Specular sets

Valérie Berthé[1], Clelia De Felice[2], Vincent Delecroix[3],
Francesco Dolce[4], Julien Leroy[5], Dominique Perrin[4],
Christophe Reutenauer[6], Giuseppina Rindone[4]

[1] CNRS, Université Paris Diderot
[2] Università degli Studi di Salerno
[3] CNRS, Université de Bordeaux
[4] Université Paris-Est
[5] Université de Liège
[6] Université du Québec à Montréal

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.

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

Contents

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-extension of w in S and edges the corresponding bi-extension 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.

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] et 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].

Acknowledgments. The authors thank Laurent Bartholdi and Pierre de la Harpe for useful indications. We want also want to thank the anonymous referees for their careful reading and for pointing out the connection between our work and the Bass-Serre theory. This work was supported by grants from Région Île-de-France, the ANR projects Dyna3S and Eqinocs.

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 $w = a_1 a_2 \cdots a_n$ with $a_i \in A$ is the word $\tilde{w} = a_n \cdots a_2 a_1$. 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

$$\begin{aligned} L_S(w) &= \{a \in A \mid aw \in S\} \\ R_S(w) &= \{a \in A \mid wa \in S\} \\ B_S(w) &= \{(a, b) \in A \times A \mid awb \in S\} \end{aligned}$$

and further

$$\ell_S(w) = \mathrm{Card}(L_S(w)), \quad r_S(w) = \mathrm{Card}(R_S(w)), \quad b_S(w) = \mathrm{Card}(B_S(w)).$$

We omit the subscript S when it is clear from the context. A word w is *right-extendable* if $r(w) > 0$, *left-extendable* if $\ell(w) > 0$ and *bi-extendable* if $b(w) > 0$. A factorial set S is called *right-extendable* (resp. *left-extendable*, resp. *bi-extendable*) if every word in S is right-extendable (resp. left-extendable, resp. bi-extendable).

A word w is called *right-special* if $r(w) \geq 2$. It is called *left-special* if $\ell(w) \geq 2$. It is called *bi-special* if it is both left-special and right-special.

For $w \in S$, we denote

$$m_S(w) = b_S(w) - \ell_S(w) - r_S(w) + 1.$$

The word w is called *weak* if $m_S(w) < 0$, *neutral* if $m_S(w) = 0$ and *strong* if $m_S(w) > 0$.

We say that a factorial set S is *neutral* if every nonempty word in S is neutral. The *characteristic* of S is the integer $\chi(S) = 1 - m_S(\varepsilon)$. Thus a neutral set of characteristic 1 is such that all words (including the empty word) are neutral. This what is called a neutral set in [7].

A set of words $S \neq \{\varepsilon\}$ is *recurrent* if it is factorial and if for any $u, w \in S$, there is a $v \in S$ such that $uvw \in S$. An infinite factorial set is said to be *uniformly recurrent* if for any word $u \in S$ there is an integer $n \geq 1$ such that u is a factor of any word of S of length n. A uniformly recurrent set is recurrent.

In [20] it is proved that the converse is true for neutral sets. As all sets we will deal with are neutral, we usually omit the term "uniformly" and just mention whenever we suppose them to be recurrent.

Theorem 2.1 ([20]) *A recurrent neutral set is uniformly recurrent*

The *factor complexity* of a factorial set S of words on an alphabet A is the sequence $p_n = \mathrm{Card}(S \cap A^n)$. Let $s_n = p_{n+1} - p_n$ and $t_n = s_{n+1} - s_n$ be respectively the first and second order differences sequences of the sequence p_n.

The following result is from [13] (see also [11], Theorem 4.5.4).

Proposition 2.2 *Let S be a factorial set on the alphabet A. One has $t_n = \sum_{w \in S \cap A^n} m(w)$ and $s_n = \sum_{w \in S \cap A^n}(r(w) - 1)$ for all $n \geq 0$.*

Let S be a bi-extendable set of words. For $w \in S$, we consider define the undirected graph $\mathcal{E}_S(w)$, or simply $\mathcal{E}(w)$ when S is clear from the context, having as set of vertices the disjoint union of $L(w)$ and $R(w)$ and edges the pairs $(a, b) \in B(w)$. This graph is called the *extension graph* of w. We sometimes denote $1 \otimes L(w)$ and $R(w) \otimes 1$ the copies of $L(w)$ and $R(w)$ used to define the set of vertices of $\mathcal{E}(w)$. We note that, since $\mathcal{E}(w)$ has $\ell(w)+r(w)$ vertices and $e(w)$ edges, the number $1 - m(w)$ is the Euler characteristic of the graph $\mathcal{E}(w)$[1].

[1]We consider here graphs as 1-dimensional complexes and thus they have no faces.

If the extension graph $\mathcal{E}(w)$ is acyclic, then $m(w) \leq 0$. Thus w is weak or neutral. More precisely, one has in this case that $c = 1 - m(w)$ is the number of connected components of the graph $\mathcal{E}(w)$.

A bi-extendable set S is called *acyclic* if for every $w \in S$, the graph $\mathcal{E}(w)$ is acyclic.

A bi-extendable set S is called a *tree set* of characteristic c if for any nonempty $w \in S$, the graph $\mathcal{E}(w)$ is a tree and if $\mathcal{E}(\varepsilon)$ is a union of c trees (the definition of tree set in [7] corresponds to a tree set of characteristic 1). Note that a tree set of characteristic c is a neutral set of characteristic c. We focus here on characteristic 1 or 2 (specular sets, that we will introduce in Section 4, are tree sets of characteristic 2 with some symmetric properties).

An infinite word is *episturmian* if the set of its factors is closed under reversal and if it contains for each n at most one word of length n which is right-special. It is a *strict episturmian* word if the set of its factors has exactly one right-special word of each length and moreover each of these words u is such that $r(u) = \mathrm{Card}(A)$ (see [4]).

A *Sturmian set* is the set of factors of a strict episturmian word. Any Sturmian set is a recurrent tree set of characteristic 1 (see [7]).

Example 2.3 Let $A = \{a, b\}$. The *Fibonacci morphism* is the morphism $f : A^* \to A^*$ defined by $f(a) = ab$ and $f(b) = a$. The *Fibonacci word* is the fixed-point $f^\omega(a)$ of the Fibonacci morphism. Its set of factors is a Sturmian set (see [26]).

2.2. Bifix codes

A *prefix code* is a set of nonempty words which does not contain any proper prefix of its elements. A *suffix code* is defined symmetrically. A *bifix code* is a set which is both a prefix code and a suffix code (see [5] for a more detailed introduction).

A *coding morphism* for a prefix code X on the alphabet A is a morphism $f : B^* \to A^*$ which maps bijectively B onto X.

Let S be a recurrent set. A prefix (resp. bifix) code $X \subset S$ is S-maximal if it is not properly contained in a prefix (resp. bifix) code $Y \subset S$. Since S is recurrent, a finite S-maximal bifix code is also an S-maximal prefix code (see [4], Theorem 4.2.2).

For example, for any $n \geq 1$, the set $X = S \cap A^n$ is an S-maximal bifix code.

Let X be a bifix code. Let Q be the set of words without any suffix in X and let P be the set of words without any prefix in X. A *parse* of a word w with respect to a bifix code X is a triple $(q, x, p) \in Q \times X^* \times P$ such that $w = qxp$. We denote by $d_X(w)$ the number of parses of a word w with respect to X. The S-degree of X, denoted $d_X(S)$, is the maximal number of parses with respect to X of a word of S.

For example, the set $X = S \cap A^n$ has S-degree n.

Let S be a recurrent set and let X be a finite bifix code. By Theorem 4.2.8 in [4], X is S-maximal if and only if its S-degree is finite. Moreover, in this case, a word $w \in S$ is such that $d_X(w) < d_X(S)$ if and only if it is an internal factor of a word of X.

The *kernel* of a bifix code X is the set of words of X which are internal factors of X.

We will use bifix codes in relation with a more general version of extension graphs (see [7]). For two sets of words X, Y and a word $w \in S$, we denote $L_S^X(w) = \{x \in X \mid xw \in S\}$, $R_S^Y(w) = \{y \in Y \mid wy \in S\}$, $B_S^{X,Y}(w) = \{(x, y) \in X \times Y \mid xwy \in S\}$. We also define $\mathcal{E}_S^{X,Y}(w)$ as the undirected graph on the set of vertices which is the disjoint union of $L_S^X(w)$ and $R_S^Y(w)$ and edges in $B_S^{X,Y}(w)$. Set further

$$\ell_S^X(w) = \mathrm{Card}(L_S^X(w)),\ r_S^Y(w) = \mathrm{Card}(R_S^Y(w)),\ b_S^{X,Y}(w) = \mathrm{Card}(B_S^{X,Y}(w)).$$

Finally, for a word w, we denote $m_S^{X,Y}(w) = b_S^{X,Y}(w) - \ell_S^X(w) - r_S^Y(w) + 1$. Note that $\mathcal{E}_S^{A,A}(w) = \mathcal{E}_S(w)$, $m_S^{A,A}(w) = m_S(w)$, and so on.

We will use below the following result.

Proposition 2.4 *Let S be a recurrent set, let $X \subset S$ be a finite S-maximal suffix code and let $Y \subset S$ be a finite S-maximal prefix code.*

1. *If $\mathcal{E}_S(x)$ is acyclic, then $\mathcal{E}_S^{X,Y}(x)$ is acyclic.*
2. *If S is neutral, then $m_S^{X,Y}(w) = m_S(w)$ for every $w \in S$.*

Proof. Statement 1 follows from Proposition 3.7 in [7]. Statement 2 is Proposition 6.2 in [20]. ∎

Observe that the condition that X (resp. Y) is an S-maximal suffix (resp. prefix) code is only necessary for Assertion 2 (for Assertion 1, X (resp. Y) may be an arbitrary suffix (resp. prefix) code). Observe also that this condition can be replaced by the condition that X (resp. Y) is an Sw^{-1}-maximal suffix code (resp. a $w^{-1}S$-maximal prefix code), where $Sw^{-1} = \{u \in S \mid uw \in S\}$ and symmetrically $w^{-1}S = \{u \in S \mid wu \in S\}$.

3. Specular groups

In this section, we introduce specular groups and we prove some properties of this family of groups. These groups are related with the notion of group acting on a tree (see, for example [33]). In particular, using the Kurosh subgroup theorem, we prove that any subgroup of a specular group is specular (Theorem 3.3).

Note that most of the results in this section are known. We state them and give some simple proofs for the sake of completeness.

3.1. Definitions

We consider an alphabet A with an involution $\theta : A \to A$, possibly with some fixed points. We also consider the group G_θ generated by A with the relations $a\theta(a) = \varepsilon$ for every $a \in A$. Thus $\theta(a) = a^{-1}$ for $a \in A$. The set A is called a *natural* set of generators of G_θ.

When θ has no fixed point, we can set $A = B \cup B^{-1}$ by choosing a set of representatives of the orbits of θ for the set B. The group G_θ is then the free group on B, denoted F_B. In general, the group G_θ is a free product of a free group and a finite number of copies of $\mathbb{Z}/2\mathbb{Z}$, that is $G_\theta = \mathbb{Z}^{*i} * (\mathbb{Z}/2\mathbb{Z})^{*j}$ where i is the number of orbits of θ with two elements and j the number of its fixed points. Such a group will be called a *specular group* of type (i, j). These groups are very close to free groups, as we will see. The integer $\mathrm{Card}(A) = 2i + j$ is called the *symmetric rank* of the specular group $\mathbb{Z}^{*i} * (\mathbb{Z}/2\mathbb{Z})^{*j}$.

Proposition 3.1 *Two specular groups are isomorphic if and only if they have the same type.*

Proof. The commutative image of a group of type (i, j) is $\mathbb{Z}^i \times (\mathbb{Z}/2\mathbb{Z})^j$ and the uniqueness of i, j follows from the fundamental theorem of finitely generated Abelian groups. ∎

Example 3.2 Let $A = \{a, b, c, d\}$ and let θ be the involution which exchanges b, d and fixes a, c. Then $G_\theta = \mathbb{Z} * (\mathbb{Z}/2\mathbb{Z})^2$ is a specular group of symmetric rank 4.

The Cayley graph of a specular group G_θ with respect to the set of natural generators A is a regular tree where each vertex has degree $\mathrm{Card}(A)$. The specular groups are actually characterized by this property (see [17]).

3.2. Subgroups

By the Kurosh subgroup theorem, any subgroup of a free product $G_1 * G_2 * \cdots * G_n$ is itself a free product of a free group and of groups conjugate to subgroups of the G_i (see [28]). Thus, we have, replacing the Nielsen-Schreier Theorem of free groups, the following result.

Theorem 3.3 *Any subgroup of a specular group is specular.*

It also follows from the Kurosh subgroup theorem that the elements of order 2 in a specular group G_θ are the conjugates of the j fixed points of θ and this number is thus the number of conjugacy classes of elements of order 2. Indeed, an element of order 2 generates a subgroup conjugate to one of the subgroups generated by the letters of order 2.

Any specular group $G = G_\theta$ has a free subgroup of index 2. Indeed, let H be the subgroup formed of the reduced words of even length. It has clearly index 2. It is free because it does not contain any element of order 2 (such an element is conjugate to a fixed point of θ and thus is of odd length).

A group having a free subgroup of finite index is called *virtually free* (see [17]). On the other hand, a finitely generated group is said to be *context-free* if, for some presentation, the set of words equivalent to ε is a context-free language. By Muller and Schupp's theorem, a finitely generated group is virtually free if and only if it is context-free [29]. Thus a specular group is context-free. One may verify this directly as follows. A context-free grammar generating the words equivalent to ε for the natural presentation of a specular group $G = G_\theta$ is the grammar with one nonterminal symbol σ and the rules

$$\sigma \to a\sigma a^{-1}\sigma \quad (a \in A), \quad \sigma \to \varepsilon.$$

The proof that this grammar generates the set of words equivalent to ε is similar to that used in [3] for the so-called Dyck-like languages.

We will need two more properties of specular groups. Both are well-known to hold for free groups.

A group G is called *residually finite* if for every element $g \neq \varepsilon$ of G, there is a morphism φ from G onto a finite group such that $\varphi(g) \neq \varepsilon$.

Every free group is residually finite. The same property holds for subgroups of a product of residually finite groups (see, for example, [33, p. 122]) Thus we have the following result (we give an alternative proof for the sake of completeness).

Proposition 3.4 *Any specular group is residually finite.*

Proof. Let K be a free subgroup of index 2 in the specular group G. Let $g \neq 1$ be in G. If $g \notin K$, then the image of g in G/K is nontrivial. Assume $g \in K$. Since K is free, it is residually finite. Let N be a normal subgroup of finite index of K such that $g \notin N$. Consider the representation of G on the right cosets of N. Since $g \notin N$, the image of g in this finite group is nontrivial. ∎

A group G is said to be *Hopfian* if any surjective morphism from G onto G is also injective. By a result of Malcev, any finitely generated residually finite group is Hopfian (see [27], p. 197). We thus deduce from Proposition 3.4 the following result.

