



Balances for fixed points of primitive substitutions

Boris Adamczewski

Institut de Mathématique de Luminy, CNRS UPR 9016, 163, Avenue de Luminy, Case 907, Marseille Cedex 09 F-13288, France

Abstract

An infinite word defined over a finite alphabet \mathcal{A} is balanced if for any pair (ω, ω') of factors of the same length and for any letter a in the alphabet

$$||\omega|_a - |\omega'|_a| \leq 1,$$

where $|\omega|_a$ denotes the number of occurrences of the letter a in the word ω . In this paper, we generalize this notion and introduce a measure of balance for an infinite sequence. In the case of fixed points of primitive substitutions, we show that the asymptotic behaviour of this measure is in part ruled by the spectrum of the incidence matrix associated with the substitution. Connections with frequencies of letters and other balance properties are also discussed.

© 2003 Published by Elsevier B.V.

Keywords: Combinatorics on words; Balance; Primitive substitution; Frequency; Infinite words

1. Introduction

A word defined over a finite alphabet \mathcal{A} is balanced if for any pair (ω, ω') of factors of the same length, we have $||\omega|_a - |\omega'|_a| \leq 1$, for each letter a . Here $|\omega|_a$ denotes the number of occurrences of the letter a in the word ω . This notion appeared for the first time in the papers of Morse and Hedlund [28,29] published in 1938 and 1940. They proved in particular that each letter of an infinite balanced word defined over a binary alphabet admits a frequency. Later, Coven and Hedlund [12] obtained that an infinite word defined over a binary alphabet is Sturmian if and only if it is aperiodic and balanced. More recently, Hubert [25] (see also a close result of Graham [23]) shows

E-mail address: adamczew@iml.univ-mrs.fr (B. Adamczewski).

that aperiodic balanced words defined over a finite alphabet of cardinality at least three are strongly connected with Sturmian words and could also be simply characterized in a geometrical way.

Another important aspect of the balanced words theory comes from Frankel's conjecture, which states that there is only one balanced word, up to a permutation of the letters, defined over a finite alphabet of cardinality m at least three, whose letters all have distinct frequencies. A proof for $m=3,4,5,6$ is given in [27,5,38,39]. In theoretical computer science, this question is closely related to optimization and job-shop problems (see for instance [22]).

All these results emphasize the fact that the balance property is a really strong constraint. This is perhaps one of the reasons for which mathematicians have tried to generalize the notion of balance, sometimes in order to make it more flexible. For instance, Berthé and Tijdeman [7] study balance properties of multi-dimensional words; in particular balanced multi-dimensional words are proved to be fully periodic.

Another possible generalization, initiated in [21] for Sturmian words, consists in estimating the difference between the number of occurrences of a word u in any pair of factors of the same length of an infinite word. We will consider such differences for fixed points of primitive substitutions in Section 5.

In a more ergodic or number theoretic point of view, a fruitful approach seems to be the extension of the notion of balance to C -balance by requiring $|\omega|_a - |\omega'|_a| \leq C$, for each letter a and for some constant C . Then, the C -balance property is closely related to the notion of bounded remainder sets (see Section 3). It was believed that an Arnoux–Rauzy sequence (see [6]) would be a natural coding of a rotation on the two-dimensional torus. Using the connection evoked above, Cassaigne et al. [9] provide a counterexample to this conjecture by proving an Arnoux–Rauzy sequence to be totally unbalanced (that is to say, not C -balanced for any C). In the same spirit, symbolic dynamical systems arising from a primitive Pisot-type substitution (with some additional technical conditions) are expected to be measure theoretically isomorphic with a minimal rotation on a torus (see for instance [32,8]).

Our motivations for this paper partly come from these results. But just as the study of the complexity is not reduced to the one of periodic, Sturmian or Epiturmian words, we want not to restrict our study to balanced or C -balanced words. We thus introduce a balance function associated with an infinite word, including in this way the notion of C -balance. In the case of fixed points of primitive substitutions, we show that the asymptotic behaviour of the balance function is in part ruled by the incidence matrix associated with the substitution. Therefore, our work is more quantitative than qualitative and thus shares the same spirit as the one done in [30], concerning the study of the complexity function.

The organization of the article is as follows. Section 2 contains most of the different definitions and notations used in the paper. We recall as well a previous result of [1] which will be fundamental for the study we propose. In Section 3 we will investigate connections between frequencies of letters and balance. The main results concerning the balance function are presented and discussed in Section 4. Section 5 deals with generalized balance properties. We will illustrate our study through some examples in Section 6. Finally, Appendix A is devoted to the presentation of a generalized

numeration system used to prove our results, and to obtain an explicit bound for the balance function of fixed points of Pisot-type substitutions.

2. Definitions and background

Symbolic sequences: A finite and non-empty set \mathcal{A} is called alphabet. The elements of \mathcal{A} are called letters. A finite word on \mathcal{A} is a finite sequence of letters and an infinite word on \mathcal{A} is a sequence of letters indexed by \mathbb{N} . The length of a finite word ω , denoted by $|\omega|$, is the number of letters it is built from. The empty word, ε , is the unique word of length 0. We denote by \mathcal{A}^* the set of finite words on \mathcal{A} and by $\mathcal{A}^{\mathbb{N}}$ the set of sequences on \mathcal{A} .

Let $U = (u_k)_{k \in \mathbb{N}}$ be a symbolic sequence defined over the alphabet \mathcal{A} . A factor of U is a finite word of the form $u_i u_{i+1} \dots u_j$, $0 \leq i \leq j$. If ω is a factor of U and a a letter, then $|\omega|_a$ is the number of occurrences of the letter a in the word ω . The *frequency* of the letter a in U is defined by

$$\lim_{N \rightarrow \infty} \frac{|u_0 u_1 \dots u_{N-1}|_a}{N},$$

when this limit exists.

We denote by $\mathcal{L}_n(U)$ the set of all the factors of length n of the sequence U and by $\mathcal{L}(U)$ the set of all the factors of the sequence U . The set $\mathcal{L}(U)$ is called the language of U . We call complexity function of U , and denote by $P_U(n)$, the function which with each positive integer n associates $\text{Card } \mathcal{L}_n(U)$. A sequence in which all the factors have an infinite number of occurrences is called *recurrent*. When these occurrences have bounded gaps, the sequence is called *uniformly recurrent*. A sequence is said *linearly recurrent* (with constant K) if there exists an integer K such that for any of its factors ω the difference between two successive occurrences is bounded by $K|\omega|$. We will also use LR instead of linearly recurrent in the following.

Substitutions and spectrum: Endowed with concatenation, the set \mathcal{A}^* is a free monoid with unit element ε . A map from \mathcal{A} to $\mathcal{A}^* \setminus \{\varepsilon\}$ can be extended by concatenation to an endomorphism of the free monoid \mathcal{A}^* and then to a map from $\mathcal{A}^{\mathbb{N}}$ to itself. A *substitution* σ on the alphabet \mathcal{A} is such a morphism satisfying

- (i) There exists $a \in \mathcal{A}$ such that a is the first letter of $\sigma(a)$;
- (ii) For all $b \in \mathcal{A}$, $\lim_{n \rightarrow +\infty} |\sigma^n(b)| = +\infty$.

Then, it is easily seen that $(\sigma^n(a))_{n \in \mathbb{N}}$ converges in $\mathcal{A}^{\mathbb{N}}$, endowed with the product of the discrete topologies on \mathcal{A} , to a sequence U . This sequence is a *fixed point* of σ , i.e., $\sigma(U) = U$. More generally, a sequence which is the image by a morphism of a fixed point of a substitution is said *substitutive*.

Given a substitution σ defined on $\mathcal{A} = \{1, 2, \dots, d\}$, we call the matrix $M_\sigma = (|\sigma(j)|_i)_{(i,j) \in \mathcal{A}^2}$ the *incidence matrix* associated with σ . The composition of substitutions corresponds to the multiplication of incidence matrices. A substitution is called *primitive* if there exists a power of its incidence matrix for which all the entries are positive. For a primitive substitution, the Perron–Frobenius theorem implies that its

incidence matrix admits a real eigenvalue greater than one and which is greater than the modulus of all the other eigenvalues (see for instance [31]). This eigenvalue is called the *Perron* eigenvalue of the substitution. In the following, we need to order the spectrum S_{M_σ} of the incidence matrix M_σ associated with a primitive substitution σ . We thus write

$$S_{M_\sigma} = \{\theta_i, 2 \leq i \leq d'\} \cup \{\theta_1 = \theta\},$$

where θ is the Perron eigenvalue of σ , d' is the number of distinct eigenvalues and

$$\forall 2 \leq i, k \leq d', \quad i < k \Rightarrow \begin{cases} |\theta_i| > |\theta_k|, \\ \text{or } |\theta_i| = |\theta_k| \text{ and } \alpha_i \geq \alpha_k, \end{cases} \quad (1)$$

where $\alpha_j + 1$ means the multiplicity of the eigenvalue θ_j in the minimal polynomial of M_σ . Furthermore, if $|\theta_i| = |\theta_k| = 1$, $\alpha_i = \alpha_k$, θ_i is not a root of unity and θ_k is a root of unity, then $i < k$.

Remark 1. In the case where two distinct eigenvalues have the same modulus and the same multiplicity in the minimal polynomial of M_σ , this way of ordering is not always well-defined. We obtain that one can give several orders satisfying our conditions. This is, in fact, not a problem because our results do not depend on the choice of such an order. Then in the following, when we will use this notation, one should understand that we have made an arbitrary choice for the corresponding order.

The balance function: We introduce here some definitions about balance properties for an infinite word.

