UNIVERSITÉ PARIS-EST

École doctorale MSTIC
Mathématiques et Sciences et Techniques de l'Information et de la Communication

Thèse de doctorat
Spécialité: Informatique

*Présenté par:*
Francesco Dolce

# Codes bifixes, combinatoire des mots et systèmes dynamiques symboliques

*Bifix codes, Combinatorics on Words and Symbolic Dynamical Systems*

RDMath IdF
Domaine d'Intérêt Majeur (DIM) en Mathématiques

îledeFrance

Soutenue le 13 septembre 2016 devant les membres du jury :

| | | |
|---|---|---|
| Président | Prof. Srečko Brlek | Université du Québec à Montréal, LACIM |
| Directeur de thèse | Prof. Dominique Perrin | Université Paris-Est, LIGM |
| Rapporteur | Prof. Fabien Durand | Université de Picardie Jules Verne |
| Rapporteuse | Prof. Edita Pelantová | Czech Technical University in Prague |
| Examinatrice | Prof. Frédérique Bassino | Université Paris 13 |
| Examinatrice | Prof. Valérie Berthé | Université Paris Diderot, IRIF |
| Examinateur | Prof. Antonio Restivo | Università degli Studi di Palermo |

## Résumé court

L'étude des ensembles de mots de complexité linéaire joue un rôle très important dans la théorie de la combinatoire des mots et dans la théorie des systèmes dynamiques symboliques. Cette famille d'ensembles comprend les ensembles de facteurs d'un mot Sturmien ou d'un mot d'Arnoux-Rauzy, d'un codage d'échange d'intervalles, d'un point fixe d'un morphisme primitif, etc.

L'enjeu principal de cette thèse est l'étude de systèmes dynamiques minimaux et de complexité linéaire, définis de façon équivalente comme ensembles factoriels de mots uniformement récurrents. Comme résultat principal nous obtenons une hiérarchie naturelle de systèmes minimaux contenant les ensembles neutres, les ensembles à extension d'arbre (*tree sets*) et les ensembles spéculaires. De plus, nous relions ces systèmes au groupe libre en utilisant les mots de retour et les bases de sous-groupes d'indice fini. Nous étudions aussi les systèmes symboliques dynamiques engendrés par les échanges d'intervalles et les involutions linéaires, ce qui nous permet d'obtenir des exemples et des interprétations géométriques des familles d'ensembles définis dans notre hiérarchie.

L'un des outils principaux utilisés est l'étude des extensions possibles d'un mot dans un ensemble, ce qui nous permet de déterminer des propriétés telles que la complexité factorielle. Dans ce manuscrit, nous définissons le graphe d'extension, un graphe non orienté associé à chaque mot $w$ dans un ensemble $S$ qui décrit les extensions possibles de $w$ dans $S$ à gauche et à droite. Dans cette thèse, nous présentons plusieurs classes d'ensembles de mots définis par les formes possibles que les graphes d'extensions des éléments dans l'ensemble peuvent avoir.

L'une des conditions les plus faibles que nous allons étudier est la condition de neutralité: un mot $w$ est neutre si le nombre de paires $(a, b)$ de lettres telles que $awb \in S$ est égal au nombre de lettres $a$ tel que $aw \in S$, plus le nombre de lettres $b$ tel que $wb \in S$, moins 1. Un ensemble tel que chaque mot non vide satisfait la condition de neutralité est appelé un ensemble neutre.

Une condition plus forte est la condition de l'arbre: un mot $w$ satisfait cette condition si son graphe d'extension est à la fois acyclique et connexe. Un ensemble est appelé un ensemble à extension d'arbre si tout mot non vide satisfait cette condition. La famille des ensembles à extension d'arbre récurrents apparaît comme fermeture naturelle de deux familles d'ensembles très importantes : les facteurs d'un mot d'Arnoux-Rauzy et les ensembles d'échange d'intervalles.

Nous présentons également les ensembles spéculaires, une sous-famille remarquable d'ensemble à extension d'arbre. Il s'agit également de sous-ensembles de groupes qui forment une généralisation naturelle des groupes libres. Ces ensembles de mots sont une généralisation abstraite des codages naturels d'échanges d'intervalles et d'involutions linéaires.

Pour chaque classe d'ensembles considérée dans cette thèse, nous montrons plusieurs résultats concernant les propriétés de fermeture (par décodage bifixe maximal ou par rapport aux mots dérivés), la cardinalité des codes bifixes et celle des mots de retour, la connexion entre mots de retour et bases du groupe libre, ainsi qu'entre les codes bifixes et les sous-groupes du groupe libre. Chacun de ces résultats est prouvé en utilisant les hypothèses les plus faibles possibles.

**Mots clés.** Informatique théorique ; combinatoire des mots ; systèmes symboliques dymaniques ; ensembles neutres ; ensembles à extension d'arbre ; ensembles spéculaires ; mots de retour ; codes bifixes ; groupe libre ; échanges d'intervalles ; involutions linéaires.

# Abstract

Sets of words of linear complexity play an important role in combinatorics on words and symbolic dynamics. This family of sets includes set of factors of Sturmian and Arnoux-Rauzy words, interval exchange sets and primitive morphic sets, that is, sets of factors of fixed points of primitive morphisms.

The leading issue of this thesis is the study of minimal dynamical systems of linear complexity, also defined equivalently as uniformly recurrent sets of words. As a main result, we obtain a natural hierarchy of minimal systems containing neutral sets, tree sets and specular sets. Moreover, we connect the minimal systems to the free group using the notions of return words and basis of subroups of finite index. Symbolic dynamical systems arising from interval exchanges and linear involutions provide us geometrical examples of this kind of sets.

One of the main tool used here is the study of possible extensions of a word in a set, that allows us to determine properties such as the factor complexity. In this manuscript we define the extension graph, an undirected graph associated to each word $w$ in a set $S$ which describes the possible extensions of $w$ in $S$ on the left and the right. In this thesis we present several classes of sets of words defined by the possible shapes that the graphs of elements in the set can have.

One of the weakest condition that we will study is the neutrality condition: a word $w$ is neutral if the number of pairs $(a, b)$ of letters such that $awb \in S$ is equal to the number of letters $a$ such that $aw \in S$ plus the number of letters $b$ such that $wb \in S$ minus 1. A set such that every nonempty word satisfies the neutrality condition is called a neutral set.

A stronger condition is the tree condition: a word $w$ satisfies this condition if its extension graph is both acyclic and connected. A set is called a tree set if any nonempty word satisfies this condition. The family of recurrent tree sets appears as a the natural closure of two known families, namely the Arnoux-Rauzy sets and the interval exchange sets.

We also introduce specular sets, a remarkable subfamily of the tree 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.

For each class of sets considered in this thesis, we prove several results concerning closure properties (under maximal bifix decoding or under taking derived words), cardinality of the bifix codes and set of return words in these sets, connection between return words and basis of the free groups, as well as between bifix codes and subgroup of the free group. Each of these results is proved under the weakest possible assumptions.



**Keywords.** Theoretical Computer Science; Combinatorics on Words; Symbolic Dynamical Systems; Neutral sets; Tree sets; Specular sets; Return words; Bifix codes; Free group; Interval Exchange Transformations; Linerar Involutions.

# Contents

# Résumé de la thèse

Dans cette thèse, nous étudions les liens entre trois sujets: la dynamique symbolique, la théorie des codes et la théorie combinatoire des groupes.

Les ensembles de mots de complexité linéaire jouent un rôle très important dans la combinatoire des mots ainsi que dans la dynamique symbolique. Cette famille d'ensembles comprend les ensembles de facteurs d'un mot Sturmien, les ensembles de facteurs d'un mot d'Arnoux-Rauzy, les ensembles de facteurs de points fixes d'un morphisme primitif et les ensembles d'échanges d'intervalles.

Ce manuscrit est consacré à l'étude de ce genre d'ensembles. Comme résultat principal, nous établissons une hiérarchie naturelle des systèmes minimaux (ensembles uniformément récurrents de mots) contenant les ensembles neutres, les ensembles à extension d'arbre et les ensembles spéculaires.

La plupart des résultats ont déjà été publiés ou soumis dans une série d'articles signés par moi et d'autres co-auteurs. Dans la Conclusion sont présentés les références à ces articles, la structure d'ensemble du travail et une note explicative sur ma contribution personnelle.

### Mots Sturmien et échanges d'intervalles

Le mots Sturmiens sont des mots infinis sur un alphabet binaire qui ont exactement $n + 1$ facteurs de longueur $n$ pour tout $n \geq 0$. Leur origine remonte à l'astronome J. Bernoulli III et leur première étude approfondie a été réalisée par Morse et Hedlund [56]. Dans un autre travail Coven et Hedlund [27] décrivent des nombreuses propriétés combinatoires des mots sturmiens.

Les mots d'Arnoux-Rauzy sont une généralisation sur un alphabet de taille arbitraire des mots Sturmiens classiques sur deux lettres (voir [42]). Un ensemble d'Arnoux-Rauzy est l'ensemble des facteurs d'un mot d'Arnoux-Rauzy. Pour plus de détails, voir [41, 52].

Les mots Sturmiens sont étroitement liés au groupe libre (voir, par exemple, [7]). Les ensembles Sturmiens satisfont, par exemple, la propriété de base indice fini, c'est-à-dire : un code bifixe fini est $S$-maximal (avec $S$ l'ensemble des mots considéré) si et seulement si il est une base d'un sous-groupe d'indice fini du groupe libre sur $A$.

Les transformations d'échange d'intervalles ont été introduites par Oseledec (voir [59]), d'après une idée d'Arnold (voir [2]). La classe des transformations régulières d'échanges d'intervalles, quant à elle, a été introduite par Keane [47] qui a montré que ces transformations sont minimales dans le sens topologique dynamique. Le codage naturel d'un échange d'intervalles donne un ensemble de mots de complexité linéaire, tel que, par exemple, le language d'un mot Sturmien (voir, par exemple [39] ou [4]). L'ensemble des facteurs de codages naturels d'une transformation régulière d'échange d'intervalles est appelé un ensemble d'échange d'intervalles. Une généralisation des échanges d'intervalles est donnée par les involutions linéaires [29] (pour d'autres généralisations, voir [63]).

On remarque que la classe des facteurs d'un mot Sturmien est contenue à la fois dans la classe des ensembles réguliers d'échanges d'intervalles et dans celle des ensembles d'Arnoux Rauzy. En plus, on peut démontrer que l'intersection de ces deux classes est réduite aux ensembles Sturmiens. Même si les deux classes ont la même complexité factorielle (c'est-à-dire, le même nombre de facteurs pour une longueur donnée), elles ont des comportements combinatoires très distincts, par exemple en ce qui concerne le comportement des facteurs spéciaux, ou les propriétés d'équilibre, etc. (voir [23, 69]).

### Ensembles neutres

Dans cette thèse, nous étudions plusieurs familles d'ensembles de mots de complexité linéaire définies par des propriétés d'un graphe $\mathcal{E}(w)$, appelé le graphe d'extension de $w$. Ce graphe décrit les possibles extensions de $w$ à droite et à gauche par une lettre de l'alphabet $A$. Un ensemble $S$ est dit neutre si la caractéristique d'Euler du graphe d'un mot non vide est égale à 1. Lens ensembles à extension d'arbre forment une famille particulière d'ensembles neutres. Ces ensembles sont tels que le graphe $\mathcal{E}(w)$ est un arbre pour tout mot non vide, et il est acyclique pour le mot vide. La caractéristique d'Euler du graphe $\mathcal{E}(\varepsilon)$ est appelé la caractéristique de $S$ et est notée $\chi(S)$. Ces ensembles ont été étudiés dans [5].

La motivation pour l'étude des ensembles neutres et à extension d'arbre est la suivante : tout d'abord, la famille des ensembles à extension d'arbres récurrents apparaît comme la fermeture naturelle de deux familles d'entropie zero, à savoir les ensembles Sturmiens et les ensembles d'échanges d'intervalles. Ensuite, la famille d'ensembles neutres peut être vue comme une généralisation naturelle des ensembles à extension d'arbre, du fait que plusieures propriétés vraies pour ces dernières sont valides aussi pour les ensembles neutres.

La complexité factorielle d'un ensemble neutre $S$ sur $k$ lettres est égale, pour $n \neq 1$ à

$$p_n = n(k - \chi(S)) + \chi(S). \tag{1}$$

Plus généralement, on prove que pour un ensemble $S$ neutre de caractéristique 1, tout code bifixe $S$-maximal fini de $S$-degré $d$ a exactement $d(\mathrm{Card}(A) - 1) + 1$ éléments. Le fait remarquable est que, pour un ensemble $S$ fixé, la cardinalité de $X$ ne dépend que de son $S$-degré. Dans le cas particulier où $X$ est l'ensemble de tous les mots de $S$ de longueur $n$, on retrouve l'équation (1).

Un autre résultat concerne l'ensemble des mots de retour sur un mot $x$ dans un ensemble factoriel $S$, notée $\mathcal{R}_S(x)$. Cela est l'ensemble des mots non vides $u$ tel que $xu$ est dans $S$ et il se termine par $x$, sans qu'aucun de ses préfixes ait la même propriété. Dans plusieurs familles d'ensembles de complexité linéaire, il est connu que l'ensemble des mots de retour sur $x$ a cardinalité fixée et indépendante de $x$. Cela a été prouvé pour les mots Sturmiens dans [45], pour les ensembles d'échanges d'intervalles dans [67] (voir aussi [17]) et pour les ensembles neutres de caractéristique 1 dans [5].

Ici, nous montrons d'abord que l'ensemble $\mathcal{CR}_S(X)$ des mots de retour complet sur un code bifixe $X$ (satisfaisant certaines hypothèses) dans un ensemble neutre récurrent $S$ sur $k$ lettres satisfait $\mathrm{Card}(\mathcal{CR}_S(X)) = \mathrm{Card}(X)+k-\chi(S)$ et que cette quantité est une borne supérieure pour $\mathrm{Card}(CR_S(X))$ pour tout ensemble neutre (Théorème 2.2.8). Le fait remarquable ici est que, pour un ensemble neutre $S$ fixé, la cardinalité de $\mathcal{CR}_S(X)$ ne dépend que de $\mathrm{Card}(X)$. Quand $X$ contient un seul élément $x$, nous avons $\mathcal{CR}_S(x) = x\mathcal{R}_S(x)$ et on récupère le résultat de [5]. En plus, lorsque $X = S \cap A^n$, alors $\mathcal{CR}_S(X) = S \cap A^{n+1}$. Cela implique que $p_{n+1} = p_n + k - \chi(S)$ et donne également l'Équation (1). Les preuves de ces formules utilisent une distribution de probabilité naturellement définie sur un ensemble neutre.

Comme corollaire du Théorème 2.2.8 nous prouvons que dans un ensemble neutre les notions de récurrence et uniforme récurrence coïncident (Corollaire 2.2.9).

Un autre résultat concerne le décodage d'un ensemble neutre par un code bifixe. Nous montrons que le décodage d'un ensemble neutre récurrent $S$ par un code bifixe $S$-maximal est encore un ensemble neutre.

**Ensembles à extension d'arbre**

Les ensembles à exstension d'arbre ont plusieures propriétés particulièrement intéressantes, concernant les groupes libres, la dynamique symbolique associée aux ensembles et les codes bifixes contenus dans ces ensembles. En particulier, les ensembles à exstension d'arbre permettent de trouver des bases du groupe libre, ou des sous-groupes du groupe libre. En effet, dans un ensemble à exstension d'arbre récurrent, les ensembles de mots de premier retour sur un mot donné sont des bases du groupe libre sur l'alphabet. Par ailleurs, les codes bifixes maximaux qui sont inclus dans un ensemble à extension d'arbre récurrent sont des bases de sous-groupes d'indice fini du groupe libre. On démontre aussi que les ensembles à extension d'arbre sont fermés par décodage bifixe maximal et par décodage par rapport aux mots de retour.

Nous étudions les ensembles des mots de premier retour contenus dans un ensemble à extension d'arbre $S$. Notre résultat principal concernant les mots de retour est que si $S$ est un ensemble à extension d'arbre récurrent, l'ensemble des mots de premier retour sur un mot de $S$ est une base du groupe libre sur $A$. Pour cela, nous utilisons les graphes de Rauzy, que sont obtenus à partir des graphes de Bruijn en utilisant comme sommets que les mots de longueur donnée dans un ensemble $S$. D'abord, nous montrons que si $S$ est un ensemble connexe récurrent, le groupe décrit par un graphe de Rauzy de $S$ avec base un de ses sommets est le groupe libre sur $A$. Ensuite, nous montrons que dans un ensemble connexe recurrent $S$ contenant $A$, l'ensemble des mots de premier retour sur un mot dans $S$ engendre le groupe libre sur $A$. La preuve utilise le fait que, dans un ensemble neutre uniformément récurrent $S$, le nombre des mots de premier retour sur un mot de $S$ est égal à $\mathrm{Card}(A)$, un résultat obtenu dans [5].

Un résultat intéressant concernant les codes bifixes dans ce contexte est qu'un ensemble $S$ est acyclique si et seulement si tout code bifixe contenu dans $S$ est une base du sous-groupe qu'il engendre. Ceci est lié à la propriété de la base d'indice fini et au Théorème 4.2.1, prouvant qu'un code bifixe fini est $S$-maximal de $S$-degré $d$ si et seulement s'il est une base d'un sous-groupe d'indice $d$. Dans le cas d'un ensemble acyclique, le sous-groupe engendré par un code bifixe peut ne pas être d'indice fini, même si le code bifixe est $S$-maximal (et même si l'ensemble $S$ est uniformément récurrent).

Nous démontrons également un résultat plus technique. On dit qu'un sous-monoïde $M$ du monoïde libre est saturé dans un ensemble $S$ si le sous-groupe $H$ du groupe libre engendré par $M$ satisfait $M \cap S = H \cap S$. Nous montrons que si $S$ est acyclique, le monoïde engendré par un code bifixe contenu dans $S$ est saturé dans $S$.

Les ensembles à extension d'arbre récurrents satisfont la propriété de la base d'indice fini. Cela généralise le résultat concernant les mots Sturmiens de [7] cité ci-dessus. Comme exemple d'une conséquence de ce résultat, si $S$ est un ensemble à extension d'arbre récurrent sur l'alphabet $A$, alors pour tout $n \geq 1$, l'ensemble $S \cap A^n$ est une base du sous-groupe formé des mots de longueur un multiple de $n$.

Notre résultat principal concernant les ensembles à extension d'arbre est que la classe des ensembles à extension d'arbre récurrents est fermée par décodage bifixe maximal. Cela signifie que si $S$ est un ensemble à extension d'arbre uniformément récurrent et $f$ un morphisme de codage pour un code bifixe $S$-maximal fini, alors $f^{-1}(S)$ est un ensemble à extension d'arbre uniformément récurrent. La famille d'ensembles réguliers d'échanges intervalles est fermée par décodage bifixe maximal tandis que la famille des ensembles Sturmiens ne l'est pas. Ainsi, ce résultat montre que la famille d'ensembles à extension d'arbre récurrents est la fermeture naturelle de la famille des ensembles Sturmiens.

La preuve de ce dernier résultat utilise la propriété de base d'indice fini des ensembles à extension d'arbre uniformément récurrents. Elle utilise également la fermeture des ensembles à extension d'arbre récurrents par décodage par rapport aux mots de retour. Cette propriété, qui est elle même intéressante en soi, généralise le fait que le mot dérivé d'un mot Sturmien est lui aussi Sturmien [45].

Nous montrons aussi deux résultats qui permettent d'obtenir d'autres exemples d'ensembles à extension d'arbre d'origine géométrique, à savoir en utilisant des transformations d'échanges d'intervalles ou des involutions linéaires. Plus précisément, nous montrons que le codage naturel d'une transformation d'échange d'intervalles sans connexions de longueur $\geq 1$ est un ensemble à extension d'arbre et que le codage naturel d'une involution linéaire sans connexions est un ensemble à extension d'arbre de caractéristique 2.

### Ensembles spéculaires

Les groupes spéculaires sont des généralisations naturelle des groupes libres: ils sont des produits libres d'un nombre fini de copies de $\mathbb{Z}$ et $\mathbb{Z}/2\mathbb{Z}$. Un ensemble spéculaire est un sous-ensemble d'un groupe spéculaire. Cela peut être vu comme une généralisation du language du codage naturel d'une involution linéaire. Plus précisément, nous considérons un alphabet $A$ avec une involution $\theta$ qui agit sur $A$, possiblement avec des points fixes, et le groupe $G_\theta$ engendré par $A$ avec relations $a\theta(a) = 1$ pour toute lettre $a \in A$. Dans ce contexte on peut considérer des mots réduits, des ensembles symétriques de mots. De plus, on peut définir les ensembles laminaires, c'est-à-dire des ensembles factoriels contenant l'inverse de tous leurs éléments. Dans le cas où $\theta$ n'a pas de point fixe, on retrouve exactement le groupe libre. On peut donc définir un ensemple spéculaire comme un ensemble laminaire tel que le graphe d'extension de tout mot non vide est un arbre et le graphe d'extension du mot vide a deux composantes connexes qui sont des arbres.

Les groupes spéculaires apparaissent à plusieurs endroits dans [30]. Ils sont appelés *free-like* dans [6]. Ces groupes sont proches des groupes libres et, en particulier, la notion de base dans ces groupes est bien définie. D'après le théorème de Kurosh pour les sous-groupes, on sait que tout sous-groupe d'un groupe spéculaire est spéculaire. Un ensemble spéculaire peut être défini comme un sous-ensemble d'un tel groupe fermé par l'inverse et défini en termes de restrictions sur les extensions de ses éléments.

De même que pour les ensembles à extension d'arbre, nous donnons pour les ensembles spéculaires deux versions du théorème du retour et du théorème de la base d'indice fini. La première affirme que l'ensemble des mots de retour sur un mot donné dans un ensemble spéculaire récurrent forme une base d'un sous-groupe d'indice 2, appelé le sous-groupe pair. La seconde caractérise les bases symétriques des sous-groupes d'indice fini de groupes spéculaires contenus dans un ensemble spéculaire $S$ comme les codes bifixes $S$-maximaux symétriques finis contenus dans $S$.

L'idée de considérer des ensembles récurrents de mots réduits fermés par inverse est également liée à la notion des mots $G$-riches présenté dans [60].

### Induction de Rauzy

Rauzy a introduit dans [61] une transformation, maintenant appelée induction de Rauzy (ou induction de Rauzy-Veech), qui agit sur les échanges d'intervalles. Cette transformation modifie une transformation d'échange d'intervalles dans un autre définie sur un intervalle plus petit. Son itération peut être considérée comme une généralisation du développement en fraction continue. L'induction consiste à prendre le premier retour de la transformation par rapport à un sous-intervalle de l'intervalle sur lequel l'échange est défini. La transformation induite d'un échange d'intervalles sur $s$ sous-intervalles est toujours un échange d'intervalles sur au plus $s+2$ intervalles. Rauzy a introduit dans [61] la définition d'admissiblité à droite pour un intervalle et il a caractérisé les intervalles admissibles à droite comme ceux qui peuvent être atteints par l'induction de Rauzy. Dans cette thèse, nous généralisons à la fois la notion d'intervalles admissibles et d'induction de Rauzy à une version bilatérale. Nous caractérisons les intervalles admissibles (Théorème 7.2.3) et montrons, en particulier, que les intervalles associés aux facteurs du codage naturel d'une transformation d'échange d'intervalles sont admissibles (Proposition 7.1.6).

De plus, nous démontrons une propriété sur les codages naturels de transformations régulières d'échange d'intervalles en disant que la famille de ces ensembles de mots est fermée par dérivation, une opération qui consiste à considérer les mots de retour sur un mot donné comme un nouvel alphabet.

### Échange d'intervalles sur un corps quadratique

Les transformations d'échange intervalles définis sur un corps quadratique ont été étudiés par Boshernitzan et Carroll ([20] et [19]). Dans ces hypothèses, ils ont montré que, en utilisant itérativement la fonction du premier retour sur l'un des intervalles échangés par la transformation, on obtient seulement un nombre fini de différentes nouvelles transformations à renormalisation près. Ce résultat étend le théorème classique de Lagrange selon lequel le développement d'un irrationnel quadratique est périodique.

Ici, nous montrons que, dans le cas d'échanges d'intervalles définis sur un corps quadratique, la famille des transformations obtenues à partir d'une transformation régulière d'échange intervalles par induction de Rauzy bilatérale est finie, à renormalisation près. De plus, nous montrons, comme conséquence, que l'ensemble d'échange d'intervalles associé à une telle transformation est l'ensemble des facteurs d'un mot morphique primitif.

### Involutions linéaires

Une involution linéaire est une isométrie par morceaux injective définie sur une paire d'intervalles. Les involutions linéaires ont été introduites par Danthony et Nogueira dans [29] et [28], en généralisant les échanges d'intervalles avec retournement(s) [57, 58] (échanges d'intervalles qui inversent l'orientation dans au moins un intervalle). Les deux auteurs ont étendu à ces transformations la notion d'induction de Rauzy (introduite dans [61]). L'étude des involutions linéaires a ensuite été développé par Boissy et Lanneau dans [18].

Le codage naturel d'une involution linéaire est l'ensemble des facteurs des mots infinis qui codent les séquences de sous-intervalles rencontrés par les orbites de la transformation. Ils sont définis sur un alphabet $A$ dont les lettres et leur inverses indexent les intervalles échangés par l'involution. Un codage naturel est donc un sous-ensemble du groupe libre $F_A$ sur l'alphabet $A$. Une propriété importante de cet ensemble est sa stabilité par inverse.

Nous étendons aux codages naturels des involutions linéaires la plupart des propriétés prouvées pour les ensembles à extension d'arbre récurrents, et donc, pour les codages naturels des échanges d'intervalles. Cette extension n'est pas triviale ni immédiate. Nous considérons les mots de retour sur l'ensemble

$\{w, w^{-1}\}$ ainsi qu'une version tronquée de cet ensemble, que nous appelons ensemble des mots de retour mixte. Dans ce contexte, nous remplaçons la base d'un sous-groupe par sa version symétrique contenant les inverses de ses éléments, appelé base symétrique.

Nous montrons, enfin, que le codage naturel d'une involution linéaire sans connexion est un ensemble spéculaire.

Le manuscrit est organisé comme suit.

Dans le Chapitre 1, nous donnons quelques notions préliminaires et des définitions. Nous introduisons les mots et les ensembles à la fois dans le monoïde libre et dans le groupe libre. Nous définissons le graphe d'extension d'un mot dans un ensemble factoriel. De plus, nous donnons les définitions de base et quelques résultats sur les codes bifixes, les automates et les mots de retour. Tous ces outils seront utilisés dans les chapitres suivants.

Les Chapitres 2 à 5 sont consacrés à l'étude des différentes classes d'ensembles ordonnés hiérarchiquement. Les plus importants sont les ensembles neutres (Chapitre 2), les ensembles à extension d'arbre (Chapitres 3 et 4) et les ensembles spéculaires (Chapitre 5).

En particulier, le Chapitre 2 est consacré aux ensembles neutres. D'abord, nous définissons les notions de faible, fort et neutre, et nous montrons que les ensembles neutres ont complexité factorielle linéaire (Proposition 2.1.3). Plus en général, nous prouvons le Théorème de la Cardinalité pour les ensembles neutres (Théorème 2.2.1) qui dit que tous les codes bifixes $S$ maximaux ayant le même $S$-degré ont la même cardinalité. Nous montrons aussi des résultats sur la cardinalité des ensembles des mots de retour (Théorème 2.2.8 et Corollaire 2.2.10) ainsi qu'une propriété de fermeture pour la famille d'ensembles neutres par décodage bifixe maximal (Théorème 2.3.1). En utilisant les résultats précédents, nous montrons également que dans le contexte des ensembles neutres (et donc pour tous les ensembles définis par une des propriétés plus fortes comme dans les chapitres suivants) les notions de récurrence et de récurrence uniforme coïncident (Corollaire 2.2.9).

Dans les Chapitres 3 et 4 nous définissons et étudions les ensembles à extension d'arbre.

Le résultat principal du Chapitre **??** est le Théorème de Retour (Théorème 3.2.5), qui dit que l'ensemble des mots de retour sur un ensemble à extension d'arbre de caractéristique 1 est une base du groupe libre. Nous donnons également un résultat de fermeture par dérivation pour les ensembles à extension d'arbre (Théorème 3.2.9) et montrons comment utiliser des fonctions dites fonctions de multiplication pour construire de nouveaux ensembles à extension d'arbre (Théorème 3.3.1). Nous terminons le chapitre avec un résultat concernant les palindromes: nous montrons que les ensembles à extension d'arbre récurrents de caractéristique 1 fermés par image miroir sont riches (Proposition 3.4.1).

Dans le Chapitre 4, consacré également aux ensembles à extension d'arbre, nous nous intéressons particulièrement à l'étude des codes bifixes dans les ensembles à extension d'arbre et à leur connexion avec les sous-groupes du groupe libre. En relaxant l'hypothèse, quand cela est possible, nous montrons le Freeness Théorème, indiquant que les codes bifixes dans les ensembles à extension d'arbre sont des bases du sous-groupe qu'ils engendrent (Théorème 4.1.1) et le Théorème de Saturation, disant que le monoïde engendré par un code bifixe fini est saturé (Théorème 4.1.2). Un autre résultat principal de ce chapitre est le Théorème de la Base d'Indice Fini, qui dit qu'un code bifixe fini dans un ensemble à extension d'arbre $S$ est $S$-maximal de $S$-degré $d$ si et seulement si il est un sous-groupe d'indice $d$ du groupe libre. Dans ce contexte, nous définissons également des bases *tame* et nous montrons que chaque ensemble récurrent à extension d'arbre de caractéristique 1 a une réprésentation $\mathcal{S}$-adique primitive, avec $\mathcal{S}$ fini et contenant seulement des automorphismes positifs. Nous terminons le chapitre en montrant plusieures propriétés de fermeture d'ensembles à extension d'arbre par décodage bifixe maximal (Théorèmes 4.3.1, 4.3.3, 4.3.5 et 4.3.17).

Dans le Chapitre 5 nous étudions les ensembles spéculaires, une famille d'ensembles à extension d'arbre de caractéristique 2 ayant, en outre, des propriétés symétriques. Nous pouvons, par exemple, définir dans ce contexte la notion de parité d'un mot. Ces ensembles sont des ensembles laminaires et ils sont liés aux groupes virtuellement libres appelés groupes spéculaires. Après avoir donné les définitions nécessaires, nous construisons une importante famille d'ensembles spéculaires, obtenue en doublant les ensembles à extension d'arbre de caractéristiques 1, et nous montrons que cette famille est $G$-riche (Proposition 5.2.26). De plus, nous donnons des versions plus précises des principaux résultats du Chapitre 4, tels que le Théorème du Retour et le Théorème de la Base d'Indice Fini (Théorèmes 5.3.11 et 5.5.1), ainsi que plusieurs résultats de cardinalité concernants les mots de retour dans ces ensembles (par exemple, Théorèmes 5.3.2, 5.3.5 et 5.3.9).

La partie du manuscrit du Chapitre 6 au Chapitre 8 est consacrée à l'étude des familles provenant de systèmes dynamiques géométriques: en particulier les échanges d'intervalles (Chapitres 6 et 7) et les involutions linéaires (Chapitre 8).

Dans le Chapitre 6, nous montrons que les ensembles factoriels résultant du codage naturel des transformations d'échange d'intervalles, sont des ensembles à extension d'arbre. Plus particulièrement, nous montrons que si la transformation est régulière, alors le language associé satisfait une propriété plus forte : l'extension d'arbre planaire (Théorème 6.1.16). En effet, cette propriété caractérise ces ensembles. Cette famille d'ensembles est fermée par décodage bifixe maximale (Théorème 6.2.11 et Corollaire 6.2.13).

Dans le Chapitre 7 nous continuons l'étude des échanges d'intervalles en introduisant l'induction à ramification, une généralisation de l'induction de Rauzy classique : une fonction qui associe à un échange d'intervalles un autre échange d'intervalles et préserve certaines de ses propriétés (par exemple la régularité). Nous donnons la définition d'admissibilité pour un sous-intervalle et nous caractérisons les semi-intervalles admissibles pour une transformation d'échange d'intervalles (Théorème 7.2.3). Enfin, nous étudions le cas d'un échange intervalles défini sur un corps quadratique. En suivant le travail de Boshernitzan et Carroll dans [19], nous montrons que, sous certaines hypothèses, il existe qu'un nombre fini de transformations obtenues par l'induction de Rauzy à ramification (Théorème 7.3.1). Nous utilisons ce résultat pour prouver que le language d'une transformation régulière d'échange d'intervalles définie sur un corps quadratique est un ensemble primitif morphique (Théorème 7.3.12).

On a vu que les échanges d'intervalles nous donnent des exemples d'ensembles à extension d'arbre. De même, dans le Chapitre 8, nous introduisons les involutions linéaires et nous montrons que le language associé à un système dynamique de ce type est un ensemble spéculaire. Dans ce chapitre, nous étudions d'abord les propriétés dynamiques des involutions linéaires, définissant certaines classes remarquables de ces systèmes, telles que les involutions linéaires orientables, les involutions linéaires cohérentes ou les involutions linéaires minimales. Par la suite, nous définissons le codage naturel d'une involution linéaire et montrons que, sous certaines hypothèses, cet ensemble est un ensemble spéculaire (Théorème 8.2.11). Nous donnons aussi des résultats concernant l'orientabilité (Proposition 8.2.5), les mots de retour mixte et les intervalles admissibles pour une involution linéaire, notion qui généralise la notion analogue vue dans le Chapitre 7 pour les échanges d'intervalles.

Enfin, nous terminons le manuscrit avec la Conclusion, où nous parlons de problèmes ouverts et de certaines directions de recherche possibles.

# Introduction

In this thesis we study the connections between three subjects: symbolic dynamics, theory of codes and combinatorial group theory.

Sets of words of linear complexity play an important role in combinatorics on words and symbolic dynamics. This family of sets includes set of factors of Sturmian and Arnoux-Rauzy words, interval exchange sets and primitive morphic sets, that is, sets of factors of fixed points of primitive morphisms.

This manuscript is devoted to the study of this kind of sets. As a main result, we establish a natural hierarchy of minimal systems (uniformly recurrent sets of words) containing neutral sets, tree sets and specular sets.

Most of the results are already been published or submitted in a series of papers by me and other authors. The references to these papers and their architecture as well as the mentions, as much as possible, of my personal contribution are presented in the Conclusion.

### Sturmian words and interval exchanges

Sturmian words are infinite words over a binary alphabet that have exactly $n+1$ factors of length $n$ for each $n \geq 0$. Their origin can be traced back to the astronomer J. Bernoulli III and their first in-depth study was done by Morse and Hedlund [56]. Another important work is the paper by Coven and Hedlund [27] which describes many combinatorial properties of Sturmian words.

Arnoux-Rauzy words are a generalization to arbitrary alphabets of the classical Sturmian words on two letters (see the survey [42]). An Arnoux-Rauzy set is the set of factors of an Arnoux-Rauzy word. For more details, see [41, 52].

Sturmian words are closely related to the free group (see, for example, [7]). Sturmian sets satisfy, for instance, the finite index basis property, in the sense that given a set $S$ of words on an alphabet $A$, a finite bifix code is $S$-maximal if and only if it is the basis of a subgroup of finite index of the free group on $A$.

Interval exchange transformations were introduced by Oseledec [59] following an earlier idea of Arnold [2]. The class of regular interval exchange transformations was introduced by Keane [47] who showed that they are minimal in the topological dynamics sense. The natural coding of interval exchange produces sequences of linear complexity, including Sturmian sequences which have been widely studied (see, for example [39] or [4] for small alphabets). The set of factors of the natural codings of a regular interval exchange transformation is called an interval exchange set. Interval exchange transformations have been generalized to transformations called linear involutions by Danthony and Nogueira in [29] (for other generalizations, see [63]).

Note that the class of factors of a Sturmian word is contained both in the class of regular interval exchange sets and of Arnoux Rauzy sets. Moreover, it can be shown that the intersection of regular interval exchange sets and the class of Arnoux-Rauzy sets is reduced to binary Sturmian sets. Indeed, Arnoux-Rauzy sets on more than two letters are not the set of factors of an interval exchange transformation with each interval labeled by a distinct letter (the construction in [3] allows one to obtain the Arnoux-Rauzy sets of 3 letters as an exchange of 7 intervals labeled by 3 letters).

Even though they have the same factor complexity (that is, the same number of factors of a given length), Arnoux-Rauzy words and codings of interval exchange transformations have a priori very distinct combinatorial behaviours, whether for the type of behaviour of their special factors, or for balance properties and deviations of Birkhoff sums (see [23, 69]).

**Neutral sets**

In this thesis, we study several families of sets of words of linear complexity defined by properties of a graph $\mathcal{E}(w)$, called the extension graph of $w$. This graph expresses the possible extensions of $w$ on both sides by a letter of the alphabet $A$. A set $S$ is neutral if the Euler characteristic of the graph of any nonempty word is equal to 1. Tree sets form a special family of neutral sets. These sets are such that the graph $\mathcal{E}(w)$ is a tree for every nonempty word and acyclic for every word. The Euler characteristic of the graph $\mathcal{E}(\varepsilon)$ is called the characteristic of $S$ and is denoted by $\chi(S)$. These sets were first considered in [5].

The motivation for studying neutral and tree sets is the following: First, the family of recurrent tree sets appears as the natural closure of two known families of languages of classical shifts of zero entropy, namely the Sturmian sets and the interval exchange sets. Next, the family of neutral sets is a naturally defined generalization of tree sets for which a number of properties true for tree sets still hold.

The factor complexity of a neutral set $S$ on $k$ letters is shown to be given for $n \neq 1$ by the formula

$$p_n = n(k - \chi(S)) + \chi(S). \tag{2}$$

More generally, we prove that under the neutrality condition of characteristic 1, any finite $S$-maximal bifix code of $S$-degree $d$ has $d(\mathrm{Card}(A) - 1) + 1$ elements (Cardinality Theorem). The remarkable feature is that, for fixed $S$, the cardinality of $X$ depends only on its $S$-degree. In the particular case where $X$ is the set of all words of $S$ of length $n$, we recover Equation (2).

Another result concerns the set of right return words to a word $x$ in a factorial set $S$, denoted by $\mathcal{R}_S(x)$. It is the set of nonempty words $u$ such that $xu$ is in

$S$ and ends with $x$ for the first time in a left to right scan. In several families of sets of linear complexity, the set of return words to $x$ is known to be of fixed cardinality independent of $x$. This was proved for Sturmian words in [45], for interval exchange sets in [67] (see also [17]) and for neutral sets of characteristic 1 in [5].

Here, we first prove that the set $\mathcal{CR}_S(X)$ of complete return words to a bifix code $X$ (satisfying additional hypotheses) in a recurrent neutral set $S$ on $k$ letters satisfies $\mathrm{Card}(\mathcal{CR}_S(X)) = \mathrm{Card}(X) + k - \chi(S)$ and that this quantity is an upper bound for $\mathrm{Card}(\mathcal{CR}_S(X))$ for every neutral set (Theorem 2.2.8). The remarkable feature here is that, for fixed $S$, the cardinality of $\mathcal{CR}_S(X)$ depends only on $\mathrm{Card}(X)$. When $X$ is reduced to one element $x$ we have $\mathcal{CR}_S(x) = x\mathcal{R}_S(x)$ and we recover the result of [5]. When $X = S \cap A^n$, then $\mathcal{CR}_S(X) = S \cap A^{n+1}$. This implies $p_{n+1} = p_n + k - \chi(S)$ and also gives Equation (2) by induction on $n$. The proofs of these formulæ use a probability distribution naturally defined on a neutral set.

As a corollary of Theorem 2.2.8 we prove that in neutral sets the notions of recurrence and uniformly recurrence coincide (Corollary 2.2.9).

Another result concerns the decoding of a neutral set by a bifix code. We prove that the decoding of any recurrent neutral set $S$ by an $S$-maximal bifix code is a neutral set.

**Tree sets**

Tree sets have particularly interesting properties relating free groups, symbolic dynamics and bifix codes. In particular tree sets allow one to exhibit bases of the free group, or of subgroups of the free group. Indeed, in a recurrent tree set, the sets of first return words to a given word are bases of the free group on the alphabet. Moreover, maximal bifix codes that are included in recurrent tree sets provide bases of subgroups of finite index of the free group. Tree sets are also proved to be closed under maximal bifix decoding and under decoding with respect to return words.

We study sets of first return words in a tree set $S$. Our main result on return words is that if $S$ is a recurrent tree set, the set of first return words to any word of $S$ is a basis of the free group on $A$ (Return Theorem). For this, we use Rauzy graphs, obtained by restricting de Bruijn graphs to the set of vertices formed by the words of given length in a set $S$. We first show that if $S$ is a recurrent connected set, the group described by any Rauzy graph of $S$ with respect to some vertex is the free group on $A$. Next, we prove that in a recurrent connected set $S$ containing $A$, the set of first return words to any word in $S$ generates the free group on $A$. The proof uses the fact that in a uniformly recurrent neutral set $S$, the number of first return words to any word of $S$ is equal to $\mathrm{Card}(A)$, a result obtained in [5].

An interesting result concerning bifix codes in this framework is that a set $S$ is acyclic if and only if any bifix code contained in $S$ is a basis of the subgroup that it generates (Freeness Theorem). This is related to the Finite Index Basis Theorem, proving that a finite bifix code is $S$-maximal of $S$-degree $d$ if and only if it is a basis of a subgroup of index $d$. The proof uses the Return Theorem. In the case of an acyclic set, the subgroup generated by a bifix code need not be of finite index, even if the bifix code is $S$-maximal (and even if the set $S$ is uniformly recurrent).

We also prove a more technical result. We say that a submonoid $M$ of the free monoid is saturated in a set $S$ if the subgroup $H$ of the free group generated by $M$ satisfies $M \cap S = H \cap S$. We prove that if $S$ is acyclic, the submonoid generated by a bifix code contained in $S$ is saturated in $S$ (Saturation Theorem). This property plays an important role in the proof of the Finite Index Basis Theorem.

Recurrent tree sets satisfy the finite index basis property. This generalizes the result concerning Sturmian words of [7] quoted above. As an example of a consequence of this result, if $S$ is a recurrent tree set on the alphabet $A$, then for any $n \geq 1$, the set $S \cap A^n$ is a basis of the subgroup formed by the words of length multiple of $n$.

Our main result concerning tree sets is that the class of recurrent tree sets is closed under maximal bifix decoding. This means that if $S$ is a uniformly recurrent tree set and $f$ a coding morphism for a finite $S$-maximal bifix code, then $f^{-1}(S)$ is a uniformly recurrent tree set. The family of regular interval exchange sets is closed under maximal bifix decoding but the family of Sturmian sets is not. Thus, this result shows that the family of recurrent tree sets is the natural closure of the family of Sturmian sets.

The proof of Maximal Bifix Decoding Theorem uses the finite index basis property of uniformly recurrent tree sets. It also uses the closure of recurrent tree sets under decoding with respect to return words. This property, which is interesting in its own, generalizes the fact that the derived word of a Sturmian word is Sturmian [45].

We also prove two results which allows one to obtain a large family of tree sets of geometric origin, namely using interval exchange transformations or linear involutions. More precisely, we prove that the natural coding of an interval exchange transformation without connections of length $\geq 1$ is a tree set and that the natural coding of a linear involution without connections is a tree set of characteristic 2.

### Specular sets

Specular groups are natural generalizations of free groups: they are free products of a finite number of copies of $\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z}$. A specular set is a subset of a specular group which generalizes the natural codings of linear involutions. More precisely, we consider an alphabet with an involution $\theta$ acting on $A$, possibly with some fixed points, and the group $G_\theta$ generated by $A$ with the relations $a\theta(a) = 1$ for every letter $a$ in $A$. We can thus consider, in this extended framework, reduced words, symmetric sets of words and define laminary sets as factorial sets containing the inverse of all their elements. In the case where $\theta$ has no fixed point, we recover the free group. A specular set is then defined as a laminary set such that the extension graph of any nonempty word is a tree and the extension graph of the empty word has two connected components which are trees.

In this manuscript, 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. They are called *free-like* in [6] and appear at several places in [30]. 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 stable by taking the inverse and defined in terms of restrictions on the extensions of its elements.

As for the tree sets, we give two versions of the First Return Theorem and the Finite Index Basis Theorem also for specular sets. 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 $S$.

The idea of considering recurrent sets of reduced words invariant by taking inverses is also connected with the notion of $G$-full words of [60].

### Rauzy induction

Rauzy introduced in [61] a transformation, now called Rauzy induction (or Rauzy-Veech induction), which operates on interval exchange transformations. This transformation changes an interval exchange transformation into another one operating on a smaller interval. Its iteration can be viewed as a generalization of the continued fraction expansion. The induction consists in taking the first return map of the transformation with respect to a subinterval of the interval on which the exchange is defined. The induced map of an interval exchange on $s$ intervals is still an interval exchange with at most $s + 2$ intervals. Rauzy introduced in [61] the definition of right-admissibility for an interval and characterized the right-admissible intervals as those which can be reached by the Rauzy induction. In this thesis, we generalize both the notion of admissible intervals and of Rauzy induction to a two-sided version. We characterize the admissible intervals (Theorem 7.2.3) and show, in particular, that intervals associated with factors of the natural coding of an interval exchange transformation are admissible (Proposition 7.1.6).

Moreover, we prove a property of the natural codings of regular interval exchange transformations saying that the family of these sets of words is closed by derivation, an operation consisting in taking the first return words to a given word as a new alphabet.

### Interval exchanges over a quadratic field

Interval exchange transformations defined over quadratic fields have been studied by Boshernitzan and Carroll ([20] and [19]). Under this hypothesis, they showed that, using iteratively the first return map on one of the intervals exchanged by the transformation, one obtains only a finite number of different new transformations up to rescaling, extending the classical Lagrange's theorem that quadratic irrationals have a periodic continued fraction expansion.

Here we prove that, in the case of interval exchanges defined over a quadratic field, the family of transformations obtained from a regular interval exchange transformation by two-sided Rauzy induction is finite up to rescaling. Moreover, we show as a consequence that the related interval exchange set is obtained as the set of factors of a primitive morphic word.

**Linear involutions**

A linear involution is an injective piecewise isometry defined on a pair of intervals. Linear involutions were introduced by Danthony and Nogueira in [29] and [28], generalizing interval exchanges with flip(s) [57, 58] (these are interval exchange transformations which reverse orientation in at least one interval). They extended to these transformations the notion of Rauzy induction (introduced in [61]). The study of linear involutions was later developed by Boissy and Lanneau in [18].

The natural coding of a linear involution is the set of factors of the infinite words that encode the sequences of subintervals met by the orbits of the transformation. They are defined on an alphabet $A$ whose letters and their inverses index the intervals exchanged by the involution. A natural coding is thus a subset of the free group $F_A$ on the alphabet $A$. An important property of this set is its stability by taking inverses.

We extend to natural codings of linear involutions most of the properties proved for recurrent tree sets, and thus, for natural codings of interval exchanges. The extension is not completely immediate. We consider return words to the set $\{w, w^{-1}\}$ and we consider a truncated version of them, that we call mixed first return words. We also have to replace the basis of a subgroup by its symmetric version containing the inverses of its elements, called a symmetric basis.

We actually prove that the natural coding of a linear involution without connection is a specular set.

The manuscript is organized as follows.

In Chapter 1 we give some preliminary notions and definitions. We introduce words and sets both in the free monoid and in the free group. We define the extension graph of a word in a factorial set. Moreover, we give the basic definitions and a few results about bifix codes, automata and return words, all tools that will be used in the following chapters.

Chapters 2 to 5 are dedicated to the study of different classes of sets, ordered hierarchicaly. The most important ones are neutral sets (Chapter 2), tree sets (Chapters 3 and 4) and specular sets (Chapter 5).

In particular, Chapter 2 is devoted to neutral sets. First we define the notions of weakness, strongness and neutrality, and we show that neutral sets have linear factor complexity (Proposition 2.1.3). More generally, we prove the Cardinality Theorem for neutral sets (Theorem 2.2.1) stating that all $S$-maximal bifix codes of the same $S$-degree have the same cardinality. We also prove some cardinality results for the set of return words (Theorem 2.2.8 and Corollary 2.2.10) and a closure property for the family of neutral sets under maximal bifix decoding (Theorem 2.3.1). Using the previous results we also show that in the framework of neutral sets (and thus for all sets defined by a stronger properties in the next chapters) the notions of recurrence and of uniformly recurrence coincide (Corollary 2.2.9).

In Chapters 3 and 4 we define and study tree sets.

In the first of the these two chapters we give the definition of the tree condition. The main result of this chapter is the Return Theorem (Theorem 3.2.5), stating that the set of return word on a tree set of characteristic 1 is a basis of the free group. We also give a closure result for tree sets under derivation (Theorem 3.2.9) and show how to use multiplying maps to construct new tree sets (Theorem 3.3.1). We close the chapter with a result about palindromes: namely we show that recurrent tree sets of characteristic 1 closed under reversal are full (Proposition 3.4.1).

In the second chapter devoted to tree sets, Chapter 4, we concentrate on the study of bifix codes in tree sets and their connection to subgroups of the free group. Relaxing the hypothesis when possible, we show the Freeness Theorem, stating that that bifix codes in tree sets are bases of the subgroup that they generate (Theorem 4.1.1), and the Saturation Theorem, stating that the submonoid generated by a finite bifix code is saturated (Theorem 4.1.2). Another main result of this chapter is the Finite Index Basis Theorem, which states that a finite bifix code in a tree set $S$ is $S$-maximal of $S$-degree $d$ if and only if it is a subgroup of index $d$ of the free group. In this context, we define also tame bases and we show that every recurrent tree set of characteristic 1 has a primitive $\mathcal{S}$-adic representation, with $\mathcal{S}$ finite and containing positive automorphisms only. We close the chapter showing several closure properties of tree sets under maximal bifix decoding (Theorems 4.3.1, 4.3.3, 4.3.5 and 4.3.17).

In Chapter 5 we study specular sets, a family of tree sets of characteristic 2 having, additionally, symmetric properties. We can, for example, define the notion of parity of a word. These sets are laminary sets and they are related to virtually free groups called specular groups. After giving the needed definitions, we show an important family of specular sets, obtained by doubling tree sets of characteristic 1, and we show that this family is $G$-full (Proposition 5.2.26). Moreover, we give more precise versions of the main results of Chapter 4, such as the First Return Theorem and the Finite Index Basis Theorem (Theorems 5.3.11 and 5.5.1), as well as several cardinality results concerning return words in these sets (e.g. Theorems 5.3.2, 5.3.5 and 5.3.9).

The part of the manuscript from Chapter 6 to Chapter 8 is devoted to the study of families arising from geometrical dynamical systems: in particular from interval exchange transformations (Chapters 6 and 7) and linear involutions (Chapter 8).

Intervar exhanges are defined in Chapter 6. Here, we show that interval exchange sets, factorial sets arising from the natural coding of interval exchange transformations, are tree sets. More in particular, we show that if the transformation is regular, then the language associated satisfies a stronger property: the planar tree condition (Theorem 6.1.16), and that actually this property characterize these sets. This family of sets is closed under maximal bifix decoding (Theorem 6.2.11 and Corollary 6.2.13).

In Chapter 7 we continue the study of interval exchanges introducing the branching induction, a generalization of the classical Rauzy induction: a map that associates to an interval exchange another intervale exchange and that preserve some of its properties (such as the regularity). We give the definition of admissibility for a sub-interval and we characterize the admissible semi-intervals for an interval exchange transformation (Theorem 7.2.3). Finally, we study the case of an interval exchange defined over a quadratic field. Following the path of Boshernitzan and Carroll in [19], we prove that under certains hypothesis, there are finitely many transformations obtained by the branching Rauzy induction (Theorem 7.3.1). We use this result to prove that the language of a regular interval exchange transformation defined over a quadratic field is a primitive morphic set (Theorem 7.3.12).

If interval exchanges give us examples of tree sets, in Chapter 8 we introduce linear involutions and we show that the language associate to a similar dynamical system satisies the specular condition. In this chapter we first study the dynamical properties of linear involutions, defining some remarkable classes of these systems, such as coherent, orientable, minimal linear involutions. Afterward, we define the natural coding of linear involutions and show that, under certain hypothesis, this set is a specular set (Theorem 8.2.11). We also give some results about orientability (Proposition 8.2.5), mixed return words in this framework and admissible interval for a linear involution, notion that generalize the analougous notion seen in Chapter 7 for interval exchanges.

Finally, we close the manuscript with the Conclusions, where we talk about some open research directions.

**Acknowledgement.** I would like to express my special appreciation and thanks to my mentor and advisor Dominique PERRIN who encouraged and advised me on both research and career; his continuous support, his patience, motivation and immense knowledge have accompanied me during my Ph.D study. *Merci beaucoup.*

I would also thank my first mentor Antonio RESTIVO who first introduced me to the world of theoretical computer science. *Grazie mille.*

A special thank to Fabien DURAND and Edita PELANTOVÁ who accepted to report this thesis. Their insightful comments and suggestions helped to improve this manuscript. *Merci*, *Dı̌ky.*

I would like to thank Frédérique Bassino, Valérie Berthé and Srečko Brlek for accepting to be members of the thesis commitee. *Merci*, *Trugarez*, *Hvala*.

A special thanks also to Giuseppina Rindone who helped me correcting several typos and other *"petites bêtises"*. *Grazie*.

In these years I had the chance to work with several brilliant coauthors. In addition to the already mentioned Dominique Perrin, Valérie Berthé and Giuseppina Rindone, I really want to thank as well Clelia De Felice, Vincent Delecroix, Julien Leroy, Christophe Reutenauer. *Grazie*, *Merci*.

Many thanks go also to my colleauges and friends at Université Paris-Est, among them: Feriel Abboud, Zakaria Chemli, Jacopo Corbetta, Pavel Heller, Sonja Hiltunen, Revekka Kyriakoglou, Vincent Penelle, Jérôme Pilliet and Tim Smith. *Thanemirthe*, *Shukraan*, *Grazie*, *Díky*, *Tack*, *Kiitos*, *Efcharistó*, *Merci*, *Thanks*.

During my permanence in the Laboratoire d'Informatique Gaspard-Monge sveral other adviced me or helped me, among them I want to show my gratitude to Marie-Pierre Béal, Claire David, Patrice Herault, Corinne Palescandalo. *Merci*.

Finally I would like to thank the people who supported me from afar in Sicily, as my family and my longtime friend Roberto Signorello. *Grazie*.

This work was supported by grants from Région Île-de-France and partially by the ANR project Eqinocs ANR-13-BS02-004 and the PHC project Tournesol 34212UG.

# Chapter 1

# Preliminaries

In this chapter we fix the notation we will use in the rest of the manuscript and we give some preliminary result.

In Section 1.2.1 we give the definitions concerning words and set of words. We discuss both about the free monoid (positive words) and about the free group. We also define the extension graph of a word in a factorial set, one of the main notion that we will develop in the next chapters.

Section 1.2 is about bifix codes. We define the degree of a bifix code, its set of internal factors and its kernel. We also give the definition of derived code and of coding morphism. This last notion is related to some of the main results of this manuscript. Moreover, we define two tranformations on the set of codes: the internal transformation that given a bifix code give us another bifix code on the same alphabet, and the composition of codes, dealing with codes on (in general) different alphabets and that given two codes allow us to construct a third one.

In Section 1.3 we introduce a few kind of automata. We show the connection between automata and free group. Moreover, we define the Rauzy graph of a set of words.

Finally, in Section 1.4 we define return words, another fundamental notion that we will use for some of the main results of this manuscript.

## 1.1 Words and sets

Let $A$ be a finite nonempty alphabet. We denote by $A^*$ the free monoid on $A$, that is the set of all finite words on $A$. We denote by $\varepsilon$ the empty word and by $A^+ = A^* \setminus \{\varepsilon\}$.

We denote by $|w|$ the length $n$ of a word $w$ and by $|w|_a$ the number of occurrences of the letter $a \in A$ in the word $w$. Of course, one has $|w| = \sum_{a \in A} |w|_a$.