Proposition 3.5 *A specular group is Hopfian.*

3.3. Monoidal basis

A word on the alphabet A is *θ-reduced* (or simply *reduced*) if it has no factor of the form $a\theta(a)$ for $a \in A$. It is clear that any element of a specular group is represented by a unique reduced word.

A subset of a group G is called *symmetric* if it is closed under taking inverses. A set X in a specular group G is called a *monoidal basis* of G if it is symmetric, if the monoid that it generates is G and if any product $x_1 x_2 \cdots x_m$ of elements of X such that $x_k x_{k+1} \neq \varepsilon$ for $1 \leq k \leq m-1$ is distinct of ε.

Example 3.6 The alphabet A is a monoidal basis of G_θ.

The previous example shows that the symmetric rank of a specular group is the cardinality of any monoidal basis (two monoidal bases have the same cardinality since the type is invariant by isomorphism by Proposition 3.1).

Let H be a subgroup of a specular group G. Let Q be a set of reduced words on A which is a prefix-closed set of representatives of the right cosets Hg of H. Such a set is traditionally called a *Schreier transversal* for H (the proof of its existence is classical in the free group and it is the same in any specular group).

Let

$$X = \{paq^{-1} \mid a \in A, p, q \in Q, pa \notin Q, pa \in Hq\}. \tag{3.1}$$

Each word x of X has a unique factorization paq^{-1} with $p, q \in Q$ and $a \in A$. The letter a is called the *central part* of x.

The following is [33, Proposition 16].

Proposition 3.7 *Let H, Q and X as above. Then X is a monoidal basis of H.*

Proof. Let us first show that X is symmetric. Let $x = paq^{-1} \in X$, then $x^{-1} = qa^{-1}p^{-1}$. We cannot have $qa^{-1} \in Q$ since otherwise $p \in Hqa^{-1}$ implies $p = qa^{-1}$ by uniqueness of the coset representative and finally $pa \in Q$.

The set X generates H as a monoid because if $x = a_1a_2 \cdots a_m \in H$ with $a_i \in A$, then $x = (a_1p_1^{-1})(p_1a_2p_2^{-1}) \cdots (p_{m-1}a_m)$ with $a_1 \cdots a_k \in Hp_k$ for $1 \le k \le m-1$ is a factorization of x in elements of $X \cup \{\varepsilon\}$.

Finally, if a product $x_1x_2 \cdots x_m$ of elements of X is equal to ε, then $x_kx_{k+1} = \varepsilon$ for some index k since the central part a never cancels in a product of two elements of X.

Thus, X is a monoidal basis of H. ∎

The set X of Proposition 3.7 is called the *Schreier basis* of the subgroup H relative to the Schreier transversal Q.

One can deduce directly Theorem 3.3 from these properties of X.

Proof of Theorem 3.3. Let H be a subgroup of a specular group G, Q be a Schreier transversal for H and X be the Schreier basis relative to Q. Let $\varphi : B \to X$ be a bijection from a set B onto X which extends to a morphism from B^* onto H. Let $\sigma : B \to B$ be the involution sending each b to c where $\varphi(c) = \varphi(b)^{-1}$.

Since the central parts never cancel, if a nonempty word $w \in B^*$ is σ-reduced then $\varphi(w) \neq \varepsilon$. This shows that H is isomorphic to the group G_σ. Thus H is specular. ∎

If H is a subgroup of index n of a specular group G of symmetric rank r, the symmetric rank s of H is

$$s = n(r-2) + 2. \tag{3.2}$$

This formula replaces Schreier's Formula (which corresponds to the case $j = 0$). It can be proved as follows. Let Q be a Schreier transversal for H and let X be the corresponding Schreier basis. The number of elements of X is

$nr - 2(n-1)$. Indeed, this is the number of pairs $(p, a) \in Q \times A$ minus the $2(n-1)$ pairs (p, a) such that $pa \in Q$ with pa reduced or $pa \in Q$ with pa not reduced. This gives Formula (3.2).

Example 3.8 Let G be the specular group of Example 3.2. Let H be the subgroup formed by the elements represented by a reduced word of even length. The set $Q = \{\varepsilon, a\}$ is a prefix-closed set of representatives of the two cosets of H. The representation of G by permutations on the cosets of H is represented in Figure 3.1.

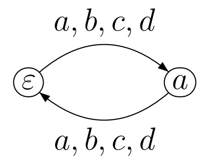

Figure 3.1: The representation of G by permutations on the cosets of H.

The monoidal basis corresponding to Formula (3.1) is

$$X = \{ab, ac, ad, ba, ca, da\}.$$

The symmetric rank of H is 6, in agreement with Formula (3.2) and H is a free group of rank 3.

Example 3.9 Let again G be the specular group of Example 3.2. Consider now the subgroup K stabilizing 1 in the representation of G by permutations on the set $\{1, 2\}$ of Figure 3.2.

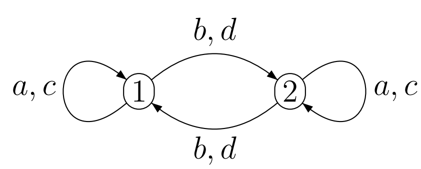

Figure 3.2: The representation of G by permutations on the cosets of K.

We choose $Q = \{\varepsilon, b\}$. The set X corresponding to Formula (3.1) is

$$X = \{a, bad, bb, bcd, c, dd\}.$$

The group K is isomorphic to $\mathbb{Z} * (\mathbb{Z}/2\mathbb{Z})^{*4}$.

The following result, which will be used later (Section 6), is a consequence of Proposition 3.5.

Proposition 3.10 *Let G be a specular group of type (i, j) and let $X \subset G$ be a symmetric set with $2i + j$ elements. If X generates G, it is a monoidal basis of G.*

Proof. Let A be a set of natural generators of G. Considering the commutative image of G, we obtain that X contains j elements of order 2. Thus there is a bijection φ from A onto X such that $\varphi(a^{-1}) = \varphi(a)^{-1}$ for every $a \in A$. The map φ extends to a morphism from G to G which is surjective since X generates G. Then φ being surjective, it also injective since G is Hopfian, and thus X is a monoidal basis of G. ∎

4. Specular sets

In this section, we introduce specular sets. We introduce odd and even words and the even code which play an important part in the sequel. We prove that the decoding of a recurrent specular set by the even code is a union of two recurrent tree sets of characteristic 1 (Theorem 4.13). We exhibit a family of specular sets obtained as the result of a transformation called doubling, starting from a tree set of characteristic 1 and invariant by reversal (Theorem 4.20). In the last part, we relate specular sets with full and G-full words, a notion linked with palindromic complexity and introduced in [32].

4.1. Definition

We assume given an involution θ on the alphabet A generating the specular group G_θ.

A symmetric bi-extendable (and thus factorial) set S of reduced words on the alphabet A is called a *laminary set* on A relative to θ (following [14] and [25]). Thus the elements of a laminary set S are elements of the specular group G_θ and the set S is contained in G_θ.

A *specular set* is a laminary set on A which is a tree set of characteristic 2. Thus, in a specular set, the extension graph of every nonempty word is a tree and the extension graph of the empty word is a union of two disjoint trees.

The following is a very simple example of a specular set.

Example 4.1 Let $A = \{a, b\}$ and let θ be the identity on A. Then the set of factors of $(ab)^\omega$ is a specular set (we denote by x^ω the word x infinitely repeated).

The next example is due to Julien Cassaigne. We frequently refer to it in next sections.

Example 4.2 Let $A = \{a, b, c, d\}$ and let S be the set of factors of the fixed point $\sigma^\omega(a)$ of the morphism σ from A^* into itself defined by

$$\sigma(a) = ab, \quad \sigma(b) = cda, \quad \sigma(c) = cd, \quad \sigma(d) = abc.$$

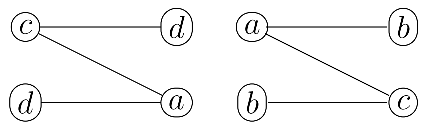

Figure 4.1: The extension graph $\mathcal{E}_S(\varepsilon)$.

The extension graph of ε is shown in Figure 4.1. It is shown in [7, Example 3.4] that S is a tree set of characteristic 2. We will see later (Example 4.23) that S is a specular set relative to the involution $\theta = (bd)$.

The following result shows in particular that in a specular set the two trees forming $\mathcal{E}(\varepsilon)$ are isomorphic since they are exchanged by the bijection $(a, b) \to (b^{-1}, a^{-1})$.

Proposition 4.3 *Let S be a specular set. Let $\mathcal{T}_0, \mathcal{T}_1$ be the two trees such that $\mathcal{E}(\varepsilon) = \mathcal{T}_0 \cup \mathcal{T}_1$. For any $a, b \in A$ and $i = 0, 1$, one has $(1 \otimes a, b \otimes 1) \in \mathcal{T}_i$ if and only if $(1 \otimes b^{-1}, a^{-1} \otimes 1) \in \mathcal{T}_{1-i}$.*

Proof. Assume that $(1 \otimes a, b \otimes 1)$ and $(1 \otimes b^{-1}, a^{-1} \otimes 1)$ are both in $\mathcal{T}_0$. Since $\mathcal{T}_0$ is a tree, there is a path from $1 \otimes a$ to $a^{-1} \otimes 1$. We may assume that this path is reduced, that is, does not use consecutively twice the same edge. Since this path is of odd length, it has the form $(u_0, v_1, u_1, \ldots, u_p, v_p)$ with $u_0 = 1 \otimes a$ and $v_p = a^{-1} \otimes 1$. Since S is symmetric, we also have a reduced path $(v_p^{-1}, u_p^{-1}, \cdots, u_1^{-1}, u_0^{-1})$ which is in $\mathcal{E}(\varepsilon)$ (for $u_i = 1 \otimes a_i$, we denote $u_i^{-1} = a_i^{-1} \otimes 1$ and similarly for v_i^{-1}) and thus in $\mathcal{T}_0$ since $\mathcal{T}_0$ and $\mathcal{T}_1$ are disjoint. Since $v_p^{-1} = u_0$, these two paths have the same origin and end. But if a path of odd length is its own inverse, its central edge has the form (x, y) with $x = y^{-1}$, as one verifies easily by induction on the length of the path. This is a contradiction with the fact that the words of S are reduced. Thus the two paths are distinct. This implies that $\mathcal{E}(\varepsilon)$ has a cycle, a contradiction. ∎

Following again the terminology of [14], we say that a laminary set S is *orientable* if there exist two factorial sets S_+, S_- such that $S = S_+ \cup S_-$ with $S_+ \cap S_- = \{\varepsilon\}$ and for any $x \in S$, one has $x \in S_-$ if and only if $x^{-1} \in S_+$ (where x^{-1} is the inverse of x in G_θ).

The following result shows in particular that for any tree set T of characteristic 1 on the alphabet B, the set $T \cup T^{-1}$ is a specular set on the alphabet $A = B \cup B^{-1}$.

Theorem 4.4 *Let S be a specular set on the alphabet A. Then, S is orientable if and only if there is a partition $A = A_+ \cup A_-$ of the alphabet A and a tree set T of characteristic 1 on the alphabet $B = A_+$ such that $S = T \cup T^{-1}$.*

Proof. The condition is trivially sufficient. Let us prove it is necessary and suppose that S is a specular set on the alphabet A which is orientable. Let (S_+, S_-) be the corresponding pair of subsets of S. The sets S_+, S_- are bi-extendable, since S is. Set $A_+ = A \cap S_+$ and $A_- = A \cap S_-$. Then $A = A_+ \cup A_-$ is a partition of A and, since S_-, S_+ are factorial, we have $S_+ \subset A_+^*$ and $S_- \subset A_-^*$. Let $\mathcal{T}_0, \mathcal{T}_1$ be the two trees such that $\mathcal{E}(\varepsilon) = \mathcal{T}_0 \cup \mathcal{T}_1$. Assume that a vertex of $\mathcal{T}_0$ is in A_+. Then all vertices of $\mathcal{T}_0$ are in A_+ and all vertices of $\mathcal{T}_1$ are in A_-. Moreover, $\mathcal{E}_{S_+}(\varepsilon) = \mathcal{T}_0$ and $\mathcal{E}_{S_-}(\varepsilon) = \mathcal{T}_1$. Thus S_+, S_- are tree sets of characteristic 1. ∎

The following result follows easily from Proposition 2.2 (see [20, Proposition 2.4] for details).

Proposition 4.5 *The factor complexity of a specular set containing the alphabet A is given by $p_0 = 1$ and $p_n = n(k-2)+2$ for $n \geq 1$ with $k = \mathrm{Card}(A)$.*

4.2. Odd and even words

We introduce a notion which plays, as we shall see, an important role in the study of specular sets. Let S be a specular set. Since a specular set is bi-extendable, any letter $a \in A$ occurs exactly twice as a vertex of $\mathcal{E}(\varepsilon)$, one as an element of $L(\varepsilon)$ and one as an element of $R(\varepsilon)$. A letter $a \in A$ is said to be *even* if its two occurrences appear in the same tree. Otherwise, it is said to be *odd*. Observe that if a specular S is recurrent, there is at least one odd letter.

Example 4.6 Let S be the set of factors of $(ab)^\omega$ as in Example 4.1. Then a and b are odd.

Example 4.7 Let S be the set of Example 4.2. The letters b, d are even, while a and c are odd.

Let S be a specular set. A word $w \in S$ is said to be *even* if it has an even number of odd letters. Otherwise it is said to be *odd*. The set of even words has the form $X^* \cap S$ where $X \subset S$ is a bifix code, called the *even code*. The set X is the set of even words without a nonempty even prefix (or suffix).

Proposition 4.8 *Let S be a recurrent specular set. The even code is an S-maximal bifix code of S-degree* 2.

Proof. Let us verify that any $w \in S$ is comparable for the prefix order with an element of the even code X. If w is even, it is in X^*. Otherwise, since S is recurrent, there is a word u such that $wuw \in S$. If u is even, then wuw is even and thus $wuw \in X^*$. Otherwise wu is even and thus $wu \in X^*$. This shows that X is S-maximal. The fact that it has S-degree 2 follows from the fact that any product of two odd letters is a word of X which is not an internal factor of X and has two parses. ∎

Example 4.9 Let S be the specular set of Example 4.2. The letters b, d are even and the letters a, c are odd. The even code is

$$X = \{abc, ac, b, ca, cda, d\}.$$

Denote by $\mathcal{T}_0, \mathcal{T}_1$ the two trees such that $\mathcal{E}(\varepsilon) = \mathcal{T}_0 \cup \mathcal{T}_1$. We consider the directed graph $\mathcal{G}$ with vertices $0, 1$ and edges all the triples (i, a, j) for $0 \le i, j \le 1$ and $a \in A$ such that $(1 \otimes b, a \otimes 1) \in \mathcal{T}_i$ and $(1 \otimes a, c \otimes 1) \in \mathcal{T}_j$ for some $b, c \in A$. The graph $\mathcal{G}$ is called the *parity graph* of S. Observe that for every letter $a \in A$ there is exactly one edge labeled a because a appears exactly once as a left (resp. right) vertex in $\mathcal{E}(\varepsilon)$.