Definition 2. Let us consider an alphabet \mathcal{A} , ω an infinite word in $\mathcal{A}^{\mathbb{N}}$, and an integer C . The word ω is said C -balanced if:

$$\forall i \in \mathcal{A}, \quad \forall (v, w) \in \mathcal{L}(\omega), \quad |v| = |w| \Rightarrow -C \leq |v|_i - |w|_i \leq C.$$

If ω is 1-balanced, we just say that ω is balanced.

In [7], Berthé and Tijdeman introduce a measure of balance for multi-dimensional words. We use in [2] a one-dimensional analogous which is defined by:

Definition 3. Let U be an infinite sequence defined over an alphabet \mathcal{A} . We define the balance function of U in the following way:

$$B_U(n) = \max_{a \in \mathcal{A}} \max_{w, w' \in \mathcal{L}_n(U)} \{||w|_a - |w'|_a|\}.$$

We thus obtain that a sequence is C -balanced if and only if its balance function is bounded by C .

The discrepancy function: Let \mathcal{A} be a finite set. Endowed with the discrete topology, \mathcal{A} is a compact set. Let us consider a probability measure μ on \mathcal{A} . A sequence

$U = (u_n)_{n \in \mathbb{N}}$ which takes its values in \mathcal{A} is said uniformly distributed with respect to the measure μ if

$$\forall a \in \mathcal{A}, \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \chi_{\{a\}}(u_k) = \mu(a),$$

where $\chi_{\{a\}}$ denotes the characteristic function of the set $\{a\}$. If U is a symbolic sequence defined over the alphabet $\mathcal{A} = \{1, 2, \dots, d\}$, whose letters have frequencies, then, we can introduce the vector of frequencies $A = (A_i)_{i=1,2,\dots,d}$ which defines a natural probability measure on \mathcal{A} for U . The existence of frequencies implies that U is uniformly distributed with respect to this probability measure. For a fixed point of a primitive substitution such a natural measure always exists (see for instance [31]).

We define the discrepancy function $D_N(\mu, U)$ of the sequence U with respect to the measure μ by

$$D_N(\mu, U) = \max_{a \in \mathcal{A}} \left| \sum_{k=0}^{N-1} (\chi_{\{a\}}(u_k) - \mu(a)) \right|.$$

Then, if U is a symbolic sequence defined over the alphabet $\mathcal{A} = \{1, 2, \dots, d\}$ whose letters have frequencies, and if $A = (A_i)_{i=1,2,\dots,d}$ denotes the associated probability vector, the discrepancy function $D_N(A, U)$ measures the speed of convergence of the vector

$$\left(\frac{|u_0 u_1 \dots u_{N-1}|_a}{N} \right)_{a \in \mathcal{A}}$$

towards the frequencies vector A .

A subset \mathcal{E} of \mathcal{A} is said a *bounded remainder set* for the sequence U with respect to the measure μ if the sequence

$$\left(\sum_{k=0}^{N-1} (\chi_{\mathcal{E}}(u_k) - \mu(\mathcal{E})) \right)_{N \in \mathbb{N}^*}$$

is bounded. Then, the discrepancy $D_N(\mu, U)$ is bounded if and only if each element of \mathcal{A} is a bounded remainder set. In the following, we will link together sequences with bounded discrepancy and C -balanced sequences.

The previous definitions, given for a finite set, come directly from the more classical notions of uniform distribution modulo 1 and discrepancy for real sequences. Let us notice that these notions can also be generalized to topological, compact or quasi-compact groups. Two important references on this subject are the books of Kuipers and Niederreiter [26] and of Drmota and Tichy [16].

Landau symbols: Let f and g be two real positive functions. We recall the definition of some Landau symbols:

$$f = O(g) \text{ if } \exists C > 0 \text{ such that } f(x) < Cg(x) \quad \forall x \in \mathbb{R}_+,$$

$$f = o(g) \text{ if } \lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0,$$

$$f = \Omega(g) \text{ if } f \neq o(g), \text{ that is } \overline{\lim} \frac{f}{g} > 0.$$

We introduce now a notation that we will use in most of our results. We will write

$$f = (O \cap \Omega)(g)$$

if both $f = O(g)$ and $f = \Omega(g)$. This a priori just means that g is, in a certain sense, an optimal asymptotic upper bound for the function f . However, in this study, the fact that $f = \Omega(g)$ will be in most of cases more significant than the relation $f = O(g)$.

Remark 4. The relation $f = (O \cap \Omega)(g)$ does of course not imply that $f \sim g$, but, in this paper we study some functions which could be oscillating, and we are essentially interested by their maximum values. So, we will sometimes write, by abuse of language, that the order of magnitude of the function f is g , as soon as the relation $f = (O \cap \Omega)(g)$ holds.

A preliminary result: In the study led in [1] on the asymptotic behaviour of the discrepancy function for fixed points of primitive substitutions, we prove the following result which lays the foundations for the study we present in this paper.

Theorem 5 (Adamczewski [1]). *Let U be a fixed point of a primitive substitution σ . Then, we have:*

- | | |
|--|--|
| (i) if $ \theta_2 < 1$, then | $D_N(A, U)$ is bounded, |
| (ii) if $ \theta_2 > 1$, then | $D_N(A, U) = (O \cap \Omega)((\log N)^{\alpha_2} N^{(\log_\theta \theta_2)}),$ |
| (iii) if $ \theta_2 = 1$ and θ_2 is not
a root of unity, then | $D_N(A, U) = (O \cap \Omega)((\log N)^{(\alpha_2+1)}),$ |
| (iv) if $ \theta_2 = 1$ and θ_2 is
a root of unity, then | |
| either $A_{\sigma, U} \neq 0$ and | $D_N(A, U) = (O \cap \Omega)((\log N)^{(\alpha_2+1)})$ |
| or $A_{\sigma, U} = 0$ and | $D_N(A, U) = (O \cap \Omega)((\log N)^{\alpha_2}),$ |

where the complex number $A_{\sigma, U}$ (which just depends on the pair (σ, U)) is defined in Appendix B and could explicitly be computed.

Remark 6. All the different cases are thus covered by Theorem 5.

3. Links between balance and discrepancy

In this section, we study the connections between the balance and the discrepancy functions for a symbolic sequence. We note the existence of some links in the general case and also their limits. In particular, we show that these two notions are strongly connected in the case of linearly recurrent sequences.

The general case: We first exhibit some links between the balance and the discrepancy functions. We prove in particular that if one of both of these functions is bounded, then the other should be bounded too. We thus obtain that problems of bounded remainder sets and C -balanced sequences are strongly connected.

Proposition 7. *Let $U = u_0u_1 \dots u_n \dots$ be a symbolic sequence defined over the alphabet \mathcal{A} . Then the two following propositions are equivalent:*

- (i) *There exists a probability measure Λ such that $D_N(\Lambda, U)$ is bounded.*
- (ii) *$B_N(U)$ is bounded.*

Proof. Let us show that (i) \Rightarrow (ii). We thus suppose there exists a constant $C > 0$ such that:

$$\forall a \in \mathcal{A}, \quad ||u_0u_1 \dots u_{N-1}| - N\Lambda_a| \leq C. \quad (2)$$

Let us consider an integer N and ω a factor of length N in U . If the integer k denotes an occurrence of ω in U , we have

$$u_0u_1 \dots u_{k-1} \dots u_{N+k-1} = u_0u_1 \dots u_{k-1}\omega.$$

Let $a \in \mathcal{A}$, (2) implies

$$||u_0u_1 \dots u_{N+k-1}|_a - (N+k)\Lambda_a| \leq C.$$

We obtain

$$|(|u_0u_1 \dots u_{k-1}|_a - k\Lambda_a) + (|\omega|_a - N\Lambda_a)| \leq C$$

and thus because of (2)

$$||\omega|_a - N\Lambda_a| \leq 2C.$$

Finally, for any pair (ω_1, ω_2) of factors of length N , we have

$$||\omega_1|_a - |\omega_2|_a| < 4C,$$

which implies that the balance function of U is bounded by $4C$.

We have now to prove (ii) \Rightarrow (i). Let us suppose that the balance function of U is bounded by a positive constant C , that is to say:

$$\forall N \in \mathbb{N} \quad \forall a \in \mathcal{A}, \quad \forall \omega, \omega' \in \mathcal{L}_N(U), \quad ||\omega|_a - |\omega'|_a| \leq C. \quad (3)$$

Let us first fix a letter a in the alphabet \mathcal{A} . For every integer k , w_k denotes a word in $\mathcal{L}_k(U)$ such that:

$$\forall w \in \mathcal{L}_k(U), \quad |w|_a \geq |w_k|_a.$$

Such a word w_k clearly exists. Then, if $n \in \mathbb{N}$, $l \in \mathbb{N}$, and $w \in \mathcal{L}_{nl}(U)$, inequality (3) implies

$$0 \leq |w|_a - l|w_n|_a \leq lC \quad \text{and} \quad 0 \leq |w|_a - n|w_l|_a \leq nC.$$

By subtraction, we obtain

$$-nC \leq n|w_l|_a - l|w_n|_a \leq lC$$

and

$$-\frac{C}{l} \leq \frac{|w_l|_a}{l} - \frac{|w_n|_a}{n} \leq \frac{C}{n}.$$

The sequence $(|w_n|_a/n)_{n \in \mathbb{N}}$ is thus convergent. Therefore, there exists a real A_a , the limit of this sequence, satisfying:

$$0 \leq A_a - \frac{|w_n|_a}{n} \leq \frac{C}{n}. \quad (4)$$

Let us remember that (3) implies for every integer n

$$0 \leq |U_n|_a - |w_n|_a \leq C,$$

that is to say,

$$0 \leq \frac{|U_n|_a}{n} - \frac{|w_n|_a}{n} \leq \frac{C}{n}.$$

By subtracting inequality (4), it follows:

$$-\frac{C}{n} \leq \frac{|U_n|_a}{n} - A_a \leq \frac{C}{n}.$$

The sequence $(|U_n|_a/n)_{n \in \mathbb{N}}$ converges thus to A_a , the frequency A_a of the letter a . We obtain finally

$$\forall n \in \mathbb{N}, \quad ||U_n|_a - nA_a| \leq C,$$

which ends the proof. \square

Remark 8. In the previous demonstration, we show in fact the following: If for any letter i , the set $\{i\}$ is a bounded remainder set (with bound C) for the sequence U with respect to the probability measure \mathcal{A} , then the balance function of U is bounded by $4C$.