The *reversal* of a word $w = a_0 a_1 \cdots a_{n-1}$ with $a_i \in A$ is the word $\tilde{w} = a_{n-1} \cdots a_1 a_0$. A word $w$ is said to be a *palindrome* if $w = \tilde{w}$.

A *factor* of a word $x$ is a word $v$ such that $x = uvw$. If $u = \varepsilon$ (resp. $w = \varepsilon$) we say that $v$ is a *prefix* (resp. *suffix*) of $x$. If both $u$ and $w$ are nonempty, we say that $v$ is an *internal factor* of $x$. A set of words on the alphabet $A$ is said to be *factorial* if it contains the factors of its elements as well as the alphabet $A$.

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 denote by $A^{\mathbb{N}}$ the set of infinite words on the alphabet $A$. The notions of factor, prefix and suffix are naturally extendend to infinite words. For a set $X \subset A^{\mathbb{N}}$, we denote by $\mathrm{Fac}(X)$ the set of factors of the words of $X$. For an infinite word $x \in A^{\mathbb{N}}$, we simply denote $\mathrm{Fac}(x)$ the set of factors of $x$.

The set $A^{\mathbb{N}}$ is equipped with a distance defined for every $u, v \in A^{\mathbb{N}}$ by $d(u, v) = 2^{-n}$ with $n = \min\{k \geq 0 \mid x_k \neq y_k\}$, with the convention that $d(x, y) = 0$ if $x = y$. With respect to this distance, the set $A^{\mathbb{N}}$ becomes a topological space, often called the *Cantor space* (see, for example, [52]).

**Example 1.1.1** Let $A = \{a, b\}$. Let $x = ab^\omega = \lim_{n\to\infty} ab^n$ be an infinite word on $A$. One has $\mathrm{Fac}(x) = \{ab^n \mid n \in \mathbb{N}\} \cup \{b^n \mid n \in \mathbb{N}\}$.

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 $uvw \in S$.

We say that an infinite word $x$ is *recurrent* if for any $u \in \mathrm{Fac}(x)$ there is a $v \in \mathrm{Fac}(x)$ such that $uvu \in \mathrm{Fac}(x)$. As well known, for any recurrent set $S$ there is a recurrent infinite word $x$ such that $S = \mathrm{Fac}(x)$ and conversely, for any recurrent infinite word $x$, the set $\mathrm{Fac}(x)$ is recurrent (see for example [48]).

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.

Given two words $u, v \in A^*$, with $u$ a prefix of $v$, we define $u^{-1}v$ as the unique word $w$ such that $uw = v$. The *residual* of a set $X \subset A^*$ with respect to a word $u$ as the set

$$u^{-1}X = \{v \in A^* \mid uv \in W\}.$$

The definitions of $vu^{-1}$ and $Xu^{-1}$ for two words $u, v$ and a set $X$ are symmetric. We will use this notion in Section 1.3

### 1.1.1 Free groups and laminary sets

We fix our notation concerning free groups (see, for example, [53]). Given an alphabet $A$ be an alphabet we denote by $A^{-1} = \{a^{-1} \mid a \in A\}$ a new alphabet called the *inverse* of $A$. Given a word $w = a_0 a_1 \cdots a_{n-1}$ its inverse is the word $w^{-1} = a_{n-1}^{-1} a_{n-2}^{-1} \cdots a_0^{-1}$.

We denote by $F_A$ the free group on the alphabet $A$. It is identified with the set of all words on the alphabet $A \cup A^{-1}$ which are *reduced*, in the sense that they do not have any factor $aa^{-1}$ or $a^{-1}a$ for $a \in A$. Sometimes we also denote by $\bar{a}$ the inverse $a^{-1}$ of a letter $a \in A$.

Note that when $u$ is a prefix of $v$, we recover the definition of $u^{-1}v$ given at the end of the previous subsection.

We extend the bijection $a \mapsto a^{-1}$ to an involution on $A \cup A^{-1}$ by defining $(a^{-1})^{-1} = a$. For any word $w$ on $A \cup A^{-1}$ there is a unique reduced word equivalent to $w$ modulo the relations $aa^{-1} \equiv a^{-1}a \equiv \varepsilon$ for $a \in A$. If $u$ is the reduced word equivalent to $w$, we say that $w$ *reduces* to $u$ and we denote $w \equiv u$. We also denote $u = \rho(w)$. The product of two elements $u, v \in F_A$ is the reduced word $w$ equivalent to $uv$, namely $\rho(uv)$.

A set of reduced words on the alphabet $A \cup A^{-1}$ is said to be *symmetric* if it contains the inverses of its elements. A symmetric factorial set of reduced words on the alphabet $A \cup A^{-1}$ is called a *laminary set* on $A$.

An infinite laminary set $S$ is called *semi-recurrent* if for any $u, w \in S$, there is a $v \in S$ such that $uvw \in S$ or $uvw^{-1} \in S$. Likewise, it is said to be *uniformly semi-recurrent* if for any word $u \in S$ there is an integer $n \geq 1$ such that for any word $w$ of length $n$ in $S$, $u$ or $u^{-1}$ is a factor of $w$. A uniformly semi-recurrent set is semi-recurrent.

Following the terminology of [26], 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 $x^{-1} \in S_+$. Note that if $S$ is a semi-recurrent orientable laminary set, then the sets $S_+, S_-$ as above are unique (up to their interchange). The sets $S_+, S_-$ are called the *components* of $S$. Moreover a uniformly recurrent and orientable laminary set is a union of two uniformly recurrent sets. Indeed, $S_+$ and $S_-$ are uniformly recurrent.

### 1.1.2 Morphisms

A *morphism* $f : A^* \to B^*$ is a monoid morphism from $A^*$ into $B^*$. If $a \in A$ is such that the word $f(a)$ begins with $a$ and if $|f^n(a)|$ tends to infinity with $n$, there is a unique infinite word denoted $f^\omega(a) = \lim_{n\to\infty} f^n(a)$ which has all words $f^n(a)$ as prefixes. It is called a *fixed point* of the morphism $f$.

A morphism $f : A^* \to A^*$ is called *primitive* if there is an integer $k$ such that for all $a, b \in A$, the letter $b$ appears in $f^k(a)$. If $f$ is a primitive morphism, the set of factors of any fixed point of $f$ is uniformly recurrent (see [41] Proposition 1.2.3 for example).

An infinite word $y$ over an alphabet $B$ is called *morphic* if there exists a morphism $f$ on an alphabet $A$, a fixed point $x = f^\omega(a)$ of $f$ and a morphism $\sigma : A^* \to B^*$ such that $y = \sigma(x)$. If $A = B$ and $\sigma$ is the identity map, we call $y$ *purely morphic*. If $f$ is primitive we say that the word is *primitive morphic*.

Extending the definition, we say that a set $\mathrm{Fac}(x)$ is *morphic* (resp. *purely morphic*, *primitive morphic*) if the infinite word $x$ is morphic (resp. purely morphic, primitive morphic).

**Example 1.1.2** Let $A = \{a, b\}$. Let $A = \{a, b\}$ and let $\varphi$ be the morphism from $A^*$ to itself defined by $\varphi : a \mapsto ab, b \mapsto a$. The *Fibonacci word*

$$x = abaababaabaababaababa \ldots$$

is the fixed point $x = \varphi^\omega(a)$ of the Fibonacci morphism. The set $\mathrm{Fac}(x)$ of factors of $x$ is called the *Fibonacci set.*

**Example 1.1.3** Let $A = \{a, b, c\}$. The *Chacon word* on three letters is the fixed point $x = f^\omega(a)$ of the morphism $f$ from $A^*$ into itself defined by $f(a) = aabc$, $f(b) = bc$ and $f(c) = abc$. Thus $x = aabcaabcbcabc\cdots$. The *Chacon set* is the set $S$ of factors of $x$. The element of lenght at most 4 are the labels of the paths starting at the root of the tree represented in Figure 1.1.

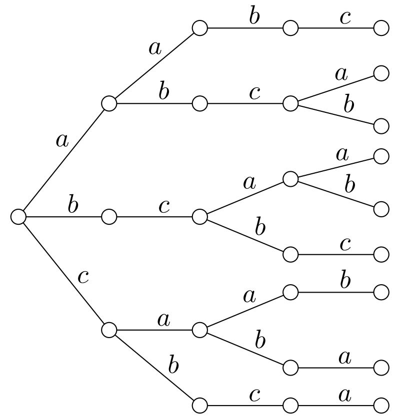

Figure 1.1: The words of lenght $\leq 4$ of the Chacon set.

### 1.1.3 Extension graphs

Let $S$ be a factorial set on the alphabet $A$. For a word $w \in S$, we define

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

and furthermore

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

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 *biextendable* if $b(w) > 0$. A factorial set $S$ is called *right-extendable* (resp. *left-extendable*, resp. *biextendable*) if every word in $S$ is right-extendable (resp. left-extendable, resp. biextendable).

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

An infinite word is *episturmian* if the set of its factors is closed under reversal and contains for each $n$ at most one word of length $n$ which is right-special (see [7] for more references). It is a *strict episturmian* word if it has exactly one right-special word of each length and moreover each right-special factor $u$ is such that $r(u) = \mathrm{Card}(A)$.

An *Arnoux-Rauzy set* is the set of factors of a strict episturmian word. Any Arnoux-Rauzy set is uniformly recurrent (see [7]).

**Example 1.1.4** The Fibonacci word defined in Example 1.1.2 is a Sturmian word (see [52]). Thus the Fibonacci set is an Arnoux-Rauzy set.

For a word $w \in S$, we define the *multiplicity*

$$m_S(w) = b_S(w) - \ell_S(w) - r_S(w) + 1. \tag{1.1}$$

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

A biextendable word $w$ is called *ordinary* if $B(w) \subset (a \times A) \cup (A \times b)$ for some $(a, b) \in B(w)$ (see [16, Chapter 4]). If $S$ is biextendable any ordinary word is neutral. Indeed, one has $B(w) = (a \times (R(w) \setminus b)) \cup ((L(w) \setminus a) \times b) \cup (a, b)$ and thus $b(w) = \ell(w) + r(w) - 1$.

**Example 1.1.5** In a Sturmian set, any word is ordinary. Indeed, for any bispecial word $w$, there is a unique letter $a$ such that $aw$ is right-special and a unique letter $b$ such that $wb$ is left-special. Then $awb \in S$ and $B(w) = (a \times A) \cup (A \times b)$.

Let $S$ be a biextendable set of words. For $w \in S$, we consider the undirected bipartite graph $\mathcal{E}_S(w)$ with vertices the disjoint union of $L_S(w)$ and $R_S(w)$ with edges the pairs $(a, b) \in B_S(w)$. This graph is called the *extension graph* of $w$. We sometimes denote by $1 \otimes L(w)$ 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 $b_S(w)$ edges, the number $1 - m(w)$ is the Euler characteristic of the graph $\mathcal{E}(w)$ [1].

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 [21, Proposition 3.5] (see also [16, Theorem 4.5.4]).

**Proposition 1.1.6** *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 $n \geq 0$.*

A classical theorem by Morse and Hedlund (see [56]) states that the set of factors of an infinite word is either eventually constant or strictly increasing. The first case corresponds to set of factors of *ultimately periodic* words, i.e. words of the form $uv^\omega$ with $u, v \in A^*$. In the other case, one has $p_n \geq n + 1$ for all $n \in \mathbb{N}$.

Arnoux-Rauzy sets are exactly factorial sets with minimal non-constant factor complexity, that is such that $p_n = n + 1$.

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

**Example 1.1.7** Let $A = \{a, b, c\}$. The Tribonacci word

$$x = abacabaabacababacabaabacaba \cdots$$

is the fixed point $x = f^\omega(a)$ of the morphism $f : A^* \to A^*$ defined by $f(a) = ab$, $f(b) = ac$, $f(c) = a$. It is a strict episturmian word (see [45]). The set $\mathrm{Fac}(x)$ of factors of $x$ is the *Tribonacci set*.

In Chapter 3 we will see that any Arnoux-Rauzy set is a recurrent neutral set of characteristic 1.

**Example 1.1.8** The Fibonacci set defined (Example 1.1.2) and the Tribonacci set (Example 1.1.7) are both neutral sets of characteristic 1. Indeed one can prove that every word, including the empty word, is neutral.

## 1.2 Bifix codes

A set of nonempty words $X \subset A^+$ is said a *code* if the relation $x_1 \cdots x_n = y_1 \cdots y_m$ with $n, m \geq 1$ and $x_1, \ldots x_n, y_1, \ldots y_m \in X$ implies $n = m$ and $x_i = y_i$ for every $1 \leq i \leq n$.

A *prefix code* is a set of nonempty words which does not contain any proper prefix of its elements. Clearly, a prefix code is a code. A *suffix code* is defined symmetrically. A *bifix code* is a set which is both a prefix code and a suffix code (see [8] for a more detailed introduction).

We denote by $X^*$ the submonoid generated by a set $X$ of words. The submonoid $M$ generated by a prefix code satisfies the following property: if $u, uv \in M$, then $v \in M$. Such a submonoid is said to be *right unitary*. The definition of a *left unitary* submonoid is symmetric and the submonoid generated by a suffix code is left unitary. Conversely, any right unitary (resp. left unitary) submonoid of $A^*$ is generated by a unique prefix code (resp. suffix code) (see [8]).

Let $S$ be a recurrent set of words. A prefix code $X \subset S$ is $S$-*maximal* if it is not properly contained in any prefix code $Y \subset S$. Note that if $X \subset S$ is an $S$-maximal prefix code, any word of $S$ is comparable for the prefix order with a word of $X$.

A set $X \subset S$ is *right $S$-complete* if any word of $S$ is a prefix of a word in $X^*$. For a factorial set $S$, a prefix code is $S$-maximal if and only if it is right $S$-complete (see [7, Proposition 3.3.2]).

**Example 1.2.1** Let $S$ be the Fibonacci set defined in Example 1.1.2. The set $X = \{a, ba\}$ is an $S$-maximal prefix code, since $X$ is right $S$-complete.

Similarly a bifix code $X \subset S$ is $S$-maximal if it is not properly contained in a bifix code $Y \subset S$. For a recurrent set $S$, a finite bifix code is $S$-maximal as a bifix code if and only if it is an $S$-maximal prefix code (see [7, Theorem 4.2.2]). For a uniformly recurrent set $S$, any finite bifix code $X \subset S$ is contained in a finite $S$-maximal bifix code ([7, Theorem 4.4.3]).

### 1.2.1 Parses and degree

A *parse* of a word $w$ with respect to a bifix code $X$ is a triple $(v, x, u)$ such that $w = vxu$ where $v$ has no suffix in $X$, $u$ has no prefix in $X$ and $x \in X^*$. We denote by $d_X(w)$ the number of parses of a word $w$ with respect to $X$. The $S$-degree of $X$, denoted by $d_X(S)$ is the maximal number of parses with respect to $X$ of a word of $S$. It can be finite or infinite.

Let $X$ be a bifix code. The number of parses of a word $w$ with respect to $X$, denoted by $\delta_X(w)$, is also equal to the number of suffixes of $w$ which have no prefix in $X$ and to the number of prefixes of $w$ which have no suffix in $X$ (see [8, Proposition 6.1.6]).

If $X$ is a prefix code, by [7, Proposition 4.1.6], for any $u \in A^*$ and $a \in A$, one has

$$\delta_X(ua) = \begin{cases} \delta_X(u) & \text{if } ua \in A^*X, \\ \delta_X(u) + 1 & \text{otherwise.} \end{cases} \tag{1.2}$$

**Example 1.2.2** Let $S$ be a recurrent set. For any integer $n \geq 1$, the set $S \cap A^n$ is an $S$-maximal bifix code of $S$-degree $n$.

The set of *internal factors* of a set of words $X$, denoted $I(X)$ is the set of words $w$ such that there exist nonempty words $u, v$ with $uwv \in X$ (recall also Section ).

Let $S$ be a set of words. A set $X \subset S$ is said to be $S$-*thin* if there is a word of $S$ which is not a factor of $X$. If $S$ is biextendable any finite set $X \subset S$ is $S$-thin. Indeed, any long enough word of $S$ is not a factor of $X$. The converse is true if $S$ is uniformly recurrent. Indeed, let $w \in S$ be a word which is not a factor of $X$. Then any long enough word of $S$ contains $w$ as a factor, and thus is not itself a factor of $X$.

Let $S$ be a recurrent set and let $X$ be a finite bifix code. By [7, Theorem 4.2.8], $X$ is $S$-maximal if and only if its $S$-degree $d_S(X)$ is finite. Moreover, in this case, a word $w \in S$ is such that $d_X(w) < d_S(X)$ if and only if it is an internal factor of $X$, that is

$$I(X) = \{w \in S \mid d_X(w) < d_S(X)\}.$$

Thus any word of $S$ which is not an internal factor of $X$ has $d_S(X)$ parses. In particular, any word of $X$ of maximal length has $d_S(X)$ parses.

The *kernel* of a bifix code $X$ is the set $K(X) = I(X) \cap X$. Thus it is the set of words of $X$ which are also internal factors of $X$. By [7, Theorem 4.3.11], a finite $S$-maximal bifix code is determined by its $S$-degree and its kernel.

**Example 1.2.3** Let $S$ be a recurrent set containing the alphabet $A$. The only $S$-maximal bifix code of $S$-degree 1 is the alphabet $A$. This is clear since $A$ is the unique $S$-maximal bifix code of $S$-degree 1 with empty kernel.

**Example 1.2.4** Let $S$ be the Fibonacci set. The set $X = \{a, baab, bab\}$ is the unique $S$-maximal bifix code of $S$-degree 2 with kernel $\{a\}$. Indeed, the word $bab$ is not an internal factor and has two parses, namely $(1, bab, 1)$ and $(b, a, b)$.

**Example 1.2.5** Let $S$ be the Fibonacci set. The set $X = \{aaba, ab, baa, baba\}$ is a $S$-maximal bifix code of $S$-degree 3 with kernel $\{ab\}$. Indeed, the word $aaba$, that is not an internal factor, has three parses, namely $(1, aaba, 1)$, $(a, ab, a)$, and $(aa, 1, ba)$.

The following result is [7, Theorem 4.3.12].

**Theorem 1.2.6** *Let $S$ be a recurrent set. A bifix code $Y \subset S$ is the kernel of some $S$-thin $S$-maximal bifix code of $S$-degree $d$ if and only if $Y$ is not $S$-maximal and $\delta_Y(y) \leq d - 1$ for all $y \in Y$.*

The following proposition allows one to embed an $S$-maximal bifix code in a maximal one of the same degree.

**Proposition 1.2.7** *Let $S$ be a recurrent set. For any $S$-thin and $S$-maximal bifix code $X$ of $S$-degree $d$, there is a thin maximal bifix code $X'$ of degree $d$ such that $X = X' \cap S$.*

*Proof.* Let $K$ be the kernel of $X$ and let $d$ be the $S$-degree of $X$. By Theorem 1.2.6, the set $K$ is not $S$-maximal and $\delta_K(y) \leq d - 1$ for any $y \in K$. Thus, applying again Theorem 1.2.6 with $S = A^*$, there is a maximal bifix code $X'$ with kernel $K$ and degree $d$. Then, by [7, Theorem 4.2.11], the set $X' \cap S$ is an $S$-maximal bifix code.

Let us show that $X \cup X'$ is prefix. Suppose that $x \in X$ and $x' \in X'$ are comparable for the prefix order. We may assume that $x$ is a prefix of $x'$ (the other case works symmetrically). If $x \in K$, then $x \in X'$ and thus $x = x'$. Otherwise, $\delta_X(x) = d$. Set $x = pa$ with $a \in A$. Then, by Equation (1.2), $\delta_X(x) = \delta_X(p)$ and thus $\delta_X(p) = d$. But since all the factors of $p$ which are in $X$ are in $K$, we have $\delta_X(p) = \delta_K(p)$. Analogously, since all factors of $p$ which are in $X'$ are in $K$, we have $\delta_K(p) = \delta_{X'}(p)$. Therefore $\delta_{X'}(p) = d$. But, since $X'$ has degree $d$, $\delta_{X'}(x) \leq d$. Then, by Equation (1.2) again, we have $\delta_{X'}(x) = d$and $x \in A^*X'$. Let $z$ be the suffix of $x$ which is in $X'$. If $x \neq x'$, then $z = x$ or $z \in K$ and in both cases $z \in X$. Since $X'$ is prefix and $X$ is suffix, this implies $z = x = x'$.

Since $X$ and $X' \cap S$ are $S$-maximal prefix codes included in $(X \cup X') \cap S$, this implies that $X = X' \cap S$. ■

**Example 1.2.8** Let $S$ and $X$ as in Example 1.2.4. Then $X' = a \cup ba^*b$ is the maximal bifix code with kernel $\{a\}$ of degree 2 such that $X' \cap S = X$.

### 1.2.2 Derived codes and coding morphisms

The following result, that we will use in Chapter 3, is the dual of [7, Theorem 4.3.7].

**Theorem 1.2.9** *Let $S$ be a recurrent set and let $X$ be a finite $S$-maximal bifix code of $S$-degree $n$. The set of nonempty proper prefixes of $X$ is a disjoint union of $n-1$ $S$-maximal suffix codes.*

Let now $S$ be a recurrent set and $X$ be a finite $S$-maximal bifix code of $S$-degree $d \geq 2$. Let us define the sets $G = (IA \cap S) \setminus I$ and $D = (AI \cap S) \setminus I$, where $I = I(X)$ and $K = K(X)$. By [7, Theorem 4.3.1] the set $X' = K \cup (G \cap D)$ is an $S$-maximal bifix code of $S$-degree $d-1$, called the *derived code* of $X$.

**Example 1.2.10** Let $S$ the Fibonacci set and $X$ be the $S$-maximal bifix code of $S$-degree 2 defined in Example 1.2.4. The kernel and the set of internal factors of $X$ are respectively $K = \{a\}$ and $I = \{\varepsilon, a, aa\}$. We have $G = \{aab, ab, b\}$, $D = \{b, ba, baa\}$ and thus the derived code is $X' = \{a, b\}$, the only $S$-maximal bifix code of $S$-degree 1.

**Example 1.2.11** Let $S$ the Fibonacci set and $X$ be the $S$-maximal bifix code of $S$-degree 3 defined in Example 1.2.5. The kernel and the set of internal factors of $X$ are respectively $K = \{ab\}$ and $I = \{\varepsilon, a, ab, b\}$. The derived code is $X' = A^2 \cap S$.

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

Let $S$ be a factorial set and let $f$ be a coding morphism for a finite bifix code $X \subset S$. The set $f^{-1}(S)$ is called a *bifix decoding* of $S$. When $X$ is an $S$-maximal bifix code, it is called a *maximal bifix decoding* of $S$.

**Example 1.2.12** Let $S$ be the Fibonacci set over the alphabet $A = \{a, b\}$ and let $B = \{\alpha, \beta, \gamma\}$. Let us consider the $S$-maximal bifix code code $X = S \cap A^2 = \{aa, ab, ba\}$ and the morphism $f : B^* \to A^*$ defined by $f : \alpha \mapsto aa, \beta \mapsto ab$ and $\gamma \mapsto ba$. Thus, the set $f^{-1}(S)$ is a maximal bifix decoding of the Fibonacci set. Moreover, we can see that $f^{-1}(S)$ is the set of factors of the infinite word $f^{-1}(x) = f^{-1}(ab\ aa\ ba\ ba\ ab\ aa \cdots) = \beta\alpha\gamma\gamma\beta\alpha\cdots$, where $x$ is the Fibonacci word defined in Example 1.1.2.

### 1.2.3 Internal transformations

In this section we describe an operation on bifix codes called internal transformation and prove a property of this transformation (Proposition 1.2.14). For a more detailed presentation see [8, Chapter 6]. It will be used in Section 2.2.

Let $X \subset S$ be a set of words and $w \in S$ a word. Let

$$G = Xw^{-1}, \qquad D = w^{-1}X, \tag{1.3}$$
$$G_0 = (wD)w^{-1} \quad D_0 = w^{-1}(Gw), \tag{1.4}$$
$$G_1 = G \setminus G_0, \qquad D_1 = D \setminus D_0. \tag{1.5}$$

Note that $Gw \cap wD = G_0w = wD_0$. Consequently $G_0^*w = wD_0^*$. The set

$$Y = (X \cup w \cup (G_1wD_0^*D_1 \cap S)) \setminus (Gw \cup wD) \tag{1.6}$$

is said to be obtained from $X$ by *internal transformation* with respect to $w$. When $Gw \cap wD = \emptyset$, the transformation takes the simpler form

$$Y = (X \cup w \cup (GwD \cap S)) \setminus (Gw \cup wD). \tag{1.7}$$

It is this form which is used in [7] to define the internal transformation.

**Example 1.2.13** Let $S$ be the Fibonacci set. Let $X = S \cap A^2$. The internal transformation applied to $X$ with respect to $b$ gives $Y = \{aa, aba, b\}$. The internal transformation applied to $X$ with respect to $a$ gives $Y' = \{a, baab, bab\}$.

The following result is proved in [7, Proposition 4.4.5] in the case $G_0 = \emptyset$.

**Proposition 1.2.14** *Let $S$ be a uniformly recurrent set and let $X \subset S$ be a finite $S$-maximal bifix code of $S$-degree $d$. Let $w \in S$ be a nonempty word such that the sets $G_1, D_1$ defined by Equation* (1.5) *are nonempty. Then the set $Y$ obtained as in Equation* (1.6) *is a finite $S$-maximal bifix code with $S$-degree at most $d$.*

*Proof.* By Proposition 1.2.7 there is a thin maximal bifix code $X'$ of degree $d$ such that $X = X' \cap S$. Let $Y'$ be the code obtained from $X'$ by internal transformation with respect to $w$. Then

$$Y' = (X' \cup w \cup (G_1'wD_0'^*D_1')) \setminus (G'w \cup wD')$$

with $G' = X'w^{-1}$, $D' = w^{-1}X'$, and $G_0' = (wD')w^{-1}$, $D_0' = w^{-1}(G'w)$, $G_1' = G' \setminus G_0'$, $D_1' = D' \setminus D_0'$. We have $G = G' \cap Sw^{-1}$, $D = D' \cap w^{-1}S$, and $D_i = D_i' \cap w^{-1}S$, $G_i = G_i' \cap Sw^{-1}$ for $i = 0, 1$. In particular $G_1 \subset G_1'$, $D_1 \subset D_1'$. Thus $G_1', D_1' \neq \emptyset$. This implies that $Y'$ is a thin maximal bifix code of degree $d$ (see [8, Proposition 6.2.8]).

Since $w \in S$, we have $Y = Y' \cap S$. By [7, Theorem 4.2.11], $Y$ is an $S$-maximal bifix code of $S$-degree at most $d$. Since $S$ is uniformly recurrent, this implies that $Y$ is finite. ∎

Note that when $G_0 = \emptyset$, the bifix code $Y$ has $S$-degree equal to $d$ (see [7, Proposition 4.4.5]). We will see in the proof of Proposition 2.2.5 another case where it is true. We have no example where it is not true.

**Example 1.2.15** Let $S$ be the Fibonacci set, and let $X = S \cap A^2$, as in Example 1.2.13. Let $w = a$. Then $Y = \{a, baab, bab\}$ is the $S$-maximal bifix code of $S$-degree 2 already considered in Example 1.2.13.

### 1.2.4 Composition of codes

We introduce the notion of composition of codes (see [8] for a more detailed presentation).

For a set $X \subset A^*$, we denote by $\mathrm{alph}(X)$ the set of letters $a \in A$ which appear in the words of $X$.

Let $Z \subset A^*$ and $Y \subset B^*$ be two finite codes with $B = \mathrm{alph}(Y)$. Then the codes $Y$ and $Z$ are *composable* if there is a bijection from $B$ onto $Z$. Since $Z$ is a code, this bijection defines an injective morphism from $B^*$ into $A^*$. If $f$ is such a morphism, then $Y$ and $Z$ are called composable *through* $f$. The set

$$X = f(Y) \subset Z^* \subset A^* \tag{1.8}$$

is obtained by *composition* of $Y$ and $Z$ (by means of $f$). We denote it by $X = Y \circ_f Z$, or by $X = Y \circ Z$ when the context permits it. Since $f$ is injective, $X$ and $Y$ are related by bijection, and in particular $\mathrm{Card}(X) = \mathrm{Card}(Y)$. The words in $X$ are obtained just by replacing, in the words of $Y$, each letter $b$ by the word $f(b) \in Z$.

**Example 1.2.16** Let $A = \{a, b\}$ and $B = \{u, v, w\}$. Let $f : B^* \to A^*$ be the morphism defined by $f(u) = aa$, $f(v) = ab$ and $f(w) = ba$. Let $Y = \{u, vu, vv, w\}$ and $Z = \{aa, ab, ba\}$. Then $Y, Z$ are composable through $f$ and $Y \circ_f Z = \{aa, abaa, abab, ba\}$.

If $Y$ and $Z$ are two composable codes, then $X = Y \circ Z$ is a code [8, Proposition 2.6.1] and if $Y$ and $Z$ are prefix (suffix) codes, then $X$ is a prefix (suffix) code. Conversely, if $X$ is a prefix (suffix) code, then $Y$ is a prefix (suffix) code.

We extend the notation alph as follows. For two codes $X, Z \subset A^*$ we denote $\mathrm{alph}_Z(X)$ the set of $z \in Z$ such that $uzv \in X$ for some $u, v \in Z^*$. The following is [8, Proposition 2.6.6].

**Proposition 1.2.17** *Let $X, Z \subset A^*$ be codes. There exists a code $Y$ such that $X = Y \circ Z$ if and only if $X \subset Z^*$ and $\mathrm{alph}_Z(X) = Z$.*

The following statement generalizes [8, Propositions 2.6.4 and 2.6.12] for prefix codes.

**Proposition 1.2.18** *Let $Y, Z$ be finite prefix codes composable through $f$ and let $X = Y \circ_f Z$.*

*(i) For every set $T$ such that $Y \subset T$ and $Y$ is a $T$-maximal prefix code, $X$ is an $f(T)$-maximal prefix code.*

*(ii) For every set $S$ such that $X, Z \subset S$, if $X$ is an $S$-maximal prefix code, $Y$ is an $f^{-1}(S)$-maximal prefix code and $Z$ is an $S$-maximal prefix code. The converse is true if $S$ is recurrent.*

*Proof.* (i) Let $w \in f(T)$ and set $w = f(v)$ with $v \in T$. Since $Y$ is $T$-maximal, there is a word $y \in Y$ which is prefix-comparable with $v$. Then $f(y)$ is prefix-comparable with $w$. Thus $X$ is $f(T)$-maximal.
(ii) Since $X$ is an $S$-maximal prefix code, any word in $S$ is prefix-comparable with some element of $X$ and thus with some element of $Z$. Therefore, $Z$ is $S$-maximal. Next if $u \in f^{-1}(S)$, $v = f(u)$ is in $S$ and is prefix-comparable with a word $x$ in $X$. Assume that $v = xt$. Then $t$ is in $Z^*$ since $v, x \in Z^*$. Set $w = f^{-1}(t)$ and $y = f^{-1}(x)$. Since $u = yw$, $u$ is prefix-comparable with $y$ which is in $Y$. The other case is similar.

Conversely, assume that $S$ is recurrent. Let $w$ be a word in $S$ of length strictly larger than the sum of the maximal length of the words of $X$ and $Z$. Since $S$ is recurrent, the set $Z$ is right $S$-complete, and consequently the word $w$ is a prefix of a word in $Z^*$. Thus $w = up$ with $u \in Z^*$ and $p$ a proper prefix of a word in $Z$. The hypothesis on $w$ implies that $u$ is longer than any word of $X$. Let $v = f^{-1}(u)$. Since $u \in S$, we have $v \in f^{-1}(S)$. It is not possible that $v$ is a proper prefix of a word of $Y$ since otherwise $u$ would be shorter than a word of $X$. Thus $v$ has a prefix in $Y$. Consequently $u$, and thus $w$, has a prefix in $X$. Thus $X$ is $S$-maximal. ■

Note that the converse of (ii) is not true if the hypothesis that $S$ is recurrent is replaced by factorial. Indeed, for $S = \{\varepsilon, a, b, aa, ab, ba\}$, $Z = \{a, ba\}$, $Y = \{uu, v\}$, $f(u) = a$ and $f(v) = ba$, one has $f^{-1}(S) = \{\varepsilon, u, uu, v\}$ and $X = \{aa, ba\}$, which is not an $S$-maximal prefix code.

Note also that when $S$ is recurrent (or even uniformly recurrent), the set $T = f^{-1}(S)$ need not be recurrent.

**Example 1.2.19** Let $S = \mathrm{Fac}\,((ab)^*)$ be the set of factors of $(ab)^*$. Let $B = \{u, v\}$ and let $f : B^* \to A^*$ be defined by $f(u) = ab$, $f(v) = ba$. Then $T = u^* \cup v^*$ which is not recurrent.

## 1.3 Automata

We denote by $\mathcal{A} = (Q, 1, T, E)$ a deterministic automaton with a set $Q$ of states, $1 \in Q$ as initial state and $T \subset Q$ as set of terminal states. The set $E$ of edges is a subset of $Q \times A \times Q$. Following the notation of [8], we usually omit the set of edges and just denote an automaton $\mathcal{A}$ as the triple $(Q, 1, T)$.

For $p \in Q$ and $w \in A^*$, we denote $p \cdot w = q$ if there is a path labeled $w$ from $p$ to the state $q$ and $p \cdot w = \emptyset$ otherwise. The automaton is *finite* when $Q$ is finite.

The set *recognized* by the automaton is the set of words $w \in A^*$ such that $i \cdot w \in T$.

All automata considered here are deterministic and we simply call them 'automata' to mean 'deterministic automata'.

The automaton $\mathcal{A}$ is *trim* if for any $q \in Q$, there is a path from 1 to $q$ and a path from $q$ to some $t \in T$.

An automaton is called *simple* if it is trim and if it has a unique terminal state which coincides with the initial state. The set recognized by a simple automaton is a right unitary submonoid. Thus it is generated by a prefix code (see Section 1.2).

An automaton $\mathcal{A} = (Q, 1, T)$ is *complete* if for any state $p \in Q$ and any letter $a \in A$, one has $p \cdot a \neq \emptyset$.

For a nonempty set $L \subset A^*$, we denote by $\mathcal{A}(L)$ the *minimal automaton* of $L$. The states of $\mathcal{A}(L)$ are the nonempty residuals $u^{-1}L$ for $u \in A^*$ (recall Section 1.2.1). For $u \in A^*$ and $a \in A$, one defines $(u^{-1}L) \cdot a = (ua)^{-1}L$. The initial state is the set $L$ itself and the terminal states are the sets $u^{-1}L$ for $u \in L$.

Let $X$ be a prefix code and let $P$ be the set of proper prefixes of $X$. The *literal automaton* of $X^*$ is the simple automaton $\mathcal{A} = (P, \varepsilon, \varepsilon)$ with transitions defined for $p \in P$ and $a \in A$ by

$$p \cdot a = \begin{cases} pa & \text{if } pa \in P\,, \\ \varepsilon & \text{if } pa \in X\,, \\ \emptyset & \text{otherwise.} \end{cases}$$

One verifies that this automaton recognizes $X^*$. Thus for any prefix code $X \subset A^*$, there is a simple automaton $\mathcal{A} = (Q, 1, 1)$ which recognizes $X^*$. Moreover, the minimal automaton of $X^*$ is simple. Note that, in general, the literal automaton is not minimal in general (see Example 1.3.1).

**Example 1.3.1** Let $X = \{aa, ab, bba, bbb\}$ a prefix code over the alphabet $A = \{a, b\}$. The literal and the minimal automata of $X^*$ are represented in Figure 1.2 (the initial state is indicated by an incoming arrow and the terminal states by a double circle).

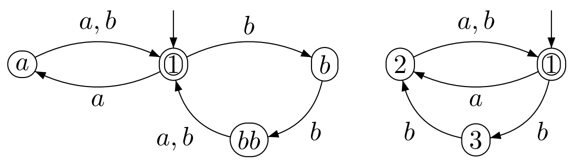


Figure 1.2: The literal and the minimal automata of $X^*$.

### 1.3.1 Groups and automata

A simple automaton $\mathcal{A} = (Q, 1, 1)$ is said to be *reversible* if for any $a \in A$, the partial map $\varphi_{\mathcal{A}}(a) : p \mapsto p \cdot a$ is injective. This condition allows to construct the *reversal* of the automaton as follows: whenever $q \cdot a = p$ in $\mathcal{A}$, then $p \cdot a = q$ in the reversal automaton. The state 1 is the initial and the unique terminal state of this automaton. Any reversible automaton is minimal [62] (but not conversely). The set recognized by a reversible automaton is a submonoid generated by a bifix code.

The following result is from [62]. We denote by $\langle X\rangle$ the subgroup of the free group $F_A$ generated by $X$.

**Proposition 1.3.2** *Let $X \subset A^+$ be a bifix code. The following conditions are equivalent.*

(i) $X^* = \langle X\rangle \cap A^*$*;*

(ii) *the minimal automaton of $X^*$ is reversible.*

A simple automaton $\mathcal{A} = (Q, 1, 1)$ is a *group automaton* if for any $a \in A$ the map $\varphi_{\mathcal{A}}(a) : p \mapsto p \cdot a$ is a permutation of $Q$. Thus, in particular, a group automaton is reversible. A finite reversible automaton which is complete is a group automaton.

The following result is proved in [7, Proposition 6.1.5].

**Proposition 1.3.3** *The following conditions are equivalent for a submonoid $M$ of $A^*$.*

*(i) $M$ is recognized by a group automaton with $d$ states.*

*(ii) $M = \varphi^{-1}(K)$, where $K$ is a subgroup of index $d$ of a group $G$ and $\varphi$ is a surjective morphism from $A^*$ onto $G$.*

*(iii) $M = H \cap A^*$, where $H$ is a subgroup of index $d$ of the free group on $A$.*

*If one of these conditions holds, the minimal generating set of $M$ is a maximal bifix code of degree $d$.*

A bifix code $Z$ such that $Z^*$ satisfies one of the equivalent conditions of Proposition 1.3.3 is called a *group code* of degree $d$.

Let $\mathcal{A} = (Q, 1, T)$ be a deterministic automaton. A *generalized path* is a sequence $(p_0, a_1, p_1, a_2, \ldots, p_{n-1}, a_n, p_n)$ with $a_i \in A \cup A^{-1}$ and $p_i \in Q$, such that for $1 \le i \le n$, one has $p_{i-1} \cdot a_i = p_i$ if $a_i \in A$ and $p_i \cdot a_i^{-1} = p_{i-1}$ if $a_i \in A^{-1}$. The *label* of the generalized path is the reduced word equivalent to $a_1a_2 \cdots a_n$. It is an element of the free group $F_A$. The set *described* by the automaton is the set of labels of generalized paths from 1 to a state in $T$. Since a path is a particular case of a generalized path, the set recognized by an automaton $\mathcal{A}$ is a subset of the set described by $\mathcal{A}$.

The set described by a simple automaton is a subgroup of $F_A$. It is called the *subgroup described* by $\mathcal{A}$.

**Example 1.3.4** Let $\mathcal{A} = (Q, 1, 1)$ be the automaton represented in Figure 1.3.

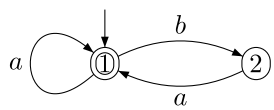


Figure 1.3: A simple automaton describing the free group on $\{a, b\}$.

The submonoid recognized by $\mathcal{A}$ is $\{a, ba\}^*$. Since $\{a, ba\}$ is a basis of the free group on $A$, the subgroup described by $\mathcal{A}$ is the free group on $A$.

The following result is [7, Proposition 6.1.3].

**Proposition 1.3.5** *Let $\mathcal{A}$ be a simple automaton and let $X$ be the prefix code generating the submonoid recognized by $\mathcal{A}$. The subgroup described by $\mathcal{A}$ is generated by $X$. If moreover $\mathcal{A}$ is reversible, then $X^* = \langle X \rangle \cap A^*$.*

For any subgroup $H$ of the free group $F_A$, the submonoid $H \cap A^*$ is right and left unitary and thus it is generated by a bifix code (see [8, Example 2.2.6]). A subgroup $H$ of $F_A$ is *positively generated* if there is a subset of $A^*$ which generates $H$. In this case, the set $H \cap A^*$ generates the subgroup $H$. Let $X$ be the bifix code which generates the submonoid $H \cap A^*$. Then $X$ generates the subgroup $H$. This shows that, for a positively generated subgroup $H$, there is a bifix code which generates $H$.

It is well known that a subgroup of finite index of the free group is positively generated (see, e.g, [7, Proposition 6.1.6]).

The following result is contained in [7, Propositions 6.1.4 and 6.1.5].

**Proposition 1.3.6** *For any positively generated subgroup $H$ of the free group on $A$, there is a unique reversible automaton $\mathcal{A}$ such that $H$ is the subgroup described by $\mathcal{A}$. The subgroup is of finite index if and only if this automaton is a finite group automaton.*

The reversible automaton $\mathcal{A}$ such that $H$ is the subgroup described by $\mathcal{A}$ is called the *Stallings automaton* of the subgroup $H$. It can also be defined for a subgroup which is not positively generated (see [46]).

The Stallings automaton of the subgroup $H$ generated by a bifix code $X \subset A^*$ can be obtained as follows. Start with the minimal automaton $\mathcal{A} = (Q, 1, 1)$ of $X^*$. Then, if there are distinct states $p, q \in Q$ and $a \in A$ such that $p \cdot a = q \cdot a$, merge $p, q$ (such a merge is called a *Stallings folding*). Iterating this operation leads to a reversible automaton which is the Stallings automaton of $H$ (see [46]).

A subgroup $H$ of the free group has finite index if and only if its Stallings automaton is a finite group automaton (see Proposition 1.3.6). In this case, the index of $H$ is the number of states of the Stallings automaton.

**Example 1.3.7** Let $X = \{aa, ab, ba\}$. The minimal automaton of $X^*$ is represented in Figure 1.4 on the left. It is not reversible because $2 \cdot a = 3 \cdot a$. Merging the states 2 and 3, we obtain the reversible automaton of Figure 1.4 on the right. It is actually a group automaton, which is the Stallings automaton of the subgroup $H = \langle X \rangle$.

Since the automaton describes the group $\mathbb{Z}/2\mathbb{Z}$, we conclude that the subgroup generated by $X$ is of index 2 in the free group on $A$. It is actually formed of the reduced words of even length.

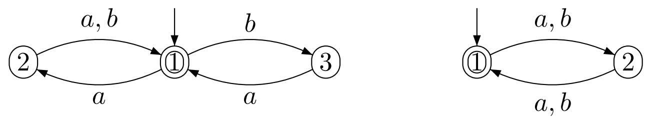


Figure 1.4: A Stallings folding.

### 1.3.2 Rauzy graphs

We first introduce the notion of a Rauzy graph (for a more detailed exposition, see [16]). Let $S$ be a factorial set. The *Rauzy graph* of $S$ of order $n \geq 0$ is the following labeled directed graph $G_n(S)$. Its vertices are the words in the set $S \cap A^n$. Its edges are the triples $(x, a, y)$ for all $x, y \in S \cap A^n$ and $a \in A$ such that $xa \in S \cap Ay$.

Let $u \in S \cap A^n$. The following properties follow easily from the definition of the Rauzy graph.

(i) For any word $w$ such that $uw \in S$, there is a path labeled $w$ in $G_n(S)$ from $u$ to the suffix of length $n$ of $uw$.

(ii) Conversely, the label of any path of length at most $n+1$ in $G_n(S)$ is in $S$.

When $S$ is recurrent, all Rauzy graph $G_n(S)$ are strongly connected. Indeed, let $u, w \in S \cap A^n$. Since $S$ is recurrent, there is a $v \in S$ such that $uvw \in S$. Then there is a path in $G_n(S)$ from $u$ to $w$ labeled $vw$ by property (i) above.

The Rauzy graph $G_n(S)$ of a recurrent set $S$ with a distinguished vertex $v$ can be considered as a simple automaton $\mathcal{A} = (Q, v, v)$ with set of states $Q = S \cap A^n$.

**Example 1.3.8** Consider again the Chacon set (see Example 1.1.3). The Rauzy graph $G_1(S)$ corresponding to the Chacon set is represented in Figure 3.11 on the left. The graph represented on the right is obtained by a Stalling folding of the graph $G_1(S)$.

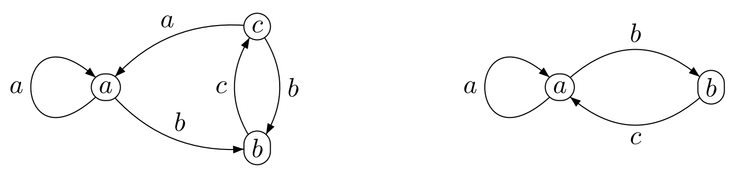


Figure 1.5: The graphs $G_1(S)$ (on the left) and the graph obtained by a Stalling folding (on the right).

We will prove in Chapter 3 that, for some particular class of sets, the group described by the Rauzy graph (seen as a simple automata) is the free group.

## 1.4 Return words

Let $S$ be a set of words over an alphabet $A$. Given a word $w \in S$, we define

$$\Gamma_S(w) = \{u \in S \mid wu \in S \cap A^+w\} \quad \text{and} \quad \Gamma'_S(w) = \{u \in S \mid uw \in S \cap wA^+\}.$$

When $S$ is recurrent, the sets $\Gamma_S(w)$ and $\Gamma'_S(w)$ are nonempty. Actually, in that case, both of them are infinite. Let

$$\mathcal{R}_S(w) = \Gamma_S(w) \setminus \Gamma_S(w)A^+ \quad \text{and} \quad \mathcal{R}'_S(w) = \Gamma'_S(w) \setminus A^+\Gamma'_S(w)$$

be respectively the set of (*first*) *right return words* and the set of (*first*) *left return words* to $w$. Thus, a right return word to $w$ in $S$ is a word $u$ such that $wu$ is a word of $S$ which ends with $w$ and has no internal factor equal to $w$.

Note that $\Gamma_S(w) \cup \{\varepsilon\} = \mathcal{R}_S(w)^* \cap w^{-1}S$. By definition, the set $\mathcal{R}_S(w)$ s a prefix code for every $w \in S$. If $S$ is recurrent, it is a $w^{-1}S$-maximal prefix code.

Note that $w\mathcal{R}_S(w) = \mathcal{R}'_S(w)w$.

**Example 1.4.1** Let $S$ be a set of words whose factors of length at most 6 are the labels of the paths starting at the root of the tree represented in Figure 6.5 (we will see in Example 6.1.13 an infinite set of words having such factors).

We have

$$\begin{array}{rcl} \mathcal{R}_S(a) & = & \{cbba, ccba, ccbba\}, \\ \mathcal{R}_S(b) & = & \{acb, accb, b\}, \\ \mathcal{R}_S(c) & = & \{bac, bbac, c\}. \end{array}$$

We colored in Figure 6.5 the words of $\alpha\mathcal{R}_S(\alpha)$ for $\alpha \in A$.

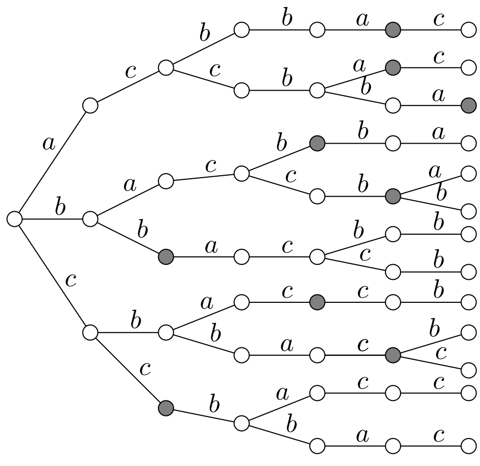

Figure 1.6: The words of length $\leq 6$ of the set $S$.

**Proposition 1.4.2** *A recurrent set $S$ is uniformly recurrent if and only if the set $\mathcal{R}_S(w)$ is finite for all $w \in S$.*

*Proof.* Assume that all sets $\mathcal{R}_S(w)$ for $w \in S$ are finite. Let $n \geq 1$. Let $N$ be the maximal length of the words in $\mathcal{R}_S(w)$ for a word $w$ of length $n$. Then, any word of length $N + n$ contains an occurrence of $w$. Indeed, assume that $u$ is a word of length $N + n$ without factor equal to $w$. Let $r$ be the word of minimal length such that $ru$ beghins with $w$ and set $ru = ws$. Then $|s| \geq N$ although $s$ is a proper prefix of a word in $\mathcal{R}_s(w)$, a contradiction.

Conversely, for $w \in S$, let $N$ be such that $w$ is a factor of any word in $S$ of length $N$. Then the words of $\mathcal{R}_S(w)$ have length at most $N$. ∎

For neutral sets we can give a more precise result. The following result has been proved in [5], generalizing a property proved for Sturmian words in [45] and for interval exchange in [67].

**Theorem 1.4.3** *Let $S$ be a uniformly recurrent neutral set containing the alphabet $A$. Then for every $w \in S$,the set $\mathcal{R}_S(w)$ has* Card$(A)$ *elements.*

One can actually prove more generally, for a uniformly recurrent set $S$, that if $S$ is strong (resp. weak, resp. neutral), then for every $w \in S$, the set $\mathcal{R}_S(w)$ has at least (resp. at most, resp. exactly) Card$(A)$ elements.

The following example shows that in a set of complexity $kn + 1$ the number of first right return words need not be equal to $k + 1$.

**Example 1.4.4** Let $S$ be the Chacon set (see Example 1.1.3). We have $\mathcal{R}_S(a) = \{a, bca, bcbca\}$ but $\mathcal{R}_S(ab) = \{caab, cbcab\}$.

Let $X \subset A^+$ be a set of words. A *complete return word* to $X$ is a word of $S$ which has a proper prefix in $X$, a proper suffix in $X$ and 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 $w \in S$, we denote by $\mathcal{CR}_S(w)$ instead of $\mathcal{CR}_S(\{w\})$. Thus $\mathcal{CR}_S(x)$ is the usual notion of a complete return word (see [36] for example).

**Example 1.4.5** Let $n \geq 1$ and let $X = S \cap A^n$. Then $\mathcal{CR}_S(X) = S \cap A^{n+1}$.

Since $\mathcal{CR}_S(w) = w\mathcal{R}_S(W)$, the sets $\mathcal{CR}_S(w)$ and $\mathcal{R}_S(w)$ have the same number of elements.

### 1.4.1 Derived sets

Let $S$ be a recurrent set and let $w \in S$. Let us consider a coding morphism for the set $\mathcal{R}_S(w)$, that is a morphism $f : B^* \to A^*$ which maps bijectively the (possibly infinite) alphabet $B$ onto $\mathcal{R}_S(w)$. The set $f^{-1}(w^{-1}S)$, denoted $\mathcal{D}_f(S)$, is called the *derived set* of $S$ with respect to $f$.

We can also define it as $\mathcal{D}_f(S) = f'^{-1}\left(Sw^{-1}\right)$, where $f' : B^* \to A^*$ is the morphism defined for $b \in B$ by $f'(b)w = wf(b)$. Note that $f'$ is a coding morphism for $\mathcal{R}'_S(w)$ (in [14] it is called the morphism *associated* with $f$).

The following result gives an equivalent definition of the derived set.

**Proposition 1.4.6** *Let $S$ be a recurrent set. For $w \in S$, let $f$ be a coding morphism for the set $\mathcal{R}_S(w)$. Then*

$$\mathcal{D}_f(S) = f^{-1}(\Gamma_S(w)) \cup \{\varepsilon\}.$$

*Proof.* Let $z \in \mathcal{D}_f(S)$. Then $f(z) \in w^{-1}S \cap \mathcal{R}_S(w)^*$ and thus $f(z) \in \Gamma_S(w) \cup \{\varepsilon\}$. Conversely, if $u \in \Gamma_S(w)$, then $u \in \mathcal{R}_S(w)^*$. Thus $u = f(z)$ for some $z \in \mathcal{D}_f(S)$, whence the result. ■

An immediate result of Proposition 1.4.6 is the following.

**Corollary 1.4.7** *If $S$ is recurrent then $\mathcal{D}_f(S)$ is recurrent.*

*Proof.* Consider two nonempty words $u, v \in \mathcal{D}_f(S)$. By Proposition 1.4.6, we have $f(u), f(v) \in \Gamma_S(w)$. Since $S$ is recurrent, there is a word $t$ such that $wf(u)twf(v) \in S$. Then $tw \in \Gamma_S(w)$ and thus $uf^{-1}(tw)v \in \mathcal{D}_f(S)$ by Proposition 1.4.6 again. This shows that $\mathcal{D}_f(S)$ is recurrent. ■

Let $S$ be a recurrent set and $x$ be an infinite word such that $S = \mathrm{Fac}(x)$. Let $w \in S$ and let $f$ be a coding morphism for the set $\mathcal{R}_S(w)$. Since $w$ appears infinitely often in $x$, there is a unique factorization $x = vwy$ with $y \in \mathcal{R}_S(w)^\omega$ and $v$ such that $vw$ has no proper prefix ending with $w$. The infinite word $f^{-1}(y)$ is called the *derived word* of $x$ relative to $f$, denoted $\mathcal{D}_f(x)$.

Since the set of factors of a recurrent infinite word is recurrent, the following Proposition, that results easily from Proposition 1.4.6 and Corollary 1.4.7, shows that the derived set of a recurrent set is recurrent.

**Proposition 1.4.8** *Let $S$ be a recurrent set and let $x$ be an infinite word such that $S = Fac(x)$. Let $w \in S$ and let $f$ be a coding morphism for the set $\mathcal{R}_S(w)$. The derived set of $S$ with respect to $f$ is the set of factors of the derived word of $x$ with respect to $f$, that is $\mathcal{D}_f(S) = Fac(\mathcal{D}_f(x))$.*

**Example 1.4.9** Let $F$ be a recurrent set having as factors of length at most 6 the set represented in Figure 6.5 (we will see such a set in Example 6.1.13).

Let $f$ be the coding morphism for the set $\mathcal{R}_S(c)$ given by $f(a) = bac$, $f(b) = bbac$, $f(c) = c$. The derived set of $S$ with respect to $f$ is represented in Figure 1.7.

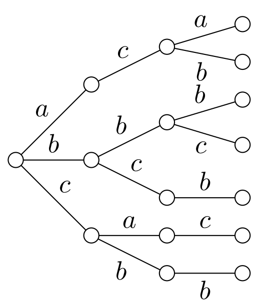


Figure 1.7: The words of length $\leq 3$ of $\mathcal{D}_f(F)$.

# Chapter 2

# Neutral sets

In this chapter we study an important family of sets: neutral sets.

Generalizing to sets the notions of weakness, strongness and neutrality seen for words in Chapter 1, we find hypotheses weak enough to contain all families of sets studied later in this manuscript, but strong enough to allow us to give several non-trivial results.

In Section 2.1, we define neutral sets and we show that such sets have linear factor complexity (Proposition 2.1.3). Moreover, we define the probability distribution of a factorial set and we give some results concerning bifix codes in neutral sets (Propositions 2.1.8, and 2.1.12).

In Section 2.2 we prove some cardinality results. In particular, in Section 2.2.1 we prove the Cardinality Theorem for neutral sets (Theorem 2.2.1), stating that all $S$-maximal bifix code of the same $S$-degree have the same cardinality, and a converse of this result (Theorem 2.2.4). Return words in neutral sets are covered in Section 2.2.2. We prove some cardinality results for the sets of return words (Theorem 2.2.8 and Corollary 2.2.10). In the same Section we also prove that recurrence and uniformly recurrence coincide in the case of neutral sets (Corollary 2.2.9).

Finally, in Section 2.3, we prove a closure property of the family of (recurrent) neutral sets under maximal bifix decoding (Theorem 2.3.1).

## 2.1 Strong, weak and neutral sets

Recall from Chapter 1 that the multiplicity of a word $w$ in a factorial set $S$ is the quantity

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

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

We say that a set $S$ is *neutral* if it is factorial and every nonempty word $w \in S$ is neutral. A factorial set is said to be *weak* (resp. *strong*) if every word in it, including $\varepsilon$, is weak or neutral (resp. strong or neutral). Note that a neutral set could be a weak set, a strong set or both. This last case is true when the empty word is also neutral.

