Recurrence function on Sturmian words: a probabilistic study

V. Berthé[1], E. Cesaratto[2], P. Rotondo[3], B. Vallée[4], and A. Viola[3]

[1] LIAFA, CNRS UMR 7089, Univ. Paris Diderot, France
[2] Conicet and Univ. Nac. of Gral. Sarmiento, Argentina
[3] Universidad de la República,Uruguay
[4] GREYC, CNRS UMR 6072, Univ. de Caen, France

Abstract. This paper is a first attempt to describe the probabilistic behaviour of a random Sturmian word. It performs the probabilistic analysis of the recurrence function which provides precise information on the structure of such a word. With each Sturmian word of slope α, we associate particular sequences of factor lengths which have a given "position" with respect to the sequence of continuants of α, we then let α to be uniformly drawn inside the unit interval $[0, 1]$. This probabilistic model is well-adapted to better understand the role of the position in the recurrence properties.

1 Introduction

The recurrence function measures the "complexity" of an infinite word and describes the possible occurrences of finite factors inside it together with the maximal gaps between successive occurrences. This recurrence function is thus widely studied, notably in the case of Sturmian words (see [3,9]) which are in a precise sense the simplest infinite words which are not eventually periodic (see e.g. [8]). With each Sturmian word is associated an irrational number α, and many of its characteristics depend on the continued fraction expansion of α. This is in particular the case for the recurrence function $n \mapsto R_\alpha(n)$, where the integer $R_\alpha(n)$ is the length of the smallest "window" which is needed for discovering the set $\mathcal{L}_\alpha(n)$ of all the finite factors of length n inside α. As this set $\mathcal{L}_\alpha(n)$ is widely used in many applications of Sturmian words (for instance quasicrystals, or digital geometry), the function $n \mapsto R_\alpha(n)$ thus intervenes very often as a pre-computation cost, and it is important to better understand this function "on average", when the real α is randomly chosen in the unit interval.
Most of the classical studies on the recurrence function deal with a *fixed* α, and the usual focus is put on *extremal* behaviours of the recurrence function. Here, we adopt a "dual" approach which is probabilistic: with each α, we associate *particular* sequences of indices n (i.e., factor lengths), which have a *given* "position" with respect to the sequence of continuants $(q_k(\alpha))_k$, we then let α be *uniformly drawn* inside the unit interval, and we perform a *probabilistic* study to better understand the role of the position in the recurrence function.
The expression of the recurrence function is recalled in Section 2. Our viewpoint and our main results are given in Section 3. Proofs are provided in Section 4.

2 The recurrence function of Sturmian words

Notation. In the sequel $\varphi = (\sqrt{5}-1)/2 = 0.6180339\ldots$ stands for the inverse of the golden ratio, and for two integers a, b, the set of integers n that satisfy $a \le n \le b$ is denoted by $[\![a, b]\!] := [a, b] \cap \mathbb{N}$.

We consider a finite set $\mathcal{A}$ of *symbols*, called *alphabet*. Let $u = (u_n)_{n\in\mathbb{N}}$ be an infinite word in $\mathcal{A}^{\mathbb{N}}$. A finite word w of length n is a factor of u if there exists an index m for which $w = u_m \ldots u_{m+n-1}$. Let $\mathcal{L}_u(n)$ stand for the set of factors of length n of u. Two functions describe the set $\mathcal{L}_u(n)$ inside the word u, namely the complexity and the recurrence function.

The (*factor*) *complexity function* of the infinite word u is defined as the sequence $n \mapsto p_u(n) := |\mathcal{L}_u(n)|$. The eventually periodic words are the simplest ones, in terms of the complexity function, and satisfy $p_u(n) \le n$ for some n. The simplest words that are not eventually periodic satisfy the equality $p_u(n) = n+1$ for each $n \ge 0$. Such words do exist, they are called *Sturmian words*. Moreover, Morse and Hedlund provided a powerful arithmetic description of Sturmian words (see also [8] for more on Sturmian words).

Proposition 1. [Morse and Hedlund][9] *Associate with a pair* $(\alpha, \beta) \in [0, 1]^2$ *the two infinite words* $\underline{\mathfrak{S}}(\alpha, \beta)$ *and* $\overline{\mathfrak{S}}(\alpha, \beta)$ *whose* n*-th symbols are respectively*

$$\underline{u}_n = \lfloor \alpha(n+1) + \beta \rfloor - \lfloor \alpha n + \beta \rfloor, \qquad \overline{u}_n = \lceil \alpha(n+1) + \beta \rceil - \lceil \alpha n + \beta \rceil.$$

Then a word $u \in \{0, 1\}^{\mathbb{N}}$ *is Sturmian if and only if it equals* $\underline{\mathfrak{S}}(\alpha, \beta)$ *or* $\overline{\mathfrak{S}}(\alpha, \beta)$ *for a pair* (α, β) *formed with an irrational* $\alpha \in]0, 1[$ *and a real* $\beta \in [0, 1[$.