In the case of the Morse sequence (see Example 24), the letters 0 and 1 correspond to bounded remainder sets (with bound $\frac{1}{2}$) with respect to the uniform probability measure on $\{0, 1\}$ (the bound $\frac{1}{2}$ being reach). Moreover, this sequence is 2-balanced (but not balanced). This shows that the upper bound we obtain in the proof of Proposition 7 is optimal.

Proposition 7 states that both balance and discrepancy functions have the same order of magnitude if one of them is bounded. We show now that it does not hold in the general case (Proposition 10). However, we prove that the order of magnitude of the discrepancy function is at most the one of the balance function (Proposition 9).

Proposition 9. Let $U = u_0 u_1 \dots u_n \dots$ be a sequence defined over the alphabet $\mathcal{A} = \{1, 2, \dots, d\}$ and such that each letter of \mathcal{A} admits a frequency in U . Let $\Lambda = (\Lambda_i)_{i \in \mathcal{A}}$ denote the frequencies vector of U . If the balance function of U satisfies:

$$B_U(N) = O(f(N)) \text{ (respectively } B_U(N) = o(f(N)))$$

then

$$D_N(\Lambda, U) = O(f(N)) \text{ (respectively } D_N(\Lambda, U) = o(f(N))).$$

Proof. The same reasoning as in the second part of the proof of Proposition 7 applies if we replace C by $f(N)$. \square

We produce now a particular sequence with both small discrepancy and extremely bad balance. We thus deduce that a converse to Proposition 9 could not hold.

Proposition 10. Let f be a real increasing unbounded function such that $f(N) = o(N)$. Then, there exists a sequence U defined over the alphabet $\{0, 1\}$ satisfying:

- (i) U has a frequency vector denoted by Λ ,
- (ii) $D_N(\Lambda, U) = O(f(N))$,
- (iii) for every integer N , $B_U(N) = N$.

Proof. Let f be a real increasing unbounded function. Let $\lfloor x \rfloor$ denote the fractional part of the real x . We introduce then the following binary sequence U defined over $\{0, 1\}$ in the following way:

$$\begin{aligned}
 U = & 01 \underbrace{00 \dots 0}_{\lfloor f(1) \rfloor \text{ times}} \underbrace{11 \dots 1}_{\lfloor f(1) \rfloor \text{ times}} 0101 \underbrace{00 \dots 0}_{\lfloor f(2) \rfloor \text{ times}} \underbrace{11 \dots 1}_{\lfloor f(2) \rfloor \text{ times}} \\
 & \dots \underbrace{0101 \dots 01}_{N \text{ times}} \underbrace{00 \dots 0}_{\lfloor f(N) \rfloor \text{ times}} \underbrace{11 \dots 1}_{\lfloor f(N) \rfloor \text{ times}} \dots
 \end{aligned}$$

We obtain that U has arbitrary long blocks of 0's and 1's because f is unbounded. This implies that

$$\forall N \in \mathbb{N}^*, \quad B_U(N) = N.$$

If U_N denotes the prefix of length N of U , we obtain then by construction of our sequence

$$|U_N|_0 - |U_N|_1 \leq \lfloor f(N) \rfloor, \tag{5}$$

since f is increasing, and thus the letters 0 and 1 have a frequency equal to $\frac{1}{2}$ in U , because $f(N) = o(N)$. By inequality (5), we have

$$||U_N|_0 - \frac{1}{2}N| \leq \frac{1}{2} \lfloor f(N) \rfloor$$

and

$$||U_N|_1 - \frac{1}{2}N| \leq \frac{1}{2} \lfloor f(N) \rfloor.$$

Then, if A means the uniform probability vector over $\{0, 1\}$, we obtain

$$D_N(A, U) \leq \frac{1}{2} \lfloor f(N) \rfloor$$

and thus

$$D_N(A, U) = O(f(N)),$$

which ends the proof. \square

Remark 11. The sequence U considered in the proof of Proposition 10 is clearly not linearly recurrent since the blocks of 0's and of 1's first occur at a rank which is not proportional to their lengths.

The case of LR sequences: It is proved in [19] that if U is a LR sequence then each letter in U admits a frequency. Hence U is uniformly distributed with respect to the natural probability measure given by its frequencies. The following result states a partial converse to Proposition 9 in the case of LR sequences.

Proposition 12. *Let U a linearly recurrent sequence (with constant K) defined over the alphabet $\mathcal{A} = \{1, 2, \dots, d\}$. Let $A = (A_i)_{i \in \mathcal{A}}$ denote the frequency vector associated with U . If there exists an increasing sublinear function f (that is $\forall(x, y), f(x + y) \leq f(x) + f(y)$) such that:*

$$D_N(A, U) = O(f(N)) \quad (\text{respectively } D_N(A, U) = o(f(N)))$$

then

$$B_U(N) = O(f(N)) \quad (\text{respectively } B_U(N) = o(f(N))).$$

Proof. Let us suppose that there exists an increasing function f satisfying

$$D_N(A, U) = O(f(N)).$$

Then, there exists a positive constant $A > 0$ such that:

$$\forall N \in \mathbb{N}^* \quad \forall a \in \mathcal{A}, \quad ||u_0 u_1 \dots u_{N-1}|_a - NA_a| < Af(N).$$

Let N be an integer and ω a factor of length N in U . Let k_ω be the first occurrence of ω in U , we have

$$u_0 u_1 \dots u_{k_\omega-1} \dots u_{N+k_\omega-1} = u_0 u_1 \dots u_{k_\omega-1} \omega.$$

The fact that U is LR with constant K implies $N + k_\omega - 1 < KN$. We thus have that for $a \in \mathcal{A}$

$$||u_0 u_1 \dots u_{N+k_\omega-1}|_a - (N + k_\omega)A_a| < Af(KN),$$

which implies

$$|(|u_0 u_1 \dots u_{k_\omega-1}|_a - k_\omega A_a) + (|\omega|_a - NA_a)| < Af(KN)$$

and thus

$$|\omega|_a - N\lambda_a < 2Af(KN),$$

since

$$|u_0u_1 \dots u_{k_N-1}|_a - k_N\lambda_a < Af(KN),$$

because f is increasing. Finally, we obtain that for any pair (ω_1, ω_2) of factors of length N :

$$|\omega_1|_a - |\omega_2|_a < 4Af(KN),$$

which implies that $B_U(N) = O(f(N))$ since f is sublinear. The same reasoning applies in the case where $D_N(\lambda, U) = o(f(N))$. \square

4. Main results

In view of the study led in Section 3, we can translate Theorem 5 in terms of balance function. Having fixed the notation in Section 2, then we obtain the following theorem.

Theorem 13. *Let U be a fixed point of a primitive substitution σ . Then, we have:*

- (i) if $|\theta_2| < 1$, then $B_U(N)$ is bounded,
- (ii) if $|\theta_2| > 1$, then $B_U(N) = (O \cap \Omega)((\log N)^{\alpha_2} N^{\log_\theta |\theta_2|})$,
- (iii) if $|\theta_2| = 1$ and θ_2 is not a root of unity, then $B_U(N) = (O \cap \Omega)((\log N)^{\alpha_2+1})$,
- (iv) if $|\theta_2| = 1$ and θ_2 is a root of unity, then
 either $A_{\sigma,U} \neq 0$ and $B_U(N) = (O \cap \Omega)((\log N)^{\alpha_2+1})$
 or $A_{\sigma,U} = 0$ and $B_U(N) = (O \cap \Omega)((\log N)^{\alpha_2})$,

where the complex number $A_{\sigma,U}$ (which just depends on the pair (σ, U)) is defined in Appendix B and could explicitly be computed.

Remark 14. This theorem implies in particular that for any fixed point of a primitive substitution U , $B_U(N) = o(N)$. We also refer the reader to (1) for a better understanding of the notation.

Proof. We have first to remark that a fixed point of a primitive substitution is linearly recurrent. This point is proved in [20]. Then, the result follows from Theorem 5 and Propositions 9 and 12. \square

We thus obtain the following characterization of the C -balanced fixed points of primitive substitutions.

Corollary 15. *A fixed point U of a primitive substitution σ has a bounded balance function if and only if one of the following holds:*

- (i) $|\theta_2| < 1$,
- (ii) $|\theta_2| = 1$, $\alpha_2 = 0$, θ_2 is a root of unity and $A_{\sigma,U} = 0$.

We bring now some direct applications of Theorem 13 and we focus on case (iv).

