Available online at www.sciencedirect.com
ScienceDirect
European Journal of Combinatorics 28 (2007) 1703–1719
European Journal of Combinatorics
www.elsevier.com/locate/ejc

Some properties of the Tribonacci sequence ☆

Bo Tan [a], Zhi-Ying Wen [b]

[a] *Department of Mathematics, Huazhong University of Science and Technology, Wuhan 430074, PR China*
[b] *Department of Mathematics, Tsinghua University, Beijing 100084, PR China*

Received 18 July 2005; accepted 4 July 2006
Available online 9 October 2006

Abstract

In this paper, we consider the factor properties of the Tribonacci sequence. We define the singular words, and then give the singular factorization and the Lyndon factorization. As applications, we study the powers of the factors and the overlap of the factors. We also calculate the free index of the sequence.

0. Introduction

The Tribonacci sequence, which is a natural generalization of the Fibonacci sequence, has been studied extensively by many authors. Rauzy [13] set up a link between the Tribonacci sequence and the distribution in $\mathbb{R}^2$ modulo $\mathbb{Z}^2$ of a certain sequence $(N\eta)_{N\in\mathbb{N}}$. Arnoux and Rauzy [2] gave the geometric representation of the so-called Arnoux–Rauzy sequence, and the Tribonacci sequence is maybe the simplest sequence amongst the Arnoux–Rauzy sequences. Sirvent [18] studied the Tribonacci substitution from the associated numeration system. Justin and Pirillo [8] introduced the epi-Sturmian sequences which also include the Arnoux–Rauzy sequence and they described the palindrome factors of the epi-Sturmian sequences.

For the Fibonacci sequence, Wen and Wen [20] studied the factor structure and gave the singular factorization; Levé and Séébold [9] studied the singular factorization of the Fibonacci sequence and its conjugates; Melançon studied the Lyndon factorization of the Fibonacci sequence [11] and the link between the singular factorization and the Lyndon factorization [12]. Cao and Wen [3] studied the factor property of the Sturmian sequences; in particular, given

☆ The first author is supported by NSFC No. 10501035, and the work is supported by the special Funds for Major State Basic Research Projects of China.

E-mail addresses: bo_tan@163.com (B. Tan), wenzy@tsinghua.edu.cn (Z.-Y. Wen).

0195-6698/$ - see front matter
doi:10.1016/j.ejc.2006.07.007

a Sturmian sequence, they calculated its free index (that is, the highest power of factor). See also [5,19]. It should be mentioned that some of the ideas and techniques are taken from these papers.

In this paper, we shall present some new properties of the factors of the Tribonacci sequence. After giving notation and listing some known facts, we study the factor structure of the Tribonacci sequence, and introduce the singular words, then give the singular factorization of the Tribonacci sequence. The Lyndon factorization is also given. As applications, we study the power property and the overlap property of the factors. We also calculate the free index of the Tribonacci sequence.

1. Notation and basic properties

Let $S = \{a, b, c\}$ be a three-letter alphabet. Let S^* and $\tilde{S}$ stand respectively for the free monoid and the free group generated by S. The empty word ε is their neutral element. Let S^ω be the set of sequences (or infinite words), indexed by $\mathbb{N}$ ($0 \in \mathbb{N}$ by convention), on S.

If $w \in S^*$ is a word, we denote by $|w|$ its length and by $|w|_a$ (resp. $|w|_b$, $|w|_c$) the number of occurrences of the letter a (resp. b, c) in it. Let $P(w)$ stand for the vector $(|w|_a, |w|_b, |w|_c)$, called the Parikh vector of w.

A word v is a factor of a word w, written $v \prec w$, if there exist $u, u' \in S^*$, such that $w = uvu'$. We say that v is a prefix (resp. suffix) of a word w, and then we write $v \lhd w$ (resp. $v \rhd w$) if there exists $u \in S^*$ such that $w = vu$ (resp. $w = uv$). The notions of prefix and factor extend in a natural way to infinite words. The language of length n of w, denoted by $\Omega_n(w)$ (Ω_n for short if there is no confusion), is the set of all factors of length n of w.

If $v \prec w$, where $w = w_0 w_1 \cdots w_n \cdots$ (with $w_i \in S$) is a finite word or a sequence, v is said to occur at place m in w if $w_m w_{m+1} \cdots w_{m+|v|-1} = v$. In this case, we also say that m is a place where u occurs in w.

Let $w = w_0 w_1 \cdots w_{n-1} \in S^*$, where $w_i \in S$. The mirror word $\overline{w}$ of w is defined to be $\overline{w} = w_{n-1} \cdots w_1 w_0$. A word w is called a palindrome if $w = \overline{w}$.

A morphism $\tau : S^* \to S^*$ is called a substitution of S^*. We denote by F_τ any one of the fixed points of τ (i.e. $\tau(F_\tau) = F_\tau$), if it exists, and by M_τ the matrix $(P(\tau(a))^t, P(\tau(b))^t, P(\tau(c))^t)$ (where the superscript ^{t} means the transposition of a vector) called the matrix of the substitution τ. A substitution is said to be primitive if its matrix is.

Let $w = w_0 w_1 \cdots w_{n-1} \in S^*$ ($w_i \in S$); we denote by w^{-1} the inverse word of w, that is $w^{-1} = w_{n-1}^{-1} \cdots w_1^{-1} w_0^{-1}$. Let $w = uv$; then $wv^{-1} := u$ and $u^{-1}w := v$ by convention.

Let $w = w_0 w_1 \cdots w_{n-1} \in S^*$ (with $w_i \in S$) and $0 \le k < |w|$; we define the kth conjugate of w by $C_k(w) := x_k \cdots x_{n-1} x_0 x_1 \cdots x_{k-1}$. The set of conjugates of w is defined by $C(w) := \{C_k(w); 0 \le k < |w|\}$. A word $w \in S^*$ is said to be primitive if $w = u^p$ with $p \in \mathbb{N}$ implies $p = 1$, in other words, if the conjugates of w are distinct (see [10] for example).

Let $w \in \tilde{S}$; we denote by $\imath_w$ the inner isomorphism $u \mapsto wuw^{-1}$, $u \in S^*$. If there exists a $w \in S^*$ such that $\phi = \imath_w\, \tau$ or $\tau = \imath_w\, \phi$, we say that ϕ is conjugate to τ and write $\phi \sim \tau$.

As usual, the substitution $\sigma = (ab, ac, a)$ (this means $\sigma(a) = ab, \sigma(b) = ac, \sigma(c) = a$) is called the Tribonacci substitution and its fixed point

$$\xi = abacabaabacababacabaabac \cdots$$

is called the Tribonacci sequence.

On the Tribonacci sequence, the following are known (see [1,2,4,6] for more details):

0. As a fixed point of a primitive substitution, it is uniformly recurrent.

1. The complexity function is $2n + 1$, that is, for any n, there are $2n + 1$ factors of length n of the Tribonacci sequence ξ.

2. The language is closed under taking the mirror.

3. There is a unique left special word (resp. right special word) of each length, that is, for any n, there is a unique factor w (resp. w') such that aw, bw and cw (resp. $w'a$, $w'b$ and $w'c$) are all factors of ξ. Moreover, every left special word is a prefix of ξ. And from 2, the right special word is the mirror of the left special word.

4. The palindrome complexity is 3 for odd length and 1 for even length, that is, for odd n, there are three palindrome factors of length n, and for even n, there is one palindrome factor of length n.

2. The factor structure of the Tribonacci sequence

In this section, we will study the factor structure of the Tribonacci sequence.

