics on Words, vol. 90. Encyclopedia of Mathematics and Its Applications. Cambridge University Press (2002)
18. Morse, M. and Hedlund, G. A.: Symbolic dynamics. Amer. J. Math. **60**, 815–866 (1938)
19. Morse, M. and Hedlund, G. A.: Symbolic Dynamics II. Sturmian trajectories. Amer. J. Math. **62**, 1–42 (1940)
20. Pytheas Fogg, N.: Substitutions in dynamics, arithmetics and combinatorics. Lecture Notes in Mathematics, **1794**,Springer-Verlag, Berlin, (2002)
21. Rotondo, P., Vallée, B: The recurrence function of a random Sturmian word. In: 2017 Proceedings of the Fourteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO), pp. 100–114, SIAM, Philadelphia, PA, (2017)