International Journal of Foundations of Computer Science

COMBINATORIAL PROPERTIES OF STURMIAN PALINDROMES

ALDO 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
aldo.deluca@unina.it

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
aledeluc@studenti.unina.it

Received (received date)
Revised (revised date)
Communicated by Editor's name

ABSTRACT

We study some structural and combinatorial properties of Sturmian palindromes, i.e., palindromic finite factors of Sturmian words. In particular, we give a formula which permits to compute in an exact way the number of Sturmian palindromes of any length. Moreover, an interesting characterization of Sturmian palindromes is obtained.

Keywords: Sturmian word; Sturmian palindrome; central word; special factor.

1. Introduction

Sturmian words have been widely studied for their theoretical importance and their applications to various fields of science. By definition, they are infinite words which are not eventually periodic and have minimal *subword complexity*. Sturmian words also enjoy some remarkable characterizations of geometrical nature (*cutting sequences*, *mechanical words*). The reader is referred to [2] and to [1] for general surveys on this subject.

In recent years, some works have investigated Sturmian words by looking at their palindromic factors. A *palindrome* is a finite word which can be read indistinctively from left to right or from right to left.

Palindromes play an essential role in the structure of Sturmian words. In fact, an important theorem of X. Droubay and G. Pirillo [12] shows that an infinite word is Sturmian if and only if it has exactly *one* palindromic factor of length n for n even, and *two* for n odd. Moreover, A. de Luca and F. Mignosi [10] proved that the set of palindromic prefixes of all standard Sturmian words is equal to the set of *central words*, i.e., words having two periods p and q which are coprime, and length $p+q-2$. Central words are words over a two letter alphabet $\{a, b\}$ and satisfy remarkable structural properties. In particular, a central word w is such that wab and wba can be factorized as a product of two palindromes. Moreover, the set St of factors of all Sturmian words is equal to the set of factors of all central words (cf. [10]).

There exist, and even they are the majority, Sturmian palindromes which are not central. For instance, there are 14 Sturmian palindromes of length 7, whereas the number of central words of the same length is only 6 (see Table 2).

In this paper we are interested in the combinatorics and in the structure of the language of Sturmian palindromes. A main theorem, proved in Section 4, gives a simple formula which permits to count for any $n \geq 0$ the Sturmian palindromes of length n. As a consequence, an interesting relation between the numbers of Sturmian palindromes of odd and even length is found. Moreover, it is shown that the number $g(n)$ of Sturmian palindromes of length n has, for all $n \geq 0$, a lower bound of the order $n^{1+\alpha}$, where $\alpha = \log_3 2$. From this result one easily derives that the number of central words of length n is a fraction of $g(n)$ which vanishes when n diverges. Furthermore, one can prove that the density of Sturmian palindromes of length n with respect to all Sturmian words having the same length, tends to 0 when n tends to infinity.

In Section 5 some structural properties of Sturmian palindromes are shown. In particular, two new characterizations of central words are given. Finally, a remarkable characterization of Sturmian palindromes is proved.

A short version of this paper was presented at the conference Developments in Language Theory, held in Palermo on July 4–8, 2005 [9].

2. Preliminaries

Let A be a finite alphabet. As usual, we denote by A^* the *free monoid* generated by A, that is the set of all words over A with the operation of concatenation. The identity element of A^* is the *empty word* ε. Let $w = a_1 a_2 \cdots a_n$ be a word, with $a_i \in A$, $1 \leq i \leq n$. The integer n is called the *length* of w, and it is denoted by $|w|$. Conventionally, the length of the empty word is 0. For any $a \in A$, $|w|_a$ will denote the number of occurrences of the letter a in w.

A word u is a *factor* of $w \in A^*$ if $w = rus$ for some words r and 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 $u \neq w$. A factor u of w is *median* if $w = rus$ with $|r| = |s|$. If u is a factor of w, w is also called an *extension* of u. The set of factors of a word w is denoted by $\mathrm{Fact}(w)$. For any $X \subseteq A^*$, one sets:

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

A factor u of w is called a *border* of w if it is both a prefix and a suffix of w.

Let $w = a_1 a_2 \cdots a_n$ be a word, with $a_i \in A$, $1 \leq i \leq n$. A positive integer p is a *period* of w if for all i, j, $1 \leq i, j \leq n$, the following condition is satisfied:

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

As is well known [14], a word w has a period $p \leq |w|$ if and only if it has a border of length $|w| - p$. The minimal period of w will be denoted by π_w; it is natural to set $\pi_\varepsilon = 1$. For any $w \neq \varepsilon$, the unique positive integer k such that $w = z^k z'$, where $|z| = \pi_w$ and $|z'| < \pi_w$, will be called the *order* of w.

Given $w = a_1 a_2 \cdots a_n$ with all $a_i \in A$, the *reversal* $w^\sim$ is the word $a_n \cdots a_1$. If $w = \varepsilon$, one sets $\varepsilon^\sim = \varepsilon$. A word $w \in A^*$ is a *palindrome* if $w = w^\sim$. The set of all palindromes of A^* is denoted by PAL_A, or simply PAL when there is no ambiguity.

An *infinite word* x is just an infinite sequence of letters:

$$x = a_1 a_2 a_3 \cdots \text{ where } a_i \in A, \text{ for all } i \geq 1 \ .$$

The product between a finite word w and an infinite one x is naturally defined as the infinite word wx. A (finite) *factor* of an infinite word x is either the empty word or any sequence $u = a_i \cdots a_j$ with $i \leq j$, i.e., a finite block of consecutive letters in x. If $i = 1$, then u is a *prefix* of x. A *suffix* of x is an infinite word y such that $x = uy$ for some $u \in A^*$. We shall denote by $\mathrm{Fact}(x)$ the set of finite factors of x.

Let x be a finite or infinite word over A. A factor u of x is a *right special* factor of x if there exist two letters $a, b \in A$, $a \neq b$, such that ua and ub are factors of x.