It is also important to study where finite factors occur inside the infinite word u. An infinite word $u \in \mathcal{A}^{\mathbb{N}}$ is *uniformly recurrent* if every factor of u appears infinitely often and with bounded gaps. More precisely, denote by $w_u(q, n)$ the minimal number of symbols u_k with $k \ge q$ which have to be inspected for discovering the whole set $\mathcal{L}_u(n)$ from the index q. Then,the integer $w_u(q, n)$ is a sort of "waiting time". Then u is uniformly recurrent if each set $\{w_u(q, n);\ \ q \in \mathbb{N}\}$ is bounded, and the *recurrence function* $n \mapsto R_u(n)$ is defined as

$$R_u(n) := \max\{w_u(q, n);\, q \in \mathbb{N}\}.$$

We then recover the usual definition: Any factor of length $R_u(n)$ of u contains all the factors of length n of u, and the length $R_u(n)$ is the smallest integer which satisfies this property. The inequality $R_u(n) \ge p_u(n) + n - 1$ thus holds.

Any Sturmian word is uniformly recurrent. Its recurrence function only depends on the slope α and is thus denoted by $n \mapsto R_\alpha(n)$. Moreover, it only depends on α via its *continuants*. We now recall this notion which plays a central role in the paper. Consider the *continued fraction expansion* of the irrational α

$$\alpha = \cfrac{1}{m_1 + \cfrac{1}{\ddots + \cfrac{1}{m_k + \cfrac{1}{\ddots}}}} = [m_1, m_2, \ldots, m_k, \ldots].$$

The positive integers m_k are called the *partial quotients*. The truncated expansion $[m_1, \ldots, m_k]$ at depth k defines a rational, and the *continuant* $q_k(\alpha)$ is the denominator of this rational. The continuant sequence satisfies $q_{-1} = 0, q_0 = 1$ and for any $k \geq 1$ the recurrence $q_k = m_k q_{k-1} + q_{k-2}$ for all k.

The following result due to Morse and Hedlund relates the recurrence function $R_\alpha(n)$ and the sequence $k \mapsto q_k(\alpha)$.

Proposition 2. [Morse and Hedlund] [9] *For any Sturmian word of slope α, the recurrence function $n \mapsto R_\alpha(n)$ is piecewise affine and satisfies*

$$R_\alpha(n) = n - 1 + q_k(\alpha) + q_{k-1}(\alpha), \quad \textit{for any } n \in [\![q_{k-1}(\alpha), q_k(\alpha) - 1]\!].$$

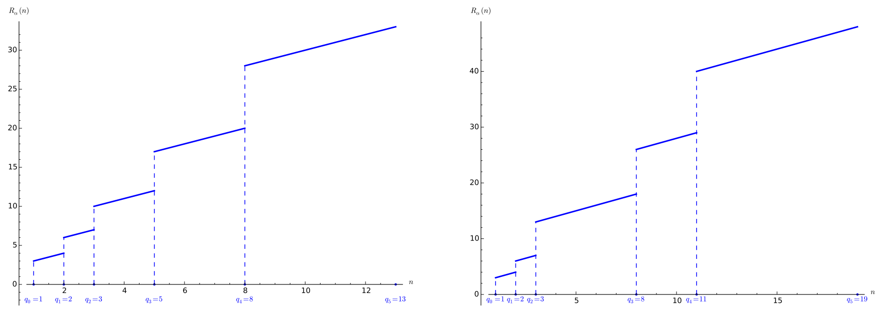

Fig. 1. Two instances of recurrence functions $n \mapsto R_\alpha(n)$ associated with $\alpha = \varphi^2$ (left) and $\alpha = 1/e$ (right), with $\varphi = (\sqrt{5} - 1)/2$ being the inverse of the golden ratio.

It is thus natural to study the quotient $S(\alpha, n) := (R_\alpha(n)+1)/n$. When n belongs to the interval $[\![q_{k-1}(\alpha), q_k(\alpha)-1]\!]$, this quotient depends itself on two quotients: the quotient $x_k(\alpha) := q_{k-1}(\alpha)/q_k(\alpha)$, and the quotient $y_k(\alpha) := n/q_k(\alpha)$, and

$$S(\alpha, n) := \frac{R_\alpha(n) + 1}{n} = 1 + \frac{1 + x_k(\alpha)}{y_k(\alpha)}. \tag{1}$$

As $y_k(\alpha)$ belongs to the interval $[x_k(\alpha), 1]$, the following bounds hold

$$2 + x_k(\alpha) \leq \frac{R_\alpha(n) + 1}{n} \leq 2 + \frac{1}{x_k(\alpha)} \tag{2}$$

(the lower bound holds for n close to $q_k(\alpha)$ whereas the upper bound is attained for $n = q_{k-1}(\alpha)$).

The ratio $x_k(\alpha)$ belongs to $]0, 1]$, and the Borel-Bernstein Theorem (see e.g. [6]) proves that $\liminf_{k\to\infty} x_k(\alpha) = 0$ for almost any irrational α. More precisely:

Proposition 3. (i) *For any irrational real α, one has*

$$\liminf_{n\to\infty} \frac{R_\alpha(n)}{n} \leq 3.$$

(ii) [Morse and Hedlund] [9] *For almost any irrational α, one has*

$$\limsup_{n\to\infty} \frac{R_\alpha(n)}{n \log n} = +\infty, \quad \textit{and } \limsup_{n\to\infty} \frac{R_\alpha(n)}{n(\log n)^{1+\varepsilon}} = 0 \textit{ for } \varepsilon > 0.$$

3 Probabilistic model and main results

The two extreme bounds in Eq.(2) may be very different, notably when $x_k(\alpha)$ is small. We wish to study the behaviour of the ratio $S(\alpha, n)$ when n is any integer in $[\![q_{k-1}(\alpha), q_k(\alpha)-1]\!]$. Eq.(1) shows the role of the quotient n/q_k (called $y_k(\alpha)$ there) and leads to the notion of *position*.