The first interesting (or surprising) point is that the three first cases of Theorem 13 just depend on θ_2 and thus on the incidence matrix associated with the substitution. In particular, we obtain that in these cases, the asymptotic behaviour of the balance function is not modified by any permutation of the letters in the definition of the substitution. As an example, the balance functions of the two fixed points of the substitutions

$$\begin{array}{ll} \sigma_1 & \sigma_2 \\ 1 \mapsto 12131234 & 1 \mapsto 11122334 \\ 2 \mapsto 12131334 & \text{and } 2 \mapsto 11123334 \\ 3 \mapsto 12242434 & 3 \mapsto 12223444 \\ 4 \mapsto 13342434 & 4 \mapsto 12333444 \end{array}$$

have the same asymptotic behaviour. A natural question is then to ask if this property holds for any fixed points of primitive substitutions. The answer is no. In fact, let us consider, as was suggested to me by J. Cassaigne, the two following substitutions defined over the alphabet $\{1,2\}$ by:

$$\begin{array}{ll} \xi_1 & \xi_2 \\ 1 \mapsto 112 & \text{and } 1 \mapsto 121, \\ 2 \mapsto 212 & 2 \mapsto 212 \end{array}$$

and let U_1 and U_2 denote, respectively, the fixed points beginning with the letter 1 of the substitutions ξ_1 and ξ_2 . It is clear that $B_{U_2}(N)$ is bounded since U_2 is a periodic sequence. On the other hand, one can show (using that $A_{\xi_1,U_1} \neq 0$; we recall its definition in Appendix B) that $B_{U_1}(N) = \Omega(\log(N))$. Even so, these two substitutions share the same incidence matrix. This example is relatively simple, but not totally convincing. One can think that if we restrict our study to non-ultimately periodic sequences, such a situation does not hold any more. However, we provide in Section 6 (Example 6) two non-ultimately periodic fixed points of substitutions sharing the same incidence matrix, corresponding, respectively, to both situations which can occur in case (iv) of Theorem 13 (see Section 6). We thus obtain that the “strange” class of substitution considered in case (ii) of Corollary 15 is really not empty. In particular, this puts an end to the hope of characterizing the non-ultimately periodic fixed points of primitive substitutions with bounded balance function just in terms of their incidence matrices.

We are now going to use Corollary 15 in order to understand what type of spectrum could have the incidence matrix associated with a primitive substitution which generates an eventually periodic sequence. This is in part motivated by a first result due to Holton and Zamboni.

Proposition 16 (Holton–Zamboni [24]). *If a fixed point of a primitive substitution τ is eventually periodic, then the incidence matrix associated with τ could not have non-zero eigenvalues of modulus less than one.*

Then, we can deduce the following.

Corollary 17. *Let τ be a primitive substitution which generates an eventually periodic fixed point over the alphabet \mathcal{A} . Then, the following holds:*

- (i) M_τ has a simple positive integer eigenvalue (its Perron eigenvalue),
- (ii) the other non-zero eigenvalues of M_τ are all roots of unity (whose algebraic degree is less than the cardinality of the set \mathcal{A} minus one),
- (iii) the non-zero eigenvalues of M_τ are all simple.

Proof. It follows from Proposition 16 that the non-zero eigenvalues of M_τ have modulus greater than or equal to one.

If M_τ has an eigenvalue of modulus greater than one which is not its Perron eigenvalue, this would imply that $|\theta_2| > 1$, and then by Corollary 15 the balance function of the eventually periodic sequence would not be bounded, hence a contradiction.

If M_τ has an eigenvalue of modulus equal to one, Corollary 15 implies that this eigenvalue is necessarily simple and a root of unity.

The minimal polynomial of θ , the Perron eigenvalue of M_τ , should divide the characteristic polynomial of M_τ . Then the algebraic conjugates of θ lie necessarily among the eigenvalues of M_τ . But, 0 could obviously neither be an algebraic conjugate of θ nor a root of unity, since θ is greater than one. We obtain finally that θ could not have any algebraic conjugate and should thus be integer.

Now, let β denote an eigenvalue of M_τ which is a root of unity and P its minimal polynomial. The degree of P is necessarily less than the cardinality of the set \mathcal{A} minus one because $(X - \theta)P$ should divide the characteristic polynomial of M_τ , concluding the proof. \square

Corollary 17 claims that the incidence matrices associated with primitive substitutions which generate eventually periodic sequences should have very specific types of spectrum composed among roots of unity and zero eigenvalues (if we forget their Perron eigenvalue). It is thus natural to ask if this result is optimal, that is to say, if roots of unity or/and zero could really be eigenvalues of such matrices. The substitution defined over $\{1, 2\}$ by

$$\begin{array}{l} \phi \\ 1 \mapsto 121 \\ 2 \mapsto 212 \end{array}$$

has spectrum $S_{M_\phi} = \{3, 1\}$, whereas the substitution defined over $\{1, 2\}$ by

$$\begin{array}{l} \psi \\ 1 \mapsto 121 \\ 2 \mapsto 121 \end{array}$$

has spectrum $S_{M_\psi} = \{3, 0\}$, and the substitution defined over $\{1, 2, 3\}$ by

$$\begin{aligned} & \tau \\ 1 & \mapsto 12 \\ 2 & \mapsto 312 \\ 3 & \mapsto 3123 \end{aligned}$$

has spectrum $S_{M_\tau} = \{3, 1, 0\}$; moreover these three substitutions generate periodic fixed points, which provides a positive answer to our question. Another consequence of Corollary 17 is that for a primitive substitution σ which generates an eventually periodic fixed point over a two or a three-letter alphabet, $S_{M_\sigma} \subset \mathbb{Z}$.

5. Application to generalized balances

In this section, we apply our results to generalized balance properties for fixed points of primitive substitutions. The balance function measures the difference between the number of occurrences of each letter in any pair of factors of the same length. We want now to introduce a similar notion but with words playing the role of letters. This generalization is inspired by Fagnot and Vuillon [21] who study generalized balances in Sturmian words.

Let U be a symbolic sequence defined over the alphabet \mathcal{A} . Then, we can define, for any positive integer n , a generalized balance function of order n for U , in the following way:

$$B_U^{(n)}(N) = \max_{u \in \mathcal{L}_n(U)} \max_{w, w' \in \mathcal{L}_N(U)} \{ ||w|_u - |w'|_u | \}, \quad (6)$$

where $|w|_u$ denotes the number of occurrences of the word u in the word w . We obtain in particular $B_U^{(1)}(N) = B_U(N)$. In view of the previous study, it is a natural question to ask if we can estimate the growth order of these generalized balance functions for fixed points of primitive substitutions. In particular, is it possible to obtain such an information in terms of the incidence matrix associated with the substitution.

In order to answer this question, we recall now a useful construction which can be found in [31]. Let σ be a primitive substitution defined over the alphabet \mathcal{A} and U an associated fixed point. For any positive integer l , \mathcal{A}_l denotes the alphabet $\{1, 2, \dots, P_U(l)\}$, where P_U is the complexity function of U . We can thus consider a map Θ_l from $\mathcal{L}_l(U)$ to \mathcal{A}_l which associates with each factor of length l its order of occurrence in U . If i denotes a letter of the alphabet \mathcal{A}_l , we can conversely associate with i a unique word $\Theta_l^{-1}(i) = w_0 w_1 \dots w_{l-1} \in \mathcal{L}_l(U)$ since Θ_l is one-to-one. If

$$\sigma(\Theta^{-1}(i)) = \sigma(w_0 w_1 \dots w_{l-1}) = y_0 y_1 \dots y_{|\sigma(w_0)|-1} y_{|\sigma(w_0)|} \dots y_{|\sigma(\Theta^{-1}(i))|-1},$$

then, we define the substitution of order l for σ by

$$\sigma_l(i) = \Theta((y_0 y_1 \dots y_{l-1})(y_1 y_2 \dots y_l) \dots (y_{|\sigma(y_0)|-1} \dots y_{|\sigma(y_0)|+l-2})). \quad (7)$$

So defined, $|\sigma_l(i)| = |\sigma(w_0)|$.

Example 18. Let U be the Fibonacci sequence, defined as the fixed point of the substitution

$$\begin{array}{l} \sigma \\ 1 \mapsto 12 \\ 2 \mapsto 1 \end{array}$$

It is well-known that U is a Sturmian sequence and thus admits three factors of length two, more precisely,

$$\mathcal{L}_2(U) = \{(12), (21), (11)\}.$$

Then, we obtain

$$\Theta^{-1}(1) = (12), \quad \Theta^{-1}(2) = (21) \quad \text{and} \quad \Theta^{-1}(3) = (11),$$

and since $\sigma(12)=121$, $\sigma(21)=112$ and $\sigma(11)=1212$, it follows that the substitution of order two for σ is defined by

$$\sigma_2(1) = 12, \quad \sigma_2(2) = 3 \quad \text{and} \quad \sigma_2(3) = 12.$$

We recall now some results about the previous construction.

Proposition 19 (Queff elec [31]). *For every positive integer l , the substitution of order l for a substitution σ admits the sequence $U_l = \sigma_l^\infty(1)$ as fixed point. Moreover, if $U = u_0u_1 \dots u_n \dots$ means the fixed point of σ , then the sequence $\Theta_l^{-1}(U_l)$ is composed by all the factors of length l of U without repetition and in the same order as in U , that is to say,*

$$\Theta_l^{-1}(U_l) = (u_0u_1 \dots u_{l-1})(u_1u_2 \dots u_l) \dots (u_nu_{n+1} \dots u_{n+l-1}) \dots$$

We can already notice that if $U_l = u_0^{(l)}u_1^{(l)} \dots u_n^{(l)} \dots$, then

$$|u_0^{(l)}u_1^{(l)} \dots u_n^{(l)}|_i = |u_0u_1 \dots u_n|_{\Theta^{-1}(i)}, \tag{8}$$

where $|\omega|_{\Theta^{-1}(i)}$ means the number of occurrences of the word $\Theta^{-1}(i)$ in ω . This implies in particular the following corollary.

Corollary 20. *The order of magnitude of the generalized balance function of order l for U is the same as that of the balance function of U_l .*

Proposition 21 (Queff elec [31]). *If σ is a primitive substitution then for every positive integer l , the substitution σ_l is primitive too and its incidence matrix M_l has the same Perron eigenvalue as the one of σ .*

The eigenvalues of M_l , $l \geq 2$, are those of M_2 with perhaps in addition the eigenvalue 0. Moreover, if P_2 is the minimal polynomial of M_2 , then there exists an integer m such that $P_l = P_2X^m$, where P_l means the minimal polynomial of M_l .

