A note on Sturmian words

Dominique Perrin, Antonio Restivo

A note on Sturmian words

Dominique Perrin[1], Antonio Restivo[2]
[1]Université Paris Est, [2] Universita di Palermo

November 4, 2011

Abstract

We describe an algorithm which, given a factor of a Sturmian word, computes the next factor of the same length in the lexicographic order in linear time. It is based on a combinatorial property of Sturmian words which is related with the Burrows-Wheeler transformation.

1 Introduction

Sturmian words are infinite words over a binary alphabet that have exactly $n+1$ factors of length n for each $n \geq 0$. Their origin can be traced back to the astronomer J. Bernoulli III. Their first in-depth study is by Morse and Hedlund [11]. Many combinatorial properties were described in the paper by Coven and Hedlund [5]. Sturmian words, also called mechanical words, are used in computer graphics as digital approximation of straight lines. See [8] for a general exposition on Sturmian words.

In this note, we describe an algorithm which, given a factor of a Sturmian word, computes the next factor of the same length in the lexicographic order in linear time. It may be used to generate the set of factors of a Sturmian word of given length in lexicographic order.

This algorithm is based on a characterization of the pairs of factors of the same length of a Sturmian word which are consecutive in the lexicographic order (Theorem 2). This result is related with several previously known results and in particular with those of [13], [7] and [3].

The characterization of consecutive pairs of factors is used to prove a combinatorial property of the set of factors of given length of a Sturmian word (Proposition 3). It says that if one orders this set lexicographically, the sequence of their last letters taken cyclically has exactly two changes from a to b or conversely. This is shown to be related to a result of Mantaci, Sciortino and the second author concerning the Burrows-Wheeler transform of a standard word [10].

The characterization is used in combination with classical algorithms on words (and, in particular, the algorithm computing the overlap of two words) to obtain a linear algorithm computing the factor of the same length immediately following a given one in the lexicographic order. The algorithm itself is used to show that one may generate the set of factors of a Sturmian word of given length in lexicographic order in quadratic time (Proposition 5).

The paper is organized as follows. In the Section 2, we recall properties of Sturmian words and their equivalence with mechanical words. In Section 3, we prove the result characterizing pairs of consecutive factors (Theorem 2). In Section 4 we introduce the notion of right border of a set of words. We prove that the right border of the set of factors of given length of a Sturmian word is conjugate to a word in a^*b^* (Proposition 3). We also relate this result with that of Mantaci, Restivo and Sciortino [10]. Finally, in Section 5, we show that one may compute in linear time the factor of the same length following a given one. The algorithm is used to generate in quadratic time the set of factors of a given length of a Sturmian word. This complements the results of [6], of [2] and of [4] who give efficient algorithms to recognize or to generate finite Sturmian words, which are by definition the finite factors of infinite Sturmian words.

2 Preliminaries

We fix the alphabet to be $A = \{a, b\}$. We denote by $|w|$ the length of a word w and by $|w|_a, |w|_b$ the number of occurrences of the letters a, b in w.

An infinite word on A is *Sturmian* if it has for all $n \geq 1$, $n + 1$ factors of length n.

A finite word w is called Sturmian if it is a factor of an infinite Sturmian word.

Example 1 Let $f : A^* \to A^*$ be the morphism defined by $f(a) = ab$ and $f(b) = a$. The infinite word $s = \lim_{n\to\infty} f^n(a)$ is a Sturmian word called the *Fibonacci word*. One has

$$s = abaababaabaab\cdots$$

The sequence of words defined by $u_n = f^{n+1}(b)$ for $n \geq -1$ is the sequence of *Fibonacci words*. One has $u_n = u_{n-1}u_{n-2}$ for $n \geq 1$.

A factor u of a Sturmian word s is *right special* if ua, ub are factors of s. There is exactly one right special factor of each length.

A set X of words over the alphabet A is *balanced* if for any $u, v \in X$ with $|u| = |v|$, one has $||u|_b - |v|_b| \leq 1$. The set of factors of length n of a Sturmian word is balanced.

