Lyndon words and singular factors of sturmian words

Guy Melançon

LaBRI, UMR 5800 CNRS, Université Bordeaux I, France

Abstract

Two different factorizations of the Fibonacci infinite word were given independently in Wen and Wen (1994) and Melançon (1996). In a certain sense, these factorizations reveal a self-similarity property of the Fibonacci word. We first describe the intimate links between these two factorizations. We then propose a generalization to characteristic sturmian words. © 1999 Elsevier Science B.V. All rights reserved.

Résumé


Keywords: Lyndon words; Sturmian words; Singular words

1. Introduction

The numerous aspects under which the combinatorial properties of the Fibonacci word have been studied are amazing. A huge set of different notions in algebraic combinatorics on words may be illustrated by using this infinite word as an example. The Fibonacci word is a well-known example of a huge family of infinite words called sturmian words. These words have been studied from many different points of view, geometrical, combinatorial, algebraic, etc. (see Remark 3.2). They naturally appear in fields such as number theory, quasicrystals, computational complexity, to name only a few (see [1]). The combinatorial structure of an infinite word is often revealed by
the study of the set of its factors: that is, the finite words appearing within it. As far as sturmian words are concerned, this is well illustrated by the work of Berstel and de Luca [2].

Wen and Wen [10] have looked at a particular set of factors of the Fibonacci word, they call singular factors. They are the consecutive factors of the Fibonacci word of lengths $F_0, F_1, F_2, \ldots$, where $(F_n)_{n \geq 0}$ is the Fibonacci sequence given by $F_0 = F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$ ($n \geq 1$). Our work started from a remark by Jean Berstel linking these singular factors to the Lyndon words appearing in the Lyndon factorization of the Fibonacci word we gave in [6] (see also [8]). Our investigation not only confirmed the remark made by Berstel but also lead us to a full description of the link between the singular factors and the Lyndon factors of the Fibonacci word. Indeed, Proposition 2.10 gives a description of the Lyndon factors in terms of singular factors and, as a consequence, new proofs of [10, Theorems 1 and 2].

A striking fact common both to our work [6] and that in [10] is the discovery of a self-similarity property of the Fibonacci word. Let us briefly describe it here. The W and W factorization in [10, Theorem 1] and the Lyndon factorization in [6, Proposition 11] rely on the computation of a set of consecutive factors of the Fibonacci word. In both cases, these factorizations are self-similar ([10, Theorems 2, 2.11]); that is, the Fibonacci word over the original alphabet \{$a, b$\} (almost) coincides with itself computed over a two-word alphabet selected among the set of singular or Lyndon factors. As far as singular factors are concerned, this is explained by a particularity they have with respect to conjugation of words, and to the fact that they are non-overlapping (see [10, Lemma 2, Property 2]).

Many results concerning the Fibonacci word (Section 2) naturally generalize to characteristic sturmian words. The Lyndon factorization of characteristic sturmian words was given in [8]. Again, this factorization is self-similar: Theorem 3.4 shows how to compute a given characteristic sturmian word using a two-word alphabet selected among its Lyndon factors. Moreover, it is possible to define singular factors of a given characteristic sturmian word (Definition 4.1) and give combinatorial properties of these words. In particular, Lemma 4.2 shows that as in the Fibonacci case, general singular factors hold a special place with respect to conjugation. Proposition 4.4 links general singular factors to Lyndon factors of a given characteristic sturmian word. This, combined with the non-overlapping property of the singular factors (Lemma 4.7), leads to a formulation of the self-similarity property in terms of the singular factors, analog to [10, Theorem 2] (Corollary 4.6).

The paper is structured as follows. Section 1 briefly describes the results in [10] concerned with the present work. More precisely, we define the singular factors of the Fibonacci word, list some of their properties and state the two main theorems in [10]. We then recall the Lyndon factorization of the Fibonacci word given in [6] and study the link between singular factors and Lyndon factors of the Fibonacci word. Expressing Lyndon words in terms of the singular factors leads to Theorem 2.11 from which we are able to deduce the two main results in [10] (Section 2.4). We then show self-similarity of the Lyndon factorization in the general case of a characteristic sturmian
word (Theorem 3.4). Lemma 4.2 and Proposition 4.3 in the last section confirm the 
words introduced in Definition 4.1 as the proper generalization for singular factors 
of characteristic sturmian words. Corollaries 4.5 and 4.6 reinforce the links with the 
Lyndon factorization (Theorem 3.4) and propose a generalization of the self-similarity 
property [10, Theorem 2].

2. Singular factors and the Fibonacci word

This first section introduces basic notations and definitions, and describes two cen-
tral results in [10]. Throughout the paper, we only consider the two-letter alphabet 
$A = \{a, b\}$. We totally order $A$ by $a < b$ and extend this order to the set $A^*$ of all 
words lexicographically. The notations we use are those usual in theoretical computer 
science (see [5]). We shall make great use of the notation $w x^{-1}$, denoting the word 
obtained from $w$ by deleting the letter $x \in A$ at the end of $w$ (if possible). Let us start 
by recalling the definition of the Fibonacci word.

**Definition 2.1.** Let $f_0 = b$, $f_1 = a$ and define $f_{n+1} = f_n f_{n-1}$, for $n \geq 1$. The words $f_n$ 
($n \geq 0$) are usually called the finite Fibonacci words. Hence, e.g., $f_2 = ab$, $f_3 = aba$, 
$f_4 = abab$, and so on. The (right) infinite word 

$$f = \lim_{n \to \infty} f_n = abaababaababaababaabab\ldots$$

is called the (infinite) Fibonacci word.

For more details on the Fibonacci word, the reader is referred to Berstel’s recent 
survey on sturmian words [1].

**Remark 2.2.** (1) The length of $f_n$ is the $n$th Fibonacci number $F_n$ (where the Fibonacci 
sequence is defined by $F_0 = F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$, for $n \geq 1$).
(2) Moreover, for all $n \geq 2$, we have $(|f_n|_a, |f_n|_b) = (F_{n-1}, F_{n-2})$.
(3) For all $n \geq 1$, the word $f_{2n}$ ends with $ab$ and the word $f_{2n+1}$ ends with $ba$.

**Definition 2.3** (Wen and Wen [10, p. 589]). Let $n \geq 2$ and suppose $f_n$ ends with $\alpha \beta$ 
(where $\alpha, \beta \in A$ and $\alpha \neq \beta$). We define the word $w_n$ by $w_n = \alpha f_n \beta^{-1}$. The word $w_n$ is 
a factor of the Fibonacci word $f$ and is called the $n$th singular factor of $f$. We also 
define $w_0 = a$, $w_1 = b$; it is useful to set $w_{-1} = e$ (the empty word).

Hence, we have $w_2 = aa$, $w_3 = bab$, $w_4 = aabaa$, $w_5 = babaabab$, and so forth.

