Some characterizations of finite Sturmian words

Aldo de Luca, Alessandro De Luca

Dipartimento di Matematica e Applicazioni "R. Caccioppoli"
Università degli Studi di Napoli Federico II
via Cintia, Monte S. Angelo
I-80126 Napoli, Italy

Abstract

In this paper we give some new characterizations of words which are finite factors of Sturmian words. An enumeration formula for primitive finite Sturmian words is given. Moreover, we provide two linear-time algorithms to recognize whether a finite word is Sturmian.

Key words: Sturmian word, standard word, central word
1991 MSC: 68R15

1 Introduction

As is well known (cf. [1]), an infinite word is *Sturmian* if for any $n \geq 0$ the number of distinct factors (or blocks) of length n is exactly $n+1$. This implies that infinite Sturmian words are over a two-letter alphabet $\{a, b\}$.

A finite word is called a *finite Sturmian word* if it is a factor of an infinite Sturmian word. A finite Sturmian word can also be defined by the following combinatorial property, usually called *balance condition* (see for instance [1]). A word w over $\{a, b\}$ is balanced if for all factors u and v of w having the same length, one has

$$||u|_a - |v|_a| \leq 1 \tag{1}$$

where $|u|_a$ and $|v|_a$ denote the number of occurrences of the letter a in u and v respectively. The set of all finite Sturmian words coincides with the set of balanced words on the alphabet $\{a, b\}$.

Email addresses: aldo.deluca@unina.it (Aldo de Luca), aledeluc@studenti.unina.it (Alessandro De Luca).

Preprint submitted to Elsevier Science *3 June 2007*

In this paper we give two new characterizations (cf. Theorems 10 and 15) of finite Sturmian words, that we summarize in the following main theorem. For a nonempty word w over any finite alphabet, we denote by z_w its fractional root, by $\pi_w = |z_w|$ its minimal period, and by R_w the minimal integer k such that w has no right special factor of length k.

Theorem 1 *Let w be a nonempty word. The following conditions are equivalent:*

(1) w is a finite Sturmian word,
(2) the fractional root of w is a conjugate of a standard Sturmian word,
(3) $\pi_w = R_{z_w^2} + 1$.

It is noteworthy that, differently from the balance condition which is concerned with the *composition* (i.e., the number of a or b) in all equally long factors of a given word w, our characterizations are based on properties of the fractional root, which determine the "*periodical structure*" of the word. From these arguments, a simple formula enumerating the finite Sturmian words which are primitive is also derived.

From the applicative point of view, the interest of Theorem 1 lies in the possibility of implementing two new and simple algorithms recognizing whether a finite word of length n is Sturmian, with time complexity $O(n)$.

2 Preliminaries

In the following A will denote an arbitrary finite alphabet and A^* the *free monoid* over A. The elements of A are called *letters* and those of A^* *words*. Let $w = a_1a_2 \cdots a_n \in A^*$ be a word, with $a_i \in A$ for $i = 1, 2, \ldots, n$. The integer n is called the *length* of w and is denoted by $|w|$. The identity element of A^* is called *empty word* and denoted by ε; the length of ε is conventionally 0. By $\mathrm{alph}(w)$ we denote the set of letters of A occurring in w.

We recall that a positive integer p is a *period* of $w = a_1 \cdots a_n$ if whenever $1 \le i, j \le n$ one has:

$$i \equiv j \pmod p \Longrightarrow a_i = a_j \ .$$

We denote by π_w the minimal period of w, assuming $\pi_\varepsilon = 1$. The word w is *unbordered* if its minimal period equals its length, that is, if $\pi_w = |w|$. If w is nonempty, the *fractional root* of w is its prefix z_w of length $|z_w| = \pi_w$.

If $w = a_1 \cdots a_n \in A^*$ is a word, the *reversal* of w is the word $w^\sim = a_n \cdots a_1$. One sets $\varepsilon^\sim = \varepsilon$. A word $w \in A^*$ is a *palindrome* if $w = w^\sim$.

