\(\alpha\)-Words and factors of characteristic sequences

Wai-Fong Chuan

Department of Mathematics, Chung-yuan Christian University, Chung Li, Taiwan 320, Taiwan, ROC

Received 31 January 1995; revised 13 February 1996

Abstract

Let \(\alpha\) be an irrational number with \(0 < \alpha < 1\). Using the continued fraction expansion of \(\alpha\), the class of \(\alpha\)-words is introduced. It contains certain sequences of words that are known to relate to the characteristic sequence \(f(\alpha)\) of \(\alpha\). When \(\alpha = (\sqrt{5} - 1)/2\), \(\alpha\)-words are precisely the Fibonacci words. In this paper, the class of \(\alpha\)-words is shown to be a subset of factors of \(f(\alpha)\), which is closed under both conjugation and reversion. The canonical palindrome factorization of unbordered \(\alpha\)-words play an important role in the determination of factors of \(f(\alpha)\). It is proved that every unbordered \(\alpha\)-word \(w\) that we obtain determines a \((|w|+1)\times |w|\) matrix \(C\) of the form

\[
C = \begin{bmatrix}
\text{circ}(w) \\
y
\end{bmatrix}
\]

such that for every \(1 \leq k \leq |w|\), the rows of the upper left \((k+1)\times k\) submatrix are distinct factors of \(f(\alpha)\) of length \(k\). As a consequence of a well-known result, this actually gives all the factors of \(f(\alpha)\) of length \(k\).

1. Introduction

Let \(\alpha\) be an irrational number with \(0 < \alpha < 1\) and let \(\alpha = [0, a_1 + 1, a_2, \ldots]\) be its continued fraction. The characteristic sequence \(f(\alpha)\) of \(\alpha\) is the infinite binary sequence \(c_0c_1c_2c_3\ldots\) whose \(n\)th term is given by

\[
c_n = [(n+1)\alpha] - [n\alpha], \quad n \geq 1.
\]

The characteristic sequence of \(\alpha = (\sqrt{5} - 1)/2 = [0, 1, 1, \ldots]\) is also called the Fibonacci word [2, 32, 36] or golden sequence [6, 11, 12].

Define a sequence of finite binary strings as follows:

\[
x_0 = 0, \quad x_1 = 0^{a_1} 1,
\]

\[
x_n = x_{n-1}a_{n-2}, \quad n \geq 2.
\]

\(^1\)Part of this research was supported by National Science Council Grant NSC-84-2121-M-033-007 and was completed while the author was visiting Massachusetts Institute of Technology.
It is known that each $x_n$ is a prefix of $f(x)$, $n \geq 1$ (see, for example, [5, 22, 28, 35, 36, 39, 44]).

On the other hand, it is also known that

$$f(x) = v_1 v_2 v_3 \cdots,$$

where

$$v_0 = 0, \quad v_1 = 0^{x_1}, \quad v_n = v_{n-1}^{x_n} v_{n-2} v_{n-1}, \quad n \geq 2$$

[42]. This factorization is used by Hendel and Monteferrante in [23]. When $x = \sqrt{2} - 1 = [0, 2, 2, \ldots]$, it is not hard to see that

$$f(x) = u_0 u_1 u_2 \cdots,$$

where

$$u_0 = 0, \quad u_1 = 10, \quad u_n = u_{n-1} u_{n-2} u_{n-1}, \quad n \geq 2.$$

For $x = (\sqrt{5} - 1)/2$, $\{x_n\}$ and $\{v_n\}$ turn out to be sequences of Fibonacci words and $v_n$ equals the mirror image of $x_n$. They have been studied by many mathematicians. (Besides the references mentioned above, see, for example, [2, 3, 7–13, 15, 24, 25, 29–32, 34, 40, 41, 44]). The class of Fibonacci words, which is introduced by Chuan [7], contains these sequences and is shown to have many interesting combinatorial properties [7–12]. For example, all Fibonacci words, arranged in a certain order, can be embedded into the golden sequence in a very nice manner [9]; the sequence $\{w_n\}$ of Fibonacci words defined by

$$w_0 = 0, \quad w_1 = 1,$$

$$w_n = \begin{cases} w_{n-1} w_{n-2} & \text{if } n \text{ is odd}, \\ w_{n-2} w_{n-1} & \text{if } n \text{ is even}, \end{cases} \quad n \geq 2,$$

arises naturally in a 3-symbol problem considered by Anderson and Chuan [1].

These results motivate the definition of $x$-words. This new class of words will contain the sequences $\{x_n\}$, $\{v_n\}$, $\{u_n\}$, $\{w_n\}$ defined above when the pair of initial words are suitably chosen and it will reduce to the class of Fibonacci words when $x = (\sqrt{5} - 1)/2$.

In Section 2, we give the definition of $x$-words associated with a finite or infinite sequence of positive integers and we prove the conjugation theorem (Theorem 2.4) of $x$-words. The number of $x$-words of order $n$ derived from a pair of distinct letters is determined. In Section 3, we prove the primitivity (Theorem 3.4) and the canonical palindrome factorization of $x$-words (Theorem 3.5).

At each level, there are other words (besides those $x$-words we defined in Section 2) which have similar structures as $x$-words and which are also factors of $f(x)$. Therefore we enlarge the class of $x$-words to include these words. However, unlike the old $x$-words, these words are not used to produce any $x$-words of higher order. We discuss this enlarged class of $x$-words in Section 4.
Unbordered $\alpha$-words are important in determining factors of $f(\alpha)$. Properties of unbordered words that are needed in later discussion are established in Section 5. In particular, we determine explicitly two unbordered $\alpha$-words for each order (Theorem 5.4).

For irrational number $\alpha$, it is well-known that for each $k \geq 1$, $f(\alpha)$ has exactly $k + 1$ factors of length $k$. In Section 6, we identify an irrational number whose continued fraction expansion is $[0, a_1 + 1, a_2, \ldots]$ with an infinite sequence $\alpha = \{a_1, a_2, \ldots\}$ and consider $\alpha$-words derived from the initial pair of words $(1, 0)$. We show that factors of $\alpha$-words are factors of $f(\alpha)$ (Lemma 6.3) and that each $\alpha$-word contains all factors of $f(\alpha)$ of length $\leq |w|$ as cyclic factors except for one factor of $f(\alpha)$ of length $|w|$ (Theorem 6.5). We determine all factors of $f(\alpha)$ of any given length using unbordered $\alpha$-words (Theorems 6.4 and 6.6).