For two real numbers α and ρ with $0 \leq \alpha \leq 1$, we define an infinite word $s_{\alpha,\rho}$ by

$$s_{\alpha,\rho}(n) = \begin{cases} a & \text{if } \lfloor \alpha(n+1) + \rho \rfloor = \lfloor \alpha n + \rho \rfloor \\ b & \text{otherwise} \end{cases}$$

A word of this form is called *mechanical*. The real number α is the *slope* of $s_{\alpha,\rho}$. The following result is from [8] (Theorem 2.1.13).

Theorem 1 *An infinite word s is Sturmian if and only if it is mechanical of irrational slope.*

The real number α such that $s = s_{\alpha,\rho}$ is called the *slope* of s. The word c_α defined by $s_{\alpha,0} = ac_\alpha$ is called the *characteristic word* of slope α. It is a Sturmian word of slope α.

A Sturmian set is the set of factors of a Sturmian word. The slope of F is the slope of a Sturmian word such that $F = F(s)$. The characteristic word associated with F is the infinite word c_α. It is the unique characteristic word x such that $F = F(x)$.

The right special words in F are the reversals of the prefixes of the characteristic word of the same slope (Proposition 2.1.23 in [8]).

Example 2 The Fibonacci word is the characteristic word of slope $\alpha = \frac{2}{3+\sqrt{5}}$. The word $s_{\alpha,0}$ is represented in Figure 1. It is the broken line joining the integer points placed just below the line $y = \alpha x$.

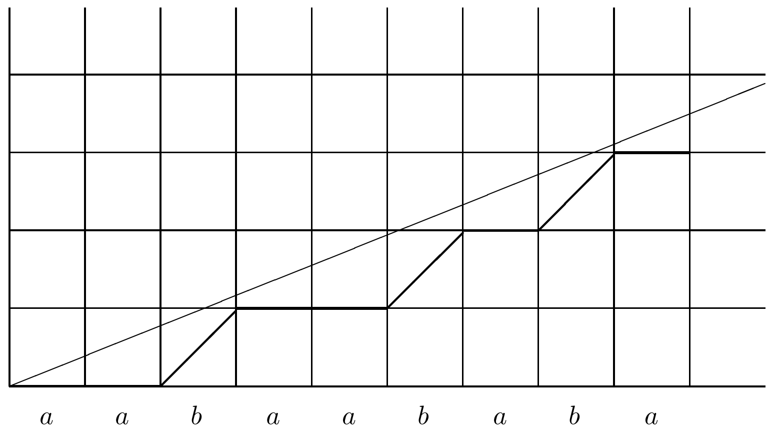

Figure 1: The word $s_{\alpha,0}$.

The following result is from [8] (see the proof of Proposition 2.1.17).

Proposition 1 *Let s be a Sturmian word with slope α. Then $w \in F(s)$ if and only if for any factor u of w one has*

$$|u|_b - 1 < \alpha|u| < |u|_b + 1. \tag{1}$$

Corollary 1 *Let F be a Sturmian set. If $ra, rba \in F$, then $rab \in F$.*

Proof Let α be the slope of F. Let u be a factor of rab. If u is a factor of ra, then Inequality (1) holds. Otherwise, u is a suffix of rab. If u is a suffix of ab, then Inequality (1) holds because $ab \in F$. Otherwise $u = tab$ where t is a suffix of r. Since $rba \in F$, Inequality (1) holds for tba and thus it holds for tab. ■

Corollary 2 *Let F be a Sturmian set. If $rabsa, rbasb, bsb \in F$, then $rabsb \in F$.*

Proof Let α be the slope of F. Let u be a factor of $rabsb$. We show that Inequality (1) holds for u and thus that $rabsb \in F$ by Proposition 1. If u is a factor of $rabs$, it is a factor of $rabsa \in F$ and thus Inequality (1) holds. Assume next that u is a suffix of $rabsb$. If $absb$ is a suffix of u, set $u = wabsb$. Since $|u|_b = |wbasb|_b$ and since $wbasb$ is a factor of $rbasb \in F$, (1) holds for u. Finally, if u is a suffix of $bsb \in F$, then Inequality (1) holds for u. ■

Note that the conclusion of Corollary 2 expresses the fact that the word $rabs$ is right special.

