The subword complexity of a two-parameter family of sequences

Aviezri S. Fraenkel, Tamar Seeman
Department of Computer Science and Applied Mathematics
The Weizmann Institute of Science
Rehovot 76100, Israel
fraenkel, tamars@wisdom.weizmann.ac.il
http://www.wisdom.weizmann.ac.il/~fraenkel, ~tamars

Jamie Simpson
School of Mathematics, Curtin University
Perth WA 6001, Australia
simpson@cs.curtin.edu.au
http://www.cs.curtin.edu.au/~simpson

RECEIVED: 4/14/2000
ACCEPTED: 2/06/2001

Abstract

We determine the subword complexity of the characteristic functions of a two-parameter family $\{A_n\}_{n=1}^{\infty}$ of infinite sequences which are associated with the winning strategies for a family of 2-player games. A special case of the family has the form $A_n = \lfloor n\alpha \rfloor$ for all $n \in \mathbb{Z}_{>0}$, where α is a fixed positive irrational number. The characteristic functions of such sequences have been shown to have subword complexity $n+1$. We show that every sequence in the extended family has subword complexity $O(n)$.

1 Introduction

Denote by $\mathbb{Z}_{\geq 0}$ and $\mathbb{Z}_{>0}$ the set of nonnegative integers and positive integers respectively. Given two heaps of finitely many tokens, we define a 2-player heap game as follows. There are two types of moves:

1. Remove any positive number of tokens from a single heap.

2. Remove $k > 0$ tokens from one heap and $l > 0$ from the other. Here k and l are constrained by the condition: $0 < k \le l < sk + t$, where s and t are predetermined positive integers.

The player who reaches a state where both heaps are empty wins. The special case $s = t = 1$ is the classical Wythoff game [15], [16], [5].

Fraenkel showed [11] that every possible position in a game of this type can be classified as either a P-position, in which the Previous player can win, or an N-position, in which the Next player can win. Thus a winning strategy involves moving from an N-position to a P-position. Let $\mathcal{P}$ denote the set of all possible P-positions in a game with given values for s and t. Let $\text{mex}\, S$ denote the least nonnegative integer in $\mathbb{Z}_{\ge 0} \setminus S$.

Then $\mathcal{P} = \bigcup_{i=0}^{\infty}\{(A_i, B_i)\}$, where for every $n \in \mathbb{Z}_{\ge 0}$,

$$A_n = \text{ mex}\,\{\{A_i : 0 \le i < n\} \cup \{B_i : 0 \le i < n\}\}, \quad B_n = sA_n + tn.$$

Thus A_n and B_n are strictly increasing sequences, with $A_0 = B_0 = 0$ and $A_1 = 1$ for all $s, t \in \mathbb{Z}_{>0}$. Denoting $A = \bigcup_{i=1}^{\infty} A_i$ and $B = \bigcup_{i=1}^{\infty} B_i$, we have $A \cup B = \mathbb{Z}_{>0}$, and $A \cap B = \emptyset$.

Fraenkel [9] generalized the classical Wythoff game ($s = t = 1$) to the case $s = 1,\ t \ge 1$, and showed that a polynomial-time-computable strategy exists for the game. The strategy is based on the Ostrowski numeration system [12], with a base computed from the simple continued fraction expansion of α, where α satisfies $A_n = \lfloor n\alpha \rfloor$ for all $n \ge 0$. Fraenkel showed [11] that such α exists if and only if $s = 1$, but that a polynomial-time-computable strategy based on a numeration system defined by certain recursion formulas [10] nevertheless exists for every $s, t \in \mathbb{Z}_{>0}$.

In this paper we investigate an additional property of the class of heap games for general $s, t \in \mathbb{Z}_{>0}$: the subword complexity of the characteristic function of A. For fixed $s, t \in \mathbb{Z}_{>0}$, define the characteristic function of A as $\chi = \chi(A) : \mathbb{Z}_{>0} \longrightarrow \{0, 1\}$, where

$$\chi(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}$$

A word w is a factor of y if there exist words u, v, possibly empty, such that $y = uwv$. Define the subword complexity function $c_{s,t} : \mathbb{Z}_{>0} \longrightarrow \mathbb{Z}_{>0}$, where $c = c_{s,t}(n) =$ number of distinct factors of length n of the infinite sequence χ. Our goal is to determine the subword complexity function c of χ for general $s, t \in \mathbb{Z}_{>0}$.

The problem of computing the subword complexity of a given sequence has been addressed in a number of earlier works. For a survey of results in this area, we refer the reader to [2] and [8]. In particular, [13] contains an analysis of the subword complexity of infinite sequences S of the form $S = f^{\omega}(b)$, where f is a morphism such that $b \in \{0, 1\}$ is a prefix of $f(b)$. For example, it is shown there that if the functions

$$f_0(n) = |f^n(0)|, \quad f_1(n) = |f^n(1)|$$

have asymptotic growth rate $\Theta(k^n)$ for some constant k, then S has linear subword complexity.

In section 2 we show that for every $s, t \in \mathbb{Z}_{>0}$, χ is generated by such a morphism f. In section 4 it is shown that both $|f^n(0)|$ and $|f^n(1)|$ have asymptotic growth rate $\Theta(k^n)$. Thus by [13], χ has linear subword complexity for all $s, t \in \mathbb{Z}_{>0}$. This is consistent with our result that for all $s, t \in \mathbb{Z}_{>0}$, $c(n+1) - c(n) \in \{1, 2\}$ for every $n \in \mathbb{Z}_{>0}$.

For every given $s, t \in \mathbb{Z}_{>0}$, the set of positive integers consists of intervals over which $c(n+1)-c(n) = 2$ for all n, alternating with intervals over which $c(n+1)-c(n) = 1$ for all n. In other words, there exist intervals of "fast growth" of $c(n)$ relative to n, alternating with intervals of relatively "slow growth". By computing $c(n)$ at the first point of every interval of fast growth, we found that the subword complexity at these points converges asymptotically to

$$E\Big((1 - \frac{s-1}{(s+t-1)\alpha}) + (1 + \frac{s-1}{(s+t-1)\alpha})n\Big),$$

where $E(x)$ denotes the closest integer to x, and $\alpha > 1$ is a constant defined below in (2). Similarly, the complexity at the first point of every interval of slow growth converges to

$$E\Big((1 + \frac{(s-1)\alpha}{(2s+t-2)\alpha - (s-1)})n + 1\Big).$$

These two limits are respectively the lower and upper bounds on the asymptotic subword complexity of χ. When $s = 1$, the intervals of fast growth of $c(n)$ are empty, so the lower and upper bounds are equivalent and we have $c(n) = n + 1$ for all $n \in \mathbb{Z}_{>0}$.