2. $\alpha$-Words

Let $A$ be an alphabet and let $A^*$ be the free monoid generated by $A$. For any word $w \in A^*$, let $|w|$ denote the length of $w$. Let $T$ (respectively, $R$) be the cyclic shift operator (respectively, reverse operator) defined by

$$
T(c_1c_2 \cdots c_m) = c_2 \cdots c_mc_1
$$

(respectively, $R(c_1c_2 \cdots c_m) = c_m \cdots c_2c_1$) where each $c_i$ is a letter in $A$, $m \geq 2$. Let $T^j(x) = T(T^{j-1}(x))$, where $j \geq 2$, $x \in A^*$. For any word $x$ and any integer $j$, $T^j(x)$ is called a conjugate of $x$. The conjugate class of a word $x \in A^*$ is the set of $y \in A^*$ such that $y$ is a conjugate of $x$.

Let $\alpha$ denote a finite or infinite sequence $a_1, a_2, \ldots$ of integers with $a_1 \geq 0$, $a_n \geq 1$, $n \geq 2$. Let $(u, v)$ be a pair of distinct initial words over an alphabet $A$. A word $w \in A^*$ is called an $\alpha$-word derived from $(u, v)$ if $w = u$ or $w = v$ or there are integers $n, r_1, \ldots, r_n$ with $1 \leq n \leq$ the number of terms of $\alpha$, $0 \leq r_i \leq a_i$, $1 \leq i \leq n$ and a sequence of words

$$
w_{-1}, w_0, \ldots, w_n
$$

such that $w_{-1} = u$, $w_0 = v$, $w_n = w$, $w_i = w_{i-1}^{r_i}w_{i-2}w_{i-1}$, $1 \leq i \leq n$. Let $r = (r_1, r_2, \ldots, r_n)$. Denote $w$ by $w_n(\alpha; u, v; r)$, with the understanding that $w$ depends only on the first $n$ terms of $\alpha$. If there is no ambiguity, we simply write $w_n(r)$ or $w_n(\alpha; r)$ for $w_n(\alpha; u, v; r)$. The positive integer $n$ is called the order of the $\alpha$-word $w_n(r)$.

In the application, $\alpha$ is an irrational number with $0 < \alpha < 1$ and with continued fraction expansion $\alpha = [0, a_1 + 1, a_2, \ldots]$. $A = \{0, 1\}$ and the pair of initial words are taken to be $(1, 0)$. In this case, we identify $\alpha$ with the sequence $\{a_1, a_2, \ldots\}$ and the words obtained are again called $\alpha$-words.

Plainly, when $\alpha = (\sqrt{5} - 1)/2 = [0, 1, 1, \ldots]$, $\alpha$-words derived from $(u, v)$ are precisely the Fibonacci words derived from $(v, u)$ in [7]. The sequences $\{x_n\}$, $\{v_n\}$, $\{u_n\}$, $\{w_n\}$ given in Section 1 are sequences of $\alpha$-words derived from $(1, 0)$. (Note: the sequence
\{u_n\} (respectively, \{w_n\}) is defined for \(z = \sqrt{2} - 1\) (respectively, \((\sqrt{5} - 1)/2\)) only.)

With the simplified notion, we have

\[
x_n = w_n(z; 0, 0, 0, \ldots, 0), \quad n \geq 1,
\]

\[
v_n = w_n(z; 0, 1, 1, \ldots, 1), \quad n \geq 2,
\]

\[
u_n = w_n(\sqrt{2} - 1; 1, 1, 1, \ldots, 1), \quad n \geq 1.
\]

\[
w_n = \begin{cases} 
    w_n((\sqrt{5} - 1)/2; 0, 1, \ldots, 0, 1, 0) & (n \text{ odd}) \\
    w_n((\sqrt{5} - 1)/2; 0, 1, \ldots, 0, 1) & (n \text{ even})
\end{cases} \quad n \geq 1.
\]

Our goals in this section are: (a) to prove the conjugation theorem (Theorem 2.4),
(b) to prove that when the pair of initial words are distinct letters, the set of nth order
\(\alpha\)-words is closed under conjugation and reversion (Theorems 2.4 and 2.7), and (c) to
count the number of distinct nth order \(\alpha\)-words (Theorem 2.5).

Throughout the rest of this section, let \(\alpha\) and \((u, v)\) be as in the definition of \(\alpha\)-words.

**Lemma 2.1.** \(\alpha\)-words of the same order are conjugates of one another. More precisely,
for \(r = (r_1, \ldots, r_n)\) with \(0 \leq r_i \leq a_i, 1 \leq i \leq n\), we have

\[
w_n(r) = T^k(w_n(0_n)),
\]

where

\[
k = r_1|v| + \sum_{i=2}^{n} |w_{i-1}(0_{i-1})|r_i,
\]

and \(0_i\) denotes the zero \(i\)-tuple.

The proof of this lemma depends on the following lemma.

**Lemma 2.2.** Let \(0 \leq r_1 \leq a_1\). Let

\[
x_{-1} = u, \quad x_0 = y_0 = v, \quad y_1 = v^{(a_1 - r_1)|uv|},
\]

and

\[
x_n = x_{n-1}x_{n-2}, \quad n \geq 1,
\]

\[
y_n = y_{n-1}y_{n-2}, \quad n \geq 2,
\]

i.e.

\[
x_n = w_n(0_n), \quad n \geq 1,
\]

\[
y_1 = w_1(r_1),
\]

\[
y_n = w_n(r_1, 0_{n-1}), \quad n \geq 2.
\]

Then \(y_n = T^{n|v|}(x_n), \quad n \geq 1\).