2.1. Elementary properties of the factors

Define the words

$$A_n = \sigma^n(a); \qquad B_n = \sigma^n(b); \qquad C_n = \sigma^n(c).$$

By convention, $A_0 = a$ and $A_{-1} = c$ (and thus $\sigma(A_{n-1}) = A_n$ for $n \geq 0$). Define the number sequence $\{f_n\}_{n\geq -1}$ as

$$f_n = f_{n-1} + f_{n-2} + f_{n-3} \ (n \geq 2) \quad \text{with } f_{-1} = f_0 = 1 \text{ and } f_1 = 2.$$

Then

$$\begin{aligned} A_n &= A_{n-1}A_{n-2}A_{n-3}, & |A_n| &= f_n; \\ B_n &= A_{n-1}A_{n-2}, & |B_n| &= f_{n-1} + f_{n-2}; \\ C_n &= A_{n-1}, & |C_n| &= f_{n-1}. \end{aligned}$$

Proposition 2.1. *$A_n A_m$ is a factor of the Tribonacci sequence for any n, m.*

Proof. Since a is the first letter of the image of each word, $A_n a$ is a factor of the Tribonacci sequence ξ, and thus $A_n A_m$ $(n \geq m)$ is also a factor of ξ (in fact $A_n A_m = \sigma^m(A_{n-m}a)$ and σ fixes ξ).

To see that $A_n A_m$ $(n < m)$ is a factor of ξ, we only need to notice the following fact: $A_n A_m$ is a suffix of $A_{n+3}A_m$. $\square$

For $n \geq 1$, let

$$D_n = A_{n-1}A_{n-2} \cdots A_2 A_1 A_0.$$

Thus $D_1 = A_0 = a$, $D_2 = A_1 A_0 = aba$. By convention, $D_0 = \varepsilon$.

Proposition 2.2. *For $n \geq 1$, D_n is a palindrome.*

Proof. This is a direct result of the following lemma and the fact that $D_n = \sigma(D_{n-1})a$ for $n \geq 1$. $\square$

Lemma 2.3. *If w is a palindrome, $\sigma(w)a$ and $a^{-1}\sigma(w)$ are palindromes.*

Proof. It is easy to see that for any word u, $\overline{\sigma(u)} = a^{-1}\sigma(\overline{u})a$. If w is a palindrome, $\overline{\sigma(w)a} = a\overline{\sigma(w)} = \sigma(\overline{w})a = \sigma(w)a$, and thus $\sigma(w)a$ is a palindrome. Likewise, $a^{-1}\sigma(w)$ is also a palindrome. $\square$

We can say more on the structure of the palindrome factors of ξ.

Lemma 2.4. *If w is a palindrome factor of the Tribonacci sequence, we have*

1. *If $a \lhd w$, then there exists a unique palindrome u factor of ξ such that $w = \sigma(u)a$.*
2. *Otherwise, a is not the prefix of w; then there exists a unique palindrome u factor of ξ such that $w = a^{-1}\sigma(u)$.*

Proof. Mathematical induction on the length of w. $\square$

Theorem 2.5. *For $n \geq 2$:*

1. *The longest common prefix of $A_{n-2}A_{n-3}A_{n-1}$ and A_n is D_{n-1}.*
2. *The longest common prefix of $A_{n-3}A_{n-1}A_{n-2}$ and A_n is D_{n-2}.*

Proof. We have the following:

Claim: For any $n \geq 2$, the word $A_{n-3}A_{n-1}$ is not a prefix of A_n and the maximal common prefix of the words A_n and $A_{n-3}A_{n-1}$ is the word D_{n-2}.

In fact, when n is small:

$$\begin{array}{lll} n=2: & A_2 = abac \quad A_{-1}A_1 = cab & D_1 = \varepsilon \\ n=3: & A_3 = abacaba \quad A_0A_2 = aabac & D_1 = a \\ n=4: & A_4 = abacabaabacab & \\ & A_1A_3 = ababacaba & D_2 = A_1A_0 = aba \\ n=5: & A_5 = abacabaabacababacabaabac & \\ & A_2A_4 = abacabacabaabacab & D_3 = A_2A_1A_0 = abacaba. \end{array}$$

If $n \geq 6$, we have

$$A_n = A_{n-3}A_{n-4}A_{n-5}\underline{A_{n-3}}A_{n-4}A_{n-2}A_{n-3}$$
$$A_{n-3}A_{n-1} = A_{n-3}A_{n-4}A_{n-5}\underline{A_{n-6}A_{n-4}}A_{n-5}A_{n-3}A_{n-4}.$$

Since $A_{n-6}A_{n-4}$ is not a prefix of A_{n-3} and their longest common prefix is D_{n-5}, we know that $A_{n-3}A_{n-1}$ is not a prefix of A_n and the longest common prefix of A_n and $A_{n-3}A_{n-1}$ is $A_{n-3}A_{n-4}A_{n-5}D_{n-5} = D_{n-2}$. The Claim follows.

1. Since $A_n = A_{n-2}A_{n-3}A_{n-4}A_{n-2}A_{n-3}$, the longest common prefix of A_n and $A_{n-2}A_{n-3}A_{n-1}$ is $A_{n-2}A_{n-3}D_{n-3} = D_{n-1}$.
2. This is a direct result of the above Claim. $\square$

Let E_n ($n \geq 1$) and F_n ($n \geq 2$) be the words defined by the equations

$$A_n = D_{n-1}E_n = D_{n-2}F_n.$$

Thus $E_1 = ab$, $E_2 = bac, \ldots$ and $F_2 = abac$, $F_3 = bacaba, \ldots$.

Proposition 2.6. *With the notation above, for $n \geq 2$, we have*

$$A_{n-2}A_{n-3}A_{n-1} = D_{n-1}\overline{E_n};$$
$$A_{n-3}A_{n-1}A_{n-2} = D_{n-2}\overline{F_n}.$$

Proof. From the definitions of D_n and E_n, we have $D_n = \sigma(D_{n-1})a$ and $E_n = a^{-1}\sigma(E_{n-1})$. We also have $\sigma(\overline{E_{n-1}}) = a\overline{\sigma(E_{n-1})}a^{-1}$ (recall that $\sigma(\overline{u}) = a\overline{\sigma(u)}a^{-1}$ for any word u).

We will show that $A_{n-2}A_{n-3}A_{n-1} = D_{n-1}\overline{E_n}$:

The cases $n = 2$ and $n = 3$ can be checked directly.

If $n > 3$, we have

$$\begin{aligned} A_{n-2}A_{n-3}A_{n-1} &= \sigma(A_{n-3}A_{n-4}A_{n-2}) = \sigma(D_{n-2}\overline{E_{n-1}}) \\ &= (D_{n-1}A_0^{-1})(a\overline{\sigma(E_{n-1})}a^{-1}) = D_{n-1}\overline{\sigma(E_{n-1})}a^{-1} \\ &= D_{n-1}\overline{E_n}. \end{aligned}$$

In the same way, we can show the second equality. □

This proposition has the following corollary.

Corollary 2.7. *We have that*

$$A_{n-1}A_n = D_nE_n,\ A_nA_{n-1} = D_n\overline{E_n}\ (n \geq 1) \text{ and } A_{n-2}A_n = D_{n-1}F_n\ (n \geq 2).$$

In the following, we will study the properties of the words D_n, E_n and F_n:

Proposition 2.8. $F_n = \overline{E_{n+1}}$ $(n \geq 2)$ *and* $E_{n+1} = \overline{E_n}A_{n-2}$ $(n \geq 1)$.

Proof. If $n \geq 2$,

$$\begin{aligned} A_{n+1} &= D_nE_{n+1}, \\ A_{n+1} &= A_nA_{n-1}A_{n-2} = A_{n-1}A_{n-2}A_{n-3}A_{n-1}A_{n-2} \\ &= A_{n-1}A_{n-2}D_{n-2}\overline{F_n} = D_n\overline{F_n}, \end{aligned}$$

and then $E_{n+1} = \overline{F_n}$ and $F_n = \overline{E_{n+1}}$.

If $n \geq 1$,

$$\begin{aligned} A_{n+1} &= D_nE_{n+1}, \\ A_{n+1} &= A_nA_{n-1}A_{n-2} = D_n\overline{E_n}A_{n-2}, \end{aligned}$$

and thus $E_{n+1} = \overline{E_n}A_{n-2}$. □

An easy induction provides the following result.

Proposition 2.9. *The lengths of* D_n *and* E_n *satisfy*

$$|D_n| = \frac{1}{2}(f_{n+1} + f_{n-1} - 3) \quad (n \geq 1);$$

$$|E_n| = \frac{1}{2}(f_n - f_{n-2} + 3) \quad (n \geq 2).$$

From here, we have that $|D_n|$ *is even if and only if* $n = 4k$ *for some integer* k *and* $|E_n|$ *is odd if and only if* $n = 4k + 2$ *for some integer* k.

