ELSEVIER

Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE @ DIRECT®

Theoretical Computer Science 304 (2003) 365–385

Theoretical Computer Science

www.elsevier.com/locate/tcs

Some properties of the factors of Sturmian sequences[☆]

Wei-Ting Cao[a], Zhi-Ying Wen[b,*]

[a] University of Illinois, 1409 W. Green Street, Urbana, IL 61801, USA
[b] Department of Applied Mathematics, School of Sciences, Tsinghua University, Beijing 100084, China

Received 20 December 2001; accepted 5 March 2003
Communicated by D.-Z. Du

Abstract

In this paper, we introduce the singular words of Sturmian sequences, which play an important role in studying the properties of the factors of Sturmian sequence. We also completely determine the powers of the factors, the overlaps of the factors and the structure of the palindromes of the factors.

MSC: primary 11b85; secondary 68R15

Keywords: Sturmian sequence; Singular word; Power; Overlap; Palindrome

1. Introduction

Sturmian sequence, as a kind of aperiodic sequences with minimal language complexity, have been studied for a long time. These sequences are related to many different objects and appear in the mathematical literature under many different names, such as rotation sequences, cutting sequences, Christoffel words, Beatty sequences, characteristic sequences, balanced sequences, and so forth. A clear exposition of early work by J. Bernouli, Christoffel, and A. Markov is given in the book by Venkov [19]. The

☆ Supported by the Special Funds for Major State Basic Research Projects of China.
* Corresponding author.
E-mail addresses: wcao1@uiuc.edu (W.-T. Cao), zwen@math.tsinghua.edu.cn, wenzy@tsinghua.edu.cn (Z.-Y. Wen).

doi:10.1016/S0304-3975(03)00213-5

term ‘Sturm’ was used by Hedlund and Morse [9] in their development of symbolic dynamics. There is much literature about properties of these sequences (see for example Series [17,8,18]). From a combinatorial point of view, they have been considered by Brown [5], Séébold [16], Mignosi [14] and Ito and Yasutomi [10] (in particular in relation with iterated morphisms). Sturmian words appear also in ergodic theory, computer graphics and quasi-crystal. For a survey, we refer the readers to Berstel [2] or Lothaire [13].

The main aim of this paper is to study the combinatorial properties of the factors of Sturmian sequences, such as powers of factors, overlaps of factors and the structure of palindrome factors. By using singular words introduced in [20], Wen and Wen studied these properties for a class of Sturmian sequences which are generated by invertible substitutions (see [20,21]). We first introduce the singular words for general Sturmian sequences, then we completely determine the powers of factors, overlaps of factors and the structure of palindrome factors. As we will see, the positive separation property of the singular words plays an important role in the studies. For example, we give a simple proof of the index of Sturmian sequences obtained by Damanik and Lenz [6], which we proved independently in 1998.

This paper is organized as follows. We first give some preliminaries in Section 2. In Section 3, we introduce the standard word A_n which is also an important class of factors. Sections 4 and 5 are dedicated to the notions and properties of singular words w_n of Sturmian sequence. We establish two decompositions of the Sturmian sequence by singular words, and prove the positive separation property of the singular words. Then in Section 7, by using singular words we study systematically the power of factors, the overlap properties of the factors and the structure of the palindrome factors.

2. Preliminaries

Let $S=\{l_1,l_2,\dots,l_k\}$ be an alphabet with k letters $l_1,l_2,\dots,l_k$. A finite string $u=u_1u_2u_3\dots u_n$ with $u_i\in S$ is called a word over S, while an infinite string $u=u_1u_2u_3\dots u_n\dots$ with $u_i\in S$ is called a sequence over S. We denote by S^* the set of all words and by S^ω the set of all sequences. The concatenation of two words $u=u_1u_2\cdots u_r$, $v=v_1v_2\cdots v_s$ is defined as $u_1u_2\cdots u_rv_1v_2\cdots v_s$ and denoted by uv. u^n is the concatenation of n copies of u. The concatenation of a word and a sequence can be defined similarly. Under the operation of concatenation, S^* forms a monoid where the neutral element is the empty word ε. The length of a word w is denoted by $|w|$ and the number of appearances of a letter $l\in S$ in a word w is denoted by $|w|_l$. $L(w)$ denotes the k-dimensional vector $(|w|_{l_1},|w|_{l_2},\dots,|w|_{l_k})$. We say a word u is a factor of another word w, written $u\prec w$, if there exist two words $v_1,v_2\in S^*$ such that $w=v_1uv_2$. In this case, we say $(|v_1|,u)$ is an occurrence of u in w. The occurrence of a word or a sequence in a sequence is defined in a similar way. If $w=uv$, we say u (resp. v) is a left (resp. right) factor of w, written $u\lhd w$ (resp. $v\rhd w$). A word u is a factor of a sequence $F\in S^\omega$ if there exist a word v and a sequence F' such that $w=vuF'$; if $v=\varepsilon$, we say u is a left factor of F, and note $u\lhd F$.

Let $w = x_1 \cdots x_n$ and $u = x_r x_{r+1} \cdots x_n \triangleright w$, we denote by wv^{-1} the word $x_1 x_2 \cdots x_{r-1}$. Throughout this paper, the expression wu^{-1} conveys this meaning. We denote by $\bar{w}$ the mirror image of w, that is, $\bar{w} = x_n x_{n-1} \cdots x_2 x_1$. If $w = \bar{w}$, the word w will be called a *palindrome*. The set of all palindromes is denoted by $\mathbb{P}$. A word $w \in S^*$ is called *primitive* if $w = u^p \Rightarrow p = 1$. Let $w \in S^*$ and $0 \leqslant k < |w|$, we define the kth conjugate of w by $C_k(w) := x_{k+1} \cdots x_{|w|} x_1 x_2 \cdots x_k$. The set of conjugates of w is defined by $C(w) := \{C_k(w);\ 0 \leqslant k < |w|\}$.

The language of length n of a sequence F, denoted by $\Omega_n(F)$, is the set of all factors of F of length n. The language of F is defined as $\Omega(F) := \bigcup_{n \geqslant 0} \Omega_n(F)$, i.e. the set of all factors of F. The complexity function of F is defined as $p_n(F) := \#\Omega_n(F)$. A sequence F over an alphabet of 2 letters is called Sturmian if $\#\Omega_n(F) = n + 1$.

Throughout this paper, we assume $S = \{a, b\}$, an alphabet with 2 letters.

Lemma 1. *The conjugates of a primitive word w are all different.*

Proof. Let $w = w_1 \cdots w_{|w|}$. Suppose to the contrary, there exists $0 \leqslant m < n \leqslant |w| - 1$ such that $C_m(w) = C_n(w)$, which means

$$w_{m+1} \cdots w_{|w|} w_1 \cdots w_m = w_{n+1} \cdots w_{|w|} w_1 \cdots w_n.$$

Let $u_1 = w_{m+1} \cdots w_n$ and $u_2 = w_{n+1} \cdots w_{|w|} w_1 \cdots w_m$, and we have $u_1 u_2 = u_2 u_1$. By Lothaire [12], there exist two integers $p, q > 0$ and a word $u_0 \in S^*$ such that $u_1 = u_0^p$ and $u_2 = u_0^q$, which implies $w = u_0^r$ with $r \geqslant 2$, contradiction. □

A sequence $F \in S^\omega$ is called a *balanced sequence* if for any $w_1, w_2 \prec F$ with $|w_1| = |w_2|$, we have $\|w_1|_a - |w_2|_a| \leqslant 1$.

Consider a line $y = \theta x + \eta$ $(x \geqslant 0)$ over the plane with θ irrational in $\mathbb{R}^+$ and η real. If the line cuts a vertical (resp. horizontal) line, we write letter a (resp. b). If it cuts lines at some lattice point, we write ab or ba. The sequence obtained is called a cutting sequence and we note $F_{\theta,\eta}$.

The following theorem says that Sturmian sequence, balanced sequence and cutting sequence are the same thing.

Theorem 1 (Ferenzy [7]). *Suppose $F \in S^\omega$, then the following assertions are equivalent*:
1. *F is a Sturmian sequence*;
2. *F is a cutting sequence*;
3. *F is a noneventually periodic balanced sequence*.

Remark 1. Let $F_1, F_2 \in S^\omega$ be two sequences over S. We say that F_1 and F_2 have the same language if $\Omega(F_1) = \Omega(F_2)$. This means F_1 and F_2 have the same set of factors. If we are only interested in the properties of the factors, we do not distinguish two sequences having the same language. It is easy to prove (see for example [7]) that for any θ and for any η_1, η_2, $\Omega(F_{\theta,\eta_1}) = \Omega(F_{\theta,\eta_2})$. Hence in this paper, we only consider the cutting sequence $F_\theta := F_{\theta,0}$.

Remark 2. It is easy to see that the sequence $F_{1/\theta}$ can be obtained by changing the letter a (resp. b) to b (resp. a) in the sequence F_θ. So, to analyze the properties of the Sturmain sequence, we only need to consider the case $\theta \in [0,1]$.

3. Standard words and their properties

Damanik and Lenz introduced the standard words by a direct manner (see for example [6]) and obtained some of their properties. We introduce them in this paper from a geometrical view and give some properties (maybe some overlaps with [6]) that will be used later.

Let $\theta \in [0,1]$ be an irrational, and consider the cutting sequence F_θ generated by the line $l_\theta : y = \theta x$ $(x \geqslant 0)$. A lattice (q,p) on the plane is called an *asymptotic point* if the vertical distance (or equivalently, the horizontal distance or orthogonal distance) from (q,p) to the line l_α is the shortest among the distances from the points whose first coordinate is not greater than q. Such points can be uniquely ordered by the first and the last coordinates. By convention we let $A_0 := (1,0)$. Suppose $A_n := (q_n, p_n)$ is the nth asymptotic point, and let Q_n be the square which contains the foot of the perpendicular from A_n to l_θ. It is easy to see that the line l_θ cuts Q_n twice. Reading from the next cutting point of the original to the second cutting point in the square Q_n, we get a word which will be called the *standard word of order n* and denoted also by A_n. By convention, we take $A_0 = a$ and $A_{-1} = b$.

In order to discuss the properties of the sequence of standard words, we collect some important and useful facts about the continued fraction which can be found in [11].