**Proof.** If \(r_1 = 0\), then the result is trivial since \(y_n = x_n\) for all \(n \geq 1\).
Now suppose \( r_1 > 0 \). Clearly, the result is true for \( n = 1, 2 \). Let \( n \geq 3 \) and suppose \( y_k = T^{r_1|v|}(x_k) \) for \( 1 \leq k < n \). Since \( a_1 \geq 1 \), it is easy to see that \( v^{a_1} \) is a prefix of \( x_n \), \( n \geq 1 \). In particular,
\[
x_{n-1} = v^{a_1}z_1, \quad x_{n-2} = v^{a_1}z_2,
\]
for some words \( z_1 \) and \( z_2 \). Therefore
\[
y_{n-1} = v^{a_1-r_1}z_1v^{r_1}, \quad y_{n-2} = v^{a_1-r_1}z_2v^{r_1},
\]
according to the inductive hypothesis. Hence
\[
y_n = y_{n-1}^{a_1}y_{n-2}
\]
\[
= (v^{a_1-r_1}z_1v^{r_1})^{a_n}(v^{a_1-r_1}z_2v^{r_1})
\]
\[
= T^{r_1|v|}(v^{r_1}(v^{a_1-r_1}z_1v^{r_1})^{a_n}(v^{a_1-r_1}z_2v^{r_1}))
\]
\[
= T^{r_1|v|}(v^{a_1}z_1)^{a_n}v^{a_1}z_2
\]
\[
= T^{r_1|v|}(x_{n-1}^{a_n}x_{n-2})
\]
\[
= T^{r_1|v|}(x_n).
\]

Proof of Lemma 2.1. We prove by induction on \( n \). Clearly, the result is true for \( n = 1 \). Now let \( n > 1 \) and assume that the result is true for \( 1 \leq k < n \). Let
\[
\alpha' = (a_2, a_3, \ldots, a_n),
\]
\[
\rho' = (r_2, r_3, \ldots, r_n),
\]
\[
u' = v,
\]
\[
v' = w_1(\alpha; u, v; r_1) = v^{a_1-r_1}uv^{r_1}.
\]
Then
\[
w_n(\alpha; u, v; r) = w_{n-1}(\alpha'; u', v'; r')
\]
\[
= T^{k_1}(w_{n-1}(\alpha'; u', v'; 0_{n-1})),
\]
where
\[
k_1 = \sum_{i=1}^{n-1} |w_{i-1}(\alpha'; u', v'; 0_{i-1})|r_{i+1}
\]
\[
= \sum_{i=1}^{n-1} |w_i(\alpha; u, v; (r_1, 0_{i-1}))|r_{i+1}
\]
\[
= \sum_{i=1}^{n-1} |w_i(\alpha; u, v; 0_i)|r_{i+1}
\]
by inductive hypothesis. By Lemma 2.2, we have
\[
w_{n-1}(\alpha'; u', v'; 0_{n-1}) = w_n(\alpha; u, v; (r_1, 0_{n-1})))
\]
\[
= T^{r_1|v|}(w_n(\alpha; u, v; 0_n)).
\]
Since \( k_1 + r_1|v| = k \), the result follows. \( \square \)
We have proved that $\alpha$-words of order $n$ are conjugates of $w_n(0_n)$. It is easily seen that the converse is not true in general. However, the converse does hold when the pair of initial words are distinct letters (Theorem 2.4(b)). To prove this fact, we need the following lemma, whose proof can be found in [20,21]. This lemma is also used to prove Corollary 2.6.

Throughout the rest of this paper, let $\{q_n\}$ be the sequence of positive integers defined by

$$q_{-1} = 1, \quad q_0 = 1, \quad q_n = a_n q_{n-1} + q_{n-2}, \quad n \geq 1.$$

**Lemma 2.3** (Zeckendorf representation). Each $k$ with $0 \leq k \leq q_n - 1$ has a unique representation

$$k = \sum_{i=1}^{n} q_{i-1} r_i,$$

where

$$0 \leq r_i \leq a_i, \quad 1 \leq i \leq n,$$

$$r_i = a_i \Rightarrow r_{i-1} = 0, \quad 2 \leq i \leq n.$$

**Theorem 2.4** (Conjugation theorem). Let $(u, v)$ be a pair of distinct letters.

(a) If $r = (r_1, r_2, \ldots, r_n)$ with $0 \leq r_i \leq a_i, 1 \leq i \leq n$, then $|w_n(r)| = q_n$, and $w_n(r) = T^k(w_n(0_n))$, where

$$k = \sum_{i=1}^{n} q_{i-1} r_i.$$

(b) For $n \geq 1$ and $0 \leq k \leq q_n - 1$, $T^k(w_n(0_n))$ is an $\alpha$-word.

(c) The set of all $\alpha$-words derived from $(u, v)$ is closed under conjugation.

(d) The set of conjugates of an $n$th order $\alpha$-word coincide with the set of $n$th order $\alpha$-words.

**Proof.** The first part of (a) is easy. The second part is a consequence of Lemma 2.1 and the first part.

(b) By Zeckendorf representation of $k$, there are integers $r_1, r_2, \ldots, r_n$ with $0 \leq r_i \leq a_i, 1 \leq i \leq n$ such that $k = \sum_{i=1}^{n} q_{i-1} r_i$. Let $r = (r_1, r_2, \ldots, r_n)$. It follows that $T^k(w_n(0_n)) = w_n(r)$ is an $\alpha$-word, according to part (a).

Part (c) follows immediately from (a) and (b); part (d) follows from (a) and (c). \qed

How many distinct $\alpha$-words of order $n$ are there? The following theorem gives an answer.

**Theorem 2.5.** Let $(u, v)$ be a pair of distinct letters. Then the conjugates of each $\alpha$-word are all distinct. In other words, there are $q_n$ $\alpha$-words of order $n$. 
Proof. This result follows immediately from part (b) of Lemma 3.2 and the primitivity of \( \alpha \)-words (Theorem 3.4) which are stated and proved in Section 3.

Corollary 2.6. Let \((u, v)\) be a pair of distinct letters and let \(n \geq 1\).

(a) Let \(r = (r_1, r_2, \ldots, r_n)\) and \(r' = (r'_1, r'_2, \ldots, r'_n)\) with \(0 \leq r_i \leq a_i\) and \(0 \leq r'_i \leq a_i\), \(1 \leq i \leq n\). Then \(w_n(r) = w_n(r')\) if and only if \(\sum_{i=1}^{n} q_{i-1}(r_i - r'_i) \equiv 0 \pmod{q_n}\).