A *standard pair* is a a pair (u, v) of words obtained starting from the pair (a, b) using the two transformations Γ, Δ defined by

$$\Gamma(u, v) = (u, uv), \quad \Delta(u, v) = (vu, v).$$

A word is *standard* if it appears as a component of a standard pair.

Example 3 The Fibonacci words are standard. Indeed, $(u_0, u_{-1}) = (a, b)$ and for $n \geq 1$, $(u_{2n+2}, u_{2n+1}) = \Delta\Gamma(u_{2n}, u_{2n+1})$.

Let $(d_1, d_2, \ldots, d_n, \ldots)$ be a sequence of integers with $d_1 \geq 0$ and $d_n > 0$ for $n > 1$. To such a sequence, we associate a sequence $(s_n)_{n \geq -1}$ of words by

$$s_{-1} = b, \quad s_0 = a, \quad s_n = s_{n-1}^{d_n} s_{n-2} \quad (n \geq 1). \tag{2}$$

The sequence $(s_n)_{n \geq -1}$ is a *standard sequence* and the sequence $(d_1, d_2, \ldots)$ is its *directive sequence*. Every standard word occurs in some standard sequence and every word occurring in a standard sequence is standard [8].

Example 4 The standard sequence associated with the sequence $(1, 1, \ldots)$ is the sequence of Fibonacci words.

We denote by $[a_0, a_1, \cdots]$ the continuous fraction

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots}}$$

and by $[a_0, a_1, \ldots, a_n]$ the corresponding truncated fraction. For an infinite continuous fraction $\alpha = [a_0, a_1, \ldots]$, the rational numbers $[a_0, a_1, \ldots, a_n]$ for $n \geq 0$ are called the *convergents* to α. It is classical that $[a_0, \ldots, a_n] = p_n/q_n$ where the integers p_n, q_n are given by $p_0 = a_0$, $p_1 = a_1 a_0 + 1$, $q_0 = 1$, $q_1 = a_1$ and the recurrence relations

$$p_n = a_n p_{n-1} + p_{n-2}, \quad q_n = a_n q_{n-1} + q_{n-2} \tag{3}$$

for $n \geq 2$.

Example 5 The continuous fraction developement of $\alpha = 2/(3 + \sqrt{5})$ is $[0, 2, 1, 1, \ldots]$. The sequence of convergents is $\frac{1}{2}, \frac{1}{3}, \frac{2}{5}, \ldots$.

Let $\alpha < 1$ be an irrational number and let F be the Sturmian set with slope α. Let $\alpha = [0, 1 + d_1, d_2, \ldots]$ be the continuous fraction expansion of the irrational α and let (s_n) be the standard sequence associated with $(d_1, d_2, \ldots)$. Then the words s_n are prefixes of the characteristic word c_α and for any $n \geq 1$, all conjugates of s_n are in F. Every element of F of length $|s_n|$ is a conjugate of s_n except one called the *singular* factor of length $|s_n|$. Moreover, a singular factor is of the form aua or bub (see [13] or Exercise 2.2.15 in [8]).

By Proposition 2.2.15 in [8], for any pair h, m of relatively prime integers with $1 \leq h < m$, there are exactly two standard words with h ocurrences of b and $m - h$ occurrences of a, which are of the form zab and zba. The words azb and bza are called *Christoffel words*.

Let $\alpha < 1$ be an irrational number and let $[0, 1 + d_1, d_2, \ldots]$ be the continuous fraction expansion of α. Let (s_n) be the standard sequence associated with the sequence $(d_1, d_2, \ldots)$. The comparison of the recurrence relations given in Equations (2) and (3) shows that the sequence of convergents p_n/q_n of the continuous fraction expansion of α is related to the sequence (s_n) of standard words by

$$p_n = |s_n|_b, \quad q_n = |s_n| \tag{4}$$

3 Consecutive factors

We will use the following characterization of consecutive elements in a Sturmian set.