Let irrational $\theta \in (0,1)$ have a continued fraction expansion $\theta = [0; a_1, a_2, \ldots, a_n, \ldots]$ with $a_n \in \mathbb{N}$, and let p_n/q_n be its nth convergent which is defined recursively by $p_{n+1} = a_{n+1} p_n + p_{n-1}$, $q_{n+1} = a_{n+1} q_n + q_{n-1}$ with $p_0 = 0$, $q_0 = 1$, $p_1 = 1$ and $q_1 = a_1$.

Proposition 1. *For any irrational* $\theta \in (0,1)$, *we have the following*:
(1) *for any* $n,m \geqslant 0$, $p_{2n}/q_{2n} < p_{2n+2}/q_{2n+2} < \theta < p_{2m+1}/q_{2m+1} < p_{2m-1}/q_{2m-1}$;
(2) *for any* $n > 0, (q_n, p_n) = 1$, *that is*, *all convergents are irreducible*;
(3) *for any rational fraction* $\frac{s}{t}$ *with* $1 \leqslant t < q_n$, $|t\theta - s| > |q_n\theta - p_n|$.

Theorem 2. *The point* (q_n, p_n) *is the nth asymptotic point of the sequence* F_θ *if and only if* p_n/q_n *is the nth convergent of the continued fraction of* θ.

Proof. Since the successive convergents of θ are also ordered by the numerator and denominator, we need only to prove that (q,p) is an asymptotic point if and only if p/q is a continued fraction convergent. By the definition of the asymptotic point, we see that (q,p) is an asymptotic point if and only if for any $s,t \in \mathbb{N}$, $1 \leqslant t < q$, $|q\theta - p| < |t\theta - s|$. Thus by Proposition 1.3, it is equivalent to say that p/q is an convergent of θ. $\square$

Proposition 2. *Under the above notations*, *we have for any* $n \in \mathbb{N}$
(1) $A_{n-1} \lhd A_n \lhd F_\theta$, *and* $ab \rhd A_{2n+1}$, $ba \rhd A_{2n}$;

(2) $d_n = a_{n+2}d_{n+1} + d_{n+2}$, *where* $d_n := |p_n - q_n\theta|$ *is the vertical distance from the asymptotic point* A_n *to the line* l_θ;
(3) $L(A_n) = (p_n, q_n)$ *and* $|A_n| = p_n + q_n$;
(4) $|A_{n+2}| = a_{n+2}|A_{n+1}| + |A_n|$.

Proof. (1) This follows directly from the definition of the standard words.

(2) By the definition of d_n and Proposition 1.1, we have $d_{2n+1} = p_{2n+1} - q_{2n+1}\theta$ and $d_{2n} = q_{2n}\theta - p_{2n}$, and the conclusion follows from the recursive relations of p_n and q_n.

(3) Because the segment OA_n cuts vertical lines p_n times and horizontal lines q_n times, we get $L(A_n) = (p_n, q_n)$, and so $|A_n| = p_n + q_n$.

(4) The conclusion is from (3) and the recursive relations of p_n and q_n. □

The following theorem gives the recursive relation of the standard words $\{A_n\}_{n\geqslant 0}$ which is very useful for us to further study the properties of Sturmian sequences.

Theorem 3. *Let* A_n *be the nth standard word of* F_θ. *Then for any* $n\geqslant 0$,

$$A_{n+1} = A_n^{a_{n+1}}A_{n-1}.$$

Proof. We prove it by induction on n.

The case $n = 0, 1$ can be checked directly.

By Proposition 2(1) and (4), $|A_{n+1}| = a_{n+1}|A_n| + |A_{n-1}|$ and $A_{n-1} \lhd A_n \lhd A_{n+1} \lhd F_\alpha$, thus for $n\geqslant 2$, we need only to prove that $A_n^{a_{n+1}+1} \lhd F_\alpha$.

First we consider the case $n = 2k$. By Proposition 2.1, $ba \rhd A_n$, $ab \rhd A_{n+1}$. Let l_θ be the associated line. Consider a_{n+1} lines $l_i : y = \theta x + id_n$ $(1\leqslant i\leqslant a_{n+1})$. We denote by S_i and T_i $(1\leqslant i\leqslant a_{n+1})$, respectively, the intersection points of l_i with y-axis and line $x = q_n$. We denote the point $(q_n, q_n\theta)$ by T_0.

Since $A_nT_{i-1} = OS_i$ $(1\leqslant i\leqslant a_{n+1})$, the cutting sequence starting from T_{i-1} is equal to the cutting sequence starting from S_i with the slope θ (here A_n is the nth asymptotic point associated with the line l_θ). On the other hand, by Proposition 2.2, $OS_i = id_n < d_{n-1}$, which implies that the nth standard word A_n is the prefix of the sequence starting from any S_i. So there exist words w_i such that $w_0 = A_nw_1$, $w_i = A_nw_{i+1}$ $(0\leqslant i\leqslant a_{n+1})$, which implies $A_n^{a_{n+1}+1} \lhd F_\theta$.

The case of n being odd can be proved in the same way (in this case, we will draw the lines $y = \theta x - id_n$). □

From now on, we will always assume that $\alpha, \beta \in S$ and $\alpha \neq \beta$.

The following proposition can be proved easily by induction.

Proposition 3. *Let* $n\geqslant 0$ *and* $\beta \rhd A_n$, *then*

$$A_nA_{n-1} = A_{n-1}A_n\beta^{-1}\alpha^{-1}\beta\alpha, \quad A_{n-1}A_n = A_nA_{n-1}\alpha^{-1}\beta^{-1}\alpha\beta.$$

The following proposition summarizes some elementary properties of the standard words.

Proposition 4. *Let F_θ be Sturmian and let A_n be the standard words of F_θ, then*
(1) *for any $n \geqslant 0$, $A_n \lhd A_{n+1}$, $A_n \rhd A_{n+2}$;*
(2) *for any $n \geqslant 0$, $A_n^2 \prec F_\theta$, $A_n A_{n+1} \prec F_\theta$;*
(3) *for any $n, m \in \mathbb{N}$, $A_n A_m \prec F_\theta$;*
(4) $a^{a_1+2} \nprec F_\theta$, $b^2 \nprec F_\theta$;
(5) *for any $n \geqslant 1$, $ba^{a_1+1}b \rhd A_{2n+1}$;*
(6) *any factor of F_θ placed between two adjacent b's is either a^{a_1} or a^{a_1+1};*
(7) *for any $n \geqslant 0$, A_n is primitive.*

Proof. (1) This is the consequence of Theorem 3.

(2) From Theorem 3, we have

$$A_{n+3} = A_{n+2}^{a_{n+3}} A_{n+1} = (A_{n+1}^{a_{n+2}} A_n)^{a_{n+3}} A_{n+1} = w_1 A_n A_{n+1} = w_1 A_n A_n w_2 \prec F_\theta,$$

where $w_1, w_2 \prec F_\alpha$. This implies $A_n A_{n+1}, A_n^2 \prec A_{n+3} \prec F_\theta$.

(3) If $m \leqslant n$, since $A_m \lhd A_n \lhd F_\theta$ and $A_n^2 \prec A_{n+3} \lhd F_\theta$, we get $A_n A_m \lhd A_n A_n \prec F_\theta$.

If $n < m$ and they have the same parity, then $A_n \rhd A_m$ by (1), and $A_n A_m \rhd A_m^2 \prec F_\theta$. If m, n have different parity, the similar discussion shows that $A_n A_m \rhd A_{m-1} A_m \prec F_\theta$.

(4) From (3) we have $ba^{a_1}b \rhd A_3 A_1$ and $A_3 A_1 \prec F_\theta$. The result follows immediately from the balance property of F_θ.

(5) By Theorem 3 and the definitions of A_{-1}, A_0, $ba^{a_1+1}b \rhd A_2 A_1 \rhd A_3 \rhd A_{2n+1}$.

(6) This follows from (4), (5) and the balance property of F_θ.

(7) If $A_n = w^k$ for some word w and integer $k > 1$, then we have $(p_n, q_n) = L(A_n) = L(w^k) = (k|w|_a, k|w|_b)$. This contradicts the fact p_n/q_n being irreducible. □

4. Singular words and their properties

In this section we study first two special kinds of factors, and as we will see, they are the powerful tools in the study of the factor properties of Sturmian sequences.

Let $\{A_n\}_{n \geqslant -1}$ be the standard words of the Sturmian sequence F_θ and $\beta \rhd A_n$, define

$$w_n := \alpha A_n \beta^{-1}, \quad P_n := \beta A_n^{a_{n+1}-1} A_{n-1} \alpha^{-1}.$$

By Proposition 4, both w_n and P_n are the factors of F_θ. The words w_n and P_n are called the *singular word* of order n of F_θ and the *adjoining word of* w_n, respectively. Since $A_{-1} = b$, $A_0 = a$, we have $w_{-1} = a$, $w_0 = b$. For convenience, we take further $A_{-2} = w_{-2} = P_{-1} = \varepsilon$. We denote by $\mathbb{S} := \mathbb{S}(F_\theta) := \bigcup_{n=-2}^{\infty} \{w_n\}$ the set of all singular words of F_θ.

The following lemma illustrates the structure of w_n and P_n.

Lemma 2. *Let $n \geqslant 0$ and $\beta \rhd A_n$, then*
(1) $\beta\alpha^{-1} w_n = \beta A_n \beta^{-1} = w_{n-1} P_{n-1}$, $w_n \alpha^{-1} \beta = P_{n-1} w_{n-1}$;
(2) $w_{n+1} = w_{n-1} P_{n-1} P_n = P_n P_{n-1} w_{n-1}$;
(3) $P_n = (w_{n-1} P_{n-1})^{a_{n+1}-1} w_{n-1}$;
(4) $w_{n+1} = (w_{n-1} P_{n-1})^{a_{n+1}} w_{n-1}$.

Proof. Since $\beta \triangleright A_n$, we have $\alpha \triangleright A_{n-1}$ and $\beta\alpha \triangleright A_{n+1}$.