A word u is a *factor* of w if $w = rus$ for some words r, s. In the special case $r = \varepsilon$ (resp., $s = \varepsilon$), we call u a *prefix* (resp., *suffix*) of w. A factor u of w is *proper* if $u \neq w$. The factor u of w is called *median* if $|r| = |s|$. The set of factors of a word w is denoted by Fact(w). For any $X \subseteq A^*$, one sets:

$$\mathrm{Fact}(X) = \bigcup_{w \in X} \mathrm{Fact}(w) \ .$$

Given two words $w, w' \in A^*$, an *overlap* of w with w' is a suffix of w which is also a prefix of w'. One says that w and w' *overlap* if there exists a nonempty overlap of w with w' or of w' with w.

A factor u of w is called *right special* if there exist two letters $a, b \in A$, $a \neq b$, such that ua and ub are both factors of w. We denote by R_w the smallest integer k such that w has no right special factor of length k (and we set $R_\varepsilon = 0$). The following noteworthy inequality (cf. [2]) relates the minimal period π_w of a word w and R_w:

$$\pi_w \geq R_w + 1 \ . \tag{2}$$

Two words $u, v \in A^*$ are *conjugate* if there exist $\lambda, \mu \in A^*$ such that $u = \lambda\mu$ and $v = \mu\lambda$. Conjugacy is an equivalence relation in A^*.

A word w is *primitive* if it cannot be written as a power $w = u^k$ with $k > 1$; we denote by $\pi(A^*)$ the set of all primitive words over A. As is well known (cf. [3]), for any nonempty word w there exists a unique primitive word u such that $w = u^k$, with $k \geq 1$. We remark that the fractional root of any nonempty word is primitive.

Suppose that the alphabet A is totally ordered. One can extend this order to A^* by the *lexicographic* order (cf. [3]). A word is called a *Lyndon* (resp., *anti-Lyndon*) word if it is primitive and minimal (resp., maximal) in its conjugacy class, with respect to the lexicographic order.

In the sequel, we shall need the two following simple lemmas; we report the proofs for the sake of completeness.

Lemma 2 *A word $w \in A^*$ has the period $p \leq |w|$ if and only if all its factors having length p are in the same conjugacy class.*

PROOF. The case $w = \varepsilon$ is trivial. Then suppose that p is a period of $w = a_1 \cdots a_n$, $a_i \in A$, $i = 1, \ldots, n$. Let u be a factor of w of length p. By the definition of period, there exists a positive integer $i \leq p$ such that $u = a_i a_{i+1} \cdots a_p a_1 a_2 \cdots a_{i-1}$, so that u is a conjugate of $a_1 a_2 \cdots a_p$.

The converse is an easy consequence of the following fact: if $x, y \in A$ and $u \in A^*$, then xu is a conjugate of uy if and only if $x = y$. Therefore, if all factors of w of length p are conjugate, one derives that $a_i = a_{i+p}$ for all i such that $1 \leq i \leq n - p$. $\square$

Lemma 3 *If w is a primitive word, then $\pi_{w^2} = |w|$.*

PROOF. By contradiction, suppose w^2 has a period $q < |w|$. Since $|w|$ is a period of w^2 and $|w^2| = 2|w| > |w| + q$, by the periodicity theorem of Fine and Wilf [3], w^2, as well as w, has also the period $d = \gcd(q, |w|)$. Thus $w = u^{|w|/d}$ for some u, which is absurd as $|w| > q$ implies $|w|/d > 1$. $\square$

