RAIRO. INFORMATIQUE THÉORIQUE ET APPLICATIONS

JACQUES JUSTIN
GIUSEPPE PIRILLO

Decimations and sturmian words

RAIRO. Informatique théorique et applications, tome 31, nº 3 (1997), p. 271-290
<http://www.numdam.org/item?id=ITA_1997__31_3_271_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 31, n° 3, 1997, pp. 271-290)

DECIMATIONS AND STURMIAN WORDS (*)

by Jacques Justin ([1]) and Giuseppe Pirillo ([2])

Abstract. – *Standard Sturmian infinite words have a curious property discovered by G. Rauzy. If in such a word we delete all occurrences of each letter, except every pth one, then we get the same infinite word. This property and several generalizations are studied here. In the last part a short and self-contained theory of Sturmian words, using only combinatorial arguments, is presented.*

Résumé. – *Les mots Sturmiens standard infinis ont une curieuse propriété découverte par G. Rauzy. Si dans un tel mot on supprime toutes les occurrences de chaque lettre sauf celles de rang multiple de p, alors on retrouve le mot infini initial. On étudie cette propriété et plusieurs généralisations. Dans la dernière partie on donne une présentation courte et autocontenue de la théorie des mots Sturmiens qui n'utilise que des arguments combinatoires.*

INTRODUCTION

Infinite Sturmian words have been studied under various names for a long time (*see* [2, 4, 9] for historical notes). They can be defined either algebraically or by combinatorial properties of their factors (the equivalence is proved in [10]). The first way is probably more powerful as it allows for instance to make use of the properties of continued fractions. However the theory can be constructed without algebra, using only combinatorics of words. This is sketched in part 4 where we give a short self-contained presentation of known results and proofs, mostly from [5], [10] and cited papers.

In order to illustrate the power of the algebraic way, we consider in Part 2 a curious property of standard Sturmian words observed by Rauzy [15]. Given any positive integer p, if, in an infinite standard Sturmian word on $\{a, b\}$, we

(*) Received February 1996, accepted June 1997.
([1]) Laboratoire d'informatique algorithmique, fondements et applications, 2, place Jussieu, 75251 Paris Cedex 05; Mailing address: J. Justin, 19, rue de Bagneux, 92330 Sceaux, France.
([2]) IAMI CNR, Viale Morgagni 67/A, 50134 Firenze, Italy, and Institut Gaspard Monge, Bâtiment IFI, Université de Marne-la-Vallée, 2, rue de la Butte Verte, 93160 Noisy-le-Grand, E-mail: pirillo@udini.math.unifi.it

keep every pth a and delete all other a, and similarly for b, then we obtain the same infinite word. This follows immediately from the description of a Sturmian word by the intersections of a line with the lines of a square grid [6, 16], description which is equivalent to the algebraic definition [6]. Rauzy also suggested in his paper that the converse is true: an infinite word invariant under all such "decimations" is standard Sturmian, and even it suffices that the word be invariant under two decimations, modulo p and q say, with p and q multiplicatively independent. We give a complete proof of the converse (Theorem 1) and show that the weaker hypothesis is not sufficient. Indeed an infinite word satisfying the weaker hypothesis either is Sturmian or can be deduced from a "periodic Sturmian" word by a suitable construction (Theorem 2). Some extensions are also given (Theorem 3 and Part 3).

1. PRELIMINARIES

1.1. Words [14, chap. 1]

Throughout this paper, $A = \{a, b\}$ will be a two-letter *alphabet*. The *free monoid* A^* generated by A is the set of the (finite) *words* on A. If $u = u(1)u(2)\cdots u(m)$, $u(i) \in A$, is a word, its *length* is $|u| = m$. Also $|u|_a$ (resp. $|u|_b$) is the number of occurrences of a (resp. b) in u. Last $\widetilde{u}$ will denote the *reversal* of u, *i.e.* the word $u(m)\cdots u(2)u(1)$. A word equal to its reversal is a *palindrome*.

In the same way an *infinite word* is a function $s : \mathbb{N}_+ \to A$ where $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$ is the set of positive integers. It is written $s = s(1)s(2)\cdots s(i)\cdots$, $s(i) \in A$. The set of infinite word is A^ω. For a finite or infinite word t, the *factor* $t(i)t(i+1)\cdots t(j)$ of t will be denoted by $t(i,j)$. If $i = 1$, this is a *left* factor. *Right* factors of finite words are defined symmetrically.

An infinite word s is *periodic* if for some $p \in \mathbb{N}_+$ and for all $i \in \mathbb{N}_+$ we have $s(i+p) = s(i)$. In this case s can be written $s = uu\cdots = u^\omega$ for some $u \in A^*$, $|u| = p$. The smallest integer p having this property is called the *period* of s. An infinite word t is *ultimately periodic* if it can be written $t = vs$ with $v \in A^*$ and $s \in A^\omega$, s periodic.

Remark 1: *Doubly infinite words*, which are not used here, are defined in the same way as functions from $\mathbb{Z}$ to A.

1.2. Infinite Sturmian words [3]

DEFINITION 1: *Let ρ, α be real numbers with $\rho \in [0, 1[$ and $\alpha \in [0, 1]$. A Sturmian word (in the wide sense) is an infinite word s given*

a) either by

$$s(n) = \begin{cases} a & \text{if } \lfloor \rho + (n+1)\alpha \rfloor - \lfloor \rho + n\alpha \rfloor = 0 \\ b & \text{if not} \end{cases}$$

b) or by

$$s(n) = \begin{cases} a & \text{if } \lceil \rho + (n+1)\alpha \rceil - \lceil \rho + n\alpha \rceil = 0 \\ b & \text{if not.} \end{cases}$$

By Sturmian word it is often (but not always) meant that α is irrational. In this case we shall say *proper Sturmian.* When α is rational, then s is periodic and will be called here *periodic Sturmian.* In particular when $\alpha = 0$ (resp. $\alpha = 1$) we get the word $aaa \cdots = a^\omega$ (resp. $bbb \cdots = b^\omega$).

DEFINITION 2: *A Sturmian word is standard (or is the characteristic sequence of α) when $\rho = 0$ in Definition 1.*

1.3. Cutting sequences [6, 16]

Let Oxy be a cartesian coordinates system for the Euclidean plane. Construct the grid consisting of lines $H_j : y = j$ and lines $V_i : x = i$, for all $i, j \in \mathbb{N}_+$. Its vertices are $M_{ij} = V_i \cap H_j$. Consider a line $D : y = \beta x$ with $0 < \beta < \infty$. Label its intersections with lines H_j by a and with lines V_i by b. If D meets H_j and V_i at the same point M_{ij}, we label it by ab or ba with the convention that all such points are labelled in the same way (this the case if and only if β is rational). The sequence of the labels (the "cutting sequence") when x grows from 0 to infinity is an infinite word s. This one is standard Sturmian, and all standard Sturmian words, except a^ω and b^ω, can be obtained in this way. This definition is equivalent to Definition 2 with $\alpha = 1/(\beta + 1)$.

More generally if we construct in the same way the cutting sequence for any line with positive slope we get a Sturmian word, and all Sturmian words given by parallel lines have the same set of factors because they correspond to the same α of Definition 1.

1.4. Now, with the same grid as in **1.3** consider a strictly increasing continuous function $f : [0, \infty[\to [0, \infty[$ such that $f(0) = 0$ and $f(i)$ is not an integer for all $i \in \mathbb{N}_+$, and let C_f be its representative curve. If we label the intersections of C_f with lines H_j and V_i as in **1.3** we get an infinite word t. We say that f, or C_f defines t or is a defining function or curve for t. For instance when β is irrational the half-line $y = \beta x$, $x \geq 0$, defines a standard proper Sturmian word.