The following proposition on the first letter and the last one of A_n and E_n is easy to check.

Proposition 2.10. *We have:*

if $n \equiv 0 \pmod 3$, *then* $a \rhd A_n$ *and* $c \lhd E_n$, $a \rhd E_n$;

if $n \equiv 1 \pmod 3$, *then* $b \rhd A_n$ *and* $a \lhd E_n$, $b \rhd E_n$;

if $n \equiv 2 \pmod 3$, *then* $c \rhd A_n$ *and* $b \lhd E_n$, $c \rhd E_n$.

2.2. *Occurrences of D_n*

Now we determine the set of places where the factor D_n occurs in ξ. It plays a very important role in the following studies.

For this, write

$$\xi = abacaba\cdots = u_0u_1u_2u_3u_4u_5u_6\cdots$$

with $u_i \in \{a,b,c\}$. Denote by $\mathcal{O}_1$ the set of places where the letter a occurs in ξ, that is, $\mathcal{O}_1 = \{i \in \mathbb{N} : u_i = a\}$. Since ξ is a fixed point of the substitution σ, we can rewrite it as follows:

$$\begin{aligned}\xi = \sigma^{n-1}(\xi) &= \sigma^{n-1}(a)\sigma^{n-1}(b)\sigma^{n-1}(a)\sigma^{n-1}(c)\sigma^{n-1}(a)\sigma^{n-1}(b)\sigma^{n-1}(a)\cdots\\ &= A_{n-1}B_{n-1}A_{n-1}C_{n-1}A_{n-1}B_{n-1}A_{n-1}\cdots.\end{aligned}$$

Define the set $\mathcal{O}_n = \{|\sigma^{n-1}(u_0u_1\cdots u_{i-1})|;\, u_i = a\}$ ($0 \in \mathcal{O}_n$ since $u_0 = a$). We can regard $\mathcal{O}_n$ as the set of the places where there is a formal (or visible) occurrence of A_{n-1} in the above factorization of ξ. Here we should point out that, when $n \geq 3$, $\mathcal{O}_n$ is not the set of all places where the word A_{n-1} occurs in ξ (merely a subset).

Theorem 2.11. *For $n \geq 1$, the set of places where the factor D_n occurs in ξ is $\mathcal{O}_n$.*

Proof. Induction on n: From $D_1 = a$ there follows the result for $n = 1$.

Suppose that $\mathcal{O}_n$ is the set of places where D_n occurs in ξ. We need to show that the set of places where D_{n+1} occurs in ξ is $\mathcal{O}_{n+1}$ which by definition is the set $\{|\sigma^n(u_0u_1\cdots u_{i-1})|;\, u_i = a\}$.

On the one hand, $D_{n+1} = \sigma(D_n)a$. Thus, by considering the image under σ (and noticing that a is a prefix of $\sigma(w)$ for any word w), at any place in $\mathcal{O}_{n+1}$ there is an occurrence of D_{n+1} in ξ.

On the other hand, suppose that D_{n+1} occurs at place t in ξ, that is

$$\xi = u_0u_1\cdots u_{t-1}u_t\cdots \ (u_i \in \{a,b,c\}) \quad \text{and} \quad u_tu_{t+1}\cdots u_{t+|D_{n+1}|-1} = D_{n+1}.$$

The following fact is readily checked: If w is a prefix of ξ with $a \rhd w$, then there is a unique prefix w' of ξ, such that $wa^{-1} = \sigma(w')$.

Since $u_t = u_{t+|D_{n+1}|-1} = a$, using the above fact twice, there exist $r, s \in \mathbb{N}$ with $s < r$ such that

$$\begin{aligned}&u_0u_1\cdots u_{t-1} = \sigma(u_0u_1\cdots u_{s-1})\\ &u_0u_1\cdots u_{t-1}D_{n+1}a^{-1} = \sigma(u_0u_1\cdots u_{s-1}u_s\cdots u_{r-1}).\end{aligned}$$

Since $D_{n+1}a^{-1} = \sigma(D_n)$ (and the substitution σ is injective), we have that $u_s\cdots u_{r-1} = D_n$, and thus s is a place where D_n occurs in ξ.

Now by the induction hypothesis, $s \in \mathcal{O}_n$. Hence there is an integer $i \geq 1$ with $u_i = a$ such that $u_0u_1\cdots u_{s-1} = \sigma^{n-1}(u_0u_1\cdots u_{i-1})$. Therefore, $u_0u_1\cdots u_{t-1} = \sigma(u_0u_1\cdots u_{s-1}) = \sigma^n(u_0u_1\cdots u_{i-1})$ and $t \in \mathcal{O}_{n+1}$. $\square$

2.3. *Structure of the factors*

Now we analyze the structure of the factors of length f_n of the Tribonacci sequence ξ.

Proposition 2.12. *For any n, a factor of the Tribonacci sequence of length f_n is either:*
1. *a conjugate word of A_n;*
2. *a factor of $A_{n-2}A_{n-3}A_{n-1}A_n$; or*
3. *a factor of $A_{n-3}A_{n-1}A_{n-2}A_n$.*

Proof. Since ξ is the fixed point of σ,

$$\xi = \sigma^n(\xi) = A_n B_n A_n C_n A_n B_n A_n \cdots.$$

Due to the fact that the words bb, bc, cb, cc are not factors of ξ, each factor of length f_n of ξ will be a factor of one of the following words:

- $A_n A_n$: the factors of length f_n of $A_n A_n$ are just the conjugates of A_n;
- $A_n C_n A_n$: since $A_n C_n A_n = A_{n-1}A_{n-2}A_{n-3}A_{n-1}A_n$, the first $|A_{n-1}| = f_{n-1}$ factors of length f_n are the conjugates of A_n; other factors are just the factors of $A_{n-2}A_{n-3}A_{n-1}A_n$;
- $A_n B_n A_n$: since $A_n B_n A_n = A_{n-1}A_{n-2}A_{n-3}A_{n-1}A_{n-2}A_n$, the first $|A_{n-1}| + |A_{n-2}|$ factors of length f_n are the conjugates of $A_{n-1}A_{n-2}A_{n-3} = A_n$; other factors are just the factors of $A_{n-3}A_{n-1}A_{n-2}A_n$.

On the other hand, the factors of the above three words are obviously factors of ξ. □

By Proposition 2.12, the set of factors of length f_n ($n \geq 2$) can be divided into the following three classes:

- $\Omega_n^0 = \{\text{conjugates of } A_n\}$.
- $\Omega_n^1 = \{\text{factor of length } f_n \text{ of the word } \alpha^{-1}\overline{E_n}D_{n-1}E_n\alpha^{-1}\}$.
  (where α is the last letter of E_n.)
  By Proposition 2.6 and the definition of E_n, $A_{n-2}A_{n-3}A_{n-1}.A_n = D_{n-1}\overline{E_n}.D_{n-1}E_n$, and thus the first $|D_{n-1}|+1$ factors of length f_n are the conjugates of $A_{n-2}A_{n-3}A_{n-1}$ (so they are also the conjugates of $A_n = A_{n-1}A_{n-2}A_{n-3}$), and the last one is just A_n. And other factors of length f_n are the factors of $\alpha^{-1}\overline{E_n}D_{n-1}E_n\alpha^{-1}$.
  Let us call the factors in Ω_n^1 the singular words of the first kind.
- $\Omega_n^2 = \{\text{factor of length } f_n \text{ of the word } \beta^{-1}E_{n+1}D_{n-2}\overline{E_{n+1}}\beta^{-1}\}$.
  (Here β is the first letter of E_{n+1}.)
  Likewise, by Proposition 2.6, the definition of F_n and Proposition 2.8, $A_{n-3}A_{n-1}A_{n-2}A_n = D_{n-2}\overline{F_n}D_{n-2}F_n = D_{n-2}E_{n+1}D_{n-2}\overline{E_{n+1}}$.
  And the factors in Ω_n^2 will be called the singular words of the second kind.