3.1 Position.

We consider a fixed sequence $(\mu_k)_k$ with values in $[0,1[$, and for each $\alpha \in \mathcal{I} := [0,1]$, and each $k \in \mathbb{N}$, we consider the real number at (barycentric) position μ_k inside the interval $[\![q_{k-1}(\alpha), q_k(\alpha)-1]\!]$, namely

$$\widetilde{n}_k^{\langle\mu_k\rangle}(\alpha) := q_{k-1}(\alpha) + \mu_k(q_k(\alpha) - q_{k-1}(\alpha)),$$

together with its integer part (which belongs to $[\![q_{k-1}(\alpha), q_k(\alpha)-1]\!]$),

$$n_k^{\langle\mu_k\rangle}(\alpha) = \lfloor \widetilde{n}_k^{\langle\mu_k\rangle}(\alpha) \rfloor = q_{k-1}(\alpha) + \lfloor \mu_k(q_k(\alpha) - q_{k-1}(\alpha)) \rfloor.$$

The subsequence $(n_k^{\langle\mu_k\rangle}(\alpha))_k$ is *the subsequence associated with the positions* μ_k. We are interested in the subsequence of $n \mapsto S(\alpha, n)$ associated with the subsequence $\{n_k^{\langle\mu_k\rangle}(\alpha),\ k \in \mathbb{N}\}$, and we then let $S_k^{\langle\mu_k\rangle}(\alpha) := S(\alpha, n_k^{\langle\mu_k\rangle}(\alpha))$, namely

$$S_k^{\langle\mu_k\rangle}(\alpha) = 1 + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{n_k^{\langle\mu_k\rangle}(\alpha)} = 1 + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{q_{k-1}(\alpha) + \lfloor \mu_k(q_k(\alpha) - q_{k-1}(\alpha)) \rfloor}. \quad (3)$$

If we drop the integer part in the expression of $S_k^{\langle\mu_k\rangle}$, we deal with the sequence $\widetilde{S}_k^{\langle\mu_k\rangle}(\alpha) := S(\alpha, \widetilde{n}_k^{\langle\mu_k\rangle}(\alpha))$, namely,

$$\widetilde{S}_k^{\langle\mu_k\rangle}(\alpha) = 1 + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{\widetilde{n}_k^{\langle\mu_k\rangle}(\alpha)} = 1 + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{q_{k-1}(\alpha) + \mu_k(q_k(\alpha) - q_{k-1}(\alpha))}, \quad (4)$$

which is expressed with the two sequences $(x_k(\alpha))_k$ and (μ_k) as

$$\widetilde{S}_k^{\langle\mu_k\rangle}(\alpha) = f_{\mu_k}(x_k(\alpha)) \quad \text{with} \quad f_\mu(x) := 1 + \frac{1+x}{x + \mu(1-x)}. \quad (5)$$

The study of the function f_μ provides a precise knowledge on the sequence $\widetilde{S}_k^{\langle\mu_k\rangle}(\alpha)$, that may be "tranfered" to the sequence $S_k^{\langle\mu_k\rangle}(\alpha)$ since the two sequences are "close enough". The following result provides such a first instance of this strategy:

Proposition 4. *Consider a sequence $(\mu_k)_k$ with $\mu_k \in [0,1]$, and let $\alpha \in [0,1] \setminus \mathbb{Q}$.*

(i) Denote by m_k the k-th partial quotient of α. Then, $x_k(\alpha) \le 1/(m_k+1)$ and

$$\widetilde{S}_k^{\langle\mu_k\rangle}(\alpha) \in \left[1 + \frac{m_k+2}{\mu_k m_k + 1}, 3\right] \quad \text{or} \quad \widetilde{S}_k^{\langle\mu_k\rangle}(\alpha) \in \left[3, 1 + \frac{m_k+2}{\mu_k m_k + 1}\right]$$

depending whether $\mu_k \in [1/2, 1]$ or $\mu_k \in [0, 1/2]$.

(*ii*) *The sequence* $S_k^{\langle\mu_k\rangle}(\alpha)$ *is bounded if* α *has bounded partial quotients or if the sequence* (μ_k) *admits a strictly positive lower bound.*

Proof. The map $f_\mu : [0,1] \to \mathbb{R}$ is strictly decreasing when $\mu \in]0,1/2[$, and strictly increasing when $\mu \in]1/2,1[$. This is the constant function equal to 3 when $\mu = 1/2$. For any $a \in]0,1[$, the image $f_\mu([a,1])$ is the interval with endpoints 3 and $f_\mu(a)$. This proves Assertion (i).
With the two inequalities

$$\widetilde{n}_k^{\langle\mu\rangle} \geq n_k^{\langle\mu\rangle} \geq q_{k-1} \geq \varphi^{1-k}, \qquad 0 \leq \widetilde{n}_k^{\langle\mu\rangle} - n_k^{\langle\mu\rangle} \leq 1,$$

we obtain the inequality

$$0 \leq S_k^{\langle\mu\rangle} - \widetilde{S}_k^{\langle\mu\rangle} = \frac{q_k + q_{k-1}}{n_k^{\langle\mu\rangle} \cdot \widetilde{n}_k^{\langle\mu\rangle}} (\widetilde{n}_k^{\langle\mu\rangle} - n_k^{\langle\mu\rangle}) \leq \frac{1}{q_{k-1}} \frac{q_k + q_{k-1}}{\widetilde{n}_k^{\langle\mu\rangle}} \leq \varphi^{k-1} \widetilde{S}_k^{\langle\mu\rangle}, \quad (6)$$

and we apply (i).

