Contents lists available at ScienceDirect



Theoretical Computer Science

www.elsevier.com/locate/tcs



tical ter Science

CrossMark

Studies on finite Sturmian words

Christophe Reutenauer

Laboratoire de combinatoire et d'informatique mathématique, Université du Québec à Montréal, Case postale 8888, succursale Centre-ville, Montréal, Québec H3C 3P8, Canada

ARTICLE INFO

Article history: Received 10 November 2014 Received in revised form 1 May 2015 Accepted 3 May 2015 Available online 7 May 2015 Communicated by D. Perrin

Keywords: Finite Sturmian words Christoffel words Sturmian sequences Palindromes Special words Rauzy graph Periods Conjugation Normal form Circular factors Enumeration Markoff numbers

ABSTRACT

Several properties of finite Sturmian words are proved. The inverse of the Richomme-Séébold bijection between factor sets of given length of Sturmian sequences and left special Sturmian words is constructed (using circular factors of Christoffel words) and studied. The Labbé bijection between Christoffel words and left special Sturmian words is introduced and studied. Factor sets are classified by two types: the two classes are characterized by conjugation and periodicity properties. The two classes are related to the Rauzy graph. A normal form for finite Sturmian words is given, following a result of de Luca and De Luca. Several classes of Sturmian words are characterized through this normal form: left or right special words, bispecial words, palindromes, central words. The normal form is used to derive a proof of the Lipatov-Mignosi formula for the number of Sturmian words of given length: to produce a linear algorithm which checks if a word is Sturmian and which computes its normal form; to construct algorithmically the contraction and completion in various classes of Sturmian words; to compute the de Luca palindromic closure of any Sturmian word and prove that it is Sturmian. The completion is related to the de Luca palindromization function and it is shown that the sets of directive words of central words of even and odd length are rational group languages. Christoffel words are characterized in several ways: an extension of a result of Chuan; an extension of a result of Droubay-Pirillo, by counting circular factors; by counting Lyndon factors. A special Christoffel word is constructed having length a given Markoff number. A bivariate count of special Sturmian words is given, using ideas of Bédaride, Domenjoud, Jamet and Rémy. Open problems are given.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Sturmians sequences appear in several areas of pure and applied mathematics. They appear, without the name, in Markoff's 1880 work on minima of quadratic forms [27,45,46,57], in the 1940 work of Morse and Hedlund in symbolic dynamics, who gave them the name *Sturmian* [52], in arithmetic under their variant called *Beatty sequences* [19] or *spectra* [17,33]. They are the nonperiodic sequences of minimal complexity [38]: for a fixed sequence, this is the function $\mathbb{N} \to \mathbb{N}$ which with *n* associates the number of factors of the sequence of length *n*; for a Sturmian sequence, this number is *n* + 1.

A factor of a Sturmian sequence is called a *finite Sturmian word* or simply a *Sturmian word* (in the present article, a word is always a finite word, and an infinite word will be called a sequence). Sturmian words appear even before Sturmian sequences, since they appear in the 1875 article of Christoffel [21] (probably one of the last mathematical article written in Latin). In Theoretical Computer Science, Sturmian words appear, under their paradigmatic example, the Fibonacci word,

http://dx.doi.org/10.1016/j.tcs.2015.05.003 0304-3975/© 2015 Elsevier B.V. All rights reserved.

E-mail address: Reutenauer.Christophe@uqam.ca.

in the precursory work of Berstel [5] and Aldo de Luca [39]; in particular, the fundamental notions of special factors and palindromes. The terminology "Christoffel words" was introduced by Jean Berstel [6], where is shown among other properties that all the discretizations of a segment, that exist in the literature, are conjugates words. See [2,38,55] for many references and results on Sturmian sequences and words.

The first part of the present article is motivated by the work [10]: a byproduct of the latter article is that the set of factors of given length n in a fixed Sturmian sequence (call such a set a *factor set*) is a basis of the subgroup of the free group F_2 which is the kernel of the homomorphism sending an element onto its algebraic length (that is, letters are sent onto 1 and their inverses onto -1) taken modulo n. Thus, the fact that a factor set is of cardinality n + 1, known since the work of Morse and Hedlund, appears as a very special case of the Schreier formula (which gives the rank of a subgroup of finite index of a free group). It was therefore tempting to give a more precise description of the factor sets. The number of factor sets of a given length n has been recently given by Richomme and Séébold [58]: it is $\varphi(1) + \cdots + \varphi(n)$, where φ is the Euler function. They show that there is a bijection between factor sets of length n and left special words of length n - 1; the bijection is defined by $F \mapsto l$, with $al, bl \in F$. We give here the inverse bijection, using the circular factors of a Christoffel word constructed from the left special word by the palindromic closure of de Luca (see Proposition 5.1). Several consequences of these bijections are derived, in particular the Labbé mapping (see Section 6).