**Remark 2.4.** We collect here some remarks from [10].
(1) The length of $w_n$ is the $n$th Fibonacci number $F_n$. Let us verify that $w_n$ is indeed 
a factor of $f$. As is known, any conjugate of a factor of $f$ is also a factor 
of $f$. Hence, the word $a f_{2n} f_{2n+1} a^{-1}$ is a factor of $f$, since it is conjugated to
Thus, $w_{2n}$ and $w_{2n+1}$ are factors of $f$ since they are consecutive factors of $a f_{2n} f_{2n+1} a^{-1} = (a f_{2n} b^{-1}) (b f_{2n+1} a^{-1})$.

(2) Note that, for all $n \geq 0$, we have

$$
(w_n|_a, w_n|_b) = \begin{cases} (|f_n|_a + 1, |f_n|_b - 1) & \text{if } n \text{ is even,} \\
(|f_n|_a - 1, |f_n|_b + 1) & \text{if } n \text{ is odd.}
\end{cases}
$$

(3) As a consequence, $w_n$ is not conjugated to $f_n$. In [10], it is shown that $w_n$ is the only factor of $f$ of length $F_n$ that is not conjugated to $f_n$ (see Lemma 4.2).

(4) Observe that the words $w_{2n}$ (resp. $w_{2n+1}$) always end with $aa$ (resp. $b$).

We are now able to formulate [10]'s first fundamental result:

**Theorem 2.5** (Wen and Wen [10, Theorem 1]). We have

$$f = \prod_{j=0}^{\infty} w_j.$$

That is,

$$f = (a)(b)(aa)(bab)(aabaa)(babaabab) \cdots.$$

The set of factors of the Fibonacci word has received great attention from a large number of authors (again, see [1]; see also [2]). From this point of view, the next fundamental result of [10] is the following:

**Theorem 2.6** (Wen and Wen [10, Theorem 2]). Two occurrences of the singular factor $w_m$ ($m \geq 0$) never overlap. Denote these occurrences by $w_m^{(1)}, w_m^{(2)}, w_m^{(3)}$, and so forth (from left to right). Then we have

$$f = \left( \prod_{j=0}^{m-1} w_j \right) (w_m^{(1)} z_1 w_m^{(2)} z_2 w_m^{(3)} z_3 \cdots),$$

where $z_k \in \{w_{m+1}, w_{m-1}\}$, for all $k \geq 1$, and $z_1 z_2 z_3 \cdots$ is the Fibonacci word over the alphabet $\{w_{m+1}, w_{m-1}\}$.

For example, with $m = 2$, we have $w_2 = aa, w_3 = bab$ and $w_1 = b$. Thus,

$$f = (a \ b)(aa \ bab \ aa \ bab \ bab \ aa \ bab \ aa \ \cdots).$$

Note that the theorem is true also for $m = 0$ (recall that $w_{-1} = \varepsilon$). The word $z_1 z_2 z_3 \cdots$ is then equal to the Fibonacci word over the "alphabet" $\{b, \varepsilon\}$.

2.1. **Lyndon factorization**

This section introduces Lyndon words and links Theorems 2.5 and 2.6 to results in [8]. **Lyndon words** are words strictly smaller than their proper right factors. Although these may be defined over an arbitrary alphabet, we shall restrict ourselves to the
two letter alphabet \( A = \{a, b\} \). We denote by \( L \) the set of Lyndon words (over \( A \)). For instance, letters are Lyndon words. The words \( ab, abb, aab, aabb \), etc., are Lyndon words. More generally, given \( u, v \in L \), we have \( uv \in L \Leftrightarrow u < v \) [4, Proposition 1.3]. Hence, e.g., \( aababb \) is a Lyndon word. For more details concerning Lyndon words, the reader is referred to [5, Ch. 5].

Any Lyndon word \( \ell \) of length at least two is a product of two Lyndon words \( u, v \) with \( u < v \). For example, we have \( aababb = (a)(ababb) \), but also \( aababb = (aab)(abb) = (aahb)(b) \). The standard factorization of \( \ell \) is obtained by taking \( v \) of maximal length. We usually denote the standard factorization of \( \ell \) by \( \ell = \ell' \ell'' \). Hence, e.g., \((aababb)' = a\) and \((aababb)'' = ababb\). The Lyndon tree associated with the Lyndon word \( \ell \) is the (planar rooted binary) tree obtained by computing, recursively down to letters, the standard factorization of \( \ell' \) and \( \ell'' \), and that of \( (\ell')' \), and \( (\ell')'' \) and so on. Fig. 1 shows the Lyndon tree associated with \( \ell = aababb \). Note that each Lyndon tree is complete, that is, every interior vertex has both a right and left son. We will only deal with complete planar rooted binary trees, having their leaves labelled by letters of \( A \), which will simply be called trees from now on.

The fundamental result concerning Lyndon words is the factorization theorem:

**Theorem 2.7** (Chen et al. [3], see also Lothaire [5]). *Any non-empty word is a unique product of non-increasing Lyndon words. That is, given any non-empty word \( w \in A^* \), there exist \( \ell_1, \ldots, \ell_n \in L \ (n \geq 1) \), with \( \ell_1 \geq \cdots \geq \ell_n \) such that \( w = \ell_1, \ldots, \ell_n \).

For example, we have \( abaababaabaabaabab = (ab)(aabab)(aaabaababaabab) \).

Theorem 2.7 extends to right infinite words. We shall not detail this extension here, but refer the reader to [9] (see also [7]). The next proposition describes the Lyndon factorization of the Fibonacci word.

**Proposition 2.8** (Melançon [8, Proposition 3.2]). Let \( \varphi : A^* \rightarrow A^* \) be the morphism defined by \( a \mapsto ab \) and \( b \mapsto ab \). Define words by \( \ell_0 = ab \) and \( \ell_{n+1} = \varphi(\ell_n) \), for \( n \geq 0 \). Then \( (\ell_n)_{n \geq 0} \) is a sequence of decreasing Lyndon words and we have

\[
f = \prod_{n=0}^{\infty} \ell_n.
\]
Thus, we have \( f = (ab)(aabayb)(aabaybabaabayb) \cdots \).

**Remark 2.9.** (1) The length of \( \ell_n \) is \( F_{2n+2} \). This is easily verified by using the morphism \( \varphi \) and by noting that \( |\varphi(w)|_a = 2|w|_a + |w|_b \) and \( |\varphi(w)|_b = |w|_a + |w|_b \).

(2) Berstel had pointed out that the Lyndon words \( \ell_0, \ell_1, \ell_2, \ldots \) were concatenation of two consecutive singular factors, i.e. \( \ell_0 = ab = (a)(b) = w_0w_1 \), \( \ell_1 = aabayb = (aa)(bab) = w_2w_3 \), etc. This is in accordance with the fact that \( |\ell_n| = F_{2n+2} = F_{2n+1} + F_{2n} = |w_{2n+1}| + |w_{2n}| \). Now, the word \( \ell_n \) is also equal to \( af_{2n}f_{2n+1}a^{-1} \), as may be directly verified. Hence, Berstel’s claim is correct since \( af_{2n}f_{2n+1}a^{-1} = (af_{2n}b)^{-1}(bf_{2n+1}a^{-1}) \), as noted in Remark 2.4(1).