In section 3 we introduce the concept of *special* words, and show how to determine the number of distinct special factors of χ of any given length. In section 4 we use the results of section 3 to determine the subword complexity of χ. The subword complexity formula is presented both for finite n, and as an asymptotic value when n approaches ∞.

2 Preliminaries

In this section, we describe a morphism, f, which generates χ for fixed s and t.

2.1 An Equivalent Sequence

A morphism h is called *non-erasing* if $h(u) \geq 1$ for every word u. See e.g. [14].

Definition. For given values of s and t, let $f : \{0,1\}^* \longrightarrow \{0,1\}^*$ be a morphism defined by the following rules:

(i) $f(0) = 1^s$ (concatenation of 1 by itself s times),

(ii) $f(1) = 1^{s+t-1}0$.

Note that the morphism thus defined is non-erasing. Further, $f(u \cdot v) = f(u) \cdot f(v)$, where "$\cdot$" (usually omitted), denotes concatenation. We use standard function iteration: $f^0(u) = u$, and $f^i(u) = f(f^{i-1}(u))$ for $i \in \mathbb{Z}_{>0}$. So also $h^i(u \cdot v) = h^i(u) \cdot h^i(v)$ for all $i \in \mathbb{Z}_{\geq 0}$.

Notation. Let ϵ denote the empty word. Then $1^0 = 0^0 = \epsilon$, and for all $i \in \mathbb{Z}_{\geq 0}$, $f^i(\epsilon) = \epsilon$.

Since f is a non-erasing morphism and $f(1) = 1x$ $(x = 1^{s+t-2}0)$, we can define $F = f^\omega(1) = 1xf(x)f^2(x)f^3(x)\cdots$, the unique infinite string of which $f(1), f^2(1), f^3(1), \ldots$ are all prefixes [6].

Theorem 1. $F = \chi(A)$.

To prove our theorem, we apply the following result.

Lemma 1. *Suppose that for some $n \in \mathbb{Z}_{>0}$, the n-th one in F is at position k. Then the n-th zero is at position $sk + tn$.*

Proof. If the n-th one is at position k, then the length k prefix contains n ones and $k-n$ zeros. Since $F = f^\omega(1)$, we can apply the morphism to this prefix to obtain another (longer) prefix of F. This prefix will contain n copies of $f(1)$ and $k-n$ of $f(0)$, and so has n zeros, $n(s+t-1)+(k-n)s$ ones, and length $n+n(s+t-1)+(k-n)s = nt+ks$. Since it ends with $f(1)$ which ends in zero, the n-th zero is in position $nt+ks$. ■

Proof of Theorem 1. We show by induction that for all $n \in \mathbb{Z}_{>0}$, the n-th one is at position A_n, and the n-th zero is at position B_n in F.

(i) $n = 1$: $A_1 = 1$, and the first one is at position 1. Thus by Lemma 1, the first zero is at position $s + t = B_1$.

(ii) $n > 1$: Suppose that for all $i < n$, the i-th one and the i-th zero are at positions A_i and B_i respectively. From the definition of the A_i sequence, A_n is the least integer distinct from A_i and B_i for all $i < n$; thus either the n-th zero or the n-th one occurs at bit position A_n, with the other bit occurring at some later position. But Lemma 1 implies that the n-th one occurs earlier than the n-th zero, so it must be in position A_n, and by the definition of B_n, the n-th zero is in position B_n.

Thus for every $x \in \mathbb{Z}_{>0}$, the bit at position x of F is a 1 if and only if $\chi(x) = 1$, where χ is the characteristic sequence of A. ■

2.2 Properties of F

For the remainder of this paper, we determine the subword complexity of χ by analyzing F. To do so, we first collect several properties of F which are implied by the rules of the generating morphism f.

Lemma 2. *F consists of isolated 0-bits separated by 1^{s+t-1} or by 1^{2s+t-1}.*

Proof. The only way to generate a 0 is as the termination of $f(1) = 1^{s+t-1}0$. Thus every 0 is preceded by 1^{s+t-1}, and is followed by either $f(1)$ or $f(0)$, so 00 is not a factor of F. Therefore every 0 is followed by either $f(1)$ or $f(0)f(1)$. If it is followed by $f(1)$, then it is separated from the next 0-bit by 1^{s+t-1}. If it is followed by $f(0)f(1)$, then it is separated from the next 0-bit by 1^{2s+t-1}. ■

Lemma 3. *If $f(x) = f(y)$ then $x = y$.*

Proof. Let $x = x_1x_2 \cdots x_m$ and $y = y_1y_2 \cdots y_n$ and suppose $f(x) = f(y)$. Then we have

$$f(x_1)f(x_2) \cdots f(x_m) = f(y_1)f(y_2) \cdots f(y_n).$$

If $f(x_m)$ ends in a zero, then it must be $f(1)$ and so $x_m = 1$. Otherwise $x_m = 0$. The same applies to y_n and so we must have $x_m = y_n$ and

$$f(x_1)f(x_2) \cdots f(x_{m-1}) = f(y_1)f(y_2) \cdots f(y_{n-1}).$$

Continuing inductively we get $x_{m-1} = y_{n-1}$ and so on, giving $x = y$. ■

Remark. If $x = f(w)$, then by Lemma 3 w is the unique inverse of x, which we denote $f^{-1}(x)$.

Notation. Denote by $|w|$ the length of the factor w, i.e., the number of its letters, counting multiplicities.

Lemma 4. *Let w be a factor of F beginning with $f(0)f(1)$ or $f(1)$, and terminating with $f(1)$. Then $f^{-1}(w)$ exists, and $|f^{-1}(w)| < |w|$.*

Proof. Suppose that the assertion holds for all w terminating with $f(1)$, with $|w| \leq n$. Let w be any factor of length $n+1$ ($n \geq |f(1)|$), beginning with $f(0)f(1)$ or $f(1)$ and terminating with $f(1)$. We consider two cases.