For example, we have

$$\Omega_2^0 = \{abac, baca, acab, caba\};$$
$$\Omega_2^1 = \{abab, baba\};$$
$$\Omega_2^2 = \{abaa, baab, aaba\}.$$

When $n = 1$, we put $\Omega_1^0 = \{ab, ba\}$, $\Omega_1^1 = \{aa\}$ and $\Omega_1^2 = \{ac, ca\}$.

Remark 2.13. Singular words were introduced in [20] for the Fibonacci case to study the factor structure of the Fibonacci sequence. In that case, given a length, there is only one singular word, but in our case, there are two classes of singular words, and thus the situation is much more complicated.

Let $s = s_0s_1s_2s_3\cdots$ be a sequence, and u be a factor of s. Suppose that u occurs in s at places p and q $(p < q)$; then the distance between the two occurrences is defined to be the difference $q - p$. If the distance between any two (distinct) occurrences of u is larger than or equal to (resp. strictly larger than) the length of u, we say that u is separated (resp. positively separated) in s.

For instance, by Corollary 2.7, both 0 and f_{n-1} are places where D_n occurs in the word $A_{n-1}A_nA_{n-1}$ (which is a factor of ξ). When $n \geq 2$, $|D_n| > f_{n-1}$ (this is implied by Proposition 2.9), and thus D_n is not separated in ξ.

Theorem 2.14. *For $n \geq 1$, any factor in Ω_n^1 or Ω_n^2 is positively separated.*

Proof. We consider the factor in Ω_n^1 first. When $n = 1$, $\Omega_1^1 = \{aa\}$, and aa is positively separated (since aaa is not a factor of ξ).

Let w be a word in Ω_n^1 $(n \geq 2)$. By definition, w is a factor of length f_n of the word $\alpha^{-1}\overline{E_n}D_{n-1}E_n\alpha^{-1}$. Comparing the lengths, we have that D_{n-1} occurs in w. By Theorem 2.11, the set of places where D_{n-1} occurs in ξ is $\mathcal{O}_{n-1}$, which can be regarded as the set of places of formal occurrences of the word A_{n-2} in the following factorization:

$$\xi = A_{n-2}B_{n-2}A_{n-2}C_{n-2}A_{n-2}B_{n-2}A_{n-2}\cdots. \tag{$*$}$$

Firstly, in the word w, the letter just before D_{n-1} is the last letter of $\overline{E_n}$ (by Proposition 2.10) which is the last letter of B_{n-2}, but not the last letter of either A_{n-2} or C_{n-2}; secondly, in w the letter just after D_{n-1} is the first letter of E_n, and a direct calculation using Corollary 2.7 gives

$$A_{n-2}B_{n-2}A_{n-2} = D_{n-1}\overline{E_{n-1}};$$
$$A_{n-2}C_{n-2}A_{n-2} = D_{n-1}E_{n-2};$$
$$A_{n-2}A_{n-2}C_{n-2} = D_{n-1}\overline{E_{n-2}},$$

and amongst the words $\overline{E_{n-1}}$, E_{n-2} and $\overline{E_{n-2}}$, only $\overline{E_{n-1}}$ has the same first letter as E_n.

Thus each occurrence of w in ξ gives a formal occurrence of $B_{n-2}A_{n-2}B_{n-2}$ in the factorization $(*)$. Consider another factorization

$$\xi = \sigma^n(\xi) = A_nB_nA_nC_nA_nB_nA_n\cdots. \tag{$\circledast$}$$

There is a one-to-one correspondence between the places where there is a formal occurrence of $B_{n-2}A_{n-2}B_{n-2}$ in $(*)$ and the places where there is a formal occurrence of C_n in $(\circledast)$.

Now in $(\circledast)$, the factor occurring just before C_n is A_n, and the factor occurring just after C_n is also A_n. And $A_nC_nA_n = A_{n-1}A_{n-2}A_{n-3}A_{n-1}A_n = A_{n-1}D_{n-1}\overline{E_n}D_{n-1}E_n$ (see the proof of Proposition 2.12 and the remarks after the definition of Ω_n^1); thus an occurrence of w appears.

Therefore there is a one-to-one correspondence between the places where w occurs in ξ and the places where there is a formal occurrence of the word C_n in the above factorization $(\circledast)$.

Now we consider the distances between two adjacent formal occurrences of C_n in $(\circledast)$: Writing

$$\xi = u_0u_1u_2u_3\cdots, \quad u_i \in \{a, b, c\},$$

we have

$$\{u_iu_{i+1}\cdots u_{j-1};\ j > i \geq 0, u_i = u_j = c \text{ and } u_k \neq c \text{ for } i < k < j\}$$
$$= \{cabaaba, cababa, caba\}.$$

Thus, in $(\circledast)$, the distance between two adjacent formal occurrences of C_n is one of the three numbers $|\sigma^n(cabaaba)| = f_{n+3}$, $|\sigma^n(cababa)| = f_{n+2} + f_{n+1}$ and $|\sigma^n(caba)| = f_{n+2}$.

Now, due to the one-to-one correspondence, the distance between two adjacent occurrences of w in ξ is also one of the three numbers f_{n+3}, $f_{n+2} + f_{n+1}$ and f_{n+2}. Each number is strictly larger than the length f_n of w; therefore w is positively separated in ξ.

Likewise, for a word v in Ω_n^2, there is a one-to-one correspondence between the places where v occurs in ξ and the places where there is a formal occurrence of the word B_n in the above factorization (⁎). Thus the distance between two adjacent occurrences of v in ξ is one of the three numbers f_{n+2}, $f_{n+1} + f_n$ and f_{n+1}, and v is positively separated in ξ. □

Remark 2.15. As we have seen in the above proof,

1. for a word w in Ω_n^1, the distances between two adjacent occurrences of w in ξ are in the set $\{f_{n+2}, f_{n+2} + f_{n+1}, f_{n+3}\}$; moreover, among any two adjacent distances, at least one is f_{n+3} (this is because, for three consecutive places i, j, m where c occurs in ξ, at least one word of $u_i u_{i+1} \cdots u_{j-1}$ and $u_j u_{j+1} \cdots u_{m-1}$ is *cabaaba*);

2. likewise, for the word w in Ω_n^2, the distances between two adjacent occurrences of w in ξ are in the set $\{f_{n+1}, f_{n+1} + f_n, f_{n+2}\}$; moreover, among any two adjacent distances, at least one is f_{n+2}.

See also Remark 3.3 for an alternative proof of certain special cases.

Now we can say more on the sets Ω_n^i $(i = 0, 1, 2)$.

Theorem 2.16. *For $n \geq 1$, the union of the sets Ω_n^i $(i = 0, 1, 2)$ is the set of factors of length f_n, and the union is disjoint. $\#\Omega_n^0 = f_n$, $\#\Omega_n^1 = |E_n| - 1$ and $\#\Omega_n^2 = |E_{n+1}| - 1$, where # denotes the cardinality of a finite set. Moreover, each Ω_n^i $(i = 0, 1, 2)$ is closed under the operation of taking the mirror, i.e. $\overline{\Omega_n^i} := \{\overline{w};\ w \in \Omega_n^i\} = \Omega_n^i$.*

Proof. We have already shown that the set of factors of length f_n is the union of Ω_n^0, Ω_n^1 and Ω_n^2. Also the cases with $n = 1, 2$ can be checked directly. So in the following we assume that $n \geq 3$.

First we study the properties of the sets Ω_n^0, Ω_n^1 and Ω_n^2 one by one.

$\underline{\Omega_n^0}$: 1. The conjugates of A_n are different from each other, and thus $\#\Omega_n^0 = f_n$.

Indeed, the Parikh vector of A_n is $P(A_n) = (|A_n|_a, |A_n|_b, |A_n|_c) = (f_{n-1}, f_{n-2}, f_{n-3})$. An easy induction shows that $\gcd(f_{n-1}, f_{n-2}, f_{n-3}) = 1$, and this implies that A_n is primitive (otherwise, if $A_n = w^m$, then $\gcd(P(A_n)) \geq m$). Thus these conjugates are distinct and $\#\Omega_n^0 = f_n$.

2. Ω_n^0 is closed under the operation of taking the mirror.