2.1 Standard and central Sturmian words

In the following, $\mathcal{A}$ will denote the binary alphabet $\mathcal{A} = \{a, b\}$. Let

$$c_0, c_1, \ldots, c_n, \ldots$$

be any sequence of integers such that $c_0 \geq 0$ and $c_i > 0$ for $i > 0$. One can define inductively the sequence of words $(s_n)_{n \geq 0}$ where

$$s_0 = b, \ s_1 = a, \ \text{and} \ s_{n+1} = s_n^{c_{n-1}} s_{n-1} \ \text{ for } n > 1 \ . \tag{3}$$

The sequence $(s_n)_{n \geq 0}$ converges to an infinite Sturmian word called *standard*. We shall denote by *Stand* the set of all the words s_n, $n \geq 0$ of any sequence $(s_n)_{n \geq 0}$ constructed by the previous rules. Any word of *Stand* is called (finite) *standard Sturmian word* or simply *standard word*.

A word w is called *central* if it is a palindromic prefix of an infinite standard Sturmian word. It has been proved (see for instance [1]) that a word w is central if and only if it has two periods p and q which are coprime and $|w| = p + q - 2$. The set of all central words is usually denoted by PER. The following remarkable relation exists between *Stand* and PER (cf. [4]):

$$Stand = \mathcal{A} \cup PER\{ab, ba\} \ , \tag{4}$$

i.e., any standard word which is not a single letter is obtained by appending ab or ba to a central word and conversely, any central word is obtained by deleting the last two letters of a standard word.

As is well known [1], the set St of finite Sturmian words is equal to the set of factors of *Stand*, i.e.,

$$St = \mathrm{Fact}(Stand) \ . \tag{5}$$

Moreover, from (4) one derives also that $St = \mathrm{Fact}(PER)$.

The following properties of central words will be used in the sequel:

Proposition 4 (see [5]) *A word w is central if and only if wab and wba are conjugate.*

Now let us suppose that the alphabet $\mathcal{A}$ is totally ordered by setting $a < b$.

Proposition 5 (see [6]) *The set $\mathcal{A} \cup aPERb$ is equal to the set of all Lyndon words which are Sturmian.*

In a similar way, the set of anti-Lyndon Sturmian words coincides with $\mathcal{A} \cup bPERa$.

Proposition 6 (see [7]) *A Sturmian word is unbordered if and only if it is a Lyndon or anti-Lyndon word.*

From Propositions 4 and 5, one derives the following interesting characterization of words conjugate of a standard word.

Proposition 7 *A primitive word, which is not a single letter, is a conjugate of a standard word if and only if the Lyndon and the anti-Lyndon words in its conjugacy class have the same proper median factor of maximal length.*

PROOF. Let z be a primitive word of length $|z| > 1$. Let s be a standard word conjugate to z. By (4), s can be written as $s = vxy$, with $v \in PER$ and $\{x, y\} = \mathcal{A}$. By Proposition 4, one derives that z is a conjugate of avb and bva. From Proposition 5, avb and bva are, respectively, a Lyndon and an anti-Lyndon word, so that the necessity is proved.

Conversely, let $z \in A^*$ and suppose that the Lyndon and the anti-Lyndon words in the conjugacy class of z can be written respectively as atb and bta, with $a, b \in A$ and $a < b$. By Proposition 4, one has that $t \in PER$, so that by (4), z is a conjugate of $tab \in Stand$. $\square$

3 Finite Sturmian words

In this section we give two characterizations of finite Sturmian words. We need some preliminary propositions. The first gives some characterizations of the words w such that $w^2 \in St$ (such words have been called *cyclic balanced* in [8]). The equivalence of some of the conditions in Proposition 8 has recently been proved in [8] (see also [9]). We report here a more direct and simple proof for the sake of completeness.

Proposition 8 *Let w be a word. The following conditions are equivalent:*

(1) $w^2 \in St$,
(2) $w^* \subseteq St$,
(3) every conjugate of w^2 *is Sturmian,*
(4) every conjugate of w *is Sturmian,*
(5) w *is a Sturmian word, and it is either non-primitive or a conjugate of a standard Sturmian word.*

PROOF. *1.* $\Rightarrow$ *2.* Let $n > 2$. Any two factors of w^n of length $k > |w|/2$ overlap, thus it suffices to verify the balance condition (1) only for factors of w^n of length $k \leq |w|/2$, which is satisfied because such words are also factors of $w^2 \in St$.