Following Eq. (1), we can note

$$S_{M_{\sigma_l}} = \{\theta_{l,i}, 2 \leq i \leq d_l\} \cup \{\theta_{l,1} = \theta_l\}$$

the spectrum of the incidence matrix associated with σ_l . Proposition 21 implies that $\theta_l = \theta$, $\theta_{l,2} = \theta_{2,2}$ and $\alpha_{l,2} = \alpha_{2,2}$, where $\alpha_{l,2}$ means the multiplicity of $\theta_{l,2}$ in the minimal polynomial of M_l . In view of Corollary 20 and Proposition 21, we can state the following result.

Theorem 22. *Let U be a fixed point of a primitive substitution σ . Then, we have for every integer $l \geq 2$:*

- (i) if $|\theta_{2,2}| < 1$, then $B_U^{(l)}(N)$ is bounded,
- (ii) if $|\theta_{2,2}| > 1$, then $B_U^{(l)}(N) = (O \cap \Omega)((\log N)^{\alpha_{2,2}} N^{\log_0 |\theta_{2,2}|})$,
- (iii) if $|\theta_{2,2}| = 1$ and $\theta_{2,2}$ is not a root of unity, then $B_U^{(l)}(N) = (O \cap \Omega)((\log N)^{\alpha_{2,2}+1})$,
- (iv) if $|\theta_{2,2}| = 1$ and $\theta_{2,2}$ is a root of unity, then $B_U(N) = O((\log N)^{\alpha_{2,2}+1})$
and $B_U(N) = \Omega((\log N)^{\alpha_{2,2}})$.

Moreover, in the case where $\theta_{2,2}$ is necessarily a root of unity, then:

- either $\forall l \geq 2$, $B_U^{(l)}(N) = (O \cap \Omega)((\log N)^{\alpha_{2,2}})$,
- or there exists an integer $m \geq 2$ such that,
for $l < m$, $B_U^{(l)}(N) = (O \cap \Omega)((\log N)^{\alpha_{2,2}})$
and for $l \geq m$, $B_U^{(l)}(N) = (O \cap \Omega)((\log N)^{\alpha_{2,2}+1})$.

Before proving Theorem 22, we need to establish the following lemma.

Lemma 23. *Let l be a positive integer, U an infinite sequence defined over the alphabet \mathcal{A} and suppose that there exists a function f such that $B_U^{(l)}(N) = \Omega(f(N))$. Then, we have $B_U^{(l+1)}(N) = \Omega(f(N))$.*

Proof. Let P_l be the projection map defined from $\mathcal{L}_{l+1}(U)$ to $\mathcal{L}_l(U)$ by:

$$P_l((w_0 w_1 \dots w_{l+1})) = (w_0 w_1 \dots w_l).$$

Then, we obtain that

$$\Theta_l \circ P_l \circ \Theta_{l+1}^{-1}(U_{l+1}) = U_l.$$

The sequence U_l is thus the image by a morphism letter-to-letter of the sequence U_{l+1} , which implies that

$$B_{U_{l+1}}(N) = \Omega(f(N)) \text{ as soon as } B_{U_l}(N) = \Omega(f(N)).$$

Finally, we obtain by Corollary 20 that

$$B_U^{(l+1)}(N) = \Omega(f(N)) \text{ as soon as } B_U^{(l)}(N) = \Omega(f(N)),$$

concluding the proof. \square

Lemma 23 points out the fact that the order of magnitude of the generalized balance functions $B_U^{(l)}(N)$ associated with a symbolic sequence U could not decrease with respect to l .

Proof of Theorem 22. Equalities (i), (ii), (iii) and (iv) come directly from Corollary 20, Proposition 21 and Theorem 13. Then, the last point of Theorem 22 is a consequence of Lemma 23. \square

If we come back to Example 18, we obtain that the incidence matrix associated with the Fibonacci substitution of order 2 admits three simple eigenvalues:

$$\left\{ \frac{\sqrt{5}-1}{2}, \frac{\sqrt{5}+1}{2}, 0 \right\}.$$

Theorem 22 implies thus that the Fibonacci sequence admits bounded balance functions of all orders. A more precise result is shown in [21] for Sturmian sequences. If U means a Sturmian sequence, then

$$B_U^{(n)}(N) \text{ is bounded by } n;$$

moreover, if the slope of U has bounded partial quotients in its continued fraction expansion, then

$$\sup_{n \in \mathbb{N}} (B_U^{(n)}(N)) \text{ is finite}$$

and an explicit bound is given.

Example 24. Let U be the Morse sequence, defined as the fixed point beginning by 1 of the substitution

$$\begin{array}{l} \sigma \\ 1 \mapsto 12 \\ 2 \mapsto 21 \end{array}$$

The incidence matrix associated with the Morse substitution admits 2 and 0 as eigenvalues. It is well-known that U has four factors of length two, more precisely,

$$\mathcal{L}_2(U) = \{(12), (22), (21), (11)\}.$$

We thus obtain

$$\Theta^{-1}(1) = (12), \quad \Theta^{-1}(2) = (22), \quad \Theta^{-1}(3) = (21) \quad \text{and} \quad \Theta^{-1}(4) = (11),$$

and since $\sigma(12)=1221$, $\sigma(22)=2121$, $\sigma(21)=2112$ and $\sigma(11)=1212$, it follows that the substitution of order two for σ is defined by

$$\sigma_2(1) = 14, \quad \sigma_2(2) = 31, \quad \sigma_2(3) = 34 \quad \text{and} \quad \sigma_2(4) = 31.$$

The incidence matrix associated with the Morse substitution of order two admits four simple eigenvalues:

$$\{2, 1, -1, 0\}.$$

Contrary to the Fibonacci case, new non-zero eigenvalues appear. One can thus think that the Morse sequence is more “well-balanced” with respect to its letters than to its factors of length two. Actually, we can show that $A_{\sigma_2, U_2} = 0$ and thus that the Morse balance function of order two is bounded too. However, and because we are in the critical case (case (iv) of Theorem 22), we can not say if the Morse balance functions of any order are bounded or not.

In this section, we have seen that all the incidence matrices associated with the substitutions of order at least two share the same spectrum (except for the zero eigenvalue). Moreover, we have exhibited an example (the Morse sequence) for which the spectrum of the substitution of order two is really distinct from the one of the initial substitution. However, we do not know any such an example for which this change of spectrum is really significant for the balance properties of the studied sequence.

6. A zoo of examples

In this section, we apply our results to some classical substitutions. We give examples of substitutions whose balance functions have the different types of growth order discussed in Theorem 13. This list does not claim of course to be exhaustive.

Pisot-type substitutions: We call Pisot-type substitution a substitution for which $|\theta_2| < 1$ (some authors require that a Pisot-type substitution has no zero eigenvalue). This class of substitutions corresponds to case (i) in Theorem 13 and contains in particular the Morse substitution (see Example 24), the Fibonacci substitution (see Example 18) and more generally Sturmian substitutions (see [13]), the Tribonacci substitution (see [32]) and more generally Arnoux–Rauzy substitutions (see [6]). Fixed points generated by all these substitutions have thus, in view of Theorem 13, a bounded balance function. However, optimal bounds are already well-known for the Morse sequence (which is 2-balanced) and Sturmian sequences (which are balanced).

Badely balanced substitutions: We call badely balanced substitution, a substitution satisfying $|\theta_2| > 1$, which corresponds to case (ii) in Theorem 13.

The Rudin–Shapiro sequence: The Rudin–Shapiro substitution is defined over the alphabet $\{1, 2, 3, 4\}$ by

$$\begin{aligned} 1 &\mapsto 12 \\ 2 &\mapsto 13 \\ 3 &\mapsto 24 \\ 4 &\mapsto 34 \end{aligned}$$

Let us denote by U the fixed point of this substitution generated by 1. The Rudin–Shapiro sequence, which is the image of the sequence U by a letter-to-letter projection, was introduced independently in [36,37] or estimating problems in harmonic analysis.

The incidence matrix associated with the Rudin–Shapiro substitution has four simple real eigenvalues:

$$\{2, \sqrt{2}, -\sqrt{2}, 0\}.$$

We thus obtain $\theta=2$, $|\theta_2|=\sqrt{2}$ and $\alpha_2=0$. Then, we have in view of Theorem 13:

$$B_U(N) = (O \cap \Omega)(\sqrt{N}).$$

A substitution related to the sum of the dyadic digits: Let us consider the substitution σ defined over the alphabet $\{1, 2, 3, 4, 5, 6\}$ by

$$\begin{aligned} 1 &\mapsto 12 \\ 2 &\mapsto 13 \\ 3 &\mapsto 26 \\ 4 &\mapsto 45 \\ 5 &\mapsto 46 \\ 6 &\mapsto 53 \end{aligned}$$

Let us denote by U the fixed point of this substitution generated by 1. This sequence is related to the sum $\sum_{n < N} (-1)^{s(3n)}$, where $s(n)$ denotes the sum of the dyadic digits of n (see [11,14]). The incidence matrix associated with σ admits six simple eigenvalues:

$$\{2, \pm\sqrt{3}, \pm 1, 0\}.$$

We thus obtain $\theta=2$, $|\theta_2|=\sqrt{3}$ and $\alpha_2=0$. Then, Theorem 13 implies:

$$B_U(N) = (O \cap \Omega)(N^{(\log_4 3)}).$$

We can note that in this case the order of magnitude of the balance function is an irrational power of N contrary to the case of the Rudin–Shapiro substitution.