3.2 Probabilistic model

Let us describe now our *probabilistic model.* We choose a sequence $(\mu_k)_k$ of positions that will be *fixed.* This defines, for each real α, a sequence of indices $n_k := n_k^{\langle\mu_k\rangle}$, and then a sequence of real numbers $k \mapsto S_k^{\langle\mu_k\rangle}(\alpha)$. When the real α is random, and uniformly drawn in the unit interval $\mathcal{I} = [0,1]$, the sequence $k \mapsto S_k^{\langle\mu_k\rangle}$ becomes a sequence of random variables, and we study the mean value and the distribution of the sequence $k \mapsto S_k^{\langle\mu_k\rangle}$ for $k \to \infty$.
For any position, the index $n_k^{\langle\mu_k\rangle}$ belongs to the interval $[\![q_{k-1}, q_k - 1]\!]$. Then, as the expectations for $\alpha \in [0,1]$ of the two extreme sequences $k \mapsto \log q_{k-1}(\alpha)$, $k \mapsto \log q_k(\alpha)$ satisfy the same estimates (see [7]), it is also the case for the expectation for $\alpha \in [0,1]$ of the sequence $k \mapsto \log n_k^{\langle\mu_k\rangle}(\alpha)$. It thus satisfies

$$\mathbb{E}[\log n_k^{\langle\mu_k\rangle}] = \frac{\pi^2}{12\log 2} k + O(1), \quad (7)$$

and it is of linear growth with respect to k.

3.3 Results for a constant position μ

We first consider the case where the sequence $(\mu_k)_k$ is a constant sequence that takes a fixed value μ, and we study the expectation and the distribution of the sequence $k \mapsto S_k^{\langle\mu\rangle}$ of random variables, when $k \to \infty$, as a function of the position μ. Theorem 1 below shows that there are two main cases:

(*a*) the case when $\mu = 0$; here, the expectations are infinite, but the functions $k \mapsto S_k^{\langle 0\rangle}$ admit a limit density;
(*b*) the case when $\mu \neq 0$; here, both the expectations and the densities have a finite limit; the case $\mu = 1/2$ is particular, as the limit density is a Dirac measure, concentrated at the value 3.

For indices n associated with parameters μ satisfying $\mu \geq \mu_0 > 0$, we exhibit a behaviour for the sequence $n \mapsto R_\alpha(n)$ which is thus "linear on average"; the "$\log n$" behaviour of Proposition 3 does not occur in this case.

Theorem 1. [Fixed position μ] *Let $\varphi = (\sqrt{5}-1)/2 < 1$. The following holds for the random variables $S_k^{\langle\mu\rangle}$.*

(i) [Expectations] *For each $\mu \in]0,1]$, their expected values $\mathbb{E}[S_k^{\langle\mu\rangle}]$ satisfy*

$$\mathbb{E}[S_k^{\langle\mu\rangle}] = 1 + \frac{1}{\log 2}\frac{|\log\mu|}{1-\mu} + O\left(\frac{\varphi^{2k}}{\mu}\right) + O\left(\varphi^k \frac{|\log\mu|}{1-\mu}\right), \tag{8}$$

with the constants in the O–term being uniform with respect to μ and k.

(ii) [Limit density] *For each $\mu \in [0,1]$ with $\mu \neq 1/2$, they admit a limit density $s^{\langle\mu\rangle}$ equal to*

$$s^{\langle\mu\rangle}(x) = \frac{1}{\log 2}\left(\frac{1}{(x-1)\,|x(1-\mu)+\mu-2|}\right)\mathbf{1}_{\mathcal{I}_\mu}(x), \tag{9}$$

where $\mathcal{I}_\mu$ is the real interval with endpoints 3 and $1+1/\mu$.
More precisely, for any $b \in I_\mu$, one has

$$\mathbb{P}\left[S_k^{\langle\mu\rangle} \leq b\right] = \int_0^b s^{\langle\mu\rangle}(x)\ dx + \frac{1}{b}O\left(\varphi^k\right),$$

where the constant of the O–term is uniform with respect to b and k. It is also uniform with respect to μ when μ satisfies $|\mu - 1/2| \geq \mu_0$ for any $\mu_0 > 0$.

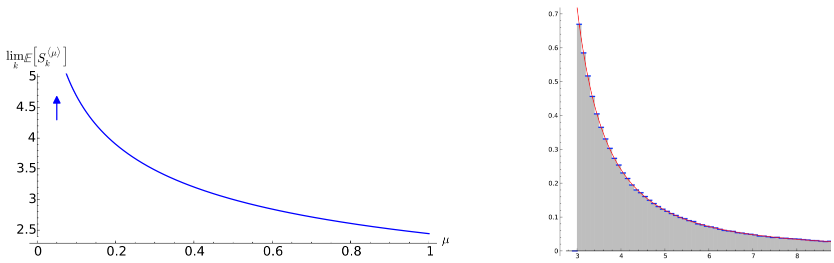

Fig. 2. On the left, the graph of $\lim_{k\to\infty}\mathbb{E}[S_k^{\langle\mu\rangle}]$ as a function of μ. On the right, the graph of the density $s^{\langle 0\rangle}$.