1.5. Decimations

The kind of decimations considered here is as follows. Let $p \in \mathbb{N}_+$ and let s be any finite or infinite word. Number $1, 2, \cdots$ the successive occurrences of a in s and similarly for b. Delete all occurrences of a except those whose number is a multiple of p, and similarly for b. We get a finite or infinite word $t = \Delta_p(s)$. When s is finite, t is a shorter word (which may be the *empty* word); when s is infinite, t is infinite. If $t = s$, then s is invariant under Δ_p. This *decimation modulo p* or *p-decimation* Δ_p is a transformation of $A^\omega \cup A^*$. The set $\{\Delta_p; p \in \mathbb{N}_+\}$ endowed with composition is a monoid isomorphic to the multiplicative structure of $\mathbb{N}_+$.

Remark 2: For $p, q \in \mathbb{N}_+$, $i, j \in \mathbb{N}$, $0 \leq i < p$, $0 \leq j < q$, we could also define a *generalized decimation* $\Delta_{(p,i,q,j)}$ by deleting all occurrences of a and b except those whose number is congruent respectively to i modulo p and to j modulo q.

Remark 3: Another perhaps more natural kind of decimation modulo p seems to have been considered here and there. When applied to s it gives $s(p)s(2p)s(3p)\cdots$. If s is Sturmian, the new word is not Sturmian in general but has some nice properties. As allusions to such decimations will occur in Remark 6 and at the end of the proof of Theorem 5 we shall call them here *blind decimations*.

Remark 4: The definitions of Sturmian words, cutting sequences and decimations may be easily extended to doubly infinite words.

2. DECIMATIONS OF STANDARD STURMIAN WORDS

2.1. The first theorem gives a characterization of standard Sturmian words

THEOREM 1: *An infinite word is invariant under all decimations if and only if it is standard Sturmian.*

Proof of the if part [15]: Let s be a standard Sturmian word. If $s = a^\omega$ or $s = b^\omega$ it is trivially invariant under decimations. If not, s can be obtained as the cutting sequence for some line $D : y = \beta x$ with $0 < \beta < \infty$. Now, given any $p \in \mathbb{N}_+$, if we delete in the grid all lines H_j except when $j \equiv 0 \bmod p$ and all lines V_i except when $i \equiv 0 \bmod p$, the cutting sequence of D with the new grid is exactly $\Delta_p(s)$. Now if we perform on D and the new grid an homothetic transformation with center O and ratio $1/p$ we get the initial configuration. So $\Delta_p(s) = s$.

Proof of the only if part: Now s is invariant under all decimations and we have to show that it is standard Sturmian. If s is a^ω or b^ω this is true. If not, a and b occur infinitely many times in s and we can write $s = u_1 u_2 u_3 \cdots$ with $u_i = a^{n_i} b$ and $n_i \geq 0$ for all $i \in \mathbb{N}_+$. Construct a defining function f for s as follows.

Let $\sigma_i = |u_1 u_2 \cdots u_i|_a = n_1 + n_2 + \cdots + n_i$, for $i \in \mathbb{N}_+$.

If $n_i > 0$ and $n_{i+1} > 0$ put $f(i) = \sigma_i + 1/2$.

If $n_i > 0$ and $n_{i+1} = 0$ put $f(i) = \sigma_i + 1/4$.

If $n_i = 0$ and $n_{i+1} > 0$ put $f(i) = \sigma_i + 3/4$.

Also put $f(0) = 0$. Let $0 = i_0 < i_1 < i_2 < \cdots$ be the sequence of the i such that $f(i)$ has been defined. Now linearly interpolate between O and point $(i_1, f(i_1))$, between points $(i_1, f(i_1))$ and $(i_2, f(i_2))$ and so on. This gives f and its representative curve C_f. Function f has the following property whose proof is postponed.

LEMMA 1: *The ratio $f(x)/x$ has a limit $\beta \in]0, \infty[$ when x approaches infinity.*

Now put $\theta = 1/(\beta + 1)$ and choose an integer $m > 1$. Let ϵ be a positive real which will be defined later. By Lemma 1 there exists a real ξ such that $(\beta - \epsilon)x < f(x) < (\beta + \epsilon)x$ for $x > \xi$. Then the curve C_f lies between lines $D' : y = (\beta - \epsilon)x$ and $D'' : y = (\beta + \epsilon)x$ for $x > \xi$. Now let p be an integer such that $p > \xi/\theta$. As s is invariant under p-decimation, the curve C_g image of C_f by the homothetic transformation with center O and ratio $1/p$ is also defining for s. This curve lies between D' and D'' for $x > \xi/p$, hence for $x \geq \theta$. For ϵ small enough we have $0 < (\beta - \epsilon)/(\beta + 1) < (\beta + \epsilon)/(\beta + 1) < 1$. So C_g, D', D'' and $D : y = \beta x$ do not cut the grid for $x \leq \theta$. Without loss of generality assume that $\beta - \epsilon$ and $\beta + \epsilon$ are irrational and consider two cases.

a) β is irrational. If ϵ is small enough, there is no vertex of the grid between D' and D'' for $0 < x \leq m$. So C_g, D', D'' and D have the same cutting sequence for $0 < x \leq m$, that is the left factor $u_1 u_2 \cdots u_m$ of s is a left factor of the standard proper Sturmian word defined by D. As m can be taken arbitrarily large, this one is s.

b) $\beta = c/d$, $c, d \in \mathbb{N}_+$, $(c, d) = 1$. Then $D : y = \beta x$ has two cutting sequences which are standard periodic Sturmian words $vxyvxyvxy \cdots = (vxy)^\omega$, with $xy = ab$ or $xy = ba$ and $|vxy|_a = c$, $|vxy|_b = d$, for some suitable $v \in A^*$. Reasoning as in case a) we see that C_g, D', D'' and D have the same cutting sequence for $x < m$ except the fact that when D passes through a vertex $M_{rd,rc}$, $r \in \mathbb{N}_+$, then the cutting sequence gives ba for line D' because D' crosses V_{rd} before H_{rc} while the cutting sequence for D'' gives ab. For C_g the corresponding letters may be ab or ba depending of the position of C_g near $M_{rd,rc}$. It follows that $s = vw_1vw_2 \cdots$ with $w_i \in \{ab, ba\}$ for $i \in \mathbb{N}_+$. But w_i corresponds to the (ic)th occurrence of a and the (id)th occurrence of b in s. So, as s is invariant by i-decimation, $w_i = w_1$ for $i \in \mathbb{N}_+$, so s is standard periodic Sturmian. $\diamond$

Proof of Lemma 1: Recall that $|u_1u_2 \cdots u_i|_a = \sigma_i$ and $|u_1u_2 \cdots u_i|_b = i$ and put $\sigma_i/i = \delta_i$ for $i \in \mathbb{N}_+$. Let p, q be any positive integers. As s is invariant by p-decimation, when we perform this one on the left factor $u_1u_2 \cdots u_{pq}$ of s we get $u_1u_2 \cdots u_q$. Consequently $\sigma_q = \lfloor \sigma_{pq}/p \rfloor$ that is $0 \leq \sigma_{pq} - p\sigma_q < p$, or $0 \leq \delta_{pq} - \delta_q < 1/q$. In the same way $0 \leq \delta_{pq} - \delta_p < 1/p$. Hence $-1/q < \delta_q - \delta_p < 1/p$. So the sequence of the δ_i, $i \in \mathbb{N}_+$, satisfies the Cauchy condition and has a limit, β say, $\beta < \infty$. Clearly also $\beta > 0$ because, as a occurs in s, we have $\delta_p > 0$ for some p and, by the inequality above, $\delta_{pq} \geq \delta_p$ for all q.

Now, as $1/4 \leq f(i) - \sigma_i \leq 3/4$, we have also, for i integer, $\lim\limits_{i \to \infty} f(i)/i = \beta$.

Last by the definition of f, $f(x)/x$ is homographic, hence monotonic, for $x \in [i - 1, i]$, $i \in \mathbb{N}_+$. So $\lim\limits_{x \to \infty} f(x)/x = \beta$. $\diamond$

Remark 5: The representation of Sturmian words by cutting sequences leads easily to the following generalization of the if part of Theorem 1.

a) If we perform a generalized decimation of the type $\Delta_{(p,i,p,j)}$, $p \in \mathbb{N}_+$, $0 \leq i, j < p$ on a (possibly doubly infinite) Sturmian word, the new word is Sturmian with the same set of factors. b) The set of all Sturmian words is globally invariant under any generalized decimation.