(3) As a consequence, Eq. (1) reproves Theorem 2.5 ([10, Theorem 1]).

(4) Note that, from the definition of the words \( \ell_n \), we find: \( w_{2n+2} = \varphi(w_{2n})b^{-1} \) and \( w_{2n+3} = b\varphi(w_{2n+1}) \) \( (n \geq 0) \).

(5) The morphism \( \varphi \) preserves the standard factorization of the words \( \ell_n \). More precisely, we have \( \ell'_{n+1} = \varphi(\ell'_n) \) and \( \ell''_{n+1} = \varphi(\ell''_n) \). This property has a geometrical interpretation: to obtain the Lyndon tree of \( \ell_{n+1} \) one only needs to replace in that of \( \ell_n \) the leaves labelled by \( a \) by the Lyndon subtree \((a,(a,b))\) and those labelled by \( b \) by the Lyndon subtree \((a,b)\) see Fig. 2.

2.2. \( L-R \) operators

Remark 2.9(2), proving \( \ell_n = w_{2n}w_{2n+1} \), may be refined. For this, we need to define operators \( L \) and \( R \) corresponding to paths in a tree. The idea we describe here is intuitively clear and is best described with pictures (see the figures), although we do need to translate it with proper notations. Let it be understood that \( L \) and \( R \) act on a given tree and let \( x \) be an interior vertex of that tree. Then, we denote by \( Lx \) (resp. \( Rx \)) the left (resp. right) son of \( x \).

We will use sequences of operators \( L \) and \( R \) always acting from the root of Lyndon trees. We will denote both the Lyndon word and the tree associated with that word by \( \ell \). For example, Fig. 3 illustrates the effect of the operator \( LRL \) over the Lyndon tree associated with the Lyndon word \( aabaabab \). Note that any sequence of \( L-R \) operators
is of the form $R^{a_0}L^{a_1} \ldots R^{a_{2n-2}}L^{a_{2n-1}}$ (with $a_0$, $a_{2n-1} \geq 0$ and $a_i > 0$ for all $1 \leq i \leq 2n - 2$) and acts from the right; that is,

$$R^{a_0}L^{a_1} \ldots R^{a_{2n-2}}L^{a_{2n-1}}, \ell = \ldots (R \cdots (R(L \cdots (L, \ell) \cdots )) \cdots)$$

For any vertex $x$ of a tree, there is a unique path going from the root down to $x$ described by a unique sequence of operators $R^{a_0}L^{a_1} \ldots R^{a_{2n-2}}L^{a_{2n-1}}$. Suppose that $x$ is an interior vertex of the tree associated with $\ell'$; then, the $L$-$R$ path going from the root down to $x$ determines a unique decomposition of $\ell'$ as a product $\ell' = uv$, with $u, v \in A^*$ non-empty. We write $R^{a_0}L^{a_1} \ldots R^{a_{2n-2}}L^{a_{2n-1}}, \ell = (u, v)$. The decomposition illustrated in Fig. 4 is precisely $(RLR).aabaaab = (aabaa, bab)$. Note that, with this convention, the identity operator gives the decomposition $(\ell', \ell'')$ cutting the Lyndon tree $\ell'$ at its root.

**Proposition 2.10.** We have: $(RL)^n.\ell_n = (w_{2n}, w_{2n+1})$ and $(RL)^{n-1}R.\ell'_n = (w_{2n}, w_{2n-1})$, for all $n \geq 1$.

The two statements are proved similarly; so we shall only prove the first one. Moreover, the proof is best understood using pictures; see the figures. We proceed by induction. Suppose $(RL)^n.\ell_n = (w_{2n}, w_{2n+1})$ and that, moreover, the left and right sons of the vertex $(RL)^n.\ell_n$ are leaves (Fig. 5).
The tree associated with \( \ell_{n+1} \) is obtained from that associated with \( \ell_n \) by replacing the leaves labelled by \( a \)'s with \( (a,(a,b)) \) and those labelled by \( b \)'s with \( (a,b) \) (cf. Remark 2.9(5)). Hence, the factorization induced by the operator \((RL)^n\) on \( \ell_{n+1} \) is \((RL)^n \ell_{n+1} = (\varphi(w_{2n}), \varphi(w_{2n+1}))\) (Fig. 6). Now, recall from Remark 2.4 that \( w_{2n} \) ends with \( aa \); thus, the left subtree attached to the vertex \((RL)^n \ell_{n+1} = (a,(a,b))\).

Since \( w_{2n+2} = \varphi(w_{2n})b^{-1} \) and \( w_{2n+3} = b \varphi(w_{2n+1}) \) (cf. Remark 2.9(4)), we see that the decomposition \((w_{2n+2}, w_{2n+3})\) is obtained by going down this left subtree following the path \( RL \). Thus \((w_{2n+2}, w_{2n+3}) = RL (RL)^n \ell_{n+1} = (RL)^{n+1} \ell_{n+1}\) (Fig. 7).

### 2.3. Self-similarity

In this section, we exhibit a self-similarity property of factorization (1) which leads, as a corollary, to a new proof of Theorem 2.6 [10, Theorem 2].
Theorem 2.11. We have
1.

\[ f = \left( \prod_{j=0}^{n-1} \ell_j \right) \varphi^n(f). \]  

Moreover, \( \varphi^n(f) \) is equal to the Fibonacci word over the alphabet \( \{ \ell'_n, \ell''_n \} \).

2. Furthermore, \( \varphi^n(f) \) is also equal to the Fibonacci word over the alphabet \( \{ \ell_n, \ell'_n \} \).

For example, with \( n = 1 \), we have \( \ell'_1 = aab \) and \( \ell''_1 = ab \). And Eq. (2) reads

\[ f = (ab)(aab)(ab)(aab)(ab)(aab)(ab)(aab)(aab) \cdots. \]

**Proof of Theorem 2.11.** An easy induction shows that, for any \( n \geq 0 \), \( \varphi^n(a) = \ell'_n \) and \( \varphi^n(b) = \ell''_n \), from which we find \( \varphi^n(ab) = \ell_n \). Let \( m \geq n \); then \( \ell'_m = \varphi^n(\varphi^{m-n}(a)) \) and \( \ell''_m = \varphi^n(\varphi^{m-n}(b)) \), so \( \ell_m = \varphi^n(\ell_{m-n}) \); this shows \( \prod_{m \geq n} \ell_m = \varphi^n(f) \). This proves part 1.

Part 2 follows from the fact that the morphism \( a \mapsto ab, \ b \mapsto a \) leaves \( f \) invariant.

To see this observe that the sequence \( (f_n)_{n \geq 1} \) is obtained using the same recurrence \( f_{n+1} = f_n f_{n-1} \) using as initial terms \( f_1 = a, \ f_2 = ab \). That is, \( f \) is equal to the Fibonacci word over the alphabet \( \{ ab, a \} \).

