Advances in Applied Mathematics 90 (2017) 27–45

Contents lists available at ScienceDirect

Advances in Applied Mathematics

www.elsevier.com/locate/yaama

The sequence of open and closed prefixes of a Sturmian word ☆

Alessandro De Luca [a], Gabriele Fici [b,*], Luca Q. Zamboni [c]

[a] *DIETI, Università degli Studi di Napoli Federico II, Italy*
[b] *Dipartimento di Matematica e Informatica, Università di Palermo, Italy*
[c] *Université Claude Bernard Lyon 1, France*

ARTICLE INFO

Article history:
Received 28 December 2016
Received in revised form 22 April 2017
Accepted 23 April 2017
Available online xxxx

MSC:
68R15

Keywords:
Sturmian word
Closed word

ABSTRACT

A finite word is closed if it contains a factor that occurs both as a prefix and as a suffix but does not have internal occurrences, otherwise it is open. We are interested in the *oc-sequence* of a word, which is the binary sequence whose n-th element is 0 if the prefix of length n of the word is open, or 1 if it is closed. We exhibit results showing that this sequence is deeply related to the combinatorial and periodic structure of a word. In the case of Sturmian words, we show that these are uniquely determined (up to renaming letters) by their oc-sequence. Moreover, we prove that the class of finite Sturmian words is a maximal element with this property in the class of binary factorial languages. We then discuss several aspects of Sturmian words that can be expressed through this sequence. Finally, we provide a linear-time algorithm that computes the oc-sequence of a finite word, and a linear-time algorithm that reconstructs a finite Sturmian word from its oc-sequence.

☆ Some of the results contained in this paper were presented at the 9th International Conference on Words, WORDS 2013 [9].

* Corresponding author.
E-mail addresses: alessandro.deluca@unina.it (A. De Luca), gabriele.fici@unipa.it (G. Fici), lupastis@gmail.com (L.Q. Zamboni).

http://dx.doi.org/10.1016/j.aam.2017.04.007

1. Introduction

In a recent paper with M. Bucci [5], the first two authors dealt with trapezoidal words (a generalization of finite Sturmian words), also with respect to the property of being closed or open. Let Σ be a finite nonempty set (the alphabet). A (finite) word $w = w[1]w[2]\cdots w[n]$ with $w[i] \in \Sigma$ is *closed* (also known as *periodic-like* [6]) if it contains a factor that occurs both as a prefix and as a suffix but does not have internal occurrences, otherwise it is *open*. For example, the words $abca$, $ababa$ and $aabaab$ are closed — any word of length 1 is closed, the empty word being a factor that occurs both as a prefix and as a suffix but does not have internal occurrences; the words ab, aab and $aaba$, instead, are open.

Given a finite or infinite word $w = w[1]w[2]\cdots$, the sequence $\mathrm{oc}(w)$ of open/closed prefixes of w, that we refer to as the *oc-sequence* of w, is the binary sequence $c(1)c(2)\cdots$ whose n-th element is 1 if the prefix of w of length n is closed, 0 if it is open. For example, if $w = abcab$, then $\mathrm{oc}(w) = 10011$.

A question that arises naturally is whether it is possible to reconstruct a word (up to renaming letters) from its oc-sequence. This is not true in general, even when the alphabet is binary. For example, the words $aaba$ and $aabb$ are not isomorphic (i.e., one cannot be obtained from the other by renaming letters), yet they have the same oc-sequence 1100. As a first result of this paper, we show that if a word is known to be Sturmian, then it can be reconstructed (up to renaming letters) from its oc-sequence. That is, Sturmian words are characterized by their oc-sequences. Moreover, we prove that the class of finite Sturmian words is a maximal element with this property in the class of binary factorial languages.

In [5], the authors investigated the structure of the sequence $\mathrm{oc}(F)$ of the Fibonacci word F. They proved that the lengths of the runs (maximal subsequences of consecutive equal elements) in $\mathrm{oc}(F)$ form the doubled Fibonacci sequence. We prove in this paper that this doubling property holds for every standard Sturmian word, and describe the sequence $\mathrm{oc}(w)$ of a standard Sturmian word w in terms of the *semicentral* prefixes of w, which are the prefixes of the form $u_n xy u_n$, where x, y are letters and $u_n xy$ is an element of the standard sequence of w. As a consequence, we show that the word $ba^{-1}w$, obtained from a standard Sturmian word w starting with letter a by replacing the first letter with a b, can be written as the infinite product of the words $(u_n^{-1}u_{n+1})^2$, $n \geq 0$. Since the words $u_n^{-1}u_{n+1}$ are reversals of standard words, this induces an infinite factorization of $ba^{-1}w$ in squares of reversed standard words.

We then show how the oc-sequence of a standard Sturmian word of slope α is related to the continued fraction expansion of α, both in terms of the convergents and of the continuants of α.

Finally, we provide a linear-time algorithm that computes the oc-sequence of a finite word, and a linear-time algorithm that reconstructs a finite Sturmian word from its oc-sequence.