Example 4.10 Let S be the specular set of Example 4.2. The parity graph of S is represented in Figure 4.2, where we assume that $\mathcal{T}_0$ is the tree on the left of Figure 4.1 and $\mathcal{T}_1$ is the tree on the right of Figure 4.1.

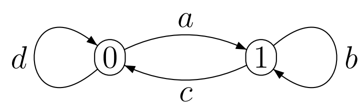

Figure 4.2: The parity graph.

Proposition 4.11 *Let S be a specular set and let $\mathcal{G}$ be its parity graph. Let $S_{i,j}$ be the set of words in S which are the label of a path from i to j in the graph $\mathcal{G}$.*

(1) The family $(S_{i,j} \setminus \{\varepsilon\})_{0 \le i,j \le 1}$ is a partition of $S \setminus \{\varepsilon\}$.
(2) For $u \in S_{i,j} \setminus \{\varepsilon\}$ and $v \in S_{k,\ell} \setminus \{\varepsilon\}$, if $uv \in S$, then $j = k$.
(3) $S_{0,0} \cup S_{1,1}$ is the set of even words.
(4) $S_{i,j}^{-1} = S_{1-j,1-i}$.

Proof. We first note that for $a, b \in A$ such that $ab \in S$, there is a path in $\mathcal{G}$ labeled ab. Since $(a, b) \in \mathcal{E}(\varepsilon)$, there is a k such that $(1 \otimes a, b \otimes 1) \in \mathcal{T}_k$. Then we have $a \in S_{i,k}$ and $b \in S_{k,j}$ for some $i, j \in \{0, 1\}$. This shows that ab is the label of a path from i to j in $\mathcal{G}$.

Let us prove by induction on the length of a nonempty word $w \in S$ that there exists a unique pair i, j such that $w \in S_{i,j}$. The property is true for a letter, by definition of the extension graph $\mathcal{E}(\varepsilon)$ and for words of length 2 by the above argument. Let next $w = ax$ be in S with $a \in A$ and x nonempty. By induction hypothesis, there is a unique pair (k, j) such that $x \in S_{k,j}$. Let b be the first letter of x. Then the edge of $\mathcal{G}$ with label b starts in k. Since ab is the label of a path, we have $a \in S_{i,k}$ for some i and thus $ax \in S_{i,j}$. The other assertions follow easily (Assertion (4) follows from Proposition 4.3). ■

Note that Assertion (4) implies that no nonempty even word is its own inverse. Indeed, $S_{0,0}^{-1} = S_{1,1}$ and $S_{1,1}^{-1} = S_{0,0}$.

Proposition 4.12 *Let S be a specular set. If $x, y \in S$ are nonempty words such that $xyx^{-1} \in S$, then y is odd.*

Proof. Let i, j be such that $x \in S_{i,j}$. Then $x^{-1} \in S_{1-j,1-i}$ by Assertion (4) of Proposition 4.11 and thus $y \in S_{j,1-j}$ by Assertion (2). Thus y is odd by Assertion (3). ■

The following result is the counterpart for recurrent specular sets of the main result of [9, Theorem 6.1] asserting that the family of (uniformly) recurrent tree sets of characteristic 1 is closed under maximal bifix decoding. Let

S be a recurrent set and let f be a coding morphism for a finite S-maximal bifix code X. The set $f^{-1}(S)$ is called a *decoding* of S by X.

Theorem 4.13 (Even code decoding Theorem) *The decoding of a recurrent specular set by the even code is a union of two recurrent tree sets of characteristic* 1*. More precisely, let S be a recurrent specular set and let f be a coding morphism for the even code. Then $f^{-1}(S_{0,0})$ and $f^{-1}(S_{1,1})$ are recurrent tree sets of characteristic* 1*.*

Proof. We show that $T_0 = f^{-1}(S_{0,0})$ is a recurrent tree set of characteristic 1. The proof for $f^{-1}(S_{1,1})$ is the same.

First, T_0 is bi-extendable, as one may easily verify. Next, since S is recurrent, it is uniformly recurrent by Theorem 2.1. Thus for every $u \in S$ there exists $n \geq 1$ such that u is a factor of any word w in S of length n. But if $u, w \in S_{0,0}$ are such that $w = \ell ur$, then $\ell, r \in S_{0,0}$. Thus T_0 is recurrent.

We now show that T_0 is a tree set of characteristic 1. Let X be the even code and set $X_0 = X \cap S_{0,0}$ and $X_1 = X \cap S_{1,1}$.

It is enough to show that $\mathcal{E}^0_S(w) = \mathcal{E}^{X_0,X_0}_S(w)$ is a tree for any $w \in S_{0,0}$. Note first that $\mathcal{E}^0_S(w) = \mathcal{E}^{X,X}_S(w)$. Indeed, for $w \in S_{0,0}$ and $x, y \in X$ such that $xwy \in S$, one has $x, y \in X_0$ and $xwy \in S_{0,0}$.

First, for any nonempty word $w \in S_{0,0}$, since $\mathcal{E}^0_S(w) = \mathcal{E}^{X,X}_S(w)$, the graph $\mathcal{E}^0_S(w)$ is a tree by Proposition 2.4.

Next, let us show that the graph $\mathcal{E}^0_S(\varepsilon)$ is a tree. First, since $\mathcal{E}_S(\varepsilon)$ is a union of two trees, it is acyclic, and thus the graph $\mathcal{E}^0_S(\varepsilon)$ is acyclic by Proposition 2.4. Next, since S is neutral, by Proposition 2.4, we have $m^{X,X}_S(\varepsilon) = m_S(\varepsilon) = -1$. This implies that $m^{X,X}_S(\varepsilon)$ is a union of two trees. Since $\mathcal{E}^{X,X}_S(\varepsilon)$ is the disjoint union of $\mathcal{E}^0_S(\varepsilon)$ and $\mathcal{E}^{X_1,X_1}_S(\varepsilon)$, this implies that each one is a tree. ∎

Example 4.14 Let S be the set of Example 4.2. Recall that it is the set of factors of the fixed point of the morphism

$$\sigma : a \mapsto ab, \quad b \mapsto cda, \quad c \mapsto cd, \quad d \mapsto abc.$$

The even code X is given in Example 4.9. Let $\Sigma = \{a, b, c, d, e, f\}$ and let g be the coding morphism for X given by

$$a \mapsto abc, \quad b \mapsto ac, \quad c \mapsto b, \quad d \mapsto ca, \quad e \mapsto cda, \quad f \mapsto d.$$

The decoding of S by X is a union of two tree sets of characteristic 1 which are the set of factors of the fixed point of the two morphisms

$$a \mapsto afbf,\ b \mapsto af,\ f \mapsto a$$

and

$$c \mapsto e,\ d \mapsto ec,\ e \mapsto ecdc.$$

These two morphisms are actually the restrictions to $\{a, b, f\}$ and $\{c, d, e\}$ of the morphism $g^{-1}\sigma g$.

4.3. Bifix codes in specular sets

Recall from Section 2 that the characteristic of a set S is given by $\chi(S) = \ell_S(\varepsilon) + r_S(\varepsilon) - b_S(\varepsilon)$.

The following result is from [20]. We will use it for specular sets.

Theorem 4.15 *Let S be a recurrent neutral set containing the alphabet A. For any finite S-maximal bifix code X of S-degree $d = d_X(S)$, one has*

$$\mathrm{Card}(X) = d(\mathrm{Card}(A) - \chi(S)) + \chi(S).$$

We can apply Theorem 4.15 to recurrent specular sets.

Theorem 4.16 (Cardinality Theorem for bifix codes) *Let S be a recurrent specular set containing the alphabet A. For any finite S-maximal bifix code X, one has*

$$\mathrm{Card}(X) = d_X(S)(\mathrm{Card}(A) - 2) + 2. \tag{4.1}$$

Proof. Since S is specular, we have $\chi(S) = 2$ and thus the statement follows directly from Theorem 4.15. ∎

Example 4.17 Let S be the specular set of Example 4.2. The even code (given in Example 4.9) is an S-maximal code of S-degree 2. We have $\mathrm{Card}(X) = 6$ in agreement with Theorem 4.15.

The following statement is a partial converse of Theorem 4.15.

Theorem 4.18 *Let S be a uniformly recurrent laminary set containing the alphabet A. If the graph $\mathcal{E}(\varepsilon)$ is acyclic and if any finite S-maximal bifix code of S-degree d has $d(Card(A) - 2) + 2$ elements, then S is specular.*

To prove Theorem 4.18, we use the following result, which can be proved in the same way as Theorem 3.12 in [8], using internal transformations.

Proposition 4.19 *Let S be a recurrent set containing the alphabet A and let $d_0 \geq 2$. If all finite S-maximal bifix codes of S-degree $d \geq d_0$ have the same cardinality, then any word of length greater than or equal to $d_0 - 1$ is neutral.*

Theorem 4.18 results from Proposition 4.19 applied with $d_0 = 2$.

4.4. Doubling maps

We now introduce a construction which allows one to build specular sets. This is a particular case of the multiplying maps introduced in [20].

A *transducer* is a labeled graph with vertices in a set Q and edges labeled in $\Sigma \times A$. The set Q is called the set of states, the set Σ is called the *input alphabet* and A is called the *output alphabet*. The graph obtained by erasing the output letters is called the *input automaton* (with an unspecified initial state). Similarly, the *output automaton* is obtained by erasing the input letters.

Let $\mathcal{A}$ be a transducer with set of states $Q = \{0, 1\}$ on the input alphabet Σ and the output alphabet A. We assume that

1. the input automaton is a group automaton, that is, every letter of Σ acts on Q as a permutation
2. the output labels of the edges are all distinct.

We define two maps $\delta_0, \delta_1 : \Sigma^* \to A^*$ corresponding to initial states 0 and 1 respectively. Thus $\delta_0(u) = v$ (resp. $\delta_1(u) = v$) if the path starting at state 0 (resp. 1) with input label u has output v. The pair $\delta_{\mathcal{A}} = (\delta_0, \delta_1)$ is called a *doubling map* and the transducer $\mathcal{A}$ a *doubling transducer*.

The *image* of a set T on the alphabet Σ by the doubling map $\delta_{\mathcal{A}}$ is the set $S = \delta_0(T) \cup \delta_1(T)$.

If $\mathcal{A}$ is a doubling transducer, we define an involution $\theta_{\mathcal{A}}$ as follows. For any $a \in A$, let (i, α, a, j) be the edge with input label α and output label a. We define $\theta_{\mathcal{A}}(a)$ as the output label of the edge starting at $1 - j$ with input label α. Thus, $\theta_{\mathcal{A}}(a) = \delta_i(\alpha) = a$ if $i + j = 1$ and $\theta_{\mathcal{A}}(a) = \delta_{1-i}(\alpha) \neq a$ if $i = j$.

Recall that the reversal of a word $w = a_1 a_2 \cdots a_n$ is the word $\tilde{w} = a_n \cdots a_2 a_1$.

One can prove by induction on the length of $y \in \Sigma^*$ that if $x = \delta_i(y)$ and if j is the end of the path starting at i and with input label y, then $x^{-1} = \delta_{1-j}(\tilde{y})$. Observe that since the input automaton is a group automaton, there is always a path starting at $1 - j$ with input label $\tilde{y}$.

A set S of words is closed under reversal if $w \in S$ implies $\tilde{w} \in S$ for every $w \in S$. By definition, any Sturmian set is closed under reversal (see [4]).

Theorem 4.20 *For any tree set T of characteristic 1 on the alphabet Σ, closed under reversal and any doubling map $\delta_{\mathcal{A}}$, the image of T by $\delta_{\mathcal{A}}$ is a specular set relative to the involution $\theta_{\mathcal{A}}$.*

Proof. Set $S = \delta_{\mathcal{A}}(T) = \delta_0(T) \cup \delta_1(T)$. By Theorem 3.1 of [20], S is a tree set of characteristic 2. By construction, it is also clear the any word in S is $\theta_{\mathcal{A}}$-reduced.

Let now prove that S is a symmetric language. Assume that $x = \delta_i(y)$ for $i \in \{0, 1\}$ and $y \in T$. Let j be the end of the path starting at i and with input label y. Since $x^{-1} = \delta_{1-j}(\tilde{y})$ and T is closed under reversal, we have $x^{-1} \in \delta_{1-j}(T)$. This shows that S is symmetric and so that it is laminary. Thus, S is a specular set. ■

We now give two examples of specular sets obtained by doubling maps (doubling the Fibonacci set).

Example 4.21 Let $\Sigma = \{\alpha, \beta\}$ and let T be the Fibonacci set over Σ. Let δ be the doubling map given by the transducer of Figure 4.3 on the left.

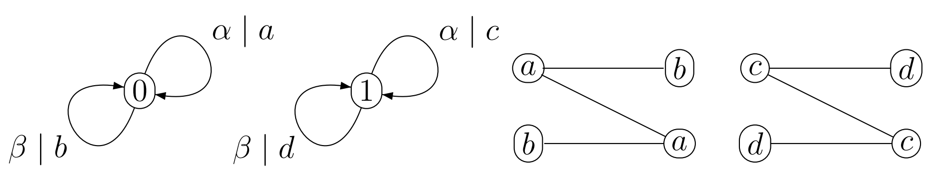

Figure 4.3: A doubling transducer and the extension graph $\mathcal{E}_S(\varepsilon)$.

Both letters in Σ act as the identity on the two states $0, 1$.

Then $\theta_{\mathcal{A}}$ is the involution defined by $\theta : a \mapsto c,\ b \mapsto d,\ c \mapsto a,\ d \mapsto b$. The image of T by δ is a specular set S on the alphabet $A = \{a, b, c, d\}$. The graph $\mathcal{E}_S(\varepsilon)$ is represented in Figure 4.4 on the right. All letters are even.

Note that the set S of Example 4.21 is not recurrent. The set S is actually just a union of two Fibonacci sets, one over the alphabet $\{a, b\}$ and the second over the alphabet $\{c, d\}$.

Example 4.22 Let $\Sigma = \{\alpha, \beta\}$ and let T be the Fibonacci set. Let δ be the doubling map given by the transducer of Figure 4.4 on the left. The letter α acts as the transposition of the two states $0, 1$, while β acts as the identity.

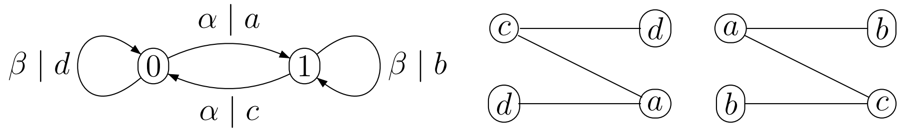

Figure 4.4: A doubling transducer and the extension graph $\mathcal{E}_S(\varepsilon)$.

Then $\theta_{\mathcal{A}}$ is the involution θ of Example 3.2 and the image of T by δ is a specular set S on the alphabet $A = \{a, b, c, d\}$. The graph $\mathcal{E}_S(\varepsilon)$ is represented in Figure 4.4 on the right.

The letters a, c are odd and b, d are even.

Note that S is the set of factors of the fixed point $g^\omega(a)$ of the morphism

$$g : a \mapsto abcab, \quad b \mapsto cda, \quad c \mapsto cdacd, \quad d \mapsto abc.$$

The morphism g is obtained by applying the doubling map to the cube f^3 of the Fibonacci morphism f in such a way that $g^\omega(a) = \delta_0(f^\omega(\alpha))$.