(b) Let

\[ S = \{ r : 0 \leq r_i \leq a_i, \text{ for } i \geq 1, \text{ and } r_i = a_i \Rightarrow r_{i-1} = 0, \text{ for } i \geq 2 \}. \]

Define \(\varphi(r) = w_n(r)\), \(r \in S\). Then \(\varphi\) is a one-to-one correspondence between \(S\) and the set of \(\alpha\)-words of order \(n\).

Proof. Part (a) follows from Theorems 2.4(a) and 2.5. Part (b) follows from part (a) and Zeckendorf representation of nonnegative integers.

Theorem 2.7. Let \((u, v)\) be a pair of distinct letters. Then the set of \(\alpha\)-words is closed under reversion. More precisely, let \((r_1, r_2, \ldots, r_n)\) satisfies \(0 \leq r_i \leq a_i, \ 1 \leq i \leq n\). Then

\[ R(w_n(r_1, r_2, \ldots, r_n)) = w_n(a_1 - r_1, a_2 - r_2, \ldots, a_n - r_n). \]

Proof. The proof is similar to the proof of part (i) of Theorem 3 of [7].

3. Canonical palindrome factorization and primitivity

In this section let \((u, v)\) be a pair of distinct letters. We will prove that \(\alpha\)-words are primitive (Theorem 3.4) and that every \(\alpha\)-word can be expressed uniquely as a product of two palindromes (Corollary 3.5). The palindrome factorization of the sequence \(\{w_n\}\) defined below is particularly important. See for example Theorems 6.4 and 6.6.

A word \(w\) is said to be a palindrome if \(w = \varepsilon\), the empty word, or \(R(w) = w\). The representations \(w_1w_2\) and \(w_2w_1\), where \(w_1\) and \(w_2\) are palindromes, are considered to be the same if \(w_2 = \varepsilon\).

For simplicity, we denote \(w_n(0_n)\) by \(x_n\), \(n \geq 1\). Let \(\{u_n\}\), \(\{v_n\}\), \(\{w_n\}\) denote sequences of words defined as follows:

\[
\begin{align*}
v_1 &= v^a, & u_1 &= u, & u_2 &= v^{a+1}, \\
v_n &= (u_{n-1}v_{n-1})^{a-1}u_{n-1}, & n &\geq 2, \\
u_n &= (u_{n-2}v_{n-2})^{a-1}u_{n-2}, & u_{n-2}v_{n-2}v_{n-1} &= v_{n-1}v_{n-2}u_{n-2}, & n &\geq 3, \\
w_n &= \begin{cases} w_n(0, a_2, 0, a_4, \ldots, 0, a_{n-1}, 0) & \text{if } n \text{ is odd,} \\
w_n(0, a_2, 0, a_4, \ldots, 0, a_n) & \text{if } n \text{ is even and } \geq 1. \\
\end{cases}
\end{align*}
\]

These notations will be used throughout the rest of this paper.

Theorem 3.1. Let \(n \geq 1\) in statements (a)--(e) and \(n \geq 2\) in (f)--(g).

(a) \(u_n\) and \(v_n\) are palindromes.
(b) \(|v_n| = q_n - q_{n-1}, \ |u_n| = q_n - 1.\)

(c) Canonical palindrome factorization of \(w_n:\)

\[
w_n = \begin{cases} u_nv_n & \text{if } n \text{ is even,} \\ v_nu_n & \text{if } n \text{ is odd.} \end{cases}
\]

(d) If \(n\) is odd (respectively, even), then \(w_n\) and \(u_{n+1}\) differ by the last (respectively, first) letter only.

(e) \(u_n\) is not an \(x\)-word if \(|u_n| \geq 2.\)

(f) \[
u_n = c_nv_1v_2 \cdots v_{n-2}v_{n-1} = v_{n-1}v_{n-2} \cdots v_2v_1c_n,
\]

where

\[
c_n = \begin{cases} v & \text{if } n \text{ is even,} \\ u & \text{if } n \text{ is odd.} \end{cases}
\]

(g) \(x_n = \tilde{u}_{n+1}d_n\) where \(\tilde{u}_{n+1}\) is a palindrome and

\[
d_n = \begin{cases} uv & \text{if } n \text{ is even,} \\ vu & \text{if } n \text{ is odd.} \end{cases}
\]

**Proof.** Part (a) and (b) follow easily from the definition.

(c) This is clearly true for \(n = 1\) and \(2.\) Suppose \(n \geq 3\) and (c) is true for all positive integers less than \(n.\) Then for odd \(n:\)

\[
w_n = w_n^{u_n}w_{n-2} = (u_{n-1}v_{n-1})^{u_n}(v_{n-2}u_{n-2})
= ((u_{n-1}v_{n-1})^{u_n-1}u_{n-1})(v_{n-1}v_{n-2}u_{n-2}) = v_nu_n,
\]

for even \(n:\)

\[
w_n = w_{n-2}w_n^{v_n} = (u_{n-2}v_{n-2})(v_{n-1}u_{n-1})^{v_n}
= u_{n-2}v_{n-2}v_{n-1}u_{n-1}(v_{n-1}u_{n-1})^{v_n-1} = u_nv_n.
\]

Part (d) is also proved by induction. The inductive step goes as follows. For odd \(n,\)

\[
w_n = w_n^{u_n}w_{n-2},
\]

\[
u_{n+1} = (u_{n-1}v_{n-1})^{u_n}u_{n-1} = w_n^{u_n}u_{n-1}
\]

by part (c). By inductive hypothesis, \(w_{n-2}\) and \(u_{n-1}\) differ by the last letter only. Therefore, so do \(w_n\) and \(u_{n+1}.\) The proof for even \(n\) is similar.

Part (e) follows from part (d) and Theorem 2.4.
(f) Since

\[ u_2 = v^{q_1+1} = uv_1, \quad u_3 = u_1v_1v_2 = uv_1v_2, \]
\[ u_n = u_{n-2}v_{n-2}v_{n-1}, \]

the result follows by induction on \( n \).