(i) w begins with $f(1)$. Then $w = f(1)w'f(1)$. By Lemma 2, if w' is nonempty, then w' begins with $f(0)f(1)$ or $f(1)$, and $|w'f(1)| = n - s - t + 1 < n$. Then by the induction hypothesis, $f^{-1}(w) = 1f^{-1}(w'f(1))$, and $|1f^{-1}(w'f(1))| < |f(1)w'f(1)|$.

(ii) w begins with $f(0)f(1)$, so $w = f(0)f(1)w'f(1)$. The argument is as in the case (i) with an extra prefix $f(0)$. ■

3 Special Words

As stated in the introduction, our goal is to determine the subword complexity function c of F. Recall that for every $n \in \mathbb{Z}_{>0}$, $c(n)$ denotes the number of distinct words of length n which are factors of F. Thus $c(1) = 2$, and for every $n \in \mathbb{Z}_{>0}$, $c(n+1) - c(n)$ is the number of length n factors of F that can be followed by both a 0 and a 1 in F.

Definition. Following standard terminology, define a factor w of F to be *special* if both $w0$ and $w1$ are factors of F. If w can be extended only by adjoining *one* of 0, 1, then w is *nonspecial*.

Remark. x is special $\iff$ every suffix of x is special.

Definition. Let $N : \mathbb{Z}_{>0} \longrightarrow \mathbb{Z}_{\geq 0}$ be the function defined as follows. For every $n \in \mathbb{Z}_{>0}$, $N(n)$ denotes the number of distinct special words of length n.

Thus for all $n \in \mathbb{Z}_{>0}$, $c(n+1) - c(n) = N(n)$, so

$$c(n) = c(1) + \sum_{i=1}^{n-1} (c(i+1) - c(i)) = 2 + \sum_{i=1}^{n-1} N(i). \tag{1}$$

To determine the subword complexity of F, therefore, we first compute $N(n)$ for every $n \in \mathbb{Z}_{>0}$.

Notation. Let x_0 denote 1^{s+t-1}, and x_1 denote 1^{2s+t-2}.

Note that $s = 1 \Longrightarrow x_0 = x_1 = 1^t$.

Remark. If $k < 2s + t - 1$ then 1^k is special. In particular, x_0 and x_1 are special.

Definition. Given x, $y \in F$, x possibly nonempty. Then x is said to be a *proper prefix* of y if $y = xu$ for some nonempty u. Similarly, x is a *proper suffix* of y if $y = ux$ for some nonempty u.

Lemma 5. *Given any word $w = bu \in F$, $b \in \{0, 1\}$. Suppose that u is special and w is nonspecial. Then either $u = 1^k$ for some $k < 2s + t - 1$, or $f(1)$ is a proper prefix of u.*

Proof. Since u is extendible in two possible ways, whereas bu is extendible in only one way, it follows that both $0u$ and $1u$ are factors of F. Suppose that u contains no 0. Then Lemma 2 implies that $|1u| < 2s+t$, so we have $u = 1^k$, $k < 2s+t-1$. On the other hand, suppose that u contains at least one 0. Then u begins with $1^m 0$ for some $m \in \mathbb{Z}_{\geq 0}$. Since $0u \in F$, Lemma 2 implies that $m \in \{2s + t - 1,\ s + t - 1\}$. But since $1u \in F$, Lemma 2 implies that $m < 2s + t - 1$. Thus $m = s + t - 1$ and u begins with $1^{s+t-1}0 = f(1)$. ■

Definition. For given s and t, define $g : \{0, 1\}^* \longrightarrow \{0, 1\}^*$, where for all $x \in \{0, 1\}^*$, $g(x) = f(x)1^{s+t-1}$.

Lemma 6. *x special $\iff$ $g(x)$ special.*

Proof.

(i) Suppose that x is special, so both $x0$ and $x1$ are factors of F. Then Lemma 2 implies that both $x01$ and $x1$, and thus $f(x)f(01)$ and $f(x)f(1)$, are factors of F. But

$$f(x)f(01) = f(x)1^s1^{s+t-1}0 = f(x)1^{s+t-1}1^s0 = g(x)1^s0,$$

and

$$f(x)f(1) = f(x)1^{s+t-1}0 = g(x)0.$$

Therefore both $g(x)1$ and $g(x)0$ are factors of F, so $g(x)$ is special.

(ii) Suppose that $g(x) = f(x)1^{s+t-1}$ is special. Then both $g(x)0$ and $g(x)1$ are factors of F. But

$$g(x)0 = f(x)1^{s+t-1}0 = f(x)f(1),$$

so by Lemma 4, $f^{-1}(f(x)f(1)) = x1$ is a factor of F.

Suppose that $x = x'0$, for some x'. Then $g(x)1 = f(x')f(0)1^{s+t} = f(x')1^{2s+t}$, contradicting Lemma 2. Thus $x = x'1$ for some x', so

$$g(x)1 = f(x')f(1)1^{s+t} = f(x')1^{s+t-1}01^{s+t}.$$

Since $s+t-1 < s+t$, Lemma 2 implies that the 01^{s+t} terminating $g(x)1$ is followed by $1^{s-1}0$, to form

$$f(x')1^{s+t-1}01^{2s+t-1}0 = f(x01).$$

Thus $f^{-1}(f(x01)) = x01$ is a factor of F, so x is special. ■

Corollary 1. *x is special $\iff$ for all $i \in \mathbb{Z}_{\geq 0}$, every suffix of $g^i(x)$ is special.*

Proof. obtain:

$$x \text{ special} \iff g(x) \text{ special} \iff g^2(x) \text{ special} \iff \cdots \iff g^i(x) \text{ special},$$

for all $i \in \mathbb{Z}_{\geq 0}$. But a word is special if and only if all of its suffixes are special, so our result follows. ■

Theorem 2. *w is special $\iff$ w is a suffix of $g^i(x_1)$ for some i.*

Proof. $i \in \mathbb{Z}_{\geq 0}$, every suffix of $g^i(x_1)$ is special.

In the other direction, suppose that w is special. If w contains no zeros, then by Lemma 2, w must have the form 1^k for some $k \leq 2s+t-2$, so w is a suffix of $x_1 = g^0(x_1)$.