In the next example (due to Julien Cassaigne), the specular set is obtained using a morphism of smaller size.

Example 4.23 Let $A = \{a, b, c, d\}$. Let T be the set of factors of the fixed point $x = f^\omega(\alpha)$ of the morphism $f : \alpha \mapsto \alpha\beta, \beta \mapsto \alpha\beta\alpha$. It is a Sturmian set. Indeed, x is the characteristic word of slope $-1 + \sqrt{2}$ (see [26]). The sequence $s_n = f^n(\alpha)$ satisfies $s_n = s_{n-1}^2 s_{n-2}$ for $n \geq 2$. The image S of T by the doubling automaton of Figure 4.4 is the set of factors of the fixed point $\sigma^\omega(a)$ of the morphism σ from A^* into itself defined by

$$\sigma(a) = ab, \quad \sigma(b) = cda, \quad \sigma(c) = cd, \quad \sigma(d) = abc.$$

Thus the set S is the same as that of Example 4.2.

Note that, when S is a specular set obtained by a doubling map using a transducer $\mathcal{A}$, the parity graph of S is the output automaton of $\mathcal{A}$ (see for instance Figures 4.2 and 4.4).

4.5. Palindromes

The notion of palindromic complexity originates in [21] where it is proved that a word of length n has at most $n+1$ palindrome factors. A word of length n is full if it has $n+1$ palindrome factors and a factorial set is *full* (or rich) if all its elements are full. By a result of [23], a recurrent set closed under reversal is full if and only if every complete return word to a palindrome in S is a palindrome (a complete return word to a set X of words of the same length is a word of S which has exactly two factors in X, one as a proper prefix and one as a proper suffix, see Section 6.1). It is known that all Sturmian sets are full [21] and also all natural codings of interval exchange defined by a symmetric permutation [1].

The fact that a tree set of characteristic 1 is full in the following result generalizes results of [21, 1].

Proposition 4.24 *Let T be a recurrent tree set of characteristic* 1*, closed under reversal. Then T is full.*

Proof. We use the following equivalent definition of full sets (see [32]): for any $x \in T$,

(i) if x is not a palindrome, it is neutral.
(ii) Otherwise, $m(x) + 1$ is equal to the number of letters a such that axa is a palindrome in T (the so-called *palindromic extensions*).

Since T is a tree set of characteristic 1, every word is neutral. We thus only have to show that every palindrome has exactly one palindromic extension. Let $x \in T$ be a palindrome. It may be verified that since x is palindrome and T is closed under reversal, the graph $\mathcal{E}_T(x)$ is closed under reversal in the sense that it contains an edge $(1 \otimes a, b \otimes 1)$ if and only if it contains the edge $(1 \otimes b, a \otimes 1)$. One may verify that, as a consequence, there is at least one $a \in A$ such that $axa \in T$. Indeed, this can be proved as follows by induction on $\mathrm{Card}(A)$. It is true if $\mathrm{Card}(A) = 1$. Otherwise, let $a \in A$ be such that $1 \otimes a$ is a leaf of $\mathcal{E}_T(x)$. Then, since the graph is closed under reversal, the vertex $a \otimes 1$ is also a leaf. Set $A' = A \setminus \{a\}$. The restriction of the graph to the vertices in A' is a tree closed under reversal, and thus the property follows by induction. But if there is another one, the graph would have a cycle. Indeed, assume that $axa, bxb \in T$. Consider a simple path γ of minimal length from one of $1 \otimes a, a \otimes 1$ to one of $1 \otimes b, b \otimes 1$. This path cannot contain the edges corresponding to axa, bxb. Using these edges and the symmetric of γ, one obtains a cycle. Thus T is full. ∎

In [32], this notion was extended to that of G-full, where G is a finite group of morphisms and antimorphisms of A^* (an antimorphism is the composition of a morphism and reversal) containing at least one antimorphism. As one of the equivalent definitions, a set S closed under G is G-full if for every $x \in S$, every complete return word to the G-orbit of x is fixed by a nontrivial element of G.

Let us consider a tree set T of characteristic 1 and a specular set S obtained as the image of T by a doubling map δ.

Let us define the antimorphism $\sigma : u \mapsto u^{-1}$ for $u \in G_\theta$. From Section 4.4 it follows that both edges (i, α, a, j) and $(1-i, \alpha, \sigma(a), 1-j)$ are in the doubling transducer. Let us define also the morphism τ obtained by replacing each letter $a \in A$ by $\tau(a)$ if there are edges (i, α, a, j) and $(1-j, \alpha, \tau(a), 1-i)$ in the doubling transducer.

We denote by $G_{\mathcal{A}}$ the group generated by the σ and τ. Actually, we have $G_{\mathcal{A}} = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

Example 4.25 Let S be the specular set defined in Example 4.21. The group $G_{\mathcal{A}}$ is generated by

$$\sigma : at \mapsto c,\ b \mapsto d,\ c \mapsto a,\ d \mapsto b,$$

and

$$\tau : a \mapsto c,\ b \mapsto d,\ c \mapsto a,\ d \mapsto b.$$

Note that, even if the images of σ and τ over the alphabet are the same, the latter is a morphism, while the first is an antimorphism. Moreover, in that case, we have $\sigma\tau = \tau\sigma : w \mapsto \tilde{w}$ for every $w \in S$.

Example 4.26 Let S be the recurrent specular set defined in Example 4.22. The group $G_{\mathcal{A}}$ is generated by the antimorphism

$$\sigma : a \mapsto a,\ b \mapsto d,\ c \mapsto c,\ d \mapsto a,$$

and the morphism

$$\tau : a \mapsto c,\ b \mapsto d,\ c \mapsto a,\ d \mapsto b.$$

We have $H = \{\mathrm{id}, \sigma, \tau, \sigma\tau\}$, where $\sigma\tau = \tau\sigma$ is the antimorphism fixing b, d and exchanging a and c.

We now connect the notions of fullness and $G_{\mathcal{A}}$-fullness, proving an analogous result of Proposition 4.24 for specular sets.

Proposition 4.27 *Let T be a recurrent tree set of characteristic 1 on the alphabet Σ, closed under reversal and let S be the image of T under a doubling map. Then S is $G_{\mathcal{A}}$-full.*

Proof. By Proposition 4.24 we know that T is full.

To show that S is $G_{\mathcal{A}}$-full, we will use several properties of the map δ_i. We note that it is injective, that it preserves prefixes and conversely: u is a prefix of v if and only if $\delta_i(u)$ is a prefix of $\delta_i(v)$. Also, for any $y \in T$ and $x = \delta_i(y)$, the images of $y, \tilde{y}$ by δ_0, δ_1 form the $G_{\mathcal{A}}$-orbit of x.

Consider $x \in S$ and a word w which is a complete return word to the $G_{\mathcal{A}}$-orbit of x. We may assume that x is a prefix of w and that $\gamma(x)$ is a prefix of w, with $\gamma \in H$. Let $y, u \in T$ and $i \in \{0, 1\}$ be such that $x = \delta_i(y)$ and $w = \delta_i(u)$. Then y is a prefix of u.

We first show that u is a palindrome. First observe that u has a suffix in the set $\{y, \tilde{y}\}$. Indeed, if $\gamma \in \{\mathrm{id}, \tau\}$ then y is a suffix of u. Otherwise, if $\gamma \in \{\sigma, \tau\sigma\}$, one has that $\tilde{y}$ is a suffix of u. Let now z be the longest palindrome prefix of u. Then y is a prefix of z since otherwise z would have a second occurrence in u (in a full set, the longest palindrome prefix of a word is unioccurrent, see [23]). Consequently $\tilde{y}$ is a suffix of z and z cannot have another occurrence of y or $\tilde{y}$ except as a prefix or a suffix (otherwise, w would have an internal factor in the $G_{\mathcal{A}}$-orbit of x). Thus z is a complete return word to $\{y, \tilde{y}\}$. Consequently, $\delta_i(z)$ is a complete return word to the $G_{\mathcal{A}}$-orbit of x and thus $\delta_i(z) = w$, which implies that $u = z$ and that u is a palindrome.

Now, the $G_{\mathcal{A}}$-orbit of any word $w = \delta_i(u)$ with u palindrome has two elements. Indeed, either w is even and $w^{-1} = \tau(w)$, or w is odd and $w^{-1} = w$. Thus such a w is fixed by a nontrivial element of $G_{\mathcal{A}}$. ∎

Example 4.28 Let S be the specular set of Example 4.21. Since it is a doubling of the Fibonacci set (which is Sturmian and thus full), it is $G_{\mathcal{A}}$-full with respect to the group $G_{\mathcal{A}}$ generated by the antimorphism σ and the morphism τ of Example 4.25. The $G_{\mathcal{A}}$-orbit of $x = a$ is the set $X = \{a, c\}$. The set of complete return words to X (see also Section 6) is given by

$$\mathcal{CR}_S(X) = \{aa, aba, cc, cdc\}.$$

The four words are palindromes and thus they are fixed by $\sigma\tau$.

As another example, consider $x = ab$. Its $G_{\mathcal{A}}$-orbit is the set $X = \{ab, ba, cd, dc\}$ and the set of complete return words to X is given by

$$\mathcal{CR}_S(X) = \{aba, baab, bab, cdc, dccd, dcd\}.$$

Each of them is a palindrome, thus is fixed by $\sigma\tau$.

Example 4.29 Let S be the specular set of Example 4.22. Since it is a doubling of the Fibonacci set (which is Sturmian and thus full), it is $G_{\mathcal{A}}$-full with respect to the group $G_{\mathcal{A}}$ generated by the map σ taking the inverse (that is fixing a, c and exchanging b and d) and the morphism τ (which exchanges a, c and b, d respectively). The $G_{\mathcal{A}}$-orbit of $x = a$ is the set $X = \{a, c\}$. We have

$$\mathcal{CR}_S(X) = \{abc, ac, ca, cda\}.$$

The four words are fixed by $\sigma\tau$. As another example, consider $x = ab$. Then $X = \{ab, bc, cd, da\}$ and $\mathcal{CR}_S(X) = \{abc, bcad, bcd, cda, dab, dacb\}$. Each of them is fixed by some nontrivial element of $G_{\mathcal{A}}$.

5. Linear involutions

In this section we define linear involutions and connections. We prove that the natural coding of a linear involution without connections is a specular set (Theorem 5.9).

5.1. Definition

Let A be an alphabet of cardinality k with an involution θ and the corresponding specular group G_θ. Note that we allow θ to have fixed points. This leads to a definition of linear involutions which is somewhat more general than the one used in [15, 10].

We consider two copies $I \times \{0\}$ and $I \times \{1\}$ of an open interval I of the real line and denote $\hat{I} = I \times \{0, 1\}$. We call the sets $I \times \{0\}$ and $I \times \{1\}$ the two *components* of $\hat{I}$. We consider each component as an open interval.

A *generalized permutation* on A of type (ℓ, m), with $\ell + m = k$, is a bijection $\pi : \{1, 2, \ldots, k\} \to A$. We represent it by a two line array

$$\pi = \begin{pmatrix} \pi(1)\ \pi(2)\ \ldots\pi(\ell) \\ \pi(\ell+1)\ \ldots\pi(\ell+m) \end{pmatrix}$$

A *length data* associated with (ℓ, m, π) is a nonnegative vector $\lambda \in \mathbb{R}_+^A = \mathbb{R}_+^k$ such that

$$\lambda_{\pi(1)} + \ldots + \lambda_{\pi(\ell)} = \lambda_{\pi(\ell+1)} + \ldots + \lambda_{\pi(k)} \text{ and } \lambda_a = \lambda_{a^{-1}} \text{ for all } a \in A.$$

We consider a partition of $I \times \{0\}$ (minus $\ell - 1$ points) in ℓ open intervals $I_{\pi(1)}, \ldots, I_{\pi(\ell)}$ of lengths $\lambda_{\pi(1)}, \ldots, \lambda_{\pi(\ell)}$ and a partition of $I \times \{1\}$ (minus $m-1$ points) in m open intervals $I_{\pi(\ell+1)}, \ldots, I_{\pi(\ell+m)}$ of lengths $\lambda_{\pi(\ell+1)}, \ldots, \lambda_{\pi(\ell+m)}$. Let Σ be the set of $k-2$ *division points* separating the intervals I_a for $a \in A$.

The *linear involution* on I relative to these data is the map $T = \sigma_2 \circ \sigma_1$ defined on the set $\hat{I} \setminus \Sigma$ as the composition of two involutions defined as follows.

(i) The first involution σ_1 is defined on $\hat{I} \setminus \Sigma$. It is such that for each $a \in A$, its restriction to I_a is either a translation or a symmetry from I_a onto $I_{a^{-1}}$.

(ii) The second involution exchanges the two components of $\hat{I}$. It is defined for $(x, \delta) \in \hat{I}$ by $\sigma_2(x, \delta) = (x, 1 - \delta)$. The image of z by σ_2 is called the *mirror image* of z.

We also say that T is a linear involution on I and relative to the alphabet A or that it is a k-linear involution to express the fact that the alphabet A has k elements.

Example 5.1 Let $A = \{a, b, c, d, a^{-1}, b^{-1}, c^{-1}, d^{-1}\}$ and

$$\pi = \begin{pmatrix} a & b & a^{-1} & c \\ c^{-1} & d^{-1} & b^{-1} & d \end{pmatrix}.$$

Let T be the 8-linear involution corresponding to the length data represented in Figure 5.1 (we represent $I \times \{0\}$ above $I \times \{1\}$) with the assumption that the restriction of σ_1 to I_a and I_d is a symmetry while its restriction to I_b, I_c is a translation.

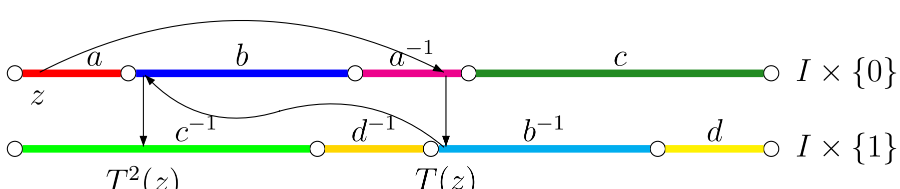

Figure 5.1: A linear involution.

We indicate on the figure the effect of the transformation T on a point z located in the left part of the interval I_a. The point $\sigma_1(z)$ is located in the right part of $I_{a^{-1}}$ and the point $T(z) = \sigma_2\sigma_1(z)$ is just below on the left of $I_{b^{-1}}$. Next, the point $\sigma_1 T(z)$ is located on the left part of I_b and the point $T^2(z)$ just below.

Thus the notion of linear involution is an extension of the notion of interval exchange transformation in the following sense. Assume that

(i) $\ell = m$,

(ii) for each letter $a \in A$, the interval I_a belongs to $I \times \{0\}$ if and only if $I_{a^{-1}}$ belongs to $I \times \{1\}$,

(iii) the restriction of σ_1 to each subinterval is a translation.