The *characteristic* of a neutral set $S$ 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.

**Example 2.1.1** The Fibonacci set defined in Example 1.1.2 is a neutral set of characteristic 1. Indeed one can prove that every word, including the empty word, is neutral.

The following example of a neutral set of characteristic larger than 1 is due to Julien Cassaigne. We will study this example more carefully in Chapter 3

**Example 2.1.2** Let $A = \{a, b, c, d\}$ and let $\sigma$ be the morphism from $A^*$ into itself defined by $\sigma : a \mapsto ab,\ b \mapsto cda,\ c \mapsto cd,\ d \mapsto abc$. Let $S$ be the set of factors of the infinite word $x = \sigma^\omega(a)$. One has $S \cap A^2 = \{ab, ac, bc, ca, cd, da\}$ and thus $m(\varepsilon) = -1$. We will see in Example 3.1.5 that every nonempty word is neutral (actually the set satisfies a stronger property). Thus $S$ is neutral of characteristic 2.

One deduces easily from Proposition 1.1.6 the following result which shows that a neutral set has linear complexity.

**Proposition 2.1.3** *The factor complexity of a factorial set $S$ is given by $p_0 = 1$ and for every $n \geq 1$ satisfies :*

*(i)* $p_n = n(\mathrm{Card}(A) - \chi(S)) + \chi(S)$ *if $S$ is neutral;*

*(ii)* $p_n \leq n(\mathrm{Card}(A) - \chi(S)) + \chi(S)$ *if $S$ is weak;*

*(iii)* $p_n \geq n(\mathrm{Card}(A) - \chi(S)) + \chi(S)$ *if $S$ is strong.*

*Proof.* Since $S$ contains the empty word and the alphabet, we have $p_0 = 1$ and $p_1 = k$. Thus $s_0 = k - 1$.

By Proposition 1.1.6 one has $t_0 = m(\varepsilon) = 1 - \chi(S)$ and $t_n = 0$ for every $n > 0$. Thus $s_n = k - \chi(S)$ for every $n > 0$. The conclusion immediately follows by induction on $n$.

The inequalities for weak and strong sets are proved in the same way. ■

We now give an example of a set of complexity $2n + 1$ on an alphabet with three letters which is not neutral.

**Example 2.1.4** Let $S$ be the Chacon set (see Example 1.1.3). The set $S$ is of complexity $2n + 1$ (see for example [41, Section 5.5.2]).

It contains strong, neutral and weak words. Indeed, $S \cap A^2 = \{aa, ab, bc, ca, cb\}$ and thus $m(\varepsilon) = 0$ showing that the empty word is neutral. Next $m(abc) = 1$ and $m(bca) = -1$, showing that $abc$ is strong while $bca$ is weak.

### 2.1.1 Probability distributions

Let $S$ be a factorial set. A *left probability distribution* on $S$ is a map $\rho : S \to [0,1]$ such that

(i) $\rho(\varepsilon) = 1$,

(ii) $\sum_{a \in L_S(w)} \rho(aw) = \rho(w)$, for any $w \in S$.

For a left probability distribution $\rho$ on $S$ and a set $X \subset S$, we denote $\rho(X) = \sum_{x \in S} \rho(x)$.

Symmetrically, a *right probability distribution* on $S$ is a map $\sigma : S \to [0,1]$ satisfying condition (i) above and

(iii) $\sum_{a \in R_S(w)} \sigma(aw) = \sigma(w)$, for any $w \in S$.

See [8, Chapters 1.11 and 13] for elementary properties of probability distributions and their connections with probability measures. Note in particular that for any $w \in S$ and $n \geq 0$ one has, as a consequence of (ii) and (iii),

$$\rho\left(A^n w \cap S\right) = \rho(w) \quad \text{and} \quad \sigma\left(w A^n \cap S\right) = \sigma(w).$$

For $w \in S$, we define

$$\rho_S(w) = b_S(w) - \ell_S(w), \quad \lambda_S(w) = b_S(w) - r_S(w).$$

Thus, when $w$ is neutral, $\rho_S(w) = r_S(w) - 1$ and $\lambda_S(w) = \ell_S(w) - 1$. The following result shows that in a biextendable neutral set, $\rho_S$ is a left probability distribution on $S$ (and $\lambda_S$ is a right probability), except for the value on $\varepsilon$ which is $\rho(\varepsilon) = b(\varepsilon) - \ell(\varepsilon) = m(\varepsilon) + r(\varepsilon) - 1 = \mathrm{Card}(A) - \chi(S)$ and can be different from 1.

**Proposition 2.1.5** *Let $S$ be a biextendable set. Then, for any $w \in S$, one has $\lambda_S(w), \rho_S(w) \geq 0$ and*

$$\sum_{a \in L_S(w)} \rho_S(aw) = \rho_S(w) + \sum_{a \in L_S(w)} m_S(aw),$$

*and*

$$\sum_{a \in R_S(w)} \lambda_S(wa) = \lambda_S(w) + \sum_{a \in R_S(w)} m_S(wa).$$

*Proof.* Since $S$ is biextendable, we have $\ell_S(w), r_S(w) \leq b_S(w)$. Thus, $\lambda_S(w), \rho_S(w) \geq 0$. Next, since by Equation 1.1 one has $b(w) - \ell(w) = m(w) + r(w) - 1$, one has

$$\begin{aligned}
\sum_{a \in L_S(w)} \rho(aw) &= \sum_{a \in L_S(w)} \big(m(aw) + r(aw) - 1\big) \\
&= \sum_{a \in L_S(w)} m(aw) + \big(b(w) - \ell(w)\big) \\
&= \sum_{a \in L_S(w)} m(aw) + \rho_S(w).
\end{aligned}$$

The proof for $\lambda_S$ is symmetric. ∎

An immediate corollary of Proposition 2.1.5 is the following.

**Corollary 2.1.6** *Let $S$ be a biextendable neutral set. Then for any $w \in S$, one has*

$$\sum_{a \in L_S(w)} \rho_S(aw) = \rho_S(w), \qquad \sum_{a \in R_S(w)} \lambda_S(wa) = \lambda_S(w).$$

If in a neutral set $S$ we have $\rho_S(\varepsilon) = 0$, then $\rho_S(x) = 0$ for all $x \in S$. Otherwise, $\rho'_S(x) = \rho_S(x)/\rho_S(\varepsilon)$ is a left probability distribution. A symmetric result holds for $\lambda_S$.

Given a set $X$, we denote by $m_S(X) = \sum_{x \in X} m(x)$. We now prove the following result. It accounts for the fact that, in an Arnoux-Rauzy set $S$, any finite $S$-maximal suffix code contains exactly one right-special word [7, Proposition 5.1.5].

**Lemma 2.1.7** *Let $S$ be a biextendable set, let $X$ be a finite $S$-maximal suffix code and let $Q$ be the set of nonempty suffixes of $X$. Then $\rho_S(X) = m(Q) + \rho_S(\varepsilon)$.*

*Proof.* The theorem is trivially true for $\mathrm{Card}(A) = 1$, so let us suppose that $\mathrm{Card}(A) \geq 2$. We show by induction on $\mathrm{Card}(X)$ that for any word $w$, we have $\rho_S(Xw \cap S) = m_S(Qw \cap S) + \rho_S(w)$. The statement follows for $w = \varepsilon$.

For $X = A$, the statement follows from Proposition 2.1.5. We may assume that that the words of $X$ do not all end with the same letter. For every $a \in A$, the set $X_a = Xa^{-1}$ is an $Sa^{-1}$-maximal suffix code. Moreover, $\mathrm{Card}(X_a) \leq \mathrm{Card}(X)$. Let $Q_a$ be the set of its nonempty suffixes. Clearly $Q = \bigcup_{a \in A} Q_a a$. Then, using the induction hypothesis for each $X_a$ and Proposition 2.1.5, we have

$$\begin{aligned}
\rho_S(Xw \cap S) &= \sum_{a \in L_S(w)} \rho_S(X_a aw \cap S) = \\
&= \sum_{a \in L_S(w)} \big(m_S(Q_a aw \cap S) + \rho_S(aw)\big) \\
&= \sum_{a \in L_S(w)} m_S(Q_a aw \cap S) + \sum_{a \in L_S(w)} m_S(aw) + \rho_S(w) \\
&= m_S(Qw \cap S) + \rho_S(w).
\end{aligned}$$

∎

Using Lemma 2.1.7 we can easily prove the following result.

**Proposition 2.1.8** *Let $S$ be a neutral set and let $X$ be a finite $S$-maximal suffix code. Then $\rho_S(X) = \mathrm{Card}(A) - \chi(S)$.*

*Proof.* The formula easily follows from Lemma 2.1.7 and the fact that in a neutral set $S$, every nonempty word $w$ satisfies $m_S(w) = 0$. ∎

**Example 2.1.9** Let $S$ be the neutral set of characteristic 2 of Example 2.1.2. The set $X = \{a, ac, b, bc, d\}$ is an $S$-maximal suffix code (its reversal is the $\tilde{S}$-maximal prefix code $\tilde{X} = \{a, b, ca, cb, d\}$). The values of $\rho_S$ on $X$ are represented in Figure 2.1 on the left. One has $\rho_S(X) = 2$, in agreement with Proposition 2.1.8.

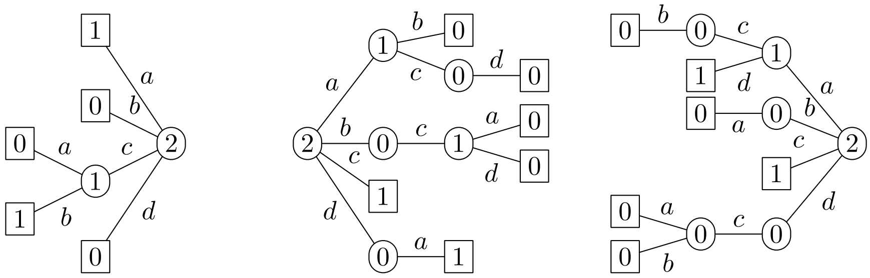

Figure 2.1: An $S$-maximal suffix code (left) and an $S$-maximal bifix code represented as a prefix code (center) and as a suffix code (right).

**Example 2.1.10** Let $S$ be the neutral set of characteristic 2 of Example 2.1.2. The set $X = \{ab, acd, bca, bcd, c, da\}$ is an $S$-maximal bifix code of $S$-degree 2 (see Figure 2.1 on the center and the right).

Note that the set on nonempty proper prefixes of $X$ is exactly the $S$-maximal suffix code of the previous example and represented in Figure 2.1 on the left.

The following statement is closely related with a similar statement concerning the average length of a bifix code (see [7, Corollary 4.3.8]).

**Lemma 2.1.11** *Let $S$ be a recurrent biextendable set and let $X$ be a finite $S$-maximal bifix code $S$ of $S$-degree $d$. Let $F = Fac(X)$. The set $P$ of proper prefixes of $X$ satisfies $\rho_S(P) = d\rho_S(\varepsilon) + \alpha(F)$, where $\alpha(w) = \delta(w)m_S(w)$ with $\delta(\varepsilon) = 0$, $\delta(w) \geq 0$ for $w \neq \varepsilon$ and $\delta(w) \geq 1$ for $w \in P \setminus \{\varepsilon\}$.*

*Proof.* By Theorem 1.2.9, we have $P \setminus \{\varepsilon\} = \cup_{i=1}^{n-1} Y_i$, where the $Y_i$ are $S$-maximal suffix codes. By Lemma 2.1.7, we have $\rho_S(Y_i) = m(Q_i) + \rho_S(\varepsilon)$, where $Q_i$ is the set of nonempty suffixes of $Y_i$. Thus $\rho_S(P) = d\rho_S(\varepsilon) + \alpha(F)$, where $\alpha(w) = \delta(w)m_S(w)$ with $\delta(w)$ the number of $i$ such that $w$ is a nonemtpy suffix of $Y_i$. ∎

We will use this consequence of Lemma 2.1.11 in the next section.

**Proposition 2.1.12** *Let $S$ be a recurrent set and let $X$ be a finite $S$-maximal bifix code of $S$-degree $d$. The set $P$ of proper prefixes of $X$ satisfies*

1. $\rho_S(P) = d(\mathrm{Card}(A) - \chi(S))$ *if* $S$ *is neutral,*

2. $\rho_S(P) \leq d(\mathrm{Card}(A) - \chi(S))$ *if* $S$ *is weak,*

3. $\rho_S(P) \geq d(\mathrm{Card}(A) - \chi(S))$ *if* $S$ *is strong,*

*Proof.* Let $F$ and $\alpha$ be as in Lemma 2.1.11. Since $\delta(w) \geq 0$ for every word $w$, the sign of of $\alpha(F)$ only depends on the values of $m_S(w)$.

In a neutral set the only word having non zero multiplicity is $\varepsilon$, thus, by Lemma 2.1.11, $\rho_S(P) = d\rho_S(\varepsilon) = d\big(\mathrm{Card}(A) - \chi(S)\big)$. The two other cases are proved in a similar way. ∎

## 2.2 Cardinality Theorems

### 2.2.1 Bifix codes

In the following we prove a result referred to as the Cardinality Theorem. This is a generalization of a result proved in [7] in the less general case of an Arnoux-Rauzy set. Since $S \cap A^n$ is an $S$-maximal bifix code of $S$-degree $n$ (see Example 1.2.2)

**Theorem 2.2.1 (Cardinality Theorem)** *Let* $S$ *be a recurrent set containing the alphabet* $A$ *and let* $X \subset S$ *be a finite* $S$*-maximal bifix code.*

1. *If* $S$ *is neutral, then* $\mathrm{Card}(X) = d_S(X)\,(\mathrm{Card}(A) - \chi(S)) + \chi(S)$.

2. *If* $S$ *is weak, then* $\mathrm{Card}(X) \leq d_S(X)\,(\mathrm{Card}(A) - \chi(S)) + \chi(S)$.

3. *If* $S$ *is strong, then* $\mathrm{Card}(X) \geq d_S(X)\,(\mathrm{Card}(A) - \chi(S)) + \chi(S)$.

Note that, for a recurrent neutral set $S$, a bifix code $X \subset S$ may be infinite since this may happen for an Arnoux-Rauzy set $S$ (see [7, Example 5.1.4]).

*Proof of Theorem 2.2.1.* Since $X$ is a finite $S$-maximal bifix code, it is an $S$-maximal prefix code (see Section 1.2). By a well-known property of trees, this implies that $\mathrm{Card}(X) = 1 + \sum_{p \in P}\big(r(p) - 1\big)$, where $P$ is the set of proper prefixes of $X$. Since $r(p) - 1 = \rho(p) - m(p)$, we have

$$\begin{aligned}
\mathrm{Card}(X) &= 1 + \sum_{p \in P}\big(r(p) - 1\big) = 1 + \sum_{p \in P}\big(\rho(p) - m(p)\big) \\
&= 1 + \rho(P) - m(P) \\
&= 1 + \big(d\rho(\varepsilon) + \alpha(F)\big) - m(P) \\
&= d\rho(\varepsilon) + \big(1 - m(\varepsilon)\big) + \big(\alpha(F) - m\,(P \setminus \{\varepsilon\})\,\big),
\end{aligned}$$

where the fourth equality and the definition of $F$ and $\alpha$ come from Lemma 2.1.11, while $d = d_S(X)$. From what we have seen in Section 2.1.1, we have $\rho(\varepsilon) =$

$\mathrm{Card}(A) - \chi(S)$, and by definition of the characteristic one has $1 - m(\varepsilon) = \chi(S)$. Thus,

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

with $\alpha = \alpha(F) - m\,(P \setminus \{\varepsilon\})$. The quantity $\alpha$ is a nonnegative linear combination of multiplicities of nonempty words, thus it is nonnegative when $S$ is strong, nonpositive when $S$ is weak and zero when $S$ is neutral, whence the result. ■

Note that we recover, as a particular case of Theorem 2.2.1 applied to the set $X$ of words of length $n$ in $S$, the fact that for a set $S$ satisfying the hypotheses of the theorem, the factor complexity is $p_0 = 1$ and $p_n = n(\mathrm{Card}(A) - \chi(S)) + \chi(S)$ (see Proposition 2.1.3).

**Example 2.2.2** Let $S$ be the neutral set of Example 2.1.2 and let $X$ be the $S$-maximal bifix code of Example 2.1.10. We have $\mathrm{Card}(X) = 2(4-2)+2 = 6$ according to Theorem 2.2.1.

The following example illustrates the necessity of the hypotheses in Theorem 2.2.1.

**Example 2.2.3** Consider again the Chacon set $S$ of Example 1.1.3. Let $X = S \cap A^4$ and let $Y, Z$ be the $S$-maximal bifix codes of $S$-degree 4 represented in Figure 2.2. The first one is obtained from $X$ by internal transformation with respect to $abc$, the second one with respect to $bca$ (for the definition of an internal transformation recall Section 1.2).

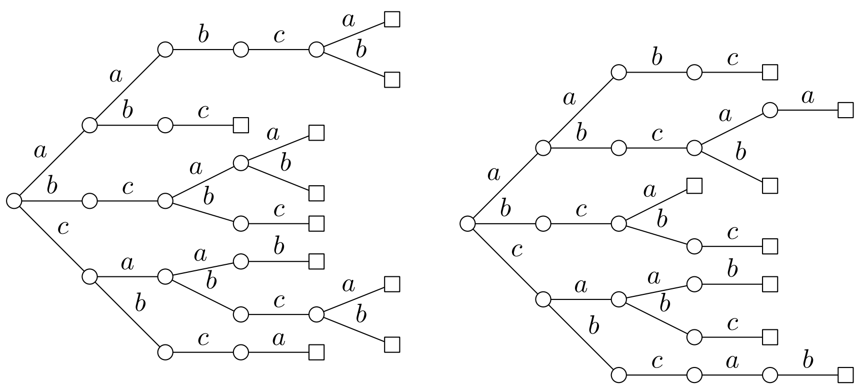

Figure 2.2: Two $S$-maximal bifix codes of $S$-degree 4.

We have $\mathrm{Card}(Y) = 10$ and $\mathrm{Card}(Z) = 8$ showing that $\mathrm{Card}(Y) - 1 > 8$ and $\mathrm{Card}(Z) - 1 < 8$, illustrating the fact that $S$ is neither strong nor weak.

The following statement is a converse of Theorem 2.2.1 for uniformly recurrent sets.

**Theorem 2.2.4** *Let $S$ be a uniformly recurrent set containing the alphabet $A$. If every finite $S$-maximal bifix code of $S$-degree $d$ has $d(Card(A)-c)+c$ elements, then $S$ is neutral of characteristic $c$.*

To prove Theorem 2.2.4, we use the following result, using internal transformations (recall Section 1.2).

**Proposition 2.2.5** *Let $S$ be a uniformly 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.*

*Proof.* We argue by contradiction. Let $w \in S$ be a word of length $n \geq d_0 - 1$ which is not neutral.

Set $X = S \cap A^{n+1}$. The set $X$ is an $S$-maximal bifix code of $S$-degree $n+1$. Let $Y$ be the code obtained by internal transformation from $X$ with respect to $w$ and defined by Equation (1.6). Note that $G = L(w)$ and $D = R(w)$. Recall that, by Proposition 1.2.14, the $S$-degree of $Y$ is at most $n+1$.

We distinguish two cases.

**Case 1.** Assume that $Gw \cap wD = \emptyset$.

The code $Y$ is defined by Equation (1.7) and we have $\mathrm{Card}(GwD \cap S) = b(w)$. Moreover, we have $D_0 = G_0 = \emptyset$. By [7, Proposition 4.4.5]) $Y$ has the same $S$-degree as $X$, that is $n+1$. This implies $\mathrm{Card}(X) = \mathrm{Card}(Y)$. On the other hand

$$\mathrm{Card}(Y) = \mathrm{Card}(X) + 1 + b(w) - \ell(w) - r(w) = \mathrm{Card}(X) + m(w).$$

Since $w$ is not neutral, we have $m(w) \neq 0$ and thus we obtain a contradiction.

**Case 2.** Assume next that $Gw \cap wD \neq \emptyset$. Then $w = a^n$ with $n > 0$ for some letter $a$ and the sets $G_0, D_0$ defined by Equation 1.4 are $G_0 = D_0 = \{a\}$. Moreover $a^{n+1} \in X$.

Since $w$ is not neutral, it is bispecial. Thus the sets $G_1, D_1$ are nonempty and the hypotheses of Proposition 1.2.14 are satisfied. Since $S$ is uniformly recurrent and since $S \neq a^*$, the set $a^* \cap S$ is finite. Set $a^* \cap S = \{1, a, \ldots, a^m\}$. Thus $m \geq n+1$. Then, $\delta_Y(a^m) = n$ since $a^m$ has $n$ suffixes which are proper prefixes of $Y$.

Let $b \in R(a^m)$. By construction, $b \neq a$. The word $a^m b$ has no suffix in $Y$. Indeed, if $a^t b \in Y$, we cannot have $t \geq n$ since $a^n \in Y$ and $Y$ is a bifix code by Proposition 1.2.14. Moreover, since all words in $Y$ except $a^n$ have length greater than $n$, $t < n$ is also impossible. Thus by Equation (1.2), we have $\delta_Y(a^m b) = \delta_Y(a^m) + 1$ and thus $\delta_Y(a^m b) = n+1$. This shows that the $S$-degree of $Y$ is $n+1$ and thus that $\mathrm{Card}(Y) = \mathrm{Card}(X)$ as in Case 1.

If $a^m$ is not neutral, Case 1 applies to $X = S \cap A^{m+1}$ and $w = a^m$. Otherwise, we can assume, without loss of generality, that $n$ is chosen maximal such that $a^n$ is not neutral.

For $n \le i \le m-2$ (there may be no such integer $i$ if $n = m-1$), since $a^{i+1}$ is neutral, we have

$$\mathrm{Card}(G_1 a^i D_1 \cap S) = b(a^i) - \ell(a^{i+1}) - r(a^{i+1}) + 1 = b(a^i) - b(a^{i+1}).$$

Moreover, $\mathrm{Card}(G_1 a^{m-1} D_1 \cap S) = b(a^{m-1}) - \ell(a^m) - r(a^m) = b(a^{m-1}) - b(a^m) - 1$ and $\mathrm{Card}(G_1 a^m D_1 \cap S) = b(a^m)$.

Thus

$$\begin{aligned}
\mathrm{Card}(G_1 a^n a^* D_1 \cap S) &= \sum_{i=n}^{m-2} \big(b(a^i) - b(a^{i+1})\big) + b(a^{m-1}) - b(a^m) - 1 + b(a^m) \\
&= b(a^n) - 1.
\end{aligned}$$

Thus $\mathrm{Card}(Y) - \mathrm{Card}(X)$ evaluates as

$$\begin{aligned}
& 1 + \mathrm{Card}(G_1 a^n a^* D_1 \cap S) - \mathrm{Card}(G a^n) - \mathrm{Card}(a^n D) + 1 \\
= \;& 1 + b(a^n) - 1 - \ell(a^n) - r(a^n) + 1 \\
= \;& m(a^n),
\end{aligned}$$

where the last $+1$ on the first line comes from the word $a^{n+1}$ counted twice in $\mathrm{Card}(G a^{n+1}) + \mathrm{Card}(a^{n+1} D)$. Since $m(a^n) \neq 0$, this contradicts the fact that $X$ and $Y$ have the same number of elements. ■

We can now prove Theorem 2.2.4.

*Proof of Theorem 2.2.4.* We first apply the statement to the $S$-maximal bifix code $X = S \cap A^2$ which has $S$-degree 2. Since $\mathrm{Card}(X) = 2(\mathrm{Card}(A) - c) + c = 2\,\mathrm{Card}(A) - c$, we conclude that $m_S(\varepsilon) = 1 - c$. On the other hand, applying Proposition 2.2.5 with $d_0 = 2$, we conclude that every nonempty word is neutral. Thus $S$ is neutral of characteristic $c$. ■

We also note that Theorem 2.2.1 can be formulated in an equivalent way using the notion of derived code of a maximal bifix code.

**Theorem 2.2.6** *Let $S$ be a recurrent neutral set, let $X$ be a finite $S$-maximal bifix code of $S$-degree $d \ge 2$ and let $X'$ be the derived code of $X$. One has*

$$\mathrm{Card}(X) = \mathrm{Card}(X') + \mathrm{Card}(A) - \chi(S). \tag{2.1}$$

*Proof.* Since $X'$ has degree $d_S(X) - 1$, by Theorem 2.2.1, we have $\mathrm{Card}(X) - \mathrm{Card}(X') = \mathrm{Card}(A) - \chi(S)$. ■

Conversely, we may prove Theorem 2.2.1 by induction on $n$, assuming Theorem 2.2.6. We just prove the case of a neutral set.

Theorem 2.2.1 holds for $n = 1$ since in this case $X = A$. Next, assume that it holds for $d - 1$. Then, by Equation (2.1), we have

$$\begin{aligned}
\mathrm{Card}(X) &= \mathrm{Card}(X') + \mathrm{Card}(A) - \chi(S) \\
&= (d-1)(\mathrm{Card}(A) - \chi(S)) + \chi(S) + \mathrm{Card}(A) - \chi(S) \\
&= d(\mathrm{Card}(A) - \chi(S)) + \chi(S).
\end{aligned}$$

**Example 2.2.7** Let $S$ be the neutral set of Example 2.1.2 and let $X$ be the $S$-maximal bifix code of Example 2.1.10. We have $X' = A$ and accordingly $\mathrm{Card}(X) = \mathrm{Card}(A) + \mathrm{Card}(A) - 2 = 6$.

### 2.2.2 Return words

Let $S$ be a factorial set of words. Recall from Section 1.4 that a complete return word to a set $X \subset S$ is a word of $S$ which has a proper prefix in $X$, a proper suffix in $X$ and no internal factor in $X$. Recall also that the set of complete return words to $X$, denoted by $\mathcal{CR}(X)$, is a bifix code and that, if $S$ is uniformly recurrent, it is finite for any finite set $X$.

**Theorem 2.2.8** *Let $S$ be a neutral set. For any finite nonempty bifix code $X \subset S$ with empty kernel, we have*

$$\mathrm{Card}(\mathcal{CR}_S(X)) \leq \mathrm{Card}(X) + \mathrm{Card}(A) - \chi(S) \tag{2.2}$$

*with equality if $S$ is recurrent.*

*Proof.* Let $P$ be the set of proper prefixes of $\mathcal{CR}_S(X)$. For $q \in P$, we define $\alpha(q) = \mathrm{Card}\{a \in A \mid qa \in P \cup \mathcal{CR}_S(X)\} - 1$. For $P' \subset P$, we set $\alpha(P') = \sum_{p \in P'} \alpha(p)$.

Since $\mathcal{CR}_S(X)$ is a finite prefix code, we have, by a well-known property of trees, $\mathrm{Card}(\mathcal{CR}_S(X)) \leq 1 + \alpha(P)$ with equality if $\mathcal{CR}_S(X)$ is nonempty (that is, if $S$ is recurrent).

Let $P'$ be the set of words in $P$ which are proper prefixes of $X$ and let $Y = P \setminus P'$. Since $P'$ is the set of proper prefixes of $X$, we have $\alpha(P') = \mathrm{Card}(X) - 1$.

Since $P \cup \mathcal{CR}_S(X) \subset S$, one has $\alpha(q) \leq \rho_S(q)$ for any $q \in P$. Moreover, if $S$ is recurrent, and since $X$ has empty kernel, any word of $S$ with a prefix in $X$ is comparable for the prefix order with a word of $\mathcal{CR}_S(X)$. This implies that for any $q \in Y$ and any $b \in R_S(q)$, one has $qb \in P \cup \mathcal{CR}_S(X)$. Consequently, we have $\alpha(q) = \rho_S(q)$ for any $q \in Y$. Thus we have shown that

$$\mathrm{Card}(\mathcal{CR}_S(X)) \leq 1 + \alpha(P') + \rho(Y) \leq \mathrm{Card}(X) + \rho(Y)$$

with equality if $S$ is recurrent. Let us show that $Y$ is a suffix code which is $S$-maximal if $S$ is recurrent. This will imply our conclusion by Proposition 2.1.8. Suppose that $q, uq \in Y$ with $u$ nonempty. Since $q$ is in $Y$, it has a proper prefix in $X$. But this implies that $uq$ has an internal factor in $X$, a contradiction. Thus $Y$ is a suffix code. Assume next that $S$ is recurrent. Consider $w \in S$. Then, for any $x \in X$, there is some $u \in S$ such that $xuw \in S$. Let $y$ be the shortest suffix of $xuw$ which has a proper prefix in $X$. Then $y \in Y$. This shows that $Y$ is an $S$-maximal suffix code. ■

Recall, from Section 1.4, that the sets $\mathcal{CR}_S(x)$ and $\mathcal{R}_S(x)$ have the same number of elements.

Since a recurrent set $S$ is uniformly recurrent if and only if the set of return words is finite (see Proposition 1.4.2), we have the following important consequence of Theorem 2.2.8.

**Corollary 2.2.9** *A recurrent neutral set is uniformly recurrent.*

*Proof.* By Theorem 2.2.8, the set $\mathcal{CR}_S(x)$ is finite for any $x \in X$. Thus, $S$ is uniformly recurrent. ∎

Another consequence of Theorem 2.2.8 is that the number of right return words to a word $x$ in a recurrent neutral set is always the same.

**Corollary 2.2.10** *Let $S$ be a recurrent neutral set. For any $x \in S$, the set $\mathcal{R}_S(x)$ has* $\mathrm{Card}(A) - \chi(S) + 1$ *elements.*

**Example 2.2.11** Consider again the neutral set $S$ of Example 2.1.2. We have $\mathrm{Card}\,(\mathcal{R}_S(a)) = \mathrm{Card}\,(\{bca, bcda, cad\}) = 4 - 2 + 1 = 3$, according to Corollary 2.2.10.

The following statement, which holds under fairly general hypotheses, shows an interesting connection between complete return words to a bifix code and the derived code (see Section 1.2). It explains the similarity between Formulae (2.1) and (2.2) (with equality).

**Proposition 2.2.12** *Let $S$ be a recurrent set. Let $X$ be a finite $S$ maximal bifix code, let $X'$ be the derived code of $X$ and let $K, K'$ be the kernels of $X$ and $X'$ respectively. Then*

$$\mathcal{CR}_S(X' \setminus K) = X \setminus K. \tag{2.3}$$

*Proof.* Let us first show the inclusion from right to left. Let $x \in X \setminus K$. Then $x$ has a proper prefix in $X' \setminus K$, namely the shortest prefix of $x$ which is not an internal factor of $X$ (see [7, Lemma 4.3.3]). Similarly, $x$ has a proper suffix which is in $X' \setminus K$. Moreover $x$ cannot have an internal factor in $X' \setminus K$. Indeed, by definition of $X'$, the words in $X' \setminus K$ are not internal factors of $X$. This shows that $x \in \mathcal{CR}_S(X' \setminus K)$.

Conversely, consider $x \in \mathcal{CR}_S(X' \setminus K)$. Let $P$ be the set of proper prefixes of $X$. Let $y$ (resp. $z$) be the proper prefix (resp. suffix) of $x$ which is in $X' \setminus K$. Since $x'$ is in $X'$, it is in $P$. We cannot have $x \in P$ since otherwise $z$ would be in $K$. Thus $x$ has a prefix $yu$ in $X$. By the first part of the proof, $yu$ has a suffix in $\mathcal{CR}_S(X' \setminus K$, and thus $x$ has an internal factor in $X' \setminus K$, a contradiction unless $x = yu$. Thus $x \in X$. ∎

If $S$ is assumed to be recurrent and neutral, Formulæ (2.1) and (2.2) (with equality) show that both sides of Equation (2.3) have the same cardinality. Thus the inclusion implies the equality.

**Example 2.2.13** Let $S$ and $X$ be as in Example 2.1.10. We have $K = \{c\}$ and

$$X \setminus K = \{ab, acd, bca, bcd, da\} = \mathcal{CR}_S(\{a, b, d\})$$

in agreement with Proposition 2.2.12.

## 2.3 Bifix decoding of neutral sets

Recall from Section 1.2 the definitions of coding morphism and maximal bifix decoding.

We prove show the following closure properties for the family of neutral sets.

**Theorem 2.3.1** *Any maximal bifix decoding of a recurrent neutral set is a neutral set with the same characteristic.*

In order to prove Theorem 2.3.1 we need some preliminary results. We also generalize the notation of left extensions, right extensions and biextensions of Section 1.2.1.

Let $S$ be a factorial set. For two sets of words $X, Y$ and a word $w \in S$, we set

$$\begin{aligned} L_S^X(w) &= \{x \in X \mid xw \in S\}, \\ R_S^Y(w) &= \{y \in A \mid wy \in S\}, \\ B_S^{X,Y}(w) &= \{(x,y) \in X \times Y \mid xwy \in S\} \end{aligned}$$

and furthermore

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

Finally, for a word $w$, we define

$$m_S^{X,Y}(w) = b_S^{X,Y}(w) - \ell_S^X(w) - r_S^Y(w) + 1.$$

Note that $B_S^{A,A}(w) = B_S(w)$, $m_S^{A,A}(w) = m_S(w)$, and so on.

**Proposition 2.3.2** *Let $S$ be a neutral set, let $X$ be a finite $S$-maximal suffix code and let $Y$ be a finite $S$-maximal prefix code. Then $m_S^{X,Y}(w) = m_S(w)$ for every $w \in S$.*

*Proof.* We use an induction on the sum of the lengths of the words in $X$ and in $Y$. If $X, Y$ contain only words of length 1, since $X$ (resp. $Y$) is an $S$-maximal suffix (resp. prefix) code, we have $X = Y = A$ and there is nothing to prove.

Assume next that one of them, say $Y$, contains words of length at least 2. Let $p$ be a nonempty proper prefix of $Y$ of maximal length. Set $Y' = (Y \setminus pA) \cup p$. If $wp \notin S$, then $m^{X,Y}(w) = m^{X,Y'}(w)$ and the conclusion follows by induction hypothesis. Thus we may assume that $wp \in S$. Then

$$m^{X,Y}(w) - m^{X,Y'}(w) = b^{X,A}(wp) - \ell^X(wp) - r^A(wp) + 1 = m^{X,A}(wp).$$

By induction hypothesis, we have $m^{X,Y'}(w) = m(w)$ and $m^{X,A}(wp) = 0$, whence the conclusion. ∎

We can now prove the main result of the section.

*Proof of Theorem 2.3.1.* Let $S$ be a recurrent neutral set and let $f : B^* \to A^*$ be a coding morphism for a finite $S$-maximal bifix code $X$. Set $U = f^{-1}(S)$. Let $v \in U \setminus \{\varepsilon\}$ and let $w = f(v)$. Then $m_U(v) = m_S^{X,X}(w)$. Since $S$ is recurrent, $X$ is an $S$-maximal suffix code and prefix code. Thus, by Proposition 2.3.2, $m_U(v) = m_S(w)$, which implies our conclusion. ∎

The following example shows that the maximal decoding of a recurrent neutral set need not be recurrent.

**Example 2.3.3** Let $S$ be the set of factors of the infinite word $(ab)^\omega$. $S$ is a recurrent neutral set of characteristic 2. The set $X = \{ab, ba\}$ is a bifix code of $S$-degree 2. Let $f : u \mapsto ab$, $v \mapsto ba$. The set $f^{-1}(S)$ is the set of factors of $u^\omega \cup v^\omega$ and it is not recurrent.

The following example shows that the class of sets of factor complexity $kn+c$ is not closed by maximal bifix decoding.

**Example 2.3.4** Let $S$ be the Chacon set and let $f : B^* \to A^*$ be a coding morphism for the $S$-maximal bifix code $Z$ of $S$-degree 4 with 8 elements of Example 2.2.3. One may verify that $\mathrm{Card}(B^2 \cap f^{-1}(S)) = \mathrm{Card}(Z^2 \cap S) = 17$. This shows that the set $f^{-1}(S)$ does not have factor complexity $7n + 1$.

# Chapter 3

# Tree sets

In this chapter we define acyclic, connected and tree sets. The last one is a particular family of neutral sets, large enough to contain well-studied families as, for example, Arnoux-Rauzy sets and interval exchange sets (that we will introduce in Chapter 6).

In Section 3.1 we give the basic definitions (acyclic sets, connected sets, trees, planar trees) as well as some examples. Moreover, we generalize the extension graphs defined in Section 1.2.1 and give conditions under which this generalized extension graphs are acyclic (Proposition 3.1.13). Using this generalization we can work with longer extensions to a given word, namely using elements of a maximal bifix code instead of letters.

Return words are the topic of Section 3.2. The main result of this section is the Return Theorem (Theorem 3.2.5), stating that the set of return words on a tree set of characteristic 1 is a basis of the free group. In Section 3.2.3 we show a closure property under derivation (Theorem 3.2.9).

In Section 3.3 we introduce a technique to construct tree sets of characteristic $c$ starting from a tree set of characteristic a divisor of $c$ (Theorem 3.3.1).

Finally, in Section 3.4 we show that a recurrent tree set of characteristic 1, closed under reversal, contains the maximal possible number of palindromic factors (Proposition 3.4.1).

## 3.1 The tree condition

Recall from Section 1.2.1 that, given a set $S$ of words and a word $w \in S$, the extension graph of $w$ is the undirected bipartite graph $\mathcal{E}_S(w)$ on the set of vertices which is the disjoint union of $L_S(w)$ and $R_S(w)$ with edges the pairs $(a, b) \in B_S(w)$. An edge $(a, b) \in B_S(w)$ goes from $a \in L_S(w)$ to $b \in R_S(w)$.

### 3.1.1 Acyclic, connected and tree sets

Let $S$ be a biextendable set. We say that $S$ satisfies the *acyclicity condition*, or simply that $S$ is *acyclic* if for every word $w \in S$, the graph $\mathcal{E}_S(w)$ is acyclic. A set $S$ satisfies the *connection condition*, or simply $S$ is *connected*, if for every word $w \in S$, the graph $\mathcal{E}_S(w)$ is connected.

**Example 3.1.1** Let $S$ be the Tribonacci set (see Example 1.1.7). The graphs $\mathcal{E}_S(\varepsilon)$ and $\mathcal{E}_S(ab)$ are represented in Figure 3.1.


Figure 3.1: The extension graphs $\mathcal{E}_S(\varepsilon)$ and $\mathcal{E}_S(ab)$ in the Tribonacci set.

Note that a biextendable set $S$ is acyclic (resp. connected) if and only if the graph $\mathcal{E}_S(w)$ is acyclic (resp. connected) for every bispecial word $w$. Indeed, if $w$ is not bispecial, then $\mathcal{E}_S(w) \subset a \times A$ or $\mathcal{E}_S(w) \subset A \times a$, for some letter $a \in A$, thus it is always acyclic and connected.

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

Similarly, if $\mathcal{E}_S(w)$ is connected, then $w$ is strong or neutral. Thus, if $S$ is an acyclic (resp. a connected) set, then $S$ is a weak (resp. strong) set.

Recall that an undirected graph is a tree if it is connected and acyclic. A biextendable set is called a *tree set* of characteristic $c$ (or equivalently it satisfies the *tree* condition) if for every nonempty $w \in S$, the graph $\mathcal{E}_S(w)$ is a tree and if $\mathcal{E}_S(\varepsilon)$ is a union of $c$ trees.

The following proposition is straigthforward.

**Proposition 3.1.2** *A tree set of characteristic $c$ is a neutral set of characteristic $c$.*

We use the same notation $\chi(S)$ for the characteristic of a tree set $S$.

The following result is easy to prove. Recall that a recurrent neutral set is uniformly recurrent (Corollary 2.2.9).

**Proposition 3.1.3** *An Arnoux-Rauzy set $S$ is a (uniformly) recurrent tree set of characteristic* 1.

*Proof.* It is known that an Arnoux-Rauzy set is uniformly recurrent (see, for example, [7]). Let us show that for every word $w$, including the empty one, the extension graph $\mathcal{E}_S(w)$ is a tree. Consider $w \in S$. If $w$ is not left-special there is a unique $a \in A$ such that $aw \in S$. Then $B_S(w) \subset \{a\} \times A$ and thus $\mathcal{E}_S(w)$ is a tree. The case where $w$ is not right-special is symmetrical. Finally, assume that $w$ is bispecial. Let $a, b \in A$ be such that $aw$ is right-special and $wb$

is left-special. Then $B_S(w) = (\{a\} \times A) \cup (A \times \{b\})$ and thus $\mathcal{E}_S(w)$ is a tree. Thus, the set is a tree set of characteristic 1. ∎

Since a tree set is neutral, we deduce from Proposition 2.1.3 the following statement.

**Proposition 3.1.4** *The factor complexity of a tree set is* $p_n = n(\mathrm{Card}(A) - \chi(S)) + \chi(S)$, *for all* $n > 0$.

We now present two examples, due to Julien Cassaigne [22]. The first one is a recurrent tree set of characteristic 2, and thus, in particular, an acyclic set.

**Example 3.1.5** Let $A = \{a, b, c, d\}$ and let $\sigma$ be the morphism from $A^*$ into itself defined by

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

Let $S$ be the set of factors of the infinite word $x = \sigma^\omega(a)$. Since $\sigma$ is primitive, $S$ is uniformly recurrent. The graph $\mathcal{E}_S(\varepsilon)$ is represented in Figure 3.2.


Figure 3.2: The graph $\mathcal{E}_S(\varepsilon)$.

It is acyclic with two connected components (and thus $m_S(\varepsilon) = -1$). We will show that for any non empty word $w \in S$, the graph $\mathcal{E}_S(w)$ is a tree. This will prove that $S$ is a tree set of characteristic 2. Actually, let $\pi$ be the morphism from $A^*$ onto $\{a, b\}^*$ defined by $\pi(a) = \pi(c) = a$ and $\pi(c) = \pi(d) = b$. The image of $x$ by $\pi$ is the Sturmian word $y$ which is the fixpoint of the morphism $\tau : a \mapsto ab,\ b \mapsto aba$. The word $x$ can be obtained back from $y$ by changing one every other letter $a$ into a $c$ and any letter $b$ after a $c$ into a $d$. Set $S' = \mathrm{Fac}(y)$. Thus every word of the set $S'$ gives rise to 2 words in $S$.

In this way every bispecial word $w$ of $S'$ gives two bispecial words $w', w''$ of $S$ and their extension graphs in $S$ are isomorphic to $\mathcal{E}_{S'}(w)$. For example, the word $ababa$ is bispecial in $S'$. It gives the bispecial words $abcda$ and $cdabc$ in $S$. Their extension graphs are shown below.

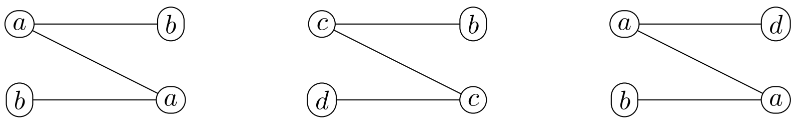


Figure 3.3: The graphs $\mathcal{E}_{S'}(ababa)$, $\mathcal{E}_S(abcda)$ and $\mathcal{E}_S(cdabc)$.

This proves that $S$ is a tree set of characteristic 2.

The second example is a recurrent set which is neutral but is not a tree set (it is actually not even acyclic).

**Example 3.1.6** Let $B = \{1, 2, 3\}$ and let $\tau : A^* \to B^*$ be defined by

$$\tau(a) = 12, \quad \tau(b) = 2, \quad \tau(c) = 3, \quad \tau(d) = 13.$$

Let $T = \tau(S)$ where $S$ is the set of Example 3.1.5. Thus $T$ is also the set of factors of the infinite word $\tau(\sigma^\omega(a))$.

The set $Y = \tau(A)$ is a prefix code. It is not a suffix code but it is *weakly suffix* in the sense that if $x, y, y' \in X$ and $x' \in X^*$ are such that $xy$ is a suffix of $x'y'$, then $y = y'$.

Let $g : \{a, c\}A^* \cap A^*\{a, c\} \to B^*$ be the map defined by

$$g(w) = \begin{cases} 3\tau(w) & \text{if } w \text{ begins and ends with } a \\ 3\tau(w)1 & \text{if } w \text{ begins with } a \text{ and ends with } c \\ 2\tau(w) & \text{if } w \text{ begins with } c \text{ and ends with } a \\ 2\tau(w)1 & \text{if } w \text{ begins with } c \text{ and ends with } c \end{cases}$$

It can be verified, using the fact that $Y$ is a prefix and weakly suffix code, that the set of nonempty bispecial words of $T$ is the union of 2, 31 and of the set $g(S)$ where $S$ is the set of nonempty bispecial words of $S$. One may verify that the words of $g(S)$ are neutral. Since the words 2, 31 are also neutral, the set $T$ is neutral. Its characteristic is $\chi(T) = 1$, as one can easily see from the extension graph of the empty word (see Figure 3.4).

It is recurrent since $S$ is recurrent and $\tau$ is a nontrivial morphism. The set $T$ is not a tree set since the graph $\mathcal{E}_T(\varepsilon)$ is neither acyclic nor connected (see Figure 3.4).

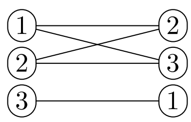


Figure 3.4: The graph $\mathcal{E}_T(\varepsilon)$.

The following is another example of a neutral set which is not a tree set.

**Example 3.1.7** Let $A = \{a, b, c\}$ and let $S$ be the set of factors of $a^*\{bc, bcbc\}a^*$. The set $S$ is biextendable. One has $S \cap A^2 = \{aa, ab, bc, cb, ca\}$. It is neutral of characteristic 1. Indeed the empty word is neutral since $b_S(\varepsilon) = \mathrm{Card}(S \cap A^2) = 5 = \ell_S(\varepsilon) + r_S(\varepsilon) - 1$. Next, the only nonempty bispecial words are $bc$ and $a^n$ for $n \geq 1$. They are neutral since $b_S(bc) = 3 = \ell_S(bc) + r_S(bc) - 1$ and $b_S(a^n) = 3 = \ell_S(a^n) + r_S(a^n) - 1$. However, $S$ is not acyclic since the graph $\mathcal{E}_S(\varepsilon)$ contains a cycle (and has two connected components, see Figure 3.5).

Note that, even if the extension graph of the empty word is the same as the one in Example 3.1.6, the two sets are different. Indeed, in this last example, the set is not recurrent.

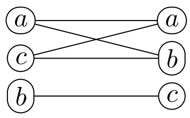

Figure 3.5: The graph $\mathcal{E}_S(\varepsilon)$.

### 3.1.2 Planar trees

Let $<_1$ and $<_2$ be two total orders on $A$. For a set $S$ and a word $w \in S$, we say that the graph $\mathcal{E}_S(w)$ is *compatible* with $<_1$ and $<_2$ if for any $(a, b), (c, d) \in B_S(w)$, one has

$$a <_2 c \Longrightarrow b \leq_1 d.$$

Thus, placing the vertices of $L_S(w)$ ordered by $<_2$ on a line and those of $R_S(w)$ ordered by $<_1$ on a parallel line, the edges of the graph may be drawn as straight noncrossing segments, resulting in a planar graph.

We say that a biextendable set $S$ is a *planar tree set* of characteristic $c$ with respect to two total orders $<_1$ and $<_2$ on $A$ if for any nonempty $w \in S$, the graph $\mathcal{E}_S(w)$ is a tree compatible with $<_1, <_2$, while $\mathcal{E}_S(w)$ is a union of $c$ trees compatible with the two orders. Obviously, a planar tree set is a tree set.

**Example 3.1.8** Let $S$ be the Fibonacci set (see Example 1.1.2). As we will prove in Section 6.1.4, $S$ is a planar tree set with respect to $<_1$ and $<_2$ on $A$ defined by: $a <_1 b$ and $b <_2 a$. The graphs $\mathcal{E}_S(\varepsilon), \mathcal{E}_S(a), \mathcal{E}_S(b)$ and $\mathcal{E}_S(ab)$ are shown in Figure 3.6.

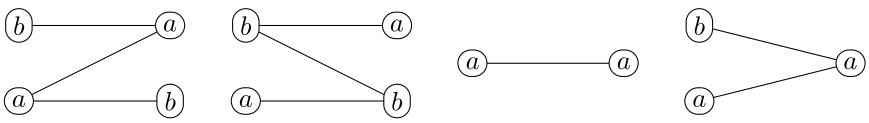

Figure 3.6: The extension graphs of $\varepsilon, a, b, ab$ in the Fibonacci set.

We will study in Chapter 6 an important family of planar tree sets containing the Fibonacci set as well as the class of Arno ux-Rauzy sets.

The following example shows that the Tribonacci set is not a planar tree set.

**Example 3.1.9** Let $S$ be the Tribonacci set (see Example 1.1.7). The words $a, aba$ and $abacaba$ are bispecial. Thus the words $ba, caba$ are right-special and the words $ab, abac$ are left-special. The graphs $\mathcal{E}_S(\varepsilon), \mathcal{E}_S(a)$ and $\mathcal{E}_S(aba)$ are shown in Figure 3.7.

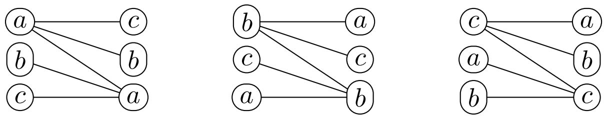

Figure 3.7: The graphs $\mathcal{E}_S(\varepsilon), \mathcal{E}_S(a)$ and $\mathcal{E}_S(aba)$ in the Tribonacci set.

One sees easily that it not possible to find two total orders on $A$ making the three graphs planar.

### 3.1.3 Generalized extension graphs

In this section we consider a variant of the extension graph.

Let $S$ be a set. For $w \in S$, and $U, V \subset S$, let $U_S(w) = \{\ell \in U \mid \ell w \in S\}$ and let $V_S(w) = \{r \in V \mid wr \in S\}$. The *generalized extension graph* of $w$ relative to $U, V$ is the following undirected graph $\mathcal{E}_S^{U,V}(w)$. The set of vertices is made of two disjoint copies of $U_S(w)$ and $V_S(w)$. The edges are the pairs $(\ell, r)$ for $\ell \in U_S(w)$ and $r \in V_S(w)$ such that $\ell wr \in S$. The extension graph $\mathcal{E}_S(w)$ defined previously corresponds to the case where $U, V = A$.

**Example 3.1.10** Let $S$ be the Fibonacci set (Example 1.1.2). Let $w = a$, $U = \{aa, ba, b\}$ and let $V = \{aa, ab, b\}$. The graph $\mathcal{E}_S^{U,V}(w)$ is represented in Figure 3.8.

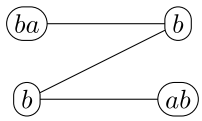


Figure 3.8: The graph $\mathcal{E}_S^{U,V}(w)$.

The following property shows that in an acyclic set, not only the extension graphs but, under appropriate hypotheses, all generalized extension graphs are acyclic.

**Proposition 3.1.11** *Let $S$ be an acyclic set. For any $w \in S$, any finite suffix code $U$ and any finite prefix code $V$, the generalized extension graph $\mathcal{E}_S^{U,V}(w)$ is acyclic.*

The proof uses the following lemma.

**Lemma 3.1.12** *Let $S$ be a biextendable set. Let $w \in S$ and let $U, V, T \subset S$. Let $\ell \in S \setminus U$ be such that $\ell w \in S$. Set $U' = (U \setminus T\ell) \cup \ell$. If the graphs $\mathcal{E}_S^{U',V}(w)$ and $\mathcal{E}_S^{T,V}(\ell w)$ are acyclic then $\mathcal{E}_S^{U,V}(w)$ is acyclic.*

*Proof.* Assume that $\mathcal{E}_S^{U,V}(w)$ contains a cycle $C$. If the cycle does not use a vertex in $U'$, it defines a cycle in the graph $\mathcal{E}_S^{T,V}(\ell w)$ obtained by replacing each vertex $t\ell$ for $t \in T$ by a vertex $t$. Since $\mathcal{E}_S^{T,V}(\ell w)$ is acyclic, this is impossible. If it uses a vertex of $U'$ it defines a cycle of the graph $\mathcal{E}_S^{U',V}(w)$ obtained by replacing each possible vertex $t\ell$ by $\ell$ (and suppressing the possible identical successive edges created by the identification). This is impossible since $\mathcal{E}_S^{U',V}(w)$ is acyclic. Thus $\mathcal{E}_S^{U,V}(w)$ is acyclic. ■

*Proof of Proposition 3.1.11.* We show by induction on the sum of the lengths of the words in $U, V$ that for any $w \in S$, the graph $\mathcal{E}_S^{U,V}(w)$ is acyclic.

Let $w \in S$. We may assume that $U = U_S(w)$ and $V = V_(w)$ and also that $U, V \neq \emptyset$. If $U, V \subset A$, the property is true since $S$ is acyclic.

Otherwise, assume for example that $U$ contains words of length at least 2. Let $u \in U$ be of maximal length. Set $u = a\ell$ with $a \in A$. Let $T = \{b \in A \mid b\ell \in U\}$. Then $U' = (U \setminus T\ell) \cup \ell$ is a suffix code and $\ell w \in S$ since $U = U_S(w)$.

By induction hypothesis, the graphs $\mathcal{E}_S^{U',V}(w)$ and $\mathcal{E}_S^{T,V}(\ell w)$ are acyclic. By Lemma 3.1.12, the graph $\mathcal{E}_S^{U,V}(w)$ is acyclic. ∎

We prove now a similar statement concerning tree sets.

**Proposition 3.1.13** *Let $S$ be a tree set. For any $w \in S$, any finite $S$-maximal suffix code $U \subset S$ and any finite $S$-maximal prefix code $V \subset S$, the generalized extension graph $\mathcal{E}_S^{U,V}(w)$ is a tree.*

The proof uses the following lemma, analogous to Lemma 3.1.12.

**Lemma 3.1.14** *Let $S$ be a biextendable set. Let $w \in S$ and let $U, V \subset S$. Let $\ell \in S \setminus U$ be such that $\ell w \in S$ and $A\ell \cap S \subset U$. Set $U' = (U \setminus A\ell) \cup \ell$. If the graphs $\mathcal{E}_S^{U',V}(w)$ and $\mathcal{E}_S^{A,V}(\ell w)$ are connected then $\mathcal{E}_S^{U,V}(w)$ is connected.*

*Proof.* Since $S$ is left extendable, there is a letter $a$ such that $a\ell w \in S$ and thus $a\ell \in U_S(w)$. We proceed by steps.

**Step 1.** As a preliminary step, let us show that for each $b \in A$ such that $b\ell w \in S$, and each $v \in V(\ell w)$, there is a path from $b\ell$ to $v$ in $\mathcal{E}_S^{U,V}(w)$. Indeed, since the graph $\mathcal{E}_S^{A,V}(\ell w)$ is connected there is a path from $b$ to $v$ in this graph. Thus, since $b\ell \in U_S(w)$, there is a path from $b\ell$ to $v$ in $\mathcal{E}_S^{U,V}(w)$.

**Step 2.** As a second step, let us show that for any $m \in U'(w) \setminus \ell$ and $v \in V_(w)$, there is a path from $m$ to $v$ in $\mathcal{E}_S^{U,V}(w)$. Indeed there is a path from $m$ to $v$ in $\mathcal{E}_S^{U',V}(w)$. For each edge of this path of the form $(\ell, s)$, $s$ is also in $V(\ell w)$ and thus, by Step 1, there is a path from $a\ell$ to $s$ in the graph $\mathcal{E}_S^{U,V}(w)$. Thus there is a path from $m$ to $v$ in $\mathcal{E}_S^{U,V}(w)$.

**Step 3.** For each $b \in A$ such that $b\ell \in U_S(w)$, for each $v \in V_(w)$, there is a path from $b\ell$ to $v$ in $\mathcal{E}_S^{U,V}(w)$. Indeed, since $\mathcal{E}_S^{A,V}(\ell w)$ is connected, there is a path from $b$ to $a$ in $\mathcal{E}_S^{A,V}(\ell w)$, thus a path from $b\ell$ to $a\ell$ in $\mathcal{E}_S^{U,V}(w)$. Then there is a path from $\ell$ to $v$ in $\mathcal{E}_S^{U',V}(w)$ and, in the same way as in Step 2, there is a path from $a\ell$ to $v$ in $\mathcal{E}_S^{U,V}(w)$.