3.4 Results when the sequence $\mu_k \to 0$

We now focus on the difficult case, when the sequence $(\mu_k)_k$ is no longer constant, and we consider a sequence $(\mu_k)_k$ of positions which tends to 0. We first consider in Theorem 2 below sequences $(\mu_k)_k$ which tend exponentially fast to 0, and we observe that the expectations are of order k. We then consider general sequences $(\mu_k)_k$ which tend to 0, and we show that the associated random variables admit a limit density, with a speed of convergence which depends on the sequence $(\mu_k)_k$.

Theorem 2. [Sequence $\mu_k \to 0$] *The following holds for the random variables $S_k^{\langle \mu \rangle}$ associated with a sequence $\mu_k \to 0$.*

(i) [Expectations] *Consider the sequence $\mu_k = \tau^k$, with $\tau \in [\varphi^2, 1[$. Then*

$$\mathbb{E}[S_k^{\langle \tau^k \rangle}] = k \frac{|\log \tau|}{\log 2} + O(1), \tag{10}$$

where the constant hidden in the O–term is uniform with respect to τ and k. For any α, and for each $\tau \in [\varphi^2, 1[$, there exists an increasing subsequence $\mathcal{N}(\alpha, \tau)$ of indices n for which

$$\mathbb{E}\left[\frac{R_\alpha(n)}{n} - \frac{12|\log \tau|}{\pi^2} \log n\right] = O(1) \qquad (n \to \infty). \tag{11}$$

For any $\tau < 1$, if μ_k is drawn uniformly in $[0, 1]$, the conditional expectation with respect to the event $[\mu_k \geq \tau^k]$ satisfies

$$\lim_{k\to\infty} \mathbb{E}\left[S_k^{\langle \mu_k \rangle} \Big| [\mu_k \geq \tau^k]\right] = 1 + \frac{\pi^2}{6 \log 2}.$$

(ii) [Limit density] *For any sequence $\mu_k \to 0$, the random variables $S_k^{\langle \mu_k \rangle}$ admit as limit density the density $s^{\langle 0 \rangle}$ equal to*

$$s^{\langle 0 \rangle}(x) = \frac{1}{\log 2} \frac{1}{(x-1)(x-2)} \mathbf{1}_{[3,\infty]}(x).$$

More precisely, for any $b \geq 3$, the probability $\mathbb{P}[S_k^{\langle \mu_k \rangle} \geq b]$ satisfies

$$\mathbb{P}[S_k^{\langle \mu_k \rangle} \geq b] = \frac{1}{\log 2} \log\left(\frac{b-1}{b-2}\right) + O(\mu_k) + \frac{1}{b} O\left(\varphi^k\right),$$

where the constants hidden in the O–term are uniform to respect to b and k. If now the sequences $(b_k)_k$ and $(\mu_k)_k$ satisfy the following three conditions ($b_k \to \infty$, $\mu_k \to 0$ with $b_k \mu_k \to 0$), then

$$\lim_{k\to\infty} b_k \cdot \mathbb{P}\left[S_k^{\langle \mu_k \rangle} \geq b_k\right] = \frac{1}{\log 2}.$$

Remark 1. The estimate (7) together with (10) yields (11). We have then exhibited a $\log n$ behaviour "on average" for the ratio $R_\alpha(n)/n$ for (an infinity of) particular subsequences n (which depend on α). On the contrary, when the position is not too small, the ratio $R_\alpha(n)/n$ remains bounded (on average).

4 Strategy for the proofs.

We begin with Theorem 1 which deals with a fixed position μ. There are three main steps in the proof of Theorem 1.

(i) We drop the integer part in the expression of $S_k^{\langle\mu\rangle}$ and deal with the sequence $\widetilde{S}_k^{\langle\mu\rangle}(\alpha)$ that can be written as $f_\mu(x_k(\alpha))$ (see (5))). This is an instance of a smooth sequence (as defined in Section 4.1 below). We express its mean value and its distribution with the k-th iterate of the Perron Frobenius operator $\mathbf{H}$.

(ii) With the spectral properties of the operator $\mathbf{H}$ (described in Section 4.2), when acting on the Banach space $BV(\mathcal{I})$ of the functions of bounded variation on the unit interval $\mathcal{I}$, we obtain the asymptotics of the expectations and the expression of the limit distribution, always for the sequence $\widetilde{S}_k^{\langle\mu\rangle}$.

(iii) We return to the initial sequence $S_k^{\langle\mu\rangle}$ with the following estimates

$$\mathbb{E}[S_k^{\langle\mu\rangle}] = \mathbb{E}[\widetilde{S}_k^{\langle\mu\rangle}]\left(1+O(\varphi^k)\right), \quad \mathbb{P}[S_k^{\langle\mu\rangle} \le b] - \mathbb{P}[\widetilde{S}_k^{\langle\mu\rangle} \le b] = O\left(\frac{\varphi^k}{b}\right), \quad (12)$$

which are refinements of Eq.(6) and will be proven in an extended version.