Then, the restriction of T to $I \times \{0\}$ is an interval exchange (and so is its restriction to $I \times \{1\}$ which is the inverse of the first one). Thus, in this case, T is a pair of mutually inverse interval exchange transformations.

It is also an extension of the notion of interval exchange with flip [30, 31]. Assume again conditions (i) and (ii), but now that the restriction of σ_1 to at least one subinterval is a symmetry. Then the restriction of T to $I \times \{0\}$ is an interval exchange with flip.

Note that for convenience we consider in this paper interval exchange transformations defined by a partition of an open interval minus $\ell - 1$ points in ℓ open intervals. The usual notion of interval exchange transformation uses a partition of a semi-interval in a finite number of semi-intervals.

A linear involution T is a bijection from $\hat{I} \setminus \Sigma$ onto $\hat{I} \setminus \sigma_2(\Sigma)$. Since σ_1, σ_2 are involutions and $T = \sigma_2 \circ \sigma_1$, the inverse of T is $T^{-1} = \sigma_1 \circ \sigma_2$.

The set Σ of division points is also the set of singular points of T and their mirror images are the singular points of T^{-1} (which are the points where T (resp. T^{-1}) is not defined). Note that these singular points z may be 'false' singularities, in the sense that T can have a continuous extension to an open neighborhood of z.

Two particular cases of linear involutions deserve attention.

A linear involution T on the alphabet A relative to a generalized permutation π of type (ℓ, m) is said to be *non-orientable* if there are indices $i, j \leq \ell$ such that $\pi(i) = \pi(j)^{-1}$ (and thus indices $i, j \geq \ell + 1$ such that $\pi(i) = \pi(j)^{-1}$). In other words, there is some $a \in A$ for which I_a and $I_{a^{-1}}$ belong to the same component of $\hat{I}$. Otherwise T is said to be *orientable*.

A linear involution $T = \sigma_2 \circ \sigma_1$ on I relative to the alphabet A is said to be *coherent* if, for each $a \in A$, the restriction of σ_1 to I_a is a translation if and only if I_a and $I_{a^{-1}}$ belong to distinct components of $\hat{I}$.

Example 5.2 The linear involution of Example 5.1 is coherent.

Linear involutions which are orientable and coherent correspond to interval exchange transformations, whereas orientable but non-coherent linear involutions are interval exchanges with flip.

Orientable linear involutions correspond to orientable laminations (see [10]), whereas coherent linear involutions correspond to orientable surfaces. Thus coherent non-orientable involutions correspond to non-orientable laminations on orientable surfaces.

5.2. Minimal involutions

A *connection* of a linear involution T is a triple (x, y, n) where x is a singularity of T^{-1}, y is a singularity of T, $n \geq 0$ and $T^n x = y$.

Example 5.3 Let us consider the linear involution T which is the same as in Example 5.1 but such that the restriction of σ_1 to I_c is a symmetry. Thus T is not coherent. We assume that $I =]0, 1[$, that $\lambda_a = \lambda_d$. Let $x = (1 - \lambda_d, 0)$ and $y = (\lambda_a, 0)$.

Then x is a singularity of T^{-1} ($\sigma_2(x)$ is the left endpoint of I_d), y is a singularity of T (it is the right endpoint of I_a) and $T(x) = y$. Thus $(x, y, 1)$ is a connection.

Example 5.4 Let T be the linear involution on $I =]0, 1[$ represented in Figure 5.2. We assume that the restriction of σ_1 to I_a is a translation whereas the restriction to I_b and I_c is a symmetry. We choose $(3-\sqrt{5})/2$ for the length of the interval I_c (or I_b). With this choice, T has no connections.

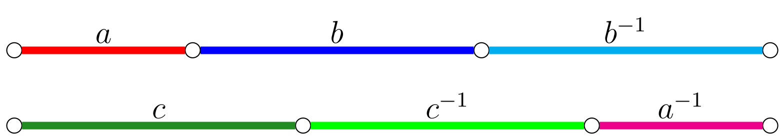

Figure 5.2: A linear involution without connections.

Let T be a linear involution without connections. Let

$$O = \bigcup_{n\geq 0} T^{-n}(\Sigma) \quad \text{and} \quad \hat{O} = O \cup \sigma_2(O) \tag{5.1}$$

be respectively the negative orbit of the singular points and its closure under mirror image. Then T is a bijection from $\hat{I} \setminus \hat{O}$ onto itself. Indeed, assume that $T(z) \in \hat{O}$. If $T(z) \in O$ then $z \in O$. Next if $T(z) \in \sigma_2(O)$, then $T(z) \in \sigma_2(T^{-n}(\Sigma)) = T^n(\sigma_2(\Sigma))$ for some $n \geq 0$. We cannot have $n = 0$ since $\sigma_2(\Sigma)$ is not in the image of T. Thus $z \in T^{n-1}(\sigma_2(\Sigma)) = \sigma_2(T^{-n+1}(\Sigma)) \subset \sigma_2(O)$. Therefore in both cases $z \in \hat{O}$. The converse implication is proved in the same way.

A linear involution T on I without connections is minimal if for any point $z \in \hat{I} \setminus \hat{O}$ the nonnegative orbit of z is dense in $\hat{I}$.

Note that when a linear involution is orientable, that is, when it is a pair of interval exchange transformations (with or without flips), the interval exchange transformations can be minimal although the linear involution is not since each component of $\hat{I}$ is stable by the action of T. Moreover, it is shown in [16] that non-coherent linear involutions are almost surely not minimal.

Example 5.5 Let us consider the non-coherent linear involution T which is the same as in Example 5.1 but such that the restriction of σ_1 to I_c is a symmetry, as in Example 5.3. We assume that $I =]0, 1[$, that $\lambda_a = \lambda_d$ and that $1/4 < \lambda_c < 1/2$ and that $\lambda_a + \lambda_b < 1/2$. Let $x = 1/2 + \lambda_c$ and $z = (x, 0)$ (see Figure 5.3). We have then $T^3(z) = z$, showing that T is not minimal. Indeed, since $z \in I_c$, we have $T(z) = (1-x, 0) = (1/2-\lambda_c, 0)$. Since $T(z) \in I_a$ we have $T^2(z) = ((\lambda_a + \lambda_b) + (\lambda_a - 1 + x), 1) = (x - \lambda_c, 1) = (1/2, 1)$. Finally, since $T^2(z) \in I_{d^{-1}}$, we obtain $(1, 0) - T^3(z) = T^2(z) - (\lambda_c, 1) = (1, 0) - z$ and thus $T^3(z) = z$.

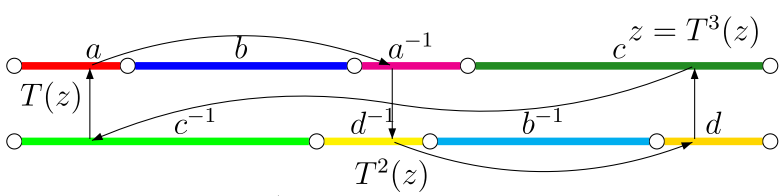

Figure 5.3: A non-coherent linear involution.

The following result (already proved in [12, Proposition 4.2] for the class of coherent involutions) is [10, Proposition 3.7]. The proof uses Keane's theorem proving that an interval exchange transformation without connections is minimal [24].

Proposition 5.6 *Let T be a linear involution without connections on I. If T is non-orientable, it is minimal. Otherwise, its restriction to each component of $\hat{I}$ is minimal.*

5.3. Natural coding

Let T be a linear involution on I, let $\hat{I} = I \times \{0, 1\}$ and let $\hat{O}$ be the set defined by Equation (5.1).

Given $z \in \hat{I} \setminus \hat{O}$, the *infinite natural coding* of T relative to z is the infinite word $\Sigma_T(z) = a_0 a_1 \dots$ on the alphabet A defined by

$$a_n = a \quad \text{if} \quad T^n(z) \in I_a.$$

We first observe that the infinite word $\Sigma_T(z)$ is reduced. Indeed, assume that $a_n = a$ and $a_{n+1} = a^{-1}$ with $a \in A$. Set $x = T^n(z)$ and $y = T(x) = T^{n+1}(z)$. Then $x \in I_a$ and $y \in I_{a^{-1}}$. But $y = \sigma_2(u)$ with $u = \sigma_1(x)$. Since $x \in I_a$, we have $u \in I_{a^{-1}}$. This implies that $y = \sigma_2(u)$ and u belong to the same component of $\hat{I}$, a contradiction.

We denote by $\mathcal{L}(T)$ the set of factors of the infinite natural codings of T. We say that $\mathcal{L}(T)$ is the *natural coding* of T.

Example 5.7 Let T be the linear involution of Example 5.4. The words of length at most 3 of $S = \mathcal{L}(T)$ are represented in Figure 5.4.

The set S can actually be defined directly as the set of factors of the substitution

$$f : a \mapsto cb^{-1}, \quad b \mapsto c, \quad c \mapsto ab^{-1}.$$

which extends to an automorphism of the free group on $\{a, b, c\}$ (see [10]).

The following is Proposition 5.3 in [10].

Proposition 5.8 *The natural coding of a linear involution is closed under taking inverses.*

We prove the following result.

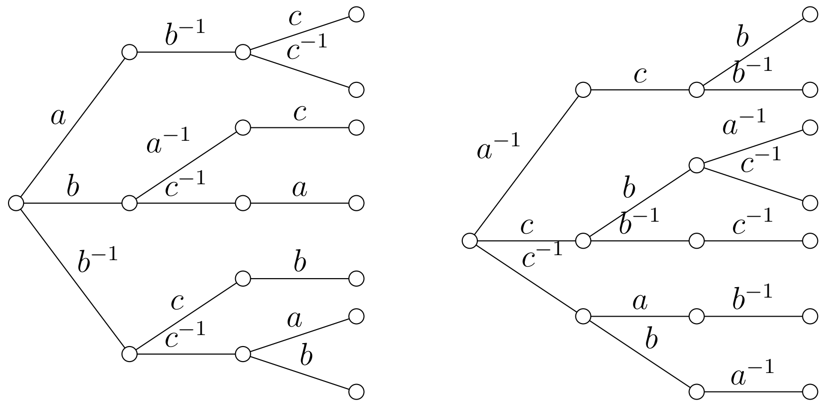

Figure 5.4: The words of length at most 3 of S.

Theorem 5.9 *The natural coding of a linear involution without connections is a specular set.*

Proof. Let T be a linear involution without connections. By Proposition 5.8, the set $\mathcal{L}(T)$ is symmetric. Since it is by definition bi-extendable and formed of reduced words, it is a laminary set. By [20, Theorem 9.5], $\mathcal{L}(T)$ is a tree set of characteristic 2. Thus $\mathcal{L}(T)$ is specular. ∎

We now present an example of a linear involution on an alphabet A where the involution θ has fixed points.

Example 5.10 Let $A = \{a, b, c, d\}$ be as in Example 3.2 (in particular, $d = b^{-1}$, $a = a^{-1}$, $c = c^{-1}$). Let T be the linear involution represented in

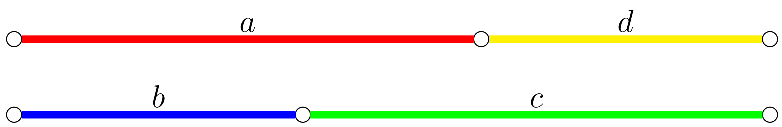

Figure 5.5: A linear involution on $A = \{a, b, c, d\}$.

Figure 5.5 with σ_1 being a translation on I_b and a symmetry on I_a, I_c. Choosing $(3 - \sqrt{5})/2$ for the length of I_b, the involution is without connections. Thus $S = \mathcal{L}(T)$ is a specular set. Let us show it is equal to the specular set obtained by the doubling transducer in Example 4.22. Indeed, consider the interval exchange V on the interval $Y =]0, 2[$ represented in Figure 5.6 on the right, which is obtained by using two copies of the interval exchange U defining the Fibonacci set (represented in Figure 5.6 on the left).

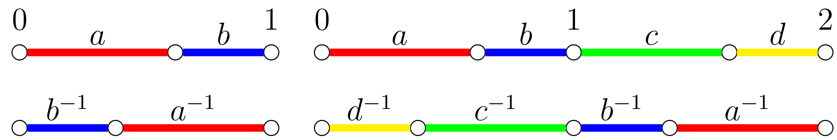

Figure 5.6: Interval exchanges U and V for the Fibonacci set and its doubling.

Let $X =]0,1[\times \{0,1\}$ and let $\alpha : Y \to X$ be the map defined by

$$\alpha(z) = \begin{cases} (z,0) & \text{if } z \in]0,1[\\ (2-z,1) & \text{otherwise.} \end{cases}$$

Then $\alpha \circ V = T \circ \alpha$ and thus $\mathcal{L}(V) = \mathcal{L}(T)$. The interval exchange V is actually the orientation covering of the linear involution T (see [10]).

6. Return words

In this section we introduce three variants of the notion of return words, namely complete, right and mixed return words. We prove several results concerning sets of return words (Theorems 6.6, 6.9, 6.12). We also 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).

6.1. Cardinality Theorems for return words

In this section, we introduce several notions of return words: complete return words, right (or left) return words and mixed return words. For each of them, we prove a cardinality theorem (Theorems 6.6, 6.9 and 6.12).

Here, when we consider a recurrent set S containing the alphabet A, we implicitly assume that all words of S are on the alphabet A.

6.1.1. Complete return words

Let S be a factorial set of words and let $X \subset S$ be a set of nonempty words. A *complete return word* to X is a word of S with a proper prefix in X, a proper suffix in X but no internal factor in X. We denote by $\mathcal{CR}_S(X)$ the set of complete return words to X.

The set $\mathcal{CR}_S(X)$ is a bifix code. If S is uniformly recurrent, $\mathcal{CR}_S(X)$ is finite for any finite set X. For $x \in S$, we denote $\mathcal{CR}_S(x)$ instead of $\mathcal{CR}_S(\{x\})$. Thus $\mathcal{CR}_S(x)$ is the usual notion of a complete return word (see [22] for example).

Example 6.1 Let S be the specular set of Example 4.22. One has

$$\begin{aligned}\mathcal{CR}_S(a) &= \{abca, abcda, acda\}\\ \mathcal{CR}_S(b) &= \{bcab, bcdacdab, bcdacdacdab\}\\ \mathcal{CR}_S(c) &= \{cabc, cdabc, cdac\}\\ \mathcal{CR}_S(d) &= \{dabcabcabcd, dabcabcd, dacd\}.\end{aligned}$$

The following result is proved in [20, Theorem 5.2].

Theorem 6.2 *Let S be a recurrent neutral set containing the alphabet A. For any finite nonempty bifix code $X \subset S$ with empty kernel, we have*

$$\mathrm{Card}(\mathcal{CR}_S(X)) = \mathrm{Card}(X) + \mathrm{Card}(A) - \chi(S). \tag{6.1}$$

As a consequence of Theorem 6.2, one has the following statement.

Corollary 6.3 *Let S be a recurrent specular set on the alphabet A. For any finite nonempty bifix code $X \subset S$ with empty kernel, one has*

$$\mathrm{Card}(\mathcal{CR}_S(X)) = \mathrm{Card}(X) + \mathrm{Card}(A) - 2.$$

The following example illustrates Corollary 6.3.