(1) The case $n=0$ can be checked easily. Suppose $n \geqslant 1$, then by the definitions of w_n, P_n and Theorem 3, we have

$$\beta\alpha^{-1}w_n = \beta A_n\beta^{-1} = \beta A_{n-1}^{a_n}A_{n-2}\beta^{-1} = \beta A_{n-1}\alpha^{-1}\alpha A_{n-1}^{a_n-1}A_{n-2}\beta^{-1} = w_{n-1}P_{n-1},$$

$$w_n\alpha^{-1}\beta = \beta A_{n-1}^{a_n}A_{n-2}\beta^{-1}\alpha^{-1}\beta = \beta A_{n-1}^{a_n-1}A_{n-2}A_{n-1}\alpha^{-1} = P_{n-1}w_{n-1}.$$

(2) As in (1), we get

$$w_{n+1} = \beta A_{n+1}\alpha^{-1} = \beta A_n^{a_{n+1}}A_{n-1}\alpha^{-1} = \beta A_n\beta^{-1}\beta A_n^{a_{n+1}-1}A_{n-1}\alpha^{-1} = w_{n-1}P_{n-1}P_n.$$

(3) The case of $n=0$ can be checked directly. For $n \geqslant 1$, we have

$$P_n = \beta A_n^{a_{n+1}-1}A_{n-1}\alpha^{-1} = (\beta A_n\beta^{-1})^{a_{n+1}-1}(\beta A_{n-1}\alpha^{-1}) = (w_{n-1}P_{n-1})^{a_{n+1}-1}w_{n-1}.$$

(4) The conclusion follows from (2) and (3). □

By induction, we can easily get the following corollary.

Corollary 1. *For any $n \geqslant -1$, $w_n, P_n \in \mathbb{P}$, that is, all words w_n and P_n are palindromes.*

Corollary 2. *The left and right factors of length $|A_{n-2k}|$ of w_n are w_{n-2k}.*

Proof. This follows directly from the fact that $w_{n-2} \triangleleft w_n$, $w_{n-2} \triangleright w_n$ by Lemma 2.4. □

Proposition 5. *Let A_n be the nth standard word of F_θ and $C(A_n)$ the set of the conjugates of A_n, then*

(1) *For $0 \leqslant k < |A_n|$, $C_k(A_n)$ is either a palindrome or a product of two palindromes. Moreover, for $0 \leqslant k \leqslant |A_{n-1}|-1$, $C_k(A_n) = uP_{n-1}v$ with $vu = w_{n-1}$; and for $|A_{n-1}| \leqslant k \leqslant |A_n|-1$, $C_k(A_n) = uw_{n-1}v$ with $vu = P_{n-1}$.*

$$C_{|A_n|-1}(A_n) = w_{n-1}P_{n-1}, \quad C_{|A_{n-1}|-1}(A_n) = P_{n-1}w_{n-1}.$$

(2) *All elements of $C(A_n)$ are different.*
(3) *$C(A_n) = \overline{C(A_n)}$, where $\overline{C(A_n)} = \{\bar{w};\ w \in C(A_n)\}$.*
(4) *$\Omega_{|A_n|}(A_nA_n) = C(A_n)$.*
(5) *$w_n \notin C(A_n)$.*
(6) *$\Omega_{|A_n|} = C(A_n) \cup w_n$.*
(7) *$\overline{\Omega_{|A_n|}} = \Omega_{|A_n|}$.*
(8) *For any $n \geqslant 2$, $\Omega_{|A_n|}(A_{n-1}A_n) = w_n \cup \{C_k(A_n);\ 0 \leqslant k \leqslant |A_{n-1}|-2\}$. In particular, as a factor, w_n appears only once in $A_{n-1}A_n$.*

Proof. (1) By Lemma 2(1), $C_{|A_n|-1}(A_n) = w_{n-1}P_{n-1}$, which is a product of two palindromes by Corollary 1. It is easy to see that a conjugate of a product of two palindromes is either a palindrome or a product of two palindromes.

Since $|A_{n-1}| = |w_{n-1}|$, the second follows directly.

(2) By Proposition 4(7), the word A_n is primitive, hence the conclusion follows from Lemma 1.

(3) By (1), for any $0 \leqslant k < |A_n|$, $C_k(A_n)$ is either a palindrome, or a product of two palindromes. If $C_k(A_n)$ is a palindrome, then $\overline{C_k(A_n)} = C_k(A_n) \in C(A_n)$; if $C_k(A_n)$ is a product of two palindromes, then there exist $u, v \in \mathbb{P}$ such that $C_k(A_n) = uv$. Thus

$$\overline{C_k(A_n)} = \overline{uv} = vu = C_{k+|u|}(A_n) \in C(A_n).$$

This proves $\overline{C(A_n)} \subset C(A_n)$, the reverse inclusion can be proved in the same way.

(4) It is obvious.

(5) By Proposition 2(3) and the definition of w_n, we have $L(w_n) = (p_n - 1, q_n + 1)$ or $(p_n + 1, q_n - 1)$. On the other hand, for any $0 \leqslant k < |A_n|$, $L(C_k(A_n)) = L(A_n) = (p_n, q_n)$. The conclusion follows.

(6) Since F_θ is Sturmian, $\#\Omega_{|A_n|}(F_\theta) = |A_n| + 1$. Thus by (2) and (5), we have $\Omega_{|A_n|} = C(A_n) \cup \{w_n\}$.

(7) The conclusion follows from (3) and Corollary 1.

(8) Assume $\alpha\beta \triangleright A_n$, then by Proposition 3, $A_{n-1}A_n = A_nA_{n-1}\alpha^{-1}\beta^{-1}\alpha\beta$ and $A_nA_{n-1} \triangleleft A_nA_n$. Therefore the first $p = |A_{n-1}| - 2$ factors of length $|A_n|$ belong to $C(A_n)$, the last factor of length $|A_n|$ is A_n, and the $(|A_{n-1}| - 1)$th factor is w_n. □

From the above, we conclude that any factor of F_θ of length $|A_n|$ must be contained either in $A_{n-1}A_n$ or in A_nA_n. By Proposition 5, we can see that the set of factors of F_θ of length $|A_n|$ consists of conjugates of A_n and the singular word w_n. The discussions below will show that, as a factor, the singular word w_n has some special properties.

Proposition 6. *Let w_n be the singular word of order n of F_θ, then*

(1) *for any $n \geqslant 1$, we have*

$$L(w_n) = \begin{cases} (p_n + 1, q_n - 1) & \text{if } n \text{ is odd}, \\ (p_n - 1, q_n + 1) & \text{otherwise}; \end{cases}$$

(2) $w_n \not\prec w_{n+1}$;

(3) *for any $n \geqslant 1$, $w_{2n+1} = a^{a_1+1}ua^{a_1+1}$, $w_{2n} = bvb$, where $u, v \in \mathbb{P}$;*

(4) *for any $n \geqslant 2$, $1 \leqslant k < |A_n|$, $C_k(w_n) \not\prec F_\theta$, i.e., any proper conjugation of w_n is not a factor of F_θ;*

(5) *for any $n \geqslant 2, w_n$ is not a product of two palindromes;*

(6) *for any $n \geqslant 2, w_n$ is primitive;*

(7) *for any $n \geqslant 0, {w_n}^2 \not\prec F_\theta$.*

Proof. (1) If n is odd, then $b \triangleright A_n$, so $L(w_n) = L(aA_nb^{-1}) = (p_n + 1, q_n - 1)$; the case of n being even can be proved in the same way.

(2) Notice that by the definition of the singular words and Theorem 3, we have

$$w_{n+1} = \alpha A_{n+1}\beta^{-1} = \alpha A_n^{a_{n+1}} A_{n-1}\beta^{-1} \prec A_n^{a_{n+1}+3}.$$

So if $w_n \prec w_{n+1}$, then $w_n \prec w_{n+1} \prec A_n^{a_{n+1}+3}$, which yields $w_n \in C(A_n)$, and contradicts Proposition 5(5).

(3) From Proposition 4(5), $a^{a_1+1}b \triangleright A_{2n+1}$ and $ba \triangleright A_{2n}$. This gives the equalities of the conclusion. The words u and v are palindromes since w_n is a palindrome.

(4) From (3), any proper conjugation of w_n contains either b^2 or a^{a_1+2} as its factor. But Proposition 4(4) says neither b^2 nor a^{a_1+2} is a factor of F_θ, contradiction.

(5) Assume that $w_n = uv$, $u, v \in \mathbb{P}$. Since $w_n \in \mathbb{P}$, $w_n = \overline{w_n} = \bar{v}\bar{u} = vu$, thus $w_n = C_{|u|}(w_n)$ which contradicts (4).

(6) Suppose that w_n is not primitive, then there exists an integer $p \geqslant 2$ such that $w_n = u^p . w_n \in \mathbb{P}$ implies $u \in \mathbb{P}$ which implies further that w_n is the product of two palindromes, contradiction.

(7) Notice that $w_0^2 = b^2$ and $w_1^2 = a^{2a_1+2}$, $w_0^2, w_1^2 \not\prec F_\theta$ by Proposition 4(4).

Now suppose $n \geqslant 2$. If $w_n^2 \prec F_\alpha$, then $C(w_n)$ will be the factors of F_θ, this contradicts (4). □

A factor $w \prec F_\theta$ is called a *special word* of F_θ if both wa and wb are factors of F_θ. The special words introduced first by Berstel [3] for studying the factor properties of Fibonacci sequence. Since $\Omega_n(F_\theta) = n + 1$, there exists a unique special word of F_θ of length n. The following theorem determines all special words.

Theorem 4. *Let $w \prec F_\theta$. Then w is a special word if and only if there exists $n \in \mathbb{N}$ such that $w \triangleleft \bar{A}_n$.*

Proof. Since $w \triangleleft \bar{A}_n \Leftrightarrow \bar{w} \triangleright A_n$, the conclusion of the part "only if" follows from Propositions 2(1) and 4(3). The part "if" is thus from the uniqueness of the special word for any length. □

5. Decompositions of Sturmian sequence by singular words

In this section, we will be able to establish two different decompositions of Sturmian sequence F_θ by singular words and their adjoining words which will be used to study the properties of the factors of F_θ.