Since the probabilistic estimates obtained in Theorem 1 are uniform with respect to μ and k, we may extend them to the case where μ depends on k, and we may study the interesting case where the sequence $(\mu_k)_k$ tends to 0 for $k \to \infty$. We then obtain the results of Theorem 2.

Remark 2. There are two error terms in the asymptotic estimates (8) of the expectations. The first one comes from the spectral gap of the Perron-Frobenius operator and the second one arises when one takes into account integer parts in the definition of $S_k^{\langle\mu\rangle}$.

4.1 Smooth sequences

The sequence $\widetilde{S}_k^{\langle\mu\rangle}$ provides an instance of a smooth sequence, defined as follows:

Definition 1. *A sequence of random variables (T_k) defined on the unit interval $\mathcal{I} = [0,1]$ is a* smooth sequence *if there exists a function $f \in BV(\mathcal{I})$ for which*

$$T_k(\alpha) = f(x_k(\alpha)) \qquad \textit{with} \quad x_k(\alpha) = \frac{q_{k-1}(\alpha)}{q_k(\alpha)} \qquad \textit{for all } \alpha \in \mathcal{I}.$$

Here, we deal with the function f_μ defined in (5), whose inverse map g_μ is

$$g_\mu : f_\mu(\mathcal{I}) \mapsto [0,1], \quad g_\mu(x) = \frac{-1-\mu+\mu x}{2-\mu-x(1-\mu)}.$$

For $\mu \in]0,1[$, the function f_μ is integrable on $\mathcal{I}$, and its L^1–norm satisfies

$$\|f_\mu\|_{L^1} = 1 + \frac{1}{1-\mu} + \frac{1-2\mu}{(1-\mu)^2}|\log\mu|, \qquad \|f_1\|_{L^1} = 5/2.$$

Moreover, always for $\mu \in]0,1[$, the function f_μ is monotonic and thus of bounded variation, with a total variation equal to $(1/\mu)|1-2\mu|$, hence $\|f_\mu\|_{BV} = O(1/\mu)$. Remark that f_0 does not belong to $BV(\mathcal{I})$.

We now recall some basic facts on the underlying dynamical system, together with the Perron-Frobenius operator, that will be useful in the sequel.

4.2 The dynamical system and the Perron-Frobenius operator

The underlying dynamical system. We consider the dynamical system $(\mathcal{I}, V)$ associated with the unit interval $\mathcal{I}$ and the Gauss map V, defined by

$$V(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor = \left\{ \frac{1}{x} \right\} \quad \text{for } x \neq 0, \qquad V(0) = 0.$$

The map V builds the continued fraction expansion of α, via the function $m(\alpha) := \lfloor 1/\alpha \rfloor$, as

$$\alpha = [m_1, m_2, \ldots, m_k, \ldots] \qquad \text{with} \quad m_{k+1}(\alpha) = m(V^k(\alpha)) \text{ for all } k \geq 0.$$

The inverse branches of V belong to the set

$$\mathcal{H} := \left\{ h_m : x \mapsto \frac{1}{m+x}; \quad m \geq 1 \right\},$$

and the inverse branches of V^k belong to the set

$$\mathcal{H}^k = \{ h_{m_1} \circ h_{m_2} \circ \ldots \circ h_{m_k} \quad : \quad m_1, \ldots, m_k \geq 1 \}\ .$$

For a k-uple $\boldsymbol{m} = (m_1, m_2, \ldots, m_k)$, let $h_{\boldsymbol{m}} := h_{m_1} \circ h_{m_2} \circ \ldots \circ h_{m_k}$. The linear fractional transformation $h_{\boldsymbol{m}}$ is expressed with two sequences of continuants $(p_k)_k, (q_k)_k$ under the form

$$h_{\boldsymbol{m}}(x) = h_{m_1} \circ h_{m_2} \circ \ldots \circ h_{m_k}(x) = \cfrac{1}{m_1 + \cfrac{1}{\ddots + \cfrac{1}{m_k + x}}} = \frac{p_{k-1}\, x + p_k}{q_{k-1}\, x + q_k}\,.$$

Remark that the continuants q_k, p_k which are just defined only depend on the k-uple $\boldsymbol{m} = (m_1, m_2, \ldots, m_k)$. However, there is no conflict with our previous definition of the sequence $q_k(\alpha)$ given in Section 2, since, for any α which belongs to the interval $h_{\boldsymbol{m}}(\mathcal{I})$, the equality $q_k(\alpha) = q_k(\boldsymbol{m})$ holds.