An infinite word is *Sturmian* if for each $n \in \mathbb{N}$ it has $n + 1$ distinct factors of length n. This implies that a Sturmian word is on a two-letter alphabet that will be denoted by $\mathcal{A} = \{a, b\}$. As is well known [2], an infinite binary word x is Sturmian if and only if for any $n \geq 0$ there is only one right special factor of x of length n.

An equivalent geometrical definition of Sturmian words can be given in terms of *cutting sequences*. In fact, a Sturmian word can be defined by considering the sequence of cuts in a squared lattice ($\mathbb{N} \times \mathbb{N}$) made by a ray having a slope which is an irrational number α. A horizontal cut is denoted by the letter b, a vertical by a, and a cut with a corner by ab or ba.

A Sturmian word represented by a ray starting from the origin is usually called *standard* or the *characteristic word* associated with the irrational α and it is often denoted by c_α. Standard Sturmian words can be equivalently defined as follows. For any sequence $d_0, d_1, \ldots, d_n, \ldots$ of integers such that $d_0 \geq 0$ and $d_i > 0$ for $i > 0$, one defines, inductively, the sequence of words $(s_n)_{n \geq 0}$ where

$$s_0 = b, \ s_1 = a, \text{ and } s_{n+1} = s_n^{d_{n-1}} s_{n-1}, \text{ for } n \geq 1 \ .$$

The sequence $(s_n)_{n \geq 0}$ converges to a limit s which is an infinite standard Sturmian word. Any standard Sturmian word can be generated in this way. If $d_i = 1$ for all $i \geq 0$, one obtains the famous *Fibonacci word*. 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 rule.

Any word of *Stand* is called *finite standard (Sturmian) word*.

2.1. Central words

A word w is called central if it has two periods p and q such that $\gcd(p, q) = 1$ and $|w| = p + q - 2$. Conventionally, the empty word ε is central (in this case, $p = q = 1$). Central words are over a two-letter alphabet. The set of all central words over $\mathcal{A} = \{a, b\}$ is usually denoted by PER. It is well known (see [10, 2]) that the set PER coincides with the set of palindromic prefixes of all standard Sturmian words. In this section we recall some properties of central words which will be useful in the following. As proved in [10], one has that

$$Stand = \mathcal{A} \cup PER\{ab, ba\} \ , \tag{1}$$

i.e., any finite standard Sturmian word which is not a single letter is obtained by appending ab or ba to a central word. Conversely, by deleting the last two letters of a standard word which is not a letter, one obtains a central word.

Let St be the set of finite Sturmian words, i.e., factors of infinite Sturmian words over the alphabet $\mathcal{A} = \{a, b\}$. We recall that for any infinite Sturmian word there exists an infinite standard Sturmian word having the same set of factors (cf. [2]). Therefore one easily derives that $St = \mathrm{Fact}(Stand) = \mathrm{Fact}(PER)$.

The following important characterization of central words holds (see for instance [5]):

Proposition 1 *A word w is central over $\mathcal{A}$ if and only if w is a power of a letter of $\mathcal{A}$ or it satisfies the equation*

$$w = w_1 ab w_2 = w_2 ba w_1$$

for some words w_1 and w_2. Moreover, in this latter case, w_1 and w_2 are central words over $\mathcal{A}$, $p = |w_1| + 2$ and $q = |w_2| + 2$ are coprime periods of w, and $\min\{p, q\}$ is the minimal period of w.

The following simple but useful lemma on central words holds (see [5]):

Lemma 2 *If a central word w has the factor x^n, with $x \in \mathcal{A}$ and $n > 0$, then x^{n-1} is a prefix (and suffix) of w.*

3. Sturmian Palindromes

In the sequel we shall be interested in the set $St \cap PAL$, whose elements will be called *Sturmian palindromes*.

One has that $PER \subseteq St \cap PAL$. However, the previous inclusion is strict since there exist non-central Sturmian palindromes, for instance *abba*.

We have seen that $St = \mathrm{Fact}(PER)$. We shall prove (cf. Corollary 4) a similar property for Sturmian palindromes.

Theorem 3 *Every palindromic factor of a standard Sturmian word c_α is a median factor of a palindromic prefix of c_α.*

The result is attributed to A. de Luca [7] by J.-P. Borel and C. Reutenauer, who gave a geometrical proof in [3]. Theorem 3 can be also obtained as a consequence of a more general result of X. Droubay, J. Justin, and G. Pirillo [11]. We shall report later a direct proof for the sake of completeness.

Corollary 4 *A word is a Sturmian palindrome if and only if it is a median factor of some central word.*

Proof. Trivially, every median factor of a palindrome is itself a palindrome. Since $St = \mathrm{Fact}(PER)$, it follows that a median factor of an element of PER is a Sturmian palindrome.

Conversely, let u be in $St \cap PAL$. By definition, there exists an infinite (standard) Sturmian word s such that $u \in \mathrm{Fact}(s)$. By Theorem 3, u is a median factor of a palindromic prefix of s. Since palindromic prefixes of standard Sturmian words are exactly the elements of PER, the result follows. □

Our proof of Theorem 3, which follows a simple argument suggested by A. Carpi [4], is based on the following results (see [7]):

Proposition 5 *If $w \in \mathrm{Fact}(x)$, where x is an infinite Sturmian word, then the reversal $w^\sim$ is a factor of x too. Moreover, if x is standard, then w is a right special factor of x if and only if $w^\sim$ is a prefix of x.*

Corollary 6 *A palindromic factor of an infinite standard Sturmian word x is a right special factor of x if and only if it is a palindromic prefix of x.*