We prove the other cases by induction on $|w|$, the start of the induction being the case above. Suppose w contains a zero and both $w0$ and $w1$ occur in F. By Lemma 2, w must end in 1^{s+t-1}, so let

$$w = w[1]w[2]\cdots w[k]1^{s+t-1},$$

where $w[j]$ denotes the j-th bit of w. Again by Lemma 2 we see that $w[k] = 0$ and therefore $w[k-s-t+1]\cdots w[k] = f(1)$. We then consider $w[k-s-t]$. If this is zero then $w[1]\cdots w[k-s-t]$ ends in $f(1)$ or a suffix of $f(1)$; otherwise it ends in $f(0)$ or a suffix of $f(0)$. Going backwards in this way we can uniquely identify w as having the form

$$w = vf(u_1)f(u_2)\cdots f(u_m)1^{s+t-1},$$

where v is a nonempty suffix of $f(0)$ or $f(1)$. Say it is a suffix of $f(u_0)$, and denote by $xvf(u_1)\cdots f(u_m)1^{s+t-1}$ the longest special suffix of $f(u_0)f(u_1)\cdots f(u_m)1^{s+t-1}$. If $|xv| < |f(u_0)|$, then by Lemma 5, $xvf(u_1)\cdots f(u_m)1^{s+t-1}$ begins with $f(1)$. But this contradicts the fact that xv is a proper suffix of $f(u_0)$.

Thus $f(u_0)f(u_1)\cdots f(u_m)1^{s+t-1} = g(u_0\cdots u_m)$ is special, so by Lemma 6, $u_0\cdots u_m$ is special. Therefore by the induction hypothesis, $u_0\cdots u_m$ is a suffix of $g^i(x_1)$ for some i. Since w is a suffix of $f(u_0)f(u_1)\cdots f(u_m)1^{s+t-1} = g(u_0\cdots u_m)$, this implies that w is a suffix of $g^{i+1}(x_1)$. ■

For every $n \in \mathbb{Z}_{>0}$, define the set

$$S_n = \{w : \ |w| = n, \text{ and for some } i \in \mathbb{Z}_{\geq 0}, \ w \text{ is a suffix of } g^i(x_1)\}.$$

Theorem 2 implies that $N(n) = |S_n|$ for all $n \in \mathbb{Z}_{>0}$.

Theorem 3.

$$N(n) = \begin{cases} 2 & \text{if for some } i \in \mathbb{Z}_{\geq 0}, \ |g^i(x_0)| < n \leq |g^i(x_1)|, \\ 1 & \text{otherwise.} \end{cases}$$

To prove the theorem, we apply several results.

Lemma 7. (a) *For all $i \in \mathbb{Z}_{>0}$ and w, $g^i(w) = f^i(w)g^{i-1}(x_0)$.*

(b) *For all $i \in \mathbb{Z}_{>0}$ and w, $g^i(w) = f^i(w)f^{i-1}(x_0)\cdots f^1(x_0)f^0(x_0)$.*

(c) *For all $i \in \mathbb{Z}_{\geq 0}$ and x, y, $g^i(xy) = f^i(x)g^i(y)$.*

Proof. By induction on i. ■

Corollary 2. (a) *Let $i, j \in \mathbb{Z}_{\geq 0}$, with $i \leq j$. Then $g^i(x_0)$ is a suffix of $g^j(x_0)$.*

(b) *Let k, $m \in \mathbb{Z}_{>0}$, with $k \leq m$. Then for all $i \in \mathbb{Z}_{\geq 0}$, $g^i(1^k)$ is a suffix of $g^i(1^m)$.*

(c) *For all $i \in \mathbb{Z}_{\geq 0}$, $|g^i(x_0)| \leq |g^i(x_1)| < |g^{i+1}(x_0)|$, with $|g^i(x_0)| = |g^i(x_1)|$ if and only if $s = 1$.*

Proof.

(a) By Lemma 7(b), $g^j(x_0) = f^j(x_0) \cdots f^{i+1}(x_0) g^i(x_0)$.

(b) Let $\ell = m - k$. Lemma 7(c) implies that

$$g^i(1^m) = g^i(1^l 1^k) = f^i(1^l) g^i(1^k).$$

(c) By Lemma 7(c), $g^i(x_1) = g^i(1^{s-1}x_0) = f^i(1^{s-1}) g^i(x_0)$. Thus $|g^i(x_0)| \leq |g^i(x_1)|$, with equality if and only if $s = 1$. Lemma 7(a) implies that $g^{i+1}(x_0) = f^{i+1}(x_0) g^i(x_0)$, so to prove that $|g^i(x_1)| < |g^{i+1}(x_0)|$, it suffices to show that $|f^i(1^{s-1})| < |f^{i+1}(x_0)|$. But this is satisfied, since $|f^i(1^{s-1})| < |f^i(x_0)| < |f^{i+1}(x_0)|$. ■

Notation. By $x \in F$, we mean that x is a factor of F.

Proof of Theorem 3. Suppose that $n \leq s + t - 1 = |g^0(x_0)|$. We show that $|S_n| = 1$. Now, $g^0(x_1) = x_1$ terminates with x_0, and Lemma 7(b) implies that for all $i \in \mathbb{Z}_{>0}$, $g^i(x_1)$ terminates with x_0, which terminates with $1^{|n|}$. Thus $1^{|n|}$ is the unique member of S_n, so $|S_n| = 1$.

Let $i \in \mathbb{Z}_{\geq 0}$. Consider the set of n satisfying

$$|g^i(x_0)| < n \leq |g^{i+1}(x_0)|.$$

Then by Corollary 2(a), for all n in the set, $n > |g^0(x_0)| = s + t - 1$.

Given some n in the set, denote by w the suffix of length n of $g^{i+1}(x_0)$. Corollary 2(a) implies that w is a suffix of $g^j(x_0)$ for all $j \geq i + 1$. Thus by Corollary 2(b), w is a suffix of $g^j(x_1)$ for all $j \geq i + 1$, so we have $w \in S_n$.

If there exists a second member of S_n distinct from w, then this member of S_n is a suffix of $g^j(x_1)$ for some $j \leq i$. Now, Corollary 2(c) implies that for every positive integer $j < i$,

$$|g^j(x_0)| \leq |g^j(x_1)| < |g^{j+1}(x_0)| \leq \cdots \leq |g^i(x_0)|.$$

Since $n > |g^i(x_0)|$, it follows that if there exists a member of S_n distinct from w, then this member is a suffix of $g^i(x_1)$. Thus $|S_n| \in [1, 2]$.