Lemma 3. *Let $\alpha \triangleright A_{n+1}$, then*

$$A_{n+1} = \left(\prod_{i=0}^{n} P_i\right)\alpha,$$

$$w_{n+1} = \beta \prod_{i=0}^{n} P_i,$$

$$w_{2n+2} = \prod_{i=n}^{0} (w_{2i}P_{2i})^{a_{2i+2}} w_0, \quad w_{2n+1} = \prod_{i=n}^{0} (w_{2i-1}P_{2i-1})^{a_{2i+1}} w_{-1}.$$

Proof. Since $\alpha \triangleright A_{n+1}$, $\beta \triangleright A_n$. By Theorem 3 and the definition of P_n, we get

$$A_{n+1}\alpha^{-1} = A_n^{a_{n+1}} A_{n-1}\alpha^{-1} = A_n\beta^{-1}(\beta A_n^{a_{n+1}-1} A_{n-1}\alpha^{-1}) = A_n\beta^{-1}P_n.$$

Repeating the above discussion, we get finally

$$A_{n+1}\alpha^{-1} = A_0 a^{-1} P_0 P_1 \cdots P_n = \prod_{i=0}^{n} P_i.$$

Since $w_{n+1} = \beta A_{n+1}\alpha^{-1}$, the second equality follows immediately. Similar arguments show that

$$\begin{aligned} w_{2n+2} &= bA_{2n+2}a^{-1} = bA_{2n+1}^{a_{2n+2}} A_{2n} a^{-1} \\ &= (bA_{2n+1}b^{-1})^{a_{2n+2}} bA_{2n}a^{-1} = (w_{2n}P_{2n})^{a_{2n+2}} w_{2n} \\ &= \prod_{i=n}^{0} (w_{2i}P_{2i})^{a_{2i+2}} w_0, \end{aligned}$$

The fourth equality can be obtained in the same way. □

From Proposition 2, we know that for any $n \in \mathbb{N}$, $A_n \lhd F_\theta$. This fact combined with Lemma 3 gives the following decomposition of F_θ with respect to P_i.

Theorem 5. $F_\theta = \prod_{i=0}^{+\infty} P_i$.

Now we introduce another decomposition of F_θ. With this decomposition, we will establish a "positive separation" property of the singular words which, as we will see, is a powerful tool in studying the combinatorial properties of the factors.

Let $\pi : S \to S^*$ be a mapping with $\pi(a) = u$ and $\pi(b) = v$, which we also denote by $\pi = (u, v)$. Let $F = x_1 x_2 \cdots x_n \in S^*$, and we define $\pi(F) = \pi(x_1)\pi(x_2)\cdots\pi(x_n)$; i.e. the word $\pi(F)$ obtained by replacing the letters a and b in F by the words u and v, respectively. We also denote by $F(u, v)$ the word $\pi(F)$. For $F \in S^\omega$, we can define $\pi(F)$ in the same way.

Assume that $\theta = [0; a_1, a_2, \ldots, a_n, \ldots]$ is the continued fraction expansion of the irrational θ and let $r_n := r_n(\theta) := [0; a_{n+1}, a_{n+2}, \ldots]$. Then $a_n + r_n$ is the nth remainder of θ.

Theorem 6. *With the above notations, we have*
(1) *for any* $n \geqslant 1$, $F_\theta = F_{r_n}(A_n, A_{n-1})$;
(2) *for any* $n \geqslant 1$,

$$F_\alpha = \left(\prod_{i=0}^{n} P_i\right) F_{r_{n+2}}((w_n P_n)^{a_{n+2}-1} w_n w_{n+1}, w_n P_n).$$

Proof. (1) Let R_k be the kth standard word of F_{r_n}, then we need only to prove

$$A_{n+k} = R_k(A_n, A_{n-1}). \tag{*}$$

We prove equality (*) by induction on k.

The cases $k = -1$ and 0 are trivial.

Suppose that (*) is true for any positive integers less than k, then

$$A_{n+k+1} = A_{n+k}^{a_{n+k+1}} A_{n+k-1} = R_k^{a_{n+k+1}}(A_n, A_{n-1}) R_{k-1}(A_n, A_{n-1}) = R_{k+1}(A_n, A_{n-1}).$$

The first and the third equalities are due to Theorem 3 and the second equality is due to the hypotheses of induction. This completes the proof of conclusion (1).

(2) Suppose $\beta \triangleright A_n$, then

$$\begin{aligned}
F_\theta &= F_{r_{n+2}}(A_{n+2}, A_{n+1}) = F_{r_{n+2}}(A_{n+1}^{a_{n+2}} A_n, A_{n+1}) \\
&= A_{n+1} F_{r_{n+2}}(A_{n+1}^{a_{n+2}-1} A_n A_{n+1}, A_{n+1}) \\
&= A_{n+1}\alpha^{-1} F_{r_{n+2}}(\alpha A_{n+1}^{a_{n+2}-1} A_n \beta^{-1}\beta A_{n+1}\alpha^{-1}, \alpha A_{n+1}\alpha^{-1}) \\
&= \left(\prod_{i=0}^{n} P_i\right) F_{r_{n+2}}(P_{n+1} w_{n+1}, w_n P_n) \\
&= \left(\prod_{i=0}^{n} P_i\right) F_{r_{n+2}}((w_n P_n)^{a_{n+2}-1} w_n w_{n+1}, w_n P_n).
\end{aligned}$$

The first two equalities are due to (1) and Theorem 3, respectively. The third and the fourth equalities can be checked by the definition of $F(u, v)$, and the last two equalities come from Lemmas 2 and 3(1). □

The decomposition in Theorem 6(2) is called *the composition of F_α with respect to the singular words of order n*.

By Lemma 3 and Theorem 6(2),

$$\begin{aligned}
\beta F_\theta &= w_{n+1} F_{r_{n+2}}((w_n P_n)^{a_{n+2}-1} w_n w_{n+1}, w_n P_n) \\
&= \prod_{i=1}^{\infty} t_i \quad (*),
\end{aligned}$$

where $t_{2i} = w_n$, $t_{2i+1} = w_{n+1}$ or P_n. This shows βF_θ is the concatenation of infinitely many copies of w_n, w_{n+1}, P_n. So for a fixed n, there are infinite many occurrences of w_n in F_θ. We denote by $w_{n,k}$ the kth occurrence of w_n indicated by (*) above, i.e. $w_{n,k} = (\sum_{i=1}^{2k-1} |t_i| - 1,\ w_n)$. Let $\mathbb{W}(n) := \{w_{n,k}\}_{k \geqslant 1}$. Now we are going to prove that w_n occurs nowhere else except at $w_{n,k}$, i.e. $\mathbb{W}(n)$ contains all occurrences of w_n in F_θ.

Suppose to the contrary, some occurrence of w_n equals none of $w_{n,k}$. Then w_n occurs in the "middle" of the following concatenations which appear in (*) above:

1. $P_n w_n$; 2. $w_n P_n$; 3. $w_{n+1} w_n$; 4. $w_n w_{n+1}$;
5. $w_n P_n w_n$ (if $|P_n| < |w_n|$); 6. P_n (if $|w_n| \leqslant |P_n|$)

The following lemma shows that cannot happen.

Lemma 4. *For any $n \geqslant 0$, we have*

(1) $w_n \not\prec P_n$; $P_n \not\prec w_n$.

This implies w_n could not be a prefix of (1), (3), (6), *or a suffix of* (2), (4), (6).

(2) *Assume $z = xy = u_1 u_2 u_3 \prec F_\theta$, where one of x, y is w_n, the other is P_n or w_{n+1}, is one of the first 4 words defined above with $0 < |u_1| < |x|$, $0 < |u_3| < |y|$, then $u_2 \notin \mathbb{S}$. This implies w_n could not be situated in the middle of the* (1)–(4).

(3) *Assume* $z = w_n P_n w_n = u_1 u_2 u_3$ *with* $0 < |u_1|, |u_3| < |w_n|$. *Then* $u_2 \notin \mathbb{S}$. *This implies* w_n *could not be situated in the middle of* (5).

Proof. (1) By Lemma 3(1), $\prod_{i=0}^{n} P_i = A_{n+1}\alpha^{-1} = \beta^{-1} w_{n+1}$. By Proposition 6(2), $w_n \not\prec w_{n+1}$, which implies $w_n \not\prec \prod_{i=0}^{n} P_i$.

If $a_{n+1} = 1$, $P_n = w_{n-1} \not\prec w_n$. If $a_n \geqslant 2$, $|P_n| > |w_n|$. This proves $P_n \not\prec w_n$.

(2) We only prove the case $z = P_n w_n$; the other four cases can be proved in the same way.

Let $z = P_n w_n$ and $\beta\alpha \triangleright A_n$, then $\alpha\beta \triangleright A_{n-1}$. By the definitions of A_n, w_n, P_n, we have

$$z = u_1 u_2 u_3 = P_n w_n = \alpha A_n^{a_{n+1}-1} A_{n-1} \beta^{-1} \beta A_n \alpha^{-1} = \alpha A_n^{a_{n+1}-1} A_{n-1} A_n \alpha^{-1}.$$

We now prove $u_2 \notin \mathbb{S}$.

(i) Since $|u_2| < |P_n w_n| = |w_{n+1}|$, $u_2 \neq w_{n+1}$.

(ii) Notice that

$$u_2 \prec \alpha^{-1} P_n w_n \beta^{-1} = A_n^{a_{n+1}-1} A_{n-1} A_n \alpha^{-1} \beta^{-1} = A_n^{a_{n+1}-1} A_n A_{n-1} \beta^{-1} \alpha^{-1} \prec A_n^{a_{n+1}+1},$$

where the second equality follows from Proposition 3. By Proposition 6(4), $w_n \notin C(A_n)$, thus $u_2 \neq w_n$.

(iii) Now we prove that for any $-1 \leqslant i < n$, $u_2 \neq w_i$. Suppose that $|u_2| = |w_i| = |A_i|$, $-1 \leqslant i < n$. If $i = n - 2k$, then

$$A_{n-2k} A_{n-2k-1} \beta^{-1} \triangleright P_n, \quad \beta A_{n-2k} \alpha^{-1} \triangleleft w_n,$$

but $|u_1| < |P_n|$, so

$$u_2 \prec A_{n-2k} A_{n-2k-1} A_{n-2k} \beta^{-1} \alpha^{-1} \prec A_{n-2k}^3$$

hence, $u_2 \in C(A_{n-2k})$. Since $w_{n-2k} \notin C(A_{n-2k})$, $u_2 \neq w_{n-2k}$.

If $i = n - 2k - 1$, we can get in the same way as above $u_2 \prec A_{n-2k-1}^3$ and so $u_2 \neq w_{n-2k-1}$.