Since Ω_n^0 consists of all the conjugates of A_n, we only need show that the mirror word $\overline{A_n}$ of A_n is a conjugate of itself.

By the definition of E_n, $A_n = D_{n-1} E_n$, thus $\overline{A_n} = \overline{E_n}\,\overline{D_{n-1}} = \overline{E_n} D_{n-1}$ (the last equality is due to Proposition 2.2). On the other hand, by Proposition 2.6, $A_{n-2} A_{n-3} A_{n-1} = D_{n-1} \overline{E_n}$. Since $A_{n-2} A_{n-3} A_{n-1}$ is a conjugate of $A_n = A_{n-1} A_{n-2} A_{n-3}$, $\overline{A_n}$ is also a conjugate of A_n.

$\underline{\Omega_n^1}$: 1. There is just one occurrence of D_{n-1} in the word $A_n \alpha^{-1}$, where α is the last letter of A_n.

$A_n \alpha^{-1}$ is a prefix of ξ; then, by Theorem 2.11, the possible place where D_{n-1} occurs in $A_n \alpha^{-1}$ is either 0 or f_{n-1}. D_{n-1} occurs at place 0, but for the second place f_{n-1}, since $f_{n-1} + |D_n - 1| = f_{n-1} + \frac{1}{2}(f_n + f_{n-2} - 3) > f_n - 1$ for $n \geq 3$ (Proposition 2.9 is applied in the equality), $A_n \alpha^{-1}$ is not too long to contain the second D_{n-1}.

2. There is just one occurrence of D_{n-1} in the word $\alpha^{-1} \overline{E_n} D_{n-1} E_n \alpha^{-1}$.

By the definition of E_n, $D_{n-1} E_n \alpha^{-1} = A_n \alpha^{-1}$, and then $\alpha^{-1} \overline{E_n} D_{n-1} = \alpha^{-1} \overline{A_n} = \overline{A_n \alpha^{-1}}$. By 1, D_{n-1} occurs in $A_n \alpha^{-1}$ just once (as a prefix). Since D_n is a palindrome, D_n occurs in $\overline{A_n \alpha^{-1}}$ just once as a suffix. So there is just one occurrence of D_{n-1} in $\alpha^{-1} \overline{E_n} D_{n-1} E_n \alpha^{-1}$.

3. $\#\Omega_n^1 = |E_n| - 1$.

As stated in the proof of Theorem 2.14, for $w \in \Omega_n^1$, D_{n-1} occurs in w, and then by 2, D_{n-1} occurs in w just once. Therefore the factors of $\alpha^{-1}\overline{E_n}D_{n-1}E_n\alpha^{-1}$ occurring in different places are distinct, and thus $\#\Omega_n^1 = |E_n| - 1$.

4. Ω_n^1 is closed under the operation of taking the mirror.

This is just because $\alpha^{-1}\overline{E_n}D_{n-1}E_n\alpha^{-1}$ is a palindrome.

Ω_n^2: 1. D_{n-2} occurs in $\beta^{-1}E_{n+1}D_{n-2}\overline{E_{n+1}}\beta^{-1}$ just twice, at the places $|E_{n+1}| - |A_{n-2}| - 1$ and $|E_{n+1}| - 1$.

By arguing as above, D_{n-1} occurs in $D_{n-2}\overline{E_{n+1}}\beta^{-1} = A_n\beta^{-1}$ just once as a prefix. Taking the mirror words, D_{n-1} occurs in $\beta^{-1}E_{n+1}D_{n-2}$ just once as a suffix. Thus there are just two occurrences of D_{n-2} in $\beta^{-1}E_{n+1}D_{n-2}\overline{E_{n+1}}\beta^{-1}$, at places $|E_{n+1}| - |A_{n-2}| - 1$ and $|E_{n+1}| - 1$.

2. $\#\Omega_n^2 = |E_{n+1}| - 1$.

In $\beta^{-1}E_{n+1}D_{n-2}\overline{E_{n+1}}\beta^{-1}$, the factor occurring just before the first D_{n-1}, i.e. D_{n-1} occurring at place $|E_{n+1}| - |A_{n-2}| - 1$ (resp. the second D_{n-1}, i.e. D_{n-1} occurring at place $|E_{n+1}| - 1$), is $\overline{E_{n+1}}\beta^{-1}$ (resp. $E_n\beta^{-1}$). By Proposition 2.10, the first letters of $\overline{E_{n+1}}\beta^{-1}$ and $E_n\beta^{-1}$ are different.

Now the class Ω_n^2 is divided into three subclasses, namely: the words containing only the first D_{n-1} as a factor; the words containing both D_{n-1}'s; and the words containing only the second D_{n-1}. Considering the places where D_{n-1} occurs, the words in the same subclass are distinct; considering the number of occurrences of D_{n-1}, the second subclass is disjoint with the others; considering the letter occurring just after D_{n-1}, the first subclass is disjoint from the third one.

Therefore, the factors of $\beta^{-1}E_{n+1}D_{n-2}\overline{E_{n+1}}\beta^{-1}$ occurring in different places are distinct, and thus $\#\Omega_n^2 = |E_{n+1}| - 1$.

3. Ω_n^2 is closed under the operation of taking the mirror.

This is just because $\beta^{-1}E_{n+1}D_{n-2}\overline{E_{n+1}}\beta^{-1}$ is a palindrome.

Finally, by considering the letter occurring just before or just after D_{n-1}, we show that the sets Ω_n^0, Ω_n^1 and Ω_n^2 are pairwise disjoint. $\square$

By the above theorem,

$$\begin{aligned}\#\Omega_{f_n} &= |\Omega_n^0| + |\Omega_n^1| + |\Omega_n^2| \\ &= f_n + \frac{1}{2}(f_n - f_{n-2} + 3) - 1 + \frac{1}{2}(f_{n+1} - f_{n-1} + 3) - 1 \\ &= 2f_n + 1\end{aligned}$$

and this checks the complexity function of the Tribonacci sequence for the length f_n.

Now we come to determining the palindrome factor of length f_n. We have that each of Ω_n^0, Ω_n^1 and Ω_n^2 contains at most one palindrome. More precisely:

1. Ω_n^0 contains one palindrome $\Leftrightarrow$ f_n is odd $\Leftrightarrow$ $n \equiv 0, 3 \pmod 4$.

In fact, if f_n is even and $w = u\overline{u} \in \Omega_n^0$, then $P(w) = 2P(u)$ which contradicts the facts that $P(w) = P(A_n) = (f_{n-1}, f_{n-2}, f_{n-3})$ and $\gcd(f_{n-1}, f_{n-2}, f_{n-3}) = 1$. On the other hand, if f_n is odd, since Ω_n^0 is closed under taking the mirror, Ω_n^0 contains one palindrome.

We will denote the palindrome factor, if any, by P_n^0.

2. Ω_n^1 contains one palindrome $\Leftrightarrow$ $|E_n|$ is even $\Leftrightarrow$ $n \equiv 0, 1, 3 \pmod 4$.

Recall that the factors of $\alpha^{-1}\overline{E_n}D_{n-1}E_n\alpha^{-1}$ occurring in different places are distinct, and $\#\Omega_n^1 = |E_n| - 1$.

Since $\alpha^{-1}\overline{E_n}D_{n-1}E_n\alpha^{-1}$ is a palindrome, the factor of length f_n occurring at place m is the mirror word of the one occurring at place $|E_n| - 2 - m$. Thus Ω_n^1 contains at most one palindrome, and it contains a palindrome only if the total number $|E_n|-1$ of words in Ω_n^1 is odd; on the other hand, when $|E_n|$ is even, denoting by L_n the prefix of length $\frac{1}{2}|E_n|$ of E_n, then $P_n^1 = \overline{L_n}D_{n-1}L_n$ is a palindrome in Ω_n^1.

We will denote the palindrome factor, if any, by P_n^1.

3. Ω_n^2 contains one palindrome $\Leftrightarrow$ $|E_{n+1}|$ is even $\Leftrightarrow$ $n \equiv 0, 2, 3 \pmod 4$.

We will denote the palindrome factor, if any, by P_n^2.