Now, by Corollary 2(c), we have $|g^i(x_0)| \leq |g^i(x_1)| < |g^{i+1}(x_0)|$. It follows that either (i) $|g^i(x_0)| < n \leq |g^i(x_1)|$, or (ii) $|g^i(x_1)| < n \leq |g^{i+1}(x_0)|$. Since $N(n) = |S_n|$ for all n, it suffices to show that $|S_n| = 2$ in case (i), and $|S_n| = 1$ in case (ii).

(i) Suppose that $|g^i(x_0)| < n \leq |g^i(x_1)|$. We assume that $s \geq 2$, because otherwise the set of n satisfying this inequality is empty. Denote by w' the length-n suffix of $g^i(x_1)$. To prove that $|S_n| = 2$, we show that w and w' are distinct.

Let $z = g^i(x_0)$. Then by Lemma 7(a) and (c), we have

$$g^i(x_1) = g^i(1^{s-1}x_0) = f^i(1^{s-1})z,$$

$$g^{i+1}(x_0) = f^{i+1}(x_0)z.$$

Thus w is a suffix of $f^{i+1}(x_0)z$, and w' is a suffix of $f^i(1^{s-1})z$. Moreover, $n > |z|$, so to prove that w and w' are distinct, it is sufficient to show that the rightmost bits of $f^i(1^{s-1})$ and $f^{i+1}(x_0)$ differ.

For every $x \in F$, if x terminates with 1, then $f(x)$ terminates with $f(1) = 1^{s+t-1}0$, which ends in 0. Similarly, if x terminates with 0, then $f(x)$ terminates with $f(0) = 1^s$, which ends in 1. Thus for all $i \in \mathbb{Z}_{\geq 0}$, the rightmost bits of $f^i(1)$ and $f(f^i(1)) = f^{i+1}(1)$ differ. But since f is a morphism, $f^{i+1}(x_0)$ terminates with $f^{i+1}(1)$, and $f^i(1^{s-1})$ terminates with $f^i(1)$. Thus the rightmost bit of $f^i(1^{s-1})$ differs from that of $f^{i+1}(x_0)$. It follows that w and w' are distinct, so for all n satisfying $|g^i(x_0)| < n \leq |g^i(x_1)|$, $|S_n| = 2$.

(ii) Suppose that $|g^i(x_1)| < n \leq |g^{i+1}(x_0)|$. Since $n > |g^i(x_1)|$, there does not exist a suffix of length n of $g^i(x_1)$. Thus w is the only member of S_n, so $|S_n| = 1$. ■

Example. Let $s = 2$, $t = 1$. Then $f(1) = 110$, $f(0) = 11$, $x_0 = 11$ and $x_1 = 111$. Thus

(i) $g(x_0) = 11011011$, so $|g(x_0)| = 8$,

(ii) $g(x_1) = 11011011011$, so $|g(x_1)| = 11$,

(iii) $g^2(x_0) = 110110111101101111011011$, so $|g^2(x_0)| = 24$.

Note that both $g(x_1)$ and $g^2(x_0)$ terminate with $g(x_0)$ — consistent with parts (a) and (b) of Corollary 2.

Now, for example, $|g(x_0)| < 10 \leq |g(x_1)|$, so we show that $|S_{10}| = 2$. The suffix of length 10 of $g^2(x_0)$ is 1111011011, which we denote w. In fact, Corollary 2 implies that $g^2(x_0)$, and thus w, is a suffix of $g^j(x_1)$ for all $j \geq 2$. Thus the suffix $w' = 1011011011$ of $g(x_1)$ is the only member of S_{10} which is distinct from w. Both w and w' terminate with $g(x_0) = 11011011$, but they differ in the bit immediately preceding $g(x_0)$, so they are distinct. Thus $|S_{10}| = 2$.

On the other hand, $|g(x_1)| < 15 \leq |g^2(x_0)|$, so we show that $|S_{15}| = 1$. The suffix of length 15 of $g^2(x_0)$ is 101101111011011, which by Corollary 2 is a suffix of $g^j(x_1)$ for all $j \geq 2$. Since $15 > |g(x_1)|$, there does not exist an additional member of S_{15} as there did in S_{10}. Thus $|S_{15}| = 1$.

4 Subword Complexity of F

In this section we determine the subword complexity $c(n)$ of F. In section 2, we defined $F = f^{\omega}(1)$, where f is a morphism such that $f(1)$ begins with 1. Thus the results of [13] indicate that the order of $c(n)$ can be determined directly from the order of the functions

$$u(n) = |f^n(1)| \quad \text{and} \quad v(n) = |f^n(0)|.$$

More precisely, if both $u(n) = \Theta(k^n)$ and $v(n) = \Theta(k^n)$, then F has linear subword complexity. For convenience, denote $u_n = u(n)$, and $v_n = v(n)$ for all $n \in \mathbb{Z}_{\geq 0}$.

Now $u_0 = 1$, and $u_1 = |1^{s+t-1}0| = s+t$. For $n \geq 2$,

$$f^n(1) = f^{n-1}(f(1)) = f^{n-1}(1^{s+t-1}0) = f^{n-1}(1^{s+t-1})f^{n-2}(1^s),$$

so $u_n = (s+t-1)u_{n-1} + su_{n-2} = ru_{n-1} + su_{n-2}$, where $r = s+t-1$. The characteristic polynomial of this recurrence is $x^2 - rx - s = 0$, which has solutions

$$\alpha = \frac{r + \sqrt{r^2+4s}}{2}, \quad \beta = \frac{r - \sqrt{r^2+4s}}{2}. \tag{2}$$

Thus for general $n \in \mathbb{Z}_{\geq 0}$, $u_n = c_1\alpha^n + c_2\beta^n$, where c_1 and c_2 are constants. Solving for c_1 and c_2, therefore, we obtain the solutions :

$$c_1 = \frac{1}{2} + \frac{r+2}{2\sqrt{r^2+4s}}, \quad c_2 = \frac{1}{2} - \frac{r+2}{2\sqrt{r^2+4s}}.$$

Now, $r = s+t-1 \geq 1$ implies that $\alpha > 1$, $-1 < \beta < 0$, and $-\frac{1}{2} < c_2 < 0$. Thus for even n, $-\frac{1}{2} < c_2\beta^n < 0$, and for odd n, $0 < c_2\beta^n < \frac{1}{2}$. It follows that

$$u_n = c_1\alpha^n + c_2\beta^n = \begin{cases} \lfloor c_1\alpha^n \rfloor, & n \text{ even} \\ \lceil c_1\alpha^n \rceil, & n \text{ odd.} \end{cases}$$

Notation. Let $E(x)$ denote the closest integer to x.