(3) This follows from (1). □

Let $F = x_1 x_2 \cdots x_n \cdots \in A^\omega$ and $u = x_n x_{n+1} \cdots x_{n+s-1}$, $v = x_{n+m} \cdots x_{n+m+t-1}$ two factors of F. We say that the occurrences (n, u) and $(n+m, v)$ are *positively separated* in F if $m > s$ and we call the word $x_{n+s} \cdots x_{n+m-1}$ the *separating factor* of the two occurrences; otherwise (if $m \leqslant s$), we say the occurrences of u and v are not positively separated.

Let $\{(p_n, u_n)\}_{n \geqslant 1}$ be a finite or infinite sequence of occurrences of factors of F. We say that the sequence $\{(p_n, u_n)\}_{n \geqslant 1}$ is positively separated in F if any two adjacent occurrences (p_n, u_n) and (p_{n+1}, u_{n+1}) are positively separated. Let $\{(p_n, u_n)\}_{n \geqslant 1}$ be a positively separated sequence in F and let $v_n \prec F$ be the separating factor situated between u_n and u_{n+1} (by convention, v_0 is the factor before u_1). We call the sequence of occurrences $\{p_n + |u_n|, v_n\}$ of separating factors $\{v_n\}_{n \geqslant 0}$ the separating sequence with respect to the sequence $\{(p_n, u_n)\}_{n \geqslant 1}$.

From Theorem 6 and Lemma 4, we get immediately

Theorem 7. *Let $n \in \mathbb{N}$ be fixed. Then*
(1) $\mathbb{W}(n) = \{w_{n,k}\}_{k \geqslant 1}$ *is the sequence of occurrences of* w_n *in* F_θ;
(2) *the sequence* $\mathbb{W}(n)$ *is positively separated in* F_θ (*we say also that the singular word* w_n *is positively separated*);
(3) *the separating sequence with respect to* $\mathbb{W}(n)$ *consists of the words* P_n *and* w_{n+1} (*except for* $v_0 := \prod_{i=0}^{n} P_i = \beta^{-1} w_{n+1}$).

The following corollary also follows directly from Lemma 4.

Corollary 3. *Let* $F_\theta = \prod_{i=0}^{+\infty} P_i$ *be the decomposition of* F_θ, *and* $w_n \prec F_\theta$ *be a singular word of order* n, *then every* w_n *must be completely contained in some* P_i.

Corollary 4. *Let* $u \prec F_\alpha$ *with* $|A_n| < |u| \leqslant |A_{n+1}|$ *for some* $n \geqslant 0$. *Suppose that* $w \prec u$ *is a singular word of the highest order contained in* u, *then*
(1) *w must be one of the following four singular words*: $w_{n-2}, w_{n-1}, w_n, w_{n+1}$;
(2) *if* $w = w_{n-2}$, *then* w *appears in* u *exactly* a_n *times*; *if* $w = w_{n-1}$, *then* w *may appear in* u *from one to* a_{n+1} *times*; *if* $w = w_n$ *or* w_{n+1}, *then* w *appears exactly once in* u.

Proof. (1) The restriction on the length of u shows w can take one of the words of w_{n-2}, w_{n-1}, w_n, and w_{n+1}.

Since $|w_{n+2}| > |A_{n+1}| \geqslant |u|$, $w \neq w_{n+2}$. Now suppose $w = w_{n-3}$. Consider the decomposition of F_α by singular words of order $n-3$ and notice that $|u| > |A_n|$, and we can see that w_{n-3} appears in u at least a_{n-1} times. Thus from Theorem 6(2), u will contain either one w_{n-2} or one w_{n-1}, and this contradicts maximality of w in u.

(2) If $w = w_{n-2}$, then
(i) the separating factor between two adjacent occurrences of w_{n-2} in u must be P_{n-2}, otherwise w_{n-1} will appear in u which contradicts the maximality w in u;
(ii) w_{n-2} appears in u at most a_n times, otherwise w_n appear in u;
(iii) w_{n-2} appears in u at least a_n times since $u > |A_n|$;
so from (i) to (iii), w_{n-2} appears in u exactly a_n times.
The other three cases can be proved by the same argument. □

Now we discuss the factor P_n. By an analogous analysis to w_n with known facts $w_{n-1} \prec P_n$, $w_{n-1} \not\prec w_n$ and $P_n \not\prec w_n$, we see that P_n is located between two adjacent w_n as a separating factor, or inside some w_{n+1}, or in the "middle" of the following seven words:

1. $P_n w_n$; 2. $w_n P_n$; 3. $P_n w_n P_n$; 4. $P_n w_n w_{n+1}$; 5. $w_n w_{n+1}$; 6. $w_{n+1} w_n$; 7. $w_{n+1} w_n P_n$.

The following lemma shows that none of above seven cases could happen.

Lemma 5. *Suppose* z *is one of the above seven words. Let* $z = u_1 u_2 u_3$, $0 < |u_1| < |B_1|$, $0 < |u_3| < |B_2|$, *where* B_1, B_2 *denote the left and right factor of* z, *respectively*, *with* $|B_1| + |B_2| \leqslant |z|$ (*for example*, *in the first case*, $B_1 = P_n$, $B_2 = w_n$), *then* $u_2 \neq P_n$.

Proof. We only prove the first case, the other cases can be proved in the same way.

Assume $z=P_nw_n$ and $u_2=P_n$. We know by Lemma 2(3) that w_{n-1} is a right factor of $u_2=P_n$. This means w_{n-1} must be in the "middle" of P_nw_n or contained in w_n. The first case contradicts Lemma 4(2) and the second case contradicts Proposition 6(2). □

Remark 3. Above discussions and Lemma 5 shows that P_n appears in F_θ either as a factor either between two adjacent w_n as a separating word, or contained in some w_{n+1}. In the latter case, the P_n is both the suffix and the prefix of w_{n+1} from Lemma 2(2).

6. Combinatorial properties of the factors of the Sturmian sequence

In this section, we discuss the combinatorial properties of the factors of the Sturmian sequence, such as the power of the factors, overlap property of the factors, and the structure of the palindrome factor. As we will see, the positive separation property of the singular words will play an important role in these studies.

6.1. Power of the factors of Sturmian sequence

Theorem 8. *Let F_θ be a Sturmian sequence with $\theta=[0;a_0,a_1,\ldots,a_n,\ldots]$. We have the following facts*:

(1) *for any $n\geqslant 1$, $w_n^2 \not\prec F_\theta$;*

(2) *for any $n\geqslant 1$ and $0\leqslant k<|A_n|$, $(C_k(A_n))^{a_{n+1}+1}\prec F_\theta$;*

(3) *for any $n\geqslant 1$ and $0\leqslant k\leqslant |A_{n-1}|-2$, $(C_k(A_n))^{a_{n+1}+2}\prec F_\theta$;*

(4) *for any $n\geqslant 1$ and $|A_{n-1}|-2<k<|A_n|$, $(C_k(A_n))^{a_{n+1}+2}\not\prec F_\theta$;*

(5) *for any $n\geqslant 1$ and $0\leqslant k<|A_n|$, $(C_k(A_n))^{a_{n+1}+3}\not\prec F_\theta$;*

(6) *let $u\prec F_\theta$ with $|A_n|<|u|<|A_{n+1}|$ for some $n\geqslant 0$. If $w_n\prec u$, then $u^2\not\prec F_\theta$;*

(7) *let $u\prec F_\theta$ with $|A_n|<|u|<|A_{n+1}|$ for some $n\geqslant 0$. If $w_n\not\prec u$ and $u\neq C_k(A_n)^t$, $0\leqslant k\leqslant |A_n|-1$, $2\leqslant t\leqslant a_{n+1}$, then $u^2\prec F_\theta$ if and only if $u=u_1(w_{n-1}P_{n-1})^kw_{n-1}u_2$ with $u_2u_1=w_n$ and $0\leqslant k\leqslant a_{n+1}-2$, $1\leqslant |u_1|,|u_2|\leqslant |A_n|-1$;*

(8) *let $u\prec F_\theta$ with $|A_n|<|u|<|A_{n+1}|$ for some $n\geqslant 0$. If $w_n\not\prec u$ and $u\neq C_k(A_n)^t$, $0\leqslant k\leqslant |A_n|-1$, $2\leqslant t\leqslant a_{n+1}$, then $u^3\not\prec F_\theta$;*

(9) *let $u\prec F_\theta$ with $|A_n|<|u|<|A_{n+1}|$ for some $n\geqslant 0$, then $u^{a_{n+1}+3}\not\prec F_\theta$.*

Proof. (1) This is due to the positive separation property of singular words (in fact, we have shown this in Proposition 6(7)).

(2) Since $A_{n+1}A_n\lhd A_{n+2}$, and $A_nA_{n+2}\prec F_\theta$, $A_nA_{n+1}A_n\prec F_\theta$ by Proposition 4(3). So Theorem 3 and Proposition 3 imply that $A_n^{a_{n+1}+1}A_nA_{n-1}\alpha^{-1}\beta^{-1}\alpha\beta\prec F_\theta$. Thus, $A_n^{a_{n+1}+2}\beta^{-1}\prec F_\theta$. Hence, for any $0\leqslant k\leqslant |A_n|-1$, we have $(C_k(A_n))^{a_{n+1}+1}\prec A_n^{a_{n+1}+2}\beta^{-1}\prec F_\theta$.

(3) As in (2), $A_n^{a_{n+1}+2}A_{n-1}\alpha^{-1}\beta^{-1}\prec F_\theta$. Since $A_{n-1}\lhd A_n$, we have for $0\leqslant k\leqslant |A_{n-1}|-2$, $C_k(A_n)^{a_{n+1}+2}\prec F_\theta$.

(4) Let $|A_{n-1}|-1\leqslant k\leqslant |A_n|-1$. By Proposition 5(1), $C_k(A_n)=v_1w_{n-1}v_2$ with $v_2v_1=P_{n-1}$. So if $C_k(A_n)^{a_{n+1}+2}\prec F_\theta$, then

$$(w_{n-1}v_2v_1)^{a_{n+1}+1}w_{n-1}=(w_{n-1}P_{n-1})^{a_{n+1}+1}w_{n-1}\prec C_k(A_n)^{a_{n+1}+2}\prec F_\theta.$$

But by Lemma 2(4), we have $w_{n+1}=(w_{n-1}P_{n-1})^{a_{n+1}}w_{n-1}$. So w_{n+1} appears twice in the word $(w_{n-1}P_{n-1})^{a_{n+1}+1}w_{n-1}$, which contradicts the positive separation property of w_{n+1}.