2.4. A new proof of Theorem 2.6

Recall that \( \ell_n = w_{2n}w_{2n+1} \) and that, by Proposition 2.10, we have \( \ell'_n = w_{2n}w_{2n-1} \); so \( \ell''_n = w_{2n-2}w_{2n-1} \) since \( \ell''_n = \ell_{n-1} \) (use Remark 2.9(5)). Denote by \( f_{(x,y)} \) the Fibonacci word over the alphabet \( \{ x, y \} \). Thus in case 1 of Theorem 2.11, Eq. (2) reads

\[ f = \left( \prod_{j=0}^{n-1} \ell_j \right) f(\ell'_n, \ell''_n) = \left( \prod_{j=0}^{n-1} w_{2j}w_{2j+1} \right) f(w_{2n}w_{2n-1}, w_{2n-2}w_{2n-1}). \]
Note that this is also equal to

\[ f = \left( \prod_{j=0}^{2n-2} w_j \right) w_{2n-1} f_{\{w_{2n}, w_{2n-1}, w_{2n-2}w_{2n-1}\}}. \]

Thus, \( f_{\{w_{2n}, w_{2n-1}, w_{2n-2}w_{2n-1}\}} \) is obtained by first forming the Fibonacci word over the alphabet \( \{w_{2n}, w_{2n-2}\} \) and then inserting the word \( w_{2n-1} \) before each occurrence of \( w_{2n} \) or \( w_{2n-2} \). This is precisely what says Theorem 2.6, for \( m=2n-1 \) odd.

In case 2 of Theorem 2.11, Eq. (2) reads

\[ f = \left( \prod_{j=0}^{n-1} e_j \right) f_{\{c_n, c_n'\}} = \left( \prod_{j=0}^{2n-1} w_j \right) f_{\{w_{2n}, w_{2n-1}, w_{2n-2}w_{2n-1}\}} \]

provides a proof for \( m=2n \) even.

3. Characteristic sturmian words

The Fibonacci word is a famous and important example of a general family of infinite words called sturmian words. Consequently, it is natural to look for a generalization of results in Sections 2 and 2.1.

**Definition 3.1.** Let \((c_n)_{n \geq 0}\) be a sequence of integers satisfying \(c_0 \geq 0\) and \(c_n > 0\), for \(n > 0\). Define \(s_0 = b, s_1 = a\) and \(s_{n+1} = s_n^c s_{n-1}\). Then \(s = \lim_{n \to \infty} s_n\) is a well-defined infinite word.

The sequence \((c_n)_{n \geq 0}\) is called the directive sequence of \(s\). Moreover, \(s\) is a characteristic sturmian word.

**Remark 3.2.** The Fibonacci word is a special case of a sturmian word having \(c_n = 1\), for all \(n \geq 0\). General sturmian words may be defined geometrically: let \(y = ax + \beta\) be a line, with \(a > 0\) irrational. Consider the grid formed by the lines \(y = p, x = q\) where \(p, q\) are integers satisfying \(p, q \geq 0\). Denote by \(a's\) and \(b's\) the horizontal and vertical crossings of the line \(y = ax + \beta\) on this grid (since \(a\) is irrational the line crosses the grid in at most one point with integer coordinates). This infinite word thus obtained is the sturmian word associated with the line \(y = ax + \beta\). One may show that two sturmian words associated to lines having equal slopes have the same set of factors. Hence, as far as factors of sturmian words are concerned, it is sufficient to study those having \(\beta = 0\). In that case, if \(a\) has its simple continued fraction equal to \([c_0, c_1, \ldots]\) then the word \(s\) in Definition 3.1 is the sturmian word associated to the line \(y = ax\). For more details, the reader may see [1].

**Remark 3.3.** Observe that \(c_0 = 0\) implies \(s_2 = s_0\); consequently, the sturmian word associated with the sequence \((c_n)_{n \geq 0}\) with \(c_0 = 0\) is obtained from the sturmian word associated to the sequence \((c_n')_{n \geq 0}\) with \(c_n' = c_{n+1}\) by exchanging all letters \(a\) and \(b\). From now on, we shall only consider sequences satisfying \(c_0 > 0\).
In [8], we gave the Lyndon factorization of any general characteristic sturmian word \( s \); more precisely, we proved

\[
s = \prod_{n=0}^{\infty} [(as_{2n+1}a^{-1})^{c_{2n}-1}as_{2n}s_{2n+1}a^{-1}]^{c_{2n+1}},
\]

where \([(as_{2n+1}a^{-1})^{c_{2n}-1}as_{2n}s_{2n+1}a^{-1}]_{n \geq 0} \) is a sequence of strictly decreasing Lyndon words. We write

\[
\ell_n = (as_{2n+1}a^{-1})^{c_{2n}-1}as_{2n}s_{2n+1}a^{-1} \quad \text{and} \quad u_n = as_{2n}s_{2n+1}a^{-1}.
\]

For instance, we have \( \ell_0 = a^\omega b, \ell_1 = (a(a^\omega b)^\omega b)^\omega (a^\omega b) \), and so forth. The word \( u_n \) is a Lyndon word. Moreover, we have \( u'_n = as_{2n+1}a^{-1} \), so that \( \ell_n = (u'_n)^{c_{2n}-1}u_n \). This a key fact when proving that \( (\ell_n)_{n \geq 0} \) is a sequence of decreasing Lyndon words (see [8]).

We shall make use of two formulas borrowed from [8, Eqs. (5) and (6)]; they are recurrence relations that describe the tree structure of \( u_n \) and \( u'_n \), hence of \( \ell_n \) (\( n \geq 1 \)). They are:

\[
u_{n+1} = (as_{2n+1}a^{-1})[(as_{2n+1}a^{-1})^{c_{2n}-1}as_{2n}s_{2n+1}a^{-1}]^{c_{2n+1}};
\]

\[
(\cdots ((u_n, \ell_n), \ell_n), \cdots, \ell_n).
\]

\[
u'_{n+1} = (as_{2n+1}a^{-1})[(as_{2n+1}a^{-1})^{c_{2n}-1}as_{2n}s_{2n+1}a^{-1}]^{c_{2n+1}};
\]

\[
(\cdots ((u'_n, \ell_n), \ell_n), \cdots, \ell_n).
\]

Moreover, we have \( \ell'_n = u'_n = as_{2n+1}a^{-1} \) and

\[
\ell_n = (u'_n (\cdots (u'_n, u_n) \cdots ))
\]

We may formulate a self-similarity property analog to Theorem 2.11.

Theorem 3.4. We have

\[
s = \left( \prod_{j=0}^{n-1} \ell^{c_{2j+1}} \right) \times \tilde{s},
\]

where \( \tilde{s} \) is the sturmian word with directive sequence \( (d_m)_{m \geq 0} \) over the alphabet \( \{u'_m, u'_n\} \), with \( d_m = c_{m+2n} \). Moreover, the word \( \tilde{s} \) is also equal to the Sturmian word with directive sequence \( (d'_m)_{m \geq 0} \) with \( d'_0 = d_0 - 1 \) and \( d'_m = d_m \) (\( m \geq 1 \)) over the alphabet \( \{u'_m, u_n\} \).