Extremely badly balanced substitutions: For any positive integer n , let us consider the substitution σ_n defined over the alphabet $\{1, 2\}$ by

$$\begin{aligned} 1 &\mapsto 1^n 2 \\ 2 &\mapsto 12^n \end{aligned}$$

and let us denote by U_n its fixed point. The incidence matrix associated with σ_n has two simple eigenvalues:

$$\{(n+1), (n-1)\}.$$

It follows that for $n \geq 3$, $\theta_{2,n} > 1$ and

$$\lim_{n \rightarrow \infty} \log_{\theta_n}(\theta_{2,n}) = 1,$$

which thus implies that

$$\forall \varepsilon > 0 \exists n \in \mathbb{N}, \text{ such that, } B_{U_n}(N) = \Omega(N^{1-\varepsilon}).$$

We thus provide examples of fixed points of primitive substitutions whose balance functions take highest growth orders as possible in view of Remark 14.

Not so badly balanced substitutions: For any pair of positive integers (n, k) , let us consider the substitution $\sigma_{(n,k)}$ defined over the alphabet $\{1, 2\}$ by

$$\begin{aligned} 1 &\mapsto 1^{(n^k)}2^{(n^k-n)} \\ 2 &\mapsto 1^{(n^k-n)}2^{(n^k)} \end{aligned}$$

and let us denote by $U_{(n,k)}$ its fixed point. The incidence matrix associated with $\sigma_{(n,k)}$ has two simple eigenvalues:

$$\{(2n^k - n), n\}.$$

It thus follows

$$\log_{\theta_{(n,k)}}(\theta_{2,(n,k)}) \underset{n \rightarrow \infty}{\sim} \frac{1}{k},$$

which implies that

$$\forall \varepsilon > 0, \exists (n, k) \in \mathbb{N} \times \mathbb{N}, \text{ such that, } B_{U_{(n,k)}}(N) = O(N^\varepsilon).$$

We thus provide examples of fixed points of badly balanced substitutions whose balance functions take smallest growth orders as possible in view of Theorem 13.

A Salem-type substitution: We call Salem-type substitution, a substitution for which θ , the Perron eigenvalue, is a Salem number. We recall that a Salem number is a real algebraic number greater than one whose all conjugates have a modulus smaller than or equal to one, one at least having a modulus equal to one.

Let us consider the substitution σ defined over the alphabet $\{1, 2, 3, 4\}$ by:

$$\begin{aligned} 1 &\mapsto 12 \\ 2 &\mapsto 14 \\ 3 &\mapsto 2 \\ 4 &\mapsto 3 \end{aligned}$$

Let us denote by U the fixed point of this substitution generated by 1. The sequence U was introduced in [24]. The characteristic polynomial of M_σ is

$$x^4 - x^3 - x^2 - x + 1 = \left(x^2 + \frac{1}{2}(-1 + \sqrt{3})x + 1\right) \left(x^2 + \frac{1}{2}(-1 - \sqrt{3})x + 1\right).$$

The roots of the first quadratic factor are

$$\beta = \frac{1 - \sqrt{13} + \sqrt{2 + 2\sqrt{13}i}}{4} \quad \text{and} \quad \gamma = \frac{1 - \sqrt{13} - \sqrt{2 + 2\sqrt{13}i}}{4},$$

which have modulus one but are not roots of unity. In this example, θ is a Salem number. We thus obtain $|\theta_2|=1$ and θ_2 is not a root of unity. Moreover, $\alpha_2=0$ because β and γ are simple eigenvalues. Then, case (iii) of Theorem 13 implies:

$$B_U(N) = (O \cap \Omega)(\log N).$$

Critical cases: We give here examples of substitutions corresponding to case (iv) in Theorem 13.

The Chacon sequence: The primitive Chacon substitution σ is defined over the alphabet $\{1, 2, 3\}$ by

$$\begin{aligned} 1 &\mapsto 1123 \\ 2 &\mapsto 23 \\ 3 &\mapsto 123 \end{aligned}$$

Let us denote by U the fixed point of this substitution generated by 1. The Chacon sequence (which is not exactly U but the image of U by a morphism) was introduced in [10]. The incidence matrix associated with the primitive Chacon substitution has three integer simple eigenvalues:

$$\{3, 1, 0\}.$$

We thus obtain $\theta_2=1$ and $\alpha_2=0$. In particular, θ_2 is a root of unity and we have thus to consider the constant $A_{\sigma,U}$. Using the algorithm described in Appendix B, we can show that $A_{\sigma,U}$ is not equal to zero, which implies

$$B_U = (O \cap \Omega)(\log(N)).$$

Substitutive Rote sequence: Let us consider the quadratic number $\alpha=(\sqrt{3}-1)/2$. We defined the sequence $U=(u_n)_{n \in \mathbb{N}}$ by

$$u_n = \begin{cases} 1 & \text{if } \{n\alpha\} \in [0, \frac{1}{2}[\\ 2 & \text{else.} \end{cases}$$

This sequence is called coding of rotation of parameters $(\alpha, \frac{1}{2})$ (see for instance [2,4,15]) and is included in the class of sequences of complexity $2n$ considered in [35]. It is shown in [33] that

$$U = \phi(X_\sigma),$$

where ϕ is the morphism defined by

$$\begin{aligned} \{1, 2, 3\} &\rightarrow \{1, 2\} \\ 1 &\mapsto 1 \\ 2 &\mapsto 122 \\ 3 &\mapsto 12 \end{aligned}$$

and X_σ is the fixed point of the substitution σ defined by

$$\begin{aligned} \{1, 2, 3\} &\rightarrow \{1, 2, 3\} \\ 1 &\mapsto 13 \\ 2 &\mapsto 13223 \\ 3 &\mapsto 1323 \end{aligned}$$

The incidence matrix associated with σ admits three simple eigenvalues:

$$\theta = 2 + \sqrt{3}, \quad 1 \quad \text{and} \quad \frac{1}{\theta} = 2 - \sqrt{3}.$$

We thus obtain $\theta_2 = 1$ and $\alpha_2 = 0$. In particular, θ_2 is a root of unity and we have thus to consider the constant $A_{\sigma,U}$. Using the algorithm described in Appendix B, we can show again that $A_{\sigma,U}$ is not equal to zero, which gives

$$B_{X_\sigma}(N) = (O \cap \Omega)(\log(N)).$$

This implies that

$$B_U(N) = O(\log(N)),$$

but not necessarily

$$B_U(N) = \Omega(\log(N)).$$

However, we show in [3] that

$$D_N(A, U) = \Omega(\log(N)),$$

where A means the uniform probability vector on $\{1, 2\}$. Proposition 7 implies finally

$$B_U(N) = \Omega(\log(N)).$$

More generally, it is shown in [2] that if U means the coding of rotation of parameters (α, β) , where α is a quadratic number and β lies in the quadratic extension of α , then

$$B_U(N) = (O \cap \Omega)(\log(N)).$$

The method used here shows that Theorem 13 could sometimes be extended to the study of substitutive sequences (which are not necessarily fixed points of substitutions).

An example with multiplicity: Let us consider the substitution σ defined over the alphabet $\{1, 2, 3, 4, 5\}$ by

$$\begin{aligned} 1 &\mapsto 1112455 \\ 2 &\mapsto 111255 \\ 3 &\mapsto 1123455 \\ 4 &\mapsto 23445 \\ 5 &\mapsto 123455 \end{aligned}$$

and let us denote by U the fixed point of this substitution. The incidence matrix associated with σ admits the following characteristic polynomial:

$$(x-1)^2(x^3 - 7x^2 + 5x - 1).$$

The Perron eigenvalue of σ is a Pisot number whose minimal polynomial is the factor of degree three in the previous expression. We thus obtain $\Theta_2 = 1$. Moreover, the minimal polynomial of M_σ is equal to its characteristic one and then $\alpha_2 = 1$. The constant $A_{\sigma,U}$ being not equal to zero, it follows from case (iv) of Theorem 13 that

$$B_U(N) = (O \cap \Omega)(\log^2(N)).$$

Two degenerated examples: We call degenerated substitution, a substitution which is not a Pisot-type substitution but which as well generates a fixed point whose balance function is bounded.

As we have already noted in Section 4, the periodic fixed point U of the substitution σ

$$\begin{aligned} 1 &\mapsto 121 \\ 2 &\mapsto 212 \end{aligned}$$

has a bounded balance function. The incidence matrix associated with σ admits yet 3 and 1 as eigenvalues. It thus follows that $\theta_2=1$, $\alpha_2=0$ and $A_{\sigma,U}=0$.

In Section 5, we introduce σ_2 , the Morse substitution of order two, defined by

$$\begin{aligned} 1 &\mapsto 14 \\ 2 &\mapsto 31 \\ 3 &\mapsto 34 \\ 4 &\mapsto 31 \end{aligned}$$

and U its fixed point beginning with 1. This substitution is degenerated, we have $\theta_2=\pm 1$, $\alpha_2=0$ and $A_{\sigma,U}=0$. It thus follows that U is a C -balanced sequence. Then, it is noticeable to see that if we consider the substitution σ' defined by

$$\begin{aligned} 1 &\mapsto 12 \\ 2 &\mapsto 13 \\ 3 &\mapsto 34 \\ 4 &\mapsto 13 \end{aligned}$$

and U' its fixed point beginning with 1, then the sequence U' satisfies

$$B_{U'}(N) = (O \cap \Omega)(\log N),$$

although both substitutions σ_2 and σ'_2 share the same incidence matrix.

Acknowledgements

I like to express my gratitude to Julien Cassaigne for many useful discussions during the preparation of this paper.

Appendix A.

Let us consider a primitive substitution σ defined over the finite alphabet $\mathcal{A}=\{1,2,\dots,d\}$, and let us suppose that U is a fixed point for σ , generated by the letter 1. Moreover, we assume that σ is a Pisot-type substitution, that is to say, $|\theta_2|<1$. Then, Theorem 13 states that $B_U(N)$ is bounded or equivalently that U is C -balanced for some constant C . The object of this appendix is to exhibit such a constant C , that is to say, to give an explicit upper bound for $B_U(N)$. We will use freely in the following the different definitions and notation introduced in Section 2.