**Step 4.** Consider now $m \in U_S(w)$ and $v \in V_(w)$. If $m \notin A\ell$, then $m \in U'(w)\backslash\ell$ and thus, by Step 2, there is a path from $m$ to $v$ in $\mathcal{E}_S^{U,V}(w)$. Next, assume that $m = b\ell$ with $b \in A$. By Step 3, there is a path from $m$ to $v$ in $\mathcal{E}_S^{U,V}(w)$. This shows that the graph $\mathcal{E}_S^{U,V}(w)$ is connected. ■

*Proof of Proposition 3.1.13.* The fact that $\mathcal{E}_S^{U,V}(w)$ is acyclic follows from Proposition 3.1.11.

We show by induction on the sum of the lengths of the words in $U, V$ that for any $w \in S$, the graph $\mathcal{E}_S^{U,V}(w)$ is connected.

Assume first that $U_S(w), V_S(w) \subset A$. Since $U$ is an $S$-maximal suffix code, we have $U_S(w) = L_S(w)$. Similarly, $V_S(w) = R_S(w)$. Thus the property is true since $S$ is a tree set.

Otherwise, assume for example that $U_S(w)$ contains words of length at least 2. Let $u \in U_S(w)$ be of maximal length. Set $u = a\ell$ with $a \in A$. Then $U' = (U \setminus A\ell) \cup \ell$ is an $S$-maximal suffix code and $\ell w \in S$ since $a\ell \in U_S(w)$. Moreover, we have $A\ell \cap S \subset U$ since $U$ is an $S$-maximal suffix code. Thus $\ell$ satisfies the hypotheses of Lemma 3.1.14.

By induction hypothesis, the graphs $\mathcal{E}_S^{U',V}(w)$ and $\mathcal{E}_S^{A,V}(\ell w)$ are connected. By Lemma 3.1.14, the graph $\mathcal{E}_S^{U,V}(w)$ is connected. ■

## 3.2 Return words in tree sets

We study sets of return words in tree sets. We first show that if $S$ is a recurrent connected set, the group described by any Rauzy graph of $S$ containing the alphabet $A$, with respect to some vertex is the free group on $A$ (Theorem 3.2.1). Next, we prove the Return Theorem, that is that in a recurrent tree set containing $A$, the set of return words to any word of $S$ is a basis of the free group on $A$ (Theorem 3.2.5).

### 3.2.1 Rauzy graphs

Recall from Section 1.3.2 that, given a factorial set $S$, the Rauzy graph of $S$ of order $n \geq 0$ is the labeled graph $G_n(S)$ with vertices the words in the set $S \cap A^n$ and edges the triples $(x, a, y)$ for all $x, y \in S \cap A^n$ and $a \in A$ such that $xa \in S \cap Ay$.

Let $G$ be a labeled graph on a set $Q$ of vertices. The group described by $G$ with respect to a vertex $v$ is the subgroup described by the simple automaton $(Q, v, v)$. We will prove the following statement.

**Theorem 3.2.1** *Let $S$ be a recurrent connected set. The group described by a Rauzy graph of $S$ with respect to any vertex is the free group on $A$.*

A *morphism* $\varphi$ from a labeled graph $G$ onto a labeled graph $H$ is a map from the set of vertices of $G$ onto the set of vertices of $H$ such that $(u, a, v)$ is an edge of $H$ if and only if there is an edge $(p, a, q)$ of $G$ such that $\varphi(p) = u$ and $\varphi(q) = v$. An *isomorphism* of labeled graphs is a bijective morphism.

The *quotient* of a labeled graph $G$ by an equivalence $\theta$, denoted $G/\theta$, is the graph with vertices the set of equivalence classes of $\theta$ and an edge from the class of $u$ to the class of $v$ labeled $a$ if there is an edge labeled $a$ from a vertex $u'$ equivalent to $u$ to a vertex $v'$ equivalent to $v$. The map from a vertex of $G$ to its equivalence class is a morphism from $G$ onto $G/\theta$.

We consider on a Rauzy graph $G_n(S)$ the equivalence $\theta_n$ formed by the pairs $(u, v)$ with $u = ax$, $v = bx$, $a, b \in L_S(x)$ such that there is a path from $a$ to $b$ in the extension graph $\mathcal{E}_S(x)$ (and more precisely from the vertex corresponding to $a$ to the vertex corresponding to $b$ in the copy corresponding to $L_S(x)$ in the bipartite graph $\mathcal{E}_S(x)$).

**Proposition 3.2.2** *If $S$ is connected, for each $n \geq 1$, the quotient of $G_n(S)$ by the equivalence $\theta_n$ is isomorphic to $G_{n-1}(S)$.*

*Proof.* The map $\varphi : S \cap A^n \to S \cap A^{n-1}$ mapping a word of $S$ of length $n$ to its suffix of length $n-1$ is clearly a morphism from $G_n(S)$ onto $G_{n-1}(S)$. If $u, v \in S \cap A^n$ are equivalent modulo $\theta_n$, then $\varphi(u) = \varphi(v)$. Thus there is a morphism $\psi$ from $G_n(S)/\theta_n$ onto $G_{n-1}(S)$. It is defined for any word $u \in S \cap A^n$ by $\psi(\bar{u}) = \varphi(u)$, where $\bar{u}$ denotes the class of $u$ modulo $\theta_n$. But since $S$ is connected, the class modulo $\theta_n$ of a word $ax$ of length $n$ has $\ell_S(x)$ elements, which is the same as the number of elements of $\varphi^{-1}(x)$. This shows that $\psi$ is a surjective map from a finite set onto a set of the same cardinality and thus that it is one-to-one. Thus $\psi$ is an isomorphism. ∎

Let $G$ be a strongly connected labeled graph. Recall from Section 1.3 that a Stallings folding at vertex $v$ relative to letter $a$ of $G$ consists in identifying the edges coming into $v$ labeled $a$ and identifying their origins. A Stallings folding does not modify the group described by the graph with respect to some vertex. Indeed, if $p \xrightarrow{a} v$, $p \xrightarrow{b} r$ and $q \xrightarrow{a} v$ are three edges of $G$, then adding the edge $q \xrightarrow{b} r$ does not change the group described since the path $q \xrightarrow{a} v \xrightarrow{a^{-1}} p \xrightarrow{b} r$ has the same label. Thus merging $p$ and $q$ does not add new labels of generalized paths.

*Proof of Theorem 3.2.1.* Let us first prove that the quotient $G_n(S)/\theta_n$ can be obtained by a sequence of Stallings foldings from the graph $G_n(S)$. Let $H$ be graph obtained by $G_n(S)$ by applying all possible Stalling foldings and let $\varphi : G_n(S) \to H$ be the natural projection, that is such that $v$ and all the other vertices merged with $v$ are sended to $\varphi(v)$. Let now $\psi : G_n(S)/\theta_n \to H$ be the map $\psi : \bar{u} \mapsto \varphi(u)$, where $\bar{u}$ is the class of $u$ modulo $\theta_n$.

The map $\psi$ is well defined. Indeed, consider $u, v \in G_n(S)$ be equivalent modulo $\theta_n$. Thus, we can write $u = ax$ and $v = bx$, with $u, v \in S \cap A^n$ and $a, b \in A$ such that $a$ and $b$ (considered as elements of $L_S(x)$), are connected by a path in $\mathcal{E}_S(x)$. Let $a_0, \ldots a_k$ and $b_1, \cdots b_k$ with $a = a_0$ and $b = a_k$ be such that $(a_i, b_{i+1})$ for $0 \leq i \leq k-1$ and $(a_i, b_i)$ for $1 \leq i \leq k$ are in $\mathcal{E}_S(x)$.

The successive Stallings foldings at $xb_1, xb_2, \ldots, xb_k$ identify the vertices $u = a_0x, a_1x, \ldots, a_kx = v$. Indeed, since $a_ixb_{i+1}, a_{i+1}xb_{i+1} \in S$, there are two edges labeled $b_{i+1}$ going out of $a_ix$ and $a_{i+1}x$ which end at $xb_{i+1}$. The Stallings folding identifies $a_ix$ and $a_{i+1}x$. By induction, we have that the two vertices $u$ and $v$ are merged in the same vertex of $H$, that is that $\varphi(u) = \varphi(v)$.

The map $\psi$ is clearly surjective. Moreover, it is a morphism from $G_n(S)/\theta_n$ onto $H$. Indeed, $(\varphi(u), a, \varphi(v))$ is an edge of $H$ if and only if there exist $u' \in \varphi^{-1}(\varphi(u))$ and $v' \in \varphi^{-1}(\varphi(v))$ such that $(u', a, v')$ is an edge of $G_n(S)$, and this implies that $(\bar{u}, a, \bar{v})$ is an edge of $G_n(S)/\theta_n$. The other direction is proved symmetrically.

Since $G_n(S)$ and $H$ are finite, the map $\psi$ is an isomorphism.

Since the Stallings foldings do not modify the group described, we deduce from Proposition 3.2.2 that the group described by the Rauzy graph $G_n(S)$ is the same as the group described by the Rauzy graph $G_0(S)$. Since $G_0(S)$ is the graph with one vertex and with loops labeled by each of the letters, it describes the free group on $A$. ∎

**Example 3.2.3** Let $S$ be the tree set obtained by decoding the Fibonacci set into blocks of length 2 (see Example 4.3.4). Set $u = aa$, $v = ab$, $w = ba$. The graph $G_2(S)$ is represented on the right of Figure 3.9.

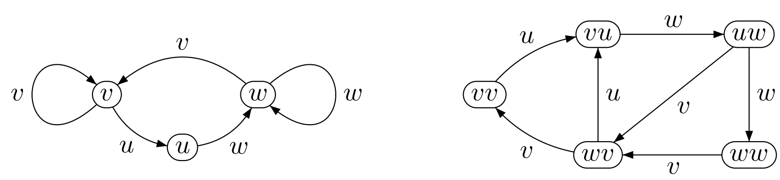


Figure 3.9: The Rauzy graphs $G_1(S)$ and $G_2(S)$ for the decoding of the Fibonacci set into blocks of length 2.

The classes of $\theta_2$ are $\{wv, vv\}$, $\{vu\}$ and $\{ww, uw\}$. The graph $G_1(S)$ is represented on the left.

The graph $G_0(S)$ is represented in Figure 3.10. The group described is the free group on 3 letters.

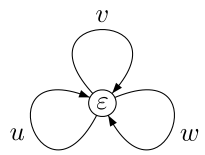


Figure 3.10: The Rauzy graph $G_0(S)$ for the decoding of the Fibonacci set into blocks of length 2.

The following example shows that Proposition 3.2.2 is false for sets which are not connected.

**Example 3.2.4** Consider again the Chacon set (see Example 1.1.3). The Rauzy graph $G_1(S)$ corresponding to the Chacon set is represented in Figure 3.11 on the left. The graph $G_1(S)/\theta_1$ is represented on the right (note that $a$ and $c$ are $\theta_1$-equivalent). It is not isomorphic to $G_0(S)$ since it has two vertices instead of one.

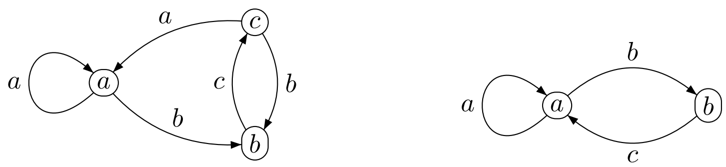


Figure 3.11: The graphs $G_1(S)$ and $G_1(S)/\theta_1$.

### 3.2.2 The Return Theorem

We can now prove the main result of this section, referred to as the Return Theorem.

**Theorem 3.2.5 (Return Theorem)** *Let $S$ be a recurrent tree set of characteristic* 1*. Then for any $w \in S$, the set $\mathcal{R}_S(w)$ is a basis of the free group on $A$.*

We first show an example of a neutral set which is not a tree set and for which Theorem 3.2.5 does not hold.

**Example 3.2.6** Consider the set $S$ of Example 3.1.6. Then, one has $\mathcal{R}_S(1) = \{2231, 31, 231\}$. This set has 3 elements, in agreement with Corollary 2.2.10 but it is not a basis of the free group on $\{1, 2, 3\}$ since it generates the same group as $\{2, 31\}$.

The proof of Theorem 3.2.5 uses Corollary 2.2.10 and the following result.

**Theorem 3.2.7** *Let $S$ be a uniformly recurrent connected set. For any $w \in S$, the set $\mathcal{R}_S(w)$ generates the free group on $A$.*

*Proof.* Since $S$ is uniformly recurrent, the set $\mathcal{R}_S(w)$ is finite. Let $n$ be the maximal length of the words in $w\mathcal{R}_S(w)$. In this way, any word in $S \cap A^n$ beginning with $w$ has a prefix in $wR_S(w)$. Moreover, recall from Property (ii) of Rauzy graphs (Section 1.3.2), that the label of any path of length $n+1$ in the Rauzy graph $G_n(S)$ is in $S$.

Let $x \in S$ be a word of length $n$ ending with $w$. Let $\mathcal{A}$ be the simple automaton defined by $G_n(S)$ with initial and terminal state $x$. Let $X$ be the prefix code generating the submonoid recognized by $\mathcal{A}$. Since the automaton $\mathcal{A}$ is simple, by Proposition 1.3.5, the set $X$ generates the group described by $\mathcal{A}$.

We show that $X \subset \mathcal{R}_S(w)^*$. Indeed, let $y \in X$. Since $y$ is the label of a path starting at $x$ and ending in $x$, the word $xy$ ends with $x$ and thus the word $wy$ ends with $w$. Let $\Gamma = \Gamma_{A^+}(w) = \{z \in A^+ \mid wz \in A^+w\}$ and let $R = \mathcal{R}_{A^+}(w) = \Gamma \setminus \Gamma A^+$. Then $R$ is a prefix code and $\Gamma \cup \{\varepsilon\} = R^*$ (see Section 1.4). Since $y \in \Gamma$, we can write $y = u_1u_2 \cdots u_m$ where each word $u_i$ is in $R$. Since $S$ is recurrent and since $x \in S$, there is $v \in S \cap A^n$ such that $vx \in S$ and thus there is a path labeled $x$ ending at the vertex $x$ by property (i) of Rauzy graphs. Thus there is a path labeled $xy$ in $G_n(S)$. This implies that for $1 \le i \le m$, there is a path in $G_n(S)$ labeled $wu_i$ (see Figure 3.12).

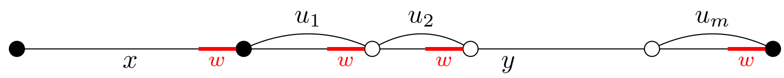


Figure 3.12: The word $xy$ in $G_n(S)$.

Assume that some $u_i$ is such that $|wu_i| > n$. Then the prefix $p$ of length $n$ of $wu_i$ is the label of a path in $G_n(S)$. This implies, by Property (ii) of Rauzy graphs, that $p$ is in $S$ and thus that $p$ has a prefix in $wR_S(w)$. But then $wu_i$ has a proper prefix in $wR_S(w)$, a contradiction. Thus we have $|wu_i| \le n$ for all $i = 1, 2, \ldots, m$. But then the $wu_i$ are in $S$ by property (i) again and thus the $u_i$ are in $\mathcal{R}_S(w)$. This shows that $y \in \mathcal{R}_S(w)^*$.

Thus the group generated by $\mathcal{R}_S(w)$ contains the group generated by $X$. But, by Theorem 3.2.1, the group described by $\mathcal{A}$ is the free group on $A$. Thus $\mathcal{R}_S(w)$ generates the free group on $A$. ∎

We illustrate the proof in the following example.

**Example 3.2.8** Let $S$ be the Fibonacci set. We have $\mathcal{R}_S(aa) = \{baa, babaa\}$. The Rauzy graph $G_7(S)$ is represented in Figure 3.13. The set recognized by the automaton obtained using $x = aababaa$ as initial and terminal state is $X^*$ with $X = \{babaa, baababaa\}$. In agreement with the proof of Theorem 3.2.7, we have $X \subset \mathcal{R}_S(aa)^*$.

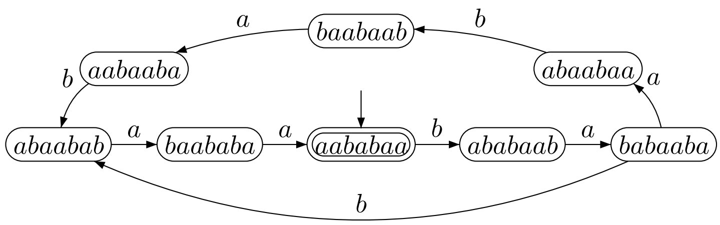


Figure 3.13: The Rauzy graph $G_7(S)$

*Proof of Theorem 3.2.5.* When $S$ is a tree set of characteristic 1, we have $\mathrm{Card}(\mathcal{R}_S(w)) = \mathrm{Card}(A)$ by Corollary 2.2.10, which implies the conclusion since any set with $\mathrm{Card}(A)$ elements generating $F_A$ is a basis of $F_A$. ∎

### 3.2.3 Derived sets of tree sets

We will use the following closure property of the family of recurrent tree sets. It generalizes the fact that the derived word of a Sturmian word is Sturmian (see [45]).

**Theorem 3.2.9** *Any derived set of a recurrent tree set of characteristic* 1 *is a recurrent tree set of characteristic* 1*.*

*Proof.* Let $S$ be a uniformly recurrent tree set. Let $v \in S$ and let $f$ be a coding morphism for $X = \mathcal{R}_S(v)$. By Theorem 3.2.5, $X$ is a basis of the free group on $A$. Thus $f : B^* \to A^*$ extends to an isomorphism from $F_B$ onto $F_A$.

Set $H = f^{-1}(v^{-1}S)$. By Proposition 1.4.6, the set $H$ is recurrent and $H = f^{-1}(\Gamma_S(v)) \cup \{\varepsilon\}$.

Consider $x \in H$ and set $y = f(x)$. Let $f'$ be the coding morphism for $X' = \mathcal{R}'_S(v)$ associated with $f$. For $a, b \in B$, we have

$$(a, b) \in B_S(x) \quad \Longleftrightarrow \quad (f'(a), f(b)) \in B_S^{X',X}(vy),$$

where $B_S^{X',X}(vy)$ is the set of edges of the generalized extension graph $\mathcal{E}^{X',X}(vy)$ (see Section 3.1.3). Indeed,

$$axb \in H \;\Leftrightarrow\; f(a)yf(b) \in \Gamma_S(v) \;\Leftrightarrow\; vf(a)yf(b) \in S \;\Leftrightarrow\; f'(a)vyf(b) \in S.$$

The set $X'$ is an $Sv^{-1}$-maximal suffix code and the set $X$ is a $v^{-1}S$-maximal prefix code. By Proposition 3.1.13 the generalized extension graph $\mathcal{E}^{X',X}(vy)$ is a tree. Thus the graph $\mathcal{E}_S(x)$ is a tree. This shows that $H$ is a tree set of characteristic 1.

Let us now prove that $H$ is (uniformly) recurrent. Consider $x \in H \setminus \varepsilon$. Set $y = f(x)$. Let us show that $\Gamma_H(x) = f^{-1}(\Gamma_S(vy))$ or equivalently $f(\Gamma_H(x)) = \Gamma_S(vy)$. Consider first $r \in \Gamma_H(x)$. Set $s = f(r)$. Then $xr = ux$ with $u, ux \in H$. Thus $ys = wy$ with $w = f(u)$.

Since $u \in H \setminus \{\varepsilon\}$, $w = f(u)$ is in $\Gamma_S(v)$, we have $vw \in A^+v \cap S$. This implies that $vys = vwy \in A^+vy \cap S$ and thus that $s \in \Gamma_S(vy)$. Conversely, consider $s \in \Gamma_S(vy)$. Since $y = f(x)$, we have $s \in \Gamma_S(v)$. Set $s = f(r)$. Since $vys \in A^+vy \cap S$, we have $ys \in A^+y \cap S$. Set $ys = wy$. Then $vwy \in A^+vy$ implies $vw \in A^+v$ and therefore $w \in \Gamma_S(v)$. Setting $w = f(u)$, we obtain $f(xr) = ys = wy \in X^+y \cap \Gamma_S(v)$. Thus $r \in \Gamma_H(x)$. This shows that $f(\Gamma_H(x)) = \Gamma_S(vy)$ and thus that $\mathcal{R}_H(x) = f^{-1}(\mathcal{R}_S(vy))$.

Since $S$ is uniformly recurrent, the set $\mathcal{R}_S(vy)$ is finite. Since $f$ is an isomorphism, $\mathcal{R}_H(x)$ is also finite, which shows that $H$ is uniformly recurrent. ■

**Example 3.2.10** Let $S$ be the Tribonacci set (see Example 1.1.7), which is is the set of factors of the infinite word $x = abacabaabacababacabaabacaba \cdots$. We have $\mathcal{R}_S(a) = \{a, ba, ca\}$. Let $g$ be the coding morphism for $\mathcal{R}_S(a)$ defined by $g(a) = a$, $g(b) = ba$, $g(c) = ca$ and let $g'$ be the associated coding morphism for $\mathcal{R}'_S(a)$. We have $f = g'\pi$ where $\pi$ is the circular permutation $\pi = (abc)$. Set $z = g'^{-1}(x)$. Since $g'\pi(x) = x$, we have $z = \pi(x)$. Thus the derived set of $S$ with respect to $a$ is the set $\pi(S)$.

## 3.3 Multiplying maps

We now introduce a construction which allows one to build tree sets of characteristic $m$ starting from a tree set of characteristic 1. We will use this method in Chapter 5 to construct a family of specular sets.

Recall from Section 1.3 the definition of automaton. 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 $\Sigma$ is called the *input alphabet* and $A$ is called the *output alphabet*. The automaton 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 $Q = \{0, 1, \ldots, m-1\}$ on the input alphabet $\Sigma$ and the output alphabet $A$. We assume that

1. the input automaton is a group automaton, that is, every letter of $\Sigma$ acts on $Q$ as a permutation,

2. the output labels of the edges are all distinct.

We define $m$ maps $\delta_k : \Sigma^* \to A^*$ corresponding to the initial state $k$, for $k = 0, 1, \ldots, m-1$. Let $\delta_k(u) = v$ if the path starting at state $k$ with input label $u$ has output $v$. An $m$-tuple $\delta = (\delta_0, \delta_1, \ldots, \delta_{m-1})$ is called a *$m$-multiplying map* and the transducer $\mathcal{A}$ a *$m$-multiplying transducer*. The *image* of a set of words $T$ on the alphabet $\Sigma$ by the $m$-multiplying map $\delta$ is the set $\delta_0(T) \cup \delta_1(T) \cup \cdots \cup \delta_{m-1}(T)$.

**Theorem 3.3.1** *For any tree set $T$ of characteristic $c$ on the alphabet $\Sigma$ and any $m$-multiplying map $\delta$, the image of $T$ by $\delta$ is a tree set of characteristic $mc$.*

*Proof.* Set $S = \delta_0(T) \cup \delta_1(T) \cup \cdots \cup \delta_{m-1}(T)$. The set $S$ is clearly biextendable since $T$ is biextendable by definition.

Let us consider a nonempty word $x = \delta_i(y)$, with $0 \leq i \leq m-1$. The graph $\mathcal{E}_S(x)$ is isomorphic to the graph $\mathcal{E}_T(y)$. Indeed, let $j$ be the end of the path with origin $i$ and input label $y$ in the $m$-multiplying transducer. For $a_i, b_j \in A$, one has $a_i x b_j \in S$ if and only if $ayb \in T$ where $a$ (resp. $b$) is the input label of the edge with output label $a_i$ (resp. $b_j$) ending in $i$ (resp. with origin $j$). Thus, $\mathcal{E}_S(x)$ is a tree for any nonempty word $x \in S$.

Finally, the graph $\mathcal{E}_S(\varepsilon)$ is, up to orientation, the union of $m$ graphs, all of them isomorphic to $\mathcal{E}_T(\varepsilon)$. Indeed, consider the map $\pi$ from $S \cap A^2$ onto $\{0, 1, \cdots, m-1\}$ which assigns to $ab \in S \cap A^2$ the state $i$ which is the end of the edge of $\mathcal{A}$ with output label $a$ (and the origin of the edge with output label $b$). Set $S_i = \pi^{-1}(i)$. We have a partition $S \cap A^2 = S_0 \cup S_1 \cup \cdots \cup S_{m-1}$ such that each graph having $S_i$ as set of edges is isomorphic to $\mathcal{E}_T(\varepsilon)$. Since $\mathcal{E}_T(\varepsilon)$ is a forest of $c$ trees, the graph $\mathcal{E}_S(\varepsilon)$ is a forest of $mc$ trees. ∎

**Example 3.3.2** Let $B = \{\alpha\}$ and let $T = \mathrm{Fac}\,(\alpha^*)$. Let $\delta$ be the doubling map given by the transducer of Figure 3.14.

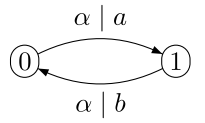


Figure 3.14: A doubling automaton.

The image of $T$ by $\delta$ is the set $S = \mathrm{Fac}\,((ab)^*)$ of Example 2.3.3. The graph $\mathcal{E}_S(\varepsilon)$ is represented in Figure 3.15.

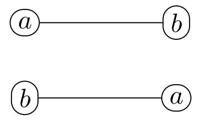


Figure 3.15: The graph $\mathcal{E}_S(\varepsilon)$.

The set $S$ is a tree set of characteristic 2 according to Theorem 3.3.1.

**Example 3.3.3** Let $B = \{\alpha, \beta\}$ and let $T$ be the Fibonacci set (see Example 1.1.2). Let $\delta$ be the doubling map given by the transducer of Figure 3.16.

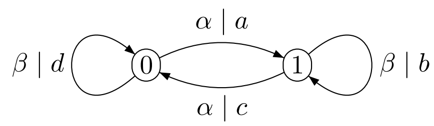


Figure 3.16: A doubling automaton.

The graph $\mathcal{E}_S(\varepsilon)$ is represented in Figure 3.2.

## 3.4 Palindromes

We close this chapter with a connection between tree sets and palindromes.

The notion of palindromic complexity originates in [35] 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 [43], a recurrent set closed under reversal is full if and only if every complete return word to a palindrome in $S$ is a palindrome. It is known that all Sturmian sets are full [35] and also all natural codings of interval exchange defined by a symmetric permutation [4].

The fact that a tree set of characteristic 1 is full in the following result generalizes results of [35, 4].

**Proposition 3.4.1** *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 [60]): 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 $\gamma$ 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 $\gamma$, one obtains a cycle. Thus $T$ is full. ■

# Chapter 4

# Bifix codes in tree sets

In this chapter we concentrate on the study of bifix codes in tree sets and their connection to subgroups of the free group. Some results are true for acyclic sets, some for tree sets of an arbitrary characteristic and others only for tree sets of characteristic 1. When it is possible we state and prove the result using the weakest hypothesis.

In Section 4.1 we prove that bifix codes in acyclic sets are bases of the subgroup that they generate (Theorem 4.1.1, referred to as the Freeness Theorem). Moreover, we prove that the submonoid generated by a finite bifix code $X$ included in an acyclic set $S$ is such that $X^* \cap S = \langle X \rangle \cap S$ (Theorem 4.1.2, referred to as the Saturation Theorem). In order to prove the Freeness and the Saturation Theorems we introduce some tools: incidence graphs (Section 4.1.1) and coset automata (Section 4.1.2).

In Section 4.2 we define the finite index basis property that connects bifix codes with subgroups, and we prove the Finite Index Basis Theorem (Theorem 4.2.1) which states that a recurrent tree set of characteristic 1 has this property. In the same section we also discuss about tame bases and $\mathcal{S}$-adic representations.

Section 4.3 is devoted to the study of maximal bifix decoding in tree sets. we states several closure properties (Theorems 4.3.1, 4.3.3, 4.3.5 and 4.3.17) showing that the stronger is the hypothesis, the stronger is the result. We also give a result about the composition of bifix codes in a tree set (Theorem 4.3.11) and introduce modular codes (Section 4.3.3).

## 4.1 Bifix codes in acyclic sets

Let $X$ be a subset of the free group. We say that $X$ is *free* if it is a basis of the subgroup $\langle X \rangle$ generated by $X$. This means that if $x_1, x_2, \ldots, x_n \in X \cup X^{-1}$ are such that $x_1 x_2 \cdots x_n$ is equivalent to $\varepsilon$, then $x_i x_{i+1}$ is equivalent to $\varepsilon$ for some $1 \leq i < n$.

We will prove the following result (Freeness Theorem).

**Theorem 4.1.1 (Freeness Theorem)** *A set $S$ is acyclic if and only if any bifix code $X \subset S$ is a free subset of the free group $F_A$.*

Let $M$ be a submonoid of $A^*$ and let $H$ be the subgroup of $F_A$ generated by $M$. Given a set of words $S$, the submonoid $M$ is said to be *saturated* in $S$ if $M \cap S = H \cap S$. Note that the inclusion $M \cap S \subset H \cap S$ is always satisfied. Thus M is saturated if taking the subgroup generated we do not have additional words of $S$.

If $M$ is generated by $X$, then $M$ is saturated in $S$ if and only if $X^* \cap S = \langle X \rangle \cap S$.

Thus, for example, the submonoid recognized by a reversible automaton is saturated in $A^*$ (Proposition 1.3.5).

We will prove the following result (Saturation Theorem).

**Theorem 4.1.2 (Saturation Theorem)** *Let $S$ be an acyclic set. The submonoid generated by a bifix code included in $S$ is saturated in $S$.*

As a preliminary to the proof of The Freeness Theorem and the Saturation Theorem, we first define, in Section 4.1.1, the incidence graph of a finite bifix code (already used in [7]). We prove a result concerning this graph, implying in particular that it is acyclic (Proposition 4.1.3).

We then define, in Section 4.1.2, the coset automaton whose states are connected components of the incidence graph. We prove that this automaton is the Stallings automaton of the subgroup $\langle X \rangle$ (Proposition 4.1.7).

Finally, in Sections 4.1.3 and 4.1.4, we prove the Freeness and the Saturation Theorem and we show some corollaries and examples.

### 4.1.1 Incidence graph

Let $X$ be a set, let $P_X$ be the set of its nonempty proper prefixes and $S_X$ be the set of its nonempty proper suffixes. Recall from [7] that the *incidence graph* of $X$ is the undirected graph $\mathcal{G}_X$ defined as follows. The set of vertices is the disjoint union of $P_X$ and $S_X$. The edges of $\mathcal{G}_X$ are the pairs $(p, s)$ for $p \in P_X$ and $s \in S_X$ such that $ps \in X$ As in any undirected graph, a connected component of $\mathcal{G}_X$ is a maximal set of vertices connected by paths.

The following result is proved in [7, Lemma 6.3.3] in the case of an Arnoux-Rauzy set. We give here a proof in the more general case of an acyclic set. We call a path *reduced* if it does not use equal consecutive edges.

**Proposition 4.1.3** *Let $S$ be an acyclic set, let $X \subset S$ be a bifix code and let $\mathcal{G}_X$ be the incidence graph of $X$. Then the following assertions hold.*

*(i) The graph $\mathcal{G}_X$ is acyclic.*

*(ii) The intersection of $P_X$ (resp. $S_X$) 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* $u_1, \ldots, u_n \in P_X$ *and* $v_1, \ldots, v_{n+1}$ *in* $S_X$*, 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* $u_1, \ldots, u_{n+1} \in P_X$ *and* $v_1, \ldots, v_n \in S_X$*, the longest common suffix of* $u_1, u_{n+1}$ *is a proper suffix of* $u_1, u_2, \ldots, u_{n+1}$*.*

*Proof.* Assertions (iii) and (iv) imply Assertions (i) and (ii). Indeed, assume that (iii) holds. Consider a reduced path $(v_1, u_1, \ldots, u_n, v_{n+1})$ in $\mathcal{G}_X$ with $u_1, \ldots, u_n \in P_X$ and $v_1, \ldots, v_{n+1}$ in $S_X$. If $v_1 = v_{n+1}$, then the longest common prefix of $v_1, v_{n+1}$ is not a proper prefix of them. Thus $\mathcal{G}_X$ is acyclic and (i) holds. Next, if $v_1$, $v_{n+1}$ are comparable for the prefix order, their longest common prefix is one of them, a contradiction with (iii) again. The assertion on $P_X$ is proved in an analogous way using Assertion (iv).

We prove (iii) and (iv) by induction on $n \geq 1$.

The assertions holds for $n = 1$. Indeed, if $u_1v_1, u_1v_2 \in X$ and if $v_1 \in S_X$ is a prefix of $v_2 \in S_X$, then $u_1v_1$ is a prefix of $u_1v_2$, a contradiction with the hypothesis that $X$ is a prefix code. The same holds symmetrically for $u_1v_1, u_2v_1 \in X$ since $X$ is a suffix code.

Let $n \geq 2$ and assume that the assertions hold for any path of length at most $2n - 2$. We treat the case of a path $(v_1, u_1, \ldots, u_n, v_{n+1})$ in $\mathcal{G}_X$ with $u_1, \ldots, u_n \in P_X$ and $v_1, \ldots, v_{n+1}$ in $S_X$. The other case is symmetric.

Let $p$ be the longest common prefix of $v_1$ and $v_{n+1}$. We may assume that $p$ is nonempty since otherwise the statement is obviously true. Any two elements of the set $U = \{u_1, \ldots, u_n\}$ are connected by a path of length at most $2n - 2$ (using elements of $\{v_2, \ldots v_n\}$). Thus, by induction hypothesis, $U$ is a suffix code. Similarly, any two elements of the set $V = \{v_1, \ldots, v_n\}$ are connected by a path of length at most $2n - 2$ (using elements of $\{u_1, \ldots u_{n-1}\}$). Thus $V$ is a prefix code. We cannot have $v_1 = p$ since otherwise, using the fact that $u_np$ is a prefix of $u_nv_{n+1}$ and thus in $S$, the generalized extension graph $\mathcal{E}^{U,V}(\varepsilon)$ would have the cycle $(p, u_1, v_2, \ldots, u_n, p)$, a contradiction since $\mathcal{E}^{U,V}(\varepsilon)$ is acyclic by Proposition 3.1.13. Similarly, we cannot have $v_{n+1} = p$.

Set $W = p^{-1}V$ and $V' = (V \setminus pW) \cup p$. Since $V$ is a prefix code and since $p$ is a proper prefix of $V$, the set $V'$ is a prefix code. Suppose that $p$ is not a proper prefix of all $v_2, \ldots, v_n$. Then there exist $i, j$ with $1 \leq i < j \leq n+1$ such that $p$ is a proper prefix of $v_i, v_j$ but not of any $v_{i+1}, \ldots, v_{j-1}$. Then $v_{i+1}, \ldots, v_{j-1} \in V'$ and there is the cycle $(p, u_i, v_{i+1}, u_{i+1}, \ldots, v_{j-1}, u_{j-1}, p)$ in the graph $\mathcal{E}^{U,V'}(\varepsilon)$. This is in contradiction with Proposition 3.1.13 because, $V'$ being a prefix code, $\mathcal{E}^{U,V'}(\varepsilon)$ is acyclic. Thus $p$ is a proper prefix of all $v_2, \ldots, v_n$. ∎

Let $X$ be a bifix code and let $P_X$ be the set of nonempty proper prefixes of $X$. Consider the equivalence $\theta_X$ on $P_X \cup \{\varepsilon\}$ which is the transitive closure of the relation formed by the pairs $p, q \in P_X \cup \{\varepsilon\}$ such that $ps, qs \in X$ for some $s \in A^+$. Such a pair corresponds, when $p, q \neq \varepsilon$, to a path $p \to s \to q$ in the incidence graph of $X$. We call the equivalence $\theta_X$ the *coset equivalence* of $X$.

Thus a class of $\theta_X$ is either reduced to the empty word or it is the intersection of $P_X$ with a connected component of the incidence graph of $X$.

The following property, proved in [7, Proposition 6.3.5], relates the equivalence $\theta_X$ with the right cosets of $H = \langle X \rangle$.

**Proposition 4.1.4** *Let $X$ be a bifix code, let $P = P_X \cup \{\varepsilon\}$ be the set of proper prefixes of $X$ and let $H$ be the subgroup generated by $X$. For any $p, q \in P$, $p \equiv q \bmod \theta_X$ implies $Hp = Hq$.*

The following result is proved in [7, Lemmas 6.3.6 and 6.4.2] in the case of an Arnoux-Rauzy set $S$. It shows that the equivalence $\theta_X$ is compatible with the transitions of the literal automaton $\mathcal{A} = (P, \varepsilon, \varepsilon)$ of $X^*$.

**Proposition 4.1.5** *Let $S$ be an acyclic set. Let $X \subset S$ be a bifix code and let $P = P_X \cup \{\varepsilon\}$ be the set of proper prefixes of $X$. Let $p, q \in P$ and $a \in A$ be such that $pa, qa \in P \cup X$. Then in the literal automaton of $X^*$, one has $p \equiv q \bmod \theta_X$ if and only if $p \cdot a \equiv q \cdot a \bmod \theta_X$.*

*Proof.* Assume first that $p \equiv q \bmod \theta_X$. We may assume that $p, q$ are nonempty. Let $(u_0, v_1, u_1, \ldots, v_n, u_n)$ be a reduced path in the incidence graph $\mathcal{G}_X$ of $X$ with $p = u_0$, $u_n = q$. The corresponding words in $X$ are $u_0v_1, u_1v_1, u_1v_2, \ldots, u_nv_n$. We may assume that the words $u_i$ are pairwise distinct, and that the $v_i$ are pairwise distinct. Moreover, since $pa, qa \in P \cup X$ there exist words $v, w$ such that $pav, qaw \in X$. Set $v_0 = av$ and $v_{n+1} = aw$.

By Proposition 4.1.3, $a$ is a proper prefix of $v_0, v_1, \ldots, v_{n+1}$. Set $v_i = av'_i$ for $0 \le i \le n+1$.

If $pa, qa \in P$, then $(u_0a, v'_1, u_1a, \ldots, v'_n, u_na)$ is a path from $pa$ to $qa$ in $\mathcal{G}_X$. This shows that $pa \equiv qa \bmod \theta_X$.

Next, suppose that $pa \in X$ and thus that $v_0 = a$. By Proposition 4.1.3, we have $w = \varepsilon$ since otherwise $v_0 = a$ is a proper prefix of $v_{n+1}$. Thus $qa \in X$ and $p \cdot a = q \cdot a$.

Conversely, if $p \cdot a \equiv q \cdot a \bmod \theta_X$, assume first that $pa, qa \in P$. Then $pa \equiv qa \bmod \theta_X$ and thus there is a reduced path $(u_0, v_1, \ldots, v_n, u_n)$ in $\mathcal{G}_X$ with $u_0 = pa$ and $u_n = qa$. By Proposition 4.1.3, $a$ is a proper suffix of $u_1, \ldots, u_n$. Set $u_i = u'_ia$. Thus $(p, av_1, u'_1, \ldots, q)$ is a path in $\mathcal{G}_X$, showing that $p \equiv q \bmod \theta_X$.

Finally, if $pa, qa \in X$, then $(p, a, q)$ is a path in $\mathcal{G}_X$ and thus $p \equiv q \bmod \theta_X$. ∎

### 4.1.2 Coset automaton

Let $S$ be an acyclic set and let $X \subset S$ be a bifix code. Let $P$ be the set of poper prefixes of $X$. We introduce a new automaton denoted $\mathcal{B}_X$ and called the *coset automaton* of $X$. Let $Q$ be the set of classes of $\theta_X$ with the class of $\varepsilon$ still denoted $\varepsilon$. The coset automaton of $X$ is the automaton $\mathcal{B}_X = (Q, \varepsilon, \varepsilon)$ with set of states $Q$ and transitions induced by the transitions of the literal automaton

$\mathcal{A} = (P, \varepsilon, \varepsilon)$ of $X^*$. Formally, or $r, s \in Q$ and $a \in A$, one has $r \cdot a = s$ in the automaton $\mathcal{B}_X$ if there exist $p$ in the class $r$ and $q$ in the class $s$ such that $p \cdot a = q$ in the automaton $\mathcal{A}$.

Observe first that the definition is consistent since, by Proposition 4.1.5, if $p \cdot a$ and $p' \cdot a$ are nonempty and $p, p'$ are in the same class $r$, then $p \cdot a$ and $p' \cdot a$ are in the same class.

Observe next that if there is a path from $p$ to $p'$ in the automaton $\mathcal{A}$ labeled $w$, then there is a path from the class $r$ of $p$ to the class $r'$ of $p'$ labeled $w$ in $\mathcal{B}_X$.

**Example 4.1.6** Let $S$ be the Fibonacci set and let

$$X = \{a, baab, babaabab, babaabaabab\}.$$

The set $X$ is an $S$-maximal bifix code of $S$-degree 3 (see [7, Example 6.3.1]). The automaton $\mathcal{B}_X$ has three states, as shown in Figure 4.1.

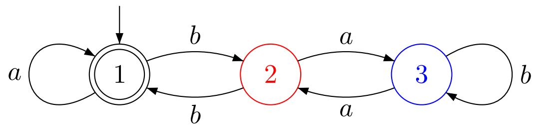


Figure 4.1: The automaton $\mathcal{B}_X$.

It is a group automaton. State 2 is the class containing $b$, and state 3 is the class containing $ba$. The bifix code generating the submonoid recognized by this automaton is $Z = a \cup b(ab^*a)^*b$.

The following result shows that the coset automaton of $X$ is the Stallings automaton of the subgroup generated by $X$ (recall Section 1.3).

**Proposition 4.1.7** *Let $S$ be an acyclic set, and let $X \subset S$ be a bifix code. The coset automaton $\mathcal{B}_X$ is reversible and describes the subgroup generated by $X$. Moreover $X \subset Z$, where $Z$ is the bifix code generating the submonoid recognized by $\mathcal{B}_X$.*

*Proof.* Let $\mathcal{A} = (P, \varepsilon, \varepsilon)$ be the literal automaton of $X^*$ and set $\mathcal{B}_X = (Q, \varepsilon, \varepsilon)$. By Proposition 4.1.5, the automaton $\mathcal{B}_X$ is reversible.

Let $Z$ be the bifix code generating the submonoid recognized by $\mathcal{B}_X$. To show the inclusion $X \subset Z$, consider a word $x \in X$. There is a path from $\varepsilon$ to $\varepsilon$ labeled $x$ in $\mathcal{A}$, hence also in $\mathcal{B}_X$. Since the path in $\mathcal{A}$ does not pass by $\varepsilon$ except at its ends and since the class of $\varepsilon$ modulo $\theta_X$ is reduced to $\varepsilon$, the path in $\mathcal{B}_X$ does not pass by $\varepsilon$ except at its ends. Thus $x$ is in $Z$.

Let us finally show that the coset automaton describes the group $H = \langle X \rangle$. By Proposition 1.3.5, the subgroup described by $\mathcal{B}_X$ is equal to $\langle Z \rangle$. Set $K = \langle Z \rangle$. Since $X \subset Z$, we have $H \subset K$. To show the converse inclusion, let us show by induction on the length of $w \in A^*$ that if, for $p, q \in P$, there is a path from the class of $p$ to the class of $q$ in $\mathcal{B}_X$ with label $w$ then $Hpw = Hq$. By

Proposition 4.1.4, this holds for $w = \varepsilon$. Next, assume that it is true for $w$ and consider $wa$ with $a \in A$. Assume that there are states $p, q, r \in P$ such that there is a path from the class of $p$ to the class of $q$ in $\mathcal{B}_X$ with label $w$, and an edge from the class of $q$ to the class of $r$ in $\mathcal{B}_X$ with the label $a$. By induction hypothesis, we have $Hpw = Hq$. Next, by definition of $\mathcal{B}_X$, there is an $s \equiv q \bmod \theta_X$ such that $s \cdot a \equiv r \bmod \theta_X$. If $sa \in P$, then $s \cdot a = sa$, and by Proposition 4.1.4, we have $Hs = Hq$ and $Hsa = Hr$. Otherwise, $sa \in X \subset H$ and $s \cdot a = r = \varepsilon$ because the class of $\varepsilon$ is a singleton and thus $Hqa = Hsa = H = Hr$. In both cases, $Hpwa = Hqa = Hsa = Hr$. This property shows that if $z \in Z$, then $Hz = H$, that is $z \in H$. Thus $Z \subset H$ and finally $H = K$. ∎

### 4.1.3 Freeness Theorems

We can now prove Theorem 4.1.1. The proof uses Proposition 4.1.3.

*Proof of the Freeness Theorem.* To prove the necessity of the condition, assume that for some $w \in S$ the graph $\mathcal{E}_S(w)$ contains a cycle $(a_1, b_1, \ldots, a_p, b_p, a_1)$ with $p \geq 2$, $a_i \in L_S(w)$ and $b_i \in R_S(w)$ for $1 \leq i \leq p$. Consider the bifix code $X = AwA \cap S$. Then $a_1wb_1, a_2wb_1, \ldots, a_pwb_p, a_1wb_p \in X$. But

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

contradicting the fact that $X$ is free.

Let us now show the converse. Assume that $S$ is acyclic and let $X \subset S$ be a bifix code. Set $Y = X \cup X^{-1}$. Let $y_1, \ldots, y_n \in Y$. We intend to show that provided $y_iy_{i+1} \not\equiv \varepsilon$ for $1 \leq i < n$, we have $y_1 \cdots y_n \not\equiv \varepsilon$. We may assume $n \geq 3$. We say that a sequence $(u_i, v_i, w_i)_{1 \leq i \leq n}$ of elements of the free group on $A$ is *admissible* with respect to $y_1, \ldots, y_n$ if the following conditions are satisfied (see Figure 4.2):

(i) $y_i = u_iv_iw_i$ for $1 \leq i \leq n$,

(ii) $u_1 = w_n = \varepsilon$ and $v_1, v_n \neq \varepsilon$,

(iii) $w_iu_{i+1} \equiv \varepsilon$ for $1 \leq i \leq n-1$,

(iv) for $1 \leq i < j \leq n$, if $v_i, v_j \neq \varepsilon$ and $v_k = \varepsilon$ for $i+1 \leq k \leq j-1$, then $v_iv_j$ is reduced.

Note that if the sequence $(u_i, v_i, w_i)_{1 \leq i \leq n}$ is admissible with respect to $y_1, \ldots, y_n$, then $y_1 \cdots y_n$ is equivalent to the word $v_1 \cdots v_n$ which is a reduced nonempty word. Thus, in particular $y_1 \cdots y_n \not\equiv \varepsilon$.

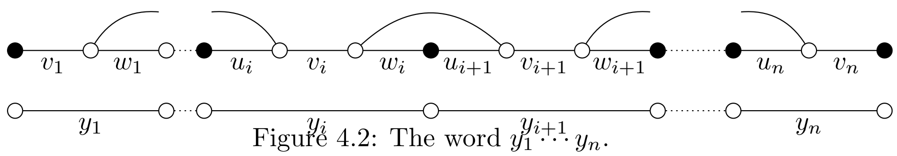


Figure 4.2: The word $y_1 \cdots y_n$.

Let us show by induction on $n$ that for any $y_1, \ldots, y_n$ such that $y_i y_{i+1} \not\equiv \varepsilon$ for $1 \leq i \leq n-1$, there exists an admissible sequence with respect to $y_1 \ldots, y_n$.

The property is true for $n = 1$. Indeed, we take $u_1 = w_1 = \varepsilon$.

Assume that the property is true for $n$. Among the possible admissible sequences with respect to the $y_1, \ldots, y_n$, we choose one such that $|v_n|$ is maximal.

Set $v_n = v'_n w'_n$ and $y_{n+1} = u_{n+1} v_{n+1}$ with $|w'_n| = |u_{n+1}|$ maximal such that $w'_n u_{n+1} \equiv \varepsilon$. Note that $v_{n+1} \neq \varepsilon$ since otherwise $y_{n+1}$ would cancel completely with $y_n$.

If $v'_n \neq \varepsilon$, the sequence

$$(\varepsilon, v_1, w_1), \ldots, (u_{n-1}, v_{n-1}, w_{n-1}), (u_n, v'_n, w'_n), (u_{n+1}, v_{n+1}, \varepsilon)$$

is admissible with respect to $y_1, \ldots, y_{n+1}$.

Otherwise, let $i$ with $1 \leq i < n$ be the largest integer such that $v_i \neq \varepsilon$. Observe that $w_i, w_{i+1}, \ldots, w_{n-1}, w'_n$ are nonempty. Indeed, if $w_j = \varepsilon$ with $i \leq j \leq n-1$, then $u_{j+1} = \varepsilon$ and thus $y_{j+1}$ cancels completely with $y_{j+2}$. Next, if $v_n = w'_n = \varepsilon$, then $y_n$ cancels completely with $y_{n-1}$.

Assume that $y_i \in X$ (the other case is symmetric). If $y_{n+1} \in X$ (and thus $n - i$ is odd), then $v_i v_{n+1}$ is reduced because they are both in $A^*$ and $v_{n+1} \neq \varepsilon$ as we have already seen. Thus the sequence

$$(\varepsilon, v_1, w_1), \ldots, (u_{n-1}, v_{n-1}, w_{n-1}), (u_n, \varepsilon, w'_n), (u_{n+1}, v_{n+1}, \varepsilon)$$

is admissible with respect to $y_1, \ldots, y_{n+1}$.

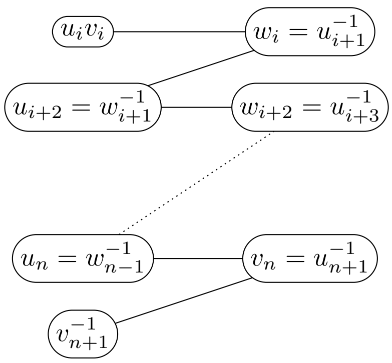


Figure 4.3: The graph $\mathcal{G}_X$.

Otherwise, let $s$ be the longest common suffix of $u_i v_i$ and $v_{n+1}^{-1}$.

There is a path in the incidence graph $\mathcal{G}_X$ from $u_i v_i$ to $v_{n+1}^{-1}$ (see Figure 4.3). By Proposition 4.1.3, $s$ is a proper suffix of $u_i v_i, w_{i+1}^{-1}, \ldots, w_{n-1}^{-1}, v_{n+1}^{-1}$. This implies that $s^{-1}$ is a proper prefix of $w_{i+1}, \ldots, w_{n-1}, v_{n+1}$.

It is not possible that $v_i$ is a suffix of $s$. Indeed, this would imply that $v_i^{-1}$ is a proper prefix of $w_{i+1}, \ldots, w_{n-1}, v_{n+1}$. But then we could change the $n - i + 1$ last terms of the sequence $(u_j, v_j, w_j)_{1 \leq j \leq n}$ into $(u_i, \varepsilon, v_i w_i)$,

$(u_{i+1}v_i^{-1}, \varepsilon, \rho(v_iw_{i+1})), \ldots, (\rho(u_nv_i^{-1}), v_iv_n, \varepsilon)$ resulting in an admissible sequence with a longer $v_n$.

Thus $s$ is a proper suffix of $v_i$. Since $s$ is a proper suffix of $v_i$ and $v_{n+1}^{-1}$, there are nonempty words $p, q \in A^*$ such that $v_i = ps$ and $v_{n+1}^{-1} = qs$. Moreover, the word $pq^{-1}$ is reduced since $s$ is the longest common suffix of $v_i$ and $v_{n+1}^{-1}$. Thus we can change the last $n - i + 2$ terms of the sequence formed by $(u_j, v_j, w_j)_{1 \le j \le n-1}$ followed by $(u_n, \varepsilon, v_n), (u_{n+1}, v_{n+1}, \varepsilon)$ into

$$(u_i, p, sw_i), (u_{i+1}s^{-1}, \varepsilon, \rho(sw_{i+1})), \ldots, (\rho(u_ns^{-1}), \varepsilon, sv_n), (u_{n+1}s^{-1}, q^{-1}, \varepsilon)$$

(see Figure 4.4).

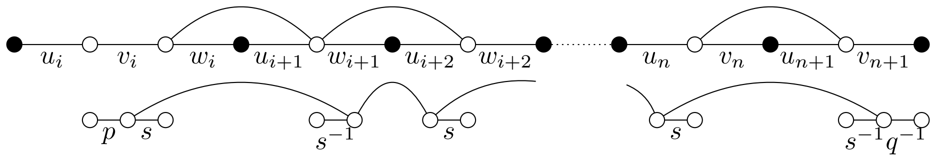


Figure 4.4: The word $y_i \cdots y_{n+1}$.

Since the word $pq^{-1}$ is reduced, the new sequence is admissible.

This shows that $y_1 \cdots y_n \not\equiv \varepsilon$ for any sequence $y_1, \ldots, y_n \in X \cup X^{-1}$ such that $y_iy_{i+1} \not\equiv \varepsilon$ for $1 \le i < n$. Thus $X$ is free. ■

We illustrate Theorem 4.1.1 in the following example.

**Example 4.1.8** Let $S$ be as in Example 3.1.5 and let $X = S \cap A^2$. We have

$$X = \{ab, ac, bc, ca, cd, da\}.$$

The set $X$ is an $S$-maximal bifix code. It is a basis of a subgroup of infinite index. Indeed, the minimal automaton of $X^*$ is represented in Figure 4.5 on the left. The Stallings automaton of the subgroup $H$ generated by $X$ is obtained by merging 3 with 4 and 2 with 5 (recall Section 1.3). It is represented in Figure 4.5 on the right. Since it is not a group automaton, the subgroup has infinite index (see Proposition 1.3.6).

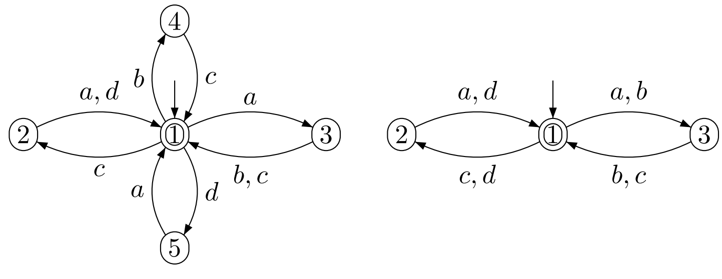


Figure 4.5: The minimal automaton of $X^*$ (on the left) and the Stallings automaton of $\langle X \rangle$ (on the right).

The set $X$ is a basis of $H$ by Theorem 4.1.1. This can also be seen by performing Nielsen transformations on the set $X$ (see [53] for example). Indeed, replacing $bc$ and $da$ by $bc(ac)^{-1}$ and $da(ca)^{-1}$, we obtain $X' = \{ab, ac, ba^{-1}, ca, cd, dc^{-1}\}$ which is Nielsen reduced. Thus $X'$ is a basis of $H$ and thus also $X$.

Note that, in agreement with Theorem 4.1.2, the two words of length 2 which are in $H$ but not in $X^*$, namely $bb$ and $dd$, are not in $S$.

Theorem 4.1.1 is false if $X$ is prefix but not bifix, as shown in the following example.

**Example 4.1.9** Let $S$ be the Fibonacci set and let $X \subset S$ be the prefix code $X = \{aa, ab, b\}$. Then $a = (ab)b^{-1}$ is in $\langle X \rangle$ and thus $X$ generates the free group on $A$. Thus $X$ is not a basis and $X^* \cap S$ is strictly included in $\langle X \rangle \cap S$ (for example $a \notin X^*$).

The proof of Theorem 4.1.1 proves not only that bifix codes in acyclic sets are free, but also that, in a sense made more precise below, the associated reductions are of low complexity.

We define the *height* of a word on $A \cup A^{-1}$ in the following recursive way. The reduced words (including the empty word) are the only words of height 0. The height of a word $w$ on $A \cup A^{-1}$ equivalent to $\varepsilon$ is the least integer $h$ such that $w$ is a concatenation of words of the form $w = uvu^{-1}$ where $u$ is a word on $A \cup A^{-1}$ and $v$ is a word of height $h - 1$ equivalent to $\varepsilon$.