2. $\Rightarrow$ *3.* This is trivial, since any conjugate of w^2 is a factor of w^3.

3. $\Rightarrow$ *4.* This is trivial too, because the square of a conjugate of w is just a conjugate of w^2.

4. $\Rightarrow$ *5.* We have to prove that if w is primitive, then it has a conjugate which is a standard word. Indeed, there exists a unique conjugate of w which is a Lyndon word, say u. Since u is Sturmian, by Proposition 5 one has that u is either a letter or a word avb with $v \in PER$. In the former case, the desired standard conjugate is u itself; in the latter case, one can take vba.

5. $\Rightarrow$ *1.* If w is not primitive, it can be written as $w = u^k$ with $u \in \mathcal{A}^*$ and $k > 1$. Hence $u^2 \in St$, and the result follows because *1.* $\Rightarrow$ *2.* and $w^2 \in \mathrm{Fact}(u^*)$. If w is primitive and v is a standard word in its conjugacy class, from equations (3) and (5) one derives that $v^2 \in St$. Since *1.* $\Rightarrow$ *3.* and w^2 is a conjugate of v^2, one has $w^2 \in St$. $\square$

Let $w, u \in A^*$ with w unbordered; the word wu is called a *Duval extension* of w if no unbordered factor of wu is longer than w.

Proposition 9 (see [10]) *Every Duval extension wu of a Sturmian unbordered word w has the period $|w|$.*

We are now in the position of giving our first characterization of finite Sturmian words.

Theorem 10 *A nonempty word is Sturmian if and only if its fractional root is a conjugate of a standard word.*

PROOF. Let w be a word. If its fractional root z_w is a conjugate of a standard word, then by Proposition 8, $z_w^* \subseteq St$, so that $w \in \mathrm{Fact}(z_w^*) \subseteq St$.

Conversely, let s be an unbordered factor of $w \in St$ of maximal length. One has $w = usv$ for suitable $u, v \in \mathcal{A}^*$. The word sv is a Duval extension of s, by the maximality of s. Since $s^\sim$ is unbordered too, and again by the maximality of s, the word $s^\sim u^\sim = (us)^\sim$ is a Duval extension of $s^\sim$. From Proposition 9, one gets that both sv and $(us)^\sim$ have the period $|s|$. This implies that also us has the period $|s|$.

By Lemma 2, all factors of us and sv having length $|s|$ are conjugates of s. Since any factor of w of length $|s|$ is either a factor of us or of sv, and s is a factor of both, we deduce from Lemma 2 that the whole w has the period $|s|$. Moreover, such period is minimal, because

$$|s| = \pi_s \leq \pi_w \leq |s| \ .$$

By Lemma 2, z_w is a conjugate of s; since s is an unbordered Sturmian word, by Proposition 6 it is a Lyndon (or anti-Lyndon) word, and therefore, by Proposition 5 it is in the set $\mathcal{A} \cup aPERb \cup bPERa$. Hence s, as well as z_w, is a conjugate of a standard word, which proves the assertion. □

Example 11 *Let w be the word aababaa. Its fractional root $z_w =$ aabab is a conjugate of the standard word ababa, so that w is Sturmian.*

Let $r =$ baabb. In the conjugacy class of its root $z_r =$ baab there is no standard word, so that r is not Sturmian.

Corollary 12 *Let w be a nonempty word and z_w be its fractional root. Then w is a finite Sturmian word if and only if so is z_w^2.*

PROOF. This is a straightforward consequence of the preceding theorem and of Proposition 8. □

The following proposition has been proved in [11]. We report here a simpler proof for the sake of completeness.