(g) By part (f) there is a palindrome \( \tilde{u}_{n+1} \) such that

\[ u_{n+1} = \begin{cases} uv_{n+1} & \text{if } n \text{ is odd}, \\ u\tilde{u}_{n+1} & \text{if } n \text{ is even}. \end{cases} \]

It follows from (d) that

\[ w_n = v\tilde{u}_{n+1}u, \quad R(w_n) = u\tilde{u}_{n+1}v. \]

Therefore by the conjugation theorem and the identities

\[ a_2q_1 + a_4q_3 + \cdots + a_nq_{n-1} = q_n - 1 \text{ if } n \text{ is even}, \]
\[ a_1q_0 + a_3q_2 + \cdots + a_nq_{n-1} = q_n - 1 \text{ if } n \text{ is odd} \]

(which can be proved easily by induction), we have

\[ x_n = \begin{cases} T(w_n) & \text{if } n \text{ is even}, \\ T(R(w_n)) & \text{if } n \text{ is odd}. \end{cases} \]

Hence \( x_n = \tilde{u}_{n+1}d_n \). \( \Box \)

The factorization \( x_n = \tilde{u}_{n+1}d_n \) in part (g) of Theorem 3.1 has been obtained by de Luca and Mignosi [18]. It confirms the primitivity of \( \varepsilon \)-words (Theorem 3.4 below).

A word \( w \) is said to be primitive if it is not a power of another word. The following lemma contains some basic facts about primitivity.

**Lemma 3.2** (Shyr [37]). (a) The set of primitive words is closed under conjugation and reversion.

(b) Let \( n \geq 2 \). A word \( w \) of length \( n \) is primitive if and only if its conjugates are all distinct.

**Lemma 3.3** (de Luca and Mignosi [18, Lemma 2]). If \( w = zd \) where \( z \) is a palindrome in \( \{u, v\}^* \) and \( d = uv \) or \( vu \), then \( w \) is primitive.

**Theorem 3.4.** \( \varepsilon \)-words are primitive.

**Proof.** This follows from Theorem 3.1, Lemma 3.3, Theorem 2.4 and Lemma 3.2(a). \( \Box \)

**Corollary 3.5** (Canonical palindrome factorization). Every \( \varepsilon \)-word has a unique representation as a product of two palindromes.
Proof. This follows from Lemma 2 of [8], Theorems 2.4 and 3.4. □

Special cases of Theorems 3.4 and 3.5 have been obtained by de Luca and Mignosi [18], Z.X. Wen and Z.Y. Wen [44] and Chuan [8]. We recall that Lemma 3.2 and Theorem 3.4 have been used to prove Theorem 2.5.

4. More α-words

For reasons which will be apparent in Section 6 we will enlarge the class of α-words derived from a pair of distinct initial words. The elements of this new class are again called α-words.

Let α and (u, v) be as in Section 2 and let \( w_{-1} = u, w_0 = v \). For
\[
\begin{align*}
  n &\geq 1, \\
  1 &\leq j \leq a_n, \\
  0 &\leq r_n \leq j, \\
  0 &\leq r_i \leq a_i, \quad 1 \leq i \leq n - 1, \\
  r &=(r_1, \ldots ,r_n),
\end{align*}
\]
define the α-word \( w_n(r|j) = w_n(\alpha; u,v;r|j) \) by
\[
 w_n(r|j) = w_{n-1}(r_1,\ldots,r_{n-1})^{j-r_n}w_{n-2}(r_1,\ldots,r_{n-2})w_{n-1}(r_1,\ldots,r_{n-1})^n.
\]
Note that
\[
 w_n(r|a_n) = w_n(r).
\]

\((n,j)\) is called the order of the α-word \( w_n(r|j) \).

In the nth level, there is nothing special about \( a_n \). Results that hold for \( a_n \) also hold for \( j \) with \( 1 \leq j \leq a_n \) (also for \( j \) larger than \( a_n \) as well). However, slight modification of notations in the statements of these results are sometimes necessary. Here are some examples.

Let \((u,v)\) be a pair of distinct letters. Theorem 2.4(a) (conjugation theorem) is replaced by: If \( 1 \leq j \leq a_n, 0 \leq r_n \leq j, 0 \leq r_i \leq a_i, 1 \leq i \leq n - 1 \) then
\[
\begin{align*}
  |w_n(r_1,\ldots,r_n|j)| &= q_{n,j}, \\
  w_n(r_1,\ldots,r_n|j) &= T^k(w_n(0_n|j)),
\end{align*}
\]
where
\[
\begin{align*}
  q_{nj} &= jq_{n-1} + q_{n-2}, \\
  k &= \sum_{i=1}^{n} q_{i-1}r_i.
\end{align*}
\]
Theorem 2.7 now reads:

\[ R(w_n(r_1, \ldots, r_n|j)) = w_n(a_1 - r_1, \ldots, a_{n-1} - r_{n-1}, j - r_n|j). \]

The following theorem is an analog of Theorem 3.1. Let

\[ v_{nj} = v_j, \quad \text{if } a_1 \geq 1 \text{ and } 1 \leq j \leq a_1, \]
\[ v_{nj} = (u_{n-1}v_{n-1})^{-1}u_{n-1}, \quad \text{if } n \geq 2 \text{ and } 1 \leq j \leq a_n, \]
\[ w_{nj} = \begin{cases} 
  w_{n-2}w_{n-1}^j & \text{(n is even)}, \\
  w_{n-1}^jw_{n-2} & \text{(n is odd)},
\end{cases} \]

where the \( u_n \)'s, \( v_n \)'s and \( w_n \)'s are as in Section 3. (Note:

\[ w_{n,0} = w_n, \quad v_{n,0} = v_n, \quad v_{n,1} = u_{n-1}, \quad q_{n,0} = q_n \]

**Theorem 4.1.** Let \( n \geq 1 \).

(a) Each \( v_{nj} \) is a palindrome.
(b) \( |v_{nj}| = |v_{n1}| = q_{n-2}, \quad |v_{nj}| = q_{n,j-1}, \quad 2 \leq j \leq a_n. \)
(c) Canonical palindrome factorization:

\[ w_{nj} = \begin{cases} 
  v_{nj}u_n & \text{if } n \text{ is odd}, \\
  u_nv_{nj} & \text{if } n \text{ is even}.
\end{cases} \]

(d) If \( n \) is odd (respectively, even), then \( v_{n,j+1} \) and \( w_{nj} \) differ by the last (respectively, first) letter only, for \( a_n \geq 2 \) and \( 1 \leq j \leq a_n - 1. \)
(e) \( v_{nj} \) is not an \( x \)-word if \( 1 \leq j \leq a_n \) and \( |v_{nj}| \geq 2. \)