We define the height of an arbitrary word $w$ on $A \cup A^{-1}$ as the least integer $h$ such that $w = z_0 v_1 z_1 \cdots v_n z_n$ with $z_0, \ldots, z_n$ equivalent to $\varepsilon$ of height at most $h$ and $v_1 \cdots v_n$ reduced.

In this way, any word on $A \cup A^{-1}$ has finite height. For example, the word $aa^{-1}cbb^{-1}$ has height 1 and $aaa^{-1}bb^{-1}a^{-1}$ has height 2.

**Proposition 4.1.10** *Let $S$ be an acyclic set and let $X \subset S$ be a bifix code. Any word $y = y_1 \cdots y_n$ with $y_i \in X \cup X^{-1}$ for $1 \le i \le n$ such that $y_i y_{i+1} \not\equiv \varepsilon$ for $1 \le i \le n-1$ has height at most 1.*

*Proof.* The proof of Theorem 4.1.1 shows that $y = z_0 v_1 z_1 \cdots z_{n-1} v_n z_n$ where

(i) $z_0, \ldots, z_n$ have height at most 1,

(ii) $v_1 \cdots v_n$ is reduced.

Thus $y$ has height at most 1. ∎

**Example 4.1.11** Let $X$ be as in Example 4.1.8. The word $bc(ac)^{-1}ab$, which reduces to $bb$, has height 1.

### 4.1.4 Saturation Theorem

We now give a proof of Theorem 4.1.2. It uses Proposition 4.1.7.

*Proof of the Saturation Theorem.* Let $S$ be an acyclic set and let $X \subset S$ be a bifix code. We have to prove that $X^* \cap S = \langle X \rangle \cap S$. Since $X^* \cap S \subset \langle X \rangle \cap S$, we only need to prove the reverse inclusion.

Consider the bifix code $Z$ generating the submonoid recognized by the coset automaton $\mathcal{B}_X$ associated to $X$. Set $Y = Z \cap S$. By Theorem 4.1.1, $Y$ is a basis of $\langle Y \rangle$.

By Proposition 4.1.7, we have $X \subset Z$ and thus $X \subset Y$.

Since any reversible automaton is minimal and since the automaton $\mathcal{B}_X$ is reversible by Proposition 4.1.7, it is equal to the minimal automaton of $Z^*$. Let $K$ be the subgroup generated by $Z$. By Proposition 1.3.2, we have $K \cap A^* = Z^*$.

This shows that

$$\langle X \rangle \cap S \subset K \cap S = K \cap A^* \cap S = Z^* \cap S = Y^* \cap S \subset Y^*.$$

The first inclusion holds because $X \subset Z$ implies $\langle X \rangle \subset K$. The last equality follows from the fact that if $z_1 \cdots z_n \in S$ with $z_1, \ldots, z_n \in Z$, then each $z_i$ is in $S$ (because $S$ is factorial) and hence in $Z \cap S = Y$. Thus $\langle X \rangle \cap S \subset Y^*$. Consider $x \in \langle X \rangle \cap S$. Then $x \equiv x_1 \cdots x_n$ with $x_i \in X \cup X^{-1}$. But since $\langle X \rangle \cap S \subset Y^*$, we have also $x = y_1 \cdots y_m$ with $y_i \in Y$. Since $X \subset Y$ and since $Y$ is free, this forces $n = m$ and $x_i = y_i$. Thus all $x_i$ are in $X$ and $x$ is in $X^*$. This shows that $\langle X \rangle \cap S \subset X^*$ which was to be proved. ∎

We note the following corollary of Theorem 4.1.2, which shows that bifix codes in acyclic sets satisfy a property which is stronger than being bifix (or more precisely that the submonoid $X^*$ satisfies a property stronger than being right and left unitary).

**Corollary 4.1.12** *Let $S$ be an acyclic set, let $X \subset S$ be a bifix code and let $H = \langle X \rangle$. For any $u, v \in S$,*

*(i) if $u, uv \in H \cap S$, then $v \in X^*$,*

*(ii) if $v, uv \in H \cap S$, then $u \in X^*$.*

*Proof.* Assume that $u, uv \in H \cap S$. Since $v \equiv u^{-1}(uv)$, we have $v \in H$. But $v \in H \cap S$ implies $v \in X^*$ by Theorem 4.1.2. This proves (i). The proof of (ii) is symmetric. ∎

We can express Corollary 4.1.12 in a different way. Let $S$ be an acyclic set and let $X \subset S$ be a bifix code. Then no nonempty word of $\langle X \rangle$ can be a proper prefix (or suffix) of a word of $X$. Indeed, assume that $u \in \langle X \rangle$ is a prefix of a word of $X$. Then $u$ is in $\langle X \rangle \cap S$ and thus in $X^*$ since $X^*$ is saturated in $S$. This implies $u = \varepsilon$ or $u \in X$.

## 4.2 Finite index basis property

In this Section we study the connection between tree sets and subgroups of the free set. The main result of the Section, namely the Finite Index Basis Theorem, is given in Section 4.2.1. In the same section we also show a converse of this theorem (Corollary 4.2.6).

In Section 4.2.2 we define tame bases and prove that in a recurrent tree set of characteristic 1 any basis of the free group is tame (Theorem 4.2.11).

Finally, we define in Section 4.2.3 $\mathcal{S}$-adic representations and show that every recurrent tree set of characteristic 1 has a primitive $\mathcal{S}$-adic representation with $\mathcal{S}$ finite and containing positive automorphisms only.

### 4.2.1 The Finite Index Basis Theorem

Let $S$ be a recurrent set containing the alphabet $A$. We say that $S$ has the *finite index basis property* if the following holds. A finite bifix code $X \subset S$ is an $S$-maximal bifix code of $S$-degree $d$ if and only if it is a basis of a subgroup of index $d$ of the free group on $A$.

We refer to the nex result as the Finite Index Basis Theorem.

**Theorem 4.2.1 (Finite Index Basis Theorem)** *A recurrent tree set of characteristic* 1 *has the finite index basis property.*

Note that the Cardinality Theorem (Theorem 2.2.1) holds for a set $S$ satisfying the finite index basis property. Indeed, by Schreier's formula a basis of a subgroup of index $d$ of a free group on $s$ generators has $(s-1)d+1$ elements. Since a tree set of characteristic 1 is in particular a neutral set of characteristic 1, the formula (1) of Theorem 2.2.1 is verified.

*Proof of the Finite Index Basis Theorem.* Let $S$ be a recurrent tree set of characteristic 1. Assume first that $X$ is a finite $S$-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 $\delta_X(u) = d$, or, equivalently, which is not an internal factor of $X$ (recall Section 1.2). Let $Q$ be the set formed of the $d$ suffixes of $u$ which are in $P$.

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 $p, p' \in Q$, one has $pv, p'v \in Hq$ for some $q \in Q$. Then $qv^{-1}$ is in $Hp \cap Hp'$ and thus $p = p'$ by the above argument.

The set $V$ is a subgroup of $F_A$. Indeed, $\varepsilon \in V$. 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 $\mathcal{R}_S(u)$, the set of right return words to $u$ in $S$, is contained in $V$. Indeed, let $q \in Q$ and $y \in \mathcal{R}_S(u)$. Since $q$ is a suffix of $u$, $qy$ is a suffix of $uy$, and since $uy$ is in $S$ (by definition of $\mathcal{R}_S(u)$), also $qy$ is in $S$. Since $X$

is an $S$-maximal bifix code, it is an $S$-maximal prefix code and thus it is right $S$-complete (recall Section 1.2). This implies that $qy$ is a prefix of a word in $X^*$ and thus there is a word $r \in P$ such that $qy \in X^*r$. We verify that the word $r$ is a suffix of $u$. Since $y \in \mathcal{R}_S(u)$, there is a word $y'$ such that $uy = y'u$. Consequently, $r$ is a suffix of $y'u$, and in fact the word $r$ is a suffix of $u$. Indeed, one has $|r| \le |u|$ since otherwise $u$ is in the set $I(X)$ of internal factors of $X$, and this is not the case. Thus we have $r \in Q$ (see Figure 4.6). Since $X^* \subset H$ and $r \in Q$, we have $qy \in HQ$. Thus $y \in V$.

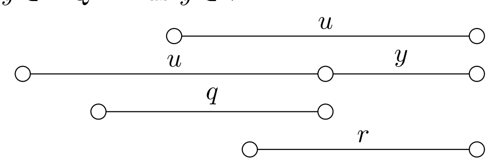


Figure 4.6: A word $y \in \mathcal{R}_S(u)$.

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 3.2.5, since $S$ is acyclic, this implies $t \in X^*$ and thus $t = \varepsilon$. Thus $p = q$. Let

$$V = \{v \in F_A \mid Qv \subset HQ\}.$$

By Theorem 3.2.5, the group generated by $\mathcal{R}_S(u)$ is the free group on $A$. Since $\mathcal{R}_S(u) \subset V$, and since $V$ is a subgroup of $F_A$, we have $V = F_A$. Thus $Qw \subset HQ$ for any $w \in F_A$. Since $\varepsilon \in Q$, we have in particular $w \in HQ$. Thus $F_A = HQ$. 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 $d$. Since $S$ is a recurrent neutral set, by Theorem 2.2.1, we have $\mathrm{Card}(X) = d(\mathrm{Card}(A) - 1) + 1$. In view of Schreier's Formula, this implies that $X$ is a basis of $H$.

Assume conversely that the finite bifix code $X \subset S$ is a basis of the group $H = \langle X \rangle$ and that $H$ has index $d$. Since $X$ is a basis of $H$, by Schreier's Formula, we have $\mathrm{Card}(X) = (\mathrm{Card}(A) - 1)d + 1$. The case $\mathrm{Card}(A) = 1$ is straightforward, thus we assume $\mathrm{Card}(A) \ge 2$. Recall that a recurrent tree set is uniformly recurrent (Corollary 2.2.9). By [7, Theorem 4.4.3], if $S$ is a uniformly recurrent set, any finite bifix code contained in $S$ is contained in a finite $S$-maximal bifix code. Thus there is a finite $S$-maximal bifix code $Y$ containing $X$. Let $e$ be the $S$-degree of $Y$. By the first part of the proof, $Y$ is a basis of a subgroup $K$ of index $e$ of the free group on $A$. In particular, it has $(\mathrm{Card}(A) - 1)e + 1$ elements. Since $X \subset Y$, we have $(\mathrm{Card}(A) - 1)d + 1 \le (\mathrm{Card}(A) - 1)e + 1$ and thus $d \le e$. On the other hand, since $H$ is included in $K$, $d$ is a multiple of $e$ and thus $e \le d$. We conclude that $d = e$ and thus that $X = Y$. ■

The following examples shows that Theorem 4.2.1 may be false for a set $S$ which does not satisfy some of the hypotheses.

The first example is a recurrent set which is not neutral.

**Example 4.2.2** Let $S$ be the Chacon set (see Example 1.1.3). We have seen that $S$ is not neutral and thus not a tree set. The set $S \cap A^2 = \{aa, ab, bc, ca, cb\}$ is an $S$-maximal bifix code of $S$-degree 2. It is not a basis since $ca(aa)^{-1}ab = cb$. Thus $S$ does not satisfy the finite index basis property.

In the second example, the set is neutral but not a tree set and is not recurrent.

**Example 4.2.3** Let $S$ be the set of Example 3.1.7. It is not a tree set (and it is not either recurrent). The set $S \cap A^2$ is the same as in the Chacon set. Thus $S$ does not satisfy the finite index basis property.

In the last example we have a recurrent set which is neutral but not a tree set.

**Example 4.2.4** Let $S$ be the set on the alphabet $B = \{1, 2, 3\}$ of Example 3.1.6. We have seen that $S$ is neutral but not a tree set.

Let $X = S \cap B^2$. We have $X = \{12, 13, 22, 23, 31\}$. The set $X$ is not a basis since $13 = 12(22)^{-1}23$. Thus $S$ does not satisfy the finite index basis property.

We close this section with a converse of Theorem 4.2.1.

**Proposition 4.2.5** *A biextendable set $S$ such that $S \cap A^n$ is a basis of the subgroup $\langle A^n \rangle$ for all $n \geq 1$ is a tree set of characteristic* 1*.*

*Proof.* Set $k = \mathrm{Card}(A) - 1$. Since $A^n$ generates a subgroup of index $n$, the hypothesis implies that $\mathrm{Card}(A^n \cap S) = kn + 1$ for all $n \geq 1$. Consider $w \in S$ and set $m = |w|$. The set $X = AwA \cap S$ is included in $Y = S \cap A^{m+2}$. Since $Y$ is a basis of a subgroup, $X \subset Y$ is a basis of the subgroup $\langle X \rangle$.

This implies that the graph $\mathcal{E}_S(w)$ is acyclic. Indeed, assume that the path $(a_1, b_1, \ldots, a_p, b_p, a_1)$ is a cycle in $\mathcal{E}_S(w)$ with $p \geq 2$, $a_i \in L_S(w)$, $b_i \in R_S(w)$ 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

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

contradicting the fact that $X$ is a basis.

Since $\mathcal{E}_S(w)$ is an acyclic graph with $\ell_S(w) + r_S(w)$ vertices and $b_S(w)$ edges, we have $b_S(w) \leq \ell_S(w) + r_S(w) - 1$. But then

$$\begin{aligned} \mathrm{Card}(A^{m+2} \cap S) = \sum_{w \in A^m \cap S} b_S(w) \quad &\leq \quad \sum_{w \in A^m \cap S} (\ell_S(w) + r_S(w) - 1) \\ &\leq \quad 2\,\mathrm{Card}(A^{m+1} \cap S) - \mathrm{Card}(A^m \cap S) \\ &\leq \quad k(m+2) + 1. \end{aligned}$$

Since $\mathrm{Card}(A^{m+2} \cap S) = k(m+2) + 1$, we have $b_S(w) = \ell_S(w) + r_S(w) - 1$ for all $w \in A^m$. This implies that $\mathcal{E}_S(w)$ is a tree for all $w \in S$, including the emptyword. Thus $S$ is a tree set of characteristic 1. ∎

**Corollary 4.2.6** *A recurrent set which has the finite index basis property is a tree set of characteristic* 1*.*

*Proof.* Let $S$ be a recurrent set having the finite index basis property. For any $n \geq 1$, the set $S \cap A^n$ is an $S$-maximal bifix code of $S$-degree $n$ (Example 1.2.2). Thus it is a basis of a subgroup of index $n$. Since it is included in the subgroup generated by $A^n$, which has index $n$, it is a basis of this subgroup. This implies that $S$ is a tree set by Proposition 4.2.5. ■

### 4.2.2 Tame bases

An automorphism $\alpha$ of the free group on $A$ is called *positive* if $\alpha(a) \in A^+$ for every $a \in A$. We say that a positive automorphism of the free group on $A$ is *tame* [1] if it belongs to the submonoid generated by the permutations of $A$ and the automorphisms $\alpha_{a,b}$, $\tilde{\alpha}_{a,b}$ defined for $a, b \in A$ with $a \neq b$ by

$$\alpha_{a,b}(c) = \begin{cases} ab & \text{if } c = a, \\ c & \text{otherwise} \end{cases} \quad \text{and} \quad \tilde{\alpha}_{a,b}(c) = \begin{cases} ba & \text{if } c = a, \\ c & \text{otherwise.} \end{cases}$$

Thus $\alpha_{a,b}$ places a letter $b$ after each $a$ and $\tilde{\alpha}_{a,b}$ places a letter $b$ before each $a$. The above automorphisms and the permutations of $A$ are called the *elementary* positive automorphisms on $A$. The monoid of positive automorphisms is not finitely generated as soon as the alphabet has at least three generators (see [64]).

A basis $X$ of the free group is *positive* if $X \subset A^+$. A positive basis $X$ of the free group is *tame* if there exists a tame automorphism $\alpha$ such that $X = \alpha(A)$.

**Example 4.2.7** The set $X = \{ba, cba, cca\}$ is a tame basis of the free group on $\{a, b, c\}$. Indeed, one has the following sequence of elementary automorphisms.

$$(b, c, a) \overset{\alpha_{c,b}}{\to} (b, cb, a) \overset{\tilde{\alpha}^2_{a,c}}{\to} (b, cb, cca) \overset{\alpha_{b,a}}{\to} (ba, cba, cca).$$

The fact that $X$ is a basis can be checked directly by the fact that $(cba)(ba)^{-1} = c$, $c^{-2}(cca) = a$ and finally $(ba)a^{-1} = b$.

The following result will play a key role in the proof of the main result of this section (Theorem 4.2.11).

**Proposition 4.2.8** *A set $X \subset A^+$ is a tame basis of the free group on $A$ if and only if $X = A$ or there is a tame basis $Y$ of the free group on $A$ and $u, v \in Y$ such that $X = (Y \setminus v) \cup uv$ or $X = (Y \setminus u) \cup uv$.*

[1] The word *tame* (as opposed to *wild*) is used here on analogy with its use in ring theory (see [24]). The tame automorphisms as introduced here should, strictly speaking, be called *positive tame* automophisms since the group of all automorphisms, positive or not, is tame in the sense that it is generated by the elementary automorphisms.

*Proof.* Assume first that $X$ is a tame basis of the free group on $A$. Then $X = \alpha(A)$ where $\alpha$ is a tame automorphism of $\langle A \rangle$. Then $\alpha = \alpha_1\alpha_2 \cdots \alpha_n$ where the $\alpha_i$ are elementary positive automorphisms. We use an induction on $n$. If $n = 0$, then $X = A$. If $\alpha_n$ is a permutation of $A$, then $X = \alpha_1\alpha_2 \cdots \alpha_{n-1}(A)$ and the result holds by induction hypothesis. Otherwise, set $\beta = \alpha_1 \cdots \alpha_{n-1}$ and $Y = \beta(A)$. By induction hypothesis, $Y$ is tame. If $\alpha_n = \alpha_{a,b}$, set $u = \beta(a)$ and $v = \beta(b) = \alpha(b)$. Then $X = (Y \setminus u) \cup uv$ and thus the condition is satisfied. The case were $\alpha_n = \tilde{\alpha}_{a,b}$ is symmetrical.

Conversely, assume that $Y$ is a tame basis and that $u, v \in Y$ are such that $X = (Y \setminus u) \cup uv$. Then, there is a tame automorphism $\beta$ of $\langle A \rangle$ such that $Y = \beta(A)$. Set $a = \beta^{-1}(u)$ and $b = \beta^{-1}(v)$. Then $X = \beta\alpha_{a,b}(A)$ and thus $X$ is a tame basis. ∎

We note the following corollary.

**Corollary 4.2.9** *A tame basis of the free group which is a bifix code is the alphabet.*

*Proof.* Assume that $X$ is a tame basis which is not the alphabet. By Proposition 4.2.8 there is a tame basis $Y$ and $u, v \in Y$ such that $X = (Y \setminus v) \cup uv$ or $X = (Y \setminus u) \cup uv$. In the first case, $X$ is not prefix. In the second one, it is not suffix. ∎

The following example is from [64].

**Example 4.2.10** The set $X = \{ab, acb, acc\}$ is a basis of the free group on $\{a, b, c\}$. Indeed, $accb = (acb)(ab)^{-1}(acb) \in \langle X \rangle$ and thus $b = (acc)^{-1}accb \in \langle X \rangle$, which implies easily that $a, c \in \langle X \rangle$. The set $X$ is bifix and thus it is not a tame basis by Corollary 4.2.9.

The following result is a remarkable consequence of Theorem 4.2.1.

**Theorem 4.2.11** *Any basis of the free group included in a recurrent tree set is tame.*

*Proof.* Let $S$ be a recurrent tree set. Let $X \subset S$ be a basis of the free group on $A$. Since $A$ is finite, $X$ is finite (and of the same cardinality as $A$). We use an induction on the sum $\lambda(X)$ of the lengths of the words of $X$. If $X$ is bifix, by Theorem 4.2.1, it is an $S$-maximal bifix code of $S$-degree 1. Thus $X = A$ (see Example 1.2.3). Next assume for example that $X$ is not prefix. Then there are nonempty words $u, v$ such that $u, uv \in X$. Let $Y = (X \setminus uv) \cup v$. Then $Y$ is a basis of the free group and $\lambda(Y) < \lambda(X)$. By induction hypothesis, $Y$ is tame. Since $X = (Y \setminus v) \cup uv$, $X$ is tame by Proposition 4.2.8. ∎

**Example 4.2.12** The set $X = \{ab, acb, acc\}$ is a basis of the free group which is not tame (see Example 4.2.10). Accordingly, the extension graph $\mathcal{E}_X(\varepsilon)$ relative to the set of factors of $X$ is not a tree (see Figure 4.7).

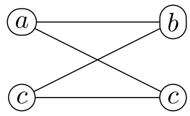


Figure 4.7: The graph $\mathcal{E}_X(\varepsilon)$.

### 4.2.3 $\mathcal{S}$-adic representations

In this section we study $\mathcal{S}$-adic representations of tree sets. This notion was introduced in [38], using a terminology initiated by Vershik and coined out by B. Host. We first recall a general construction allowing to build $\mathcal{S}$-adic representations of any recurrent aperiodic set (Proposition 4.2.14) which is based on return words. Using Theorem 4.2.11, we show that this construction actually provides $\mathcal{S}_e$-representations of recurrent tree sets (Theorem 4.2.15), where $\mathcal{S}_e$ is the set of elementary positive automorphisms of the free group on $A$.

Let $\mathcal{S}$ be a set of morphisms and $\mathbf{h} = (\sigma_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{S}^{\mathbb{N}}$ with $\sigma_n : A_{n+1}^* \to A_n^*$ and $A_0 = A$.

We let $T_{\mathbf{h}}$ denote the set of words $\bigcap_{n\in\mathbb{N}} \mathrm{Fac}(\sigma_0 \cdots \sigma_n(A_{n+1}^*))$. We call a factorial set $T$ an *$\mathcal{S}$-adic set* if there exists $\mathbf{h} \in \mathcal{S}^{\mathbb{N}}$ such that $T = T_{\mathbf{h}}$. In this case, the sequence $\mathbf{h}$ is called an *$\mathcal{S}$-adic representation* of $T$.

**Example 4.2.13** Any Arnoux-Rauzy set is $\mathcal{S}$-adic with a finite set $\mathcal{S}$. This results from the fact that any Arnoux-Rauzy word is obtained by iterating a sequence of morphism of the form $\psi_a$ for $a \in A$ defined by $\psi_a(a) = a$ and $\psi_a(b) = ab$ for $b \neq a$ (see [3] or [7]).

A sequence of morphisms $(\sigma_n)_{n\in\mathbb{N}}$ is said to be *everywhere growing* if $\min_{a\in A_n} |\sigma_0 \cdots \sigma_{n-1}(a)|$ goes to infinity as $n$ increases. A sequence of morphisms $(\sigma_n)_{n\in\mathbb{N}}$ is said to be *primitive* if for all $r \geq 0$ there exists $s > r$ such that all letters of $A_r$ occur in all images $\sigma_r \cdots \sigma_{s-1}(a)$, $a \in A_s$. Obviously any primitive sequence of morphisms is everywhere growing.

A uniformly recurrent set $T$ is said to be *aperiodic* if it contains at least one right-special factor of each length. The next (well-known) proposition provides a general construction to get a primitive $\mathcal{S}$-adic representation of any aperiodic uniformly recurrent set $T$.

It complements the main result of [38] asserting that any minimal symbolic system on a finite alphabet $A$ with at most linear factor complexity has an everywhere growing $\mathcal{S}$-adic representation with $\mathcal{S}$ finite.

**Proposition 4.2.14** *An aperiodic factorial set $T \subset A^*$ is uniformly recurrent if and only if it has a primitive $\mathcal{S}$-adic representation for some (possibly infinite) set $\mathcal{S}$ of morphisms.*

*Proof.* Let $\mathcal{S}$ be a set of morphisms and $\mathbf{h} = (\sigma_n : A_{n+1}^* \to A_n^*)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ be a primitive sequence of morphisms such that $T = \bigcap_{n\in\mathbb{N}} \mathrm{Fac}(\sigma_0 \cdots \sigma_n(A_{n+1}^*))$. Consider a word $u \in T$ and let us prove that $u \in \mathrm{Fac}(v)$ for all long enough $v \in T$. The sequence $\mathbf{h}$ being everywhere growing, there is an integer $r > 0$

such that $\min_{a\in A_r} |\sigma_0\cdots\sigma_{r-1}(a)| > |u|$. As $T = \bigcap_{n\in\mathbb{N}} \mathrm{Fac}(\sigma_0\cdots\sigma_n(A_{n+1}^*))$, there is an integer $s > r$, two letters $a, b \in A_r$ and a letter $c \in A_s$ such that $u \in \mathrm{Fac}(\sigma_0\cdots\sigma_{r-1}(ab))$ and $ab \in \mathrm{Fac}(\sigma_r\cdots\sigma_{s-1}(c))$. The sequence $\mathbf{h}$ being primitive, there is an integer $t > s$ such that $c$ occurs in $\sigma_s\cdots\sigma_{t-1}(d)$ for all $d \in A_t$. Thus $u$ is a factor of all words $v \in T$ such that $|v| \geq 2\max_{d\in A_t} |\sigma_0\cdots\sigma_{t-1}(d)|$ and $T$ is uniformly recurrent.

Let us prove the converse. Let $(u_n)_{n\in\mathbb{N}} \in T^{\mathbb{N}}$ be a non-ultimately periodic sequence such that $u_n$ is suffix of $u_{n+1}$. By assumption, $T$ is uniformly recurrent so $\mathcal{R}_T(u_{n+1})$ is finite for all $n$. The set $T$ being aperiodic, $\mathcal{R}_T(u_{n+1})$ also has cardinality at least 2 for all $n$. For all $n$, let $A_n = \{0, \ldots, \mathrm{Card}(\mathcal{R}_T(u_n))-1\}$ and let $\alpha_n : A_n^* \to A^*$ be a coding morphism for $\mathcal{R}_T(u_n)$. The word $u_n$ being suffix of $u_{n+1}$, we have $\alpha_{n+1}(A_{n+1}) \subset \alpha_n(A_n^+)$. Since $\alpha_n(A_n) = \mathcal{R}_T(u_n)$ is a prefix code, there is a unique morphism $\sigma_n : A_{n+1}^* \to A_n^*$ such that $\alpha_n\sigma_n = \alpha_{n+1}$. For all $n$ we get $\mathcal{R}_T(u_n) = \alpha_0\sigma_0\sigma_1\cdots\sigma_{n-1}(A_n)$ and $T = \bigcap_{n\in\mathbb{N}} \mathrm{Fac}(\alpha_0\sigma_0\cdots\sigma_n(A_{n+1}^*))$. Without loss of generality, we can suppose that $u_0 = \varepsilon$ and $A_0 = A$. In that case we get $\alpha_0 = \mathrm{id}$ and the set $\mathcal{S}$ thus has an $\mathcal{S}$-adic representation with $\mathcal{S} = \{\sigma_n \mid n \in \mathbb{N}\}$.

Let us show that $\mathbf{h} = (\sigma_n)_{n\in\mathbb{N}}$ is everywhere growing. If not, there is a sequence of letters $(a_n \in A_n)_{n\geq N}$ such that $\sigma_n(a_{n+1}) = a_n$ for all $n \geq N$ for some $N \geq 1$. This means that the word $v = \sigma_0\cdots\sigma_n(a_n) \in T$ is a return word to $u_n$ for all $n \geq N$. The sequence $(|u_n|)_{n\in\mathbb{N}}$ being unbounded, the word $v^k$ belongs to $T$ for all positive integers $k$, which contradicts the uniform recurrence of $T$.

Let us show that $\mathbf{h}$ is primitive. The set $T$ being uniformly recurrent, for all $n \in \mathbb{N}$ there exists $N_n$ such that all words of $T \cap A^{\leq n}$ occur in all words of $T \cap A^{\geq N_n}$. Let $r \in \mathbb{N}$ and let $u = \sigma_0\cdots\sigma_{r-1}(a)$ for some $a \in A_r$. Let $s > r$ be an integer such that $\min_{b\in A_s} |\sigma_0\cdots\sigma_{s-1}(b)| \geq N_{|u|}$. Thus $u$ occurs in $\sigma_0\cdots\sigma_{s-1}(b)$ for all $b \in A_s$. As $\sigma_0\cdots\sigma_{s-1}(A_s) \subset \sigma_0\cdots\sigma_{r-1}(A_r^+)$ and as $\sigma_0\cdots\sigma_{r-1}(A_r) = \mathcal{R}_T(u_r)$ is a prefix code, the letter $a \in A_r$ occurs in $\sigma_r\cdots\sigma_{s-1}(b)$ for all $b \in A_r$. ■

Note that even for uniformly recurrent sets with linear factor complexity, the set of morphisms $\mathcal{S} = \{\sigma_n \mid n \in \mathbb{N}\}$ considered in Proposition 4.2.14 is usually infinite as well as the sequence of alphabets $(A_n)_{n\in\mathbb{N}}$ is usually unbounded (see [37]). For tree sets $T$, the next theorem significantly improves the only if part of Proposition 4.2.14. For such sets, the set $\mathcal{S}$ can be replaced by the set $\mathcal{S}_e$ of elementary positive automorphisms. In particular, $A_n$ is equal to $A$ for all $n$. The following theorem also improve the main result of [38], because under our hypothesi, we obtain the primitivity of the representation.

**Theorem 4.2.15** *If $T$ is a recurrent tree set of characteristic 1 over an alphabet $A$, then it has a primitive $\mathcal{S}_e$-adic representation.*

*Proof.* For any non-ultimately periodic sequence $(u_n)_{n\in\mathbb{N}} \in T^{\mathbb{N}}$ such that $u_0 = \varepsilon$ and $u_n$ is suffix of $u_{n+1}$, the sequence of morphisms $(\sigma_n)_{n\in\mathbb{N}}$ built in the proof of Proposition 4.2.14 is a primitive $\mathcal{S}$-adic representation of $T$ with $\mathcal{S} = \{\sigma_n \mid$

$n \in \mathbb{N}\}$. Therefore, all we need to do is to consider such a sequence $(u_n)_{n\in\mathbb{N}}$ such that $\sigma$ is tame for all $n$.

Let $u_1 = a^{(0)}$ be a letter in $A$. Set $A_0 = A$ and let $\sigma_0 : A_1^* \to A_0^*$ be a coding morphism for $\mathcal{R}_T(u_1)$. By Theorem 3.2.5, the set $\mathcal{R}_T(u_1)$ is a basis of the free group on $A$. By Theorem 4.2.11, the morphism $\sigma_0 : A_1^* \to A_0^*$ is tame ($A_0 = A$). Let $a^{(1)} \in A_1$ be a letter and set $u_2 = \sigma_0(a^{(1)})$. Thus $u_2 \in \mathcal{R}_T(u_1)$ and $u_1$ is a suffix of $u_2$. By Theorem 3.2.9, the derived set $T^{(1)} = \sigma_0^{-1}(T)$ is a (uniformly) recurrent tree set on the alphabet $A$. We thus reiterate the process with $a^{(1)}$ and we conclude by induction with $u_n = \sigma_0 \cdots \sigma_{n-2}(a^{(n-1)})$ for all $n \geq 2$. ■

The converse of Theorem 4.2.15 is not true, as shown by Example 4.2.16 below.

**Example 4.2.16** Let $A = \{a, b, c\}$ and let $f : a \mapsto ac, b \mapsto bac, c \mapsto cb$. The set $S$ of factors of the fixed point $f^\omega(a)$ is not a tree set since $bb, bc, cb, cc \in S$ and thus $G_S(\varepsilon)$ has a cycle although $f$ is a tame automorphism since $f = \alpha_{a,c}\alpha_{c,b}\alpha_{b,a}$.

In the case of a ternary alphabet, a characterization of tree sets by their $\mathcal{S}$-adic representation can be proved [50], showing that there exists a Büchi automaton on the alphabet $\mathcal{S}_e$ recognizing the set of $\mathcal{S}$-adic representations of recurrent tree sets.

## 4.3 Bifix decoding of tree sets

In this Section we introduce several results concerning maximal bifix decoding. In Section 4.3.1 we prove that the family of acyclic sets are closed under maximal bifix decoding (Theorem 4.3.1), and that the same closure property is true for tree set, provided the original set is recurrent (Theorem 4.3.3). Moreover, we focus in the case of a tree set of characteristic 1, proving that in that case the recurrence is preserved (Theorem 4.3.5).

Composition of bifix codes is treated in Section 4.3.2, while modular codes are introduced in Section 4.3.3. In this last section we also consider the case of a decoding under a special family of maximal bifix codes (Theorem 4.3.17).

### 4.3.1 Maximal bifix decoding

In this section we prove the counterpart of Theorem 2.3.1 for acyclic and tree sets.

Recall, from Section 1.2 that given a coding morphism $f$ for a finite ($S$-maximal) bifix code $X \subset S$, the set $f^{-1}(S)$ is a (maximal) bifix decoding of $S$.

**Theorem 4.3.1** *Any biextendable set which is the bifix decoding of an acyclic set is acyclic.*

*Proof.* Let $S$ be an acyclic set and let $f : B^* \to A^*$ be a coding morphism for a finite bifix code $X \subset S$ such that $f^{-1}(S)$ is biextendable. Let $u \in f^{-1}(S)$ and let $v = f(u)$. Since $X$ is a finite bifix code, it is both a suffix code and a prefix code. Thus the generalized extension graph $\mathcal{E}^{X,X}(v)$ is acyclic by Proposition 3.1.11. Since $\mathcal{E}(u)$ is isomorphic with $\mathcal{E}^{X,X}(v)$, it is also acyclic. Thus $f^{-1}(S)$ is acyclic. ∎

The previous statement is not satisfactory because of the assumption that $f^{-1}(S)$ is biextendable which is added to obtain the conclusion. The following example shows that the condition is necessary.

**Example 4.3.2** Let $S$ be the Fibonacci set and let $f$ be the coding morphism for $X = \{aa, ab\}$ defined by $f(u) = aa$, $f(v) = ab$. Then $f^{-1}(S)$ is the finite set $\{u, v, vu, vv, vvu\}$ and thus not biextendable. Note however that for any $w \in f^{-1}(S)$, the graph $\mathcal{E}(w)$ is acyclic.

One may verify that a sufficient condition for $f^{-1}(S)$ to be biextendable is that $X$ is an $S$-maximal prefix code and an $S$-maximal suffix code (when $S$ is recurrent, this is equivalent to the fact that $X$ is an $S$-maximal bifix code).

The following result is a consequence of Proposition 3.1.13.

**Theorem 4.3.3** *Any maximal bifix decoding of a recurrent tree set is a tree set with the same characteristic.*

*Proof.* Let $S$ be a recurrent tree set of characteristic $c$ and let $f : B^* \to A^*$ be a coding morphism for a finite $S$-maximal bifix code $X$. By definition $S$ is acyclic. By Theorem 4.3.1, the set $U = f^{-1}(S)$ is also acyclic. From Proposition 2.3.2, we have that $m_U(f^{-1}(w)) = m_S^{X,X}(w) = m_S(w)$ for every $w \in S$. Thus $m_U(u) = 0$ for every nonempty word $u$ and $m_U(\varepsilon) = \chi(S)$. By an elementary result of graph theory it follows that $\mathcal{E}_U(u)$ is a tree for every nonempty $u \in U$ and $\mathcal{E}_U(\varepsilon)$ is a forest of $\chi(S)$ trees. Hence $U$ is a tree set of characteristic $\chi(U) = \chi(S)$. ∎

**Example 4.3.4** Let $S$ be the Fibonacci set and let $X = A^2 \cap S = \{aa, ab, ba\}$. Let $B = \{u, v, w\}$ and let $f$ be the coding morphism for $X$ defined by $f(u) = aa$, $f(v) = ab$ and $f(w) = ba$. Then the set $f^{-1}(S)$ is a recurrent tree set (we will see in Chapter 6 that it is actually a regular interval exchange set).

Note that, in general, the maximal bifix decoding of a recurrent tree set is not recurrent anymore. Anyway, for tree sets of characteristic 1 we can prove a stronger result.

**Theorem 4.3.5** *The family of recurrent tree sets of characteristic* 1 *is closed under maximal bifix decoding.*

In Chapter 6 we will see an analogous for the family of planar tree sets of characteristic 1 (Corllary 6.2.13). Another important result concerning maximal bifix decoding of tree sets is given at the end of the section (Theorem 4.3.17).

In order to prove Theorem 4.3.5, we prove first some preliminary results concerning the restriction to a recurrent tree set of a morphism onto a finite group (Propositions 4.3.7 and 4.3.9). Recall from Section 1.3 that a group code of degree $d$ is a bifix code $X$ such that $X^* = \varphi^{-1}(H)$ for a surjective morphism $\varphi : A^* \to G$ from $A^*$ onto a finite group $G$ and a subgroup $H$ of index $d$ of $G$.

The following result is stated for an Arnoux-Rauzy set $S$ in [7, Theorem 7.2.5] but the proof only uses the fact that $S$ is uniformly recurrent and satisfies the finite index basis property. We reproduce the proof for the sake of clarity.

**Theorem 4.3.6** *Let $Z \subset A^+$ be a group code of degree $d$. For every recurrent tree set $S$ of characteristic 1, the set $X = Z \cap S$ is a basis of a subgroup of index $d$ of $F_A$.*

*Proof.* By [7, Theorem 4.2.11], the code $X$ is an $S$-maximal bifix code of $S$-degree $e \leq d$. Since $S$ is recurrent, by [7, Theorem 4.4.3], $X$ is finite. By Theorem 4.2.1, $X$ is a basis of a subgroup of index $e$. Since $\langle X \rangle \subset \langle Z \rangle$, the index $e$ of the subgroup $\langle X \rangle$ is a multiple of the index $d$ of the subgroup $\langle Z \rangle$. Since $e \leq d$, this implies that $e = d$. ∎

As an example of this result, if $S$ is a recurrent tree set, then $S \cap A^n$ is a basis of the subgroup of the free group which is the kernel of the morphism onto $\mathbb{Z}/n\mathbb{Z}$ sending any letter to 1.

**Proposition 4.3.7** *Let $S$ be a recurrent tree set of characteristic 1 and let $\varphi : A^* \to G$ be a morphism from $A^*$ onto a finite group $G$. Then $\varphi(S) = G$.*

*Proof.* Let $1_G$ be the identity element of the group $G$. Since the submonoid $\varphi^{-1}(1_G)$ is right and left unitary, there is a bifix code $Z$ such that $Z^* = \varphi^{-1}(1_G)$. Let $X = Z \cap S$. By Theorem 4.3.6, $X$ is a basis of a subgroup of index $\mathrm{Card}(G)$. Let $x$ be a word of $X$ of maximal length (since $X$ is a basis of a subgroup of finite index, it is finite). Then $x$ is not an internal factor of $X$ and thus it has $\mathrm{Card}(G)$ parses. Let $S(x)$ be the set of suffixes of $x$ which are prefixes of $X$. If $s, t \in S(x)$, then they are comparable for the suffix order. Assume for example that $s = ut$. If $\varphi(s) = \varphi(t)$, then $u \in X^*$ which implies $u = \varepsilon$ since $s$ is a prefix of $X$. Thus all elements of $S(x)$ have distinct images by $\varphi$. Since $S(x)$ has $\mathrm{Card}(G)$ elements, this forces $\varphi(S(x)) = G$ and thus $\varphi(S) = G$ since $S(x) \subset S$. ∎

We illustrate the proof on the following example.

**Example 4.3.8** Let $A = \{a, b\}$ and let $\varphi$ be the morphism from $A^*$ onto the symmetric group $G$ on 3 elements defined by $\varphi(a) = (12)$ and $\varphi(b) = (13)$. We denote by (1) the identity permutation. Let $Z$ be the group code such that $Z^* = \varphi^{-1}\big((1)\big)$. The group automaton corresponding to the regular representation of $G$ is represented in Figure 4.8 (this automaton has $G$ as set of states and $g \cdot a = g\varphi(a)$ for every $g \in G$ and $a \in A$).

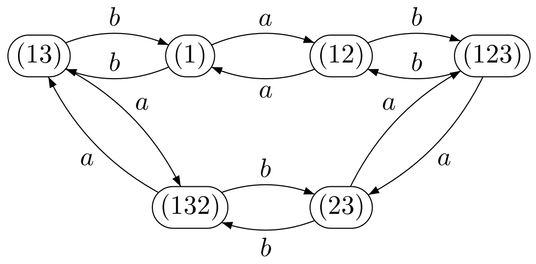


Figure 4.8: The group automaton corresponding to the regular representation of $G$.

Let $S$ be the Fibonacci set. The code $X = Z \cap S$ is represented in Figure 4.9.

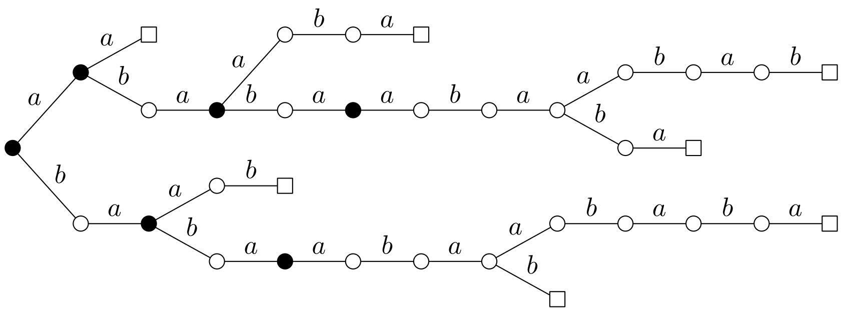


Figure 4.9: The code $X = Z \cap S$.

The word $w = ababa$ is not an internal factor of $X$. All its 6 suffixes (indicated in black in Figure 4.9) are proper prefixes of $X$ and their images by $\varphi$ are the 6 elements of the group $G$.

**Proposition 4.3.9** *Let $S$ be a recurrent tree set of characteristic* 1 *and let $\varphi : A^* \to G$ be a morphism from $A^*$ onto a finite group $G$. For any $w \in S$, one has $\varphi(\Gamma_S(w) \cup \{\varepsilon\}) = G$.*

*Proof.* Let $\alpha : B^* \to A^*$ be a coding morphism for $\mathcal{R}_S(w)$. Then $\beta = \varphi \circ \alpha : B^* \to G$ is a morphism from $B^*$ to $G$. By Theorem 3.2.5, the set $\mathcal{R}_S(w)$ is a basis of the free group on $A$. Thus $\langle \alpha(B) \rangle = F_A$. This implies that $\beta(F_B) = G$. Thus $\beta(B)$ generates $G$. Since $G$ is a finite group, $\beta(B^*)$ is a subgroup of $G$ and thus $\beta(B^*) = G$. By Theorem 3.2.9, the set $H = \alpha^{-1}(w^{-1}S)$ is a recurrent tree set. Thus $\beta(H) = G$ by Proposition 4.3.7. This implies that $\varphi(\Gamma_S(w) \cup \{\varepsilon\}) = G$. ■

We can now prove Theorem 4.3.5

*Proof of Theorem 4.3.5.* Let $S$ be a recurrent tree set of characteristic 1 and let $f : B^* \to A^*$ be a coding morphism for a finite $S$-maximal bifix code $X$. Set $T = f^{-1}(S)$. By Theorem 4.3.3 we know that $T$ is a tree set. We now prove that $T$ is recurrent.

Let $r, s \in T$. Since $S$ is recurrent, there exists $u \in S$ such that $f(r)uf(s) \in S$. Set $t = f(r)uf(s)$. Let $G$ be the representation of $F_A$ on the right cosets of $\langle X \rangle$. Let $\varphi : A^* \to G$ be the natural morphism from $A^*$ onto $G$. By Proposition 4.3.9, we have $\varphi(\Gamma_S(t) \cup \{\varepsilon\}) = G$. Let $v \in \Gamma_S(t)$ be such that $\varphi(v)$ is the inverse of $\varphi(t)$. Then $\varphi(tv)$ is the identity of $G$ and thus $tv \in \langle X \rangle$.

Since $S$ is a tree set, it is acyclic and thus $X^*$ is saturated in $S$ by the Saturation Theorem (Theorem 4.1.2). Thus $X^* \cap S = \langle X \rangle \cap S$. This implies that $tv \in X^*$. Since $tv \in A^*t$, we have $f(r)uf(s)v = f(r)qf(s)$ and thus $uf(s)v = qf(s)$ for some $q \in S$. Since $X^*$ is right unitary, $f(r), f(r)uf(s)v \in X^*$ imply $uf(s)v = qf(s) \in X^*$. In turn, since $X^*$ is left unitary, $qf(s), f(s) \in X^*$ imply $q \in X^*$ and thus $q \in X^* \cap S$. Let $w \in T$ be such that $f(w) = q$. Then $rws$ is in $T$. This shows that $T$ is recurrent. ■

The following example shows that the condition that $S$ is a tree set of characteristic 1 is necessary.

**Example 4.3.10** Let $S = \mathrm{Fac}\,((ab)^*))$ and $f$ be as in Example 2.3.3 (see also Example 3.3.2). It is easy to see that $S$ is a tree set of characteristic 2. Let $X = \{ab, ba\}$. The set $X$ is a finite $S$-maximal bifix code. It follows from Example 2.3.3 that the maximal bifix decoding $f^{-1}(S)$ is not recurrent.

### 4.3.2 Composition of bifix codes

In this section proving a result showing that in a recurrent tree set, the degrees of the terms of a composition of maximal bifix codes are multiplicative (Theorem 4.3.11). The following result is proved in [8, Proposition 11.1.2] for a more general class of codes (including all finite codes and not only finite bifix codes), but in the case of $S = A^*$.

**Theorem 4.3.11** *Let $S$ be a recurrent tree set and let $X, Z \subset S$ be finite bifix codes such that $X$ decomposes into $X = Y \circ_f Z$ where $f$ is a coding morphism for $Z$. Set $T = f^{-1}(S)$. Then $X$ is an $S$-maximal bifix code if and only if $Y$ is a $T$-maximal bifix code and $Z$ is an $S$-maximal bifix code. Moreover, in this case*

$$d_X(S) = d_Y(T)d_Z(S). \tag{4.1}$$

*Proof.* Assume first that $X$ is an $S$-maximal bifix code. By Proposition 1.2.18 (ii), $Y$ is a $T$-maximal prefix code and $Z$ is an $S$-maximal prefix code. This implies that $Y$ is a $T$-maximal bifix code and that $Z$ is an $S$-maximal bifix code.

The converse also holds by Proposition 1.2.18.

To show Formula (4.1), let us first observe that there exist words $w \in S$ such that for every parse $(v, x, u)$ of $w$ with respect to $X$, the word $x$ is not a factor of $X$. Indeed, let $n$ be the maximal length of the words of $X$. Assume that the length of $w \in S$ is larger than $3n$. Then if $(v, x, u)$ is a parse of $w$, we have $|u|, |v| < n$ and thus $|x| > n$. This implies that $x$ is not a factor of $X$.

Next, we observe that by Theorem 4.3.3, the set $T$ is a recurrent tree set.

Let $w \in S$ be a word with the above property. Let $\Pi_X(w)$ denote the set of parses of $w$ with respect to $X$ and $\Pi_Z(w)$ the set of its parses with respect to $Z$. We define a map $\varphi : \Pi_X(w) \to \Pi_Z(w)$ as follows. Let $\pi = (v, x, u) \in \Pi_X(w)$. Since $Z$ is a bifix code, there is a unique way to write $v = sy$ and $u = zr$ with $s \in A^* \setminus A^*Z$, $y, z \in Z^*$ and $r \in A^* \setminus ZA^*$. We set $\varphi(\pi) = (s, yxz, r)$. The triples $(y, x, z)$ are in bijection with the parses of $f^{-1}(yxz)$ with respect to $Y$. Since $x$ is not a factor of $X$ by the hypothesis made on $w$, and since $T$ is recurrent, there are $d_Y(T)$ such triples. This shows Formula (4.1). ∎

**Example 4.3.12** Let $S$ be the Fibonacci set. Let $B = \{u, v, w\}$ and $A = \{a, b\}$. Let $f : B^* \to A^*$ be the morphism defined by $f(u) = a$, $f(v) = baab$ and $f(w) = bab$. Set $T = f^{-1}(S)$. The words of length at most 3 of $T$ are represented on Figure 4.10.

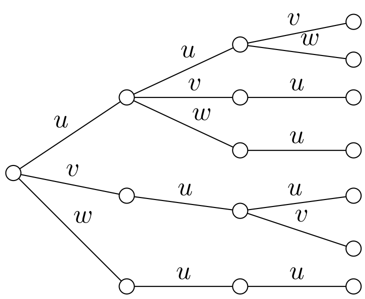


Figure 4.10: The words of length at most 3 in $T$.

The set $Z = f(B)$ is an $S$-maximal bifix code of $S$-degree 2 (it is the unique $S$-maximal bifix code of $S$-degree 2 with kernel $\{a\}$). Let $Y = \{uu, uvu, uw, v, wu\}$, which is a $T$-maximal bifix code of $T$-degree 2 (it is the unique $T$-maximal bifix code of $T$-degree 2 with kernel $\{v\}$).

The code $X = f(Y)$ is the $S$-maximal bifix code of $S$-degree 4 shown on Figure 4.11.

The following example shows that Formula (4.1) does not hold if $S$ is not a tree set of characteristic 1.

**Example 4.3.13** Let $S = \mathrm{Fac}(ab)^*$ (see Example 2.3.3). Let $Z = \{ab, ba\}$ and let $X = \{abab, ba\}$. We have $X = Y \circ_f Z$ for $B = \{u, v\}$, $f : B^* \to A^*$ defined by $f(u) = ab$ and $f(v) = ba$ with $Y = \{uu, v\}$. The codes $X$ and $Z$ are $S$-maximal bifix codes and $d_Z(S) = 2$. We have $d_X(S) = 3$ since $abab$ has three parses. Thus $d_Z(S)$ does not divide $d_X(S)$.

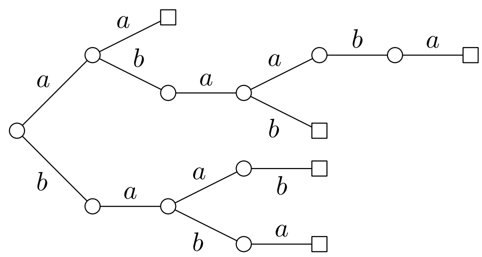


Figure 4.11: An $S$-maximal bifix code of $S$-degree 4.

### 4.3.3 Modular codes

For some special bifix code, we can give a more precise description of the bifix decoding and of Theorem 4.3.3.

Let $S$ be a tree set of characteristic $c$. Since $S$ is biextendable, 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 $R(\varepsilon)$.

Denote by $\mathcal{T}_0, \ldots \mathcal{T}_{c-1}$ the $c$ trees such that $E(\varepsilon) = \mathcal{T}_0 \cup \cdots \cup \mathcal{T}_{c-1}$. We define the *modular weight* of a letter $a$ as $\|a\| = j - i \pmod{c}$, where $\mathcal{T}_i$ is the tree containing $a$ as a left extension and $\mathcal{T}_j$ the tree containing $a$ as a right extension.

Given a word $w = a_0 a_1 \cdots a_m$, we define the *modular weight* of $w$ as $\|w\| = \sum_{k=0}^{m} \|a_k\| \pmod{c}$.

Note that the modular weight of a word depends on the choice of the order for the trees $\mathcal{T}_i$.

The set of words having modular weight equal to zero has the form $X^* \cap S$ for some special bifix code $X \subset S$ called the *modular code*. The set $X$ is the set of words having modular weight 0 such that all nonempty proper prefixes (or suffixes) have positive modular weight. It is easy to see that $X$ is actually a $S$-maximal bifix code.

Another way to define the modular code is by using the *modular graph*. This graph is defined as the directed graph $\mathcal{G}$ with vertices $0, 1, \ldots, c-1$ and edges all triples $(i, a, j)$ for $0 \leq i, j \leq c-1$ 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 $b, c \in A$. Observe that for every letter $a \in A$ there is exactly one edge labeled $a$ because $a$ appears exactly one as a left (resp. right) vertex in $\mathcal{E}(\varepsilon)$.

Note that, when $S$ is a tree set of characteristic $c$ obtained by a multiplying map using a transducer $\mathcal{A}$ (recall Section 3.3), the modular graph of $S$ is the output automaton of $\mathcal{A}$.

**Example 4.3.14** Let $S$ be the tree set of characteristic 2 of Example 2.3.3 (see also Example 3.3.2). The modular graph of $S$ is represented in Figure 4.12. It is the output automaton of the 2-multiplying transducer of Figure 3.14.

**Example 4.3.15** Let $S$ be the tree set of characteristic 2 of Example 5.2.22.

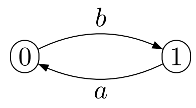


Figure 4.12: The modular graph of Fac $((ab)^*)$.

The modular graph of $S$ is represented in Figure 4.13. It is the output automaton of the 2-multiplying transducer of Figure 3.16.

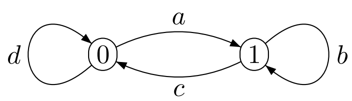


Figure 4.13: The modular graph of $S$.

**Proposition 4.3.16** *Let $S$ be a tree set of characteristic $c$ and let $\mathcal{G}$ be its modular 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 $(S_{i,j} \setminus \{\varepsilon\})_{0\leq i,j\leq c-1}$ 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) $\|w\| = 0$ if and only if $w \in S_{k,k}$ for some $0 \leq k \leq c-1$.*

*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 $0 \leq i, j \leq c-1\}$. 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. ■

Note that point (3) of Proposition 4.3.16 says that the modular code does not depend on the choice of the order of the states in the modular graph (or of the trees $\mathcal{T}_i$ in $\mathcal{E}(\varepsilon)$).

The following theorem improves Theorem 4.3.3 in the case of a bifix decoding by the modular code.

**Theorem 4.3.17** *The decoding of a recurrent tree set $S$ of characteristic $c$ by the modular code is a union of $c$ recurrent tree sets of characteristic* 1. *More precisely, if $f$ is the coding morphism for the modular code, then $f^{-1}(S_{0,0})$, $f^{-1}(S_{1,1})$, …, $f^{-1}(S_{c-1,c-1})$ are recurrent tree sets of characteristic* 1.

*Proof.* Let us define $T_k = f^{-1}(S_{k,k})$ for every $0 \le k \le c-1$. Fixed a $k$, we show that $T_k$ is a recurrent tree set of characteristic 1.

First, it is easy to verify that $T_k$ is biextendable.

Next, since $S$ is recurrent, for every $u, v \in S_{k,k} \subset S$ there exists a $w \in S$ such that $uwv \in S$. From point (2) of Proposition 4.3.16 follows that $w \in S_{k,k}$. Thus $T_k$ is recurrent.

Let now $X$ be the modular code and set $X_k = X \cap S_{k,k}$. In order to prove that $T_k$ is a tree set it is enough to show that $\mathcal{E}_{S_{k,k}}(w) = \mathcal{E}_S^{X_k,X_k}(w)$ is a tree for any $w \in S_{k,k}$. Note first that $\mathcal{E}_{S_{k,k}}(w) = \mathcal{E}_S^{X,X}(w)$ for any $w \in S_{k,k} \setminus \{\varepsilon\}$. Indeed, for $w \in S_{k,k}$ and $x, y \in X$ such that $xwy \in S$, one has $x, y \in X_k$ and thus $xwy \in S_{k,k}$.

According to Proposition 3.1.13, the graph $\mathcal{E}_S^{X,X}(w)$ is a tree for any word $w \in S \setminus \{\varepsilon\}$, whence the result.

Next, let us show that the graph $\mathcal{E}_S^{X_k,X_k}(\varepsilon)$ is also a tree. First, since a tree set is acyclic, the graph $\mathcal{E}_S^{X,X}(\varepsilon)$ is acyclic by Proposition 3.1.11 and so is its subgraph $\mathcal{E}_S^{X_k,X_k}$.

Let us prove that for every $x, y \in S_{k,k}$ there is a path in $\mathcal{E}_S^{X_k,X_k}(\varepsilon)$ from $x$ to $y$.