Example 6.4 Let S be the specular set on the alphabet $A = \{a, b, c, d\}$ of Example 4.2. We have

$$\mathcal{CR}_S(\{a, b\}) = \{ab, acda, bca, bcda\}.$$

It has four elements in agreement with Corollary 6.3.

We note that when X is a finite S-maximal bifix code of S-degree d with kernel $K(X)$, the set $\mathcal{CR}_S(X)$ has the following property. For any set K such that $K(X) \subset K \subset X$ with $K \neq X$, the set $Y = K \cup \mathcal{CR}_S(X \setminus K)$ is an S-maximal bifix code of S-degree $d_S(X)+1$. The code X is the derived code of Y (see [4, Section 4.3]). This gives a connection between Equations (4.1) and (6.1). By Equation (4.1), we have

$$\mathrm{Card}(Y) = (d+1)(\mathrm{Card}(A) - \chi(S)) + \chi(S) = \mathrm{Card}(X) + \mathrm{Card}(A) - \chi(S).$$

Thus

$$\begin{aligned}
\mathrm{Card}(\mathcal{CR}_S(X \setminus K)) &= \mathrm{Card}(Y) - \mathrm{Card}(K) \\
&= \mathrm{Card}(X) - \mathrm{Card}(K) + \mathrm{Card}(A) - \chi(S) \\
&= \mathrm{Card}(X \setminus K) + \mathrm{Card}(A) - \chi(S)
\end{aligned}$$

which is Formula (6.1) since $X \setminus K$ is a bifix code with empty kernel.

6.1.2. Right return words

Let S be a factorial set. For any nonempty word $x \in S$, a *right return word* to x in S is a word w such that xw is a complete return word to x. One defines symmetrically the *left return words* to $x \in S$ as the words w such that wx is a complete return word. We denote by $\mathcal{R}_S(x)$ the set of right return words to x in S and by $\mathcal{R}'_S(x)$ the corresponding set of left return words.

Note that when S is a laminary set $\mathcal{R}_S(x)^{-1} = \mathcal{R}'_S(x^{-1})$.

Proposition 6.5 *Let S be a specular set and let $x \in S$ be a nonempty word. All the words of $\mathcal{R}_S(x)$ are even.*

Proof. If $w \in \mathcal{R}_S(x)$, we have $xw = vx$ for some $v \in S$. If x is odd, assume that $x \in S_{0,1}$. Then $w \in S_{1,1}$. Thus w is even. If x is even, assume that $x \in S_{0,0}$. Then $w \in S_{0,0}$ and w is even again. ■

Theorem 6.6 (Cardinality Theorem for right return words) *Let S be a recurrent specular set. For any $x \in S$, the set $\mathcal{R}_S(x)$ has* $\mathrm{Card}(A) - 1$ *elements.*

Proof. This follows directly from Corollary 6.3 with $X = \{x\}$ since $\mathrm{Card}(\mathcal{R}_S(x)) = \mathrm{Card}(\mathcal{CR}_S(x))$. ■

Example 6.7 Let S be the specular set of Example 4.22. We have

$$\begin{aligned}
\mathcal{R}_S(a) &= \{bca, bcda, cda\}, \\
\mathcal{R}_S(b) &= \{cab, cdacdab, cdacdacdab\}, \\
\mathcal{R}_S(c) &= \{abc, dabc, dac\}, \\
\mathcal{R}_S(d) &= \{abcabcd, abcabcabcd, acd\}.
\end{aligned}$$

It is shown in [7] that if S is a (uniformly) recurrent tree set of characteristic 1 containing the alphabet B, then for any $x \in S$, one has $\mathrm{Card}(\mathcal{R}_S(x)) = \mathrm{Card}(B)$. The relation with Theorem 6.6 is as follows. Let X be the even code and let $X_0 = X \cap S_{0,0}, X_1 = X \cap S_{1,1}$. Thus $X = X_0 \cup X_1$.

One has $\mathrm{Card}(X_0) = \mathrm{Card}(A) - 1$ by Theorem 4.16 (indeed, $\mathrm{Card}(X) = 2\,\mathrm{Card}(A) - 2$ and $\mathrm{Card}(X_0) = \mathrm{Card}(X_1)$).

Let f be a coding morphism for X. Then for any $x \in S_{0,0}$, the set $\mathcal{R}_S(x)$ is in bijection, via the decoding by X_0, with the set of right return words to $f^{-1}(x)$. Since $f^{-1}(S_{0,0})$ is a tree set on $B_0 = f^{-1}(X_0)$, the set $\mathcal{R}_S(x)$ has $\mathrm{Card}(A) - 1$ elements, in agreement with Theorem 6.6.

6.1.3. Mixed return words

Let S be a laminary set. For $w \in S$ such that $w \neq w^{-1}$, we consider complete return words to the set $X = \{w, w^{-1}\}$.

Example 6.8 Let T be the linear involution of Example 5.4. We have

$$\begin{aligned}
\mathcal{CR}_S(\{a, a^{-1}\}) &= \{ab^{-1}cba^{-1}, ab^{-1}cbc^{-1}a, a^{-1}cb^{-1}c^{-1}a,\\
&\qquad ab^{-1}c^{-1}ba^{-1}, a^{-1}cbc^{-1}a, a^{-1}cb^{-1}c^{-1}ba^{-1}\}\\
\mathcal{CR}_S(\{b, b^{-1}\}) &= \{ba^{-1}cb, ba^{-1}cb^{-1}, bc^{-1}ab^{-1}, b^{-1}cb, b^{-1}c^{-1}ab^{-1}, b^{-1}c^{-1}b\},\\
\mathcal{CR}_S(\{c, c^{-1}\}) &= \{cba^{-1}c, cbc^{-1}, cb^{-1}c^{-1}, c^{-1}ab^{-1}c, c^{-1}ab^{-1}c^{-1}, c^{-1}ba^{-1}c\}.
\end{aligned}$$

Theorem 6.9 *Let S be a recurrent specular set containing the alphabet A. For any $w \in S$ such that $w \neq w^{-1}$, the set of complete return words to $\{w, w^{-1}\}$ has $\mathrm{Card}(A)$ elements.*

Proof. The statement results directly of Corollary 6.3. ∎

Example 6.10 Let S be the specular set of Example 4.22. In view of the values of $\mathcal{CR}_S(b)$ and $\mathcal{CR}_S(d)$ given in Example 6.1, we have

$$\mathcal{CR}_S(\{b, d\}) = \{bcab, bcd, dab, dacd\}.$$

Two words u, v are said to *overlap* if a nonempty suffix of one of them is a prefix of the other. In particular a nonempty word overlaps with itself.

We now consider the return words to $\{w, w^{-1}\}$ with w such that w and w^{-1} do not overlap. This is true for every w in a laminary set S where the involution θ has no fixed point, in particular when S is the natural coding of a linear involution. In this case, the group G_θ is free and for any $w \in S$, the words w and w^{-1} do not overlap.

With a complete return word u to $\{w, w^{-1}\}$, we associate a word $N(u)$ obtained as follows. If u has w as prefix, we erase it and if u has a suffix w^{-1}, we also erase it. Note that these two operations can be made in any order since w and w^{-1} cannot overlap.

The *mixed return words to* w are the words $N(u)$ associated with complete return words u to $\{w, w^{-1}\}$. We denote by $\mathcal{MR}_S(w)$ the set of mixed return words to w in S.

Note that $\mathcal{MR}_S(w)$ is symmetric and that $w\mathcal{MR}_S(w)w^{-1} = \mathcal{MR}_S(w^{-1})$. Note also that if S is orientable, then

$$\mathcal{MR}_S(w) = \mathcal{R}_S(w) \cup \mathcal{R}_S(w)^{-1} = \mathcal{R}_S(w) \cup \mathcal{R}'_S(w^{-1}).$$

The reason for this definition comes from the fact that, when S is the natural coding of a linear involution, we are interested in the transformation induced on $I_w \cup \sigma_2(I_w)$, where $I_w = I_{b_0} \cap T^{-1}(I_{b_1}) \cap \ldots \cap T^{-m+1}(I_{b_{m-1}})$ for a word $w = b_0 b_1 \cdots b_{m-1}$ (see [10]). The natural coding of a point in I_w begins with w while the natural coding of a point z in $\sigma_2(I_w)$ 'ends' with w^{-1} in the sense that the natural coding of $T^{-|w|}(z)$ begins with w^{-1}.

Example 6.11 Let T be the linear involution of Example 5.4. We have

$$\begin{aligned}
\mathcal{MR}_S(a) &= \{b^{-1}cb, b^{-1}cbc^{-1}a, a^{-1}cb^{-1}c^{-1}a, b^{-1}c^{-1}b, a^{-1}cbc^{-1}a, a^{-1}cb^{-1}c^{-1}b\}\\
\mathcal{MR}_S(b) &= \{a^{-1}cb, a^{-1}c, c^{-1}a, b^{-1}cb, b^{-1}c^{-1}a, b^{-1}c^{-1}b\},\\
\mathcal{MR}_S(c) &= \{ba^{-1}c, b, b^{-1}, c^{-1}ab^{-1}c, c^{-1}ab^{-1}, c^{-1}ba^{-1}c\}.
\end{aligned}$$

Observe that any uniformly recurrent biinfinite word x such that $F(x) = S$ can be uniquely written as a concatenation of mixed return words (see Figure 6.1). Note that successive occurrences of w may overlap but that successive occurrences of w and w^{-1} cannot.

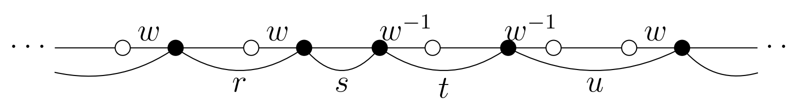

Figure 6.1: A uniformly recurrent infinite word factorized as an infinite product $\cdots rstu \cdots$ of mixed return words to w.

We have the following cardinality result.

Theorem 6.12 (Cardinality Theorem for mixed return words) *Let S be a recurrent specular set on the alphabet A. For any $w \in S$ such that w, w^{-1} do not overlap, the set $\mathcal{MR}_S(w)$ has* $\mathrm{Card}(A)$ *elements.*

Proof. This is a direct consequence of Theorem 6.9 since $\mathrm{Card}(\mathcal{MR}_S(w)) = \mathrm{Card}(\mathcal{CR}_S(\{w, w^{-1}\})$ when w and w^{-1} do not overlap. ∎

Note that the bijection between $\mathcal{CR}_S(w, w^{-1})$ and $\mathcal{MR}_S(w)$ is illustrated in Figure 6.1.

Example 6.13 Let S be the specular set of Example 4.22. The value of $\mathcal{CR}_S(b, d)$ is given in Example 6.10. Since b, d do not overlap,

$$\mathcal{MR}_S(b) = \{cab, c, dac, dab\}$$

has four elements in agreement with Theorem 6.12.

As a corollary, we obtain the following result.

Corollary 6.14 *Let S be the natural coding of a linear involution without connections on the alphabet A. For any $w \in S$, the set $\mathcal{MR}_S(w)$ has* $\mathrm{Card}(A)$ *elements.*

6.2. First Return Theorem

By [7, Theorem 4.5], the set of right return words to a given word in a recurrent tree set of characteristic 1 containing the alphabet A is a basis of the free group on A. We will see a counterpart of this result for recurrent specular sets.

Let S be a specular set. The *even subgroup* is the group formed by the even words. It is a subgroup of index 2 of G_θ with symmetric rank $2(\mathrm{Card}(A) - 1)$ by (3.2) generated by the even code. Since no even word is its own inverse (by Proposition 4.11), it is a free group. Thus its rank is $\mathrm{Card}(A) - 1$.

Theorem 6.15 (First Return Theorem) *Let S be a recurrent specular set. For any $w \in S$, the set of right return words to w is a basis of the even subgroup.*

Proof. We first consider the case where w is even. Let $f : B^* \to A^*$ be a coding morphism for the even code $X \subset S$. Consider the partition $(S_{i,j})$, as in Proposition 4.11, and set $X_0 = X \cap S_{0,0}$, $X_1 = X \cap S_{1,1}$. By Theorem 4.13, the set $f^{-1}(S)$ is the union of the two recurrent tree sets of characteristic 1, $T_0 = f^{-1}(S_{0,0})$ and $T_1 = f^{-1}(S_{1,1})$ on the alphabets $B_0 = f^{-1}(X_0)$ and $B_1 = f^{-1}(X_1)$ respectively. We may assume that $w \in S_{0,0}$. Then $\mathcal{R}_S(w)$ is the image by f of the set $R = \mathcal{R}_{T_0}(f^{-1}(w))$. By [7, Theorem 4.5], the set R is a basis of the free group on B_0. Thus $\mathcal{R}_S(w)$ is a basis of the image of F_{B_0} by f, which is the even subgroup.

Suppose now that w is odd. Since the even code is an S-maximal bifix code, there exists an odd word u such that $uw \in S$. Then $\mathcal{R}_S(uw) \subset \mathcal{R}_S(w)^*$. By what precedes, the set $\mathcal{R}_S(uw)$ generates the even subgroup and thus the group generated by $\mathcal{R}_S(w)$ contains the even subgroup. Since all words in $\mathcal{R}_S(w)$ are even, the group generated by $\mathcal{R}_S(w)$ is contained in the even subgroup, whence the equality. We conclude by Theorem 6.6. ∎

Example 6.16 Let S be the specular set of Example 4.22. The sets of right return words to a, b, c, d are given in Example 6.7. Each one is a basis of the even subgroup.

Concerning mixed return words, we have the following statement.

Theorem 6.17 *Let S be a recurrent specular set. For any $w \in S$ such that w, w^{-1} do not overlap, the set $\mathcal{MR}_S(w)$ is a monoidal basis of the group G_θ.*

Proof. Since w and w^{-1} do not overlap, we have $\mathcal{R}_S(w) \subset \mathcal{MR}_S(w)^*$. Thus, by Theorem 6.15, the group $\langle \mathcal{MR}_S(w) \rangle$ contains the even subgroup. But $\mathcal{MR}_S(w)$ always contains odd words. Indeed, assume that $w \in S_{i,j}$. Then $w^{-1} \in S_{1-j,1-i}$ and thus any $u \in \mathcal{MR}_S(w)$ such that $wuw^{-1} \in S$ is odd. Since the even group is a maximal subgroup of G_θ, this implies that $\mathcal{MR}_S(w)$ generates the group G_θ. Finally since $\mathcal{MR}_S(w)$ has $\mathrm{Card}(A)$ elements by Theorem 6.12, we obtain the conclusion by Proposition 3.10. ∎

Example 6.18 Let S be the specular set of Example 4.22. We have seen in Example 6.13 that

$$\mathcal{MR}_S(b) = \{c, cab, dab, dac\}.$$

This set is a monoidal basis of G_θ in agreement with Theorem 6.17.

Since, in the free group, a reduced word w and its inverse do not overlap, we have the following corollary of Theorem 6.17 in the case where the involution θ has no fixed points. A geometric proof and interpretation is given in [10].