(5) The conclusion can be obtained by the same discussion as in (4).

(6) Assume $w_n \prec u$ with $|A_n|<|u|<|A_{n+1}|$ and $u^2 \prec F_\theta$. Let $u=v_1w_nv_2$, then $u^2=v_1w_nv_2v_1w_nv_2 \prec F_\theta$. From Theorem 7, the length of the word between two w_n is at least $|P_n|$, i.e. $|v_2v_1|\geqslant|P_n|$, so $|u|\geqslant|w_n|+|P_n|=|A_{n+1}|$, which contradicts our hypothesis $|u|<|A_{n+1}|$.

(7) Suppose that w is the singular word of the maximum order appearing in u. Corollary 4 implies that w will be one of the following four words: w_{n+1}, w_n, w_{n-1} and w_{n-2}. By the hypotheses, $w\neq w_{n+1}$ and $u\neq w_n$, so w must be w_{n-1} or w_{n-2}.

We prove first the part "if".

Assume first $w=w_{n-2}$, and by Corollary 4, w_{n-2} appears in u exactly a_n times. Since w_{n-1} does not occur in u, all separating words of w_{n-2} in u are P_{n-2}. So we have

$$u=v_1(w_{n-2}P_{n-2})^{a_n-1}w_{n-2}v_2=v_1P_{n-1}v_2.$$

We have the following two facts:

(i) $|v_1|,|v_2|<|A_{n-1}|$. Otherwise, $|v_1|\geqslant|A_{n-1}|$, and $w_{n-1}\triangleright v_1$ or $w_{n-2}P_{n-2}\triangleright v_1$. The first contradicts the fact w_{n-1} does not occur in u and the second contradicts the fact that w_{n-2} only appears a_n times.

(ii) $|v_1v_2|>|A_{n-1}|$ since $|u|=|P_{n-1}|+|v_2v_1|>|A_n|=|P_{n-1}|+|A_{n-1}|$.

By hypothesis

$$u^2=v_1(w_{n-2}P_{n-2})^{a_n-1}w_{n-2}v_2v_1(w_{n-2}P_{n-2})^{a_n-1}w_{n-2}v_2 \prec F_\theta,$$

the word v_2v_1 is the word between two w_{n-2}. But $|A_{n-1}|<|v_2v_1|<2|A_{n-1}|$ by (i) and (ii), we must have that $v_2v_1=P_{n-2}w_{n-2}P_{n-2}$, thus $(w_{n-2}P_{n-2})^{a_n+1}w_{n-2}\prec u^2\prec F_\theta$, which contradicts the positively separation property of $w_n(=(w_{n-2}P_{n-2})^{a_n}w_{n-2})$. This proves $w\neq w_{n-2}$.

Now we assume $w=w_{n-1}$. Since $w_n\not\prec u$, all separating words in u with respect to w_{n-1} are P_{n-1}, and we can write $u=v_1(w_{n-1}P_{n-1})^{s_1}w_{n-1}v_2$ for some $s_1\geqslant 0$ with $w_{n-1}\not\prec v_1, w_{n-1}\not\prec v_2$. By the hypotheses,

$$u^2=v_1(w_{n-1}P_{n-1})^{s_1}w_{n-1}v_2v_1(w_{n-1}P_{n-1})^{s_1}w_{n-1}v_2 \prec F_\theta,$$

thus v_2v_1 is situated between two w_{n-1}.

If v_2v_1 does not contain w_n, it must contain only P_{n-1} and w_{n-1}, which gives $v_2v_1=(P_{n-1}w_{n-1})^{s_2}P_{n-1}$ for some $s_2\geqslant 0$. So $C_{|v_1|}u=(w_{n-1}P_{n-1})^{s_1+s_2+1}=\beta^{-1}A_n^{s_1+s_2+1}\beta$, where the second equality is due to Lemma 2(1). Hence there exists $0\leqslant k\leqslant|A_{n-1}|-1$ such that $u=(C_k(A_n))^t$, which contradicts the hypotheses.

So we must have v_2v_1 contains w_n. Consequently, it contains only one such word, otherwise $u\geqslant|A_{n+1}|$. Hence by similar discussion as above, we have $v_2v_1=(P_{n-1}w_{n-1})^{s_3}w_n(w_{n-1}P_{n-1})^{s_4}$.

Since $w_n\not\prec u$, then $u=u_1(w_{n-1}P_{n-1})^{s_1+s_2+s_4}w_{n-1}u_2$ and $u_2u_1=w$, $0\leqslant s_1+s_2+s_4\leqslant a_{n+1}-2$. This finishes the proof of the "if" part.

Now we prove in the following the "only if" part.

Assume $u=u_1(w_{n-1}P_{n-1})^k w_{n-1}u_2$ with $u_2u_1=w_n$ and $0\leqslant k\leqslant a_{n+1}-2$, $1\leqslant|u_1|,|u_2|\leqslant |A_n|-1$. Then

$$u^2=u_1(w_{n-1}P_{n-1})^k w_{n-1}w_n(w_{n-1}P_{n-1})^k w_{n-1}u_2\prec P_nw_nP_n\prec F_\theta.$$

(8) Suppose that $u^3\prec F_\theta$, then $u^2\prec F_\theta$. By conclusion (7), $u=u_1(w_{n-1}P_{n-1})^k w_{n-1}u_2$ with $u_2u_1=w_n$ and $0\leqslant k\leqslant a_{n+1}-2$, therefore $w_n(w_{n-1}P_{n-1})^k w_{n-1}w_n\prec u^3\prec F_\theta$. The positive separating property of w_n shows $|(w_{n-1}P_{n-1})^k w_{n-1}|\geqslant|P_n|$. Hence $k\geqslant a_{n+1}-1$ by Lemma 2(3), and this is a contradiction.

(9) This follows from conclusions (6)–(8). □

Remark 4. Theorem 8(2) shows that, although each conjugation of the standard word A_n appears in F_α infinitely many times, the conjugates are not necessary to be positively separated. This is an essential difference between singular words and standard words.

Now we study the highest order of the repetition in the Sturmain sequence.

Let $r>1$ be a rational, we say the sequence $F\in S^\omega$ contains a *repetition of order* r, if there exist two factors $z,x\prec F$ such that

$$z\lhd x^{[r]+1}\quad\text{and}\quad\frac{|z|}{|x|}=r.$$

In this case we write $z=x^r$ (note that x^r is well defined if and only if $k|x|$ is an integer). Above definition is equivalent to that $z=(uv)^{[r]}u$ with $|u|/(|u|+|v|)=\{r\}$.

Define the free index $\mathbf{FI}(F)$ of the sequence F as follows:

$$\mathbf{FI}(F)=\sup\{r\in\mathbb{Q}:\ F \text{ contains a repetition of order } r\}.$$

The following theorem is proved by Damanik and Lenz [6] (for the related results, see also Berstel [4], Mignosi and Pirillo [15] and Vandeth [22]). Here we give a simple proof of this result using singular words.

Theorem 9. *Suppose that* $\theta=[0;a_1,a_2,\ldots]$ *is the continued fraction expansion of* θ. *Then*

$$\mathbf{FI}(F_\theta)=2+\sup_{n\geqslant 0}\left\{a_{n+1}+\frac{|A_{n-1}|-2}{|A_n|}\right\}.$$

Proof. For any factor $u\prec F_\theta$, define the index of u by $ind(u)=\max\{r\in\mathbb{Q}:\ u^r\prec F_\theta\}$, which yields immediately $\mathbf{FI}(F_\theta)=\sup_{u\prec F_\theta} ind(u)$.

By Proposition 4(4), $ind(b)=1$; since $|A_{-1}|=|A_0|=1$, $ind(a)=1+a_1=2+(a_1+(|A_{-1}|-2)/|A_0|)$. Suppose $|A_n|<|u|<|A_{n+1}|$ for some $n\geqslant 0$. If $u=C_k(A_n)^t$ for some t, then $ind(u)=(1/t)ind(C_k(A_n))$; if $u\neq C_k(A_n)^t$ $(0\leqslant k\leqslant|A_n|-1, 2\leqslant t\leqslant a_{n+1})$, then $ind(u)<3$ from Theorem 8(7) and (8). So we only need to consider the word u of length $|A_n|$ for some n. If $u=w_n$ is a singular word, $ind(u)<2$ since $w_n^2\not\prec F_\theta$. Now fix $n\geqslant 1$ and we are going to determine $\max\{ind(u):\ u=C_k(A_n)$ for some $k\}$. In fact, we only need to find the maximum length of the word $x\prec F_\theta$ which is a factor of the infinite sequence $A_n^\omega := A_nA_nA_n\cdots$. Since the singular word w_n is not a conjugate of A_n, w_n is not a factor of x. By Theorem 7, w_n is positively separated by the separating words P_n and w_{n+1}. So the possible maximum length of x is $|A_n|+|A_{n+1}|+|A_n|-2$. In this case $x=\alpha^{-1}w_nw_{n+1}w_n\alpha^{-1}$ with α the first letter of w_n. So by Theorem 3 and Proposition 3, we get

$$\begin{aligned} x &= \alpha^{-1}w_nw_{n+1}w_n\alpha^{-1} \\ &= A_nA_{n+1}A_n\beta^{-1}\alpha^{-1} = A_n^{a_{n+1}+1}A_{n-1}A_n\beta^{-1}\alpha^{-1} \\ &= A_n^{a_{n+1}+2}A_{n-1}\alpha^{-1}\beta^{-1} = A_n^{2+a_{n+1}+\frac{|A_{n-1}|-2}{|A_n|}}, \end{aligned}$$

which yields the conclusion of the theorem. □

Remark 5. (1) It is easy to see that the above theorem is also true for $\theta>1$. (In this case we will take the continued fraction expression of θ as $[a_1;a_2,\ldots,a_n,\ldots]$.)

(2) Denote c_μ the Sturmian sequence $\{\chi_{[1-\mu,1)}(n\mu \bmod 1)\}_{n\geqslant 1}$. Then $F_\theta=\pi(c_\mu)$ if and only if $\mu=\theta/(1+\theta)$, where π is projection: $\pi(0)=a$, $\pi(1)=b$.