In fact, this checks the palindrome complexity for the length f_n.

3. Some factorizations of the Tribonacci sequence

In this section, we will give some factorizations of the Tribonacci sequence.

Theorem 3.1.

$$\xi = \prod_0^\infty \overline{A_i} = \overline{A_0}\ \overline{A_1}\ \overline{A_2} \cdots .$$

Proof. By Proposition 2.2 and the definition of D_n, $D_n = \overline{A_0}\ \overline{A_1}\ \overline{A_2} \cdots \overline{A_{n-1}}$. By Theorem 2.5, for any n, D_n is a prefix of A_{n+1}, and A_{n+1} is a prefix of ξ, and thus D_n is also a prefix of ξ, and the theorem follows. $\square$

If $n \not\equiv 2 \pmod 4$, there is a palindrome singular factor P_n^1 of the first kind. We will give a factorization of ξ according to P_n^1.

By Theorem 2.14, P_n^1 is positively separated, then we can rewrite the Tribonacci sequence:

$$\xi = Z_0\ P_n^1\ Z_1\ P_n^1\ Z_2\ P_n^1\ Z_3\ P_n^1\ Z_4\ P_n^1\ Z_5\ P_n^1\ Z_6 \cdots$$

where P_n^1 is not a factor of any Z_k.

Since $|E_n|$ is even, we can write $E_n = L_n R_n$ with $|L_n| = |R_n| = \frac{1}{2}|E_n|$. Then $P_n^1 = \overline{L_n} D_{n-1} L_n$. Put

$$\begin{aligned}
\Xi &= A_n B_n A_{n-1} D_{n-1} \overline{R_n};\\
\Delta &= R_n B_n A_n A_n B_n A_{n-1} D_{n-1} \overline{R_n} = R_n B_n A_n \Xi;\\
\Theta &= R_n B_n A_n B_n A_{n-1} D_{n-1} \overline{R_n} = R_n B_n \Xi;\\
\Lambda &= R_n B_n A_{n-1} D_{n-1} \overline{R_n} = R_n \Xi.
\end{aligned}$$

Theorem 3.2. *With the notation above, we have*

1. *Δ, Θ and Λ are palindromes;*

2. *$Z_0 = \Xi$; for any $n \geq 1$, $Z_i \in \{\Delta, \Theta, \Lambda\}$; moreover the sequence $Z = Z_1 Z_2 Z_3 \cdots$ is the Tribonacci sequence over the alphabet $\{\Delta, \Theta, \Lambda\}$.*

Proof. 1. Using the definitions of words E_n and F_n, and results 2.6, 2.7 and 2.8 we have

$$\begin{aligned}
\Delta &= R_n B_n A_n A_n B_n A_{n-1} D_{n-1} \overline{R_n}\\
&= R_n A_{n-1} A_{n-2} A_n A_{n-1} A_{n-2} A_{n-3} A_{n-1} A_{n-2} A_{n-1} D_{n-1} \overline{R_n}\\
&= R_n D_{n-1} \overline{E_{n-1}} D_{n-2} \overline{E_{n+1}} D_{n-2} E_{n-1} D_{n-4} \overline{E_{n-1}} D_{n-2} E_{n+1} D_{n-2} E_{n-1} D_{n-1} \overline{R_n};\\
\Theta &= R_n B_n A_n B_n A_{n-1} D_{n-1} \overline{R_n}\\
&= R_n A_{n-1} A_{n-2} A_n A_{n-1} A_{n-2} A_{n-1} D_{n-1} \overline{R_n}
\end{aligned}$$

$$= R_n D_{n-1} \overline{E_{n-1}} D_{n-1} E_{n-1} D_{n-3} \overline{E_n} D_{n-1} E_{n-1} D_{n-1} \overline{R_n};$$

$$\begin{aligned} \Lambda &= R_n B_n A_{n-1} D_{n-1} \overline{R_n} \\ &= R_n A_{n-1} A_{n-2} A_{n-1} D_{n-1} \overline{R_n} \\ &= R_n D_{n-1} \overline{E_{n-1}} D_{n-2} E_{n-1} D_{n-1} \overline{R_n}. \end{aligned}$$

2. Put $w = P_n^1$ in the proof of Theorem 2.14. We have shown that there is a one-to-one correspondence between the places where P_n^1 occurs in ξ and the places where there is a formal occurrence of the word C_n in the following factorization:

$$\xi = \sigma^n(\xi) = A_n B_n A_n C_n A_n B_n A_n \cdots. \tag{$\circledast$}$$

Then $Z_i \in \{\Delta, \Theta, \Lambda\}$ for $n \geq 1$. On the other hand, since

$$\begin{aligned} A_{n+3} &= \Xi\ P_n^1\ \Delta\ \Xi^{-1}; \\ B_{n+3} &= \Xi\ P_n^1\ \Theta\ \Xi^{-1}; \\ C_{n+3} &= \Xi\ P_n^1\ \Lambda\ \Xi^{-1}, \end{aligned} \tag{$\diamond$}$$

$Z = Z_1 Z_2 Z_3 \cdots$ is the Tribonacci sequence over the alphabet $\{\Delta, \Theta, \Lambda\}$. □

The following example illustrates the factorization of ξ according to $P_1^1 = aa$:
$abacab(aa)bacababacab(aa)abcabacab(aa)bacababacab(aa)bacab(aa)$
$bacababacab(aa)bacababacab(aa)abcabacab(aa)bacababacab\cdots$.

If $n \not\equiv 1 (\mathrm{mod}\ 4)$, there is a palindrome singular factor P_n^2 of the second kind. We can also get a factorization of ξ according to P_n^2. Moreover if we consider the factor of length $|B_n|$ of ξ, we can define the singular words in the same way and get similar factorizations.

Remark 3.3. From ($\diamond$), $|P_n^1| + |\Delta| = |A_{n+3}| = f_{n+3}$, $|P_n^1| + |\Theta| = f_{n+2} + f_{n+1}$ and $|P_n^1| + |\Lambda| = f_{n+2}$, and thus the distances between two adjacent occurrences of P_n^1 are in the set $\{f_{n+3}, f_{n+2} + f_{n+1}, f_{n+2}\}$. This checks again Remark 2.15.1 for P_n^1. The situation for the word P_n^2 is similar.

Now we give the Lyndon factorization of ξ. We totally order S by $a < b < c$ and extend this order to the set S^* of all words lexicographically. Lyndon words are defined as primitive words, which are minimal in the class of all their conjugates. The reader is referred to [10] for more information and the proof of the following theorem.

Theorem 3.4 (*Lyndon*). *Any non-empty word is a unique product of non-increasing Lyndon words.*

In [17] the authors defined Lyndon sequences as those sequences which have infinitely many prefixes being Lyndon words. And they proved the following generalization of the Lyndon theorem.

Theorem 3.5. *Any sequence s can be uniquely factorized in one of the following forms:*
(1) $s = l_0 l_1 l_2 \cdots$, with $l_0 \geq l_1 \geq l_2 \geq \cdots$, are Lyndon words;
(2) $s = l_0 l_1 l_2 \cdots l_k l_{k+1}$, with $l_0 \geq l_1 \geq l_2 \geq \cdots \geq l_k$, are Lyndon words, and $l_{k+1} < l_k$ is a Lyndon sequence.

The following theorem gives the Lyndon factorization of ξ.

Theorem 3.6. *Define the substitution $\varphi = (aabacab, abacab, acab)$;*

$$\xi = \prod_{i=0}^{\infty} ([\varphi^i(abac)]\,[\varphi^i(ab)])$$

is the Lyndon factorization of the Tribonacci sequence. That is

1. *For any n, both $\varphi^n(abac)$ and $\varphi^n(ab)$ are Lyndon words and*

$$abac > ab > \varphi(abac) > \varphi(ab) > \varphi^2(abac) > \varphi^2(ab) > \cdots.$$

2. $\xi = [abac][ab][\varphi(abac)][\varphi(ab)][\varphi^2(abac)][\varphi^2(ab)]\cdots$.

Proof. 1. It is easy to check.