A noteworthy case of a) is the following. Let s be a standard Sturmian word and $s' = \Delta_{(2,0,2,1)}(s)$, $s'' = \Delta_{(2,1,2,0)}(s)$, $s''' = \Delta_{(2,1,2,1)}(s)$. Then if w' (resp. w'', w''') is any left factor of s' (resp s'', s''') then $\widetilde{w}'aw'$ (resp. $\widetilde{w}''bw''$, $\widetilde{w}'''w'''$) is a palindrome factor of s. This follows from the fact that if s is the cutting sequence of line D, then s' for example is the cutting sequence of the parallel D' to D passing through point (1/2,0). Consequently if we extend the grid to the whole plane, the symmetry of the figure with respect to point $(1/2, 0)$ shows that $\widetilde{w}'aw'$ is a factor of the cutting sequence of D' with the extended grid, hence is a factor of s.

2.2. Here we generalize Theorem 1 to the case where s is invariant by a subset of all decimations. Recall that two integers, x, y are *multiplicatively independent* if $x^\lambda = y^\mu$, $\lambda, \mu \in \mathbb{N}$ implies $\lambda = \mu = 0$.

Let H be a submonoid of the multiplicative structure of $\mathbb{N}_+$. We define an equivalence relation as follows

$$p \equiv q \bmod H \quad \text{if and only if} \quad pH \cap qH \neq \emptyset$$

We have

THEOREM 2: *Given a multiplicative submonoid H of $\mathbb{N}_+$ containing at least two multiplicatively independent elements, an infinite word s is invariant under at least all p-decimations such that $p \in H$ if and only if either s is standard Sturmian or it can be deduced from a standard periodic Sturmian word t, different from a^ω and b^ω, as follows. Let $t = (vab)^\omega$ with $|vab|$ the period of t. Then $s = vw_1vw_2vw_3\cdots$ with $w_i \in \{ab, ba\}$ for $i \in \mathbb{N}_+$ and, for all $i, j \in \mathbb{N}_+$, $i \equiv j \bmod H$ implies $w_i = w_j$.*

Proof of the if part: If s is standard Sturmian, it is invariant by all p-decimations, by Theorem 1. If s has the second form given in Theorem 2, then t is the cutting sequence of some line $D : y = \beta x$, with $\beta = c/d$, $(c, d) = 1$, $c, d \in \mathbb{N}_+$. Line D passes through all vertices $M_{di,ci}$, $i \in \mathbb{N}_+$. The first one being M_{dc} we have $|vab|_a = c$, $|vab|_b = d$. Give $M_{di,ci}$ the new label w_i, for $i \in \mathbb{N}_+$. Then the new cutting sequence is s. Now, given $p \in H$, if we delete all lines H_j except when $j \equiv 0 \bmod p$, and similarly for lines V_i, the remaining cutting sequence, that is $\Delta_p(s)$, will be equal to s if $M_{di,ci}$ and $M_{dip,cip}$ are labeled in the same way for all $i \in \mathbb{N}_+$ and this is true because $i \equiv ip \bmod H$.

Proof of the only if part: The proof follows that of the only if part of Theorem 1. Function f is defined in the same way and has the following property whose proof is postponed.

LEMMA 2: *We have $\lim_{x \to \infty} f(x)/x = \beta$ for some $\beta \in]0, \infty[$.*

If β is irrational, then we get that β is a standard proper Sturmian word as in case a) in the proof of Theorem 1.

If β is rational, we get that $s = vw_1vw_2\cdots$ with the notations in the statement of Theorem 2. Then let $i, j \in \mathbb{N}_+$ be such that $i \equiv j \bmod H$.

Then $ip = jp'$ for some $p, p' \in H$. As $\Delta_p(s) = s$ by hypothesis we must have $w_{ip} = w_i$ and similarly $w_{jp'} = w_j$, whence $w_i = w_j$ and this achieves the proof in the case β is rational.

Remark 6: Here is the simplest example of a non Sturmian word invariant by 2- and 3- decimation. Taking for v the empty word we put $s = w_1 w_2 w_3 \cdots$ with $w_i = ab$ if $i \in \{2^\lambda 3^\mu; \lambda, \mu \in \mathbb{N}\}$ and $w_i = ba$ if not. So $s = (ab)^4 baabba(ab)^2(ba)^2 ab(ba)^3 \cdots$. Note also that if we replace ab by x and ba by y we get an infinite word $xxxxyxy \cdots$ on $\{x, y\}$ which is invariant by the blind decimations (*see* Remark 3), modulo 2 and modulo 3.

Proof of Lemma 2: With the same notations as in the proof of Lemma 1, we show in the same way that the sequence (δ_h), $h \in H$ has a limit β. Now let $\epsilon > 0$ be arbitrary small and let p, q be two multiplicatively independent elements of H. As $\{p^\lambda q^\mu; \lambda, \mu \in \mathbb{Z}\}$ is dense in $\mathbb{R}_+$ there exist $e, g, h, k \in \mathbb{N}$ such that $1 < p^{-e} q^g < 1 + \epsilon$ and $1 < p^h q^{-k} < 1 + \epsilon$. Now, for any integer $i \geq p^e q^k$, let $m = p^x q^y$, $x, y \in \mathbb{N}$ be maximal such that $m \leq i$. We have $x \geq e$ or $y \geq k$. Suppose for instance $x \geq e$. Then $m < p^{x-e} q^{y+g} < m(1 + \epsilon)$.

Let $n = p^{x-e} q^{y+g}$. As m is maximal we have $m \leq i < n < m(1 + \epsilon)$. So, considering the left factors $u_1 u_2 \cdots u_m$ and so on of s we have

$$\sigma_m \leq \sigma_i \leq \sigma_n, \qquad \text{that is} \qquad m\delta_m \leq i\delta_i \leq n\delta_n,$$

whence $\delta_m/(1 + \epsilon) < \delta_i < (1 + \epsilon)\delta_n$. Consequently

$$\beta/(1 + \epsilon) \leq \lim\ \inf\ \delta_i \leq \lim\ \sup\ \delta_i \leq (1 + \epsilon)\beta.$$

As ϵ is arbitrarily small it follows $\lim_{i \to \infty} \delta_i = \beta$, whence $\lim_{x \to \infty} f(x)/x = \beta$.

2.3. It remains to consider the case where H, as defined in Theorem 2.2, does not contain multiplicatively independent elements. In this case the proof of Lemma 2 does not work and this property of f must be introduced as an hypothesis or, equivalently, s must have a *density*, with the following definition.

DEFINITION 3: *An infinite word s has density γ if $|s(1, m)|_a/m$ has limit γ as m approaches infinity.*

Then we have

THEOREM 3: *Let p be a positive integer and $H = \{p^\lambda; \lambda \in \mathbb{N}\}$. Then*

a) an infinite word is standard proper Sturmian if and only if it has an irrational density and is invariant under Δ_p;

b) an infinite word can be deduced from a standard periodic Sturmian word in the way indicated in Theorem 2 if and only if it has a rational density different from 0 and 1 and is invariant under Δ_p.

Remark 7: It is easy to construct an infinite word without density which is invariant under some p-decimation. For instance [15], with $p = 2$, this is the case of $aba^2b^2a^4b^4a^8b^8 \cdots$.

Remark 8: An example of case b) of Theorem 3 is the well-known Thue-Morse infinite word $abbabaab \cdots$ which is obtained from a by iterating the substitution: $a \longmapsto ab$, $b \longmapsto ba$. It is invariant under Δ_4 (not difficult to see), has density 1 and has the form given in the Theorem with v being the empty word.