If $\theta=[0;a_1,a_2,\ldots,a_n,\ldots]<1$, then $\theta/(1+\theta)=1/(1+1/\theta)=[0;a_1+1,a_2,\ldots,a_n,\ldots]$; if $\theta=[a_1;a_2,\ldots,a_n,\ldots]>1$, then $\theta/(1+\theta)=[0;1,a_1,a_2,\ldots,a_n,\ldots]$. Thus, a simple computation shows the equivalence between the result in [6] and Theorem 9.

From Theorem 9, we get immediately

Corollary 5. *Suppose that F_θ is a Sturmian sequence. Then*

$$\sup\{p;\ \exists w\prec F_\theta \text{ such that } w^p\prec F_\theta\} = \max\left\{1+a_1, 2+\sup_{n\geqslant 2}\{a_n\}\right\}.$$

In particular, if $\sup_{n\geqslant 1}\{a_n\}=\infty$, then for any $k\in\mathbb{N}$, there exists a factor $w\prec F_\theta$ such that $w^k\prec F_\theta$.

Example 1 (Mignosi and Pirillo [15], Wen Zhi-Xiong and Wen Zhi-Ying [21]). *Let $\theta=\frac{\sqrt{5}-1}{2}=[0;1,1,1,\ldots]$ be the golden number, then $\mathbf{FI}(F_\theta)=\frac{5+\sqrt{5}}{2}$, and for any factor $u\prec F_\theta$, $u^4\not\prec F_\theta$.*

6.2. *Overlap property of the factors*

Suppose $u\prec F\in S^\omega$. If there exist words x,y and z such that $u=xy=yz$ and $u^*(y):=uz=xyz\prec F$, then we say that the word u *have overlap* with the *overlap factor* y (or *overlap length* $|y|$), and the word $u^*(y)$ is called the *overlap of* u with the overlap factor y. We denote by $\mathbb{O}(F):=\mathbb{O}$ the set of factors of F having overlap. The structure of $\mathbb{O}$ for the Fibonacci sequence and the Morse sequence have been studied, respectively, in [21] and [1].

From Theorem 7, we have

Proposition 7. *For any $n \geqslant 1$, w_n has no overlap.*

Lemma 6. *Let $u \prec F_\theta$ with $|A_n| < |u| \leqslant |A_{n+1}|$ for some $n \geqslant 0$. Then $w_n \not\prec u$ if and only if $u \prec \alpha^{-1} w_n w_{n+1} w_n \alpha^{-1}$ with $\alpha \lhd w_n$.*

Proof. By Theorem 7, if $w_n \not\prec u$, u is a factor of either $\alpha^{-1} w_n P_n w_n \alpha^{-1}$ or $\alpha^{-1} w_n w_{n+1} w_n \alpha^{-1}$. By Lemma 2(1), $\alpha^{-1} w_n P_n = \beta^{-1} w_{n+1}$, which means

$$\alpha^{-1} w_n P_n w_n \alpha^{-1} = \beta^{-1} w_{n+1} w_n \alpha^{-1} \prec \alpha^{-1} w_n w_{n+1} w_n \alpha^{-1}.$$

On the other hand, if $u \prec \alpha^{-1} w_n w_{n+1} w_n \alpha^{-1}$, then $w_n \not\prec u$ from the positive separation property of w_n, and the result follows. □

Lemma 7. *Let $w = \alpha^{-1} w_n w_{n+1} w_n \alpha^{-1}$ with $\alpha \lhd w_n$. If $u \prec w$ with $|A_n| < |u| \leqslant |A_{n+1}|$ and $u \neq w_{n+1}$, then u has overlap.*

Proof. From Lemma 2(1) and (4), we have

$$w = \alpha^{-1} w_n w_{n+1} w_n \alpha^{-1} = \beta^{-1} (w_{n-1} P_{n-1})^{a_{n+1}+2} w_{n-1} \beta^{-1}.$$

By Lemma 6, $w_n \not\prec u$. From the hypotheses of the lemma, we see that if u contains the word w_{n-1}, then it contains the words w_{n-1} at most a_{n+1} times.

(i) Suppose u contains words w_{n-1} for i times ($1 \leqslant i \leqslant a_{n+1}$) and $u = s_2 (w_{n-1} P_{n-1})^{i-1} w_{n-1} t_1$. Then $s_2 \rhd \beta^{-1} w_{n-1} P_{n-1}$ and $t_1 \lhd P_{n-1} w_{n-1} \beta^{-1}$, where s_2, t_1 may be ε. Let $w_{n-1} P_{n-1} = s_1 s_2$, $P_{n-1} w_{n-1} = t_1 t_2$, then

$$\begin{aligned} u &= s_2 (w_{n-1} P_{n-1})^{i-1} w_{n-1} t_1 \\ &= (s_2 s_1)(s_2 (w_{n-1} P_{n-1})^{i-2} w_{n-1} t_1) \\ &= (s_2 (w_{n-1} P_{n-1})^{i-2} w_{n-1} t_1)(t_2 t_1), \end{aligned}$$

so $u = xy = yz$ with $x = s_2 s_1$, $y = s_2 (w_{n-1} P_{n-1})^{i-2} w_{n-1} t_1$ and $z = t_2 t_1$. Hence to prove that u has overlap, it suffices to prove that $u t_2 t_1 \prec F_\infty$. In fact,

$$\begin{aligned} u t_2 t_1 &= s_2 (w_{n-1} P_{n-1})^{i-1} w_{n-1} t_1 t_2 t_1 = s_2 (w_{n-1} P_{n-1})^{i} w_{n-1} t_1 \\ &\prec \beta^{-1} w_{n-1} P_{n-1} ((w_{n-1} P_{n-1})^{i} w_{n-1}) P_{n-1} w_{n-1} \beta^{-1} \prec w \prec F_\theta. \end{aligned}$$

(ii) Suppose u contains no copies of w_{n-1}. Because $|u| > |w_n|$, $u = s P_{n-1} t \prec w_{n-1} P_{n-1} w_{n-1}$, $s \rhd \beta^{-1} w_{n-1}$, $t \lhd w_{n-1} \beta^{-1}$, hence $|s| + |t| > |w_{n-1}|$. This means there exists a word v_0, such that $t = t' v_0$, $s = v_0 s'$, $w_{n-1} = t' v_0 s'$, and $u = v_0 s' P_{n-1} t' v_0$ with $h|u| > |v_0 s' P_{n-1} t'|$. Since

$$\begin{aligned} u s' P_{n-1} t' v_0 &= v_0 s' P_{n-1} t' v_0 s' P_{n-1} t' v_0 \\ &= v_0 s' P_{n-1} t' u = s P_{n-1} w_{n-1} P_{n-1} t \end{aligned}$$

$$\prec \beta^{-1}w_{n-1}P_{n-1}w_{n-1}P_{n-1}w_{n-1}\beta^{-1}$$
$$\prec w,$$

which shows that u has overlap with overlap factor v_0. □

Theorem 10. *Let $u \prec F_\theta$ with $|A_n| < |u| \leqslant |A_{n+1}|$ and $u \neq w_{n+1}$. Then*

$$u \notin \mathbb{O} \Leftrightarrow w_n \prec u.$$

Proof. Suppose that $w_n \prec u$ and u has overlap. Then w_n will appear twice in the overlap of u. By Theorem 7, any word between two adjacent singular words w_n must be either P_n or w_{n+1}, and it follows that $|u| > |w_{n+1}|$. This contradiction proves that u has no overlap and we prove the implication $w_n \prec u \Rightarrow u \notin \mathbb{O}$. The opposite implication $u \notin \mathbb{O} \Rightarrow w_n \prec u$ follows directly from Lemmas 6 and 7. □

Remark 6. If a word $w \prec F_\theta$ has overlap, then the overlap factor does not need to be unique. For example, let $a_{n+1} = 2$, and let

$$w := \beta^{-1}(w_{n-1}P_{n-1})^4 w_{n-1}\beta^{-1} = \alpha^{-1}w_n w_{n+1} w_n \alpha^{-1} \prec F_\alpha.$$

Then the word $u = \beta^{-1}(w_{n-1}P_{n-1})^2 w_{n-1}\beta^{-1}$ has two overlaps w and $w' := u\beta P_{n-1}w_{n-1}\beta^{-1}$, and the corresponding overlap factors are $\beta P_{n-1}w_{n-1}\beta^{-1}$ and $\beta(P_{n-1}w_{n-1})^2\beta^{-1}$, respectively.

Corollary 6. *Let F_θ be a Sturmian sequence. We have the following*:
(1) *for any $n \geqslant 1$ and $0 \leqslant k \leqslant |A_{n-1}| - 2$, $C_k(A_n)^{a_{n+1}+1} \in \mathbb{O}$;*
(2) *for any $n \geqslant 1$ and $|A_{n-1}| - 1 \leqslant k \leqslant |A_n| - 1$, $C_k(A_n)^{a_{n+1}+1} \notin \mathbb{O}$;*
(3) *for any $n \geqslant 1$, $0 \leqslant k \leqslant |A_{n-1}| - 2$, and $a_{n+2} \geqslant 2$, $C_k(A_n)^{a_{n+1}+2} \notin \mathbb{O}$;*
(4) *for any $n \geqslant 1$, $0 \leqslant k \leqslant |A_{n-1}| - 2$, and $a_{n+2} = 1$, $C_k(A_n)^{a_{n+1}+2} \in \mathbb{O}$.*

Proof. (1) First, we have $|A_{n+1}| < |A_n^{a_{n+1}+1}| < |A_{n+2}|$. If $0 \leqslant k \leqslant |A_{n-1}| - 2$, then by Proposition 5(1) we can write that $C_k(A_n) = uP_{n-1}v$ with $vu = w_{n-1}$ and $|u|, |v| < |w_{n-1}|$. Now

$$C_k(A_n)^{a_{n+1}+1} = (uP_{n-1}v)^{a_{n+1}+1} = uP_{n-1}(w_{n-1}P_{n-1})^{a_{n+1}-1}w_{n-1}P_{n-1}v,$$

but we also have $w_{n+1} = (w_{n-1}P_{n-1})^{a_{n+1}}w_{n-1}$. The positive separation property of w_{n-1} shows $w_{n+1} \not\prec C_k(A_n)^{a_{n+1}+1}$ and the result follows from Theorem 10.

