Theoretical Computer Science 635 (2016) 16–34

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Abelian powers and repetitions in Sturmian words ☆

Gabriele Fici [a,*], Alessio Langiu [b], Thierry Lecroq [c], Arnaud Lefebvre [c], Filippo Mignosi [d], Jarkko Peltomäki [e], Élise Prieur-Gaston [c]

[a] Dipartimento di Matematica e Informatica, Università di Palermo, Italy
[b] Department of Informatics, King's College London, UK
[c] LITIS EA4108, FR CNRS 3638 Normastic, Normandie Université, Université de Rouen, France
[d] DISIM, Università dell'Aquila, Italy
[e] Turku Centre for Computer Science, TUCS, Department of Mathematics and Statistics, University of Turku, Finland

ARTICLE INFO

Article history:
Received 12 January 2016
Received in revised form 15 April 2016
Accepted 28 April 2016
Available online 6 May 2016
Communicated by D. Perrin

Keywords:
Sturmian word
Abelian power
Abelian period
Lagrange constant
Critical exponent

ABSTRACT

Richomme, Saari and Zamboni (2011) [39] proved that at every position of a Sturmian word starts an abelian power of exponent k for every $k > 0$. We improve on this result by studying the maximum exponents of abelian powers and abelian repetitions (an abelian repetition is an analogue of a fractional power) in Sturmian words. We give a formula for computing the maximum exponent of an abelian power of abelian period m starting at a given position in any Sturmian word of rotation angle α. By considering all possible abelian periods m, we recover the result of Richomme, Saari and Zamboni.

As an analogue of the critical exponent, we introduce the abelian critical exponent $A(s_\alpha)$ of a Sturmian word s_α of angle α as the quantity $A(s_\alpha) = \limsup k_m/m = \limsup k'_m/m$, where k_m (resp. k'_m) denotes the maximum exponent of an abelian power (resp. of an abelian repetition) of abelian period m (the superior limits coincide for Sturmian words). We show that $A(s_\alpha)$ equals the Lagrange constant of the number α. This yields a formula for computing $A(s_\alpha)$ in terms of the partial quotients of the continued fraction expansion of α. Using this formula, we prove that $A(s_\alpha) \geq \sqrt{5}$ and that the equality holds for the Fibonacci word. We further prove that $A(s_\alpha)$ is finite if and only if α has bounded partial quotients, that is, if and only if s_α is β-power-free for some real number β.

Concerning the infinite Fibonacci word, we prove that: *i*) The longest prefix that is an abelian repetition of period F_j, $j > 1$, has length $F_j(F_{j+1} + F_{j-1} + 1) - 2$ if j is even or $F_j(F_{j+1} + F_{j-1}) - 2$ if j is odd, where F_j is the jth Fibonacci number; *ii*) The minimum abelian period of any factor is a Fibonacci number. Further, we derive a formula for the minimum abelian periods of the finite Fibonacci words: we prove that for $j \geq 3$ the Fibonacci word f_j, of length F_j, has minimum abelian period equal to $F_{\lfloor j/2 \rfloor}$ if $j = 0, 1, 2 \mod 4$ or to $F_{1+\lfloor j/2 \rfloor}$ if $j = 3 \mod 4$.

☆ Some of the results in this paper have been presented without proofs at the 17th International Conference on Developments in Language Theory, DLT 2013 [19].

* Corresponding author.
E-mail addresses: gabriele.fici@unipa.it (G. Fici), alessio.langiu@kcl.ac.uk (A. Langiu), thierry.lecroq@univ-rouen.fr (T. Lecroq), arnaud.lefebvre@univ-rouen.fr (A. Lefebvre), filippo.mignosi@di.univaq.it (F. Mignosi), jspelt@utu.fi (J. Peltomäki), elise.prieur@univ-rouen.fr (É. Prieur-Gaston).

http://dx.doi.org/10.1016/j.tcs.2016.04.039
0304-3975/

1. Introduction

Sturmian words are infinite words having exactly $n+1$ distinct factors of each length $n \geq 0$. By the celebrated theorem of Morse and Hedlund [33], they are the aperiodic binary words with minimal factor complexity. Every Sturmian word is characterized by an irrational number α and a real number ρ called the *angle* and the *initial point* respectively. The Sturmian word $s_{\alpha,\rho}$ is defined by rotating the point ρ by the angle α in the torus $I = \mathbb{R}/\mathbb{Z} = [0,1)$ and by writing a letter **b** when the point falls in the interval $[0, 1-\alpha)$ and a letter **a** when the point falls in the complement. The Fibonacci word $f =$ **abaababaabaababaabab** $\cdots$ is a well-known Sturmian word obtained by taking both the angle and the initial point equal to $\varphi - 1$, where $\varphi = (1+\sqrt{5})/2$ is the Golden Ratio. The Fibonacci word f is also the limit of the sequence of finite Fibonacci words f_n, defined by $f_0 = \mathbf{b}$, $f_1 = \mathbf{a}$ and $f_j = f_{j-1} f_{j-2}$ for every $j > 1$, that are the natural counterpart of the Fibonacci numbers in the setting of words.

Sturmian words have several equivalent definitions and a lot of combinatorial properties that make them well-studied objects in discrete mathematics and theoretical computer science. In fact, there exists a huge bibliography on Sturmian words (see for instance the survey papers [6,7], [29, Chap. 2], [38, Chap. 6] and the references therein).