3. FURTHER RESULTS

Now we study two extensions of Theorem 1. First we shall say that an infinite sequence (s_i) of infinite words s_i, $i \in \mathbb{N}_+$ converges towards the infinite word t if given any $m \in \mathbb{N}_+$ there exists i_0 such that, for all $i > i_0$, s_i and t have a common left factor of length at least m. We have then

THEOREM 4: *Let an infinite word s have density γ. Then*

a) if γ is irrational the sequence $(\Delta_p(s))_{p \in \mathbb{N}_+}$ converges towards the standard proper Sturmian word with density γ;

b) if $\gamma \in]0,1[$ is rational, let $(vab)^\omega$ be one of the two standard periodic words with density γ, then there exists an infinite word $t = vx_1y_1vx_2y_2 \cdots$ with $x_iy_i \in \{ab, ba\}$ and an infinite subsequence of $(\Delta_p(s))_{p \in \mathbb{N}_+}$ which converges towards t;

c) if $\gamma = 0$ (resp. $\gamma = 1$) the sequence $(\Delta_p(s))_{p \in \mathbb{N}_+}$ converges towards b^ω (resp. a^ω).

Proof: The proof follows that of the only if part of Theorem 1.

If $\gamma = 0$ (resp. $\gamma = 1$) then, clearly, the sequence of the $\Delta_p(s)$ converges towards b^ω (resp. a^ω). If not, we put $\beta = \gamma/(1-\gamma)$ and construct a defining function f for s as in the proof of the only if part of Theorem 1. As $\lim_{x \to \infty} f(x)/x = \beta$, reasoning in the same way when β is irrational, we get part a) of Theorem 4. When β is rational let $t' = (vab)^\omega$, where $|vab|$ is the period, one of the two standard periodic Sturmian words with cutting sequence $y = \beta x$. Then given any $m \in \mathbb{N}_+$ we get that for p large enough $\Delta_p(s)$ has a left factor of the form $vx_{p,1}y_{p,1}vx_{p,2}y_{p,2} \cdots$ with

$x_{p,i}y_{p,i} \in \{ab, ba\}$. Now for infinitely many values of p, $x_{p,1}y_{p,1}$ has the same value x_1y_1 and, among these values of p, for infinitely many, $x_{p,2}y_{p,2}$ has the same value x_2y_2. Continuing this way we get an infinite word $t = vx_1y_1vx_2y_2 \cdots x_iy_i \in \{ab, ba\}$ and a subsequence of $(\Delta_p(s))_{p\in\mathbb{N}_+}$ converging towards it. $\diamondsuit$

The proof of Theorem 5 will make use of a famous Theorem of van der Waerden in the following form [14, chap. 3].

THEOREM (van der Waerden): *Given a finite alphabet A and a positive integer n, there exists an integer m such that each word on A with length at least m contains an arithmetic cadence of order n, that is a factor of the form $xu_1xu_2 \cdots xu_{n-1}x$ for some $x \in A$ and $u_i \in A^*$ with $|u_1| = |u_2| = \cdots |u_{n-1}|$.*

Given a finite or infinite word s, denote by $\mathrm{Fac}(s)$ the set of its factors.

DEFINITION 4: *An infinite word s is uniformly recurrent if for any $u \in Fac(s)$ there exists an integer m such that $u \in Fac(v)$ whenever $v \in Fac(s)$ and $|v| \geq m$.*

We have then

THEOREM 5: *Let s be an infinite word. Then s is Sturmian if and only if it is uniformly recurrent, it has a density, and for each $p \in \mathbb{N}_+$, $\Delta_p(Fac(s)) \subset Fac(s)$.*

Proof of the only if part: It is well known (and easily deduced from Definition 1) that any Sturmian word is uniformly recurrent and has a density. It remains to show that when s is Sturmian, $\Delta_p(\mathrm{Fac}(s)) \subset \mathrm{Fac}(s)$. Let $u \in \mathrm{Fac}(s)$. Then for some i, u is a left factor of $t = s(i)s(i+1)s(i+2)\cdots$. Hence $\Delta_p(u)$ is a left factor of $\Delta_p(t)$. So $\Delta_p(u) \in \mathrm{Fac}(\Delta_p(t)$. Clearly t is Sturmian and $\mathrm{Fac}(t) = \mathrm{Fac}(s)$. Also, by a) of Remark 5, $\mathrm{Fac}(\Delta_p(t)) = \mathrm{Fac}(t)$. So $\Delta_p(u) \in \mathrm{Fac}(s)$, whence $\Delta_p(\mathrm{Fac}(s)) \subset \mathrm{Fac}(s)$.

Proof of the if part: Let γ be the density of s. If γ is irrational then, by Theorem 4, $(\Delta_p(s))_{p\in\mathbb{N}_+}$ converges towards the standard proper Sturmian word, t say, with density γ. Observe that for each p the left factors of $\Delta_p(s)$ belong to $\Delta_p(\mathrm{Fac}(s))$, hence to $\mathrm{Fac}(s)$. It follows that $\mathrm{Fac}(t) \subset \mathrm{Fac}(s)$. Now let $u \in \mathrm{Fac}(s)$. Then there exists m such that u occurs in every factor v of s with length at least m. Take for such v a factor of t. It follows that $u \in \mathrm{Fac}(v)$, hence $u \in \mathrm{Fac}(t)$, whence $\mathrm{Fac}(t) = \mathrm{Fac}(s)$. Consequently s is proper Sturmian.

If $\gamma \in]0,1[$ is rational, there exist (Theorem 4) an infinite word $t = vx_1y_1vx_2y_2\cdots$, $x_iy_i \in \{ab, ba\}$, and an infinite subset H of $\mathbb{N}_+$ such that $(\Delta_p(s))_{p\in H}$ converges towards t. As before, we show that $\mathrm{Fac}(t) = \mathrm{Fac}(s)$. Now put $vab = X, vba = Y$. Then t can be written $t = z_1z_2z_3\cdots$, $z_i \in \{X, Y\}$. Now, by the Theorem of van der Waerden, for any $n \geq 2$ there exists m such that any factor $z_i\cdots z_{i+m}$ of t contains a cadence of order n. This means that for any $i \in \mathbb{N}_+$ there exist $z \in \{X, Y\}$ and $j, r \in \mathbb{N}_+$ such that $i \leq j < j + (n-1)r \leq i + m$ and $z_j = z_{j+r} = \cdots = z_{j+(n-1)r} = z$.

Take $i \geq m$ and consider the factor $w = z_{j-r+1}z_{j-r+2}\cdots z_{j+(n-1)r}$ of t. By r-decimation of w we get

$$\Delta_r(w) = z_jz_{j+r}\cdots z_{j+(n-1)r} = z^n$$

(Indeed applying r-decimation to w is equivalent to applying blind decimation modulo r to w considered as a words on $\{X, Y\}$. This is a generalization of what was observed in Remark 6).

As $w \in \mathrm{Fac}(s)$ we have $\Delta_r(w) \in \mathrm{Fac}(s)$. So, as n is arbitrarily large, there are in s occurrences of arbitrary large powers of vab (or vba). As s is uniformly recurrent, this implies that $s = s_0(vab)^\omega$ for some right factor s_0 of vab. Consequently s is periodic Sturmian.

Last when $\gamma = 0$ (case $\gamma = 1$ is similar) $s = b^\omega$ because if a occurs in s, it occurs in each factor of length m for some $m \in \mathbb{N}_+$ because s is uniformly recurrent, whence $|s(1, km)| \geq k$, *i.e.* $|s(1, km)|/m \geq 1/m$ for all $k \in \mathbb{N}_+$ and $\gamma \geq 1/m$, a contradiction.

Remark 9: It is not possible to delete the hypothesis that s is uniformly recurrent in the if part of Theorem 5. Indeed let $w_1, w_2, \cdots$ be an enumeration of all the words and let t be an infinite word with arbitrary density γ, irrational for instance. Let $t = g_1g_2\cdots$, $g_i \in A^*$, be a factorization of t and put $s = g_1w_1g_2w_2\cdots$. Clearly if the $|g_i|$ increase with sufficient rapidity then s has density γ. Also $\mathrm{Fac}(s) = A^*$ whence, for all p, $\Delta_p(\mathrm{Fac}(s)) \subset \mathrm{Fac}(s)$. However, s is not Sturmian.