Again, \( \{u'_n, u'_n\} \) and \( \{u'_n, u_n\} \) may be considered as alphabets (codes) (cf. the proof of Theorem 2.11). Denote by \( t_1 \) the sturmian word over \( \{a, b\} \) with directive sequence.
Denote by \((d_m)_{m \geq 0}\) the Lyndon words in the Lyndon factorisation of \(t_1\) and consider the morphism \(\gamma\) sending \(a \mapsto u'_n\) and \(b \mapsto u''_n\). We claim that \(\ell_{m+2n} = \gamma(\ell_{m}^{(1)}).\) This is easily shown by induction using Eqs. (12)–(14) and proves the first statement. Similarly, consider the sturmian word \(t_2\) with directive sequence \((d_m)_{m \geq 0}\) and denote by \((d_m^{(2)})_{m \geq 0}\) the Lyndon words in the Lyndon factorization of \(t_2\). Again, an induction shows that \(\ell_{m+2n} = \theta(\ell_{m}^{(2)}\) where \(\theta\) is the morphism sending \(a \mapsto u'_n\) and \(b \mapsto u_n\). This proves the second statement.

4. General singular factors

This section introduces general singular factors (of a given characteristic sturmian word) and contains results generalizing those in [10].

**Definition 4.1.** Suppose the sequence \((c_n)_{n \geq 0}\) is given. Let \(n \geq 2\) and suppose \(s_n\) ends with \(\alpha \beta\) (where \(\alpha, \beta \in A\) and \(\alpha \neq \beta\)). We define the word \(w_n\) by \(w_n = \alpha s_n \beta^{-1}\). We also define \(w_0 = a\), \(w_1 = b\).

Hence, e.g., \(w_2 = a^{s_2^0+1}\), \(w_3 = b(a^{s_3^0} b)^{s_3^0}\), and so on (since \(s_2 = s_1^0 \alpha s_0 = a^{s_0} b, s_3 = s_1^0 \alpha s_1 = (a^{s_2^0} b)^{s_2^0} a\), etc.). Let us first verify that \(w_n\) is indeed a factor of \(s\). Again, any conjugate of the word \(s_n\) is a factor of \(s\). So, the fact that \(w_{2n}\) and \(w_{2n+1}\) are factors of \(s\) follows from \(a s_{2n} s_{2n+1} a^{-1} = (a s_{2n} b^{-1}) (b s_{2n+1} a^{-1})\). Observe also that \(u_n = w_{2n} w_{2n+1}\).

4.1. Properties of general singular factors

Section 4.1 contains results that generalize those given in [10, Lemma 2, Property 2] for the Fibonacci word and confirms the words \(w_n\) as the proper generalization of singular factors of the word \(s\) (associated with \((c_n)_{n \geq 0}\) at the time of writing we were not able to determine if the authors of [10] had already proposed a generalization of their work, and if so, whether their methods compare to the ones we expose in this subsection. Denote by \(q_n\) the length of the word \(s_n\). That is, we have \(q_0 = q_1 = 1\) and \(q_{n+1} = c_n q_n + q_{n-1}\). Denote by \(\mathcal{C}_k(u)\) the conjugate of order \(k\) of the word \(u\). That is, if \(u = u_0 u_0\) with \(|u_0| = k\), then \(\mathcal{C}_k(u) = v_0 u_0\). Note that indices are taken mod \(|u|\), so we may allow negative indices and write, for instance, \(\mathcal{C}_{-1}(u) = u_0 z\) if \(u = z u_0\), with \(z \in A\). Observe also that \(\mathcal{C}_{-1}(u^p) = \mathcal{C}_{-1}(u)^p\).

**Lemma 4.2.** (1) The factor \(w_n\) is not a proper conjugate of \(s_n\).

(2) The set of factors of length \(q_n\) of \(s_{n-1} s_n\) is equal to \(\{\mathcal{C}_k(s_n) | 0 \leq k \leq q_{n-1} - 2\} \cup \{w_n\}\).

The first statement is clear since

\[
(|w|_a, |w|_b) = \begin{cases} 
(|s|_a - 1, |s|_b + 1) & \text{if } n \text{ is odd}, \\
(|s|_a + 1, |s|_b - 1) & \text{if } n \text{ is even}.
\end{cases}
\]
Suppose $n$ is given and that $s_{n+1}$ ends with $\alpha \beta$ ($\alpha, \beta \in A$, $\alpha \neq \beta$). We claim

$$s_{n+1} \alpha^{-1} \beta^{-1} \alpha \beta = s_n c_{n-1} s_{n+1} s_n,$$

$$s_n s_{n+1} \alpha^{-1} \beta^{-1} \alpha \beta = s_{n+1} s_n.$$

(6)

Proceed by induction. First, we have

$$s_{n+1} \alpha^{-1} \beta^{-1} \alpha \beta = s_n c_{n-1} s_{n-1} \alpha^{-1} \beta^{-1} \alpha \beta$$

$$= s_n c_{n-1} (s_{n-1} s_{n-2}) s_{n-1} \alpha^{-1} \beta^{-1} \alpha \beta$$

$$= s_n c_{n-1} s_{n-1} s_{n-1} s_{n-2}$$

(we use the induction and apply second equality (6))

$$= s_n c_{n-1} s_{n-1} s_{n-1} s_{n-2}$$

$$= s_n c_{n-1} s_{n-1} s_n.$$ Second, we use the equality just proved,

$$s_n s_{n+1} \alpha^{-1} \beta^{-1} \alpha \beta = s_n (s_n c_{n-1} s_{n-1} s_n) = s_{n+1} s_n.$$

We may now prove point 2. Suppose $s_n$ ends with $\alpha \beta$; applying Eq. (6) we find $s_{n-1} s_n \alpha^{-1} \beta^{-1} \alpha \beta = s_n s_{n-1}$. Thus, the first factors (of length $q_n$) of $s_{n-1} s_n$ are the conjugates $C_k(s_n)$ with $1 \leq k \leq q_{n-1} - 2$. The next factor is just $\beta s_n \alpha^{-1} = w_n$. The last one is $C_{q_n}(s_n) = C_0(s_n) = s_n$.