Proposition 13 *Let w be a word. If $\pi_w = R_w + 1$, then w is Sturmian.*

PROOF. Let $w \in A^*$. If $\pi_w = 1$, the result is trivially true. Thus we assume $\pi_w = R_w + 1 > 1$, so that there exists a right special factor s of w such that $|s| = \pi_w - 2$. Hence, there exist letters $a, b \in A$ such that $a \neq b$ and $sa, sb \in \mathrm{Fact}(w)$. The words sa and sb cannot be both suffixes of w, so we suppose, without loss of generality, that sa is not. Therefore one has either $saa \in \mathrm{Fact}(w)$ or $sac \in \mathrm{Fact}(w)$ with $c \neq a$. Since $|saa| = |sac| = \pi_w$, these two possibilities imply, respectively:

$$w \in \mathrm{Fact}((saa)^*) \tag{6}$$

or

$$w \in \mathrm{Fact}((sac)^*) \ . \tag{7}$$

We first show that (6) cannot hold. By contradiction, assume that it does hold. Since sb is a factor of w, it has to be a factor of $saas$ as well. We clearly have $sb \neq sa$, thus there exist $u, v \in A^*$ and $x \in A$ such that $saas = uxsbv$. The words u and v are respectively a prefix and a suffix of s, and $|u| + |v| = |saas| - |xsb| = 2|s| + 2 - |s| - 2 = |s|$. Therefore $s = uv$ and $vaau = xuvb$. But this is a contradiction, because $|vaau|_a > |xuvb|_a$.

Equation (7) is then satisfied. Let $u = sacsa$. The word $sb \in \mathrm{Fact}(w)$ has to be a factor of u; since sb is not a suffix of u, one has either $sba \in \mathrm{Fact}(u)$ or $sbx \in \mathrm{Fact}(u)$, with $x \neq a$. By Lemma 2, the latter is impossible, because $|sab| = |sbx| = \pi_w$ is a period of u, and $|sab|_a > |sbx|_a$. Thus sba is a factor of u, and by Lemma 2 it is a conjugate of sac. Therefore $c = b$; by Proposition 4 and equation (4) one derives that sab is a standard word of length π_w. By Lemma 2, z_w is a conjugate of sab, so that by Theorem 10 one obtains $w \in St$. □

For any word w, a factor u of w is called *left special* if there exist two distinct letters a and b such that au and bu are factors of w. One can define L_w as the minimal integer k for which w has no left special factor of length k. We remark that, by symmetrical arguments, one can prove a result analogous to Proposition 13, namely, if $\pi_w = L_w + 1$, then $w \in St$.

Example 14 *The word $w = abbab$ has minimal period $\pi_w = 3$ and $R_w = 2$, therefore it is Sturmian. The word $v = aabba$ is not Sturmian, and indeed $\pi_v = 4 > 3 = R_v + 1 = L_v + 1$. However, for $u = aabab \in St$ one has $\pi_u = 5 > 4 = \max\{R_u, L_u\} + 1$.*

Theorem 15 *A finite nonempty word w is Sturmian if and only if*

$$\pi_w = R_{z_w^2} + 1 \ . \tag{8}$$

PROOF. Assume (8) holds. By Lemma 3, one has $\pi_{z_w^2} = |z_w| = \pi_w = R_{z_w^2} + 1$, so that from Proposition 13 it follows $z_w^2 \in St$. As $w \in \mathrm{Fact}(z_w^*)$, one obtains $w \in St$ by Proposition 8.