Theorem 2 *Let F be a Sturmian set. Two words u, v of F of the same length are consecutive in the lexicographic order if and only if $u = rabs$ and $v = rbas$ or if $u = ra$ and $v = rb$.*

This result can be obtained from previously known properties of Sturmian words. We first describe these connections and then give a direct proof to make this paper self-contained.

The fact that the condition is sufficient is a direct consequence of the balance property of standard words (this argument is detailed below in the direct proof of the theorem). To see that it is necessary, consider a Christoffel word of length n of the form $w = aub$. Then two conjugates of w which are consecutive for the lexicographic order are of the form $rabs$ and $rbas$ (Corollary 5.1 in [3]). This gives Theorem 2 since any factor of a Sturmian word is a factor of some standard factor.

Another deduction of Theorem 2 can be made using results from [7] and a result from [12] asserting that if axb and bxa are Sturmian, then they are conjugate. It follows then that if $x = rs$, then $rabs$ and $rbas$ are Sturmian whence our statement.

Note that Theorem 2 is not true for consecutive Sturmian words (not necessarily factors of the same Sturmian word). Indeed, for any $n \geq 2$, ab^n and ba^n are consecutive Sturmian words of length $n+1$ but they are not of the indicated form.

Proof of Theorem 2. Assume first that $u = rabs$ and $v = rbas$. If u, v are not consecutive, we have $s = taw$ with $rabtb \in F$. Since $v = rbas$, the word ata is in F. But the fact that btb and ata are in F contradicts the balance property. Thus u, v are consecutive. This shows that the condition is sufficient.

Let us prove the converse. We may assume that $bb \notin F$. Since u, v are consecutive, we have $u = raw$ and $v = rbx$ for some $r \in F$ and $w, x \in F$ of the same length. We may assume that w, x are not empty. Since $bb \notin F$, we have $x = az$. Assume that the first letter of w is a. By Corollary 1, we have $rab \in F$. Since rab is a prefix of a word w of F of length n and since $u < w < v$, this contradicts the hypothesis that u, v are consecutive. Thus the first letter of w is b. Set $w = by$. We have therefore

$$u = raby, \quad v = rbaz.$$

Let us show by induction on the length of y, z that $y = z$.

The property is true if y, z are empty. Assume the contrary. Then, by induction hypothesis, we have $y, z \in sA$ for some word s.

Let us first show that $y = sa$ and $z = sb$ is impossible. Since s is right special, one of as, bs has to be right special. If bs is right special, then $rabs$ is right special by Corollary 2. But then $u = rabsa < rabsb < rbasb = v$ contrary to the assumption that u, v are consecutive. Similarly, if as is right special, we obtain $rbasa \in F$ by the symmetric statement of Corollary 2 obtained by exchanging a and b. Thus $u < rbasa < rbasb$, a contradiction.

Assume next that $y = sb$ and $z = sa$. We would then have $asa, bsb \in F$, a contradiction with the balance property of F.

Thus either $y = sa, z = sa$ or $y = sb, z = sb$ and thus $y = z$. ∎

The following statement is a direct consequence of Theorem 2.

Corollary 3 *Let F be a Sturmian set and let $n \geq 1$. For any word u in $F \cap A^n$ which is not maximal for the lexicographic order in $F \cap A^n$, there is a prefix r of u such that*

(i) *r is right special,*

(ii) *$u = rabs$ or $u = ra$.*

The element of $F \cap A^n$ following u is the word $rbas$ (or rb if $u = ra$) where r is the longest prefix of u such that conditions (i) *and* (ii) *hold.*

The longest prefix of u such that conditions (i) and (ii) hold is called the *principal prefix* of u.

We add the following statement to Theorem 2. It complements the description of the cycle formed by consecutive elements with the first following the last one.