Corollary 6.19 *Let S be the natural coding of a linear involution without connections on the alphabet $A = B \cup B^{-1}$. For any $w \in S$, the set $\mathcal{MR}_S(w)$ is a monoidal basis of F_B.*

Example 6.20 Let T be the linear involution of Example 5.4. We have seen in Example 6.11 that $\mathcal{MR}_S(b) = \{a^{-1}cb, a^{-1}c, c^{-1}a, b^{-1}cb, b^{-1}c^{-1}a, b^{-1}c^{-1}b\}$. It is a monoidal basis of the free group on $\{a, b, c\}$.

7. Freeness and Saturation Theorems

In this section we consider two notions concerning sets of generators of a subgroup H in a specular group, namely free subsets and the set of prime words with respect to H. 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). To prove the last result we use the notion of coset automaton.

7.1. Freeness Theorem

Let θ be an involution on A and let G_θ be the corresponding specular group. A symmetric set X is free if it is a monoidal basis of a subgroup H of the group G_θ. Thus a symmetric set $X \subset G_\theta$ is *free* if for $x_1, x_2, \ldots, x_n \in X$, the product $x_1 x_2 \cdots x_n$ cannot reduce to 1 unless $x_i = x_{i+1}^{-1}$ for some i with $1 \le i < n$.

The following is essentially Theorem 5.1 in [7].

Theorem 7.1 (Freeness Theorem) *A laminary set S is acyclic if and only if any symmetric bifix code $X \subset S$ is free.*

The proof is identical with that of Theorem 5.1 in [7], using the *incidence graph* of a set X, which is the undirected graph $\mathcal{G}_X$ defined as follows. Let P be the set of proper prefixes of X and let Q be the set of its proper suffixes. Set $P' = P \setminus \{\varepsilon\}$ and $Q' = Q \setminus \{\varepsilon\}$. The set of vertices of $\mathcal{G}_X$ is the disjoint union of P' and Q'. The edges of $\mathcal{G}_X$ are the pairs (p, q) for $p \in P'$ and $q \in Q'$ such that $pq \in X$. As for the extension graph, we sometimes denote $1 \otimes P', Q' \otimes 1$ the copies of P', Q' used to define the set of vertices of $\mathcal{G}_X$.

Example 7.2 Let S be a laminary set and let $X = S \cap A^2$ be the bifix code formed of the words of S of length 2. The incidence graph of X is identical with the extension graph $\mathcal{E}(\varepsilon)$.

The following statement is proved in [7, Proposition 5.6]. Recall that a path in an undirected graph is reduced if it does not use twice consecutively the same edge.

Proposition 7.3 *Let S be an acyclic set. For any bifix code $X \subset S$, the following assertions hold.*

(i) *The incidence graph $\mathcal{G}_X$ is acyclic.*
(ii) *The intersection of $P' = P \setminus \{\varepsilon\}$ (resp. $Q' = Q \setminus \{\varepsilon\}$) with each connected component of $\mathcal{G}_X$ is a suffix (resp. prefix) code.*
(iii) *For every reduced path $(v_1, u_1, \ldots, u_n, v_{n+1})$ in $\mathcal{G}_X$ with $u_1, \ldots, u_n \in P'$ and $v_1, \ldots, v_{n+1} \in Q'$, the longest common prefix of v_1, v_{n+1} is a proper prefix of all $v_1, \ldots, v_n, v_{n+1}$.*
(iv) *Symmetrically, for every reduced path $(u_1, v_1, \ldots, v_n, u_{n+1})$ in $\mathcal{G}_X$ with $u_1, \ldots, u_{n+1} \in P'$ and $v_1, \ldots, v_n \in Q'$, the longest common suffix of u_1, u_{n+1} is a proper suffix of $u_1, u_2, \ldots, u_{n+1}$.*

7.2. Cosets

Let X be a symmetric set. We use the incidence graph to define an equivalence relation γ_X on the set P of proper prefixes of X, called the *coset equivalence* of X, as follows. It is the relation defined by $p \equiv q \bmod \gamma_X$ if there is a path (of even length) from $1 \otimes p$ to $1 \otimes q$ or a path (of odd length) from $1 \otimes p$ to $q^{-1} \otimes 1$ in the incidence graph $\mathcal{G}_X$. It is easy to verify that, since X is symmetric, γ_X is indeed an equivalence. The class of the empty word ε is reduced to ε. This definition is an extension to symmetric sets of the equivalence denoted θ_X introduced in [4].

The following statement is the generalization to symmetric bifix codes of Proposition 6.3.5 in [4]. We denote by $\langle X \rangle$ the subgroup generated by X.

Proposition 7.4 *Let X be a symmetric bifix code and let P be the set of its proper prefixes. Let γ_X be the coset equivalence of X and let $H = \langle X \rangle$. For any $p, q \in P$, if $p \equiv q \bmod \gamma_X$, then $Hp = Hq$.*

Proof. Assume that there is a path of even length from p to q. If the path has length 2, then we have $pr, qr \in X$ for some suffix r of X. This implies

$pq^{-1} \in H$ and thus $Hp = Hq$. The general case follows by induction. In the case where there is a path of odd length from p to q^{-1}, there is a path of even length from p to r and an edge from r to q^{-1} for some $r \in P$. Then $Hp = Hr$ by the preceding argument. Since $rq^{-1} \in X$, we have $Hr = Hq$ and the conclusion follows. ■

We now use the coset equivalence γ_X to define the *coset automaton* $\mathcal{C}_X$ of a symmetric bifix code X as follows. The vertices of $\mathcal{C}_X$ are the equivalence classes of γ_X. We denote by $\hat{p}$ the class of p. There is an edge labeled $a \in A$ from s to t if for some $p \in s$ and $q \in t$ (that is, $s = \hat{p}$ and $t = \hat{q}$), one of the following cases occurs (see Figure 7.1):

(i) $pa \in P$ and $pa \equiv q \bmod \gamma_X$,
(ii) or $pa \in X$ and $q = \varepsilon$.

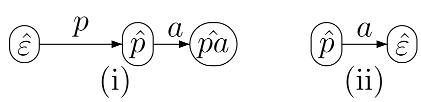

Figure 7.1: The edges of the coset automaton.

Proposition 7.5 *Let X be a symmetric bifix code, let P be its set of proper prefixes and let $H = \langle X \rangle$. If for $p, q \in P$ and a word $w \in A^*$ there is a path labeled w from the class $\hat{p}$ to the class $\hat{q}$, then $Hpw = Hq$.*

Proof. Assume first that w is a letter $a \in A$. It is easy to verify using Proposition 7.4 that in the two cases of the definition of an edge $(\hat{p}, a, \hat{q})$, one has $Hpa = Hq$. Since the coset does not depend on the representative in the class, this implies the conclusion. The general case follows easily by induction. ■

Let A be an alphabet with an involution θ. A directed graph with edges labeled in A is called *symmetric* if there is an edge from p to q labeled a if and only if there is an edge from q to p labeled a^{-1}.

If $\mathcal{G}$ is a symmetric graph and v is a vertex of $\mathcal{G}$, the set of reductions of the labels of paths from v to v is a subgroup of G_θ called the subgroup *described* by $\mathcal{G}$ with respect to v.

A symmetric graph is called *reversible* if for every pair of edges of the form $(v, a, w), (v, a, w')$, one has $w = w'$ (and the symmetric implication since the graph is symmetric).

Proposition 7.6 *Let S be a specular set and let $X \subset S$ be a finite symmetric bifix code. The coset automaton $\mathcal{C}_X$ is reversible. Moreover the subgroup described by $\mathcal{C}_X$ with respect to the class of the empty word is the group generated by X.*

Proof. It is easy to verify that the words of X are labels of paths from $\hat{\varepsilon}$ to $\hat{\varepsilon}$ which do not pass by $\hat{\varepsilon}$ in between. Thus the group described by $\mathcal{C}_X$ with respect to $\hat{\varepsilon}$ contains $H = \langle X \rangle$.

By Proposition 7.5, if there is a path from the class of p to the class of q labeled w, then $Hpw = Hq$. Thus if w belongs to the group described by $\mathcal{C}_X$ (w.r.t. $\hat{\varepsilon}$), it is in H. We have thus proved that the coset automaton describes H.

Let us show now that $\mathcal{C}_X$ is reversible. First, it is symmetric since X is symmetric. Let us show that if (v, a, w) and (v, a, w') are edges of $\mathcal{C}_X$, then $w = w'$. Consider $p, p' \in P$ such that $p \equiv p' \bmod \gamma_X$. Assume that there is an edge labeled a from $\hat{p} = \hat{p'}$ to $\hat{q}$ and to $\hat{q'}$.

Case 1. Suppose that $pa, p'a \in P$. We have to show that $pa \equiv p'a \bmod \gamma_X$. Let u, v be such that $pau, p'av \in X$. It is not possible that there exists a path of odd length from p to p'^{-1} in the incidence graph $\mathcal{G}_X$. Indeed, assume that $p \in S_{i,j}$ and $a \in S_{j,k}$. Let $(p, u_1, \ldots, u_{2m}, p'^{-1})$ with $m \geq 0$ be a path of odd length from p to p'^{-1}. Then each u_{2t} for $1 \leq t \leq m$ is in $S_{i_t,j}$ and each u_{2t+1} for $0 \leq t \leq m-1$ is in S_{j,ℓ_t} for some $i_t, \ell_t \in \{0, 1\}$. Then $p'^{-1} \in S_{j,\ell_m}$ and thus $p' \in S_{1-\ell_m,1-j}$. But then we cannot have $p'a \in S$. Thus there is a path of even length from p to p' in $\mathcal{G}_X$. This implies that there is a path of even length of the form $(au, p, \ldots, p', av)$. Thus by Proposition 7.3 (iii), there is a path of even length from pa to $p'a$. This implies that $pa \equiv p'a \bmod \gamma_X$.

Case 2. Assume now that $pa \in P$ and $p'a \in X$. For the same reason as in Case 1, there cannot exist a path of odd length from p to p'. Thus there is a path of even length from p to p'. By Proposition 7.3 (iii), this is not possible since otherwise we would have for some word u, a path $(au, p, \ldots, p', a)$ and a is not a proper prefix of the last term of the sequence.

The case where $pa \in X$ and $p'a \in P$ is symmetrical. Finally, if $pa, p'a \in X$, we have $q = q' = \varepsilon$.

This shows that if (v, a, w) and (v, a, w') are edges of $\mathcal{C}_X$, then $w = w'$. Since $\mathcal{C}_X$ is symmetric, it follows that if (v, a, w) and (v', a, w) are edges of $\mathcal{C}_X$, then $v = v'$. Thus $\mathcal{C}_X$ is reversible. ∎

Example 7.7 Let T be the linear involution of Example 5.4 and let $S = \mathcal{L}(T)$. Let X be the set of words of length 3 of S (see Figure 5.4), which is a symmetric bifix code. The incidence graph $\mathcal{G}_X$ is represented in Figure 7.2. The coset automaton $\mathcal{C}_X$ is represented in Figure 7.3 (we only represent one

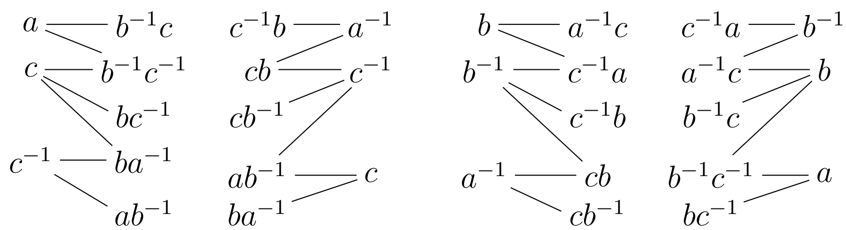

Figure 7.2: The incidence graph of X.

of the edges labeled a and a^{-1}, the other one is understood). The vertex 2 is the class corresponding to the first two trees in Figure 7.2. The vertex 3 corresponds to the two last ones.

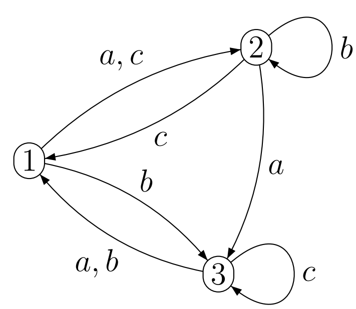

Figure 7.3: The coset automaton.

7.3. Saturation Theorem

Let H be a subgroup of the specular group G_θ and let S be a specular set on A relative to θ. The set of *prime* words in S with respect to H is the set of nonempty words in $H \cap S$ without a proper nonempty prefix in $H \cap S$. Note that the set of prime words with respect to H is a symmetric bifix code. One may verify that it is actually the unique bifix code X such that $X \subset S \cap H \subset X^*$.

The following statement is a generalization of Theorem 5.2 in [7] (Saturation Theorem).

Theorem 7.8 (Saturation Theorem) *Let S be an acyclic laminary set. Any finite symmetric bifix code $X \subset S$ is the set of prime words in S with respect to the subgroup $\langle X \rangle$. Moreover $\langle X \rangle \cap S = X^* \cap S$.*

Proof. Let $H = \langle X \rangle$ and let $Y \subset S$ be the set of prime words with respect to H. Then Y is a symmetric bifix code and thus it is free by Theorem 7.1. Since, by Proposition 7.6, the coset automaton $\mathcal{C}_X$ is reversible, any reduced word is the label of at most one reduced path in $\mathcal{C}_X$. Since any word of X is the label of a reduced path from $\hat{\varepsilon}$ to $\hat{\varepsilon}$ in $\mathcal{C}_X$ which does not pass by $\hat{\varepsilon}$ in-between, this implies that $X \subset Y$. But any $y \in Y$ is the reduction of some product $x_1 x_2 \cdots x_n$ with $x_i \in X$. Since Y is free and contains X, this implies $n = 1$ and $y \in X$. Thus $X = Y$.

The last assertion follows from the fact that, since X is the set of prime words in S with respect to H, one has $H \cap S \subset X^*$. ∎

Note that the hypothesis that X is symmetric is necessary, as shown in the following example.

Example 7.9 Let $A = \{a, b, a^{-1}, b^{-1}\}$. Let S be the set of factors of $(ab^{-1})^\omega \cup (a^{-1}b)^\omega$ (we denote as usual by x^ω the infinite word $xxx\cdots$). Then S is an acyclic laminary set. The set $X = \{a, ba^{-1}\}$ is a bifix code but it is not the set of prime words with respect to $\langle X \rangle$ since $b \in \langle X \rangle \cap S$.

8. Bifix codes and monoidal bases

In this section we prove the Finite Index Basis Theorem (Theorem 8.1) and a converse (Theorem 8.6).

8.1. Finite Index Basis Theorem

The following result is the counterpart for specular sets of the result holding for recurrent tree sets of characteristic 1 (see [8, Theorem 4.4]). The proof is very similar to that of Theorem 4.4 in [8] and we omit some details.

Theorem 8.1 (Finite Index Basis Theorem) *Let S be a recurrent specular set and let $X \subset S$ be a finite symmetric bifix code. Then X is an S-maximal bifix code of S-degree d if and only if it is a monoidal basis of a subgroup of index d.*

The following result is a complement to Theorem 4.4.3 in [4], asserting that if S is a recurrent set, any finite bifix code $X \subset S$ is contained in a finite S-maximal bifix code Z. It shows that when X is symmetric, then Z can be chosen symmetric.

