7th International Conference on
Combinatorics, Cryptography, Computer Science and Computing
November. 16-17, 2022

The palindromic Ziv-Lempel factorization of Fibonacci word

Morteza Mohammad-Noori[1]
School of Mathematics, Statistics and Computer Scienec, College of Science, University of Tehran, Tehran, Iran

Abstract

Based on the concept palindromic Ziv-Lempel factorization of an infinite word, we apply such a factorization on the infinite Fibonacci word and we present a closed form for it. This paper is based on a joint work with M. Jahannia, N. Rampersad and M. Stipulanti.

Keywords: Ziv-Lempel factorization, Fibonacci word, Singular words

AMS Mathematical Subject Classification [2010]: 05-06, 68R15

1 Introduction

Among factorizations of words used in text compression, the Ziv-Lempel [10] and Crochemore [4] factorizations are very well-known. As in [8] we consider infinite words and we use new variations of these factorizations based on the requirement that each factor is palindromic. In this paper we start indexing words at 0, thus, if we denote the *length* of a finite word u by $|u|$, then we can write $u = u_0 \cdots u_{|u|-1}$, where each u_i is an element of the given alphabet A. We denote infinite words by bold characters. For an infinite word $\mathbf{w}$ and a finite word u, we say that there is an occurrence of u at position j in $\mathbf{w}$ if $\mathbf{w} = pu\mathbf{w}'$ for some word p of length j and some infinite word $\mathbf{w}'$.

Given an infinite word $\boldsymbol{w}$, the *Ziv–Lempel* or *z-factorization* of $\boldsymbol{w}$ is the factorization

$$z(\boldsymbol{w}) = (z_1, z_2, z_3, \ldots)$$

where z_i is the shortest prefix of $z_i z_{i+1} z_{i+2} \cdots$ such that there is no occurrence of z_i in $\boldsymbol{w}$ at any position $j < |z_1 z_2 \cdots z_{i-1}|$. For instance, if $\boldsymbol{f}$ is the Fibonacci word, we have

$$z(\boldsymbol{f}) = (0, 1, 00, 101, 00100, 10100101, \ldots).$$

Note that if $\boldsymbol{w}$ is ultimately periodic the z-factorization is not well-defined, since eventually there will be no factors that do not occur previously in $\boldsymbol{w}$. We are not interested in ultimately periodic words in this paper and will therefore ignore this possibility and assume that any infinite word considered in this paper is aperiodic.

[1]speaker

In the context of combinatorics on words, the z-factorization has been computed for certain important families of words. Berstel and Savelli [2] observed that the z-factorization of the Fibonacci word coincides with the *singular factorization* of the Fibonacci word introduced by Wen and Wen [14]. Fici [5] has given an excellent survey of these and other factorizations of the Fibonacci word. Ghareghani, Mohammad–noori, and Sharifani [6] determined the z-factorization of standard episturmian words (epistandard words). Constantinescu and Ilie [3] used the z-factorization to define the *Lempel–Ziv complexity* of an infinite word.

In [8] we have introduced the *palindromic z-factorization* $pz(\boldsymbol{w})$ by requiring that each of the factors in the previous definitions be palindromes. That is, the *palindromic z-factorization* of $\boldsymbol{w}$ is the factorization

$$pz(\boldsymbol{w}) = (z_1, z_2, z_3, \ldots)$$

where z_i is the shortest *palindromic* prefix of $z_i z_{i+1} z_{i+2} \cdots$ such that there is no occurrence of z_i in $\boldsymbol{w}$ at any position $j < |z_1 z_2 \cdots z_{i-1}|$. The palindromic z-factorization may not exist for certain infinite words $\boldsymbol{w}$. For instance, if $\boldsymbol{w}$ only contains palindromes of bounded length, then the palindromic z-factorization will not exist. This type of factorization is therefore only interesting when applied to infinite words with arbitrarily long palindromic factors. For instance, if $\boldsymbol{f}$ is the Fibonacci word, we have