There are mainly two approaches to Sturmian words: one is purely combinatorial, while the other uses techniques from elementary number theory to derive correspondences between the finer arithmetic properties of the irrational α and the factors of the Sturmian words of angle α. In the language of computer science, such correspondences are called *semantics*. In this paper, we aim at building upon such approach by showing new semantics allowing us to give new and tight results on the abelian combinatorics of Sturmian words. Indeed, this approach extends to the abelian setting the well-known fruitful semantics that in the last decades have allowed researchers to derive deep and important results on the combinatorics of infinite words from the theory of codings of rotations and continued fractions of irrationals. Interestingly these semantics also allowed researchers to shed new light on consolidated theories by exploiting the opposite direction. A remarkable example of this is represented by the work of B. Adamczewski and Y. Bugeaud [1,2].

Concerning the maximum exponent of repetitions in Sturmian words, there exists a vast bibliography (see for example [5,8,10,16,25,27,42] and the references therein), which stems from the seminal work on the Fibonacci word presented in [32]. Indeed, the study of repetitions in words is a classical subject both from the combinatorial and the algorithmic point of view. Repetitions are strictly related to the notion of periodicity. Recall that a word w of length $|w|$ has a *period* $p > 0$ if $w_i = w_{i+p}$ for every $1 \leq i \leq |w| - p$, where w_i is the letter in the position i of w. The *exponent* of w is the ratio $|w|/\pi_w$ between its length $|w|$ and its *minimum period* π_w. When studying the degree of repetitiveness of a word, we are often interested in the factors whose exponent is at least 2, called *repetitions*. Repetitions whose exponent is an integer are called *integer powers* since a word w with integer exponent $k \geq 2$ can be written as $w = u^k$, i.e., w is the concatenation of k copies of a non-empty word u of length π_w. If instead k is not an integer, then the word w is a *fractional power*. In this case we can write $w = u^{\lfloor k \rfloor} u'$, where u' is the prefix of u of length $\pi_w(k - \lfloor k \rfloor)$. For example, the word $w = aabaaba$ is a 7/3-power since it has minimum period 3 and length 7. A good reference on periodicity is [29, Chap. 8].

A measure of repetitiveness of an infinite word is given by the supremum of the exponents of its factors, called the *critical exponent* of the word. If this supremum β is finite, then the word is said to be β^+-power-free (or simply β-power-free if β is irrational, so there are no factors with exponent β). For example, the critical exponent of the Fibonacci word f is $2+\varphi$ [32], so f is $(2+\varphi)$-power-free. In general, a Sturmian word $s_{\alpha,\rho}$ is β-power-free for some β if and only if the continued fraction expansion of α has bounded partial quotients [31]. The critical exponent of $s_{\alpha,\rho}$ can be explicitly determined by a formula involving these partial quotients [10,16,25,35].

Recently, the extension of these notions to the so-called abelian setting has received a lot of interest. Abelian properties of words have been studied since the very beginning of formal language theory and combinatorics on words. The notion of the Parikh vector of a word (see later for its definition) has become a standard and is often used without an explicit reference to the original 1966 paper by Parikh [34]. Abelian powers were first considered in 1957 by Erdős [18] as a natural generalization of usual integer powers. Research concerning abelian properties of words and languages developed afterwards in different directions. For instance, there is an increasing interest in abelian properties of words linked to periodicity; see, for example, [4,11,12,17,37,39,41].

Recall that the Parikh vector $\mathcal{P}_w$ of a finite word w enumerates the total number of each letter of the alphabet in w. Therefore, two words have the same Parikh vector if and only if one can be obtained from the other by permuting letters. An *abelian decomposition* of a word w is a factorization $w = u_0 u_1 \cdots u_{j-1} u_j$, where $j \geq 2$, the words $u_1, \ldots, u_{j-1}$ have the same Parikh vector $\mathcal{P}$ and the Parikh vectors of the words u_0 and u_j are contained in $\mathcal{P}$ (that is, they are component-wise less or equal to $\mathcal{P}$ but not equal to $\mathcal{P}$). The sum m of the components of the Parikh vector $\mathcal{P}$ (that is, the length of u_1) is called an *abelian period* of w (cf. [12]). The words u_0 and u_j, the first and the last factor of the decomposition, are respectively called the *head* and the *tail* of the abelian decomposition. Notice that different abelian decompositions can give the same abelian period. For example, the word $w = abab$ has an abelian period 2 with $u_0 = \varepsilon$ (the empty word), $u_1 = u_2 = ab$ and $u_3 = \varepsilon$ or with $u_0 = a$, $u_1 = ba$ and $u_2 = b$. The *abelian exponent* of w is the ratio $|w|/\mu_w$ between its length $|w|$ and its minimum abelian period μ_w. We say that a word w is an *abelian repetition* of period m and exponent k if it has an abelian decomposition $w = u_0 u_1 \cdots u_{j-1} u_j$ with $j \geq 3$ (so that u_1 and u_2 exist and are nonempty) such that $|u_1| = m$ and $|w|/m = k$. If we are uninterested in the period and exponent, then we simply call w an abelian repetition. An *abelian power* (also known as a *weak repetition* [13]) is a word w that has an abelian decomposition with empty head and empty tail. Let m be the abelian period of w corresponding to the decomposition. Then we say that the word w is an