Proposition 2 *Let F be a Sturmian set and let $n \geq 1$. The first and the last elements of $F \cap A^n$ have the form $u = as, v = bs$.*

Proof Let us prove the assertion by induction on n. It is true for $n = 1$. Next, suppose that $n \geq 2$. By induction hypothesis, the first and the last elements of $F \cap A^{n-1}$ have the form as, bs. Thus $u \in asA$ and $v \in bsA$. We cannot have $u = asa$ and $v = bsb$ by the balance property. We cannot have either $u = asb$ and $v = bsa$. Indeed, either as or bs is right special and thus either asa or bsb is in F, a contradiction. This proves that $u = asa, v = bsa$ or $u = asb, v = bsb$. ∎

Note that the word s in Proposition 2 is the left special factor of length $n - 1$. In a similar way, the word s of Theorem 2 is left special.

Note additionally that proposition 2 can also be deduced from the fact that if n is the length of a standard factor of a Sturmian word, the lexicographic minimal and maximal elements of its conjugacy class have the form aub and bua (see [3] for example).

Example 6 Let F be the Fibonacci set. The words of F of length at most 10 appear in Figure 2. Note that the first and the last element have the form as, bs where s is the prefix of length 9 of the Fibonacci word.

4 Right borders

Let X be a finite set of nonempty words, let $(u_0, u_1, \ldots, u_n)$ be the list of elements of X in increasing lexicographic order and let a_i be the last letter of u_i for $0 \leq i \leq n$. The *right border* of X is the word $a_0 a_1 \cdots a_n$. We use Theorem 2 to prove the following result.

Proposition 3 *For $n \geq 1$, the right border of the set $F \cap A^n$ is conjugate to a word in $a^* b^*$.*

Proof Set $F \cap A^n = \{u_0, u_1, \ldots, u_n\}$ and let $v_n = a_0 a_1 \cdots a_n$ be the right border of $F \cap A^n$. For $n = 1$, we have $a_0 = a$ and $a_1 = b$ and thus $v_1 = ab$. Assuming that $bb \notin F$, we have $F \cap A^2 = \{aa, ab, ba\}$ and thus $v_2 = aba$. Thus the property holds for $n \leq 2$.

Assume now that $n \geq 2$. For $0 \leq i \leq n$, let w_i, z_i be the prefixes of u_i of length $n-1, n-2$ respectively. Let i_0 be the least index such that w_i is right special and i_1 be the least index such that z_i is right special.

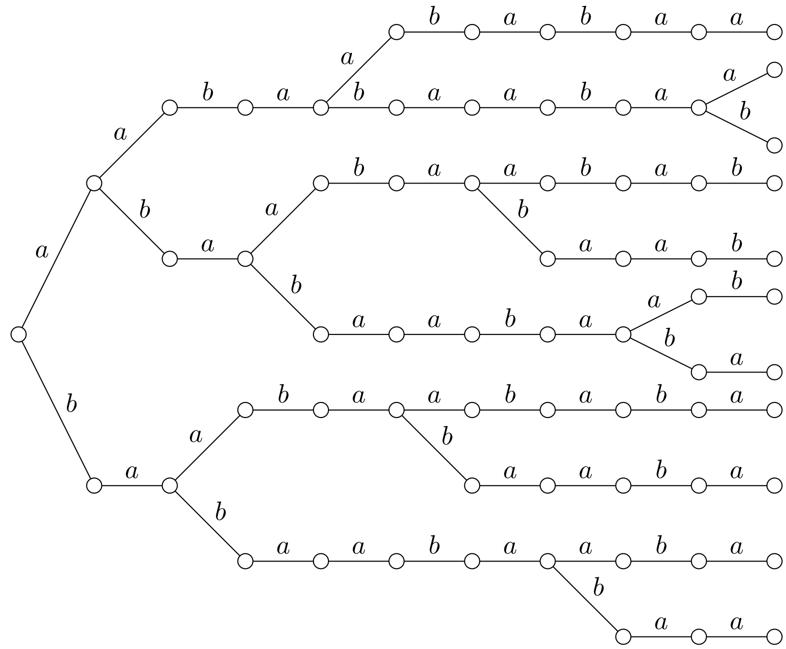

Figure 2: The Fibonacci set

We claim that $a_i = a_{i+1}$ for all indices i such that $0 \leq i < n$ and $i \neq i_0, i_1$. Indeed, by Theorem 2, we have $u_i = rabs$ and $u_{i+1} = rbas$ with $s \neq 1$ and thus a_i, a_{i+1} are equal to the last letter of s. Moreover $a_0 = a_n$ by Proposition 2. We have

