数理解析研究所講究録
910 巻 1995 年 75-81

Some remarks on Fibonacci infinite word

Giuseppe Pirillo

IAMI CNR viale Morgagni 67/A 50134 FIRENZE ITALIA

Abstract. *This paper concerns some factorizations of the Fibonacci infinite word. In particular, we recall two of them which use prefix codes and we present another one whose factors are never twice repeated and all belong to a biprefix code.*

Terminology and notations are those currently used in theoretical computer science [1, 2, 4, 5]. In this paper we use only the two letter *alphabet* $A^{+} = \{a, b\}$. We call (finite) *words* the elements of the *free monoid* A^{*}; we denote by 1 the *empty word*, by A^{+} the *free semigroup* on A and by $|u|$ the *length* of a word u. We consider a word u of length $k \geq 1$ as a map $u : \{0, 1, \ldots, k-1\} \rightarrow A$ and we write $u = u(0) \ldots u(i) \ldots u(k-1)$. We say that a word u is a *factor* of a word v if there exist two words $u', u'' \in A^{*}$ such that $v = u'uu''$. When $u' = 1$ (resp. $u'' = 1$) we say that u is a *left factor* (resp. *right factor*) of v.

A *right infinite word* on A is a map g from the set of non-negative integers into A and we write it as an infinite sequence:

$$g = g(0)g(1) \ldots g(i) \ldots$$

We say that a word u is a *factor* of g if there exist a word u' and a right infinite word g' such that $g = u'ug'$. If $u' = 1$ we say that u is a *left factor* of g. We say that a right infinite word g is *ultimately periodic* if there exists $p \geq 1$ such that $g(j+p) = g(j)$ for each $j \geq i$ for some $i \geq 0$. Let i, j be integers such that $0 \leq i \leq j$ and g be a right infinite word; we denote by $g(i,j)$ the word $g(i)\dots g(j)$.

Definition. *We say that a subset X of a free semigroup A^+ is a code over A if for all $n, m \geq 1$ and $x_1, \dots, x_n, x'_1, \dots, x'_m \in X$ the condition*

$$x_1 \dots x_n = x'_1 \dots x'_m$$

implies

$$n = m$$

and implies, for $i \in \{1, \dots n\}$,

$$x_i = x'_i.$$

Definition. *We say that a subset X of a free semigroup A^+ is a prefix set (resp. suffix set) if, for all $u, v \in X$, the condition u is a left factor (resp. right factor) of v implies $u = v$. We say that X is biprefix if it is both prefix and suffix.*

Clearly, a prefix or suffix or biprefix subset X is a code (see [1]). So we speak about *prefix*, *suffix* or *biprefix code.*

Now, let $\varphi : \{a,b\}^* \to \{a,b\}^*$ be the morphism whose restriction to $\{a,b\}$ is given by $\varphi(a) = ab$, $\varphi(b) = a$. Let us define the n–th Fibonacci finite word

f_n in the following way: $f_0 = b$ and, for each $n \geq 0$,

$$f_{n+1} = \varphi(f_n).$$

In particular, we have: $f_1 = a$, $f_2 = ab$, $f_3 = aba$, $f_4 = abaab$, $f_5 = abaababa$, $f_6 = abaababaabaab$, $f_7 = abaababaabaababaababa\ldots$. It is clear that, for each $n \geq 2$, f_n is the product (juxtaposition) $f_{n-1}f_{n-2}$ of f_{n-1} and f_{n-2}. Also, for each $n \geq 0$, $|f_n|$ is the n–th element F_n of the sequence of Fibonacci numbers 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377 ....

We note that, for each $n \geq 1$, f_n is a left factor of f_{n+1}. So there exists an unique infinite word, namely the Fibonacci infinite word f, such that, for each $n \geq 1$, f_n is a left factor of f (see, [2, 4]). Recall that the Fibonacci infinite word f is also the *Sturmian word* associated with the golden ratio $\Phi = (\sqrt{5}+1)/2$. We have

$$f = abaababaabaababaababaabaababaabaababaababa\ldots.$$

For each $n \geq 2$, we denote by g_n the product $f_{n-2}f_{n-1}$ and by h_n the longest common left factor of f_n and g_n. In particular, we have: $g_2 = ba$, $g_3 = aab$, $g_4 = ababa$, $g_5 = abaabaab$, ... and $h_2 = 1$, $h_3 = a$, $h_4 = aba$, $h_5 = abaaba$ .... We note that if $f(i) = b$ then $i > 0$ and $f(i-1) = f(i+1) = a$ and if $f(i, i+1) = aa$ then $i > 0$ and $f(i-1) = f(i+2) = b$; in other words, bb and aaa are not factors of f.

Lemma 1 belongs to the folklore (see for example [2, 3, 4]), it is very simple and states the *near–commutative property* (see [4]).