Conversely, let $w \in St$. The result is trivial if $\pi_w = 1$, so assume $|z_w| > 1$. By Theorem 10, z_w is a conjugate of a standard word. Since all conjugates of z_w are factors of z_w^2, by (4) and Proposition 4 there exists $v \in PER$ such that vab and vba are factors of z_w^2, of length π_w. This means that v is a right special factor of z_w^2 of length $\pi_w - 2$; thus $R_{z_w^2} \geq \pi_{z_w^2} - 1$. By (2), one has $\pi_{z_w^2} \geq R_{z_w^2} + 1$, hence $\pi_w = \pi_{z_w^2} = R_{z_w^2} + 1$ as desired. □

We remark that in the case of palindromes, condition (8) in the preceding theorem can be replaced by the equation $\pi_w = R_w + 1$. This is a consequence of Proposition 13 and of a property of Sturmian palindromes proved in [12].

Proposition 16 *Let w be a word having minimal period $\pi_w > 1$ and v be its shortest prefix such that $\pi_v = \pi_w$. Let ux ($x \in A$) be the suffix of v of length $\pi_w - 1$. One has $w \in St$ if and only if there exists a letter $y \neq x$ such that uy is a factor of z_w^2.*

PROOF. If $uy \in \mathrm{Fact}(z_w^2)$, then u is a right special factor of z_w^2 of length $\pi_w - 2$, so that $\pi_w \leq R_{z_w^2} + 1$. By (2) one has $\pi_w = \pi_{z_w^2} \geq R_{z_w^2} + 1$; thus $\pi_w = R_{z_w^2} + 1$ and by Theorem 15 it follows $w \in St$.

Conversely, as shown in the proof of Theorem 10, any word of St has an unbordered factor of maximal length, whose value is the minimal period of the word. Therefore, one can write v as $v = tx$ with $x \in \mathcal{A}$ and $\pi_t < \pi_w$ and t cannot have unbordered factors of length π_w since the maximal length of these factors is π_t. Since $v \in St$, it has an unbordered factor r of maximal length $|r| = \pi_v = \pi_w$. This factor has to be necessarily a suffix of v. Since r is unbordered and $|r| = \pi_w > 1$, from Propositions 5 and 6 one has $r = yux$ with $u \in PER$ and $\{x, y\} = \mathcal{A}$. By Lemma 2, z_w is conjugate of yux and, by Proposition 4, of xuy. Since $xuy \in \mathrm{Fact}(z_w^2)$, the result follows. □

Example 17 *Let $w = aababaa \in St$. One has $\pi_w = 5$, $z_w^2 = aababaabab$, and $R_{z_w^2} = 4$, so that $\pi_w = R_{z_w^2} + 1$. The shortest prefix v of w such that $\pi_v = \pi_w$ is $v = aabab$. Its suffix of length $\pi_w - 1$ is $ub = abab$, and $ua = abaa$ is a factor of z_w^2.*

Let $r = baabb \notin St$. One has $\pi_r = 4$, $z_r^2 = baabbaab$, and $R_{z_r^2} = 2$, so that $\pi_r > R_{z_r^2} + 1$. In this case, the shortest prefix v such that $\pi_v = \pi_r$ is $v = r$. The suffix ub of v of length 3 is abb, and $aba \notin \mathrm{Fact}(z_r^2)$.

3.1 Enumeration of primitive Sturmian words

As an application of preceding results, we give a formula which counts for any $n > 1$ the finite primitive Sturmian words of length n. We need the following:

Lemma 18 *The number of words of length $n > 0$ which are conjugate of standard Sturmian words is 2 if $n = 1$ and $n\phi(n)$ for $n > 1$, where ϕ is Euler's totient function.*

PROOF. For $n = 1$ the result is trivial since the only two words conjugate of standard words are a and b. Let us suppose $n > 1$. As is well known (see for instance [1]), the number of standard words of length $n > 1$ is given by $2\phi(n)$. If s is a standard word, by (4) we can write $s = vxy$ with $\{x, y\} = \{a, b\}$ and $v \in PER$. By Proposition 4, $s' = vyx \in Stand$ is a conjugate of s. In the conjugacy class of s there is no other standard word. Indeed, if $t = uxy$ is a conjugate of s, with $u \in PER$, then $|t|_a = |s|_a$ and $|t|_b = |s|_b$, so that t and s have the same "slope"; from this it follows that $u = v$ (see for instance [6,1]). Hence, in each conjugacy class of a standard word of length $n > 1$ there are exactly two standard words. Thus, the number of these conjugacy classes is $\phi(n)$. Since any standard word is primitive, in any class there are n words. From this the assertion follows. $\square$