For every $n \in \mathbb{Z}_{>0}$,

$$u_n = E(c_1\alpha^n). \tag{3}$$

Also, for all $n \geq 1$, $f^n(0) = f^{n-1}(f(0)) = f^{n-1}(1^s)$, so $v_n = su_{n-1}$. Thus both $u_n = \Theta(\alpha^n)$ and $v_n = \Theta(\alpha^n)$, so F has linear subword complexity [13].

We now determine a more precise formula for $c(n)$, using equation (1):

$$c(n) = 2 + \sum_{i=1}^{n-1} N(i) = n + 1 + \sum_{i=1}^{n-1}(N(i) - 1).$$

Theorem 3 implies that this is equivalent to

$$c(n) = n + 1 + k, \tag{4}$$

where k is the number of integers $m < n$ such that $|g^i(x_0)| < m \leq |g^i(x_1)|$ for some $i \in \mathbb{Z}_{\geq 0}$.

Definition. For every $i \in \mathbb{Z}_{\geq 0}$, let I_i denote the interval of integers m satisfying $|g^i(x_0)| < m \leq |g^i(x_1)|$, and let $|I_i|$ denote its length.

Lemma 8. *Let $n \in \mathbb{Z}_{>0}$. If $n-1 \leq |x_0|$, or $|g^i(x_1)| \leq n-1 \leq |g^{i+1}(x_0)|$ for some $i \in \mathbb{Z}_{\geq 0}$, then*

$$c(n) = n + 1 + (s-1)\sum_{j=0}^{i} u_j,$$

where i is the minimal integer such that $|g^{i+1}(x_0)| \geq n-1$.

Proof. let k be the number of integers $m < n$ such that for some $j \in \mathbb{Z}_{\geq 0}$, $m \in I_j$. Then $k = \sum_{j=0}^{i} |I_j|$. Now, for every $j \in \mathbb{Z}_{\geq 0}$,

$$|I_j| = |g^j(x_1)| - |g^j(x_0)|.$$

Thus $|I_0| = |x_1| - |x_0| = (s-1)u_0$. Similarly, for $j \geq 1$, Lemma 7(c) implies that

$$|I_j| = |g^j(x_1)| - |g^j(x_0)| = |f^j(1^{s-1})| = (s-1)u_j.$$

Thus $k = \sum_{j=0}^{i} |I_j| = (s-1)\sum_{j=0}^{i} u_j$, so our result follows from (4). ■

Remark. If $s = 1$, then $x_0 = x_1$, so for every $m \in \mathbb{Z}_{>0}$, either $m \leq |x_0|$ or $|g^i(x_1)| = |g^i(x_0)| \leq m \leq |g^{i+1}(x_0)|$ for some $i \in \mathbb{Z}_{\geq 0}$. Thus Lemma 8 implies that in the case $s = 1$, $c(n) = n + 1$ for all $n \in \mathbb{Z}_{>0}$.

Applying Lemma 8, we analyze $c(n)$ for two different classes of n:
(a) $n = |g^i(x_0)| + 1$ for some $i \in \mathbb{Z}_{\geq 0}$, (b) $n = |g^i(x_1)| + 1$ for some $i \in \mathbb{Z}_{\geq 0}$. To do so, we define functions $L_i(n)$ and $U_i(n)$, which are dependent on i, and show that $c(n) = L_i(n)$ and $c(n) = U_i(n)$ in cases (a) and (b) respectively.

Lemma 9. *For every $i \in \mathbb{Z}_{\geq 0}$, $|g^i(x_0)| = (s+t-1)\sum_{j=0}^{i} u_j$.*

Proof. Lemma 7(b) implies that

$$|g^i(x_0)| = \sum_{j=0}^{i} |f^j(x_0)| = (s+t-1)\sum_{j=0}^{i} |f^j(1)| = (s+t-1)\sum_{j=0}^{i} u_j. \quad \blacksquare$$

Let $n = |g^i(x_0)| + 1$ for some $i \in \mathbb{Z}_{\geq 0}$. Then by Lemmas 8 and 9, we have

$$c(n) = \Big(1 - \frac{(s-1)\sum_{j=0}^{i-1} u_j}{(s+t-1)\sum_{j=0}^{i} u_j}\Big) + \Big(1 + \frac{(s-1)\sum_{j=0}^{i-1} u_j}{(s+t-1)\sum_{j=0}^{i} u_j}\Big)n\ .$$

But (3) implies that for all $m \in \mathbb{Z}_{\geq 0}$,

$$\sum_{j=0}^{m} u_j = \sum_{j=0}^{m} E(c_1\alpha^j) \approx c_1\frac{\alpha^{m+1}-1}{\alpha-1},$$

so it follows that

$$c(n) \approx \Big(1 - \frac{(s-1)(\alpha^i-1)}{(s+t-1)(\alpha^{i+1}-1)}\Big) + \Big(1 + \frac{(s-1)(\alpha^i-1)}{(s+t-1)(\alpha^{i+1}-1)}\Big)n\ .$$

Definition. For every $i \in \mathbb{Z}_{\geq 0}$,

$$L_i(n) = E\Big((1 - \frac{(s-1)(\alpha^i-1)}{(s+t-1)(\alpha^{i+1}-1)}) + (1 + \frac{(s-1)(\alpha^i-1)}{(s+t-1)(\alpha^{i+1}-1)})n\Big).$$

Theorem 4. *If $n = |g^i(x_0)| + 1$ for some $i \in \mathbb{Z}_{\geq 0}$, then $c(n) = L_i(n)$.*

The proof of Theorem 4 depends on the following two lemmas, which we leave to the reader to verify. Recall that for all $i \in \mathbb{Z}_{\geq 0}$, $u_i = c_1\alpha^i + c_2\beta^i$, where $-1 < \beta < 0$, and $-\frac{1}{2} < c_2 < 0$.