Proof of Theorem 3. By contradiction, let $c_\alpha = \lambda u x$, where u is a palindrome that is not a median factor of any palindromic prefix of c_α, and $\lambda \in \mathcal{A}^*$ has minimal length for such condition. Since u cannot be a prefix of c_α, we have $|\lambda| \geq 1$. Thus we can assume, without loss of generality, $\lambda = \lambda' a$. Now let z be the first letter of x, so that $x = zx'$. Suppose first $z = a$. The palindrome aua is not a median factor of a palindromic prefix of c_α, otherwise so would be u. But $c_\alpha = \lambda' auax'$ with $|\lambda'| < |\lambda|$, and this contradicts the minimality of $|\lambda|$. Therefore $z = b$, and then aub and $bua = (aub)^\sim$ are factors of c_α, in view of Proposition 5. This means in particular that u is a right special factor of c_α. Corollary 6 then implies that u is a prefix of c_α, a contradiction. □

4. Enumeration of Sturmian Palindromes

In this section we shall give an explicit formula for the enumeration function of $St \cap PAL$. We start by recalling some basic facts (see [10, 7]):

Proposition 7 *Let w be a word. The following conditions are equivalent:*

(i) $w \in PER$,
(ii) awb and bwa are Sturmian,
(iii) awa, awb, bwa, and bwb are all Sturmian.

Proposition 8 *If wa and wb are Sturmian words, then there exists a letter $x \in \mathcal{A}$ such that xwa and xwb are both Sturmian.*

We now prove two easy consequences (see also [7]):

Proposition 9 *Let $w \in \mathcal{A}^*$ be a palindrome. If wa and wb are Sturmian, then w is central.*

Proof. From the previous proposition, there exists a letter $x \in \mathcal{A}$ such that xwa and xwb are both Sturmian. Without loss of generality, we may suppose $x = a$, so that $awb \in St$. Therefore $(awb)^{\sim} = bwa$ is Sturmian too, thus by Proposition 7, w is central. □

Lemma 10 *Let w be a Sturmian palindrome. If w is not central, then there exists a unique letter $x \in \mathcal{A}$ such that xwx is Sturmian.*

Proof. If awa and bwb are both Sturmian, then w is central by Proposition 9, a contradiction. Since by Corollary 4 the word w is a (proper) median factor of some central word, there exists a unique letter $x \in \mathcal{A}$ such that xwx is Sturmian. □

Now let us introduce the function $g : \mathbb{N} \to \mathbb{N}$ defined for all $n \geq 0$ as

$$g(n) := \mathrm{card}(St \cap PAL \cap \mathcal{A}^n) \ .$$

For any $n \geq 0$, $g(n)$ gives the number of Sturmian palindromes of length n.

Theorem 11 *For any $n \geq 0$, the number $g(n)$ of Sturmian palindromes of length n is given by*

$$1 + \sum_{i=0}^{\lceil n/2 \rceil - 1} \phi(n - 2i) \ , \tag{2}$$

where ϕ is Euler's totient function. Equivalently, for any $n \geq 0$

$$g(2n) = 1 + \sum_{i=1}^{n} \phi(2i) \quad \text{and} \quad g(2n+1) = 1 + \sum_{i=0}^{n} \phi(2i+1) \ .$$

Proof. Given $w \in St \cap PAL$, at least one of its extensions awa and bwb is Sturmian. Indeed, according to Lemma 10, if $w \notin PER$, then exactly one of these extensions is in St. If $w \in PER$, then from Proposition 7, both awa and bwb are Sturmian palindromes. Since the number of central words of length n is $\phi(n+2)$ (see [10]), one gets:

$$g(n + 2) = g(n) + \phi(n + 2)$$

and this implies the desired formula, because $g(0) = 1$ and $g(1) = 2$. □

We define a function $f : \mathbb{N} \to \mathbb{N}$ by setting for $n \geq 0$:

$$f(2n) = 1 + \frac{n(n+1)}{2} \quad \text{and} \quad f(2n+1) = 2 + n(n+1) \ .$$

It is easy to verify that $g(n) \leq f(n)$ for all $n \geq 0$. Moreover, for any $n \geq 0$ we set

$$h(n) = \text{card}(PER \cap \mathcal{A}^n) = \phi(n+2) \ .$$

In Table 1 we list the values of the functions g, f, and h for $0 \leq n \leq 17$. As an example, in Table 2 we list all 14 Sturmian palindromes of length 7. The six central words in it are underlined.

Table 1. The functions g, f, and h.

n	$g(n)$	$f(n)$	$h(n)$	n	$g(n)$	$f(n)$	$h(n)$
0	1	1	1	9	20	22	10
1	2	2	2	10	14	16	4
2	2	2	2	11	30	32	12
3	4	4	4	12	18	22	6
4	4	4	2	13	42	44	8
5	8	8	6	14	24	29	8
6	6	7	4	15	50	58	16
7	14	14	6	16	32	37	6
8	10	11	4	17	66	74	18

Table 2. Sturmian palindromes of length 7 (central words are underlined).

$\underline{aaaaaaa}$	$\underline{bbbbbbb}$
$\underline{aaabaaa}$	$\underline{bbbabbb}$
$aababaa$	$bbababb$
$abaaaba$	$babbbab$
$\underline{abababa}$	$\underline{bababab}$
$abbabba$	$baabaab$
$abbbbba$	$baaaaab$

The following proposition relates the numbers of Sturmian palindromes of odd and even length.

Proposition 12 *For any $n > 0$ one has*

$$g(2n-1) = g(4n) - 2g(2n) + 2 \ .$$

Proof. From Theorem 11 one has

$$g(4n) = 1 + \sum_{i=1}^{2n} \phi(2i) \ .$$

As is well known (see for instance [13]), for any $n > 0$ one has $\phi(2n) = \phi(n)$ for odd n and $\phi(2n) = 2\phi(n)$ for even n. Thus we can write

$$\begin{aligned} g(4n) &= 1 + 2 \sum_{\substack{i \text{ even} \\ i \leq 2n}} \phi(i) + \sum_{\substack{i \text{ odd} \\ i < 2n}} \phi(i) \\ &= 1 + 2 \sum_{k=1}^{n} \phi(2k) + \sum_{k=0}^{n-1} \phi(2k+1) \\ &= g(2n-1) + 2(g(2n)-1) \ , \end{aligned}$$

which concludes the proof. □

Now we consider the problem of finding lower bounds for the number of Sturmian palindromes of any length. We premise the following simple lemma, whose proof is reported in the Appendix for the sake of completeness.

