Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Specular sets

Valérie Berthé [a], Clelia De Felice [b], Vincent Delecroix [c], Francesco Dolce [d,*], Julien Leroy [e], Dominique Perrin [f], Christophe Reutenauer [d], Giuseppina Rindone [f]

[a] CNRS, Université Paris Diderot, France
[b] Università degli Studi di Salerno, Italy
[c] CNRS, Université de Bordeaux, France
[d] Université du Québec à Montréal, Canada
[e] Université de Liège, Belgium
[f] Université Paris-Est, France

ARTICLE INFO

Article history:
Received 30 May 2016
Received in revised form 28 February 2017
Accepted 2 March 2017
Available online xxxx

Keywords:
Tree sets
Return words
Bifix codes
Linear involutions
Specular groups
Free group

ABSTRACT

We introduce specular sets. These are subsets of groups which form a natural generalization of free groups. These sets of words are an abstract generalization of the natural codings of interval exchanges and of linear involutions. We consider two important families of sets contained in specular sets: sets of return words and maximal bifix codes. For both families we prove several cardinality results as well as results concerning the subgroup generated by these sets.

1. Introduction

We have studied in a series of papers initiated in [4] the links between uniformly recurrent languages, subgroups of free groups and bifix codes. In this paper, we continue this investigation in a situation which involves groups which are not free anymore. These groups, named here specular, are free products of a free group and of a finite number of cyclic groups of order two. These groups are close to free groups and, in particular, the notion of a basis in such groups is clearly defined. It follows from the Kurosh subgroup theorem that any subgroup of a specular group is specular.

A specular set is a subset of such a group which generalizes the natural codings of linear involutions studied in [10].

The extension graph of a word w with respect to a set of words S is the bipartite graph with vertices the disjoint union of left- and right-extensions of w in S, and edges the corresponding bi-extensions in S.

A specular set can be seen as a set of words stable by taking the inverse and defined in terms of restrictions on the extensions of its elements. More precisely, a specular set has the property that the extension graph of every nonempty word is a tree and the extension graph of the nonempty word is a union of two disjoint trees.

* Corresponding author.
E-mail address: francesco.dolce@lacim.ca (F. Dolce).

http://dx.doi.org/10.1016/j.tcs.2017.03.001
0304-3975/© 2017 Published by Elsevier B.V.

Specular sets extend the notion of tree sets developed in [7] and [20] that encompass Sturmian words, Arnoux–Rauzy words or else natural codings of interval exchanges. Tree sets have striking combinatorial and algebraic properties that we extend here.

The main results of this paper are Theorems 6.15 and 8.1, referred to as the First Return Theorem and the Finite Index Basis Theorem. The first one asserts that the set of return words to a given word in a recurrent specular set is a basis of a subgroup of index 2, called the even subgroup. The last one characterizes the symmetric bases of subgroups of finite index of specular groups contained in a specular set S as the finite S-maximal symmetric bifix codes contained in S. This generalizes the analogous result proved initially in [4] for Sturmian sets and extended in [8] to a more general class of sets, containing both Sturmian sets and interval exchange sets.

There are two interesting features of the subject of this paper.

In the first place, some of the statements concerning the natural codings of linear involutions can be proved using geometric methods, as shown in a separate paper [10]. This provides an interesting interpretation of the groups playing a role in the natural codings (these groups are generated either by return words or by maximal bifix codes) as fundamental groups of some surfaces. The methods used here are, however, purely combinatorial.

In the second place, the abstract notion of specular set gives rise to groups called here specular. These groups are natural generalizations of free groups, and are free products of a finite number of copies of $\mathbb{Z}$ and of $\mathbb{Z}/2\mathbb{Z}$. They are called *free-like* in [2], appear at several places in [17] and are well-known in the Bass–Serre theory (see [33,18]).

The idea of considering recurrent sets of reduced words invariant by taking inverses is connected with the notion of G-full words of [32] (see Section 4.5).

The paper is organized as follows. In Section 2, we recall some notions concerning words, extension graphs and bifix codes. We define the notion of characteristic which is the Euler characteristic of the extension graph of the empty word. We consider tree sets of characteristic 1 or 2 (tree sets of characteristic 1 are introduced in [7], while the case of arbitrary characteristic is treated in [20]).

In Section 3, we introduce specular groups, which form a family with properties very close to free groups. We deduce from the Kurosh subgroup theorem that any subgroup of a specular group is specular (Theorem 3.3). Actually (as pointed out to us by a referee), specular groups can be studied as groups acting on trees as developed in the Bass–Serre theory [33].

In Section 4 we introduce specular sets. We recall several results from [19] and [20] concerning the cardinality of some sets included in neutral sets, namely bifix codes (Theorems 4.15 and 4.16). We give a construction which allows to build specular sets from a tree set of characteristic 1 using a transducer called doubling transducer (Theorem 4.20). We make a connection with the notion of G-full words introduced in [32] and related to the palindromic complexity of [21].

In Section 5 we recall the definition of a linear involution introduced in [15] and we show that the natural coding of a linear involution without connections is a specular set (Theorem 5.9).

In Section 6 we introduce three variants of the notion of set of return words. We prove several cardinality results concerning these sets (Theorems 6.6, 6.9, 6.12). We prove that the set of return words to a given word forms a basis of the even subgroup (Theorem 6.15 referred to as the First Return Theorem) and that the mixed return words form a monoidal basis of the specular group (Theorem 6.17).

In Section 7 we prove several results concerning subgroups generated by bifix codes. We prove that a set closed by taking inverses is acyclic if and only if any symmetric bifix code is free (Theorem 7.1). Moreover, we prove that in such a set, for any finite symmetric bifix code X, the free monoid X^* and the free subgroup $\langle X\rangle$ have the same intersection with S (Theorem 7.8).

Finally, in Section 8, we prove the Finite Index Basis Theorem (Theorem 8.1) and a converse (Theorem 8.6).

This paper is an extended version of a conference paper [6].

2. Preliminaries

In this section, we first recall some notions on sets of words including recurrent, uniformly recurrent and tree sets. We also recall some definitions and properties concerning bifix codes.

2.1. Extension graphs

Let A be a finite alphabet. We denote by A^* the free monoid on A. We denote by ε the empty word. The *reversal* of a word $w = a_1a_2\cdots a_n$ with $a_i \in A$ is the word $\tilde{w} = a_n \cdots a_2a_1$. A word w is said to be a *palindrome* if $w = \tilde{w}$.

A set of words on the alphabet A is said to be *factorial* if it contains the alphabet A and all the factors of its elements.

An *internal factor* of a word x is a word v such that $x = uvw$ with u, w nonempty.

Let S be a set of words on the alphabet A. For $w \in S$, we denote

$$L_S(w) = \{a \in A \mid aw \in S\}$$

$$R_S(w) = \{a \in A \mid wa \in S\}$$

$$B_S(w) = \{(a,b) \in A \times A \mid awb \in S\}$$