$$v_n = \begin{cases} a^{i_0+1}b^{i_1-i_0}a^{n-i_1} & \text{if } i_0 < i_1 \\ b^{i_1+1}a^{i_0-i_1}b^{n-i_0} & \text{otherwise} \end{cases}$$

This proves that v_n is conjugate to a word in a^*b^*. ∎

Note that Proposition 3 can also be stated in an equivalent way as follows. Let M be the $n \times n$ matrix with elements in A such that its i-th row is, for $1 \leq i \leq n$, the i-th element of $F \cap A^n$ in lexicographic order (we identify a word with an n-vector with elements in A). Then Proposition 3 is equivalent to the assertion that *each column of M is conjugate to a word in a^*b^**. Indeed, let u_j be the j-th element of $F \cap A^i$ in the lexicographic order and let n_j be the number of elements in $F \cap A^n$ which have u_j as a prefix. Then the i-th column c_i of M is obtained from the border w_i of the set $F \cap A^i$ by repeating n_j times the j-th letter of w_i. The fact that w_i is conjugate to a word in a^*b^* implies clearly that c_i has the same property.

Example 7 Let F be the Fibonacci set. The set of words of F of length 10 is listed in Table 1. For each word except for the maximal one, we have indicated its principal prefix. Note that the lengths of the principal prefixes are all distinct, as a consequence of the fact that a principal prefix is right special.

Proposition 3 is related with another result proved in [10] that we introduce now.

a	a	b	a	a	b	a	b	a	a
a	a	b	a	b	a	a	b	a	a
a	a	b	a	b	a	a	b	a	b
a	b	a	a	b	a	a	b	a	b
a	b	a	a	b	a	b	a	a	b
a	b	a	b	a	a	b	a	a	b
a	b	a	b	a	a	b	a	b	a
b	a	a	b	a	a	b	a	b	a
b	a	a	b	a	b	a	a	b	a
b	a	b	a	a	b	a	a	b	a
b	a	b	a	a	b	a	b	a	a

Table 1: The factors of length 10 of the Fibonacci word

a	a	b	r	a	c	a	d	a	b	r
a	b	r	a	a	b	r	a	c	a	d
a	b	r	a	c	a	d	a	b	r	a
a	c	a	d	a	b	r	a	a	b	r
a	d	a	b	r	a	a	b	r	a	c
b	r	a	a	b	r	a	c	a	d	a
b	r	a	c	a	d	a	b	r	a	a
c	a	d	a	b	r	a	a	b	r	a
d	a	b	r	a	a	b	r	a	c	a
r	a	a	b	r	a	c	a	d	a	b
r	a	c	a	d	a	b	r	a	a	b

Table 2: The Burrows-Wheeler transform

The *Burrows-Wheeler transform* of a primitive word w is the word denoted $T(w)$ which is the right border of the set of conjugates of w.

Example 8 We have $T(abracadabra) = rdarcaaaabb$ as shown in Table 2

The following result appears in [10] (Theorem 9).

Theorem 3 *One has $T(w) = b^p a^q$ with p, q relatively prime if and only if w is a conjugate of a standard word.*

Example 9 The Fibonacci word $w = abaababa$ is standard and $T(w) = b^3a^5$.

Let F be Sturmian set of slope α. Let (s_n) be the standard sequence associated with the characteristic word c_α. For any integer m of the form $|s_n|$, the set $F \cap A^m$ is the union of the set X of conjugates of s_n and of the singular factor of length m. Since the right border of X and

$F \cap A^m$ differ by one letter, the fact that the first is conjugate to a word in a^*b^* (the only if part of Theorem 3) follows also from Proposition 3.