Lemma 13 *The Euler totient function has the following lower bounds:*

$$\phi(n) \geq n^{\alpha} \ \textit{for odd } n \ \textit{ and } \phi(n) \geq 2^{-\alpha} n^{\alpha} \ \textit{for even } n \ ,$$

where $\alpha = \log_3 2 = 0.6309\ldots$.

Proposition 14 *Let* $\beta = \frac{1}{2(1+\alpha)}$. *For* $n \geq 0$ *one has:*

$$g(2n+1) \geq (2-\beta) + \beta(2n+1)^{1+\alpha} \tag{3}$$

and

$$g(2n) \geq 1 + \frac{1}{\alpha} n^{1+\alpha} \ . \tag{4}$$

Proof. By Lemma 13, we can write

$$g(2n+1) = 1 + \sum_{i=0}^{n} \phi(2i+1) \geq 2 + \sum_{i=1}^{n} (2i+1)^{\alpha} \ .$$

Approximating the sum with an integral, one has

$$\sum_{i=1}^{n} (2i+1)^{\alpha} \geq \int_{0}^{n} (2x+1)^{\alpha} dx = \beta(2n+1)^{1+\alpha} - \beta \ ,$$

so that (3) follows.

By Lemma 13, we can write

$$g(2n) = 1 + \sum_{i=1}^{n} \phi(2i) \geq 1 + \sum_{i=1}^{n} i^{\alpha} \geq 1 + \int_{0}^{n} x^{\alpha} dx = 1 + \frac{1}{\alpha} n^{1+\alpha} \ ,$$

so that (4) follows. □

As a consequence, one derives that

$$g(n) = \Omega(n^{1+\alpha}) \ .$$

From this result we can prove that the *density* $h(n)/g(n)$ of central words of length n with respect to all Sturmian palindromes of length n, vanishes when n diverges.

Proposition 15 *The following holds:*

$$\lim_{n\to\infty} \frac{h(n)}{g(n)} = 0 \ .$$

Proof. We recall that $h(n) = \phi(n+2) \leq n+1$ for all $n \geq 0$. Since $g(n) = \Omega(n^{1+\alpha})$, i.e., $g(n) \geq dn^{1+\alpha}$ for all $n \geq 0$ and some positive constant d, it follows that for any $n > 0$ one has

$$\frac{h(n)}{g(n)} \leq \frac{n+1}{dn^{1+\alpha}} \ .$$

As the right hand side of last equation vanishes when n diverges, the assertion follows. □

Now let us recall (cf. [2]) that for any $n \geq 0$ the number $st(n) = \text{card}(St \cap \mathcal{A}^n)$ of all finite Sturmian words of length n is given by the following formula:

$$st(n) = 1 + \sum_{i=1}^{n} (n-i+1)\phi(i) \ .$$

We shall prove that the *density* $g(n)/st(n)$ of Sturmian palindromes of length n with respect to all Sturmian words of length n, tends to 0 when n tends to infinity. The proof is based on the following lemma whose proof is reported in the Appendix.

Lemma 16 $st(n) = \Omega(n^{2+\alpha})$.

Proposition 17 *The following holds:*

$$\lim_{n\to\infty} \frac{g(n)}{st(n)} = 0 \ .$$

Proof. From the definition one has that for any n, $g(n) \leq 1 + \Phi(n)$, where

$$\Phi(n) = \sum_{i=1}^{n} \phi(i) \ .$$

As is well known (cf. [13]), $\Phi(n) = O(n^2)$, so that by the previous lemma one has

$$\frac{g(n)}{st(n)} \leq \frac{cn^2}{dn^{2+\alpha}}$$

for all $n > 0$ and some constants $c, d > 0$. Since the right hand side in the previous equation vanishes when n diverges, the result follows. □

Let us observe that, by using the well known result (see for instance [13, Theorem 327]),

$$\lim_{n\to\infty} \frac{\phi(n)}{n^{1-\delta}} = \infty \quad \text{for any } \delta > 0 \ ,$$

one easily derives a sharper asymptotic lower bound for the function g, i.e.,

$$g(n) = \Omega(n^{2-\delta})$$

when n diverges.

5. Structural properties

We have seen in Section 3 that a Sturmian palindrome is a median factor of a central word. In this section we shall give some further results concerning the structure of Sturmian palindromes.

The following lemma is well known (see for instance [7]). We report a proof for the sake of completeness.

Lemma 18 *A palindrome $w \in A^*$ has a period $p \leq |w|$ if and only if it has a palindromic prefix (suffix) of length $|w| - p$.*

Proof. If w has a period $p \leq |w|$, then it has a border v of length $|w| - p$, so that we can write $w = \lambda v = v\mu$ for some words λ and μ. Since w is a palindrome, one has

$$w = v\mu = v^{\sim}\lambda^{\sim} \ .$$

Therefore, $v = v^{\sim}$. Conversely, if the palindrome w has the palindromic prefix v, one has

$$w = v\mu = \mu^{\sim}v \ ,$$

so that v is a border of w and $|w| - |v|$ is a period of w. □

Proposition 19 *A palindrome $w \in A^*$ with minimal period $\pi_w > 1$ can be uniquely represented as*

$$w = w_1xyw_2 = w_2yxw_1^{\sim}$$

with $x, y \in A$, w_2 the longest proper palindromic suffix of w, and $|w_1xy| = \pi_w$. The word w is not central if and only if either $w_1 \notin PAL$ or $w = (w_1xx)^kw_1$ where $k \geq 1$ is the order of w.