Before continuing on with properties of the words $w_n$, we need to introduce words $v_n$, $n \geq 0$, defined by

$$v_n = \alpha s_{n+1} s_n \beta^{-1},$$

(7)

where $\alpha$ and $\beta$ have appropriate values according to the parity of $n$. That is, $v_n$ differs from $w_{n+2}$ by a factor $s_{n+1}$. It is useful to set $v_{-1} = \varepsilon$ (cf. Corollary 4.6). Observe that

$$v_n = C_{-1}(s_{n+1}) c_{n-1} w_n$$

(8)

(so $v_n = w_n$ if $c_n = 1$). Similarly, note also that

$$w_{n+1} = C_{-1} (s_n) c_{n-1} w_{n-1}.$$  

(9)

Proposition 4.3. (1) For all $n \geq 0$, the words $v_n$ and $w_n$ are palindromes and we have

$$v_n = (w_n v_{n-1})^{c_{n-1}} w_n = w_n (v_{n-1} w_n)^{c_{n-1}} \quad (n \geq 1),$$

(10)

$$w_n = (w_{n-2} v_{n-3})^{c_{n-2}} w_{n-2} = w_{n-2} (v_{n-3} w_{n-2})^{c_{n-2}} \quad (n \geq 3).$$

(11)

Moreover, for all $n \geq 2$, $v_n = v_{n-2} w_{n-1} \alpha^{-1} \alpha = \alpha^{-1} w_{n-1} v_{n-2}$, where $\alpha = a$ if $n$ is even and $\alpha = b$ if $n$ is odd.

(2) The words $v_n, w_n$ start and end with $a^{c_{n+1}}$ if $n$ is even, and with $b$ if $n$ is odd ($n \geq 2$).
As a consequence, for any $n \geq 2$, no proper conjugate of $v_n$ or $w_n$ is a factor of $s$. Moreover, the words $v_n^2$ and $w_n^2$ are not factors of $s$.

3. For all $n \geq 0$, the word $w_n$ is not a factor of the word $w_{n+1}$.

4. The word $w_n$ is not the product of two non-empty palindromes. As a consequence, it is primitive.

5. We have

$$C_{q_n-1}(s_n) = \mathcal{C}_{q_n-1}(s_{n-1})^{c_{n-2}-1} w_{n-2} w_{n-1},$$

$$C_{q_n-1}(s_n) = w_{n-1} \mathcal{C}_{q_n-1}(s_{n-1})^{c_{n-2}-1} w_{n-2}.$$

Consequently,

the word $w_{n-2}$ is a factor of $\mathcal{C}_k(s_n)$ if and only if $0 \leq k \leq c_{n-2}q_{n-1} - 1$;

the word $w_{n-1}$ is a factor of $\mathcal{C}_k(s_n)$ if and only if $q_{n-1} - 1 \leq k \leq q_n - 1$.

6. We have $w_n = \alpha \cdot (\prod_{k=0}^{n-2} v_k) = \prod_{k=0}^{n-2} v_{n-2-k} \cdot \alpha$, where $\alpha = a$ if $n$ is even and $\alpha = b$ if $n$ is odd ($n \geq 2$).

1. We claim that for any $p \geq 0$, $\mathcal{C}_{q_n-1}(s_{n-1})^p w_{n-2}$ is a palindrome and prove it by induction on $n$. The claim is trivially true for $n = 2$ since $w_0 = s_1 = a$. We have

$$\mathcal{C}_{q_n-1}(s_n)^p w_{n-1} = (\alpha s_n^{c_n-2} s_{n-2} \alpha^{-1})^p \cdot w_{n-1} = (\alpha s_{n-1} s_{n-2} \alpha^{-1})^p \cdot w_{n-1}$$

$$= \prod_{k=0}^{n-2} v_{n-2-k} \cdot \alpha$$

So, the claim is proved and the fact that $v_n$ and $w_n$ are palindromes follows from Eqs. (8) and (9). Finally, for Eq. (10), observe that by virtue of Eq. (8) it suffices to show $\mathcal{C}_{q_n-1}(s_{n+1}) = w_n v_{n-1}$. This follows from

$$\mathcal{C}_{q_n-1}(s_{n+1}) = \alpha s_n \beta^{-1} (\beta s_n^{c_n-1} \alpha^{-1} s_{n+1} \alpha^{-1}).$$

Eq. (11) is proved similarly. As for the last equality, we use Eq. (7) together with the preceding result to get

$$w_n = \alpha s_n \beta^{-1} = \alpha (s_n \beta^{-1} \alpha^{-1} \beta \alpha) \alpha^{-1} \beta^{-1} \alpha$$

$$= (\alpha (s_{n-1} \beta^{-1}) \alpha^{-1}) (\beta s_{n-1} \alpha^{-1}) \beta^{-1} \alpha$$

$$= v_{n-2} w_{n-1} \beta^{-1} \alpha.$$

The last equality follows from the fact that $w_n$ is a palindrome.

2. Recall that $c_0 > 0$ (cf. Remark 3.3): It is easy to observe that $s_n$ starts with $\alpha^a$, for $n \geq 2$, and that it ends with $b^a$ for $n$ even, and with $a$ for $n$ odd. Hence, the first statement follows from the definitions for $w_n$ and $v_n$ and from the fact that they are palindromes. The other statements are immediate since $\alpha^{a+2}$ and $bb$ are not factors of $s$, since they are not factors of $s_n$, for any $n \geq 0$.

3. First observe that $w_n$ is not a factor of $s_{n+1}$. Indeed, this follows from Lemma 4.2 (1) since any factor of $s_{n+1} = s_n \alpha s_{n-1}$ is conjugated to $s_n$. Suppose that $s_{n+1} \beta$ ends with $\alpha$. Then $w_n$ is not a factor of $s_{n+1} \beta^{-1}$. Now, since $w_n = \beta s_n \alpha^{-1}$ and $w_{n+1} = \alpha s_{n+1} \beta^{-1}$ (with $\alpha, \beta \in \{a, b\}$ and $\alpha \neq \beta$), we find that $w_n$ cannot be a factor of $w_{n+1}$.

4. Suppose that $w_n = uv$ with $u, v$ palindromes. Since $w_n$ is a palindrome, we would have $w_n = uu$ contradicting Lemma 4.2(1). If $w_n$ is not primitive, then there exists an
integer $p \geq 2$ and a non-empty word $u$ such that $w_n = u^p$. Then, by point 1, $u$ must be a palindrome; but this contradict point 2 (with $v = u^{p-1}$).

5. The equalities are easily derived from

$$s_n = s_{n-1}^a s_{n-2}$$

$$= (s_{n-1} \alpha^{-1}) (s_{n-2} \beta^{-1} \gamma^{-1}) (\gamma s_{n-2} \beta^{-1}) \beta.$$  

The last part of the statement then follows from points 2 and 3.

6. First note that $s_n = s_1 s_0 s_2 s_1 \cdots s_{n-1} s_{n-2}$. So,

$$w_n = \alpha s_n \beta^{-1}$$

$$= \alpha (s_1 \alpha^{-1} s_0 \beta^{-1} \gamma^{-1}) \alpha (s_2 \alpha^{-1} s_1 \alpha^{-1}) \cdots \alpha (s_{n-1} \alpha^{-1} s_{n-2} \beta^{-1})$$

(with appropriate values for $\alpha$ and $\beta$ according to the parity of $n$). So the identity follows from definition (7) of the words $v_n$. The second identity follows from point 1.

