Journal of Combinatorial Theory, Series A 145 (2017) 36–56

Contents lists available at ScienceDirect

Journal of Combinatorial Theory, Series A

www.elsevier.com/locate/jcta

Cyclic complexity of words ☆

Julien Cassaigne [a], Gabriele Fici [b], Marinella Sciortino [b], Luca Q. Zamboni [c]

[a] *Aix-Marseille Univ, Institut de Mathématiques de Marseille, CNRS UMR 7373, Campus de Luminy, F-13288 Marseille cedex 9, France*

[b] *Dipartimento di Matematica e Informatica, Università di Palermo, Via Archirafi 34, 90123 Palermo, Italy*

[c] *Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, F-69622 Villeurbanne cedex, France*

ARTICLE INFO

Article history:
Received 29 May 2015
Available online 12 August 2016

Keywords:
Cyclic complexity
Factor complexity
Sturmian words

ABSTRACT

We introduce and study a complexity function on words $c_x(n)$, called *cyclic complexity*, which counts the number of conjugacy classes of factors of length n of an infinite word x. We extend the well-known Morse–Hedlund theorem to the setting of cyclic complexity by showing that a word is ultimately periodic if and only if it has bounded cyclic complexity. Unlike most complexity functions, cyclic complexity distinguishes between Sturmian words of different slopes. We prove that if x is a Sturmian word and y is a word having the same cyclic complexity of x, then up to renaming letters, x and y have the same set of factors. In particular, y is also Sturmian of slope equal to that of x. Since $c_x(n) = 1$ for some $n \geq 1$ implies x is periodic, it is natural to consider the quantity $\liminf_{n\to\infty} c_x(n)$. We show that if x is a Sturmian word, then $\liminf_{n\to\infty} c_x(n) = 2$. We prove however that this is not a characterization of Sturmian words by exhibiting a restricted class of Toeplitz words, including the period-doubling word, which also verify this same condition on the

☆ Some of the results in this paper were presented at the 39th International Symposium on Mathematical Foundations of Computer Science, MFCS 2014 [6].

E-mail addresses: julien.cassaigne@math.cnrs.fr (J. Cassaigne), gabriele.fici@unipa.it (G. Fici), marinella.sciortino@unipa.it (M. Sciortino), zamboni@math.univ-lyon1.fr (L.Q. Zamboni).

http://dx.doi.org/10.1016/j.jcta.2016.07.002

limit infimum. In contrast we show that, for the Thue–Morse word t, $\liminf_{n\to\infty} c_t(n) = +\infty$.

1. Introduction

The factor complexity $p_x(n)$ of an infinite word $x = x_0x_1x_2\cdots \in A^{\mathbb{N}}$ (with each x_i belonging to a finite nonempty alphabet A) counts the number of distinct factors $x_ix_{i+1}\cdots x_{i+n-1}$ of length n occurring in x. It provides a measure of the extent of randomness of the word x and more generally of the subshift generated by x. Periodic words have bounded factor complexity while digit expansions of normal numbers have full complexity. A celebrated result of Hedlund and Morse in [16] states that every non-periodic word contains at least $n+1$ distinct factors of each length n. Moreover, there exist words satisfying $p_x(n) = n+1$ for each $n \geq 1$. These words are called Sturmian words, and in terms of their factor complexity, are regarded to be the simplest non-periodic words.

Sturmian words admit many different characterizations of combinatorial, geometric and arithmetic nature. In the 1940's, Hedlund and Morse showed that each Sturmian word is the symbolic coding of the orbit of a point x on the unit circle under a rotation by an irrational angle θ, called the slope, where the circle is partitioned into two complementary intervals, one of length θ and the other of length $1-\theta$. Conversely, each such coding defines a Sturmian word. It is well known that the dynamical/ergodic properties of the system, as well as the combinatorial properties of the associated Sturmian word, hinge on the arithmetical/Diophantine qualities of the slope θ given by its continued fraction expansion. Sturmian words arise naturally in various branches of mathematics including combinatorics, algebra, number theory, ergodic theory, dynamical systems and differential equations. They also have implications in theoretical physics as 1-dimensional models of quasi-crystals.

Other measures of complexity of words have been introduced and studied in the literature, including abelian complexity, maximal pattern complexity, k-abelian complexity, binomial complexity, periodicity complexity, minimal forbidden factor complexity and palindromic complexity. With respect to most word complexity functions, Sturmian words are characterized as those non-periodic words of lowest complexity. One exception to this occurs in the context of maximal pattern complexity introduced by Kamae in [11]. In this case, while Sturmian words are *pattern Sturmian*, meaning that they have minimal maximal pattern complexity amongst all non-periodic words, they are not the only ones. In fact, a certain restricted class of Toeplitz words which includes the period-doubling word are also known to be pattern Sturmian (see [12]). On the other hand, the Thue–Morse word, while closely connected to the period-doubling word, is known to have full maximal pattern complexity (see Example 1 in [11]).