$$pz(\boldsymbol{f}) = (0, 1, 00, 101, 00100, 10100101, \ldots).$$

The main result of this paper states that $pz(\boldsymbol{f})$ and $z(\boldsymbol{f})$ are the same, and in fact are equal to the *singular factorization* of $\boldsymbol{f}$ (which we define later). In the sequel we mention some definitions and notation from combinatorics on words. Let A be a finite *alphabet*, i.e., a finite set made of *letters*. A *(finite) word* w over A is a finite sequence of letters belonging to A. If $w = w_0 w_1 \cdots w_n \in A^*$ with $n \geq 0$ and $w_i \in A$ for all $i \in \{0, \ldots, n\}$, then the *length* $|w|$ of w is $n + 1$, i.e., it is the number of letters that w contains. We let ε denote the empty word. This special word is the neutral element for concatenation of words, and its length is set to be 0. The set of all finite words over A is denoted by A^*, and we let $A^+ = A^* \setminus \{\varepsilon\}$ denote the set of non-empty finite words over A. An *infinite word* $\boldsymbol{w}$ over A is any infinite sequence over A. The set of all infinite words over A is denoted by A^ω. Note that in this paper infinite words are written with bold letters.

A finite word $w \in A^*$ is a *prefix* (resp., *suffix*) of another finite word $z \in A^*$ if there exists $u \in A^*$ such that $z = wu$ (resp., $z = uw$). The word $w \in A^*$ is said to be a *factor* of $z \in A^*$ if there exist $u, v \in A^*$ such that $z = uwv$. If $z = xy$ is a finite word over A, we write $x^{-1}z = y$ and $zy^{-1} = x$. Observe that if $z = xyt$ with $t, x, y, z \in A^*$, then $(xy)^{-1}z = y^{-1}(x^{-1}z)$ and $z(yt)^{-1} = (zt^{-1})y^{-1}$. In particular, for any words $u, v \in A^*$, we have $(uv)^{-1} = v^{-1}u^{-1}$. In the same way, a finite word $w \in A^*$ is a *prefix* of an infinite word $\boldsymbol{z} \in A^\omega$ if there exist $\boldsymbol{u} \in A^\omega$ such that $z = w\boldsymbol{u}$. The word $w \in A^*$ is said to be a *factor* of $z \in A^\omega$ if there exist $u \in A^*$ and $\boldsymbol{v} \in A^\omega$ such that $\boldsymbol{z} = uw\boldsymbol{v}$. Let $w = w_0 w_1 \cdots w_n \in A^*$ with $n \geq 0$ and $w_i \in A$ for all $i \in \{0, \ldots, n\}$. The *mirror image*, or *reversal*, of w is the word $w^R = w_n w_{n-1} \cdots w_0$ over A, i.e., the word obtained by reading w from right to left. We say that a word w over A is a *palindrome* if $w^R = w$.

A *factorization* of a finite word $w \in A^*$ is a finite sequence $(x_n)_{0 \leq n \leq m}$ of finite words over A such that

$$w = \prod_{n=0}^{m} x_n.$$

Similarly, a *factorization* of an infinite word $\boldsymbol{w} \in A^\omega$ is a sequence $(x_n)_{n \geq 0}$ of finite words over A such that

$$\boldsymbol{w} = \prod_{n \geq 0} x_n.$$

A *morphism* on A is a map $\sigma : A^* \to A^*$ such that for all $u, v \in A^*$, we have $\sigma(uv) = \sigma(u)\sigma(v)$. In order to define a morphism, it suffices to provide the image of letters belonging to A. A morphism is said to be *prolongable* on a letter $a \in A$ if $\sigma(a) = au$ with $u \in A^+$ and σ is non-erasing, i.e., the image of no letter is the empty word. If σ is prolongable on a, then $\sigma^n(a)$ is a proper prefix of $\sigma^{n+1}(a)$ for all $n \geq 0$. Therefore, the sequence $(\sigma^n(a))_{n\geq 0}$ of finite words defines an infinite word $\boldsymbol{w}$ that is a fixed point of σ.

In combinatorics on words, given an alphabet A, a set $X \subset A^+$ of non-empty words is a *code* on A if any word $w \in A^*$ has at most one factorization using words of X. For more on this topic, see, for instance, [11, Chapter 6]. The following result can be found in [11, Chapter 6].