4.2. Self-similarity revisited

This section links Lyndon factors $u_n$ and $\ell_n$ to singular factors $w_n$ of a given sturmian word $s$, and states corollaries generalizing results in [10] (and in Section 2.2).

**Proposition 4.4.** For all $n \geq 1$, we have

$$R^n L^c_1 \cdots R^n L^{c_{2n-1}} . u_n = (w_{2n}, w_{2n+1}).$$  

(12)

$$R^n L^c_1 \cdots R^n L^{c_{2n-1}} . u_n' = (w_{2n}, v_{2n-1}).$$  

(13)

$$R^n L^c_1 \cdots L^{c_{2n-1}} R^{c_{2n-1}} . \ell_n = (v_{2n}, w_{2n+1}).$$  

(14)

We shall make use of Eqs. (3)–(5) introduced earlier. We proceed by induction. Eq. (3) leads to

$$R^n L^c_1 \cdots R^n L^{c_{2n-1}} . u_{n+1} = (u_n' x_0, y_0 \ell_{c_{2n-1}}')$$

with $(x_0, y_0) = R^n L^c_1 \cdots L^{c_{2n-1}} R^{c_{2n-1}} . \ell_n$. Consequently, the result follows by induction together with the identities $u_n' v_{2n} = (w_{2n} v_{2n-1}) v_{2n} = w_{2n+2}$ and $w_{2n+1} \ell_{c_{2n+1}} = w_{2n+1}$ ($v_{2n} w_{2n+1} \gamma_{2n+1} = w_{2n+3}$) (use Proposition 4.3.1).

Similarly, Eq. (4) gives $R^n L^c_1 \cdots R^n L^{c_{2n-1}} . u_n' = (u_n' x_0, y_0 \ell_{c_{2n-1}}')$ with $(x_0, y_0) = R^n L^c_1 \cdots R^{c_{2n-1}} . \ell_n$. Hence, induction together with Proposition 4.3.1 applied to $u_n' v_{2n} = w_{2n} v_{2n-1} v_{2n} = w_{2n+2}$ and $w_{2n+1} (v_{2n} w_{2n+1} \gamma_{2n+1} = w_{2n+1}$ gives the result.

Finally, using Eq. (5) we find

$$R^n L^c_1 \cdots L^c_{2n-1} R^{c_{2n+1}} . \ell_{n+1} = (u_{c_{2n+1}} x_0, y_0)$$

with $(x_0, y_0) = R^n L^c_1 \cdots R^n L^{c_{2n}} . u_{n+1}$. The result follows by induction using the identity $v_{2n+2} = u_{c_{2n+2}} w_{2n+2}$.
As a corollary, Eq. (14) leads to an identity generalizing that in Theorem 2.5.

**Corollary 4.5.** We have

\[
s = \prod_{j=0}^{\infty} (v_{2j} w_{2j+1})^{c_{2j+1}} = \prod_{j=0}^{\infty} v_j.
\]

The first equality is clear. The second one follows from Proposition 4.3.1. Indeed, we have \((v_{2j} w_{2j+1})^{c_{2j+1}} = v_{2j} w_{2j+1}(v_{2j} w_{2j+1})^{c_{2j+1}-1} = v_{2j} v_{2j+1}\).

The next corollary generalizes the identity given in [10, Theorem 2] (Theorem 2.6), although we still have to prove the nonoverlapping property for the words \(w_n\). Let \((d_m)_{m \geq 0}\) and \((d'_m)_{m \geq 0}\) be as in Theorem 3.4.

**Corollary 4.6.** The sturmian word \(s\) may be written as

\[
s = \prod_{j=0}^{n-1} (v_{2j} w_{2j+1})^{c_{2j+1}} (z_1 w_{2n-1}^{(1)} z_2 w_{2n-1}^{(2)} z_3 \cdots),
\]

where \(z_1 z_2 z_3 \cdots\) is the sturmian word with directive sequence \((d_m)_{m \geq 0}\) over the alphabet \(\{w_n, v_{2n-2}\}\) and \(w_{2n-1}^{(i)} = v_{2n-1}\) if \(z_i = w_{2n}\) and \(w_{2n-1}^{(i)} = w_{2n-1}\) if \(z_i = v_{2n-2}\), for all \(i \geq 1\);

2. or,

\[
s = \prod_{j=0}^{n-1} (v_{2j} w_{2j+1})^{c_{2j+1}} (w_{2n}^{(0)} z_1 w_{2n}^{(1)} z_2 w_{2n}^{(2)} z_3 \cdots)
\]

where \(z_1 z_2 z_3 \cdots\) is the sturmian word with directive sequence \((d'_m)_{m \geq 0}\) over the alphabet \(\{v_{2n-1}, w_{2n+1}\}\) and \(w_{2n}^{(i)} = w_{2n}\) for all \(i \geq 0\).

Moreover, given \(n \geq 0\), any two occurrences of \(w_n\) in \(s\) are separated either by \(v_{n-1}\) or by \(w_{n+1}\). Consequently, these above expansions may be obtained by locating the non-overlapping occurrences of \(w_n\) in \(s\).

Let us first look at an example, with \(n = 1\), to illustrate case 2 of the corollary. We have \(w_2 = a^{\infty+1}\), \(v_1 = b(a^{\infty} b)^{c_1-1}\) and \(w_3 = b(a^{\infty} b)^{c_1}\). We compute \(s_5\) as an approximation for \(s\):

\[
s = (((a^{\infty} b)^{c_1} a)^{c_2} a^{\infty} b)^{c_3} (a^{\infty} b)^{c_1} a \cdots.
\]

Writing this as

\[
(a^{\infty} b)^{c_1} \cdot (((a^{\infty+1} b(a^{\infty} b)^{c_1-1})^{c_2-1} a^{\infty+1} b(a^{\infty} b)^{c_1})^{c_3} a \cdots
\]

\[
= (v_0 w_1)^{c_1} \cdot ((w_2 v_1)^{d_1} w_2 v_3)^{d_1} a \cdots
\]

we get the beginning of the expansion predicted by Corollary 4.6.
Proof of Corollary 4.6. Statements 1 and 2 in the corollary are proved as in Section 2.4, by rewriting the identity in Theorem 3.4 using Eqs. (12)–(14) of Proposition 4.4. Indeed, we have by virtue of Theorem 3.4

\[ s = \left( \prod_{j=0}^{n-1} \ell_j^{e_{j+1}} \right) \times \tilde{s}_{\{u'_n, u''_n\}}, \]

where \( \tilde{s}_{\{u'_n, u''_n\}} \) denotes the sturmian word associated with \((d_m)_{m \geq 0}\) over the alphabet \(\{u'_n, u''_n\}\). Observe that \(u''_n = \ell_{n-1}\) (cf. Eq. (5)). Hence, according to Proposition 4.4, this is equal to

\[ \left( \prod_{j=0}^{n-1} (v_{2j}w_{2j+1})^{e_{j+1}} \right) \times \tilde{s}_{\{w_{2n-1}, v_{2n-2}, w_{2n-1}\}}. \]

That is, \( \tilde{s}_{\{w_{2n-1}, v_{2n-2}, w_{2n-1}\}} \) is obtained by first forming the sturmian word associated with \((d_m)_{m \geq 0}\) over the alphabet \(\{w_{2n}, v_{2n-2}\}\) and then insert \(v_{2n-1}\) (resp. \(w_{2n-1}\)) after each occurrence of \(w_{2n}\) (resp. \(v_{2n-2}\)). This is precisely what says part 1 of the corollary. Using part 2 of Theorem 3.4 gives a proof for the second statement.