(2) If $|A_{n-1}| - 1 \leqslant k \leqslant |A_n| - 1$, then $C_k(A_n) = uw_{n-1}v$ and $vu = P_{n-1}$, which implies $C_k(A_n)^{a_{n+1}+1} = u(w_{n-1}P_{n-1})^{a_{n+1}}w_{n-1}v = uw_{n+1}v$.

(3) In this case, $|A_{n+1}| < |A_n^{a_{n+1}+2}| < |A_{n+2}|$ and we can show that $w_{n+1} \prec C_k(A_n)^{a_{n+1}+2}$ in the same way as above.

(4) We have in this case $|A_{n+2}| < |A_n^{a_{n+1}+2}| < |A_{n+3}|$. Since w_{n+2} is not a factor of $C_k(A_n)^{a_{n+1}+2}$, we have $C_k(A_n)^{a_{n+1}+2} \in \mathbb{O}$ by Theorem 10. □

6.3. The Palindrome factors

In this subsection, we study the structures of the palindrome factors of Sturmian sequences. We recall the following basic facts: both w_n and P_n are palindromes; w_n is positively separated by the separating factors w_{n+1} and P_n; the words $w_{n-1}P_{n-1}$ and w_n differ merely by the first letter; $P_n \lhd w_n$ and $P_n \rhd w_n$.

Lemma 8. *Let* $u \prec F_\theta$ *with* $|A_n| < |u| \leqslant |A_{n+1}|$ *for some* $n \geqslant 0$. *If* $w_n \not\prec u$ *and* $P_n \not\prec u$, *then* $u \prec w_{n+1}$.

Proof. Since $w_n \not\prec u$, $u \prec \alpha^{-1} w_n w_{n+1} w_n \alpha^{-1}$ from Lemma 6. Because $|u| \leqslant |A_{n+1}|$, we have either $u \prec \alpha^{-1} w_n w_{n+1}$ or $u \prec w_{n+1} w_n \alpha^{-1}$. First suppose $u \prec \alpha^{-1} w_n w_{n+1}$. Then Lemma 2(1), we have

$$u \prec \alpha^{-1} w_n w_{n+1} = \beta^{-1}(w_{n-1}P_{n-1})^{a_{n+1}+1} w_{n-1} \prec (w_{n-1}P_{n-1})^{a_{n+1}+1} w_{n-1}.$$

Since $P_n = (w_{n-1}P_{n-1})^{a_{n+1}-1} w_{n-1} \not\prec u$, we have $u \prec (w_{n-1}P_{n-1})^{a_{n+1}} w_{n-1} = w_{n+1}$.

The case $u \prec w_{n+1} w_n \alpha^{-1}$ can be proved similarly. □

Theorem 11. *Let* $u \in \mathbb{P}$ *with* $|A_n| < |u| \leqslant |A_{n+1}|$ *for some* $n \geqslant 0$, *then* $u \prec F_\theta$ *if and only if u is one of the following forms*:

(1) $u = x w_n \bar{x}$ *with* $x \rhd P_n$ *and* $|x| \leqslant \frac{1}{2}|P_n|$;

(2) $u = x P_n \bar{x}$ *with* $x \rhd w_n$ *and* $|x| \leqslant \frac{1}{2}|w_n|$;

(3) $u = x(w_{n-1}P_{n-1})^k w_{n-1}\bar{x}$, *where* $x \rhd P_{n-1}$, $0 \leqslant k \leqslant a_{n+1}-1$. *Moreover if* $k = 0$ *then* $|x| > \frac{1}{2}|P_{n-1}|$;

(4) $u = x(P_{n-1}w_{n-1})^k P_{n-1}\bar{x}$, *where* $x \rhd w_{n-1}$, $0 \leqslant k \leqslant a_{n+1}-1$. *Moreover if* $k = 0$ *then* $|x| > \frac{1}{2}|w_{n-1}|$.

Proof. The part "if" is ready to check by noting that $P_n w_n P_n \prec F_\theta$ and $w_n P_n w_n \prec F_\theta$ for any $n \in \mathbb{N}$.

Now suppose $u \in \mathbb{P}$ is a factor of F_α with $|A_n| < |u| \leqslant |A_{n+1}|$ for some $n \geqslant 0$.

(i) Suppose $w_n \prec u$, and we write $u = x w_n y$. Then x is a right factor of either P_n or w_{n+1} by Theorem 7(3) Since $|x| \leqslant |u| - |w_n| < |A_{n+1}| - |w_n| = |P_n|$ and P_n is a right factor of w_{n+1}, we get $x \rhd P_n$. In the same way $y \lhd P_n$. Since $u \in \mathbb{P}$, $u = \bar{u} = \overline{x w_n y} = \bar{y} w_n \bar{x}$. The positive separation property of w_n shows that w_n has only one occurrence in u. So we have $\bar{y} = x$ which yields $|x| = |y| < \frac{1}{2}|P_n|$. Conclusion (1) is proved.

(ii) Suppose $P_n \prec u$, $u \neq w_{n+1}$ and write $u = x P_n y$. We conclude $x \rhd w_n$. In fact, by noting that $w_{n-1} \lhd P_n$, and $|x| \leqslant |u| - |P_n| \leqslant |A_{n+1}| - |P_n| = |w_n| = |w_{n-1}P_{n-1}|$, we have that either $x \rhd w_n$ or $x \rhd w_{n-1}P_{n-1}$ due to the positive separation property of w_{n-1}. Since $u \neq w_{n+1}$, $x \neq w_{n-1}P_{n-1}$, we have $x \rhd \beta^{-1} w_{n-1}P_{n-1} = \alpha^{-1} w_n$ by Lemma 2(1), and further $x \rhd w_n$. In the same way $y \lhd w_n$. But $w_{n-1} \not\prec w_n$, so $w_{n-1} \not\prec x$ and $w_{n-1} \not\prec y$. Since $u \in \mathbb{P}$, $u = \bar{u} = \overline{x P_n y} = \bar{y} P_n \bar{x}$. Above analysis shows $w_{n-1} \not\prec \bar{x}$ and $w_{n-1} \not\prec \bar{y}$. Because w_{n-1} is both left factor and right factor of P_n, we have $\bar{y} = x$ by the positive separation property of w_{n-1}. We prove thus assertion (2) of the theorem.

Now if neither (i) nor (ii) holds, then $u \prec w_{n+1}$ by Lemma 8. By using the fact $w_{n+1} = (w_{n-1}P_{n-1})^{a_{n+1}} w_{n-1}$, and by an almost same discussion as above, we get the assertions either (3) or (4), which finishes the proof of the theorem. □

From Theorems 10 and 11, we get

Corollary 7. *Let* $u \prec F_\theta$ *with* $|A_n| < |u| \leqslant |A_{n+1}|$ *for some* $n \geqslant 0$, *then u is a palindrome without overlap if and on if* $u = xw_n\bar{x}$ *with* $x \triangleright P_n$ *and* $|x| \leqslant \frac{1}{2}|P_n|$.

Acknowledgements

The authors thank Prof. J.-P. Allouche who showed us Refs. [6,22], and Dr. Bo Tan for the useful discussions and a simplified proof of Theorem 9. The second author thanks The Morning-Side Center of Mathematics for their partial support.

References

[1] J.-P. Allouche, J. Peyriere, Wen Zhi-Xiong, Wen Zhi-Ying, Some properties of the Hankel determinant associated with the Thue-Morse sequences, Ann Inst. Fourier 48 (1) (1998) 1–27.
[2] J. Berstel, Recent results in Sturmian words, in: J. Dassow, A. Salomaa (Eds.), Developments in Language Theory, World Scientific, Singapore, 1966, pp. 13–24.
[3] J. Berstel, Mot de Fibonacci, Séminaire d'informatique théorique, L.I.T.P., Paris, Année 1980/1981, pp. 57–78.
[4] J. Berstel, On the Ondex of Sturmian Words, Jewels are Forever, Springer, Berlin, 1999, pp. 287–294.
[5] Tom C. Brown, A characterization of the quadratic irrationals, Canad. Math. Bull. 34 (1991) 36–41.
[6] D. Damanik, D. Lenz, The index of Sturmian sequences, preprint.
[7] S. Ferenzy, Symbolic Dynamics, CIMPA Summer School Lectures, 1996.
[8] A.S. Fraenkel, M. Mushkin, U. Tassa, Determination of $\lfloor n\theta \rfloor$ by its sequence of differences, Canad. Math. Bull. 21 (1978) 441–446.
[9] G.A. Hedlund, M. Morse, Sturmian sequences, Amer. J. Math. 61 (1940) 1–42.
[10] S. Ito, S. Yasutomi, On continued fractions, substitutions and characteristic sequences, Japan. J. Math. 16 (1990) 287–306.
[11] A.Ya. Khintchine, Continued Fractions, P. Noordhoff, Groningen, 1963, pp. 31–33.
[12] M. Lothaire, Combinatorics on words, in: Encyclopedia of Mathematics and its Applications, Vol. 17, Addison-Wesley, Reading, MA, 1983.
[13] M. Lothaire, Algebraic combinatorics on words, Cambridge Univ. Press, Cambridge, 2002.
[14] F. Mignosi, On the number of factors of Sturmian words, Theoret. Comput. Sci. 82 (1991) 71–84.
[15] F. Mignosi, G. Pirillo, Repetitions of Fibonacci infinite word, RAIRO Inform. Theor. Appl. 26 (1992) 199–204.
[16] P. Séébold, Fibonacci morphisms and Sturmian words, Theoret. Comput. Sci. 88 (1991) 367–384.
[17] C. Series, The Geometry of Markoff numbers, Math. Intell. 7 (1985) 20–29.
[18] K.B. Stolarsky, Beatty sequences, continued fractions, and certain shift operators, Canad. Math. Bull. 19 (1976) 473–482.
[19] B.A. Venkov, Elementary Number Theory, Wolter-Noordhoff, Groningen, 1970.
[20] Wen Zhi-Xiong, Wen Zhi-Ying, Some studies of factors of infinite words generated by invertible substitution, in: A. Barlotti, M. Delest, R. Pinzani (Eds.), Proc. Fifth Conf. Formal Power Series and Algebraic Combinatorics, 1993, pp. 455–466.
[21] Wen Zhi-Xiong, Wen Zhi-Ying, Some properties of the singular words of the Fibonacci word, European J. Combin. 15 (1994) 587–598.
[22] D. Vandeth, Sturmian words and words with a critical exponent, Theoret. Comput. Sci. 242 (2000) 283–300.