Proposition 1.1. *Let A, B be two finite alphabets, and let $\sigma : A^* \to B^*$ be an injective morphism. If $X \subset A^+$ is a code on A, then $\sigma(X)$ is a code on B.*

In the following definition, we mention a new factorization of interest.

Definition 1.2. Let $\boldsymbol{w}$ be an infinite word over A. The *palindromic Ziv–Lempel* or *palindromic z-factorization* of $\boldsymbol{w}$ is the factorization

$$pz(\boldsymbol{w}) = (z_1, z_2, z_3, \ldots)$$

where z_i is the shortest *palindromic* prefix of $z_i z_{i+1} z_{i+2} \cdots$ such that there is no occurrence of z_i in $\boldsymbol{w}$ at any position $j < |z_1 z_2 \cdots z_{i-1}|$.

2 Main results

Before establishing the two palindromic factorizations of the Fibonacci word, we mention some definitions and necessary results. Some of them are well known and can be found in [5, 14, 8]. In the following definition, we follow the lines of [5].

Definition 2.1. Let $\boldsymbol{f}$ be the (infinite) Fibonacci word, i.e., the fixed point of the morphism $\varphi : 0 \mapsto 01, 1 \mapsto 0$, starting with 0. For all $n \geq 0$, define the finite word $h_n = \varphi^n(0)$ to be the nth iteration of φ on 0. The first few words of the sequence $(h_n)_{n\geq 0}$ are $0, 01, 010, 01001$. It is well known that the Fibonacci word $\boldsymbol{f}$ is the limit of $(h_n)_{n\geq 0}$. Let $(p_n)_{n\geq 3}$ be the sequence of the palindromic prefixes of $\boldsymbol{f}$, which are also called *central words*. The first few terms of this sequence are $\varepsilon, 0, 010, 010010, \ldots$. The *singular words* $(\hat{f}_n)_{n\geq 1}$ satisfy $\hat{f}_1 = 0$, $\hat{f}_2 = 1$ and, for all $n \geq 1$, $\hat{f}_{2n+1} = 0p_{2n+1}0$ and $\hat{f}_{2n+2} = 1p_{2n+2}1$. The first few singular words are $0, 1, 00, 101, 00100$.

The following properties of the singular words can be found in [14].

Proposition 2.2. *Let $(F_n)_{n\geq 0}$ be the sequence of Fibonacci numbers with initial conditions $F_0 = 0$ and $F_1 = 1$.*

(1) For all $n \geq 1$, $\hat{f}_n$ is a palindrome.

(2) For all $n \geq 1$, $|\hat{f}_n| = F_n$.

(3) For all $n \geq 4$, $\hat{f}_n = \hat{f}_{n-2}\hat{f}_{n-3}\hat{f}_{n-2}$.

(4) For all $n \geq 1$, $\hat{f}_n$ is not a factor of $\hat{f}_{n+1}$.

(5) *For all* $n \geq 1$, $\hat{f}_n$ *is not a factor of* $\prod_{m=1}^{n-1} \hat{f}_m$.

(6) *Let* $n \geq 1$ *and let* $\hat{f}_{n+1} = wa$ *where* $w \in \{0,1\}^*$ *and* $a \in \{0,1\}$. *If* $\hat{f}'_{n+1} = w\overline{a}$ *with* $\overline{a} = 1 - a$, *then* $\hat{f}_{n+2} = \hat{f}_n \hat{f}'_{n+1}$.

(7) *Let* $n \geq 3$ *and define* α_n *to be* 0 *if* n *is odd, or* 1 *if* n *is even. Then* $\hat{f}_n = \alpha_n \prod_{m=1}^{n-2} \hat{f}_m$.

The following result can be found in [5]. Note that the first factorization of the Fibonacci word $\boldsymbol{f}$ also appears in [14].