Theorem 8.2 *Let S be a recurrent laminary set. Any finite symmetric bifix code $X \subset S$ is contained in a finite symmetric S-maximal bifix code.*

Proof. Let $X \subset S$ be a finite symmetric bifix code which is not S-maximal. Since X is finite, the number $d = \max\{d_X(w) \mid w \in X\}$ is finite. By Theorem 4.3.12 of [4], X is the kernel of some S-maximal bifix code Z of S-degree $d+1$. Since S is recurrent, by Theorem 4.4.3 of [4], Z is finite. Let us show that Z is symmetric. Indeed, we have by Theorem 4.3.11 in [4], $d_Z(w) = \min\{d+1, d_X(w)\}$. Since X is symmetric, we have $d_X(w) = d_X(w^{-1})$ for any $w \in S$. Indeed, (q, x, p) is a parse of w if and only if (p^{-1}, x^{-1}, q^{-1}) is a parse of w^{-1}. Thus $d_Z(w) = d_Z(w^{-1})$. This implies that Z is symmetric. ∎

Proof of Theorem 8.1. Assume first that X is a finite symmetric S-maximal bifix code of S-degree d. Let P be the set of proper prefixes of X. Let H be the subgroup generated by X.

Let $u \in S$ be a word such that $d_X(u) = d$, or, equivalently, which is not an internal factor of X. Since u can be replaced by any of its right extensions, we may assume that u is odd. Let Q be the set formed of the d suffixes of u which are in P.

Let us first show that the cosets Hq for $q \in Q$ are disjoint. Indeed, $Hp \cap Hq \neq \emptyset$ implies $Hp = Hq$. Any $p, q \in Q$ are comparable for the suffix order. Assuming that q is longer than p, we have $q = tp$ for some $t \in P$. Then $Hp = Hq$ implies $Ht = H$ and thus $t \in H \cap S$. By Theorem 7.8, since S is acyclic and X is symmetric, this implies $t \in X^*$ and thus $t = \varepsilon$. Thus $p = q$.

Let

$$V = \{v \in G_\theta \mid Qv \subset HQ\}$$

where the products Qv and HQ are understood in the group G_θ (that is, with reduction).

For any $v \in V$ the map $p \mapsto q$ from Q into itself defined by $pv \in Hq$ is a permutation of Q. Indeed, suppose that for $p, q \in Q$, one has $pv, qv \in Hr$ for some $r \in Q$. Then rv^{-1} is in $Hp \cap Hq$ and thus $p = q$ by the above argument.

The set V is a subgroup of G_θ. Indeed, $1 \in V$. Next, let $v \in V$. Then for any $q \in Q$, since v defines a permutation of Q, there is a $p \in Q$ such that $pv \in Hq$. Then $qv^{-1} \in Hp$. This shows that $v^{-1} \in V$. Next, if $v, w \in V$, then $Qvw \subset HQw \subset HQ$ and thus $vw \in V$.

We show that the set $\mathcal{R}_S(u)$ is contained in V. Let $y \in \mathcal{R}_S(u)$. Since uy ends with u, and since u is not an internal factor of X, for any $p \in Q$, we have $py = xq$ for some $x \in X^*$ and $q \in Q$. Therefore $y \in V$.

By Theorem 6.15, the group generated by $\mathcal{R}_S(u)$ is the even subgroup. Thus V contains the even subgroup. But V contains odd words. Indeed, let $v \in S$ be such that $uvu^{-1} \in S$. Then v is odd by Proposition 4.12. Moreover, for any $p \in Q$ there is some $q \in Q$ such that $pvq^{-1} \in X^*$. This implies that $pv \in X^*q$ and thus v is in V. Since the even subgroup is of index 2, it is maximal in G_θ and we conclude that $V = G_\theta$.

Thus $Qw \subset HQ$ for any $w \in G_\theta$. Since $\varepsilon \in Q$, we have in particular $w \in HQ$ for any $w \in G_\theta$. Thus $G_\theta = HQ$. Since $\mathrm{Card}(Q) = d$, and since the right cosets Hq for $q \in Q$ are pairwise disjoint, this shows that H is a subgroup of index d. By Theorem 4.16, we have $\mathrm{Card}(X) - 2 = d(\mathrm{Card}(A) - 2)$. But since X generates H, and since X contains the inverses of its elements, this implies by Proposition 3.10 that X is a monoidal basis of H.

Assume conversely that the finite bifix code $X \subset F$ is a monoidal basis of the group $H = \langle X \rangle$ and that $\langle X \rangle$ has index d. Since X is a monoidal basis, by Schreier's Formula, we have $\mathrm{Card}(X) = (k-2)d + 2$, where $k = \mathrm{Card}(A)$. The case $k = 1$ is straightforward; thus we assume $k \geq 2$. By Theorem 8.2, there is a finite symmetric S-maximal bifix code Y containing X. Let e be the S-degree of Y. By the first part of the proof, Y is a monoidal basis of a subgroup K of index e of G_θ. In particular, it has $(k-2)e + 2$ elements. Since $X \subset Y$, we have $(k-2)d + 2 \leq (k-2)e + 2$ and thus $d \leq e$. On the other hand, since H is included in K, d is a multiple of e and thus $e \leq d$. We conclude that $d = e$ and thus that $X = Y$. ∎

Note that when X is not symmetric, the index of the subgroup generated by X may be different of $d_S(X)$, as shown in the following example.

Example 8.3 Let T be as in Example 5.4 and let $S = \mathcal{L}(T)$. The set $X = \{a, ba^{-1}, bc^{-1}, b^{-1}c, b^{-1}c^{-1}, a^{-1}c, cb, cb^{-1}, c^{-1}ab^{-1}, c^{-1}b\}$ is an S-maximal bifix code of S-degree 2. Since $b, c \in \langle X \rangle$, the group generated by X is the free group on A.

The following consequence of Theorem 8.1 is the counterpart for specular sets of Theorem 5.10 in [9]. We give in [10] a geometric proof and interpretation of Theorem 8.4 for the natural coding of a linear involution.

Theorem 8.4 *Let S be a recurrent specular set. For any subgroup H of finite index of the group G_θ, the set of prime words in S with respect to H is a monoidal basis of H.*

Proof. Let X be the set of prime words in S with respect to H. The set X is a symmetric bifix code and the number of parses of a word of S is at most equal to the index d of H in G_θ. Indeed, let (v, x, u) and (v', x', u') be two parses of a word $w \in S$. If v, v' are in the same left coset of H, then the two interpretations are equal. Indeed, assume that $|v| \geq |v'|$ and set $v = v's$. Then $s \in H$ and thus $s \in X^*$, which implies $s = 1$ by definition of a parse. Therefore X is an S-maximal bifix code by [4, Theorem 4.2.8].

By Theorem 8.1, X is a monoidal basis of a subgroup K of index e. Since $K \subset H$, the index of K is a multiple of the index of H. Since $e \leq d$, we conclude that $e = d$ and that $K = H$. ∎

We illustrate Theorem 8.4 with the following interesting example.

Example 8.5 Let T be as in Example 5.4 and let $S = \mathcal{L}(T)$. Let G be the group of even words in F_A. It is a subgroup of index 2. The set of prime words in S with respect to G is the set $Y = X \cup X^{-1}$ with

$$X = \{a, ba^{-1}c, bc^{-1}, b^{-1}c^{-1}, b^{-1}c\}.$$

Actually, the transformation induced by T on the set $I \times \{0\}$ (the upper part of $\hat{I}$ in Figure 5.2) is the interval exchange transformation represented in Figure 8.1. Its upper intervals are the I_x for $x \in X$. This corresponds to

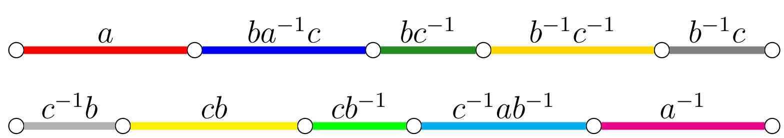

Figure 8.1: The transformation induced on the upper level.

the fact that the words of X correspond to the first returns to $I \times \{0\}$ while the words of X^{-1} correspond to the first returns to $I \times \{1\}$.

8.2. A converse of the Finite Index Basis Theorem

The following is a converse of Theorem 8.1.

Theorem 8.6 *Let S be a recurrent laminary set of factor complexity $p_n = n(\mathrm{Card}(A)-2)+2$. If $S \cap A^n$ is a monoidal basis of the subgroup $\langle A^n \rangle$ for all $n \geq 1$, then S is a specular set.*

Proof. Consider $w \in S$ and set $m = |w|$. The set $X = (AwA \cup Aw^{-1}A) \cap S$ is closed by taking inverses and it is included in $Y = S \cap A^{m+2}$. Since Y is a monoidal basis of a subgroup, $X \subset Y$ is a monoidal basis of the subgroup $\langle X \rangle$.

This implies that the graph $\mathcal{E}(w)$ is acyclic. Indeed, assume that the path $(a_1, b_1, \ldots, a_p, b_p, a_1)$ is a cycle in $\mathcal{E}(w)$ with $p \geq 2$, $a_i \in L(w)$, $b_i \in R(w)$ for $1 \leq i \leq p$ and $a_1 \neq a_p$. Then $a_1wb_1, a_2wb_1, \ldots, a_pwb_p, a_1wb_p \in X$. But

$$a_1wb_1(a_2wb_1)^{-1}a_2wb_2 \cdots a_pwb_p(a_1wb_p)^{-1} = \varepsilon,$$

with $a_jwb_j(a_{j+1}wb_j)^{-1} = a_ja_{j+1}^{-1} \neq \varepsilon$ (otherwise $a_j = a_{j+1}$), contradicting the fact that X is a monoidal basis.

Since $p_n = n(\mathrm{Card}(A)-2)+2$, we have $s_n = \mathrm{Card}(A) - 2$ and $t_n = 0$ for all $n > 0$. By Proposition 2.2, it implies that $m(w) = 0$ for all nonempty words w. Since $\mathcal{E}(w)$ is acyclic, we conclude that $\mathcal{E}(w)$ is a tree.

Finally, since $\mathcal{E}(\varepsilon)$ is acyclic, and since $m(\varepsilon) = -1$, the graph $\mathcal{E}(\varepsilon)$ has two connected components which are trees. ∎

References

[1] Peter Baláži, Zuzana Masáková, and Edita Pelantová. Factor versus palindromic complexity of uniformly recurrent infinite words. *Theoret. Comput. Sci.*, 380(3):266–275, 2007.

[2] Laurent Bartholdi. Growth of groups and wreath products. http://arxiv.org/abs/1512.07044v1.

[3] Jean Berstel. *Transductions and context-free languages*, volume 38 of *Leitfäden der Angewandten Mathematik und Mechanik [Guides to Applied Mathematics and Mechanics]*. B. G. Teubner, Stuttgart, 1979.

[4] Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012.

[5] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2009.

[6] Valérie Berthé, Clelia De Felice, Vincent Delecroix, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Specular sets. In *Combinatorics on Words*, number 9304 in Springer LNCS, pages 210–222. 2015.

[7] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Acyclic, connected and tree sets. *Monatsh. Math.*, 176:521–550, 2015.

[8] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. The finite index basis property. *J. Pure Appl. Algebra*, 219:2521–2537, 2015.

[9] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. *Dicrete Math.*, 338:725–742, 2015.

[10] Valérie Berthé, Vincent Delecroix, Francesco Dolce, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Return words of linear involutions and fundamental groups. *Ergodic Theory Dyn. Syst.*, 2015. to appear.

[11] Valérie Berthé and Michel Rigo. *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.* Cambridge Univ. Press, Cambridge, 2010.

[12] Corentin Boissy and Erwan Lanneau. Dynamics and geometry of the Rauzy-Veech induction for quadratic differentials. *Ergodic Theory Dynam. Systems*, 29(3):767–816, 2009.

[13] Julien Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin*, 4(1):67–88, 1997. Journées Montoises (Mons, 1994).

[14] Thierry Coulbois, Arnaud Hilion, and Martin Lustig. $\mathbb{R}$-trees and laminations for free groups. I. Algebraic laminations. *J. Lond. Math. Soc. (2)*, 78(3):723–736, 2008.

[15] Claude Danthony and Arnaldo Nogueira. Involutions linéaires et feuilletages mesurés. *C. R. Acad. Sci. Paris Sér. I Math.*, 307(8):409–412, 1988.

[16] Claude Danthony and Arnaldo Nogueira. Measured foliations on nonorientable surfaces. *Ann. Sci. École Norm. Sup. (4)*, 23(3):469–494, 1990.

[17] Pierre de la Harpe. *Topics in geometric group theory*. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 2000.

[18] Warren Dicks and M. J. Dunwoody. *Groups acting on graphs*, volume 17 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1989.

[19] Francesco Dolce and Dominique Perrin. Enumeration formulæ in neutral sets. In *Developments in Language Theory - 19th International Conference, DLT 2015, Liverpool, UK, July 27-30, 2015, Proceedings.*, pages 215–227, 2015.

[20] Francesco Dolce and Dominique Perrin. Neutral and tree sets of arbitrary characteristic. *Theoret. Comput. Sci.*, 658(part A):159–174, 2017.

[21] Xavier Droubay, Jacques Justin, and Giuseppe Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255(1-2):539–553, 2001.

[22] Fabien Durand. A characterization of substitutive sequences using return words. *Discrete Mathematics*, 179(1-3):89–101, 1998.

[23] Amy Glen, Jacques Justin, Steve Widmer, and Luca Q. Zamboni. Palindromic richness. *European J. Combin.*, 30(2):510–531, 2009.

[24] Michael Keane. Interval exchange transformations. *Math. Z.*, 141:25–31, 1975.

[25] Luis-Miguel Lopez and Philippe Narbel. Lamination languages. *Ergodic Theory Dynam. Systems*, 33(6):1813–1863, 2013.

[26] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

[27] Roger C. Lyndon and Paul E. Schupp. *Combinatorial Group Theory*. Classics in Mathematics. Springer-Verlag, 2001. Reprint of the 1977 edition.

[28] Wilhelm Magnus, Abraham Karrass, and Donald Solitar. *Combinatorial group theory*. Dover Publications, Inc., Mineola, NY, second edition, 2004.

[29] David E. Muller and Paul E. Schupp. Groups, the theory of ends, and context-free languages. *J. Comput. System Sci.*, 26(3):295–310, 1983.

[30] Arnaldo Nogueira. Almost all interval exchange transformations with flips are nonergodic. *Ergodic Theory Dynam. Systems*, 9(3):515–525, 1989.

[31] Arnaldo Nogueira, Benito Pires, and Serge Troubetzkoy. Orbit structure of interval exchange transformations with flip. *Nonlinearity*, 26(2):525–537, 2013.

[32] Edita Pelantová and Štěpán Starosta. Palindromic richness for languages invariant under more symmetries. *Theoret. Comput. Sci.*, 518:42–63, 2014.

[33] Jean-Pierre Serre. *Trees*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2003. Translated from the French original by John Stillwell, Corrected 2nd printing of the 1980 English translation.