4. STURMIAN WORDS WITHOUT ALGEBRA

In order to present a complete theory it would be necessary to consider Sturmian words in the wide sense and also doubly infinite Sturmian words. This is done in details in [5, 10] but, as periodicity introduces some complications in the presentation, we limit ourselves, here, to proper Sturmian words.

Hereafter the alphabet remains $A = \{a, b\}$. If $y \in A$, we denote by $\overline{y}$ the other letter of A. For an infinite word s, $F_n(s)$ will denote the set $A^n \cap \mathrm{Fac}(s)$ of all its factors of length n. Recall [1].

DEFINITION 5: *A factor u of the infinite word s is special if ua and ub are factors of s.*

Clearly, the number of special factors of length n is $|F_{n+1}| - |F_n(s)|$. The following proposition is well known.

PROPOSITION 1: *Let s be an infinite word, then*

i) *if for some $m \in \mathbb{N}_+$, $|F_m(s)| \leq m$, then s is ultimately periodic with period $p \leq m$ and, for all $n \geq m$, $|F_n(s)| \leq m$;*

ii) *if s is periodic with period p then $|F_n(s)| > n$ for $n < p$ and $|F_n(s)| = p$ for $n \geq p$.*

Proof of i): Let q be minimal such that $|F_q(s)| \leq q$. Then $|F_{q-1}(s)| > q-1$ whence $|F_{q-1}(s)| = q$ and $|F_q(s)| = q$. If for some $n > q$ we have $|F_{n-1}(s)| = q$ and $|F_n(s)| > q$ then $|F_{n-1}(s)|$ must contain at least one special factor, x, say. But then the right factor of length $q-1$ of x is also a special factor. But $F_{q-1}(s)$ contains no special factor as $|F_{q-1}(s)| = |F_q(s)|$. Consequently, for all $n \geq q$ we have $|F_n(s)| = q$. Now, given any $n \geq q-1$, consider the factors $w = s(t, t+n-1)$ of s for $1 \leq t \leq q+1$. As $|F_n(s)| = q$, two of them must be equal, say $w_t = w_r$ for some $t < r$ in $[1, q+1]$. So $s(i) = s(i+r-t)$ for all $i \in [t, t+n-1]$. Now, for infinitely many values of n, the pair (t, r) is the same. Consequently for these values of t and r we have $s(i) = s(i+r-t)$ for all $i \geq t$, so s is ultimately periodic with period at most $r - t \leq m$.

Proof of ii):

As s is periodic with period p, then $s = u^\omega$ for some word u of length p. If for some $n < p$ we have $|F_n(s)| \leq n$, then s is (ultimately) periodic with period at most n, which is impossible. So $|F_n(s)| > n$ for $n < p$. Also, by part i), $|F_n(s)| = p$ for $n \geq p$. $\diamondsuit$

DEFINITION 6 [3]: *A finite or infinite word s is balanced if for all $u, v \in Fac(s)$, $|u| = |v|$ implies $||u|_a - |v|_a| \leq 1$.*

PROPOSITION 2: *If the infinite word s is balanced, then, for all $n \in \mathbb{N}_+$, $|F_n(s)| \leq n+1$.*

Proof: If the conclusion is false, let q be minimal such that $|F_q(s)| \neq q+1$. As $|F_0(s)| = 1$ we have $q > 0$. Also $|F_q(s)| \geq |F_{q-1}(s)| = q$. Suppose first that $|F_q(s)| \geq q+2$. Then $F_{q-1}(s)$ contains at least two special factors, u and v, say. So $q-1 \geq 1$. Put $u = xu'$, $v = yv'$, $x, y \in A$, $u', v' \in F_{q-2}(s)$. As u' and v' are special factors of s they are equal because $|F_{q-1}(s)| - |F_{q-2}(s)| = 1$. So $u = xu', v = yu'$, $x \neq y$. Hence $xu'x, yu'y \in F(s)$, that is s is not balanced contrarily to the hypothesis. So we must have $|F_q(s)| = q$ whence, by Proposition 1 and by the definition of q, $|F_n(s)| \leq n+1$ for all $n \in \mathbb{N}_+$. $\diamondsuit$

Proposition 3 [5]: *For an infinite word s the following conditions are equivalent*

i) *s has exactly one special factor of length n, for all $n \in \mathbb{N}_+$;*

ii) $|F_n(s)| = n+1$, *for all $n \in \mathbb{N}_+$;*

iii) *s is balanced and is not ultimately periodic.*

Proof: i) $\Longleftrightarrow$ ii) is trivial as the number of special factors of length n is $|F_{n+1}(s)| - |F_n(s)|$.

i) $\Longrightarrow$ iii). Suppose by contradiction that s is ultimately periodic, say $s = ut$ with $u \in A^*$ and t infinite periodic. We have $|F_n(s)| \leq |u| + |F_n(t)|$ for all $n \in \mathbb{N}_+$. So, by ii) of Proposition 1, $|F_n(s)|$ is bounded in contradiction with $|F_n(s)| = n+1$.

Now, suppose by contradiction that s is not balanced and let n be minimal such that for some $u, v \in F_n(s)$ we have $|u|_a - |v|_a \geq 2$. Clearly $n \geq 2$ because $n = 2$ would imply $F_2(s) = \{aa, bb, ab\}$ or $F_2(s) = \{aa, bb, ba\}$ and s would be $aa \cdots ab^\omega$ or $bb \cdots ba^\omega$. So, by the minimality of n, $u = au'a, v = bv'b$ with $|u'|_a - |v'|_a \geq 0$. Let $u' = x_1x_2 \cdots x_{n-2}, v' = y_1y_2 \cdots y_{n-2}, x_i, y_i \in A$. If $u' \neq v'$ let k be minimal such that $x_k \neq y_k$. If $x_k = a, y_k = b$, then $|ax_1 \cdots x_k|_a - |by_1 \cdots y_k|_a = 2$ and n is not minimal. If $x_k = b$, $y_k = a$, then $|ax_1 \cdots x_k|_a = |by_1 \cdots y_k|_a$, whence, deleting these two left factors of u and v, we see that n is not minimal. Consequently $u' = v'$. Now if u' is not a palindrome, let k be minimal such that $x_k \neq x_{n-1-k}$. If $x_k = a$ and $x_{n-1-k} = b$ for instance, we have $|ax_1 \cdots x_k|_a - |x_{n-1-k} \cdots x_{n-2}b|_a = 2$ and n is not minimal. The case $x_k = b$ and $x_{n-1-k} = a$ is similar. Consequently u' is a palindrome.

Now, as $u'a$ and $u'b$ are factors of s, u' is special. As, by i), there is exactly one special factor of each length, either au' or bu' is special. Suppose for instance au' is special. Then $bu'a$ is not a factor of s. Let $s(i, i+n-2)$ be an occurrence of bu' in s and put $w = s(i, i+2n-3)$.

We can write $w = x_0x_1 \cdots x_{2n-3}$ where $x_k \in A, x_0 = b, x_{n-1} = b$ and $x_1x_2 \cdots x_{n-2} = u'$. We claim that au' is not a factor of w. Suppose by contradiction that $au' = x_kx_{k+1} \cdots x_{k+n-2}$ for some $k \in [1, n-1]$. Then we have $x_k = a$ and $x_{n-1-k} = x_{n-1} = b$ in contradiction with the fact that u' is palindrome. Consequently, as $au' \notin Fac(w)$, all $s(i+k, i+k+n-2)$, $0 \leq k \leq n-1$ are non-special. As there are exactly $n-1$ non-special factors of length $n-1$, we have, for some $r < t$ in $[0, n-1]$, $s(i+r, i+r+n-2) = s(i+t, i+t+n-2)$.