2. Define $\eta = (abac)(ab)(\varphi(abac))(\varphi(ab))(\varphi^2(abac))(\varphi^2(ab))\cdots$; we need show $\eta = \xi$. Notice that $\eta = (abacab)\varphi(\eta)$ and $\sigma^3 = \iota_{abacab}\,\varphi$; we have

$$\sigma^3(\eta) = \iota_{abacab}\,\varphi(\eta) = \iota_{abacab}((abacab)^{-1}\eta) = \eta,$$

and thus η is a fixed point of σ and $\eta = \xi$. □

The reader is referred to [7,9,12,14–16] for similar results on the Thue–Morse sequence, Sturmian sequences, and generalizations of those sequences.

4. Applications

In this section, we discuss the combinatorial properties of the factors of the Tribonacci sequence, such as the power of the factor, and the overlap property. As we will see, the positively separated property of the singular words plays an important role in these studies.

4.1. Power property of the factors

Theorem 4.1. *For $n \geq 3$, the maximal common factor of both the Tribonacci sequence and the sequence $A_n^\infty := A_nA_nA_n\cdots$ is $A_nA_nA_nD_{n-3}$.*

Proof. In fact, we need to find the longest factor of ξ which has no singular factor of length f_n. We will denote this word as M.

Recall (from the proof of Theorem 2.14) that, for a singular word w of the second kind, there is a one-to-one correspondence between the places where w occurs in ξ and the places where there is a formal occurrence of the word B_n in the following factorization:

$$\xi = \sigma^n(\xi) = A_nB_nA_nC_nA_nB_nA_n\cdots. \qquad (\divideontimes)$$

Since M has no singular factor of the second kind, we know that M is a factor of one of the following words:

1. $A_{n-2}B_{n-2}A_{n-2}C_{n-2}A_{n-2}B_{n-2}A_{n-2}B_{n-2}A_{n-2}C_{n-2}A_{n-2}B_{n-2}A_{n-2}D_{n-2}$;
2. $A_{n-2}B_{n-2}A_{n-2}C_{n-2}A_{n-2}B_{n-2}A_{n-2}C_{n-2}A_{n-2}B_{n-2}A_{n-2}D_{n-2}$;
3. $A_{n-2}B_{n-2}A_{n-2}C_{n-2}A_{n-2}B_{n-2}A_{n-2}D_{n-2}$.

But the word 1 equals $A_{n-1}D_{n-1}\overline{E_n}D_{n-1}E_nA_{n-1}A_{n-2}D_{n-2}$ and it contains a singular factor of the first kind. Then the word 2 which equals $A_nA_nA_nD_{n-3}$ is the desired word. □

The next theorem considers the power property of the factor of length f_n. Recall that if $A_n = u_0 u_1 \cdots u_{f_n-1}$ with $u_i \in \{a, b, c\}$, then for $k = 0, 1, \ldots f_n - 1$, $C_k(A_n) = u_k \cdots u_{f_n-1} u_0 \cdots u_{k-1}$.

Theorem 4.2. *Let w be a factor of length f_n of the Tribonacci sequence.*

1. *If $w \in \Omega_n^0$, that is, w is conjugate to A_n, then $w^2 \prec \xi$.*
For $n = 0, 1, 2$, $w^3 \not\prec \xi$.
For $n \geq 3$, if $0 \leq k \leq |D_{n-3}|$, then $[C_k(A_n)]^3 \prec \xi$; if $|D_{n-3}| + 1 \leq k \leq f_n - 1$, then $[C_k(A_n)]^3 \not\prec \xi$.
2. *If $w \in \Omega_n^1 \cup \Omega_n^2$, that is, w is a singular word, then $w^2 \not\prec \xi$.*

Proof. 1. Let us consider the case when $n \geq 3$ and $|D_{n-3}| + 1 \leq k \leq f_n - 1$. Since

$$A_n = A_{n-2} B_{n-2} A_{n-2} A_{n-3} = D_{n-3} E_{n-2} B_{n-2} D_{n-2} \overline{E_{n-2}},$$

and $|D_{n-3} E_{n-2}| = f_{n-2}$, $C_k(A_n)$ has a factor w from the set Ω_{n-2}^1.

Assume that $[C_k(A_n)]^3 \prec \xi$; then w occurs in $[C_k(A_n)]^3$ three times, and both of the distances between two adjacent occurrences of w are f_n and this contradicts Remark 2.15.

The other cases are either trivial or a corollary of Theorem 4.1.

2. This is by the positively separated property of the singular word (Theorem 2.14). □

Now we consider the factors of other lengths, and we need the following Fine and Wilf theorem (See [10] for example):

Theorem 4.3. *Let $x, y \in S^*$, $n = |x|$, $m = |y|$, $d = \gcd(n, m)$. If two powers x^p and y^q of x and y have a common prefix of length at least equal to $n + m - d$, then x and y are powers of the same word.*

Theorem 4.4. *Let w be a factor of ξ, $f_n < |w| < f_{n+1}$; then $w^3 \not\prec \xi$.*

Proof. If w has a singular factor of length f_n, then from Remark 2.15 and the fact $|w| < f_{n+1}$, $w^2 \not\prec \xi$.

Otherwise any factor of length f_n of w is conjugate to A_n; then from $|w| < f_{n+1} < 2f_n$, w is a factor of A_n^3.

Assume that $w^3 \prec \xi$; then w^3 has no singular factor of length f_n. In fact, if w^3 has a factor $u \in \Omega_n^1 \cup \Omega_n^2$, then u is also a factor of w^2 since $|w| > |u|$ and the distance between two occurrences of u in w^3 is $|w| < f_{n+1}$, which contradicts Remark 2.15.

Thus any factor of length f_n of w^3 is conjugate to A_n, and w^3 is a factor of A_n^9. Then there is a conjugate word v of A_n such that w^3 is the prefix of v^9. Apply Theorem 4.3 with $x = w$ and $y = v$; we know that w and v are powers of the same word. But as a conjugate word of the primitive word A_n, v is primitive, and hence w is a power of v. This is a contradiction to the fact $|v| = f_n < |w| < f_{n+1} < 2f_n = 2|v|$. □

Now we study the highest order of the repetitions in the Tribonacci sequence.

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

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

and in this case we write that $z = x^r$.

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

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

Theorem 4.5. *We have*

$$\mathbb{FI}(\xi) = 3 + \frac{1}{2}(\theta^2 + \theta^4) \approx 3.19148,$$

where $\theta \approx 0.54368$ *is the unique real root of the equation* $\theta^3 + \theta^2 + \theta = 1$.

Proof. For any factor $w \prec \xi$, we define the index of w in ξ by $\operatorname{ind}(w) = \sup\{r \in \mathbb{Q} : w^r \prec \xi\}$, and the definition yields $\mathbb{FI}(\xi) = \sup\{\operatorname{ind}(w) : w \prec \xi\}$.

By Theorem 4.4, if $f_n < |w| < f_{n+1}$, $\operatorname{ind}(w) < 3$.

By Theorem 4.2, if w is a singular word, $\operatorname{ind}(w) < 2$.

By Theorem 4.1, $\operatorname{ind}(A_n) = 3 + \frac{|D_{n-3}|}{|A_n|}$ and $\operatorname{ind}(C_k(A_n)) \le \operatorname{ind}(A_n)$.

Hence we have

$$\mathbb{FI}(\xi) = \sup\left\{3 + \frac{|D_{n-3}|}{|A_n|}\right\} = 3 + \sup\left\{\frac{f_{n-2} + f_{n-4} + 3}{2f_n}\right\},$$

and the theorem follows. □

One can also obtain the result using Theorem 5.2 of [8].

4.2. Overlap property of factors

Suppose $u \prec s \in S^\omega$. If there exist non-empty words x, y and z such that $u = xy = yz$ and $u^*(y) := uz = xyz \prec s$, then we say that the word u has overlap, with the overlap factor y (or overlap length $|y|$); the word $u^*(y)$ is called the overlap of u with the overlap factor y.

Recall that the set of places where the word D_{n-1} occurs in ξ is $\mathcal{O}_{n-1}$, and since $A_n = D_{n-1}E_n$, the set of the occurrences of A_n is a subset of $\mathcal{O}_{n-1}$.