Proof. Since $\pi_w > 1$, it follows by Lemma 18 that w can be uniquely factorized as $w = w_1xyw_2$ where w_2 is the longest proper palindromic suffix of w, $x, y \in A$, and $|w_1xy| = \pi_w$. Since w is a palindrome, we can write

$$w = w_1xyw_2 = w_2yxw_1^{\sim} \ .$$

When $\pi_w > 1$, by Proposition 1, w is central if and only if $w_1 \in PAL$ and $x \neq y$. Therefore, in the case $w_1 \in PAL$, w is not central if and only if $w = w_1xxw_2 = w_2xxw_1$. The word w has the two periods

$$\pi_w = |w_1xx| \text{ and } q = |w_2xx| \tag{5}$$

and length $\pi_w + q - 2$. Thus $w \notin PER$ if and only if $d = \gcd(\pi_w, q) > 1$. Since $|w| \geq \pi_w + q - d$, by Fine and Wilf's theorem (cf. [14]) w has the period $d = \pi_w$. This occurs if and only if $q = k\pi_w$ with $k \geq 1$. From (5) this condition is equivalent to the statement $w_2xx = (w_1xx)^k$, i.e., $w = (w_1xx)^kw_1$. □

Example 20 Let $w = aaabaaaaaabaaa \in St \cap PAL$, with $\pi_w = 7$. The word w can be factorized as $(aaaba)aa(aaabaaa)$, where $aaabaaa$ is the longest proper palindromic suffix of w, $|aaaba| = \pi_w - 2 = 5$. The prefix $aaaba$ is not a palindrome, thus w is not central.

Let $v = abaababababaaba \in St \cap PAL$. We factorize v as

$$v = (abaabab)ab(abaaba)$$

where $abaaba$ is the longest proper palindromic suffix of v. Also in this case $abaabab$ is not a palindrome, so that $w \notin PER$.

Let $u = abbabbabba \in St \cap PAL$. We factorize u as $(a)bb(abbabba)$, where $abbabba$ is the longest palindromic suffix of u. In this case, the prefix a is a palindrome, and $u = (abb)^3a$. Hence u is not central.

Lemma 21 *If $w = w_1xyw_2 = w_2yxw_1^{\sim}$, where w_2 is the longest proper palindromic suffix of w and $x, y \in A$, then $w' = ywy$ has the minimal period $\pi_{w'} = \pi_w$.*

Proof. Since w is a factor of w', one has $\pi_{w'} \geq \pi_w$. The word yw_2y is a palindromic proper suffix of $w' = yw_1xyw_2y$, so that by Lemma 18 the word w' has the period $|yw_1x|$. Hence, $\pi_{w'} \leq |yw_1x| = |w_1xy| = \pi_w$. Thus $\pi_w = \pi_{w'}$. □

The next lemma is essentially a restatement of Lemma 2 in [6]. Note that its first part is an obvious consequence of Lemma 21.

Lemma 22 *Let $w = w_1xyw_2 = w_2yxw_1 \in PER$, with $|w_2| > |w_1|$ and $\{x, y\} = \mathcal{A}$. The word $v = ywy$ has minimal period $\pi_v = \pi_w$, whereas $v' = xwx = xw_1xyw_2x$ has minimal period $\pi_{v'} = |w_2| + 2 = |w| - \pi_w + 2$.*

Let $w \in (St \cap PAL) \setminus PER$. We denote by u the (unique) *shortest* median extension of w in PER, and by v the *longest* central median factor of w. Note that also v is unique. For instance, for the Sturmian palindrome $w = baaabaaab$ one has $u = aawaa$ and $v = aaabaaa$.

Theorem 23 *Let $w \in (St \cap PAL) \setminus PER$. With the preceding notation, one has $\pi_u = \pi_w$. Moreover, either $\pi_w = \pi_v$ or $\pi_w = |v| - \pi_v + 2$.*

Proof. We consider first the case that $\pi_v = 1$, so that $v = x^n$ with $x \in \mathcal{A}$ and $n = |v|$. In such a case w has also the median palindromic factor $v_1 = yx^ny$, where $\{x, y\} = \mathcal{A}$ (recall that v is the longest central median factor of w). Moreover, $n = |v|$ is at least 2, otherwise v_1 would be equal to $yxy \in PER$. One has $\pi_{v_1} = |yx^n| = n + 1 = |v| - \pi_v + 2$. Now we define, for $2 \leq i \leq n$:

$$v_i = xv_{i-1}x = x^{i-1}yx^nyx^{i-1} = (x^{i-1}yx^{n-i+1})(x^{i-1}yx^{i-1}) \ . \qquad (6)$$

The word $v_n = x^{n-1}yx^nyx^{n-1}$ is central, whereas for $i < n$ one has, by Lemma 2, that $v_i \notin PER$. From Lemma 10 it follows that the words v_i are the *only* Sturmian extensions of v_1 which are median factors of v_n. Since for $i < n$ one has $v_i \notin PER$, one derives that $w = v_k$ for some $1 \leq k < n$, and $u = v_n$. As shown in (6), by

Lemma 21 all the v_i's have the same minimal period, for $1 \le i \le n$. The result in this case follows: $\pi_w = \pi_u = |v| - \pi_v + 2$.

Now let us assume $\pi_v > 1$. One has $v = w_1xyw_2 = w_2yxw_1$, with $w_1, w_2 \in PAL$ and $x \neq y$. We suppose $|w_1| < |w_2|$, so that $\pi_v = |w_1| + 2$. From the definition of v, it follows that there exists a letter $z \in \mathcal{A}$ such that $v_1 = zvz$ is a median factor of w which is not central. By Lemma 22, we have $\pi_{v_1} = \pi_v$ if $z = y$, or else $\pi_{v_1} = |v| - \pi_v + 2$ if $z = x$.