The mirror property (described for instance in [1]) relates the coefficients of $h = h_{m_1} \circ h_{m_2} \circ \ldots \circ h_{m_k}$ and those of its mirror $\widehat{h} := h_{m_k} \circ h_{m_{k-1}} \circ \ldots \circ h_{m_1}$:

$$h(y) = \frac{p_{k-1}y + p_k}{q_{k-1}y + q_k} \qquad \Longrightarrow \qquad \widehat{h}(y) = \frac{p_{k-1}y + q_{k-1}}{p_k y + q_k}.$$

Perron-Frobenius operator. When the unit interval is endowed with a density f, after one iteration of V, it is endowed with the density

$$\mathbf{H}[f](x) := \sum_{h \in \mathcal{H}} |h'(x)| \cdot f \circ h(x),$$

and after k iterations of V, with the density

$$\mathbf{H}^k[f](x) = \sum_{h \in \mathcal{H}^k} |h'(x)| \cdot f \circ h(x).$$

The operator $\mathbf{H}$ is called the Perron Frobenius operator.
Now, at $x = 0$, the two maps h and $\widehat{h}$ satisfy $|h'(0)| = |\widehat{h}'(0)| = 1/q_k^2$, and the equality $q_{k-1}/q_k = \widehat{h}(0)$ holds. With this remark, the k-th iterate $\mathbf{H}^k$ generates the continuants q_k,

$$\mathbf{H}^k[f](0) = \sum_{h \in \mathcal{H}^k} \frac{1}{q_k^2} f\left(\frac{p_k}{q_k}\right) = \sum_{h \in \mathcal{H}^k} \frac{1}{q_k^2} f\left(\frac{q_{k-1}}{q_k}\right). \tag{13}$$

We now summarize some classical spectral properties of the operator $\mathbf{H}$ (see e.g. [6] or [2]). When acting on the Banach space $BV(\mathcal{I})$ of functions of bounded variation, the operator $\mathbf{H}$ admits a unique dominant eigenvalue $\lambda = 1$, with an eigenfunction proportional to $\psi(x) = 1/(1+x)$, and it has a subdominant spectral radius equal to φ^2. Moreover, the adjoint $\mathbf{H}^*$ has an eigenmeasure proportional to the Lebesgue measure. Then, for any $g \in BV(\mathcal{I})$, the iterate $\mathbf{H}^k[g]$ decomposes as

$$\mathbf{H}^k[g](x) = \frac{1}{\log 2} \frac{1}{1+x} \cdot \int_{\mathcal{I}} g(x) dx + O(\varphi^{2k}) \|g\|_{BV}. \tag{14}$$

4.3 Smooth random variables and Perron-Frobenius operator

We now perform Step (i) in the proof of Theorem 1. The following lemma (inspired by [5]) expresses the expectation and distribution of smooth sequences in terms of the Perron-Frobenius operator $\mathbf{H}$.

Lemma 1. *Assume that (T_k) is a smooth sequence associated with the function f. Then, the expected value $\mathbb{E}[T_k]$ and the distribution of the random variable (T_k) are both expressed with the k-th iterate of the Perron-Frobenius operator $\mathbf{H}$:*

$$\mathbb{E}[T_k] = \mathbf{H}^k \left[f(x) \cdot \frac{1}{1+x} \right](0), \quad \mathbb{P}[T_k \in J] = \mathbf{H}^k \left[\mathbf{1}_J \circ f(x) \cdot \frac{1}{1+x} \right](0),$$

where J is a subinterval of $\mathcal{I}$.

Proof. For each index k, consider the family of linear fractional transformations $h \in \mathcal{H}^k$. The intervals $h(\mathcal{I})$ form a partition of the interval $\mathcal{I}$, and the length of the interval $h(\mathcal{I})$ is expressed as a function of the continuants q_k, as

$$|h(\mathcal{I})| = \frac{1}{q_k(q_k + q_{k-1})} = \frac{1}{q_k^2} \left(\frac{1}{1 + \dfrac{q_{k-1}}{q_k}} \right).$$

Moreover $T_k(\alpha)$ is constant on the interval $h(\mathcal{I})$, and equal to $f(q_{k-1}/q_k)$. Finally

$$\mathbb{E}[T_k] := \int_{\mathcal{I}} T_k(\alpha) \mathrm{d}\alpha = \sum_{h \in \mathcal{H}^k} \frac{1}{q_k^2} \ell\left(\frac{q_{k-1}}{q_k}\right) \qquad \text{with} \quad \ell(x) = \frac{1}{1+x} f(x).$$

With Relation (13), the last expression is exactly $\mathbf{H}^k[\ell](0)$.
We now consider, for any $J \subset \mathbb{R}$, the probability $\mathbb{P}[T_k \in J] = \mathbb{E}[\mathbf{1}_J \circ T_k]$. Using the same transforms as above (now applied to the function $\mathbf{1}_J \circ f(x)$) yields

$$\mathbb{P}[T_k \in J] = \mathbf{H}^k \left[\mathbf{1}_J \circ f(x) \cdot \frac{1}{1+x} \right] (0).$$

4.4 Asymptotic study of smooth variables

We now perform Step (ii) in the proof of Theorem 1. Since the probabilistic characteristics of the random variable T_k are expressed with the k-th iterate of the Perron Frobenius operator $\mathbf{H}$, their asymptotics will be related to the dominant spectral properties of this operator when it acts on the Banach space $BV(\mathcal{I})$ of the functions of bounded variation on the unit interval, and we use the decomposition (14).