Theorem 4.6. *Let w be a factor of length f_n of the Tribonacci sequence.*

1. *If $w \in \Omega_n^0$:*

For $n = 0, 1, 2$, w has no overlap.

For $n \ge 3$, if $0 \le k \le |D_{n-3}|$, $C_k(A_n)$ has overlap with overlap length f_{n-3} and $f_{n-3}+f_{n-2}$; if $|D_{n-3}| + 1 \le k \le f_n - 1$, $C_k(A_n)$ has no overlap.

2. *If $w \in \Omega_n^1 \cup \Omega_n^2$, w has no overlap.*

Proof. 1. The results for $n = 0, 1, 2$ can be easily checked.

By Theorem 2.5, D_{n-1} is a prefix of A_n. Hence by Theorem 2.11, the set of places where A_n occurs in ξ is a subset of $\mathcal{O}_{n-1}$, and thus the possible overlaps of A_n occur in the following words:

1. $A_{n-2}B_{n-2}A_{n-2}A_{n-2}B_{n-2}A_{n-2}C_{n-2}A_{n-2}$;
2. $A_{n-2}B_{n-2}A_{n-2}C_{n-2}A_{n-2}B_{n-2}A_{n-2}$.

For the word (1), we have

$$\begin{aligned}
&A_{n-2}B_{n-2}A_{n-2}A_{n-2}B_{n-2}A_{n-2}C_{n-2}A_{n-2} \\
&\quad = A_n D_{n-3}\alpha \cdots \\
&\quad = (A_{n-1}A_{n-2})A_n D_{n-3}\beta \cdots
\end{aligned}$$

where α and β are two distinct letters and $\cdots$ denotes the rest of the representations.

For the word (2), we have that

$$A_{n-2}B_{n-2}A_{n-2}C_{n-2}A_{n-2}B_{n-2}A_{n-2} = A_n D_{n-3}\alpha \cdots = (A_{n-1})A_n D_{n-3}\beta \cdots.$$

Then we know that for $0 \le k \le |D_{n-3}|$, $C_k(A_n)$ has overlap with overlap length f_{n-3} and $f_{n-3} + f_{n-2}$.

If $|D_{n-3}| + 1 \le k \le f_n - 1$, as shown in the proof of Theorem 4.2.1, $C_k(A_n)$ has a factor w which is in Ω^1_{n-2}. If $C_k(A_n)$ has an overlap y, then the word w will occur in $C_k(A_n)^*(y)$ twice and the distance between the two occurrences is less than $|C_k(A_n)| = f_n$; this is a contradiction to Remark 2.15.

2. By the positively separated property of the singular word. □

Theorem 4.7. *Let w be a factor of ξ, $f_n < |w| < f_{n+1}$; then w has overlap if and only if w is a factor of $A_n^2 D_{n-3}$.*

Proof. If $w \prec A_n^2 D_{n-3}$, then by Theorem 4.1 and $|w| > f_n$, w has overlap.

Suppose that $w \not\prec A_n^2 D_{n-3}$ and w has overlap.

If w has a singular factor of length f_n, then from Remark 2.15 and the fact $|w| < f_{n+1}$, w has no overlap. Thus any factor of length f_n of w is conjugate to A_n, and from $|w| < f_{n+1} < 2f_n$, w is a factor of A_n^3.

Since $w \not\prec A_n^2 D_{n-3}$ and $w \prec A_n^3$, then

$$w = w_1 A_n D_{n-3} w_2 = w_1 D_n w_2,$$

where w_1, w_2 are non-empty words and $w_1 \rhd A_n, w_2 \lhd E_{n-2}B_{n-2}B_{n-1}$ (note that $A_n = D_{n-3}E_{n-2}B_{n-2}B_{n-1}$).

By Theorem 2.11, the set of places where D_n occurs is just the set of places where there is a formal occurrence of A_{n-1} in

$$\xi = A_{n-1}A_{n-2}A_{n-3}A_{n-1}A_{n-2}A_{n-1}A_{n-2}A_{n-3}A_{n-1}\cdots. \qquad (\heartsuit)$$

Moreover, since the factor D_n occurs just after w_1 in the word w, and since (by Proposition 2.10) only A_{n-3} amongst the words A_{n-1}, A_{n-2} and A_{n-3} has the same last letter as w_1, the occurrence of w will give a formal occurrence of the word $A_{n-3}A_{n-1}$ in $(\heartsuit)$. Likewise, from the fact that, in the word w, the factor occurring just after D_n is w_2, we know that the occurrence of w will correspond to the occurrence of the word $A_{n-3}A_{n-1}A_{n-2}A_{n-3}$. Then the distance of the two adjacent occurrences of w's in ξ is larger than the length of w, and it is impossible that w has overlap. □

Acknowledgement

The authors thank the anonymous referees for a lot of useful suggestions.

References

[1] J.P. Allouche, M. Baake, J. Cassaigne, D. Damanik, Palindrome complexity, Theoret. Comput. Sci. 292 (2003) 9–31.

[2] P. Arnoux, G. Rauzy, Représentation Géométriques de suites de complexité $2n + 1$, Bull. Soc. Math. France 119 (1991) 199–215.

[3] W.T. Cao, Z.Y. Wen, Some properties of the factors of Sturmian sequences, Theoret. Comput. Sci. 304 (2003) 365–385.

[4] N. Chekhova, P. Hubert, M. Ali, Propriétés combinatoires, ergodiques et arithmétiques de la substitution de Tribonacci, J. Théor. Nombres Bordeaux 13 (2) (2001) 371–394.

[5] D. Damanik, D. Lenz, Powers in Sturmian sequences, European J. Combin. 24 (4) (2003) 377–390.

[6] N.P. Fogg, Substitutions in Dynamics, Arithmetics, and Combinatorics, in: V. Berthé, S. Ferenczi, C. Mauduit, A. Siegel (Eds.), in: Lecture Notes in Mathematics, vol. 1794, Springer, 2002.

[7] A. Ido, G. Melançon, Lyndon factorization of the Thue–Morse word and its relatives, Discrete Math. 1 (1997) 43–52.

[8] J. Justin, G. Pirillo, Episturmian words and episturmian morphism, Theoret. Comput. Sci. 276 (2002) 281–313.

[9] F. Levé, P. Séébold, Conjugation of standard morphisms and a generalization of singular words, Bull. Belg. Math. Soc. 10 (2003) 737–748.

[10] M. Lothaire, Combinatorics on Words, in: Encyclopedia of Mathematics and its Applications, vol. 17, Addison-Wesley, 1983.

[11] G. Melançon, Lyndon factorization of Sturmian words, in: D. Stanton, P. Leroux (Eds.), 8th Int. Conf. on Formal Power Series and Algebraic Combinatorics, FPSAC'96, Discrete Math. 210 (2000) 137–149 (special issue).

[12] G. Melançon, Lyndon words and singular factors of Sturmian words, Theoret. Comput. Sci. 218 (1999) 41–59.

[13] G. Rauzy, Nombres algébriques et substitutions, Bull. Soc. Math. France 110 (1982) 147–178.

[14] G. Richomme, Lyndon morphisms, Bull. Belg. Math. Soc. 10 (2003) 761–785.

[15] G. Richomme, Conjugacy of morphisms and Lyndon decomposition of standard Sturmian words, in: S. Brlek, C. Reutenauer (Eds.), 5th Int. Conf. on Words, vol. 36, Publications du LaCIM, 2005, pp. 431–451.

[16] P. Séébold, Lyndon factorization of the Prouhet words, Theoret. Comput. Sci. 307 (2003) 179–197.

[17] R. Siromoney, L. Mattew, V.R. Dare, K.G. Subramanian, Infinite Lyndon words, Inform. Process. Lett. 50 (1994) 101–104.

[18] V.F. Sirvent, Semigroups and the self-similar structure of the flipped Tribonacci substitution, Appl. Math. Lett. 12 (1999) 25–29.

[19] D. Vandeth, Sturmian words and words with a critical exponent, Theoret. Comput. Sci. 242 (2000) 283–300.

[20] Z.X. Wen, Z.Y. Wen, Some properties of the singular words of the Fibonacci word, European J. Combin. 15 (1994) 587–598.