**Proof.** Part (a) follows by induction.
Part (b) follows from Theorem 3.1(b).
Part (c) follows from the proof of Theorem 3.1(c) with \( j \) in place of \( a_n. \)
(d) Let \( n \) be odd. By part (b) we have

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

Therefore the result follows from Theorem 3.1(d). The proof for even \( n \) is similar.
(e) First note that, according to Theorem 3.1(d), \( v_{n1} = u_{n-1} \) is not an \( x \)-word if \( |v_{n1}| \geq 2. \) Next let \( a_n \geq 2 \) and \( 2 \leq j \leq a_n. \) Then the result follows from part (d) and the conjugation theorem. \( \square \)
5. Unbordered words

A word \( z \in \mathcal{A}^* \) is said to be a \textit{border} of a word \( w \in \mathcal{A}^* \) if \( 0 < |z| < |w| \) and there are words \( x \) and \( y \) in \( \mathcal{A}^* \) such that

\[ w = xz = zy. \]

A word \( w \in \mathcal{A}^* \) is said to be \textit{bordered} if it has a border, or equivalently, if \( w = zxz \) for some \( z \) and \( x \) in \( \mathcal{A}^* \) with \( z \neq \varepsilon \). \( w \) is said to be \textit{unbordered} if it is not bordered. Bordered (resp., unbordered) word is also called \textit{overlapping} (resp., \textit{nonoverlapping}, \textit{d-primitive}, or \textit{primary}) word. Border is also called \textit{bifix}. (See [19, 27, 30, 37, 43].)

As will be seen in the next section, unbordered \( \alpha \)-words play an important role in the determination of factors of \( f(\alpha) \). In this section we prove some lemmas about unbordered words and we show that the \( \alpha \)-words \( w_nj \) and \( R(w_nj) \) derived from a initial pair of distinct letters, defined as in Section 4, are unbordered.

Lemma 5.1. Let \( w \) be an unbordered word having length \( n \geq 2 \). Then

(a) \( R(w) \) is unbordered.

(b) \( w \) is primitive.

(c) The conjugates \( w, T^{-1}(w), \ldots, T^{-n+1}(w) \) are distinct.

Proof. (a) and (b) clearly hold. Since (b) \( \Leftrightarrow \) (c), (c) also holds. \( \square \)

Characterizations of unbordered words are given in Lemmas 5.2 and 5.3.

Lemma 5.2. Let \( |w| = n \geq 2 \). Then (a) \( w \) is unbordered if and only if (b) for each \( k \) with \( 1 \leq k \leq n - 1 \), the prefixes of \( w, T^{-1}(w), \ldots, T^{-k}(w) \) of length \( k \) are distinct.

Proof. (a) \( \Rightarrow \) (b): We prove that if (b) does not hold, then \( w \) has a border. Let \( k \) be the smallest integer with \( 1 \leq k \leq n - 1 \) such that the prefixes of \( w, T^{-1}(w), \ldots, T^{-k}(w) \) of length \( k \) are not distinct. Let \( i \) and \( j \) be such that \( 0 \leq i < j < k \) and \( T^{-i}(w) \) and \( T^{-j}(w) \) have a common prefix of length \( k \). Clearly, \( j = k \), for otherwise \( T^{-i}(w) \) and \( T^{-j}(w) \) would have a prefix of length \( j \) in common, contradicting the minimality of \( k \).

Write

\[ T^{-i}(w) = yzx, \]
\[ T^{-k}(w) = y_1x_1, \]

where \( y, z, x, y_1, x_1 \) are words with \( |y| = i, |y_1| = k \), \( y_1 = yz \). Then

\[ w = zxy = x_1y_1 = x_1yz. \]

Thus \( z \) is a border of \( w \) because

\[ 0 < |z| = k - i \leq k \leq n - 1. \]
(b) ⇒ (a): If \( w \) has a border \( z \), then clearly \( z \) is a common prefix of \( w \) and \( T^{-k}(w) \) having length \( k \), where \( k = |z| \). □

A circulant matrix of order \( n \) over an alphabet \( A \) is an \( n \times n \) matrix of the form

\[
\begin{bmatrix}
c_1 & c_2 & \cdots & c_n \\
c_n & c_1 & \cdots & c_{n-1} \\
\vdots & \vdots & \ddots & \vdots \\
c_2 & c_3 & \cdots & c_1
\end{bmatrix}
\]

where \( c_1, c_2, \ldots, c_n \in A \). We denote this matrix by \( \text{circ}(c_1, c_2, \ldots, c_n) \) or \( \text{circ}(w) \) where \( w = c_1 c_2 \cdots c_n \). Note that [14]:

(a) the \((k,j)\)th entry of \( \text{circ}(c_1, c_2, \ldots, c_n) \) is \( c_{j-k+1} \) with subscripts modulo \( n \);
(b) the \( k \)th row of \( \text{circ}(w) \) is \( T^{-k+1}(w) \).

From (b) it is clear that Lemma 5.2 can be restated as follows:

**Lemma 5.3.** Let \( |w| = q \geq 2 \). Then \( w \) is unbordered if and only if for each \( k \) with \( 1 \leq k \leq q - 1 \), the rows of the upper left \((k+1) \times k\) submatrix of \( \text{circ}(w) \) are distinct words over \( A \).

Consider the word \( w = 00101 \) over the alphabet \( A = \{0,1\} \). We have

\[
\text{circ}(w) = \begin{bmatrix}
0 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0
\end{bmatrix}.
\]

For each \( 1 \leq k \leq 4 \), the rows of the upper left \( j \times k \) submatrix of \( \text{circ}(w) \) are distinct words over \( A \) if \( j = k + 1 \) (or less). However, this is no longer true if \( j > k + 1 \). This shows that \( j = k + 1 \) in Lemma 5.3 is optimal.

The existence of unbordered \( \alpha \)-words is guaranteed by a result of [19, 38] which asserts that every conjugate class of primitive words contains unbordered words. For example, take the Lyndon word in this conjugate class. By definition a Lyndon word is a primitive word that is minimal in the lexicographic order in its conjugate class. However, in the case that we are dealing with, two unbordered \( \alpha \)-words of each order can be determined explicitly in the following theorem which together with Lemma 5.1 gives another proof for the primitivity of \( \alpha \)-words.