Using Lemma 21, we shall now define a sequence of palindromes with the same minimal period as v_1. Let us first suppose that $z = y$, so that $v_1 = yw_1xyw_2y$. We set $v_2 = xv_1x = (xyw_1)(xyw_2yx)$. Moreover, if $w_1 = p_1p_2\cdots p_k$ with $p_j \in \mathcal{A}$ for $1 \le j \le k$, we set $v_i = p_{k-i+3}v_{i-1}p_{k-i+3}$ for $i \ge 3$, so that

$$\begin{aligned} v_3 &= p_kv_2p_k = (p_kxyp_1\cdots p_{k-1})(p_kxyw_2yxp_k) \ , \\ &\vdots \\ v_{k+2} &= p_1v_{k+1}p_1 = p_1\cdots p_kxyw_1xyw_2yxp_k\cdots p_1 = w_1xyw_1xyw_2yxw_1^{\sim} \ . \end{aligned}$$

Since $w_1 = w_1^{\sim}$, the last equation can be written as

$$v_{k+2} = (w_1)xy(w_1xyw_2yxw_1) = (w_1xyw_2yxw_1)yx(w_1)$$

showing, by Proposition 19, that the word v_{k+2} is central, so that for any $i \le \pi_v = k+2$ one has $v_i \in St \cap PAL$.

Let $s \le k+2$ be the minimal integer such that $v_s \in PER$. Since for $i < s$ one has $v_i \notin PER$, one derives from Lemma 10 that $u = v_s$ and $w = v_r$ for some integer $r < s$. Hence $\pi_w = \pi_{v_s} = \pi_u$, and in this case $\pi_w = \pi_v$.

The case $z = x$ is similarly dealt with, but interchanging the roles of w_1 and w_2. Thus one assumes $w_2 = q_1\cdots q_k$, $q_j \in \mathcal{A}$, $1 \le j \le k$, and defines v_i as $q_{k-i+3}v_{i-1}q_{k-i+3}$ for $i \ge 3$, starting from $v_2 = yv_1y = (yxw_2)(yxw_1xy)$ and ending with

$$v_{k+2} = w_2yxw_2yxw_1xyw_2 \in PER \ .$$

Therefore there exist integers r and s such that $1 \le r < s \le k+2 = |v| - \pi_v + 2$, $w = v_r$, and $u = v_s$, so that $\pi_w = \pi_u$ and $\pi_w = \pi_{v_1} = |v| - \pi_v + 2$. □

Example 24 Let $w = baaabaaab \in St \cap PAL$. Following the notations of Theorem 23, one has $v = aaabaaa$, $v_1 = w$, and $u = v_3 = aabaaabaaabaa$. Thus $\pi_w = \pi_u = \pi_v = 4$.

Let $w = babbbbab$. In this case we have $v = bbbb$, $w = v_2$, and $u = v_4 = bbbabbbbabbb$, so that $\pi_w = \pi_u = 5 = |v| + 1 = |v| - \pi_v + 2$.

For any word $w \in A^*$, we denote by R_w the minimal integer k such that there is no right special factor of w of length k, and by K_w the length of the shortest unrepeated suffix of w. Conventionally, one assumes $R_\varepsilon = K_\varepsilon = 0$.

There exist some relations among the parameters R_w, K_w, π_w, and $|w|$. The following lemma synthesizes some results proved in [8, Corollary 5.3, Propositions 4.6 and 4.7] which will be useful in the sequel.

Lemma 25 *Let $w \in A^*$. One has:*

$$|w| \geq R_w + K_w \ \text{ and } \pi_w \geq R_w + 1 \ .$$

Moreover, the following holds:

- *if $\pi_w = R_w + 1$, then $|w| = R_w + K_w$,*
- *if $|w| = R_w + K_w$, then for any n there exists at most one right special factor of w of length n.*

The following theorem gives a further criterion, different from Proposition 19, to discriminate whether a palindrome is central or not.

Theorem 26 *Let $w \in A^*$ be a palindrome with $\pi_w > 1$. Then w is central if and only if its prefix of length $\pi_w - 2$ is a right special factor of w.*

Proof. From Proposition 19, we can write

$$w = w_1 xy w_2 = w_2 yx w_1^{\sim} \tag{7}$$

where $x, y \in A$, w_2 is the longest proper palindromic suffix of w, $|w_1| = \pi_w - 2$, and w is central if and only if $w_1 \in PAL$ and $x \neq y$. Therefore we have to prove that w_1 is a right special factor of w if and only if $w_1 = w_1^{\sim}$ and $x \neq y$.

Indeed, assume that these two latter conditions are satisfied. Since $w_1^{\sim} = w_1$ and w_2 is the longest proper palindromic suffix (and prefix) of w, one has that w_1 is a border of w_2. This implies, from (7), that w_1 is a right special factor of w.

Conversely, suppose w_1 is a right special factor of w. Let us first prove that $w_1 \in PAL$. By hypothesis, we have $\pi_w - 2 = |w_1| \leq R_w - 1$, that is $R_w \geq \pi_w - 1$. By Lemma 25 one has $\pi_w \geq R_w + 1$, so that $\pi_w = R_w + 1$. This implies $|w| = R_w + K_w$, again by Lemma 25. The suffix $w_1^{\sim}$ of w is repeated, because w_1 is a right special factor of w, which is a palindrome. This leads to

$$\pi_w - 2 = |w_1^{\sim}| \leq K_w - 1$$

and thus to $|w| = R_w + K_w \geq 2\pi_w - 2$. If $|w| = 2\pi_w - 2$, then $|w_1| = |w_2|$ so that one derives $w_1 = w_2 \in PAL$. If $|w| \geq 2\pi_w - 1$, then w has the prefix $w_1 xy w_1 x$, so that $y w_1 x \in \mathrm{Fact}(w)$. Since w_1 is a right special factor of w, there exists a letter $z \neq x$ such that $w_1 z \in \mathrm{Fact}(w)$. Moreover, since $w_1 z$ is not a prefix, there exists a letter y' such that $y' w_1 z \in \mathrm{Fact}(w)$. One has $y \neq y'$, for otherwise $y w_1$ would be a right special factor of w of length $\pi_w - 1 = R_w$, which is a contradiction. As w is a palindrome, the words $x w_1^{\sim} y$ and $z w_1^{\sim} y'$ are factors of w too, so that $w_1^{\sim}$ is a right special factor of w. By Lemma 25, one obtains $w_1 = w_1^{\sim}$. Therefore we get $w_1 \in PAL$ again.