Proposition 19 *For any $n > 1$, the number of primitive finite Sturmian words of length n is given by:*

$$\sum_{i=1}^{n}(n+1-i)\phi(i) - \sum_{\substack{d|n \\ d\neq n}} d\phi(d) \ .$$

PROOF. Let w be a non-primitive Sturmian word of length $n > 1$. The word w can be written uniquely as $w = u^k$, with $u \in \pi(\mathcal{A}^*)$ and $k > 1$. Moreover, from Lemma 3 one has $z_w = u$; by Theorem 10, u is a conjugate of a standard word. Since $|w| = k|u|$, the integer $|u|$ is a proper divisor of n. Conversely, if u is a conjugate of a standard word, then by Proposition 8 one has that $u^k \in St$ for any k.

The number of primitive Sturmian words of length n is then obtained by subtracting from $\mathrm{card}(St \cap \mathcal{A}^n)$ the number of words conjugate of a standard word whose length is a proper divisor of n. It is well known (see for instance [1]) that the number of all finite Sturmian words of length n is given by the following formula:

$$\mathrm{card}(St \cap \mathcal{A}^n) = 1 + \sum_{i=1}^{n}(n+1-i)\phi(i) \ .$$

From Lemma 18 it follows

$$\mathrm{card}(St \cap \pi(\mathcal{A}^*) \cap \mathcal{A}^n) = 1 + \sum_{i=1}^{n}(n+1-i)\phi(i) - \left(\sum_{\substack{d|n \\ d\neq n}} d\phi(d) + 1 \right)$$

which proves the assertion. □

4 Algorithms recognizing finite Sturmian words

The problem of finding efficient algorithms for testing whether a finite word is Sturmian is of fundamental importance in discrete geometry for several applications such as pattern recognition, image processing, and computer graphics. Several linear-time algorithms have been found by different authors, using various concepts and techniques (cf. [13] and references therein). In particular, in [14] a linear algorithm which uses methods of elementary number theory is given, and in [15,16] linear algorithms based on methods of discrete geometry are provided. In these latter works an essential role is played by a suitable representation of finite Sturmian words by triplets of integers introduced in [17].

In this section, we give two new and simple linear algorithms for the recognition of Sturmian words, which are based on the combinatorial results on words obtained in the previous section.

A first algorithm to recognize whether a word w of length n is Sturmian can be carried out, by Proposition 7 and Theorem 10, in the following three steps.

(1) Determine the fractional root z_w of w.
(2) Compute the Lyndon word ℓ and the anti-Lyndon word ℓ' in the conjugacy class of z_w.
(3) Compare ℓ and ℓ' and check whether they have the same proper median factor of maximal length.

Step 1 can be executed in linear time; in fact, there exists an algorithm to determine the minimal period π_w (as well as the minimal periods of all prefixes of w) which runs in linear time [18]. Therefore, also the fractional root z_w can be generated in linear time. As regards to step 2, to determine the Lyndon word in the conjugacy class of z_w requires $O(|z_w|)$ time (see [18]). The same occurs for the anti-Lyndon word. Step 3 trivially requires $O(|z_w|)$ time. In conclusion, the preceding algorithm allows one to recognize whether a word is Sturmian or not in linear time.

A second algorithm can be developed as follows, by using Proposition 16.

(1) Determine the fractional root z_w of w.
(2) If $|z_w| = 1$, then $w \in St$; if $\mathrm{alph}(z_w)$ contains more than two letters, then $w \notin St$.
(3) Find the shortest prefix v of w such that $\pi_v = \pi_w$.