Lemma 10. *For every $i \in \mathbb{Z}_{\geq 0}$, $0 < \sum_{j=0}^{i} \beta^j \leq 1$, with equality to 1 if and only if $i = 0$.*

Lemma 11. *$s|c_2| < 1/2$.*

Proof of Theorem 4. If $s = 1$, then for all n, $L_i(n) = n + 1$. But Lemma 8 implies that $c(n) = n + 1$ for all n, so we are done. To prove the theorem for $s \geq 2$, we first note that if z is an integer, then given any real number x, proving that $E(x) = z$ is equivalent to proving that $|z - E(x)| < \frac{1}{2}$.

Let $s \geq 2$, and suppose that $n = |g^i(x_0)| + 1$ for some $i \in \mathbb{Z}_{\geq 0}$. Then letting $n' = n - 1 = |g^i(x_0)|$, we have $N(n') = 1$ by Theorem 3, so $c(n') = c(n) - 1$. Thus it suffices to show that

$$c(n') = E\Big(- \frac{(s-1)(\alpha^i - 1)}{(s+t-1)(\alpha^{i+1} - 1)} + (1 + \frac{(s-1)(\alpha^i - 1)}{(s+t-1)(\alpha^{i+1} - 1)})n\Big)$$

$$= n' + 1 + E\Big(\frac{(s-1)(\alpha^i - 1)}{(s+t-1)(\alpha^{i+1} - 1)}n'\Big).$$

Lemma 8 implies that this is equivalent to proving that

$$(s-1)\sum_{j=0}^{i-1} u_j = E\Big(\frac{(s-1)(\alpha^i - 1)}{(s+t-1)(\alpha^{i+1} - 1)}n'\Big),$$

or equivalently,

$$(s-1)\Big|\sum_{j=0}^{i-1} u_j - \frac{(\alpha^i - 1)}{(s+t-1)(\alpha^{i+1} - 1)}n'\Big| < \frac{1}{2}. \tag{5}$$

Now $u_j = c_1\alpha^j + c_2\beta^j$, and by Lemma 9, $n' = |g^i(x_0)| = (s+t-1)\sum_{j=0}^{i} u_j$. Therefore

$$(s-1)\Big|\sum_{j=0}^{i-1} u_j - \frac{(\alpha^i - 1)}{(s+t-1)(\alpha^{i+1} - 1)}n'\Big| = (s-1)\Big|c_2(\sum_{j=0}^{i-1}\beta^j - \frac{\alpha^i - 1}{\alpha^{i+1} - 1}\sum_{j=0}^{i}\beta^j)\Big|.$$

Thus by Lemmas 10 and 11, equation (5) is satisfied. ■

Let $n = |g^i(x_1)| + 1 = |g^i(x_0)| + |I_i| + 1$ for some $i \in \mathbb{Z}_{\geq 0}$. Lemma 9 implies that $|g^i(x_0)| = (s+t-1)\sum_{j=0}^{i} u_j$. Similarly, in the proof of Lemma 8, we saw that $|I_i| = (s-1)u_i$. Therefore,

$$n = (s+t-1)\sum_{j=0}^{i} u_j + (s-1)u_i + 1. \tag{6}$$

Now, Lemma 8 implies that

$$c(n) = n + 1 + \frac{(s-1)\sum_{j=0}^{i} u_j}{(s+t-1)\sum_{j=0}^{i} u_j + (s-1)u_i + 1}n\ .$$

Thus by (3), we have

$$c(n) \approx n + 1 + E\Big(\frac{c_1(s-1)\sum_{j=0}^{i}\alpha^j}{c_1(s+t-1)\sum_{j=0}^{i}\alpha^j + c_1(s-1)\alpha^i + 1}n\Big). \tag{7}$$

Definition. For every $i \in \mathbb{Z}_{\geq 0}$,

$$U_i(n) = n + 1 + E\Big(\frac{(s-1)(\alpha^{i+1}-1)}{(s+t-1)(\alpha^{i+1}-1) + (s-1)(\alpha^{i+1}-\alpha^i) + (\alpha-1)/c_1} n\Big). \qquad (8)$$

Theorem 5. *If $n = |g^i(x_1)| + 1$ for some $i \in \mathbb{Z}_{\geq 0}$, then $c(n) = U_i(n)$.*

Proof. If $s = 1$, then for all n, $U_i(n) = n+1$. But Lemma 8 implies that $c(n) = n+1$ for all n, so we are done. To prove the theorem for $s \geq 2$, we first note that the two formulas presented in equations (7) and (8) are equivalent.

Now, from Lemma 8 we have $c(n) = n + 1 + (s-1)\sum_{j=0}^{i} u_j$, so it suffices to prove that

$$(s-1)\sum_{j=0}^{i} u_j = E\Big(\frac{c_1(s-1)\sum_{j=0}^{i}\alpha^j}{c_1(s+t-1)\sum_{j=0}^{i}\alpha^j + c_1(s-1)\alpha^i + 1} n\Big),$$

or equivalently, that

$$(s-1)\Big|\sum_{j=0}^{i} u_j - \frac{c_1\sum_{j=0}^{i}\alpha^j}{c_1(s+t-1)\sum_{j=0}^{i}\alpha^j + c_1(s-1)\alpha^i + 1} n\Big| < \frac{1}{2}. \qquad (9)$$

Since $n = |g^i(x_1)| + 1$, equation (6) implies that (9) is equivalent to

$$(s-1)\Big|\sum_{j=0}^{i} u_j - \frac{c_1\sum_{j=0}^{i}\alpha^j((s+t-1)\sum_{j=0}^{i}u_j + (s-1)u_i + 1)}{c_1(s+t-1)\sum_{j=0}^{i}\alpha^j + c_1(s-1)\alpha^i + 1}\Big| < \frac{1}{2}.$$

This inequality can be verified by substituting $c_1\alpha^j + c_2\beta^j$ for u_j and applying Lemmas 10 and 11. ■

Theorems 4 and 5 imply that for every $i \in \mathbb{Z}_{\geq 0}$, $c(n) = L_i(n)$ and $c(n) = U_i(n)$ at, respectively, the beginning and end points of interval I_i. But Theorem 3 implies that $c(n)$ increases in steps of 2 throughout the entire interval, so it follows that for all $i \in \mathbb{Z}_{\geq 0}$,

$$|g^i(x_0)| < n \leq |g^{i+1}(x_0)| \Longrightarrow L_i(n) \leq c(n) \leq U_i(n).$$

Thus denoting

$$L(n) = \lim_{i\to\infty} L_i(n), \quad \text{and} \quad \mathrm{U(n)} = \lim_{\mathrm{i}\to\infty} \mathrm{U_i(n)},$$

$L(n)$ and $U(n)$ are, respectively, the lower and upper bounds of the asymptotic subword complexity of F as n approaches ∞. The precise formulas for these bounds are stated in Theorem 6, which we leave to the reader to verify.