If $x, y \in A$, then there is a path from $x$ to $y$ in $\mathcal{E}(\varepsilon)$ and thus a there is a path from $x$ to $y$ in $\mathcal{E}_S^{X_k,X_k}(\varepsilon)$ obtained by replacing an edge $(a, b) \in A \times A$ of the path by an edge $(z, t)$ in $X_S^{X_k,X_k} \times X_S^{X_k,X_k}$ such that $z$ ends with $a$ and $t$ begins with $b$.

Otherwise, assume for example that $y = au$ with $u$ nonempty. Set $Y = \{v \in S \mid av \in X_k\}$. Since $Y$ is an $a^{-1}S$-maximal prefix code, by 3.1.13, the graph $\mathcal{E}_S^{X_k,Y}(a)$ is a tree. Since $u \in Y$, there is a path in $\mathcal{E}_S^{X_k,Y}(a)$ from $x$ to $u$. This implies that there is a path from $x$ to $y$ in $\mathcal{E}_S^{X_k,X_k}(\varepsilon)$. Thus $\mathcal{E}_S^{X_k,X_k}(\varepsilon)$ is connected. ∎

**Example 4.3.18** Let $S$ and $f$ be as in Examples 2.3.3 and 4.3.10. One has $f^{-1}(S_{0,0}) = \mathrm{Fac}(u^\omega)$ and $f^{-1}(S_{1,1}) = \mathrm{Fac}(v^\omega)$. Both are recurrent tree sets of characteristic 1, according with Theorem 4.3.17.

# Chapter 5

# Specular sets

In this chapter, we introduce specular groups and specular sets. Specular groups 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.

A specular set is a subset of such a group stable by taking the inverse and defined in terms of restrictions on the extensions of its elements.

The main results of this chapter are Theorems 5.3.11 and 5.5.1, referred to as the First Return Theorem and the Finite Index Basis Theorem for specular sets. 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 $S$. These generalize the analogous results proved for tree sets in Chapter 3 (Theorems 3.2.5 and 4.2.1).

The idea of considering recurrent sets of reduced words invariant by taking inverses is connected with the notion of $G$-full words of [60] (see Section 5.2.4).

This chapter is organized as follows. In Section 5.1, 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 5.1.3).

In Section 5.2 we introduce specular sets. We introduce odd and even words and the even code. We prove that the decoding of a recurrent specular set by this code is a union of two recurrent tree sets of characteristic 1 (Theorem 5.2.15). Moreover, we give a construction which allows to build specular sets from a tree set of characteristic 1 using a multiplying transducer, called doubling transducer (Theorem 5.2.20). We finally make a connection with the notion of $G$-full words introduced in [60] and related to the palindromic complexity of [35].

In Section 5.3 we prove several cardinality results concerning sets of return words on a specular set (Theorems 5.3.2, 5.3.5, 5.3.9). We also prove that the set of return words to a given word forms a basis of the even subgroup (Theorem 5.3.11 referred to as the First Return Theorem for specular sets) and that the mixed return words form a monoidal basis of the specular group (Theorem 5.3.13).

In Section 5.4 we prove several results concerning subgroups generated by bifix codes. Namely, we prove give new versions of the Freeness Theorem and of the Saturation Theorem for specular sets (Theorems 5.4.1 and 5.4.6).

Finally, in Section 5.5, we prove a version of the Finite Index Basis Theorem and a converse for specular sets (Theorem 5.5.1 and Theorem 5.5.6).

## 5.1 Specular groups

In this section, we introduce specular groups and we prove some properties of this family of groups. In particular, using the Kurosh subgroup theorem, we prove that any subgroup of a specular group is specular (Theorem 5.1.3).

We consider an alphabet $A$ with an involution $\theta : A \to A$, possibly with some fixed points. We also consider the group $G_\theta$ 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 $G_\theta$.

When $\theta$ has no fixed point, we can set $A = B \cup B^{-1}$ by choosing a set of representatives of the orbits of $\theta$ for the set $B$. The group $G_\theta$ is then the free group on $B$, denoted $F_B$. In general, the group $G_\theta$ 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 $\theta$ 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 5.1.1** *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 5.1.2** Let $A = \{a, b, c, d\}$ and let $\theta$ 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 $G_\theta$ 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 [30]).

### 5.1.1 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 [54]). Thus, we have, replacing the Nielsen-Schreier Theorem of free groups, the following result.

**Theorem 5.1.3** *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 $G_\theta$ are the conjugates of the $j$ fixed points of $\theta$ 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 $\theta$ and thus is of odd length).

A group having a free subgroup of finite index is called *virtually free* (see [30]).

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

**Proposition 5.1.4** *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 [53, p. 197]). We thus deduce from Proposition 5.1.4 the following result.

**Proposition 5.1.5** *A specular group is Hopfian.*

### 5.1.2 Monoidal basis

A word on the alphabet $A$ is *$\theta$-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* (with respect to $\theta$) if it is closed under taking inverses (with respect to $\theta$). 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 $\varepsilon$.

**Example 5.1.6** The alphabet $A$ is a monoidal basis of $G_\theta$.

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 5.1.1).

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

$$X = \{paq^{-1} \mid a \in A, p, q \in Q, pa \notin Q, pa \in Hq\}. \tag{5.1}$$

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 set $X$ is a monoidal basis of $H$, called the *Schreier basis* relative to $Q$.

**Proposition 5.1.7** *Let $H$ and $Q$ be as above and let $X$ be a Schreier basis relative to $Q$. Then $X$ is closed by taking inverses.*

*Proof.* 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 $pa \in Q$. It 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 $1 \leq k \leq m-1$ 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 $\varepsilon$, then $x_kx_{k+1} = 1$ for some index $k$ since the central part $a$ never cancels in a product of two elements of $X$. ■

One can deduce directly Theorem 5.1.3 from these properties of $X$.

*Proof of Theorem 5.1.3.* 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 $\sigma$-reduced then $\varphi(w) \neq \varepsilon$. This shows that $H$ is isomorphic to the group $G_\sigma$. 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

$$s = n(r-2) + 2. \tag{5.2}$$

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 (5.2).

**Example 5.1.8** Let $G$ be the specular group of Example 5.1.2. 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 5.1.

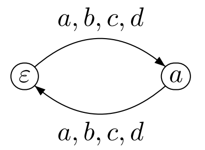


Figure 5.1: The representation of $G$ by permutations on the cosets of $H$.

The monoidal basis corresponding to Formula (5.1) is $X = \{ab, ac, ad, ba, ca, da\}$. The symmetric rank of $H$ is 6, in agreement with Formula (5.2) and $H$ is a free group of rank 3.

**Example 5.1.9** Let again $G$ be the specular group of Example 5.1.2. Consider now the subgroup $K$ stabilizing 1 in the representation of $G$ by permutations on the set $\{1, 2\}$ of Figure 5.2.

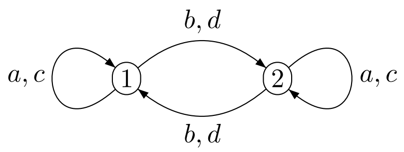


Figure 5.2: The representation of $G$ by permutations on the cosets of $K$.

We choose $Q = \{\varepsilon, b\}$. The set $X$ corresponding to Formula (5.1) 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 5.3), is a consequence of Proposition 5.1.5.

**Proposition 5.1.10** *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 $\varphi$ from $A$ onto $X$ such that $\varphi(a^{-1}) = \varphi(a)^{-1}$ for every $a \in A$. The map $\varphi$ extends to a morphism from $G$ to $G$ which is surjective since $X$ generates $G$. Then $\varphi$ being surjective, it also injective since $G$ is Hopfian, and thus $X$ is a monoidal basis of $G$. ∎

## 5.2 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 5.2.15). 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 5.2.20). In the last part, we relate specular sets with full and $G$-full words, a notion linked with palindromic complexity and introduced in [60].

We assume given an involution $\theta$ on the alphabet $A$ generating the specular group $G_\theta$.

A symmetric biextendable (and thus factorial) set $S$ of reduced words on the alphabet $A$ is called a *laminary set* on $A$ relative to $\theta$ (following [26] and [51]). Thus the elements of a laminary set $S$ are elements of the specular group $G_\theta$ and the set $S$ is contained in $G_\theta$.

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 5.2.1** Let $A = \{a, b\}$ and let $\theta$ be the identity on $A$. Then the set of factors of $(ab)^\omega$ is a specular set.

**Example 5.2.2** Let $S$ be the set defined in Example 3.1.5. The set $S$ is a tree set of characteristic 2. The extension graph of $\varepsilon$ is shown in Figure 3.2.

We will see later (Example 5.2.23) that $S$ is a specular set relative to the involution $\theta$ fixing $a, c$ and exchanging $b$ and $d$.

**Example 5.2.3** The set $S$ be 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 $\{a, b, c\}$. The set $S$ is a specular set (it is actually the natural coding of a linear involution, as we will see in Example 8.1.4).

The words of length at most 3 of $S = \mathcal{L}(T)$ are represented in Figure 5.3.

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 5.2.4** *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 $(1 \otimes b^{-1}, a^{-1} \otimes 1) \in \mathcal{T}_{1-i}$.*

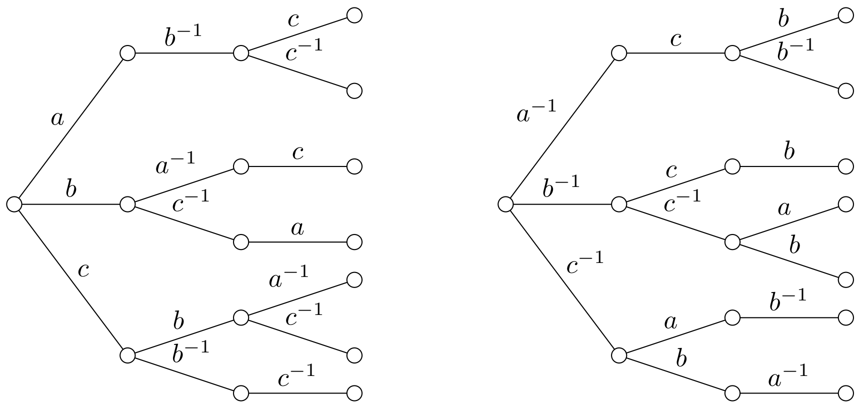


Figure 5.3: The words of length at most 3 of $S$.

*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. ■

**Example 5.2.5** Let $S$ be the specular set of Example 5.2.3. The extension graph of the empty word of $S$ is represented in Figure 5.4.

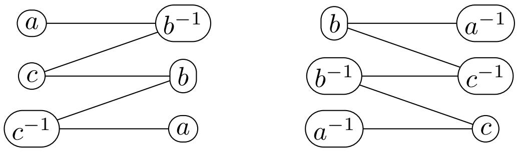


Figure 5.4: The extension graphs $\mathcal{E}_S(\varepsilon)$.

Recall from Chapter 1 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 $G_\theta$).

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 5.2.6** *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 biextendable, 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. ■

Since a specular set is, in particular, a tree set of characteristic 2, we have the following immediate consequence of Proposition 3.1.4.

**Proposition 5.2.7** *The factor complexity of a specular set is given by $p_0 = 1$ and $p_n = n(\mathrm{Card}(A) - 2) + 2$ for $n \geq 1$.*

### 5.2.1 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 biextendable, 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 5.2.8** Let $S$ be the set of factors of $(ab)^\omega$ as in Example 5.2.1. Then $a$ and $b$ are odd.

**Example 5.2.9** Let $S$ be the set of Example 5.2.2. 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). Note that, since a specular set is in particular a tree set of characteristic 2, the even code coincides with the modular code seen in Section 4.3.3.

**Proposition 5.2.10** *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 5.2.11** Let $S$ be the specular set of Example 5.2.2 (see also Example 5.2.9). 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 $\mathcal{E}(\varepsilon)$.

Note that the parity graph of a specular set $S$ coincides with the modular graph defined in Section 4.3.3.

**Example 5.2.12** Let $S$ be the specular set of Example 5.2.2. The parity graph of $S$ is represented in Figure 4.13, where we assume that $\mathcal{T}_0$ is the tree on the left of Figure 3.2 and $\mathcal{T}_1$ is the tree on the right of Figure 3.2.

The following result is an easy generalization of Proposition 4.3.16.

**Proposition 5.2.13** *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.* Assertsions (1)-(3) follow from Proposition 4.3.16, while assertion (4) follows from Proposition 5.2.4). ■

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 5.2.14** *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 5.2.13 and thus $y \in S_{j,1-j}$ by Assertion (2). Thus $y$ is odd by Assertion (3). ∎

The following result is just Theorem 4.3.17 applied to a specular set.

**Theorem 5.2.15 (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.

**Example 5.2.16** Let $S$ be the set of Example 3.1.6. Recall that it is the set of factors of the fixed point of the morphism $\sigma(a) = ab$, $\sigma(b) = cda$, $\sigma(c) = cd$, $\sigma(d) = abc$. The even code $X$ is given in Example 5.2.11.

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 $g^{-1}\sigma g$.

### 5.2.2 Bifix codes in specular sets

Recall from Chapter 1 that the characteristic of a set $S$ is given by $\chi(S) = \ell_S(\varepsilon) + r_S(\varepsilon) - b_S(\varepsilon)$.

Applying Theorem 2.2.1 to recurrent specular sets we have the following result.

**Theorem 5.2.17 (Cardinality Theorem for bifix codes)** *Let $S$ be a recurrent specular set. For any finite $S$-maximal bifix code $X$, one has*

$$\mathrm{Card}(X) = d_X(S)(\mathrm{Card}(A) - 2) + 2. \tag{5.3}$$

**Example 5.2.18** Let $S$ be the specular set of Example 3.1.6. The even code (given in Example 5.2.11) is an $S$-maximal code of $S$-degree 2. We have $\mathrm{Card}(X) = 6$ in agreement with Theorem 5.2.17.

The following statement is a partial converse of Theorem 5.2.17.

**Theorem 5.2.19** *Let $S$ be a uniformly recurrent laminary set. 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.*

Theorem 5.2.19 results from Proposition 2.2.5 applied with $d_0 = 2$.

### 5.2.3 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 Section 3.3.

Let $Q = \{0,1\}$. We call *doubling map* a 2-multiplying map $\delta_{\mathcal{A}} = (\delta_0, \delta_1)$ with respect to a transducer $\mathcal{A}$, called *doubling transducer*.

By Theorem 3.3.1, the image of a tree set of characteristic 1 by a doubling map is a tree set of characteristic 2. We will show that it is actually a specular set.

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

Recall that a set $S$ of words is closed under reversal if $w \in S$ implies $\tilde{w} \in S$ for every $w \in S$.

**Theorem 5.2.20** *For any tree set $T$ of characteristic 1 on the alphabet $\Sigma$, 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 $S = \delta_0(T) \cup \delta_1(T)$. By Theorem 3.3.1, $S$ 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 5.2.21** Let $\Sigma = \{\alpha, \beta\}$ and let $T$ be the Fibonacci set over $\Sigma$ (see Example 1.1.2). Let $\delta$ be the doubling map given by the transducer of Figure 5.5 on the left.

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

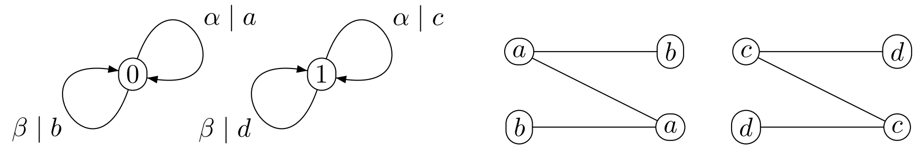


Figure 5.5: A doubling transducer (on the left) and the extension graph $\mathcal{E}_S(\varepsilon)$ (on the right).

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 $\delta$ is a specular set $S$ on the alphabet $A = \{a, b, c, d\}$. The graph $\mathcal{E}_S(\varepsilon)$ is represented in Figure 5.5 on the right. All letters are even.

Note that the set $S$ of Example 5.2.21 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 5.2.22** Let $\Sigma = \{\alpha, \beta\}$ and let $T$ be the Fibonacci set. Let $\delta$ be the doubling map given by the transducer of Figure 3.16 on the left. The letter $\alpha$ acts as the transposition of the two states $0, 1$, while $\beta$ acts as the identity.

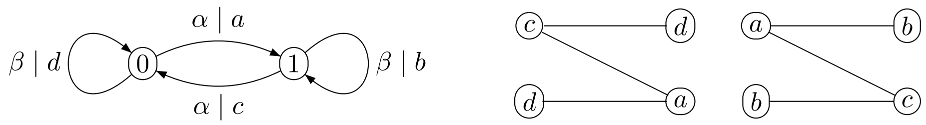


Figure 5.6: A doubling transducer and the extension graph $\mathcal{E}_S(\varepsilon)$.

Then $\theta_{\mathcal{A}}$ is the involution $\theta$ of Example 5.1.2 and the image of $T$ by $\delta$ is a specular set $S$ on the alphabet $A = \{a, b, c, d\}$. The graph $\mathcal{E}_S(\varepsilon)$ is represented in Figure 3.16 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 5.2.23** 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 [52]). 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 3.16 is the set of factors of the fixed point $\sigma^\omega(a)$ of the morphism $\sigma$ 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 3.1.5 The set $S$ is a specular set relative to the involution $\theta$ fixing $a, c$ and exchanging $b$ and $d$.

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 4.13 and 3.16).

### 5.2.4 $G$-Palindromes

We discussed at the end of Chapter 3 the connection between tree sets and palindromes. In particular we proved that a recurrent tree set of characteristic 1 closed under reversal is full (Proposition 3.4.1).

In [60], this notion of full set 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 $\delta$.

Let us denote by $\sigma$ the antimorphism $u \mapsto u^{-1}$ for $u \in G_\theta$. From Section 5.2.3 it follows that both edges $(i, \alpha, a, j)$ and $(1-i, \alpha, \sigma(a), 1-j)$ are in the doubling transducer. Let us define also the morphism $\tau$ obtained by replacing each letter $a \in A$ by $\tau(a)$ if there are edges $(i, \alpha, a, j)$ and $(1-j, \alpha, \tau(a), 1-i)$ in the doubling transducer.

We denote by $G_{\mathcal{A}}$ the group generated by the $\sigma$ and $\tau$. Actually, we have $G_{\mathcal{A}} = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. Indeed, one has $\sigma\tau = \tau\sigma$.

**Example 5.2.24** Let $S$ be the specular set defined in Example 5.2.21. The group $G_{\mathcal{A}}$ is generated by

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

and

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

Note that, even if the images of $\sigma$ and $\tau$ 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 5.2.25** Let $S$ be the recurrent specular set defined in Example 5.2.22. 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 $= \{\mathrm{id}, \sigma, \tau, \sigma\tau\}$, 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 3.4.1 for specular sets.

**Proposition 5.2.26** *Let $T$ be a recurrent tree set of characteristic* 1 *on the alphabet $\Sigma$, 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 3.4.1 we know that $T$ is full.

To show that $S$ is $G_{\mathcal{A}}$-full, we will use several properties of the map $\delta_i$. 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 $\delta_0, \delta_1$ 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 [43]). 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 5.2.27** Let $S$ be the specular set of Example 5.2.21. 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 $\sigma$ and the morphism $\tau$ of Example 5.2.24. 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 1.4) 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 5.2.28** Let $S$ be the specular set of Example 5.2.22. 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 $\sigma$ taking the inverse (that is fixing $a, c$ and exchanging $b$ and $d$) and the morphism $\tau$ (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 $G_{\mathcal{A}}$.

## 5.3 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 5.3.5, 5.3.2, 5.3.9). We also prove that the set of return words to a given word forms a basis of the even subgroup (Theorem 5.3.11 referred to as the First Return Theorem) and that the mixed return words form a monoidal basis of the specular group (Theorem 5.3.13).

### 5.3.1 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 5.3.5, 5.3.2 and 5.3.9).

#### Complete return words

Let $S$ be a factorial set of words and let $X \subset S$ be a set of nonempty words. Recall from Section 1.4 that 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 $X$. The set $\mathcal{CR}_S(X)$ of complete return words to $X$ is a bifix code. If $S$ is uniformly recurrent, $\mathcal{CR}_S(X)$ is finite for any finite set $X$.

**Example 5.3.1** Let $S$ be the specular set of Example 5.2.22. One has

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

A direct consequence of Theorem 2.2.8 is the following.

**Theorem 5.3.2** *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 Theorem 5.3.2.

**Example 5.3.3** Let $S$ be the specular set on the alphabet $A = \{a, b, c, d\}$ of Example 5.2.2. We have

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

It has four elements in agreement with Theorem 5.3.2.

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 [7, Section 4.3]). This gives a connection between Equations (5.3) and (2.2). Indeed, by Equation (5.3), 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 (2.2) since $X \setminus K$ is a bifix code with empty kernel.

### Right return words

Let $S$ be a factorial set. For any nonempty word $x \in S$, we defined in Section 1.4 a right return word to $x$ in $S$ as a word $w$ such that $xw$ is a complete return word to $x$. We also denoted by $\mathcal{R}_S(x)$ the set of right return words to $x$ in $S$.

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

**Proposition 5.3.4** *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 5.3.5 (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 Theorem 5.3.2 with $X = \{x\}$, since $\mathrm{Card}(\mathcal{R}_S(x)) = \mathrm{Card}(\mathcal{CR}_S(x))$. ∎

**Example 5.3.6** Let $S$ be the specular set of Example 5.2.22. We have

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

By Theorem 3.2.5, if $S$ is a recurrent tree set of characteristic 1 on the alphabet $B$, then for any $x \in S$, one has $\mathrm{Card}(\mathcal{R}_S(x)) = \mathrm{Card}(B)$. The relation with Theorem 5.3.5 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 5.3.5 (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 5.3.5.

### 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 $X = \{w, w^{-1}\}$.

**Theorem 5.3.7** *Let $S$ be a recurrent specular set. 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 Theorem 5.3.2. ∎

**Example 5.3.8** Let $S$ be the specular set of Example 5.2.22. In view of the values of $\mathcal{CR}_S(b)$ and $\mathcal{CR}_S(d)$ given in Example 5.3.1, 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 $\theta$ has no fixed point (in particular when $S$ is the natural coding of a linear involution, as we will see in Chapter 8). In this case, the group $G_\theta$ 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 case where $S$ is the natural coding of a linear involution, as we will see in Chapter 8.

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 5.7). Note that successive occurrences of $w$ may overlap but that successive occurrences of $w$ and $w^{-1}$ cannot.

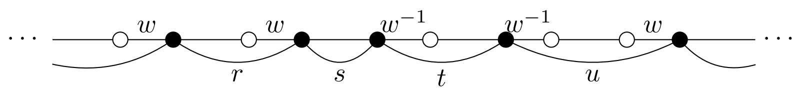


Figure 5.7: A uniformly recurrent infinite word factorized as an infinite product $\cdots rstu \cdots$ of mixed return words to $w$.

We have the following cardinality result.

**Theorem 5.3.9 (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 5.3.7 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 5.7.

**Example 5.3.10** Let $S$ be the specular set of Example 5.2.22. The value of $\mathcal{CR}_S(b, d)$ is given in Example 5.3.8. Since $b, d$ do not overlap,

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

has four elements in agreement with Theorem 5.3.9.

### 5.3.2 First Return Theorem for specular sets

By Theorem 3.2.5, the set of right return words to a given word in a recurrent tree set of characteristic 1 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 $G_\theta$ with symmetric rank $2(\mathrm{Card}(A) - 1)$ by (5.1) generated by the even code. Since no even word is its own inverse (by Proposition 5.2.13), it is a free group. Thus its rank is $\mathrm{Card}(A) - 1$.

**Theorem 5.3.11 (First Return Theorem for specular sets)** *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 5.2.13, and set $X_0 = X \cap S_{0,0}$, $X_1 = X \cap S_{1,1}$. By Theorem 5.2.15, 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 Theorem 3.2.5, 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 5.3.5. ∎

**Example 5.3.12** Let $S$ be the specular set of Example 5.2.22. The sets of right return words to $a, b, c, d$ are given in Example 5.3.6. Each one is a basis of the even subgroup.

Concerning mixed return words, we have the following statement.

**Theorem 5.3.13** *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 $G_\theta$.*

*Proof.* Since $w$ and $w^{-1}$ do not overlap, we have $\mathcal{R}_S(w) \subset \mathcal{MR}_S(w)^*$. Thus, by Theorem 5.3.11, 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 $G_\theta$, this implies that $\mathcal{MR}_S(w)$ generates the group $G_\theta$. Finally since $\mathcal{MR}_S(w)$ has $\mathrm{Card}(A)$ elements by Theorem 5.3.9, we obtain the conclusion by Proposition 5.1.10. ∎

**Example 5.3.14** Let $S$ be the specular set of Example 5.2.22. We have seen in Example 5.3.10 that

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

This set is a monoidal basis of $G_\theta$ in agreement with Theorem 5.3.13.

## 5.4 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 5.4.1). 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 5.4.6).

We can see these two results as a generalization of the Freeness Theorem and Saturation Theorem (Theorems 4.1.1 and 4.1.2) in the case of a specular set. Indeed, when the involution $\theta$ is the identity we recover the original results of Chapter 4.

### 5.4.1 Freeness Theorem

Let $\theta$ be an involution on $A$ and let $G_\theta$ be the corresponding specular group. A symmetric set $X$ is free if it is a monoidal basis of a subgroup $H$ of the group $G_\theta$. 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 $1 \leq i < n$ (see also Section 4.1.

The following is a consequence of Theorem 4.1.1.

**Theorem 5.4.1 (Freeness Theorem for laminary sets)** *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 4.1.1, using the incidence graph $\mathcal{G}_X$ of a bifix code $X$.

### 5.4.2 Cosets

Let $X$ be a symmetric set with respect to an involution $\theta$. Recall from Section 4.1.1 the definition of incidence graph $\mathcal{G}_X$. The set of vertices of $\mathcal{G}_X$ is the disjoint union of the set $P_X$ of nonempty proper prefixes of $X$ and the set $S_X$ of nonempty proper suffixes of $X$. As for extension graphs (see Chapter 1), we use the notation $1 \otimes w$ for a vertex $w \in P_X$ and $w \otimes 1$ for a vertex $w \in S_X$.

We define an equivalence relation $\gamma_X$ on the set $P$ of proper prefixes of $X$, called the *$\theta$-coset equivalence*, or simply *coset automaton* when $\theta$ is understood, 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 graph $\mathcal{G}_X$. It is easy to verify that, since $X$ is symmetric, $\gamma_X$ is indeed an equivalence. The class of the empty word $\varepsilon$ is reduced to $\{\varepsilon\}$. This definition is an extension to symmetric sets of the equivalence denoted $\theta_X$ defined in Section 4.1.1. Indeed, when the involution is just the identity, the two equivalence relations coincide for all elements in $A^*$.

The following statement is the generalization to symmetric bifix codes of Proposition 4.1.4. We denote by $\langle X \rangle$ the subgroup generated by $X$.

**Proposition 5.4.2** *Let $X$ be a symmetric bifix code and let $P$ be the set of its proper prefixes. Let $\gamma_X$ 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 $\gamma_X$ to define the *$\theta$-coset automaton*, or simply *coset automaton* when $\theta$ is understood, $\mathcal{C}_X$ of a symmetric bifix code $X$ as follows. The vertices of $\mathcal{C}_X$ are the equivalence classes of $\gamma_X$. 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 5.8):

(i) $pa \in P$ and $pa \equiv q \bmod \gamma_X$,

(ii) or $pa \in X$ and $q = \varepsilon$.

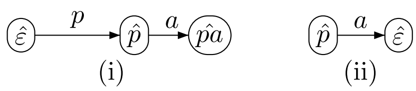


Figure 5.8: The edges of the coset automaton.

Note that, when the involution $\theta$ is the identity, the coset automaton $\mathcal{C}_X$ coincides with the automaton $\mathcal{B}_X$ defined in Section 4.1.2.

**Proposition 5.4.3** *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 5.4.2 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 $\theta$. 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 $G_\theta$ 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).

The following proposition is a generalization to specular sets of Proposition 4.1.7.

**Proposition 5.4.4** *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 5.4.3, 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 $S_{j,\ell_t}$ 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 4.1.3 (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 4.1.3 (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 5.4.5** Let $S$ be the specular set of Example 5.2.3. Let $X$ be the set of words of length 3 of $S$ (see Figure 5.3), which is a symmetric bifix code. The incidence graph $\mathcal{G}_X$ is represented in Figure 5.9.

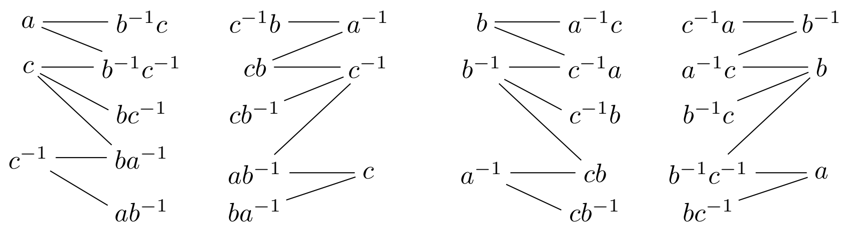


Figure 5.9: The incidence graph of $X$.

The coset automaton $\mathcal{C}_X$ is represented in Figure 5.10 (we only represent one 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 5.9. The vertex 3 corresponds to the two last ones.

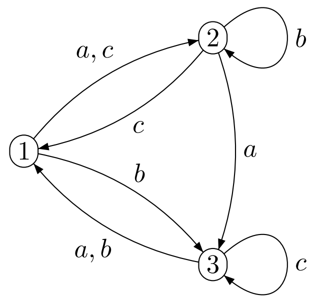


Figure 5.10: The coset automaton.

### 5.4.3 Saturation Theorem

Let $H$ be a subgroup of the specular group $G_\theta$ and let $S$ be a specular set on $A$ relative to $\theta$. 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 the Saturation Theorem (Theorem 4.1.2).

**Theorem 5.4.6 (Saturation Theorem for laminary sets)** *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 5.4.1. Since, by Proposition 5.4.4, 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 $\hat{\varepsilon}$ inbetween, 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 5.4.7** 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 $x^\omega$ 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 $b \in \langle X \rangle \cap S$.

## 5.5 The Finite Index Basis property

In this section we prove a cunterpart of the Finite Index Basis Theorem for specular sets (Theorem 5.5.1) and a converse (Theorem 5.5.6).

### 5.5.1 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 Theorem 4.2.1). The proof is very similar to that of Theorem 4.2.1 and we omit some details.

**Theorem 5.5.1 (Finite Index Basis Theorem for specular sets)** *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 [7, Theorem 4.4.3], 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 5.5.2** *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 [7, Theorem 4.3.12], $X$ is the kernel of some $S$-maximal bifix code $Z$ of $S$-degree $d+1$. Since $S$

is recurrent, by [7, Theorem 4.4.3], $Z$ is finite. Let us show that $Z$ is symmetric. Indeed, we have by [7, Theorem 4.3.11], $d_Z(w) = \min\{d+1, d_X(w)\}$. 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 5.5.1.* 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 5.4.6, 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 $G_\theta$ (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 $G_\theta$. Clearly, $\varepsilon = 1_{G_\theta} \in V$. 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 5.3.11, 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 5.2.14. 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 $G_\theta$ 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 5.3.2, 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 5.1.10 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 5.5.2, 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 $G_\theta$. 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 5.5.3** Let $S$ be the specular set of Example 5.2.3. 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 5.5.1 is the counterpart for specular sets of Theorem 4.3.6.

**Theorem 5.5.4** *Let $S$ be a recurrent specular set. For any subgroup $H$ of finite index of the group $G_\theta$, 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 $G_\theta$. 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 [7, Theorem 4.2.8].

By Theorem 5.5.1, $X$ 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 5.5.4 with the following interesting example.

**Example 5.5.5** Let $S$ be the specular set of Example 5.2.3. 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

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

### 5.5.2 A converse of the Finite Index Basis Theorem

The following is a converse of Theorem 5.5.1. It is also the counterpart for specular sets of Corollary 4.2.6.

**Theorem 5.5.6** *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 1.1.6, 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. ∎

# Chapter 6

# Interval exchanges

In this chapter we study interval exchange sets. These sets are a particular example of tree sets arising from a family of dynamical system called interal exchange transformations.

In Section 6.1 we introduce interval exchange transformations and interval exchange sets. We concentrate on the study of minimal and regular interval exchanges, showing the connection between these two families (Theorem 6.1.6). We prove that an interval exchange set satisfying some natural condition is a planar tree set (Theorem 6.1.16). This generalize a result from Ferenczi and Zamboni (see [39]).

In Section 6.2 we study the connection between regular interval exchange sets and bifix codes. Given an interval exchange, we define a transformation associated to a maximal bifix decoding and we prove that this transformation is regular provided the original one was regular (Theorem 6.2.10). We finally prove that the family of regular interval exchange sets is closed under maximal bifix decoding (Theorem 6.2.11) and, as a corollary, so is the family of recurrent planar tree sets of characteristic 1 (Corollary 6.2.13).

## 6.1 Interval exchange transformations

Let us recall the definition of an interval exchange transformation (see [25], [68] or [66] for a more detailed presentation).

A *semi-interval* is a nonempty subset of the real line of the form $[\alpha, \beta[ = \{z \in \mathbb{R} \mid \alpha \leq z < \beta\}$. Thus it is a left-closed and right-open interval. For two semi-intervals $\Delta, \Gamma$, we denote $\Delta < \Gamma$ if $x < y$ for any $x \in \Delta$ and $y \in \Gamma$.

Let $A$ be a finite, nonempty and ordered alphabet. Given an order $<$ on $A$, a partition $(I_a)_{a \in A}$ of a semi-interval $[\ell, r[$ in semi-intervals is *ordered* if $a < b$ implies $I_a < I_b$.

Let now $<_1$ and $<_2$ be two total orders on $A$. Let $(I_a)_{a \in A}$ be a partition of $[\ell, r[$ in semi-intervals ordered for $<_1$. Let $\lambda_a$ be the length of $I_a$. Let $\mu_a = \sum_{b \leq_1 a} \lambda_b$ and $\nu_a = \sum_{b \leq_2 a} \lambda_b$. Set $\alpha_a = \nu_a - \mu_a$. The *interval exchange* *transformation* relative to $(I_a)_{a\in A}$ is the map $T : [\ell, r[\to [\ell, r[$ defined by

$$T(z) = z + \alpha_a \quad \text{if } z \in I_a.$$

Observe that the restriction of $T$ to $I_a$ is a translation onto $J_a = T(I_a)$, that $\mu_a$ is the right boundary of $I_a$ and that $\nu_a$ is the right boundary of $J_a$. We additionally denote by $\gamma_a$ the left boundary of $I_a$ and by $\delta_a$ the left boundary of $J_a$. Thus

$$I_a = [\gamma_a, \mu_a[, \quad J_a = [\delta_a, \nu_a[.$$

Since $a <_2 b$ implies $J_a <_2 J_b$, the family $(J_a)_{a\in A}$ is a partition of $[\ell, r[$ ordered for $<_2$. In particular, the transformation $T$ defines a bijection from $[\ell, r[$ onto itself.

An interval exchange transformation relative to $(I_a)_{a\in A}$ is also said to be on the alphabet $A$. The values $(\alpha_a)_{a\in A}$ are called the *translation values* of the transformation $T$.

**Example 6.1.1** Let $R$ be the interval exchange transformation corresponding to $A = \{a, b\}$, $a <_1 b$, $b <_2 a$, $I_a = [0, 1-\alpha[$, $I_b = [1-\alpha, 1[$ (see Figure 6.1).

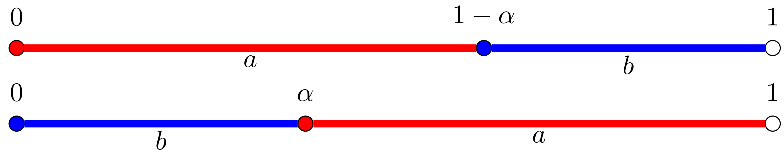


Figure 6.1: A rotation.

The transformation $R$ is the rotation of angle $\alpha$ on the semi-interval $[0, 1[$ defined by $R(z) = z + \alpha \bmod 1$.

Since $<_1$ and $<_2$ are total orders, there exists a unique permutation $\pi$ of $A$ such that $a <_1 b$ if and only if $\pi(a) <_2 \pi(b)$. Conversely, $<_2$ is determined by $<_1$ and $\pi$ and $<_1$ is determined by $<_2$ and $\pi$. The permutation $\pi$ is said to be *associated* to $T$.

Set $A = \{a_1, a_2, \ldots, a_s\}$ with $a_1 <_1 a_2 <_1 \cdots <_1 a_s$. The pair $(\lambda, \pi)$ formed by the family $\lambda = (\lambda_a)_{a\in A}$ and the permutation $\pi$ determines the map $T$. We will also denote $T$ as $T_{\lambda,\pi}$. The transformation $T$ is also said to be an $s$-interval exchange transformation.

It is easy to verify that the family of $s$-interval exchange transformations is closed by taking inverses.

**Example 6.1.2** Let $T = R^2$ where $R$ is the rotation of Example 6.1.1. The transformation $T$, represented in Figure 6.2 is a 3-interval exchange transformation. One has $A = \{a, b, c\}$ with $a <_1 b <_1 c$ and $b <_2 c <_2 a$. The associated permutation is the cycle $\pi = (abc)$.

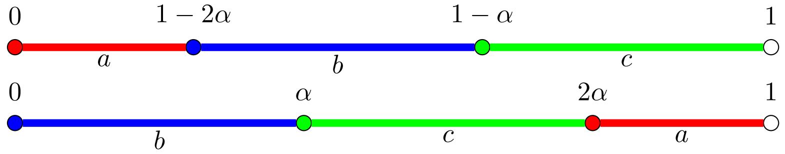

Figure 6.2: A 3-interval exchange transformation.

**Example 6.1.3** Let $A = \{a, b, c\}$. Consider the rotation of angle $\alpha$ with $\alpha$ irrational as in Example 6.1.1, but as a 3-transformation relative to the partition $(I_a)_{a\in A}$ of the interval $]0, 1[$, where

$$I_a = ]0, 1 - 2\alpha[, \quad I_b = ]1 - 2\alpha, 1 - \alpha[ \quad \text{and} \quad I_c = ]1 - \alpha, 1[,$$

while

$$J_c = ]0, \alpha[, \quad J_a = ]\alpha, 1 - \alpha[ \quad \text{and} \quad J_b = ]1 - \alpha, 1[$$

(see Figure 6.3). Then, for each letter $a$, the restriction to $I_a$ is a translation to $J_a$. Note that one has $a <_1 b <_1 c$ and $c <_2 a <_2 b$.

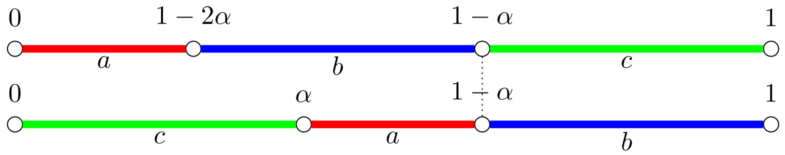

Figure 6.3: A 3-interval exchange transformation.

### 6.1.1 Regular interval exchanges

The *orbit* of a point $z \in [\ell, r[$ is the set $\mathcal{O}(z) = \{T^n(z) \mid n \in \mathbb{Z}\}$. The transformation $T$ is said to be *minimal* if for any $z \in [\ell, r[$, the orbit of $z$ is dense in $[\ell, r[$.

From now on, set $\gamma_i = \gamma_{a_i}$, $\delta_i = \delta_{a_i}$, $\mu_i = \mu_{a_i}$ and $\nu_i = \nu_{a_i}$. The points $0 = \gamma_1, \mu_1 = \gamma_2, \ldots, \mu_{s-1} = \gamma_s$ form the set of *separation points* of $T$, denoted $\mathrm{Sep}(T)$. Note that the transformation $T$ has at most $s - 1$ *singularities* (that is points at which it is not continuous), which are among the nonzero separation points $\gamma_2, \ldots, \gamma_s$.

A *connection* of an interval exchange transformation $T$ is a triple $(x, y, n)$ where $x$ is a singularity of $T^{-1}$, $y$ is a singularity of $T$, $n \geq 0$ and $T^n(x) = y$. We also say that $(x, y, n)$ is a connection of length $n$ ending in $y$. When $n = 0$, we say that $x = y$ is a connection.

**Example 6.1.4** Let $T$ be the transformation of Example 6.1.3. The point $\gamma_c$ is a connection of length 0. This connection is represented with a dotted line in Figure 6.3.

Let $T$ be an interval exchange transformation with exactly $c$ connections all of length 0. Denote by $\gamma_{k_0} = \ell$ and $\gamma_{k_1}, \ldots, \gamma_{k_c}$ the $c$ connections of $T$. For every $0 \leq i < c$ the interval $]\gamma_{k_i}, \gamma_{k_{i+1}}[$ is called a *component* of $I$.

**Example 6.1.5** Consider again the transformation $T$ of Example 6.1.3. The two components of $]0,1[$ are the two intervals $]0,1-\alpha[$ and $]1-\alpha,1[$.

An interval exchange transformation $T_{\lambda,\pi}$ is called *regular* if the orbits of the nonzero separation points $\gamma_2,\ldots,\gamma_s$ are infinite and disjoint. Note that the orbit of 0 cannot be disjoint from the others since one has $T(\gamma_i)=0$ for some $i$ with $2\le i\le s$. This condition is sometimes called *idoc*, where idoc stands for *infinite disjoint orbit condition*).

Equivalently, an interval exchange is called regular if it has no conection (see [18]).

As an example, the 2-interval exchange transformation of Example 6.1.1 which is the rotation of angle $\alpha$ is regular if and only if $\alpha$ is irrational.

The following result is due to Keane [47].

**Theorem 6.1.6 (Keane)** *A regular interval exchange transformation is minimal.*

The converse is not true. Indeed, the transformation of Example 6.1.3. The transformation is minimal as any rotation of irrational angle but it is not regular since $\mu_1=1-2\alpha$, $\mu_2=1-\alpha$ and thus $\mu_2=T(\mu_1)$.

**Example 6.1.7** Let $T$ be the 3-interval exchange transformation of Example 6.1.2 with $\alpha=(3-\sqrt{5})/2$. The transformation $T$ is regular since $\alpha$ is irrational. Note that $1-\alpha$ is a separation point which is not a singularity since $T$ is also a 2-interval exchange transformation.

The following necessary condition for minimality of an interval exchange transformation is useful. A permutation $\pi$ of an ordered set $A$ is called *decomposable* if there exists an element $b\in A$ such that the set $B$ of elements strictly less than $b$ is nonempty and such that $\pi(B)=B$. Otherwise it is called *indecomposable*. If an interval exchange transformation $T=T_{\lambda,\pi}$ is minimal, the permutation $\pi$ is indecomposable. Indeed, if $B$ is a set as above, the set of orbits of the points in the set $S=\cup_{a\in B}I_a$ is closed and strictly included in $[\ell,r[$. The following example shows that the indecomposability of $\pi$ is not sufficient for $T$ to be minimal.

**Example 6.1.8** Let $A=\{a,b,c\}$ and $\lambda$ be such that $\lambda_a=\lambda_c$. Let $\pi$ be the transposition $(ac)$. Then $\pi$ is indecomposable but $T_{\lambda,\pi}$ is not minimal since it is the identity on $I_b$.

The iteration of an $s$-interval exchange transformation is, in general, an interval exchange transformation operating on a larger number of semi-interval.

**Proposition 6.1.9** *Let $T$ be a regular $s$-interval exchange transformation. Then, for any $n\ge 1$, $T^n$ is a regular $n(s-1)+1$-interval exchange transformation.*

*Proof.* Since $T$ is regular, the set $\cup_{i=0}^{n-1} T^{-i}(\mu)$ where $\mu$ runs over the set of $s-1$ nonzero separation points of $T$ has $n(s-1)$ elements. These points partition the interval $[\ell, r[$ in $n(s-1)+1$ semi-intervals on which $T$ is a translation. ∎

We close this subsection with a lemma that will be useful in the nex chapter.

Let $T$ be an interval exchange transformation relative to a partition $(I_i)_{i=1}^s$ and let $(\alpha_i)_{j=1}^s$ be the translations values of $T$. We say that $\alpha_{m_1} + \alpha_{m_2} + \ldots + \alpha_{m_m}$ is an *$m$-translation value* of $T$ if there exists a point $z_0 \in I_{m_1} \cap T^{-1}(I_{m_2}) \cap \cdots \cap T^{-m+1}(I_{m_m})$. Roughly speaking, iterating $T$ we can start from $I_{m_1}$ and arrive to $I_{m_m}$ in exactly $m$ steps, passing (in order) through $I_{m_2}, \ldots I_{m_{m-1}}$.

Moreover, $\alpha_{m_1} + \alpha_{m_2} + \ldots + \alpha_{m_m}$ is one of the translation values of the transformation $T^m$ (namely the one corresponding to the semi-interval containing the point $z_0$).

Note that when $T$ is minimal, every $m$-translation value of $T$, with $m > 0$, is different from zero.

**Lemma 6.1.10** *Let $T$ be a minimal interval exchange transformation over an interval $I$. For every $N > 0$ there exists an $\varepsilon > 0$ such that for every $z \in I$ and for every $n > 0$, one has*

$$|T^n(z) - z| < \varepsilon \quad \Longrightarrow \quad n \geq N.$$

*Proof.* Let $\alpha_1, \alpha_2, \ldots, \alpha_s$ be the translation values of $T$. For every $N > 0$ it is sufficient to choose

$$\varepsilon = \min \left\{ \left| \sum_{j=1}^{M} \alpha_{i_j} \right| \;\; | \quad M \leq N \;\text{ and }\; \sum_{j=1}^{M} \alpha_{i_j} \in V_M(T) \right\}.$$

where $V_M(T)$ denotes the set of $M$-translation values of $T$. ∎

### 6.1.2 Natural coding

Let $T$ be an interval exchange transformation relative to $(I_a)_{a \in A}$. For a given real number $z \in [\ell, r[$, the *natural coding* of $T$ relative to $z$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots$ on the alphabet $A$ defined by

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

**Example 6.1.11** Let $\alpha = (3 - \sqrt{5})/2$ and let $R$ be the rotation of angle $\alpha$ on $[0, 1[$ as in Example 6.1.1. The natural coding of $R$ relative to $\alpha$ is the Fibonacci $x = abaab \cdots$ definen in Example 1.1.2.

For a word $w = b_0 b_1 \cdots b_{m-1}$, let $I_w$ be the set

$$I_w = I_{b_0} \cap T^{-1}(I_{b_1}) \cap \ldots \cap T^{-m+1}(I_{b_{m-1}}). \tag{6.1}$$

Note that each $I_w$ is a semi-interval. Indeed, this is true if $w$ is a letter. Next, assume that $I_w$ is a semi-interval. Then for any $a \in A$, $T(I_{aw}) = T(I_a) \cap$

$I_w$ is a semi-interval since $T(I_a)$ is a semi-interval by definition of an interval exchange transformation. Since $I_{aw} \subset I_a$, $T(I_{aw})$ is a translate of $I_{aw}$, which is therefore also a semi-interval. This proves the property by induction on the length. The semi-interval $I_w$ is the set of points $z$ such that the natural coding of the transformation relative to $z$ has $w$ as a prefix, that is for any $n \geq 0$

$$a_n a_{n+1} \cdots a_{n+m-1} = w \iff T^n(z) \in I_w. \tag{6.2}$$

Set $J_w = T^m(I_w)$. Thus

$$J_w = T^m(I_{b_0}) \cap T^{m-1}(I_{b_1}) \cap \ldots \cap T(I_{b_{m-1}}). \tag{6.3}$$

In particular, we have $J_a = T(I_a)$ for $a \in A$. Note that each $J_w$ is a semi-interval. Indeed, this is true if $w$ is a letter. Next, for any $a \in A$, we have $T^{-1}(J_{wa}) = J_w \cap I_a$. This implies as above that $J_{wa}$ is a semi-interval and proves the property by induction. We set by convention $I_\varepsilon = J_\varepsilon = [0, 1[$. Then one has for any $n \geq 0$

$$a_n a_{n+1} \cdots a_{n+m-1} = w \iff T^n(z) \in I_w \tag{6.4}$$

and

$$a_{n-m} a_{n-m+1} \cdots a_{n-1} = w \iff T^n(z) \in J_w \tag{6.5}$$

Let $(\alpha_a)_{a \in A}$ be the translation values of $T$. Note that for any word $w$,

$$J_w = I_w + \alpha_w \tag{6.6}$$

with $\alpha_w = \sum_{j=0}^{m-1} \alpha_{b_j}$ as one may verify by induction on $|w| = m$. Indeed it is true for $m = 1$. For $m \geq 2$, set $w = ua$ with $a = b_{m-1}$. One has $T^m(I_w) = T^{m-1}(I_w) + \alpha_a$ and $T^{m-1}(I_w) = I_w + \alpha_u$ by the induction hypothesis and the fact that $I_w$ is included in $I_u$. Thus $J_w = T^m(I_w) = I_w + \alpha_u + \alpha_a = I_w + \alpha_w$. Equation (6.6) shows in particular that the restriction of $T^{|w|}$ to $I_w$ is a translation.

Note that the semi-interval $J_w$ is the set of points $z$ such that the natural coding of $T^{-|w|}(z)$ has $w$ as a prefix.

### 6.1.3 Interval exchange sets

Let $T$ be an interval exchange set. The set $\mathcal{L}(T) = \mathrm{Fac}\left(\bigcup_{z \in [\ell, r[} \Sigma_T(z)\right)$ is called the *interval exchange set* relative to $T$. An interval exchange set is clearly biextendable.

If $T$ is minimal, one has $w \in \mathrm{Fac}(\Sigma_T(z))$ if and only if $I_w \neq \emptyset$. Thus the set $\mathrm{Fac}(\Sigma_T(z))$ does not depend on $z$ and we have $\mathcal{L}(T) = \mathrm{Fac}\,(\Sigma_T(z))$ for all $z$ (as for Sturmian words, see [52]). Since this set depends only on $T$, we denote it by $\mathcal{L}(T)$. When $T$ is regular (resp. minimal), such a set is called a *regular interval exchange set* (resp. a minimal interval exchange set).

Let $X$ be the closure of the set of all $\Sigma_T(z)$ for $z \in [\ell, r[$ and let $S$ be the shift on $X$ defined by $S(x) = y$ with $y_n = x_{n+1}$ for $n \geq 0$. The pair $(X, S)$ is a *symbolic dynamical system*, formed of a topological space $X$ and a continuous transformation $S$. Such a system is said to be minimal if the only closed subsets invariant by $S$ are $\emptyset$ or $X$. It is well-known that $(X, S)$ is minimal if and only if the language of $S$, denoted by $\mathcal{L}(S)$, is uniformly recurrent (see for example [52, Theorem 1.5.9]).

Then we have the following commutative diagram of Figure 6.4.

$$\begin{tikzcd}
{[\ell, r[} \arrow[r, "T"] \arrow[d, "\Sigma_T"] & {[\ell, r[} \arrow[d, "\Sigma_T"] \\
X \arrow[r, "S"] & X
\end{tikzcd}$$

Figure 6.4: A commutative diagram.

The map $\Sigma_T$ is neither continuous nor surjective. This can be corrected by embedding the interval $[\ell, r[$ into a larger space on which $T$ is a homeomorphism (see [47] or [16, page 349]). However, if the transformation $T$ is minimal, the symbolic dynamical system $(X, S)$ is minimal (see [16, page 392]). Thus, we obtain the following statement.

**Proposition 6.1.12** *For any minimal interval exchange transformation $T$, the set $\mathcal{L}(T)$ is uniformly recurrent.*

Note that for a regular interval exchange transformation $T$, the map $\Sigma_T$ is injective (see [47, page 30]).

**Example 6.1.13** Let $T$ be the transformation of Example 6.1.7. Since $T$ is minimal, the set $\mathcal{L}(T)$ is uniformly recurrent. The words of length at most 5 of the set $S = \mathrm{Fac}(T)$ are represented in Figure 6.5 on the left.

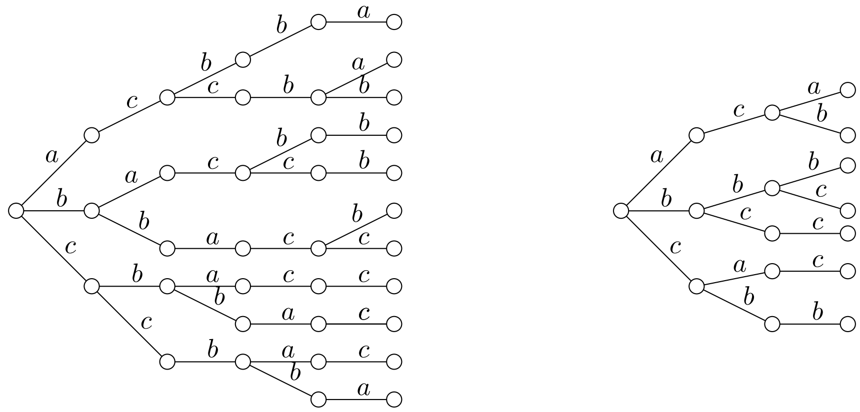

Figure 6.5: The words of length $\leq 5$ of the set $S$ and the words of length $\leq 3$ of its derived set.

Since $T = R^2$, where $R$ is the transformation of Example 6.1.1, the natural coding of $T$ relative to $\alpha$ is the infinite word $y = \gamma^{-1}(x)$ where $x$ is the Fibonacci word and $\gamma$ is the morphism defined by $\gamma(a) = aa$, $\gamma(b) = ab$, $\gamma(c) = ba$. One has

$$y = baccbaccbbacbbacbbacc\cdots \tag{6.7}$$

Actually, the word $y$ is the fixed point $g^\omega(b)$ of the primitive morphism

$$g : a \mapsto baccb\ b \mapsto bacc\ c \mapsto bacb.$$

This follows from the fact that the cube of the Fibonacci morphism $f : a \mapsto ab, b \mapsto a$ sends each letter on a word of odd length and thus sends words of even length on words of even length.

In Section 7.3 we will give a sufficient condition for an interval exchange set to be primitive morphic (Theorem 7.3.12).

The following is an elementary property of the intervals $I_u$ which will be used below. We denote by $<_1$ the lexicographic order on $A^*$ induced by the order $<_1$ on $A$.

**Proposition 6.1.14** *One has $I_u < I_v$ if and only if $u <_1 v$ and $u$ is not a prefix of $v$.*

*Proof.* For a word $u$ and a letter $a$, it results from (6.1) that $I_{ua} = I_u \cap T^{-|u|}(I_a)$. Since $(I_a)_{a\in A}$ is an ordered partition, this implies that $(T^{|u|}(I_u) \cap I_a)_{a\in A}$ is an ordered partition of $T^{|u|}(I_u)$. Since the restriction of $T^{|u|}$ to $I_u$ is a translation, this implies that $(I_{ua})_{a\in A}$ is an ordered partition of $I_u$. Moreover, for two words $u, v$, it results also from (6.1) that $I_{uv} = I_u \cap T^{-|u|}(I_v)$. Thus $I_{uv} \subset I_u$.

Assume that $u <_1 v$ and that $u$ is not a prefix of $v$. Then $u = pas$ and $v = pbt$ with $p \in A^*$ and $a, b$ two letters such that $a <_1 b$. Then we have $I_{pa} < I_{pb}$, with $I_u \subset I_{pa}$ and $I_v \subset I_{pb}$ whence $I_u < I_v$.

Conversely, assume that $I_u < I_v$. Since $I_u \cap I_v = \emptyset$, the words $u, v$ cannot be comparable for the prefix order. Set $u = pas$ and $v = pbt$ with $a, b$ two distinct letters. If $b <_1 a$, then $I_v < I_u$ as we have shown above. Thus $a <_1 b$ which implies $u <_1 v$. ■

We denote by $<_2$ the order on $A^*$ defined by $u <_2 v$ if $u$ is a proper suffix of $v$ or if $u = waz$ and $v = tbz$ with $a <_2 b$. Thus $<_2$ is the lexicographic order on the reversal of the words induced by the order $<_2$ on the alphabet.