Hence, we may concentrate on the last statement concerning the non-overlapping property of the word \(w_n\), which follows from the next lemma. We will say that a word \(u \) overlaps the product \(xy\) if \(xy = x'y'\) where \(x', y'\) non-empty are such that \(|x'| < |x|\) and \(|y'| < |y|\) (where \(x, y, x', y'\) are words).

Lemma 4.7. (1) The word \(w_n\) is not a factor of the word \(v_{n-1}\).

(2) The word \(w_n\) does not overlap neither \(w_nv_{n-1}\), nor \(v_{n-1}w_n\).

(3) The word \(w_n\) does not overlap neither \(w_nv_{n+1}\), nor \(w_{n+1}w_n\).

(4) The product \(\prod_{k=0}^{n-1} (v_{2k}w_{2k+1})^{e_{k+1}}\) may be uniquely expressed in terms of \(v_{2n-1}\), \(w_{2n}\) and \(w_{2n+1}\) only. Moreover, this expression is completely determined by locating the occurrences of \(w_{2n}\) in \((v_{2k}w_{2k+1})^{e_{k+1}}\).

1. Suppose on the contrary that \(w_n\) is a factor of \(v_{n-1}\). Then, it is a factor of \(\alpha(\prod_{k=0}^{n-2} v_k)v_{n-1}\). But this last expression is equal to \(w_{n+1}\), by Proposition 4.3(6) (with appropriate value for \(\alpha\)); so we get a contradiction since \(w_n\) is not a factor of \(w_{n+1}\), by Proposition 4.3(6).

2. We have \(w_nv_{n-1} = \alpha(\prod_{k=0}^{n-2} v_k)v_{n-1}\) and \(v_{n-1}w_n = v_{n-1}(\prod_{k=0}^{n-2} v_{n-2-k})(\alpha)\), by Proposition 4.3(6). If \(w_n\) were to overlap \(w_nv_{n-1}\) or \(v_{n-1}w_n\), then it would be a factor of \((\prod_{k=0}^{n-2} v_k)v_{n-1}\) or \(v_{n-1}(\prod_{k=0}^{n-2} v_{n-2-k})(\alpha)\) hence of \(w_{n+1}\). But, again, this contradicts Proposition 4.3(6).

3. Write

\[ w_nw_{n+1} = \left( \alpha \prod_{k=0}^{n-2} v_k \right) \left( \prod_{k=0}^{n-1} v_{n-1-k} \beta \right) = \alpha \left( \prod_{k=0}^{n-2} v_k \right) v_{n-1} \left( \prod_{k=0}^{n-2} v_{n-2-k} \right) \beta. \]
(with appropriate values for \(\alpha, \beta\) where \(\alpha \neq \beta\)). Observe that \(|v_{n-1}| = q_n + q_{n-1}\). Then, either \(c_{n-1} \geq 2\) and then \(|v_{n-1}| = |w_n|\) or \(c_{n-1} = 1\) and \(v_{n-1} = w_n - 1\). Suppose first that \(v_{n-1} = w_n - 1\); then we cannot have \(w_n = x w_{n-1} y\) since, by Proposition 4.3, \(w_{n-1}\) is not a factor of \(w_n\). Suppose now that \(|v_{n-1}| > |w_n|\). Then, if \(w_n\) were to overlap \(w_n w_{n+1}\) it would have to be a factor of \(\prod_{k=0}^{n-2} v_k v_{n-1}\) or \(v_{n-1}(\prod_{k=0}^{n-2} v_{n-2-k})\), hence of \(w_n\). So we may conclude as in case 2. We only need to exchange \(\alpha\) and \(\beta\) to obtain a proof for \(w_n w_{n+1}\).

4. We first prove the existence and unicity of the expansion for any factor \((v_{2k} w_{2k+1})^{2k+1}\). We proceed by induction together with Eqs. (10) and (11) of Proposition 4.3(1) to show, in addition, that \(v_{2k}, v_{2k}, w_{2k} \text{ and } w_{2k+1}\) may be expressed in terms of \(v_{2n-1}, v_{2n}, w_{2n+1}\) only, and that, moreover,

- \(v_{2k}\) and \(v_{2k}\) start and end with \(w_{2n}\),
- \(v_{2k-1}\) and \(w_{2k+1}\) start and end with \(w_{2n+1}\).

For \(k = n\) we have

\[
(v_{2n} w_{2n+1})^{2n+1} = [(w_{2n} v_{2n-1})^{2n-1} w_{2n} w_{2n+1}]^{2n+1}.
\]

Recall that by virtue of Proposition 4.3(2), each occurrence of \(w_{2n}\) in the expansion is necessarily followed by either \(v_{2n-1}\) or \(w_{2n+1}\). The unicity of the expansion follows from points 1–3. Indeed, since \(|v_{2n-1}| = q_n + q_{n-1} = q_n \neq q_{2n+1} = w_{2n+1}|\), any other expansion would provide a situation where either \(w_{2n}\) is a factor of \(v_{2n-1}\), or else overlaps one of \(w_{2n} v_{2n-1}\), \(v_{2n-1} w_{2n}\), \(w_{2n} w_{2n+1}\) or \(w_{2n+1} w_{2n}\). We use Proposition 4.3(1) and compute, for \(k \geq n\):

\[
(v_{2k+2} w_{2k+3})^{2k+3} = [(w_{2k+2} v_{2k+1})^{2k+1} w_{2k+2} (v_{2k+1} w_{2k+1})^{2k+1} w_{2k+1}]^{2k+3}.
\]

Combine this with Eqs. (11) and (10) applied to \(w_{2k+2}\) and \(v_{2k+1}\). This shows the existence of the predicted expansion. Unicity again follows from points 1–3. The fact that the expansion is unique for each factor \((v_{2k} w_{2k+1})^{2k+1}\) implies, by virtue of points 1–3 again, that it is unique also for the infinite product \(\prod_{k \geq n} (v_{2k} w_{2k+1})^{2k+1}\). Hence part 4 of the lemma is established, which ends the proof of Corollary 4.6.

References

Special Issue for FPSAC’96, 8th Int. Conf. on Formal Power Series and Algebraic Combinatorics, to
appear.
[10] Z.-X. Wen, Z.-Y. Wen, Some properties of the singular words of the Fibonacci word, European