In order to describe factor sets, we classify them into two classes. A factor set is of type C if it is the union of a conjugation class and another word; factor sets of this type appear already in the work of Zhi Xiong Wen and Zhi Ying Wen [62]. The other factor sets are called of type P. The interesting fact is that one may characterize these classes; in particular, type P is characterized by a property of periodicity (Proposition 7.3 and Corollary 7.5); the classes of each type may be counted, refining the Richomme–Séébold count, Corollary 7.4. Another result shows that this classification appears in the factor (Rauzy) graph, see Section 10.

In the middle part of the article, we give a normal form for Sturmian words, following a result of de Luca and De Luca [41]: they show that the minimal period of a Sturmian word is the length of some Christoffel factor of this word, a result in the spirit of a conjecture of Duval [30]. This normal form (Proposition 11.1) is very natural and rests on the palindromic factorization of the corresponding Christoffel word. It has many applications. The first one is a proof of the Lipatov–Mignosi formula giving the number of Sturmian words of a given length (Corollary 12.2); we use for its proof an amusing lemma on the number of matrices in $SL_2(\mathbb{N})$ of given total sum (Lemma 12.3). As another application of the normal form, we give characterizations of several classes of Sturmian words: palindromes, special, bispecial, central (Proposition 11.2). Moreover, we use this to characterize the completion and contraction in several classes of Sturmian words: for example, compute the longest palindrome which is the median factor of a given Sturmian palindrome w and the shortest Sturmian palindrome of which w is a median factor, reproving a result of de Luca and De Luca [42], and similar results for Sturmian and special Sturmian words (Section 14). A consequence of these constructions is the result, which seems new, that the palindromic closure of a Sturmian word is Sturmian; there exist several algorithms in the literature (see for instance [41] and the references therein), but the present one seems conceptually particularly simple: it rests on Duval's algorithm [31] which gives in linear time the factorization of a word into Lyndon words, and it is therefore linear (Section 13).

In the sections which follow, I study several properties of Christoffel words and their conjugation classes (Christoffel classes). Using the notion of *moment* of a word, Wai-Fong Chuan characterizes Christoffel classes [24,25], see also Theorem 16.1; we give a variant of her result, by using the area under the geometric realization of a word, see Proposition 16.2. In the next section, we extend to finite words the characterization by Xavier Droubay and Giuseppe Pirillo of Sturmian sequences, by their palindrome count [28], see also Theorem 17.1; the proof of this finitary version of their result is somewhat similar to their's, see Proposition 17.2. In the next section, I give several results on Christoffel factors of Christoffel words: their position, their number and the number of their occurrences, see Proposition 18.1, Corollary 18.4 and Corollary 18.5. After that, I give a characterization of Markoff numbers, Proposition 19.1, which states that such a number is the length of a special Christoffel word, whose directive word is antipalindromic; it allows to state a problem equivalent to the Markoff injectivity conjecture.

In Section 20, I want to compute the bivariate symmetric function associated with several classes of Sturmian words: all Sturmian words, palindromes, special and bispecial words. This is motivated by the fact that a bivariate count of Sturmian words and of Sturmian palindromes has been performed by the authors of [3]; they give among other results a recursive formula to count the number of Sturmian words, and of Sturmian palindromes, with a given number of *a*'s and *b*'s. Their method may be used for the set of special words, in particular using their function θ , already used in Section 15. In Section 21, tables of numbers are given and in Section 22, several open problems are given. In particular, find a closed formula for the bivariate symmetric function for the classes of Sturmian words; I could do this only for central words, Christoffel words, and bispecial words (following a formula of Fici [32]); the case of all Sturmian words (that is, find a bivariate symmetric function analogue of the Lipatov–Mignosi formula), special words and palindromes is open.

2. Periods and periodic patterns

Recall that a word $w = a_1 \cdots a_n$ has the *period* p if whenever $i, i + p \in \{1, \dots, n\}$, one has $a_i = a_{i+p}$. The period is *nontrivial* if p < n and we say that w is *periodic* if it has a nontrivial period p.

Download English Version:

https://daneshyari.com/en/article/434045

Download Persian Version:

https://daneshyari.com/article/434045

Daneshyari.com