(4) Take the suffix ux of v of length $\pi_w - 1$, with $x \in \mathrm{alph}(z_w)$.
(5) Verify if uy (with $y \in \mathrm{alph}(z_w)$, $y \neq x$) is a factor of z_w^2.

As we have already discussed, steps 1 and 3 can be executed in linear time. Steps 2 and 4 trivially require $O(n)$ time, and step 5 can be carried out by a linear-time pattern matching procedure (see for instance [18]). In conclusion, the proposed algorithm runs in linear time.

Acknowledgements

We wish to thank Jean Berstel and Arturo Carpi for their suggestions and helpful comments. The work for this paper was partially supported by the Italian Ministry of Education under project COFIN 2003 "Formal Languages and Automata: methods, models and applications" and by the Faculty of Science of the University of Rome "La Sapienza" under project "Linguaggi formali e automi: teoria e applicazioni".

References

[1] J. Berstel, P. Séébold, Sturmian words, in: M. Lothaire (Ed.), Algebraic Combinatorics on Words, Cambridge University Press, Cambridge UK, 2001, Ch. 2.

[2] A. de Luca, On the combinatorics of finite words, Theoretical Comput. Sci. 218 (1999) 13–39.

[3] M. Lothaire, Combinatorics on Words, Addison-Wesley, Reading MA, 1983.

[4] A. de Luca, F. Mignosi, Some combinatorial properties of Sturmian words, Theoretical Comput. Sci. 136 (1994) 361–385.

[5] G. Pirillo, A curious characteristic property of standard Sturmian words, in: G.-C. Rota, H. Crapo, D. Senato (Eds.), Algebraic combinatorics and computer science, Springer Italia, Milan, 2001, pp. 541–546.

[6] J. Berstel, A. de Luca, Sturmian words, Lyndon words and trees, Theoretical Comput. Sci. 178 (1997) 171–203.

[7] T. Harju, D. Nowotka, Minimal Duval extensions, Internat. J. Found. Comput. Sci. 15 (2004) 349–354.

[8] W.-F. Chuan, Moments of conjugacy classes of binary words, Theoretical Comput. Sci. 310 (2004) 273–285.

[9] O. Jenkinson, L. Q. Zamboni, Characterisations of balanced words via orderings, Theoretical Comput. Sci. 310 (2004) 247–271.

[10] F. Mignosi, L. Q. Zamboni, A note on a conjecture of Duval and Sturmian words, Theoret. Inform. Appl. 36 (2002) 1–3.

[11] A. de Luca, A. De Luca, Combinatorial properties of Sturmian palindromes, preprint (2005).

[12] A. de Luca, A. De Luca, Palindromes in Sturmian Words, in: DLT 2005, vol. 3572 of LNCS, 2005, pp. 199–208.

[13] P. Bhattacharya, R. A. Melter, A. Rosenfeld (Eds.), Vision Geometry, Vol. 119 of Contemporary Mathematics, AMS, Providence, RI, 1991.

[14] M. Boshernitzan, A. S. Fraenkel, A linear algorithm for nonhomogeneous spectra of numbers, J. Algorithms 5 (1984) 187–198.

[15] J. Berstel, M. Pocchiola, Random generation of finite Sturmian words, Discrete Mathematics 153 (1996) 29–39.

[16] M. Lindenbaum, A. Bruckstein, On recursive $o(n)$ partitioning of a digitized curve into digital straight segments, IEEE Trans. Pattern Anal. Machine Intelligence 15 (1993) 949–953.

[17] M. Lindenbaum, J. Koplowitz, A new parametrization of digital straight lines, IEEE Trans. Pattern Anal. Machine Intelligence 13 (1991) 847–852.

[18] M. Lothaire, Applied Combinatorics on Words, Cambridge University Press, Cambridge UK, 2005.