If $f=(f(i))_{i \in \mathcal{A}} \in \mathbb{C}^d$ and $N \in \mathbb{N}^*$, then we define:

$$S_U^f(N) = \sum_{i=1}^d |u_0 u_1 \dots u_{N-1}|_i f(i).$$

Just as, if $\omega \in \mathcal{A}^*$, we define:

$$S^f(\omega) = \sum_{i=1}^d |\omega|_i f(i).$$

For $1 \leq i \leq d-1$, we introduce the vectors f_i , defined by

$$f_i(j) = \begin{cases} 1 & \text{if } j = i, \\ \frac{\lambda_i}{\lambda_i - 1} & \text{else,} \end{cases}$$

where $\lambda = (\lambda_i)_{i \in \mathcal{A}}$ is the normalized eigenvector associated with θ , the Perron eigenvalue of M_σ . We can note that f_i is well-defined because $0 < \lambda_i < 1$ (it comes from the Perron–Frobenius theorem). Then, it is relatively easy to see that

$$D_N(\lambda, U) = \max_{i=1,2,\dots,d-1} (1 - \lambda_i) |S_U^{f_i}(N)|,$$

and thus in view of Remark 8,

$$B_U(N) \leq \left\lceil 4 \max_{i=1,2,\dots,d-1} (1 - \lambda_i) |S_U^{f_i}(N)| \right\rceil, \quad (\text{A.1})$$

where $\lfloor x \rfloor$ denotes the integer part of the real x .

Now, for any word $m \in \mathcal{A}^*$, let us introduce the vector $L(m) = (|m|_i)_{i \in \mathcal{A}}$. Then, we have

$$L(\sigma(m)) = M_\sigma(L(m)), \quad (\text{A.2})$$

where M_σ denotes the incidence matrix of σ . In this way, if i and j are fixed in \mathcal{A} , the sequence $(|\sigma^n(j)|_i)_{n \in \mathbb{N}}$ satisfies a linear recurrence whose coefficients are those of the minimal polynomial of M_σ . Therefore, there exist complex numbers $\lambda_{i,j}^{k,l}$ and $\lambda_{i,j}$ such that for every $n \in \mathbb{N}$, we have

$$|\sigma^n(j)|_i = \lambda_{i,j} \theta^n + \sum_{k=2}^{d'} \left(\sum_{l=0}^{\alpha_k} \lambda_{i,j}^{k,l} n^l \theta_k^n \right). \quad (\text{A.3})$$

Let us notice that Eqs. (A.2) and (A.3) imply that, for each letter j , the vector $(\lambda_{i,j})_{i \in \mathcal{A}}$ is an eigenvector of M_σ associated with the Perron eigenvalue θ . Thus, there exists a complex number ε_j such that $\lambda_{i,j} = \varepsilon_j \lambda_j$. Then, for any vector $f = (f(i))_{i=1,2,\dots,d} \in \mathbb{C}^d$ lying in the orthogonal vector space of λ , it follows:

$$\begin{aligned} S^f(\sigma^n(j)) &= \sum_{i=1}^d |\sigma^n(j)|_i f(i) \\ &= \underbrace{\left(\sum_{i=1}^d \lambda_{i,j} f(i) \right)}_0 \theta^n + \sum_{k=2}^{d'} \left(\sum_{l=0}^{\alpha_k} \left(\sum_{i=1}^d \lambda_{i,j}^{k,l} f(i) \right) n^l \theta_k^n \right). \end{aligned} \quad (\text{A.4})$$

In order to make the following more friendly readable, let us introduce, for any word $\omega = \omega_1 \omega_2 \dots \omega_m$ defined over \mathcal{A} and any such a vector f , the notation:

$$F_{f,k,l}(\omega) = \sum_{j=1}^m \left(\sum_{i=1}^d \lambda_{i,\omega_j}^{k,l} f(i) \right). \quad (\text{A.5})$$

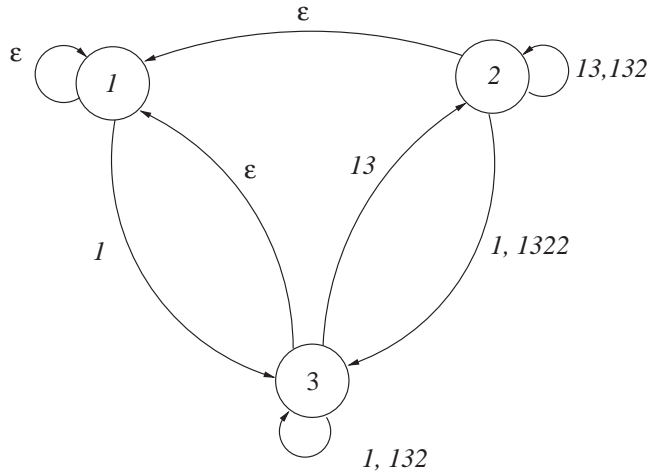


Fig. 1. Example of a prefix automaton in the case of the substitution $1 \mapsto 13, 2 \mapsto 13223, 3 \mapsto 123$.

We present now a generalized numeration system associated with a substitution, introduced simultaneously by Dumont and Thomas [17], and Rauzy [34].

Definition 25. Let σ be a substitution and let us suppose that U is a fixed point for σ generated by the letter 1. The subset of \mathcal{A}^* composed by the proper prefixes of the images by σ of the letters will be denoted by $Pref_\sigma$. The prefix automaton associated with the pair (σ, U) is defined in the following way:

- \mathcal{A} is the set of states,
- $Pref_\sigma$ is the set of labels,
- there is a transition from the state i to the state j labelled by the word m if mj is a prefix of $\sigma(i)$ (Fig. 1).

An *admissible labelled path* \mathcal{C} in the prefix automaton associated with a pair (σ, U) will be denoted by

$$((i_0, i_1, E_0), (i_1, i_2, E_1), \dots, (i_{n-1}, i_n, E_{n-1})),$$

$i_j \in \mathcal{A}$ for $0 \leq j \leq n$, $E_j \in Pref_\sigma$ for $0 \leq j \leq n - 1$. The positive integer n is the length of the path.

The main theorem concerning the prefix automaton is the following.

Theorem 26 (Dumont and Thomas [17], Rauzy [34]). *With the previous notation, if U is a fixed point generating by the letter 1 of the substitution σ , then we have:*

- (i) *for every positive integers N , there exists a unique admissible path in the prefix automaton associated with the pair (σ, U) , starting from 1 and labelled by the*

sequence $(E_0, E_1 \dots E_n)$, such that $E_0 \neq \varepsilon$ and

$$U_N = \sigma^n(E_0)\sigma^{n-1}(E_1) \dots E_n,$$

where U_N denotes the prefix of U of the length N .

(ii) Conversely, to any such a path, there corresponds a unique prefix of U , given by the above formula.

Let us consider a positive integer N . Following Theorem 26, there exists a unique admissible path in the prefix automaton associated with the pair (σ, U) , starting from 1 and labelled by the sequence $(E_0, E_1, \dots, E_{n_N})$, $E_0 \neq \varepsilon$, such that:

$$U_N = \sigma^{n_N}(E_0)\sigma^{n_N-1}(E_1) \dots E_{n_N}.$$

We thus obtain that for any vector $f \in \mathbb{C}^d$,

$$S_U^f(N) = \sum_{m=0}^{n_N} S_U^f(\sigma^k(E_{n_N-m})),$$

and in view of equalities (A.4) and (A.5),

$$S_U^f(N) = \sum_{m=0}^{n_N} \left(\sum_{k=2}^{d'} \sum_{l=0}^{\alpha_k} F_{f,k,l}(E_{n_N-m}) m^l \theta_k^m \right).$$

We can thus consider the finite quantity

$$M_{\sigma,U} = \max_{1 \leq i < d} \max_{2 \leq k \leq d'} \max_{0 \leq l \leq \alpha_k} \max_{E \in \text{Pref}_\sigma} (1 - A_i) |F_{f,i,k,l}(E)|.$$

Then, it follows from (A.1) that

$$B_U(N) \leq \left[4M_{\sigma,U} \sum_{k=2}^{d'} \sum_{l=0}^{\alpha_k} \left(\sum_{m=0}^{+\infty} m^l |\theta_k|^m \right) \right],$$

and thus

$$B_U(N) \leq \left[4M_{\sigma,U}(d-1) \sum_{l=0}^{d-1} \left(\sum_{m=0}^{+\infty} m^l |\theta_2|^m \right) \right],$$

which implies that

$$B_U(N) \leq \left[4M_{\sigma,U}(d-1) \sum_{l=0}^{d-1} \frac{l!}{(1 - |\theta_2|)^{l+1}} \right], \quad (\text{A.6})$$

because for any real $0 \leq x < 1$,

$$\sum_{n=0}^{+\infty} n^l x^n \leq \sum_{n=0}^{+\infty} (n+l)(n+l-1) \dots (n+1)x^n = \frac{l!}{(1-x)^{l+1}}.$$

The last upper bound, given in (A.6), does not depend on N and provides thus a closed formula to find an explicit upper bound for the balance function associated with a fixed

point of a Pisot-type substitution. This upper bound is certainly not optimal, because we wanted to exhibit a general formula. However, for any given substitution of Pisot type, the method evoked above could be used to find a really more precise result.

Appendix B.

The object of this appendix is to give a definition of the complex number $A_{\sigma,U}$ used in Theorem 13 and Corollary 15. The meaning of $A_{\sigma,U}$ is strongly connected with the notion of elementary loop in the prefix automaton associated with the pair (σ, U) . We recall now the definition of an elementary loop, for a definition of the prefix automaton associated with a substitution the reader is referred to Appendix A.