We denote by $\pi$ the morphism from $A^*$ onto itself which extends to $A^*$ the permutation $\pi$ on $A$. Then $u <_2 v$ if and only if $\pi^{-1}(\tilde{u}) <_1 \pi^{-1}(\tilde{v})$, where $\tilde{u}$ denotes the reversal of the word $u$.

The following statement is the analogue of Proposition 6.1.14.

**Proposition 6.1.15** *Let $T_{\lambda,\pi}$ be an interval exchange transformation. One has $J_u < J_v$ if and only if $u <_2 v$ and $u$ is not a suffix of $v$.*

*Proof.* Let $(I'_a)_{a\in A}$ be the family of semi-intervals defined by $I'_a = J_{\pi(a)}$. Then the interval exchange transformation $T'$ relative to $(I'_a)$ with translation values $-\alpha_a$ is the inverse of the transformation $T$. The semi-intervals $I'_w$ defined by Equation (6.1) with respect to $T'$ satisfy $I'_w = J_{\pi(\tilde{w})}$ or equivalently $J_w = I'_{\pi^{-1}(\tilde{w})}$. Thus, $J_u < J_v$ if and only if $I'_{\pi^{-1}(\tilde{u})} < I'_{\pi^{-1}(\tilde{v})}$ if and only if (by Proposition 6.1.14) $\pi^{-1}(\tilde{u}) <_1 \pi^{-1}(\tilde{v})$ or equivalently $u <_2 v$. ∎

### 6.1.4 Planar tree sets

Recall from Chapter 3 that a tree set $S$ is called a planar tre set with respect to two orders $<_1$ and $<_2$ if for for any $w \in S$ the graph $\mathcal{E}(w)$ is compatible with $<_1$ and $<_2$ (see Section 3.1.1), that is if for any $(a, b), (c, d) \in B(w)$, one has

$$a <_2 c \Longrightarrow b \leq_1 d.$$

Let us consider the two orders $<_1$ and $<_2$ on $A^*$ defined in Section 6.1.3.

The following result is a generalization of a result from [39] with a converse (see below).

**Theorem 6.1.16** *Let $T$ be an interval exchange transformation with exactly $C$ connections, all of length $0$. Then $\mathcal{L}(T)$ is a planar tree set of characteristic $C + 1$ with respect to $<_1$ and $<_2$.*

In order to prove Theorem 6.1.16 we need some preliminary result.

**Lemma 6.1.17** *Let $T$ be an interval exchange transformation. For every nonempty word $w$ and letter $a \in A$, one has*

*(i)* $a \in L(w) \Longleftrightarrow I_w \cap J_a \neq \emptyset$,

*(ii)* $a \in R(w) \Longleftrightarrow I_a \cap J_w \neq \emptyset$.

*Proof.* A letter $a$ is in the set $L(w)$ if and only if $aw \in \mathcal{L}(T)$. As we have seen before, this is equivalent to $J_{aw} \neq \emptyset$. One has $J_{aw} = T(I_{aw}) = T(I_a) \cap I_w = J_a \cap I_w$, whence point (i). Point (ii) is proved symmetrically. ∎

We say that a path in a graph is *reduced* if it does not use twice consecutively the same edge.

**Lemma 6.1.18** *Let $T$ be an interval exchange transformation over $I$ without connection of length $\geq 1$. Let $w \in \mathcal{L}(T)$ and $a, b \in L(w)$ (resp. $a, b \in R(w)$). Then $1 \otimes a, 1 \otimes b$ (resp. $a \otimes 1, b \otimes 1$) are in the same connected component of $\mathcal{E}(w)$ if and only if $J_a, J_b$ (resp. $I_a, I_b$) are in the same component of $I$.*

*Proof.* Let $a \in L(w)$. Since the set $\mathcal{L}(T)$ is biextendable, there exists a letter $c$ such that $(1 \otimes a, c \otimes 1) \in \mathcal{E}(w)$. Using the same reasoning as that in Lemma 6.1.17, one has $J_a \cap I_{wc} \neq \emptyset$. Since $I_{wc} \subset I_w$, one has in particular

$J_a \cap I_w \neq \emptyset$. This proves that $J_a$ and $I_w$ belong to the same component of $I$ for every $a \in L(w)$.

Conversely, suppose that $a, b \in L(w)$ are such that $J_a$ and $J_b$ belong to the same component of $I$. We may assume that $a <_2 b$. Then, there is a reduced path $(1 \otimes a_1, b_1 \otimes 1, \ldots, b_{n-1} \otimes 1, 1 \otimes a_n)$ in $\mathcal{E}(w)$ (see Figure 6.6) with $a = a_1$, $b = a_n$, $a_1 <_2 \cdots <_2 a_n$ and $wb_1 <_1 \cdots <_1 wb_{n_1}$. Indeed, by hypothesis, we have no connection of length $\geq 1$. Thus, for every $1 \leq i < n$, one has $J_{a_i} \cap I_{wb_i} \neq \emptyset$ and $J_{a_{i+1}} \cap I_{wb_i} \neq \emptyset$. Therefore, $a$ and $b$ are in the same connected component of $\mathcal{E}(w)$.

The symmetrical statement is proved similarly. ∎

We can now prove the main result of this section.

*Proof of Theorem 6.1.16.* Let us first prove that for any $w \in \mathcal{L}(T)$, the graph $\mathcal{E}(w)$ is acyclic. Assume that $(1 \otimes a_1, b_1 \otimes 1, \ldots, 1 \otimes a_n, b_n \otimes 1)$ is a reduced path in $\mathcal{E}(w)$ with $a_1, \ldots, a_n \in L(w)$ and $b_1, \ldots, b_n \in R(w)$. Suppose that $n \geq 2$ and that $a_1 <_2 a_2$. Then one has $a_1 <_2 \cdots <_2 a_n$ and $wb_1 <_1 \cdots <_1 wb_n$ (see Figure 6.6). Thus one cannot have an edge $(a_1, b_n)$ in the graph $\mathcal{E}(w)$.

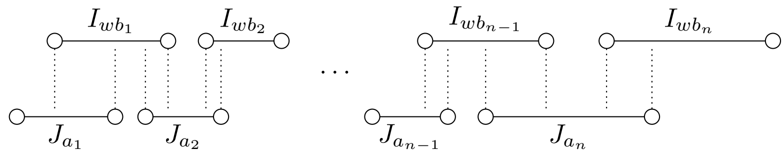


Figure 6.6: A path from $a_1$ to $a_n$ in $\mathcal{E}(w)$.

Let us now prove that the extension graph of the empty word is a union of $C + 1$ trees. Let $a, b \in A$. If $J_a$ and $J_b$ are in the same component of $I$, then $1 \otimes a, 1 \otimes b$ are in the same connected component of $\mathcal{E}(\varepsilon)$ by Lemma 6.1.18. Thus $\mathcal{E}(\varepsilon)$ is a union of $C + 1$ trees.

If $w \in \mathcal{L}(T)$ is a nonempty word and $a, b \in L(w)$, then $J_a$ and $J_b$ are in the same component of $I$, by Lemma 6.1.17, and thus $a$ and $b$ are in the same connected component of $\mathcal{E}(w)$ by Lemma 6.1.18. Thus $\mathcal{E}(w)$ is a tree.

Finally, the set $\mathcal{L}(T)$ is compatible with the orders $<_1$ and $<_2$. Indeed, let $(a, b), (c, d) \in B(w)$ for a word $w \in \mathcal{L}(T)$. Let us suppose that $a <_2 c$. By Proposition 6.1.15, one has $J_a < J_c$.

Moreover, by Lemma 6.1.17, one has $I_{wb} \cap J_a \neq \emptyset$ and $I_{wd} \cap J_c \neq \emptyset$. This implies that $I_{wb} < I_{wd}$ (see Figure 6.6). By Proposition 6.1.14, one has $wb <_1 wd$. Hence $b <_1 d$. ∎

**Example 6.1.19** Let $T$ be the interval exchange transformation of Example 6.1.3. The set $\mathcal{L}(T)$ is a tree set of characteristic 2. In Figure 6.7 are represented the extension graphs of the empty word (left) and of the letters $a$ (center) and $b$ (right).

By Theorem 6.1.16, a regular interval exchange set is a planar tree set of characteristic 1, and thus in particular a tree set of characteristic 1.

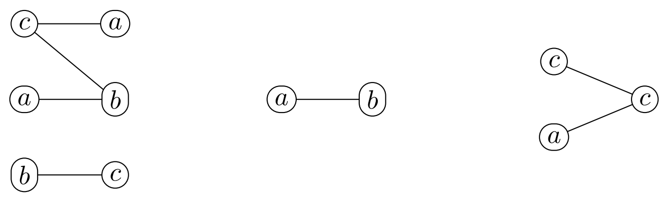

Figure 6.7: The extension graphs of $\varepsilon$ (left), $a$ (center) and $b$ (right).

The main result of [39] states that a (uniformly) recurrent set $S$ on an alphabet $A$ is a regular interval exchange set if and only if $A \subset S$ and there exist two orders $<_1$ and $<_2$ on $A$ such that the following conditions are satisfied for any word $w \in S$.

(i) The set $L(w)$ (resp. $R(w)$) is formed of consecutive elements for the order $<_2$ (resp. $<_1$).

(ii) For $(a, b), (c, d) \in B(w)$, if $a <_2 c$, then $b \leq_1 d$.

(iii) If $a, b \in L(w)$ are consecutive for the order $<_2$, then the set $R(aw) \cap R(bw)$ is a singleton.

It is easy to see that a biextendable set $S$ containing $A$ satisfies (ii) and (iii) if and only if it is a planar tree set of character of characteristic 1. Actually, in this case, it automatically satisfies also condition (i). Indeed, let us consider a word $w$ and $a, b, c \in A$ with $a <_1 b <_1 c$ such that $wa, wc \in S$ but $wb \notin S$. Since $b \in S$ there is a (possibly empty) suffix $v$ of $w$ such that $vb \in S$. We choose $v$ of maximal length. Since $wb \notin S$, we have $w = uv$ with $u$ nonempty. Let $d$ be the last letter of $u$. Then we have $dva, dvc \in S$ and $dvb \notin S$. Since $\mathcal{E}(v)$ is a tree and $b \in R(v)$, there is a letter $e \in L(v)$ such that $evb \in S$. But $e <_2 d$ and $d <_2 e$ are both impossible since $\mathcal{E}(v)$ is compatible with $<_2$ and $<_1$. Thus we reach a contradiction.

This shows that the original reformulation of the main result of [39] is equivalent to the following one.

**Theorem 6.1.20 (Ferenczi, Zamboni)** *A set $S$ is a regular interval exchange set on the alphabet $A$ if and only if it is a recurrent planar tree set of characteristic 1.*

We have already seen that the Tribonacci set is a tree set which is not a planar tree set (Example 3.1.9). The next example shows that there are recurrent tree sets which are neither Sturmian nor regular interval exchange sets.

**Example 6.1.21** Let $S$ be the Tribonacci set on the alphabet $A = \{a, b, c\}$ and let $f : \{x, y, z, t, u\}^* \to A^*$ be the coding morphism for $X = S \cap A^2$ defined by $f(x) = aa$, $f(y) = ab$, $f(z) = ac$, $f(t) = ba$, $f(u) = ca$. By Theorem 4.3.5, the set $W = f^{-1}(S)$ is a recurrent tree set of characteristic 1.

It is not Sturmian since $y$ and $t$ are two right-special words. It is not either a regular interval exchange set. Indeed, for any right-special word $w$ of $W$, one has $\mathrm{Card}(R(w)) = 3$. This is not possible in a regular interval exchange set $T$ since, $\Sigma_T$ being injective, the length of the interval $J_w$ tends to 0 as $|w|$ tends to infinity and it cannot contain several separation points. It can of course also be verified directly that $W$ is not a planar tree set.

## 6.2 Bifix codes and interval exchanges

In this section we study the connection between regular interval exchange sets and bifix codes. Firstly, we introduce in Section 6.2.1 a result concerning an invariant probability distribution on an interval exchange set (Proposition 6.2.2). We use this result to show that we can refine the partition of subintervals $(I_a)_{a\in A}$ defining an interval exchange (Proposition 6.2.3).

In Section 6.2.2 we generalize this result also for the subintervals $(J_a)_{a\in A}$ (Proposition 6.2.5). Next, given an interval exchange $T$, we define a transformation $T_f$ associated to a maximal bifix decoding and we show the connection of the natural codings with respect to $T$ and $T_f$ (Proposition 6.2.8).

In Section 6.2.3 we prove that $T_f$ is regular provided the original transformation $T$ is regular (Theorem 6.2.10). Moreover, we prove that the family of regualar interval exchange sets is closed under maximal bifix decoding (Theorem 6.2.11) and, as a corollary, so is the family of recurrent planar tree sets of characteristic 1 (Corollary 6.2.13).

Finally, in Section 6.2.4, we use the Finite Index Basis Theorem 4.2.1 to define interval exchanges on a stack and to give an alternative proof of Theorem 6.2.10.

### 6.2.1 Prefix and bifix codes

Recall from Section 1.2 the definition of prefix, suffix and bifix code. Following the terminology of Section 2.1.1, we define a (left and right) *invariant probability distribution* on an alphabet $A^*$ a map $\lambda : A^* \to [0, 1]$ such that $\lambda(\varepsilon) = 1$ and, for any word $w$

$$\sum_{a\in A} \lambda(aw) = \sum_{a\in A} \lambda(wa) = \lambda(w). \tag{6.8}$$

Let $T_{\lambda,\pi}$ be an interval exchange transformation on an interval $[\ell, r[$ For any word $w \in A^*$, denote by $|I_w|$ the length of the semi-interval $I_w$ defined by Equation (6.1). Set $\lambda(w) = |I_w| / (r - \ell)$. Then $\lambda(\varepsilon) = 1$ and for any word $w$, Equation (6.8) holds and thus $\lambda$ is an invariant probability distribution.

The fact that $\lambda$ is an invariant probability measure is equivalent to the fact that the Lebesgue measure on $[\ell, r[$ is invariant by $T$. It is known that almost all regular interval exchange transformations have no other invariant probability measure (and thus are uniquely ergodic, see [16] for references).

**Example 6.2.1** Let $S$ be the set of factors of the Fibonacci word (see Example 1.1.2). As seen in Example 6.1.11, it is an interval exchange set relative to the rotation $R$ defined in Example 6.1.1. The values of the map $\lambda$ on the words of length at most 4 in $S$ are indicated in Figure 6.8.

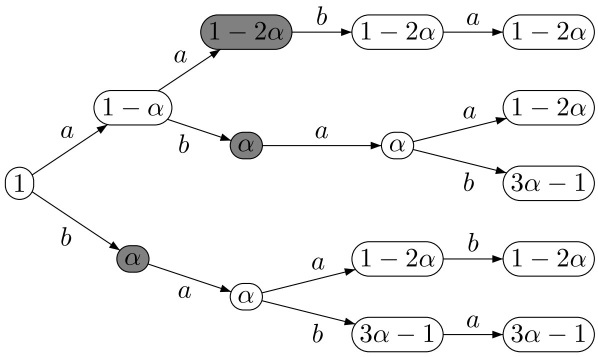


Figure 6.8: The invariant probability distribution on the Fibonacci set.

The following result is a particular case of [7, Proposition 3.3.4].

**Proposition 6.2.2** *Let $T$ be a minimal interval exchange transformation, let $S = \mathcal{L}(T)$ and let $\lambda$ be an invariant probability distribution on $S$. For any finite $S$-maximal prefix code $X$, one has $\sum_{x\in X} \lambda(x) = 1$.*

The following statement is connected with Proposition 6.2.2.

**Proposition 6.2.3** *Let $T$ be a minimal interval exchange transformation relative to $(I_a)_{a\in A}$, let $S = \mathcal{L}(T)$ and let $X$ be a finite $S$-maximal prefix code ordered by $<_1$. The family $(I_w)_{w\in X}$ is an ordered partition of $[\ell, r[$.*

*Proof.* By Proposition 6.1.14, the sets $(I_w)$ for $w \in X$ are pairwise disjoint. Let $\pi$ be the invariant probability distribution on $S$ defined by $\pi(w) = |I_w|/(r - \ell)$. By Proposition 6.2.2, we have $\sum_{w\in X} \pi(w) = 1$. Thus the family $(I_w)_{w\in X}$ is a partition of $[\ell, r[$. By Proposition 6.1.14 it is an ordered partition. ■

**Example 6.2.4** Let $R$ be the rotation of angle $\alpha = (3 - \sqrt{5})/2$. The set $S = \mathcal{L}(T)$ is the Fibonacci set. The set $X = \{aa, ab, b\}$ is an $S$-maximal prefix code (see the grey nodes in Figure 6.8).The partition of $[0, 1[$ corresponding to $X$ is

$$I_{aa} = [0, 1 - 2\alpha[, \quad I_{ab} = [1 - 2\alpha, 1 - \alpha[, \quad I_b = [1 - \alpha, 1[.$$

The values of the lengths of the semi-intervals (the invariant probability distribution) can also be read on Figure 6.8.

A symmetric statement holds for an $S$-maximal suffix code, namely that the family $(J_w)_{w\in X}$ is an ordered partition of $[\ell, r[$ for the order $<_2$ on $X$.

### 6.2.2 Maximal bifix codes

The following result shows that bifix codes have a natural connection with interval exchange transformations.

**Proposition 6.2.5** *If $X$ is a finite $S$-maximal bifix code, with $S$ as in Proposition 6.2.3, the families $(I_w)_{w\in X}$ and $(J_w)_{w\in X}$ are ordered partitions of $[\ell, r[$, relatively to the orders $<_1$ and $<_2$ respectively.*

*Proof.* This results from Proposition 6.2.3 and its symmetric and from the fact that, since $S$ is recurrent, a finite $S$-maximal bifix code is both an $S$-maximal prefix code and an $S$-maximal suffix code. ∎

Let $T$ be a regular interval exchange transformation relative to $(I_a)_{a\in A}$. Let $(\alpha_a)_{a\in A}$ be the translation values of $T$. Set $S = \mathcal{L}(T)$. Let $X$ be a finite $S$-maximal bifix code on the alphabet $A$.

Let $T_X$ be the transformation on $[\ell, r[$ defined by

$$T_X(z) = T^{|u|}(z) \quad \text{if} \quad z \in I_u$$

with $u \in X$. The transformation is well-defined since, by Proposition 6.2.5, the family $(I_u)_{u\in X}$ is a partition of $[\ell, r[$.

Let $f : B^* \to A^*$ be a coding morphism for $X$. Let $(K_b)_{b\in B}$ be the family of semi-intervals indexed by the alphabet $B$ with $K_b = I_{f(b)}$. We consider $B$ as ordered by the orders $<_1$ and $<_2$ induced by $f$. Let $T_f$ be the interval exchange transformation relative to $(K_b)_{b\in B}$. Its translation values are $\beta_b = \sum_{j=0}^{m-1} \alpha_{a_j}$ for $f(b) = a_0a_1 \cdots a_{m-1}$. The transformation $T_f$ is called the *transformation associated* with $f$.

**Proposition 6.2.6** *Let $T$ be a regular interval exchange transformation relative to $(I_a)_{a\in A}$ and let $S = \mathcal{L}(T)$. If $f : B^* \to A^*$ is a coding morphism for a finite $S$-maximal bifix code $X$, one has $T_f = T_X$.*

*Proof.* By Proposition 6.2.5, the family $(K_b)_{b\in B}$ is a partition of $[\ell, r[$ ordered by $<_1$. For any $w \in X$, we have by Equation (6.6) $J_w = I_w + \alpha_w$ and thus $T_X$ is the interval exchange transformation relative to $(K_b)_{b\in B}$ with translation values $\beta_b$. ∎

In the sequel, under the hypotheses of Proposition 6.2.6, we consider $T_f$ as an interval exchange transformation. In particular, the natural coding of $T_f$ relative to $z \in [\ell, r[$ is well-defined.

**Example 6.2.7** Let $S$ be the Fibonacci set. It is the set of factors of the Fibonacci word, which is a natural coding of the rotation $R$ of angle $\alpha = (3 - \sqrt{5})/2$ relative to $\alpha$ (see Example 6.1.11). Let $X = \{aa, ab, ba\}$ and let $f$ be the coding morphism defined by $f(u) = aa$, $f(v) = ab$, $f(w) = ba$. The two partitions of $[0, 1[$ corresponding to $T_f$ are

$$I_u = [0, 1 - 2\alpha[, \quad I_v = [1 - 2\alpha, 1 - \alpha[ \quad I_w = [1 - \alpha, 1[$$

and

$$J_v = [0, \alpha[, \quad J_w = [\alpha, 2\alpha[ \quad J_u = [2\alpha, 1[.$$

The transformation $T_f$ is the same as the one represented in Figure 6.2 where $u, v, w$ instead of, respectively, $a, b, c$.

It is actually a representation on 3 intervals of the rotation of angle $2\alpha$. Note that the point $z = 1 - \alpha$ is a separation point which is not a singularity of $T_f$.

The first row of Table 6.1 gives the two orders on $X$. The next two rows give the two orders for each of the two other $S$-maximal bifix codes of $S$-degree 2 (there are actually exactly three $S$-maximal bifix codes of $S$-degree 2 in the Fibonacci set, see [7]).

| $(X, <_1)$ | $(X, <_2)$ |
|---|---|
| $aa, ab, ba$ | $ab, ba, aa$ |
| $a, baab, bab$ | $bab, baab, a$ |
| $aa, aba, b$ | $b, aba, aa$ |

Table 6.1: The two orders on the three $S$-maximal bifix codes of $S$-degree 2.

Let $T$ be a minimal interval exchange transformation on the alphabet $A$. Let $x$ be the natural coding of $T$ relative to some $z \in [\ell, r[$. Set $S = \mathrm{Fac}(x)$. Let $X$ be a finite $S$-maximal bifix code. Let $f : B^* \to A^*$ be a morphism which maps bijectively $B$ onto $X$. Since $S$ is recurrent, the set $X$ is an $S$-maximal prefix code. Thus $x$ has a prefix $x_0 \in X$. Set $x = x_0 x'$. In the same way $x'$ has a prefix $x_1$ in $X$. Iterating this argument, we see that $x = x_0 x_1 \cdots$ with $x_i \in X$. Consequently, there exists an infinite word $y$ on the alphabet $B$ such that $x = f(y)$. The word $y$ is the *decoding* of the infinite word $x$ with respect to $f$.

**Proposition 6.2.8** *The decoding of $x$ with respect to $f$ is the natural coding of the transformation associated with $f$ relative to $z$: $\Sigma_T(z) = f\left(\Sigma_{T_f}(z)\right)$.*

*Proof.* Let $y = b_0 b_1 \cdots$ be the decoding of $x$ with respect to $f$. Set $x_i = f(b_i)$ for $i \geq 0$. Then, for any $n \geq 0$, we have

$$T_f^n(z) = T^{|u_n|}(z) \tag{6.9}$$

with $u_n = x_0 \cdots x_{n-1}$ (note that $|u_n|$ denotes the length of $u_n$ with respect to the alphabet $A$). Indeed, this is is true for $n = 0$. Next $T_f^{n+1}(z) = T_f(t)$ with $t = T_f^n(z)$. Arguing by induction, we have $t = T^{|u_n|}(z)$. Since $x = u_n x_n x_{n+1} \cdots$, $t$ is in $I_{x_n}$ by (6.2). Thus by Proposition 6.2.6, $T_f(t) = T^{|x_n|}(t)$ and we obtain $T_f^{n+1}(z) = T^{|x_n|}(T^{|u_n|}(z)) = T^{|u_{n+1}|}(z)$ proving (6.9). Finally, for $u = f(b)$ with $b \in B$,

$$b_n = b \iff x_n = u \iff T^{|u_n|}(z) \in I_u \iff T_f^n(z) \in I_u = K_b$$

showing that $y$ is the natural coding of $T_f$ relative to $z$. ∎

**Example 6.2.9** Let $T, \alpha, X$ and $f$ be as in Example 6.2.7. Let $x = abaababa\cdots$ be the Fibonacci word. We have $x = \Sigma_T(\alpha)$. The decoding of $x$ with respect to $f$ is $y = vuwwv\cdots$.

### 6.2.3 Maximal bifix decoding

The following result shows that, for the coding morphism $f$ of a finite $S$-maximal bifix code, the map $T \mapsto T_f$ preserves the regularity of the transformation.

**Theorem 6.2.10** *Let $T$ be a regular interval exchange transformation and let $S = \mathcal{L}(T)$. For any finite $S$-maximal bifix code $X$ with coding morphism $f$, the transformation $T_f$ is regular.*

*Proof.* Set $A = \{a_1, a_2, \ldots, a_s\}$ with $a_1 <_1 a_2 <_1 \cdots <_1 a_s$. We denote $\delta_i = \delta_{a_i}$. By hypothesis, the orbits of $\delta_2, \ldots, \delta_s$ are infinite and disjoint. Set $X = \{x_1, x_2, \ldots, x_t\}$ with $x_1 <_1 x_2 <_1 \cdots <_1 x_t$. Let $d$ be the $S$-degree of $X$.

For $x \in X$, denote by $\delta_x$ the left boundary of the semi-interval $J_x$. For each $x \in X$, it follows from Equation (6.3) that there is an $i \in \{1, \ldots, s\}$ such that $\delta_x = T^k(\delta_i)$ with $0 \leq k < |x|$. Moreover, we have $i = 1$ if and only if $x = x_1$. Since $T$ is regular, the index $i \neq 1$ and the integer $k$ are unique for each $x \neq x_1$. And for such $x$ and $i$, by (6.5), we have $\Sigma_T(\delta_i) = u\Sigma_T(\delta_x)$ with $u$ a proper suffix of $x$.

We now show that the orbits of $\delta_{x_2}, \ldots, \delta_{x_t}$ for the transformation $T_f$ are infinite and disjoint. Assume that $\delta_{x_p} = T_f^n(\delta_{x_q})$ for some $p, q \in \{2, \ldots, t\}$ and $n \in \mathbb{Z}$. Interchanging $p, q$ if necessary, we may assume that $n \geq 0$. Let $i, j \in \{2, \ldots, s\}$ be such that $\delta_{x_p} = T^k(\delta_i)$ with $0 \leq k < |x_p|$ and $\delta_{x_q} = T^\ell(\delta_j)$ with $0 \leq \ell < |x_q|$. Since $T^k(\delta_i) = T_f^n(T^\ell(\delta_j)) = T^{m+\ell}(\delta_j)$ for some $m \geq 0$, we cannot have $i \neq j$ since otherwise the orbits of $\delta_i, \delta_j$ for the transformation $T$ intersect. Thus $i = j$. Since $\delta_{x_p} = T^k(\delta_i)$, we have $\Sigma_T(\delta_i) = u\Sigma_T(\delta_{x_p})$ with $|u| = k$, and $u$ a proper suffix of $x_p$. And since $\delta_{x_p} = T_f^n(\delta_{x_q})$, we have $\Sigma_T(\delta_{x_q}) = x\Sigma_T(\delta_{x_p})$ with $x \in X^*$. Since on the other hand $\delta_{x_q} = T^\ell(\delta_i)$, we have $\Sigma_T(\delta_i) = v\Sigma_T(\delta_{x_q})$ with $|v| = \ell$ and $v$ a proper suffix of $x_q$. We obtain

$$\begin{aligned} \Sigma_T(\delta_i) &= u\Sigma_T(\delta_{x_p}) \\ &= v\Sigma_T(\delta_{x_q}) = vx\Sigma_T(\delta_{x_p}). \end{aligned}$$

Since $|u| = |vx|$, this implies $u = vx$. But since $u$ cannot have a suffix in $X$, $u = vx$ implies $x = \varepsilon$ and thus $n = 0$ and $p = q$. This concludes the proof. ■

Let $f$ be a coding morphism for a finite $S$-maximal bifix code $X \subset S$. The set $f^{-1}(S)$ is called a *maximal bifix decoding* of $S$ (see Chapter 1).

**Theorem 6.2.11** *The family of regular interval exchange sets is closed under maximal bifix decoding.*

*Proof.* Let $T$ be a regular interval exchange transformation over $[\ell, r[$ and let $S = \mathcal{L}(T)$. By Theorem 6.2.10, $T_f$ is a regular interval exchange transformation. We show that $f^{-1}(S) = \mathcal{L}(T_f)$, which implies the conclusion.

Let $x = \Sigma_T(z)$ for some $z \in [\ell, r[$ and let $y = f^{-1}(x)$. Then $S = \mathcal{L}(x)$ and $\mathcal{L}(T_f) = \mathrm{Fac}(y)$. For any $w \in \mathrm{Fac}(y)$, we have $f(w) \in \mathrm{Fac}(x)$ and thus $w \in f^{-1}(S)$. This shows that $\mathcal{L}(T_f) \subset f^{-1}(S)$. Conversely, let $w \in f^{-1}(S)$ and let $v = f(w)$. Since $S = \mathrm{Fac}(x)$, there is a word $u$ such that $uv$ is a prefix of $x$. Set $z' = T^{|u|}(z)$ and $x' = \Sigma_T(z')$. Then $v$ is a prefix of $x'$ and $w$ is a prefix of $y' = f^{-1}(x')$. Since $T_f$ is regular, it is minimal and thus $\mathrm{Fac}(y') = \mathcal{L}(T_f)$. This implies that $w \in \mathcal{L}(T_f)$. ■

We illustrate the proof of Theorem 6.2.10 in the following example.

**Example 6.2.12** Let $T$ be the rotation of angle $\alpha = (3 - \sqrt{5})/2$ (see Example 6.1.1). The set $S = \mathcal{L}(T)$ is the Fibonacci set. Let $X = \{a, baab, babaabaabab, babaabab\}$. The set $X$ is an $S$-maximal bifix code of $S$-degree 3 (see [7]). The values of the $\mu_{x_i}$ (which are the right boundaries of the intervals $I_{x_i}$) and $\delta_{x_i}$ are represented in Figure 6.9.

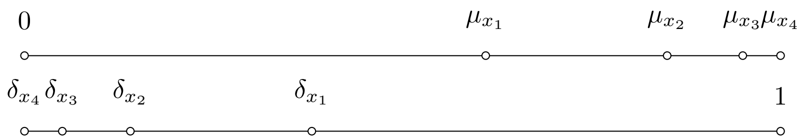


Figure 6.9: The transformation associated with a bifix code of $S$-degree 3.

The infinite word $\Sigma_T(0)$ is represented in Figure 6.10. The value indicated on the word $\Sigma_T(0)$ after a prefix $u$ is $T^{|u|}(0)$. The three values $\delta_{x_4}, \delta_{x_2}, \delta_{x_3}$ correspond to the three prefixes of $\Sigma_T(0)$ which are proper suffixes of $X$.

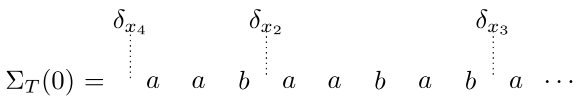


Figure 6.10: The infinite word $\Sigma_T(0)$.

A consequence of Theorem 6.2.11 is the following result.

**Corollary 6.2.13** *The family of recurrent planar tree sets of characteristic 1 is closed under maximal bifix decoding.*

*Proof.* The result easily follows from Theorems 4.3.5 and 6.2.11. ■

The following example shows that Theorem 6.2.11 is not true when $X$ is not bifix.

**Example 6.2.14** Let $S$ be the Fibonacci set and let $X = \{aa, ab, b\}$. The set $X$ is an $S$-maximal prefix code. Let $B = \{u, v, w\}$ and let $f$ be the coding morphism for $X$ defined by $f(u) = aa$, $f(v) = ab$, $f(w) = b$. The set $W = f^{-1}(S)$ is not an interval exchange set. Indeed, we have $vu, vv, wu, wv \in W$. This implies that both $J_v$ and $J_w$ meet $I_u$ and $I_v$, which is impossible in an interval exchange transformation.

### 6.2.4 Subgroups of finite index

Let $S$ be a recurrent set containing the alphabet $A$. Recall from Chapter 3 that $S$ has the *finite index basis property* if the following holds: a finite bifix code $X \subset S$ is an $S$-maximal bifix code of $S$-degree $d$ if and only if it is a basis of a subgroup of index $d$ of the free group $F_A$.

Since a regular interval exchange set is a tree set of characteristic 1, we have the following immediate consequence of Theorem 4.2.1.

**Theorem 6.2.15** *A regular interval exchange set has the finite index basis property.*

We use Theorem 6.2.15 to give another proof of Theorem 6.2.10. For this, we recall the following notion.

Let $T$ be an interval exchange transformation on $I = [\ell, r[$ relative to $(I_a)_{a \in A}$. Let $G$ be a transitive permutation group on a finite set $Q$. Let $\varphi : A^* \to G$ be a morphism and let $\psi$ be the map from $I$ into $G$ defined by $\psi(z) = \varphi(a)$ if $z \in I_a$. The *skew product* of $T$ and $G$ is the transformation $U$ on $I \times Q$ defined by

$$U(z, q) = (T(z), q\psi(z))$$

(where $q\psi(z)$ is the result of the action of the permutation $\psi(z)$ on $q \in Q$). Such a transformation is equivalent to an interval exchange transformation via the identification of $I \times Q$ with an interval obtained by placing the $d = \mathrm{Card}(Q)$ copies of $I$ in sequence. This is called an *interval exchange transformation on a stack* in [19] (see also [65]). If $T$ is regular, then $U$ is also regular.

Let $T$ be a regular interval exchange transformation and let $S = \mathcal{L}(T)$. Let $X$ be a finite $S$-maximal bifix code of $S$-degree $d = d_X(S)$. By Theorem 6.2.15, $X$ is a basis of a subgroup $H$ of index $d$ of $F_A$. Let $G$ be the representation of $F_A$ on the right cosets of $H$ and let $\varphi$ be the natural morphism from $F_A$ onto $G$. We identify the right cosets of $H$ with the set $Q = \{1, 2, \ldots, d\}$ with 1 identified to $H$. Thus $G$ is a transitive permutation group on $Q$ and $H$ is the inverse image by $\varphi$ of the permutations fixing 1.

The transformation induced by the skew product $U$ on $I \times \{1\}$ is clearly equivalent to the transformation $T_f = T_X$ where $f$ is a coding morphism for the $S$-maximal bifix code $X$. Thus $T_X$ is a regular interval exchange transformation.

**Example 6.2.16** Let $T$ be the rotation of Example 6.1.11. Let $Q = \{1, 2, 3\}$ and let $\varphi$ be the morphism from $A^*$ into the symmetric group on $Q$ defined by $\varphi(a) = (23)$ and $\varphi(b) = (12)$. The transformation induced by the skew product of $T$ and $G$ on $I \times \{1\}$ corresponds to the bifix code $X$ of Example 6.2.12. For example, we have $U : (1-\alpha, 1) \to (0, 2) \to (\alpha, 3) \to (2\alpha, 2) \to (3\alpha - 1, 1)$ (see Figure 6.11) and the corresponding word of $X$ is $baab$.

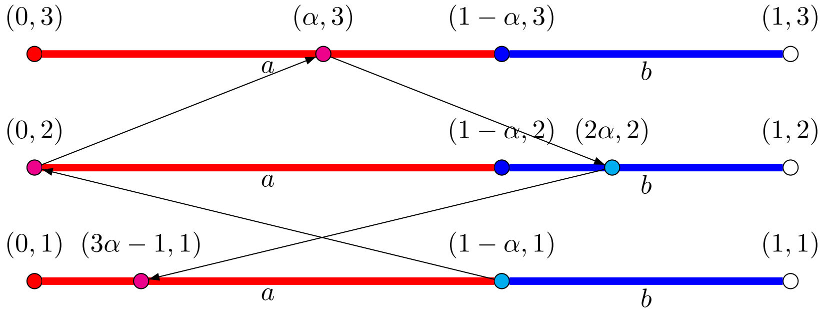


Figure 6.11: The transformation $U$.

# Chapter 7

# Branching Rauzy induction

In this chapter we continue our study of interval exchanges started in Chapter 6. The main tool introduced here is the branching Rauzy induction, a generalization of the one-side Rauzy induction defined in [61].

In Section 7.1 we introduce the definition of admissible semi-interval. This generalize in a natural way the notion of (one-side) admissibility introduced in [61]. We show that all semi-intervals of the form $I_w$ and $J_w$ are admissible (Proposition 7.1.6) and that the induction on any admissible semi-interval preserve regularity (Theorem 7.1.7). Moreover, we prove that the family of regular $s$-interval exchanges is closed by derivation (Theorem 7.1.9).

In Section 7.2 we define the branching Rauzy induction which operates on regular interval transformations. This trasformation generalize the (one-sided) Rauzy induction defined in [61]

We recall the results concerning this classical case (Theorems 7.1.3 and 7.2.1) and we generalize them the branching case (Theorems 7.1.7 and 7.2.3). In particular we characterize the admissible semi-intervals for an interval exchange transformation (Theorem 7.2.3).

Finally, in Section 7.3 we study the case of an interval exchange defined over a qaudratic field. Following the path of Boshernitzan and Carroll in [19], we prove that under certains hypothesis, there are finitely many transformations obtained by the branching Rauzy induction (Theorem 7.3.1). We use this result to prove that the language of a regular interval exchange transformation defined over a quadratic field is a primitive morphic set (Theorem 7.3.12).

## 7.1 Induced transformations and admissible intervals

In this section we define the trasformation induced by an interval exchange on a sub-interval. We also introduce the definition of admissibility for an interval. It generalizes in a natural way the notion of admissibility defined in [61].

We show that all semi-intervals of the form $I_w$ and $J_w$ are admissible (Proposition 7.1.6) and that the induction on any admissible semi-interval preserve regularity (Theorem 7.1.7).

We close this section with a closure property, namely we prove that the family of regular $s$-interval exchanges is closed by derivation (Theorem 7.1.9).

### 7.1.1 Induced transformations

Let $T$ be a minimal interval exchange transformation. Let $I \subset [\ell, r[$ be a semi-interval. Since $T$ is minimal, for each $z \in [\ell, r[$ there is an integer $n > 0$ such that $T^n(z) \in I$.

The *transformation induced* by $T$ on $I$ is the transformation $S : I \to I$ defined for $z \in I$ by $S(z) = T^n(z)$ with $n = \min\{n > 0 \mid T^n(z) \in I\}$. We also say that $S$ is the *first return map* (of $T$) on $I$. The semi-interval $I$ is called the *domain* of $S$, denoted $D(S)$.

**Example 7.1.1** Let $T$ be the transformation of Example 6.1.7. Let $I = [0, 2\alpha[$. The transformation induced by $T$ on $I$ is

$$S(z) = \begin{cases} T^2(z) & \text{if } 0 \leq z < 1 - 2\alpha \\ T(z) & \text{otherwise.} \end{cases}$$

Let $T = T_{\lambda,\pi}$ be an interval exchange transformation relative to $(I_a)_{a \in A}$. For $\ell < t < r$, the semi-interval $[\ell, t[$ is *right admissible* for $T$ if there is a $k \in \mathbb{Z}$ such that $t = T^k(\gamma_a)$ for some $a \in A$ and

(i) if $k > 0$, then $t < T^h(\gamma_a)$ for all $h$ such that $0 < h < k$,

(ii) if $k \leq 0$, then $t < T^h(\gamma_a)$ for all $h$ such that $k < h \leq 0$.

We also say that $t$ itself is right admissible. Note that all semi-intervals $[\ell, \gamma_a[$ with $\ell < \gamma_a$ are right admissible. Similarly, all semi-intervals $[\ell, \delta_a[$ with $\ell < \delta_a$ are right admissible.

**Example 7.1.2** Let $T$ be the interval exchange transformation of Example 6.1.7. The semi-interval $[0, t[$ for $t = 1 - 2\alpha$ or $t = 1 - \alpha$ is right admissible since $1 - 2\alpha = \gamma_b$ and $1 - \alpha = \gamma_c$. On the contrary, for $t = 2 - 3\alpha$, it is not right admissible because $t = T^{-1}(\gamma_c)$ but $\gamma_c < t$ contradicting (ii).

The following result is Theorem 14 in [61].

**Theorem 7.1.3 (Rauzy)** *Let $T$ be a regular $s$-interval exchange transformation and let $I$ be a right admissible interval for $T$. The transformation induced by $T$ on $I$ is a regular $s$-interval exchange transformation.*

**Example 7.1.4** Consider again the transformation of Example 6.1.7. The transformation induced by $T$ on the semi-interval $I = [0, 2\alpha[$ is the 3-interval exchange transformation represented in Figure 7.1.

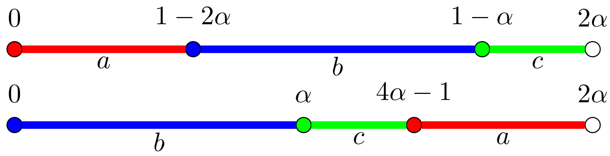

Figure 7.1: The transformation induced on $I$.

The notion of left admissible interval is symmetrical to that of right admissible. For $\ell < t < r$, the semi-interval $[t, r[$ is *left admissible* for $T$ if there is a $k \in \mathbb{Z}$ such that $t = T^k(\gamma_a)$ for some $a \in A$ and

(i) if $k > 0$, then $T^h(\gamma_a) < t$ for all $h$ such that $0 < h < k$,

(ii) if $k \leq 0$, then $T^h(\gamma_a) < t$ for all $h$ such that $k < h \leq 0$.

We also say that $t$ itself is left admissible. Note that, as for right induction, the semi-intervals $[\gamma_a, r[$ and $[\nu_a, r[$ are left admissible. The symmetrical statements of Theorem 7.1.3 also hold for left admissible intervals.

### 7.1.2 Admissible semi-intervals

Let now generalize the notion of admissibility to a two-sided version. For a semi-interval $I = [u, v[ \subset [\ell, r[$, we define the following functions on $[\ell, r[$:

$$\rho^+_{I,T}(z) = \min\{n > 0 \mid T^n(z) \in ]u, v[\}, \quad \rho^-_{I,T}(z) = \min\{n \geq 0 \mid T^{-n}(z) \in ]u, v[\}.$$

We then define three sets. First, let

$$E_{I,T}(z) = \{k \mid -\rho^-_{I,T}(z) \leq k < \rho^+_{I,T}(z)\}.$$

Next, the set of *neighbors* of $z$ with respect to $I$ and $T$ is

$$N_{I,T}(z) = \{T^k(z) \mid k \in E_{I,T}(z)\}.$$

The set of *division points* of $I$ with respect to $T$ is the finite set

$$\operatorname{Div}(I, T) = \bigcup_{i=1}^{s} N_{I,T}(\gamma_i).$$

We now formulate the following definition. For $\ell \leq u < v \leq r$, we say that the semi-interval $I = [u, v[$ is *admissible* for $T$ if $u, v \in \operatorname{Div}(I, T) \cup \{r\}$.

Note that a semi-interval $[\ell, v[$ is right admissible if and only if it is admissible and that a semi-interval $[u, r[$ is left admissible if and only if it is admissible. Note also that $[\ell, r[$ is admissible.

Note also that for a regular interval exchange transformation relative to a partition $(I_a)_{a \in A}$, each of the semi-intervals $I_a$ (or $J_a$) is admissible although only the first one is right admissible (and the last one is left admissible). Actually, we can prove that for every word $w$, the semi-intervals $I_w$ and $J_w$ are admissible. In order to do that, we need the following Lemma.

**Lemma 7.1.5** *Let $T$ be a $s$-interval exchange transformation on the semi-interval $[\ell, r[$. For any $k \geq 1$, the set $P_k = \{T^h(\gamma_i) \mid 1 \leq i \leq s,\ 1 \leq h \leq k\}$ is the set of $(s-1)k+1$ left boundaries of the semi-intervals $J_y$ for all words $y \in F(T) \cap A^k$.*

*Proof.* Let $Q_k$ be the set of left boundaries of the intervals $J_y$ for $|y| = k$. Since $\mathrm{Card}(\mathcal{L}(T) \cap A^k) = (s-1)k+1$ by Proposition 6.1.9, we have $\mathrm{Card}(Q_k) = (s-1)k+1$. Since $T$ is regular the set $R_k = \{T^h(\gamma_i) \mid 2 \leq i \leq s,\ 1 \leq h \leq k\}$ is made of $(s-1)k$ distinct points. Moreover, since

$$\gamma_1 = T(\gamma_{\pi(1)}),\ T(\gamma_1) = T^2(\gamma_{\pi(1)}), \ldots, T^{k-1}(\gamma_1) = T^k(\gamma_{\pi(1)}),$$

we have $P_k = R_k \cup \{T^k(\gamma_1)\}$. This implies $\mathrm{Card}(P_k) \leq (s-1)k+1$. On the other hand, if $y = b_0 \cdots b_{k-1}$, then $J_y = \cap_{i=0}^{k-1} T^{k-i}(I_{b_i})$. Thus the left boundary of each $J_y$ is the left boundary of some $T^h(I_a)$ for some $h$ with $1 \leq h \leq k$ and some $a \in A$. Consequently $Q_k \subset P_k$. This proves that $\mathrm{Card}(P_k) = (s-1)k+1$ and that consequently $P_k = Q_k$. ∎

A dual statement holds for the semi-intervals $I_y$.

**Proposition 7.1.6** *Let $T$ be a $s$-interval exchange transformation on the semi-interval $[\ell, r[$. For any $w \in \mathcal{L}(T)$, the semi-interval $J_w$ is admissible.*

*Proof.* Set $|w| = k$ and $J_w = [u, v[$. By Lemma 7.1.5, we have $u = T^g(\gamma_i)$ for $1 \leq i \leq s$ and $1 \leq g \leq k$. Similarly, we have $v = r$ or $v = T^d(\gamma_j)$ for $1 \leq j \leq s$ and $1 \leq d \leq k$.

For $1 < h < g$, the point $T^h(\gamma_i)$ is the left boundary of some semi-interval $J_y$ with $|y| = k$ and thus $T^h(\gamma_i) \notin J_w$. This shows that $g \in E_{J_w,T}(\gamma_i)$ and thus that $u \in \mathrm{Div}(J_w, T)$.

If $v = r$, then $v \in \mathrm{Div}(J_w, T)$. Otherwise, one shows in the same way as above that $v \in \mathrm{Div}(J_w, T)$. Thus $J_w$ is admissible. ∎

Note that the same statement holds for the semi-intervals $I_w$ instead of the semi-intervals $J_w$ (using the dual statement of Lemma 7.1.5).

It can be useful to reformulate the definition of a division point and of an admissible pair using the terminology of graphs. Let $G(T)$ be the graph with vertex set $[\ell, r[$ and edges the pairs $(z, T(z))$ for $z \in [\ell, r[$. Then, if $T$ is minimal and $I$ is a semi-interval, for any $z \in [\ell, r[$, there is a path $P_{I,T}(z)$ such that its origin $x$ and its end $y$ are in $I$, $z$ is on the path, $z \neq y$ and no vertex of the path except $x, y$ are in $I$ (actually $x = T^{-n}(z)$ with $n = \rho^-_{I,T}(z)$ and $y = T^m(z)$ with $m = \rho^+_{I,T}(z)$). Then the division points of $I$ are the vertices which are on a path $P_{I,T}(\gamma_i)$ but not at its end (see Figure 7.2).

The following is a generalization of Theorem 7.1.3. Recall that $\mathrm{Sep}(T)$ denotes the set of separation points of $T$, i.e. the points $\gamma_1 = 0, \gamma_2, \ldots, \gamma_s$ (which are the left boundaries of the semi-intervals $I_1, \ldots, I_s$).

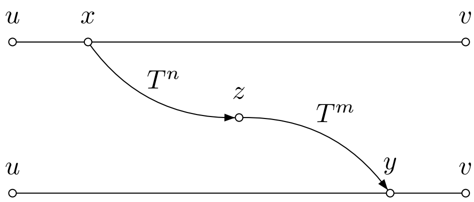


Figure 7.2: The neighbors of $z$ with respect to $I = [u, v[$.

**Theorem 7.1.7** *Let $T$ be a regular $s$-interval exchange transformation on $[\ell, r[$. For any admissible semi-interval $I = [u, v[$, the transformation $S$ induced by $T$ on $I$ is a regular $s$-interval exchange transformation with separation points* $\mathrm{Sep}(S) = \mathrm{Div}(I, T) \cap I$.

*Proof.* Since $T$ is regular, it is minimal. Thus for each $i \in \{2, \ldots, s\}$ there are points $x_i, y_i \in ]u, v[$ such that there is a path from $x_i$ to $y_i$ passing by $\gamma_i$ but not containing any point of $I$ except at its origin and its end. Since $T$ is regular, the $x_i$ are all distinct and the $y_i$ are all distinct.

Since $I$ is admissible, there exist $g, d \in \{1, \ldots, s\}$ such that $u \in N_{I,T}(\gamma_g)$ and $v \in N_{I,T}(\gamma_d)$. Moreover,since $u$ is a neighbor of $\gamma_g$ with respect to $I$, $u$ is on the path from $x_g$ to $y_g$ (it can be either before or after $\gamma_g$). Similarly, $v$ is on the path from $x_d$ to $y_d$ (see Figure 7.3 where $u$ is before $\gamma_g$ and $v$ is after $\gamma_d$).

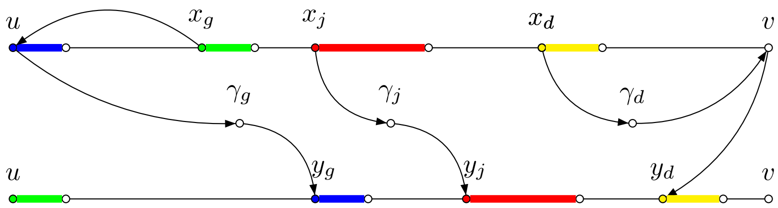


Figure 7.3: The transformation induced on $[u, v[$.

Set $x_1 = y_1 = u$. Let $(I_j)_{1 \le j \le s}$ be the partition of $I$ in semi-intervals such that $x_j$ is the left boundary of $I_j$ for $1 \le j \le s$. Let $J_j$ be the partition of $I$ such that $y_j$ is the left boundary of $J_j$ for $1 \le j \le s$. We will prove that

$$S(I_j) = \begin{cases} J_j & \text{if } j \neq 1, g \\ J_1 & \text{if } j = g \\ J_g & \text{if } j = 1 \end{cases}$$

and that the restriction of $S$ to $I_j$ is a translation.

Assume first that $j \neq 1, g$. Then $S(x_j) = y_j$. Let $k$ be such that $y_j = T^k(x_j)$ and denote $I'_j = I_j \setminus x_j$. We will prove by induction on $h$ that for $0 \le h \le k-1$, the set $T^h(I'_j)$ does not contain $u, v$ or any $x_i$. It is true for $h = 0$. Assume that it holds up to $h < k - 1$.

For any $h'$ with $0 \le h' \le h$, the set $T^{h'}(I'_j)$ does not contain any $\gamma_i$. Indeed, otherwise there would exist $h''$ with $0 \le h'' \le h'$ such that $x_i \in T^{h''}(I'_j)$, a contradiction. Thus $T$ is a translation on $T^{h'}(I_j)$. This implies that $T^h$ is a translation on $I_j$. Note also that $T^h(I'_j) \cap I = \emptyset$. Assume the contrary. We first observe that we cannot have $T^h(x_j) \in I$. Indeed, $h < k$ implies that $T^h(x_j) \notin ]u, v[$. And we cannot have $T^h(x_j) = u$ since $j \neq g$. Thus $T^h(I'_j) \cap I \neq \emptyset$ implies that $u \in T^h(I'_j)$, a contradiction.

Suppose that $u = T^{h+1}(z)$ for some $z \in I'_j$. Since $u$ is on the path from $x_g$ to $y_g$, it implies that for some $h'$ with $0 \le h' \le h$ we have $x_g = T^{h'}(z)$, a contradiction with the induction hypothesis. A similar proof (using the fact that $v$ is on the path from $x_d$ to $y_d$) shows that $T^{h+1}(I'_j)$ does not contain $v$. Finally suppose that some $x_i$ is in $T^{h+1}(I'_j)$. Since the restriction of $T^h$ to $I_j$ is a translation, $T^h(I_j)$ is a semi-interval. Since $T^{h+1}(x_j)$ is not in $I$ the fact that $T^{h+1}(I_j) \cap I$ is not empty implies that $u \in T^h(I_j)$, a contradiction.

This shows that $T^k$ is continuous at each point of $I'_j$ and that $S = T^k(x)$ for all $x \in I_j$. This implies that the restriction of $S$ to $I_j$ is a translation into $J_j$.

If $j = 1$, then $S(x_1) = S(u) = y_g$. The same argument as above proves that the restriction of $S$ to $I_1$ is a translation form $I_1$ into $J_g$. Finally if $j = g$, then $S(x_g) = x_1 = u$ and, similarly, we obtain that the restriction of $S$ to $I_g$ is a translation into $I_1$.

Since $S$ is the transformation induced by the transformation $T$ which is one to one, it is also one to one. This implies that the restriction of $S$ to each of the semi-intervals $I_j$ is a bijection onto the corresponding interval $J_j, J_1$ or $J_g$ according to the value of $j$.

This shows that $S$ is an $s$-interval exchange transformation. Since the orbits of the points $x_2, \cdots, x_s$ relative to $S$ are included in the orbits of $\gamma_2, \ldots, \gamma_s$, they are infinite and disjoint. Thus $S$ is regular.

Let us finally show that $\mathrm{Sep}(S) = \mathrm{Div}(I, T) \cap I$. We have $\mathrm{Sep}(S) = \{x_1, x_2, \ldots, x_s\}$ and $x_i \in N_{I,T}(\gamma_i)$. Thus $\mathrm{Sep}(S) \subset \mathrm{Div}(I, T) \cap I$. Conversely, let $x \in \mathrm{Div}(I, T) \cap I$. Then $x \in N_{I,T}(\gamma_i) \cap I$ for some $1 \le i \le s$. If $i \neq 1, g$, then $x = x_i$. If $i = 1$, then either $x = u$ (if $u = \ell$) or $x = x_{\pi(1)}$ since $\gamma_1 = T(\gamma_{\pi(1)})$. Finally, if $i = g$ then $x = u$ or $x = x_g$. Thus $x \in \mathrm{Sep}(S)$ in all cases. ■

Note that for any $s$-interval exchange transformation on $[\ell, r[$ and any semi-interval $I$ of $[\ell, r[$, the transformation $S$ induced by $T$ on $I$ is an interval exchange transformation on at most $s + 2$-intervals (see [25, Chapter 5]). Actually, it follows from the proof of [25, Lemma 2] that, if $T$ is regular and $S$ is an $s$-interval exchange transformation with separation points $\mathrm{Sep}(S) = \mathrm{Div}(I, T) \cap I$, then $I$ is admissible. Thus the converse of Theorem 7.1.7 is also true.

### 7.1.3 Derived sets

In the following we will prove a closure property of the family of regular interval exchange sets. The same property holds for Sturmian sets (see [45]) and for recurrent tree sets of characteristic 1 (Theorem 3.2.9).

**Lemma 7.1.8** *Let $T$ be a regular interval exchange transformation and let $F = \mathcal{L}(T)$. For $w \in F$, let $S$ be the transformation induced by $T$ on $J_w$. One has $x \in \mathcal{R}_F(w)$ if and only if*

$$\Sigma_T(z) = x\Sigma_T(S(z))$$

*for some $z \in J_w$.*

*Proof.* Assume first that $x \in \mathcal{R}_F(w)$. Then for any $z \in J_w \cap I_x$, we have $S(z) = T^{|x|}(z)$ and

$$\Sigma_T(z) = x\Sigma_T(T^{|x|}(z)) = x\Sigma_T(S(z)).$$

Conversely, assume that $\Sigma_T(z) = x\Sigma_T(S(z))$ for some $z \in J_w$. Then $T^{|x|}(z) \in J_w$ and thus $wx \in A^*w$ which implies that $x \in \Gamma_F(w)$. Moreover $x$ does not have a proper prefix in $\Gamma_F(w)$ and thus $x \in \mathcal{R}_F(w)$. ∎

Since a regular interval exchange set is recurrent, the previous lemma says that the natural coding of a point in $J_w$ is a concatenation of first return words to $w$. Moreover, note also that $T^n(z) \in J_w$ if and only if the prefix of length $n$ of $\Sigma_T(z)$ is a return word to $w$.