Lemma 1. *For each $n \geq 2$, i) $f_n = f_{n-1}f_{n-2} = f_{n-2}g_{n-1} = h_n xy$ and*

$g_n = f_{n-2}f_{n-1} = f_{n-1}g_{n-2} = h_nyx$, *where* $x, y \in \{a, b\}$, $x \neq y$ *and if* n *is even then* $xy = ab$ *and if* n *is odd then* $xy = ba$.

In March 1994 in Leipzig during the workshop *Logic and combinatorics of unary functions and related structures* and in July 1994 in Prato during the *Incontro di Combinatoria Algebrica* we announced the amusing properties of the following two factorizations of f: in the first one (resp. in the second one) the lenghts of the factors are progressively given by the Fibonacci numbers of odd index (resp. even index).

Proposition 1. *Let*

$$f = u_0u_1 \dots u_i \dots$$

be the factorization of f *such that* $|u_i| = F_{2i+1}$. *Then* $\{u_i \mid i \geq 0\}$ *is a prefix code.*

Proposition 2. *Let*

$$f = v_0v_1 \dots v_i \dots$$

be the factorization of f *such that* $|v_i| = F_{2(i+1)}$. *Then* $\{v_i \mid i \geq 0\}$ *is a prefix code.*

In [6] we proved the following result:

Theorem. *Let* g *be a right infinite word. If* g *is not ultimately periodic then there exists an infinite set* $\{h_i \mid i \geq 0\}$ *of words such that* $g = h_0h_1 \dots h_i \dots$, $\{h_i \mid i \geq 1\}$ *is a biprefix code and* $h_i \neq h_j$ *for positive integers*

$i \neq j$.

There are several biprefix factorizations of the Fibonacci infinite word "starting" from the beginning.

Proposition 3. *Let $n \geq 4$. Let $f = w_0 w_1 \dots w_i \dots$ be the factorization of f such that $|w_0| = F_n + F_{n-2} - 1$ and, for $i \geq 1$, $|w_i| = F_{n+2(i-2)-1} + 2F_{n+2(i-1)}$. Then $\{w_i \mid i \geq 0\}$ is a biprefix code and $w_i \neq w_j$ for positive integers $i \neq j$.*

Suppose now that the alphabet $\{a, b\}$ is endowed with a total order and consider on $\{a, b\}^+$ the *lexicographic order* induced by it.

We say that a word x is *n-divided* if it admits an n-divided factorization, i.e. a factorization

$$x = x_1 x_2 \dots x_n$$

such that, for each $i \in \{1, \dots n\}$, $x_i \in \{a, b\}^+$, and that, for each non trivial σ in the symmetric group S_n, $x_{\sigma(1)} x_{\sigma(2)} \cdots x_{\sigma(n)}$ strictly precedes x in the lexicographic order.

We say that an infinite word t is *ω-divided* if it admits a factorization $t = t_0 t_1 t_2 \dots t_i \dots$ such that for each $i \geq 0$ and, for each $n \geq 2$,

$$t_i \dots t_{i+n-1}$$

is an n-divided factorization.

Remark 1. The factorizations

$$f = u_0 u_1 \dots u_i \dots$$

$$= (a)(baa)(babaabaa)(babaababaabaababaabaa)\dots$$

and

$$f = v_0 v_1 \dots v_i \dots$$

$$= (ab)(aabab)(aabaababaabab)(aabaababaabaababaababaabaababaaaba)\dots$$

are ω-divisions respectively for the orders $a > b$ and $b > a$.

Remark 2. The factorizations $f = w_0 w_1 \dots w_i \dots$ are also ω-divisions. For example, for $n = 4$, we have

$$f = (abaaba)(baabaababaa)(babaabaababaabaababaababaabaa)\dots$$

and this is an ω-division for the order $a > b$,

Remark 3. At the origin of many other there is the following interesting factorization of f which holds for each $n \geq 1$:

$$f = f_n f_{n-1} f_n f_{n+1} f_{n+2} f_{n+3} f_{n+4} \cdots f_{n+i} \cdots$$

and, for example, the factorizations of Propositions 1-3 are strictly connected with it.

Remark 4. The proofs of Propositions 1-3 are based on Lemma 1 and are left to the reader.

REFERENCES

[1] J. Berstel and D. Perrin, *Theory of codes*, Academic Press, 1985.

[2] D. E. Knuth, *The Art of Computer Programming*, Addison–Wesley, Reading, Mass., 1968.

[3] D. E. Knuth, *Sequences with precisely k+1 k–blocks*, Solution to problem E2307, Amer. Math. Monthly, 79(1972), 773-774.

[4] D. E. Knuth, J. H. Morris and V. R. Pratt, *Fast pattern matching in strings*, SIAM J. Comput., 6, 1977, 323–350.

[5] M. Lothaire, *Combinatorics on words*, Addison-Wesley, 1983.

[6] G. Pirillo, *Infinite words and biprefix codes*, Information Processing Letters, 50 (1994) 293-295.