Theorem 6. *For every $n \in \mathbb{Z}_{>0}$,*

$$L(n) = E\Big((1 - \frac{s-1}{(s+t-1)\alpha}) + (1 + \frac{s-1}{(s+t-1)\alpha})n\Big),$$

$$U(n) = E\Big((1 + \frac{(s-1)\alpha}{(2s+t-2)\alpha - (s-1)})n + 1\Big).$$

5 Conclusion

For every $s, t \in \mathbb{Z}_{>0}$, the subword complexity c of F is linear in n. We presented tight upper and lower bounds for $c(n)$, both for finite n, and as an asymptotic value as n approaches infinity. If $s = 1$, the lower and upper bounds are equivalent and we have $c(n) = n + 1$ for all $n \in \mathbb{Z}_{>0}$. This property follows from the fact that $c(n+1) - c(n) = N(n) = 1$ for all $n \in \mathbb{Z}_{>0}$.

If $s \geq 2$, however, the set of positive integers consists of intervals of integers n satisfying $N(n) = 1$, alternating with intervals of n over which $N(n) = 2$. Thus as n increases, $c(n)$ grows alternately slower and faster, depending on the type of interval in which n is located. In this case, therefore, the lower and upper bounds of $c(n)$ are distinct.

This difference between the cases $s = 1$ and $s \geq 2$ is related to a property of the infinite sequences A characterized by F. The sequence A is defined to be a *Beatty sequence* if there exist real α, β such that $A_n = \lfloor n\alpha + \beta \rfloor$ for all $n \in \mathbb{Z}_{>0}$. It turns out that A is a Beatty sequence if and only if $s = 1$, in which case we have $\alpha = (2 - t + \sqrt{t^2+4})/2$, $\beta = 0$ [11]. Since $\sqrt{t^2+4}$ is irrational for all $t \in \mathbb{Z}_{>0}$, it follows that α is irrational. In [1] it was shown that every Beatty sequence with irrational α has a characteristic sequence with subword complexity $c(n) = n + 1$ for all $n \in \mathbb{Z}_{>0}$. Our result is consistent with this.

The case $s = 1$ differs from $s \geq 2$ also in that A is spectral, that is, $|(A_{k+i} - A_k) - (A_{j+i} - A_j)| \leq 1$ for every $i, j, k \in \mathbb{Z}_{>0}$, if and only if $s = 1$ [11]. This is because the set of spectral sequences is precisely the set of Beatty sequences [4]. This motivates the question of whether A has a similar property when $s \geq 2$. Perhaps associated with some of the A sequences for $s \geq 2$ is an integer $m \geq 2$ such that for all i, j, k, $|(A_{k+i} - A_k) - (A_{j+i} - A_j)| \leq m$.

Another related observation is to estimate, for fixed $m \in \mathbb{Z}_{>0}$ and for every $n \in \mathbb{Z}_{>0}$, the number of increasing sequences of length n such that for all i, j, k with $1 \leq j, k, j+i, k+i \leq n$, $|(A_{k+i} - A_k) - (A_{j+i} - A_j)| \leq m$. For $m = 1$, the number of such words of length n is Euler's totient function, which is polynomial in n [3],[7]. For $m = 2$, however, R. Tijdeman observed (private communication), that the number of such words is exponential in n, since the entire set of length-n words with a characteristic sequence beginning with $\{01, 10\}^{\lfloor n/2 \rfloor}$ exhibits this property.

References

[1] P. Alessandri and V. Berthé, Three distance theorems and combinatorics on words, *L'Enseignement Mathématique* **44** (1998) 103–132.

[2] J.-P. Allouche, Sur la complexité des suites infinies, *Bull. Belg. Math. Soc.* **1** (2) (1994) 133–143.

[3] J. Berstel and M. Pocchiola, A geometric proof of the enumeration formula for Sturmian words, *Internat. J. Algebra Comput.* **3** (1993) 349–355.

[4] M. Boshernitzan and A.S. Fraenkel, A linear algorithm for nonhomogeneous spectra of numbers, *J. of Algorithms* **5** (1984) 187–198.

[5] H.S.M. Coxeter, The golden section, phyllotaxis and Wythoff's game, *Scripta Math.* **19** (1953) 135–143.

[6] K. Culik II and A. Salomaa, On infinite words obtained by iterating morphisms, *Theoret. Comput. Sci.* **19** (1982) 29–38.

[7] A. de Luca and F. Mignosi, Some combinatorial properties of Sturmian words, *Theoret. Comput. Sci.* **65** (1994) 361–385.

[8] S. Ferenczi, Complexity of sequences and dynamical systems, *Discr. Math.* **206** (1999) 145–154.

[9] A.S. Fraenkel, How to beat your Wythoff games' opponents on three fronts, *Amer. Math. Monthly* **89** (1982) 353–361.

[10] A.S. Fraenkel, Systems of numeration, *Amer. Math. Monthly* **92** (1985) 105–114.

[11] A.S. Fraenkel, Heap games, numeration systems and sequences, *Annals of Combinatorics* **2** (1998) 197–210.

[12] A. Ostrowski, Bemerkungen zur Theorie der diophantischen Approximationen, *Abh. Math. Sem. Hamburg* **1** (1922) 77–98.

[13] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, Lecture Notes in Computer Science vol. **172**, 1984, pp. 380–389.

[14] P. Roth, Every binary pattern of length six is avoidable on the two-letter alphabet, *Acta Inform.* **29** (1992) 95–107.

[15] W.A. Wythoff, A modification of the game of Nim, *Nieuw Arch. Wiskunde* **8** (1907) 199–202.

[16] A.M. Yaglom and I.M. Yaglom, *Challenging Mathematical Problems with Elementary Solutions*, Vol. II, Holden-Day, San Francisco, translated by J. McCawley, Jr., revised and edited by B. Gordon, 1967.