Definition 27. Let σ be a substitution and let us suppose that U is a fixed point for σ generated by the letter 1. We call elementary loop any admissible labelled path $((i_0, i_1, E_0), \dots, (i_{n-1}, i_n, E_{n-1}))$ in the prefix automaton associated with the pair (σ, U) , satisfying the following conditions:

- $i_0 = i_n$,
- $\forall 0 \leq j < k < n, i_k \neq i_j$.

We will denote by $\mathcal{E}l(\sigma, U)$ the set composed by all the elementary loops in the prefix automaton associated with the pair (σ, U) .

Remark 28. Since \mathcal{A} and $Pref_\sigma$ are finite sets, $\mathcal{E}l(\sigma, U)$ is finite too.

Let us consider a primitive substitution σ defined over the finite alphabet $\mathcal{A} = \{1, 2, \dots, d\}$, and let us suppose that U is a fixed point for σ generated by the letter 1. Moreover, we first assume that σ satisfies $\theta_2 = 1$, whatever the way one orders the spectrum of M_σ (satisfying of course (1)). For every admissible labelled path in the prefix automaton $\mathcal{C} = ((i_0, i_1, E_0), \dots, (i_{n-1}, i_n, E_{n-1}))$ we introduce:

$$F_{f,2,\alpha_2}(\mathcal{C}) = \sum_{j=0}^{n-1} F_{f,2,\alpha_2}(E_j),$$

where $F_{f,2,\alpha_2}(E_j)$ is defined following Equality (13) given in Appendix A. Then, we can consider for any vector $f \in \mathbb{C}^d$, the quantity

$$A_{\sigma,U}^f = \max\{|F_{f,2,\alpha_2}(B)|, B \in \mathcal{E}l(\sigma, U)\}.$$

For $1 \leq i \leq d - 1$, we consider, as in Appendix A, the vectors f_i , defined by

$$f_i(j) = \begin{cases} 1 & \text{if } j = i, \\ \frac{A_i}{A_i - 1} & \text{else.} \end{cases}$$

The family of vectors f_i provides a canonical base of the orthogonal vector space of the eigenvector λ . Then, we can define a complex number, denoted by $A_{\sigma,U}$, just depending on the pair (σ, U) , by

$$A_{\sigma,U} = \max\{A_{\sigma,U}^{f_j}, 1 \leq j < d\}.$$

Now, let us assume that σ is as in case (iv) of Theorem 13, that is to say θ_2 is a root of unity. This implies (see (1)) that all the eigenvalues of M_σ whose modulus is equal to one and whose multiplicity is equal to α_2 are roots of unity. Thus, there exists a minimal integer n_0 such that σ^{n_0} satisfies the condition required at the beginning of this appendix, that is to say:

for all the ways of ordering the spectrum of $M_{\sigma^{n_0}}$, $\theta_2 = 1$. Then, we can in this case associate a complex number with the pair (σ, U) by putting

$$A_{\sigma, U} = A_{\sigma^{n_0}, U},$$

where $A_{\sigma^{n_0}, U}$ is obtained following the previous construction. We thus have described a way to compute $A_{\sigma, U}$, as soon as the pair (σ, U) satisfies conditions (iv) of Theorem 13.

Note: Since this paper was written, the author extended Theorem 5 mainly using the notion of return words and the fact proved in [18] that a fixed point of a primitive substitution admits only a finite number of derivative sequences. Then, defining a new generalized balance function by $\mathcal{B}_N(U) = \sup_{u \in \mathcal{L}(U)} \max_{w, w' \in \mathcal{L}_N(U)} \{|w|_u - |w'|_u\}$ (note that $\mathcal{B}_N(U) = \sup_{n \in \mathbb{N}^*} B_N^{(n)}$), an analog of Theorem 22 could be obtained for $\mathcal{B}_N(U)$.

References

- [1] B. Adamczewski, Symbolic discrepancy and self-similar dynamics, preprint, 2002.
- [2] B. Adamczewski, Codages de rotations et phénomènes d'autosimilarités, *J. Théor. Nombres Bordeaux* 14 (2002) 351–386.
- [3] B. Adamczewski, Répartitions des suites $(n\alpha)_{n \in \mathbb{N}}$ et substitutions, *Acta Arith.*, to appear.
- [4] P. Alessandri, V. Berthé, Three distance theorems and combinatorics on words, *Enseign. Math.* 44 (2) (1998) 103–132.
- [5] E. Altman, B. Gaujal, A. Hordijk, Balanced sequences and optimal routing, *J. ACM* 47 (2000) 752–775.
- [6] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité $2n + 1$, *Bull. Soc. Math. France* 119 (1991) 199–215.
- [7] V. Berthé, R. Tijdeman, Balance properties of multi-dimensional words, *Theoret. Comput. Sci.* 273 (2002) 197–224.
- [8] V. Canterini, A. Siegel, Geometric representation of substitutions of Pisot type, *Trans. Amer. Math. Soc.* 353 (2001) 5121–5144 (electronic).
- [9] J. Cassaigne, S. Ferenczi, L.Q. Zamboni, Imbalances in Arnoux–Rauzy sequences, *Ann. Inst. Fourier (Grenoble)* 50 (2000) 1265–1276.
- [10] R.V. Chacon, Weakly mixing transformations which are not strongly mixing, *Proc. Amer. Math. Soc.* 22 (1969) 559–562.
- [11] J. Coquet, A summation formula related to the binary digits, *Invent. Math.* 73 (1983) 107–115.
- [12] E.M. Coven, G.A. Hedlund, Sequences with minimal block growth, *Math. Systems Theory* 7 (1973) 138–153.
- [13] D. Crisp, W. Moran, A. Pollington, P. Shiue, Substitution invariant cutting sequences, *J. Théor. Nombres Bordeaux* 5 (1993) 123–137.
- [14] F.M. Dekking, On the distribution of digits in arithmetic sequences, in: *Seminar on Number Theory, 1982–1983* (Talence, 1982/1983), pages Exp. No. 32, 12. Univ. Bordeaux I, Talence, 1983.
- [15] G. Didier, Codages de rotations et fractions continues, *J. Number Theory* 71 (1998) 275–306.
- [16] M. Drmota, R.F. Tichy, *Sequences, Discrepancies and Applications*, Springer, Berlin, 1997.
- [17] J.-M. Dumont, A. Thomas, Systèmes de numération et fonctions fractales relatifs aux substitutions, *Theoret. Comput. Sci.* 65 (1989) 153–169.

- [18] F. Durand, A characterization of substitutive sequences using return words, *Discrete Math.* 179 (1998) 89–101.
- [19] F. Durand, Linearly recurrent subshifts have a finite number of non-periodic subshift factors, *Ergodic Theory Dynam. Systems* 20 (2000) 1061–1078.
- [20] F. Durand, B. Host, C. Skau, Substitutional dynamical systems, Bratteli diagrams and dimension groups, *Ergodic Theory Dynam. Systems* 19 (1999) 953–993.
- [21] I. Fagnot, L. Vuillon, Generalized balances in Sturmian words, *Discrete Appl. Math.* 121 (2002) 83–101.
- [22] B. Gaujal, Convexité discrète et régularité; application au contrôle des systèmes à évènements discrets, Habilitation à diriger des recherches, LORIA et INRIA, 2001.
- [23] R.L. Graham, Covering the positive integers by disjoint sets of the form $\{[n\alpha + \beta]; n = 1, 2, \dots\}$, *J. Combin. Theory Ser. A* 15 (1973) 354–358.
- [24] C. Holton, L.Q. Zamboni, Geometric realizations of substitutions, *Bull. Soc. Math. France* 126 (1998) 149–179.
- [25] P. Hubert, Suites équilibrées, *Theoret. Comput. Sci.* 242 (2000) 91–108.
- [26] L. Kuipers, H. Niederreiter, *Uniform Distribution of Sequences*, Wiley-Interscience Wiley, New York, 1974, *Pure Appl. Math.*
- [27] R. Morikawa, On eventually covering families generated by the bracket function, *Bull. Fac. Liberal Arts Nagasaki Univ.* 23 (1982/83) 17–22.
- [28] M. Morse, G.A. Hedlund, Symbolic dynamics, *Amer. J. Math.* 60 (1938) 815–866.
- [29] M. Morse, G.A. Hedlund, Symbolic dynamics II. Sturmian trajectories, *Amer. J. Math.* 62 (1940) 1–42.
- [30] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, in: *Automata, Languages and Programming* (Antwerp, 1984), Springer, Berlin, 1984, pp. 380–389.
- [31] M. Queffélec, *Substitution Dynamical Systems—Spectral Analysis*, in: *Lecture Notes in Mathematics*, Vol. 1294, Springer, Berlin, 1987.
- [32] G. Rauzy, Nombres algébriques et substitutions, *Bull. Soc. Math. France* 110 (1982) 147–178.
- [33] G. Rauzy, Des mots en arithmétique, in: *Avignon Conference on Language Theory and Algorithmic Complexity* (Avignon, 1983), Univ. Claude-Bernard, Lyon, 1984, pp. 103–113.
- [34] G. Rauzy, Sequences defined by iterated morphisms, in: *Sequences* (Naples/Positano, 1988), Springer, New York, 1990, pp. 275–286.
- [35] G. Rote, Sequences with subword complexity $2n$, *J. Number Theory* 46 (1994) 196–213.
- [36] W. Rudin, Some theorems on Fourier coefficients, *Proc. Amer. Math. Soc.* 10 (1959) 855–859.
- [37] H. Shapiro, Extremal problems for polynomials and power series, Ph.D. Thesis, Massachusetts Institute of Technology, 1951.
- [38] R. Tijdeman, Exact covers of balanced sequences and Fraenkel’s conjecture, in: *Algebraic Number Theory and Diophantine Analysis* (Graz, 1998), de Gruyter, Berlin, 2000, pp. 467–483.
- [39] R. Tijdeman, Fraenkel’s conjecture for six sequences, *Discrete Math.* 222 (2000) 223–234.