Note also that the singular factor is always the first or the last element word in $F \cap A^m$. Let indeed $s_n = p_n xy$ with $x, y \in A$. Then p_n is palindromic and the singular factor of length $m = |s_n|$ is $w_n = x s_n y^{-1} = x p_n x$. For $x = a$, w_n is the first element of $F \cap A^m$ and for $x = b$, it is the last one.

Example 10 The set Y of factors of length 8 of the Fibonacci word is represented in Table 3. The associated standard sequence is the Fibonacci sequence $(b, a, ab, aba, abaab, abaababa, \ldots)$. Thus the set Y is formed of the set X of conjugates of *abaababa* and the singular factor *babaabab* (represented in boldface). The right border of Y is b^3a^5b while the right border of X is b^3a^5.

a	a	b	a	a	b	a	b
a	a	b	a	b	a	a	b
a	b	a	a	b	a	a	b
a	b	a	a	b	a	b	a
a	b	a	b	a	a	b	a
b	a	a	b	a	a	b	a
b	a	a	b	a	b	a	a
b	a	b	a	a	b	a	a
b	**a**	**b**	**a**	**a**	**b**	**a**	**b**

Table 3: The factors of length 8 of the Fibonacci word

5 Generating Sturmian sets

We show in this section that Theorem 2 can be used to compute in linear time the factor of the same length of a Sturmian word which follows a given one in the lexicographic order. We use this algorithm to generate the set of elements of a Sturmian set F of length n in time $O(n^2)$.

We assume that the Sturmian set F is given by a function Characteristic(α, n) which returns the prefix of length n of the characteristic word associated with F. Thus the right special factor of length n is the reversal of Characteristic(α, n). The implementation of the function Characteristic(α, n) is a standard task (see [1]). Its complexity is linear in n. The implementation can be done from some representation of the real number α as follows.

Characteristic(α, n)

$y \leftarrow \alpha$
for $i \leftarrow 1$ **to** n **do**
  $z \leftarrow y$
  $y \leftarrow y + \alpha$
  if $\lfloor y \rfloor = \lfloor z \rfloor$ **then**
    $s[i] \leftarrow a$
  else $s[i] \leftarrow b$
  return s

The implementation can also be done using only integers. One starts with a rational approximation v/u of α and uses the following algorithm (reproduced from [1] with a shift by 1 of the indices and a minor correction).

```
Characteristic(u, v, n)
d ← v
for i ← 1 to n do
    if d + v < u then
        d ← d + v
        s[i] ← a
    else d ← d + v − u
        s[i] ← b
    return s
```

Both algorithms give the same result when v/u is a continued fraction approximation of α and $n \leq u-2$. Indeed, for any pair u, v of relatively prime integers with $1 \leq v < u$, the word az with $z =$Characteristic$(u, v, u-1)$ is a Christoffel word. Thus Characteristic(u, v, n) is the prefix of length n of the standard prefix of length u of the characteristic word.

Example 11 Let F be the Fibonacci set, which has slope $\alpha = 2/(3+\sqrt{5})$. The continued fraction expansion of α is $[0, 2, 1, 1, \ldots]$ and the corresponding sequence of convergents is $1/2, 1/3, 2/5, 3/8, 5/13, \ldots$. The sequence of values of d and s in the execution of Characteristic$(13, 5, 11)$ are represented below.

i	1	2	3	4	5	6	7	8	9	10	11
d	10	2	7	12	4	9	1	6	11	3	8
s	a	b	a	a	b	a	b	a	a	b	a

The computation of the principal prefix of a word u is realized by the following algorithm.

```
PrincipalPrefix(u)
n ← Length(u)
x ← Reverse(Characteristic(α, n − 1))
if u = Cat(x, a) then
    return x
x ← Suffix(x, n − 2)
if u = Cat(x, ab) then
    return x
x ← Cat(x, ab)
t ← Overlap(x, u)
m ← Length(t)
if m ≤ 1 then
    return −1
else r ← Prefix(t, m − 2)
    return r
```

In the fuction PrincipalPrefix, we use the following functions.