Lemma 2. *The following asymptotics hold, for any smooth sequence (T_k) relative to a function $f \in BV(\mathcal{I})$:*

$$\mathbb{E}[T_k] = \frac{1}{\log 2} \int_{\mathcal{I}} f(x) \cdot \frac{1}{1+x} dx + O(\varphi^{2k} \|f\|_{BV}),$$

$$\mathbb{P}[T_k \in J] = \frac{1}{\log 2} \int_{\mathcal{I}} \mathbf{1}_J \circ f(x) \cdot \frac{1}{1+x} dx + O(\varphi^{2k}),$$

where J is a subinterval of $\mathcal{I}$. If moreover the function f is of class $\mathcal{C}^1$ and monotonic, with an inverse function g, the random variable T_k admits a limit density; for any interval $[a,b] \subset f(\mathcal{I})$, one has

$$\mathbb{P}[T_k \in [a,b]] = \frac{1}{\log 2} \int_a^b \frac{|g'(u)|}{1+g(u)} du + O(\varphi^{2k}) = \frac{1}{\log 2} \left| \log \frac{1+g(a)}{1+g(b)} \right| + O(\varphi^{2k}).$$

Proof. This is just an easy application of the decomposition (14). For the distribution, the norm $\|\mathbf{1}_J \circ f \cdot \psi\|_{BV}$ admits an upper bound which neither depend on the function f nor on the interval J.

The previous lemma entails the following asymptotics for the probabilistic characteristics of the sequence $\widetilde{S}_k^{\langle\mu\rangle}$. Recall that the density $s^{\langle\mu\rangle}$ is defined in (9).

Lemma 3. *For $\mu \in]0,1]$, the two following asymptotic estimates, namely*

$$\mathbb{E}[\widetilde{S}_k^{\langle\mu\rangle}] = 1 + \frac{1}{\log 2} \frac{|\log \mu|}{1-\mu} + O\left(\frac{\varphi^{2k}}{\mu}\right), \quad \mathbb{P}\left[\widetilde{S}_k^{\langle\mu\rangle} \in J\right] = \int_J s^{\langle\mu\rangle}(x) dx + O(\varphi^{2k}).$$

The second estimate also holds for $\mu = 0$.

Proof. This is just the application of the previous lemma for $f := f_\mu$. The function f_μ belongs to $BV(\mathcal{I})$ for $\mu > 0$, with norm $\|f_\mu\|_{BV} = O(1/\mu)$.

This ends Step (ii) of the proof of Theorem 1. The estimates (12) needed in Step (iii) will be proven in the extended version, together with some hints for Theorem 2.

5 Conclusion

With a Sturmian word of slope α, and a sequence (μ_k), we have associated a sequence of indices n_k (lengths of factors) defined by their barycentric position μ_k inside $[\![q_{k-1}(\alpha), q_k(\alpha) - 1]\!]$. We then have elucidated the role played by the position in the behaviour of the recurrence of a random Sturmian word.
We plan to extend our probabilistic study in three directions, and consider three probabilistic models. The first two models were already dealt with in [4] for the study of Kronecker sequences, and the last one has been considered in [11].

Reals with bounded partial quotients. This type of slope α gives rise to Sturmian words whose recurrence function is proven to be linear (see Lemma 4). For a bound M, we restrict α to the set $\mathcal{R}^{[M]}$ of numbers whose partial quotients are at most M, endowed with the Hausdorff measure, and we wish to observe the transition when $M \to \infty$.

Rational numbers. This type of slope α gives rise to periodic words, and occurs for Christoffel words. For a bound N, we restrict α to the set $\mathcal{Q}_{[N]}$ of rationals with denominator at most N, endowed with the uniform distribution, and we wish to observe the transition when $N \to \infty$. This will explain how a periodic word "becomes" Sturmian.

Quadratic irrationals. This type of slope α occurs for substitutive Sturmian words. There is a natural notion of size associated with such numbers α, closely related to the period of their continued fraction expansion, and we wish to observe the transition when the size tends to ∞.

References

1. B. Adamczewski and J.-P. Allouche, *Reversals and palindromes in continued fractions*, heoret. Comput. Sci. **380** (2007), 220–237.
2. J. Bourdon, B. Daireaux and B. Vallée, *Dynamical analysis of α-Euclidean Algorithms*, Journal of Algorithms **44** (2002), 246–285.
3. J. Cassaigne, *Limit values of the recurrent quotient of Sturmian sequences*, Theoret. Comput. Sci. **218** (1999), 3–12.
4. E. Cesaratto and B. Vallée, *Pseudo-randomness of a random Kronecker sequence. An instance of dynamical analysis*, to appear as a chapter in the book *Combinatorics, Words and Symbolic Dynamics*, (ed : Valérie Berthé and Michel Rigo).
5. P. Flajolet and B. Vallée, *Continued fraction algorithms, functional operators, and structure constants*, Theoret. Comput. Sci. **94** (1998), 1–34.
6. M. Iosifescu and C. Kraaikamp, *Metrical Theory of Continued Fractions*, Collection Mathematics and Its Applications, Vol 547, Kluwer Academic Press, 2002.
7. P. Lévy, *Sur le développement en fraction continue d'un nombre choisi au hasard*, Compositio Mathematica **3** (1936), 286–303.
8. Lothaire, M., *Algebraic Combinatorics on Words*, vol. 90, Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 2002.
9. M. Morse and G. Hedlund, *Symbolics dynamics II. Sturmian trajectories*, American Journal of Mathematics **62** (1940), 1–42.
10. B. Vallée, *Dynamique des fractions continues à contraintes périodiques*, Journal of Number Theory **72** (1998), no. 2, 183–235.

11. B. Vallée, *Dynamique des fractions continues à contraintes périodiques*, Journal of Number Theory **72** (1998), no. 2, 183–235.