Observing that any non-special factor z of s has only one prolongation, za or zb, in s, we get that $s(i+r+n-1) = s(i+t+n-1)$. Also $s(i+r+1, i+r+n-1) = s(i+t+1, i+t+n-1)$. So all factors of length $n-1$ of $s(i+r, i+t+n-1)$ are non-special. Continuing this way we get $s(i+r+n) = s(i+t+n)$ and so on, whence for all $m \geq r$ $s(i+m) = s(i+m+t-r)$. So s is ultimately periodic, a contradiction. Consequently s is balanced.

iii) $\Longrightarrow$ i). By Proposition 2, $|F_n(s)| \leq n+1$ for all $n \in \mathbb{N}_+$. As s is not ultimately periodic, by i) of Proposition 1, $|F_n(s)| = n+1$. $\Diamond$

DEFINITION 7: *An infinite word s is proper Sturmian if it satisfies the (equivalent) conditions of Proposition* 3.

This combinatorial definition is equivalent with the algebraic one (Definition 1 with α irrational) as proved in [10]. That an infinite word satisfying the algebraic definition, with α irrational, has the properties stated in Proposition 3 is easy to see, and is recalled in [12] for instance. This follows from the fact that $\{n\alpha - \lfloor n\alpha \rfloor; n \in \mathbb{N}_+\}$ is dense in $[0, 1]$. For proving the converse, the first step [10] is the proof that an infinite word having the properties of Proposition 3 has a density (*see* Proposition 8 below).

PROPOSITION 4: *If s is proper Sturmian, then $u \in Fac(s)$ implies $\widetilde{u} \in Fac(s)$.*

Proof: Remark first that if w is any factor of s, then w occurs infinitely many times in s, because otherwise we should have $s = zt$, for some $z \in A^*$, $t \in A^\omega$ with $|\mathrm{Fac}_{|z|}(t)| \leq |z|$, and then t, hence s, would be ultimately periodic.

Now, by contradiction, let n be minimal such that $|u| = n$, $u \in \mathrm{Fac}(s)$ and $\widetilde{u} \notin \mathrm{Fac}(s)$. Clearly $n > 1$, so let $u = xvy$, $x, y \in A, v \in F_{n-2}(s)$. We have $\widetilde{v}x, y\widetilde{v} \in \mathrm{Fac}(s)$ and $y\widetilde{v}x \notin \mathrm{Fac}(s)$. Hence $\overline{y}\widetilde{v}x, y\widetilde{v}\overline{x} \in \mathrm{Fac}(s)$. So $\widetilde{v}$ is special. Also the reversal of $\overline{y}\widetilde{v}$ and $y\widetilde{v}$, that is $v\overline{y}$ and vy are factors of s, so v is special. Consequently as there is one special factor of length $n-2$, $v = \widetilde{v}$, that is v is a palindrome. So $\widetilde{u} \notin Fac(s)$ implies $x \neq y$, whence $xvx \in \mathrm{Fac}(s)$ and $yvy \in \mathrm{Fac}(s)$ and s is not balanced, a contradiction with Definition 7. $\diamondsuit$

PROPOSITION 5: *If s is proper Sturmian, it is uniformly recurrent (see Definition 4).*

Proof: If s is not uniformly recurrent there exist a factor u of s and an infinite sequence $(g_i)_{i \in \mathbb{N}_+}$ of factors of s of strictly increasing length such that $u \notin \mathrm{Fac}(g_i)$. Infinitely many of the g_i have the same leftmost letter, x say. Among them infinitely many have the same left factor of length 2, xx' say. Continuing this well-known construction we get an infinite word t such that $\mathrm{Fac}(t) \subset \mathrm{Fac}(s)$ and $u \notin \mathrm{Fac}(t)$. It follows that $|F_{|u|}(t)| \leq |u|$ whence t is ultimately periodic by i) of Proposition 1. Without loss of generality we may assume that t is periodic, with period p, say. Let m be minimal such that $s(1, m)$ is not a factor of t (such a m exists as s is not periodic), and let $M = \max(m, p)$. Let for some $h \in \mathbb{N}_+$, $s(h, h+M-1)$ be a factor of t. Then, as s is not periodic there exist integers i and q with $1 < i \leq h \leq h+M-1 \leq i+q-1$, such that $s(i, i+q-1) \in \mathrm{Fac}(t)$ and $s(i-1, i+q-1) \notin \mathrm{Fac}(t)$ and $s(i, i+q) \notin \mathrm{Fac}(t)$. Put $s(i-1, i+q) = ywz$, with $y, z \in A$ and $w = x_1 x_2 \cdots x_q$, $x_i \in A$. Clearly, $q \geq M+h-i \geq M \geq p$. As t has period p, $x_p w x_{q+1-p} \in \mathrm{Fac}(t)$. So $y \neq x_p$ and $z \neq x_{q+1-p}$. So w is a special factor of s. Also, as, $x_p w$ and yw are factors of s, by Proposition 4, $\widetilde{w} x_p, \widetilde{w} y \in \mathrm{Fac}(s)$. So $\widetilde{w}$ is special and $w = \widetilde{w}$, whence $x_p = x_{q+1-p}$ and $y = z$. So we have $ywy, \overline{y}w\overline{y} \in \mathrm{Fac}(s)$, in contradiction with property of s to be balanced. $\diamondsuit$

PROPOSITION 6: *If the infinite words s and t are uniformly recurrent (proper Sturmian, for instance), then either $Fac(s) = Fac(t)$ or $Fac(s) \cap Fac(t)$ is finite.*

Proof: Suppose $\mathrm{Fac}(s) \cap \mathrm{Fac}(t)$ is infinite and let u be any factor of s. As s is uniformly recurrent, we have, for any sufficiently long $w \in \mathrm{Fac}(s) \cap \mathrm{Fac}(t)$, $u \in \mathrm{Fac}(w)$, whence $u \in \mathrm{Fac}(t)$. So $\mathrm{Fac}(s) \subset \mathrm{Fac}(t)$, and conversely, so $\mathrm{Fac}(s) = \mathrm{Fac}(t)$. $\diamondsuit$

PROPOSITION 7: *Given a proper Sturmian word s, let u be a word such that $u \notin Fac(s)$ and $Fac(u) \setminus \{u\} \subset Fac(s)$. Then there exists a factor v of s such that $|u| = |v|$ and $||u|_a - |v|_a| > 1$*

Proof: Clearly $|u| > 1$ because $a, b \in \mathrm{Fac}(s)$. So let $u = xwy$, $x, y \in A, w \in A^*$. As xw, wy are factors of s, the same holds for $xw\overline{y}$ and $\overline{x}wy$. As s is balanced it follows that $x = y$. Also w is special because $wy, w\overline{y} \in \mathrm{Fac}(s)$.

Now, the special factor of length $|w|+1$ must be xw or $\overline{x}w$. As $xwx = u$ is not a factor of s, $\overline{x}w$ is special, whence $\overline{x}w\overline{x} \in \mathrm{Fac}(s)$. So putting $\overline{x}w\overline{x} = v$ we have $||u|_a - |v|_a| = 2$. $\diamondsuit$

PROPOSITION 8: i) *A proper Sturmian word has a density;* ii) *two proper Sturmian words have the same density if and only if they have the same set of factors.*

Proof of i): Let s be proper Sturmian. Put $M(n) = \mathrm{Sup}\{|u|_a; u \in F_n(s)\}$ and $m(n) = \mathrm{Inf}\{|u|_a; u \in F_n(s)\}$. As s is balanced, $M(n) - m(n) \leq 1$. Also, if v is the special factor of length $n-1$, $va, vb \in F_n(s)$. So $M(n) - m(n) = 1$ for all $n \in \mathbb{N}_+$. Put $D(n) = M(n)/n$ and $d(n) = m(n)/n$. Let $p, q \in \mathbb{N}_+$. As any factor of length pq is the product of p factors of length q we have $pm(q) \leq m(pq) \leq pM(q)$, or $d(q) \leq d(pq) \leq d(q) + 1/q$. Similarly $d(p) \leq d(pq) \leq d(p) + 1/p$. It follows $-1/p \leq d(p) - d(q) \leq 1/q$. So the sequence $(d(p))_{p \in \mathbb{N}_+}$ satisfies the Cauchy condition and has a limit, γ say. Clearly $(D(p))_{p \in \mathbb{N}_+}$ has the same limit. Now, if u_n is the left factor of s of length n, $d(n) \leq |u_n|_a/n \leq D(n)$. So $\lim\limits_{n \to \infty} |u_n|_a/n = \gamma$, that is s has density γ.