Proposition 2.3. *We have the following two factorizations of the Fibonacci word*

$$\begin{aligned} \boldsymbol{f} &= \prod_{n\geq 1} \hat{f}_n & (1) \\ &= 0 \cdot 1 \cdot 00 \cdot 101 \cdot 00100 \cdot 10100101 \cdots \\ &= 010 \prod_{n \geq 2} \hat{f}_{n-1} \hat{f}_n \hat{f}_{n-1} & (2) \\ &= 010 \cdot (0 \cdot 1 \cdot 0) \cdot (1 \cdot 00 \cdot 1) \cdot (00 \cdot 101 \cdot 00) \cdot (101 \cdot 00100 \cdot 101) \cdots . \end{aligned}$$

Moreover, the Ziv–Lempel factorization of the Fibonacci word is given by the sequence of singular words, i.e.,

$$z(\boldsymbol{f}) = (\hat{f}_1, \hat{f}_2, \hat{f}_3, \ldots).$$

As a matter of fact, the palindromic z-factorization of $\boldsymbol{f}$ is easily deduced from the previous result, as shown in the sequel.

Definition 2.4. For all $n \geq 2$, define

$$g_n := 010 \prod_{2 \leq m \leq n-1} \hat{f}_{m-1} \hat{f}_m \hat{f}_{m-1}.$$

From (2), observe that, for all $n \geq 2$, we have

$$\boldsymbol{f} = g_n \cdot (\hat{f}_{n-1} \hat{f}_n \hat{f}_{n-1}) \cdot \prod_{m \geq n+1} \hat{f}_{m-1} \hat{f}_m \hat{f}_{m-1}.$$

Interestingly, the prefix g_n of $\boldsymbol{f}$ can be factorized as a particular product of singular words.

Proposition 2.5. *(Proposition 7 of [8]) For all* $n \geq 2$, *we have*

$$g_n = \hat{f}_1 \hat{f}_2 \cdots \hat{f}_{n-1} \hat{f}_n \hat{f}_{n-1} \hat{f}_{n-2}. \tag{3}$$

Below we present the palindromic z-factorization of the Fibonacci word, which easily follows from already known results.

Theorem 2.6. *(Theorem 8 of [8]) The palindromic* z*-factorization of the Fibonacci word* $\boldsymbol{f}$ *is*

$$pz(\boldsymbol{f}) = (\hat{f}_1, \hat{f}_2, \hat{f}_3, \ldots).$$

Acknowledgment

The author would like to appreciate organizers of the 7-th international conference on Combinatorics, Cryptography, Computer Science and Computing.

References

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

[2] J. Berstel and A. Savelli, "Crochemore factorization of Sturmian and other infinite words". In Proc. MFCS'06, LNCS 4162, Springer, 2006, pp. 157–166.

[3] S. Constantinescu, L. Ilie, "The Lempel–Ziv complexity of fixed points of morphisms", SIAM J. Discrete Math. 21 (2007), 466–481.

[4] M. Crochemore, "Recherche linéaire d'un carré dans un mot", C. R. Math. Acad. Sci. Paris 296 (1983), 781–784.

[5] G. Fici, "Factorizations of the Fibonacci infinite word", J. Integer Seq. 18 (2015), no. 9, Article 15.9.3, 14 pp..

[6] N. Ghareghani, M. Mohammad–Noori, P. Sharifani, "On z-factorization and c-factorization of standard episturmian words", Theoret. Comput. Sci. 412 (2011), 5232–5238.

[7] A. Glen, "On Sturmian and Episturmian Words, and Related Topics", Ph.D. Thesis (2006).

[8] M. Jahannia, M. Mohammad-Noori, N.Rampersad, M. Stipulanti, Theoret. Comput. Sci. 790 (2019), 16–40.

[9] J. Justin and G. Pirillo, "Episturmian words and episturmian morphisms", Theoret. Comput. Sci. 276 (2002), no. 1-2, 281–313.

[10] A. Lempel and J. Ziv, "On the complexity of finite sequences", IEEE Transactions in Information Theory IT-22 (1976), 75–81.

[11] M. Lothaire, "Algebraic Combinatorics on Words", Encyclopedia of Mathematics and Its Applications, Cambridge University Press, vol. 90, 2002.

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

[13] B. Tan and Z.-Y. Wen, "Some properties of the Tribonacci sequence", European J. Combin. 28 (2007), no. 6, 1703–1719.

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

e-mail: `mmnoori@ut.ac.ir`