**Theorem 5.4.** Let \( n \geq 1 \), \( a_n \geq 1 \) and \( 1 \leq j \leq a_n \). Then both \( w_{nj} \) and \( R(w_{nj}) \) are unbordered.

This theorem follows immediately from the following lemma.
Lemma 5.5 (Ehrenfeucht and Silberger [19] and Hsu et al. [27]). If \( w \) is an unbordered word, then \( wx^k \) and \( y^kw \) are unbordered for every integer \( k \geq 1 \), proper prefix \( y \) and suffix \( x \) of \( w \).

6. Factors of \( f(\alpha) \)

The main goal of this section is to determine the factors of any given characteristic sequence \( f(\alpha) \) using unbordered \( \alpha \)-words derived from the pair of initial letters \((1,0)\).

Let \( \alpha \) be as in Section 2 and let \((u,v)\) be a pair of distinct letters. For simplification, we write

\[
x_{-1} = u, \quad x_0 = v, \quad x_n = w_n(0n),
\]

\[
x_{nj} = w_n(0n|j), \quad 1 \leq j \leq a_n, \quad n \geq 1.
\]

Lemma 6.1. (a) \( x_n^2 \) is a factor of \( x_{n+3} \), \( n \geq 1 \).

(b) \( \alpha \)-words of order \((n,j)\) is a factor of \( x_{n+3} \), \( n \geq 1 \), \( 1 \leq j \leq a_n \).

(c) Let \( n \geq 1 \) be even (resp., odd). Then \( w_n, u_n \) and \( v_{n-1} \) (resp., \( R(w_n), u_n \) and \( v_{n-1} \)) are prefixes of \( vx_n \) (resp., \( ux_n \)).

Proof. (a) Since

\[
x_{n+3} = (x_{n+1}^a x_n) x_{n+1},
\]

it follows that \( x_n x_{n+1} \) and hence \( x_n^2 \) is a factor of \( x_{n+3} \).

(b) By part (a), \( x_n^2 = x_{n-1}^a x_{n-2} x_{n-1} x_{n-2} \) is a factor of \( x_{n+3} \), \( n \geq 1 \). By the conjugation theorem (Section 4), \( \alpha \)-words of order \((n,j)\) are conjugates of \( x_{nj} \). Therefore, they are of the form \( vx_{n-1}^j x_{n-2} x_{n-1}^j u \), where \( 0 \leq k \leq j \), \( uv = x_{n-1} \) or of the form \( vx_{n-1}^j u \) where \( uv = x_{n-2} \). Since \( x_{n-2} \) is a prefix of \( x_{n-1} \), the result follows.

(c) Let \( n \) be even. As in the proof of Theorem 3.1(g), we have \( w_n = T^{-1}(x_n) \). Hence there is a word \( z_n \) such that

\[
x_n = z_n v, \quad w_n = uz_n.
\]

It follows that \( w_n \) is a prefix of \( vx_n \). By Theorem 3.1(c) and the definition of \( u_n \), the assertion on \( u_n \) and \( v_{n-1} \) holds. The second assertion is proved similarly. \( \square \)

Let \( \alpha \) be an irrational number with \( 0 < \alpha < 1 \) and continued fraction \([0, a_1 + 1, a_2, \ldots] \).

We identify \( \alpha \) with the infinite sequence \( \{a_1, a_2, \ldots\} \). In the rest of this section, we consider \( \alpha \)-words derived from \((1,0)\) only.

Lemma 6.2. Factors of \( \alpha \)-words are factors of \( f(\alpha) \).

Proof. This is an immediate consequence of Lemma 6.1 and the well-known fact that \( x_n \) is a prefix of \( f(\alpha) \). \( \square \)
Theorem 6.3. Let $n \geq 1$, $a_n \geq 1$ and $1 \leq j \leq a_n$. Let
\[ y_{nj} = \begin{cases} v_{n,j+1} & \text{if } 1 \leq j < a_n, \\ u_{n+1} & \text{if } j = a_n, \end{cases} \]
\[ C_{nj} = \begin{bmatrix} \circ(wn_j) \\ y_{nj} \end{bmatrix}, \]
\[ L_{nj} = \begin{bmatrix} \circ(R(wn_j)) \\ y_{nj} \end{bmatrix}. \]

Then for each $1 \leq k \leq q_{nj}$, the rows of the upper left $(k+1) \times k$ submatrix of $C_{nj}$ (resp., $L_{nj}$) are precisely the factors of $f(\alpha)$ of length $k$.

Proof. By Theorems 3.1, 4.1 and Lemma 6.2, $y_{nj}$ is a factor of $f(\alpha)$ of length $q_{nj}$ which is not an $\alpha$-word. Now the assertion follows from Theorem 5.4, Lemma 5.3 and Lemma 6.2. □

A word $x \in A^*$ is said to be a cyclic factor of a word $w \in A^*$ if $|x| \leq |w|$ and $x$ is a factor of $w^2$.

Theorem 6.4. Let $w$ be an $\alpha$-word. Then:
(a) $w$ contains all factors of $f(\alpha)$ of length $\leq |w|$ as cyclic factors, except for a (palindrome) factor of $f(\alpha)$ of length $|w|$;
(b) $w$ has $(|w|^2 + 3|w| - 2)/2$ distinct cyclic factors. These cyclic factors are all factors of $f(\alpha)$.

Proof. According to Theorem 5.4, there is an unbordered $\alpha$-word $z$ in the conjugate class of $w$. Since conjugates of $z$ as well as their factors are factors of $w^2$, the result follows from Lemmas 5.2 and 6.2.

Let $w$ be any factor of $f(\alpha)$ and let $S(w)$ be the number of cyclic factors of $w$ which are factors of $f(\alpha)$. In general,
\[ S(w) \leq (|w|^2 + 3|w| - 2)/2 \]
because for each $1 \leq k < |w|$, $w$ can contain at most $k + 1$ factors of $f(\alpha)$ as cyclic factors. Theorem 6.4 implies that equality holds for $\alpha$-words. □

According to Theorems 3.1 and 4.1, each $y_{nj}$ differs from $w_{nj}$ only by the last (resp., first) letter if $n$ is odd (resp., even). Note that $y_{nj}$ and $R(w_{nj})$ also differ by one letter only, namely, the first (resp., last) letter if $n$ is odd (resp., even).