Now, with a view to part ii), we prove that for all $p \in \mathbb{N}_+$, $d(p) < \gamma < D(p)$. As $d(p) \leq d(kp)$ for any arbitrarily large integer k, we have $d(p) \leq \gamma$ and similarly $\gamma \leq D(p)$. Suppose, for instance, by contradiction, that $\gamma = D(p)$. Let u be a word of length p such that $|u|_a = m(p)$. As s is (uniformly) recurrent, there are infinitely many occurrences of u, and among them we can choose two that occur at positions congruent modulo p. We then have a factor $w = uvu$ with $|v| = kp$ for some k, for which $M((k+2)p) - 1 = m((k+2)p) \leq |w|_a = 2|u|_a + |v|_a \leq 2m(p) + M(kp) \leq 2M(p) - 2 + kM(p)$. Therefore $M((k+2)p) < (k+2)M(p)$ hence $D(p) > D((k+2)p) \geq \gamma$, and this is a contradiction.

Proof of ii): If s and t are proper Sturmian, denote by M_s, M_t the functions M relative to s and t respectively, an so on. Clearly, if $\mathrm{Fac}(s) = \mathrm{Fac}(t)$ then $D_s(n) = D_t(n)$ for all n, so s and t have the same density.

Conversely, if s and t have density γ, we have, for all $n \in \mathbb{N}_+$, $d_s(n) < \gamma < D_s(n)$ *i.e.* $m_s(n) < n\gamma < M_s(n)$, and similarly $m_t(n) < n\gamma < M_t(n)$. So $m_s(n) = m_t(n)$ and $M_s(n) = M_t(n)$ for all $n \in \mathbb{N}_+$. It follows that if $u \in \text{Fac}(s)$ and $v \in \text{Fac}(t)$ with $|u| = |v|$, then $||u|_a - |v|_a| \leq 1$. Now, suppose by contradiction, that $\text{Fac}(s) \neq \text{Fac}(t)$. Then, by Proposition 6, there exists a factor w of s and t and a letter x such that $wx \in \text{Fac}(s)$ and $wx \notin \text{Fac}(t)$. So, by Proposition 7, we have, for some $v \in \text{Fac}(t)$, $|v| = |wx|$ and $||v|_a - |wx|_a| > 1$, a contradiction. $\diamondsuit$

The following definition is equivalent with Definition 2.

DEFINITION 8: *A proper Sturmian word is standard if each of its left factors is the reversal of a special factor.*

PROPOSITION 9: *Given a proper Sturmian word there is exactly one standard proper Sturmian word with the same set of factors.*

Proof: This follows trivially from Propositions 3i), 4, 5 and from the fact that the special factor of length $n - 1$ is a right factor of the special factor of length n. $\diamondsuit$

The last proposition allows to recover by combinatorial arguments a description of standard Sturmian words given in [17, pp. 65-68] where it is shown that $[0; q_0, q_1, \ldots]$, with the q_i as below, is the simple continued fraction representing α of Definition 1 (*see* also [4] and a construction due to Rauzy [13, 15]).

PROPOSITION 10: *Let s be a non ultimately periodic infinite word. Then the following conditions are equivalent*

i) *s is standard proper Sturmian;*

ii) *there exists a sequence $(q_i)_{i \in \mathbb{N}}$ of positive integers such that, putting $X_0 = a$, $Y_0 = b$ and, for all $i \in \mathbb{N}$, $X_{i+1} = X_i^{q_i - 1} Y_i$, $Y_{i+1} = X_i^{q_i} Y_i$, we have $s = X_1 X_2 X_3 \cdots$.*

Proof: **Proof. i) $\Longrightarrow$ ii).** Let $q_0 = \text{Sup}\{n \in \mathbb{N}; a^n \in \text{Fac}(s)\}$. If $q_0 = 1$ then $aa \notin \text{Fac}(s)$, whence $ba^{q_0 - 1} b = bb \in \text{Fac}(s)$. If $q_0 > 1$ then $ba^n b \notin \text{Fac}(s)$ for $n < q_0 - 1$ because s is balanced; also, as $bb \notin \text{Fac}(s)$ and s is not ultimately periodic, $ba^{q_0 - 1} b \in \text{Fac}(s)$. Clearly $ba^{q_0 - 1}$ is a special factor of s, hence $a^{q_0 - 1} b$ is a left factor of s because s is standard. So, putting $X_1 = a^{q_0 - 1} b$, $Y_1 = a^{q_0} b$, we have $s = X_1 Z_1 Z_2 \cdots$, where $Z_i \in \{X_1, Y_1\}$ for $i \in \mathbb{N}_+$. Let $s_1 = Z_1 Z_2 \cdots$, so that $s = X_1 s_1$. We shall show that s_1, considered as an infinite word on the alphabet $A_1 = \{X_1, Y_1\}$ is standard proper Sturmian. First, $s_1 \in A_1^\omega$ is not ultimately periodic because, otherwise, $s \in A^\omega$ would be ultimately periodic. Now if s_1 is not balanced, reasoning as in the proof of Proposition 3, we find $w \in A_1^\omega$ such that $X_1 W X_1, Y_1 W Y_1 \in \mathrm{Fac}(s_1)$. It follows that, for some $Z, Z' \in A_1$, ZX_1WX_1 and $Z'Y_1WY_1$ are factors of $X_1 s_1 \in A_1^\omega$. Consequently $ba^{q_0-1}bWa^{q_0-1}b$ and $a^{q_0}bWa^{q_0}$ are factors of s, which is impossible as s is balanced. So s_1, considered as an infinite word on A_1, is proper Sturmian. Now let $W \in A_1^*$ be any left factor of s_1. Then, $X_1 W \in A^\omega$ is a left factor of s, and as s is standard proper Sturmian, $bX_1W, baX_1W \in \mathrm{Fac}(s)$. So $X_1 W \in A_1^*$ and $Y_1 W \in A_1^*$ are factors of s_1, whence $\widetilde{W}$ (with respect to A_1) is special, and $s_1 \in A_1^\omega$ is standard proper Sturmian.

Now repeating the argument on $s_1 \in A_1^\omega$ with $q_1 = \mathrm{Sup}\{n \in \mathbb{N}_+; X_1^n \in \mathrm{Fac}(s_1)\}$ and $X_2 = X_1^{q_1-1}Y_1$, $Y_2 = X^{q_1}Y_1$, we get $s_1 = X_2 s_2$ where s_2 is a standard proper Sturmian word on the alphabet $A_2 = \{X_2, Y_2\}$. Continuing this way, we get $s = X_1 s_1 = X_1 X_2 s_2 = \cdots = X_1 X_2 X_3 \cdots$.

ii) $\Longrightarrow$ i). As $s = X_1 X_2 X_3 \cdots$, we can define $s_i \in A^\omega$ by $s = X_1 s_1$ and $s_i = X_{i+1} s_{i+1}$ for $i \in \mathbb{N}_+$. As all X_i, Y_i are products of factors equal to X_1, Y_1, we have $s = Z_1 Z_2 \cdots$, $Z_j \in A_1 = \{X_1, Y_1\}$. Observing that A_1 is a prefix code [14, chap 1] or simply that the occurrences of b in s are markers indicating the positions of factors X_1, Y_1, we are allowed to consider s_1 and $X_1 s_1$ as infinite words on the alphabet A_1. In the same way, s_i and $X_i s_i$ may be considered as infinite words on $A_i = \{X_i, Y_i\}$ (when useful, we precise the alphabet by writing "on A_i" or "$\in A_i^\omega$" or "$\in A_i^*$").

It follows also that if bfb (resp. fb), $f \in A^*$, is a factor (resp. a left factor) of s, then there exists a unique $W \in A_1^*$ such that $fb = W$ and, moreover that W is a factor (resp. a left factor) of s_1 (resp. of $X_1 s_1$) $\in A_1^\omega$. Observe also that X_1 and Y_1 are factors of X_i, for all $i \geq 3$, whence it follows that a and b occur infinitely many times in s and that $ba^p b \in \mathrm{Fac}(s)$ if and only if $p = q_0$ or $p = q_0 - 1$. Obviously, similar remarks apply to all s_i.