We shall now prove that $x \neq y$. By contradiction, suppose w has the factorization

$$w = (w_1 xx)^k w_1 \text{ , with } k \geq 1$$

as granted by Proposition 19. Since w_1 is a right special factor of w, one has $w_1 z \in \mathrm{Fact}(w)$ for a suitable letter $z \neq x$. Thus we have either $w_1 z = x w_1$ or $w_1 z =$

v_2xxv_1z, where v_1z is a prefix of w_1 and v_2 is a suffix of w_1. Since $|w_1| = |w_1z| - 1$, we can write $w_1 = v_1z\alpha v_2$, with $\alpha \in A$. The first case is impossible since w_1 is a palindrome and $x \neq z$. In the latter case, one obtains:

$$v_1 z\alpha v_2 = w_1 = w_1^{\sim} = v_1^{\sim} xxv_2^{\sim}$$

which is absurd again, because $x \neq z$. □

Example 27 The word $w = baab$ is a Sturmian palindrome of minimal period $\pi_w = 3$. Its prefix of length 1 is not a right special factor, hence $w \notin PER$. The word $v = abababbababa$ is a Sturmian palindrome having minimal period 7, and its prefix $ababa$ of length 5 is not right special. Therefore $v \notin PER$. On the contrary, the word $u = aabaabaa$ has minimal period 3, and its prefix of length 1 is a right special factor, so that u is central.

Proposition 28 *Let $w \in A^*$. If $\pi_w = R_w + 1$, then w is Sturmian.*

Proof. 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 $|s| = \pi_w - 2$. Thus there exist letters $a, b \in A$, $a \neq b$, such that $sa, sb \in \mathrm{Fact}(w)$. 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 $c \in A$ and $c \neq a$. Since $|saa| = |sac| = \pi_w$, these two possibilities imply, respectively:

$$w \in \mathrm{Fact}((saa)^*) \tag{8}$$

or

$$w \in \mathrm{Fact}((sac)^*)\ . \tag{9}$$

We first show that (8) 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 (9) is then satisfied. We shall prove that s is a central word over $\mathcal{A} = \{a, b\}$, and then the assertion will follow directly from (9) and from (1), since every power of a finite standard word is Sturmian.

Indeed, $sb \in \mathrm{Fact}(w)$ is a factor of $sacs$. Since $sb \neq sa$, there exist $u, v \in A^*$, $x \in A$ such that $sacs = uxsbv$. As before, one has $|u| + |v| = |s|$, so that $s = uv$ and then $vacu = xuvb$. The letter x has to be equal to a in order to have $|vacu|_a = |xuvb|_a$. Hence $vacu = auvb$. Since $|vacu|_b = |auvb|_b$, one derives $c = b$ and

$$vabu = auvb\ . \tag{10}$$

If $u = \varepsilon$, one obtains $va = av$, so that $s = v \in a^* \subset PER$. Similarly,

$$v = \varepsilon \Longrightarrow bu = ub \Longrightarrow s = u \in b^* \subset PER\ .$$

Finally, if $|u|$ and $|v|$ are both positive, then by (10) there exist $u', v' \in A^*$ such that $v = av'$ and $u = u'b$, and

$$s = uv = u'bav' = v'abu' \ ,$$

whence $s \in PER$, by Proposition 1. □

The converse of the preceding proposition is in general not true, as shown in the Example 30. However, the result is true for Sturmian palindromes as the next theorem shows.

Theorem 29 *A palindrome $w \in A^*$ is Sturmian if and only if $\pi_w = R_w + 1$.*

Proof. By Proposition 28, the condition is sufficient. Necessity is trivially true if $\pi_w = 1$. By Lemma 25 one has $\pi_w \geq R_w + 1$ for any word w. Hence, if $\pi_w > 1$ the condition $\pi_w = R_w + 1$ is equivalent to the existence of a right special factor s of w of length $|s| = \pi_w - 2$.

We prove that every Sturmian palindrome w such that $\pi_w \geq 2$ has such a factor. If w is central, the result follows directly from Theorem 26. Thus we suppose $w \notin PER$, and as in Theorem 23 we denote by v the central median factor of w of maximal length.

If $\pi_v = 1$, then there exists a letter $x \in \mathcal{A}$ and an integer $n \geq 1$ such that $v = x^n$. From the maximality condition, one derives that $n > 1$. In this case, by Theorem 23 one derives $\pi_w = |v| + 1 = n + 1$ and $yx^n y \in \mathrm{Fact}(w)$, where $\{x, y\} = \mathcal{A}$; therefore x^{n-1} is the desired right special factor of w, of length $n - 1 = \pi_w - 2$.

If $\pi_v > 1$, by using Proposition 1 we can write v as $v_1xyv_2 = v_2yxv_1$, with $\pi_v = |v_1xy|$. By Theorem 23, one has either $\pi_w = \pi_v$ or $\pi_w = |v| - \pi_v + 2$. In the first case, the result is a consequence of Theorem 26. Indeed, the prefix v_1 of the central word v, whose length is $\pi_v - 2 = \pi_w - 2$, is a right special factor of v, and then of w. In the latter case (that is, if $\pi_w = |v| - \pi_v + 2$), one derives that the word $xvx = xv_1xyv_2x = xv_2yxv_1x$ is a factor of w, so that v_2 is a right special factor of w, of length $|v| - \pi_v = \pi_w - 2$. □

Example 30 The word $u = ababaa$ is not a palindrome, but $\pi_u = 5 = R_u + 1$, thus it is Sturmian. However, the word $v = aabab \in St$ has $\pi_v = 5 > 3 = R_v + 1$. Let $w = abba \in St \cap PAL$. One has $\pi_w = 3 = R_w + 1$. The palindrome $s = aabbaa$ is not Sturmian. One has $\pi_s = 4 > 3 = R_s + 1$.