This observation and Theorem 6.4 imply the following theorem.

Theorem 6.5. Let $n \geq 1$, $a_n \geq 1$ and $1 \leq j \leq a_n$. Let
\[ w = \begin{cases} w_{nj}y_{nj} \text{ or } y_{nj}R(w_{nj}) & \text{if } n \text{ is odd,} \\ y_{nj}w_{nj} \text{ or } R(w_{nj})y_{nj} & \text{if } n \text{ is even.} \end{cases} \]
Write $w = c_1 c_2 \cdots c_{2q}$ where $q = |w_{nj}| = q_{nj}$, $c_i \in \{0, 1\}$, $1 \leq i \leq q$. For each $k$ with $1 \leq k \leq q$,

$$c_{q-h+1}c_{q-h+2} \cdots c_{q-h+k}, \quad 0 \leq h \leq k,$$

are the factors of $f(x)$ of length $k$.

Observe that the word $w$ in Theorem 6.4 is itself a factor of $f(x)$ since $w$ or $R(w)$ is a suffix of $T^q(w_{n+2,1})$ (resp., $T^q(R(w_{n+2,1}))$ if $n$ is even (resp., odd), where $q = q_{nj}$. Also, if $n$ is odd (resp., even), then $y_nR(w_n)$ (resp., $y_n w_n$) is a prefix of $0f(x)$ (resp., $1f(x)$).

We conclude this section by giving two interesting corollaries.

For each integer $k$ with $1 \leq k \leq q_{nj}$, denote the upper left $(k+1) \times k$ submatrix of $C_{nj}$ (resp., $L_{nj}$) by $C_{nj}(k)$ (resp., $L_{nj}(k)$). We will see in the following corollary that $C_{nj}(k)$ and $C_{mi}(k)$ are either equal or differ by exactly two entries. The proof is left to the reader.

**Corollary 6.6.** Let $|w_{nj}| < |w_{mi}|$.

(a) If either (i) $1 \leq k < |w_{nj}|$ or (ii) $k = |w_{nj}|$ and $n$ is even, then $C_{nj}(k) = C_{mi}(k)$.

(b) If $k = |w_{nj}|$ and $n$ is odd, then $C_{nj}(k)$ equals $C_{mi}(k)$ except the first and last rows interchanged. In other words, $C_{nj}(k)$ and $C_{mi}(k)$ differ by exactly two entries, namely the $(1,k)$-entry and the $(k+1,k)$-entry.

The conclusion is true for $L_{nj}(k)$ and $L_{mi}(k)$ if in (a)(ii) the condition ‘$n$ is even’ is replaced by ‘$n$ is odd’ and in (b) the condition ‘$n$ is odd’ is replaced by ‘$n$ is even’.

Let $\alpha$ and $\alpha'$ be irrational numbers with continued fractions

$$\alpha = [0, a_1 + 1, a_2, \ldots], \quad \alpha' = [0, b_1 + 1, b_2, \ldots].$$

Denote the $\alpha$-words associated with $\alpha$ by the usual notation and those associated with $\alpha'$ using primes. Denote the set of all factors of $f(\alpha)$ (resp., $f(\alpha')$) of length $k$ by $\mathcal{F}_k$ (resp., $\mathcal{F}'_k$). Let $q_{nj}$ be defined as before for $\alpha$.

**Corollary 6.7.** Let $n \geq 2$, $a_i = b_i$ for $1 \leq i \leq n - 1$ and $a_n < b_n$. If $q = q_{n-1} + q_n$ then $\mathcal{F}_k = \mathcal{F}'_k$ for $1 \leq k < q$ and $\mathcal{F}_q \neq \mathcal{F}'_q$.

**Proof.** Clearly, $w_{ij} = w'_{ij}$ for all $1 \leq i \leq n - 1$ and all $j$ involved. For $1 \leq j \leq a_n$, it is also clear that $w_{nj} = w'_{nj}$. Now

$$w'_{n,a_n+1} = \begin{cases} 
(w_{n-1})^{a_n+1}w_{n-2} & \text{if } n \text{ is odd}, \\
(w_{n-2}w_{n-1})^{a_n+1} & \text{if } n \text{ is even} 
\end{cases}$$

$$= \begin{cases} 
w_{n-1}w_{n-2}^{a_n} & \text{if } n \text{ is odd}, \\
w_{n-2}w_{n-1}^{a_n} & \text{if } n \text{ is even} 
\end{cases}$$

$$= w_{n+1,1}.$$
Let $w = w_{n,a+1} = w_{n+1,1}$ and let $q = |w| = q_{n-1} + q_n$. By Theorem 6.3

$$
\mathcal{F}_k = \mathcal{F}_k' \text{ for } 1 \leq k < q,
$$
$$
\mathcal{F}_q = \{w, T(w), \ldots, T^{q-1}(w), y\},
$$
$$
\mathcal{F}_q' = \{w, T(w), \ldots, T^{q-1}(w), y'\},
$$

where $y = y_{n+1,1}$ (resp., $y' = y'_{n+1,1}$) is a palindrome which differ from $w$ by one letter only. More precisely, expressing $w$ as $w = \alpha u \beta$, where $u$ is a palindrome, we find

$$
y = \begin{cases} 
1u1 & \text{if } n \text{ is odd}, \\
0u0 & \text{if } n \text{ is even},
\end{cases}
$$

$$
y' = \begin{cases} 
0u0 & \text{if } n \text{ is odd}, \\
1u1 & \text{if } n \text{ is even}.
\end{cases}
$$

Hence $y \neq y'$ and so $\mathcal{F}_q \neq \mathcal{F}_q'$.

To make Theorems 6.3 and 6.5 useful in practice, we must find an efficient way to generate $w_{nj}$ or $R(w_{nj})$. This is the main theme of [13]. Also it is shown in [13] that the $w_{nj}$'s and $R(w_{nj})$'s are the only unbordered factors of $f(x)$.

Acknowledgements

The author wishes to express her gratitude to the referees for their valuable comments and for bringing Ref. [44] to her attention.

References

[20] A.S. Fraenkel, J. Levitt and M. Shimshoni, Characterization of the set of values \( f(x) = \lfloor nx \rfloor, \) \( n = 1, 2, \ldots \), Discrete Math. 2 (1972) 335–345.