First, we show that s is not ultimately periodic. As a and b occur infinitely many times in s, if s is ultimately periodic with period p, say, we have $s = f(bg)^\omega$, $f, g \in A^*$, $|bg| = p$. Then there exist $U, V \in A_1^*$ such that $fb = U$, $gb = V$ and $X_1 s_1 = UV^\omega \in A_1^\omega$. Letting $|V|$ be the length of V with respect to A_1, we have $|V|q_0 + |V|_{Y_1} = |gb| = p$. So $|V| \leq p$, with equality only if $V = X_1^m$ for some $m \in \mathbb{N}_+$ and $q_0 = 1$, that is $X_1 = b$, whence $s = fb^\omega$, an impossible case. As the period p_1 of $s_1 \in A_1^\omega$ is a divisor of $|V|$ we get $p_1 < p$. Similarly all $s_i \in A_i^\omega$ are ultimately periodic with strictly decreasing periods. This being impossible, we conclude that s and all s_i are not ultimately periodic.

Now we prove that s is balanced. For each $i \in \mathbb{N}_+$, if $s_i \in A_i^\omega$ is balanced, let $n_i = \infty$ and if s_i is not, let n_i be minimal such that there exist $U, V \in F_{n_i}(s_i)$ with $|U|_{X_i} - |V|_{X_i} > 1$. Suppose by contradiction that s is not balanced. Then there exist n minimal and $u, v \in \mathrm{Fac}(s)$ with $n = |u| = |v|$ and $|u|_a - |v|_a > 1$. As previously we get $u = afa$, $v = bfb$, $f \in A^*$. Also f contains at least one occurrence of b because if $f = a^p$, then $p + 2 \leq q_0$ and $p \geq q_0 - 1$, a contradiction. So f has a left factor of the form $a^p b$ and, as $af, bf \in \mathrm{Fac}(s)$, $p = q_0 - 1$. Similarly, $ba^{q_0 - 1}$ is a right factor of f. As $bfb \in \mathrm{Fac}(s)$, we have then $fb = W \in A_1^*$, $W \in \mathrm{Fac}(s_1)$ and $W = X_1 H X_1$ for some $H \in A_1^*$. Also $afaa, aafa \notin \mathrm{Fac}(s)$ because $a^{q_0+1} \notin \mathrm{Fac}(s)$. So $afab \in \mathrm{Fac}(s)$. Moreover as $a^{q_0} b$ is not a left factor of s, $bafab \in \mathrm{Fac}(s)$. Hence $afab = aX_1 a^{q_0} b = Y_1 H Y_1$ is a factor of $s_1 \in A_1^*$. Consequently, by the definition of n_1, $|W| \geq n_1$. As $|fb| \geq |W| q_0$ we get $n \geq n_1 + 1$. Similarly $n_1 > n_2$ and so on. This being impossible, s is balanced and similarly all s_i are.

Consequently, s and s_i are proper Sturmian. It remains to show that s is standard, that is $af, bf \in \mathrm{Fac}(s)$ for any left factor f of s. It suffices to consider the left factor $f = X_1 X_2 \cdots X_{h+1}$ for arbitrary $h \in \mathbb{N}$. We claim that, considering $F = X_2 X_3 \cdots X_{h+1}$ as a word on A_1, if $X_1 F, Y_1 F$ are factors of $s_1 \in A_1^\omega$, then $af, bf \in \mathrm{Fac}(s)$. As $Y_1 F = aX_1 X_2 \cdots X_{h+1} = af \in A^*$ we have the result for af. Also, as s_1 is uniformly recurrent, $Z X_1 F \in \mathrm{Fac}(s_1)$ for some $Z \in A_1$. Hence $bX_1 F = bf \in \mathrm{Fac}(s)$.

Repeating this argument, we get that $af, bf \in \mathrm{Fac}(s)$ if $X_h X_{h+1}$ and $Y_h X_{h+1}$, considered as words on X_h, are factors of $s_h \in A_h^\omega$, and this can be verified as follows. By a remark above, $Y_h X_h^p Y_h$ is a factor of s_h if and only if $p = q_h$ or $p = q_h - 1$. So $X_h X_{h+1} = X_h^{q_h} Y_h$ and $Y_h X_{h+1} = Y_h X_h^{q_h - 1} Y_h$ are factors of s_h.

In conclusion, s is a standard proper Sturmian word. $\diamondsuit$

OTHER RESULTS

Among the features of Sturmian words that have been or can be studied without the help of the algebraic definition, let us mention: the description of the palindrome factors [8], the fact that, if s is proper Sturmian, the Rees quotient monoid $A^*/(A^* \setminus \mathrm{Fac}(s))$ has the weak permutation property P_4^* (*see* [11]), the particular case of the Fibonacci word, and above all, two important subjects, the characterization of the morphisms of A^* that leave the set of all Sturmian words globally invariant [2], and the properties of the "finite Sturmian words", that is the factors of (infinite) Sturmian words [7, 9].

AKNOWLEDGEMENT

We thank the referees for useful remarks, in particular a short cut in a proof.

REFERENCES

1. J. Berstel, *Fibonacci words - a survey*, in G. Rosenberg, A. Salomaa eds., The Book of L, Springer-Verlag, Berlin, 1986.
2. J. Berstel and P. Séébold, *A characterization of Sturmian morphisms*, in A. Borzyskowski, S. Sokolowski eds., MFCS '93, Lect. Notes Comp. Sci. Springer-Verlag, 1993.
3. J. Berstel and P. Séébold, A remark on morphic Sturmian words, *R.A.I.R.O., I.T.* 1994, *28, 3-4*, pp. 255-263.
4. T.C. Brown, Description of the Characteristic Sequence of an irrational, *Canad. Math. Bull.* , 1993, *36, 1*, pp. 15-21.
5. E. M. Coven and G. A. Hedlund, Sequences with Minimal Block Growth, *Math. System Theory* , 1973, *7, 2*, pp. 138-153.
6. D. Crisp, W. Moran, A. Pollington and P. Shive, Substitution invariant cutting sequences, *Journal de Théorie des Nombres de Bordeaux* , 1993, *5*, pp. 123-137.
7. A. de Luca and F. Mignosi, *Some Combinatorial Properties of Sturmian Words, Theoretical Computer Science 136*, 1994, pp. 361-385.
8. X. Droubay, Palindromes in the Fibonacci word, *Inf. Proc. Lett.* , 1995, *55*, pp. 217-221.
9. S. Dulucq and D. Gouyou-Beauchamps, Sur les facteurs des suites de Sturm, *Theoretical Computer Science* , 1990, *71*, pp. 381-400.
10. G. A. Hedlund and M. Morse, Symbolic Dynamics II. Sturmian Trajectories, *Amer. J. Math.*, 1940, *62*, pp. 1-42.
11. J. Justin and G. Pirillo, On a combinatorial property of Sturmian words, *Theoretical Computer Science* , 1996, *154*, pp. 387-394.
12. F. Mignosi, Infinite words with linear subword complexity, *Theoretical Computer Science* , 1989, *65*, pp. 221-242.
13. F. Mignosi and P. Séébold, Morphismes sturmiens et règles de Rauzy, *Journal de Théorie des Nombres de Bordeaux* , 1993, *5*, pp. 221-223.
14. M. Lothaire, *Combinatorics on Words*, Addison-Wesley Pub. Co., London, 1983.
15. G. Rauzy, *Mots infinis en arithmétique*, in Automata on Infinite Words, M. Nivat, D. Perrin Editors, *Lect. Notes in Computer Science,* vol. 192, Springer-Verlag, Berlin 1985.
16. C. Series, The geometry of Markoff numbers, *Math. Intelligencer* , 1985, *7*, pp. 20-29.
17. B. A. Venkov, *Elementary Number Theory* Translated and edited by H. Alderson, Wolters-Noordhoff, Groningen, 1970.