A factor u of a word $w \in A^*$ 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 results analogous to Proposition 28 and Theorem 29, namely, *if $\pi_w = L_w + 1$, then $w \in St$*, and *a palindrome $w \in A^*$*

is Sturmian if and only if $\pi_w = L_w + 1$.

Acknowledgments

We thank Arturo Carpi and Maurizio Laporta for some helpful suggestions. 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.-P. Allouche and J. Shallit, *Automatic Sequences* (Cambridge University Press, Cambridge UK, 2003), chapters 9–10.
2. J. Berstel and P. Séébold, "Sturmian words", in M. Lothaire, *Algebraic Combinatorics on Words* (Cambridge University Press, Cambridge UK, 2002), chapter 2.
3. J.-P. Borel and C. Reutenauer, "Palindromic factors of billiard words", *Theoretical Computer Science*, **340** (2005) 334–348.
4. A. Carpi, Private communication (2005).
5. A. Carpi and A. de Luca, "Codes of central Sturmian words", *Theoretical Computer Science*, **340** (2005) 220–239.
6. F. D'Alessandro, "A combinatorial problem on trapezoidal words", *Theoretical Computer Science*, **273** (2002) 11–33.
7. A. de Luca, "Sturmian words: structure, combinatorics, and their arithmetics", *Theoretical Computer Science*, **183** (1997) 45–82.
8. A. de Luca, "On the combinatorics of finite words", *Theoretical Computer Science*, **218** (1999) 13–39.
9. A. de Luca and A. De Luca, "Palindromes in Sturmian words", in *DLT 2005, Proc. 9th Int. Conf. Developments in Language Theory*, eds. C. De Felice and A. Restivo, vol. 3572 of LNCS (Springer, Berlin, 2001), pp. 199–208.
10. A. de Luca and F. Mignosi, "Some combinatorial properties of Sturmian words", *Theoretical Computer Science*, **136** (1994) 361–385.
11. X. Droubay, J. Justin, and G. Pirillo, "Episturmian words and some constructions of de Luca and Rauzy", *Theoretical Computer Science*, **255** (2001) 539–553.
12. X. Droubay and G. Pirillo, "Palindromes and Sturmian words", *Theoretical Computer Science*, **223** (1999) 73–85.
13. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (Oxford University Press, Oxford UK, 1979).
14. M. Lothaire, *Combinatorics on Words* (Addison-Wesley, Reading MA, 1983).

Appendix A:

Proof of Lemma 13. The case $n = 1$ is trivial. Let us factorize an integer $n > 1$ uniquely as $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$, where for $1 \le i \le r$ the p_i are primes, $k_i \ge 1$, and $p_1 < p_2 < \cdots < p_r$. As is well known (cf. [13]), Euler's function ϕ is related to the primes p_i by the following relation:

$$\phi(n) = n \prod_{i=1}^{r} \frac{p_i - 1}{p_i} \quad . \tag{A.1}$$

Let us first suppose that n is *odd*, so that $p_1 \geq 3$. By (A.1) one derives

$$\phi(n) \geq n \left(\frac{2}{3}\right)^r \text{ and } n \geq 3^r \quad ,$$

so that $r \leq \log_3 n$ and $\phi(n) \geq n(2/3)^{\log_3 n} = n^{\log_3 2} = n^\alpha$.

Now suppose n *even*. We can write $n = 2^k m$ with m odd and $k \geq 1$. From the multiplicative property of ϕ, one has $\phi(n) = \phi(2^k)\phi(m) = 2^{k-1}\phi(m)$. From the preceding result, one has

$$\phi(m) = \phi\left(\frac{n}{2^k}\right) \geq \left(\frac{n}{2^k}\right)^\alpha \quad ,$$

so that

$$\phi(n) \geq 2^{k-1}\left(\frac{n}{2^k}\right)^\alpha = 2^{k(1-\alpha)-1} n^\alpha \geq 2^{-\alpha} n^\alpha \quad . \qquad \square$$

Proof of Lemma 16. By Lemma 13 one has $\phi(n) \geq 2^{-\alpha} n^\alpha$ for any $n > 0$, so that

$$st(n) \geq 1 + 2^{-\alpha} \sum_{i=1}^{n} (n - i + 1) i^\alpha = 1 + 2^{-\alpha}(n+1) \sum_{i=1}^{n} i^\alpha - 2^{-\alpha} \sum_{i=1}^{n} i^{1+\alpha} \quad .$$

Since

$$\sum_{i=1}^{n} i^\alpha \geq \int_0^n x^\alpha dx \text{ and } \sum_{i=1}^{n} i^{1+\alpha} \leq \int_0^{n+1} x^{1+\alpha} dx \quad ,$$

one obtains

$$st(n) \geq 2^{-\alpha} \frac{n+1}{\alpha+1} n^{1+\alpha} - 2^{-\alpha} \frac{(n+1)^{2+\alpha}}{\alpha+2} = 2^{-\alpha}(n+1) n^{1+\alpha} r(n)$$

where

$$r(n) = \frac{1}{\alpha+1} - \frac{1}{\alpha+2}\left(\frac{n+1}{n}\right)^{2+\alpha} \quad .$$

The function r is strictly increasing and it satisfies, for all $n \geq 3$, the inequality

$$0 < r(n) < \frac{1}{(\alpha+1)(\alpha+2)} \quad ,$$

so that $st(n) \geq 2^{-\alpha}(n+1) n^{1+\alpha}\, r(3)$. Therefore there exists a constant $d > 0$ such that for all $n \geq 0$, $st(n) \geq d n^{2+\alpha}$, i.e., $st(n) = \Omega(n^{2+\alpha})$. $\square$