- Reverse(x) which returns the reversal of the word x,
- Length(u) which returns the length of a word u,
- Cat(x, y) to concatenate two words,
- Overlap(x, y) which returns the longest proper suffix of x which is also a proper prefix of y,
- Prefix(x, n) and Suffix(x, n) which return respectively the prefix or the suffix of the word x of length n.

The following result is a direct consequence of Corollary 3.

Proposition 4 *The function* PrincipalPrefix(u) *returns the principal prefix of u if u is not maximal in the set of elements of F of the same length and -1 otherwise.*

It is well-known that the function Overlap(x, y) can be implemented to compute its result in time $O(|x| + |y|)$ (see [9] for example). Thus the function Principalprefix(u) returns its value in time $O(|u|)$.

Using the function PrincipalPrefix, it is easy to obtain a function which computes the next element of the same length in the set F.

```
Next(u)
n ← Length(u)
r ← PrincipalPrefix(u)
m ← Length(r)
if n = m + 1 then
      return Cat(r, b)
else s ← Suffix(u, n − m − 2)
      return Cat(r, ba, s)
```

Finally, the following function generates all elements of F of length n by visiting them in turn.

```
Sturm(n)
x ← Characteristic(α, n − 1)
u ← Cat(a, x)
v ← Cat(b, x)
while u ≠ v do
      Visit(u)
      u ← Next(u)
      Visit(v)
```

Since Next operates in linear time, we obtain the result announced earlier.

Proposition 5 *The algorithm* Sturm *generates the elements of length n of a Sturmian set in lexicographic order in quadratic time $O(n^2)$.*

Note that the algorithm is quadratic in n but actually linear in the size of the set of elements of length n of a Sturmian set F since $\mathrm{Card}(F \cap A^n) = n(n + 1)$.

Acknowledgments The authors wish to thank Jean Berstel for his suggestions on the early versions of this paper and the referees for improving its quality.

References

[1] Jean Berstel. Tracé de droites, fractions continues et distances disccrétes. In M. Lothaire, editor, *Mots*, pages 298–309. Hermes, 1990.

[2] Jean Berstel and Michel Pocchiola. Random generation of finite Sturmian words. *Discrete Math.*, 153(1-3):29–39, 1996.

[3] Jean-Pierre Borel and Christophe Reutenauer. On Christoffel classes. *ITA*, 40(1):15–27, 2006.

[4] Michael Boshernitzan and Aviezri S. Fraenkel. A linear algorithm for nonhomogeneous spectra of numbers. *J. Algorithms*, 5(2):187–198, 1984.

[5] Ethan M. Coven and Gustav A. Hedlund. Sequences with minimal block growth. *Math. Systems Theory*, 7:138–153, 1973.

[6] Aldo De Luca and Alessandro De Luca. Some characterizations of finite Sturmian words. *Theoret. Comput. Sci.*, 356(1-2):118–125, 2006.

[7] Oliver Jenkinson and Luca Q. Zamboni. Characterisations of balanced words via orderings. *Theoret. Comput. Sci.*, 310(1-3):247–271, 2004.

[8] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

[9] M. Lothaire. *Applied Combinatorics on Words*, volume 105 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005.

[10] Sabrina Mantaci, Antonio Restivo, and Marinella Sciortino. Burrows-Wheeler transform and Sturmian words. *Inform. Process. Lett.*, 86(5):241–246, 2003.

[11] Marston Morse and Gustav A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.*, 62:1–42, 1940.

[12] Giuseppe Pirillo. A curious characteristic property of standard Sturmian words. In H. Crapo and D. Senato, editors, *Algebraic Combinatorics and Computer Science*, pages 541–546. Springer, 2001.

[13] Zhi Xiong Wen and Zhi Ying Wen. Some properties of the singular words of the Fibonacci word. *European J. Combin.*, 15(6):587–598, 1994.