We have thus the following result, who is a counterpart for interval exchange sets of Theorem 3.2.9.

**Theorem 7.1.9** *Any derived set of a regular $s$-interval exchange set is a regular $s$-interval exchange set.*

*Proof.* Let $T$ be a regular $s$-interval exchange transformation and let $F = \mathcal{L}(T)$.

Let $w \in F$. Since the semi-interval $J_w$ is admissible according to Proposition 7.1.6, the transformation $S$ induced by $T$ on $J_w$ is, by Theorem 7.1.7, an $s$-interval exchange transformation. The corresponding partition of $J_w$ is the family $(J_{wx})_{w \in \mathcal{R}_\mathcal{F}(w)}$.

Using Lemma 7.1.8 and the observation following, it is clear that $\Sigma_T(z) = f(\Sigma_S(z))$, where $z$ is a point of $J_w$ and $f : A^* \to \mathcal{R}_F(w)^*$ is a coding morphism for $\mathcal{R}_F(w)$.

Set $x = \Sigma_T(T^{-|w|}(z))$ and $y = \Sigma_T(z)$. Then $x = wy$ and thus $\Sigma_S(z) = \mathcal{D}_f(x)$. By Proposition 1.4.8, this shows that the derived set of $F$ with respect to $f$ is $\mathcal{L}(S)$. ∎

Theorem 3.2.9 implies, in particular that $\mathrm{Card}(\mathcal{R}_F(w)) = \mathrm{Card}(A)$, accordingly with Corollary 2.2.10 (see also [67] and [5]).

## 7.2 Rauzy induction

In this section we describe the transformation called Rauzy induction defined in [61] which operates on regular interval transformations and recall the results concerning this transformation (Theorems 7.1.3 and 7.2.1). We also introduce a branching version of this transformation and generalize Rauzy's results to the two-sided case (Theorems 7.1.7 and 7.2.3). In particular we characterize in Theorem 7.2.3 the admissible semi-intervals for an interval exchange transformation.

### 7.2.1 One-side Rauzy induction

Let $T = T_{\lambda,\pi}$ be a regular $s$-interval exchange transformation on $[\ell, r[$. Set $Z(T) = [\ell, \max\{\gamma_s, \delta_{\pi(s)}\}[$.

Note that $Z(T)$ is the largest semi-interval which is right-admissible for $T$. We denote by $\psi(T)$ the transformation induced by $T$ on $Z(T)$.

The following result is Theorem 23 in [61].

**Theorem 7.2.1 (Rauzy)** *Let $T$ be a regular interval exchange transformation. A semi-interval $I$ is right admissible for $T$ if and only if there is an integer $n \geq 0$ such that $I = Z(\psi^n(T))$. In this case, the transformation induced by $T$ on $I$ is $\psi^{n+1}(T)$.*

The map $T \mapsto \psi(T)$ is called the *right Rauzy induction*. There are actually two cases according to $\gamma_s < \delta_{\pi(s)}$ (Case 0) or $\gamma_s > \delta_{\pi(s)}$ (Case 1). We cannot have $\gamma_s = \delta_{\pi(s)}$ since $T$ is regular.

In Case 0, we have $Z(T) = [\ell, \delta_{\pi(s)}[$ and for any $z \in Z(T)$,

$$\psi(T)(z) = \begin{cases} T^2(z) & \text{if } z \in I_{a_{\pi(s)}} \\ T(z) & \text{otherwise.} \end{cases}$$

The transformation $S = \psi(T)$ is the interval exchange transformation relative to $(K_a)_{a\in A}$ with $K_a = I_a \cap Z(T)$ for all $a \in A$. Note that $K_a = I_a$ for $a \neq a_s$. The translation values $\beta_a$ are defined as follows, denoting $\alpha_i, \beta_i$ instead of $\alpha_{a_i}, \beta_{a_i}$,

$$\beta_i = \begin{cases} \alpha_{\pi(s)} + \alpha_s & \text{if } i = \pi(s) \\ \alpha_i & \text{otherwise.} \end{cases}$$

In summary, in Case 0, the semi-interval $J_{a_{\pi(s)}}$ is suppressed, the semi-interval $J_{a_s}$ is split into $S(K_{a_s})$ and $S(K_{a_{\pi(s)}})$. The left boundaries of the semi-intervals $K_a$ are the left boundaries of the semi-intervals $I_a$. The transformation is represented in Figure 7.4, in which the left boundary of the semi-interval $S(K_{a_{\pi(s)}})$ is denoted $\delta'_{\pi(s)}$.

In Case 1, we have $Z(T) = [\ell, \gamma_s[$ and for any $z \in Z(T)$,

$$\psi(T)(z) = \begin{cases} T^2(z) & \text{if } z \in T^{-1}(I_{a_s}) \\ T(z) & \text{otherwise.} \end{cases}$$

The transformation $S = \psi(T)$ is the interval exchange transformation relative to $(K_a)_{a\in A}$ with

$$K_a = \begin{cases} T^{-1}(I_a) & \text{if } a = a_s \\ T^{-1}(T(I_a) \cap Z(T)) & \text{otherwise.} \end{cases}$$

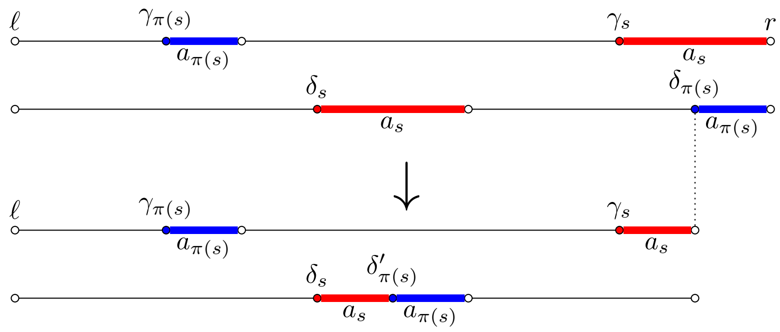


Figure 7.4: Case 0 in Rauzy induction.

Note that $K_a = I_a$ for $a \neq a_s$ and $a \neq a_{\pi(s)}$. Moreover $K_a = S^{-1}(T(I_a) \cap Z(T))$ in all cases. The translation values $\beta_i$ are defined by

$$\beta_i = \begin{cases} \alpha_{\pi(s)} + \alpha_s & \text{if } i = s \\ \alpha_i & \text{otherwise.} \end{cases}$$

In summary, in Case 1, the semi-interval $I_{a_s}$ is suppressed, the semi-interval $I_{a_{\pi(s)}}$ is split into $K_{a_{\pi(s)}}$ and $K_{a_s}$. The left boundaries of the semi-intervals $S(K_a)$ are the left boundaries of the semi-intervals $J_a$. The transformation is represented in Figure 7.5, where the left boundary of the semi-interval $K_{a_s}$ is denoted $\gamma'_s$.

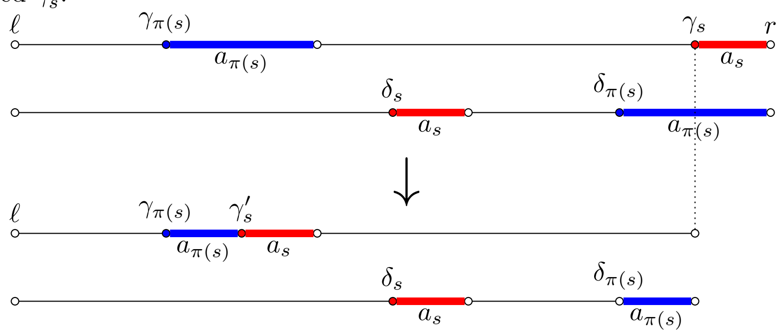


Figure 7.5: Case 1 in Rauzy induction.

**Example 7.2.2** Consider again the transformation $T$ of Example 6.1.7. Since $Z(T) = [0, 2\alpha[$, the transformation $\psi(T)$ is represented in Figure 7.1. The transformation $\psi^2(T)$ is represented in Figure 7.6.

The symmetrical notion of *left Rauzy induction* is defined similarly.

Let $T = T_{\lambda,\pi}$ be a regular $s$-interval exchange transformation on $[\ell, r[$. Set $Y(T) = [\min\{\mu_1, \nu_{\pi(1)}\}, r[$. We denote by $\varphi(T)$ the transformation induced by $T$ on $Y(T)$. The map $T \mapsto \varphi(T)$ is called the *left Rauzy induction.*

Note that one has also $Y(T) = [\min\{\gamma_2, \delta_{\pi(2)}\}, r[$.

The symmetrical statements of Theorem 7.2.1 also hold for left admissible intervals.

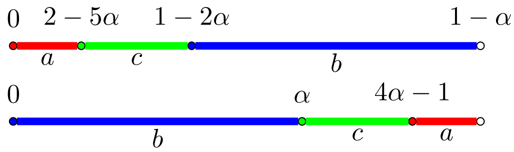


Figure 7.6: The transformation $\psi^2(T)$.

### 7.2.2 Branching induction

The following is a generalization of Theorem 7.2.1.

**Theorem 7.2.3** *Let $T$ be a regular $s$-interval exchange transformation on $[\ell, r[$. A semi-interval $I$ is admissible for $T$ if and only if there is a sequence $\chi \in \{\varphi, \psi\}^*$ such that $I$ is the domain of $\chi(T)$. In this case, the transformation induced by $T$ on $I$ is $\chi(T)$.*

We first prove the following lemmas, in which we assume that $T$ is a regular $s$-interval exchange transformation on $[\ell, r[$. Recall that $Y(T), Z(T)$ are the domains of $\varphi(T), \psi(T)$ respectively.

**Lemma 7.2.4** *If a semi-interval $I$ strictly included in $[\ell, r[$ is admissible for $T$, then either $I \subset Y(T)$ or $I \subset Z(T)$.*

*Proof.* Set $I = [u, v[$. Since $I$ is strictly included in $[\ell, r[$, we have either $\ell < u$ or $v < r$. Set $Y(T) = [y, r[$ and $Z(T) = [\ell, z[$.

Assume that $v < r$. If $y \leq u$, then $I \subset Y(T)$. Otherwise, let us show that $v \leq z$. Assume the contrary. Since $I$ is admissible, we have $v = T^k(\gamma_i)$ with $k \in E_{I,T}(\gamma_i)$ for some $i$ with $1 \leq i \leq s$. But $k > 0$ is impossible since $u < T(\gamma_i) < v$ implies $T(\gamma_i) \in ]u, v[$, in contradiction with the fact that $k < \rho_I^+(\gamma_i)$. Similarly, $k \leq 0$ is impossible since $u < \gamma_i < v$ implies $\gamma_i \in ]u, v[$. Thus $I \subset Z(T)$.

The proof in the case $\ell < u$ is symmetric. ∎

The next lemma is the two-sided version of Lemma 22 in [61].

**Lemma 7.2.5** *Let $T$ be a regular $s$-interval exchange transformation on $[\ell, r[$. Let $J$ be an admissible semi-interval for $T$ and let $S$ be the transformation induced by $T$ on $J$. A semi-interval $I \subset J$ is admissible for $T$ if and only if it is admissible for $S$. Moreover* $\mathrm{Div}(J, T) \subset \mathrm{Div}(I, T)$.

*Proof.* Set $J = [t, w[$ and $I = [u, v[$. Since $J$ is admissible for $T$, the transformation $S$ is a regular $s$-interval exchange transformation by Theorem 7.1.7.

Suppose first that $I$ is admissible for $T$. Then $u = T^g(\gamma_i)$ with $g \in E_{I,T}(\gamma_i)$ for some $1 \leq i \leq s$, and $v = T^d(\gamma_j)$ with $d \in E_{I,T}(\gamma_j)$ for some $1 \leq j \leq s$ or $v = r$.

Since $S$ is the transformation induced by $T$ on $J$ there is a separation point $x$ of $S$ of the form $x = T^m(\gamma_i)$ with $m = -\rho^-_{J,T}(\gamma_i)$ and thus $m \in E_{J,T}(\gamma_i)$. Thus $u = T^{g-m}(x)$.

Assume first that $g - m > 0$. Since $u, x \in J$, there is an integer $n$ with $0 < n \le g - m$ such that $u = S^n(x)$.

Let us show that $n \in E_{I,S}(x)$. Assume by contradiction that $\rho^+_{I,S}(x) \le n$. Then there is some $k$ with $0 < k \le n$ such that $S^k(x) \in ]u, v[$. But we cannot have $k = n$ since $u \notin ]u, v[$. Thus $k < n$.

Next, there is $h$ with $0 < h < g - m$ such that $T^h(x) = S^k(x)$. Indeed, setting $y = S^k(x)$, we have $u = T^{g-m-h}(y) = S^{n-k}(y)$ and thus $h < g - m$. If $0 < h \le -m$, then $T^h(x) = T^{m+h}(\gamma_i) \in I \subset J$ contradicting the hypothesis that $m \in E_{J,T}(\gamma_i)$. If $-m < h < g - m$, then $T^h(x) = T^{m+h}(\gamma_i) \in I$, contradicting the fact that $g \in E_{I,T}(\gamma_i)$. This shows that $n \in E_{I,S}(x)$ and thus that $u \in \mathrm{Div}(I, S)$.

Assume next that $g-m \le 0$. There is an integer $n$ with $g-m \le n \le 0$ such that $u = S^n(x)$. Let us show that $n \in E_{I,S}(x)$. Assume by contradiction that $n < -\rho^-_{I,S}(x)$. Then there is some $k$ with $n < k < 0$ such that $S^k(x) = T^h(x)$. Then $T^h(x) = T^{h+m}(\gamma_i) \in I$ with $g < h + m < m$, in contradiction with the hypothesis that $m \in E_{I,T}(\gamma_i)$.

We have proved that $u \in \mathrm{Div}(I, S)$. If $v = r$, the proof that $I$ is admissible for $S$ is complete. Otherwise, the proof that $v \in \mathrm{Div}(I, S)$ is similar to the proof for $u$.

Conversely, if $I$ is admissible for $S$, there is some $x \in \mathrm{Sep}(S)$ and $g \in E_{I,S}(x)$ such that $u = S^g(x)$. But $x = T^m(\gamma_i)$ and since $u, x \in J$ there is some $n$ such that $u = T^n(\gamma_i)$.

Assume for instance that $n > 0$ and suppose that there exists $k$ with $0 < k < n$ such that $T^k(\gamma_i) \in ]u, v[$. Then, since $I \subset J$, $T^k(\gamma_i)$ is of the form $S^h(x)$ with $0 < h < g$ which contradicts the fact that $g \in E_{I,S}(x)$. Thus $n \in E_{I,T}(\gamma_i)$ and $u \in \mathrm{Div}(I, T)$.

The proof is similar in the case $n \le 0$.

If $v = r$, we have proved that $I$ is admissible for $T$. Otherwise, the proof that $v \in \mathrm{Div}(I, T)$ is similar.

Finally, assume that $I$ is admissible for $T$ (and thus for $S$). For any $\gamma_i \in \mathrm{Sep}(T)$, one has

$$\rho^-_{I,T}(\gamma_i) \ge \rho^-_{J,T}(\gamma_i) \quad \text{and} \quad \rho^+_{I,T}(\gamma_i) \ge \rho^+_{J,T}(\gamma_i)$$

showing that $\mathrm{Div}(J, T) \subset \mathrm{Div}(I, T)$. ■

The last lemma is the key argument to prove Theorem 7.2.3. It is a branching version of the argument used by Rauzy in [61].

**Lemma 7.2.6** *For any admissible interval $I \subset [\ell, r[$, the set $\mathcal{F}$ of sequences $\chi \in \{\varphi, \psi\}^*$ such that $I \subset D(\chi(T))$ is finite.*

*Proof.* The set $\mathcal{F}$ is suffix-closed. Indeed it contains the empty word because $[\ell, r[$ is admissible. Moreover, for any $\xi, \chi \in \{\varphi, \psi\}^*$, one has $D(\xi\chi(T)) \subset D(\chi(T))$ and thus $\xi\chi \in \mathcal{F}$ implies $\chi \in \mathcal{F}$.

The set $\mathcal{F}$ is finite. Indeed, by Lemma 7.2.5, applied to $J = D(\chi(T))$, for any $\chi \in \mathcal{F}$, one has $\mathrm{Div}(D(\chi(T)), T) \subset \mathrm{Div}(I, T)$. In particular,the boundaries of $D(\chi(T))$ belong to $\mathrm{Div}(I, T)$. Since $\mathrm{Div}(I, T)$ is a finite set, this implies that there is a finite number of possible semi-intervals $D(\chi(T))$. Thus there is is no infinite word with all its suffixes in $\mathcal{F}$. Since the sequences $\chi$ are binary, this implies that $\mathcal{F}$ is finite. ∎

*Proof of Theorem 7.2.3.* We first prove by induction on the length of $\chi$ that the domain $I$ of $\chi(T)$ is admissible and that the transformation induced by $T$ on $I$ is $\chi(T)$. It is true for $|\chi| = 0$ since $[\ell, r[$ is admissible and $\chi(T) = T$. Next, assume that $J = D(\chi(T))$ is admissible and that the transformation induced by $T$ on $J$ is $\chi(T)$. Then $D(\varphi\chi(T))$ is admissible for $\chi(T)$ since $D(\varphi\chi(T)) = Y(\chi(T))$. Thus $I = D(\varphi\chi(T))$ is admissible for $T$ by Lemma 7.2.5 and the transformation induced by $T$ on $I$ is $\varphi\chi(T)$. The same proof holds for $\psi\chi$.

Conversely, assume that $I$ is admissible. By Lemma 7.2.6, the set $\mathcal{F}$ of sequences $\chi \in \{\varphi, \psi\}^*$ such that $I \subset D(\chi(T))$ is finite.

Thus there is some $\chi \in \mathcal{F}$ such that $\varphi\chi, \psi\chi \notin \mathcal{F}$. If $I$ is strictly included in $D(\chi(T))$, then by Lemma 7.2.4 applied to $\chi(T)$, we have $I \subset Y(\chi(T)) = D(\varphi\chi(T))$ or $I \subset Z(\chi(T)) = D(\psi\chi(T))$, a contradiction. Thus $I = D(\chi(T))$. ∎

We close this subsection with a result concerning the dynamics of the branching induction.

**Theorem 7.2.7** *For any sequence $(T_n)_{n\geq 0}$ of regular interval exchange transformations such that $T_{n+1} = \varphi(T_n)$ or $T_{n+1} = \psi(T_n)$ for all $n \geq 0$, the length of the domain of $T_n$ tends to $0$ when $n \to \infty$.*

*Proof.* Assume the contrary and let $I$ be an open interval included in the domain of $T_n$ for all $n \geq 0$. The set $\mathrm{Div}(I, T) \cap I$ is formed of $s$ points. For any pair $u, v$ of consecutive elements of this set, the semi-interval $[u, v[$ is admissible. By Lemma 7.2.6, there is an integer $n$ such that the domain of $T_n$ does not contain $[u, v[$, a contradiction. ∎

### 7.2.3 Equivalence relation

Let $[\ell_1, r_1[$, $[\ell_2, r_2[$ be two semi-intervals of the real line. Let $T_1 = T_{\lambda,\pi_1}$ be an $s$-interval exchange transformation relative to a partition of $[\ell_1, r_1[$ and $T_2 = T_{\mu,\pi_2}$ another $s$-interval exchange transformations relative to $[\ell_2, r_2[$. We say that $T_1$ and $T_2$ are *equivalent* if $\pi_1 = \pi_2$ and $\lambda = c\mu$ for some $c > 0$. Thus, two interval exchange transformations are equivalent if we can obtain the second from the first by a rescaling following by a translation. We denote by $[T_{\lambda,\pi}]$ the equivalence class of $T_{\lambda,\pi}$.

**Example 7.2.8** Let $S = T_{\mu,\pi}$ be the 3-interval exchange transformation on a partition of the semi-interval $[2\alpha, 1[$, with $\alpha = (3-\sqrt{5})/2$, represented in Figure 7.7. $S$ is equivalent to the transformation $T = T_{\lambda,\pi}$ of Example 6.1.7, with length vector $\lambda = (1-2\alpha, \alpha, \alpha)$ and permutation the cycle $\pi = (132)$. Indeed the length vector $\mu = (8\alpha-3, 2-5\alpha, 2-5\alpha)$ satisfies $\mu = \frac{2-5\alpha}{\alpha}\lambda$.

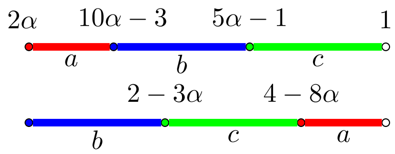


Figure 7.7: The transformation $S$.

Note that if $T$ is a minimal (resp. regular) interval exchange transformation and $[S] = [T]$, then $S$ is also minimal (resp. regular).

For an interval exchange transformation $T$ we consider the directed labeled graph $\mathcal{IG}(T)$, called the *induction graph* of $T$, defined as follows. The vertices are the equivalence classes of transformations obtained starting from $T$ and applying all possible $\chi \in \{\psi, \varphi\}^*$. There is an edge labeled $\psi$ (resp. $\varphi$) from a vertex $[S]$ to a vertex $[U]$ if and only if $U = \psi(S)$ (resp $\varphi(S)$) for two transformations $S \in [S]$ and $U \in [U]$.

**Example 7.2.9** Let $\alpha = \frac{3-\sqrt{5}}{2}$ and $R$ be a rotation of angle $\alpha$. By Example 6.1.1, $R$ is a 2-interval exchange transformation on $[0, 1[$ relative to the partition $[0, 1 - \alpha[$, $[1 - \alpha, 1[$. The induction graph $\mathcal{IG}(R)$ of the transformation is represented in the left of Figure 7.9.

Note that for a 2-interval exchange transformation $T$, one has $[\psi(T)] = [\varphi(T)]$, whereas in general the two transformations are not equivalent.

The induction graph of an interval exchange transformation can be infinite. A sufficient condition for the induction graph to be finite is given in Section 7.3.

Let now introduce a variant of this equivalence relation (and of the related graph). We consider the case of two transformations "equivalent" up to reflection (and up to the separation points). This choice allows us to obtain the same natural coding for an interval exchange transformation relative to a point, and for the mirror transformation relative to the specular point (with respect to the midpoint of the of the interval).

For an $s$-interval exchange transformation $T = T_{\lambda,\pi}$, with length vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_s)$, we define the *mirror transformation* $\widetilde{T} = T_{\widetilde{\lambda},\tau\circ\pi}$ of $T$, where $\widetilde{\lambda} = (\lambda_s, \lambda_{s-1}, \ldots, \lambda_1)$ and $\tau : i \mapsto (s - 1 + 1)$ is the permutation that reverses the names of the semi-intervals.

Given two interval exchange transformations $T_1$ and $T_2$ on the same alphabet relative to two partitions of two semi-intervals $[\ell_1, r_1[$ and $[\ell_2, r_2[$ respectively, we say that $T_1$ and $T_2$ are *similar* either if $[T_1] = [T_2]$ or $[T_1] = \widetilde{[T_2]}$. Clearly, similarity is also an equivalent relation. We denote by $\langle T \rangle$ the class of transformations similar to $T$.

**Example 7.2.10** Let $T$ be the interval exchange transformation of Example 6.1.7. The transformation $U = \varphi^6(T)$ is represented in Figure 7.8 (see also Example 7.2.18). It is easy to verify that $U$ is similar to the transformation $S$ of Example 7.2.8. Indeed, we can obtain the second transformation (up to the separation points and the end points) by taking the mirror image of the domain.

Note that the order of the labels, i.e. the order of the letters of the alphabet, may be different from the order of the original transformation.

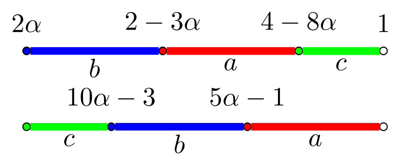


Figure 7.8: The transformation $U$.

As of the equivalence relation, also similarity preserves minimality and regularity.

Let $T$ be an interval exchange transformation. We denote by

$$\mathcal{S}(T) = \bigcup_{n \in \mathbb{Z}} T^n\big(\operatorname{Sep}(T)\big)$$

the union of the orbits of the separation points. Let $S$ be an interval exchange transformation similar to $T$. Thus, there exists a bijection $f : D(T) \setminus \mathcal{S}(T) \to D(S) \setminus \mathcal{S}(S)$. This bijection is given by an affine transformation, namely a rescaling following by a translation if $T$ and $S$ are equivalent and a rescaling following by a translation and a reflection otherwise. By the previous remark, if $T$ is a minimal exchange interval transformation and $S$ is similar to $T$, then the two interval exchange sets $\mathcal{L}(T)$ and $\mathcal{L}(S)$ are equal up to permutation, that is there exists a permutation $\pi$ such that one for every $w = a_0 a_1 \cdots a_{n-1} \in \mathcal{L}(T)$ there exists a unique word $v = b_0 b_1 \cdots b_{n-1} \in \mathcal{L}(S)$ such that $b_i = \pi(a_i)$ for all $i = 1, 2, \dots n-1$.

In a similar way as before, we can use the similarity in order to construct a graph. For an interval exchange transformation $T$ we define $\widetilde{\mathcal{IG}}(T)$ the *modified induction graph* of $T$ as the directed (unlabeled) graph with vertices the similar classes of transformations obtained starting from $T$ and applying all possible $\chi \in \{\psi, \varphi\}^*$ and an edge from $\langle S \rangle$ to $\langle U \rangle$ if $U = \psi(S)$ or $U = \varphi(S)$ for two transformations $S \in \langle S \rangle$ and $U \in \langle U \rangle$.

Note that this variant appears naturally when considering the Rauzy induction of a 2-interval exchange transformation as a continued fraction expansion.

There exists a natural bijection between the closed interval $[0,1]$ of the real line and the set of 2-interval exchange transformation given by the map $x \mapsto T_{\lambda,\pi}$ where $\pi = (12)$ and $\lambda = (\lambda_1, \lambda_2)$ is the length vector such that $x = \frac{\lambda_1}{\lambda_2}$.

In this view, the Rauzy induction corresponds to the Euclidean algorithm (see [55] for more details), i.e. the map $\mathcal{E} : \mathbb{R}_+^2 \to \mathbb{R}_+^2$ given by

$$\mathcal{E}(\lambda_1, \lambda_2) = \begin{cases} (\lambda_1 - \lambda_2, \lambda_2) & \text{if } \lambda_1 \geq \lambda_2 \\ (\lambda_1, \lambda_2 - \lambda_2) & \text{otherwise.} \end{cases}$$

Applying iteratively the Rauzy induction starting from $T$ corresponds then to the continued fraction expansion of $x$.

**Example 7.2.11** Let $\alpha$ and $R$ be as in Example 7.2.9. The extension graph $\mathcal{IG}(R)$ and the modified induction graph $\widetilde{\mathcal{IG}}(R)$ of the transformation Are represented rispectively on the left and on the right of Figure 7.9. Note that the ratio of the two lengths of the semi-intervals exchanged by $T$ is

$$\frac{1-\alpha}{\alpha} = \frac{1+\sqrt{5}}{2} = \phi = 1 + \frac{1}{1 + \frac{1}{1+\cdots}}.$$

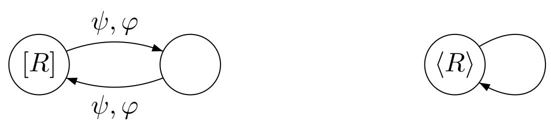


Figure 7.9: Induction graph and modified induction graph of the rotation $R$ of angle $\alpha$= $(3-\sqrt{5})/2$.

### 7.2.4 Induction and automorphisms

Let $T = T_{\lambda,\pi}$ be a regular interval exchange on $[\ell, r[$ relative to $(I_a)_{a\in A}$. Set $A = \{a_1, \ldots, a_s\}$. Recall now from Section 6.1.2 that for any $z \in [\ell, r[$, the natural coding of $T$ relative to $z$ is the infinite word $\Sigma_T(z) = b_0 b_1 \cdots$ on the alphabet $A$ with $b_n \in A$ defined for $n \geq 0$ by $b_n = a$ if $T^n(z) \in I_a$.

Denote by $\eta_1$, $\eta_2$ the morphisms from $A^*$ into itself defined by

$$\eta_1(a) = \begin{cases} a_{\pi(s)} a_s & \text{if } a = a_{\pi(s)} \\ a & \text{otherwise} \end{cases}, \qquad \eta_2(a) = \begin{cases} a_{\pi(s)} a_s & \text{if } a = a_s \\ a & \text{otherwise} \end{cases}.$$

The morphisms $\eta_1, \eta_2$ extend to automorphisms of the free group on $A$.

The following result already appears in [44]. We give a proof for the sake of completeness.

**Proposition 7.2.12** *Let $T$ be a regular interval exchange transformation on the alphabet $A$ and let $S = \psi(T)$, $I = Z(T)$. There exists an automorphism $\eta$ of the free group on $A$ such that $\Sigma_T(z) = \eta(\Sigma_S(z))$ for any $z \in I$.*

*Proof.* Assume first that $\gamma_s < \delta_{\pi(s)}$ (Case 0). We have $Z(T) = [\ell, \delta_{\pi(s)}[$ and for any $x \in Z(T)$,

$$S(z) = \begin{cases} T^2(z) & \text{if } z \in K_{a_{\pi(s)}} = I_{a_{\pi(s)}} \\ T(z) & \text{otherwise.} \end{cases}$$

We will prove by induction on the length of $w$ that for any $z \in I$, $\Sigma_S(z) \in wA^*$ if and only if $\Sigma_T(z) \in \eta_1(w)A^*$. The property is true if $w$ is the empty word. Assume next that $w = av$ with $a \in A$ and thus that $z \in I_a$. If $a \neq a_{\pi(s)}$, then $\eta_1(a) = a$, $S(z) = T(z)$ and

$$\Sigma_S(z) \in avA^* \Leftrightarrow \Sigma_S(S(z)) \in vA^* \Leftrightarrow \Sigma_T(T(z)) \in \eta_1(v)A^* \Leftrightarrow \Sigma_T(z) \in \eta_1(w)A^*.$$

Otherwise, $\eta_1(a) = a_{\pi(s)}a_s$, $S(z) = T^2(z)$. Moreover, $\Sigma_T(z) = a_{\pi(s)}a_s\Sigma_T(T^2(z))$ and thus

$$\Sigma_S(z) \in avA^* \Leftrightarrow \Sigma_S(S(z)) \in vA^* \Leftrightarrow \Sigma_T(T^2(z)) \in \eta_1(v)A^* \Leftrightarrow \Sigma_T(z) \in \eta_1(w)A^*.$$

If $\delta_{\pi(s)} < \gamma_s$ (Case 1), we have $Z(T) = [\ell, \gamma_s[$ and for any $z \in Z(T)$,

$$S(z) = \begin{cases} T^2(z) & \text{if } z \in K_{a_s} = T^{-1}(I_{a_s}) \\ T(z) & \text{otherwise.} \end{cases}$$

As in Case 0, we will prove by induction on the length of $w$ that for any $z \in I$, $\Sigma_S(z) \in wA^*$ if and only if $\Sigma_T(z) \in \eta_2(w)A^*$.

The property is true if $w$ is empty. Assume next that $w = av$ with $a \in A$. If $a \neq a_s$, then $\eta_2(a) = a$, $S(z) = T(z)$ and $z \in K_a \subset I_a$. Thus

$$\Sigma_S(z) \in avA^* \Leftrightarrow \Sigma_S(S(z)) \in vA^* \Leftrightarrow \Sigma_T(T(z)) \in \eta_2(v)A^* \Leftrightarrow \Sigma_T(z) \in \eta_2(w)A^*.$$

Next, if $a = a_s$, then $\eta_2(a) = a_{\pi(s)}a_s$, $S(z) = T^2(z)$ and $z \in K_{a_s} = T^{-1}(I_{a_s}) \subset I_{a_{\pi(s)}}$. Thus

$$\Sigma_S(z) \in avA^* \Leftrightarrow \Sigma_S(S(z)) \in vA^* \Leftrightarrow \Sigma_T(T^2(z)) \in \eta_2(v)A^* \Leftrightarrow \Sigma_T(z) \in \eta_2(w)A^*.$$

where the last equivalence results from the fact that $\Sigma_T(z) \in a_{\pi(s)}a_sA^*$. This proves that $\Sigma_T(z) = \eta_2(\Sigma_S(z))$. ■

**Example 7.2.13** Let $T$ be the transformation of Example 6.1.7. The automorphism $\eta_1$ is defined by

$$\eta_1(a) = ac, \quad \eta_1(b) = b, \quad \eta_1(c) = c.$$

The right Rauzy induction gives the transformation $S = \psi(T)$ computed in Example 7.1.4. One has $\Sigma_S(\alpha) = bacba\cdots$ and $\Sigma_T(\alpha) = baccbac\cdots = \eta_1(\Sigma_S(\alpha))$.

We state the symmetrical version of Proposition 7.2.12 for left Rauzy induction. The proof is analogous.

**Proposition 7.2.14** *Let $T$ be a regular interval exchange transformation on the alphabet $A$ and let $S = \varphi(T)$, $I = Y(T)$. There exists an automorphism $\eta$ of the free group on $A$ such that $\Sigma_T(z) = \eta(\Sigma_S(z))$ for any $z \in I$.*

Combining Propositions 7.2.12 and 7.2.14, we obtain the following statement.

**Theorem 7.2.15** *Let $T$ be a regular interval exchange transformation. For $\xi \in \{\varphi, \psi\}^*$, let $S = \xi(T)$ and let $I$ be the domain of $S$. There exists an automorphism $\eta$ of the free group on $A$ such that $\Sigma_T(z) = \eta(\Sigma_S(z))$ for all $z \in I$.*

*Proof.* The proof follows easily by induction on the length of $\xi$ using Propositions 7.2.12 and 7.2.14. ■

Note that if the transformations $T$ and $S = \xi(T)$, with $\xi \in \{\psi, \varphi\}^*$ , are equivalent, then there exists a point $z_0 \in D(S) \subseteq D(T)$ such that $z_0$ is a fixed point of the isometry that transforms $D(S)$ into $D(T)$ (if $\xi$ is different from the identity map, this point is unique). In that case one has $\Sigma_S(z_0) = \Sigma_T(z_0) = \eta\left(\Sigma_S(z_0)\right)$ for an appropriate automorphism $\eta$, i.e. $\Sigma_T(z_0)$ is a fixed point of an appropriate automorphism.

By Theorem 3.2.5, every set of return words in a regular interval exchange set is a basis of the free group. We give now a proof of this result using the branching Rauzy induction.

**Corollary 7.2.16** *Let $T$ be a regular interval exchange transformation and set $F = \mathcal{L}(T)$. For $w \in F$, the set $\mathcal{R}_F(w)$ is a basis of the free group on $A$.*

*Proof.* By Proposition 7.1.6, the semi-interval $J_w$ is admissible. By Theorem 7.2.3 there is a sequence $\xi \in \{\varphi, \psi\}^*$ such that $D(\xi(T)) = J_w$. Moreover, the transformation $S = \xi(T)$ is the transformation induced by $T$ on $J_w$. By Theorem 7.2.15 there is an automorphism $\eta$ of the free group on $A$ such that $\Sigma_T(z) = \eta(\Sigma_S(z))$ for any $z \in J_w$.

By Lemma 7.1.8, we have $x \in \mathcal{R}_F(w)$ if and only if $\Sigma_T(z) = x\Sigma_T(S(z)))$ for some $z \in J_w$. This implies that $\mathcal{R}_F(w) = \eta(A)$. Indeed, for any $z \in J_w$, let $a$ is the first letter of $\Sigma_S(z)$. Then

$$\Sigma_T(z) = \eta(\Sigma_S(z)) = \eta(a\Sigma_S(S(z))) = \eta(a)\eta(\Sigma_S(Sz)) = \eta(a)\Sigma_T(S(z)).$$

Thus $x \in \mathcal{R}_F(w)$ if and only if there is $a \in A$ such that $x = \eta(a)$. This proves that the set $\mathcal{R}_F(w)$ is a basis of the free group on $A$. ■

We illustrate the this result with the following examples.

**Example 7.2.17** We consider again the transformation $T$ of Example 6.1.7 and set $F = \mathcal{L}(T)$. We have $R_F(c) = \{bac, bbac, c\}$ (see Example 1.4.1). We represent in Figure 7.10 the sequence $\xi$ of Rauzy inductions such that $J_c$ is the

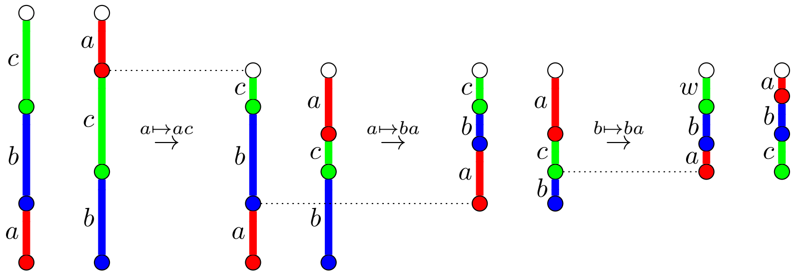


Figure 7.10: The sequence $\xi \in \{\varphi, \psi\}^*$

domain of $\xi(T)$ (where we represent the interval exchanges vertically instead that orizontally as usual).

The sequence is composed of a right induction followed by two left inductions. We have indicated on each edge the associated automorphism (indicating only the image of the letter which is modified). We have $\xi = \varphi^2\psi$ and the resulting composition $\eta$ of automorphisms gives

$$\eta(a) = bac, \quad \eta(b) = bbac, \quad \eta(c) = c.$$

Thus $\mathcal{R}_F(c) = \eta(A)$.

**Example 7.2.18** Let $T$ and $F$ be as in the preceding example. Let $U$ be the transformation induced by $T$ on $J_a$. We have $U = \varphi^6(T)$ and a computation shows that for any $z \in J_a$, $\Sigma_T(z) = \eta(\Sigma_U(z))$ where $\eta$ is the automorphism of the free group on $A = \{a, b, c\}$ which is the coding morphism for $\mathcal{R}_F(a)$ defined by:

$$\eta(a) = ccba, \quad \eta(b) = cbba, \quad \eta(c) = ccbba.$$

One can verify that $\mathcal{L}(U) = \mathcal{L}(S)$, where $S$ is the transformation obtain from $T$ by permuting the labels of the intervals according to the permutation $\pi = (acb)$.

Note that $\mathcal{L}(U) = \mathcal{L}(S)$ although $S$ and $U$ are not identical, even up to rescaling the intervals. Actually, the rescaling of $U$ to a transformation on $[0, 1[$ corresponds to the mirror image of $S$, obtained by taking the image of the intervals by a symmetry centered at $1/2$.

Note that in the above examples, all lengths of the intervals belong to the quadratic number field $\mathbb{Q}[\sqrt{5}]$.

In the next Section we will prove that if a regular interval exchange transformation $T$ is defined over a quadratic field, then the family of transformations obtained from $T$ by the Rauzy inductions contains finitely many distinct transformations up to rescaling.

## 7.3 Interval exchanges over a quadratic field

An interval exchange transformation is said to be defined over a set $Q \subset \mathbb{R}$ if the lengths of all exchanged semi-intervals belong to $Q$.

The following is proved in [19]. Let $T$ be a minimal interval exchange transformation on semi-intervals defined over a quadratic number field. Let $(T_n)_{n\geq 0}$ be a sequence of interval exchange transformation such that $T_0 = T$ and $T_{n+1}$ is the transformation induced by $T_n$ on one of its exchanged semi-intervals $I_n$. Then, up to rescaling all semi-intervals $I_n$ to the same length, the sequence $(T_n)$ contains finitely many distinct transformations. In the same paper, an extension to the right Rauzy induction is suggested (but not completly developed).

In this section we generalize this results and prove that, under the above hypothesis on the lengths of the semi-intervals and up to rescaling and translation, there are finitely many transformations obtained by the branching Rauzy induction defined in Section 7.2.

**Theorem 7.3.1** *Let $T$ be a regular interval exchange transformation defined over a quadratic field. The family of all induced transformation of $T$ over an admissible semi-interval contains finitely many distinct transformations up to equivalence.*

The proof of the Theorem 7.3.1 is based on the fact that for each minimal interval exchange transformation defined over a quadratic field, a certain measure of the arithmetic complexity of the admissible semi-intervals is bounded.

### 7.3.1 Complexities

Let $T$ be an interval exchange transformation on a semi-interval $[\ell, r[$ defined over a quadratic field $\mathbb{Q}[\sqrt{d}]$, where $d$ is a square free integer $\geq 2$. Without loss of generality, one may assume, by replacing $T$ by an equivalent interval exchange transformation if necessary, that $T$ is defined over the ring $\mathbb{Z}[\sqrt{d}] = \{m + n\sqrt{d} \mid m, n \in \mathbb{Z}\}$ and that all $\gamma_i$ and $\alpha_i$ lie in $\mathbb{Z}[\sqrt{d}]$ (replacing $[\ell, r[$ if necessary by its equivalent translate with $\gamma_0 = \ell \in \mathbb{Z}[\sqrt{d}]$).

For $z = m + n\sqrt{d}$, let define $\Psi(z) = \max(|m|, |n|)$.

Let $\mathcal{A}([\ell, r[)$ be the algebra of subsets $X \subset [\ell, r[$ which are finite unions $X = \bigcup_j I_j$ of semi-intervals defined over $\mathbb{Z}[\sqrt{d}]$, that is $I_j = [\ell_j, r_j[$ for some $\ell_j, r_j \in \mathbb{Z}[\sqrt{d}]$. Note that the algebra $\mathcal{A}([\ell, r[)$ is closed under taking finite unions, intersections and passing to complements in $[\ell, r[$.

Set $\partial(X)$ the boundary of $X$ and $|X|$ the Lebesgue measure of $X$. Given a subset $X \in \mathcal{A}([\ell, r[)$, we define the *complexity* of $X$ as $\Psi(X) = \max\{\Psi(z) \mid z \in \partial(X)\}$ and the *reduced complexity* of $X$ as $\Pi(X) = |X|\,\Psi(X)$.

A key tool to prove Theorem 7.3.1 is the following result proved in [19, Theorem 3.1].

**Theorem 7.3.2 (Boshernitzan)** *Let $T$ be a minimal interval exchange transformation on an interval $[\ell, r[$ defined over a quadratic number field. Assume* *that $(J_n)_{\geq 1}$ is a sequence of semi-intervals of $[\ell, r[$ such that the set $\{\Pi(J_n) \mid n \geq 1\}$ is bounded. Then the sequence $T_n$ of interval exchange transformations obtained by inducing $T$ on $J_n$ contains finitely many distinct equivalence classes of interval exchange transformations.*

Thus, in order to prove Theorem 7.3.1, it is sufficient to show that the reduced complexity of every admissible semi-interval is bounded.

The following Proposition is proved in [19, Proposition 2.1]. It shows that the complexity of a subset $X$ and of its image $T(X)$ differ at most by a constant that depends only on $T$.

**Proposition 7.3.3** *There exists a constant $\kappa = \kappa(T)$ such that for every $X \in \mathcal{A}([\ell, r[)$ and $z \in [\ell, r[$ one has $|\Psi(T(X)) - \Psi(X)| \leq \kappa$ and $\Psi(T(z) - z) \leq \kappa$. Moreover, one has $\Psi(\gamma) \leq \kappa$ and $\Psi(T(\gamma)) \leq \kappa$ for every separation point $\gamma$.*

Clearly, by Proposition 7.3.3, one also has $|\Psi(T^{-1}(X)) - \Psi(X)| \leq \kappa$ for every $X \in \mathcal{A}([\ell, r[)$ and $\Psi(T^{-1}(z) - z) \leq \kappa$ for every $z \in [\ell, r[$.

Although it is not necessary for our purposes, we can improve the approximation of the reduced complexity of a nonempty subset $X \in \mathcal{A}([\ell, r|)$ by the following proposition. This result, proved in [19, Proposition 2.4], determines a lower bound on $\Pi(X)$.

**Proposition 7.3.4** *Let $X \in \mathcal{A}([\ell, r[)$ be a subset composed of $n$ disjoints semi-intervals. Then $\Pi(X) > n/(4\sqrt{d})$.*

### 7.3.2 Return times

Let $T$ be an interval exchange transformation. For a subset $X \in \mathcal{A}([\ell, r[)$ we define the *maximal positive return time* and *maximal negative return time* for $T$ on $X$ by the functions

$$\rho^+(X) = \min\left\{n \geq 1 \,|\, T^n(X) \subset \textstyle\bigcup_{i=0}^{n-1} T^i(X)\right\},$$

and

$$\rho^-(X) = \min\left\{m \geq 1 \,|\, T^m(X) \subset \textstyle\bigcup_{i=0}^{m-1} T^{-i}(X)\right\}.$$

We also define the *minimal positive return time* and the *minimal negative return time* as

$$\sigma^+(X) = \min\left\{n \geq 1 \,|\, T^n(X) \cap X \neq \emptyset\right\},$$

and

$$\sigma^-(X) = \min\left\{m \geq 1 \,|\, T^{-m}(X) \cap X \neq \emptyset\right\}.$$

If $T$ is minimal, it is clear that for every $X \in \mathcal{A}([\ell, r[)$, one has

$$[\ell, r[ = \textstyle\bigcup_{i=0}^{\rho^+(X)-1} T^i(X) = \bigcup_{i=0}^{\rho^-(X)-1} T^{-i}(X).$$

Note that when $J$ is a semi-interval, we have

$$\rho^+(J) = \max_{z \in J} \rho^+_{J,T}(z) \quad \text{and} \quad \sigma^+(J) = \min_{z \in J} \rho^+_{J,T}(z).$$

Symmetrically

$$\rho^-(J) = \max_{z \in J} \rho^-_{J,T}(z) + 1 \quad \text{and} \quad \sigma^-(J) = \min_{z \in J} \rho^-_{J,T}(z) + 1.$$

Let $\zeta, \eta$ be two functions. We write $\zeta \in O(\eta)$ if there exists a constant $C$ such that $|\zeta| \leq C|\eta|$. We write $\zeta \in \Theta(\eta)$ if one has both $\zeta \in O(\eta)$ and $\eta \in O(\zeta)$. Note that $\Theta$ is an equivalence relation, that is $\zeta \in \Theta(\eta) \Leftrightarrow \eta \in \Theta(\zeta)$.

Boshernitzan and Carroll give in [19] two upper bounds for $\rho^+(X)$ and $\sigma^+(X)$ for a subset $X$ (Theorems 2.5 and 2.6 respectively) and a more precise estimation when the subset is a semi-interval (Theorem 2.8). Some slight modifications of the proofs can be made so that the results hold also for $\rho^-$ and $\sigma^-$. We summarize these estimates in the following theorem.

**Theorem 7.3.5** *For every $X \in \mathcal{A}([\ell, r[)$ one has $\rho^+(X), \rho^-(X) \in O(\Psi(X))$ and $\sigma^+(X), \sigma^-(X) \in O\left(1/|X|\right)$. Moreover, if $T$ is minimal and $J$ is a semi-interval, then $\rho^+(J) \in \Theta\left(\rho^-(J)\right) = \Theta\left(\sigma^+(J)\right) = \Theta\left(\sigma^-(J)\right) = \Theta\left(1/|J|\right)$.*

An immediate corollary of Theorem 7.3.5 is the following

**Corollary 7.3.6** *Let $T$ be minimal and assume that*

$$\{T^i(z) \mid -m+1 \leq i \leq n-1\} \cap J = \emptyset$$

*for some point $z \in [\ell, r[$, some semi-interval $J \subset [\ell, r[$ and some integers $m, n \geq 1$. Then $|J| \in O\left(1/\max\{m,n\}\right)$.*

*Proof.* By the hypothesis, $z \notin \bigcup_{i=0}^{n-1} T^{-i}(J)$, then we have $\rho^-(J) \geq n$. By Theorem 7.3.5, we obtain $|J| \in \Theta\left(1/\rho^-(J)\right) \subseteq O\left(1/n\right)$. Symmetrically, since $\rho^+(J) \geq m$, one has $|J| \in O\left(1/m\right)$. Then

$$|J| \in O\left(\min\left\{\frac{1}{m}, \frac{1}{n}\right\}\right) = O\left(\frac{1}{\max\{m,n\}}\right).$$

■

### 7.3.3 Reduced complexity of admissible semi-intervals

In order to obtain Theorem 7.3.1, we prove some preliminary results concerning the reduced complexity of admissible semi-intervals.

Let $T$ be an $s$-interval exchange transformation. Recall from Section 6.1.1 that we denote by $\mathrm{Sep}(T) = \{\gamma_i \mid 0 \leq i \leq s-1\}$ the set of separation points. For every $n \geq 0$ define $\mathcal{S}_n(T) = \bigcup_{i=0}^{n-1} T^{-i}\big(\mathrm{Sep}(T)\big)$ with the convention $\mathcal{S}_0 = \emptyset$.

Since $\mathrm{Sep}(T^{-1}) = T\big(\mathrm{Sep}(T)\big)$, one has

$$\mathcal{S}_n(T^{-1}) = T^{n-1}\big(\mathcal{S}_n(T)\big).$$

Given two integers $m, n \geq 1$, we can define

$$\mathcal{S}_{m,n} = \mathcal{S}_m(T) \cup \mathcal{S}_n(T^{-1}).$$

An easy calculation shows that

$$\mathcal{S}_{m,n}(T) = \bigcup_{i=-m+1}^{n} T^i\big(\mathrm{Sep}(T)\big).$$

Observe also that $\mathcal{S}_{m,n}(T) = T^n\big(\mathcal{S}_{m+n}(T)\big) = T^{-m+1}\big(\mathcal{S}_{m+n}(T)\big)$.

Denote by $\mathcal{V}_{m,n}(T)$ the family of semi-intervals whose endpoints are in $\mathcal{S}_{m,n}(T)$. Put $\mathcal{V}(T) = \bigcup_{m,n\geq 0} \mathcal{V}_{m,n}(T)$. Every admissible semi-interval belongs to $\mathcal{V}(T)$, while the converse is not true.

**Theorem 7.3.7** $\Pi(J) \in \Theta(1)$ *for every semi-interval $J$ admissible for $T$.*

*Proof.* Let $m, n$ be the two minimal integers such that $J = [t, w[ \in \mathcal{V}_{m,n}(T)$. Then $t, w \in \{T^m(\gamma_i) \,|\, 1 \leq i \leq s\} \cup \{T^{-n}(\gamma_i) \,|\, 1 \leq i \leq s\}$. Suppose, for instance, $t = T^M(\gamma)$, with $M = \max\{m, n\}$ and $\gamma$ a separation point. The other cases (namely, $t = T^{-M}(\gamma)$, $w = T^M(\gamma)$ or $w = T^{-M}(\gamma)$) are proved similarly.

The only semi-interval in $\mathcal{V}_{0,0}(T)$ is $[\ell, r[$ and clearly in this case the theorem is verified.

Suppose then that $J \in \mathcal{V}_{m,n}(T)$ for some nonnegative integers $m, n$ with $m + n > 0$. We have $\Psi(J) = \max\{\Psi(t), \Psi(w)\} \leq M\kappa$ where $\kappa$ is the constant introduced in Proposition 7.3.3. Moreover, by the definition of admissibility one has $\{T^j(\gamma) \,|\, 1 \leq j \leq M\} \cap J = \emptyset$. Thus, by Corollary 7.3.6 we have $|J| \in O(1/M)$. Then $\Pi(J) = |J|\;\Psi(J) \in O(1)$. By Proposition 7.3.4 we have $\Pi(J) > 1/(4\sqrt{d})$. This concludes the proof. ∎

Denote by $\mathcal{U}_{m,n}(T)$ the family of semi-intervals partitioned by $\mathcal{S}_{m,n}(T)$. Clearly $\mathcal{V}_{m,n}(T)$ contains $\mathcal{U}_{m,n}(T)$. Indeed every semi-interval $J \in \mathcal{V}_{m,n}(T)$ is a finite union of contiguous semi-intervals belonging to $\mathcal{U}_{m,n}(T)$.

Note that $\mathcal{U}_{m,0}(T)$ is the family of semi-intervals exchanged by $T^m$, while $\mathcal{U}_{0,n}(T)$ is the family of semi-intervals exchanged by $T^{-n}$.

Put

$$\mathcal{U}(T) = \bigcup_{m,n\geq 0} \mathcal{U}_{m,n}(T).$$

Using Theorem 7.3.7 we easily deduce the following corollary, which is a generalization of Theorem 2.11 in [19].

**Corollary 7.3.8** $\Pi(J) \in \Theta(1)$ *for every semi-interval $J \in \mathcal{U}(T)$.*

We are now able to prove Theorem 7.3.1.

*Proof of Theorem 7.3.1.* By Theorem 7.2.3, every admissible semi-interval can be obtained by a finite sequence $\xi$ of right and left Rauzy inductions. Thus we can enumerate the family of all admissible semi-intervals. The conclusion easily follows from Theorem 7.3.2 and Theorem 7.3.7. ∎

An immediate corollary of Theorem 7.3.1 is the following.

**Corollary 7.3.9** *Let $T$ be a regular interval exchange transformation defined over a quadratic field. Then the induction graph $\mathcal{IG}(T)$ and the modified induction graph $\widetilde{\mathcal{IG}}(T)$ are finite.*

**Example 7.3.10** Let $T$ be the regular interval exchange transformation of Example 6.1.7. The modified induction graph $\widetilde{\mathcal{IG}}(T)$ is represented in Figure 7.11. The transformation $T$ belongs to the similarity class $\langle T_1 \rangle$ as well as transformations $S$ of Example 7.2.8 and $U$ of Example 7.2.10. The transformations $\psi(T)$ and $\psi^2(T)$ of Example 7.2.2 belongs respectively to classes $\langle T_2 \rangle$ and $\langle T_4 \rangle$, while the two last transformations of Figure 7.10, namely $\varphi\psi(T)$ and $\varphi^2\psi(T)$, belongs respectively to $\langle T_5 \rangle$ and $\langle T_7 \rangle$. Finally, the left Rauzy induction sequence from $T$ to $U = \varphi^6(T)$ corresponds to the loop $\langle T_1 \rangle \to \langle T_3 \rangle \to \langle T_4 \rangle \to \langle T_6 \rangle \to \langle T_7 \rangle \to \langle T_8 \rangle \to \langle T_1 \rangle$ in $\widetilde{\mathcal{IG}}(T)$.

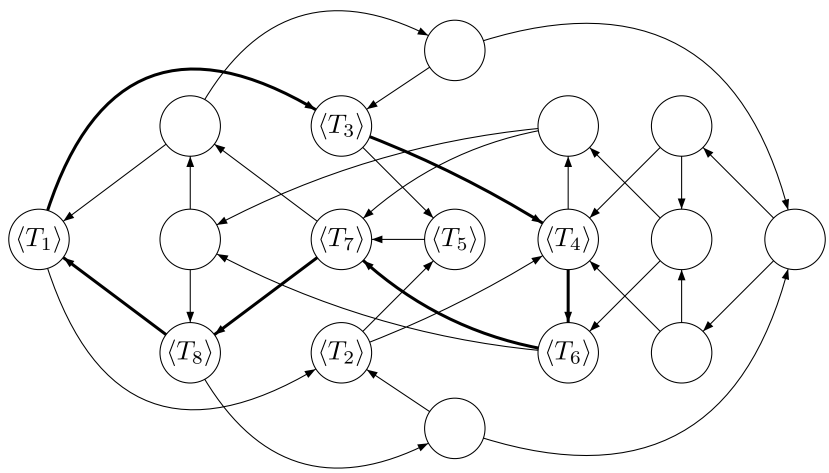


Figure 7.11: Modified induction graph of the transformation $T$.

### 7.3.4 Primitive morphic sets

In this final subsection we show an important property of interval exchange transformations defined over a quadratic field, namely that the related interval exchange sets are primitive morphic. Let prove first the following result.

**Proposition 7.3.11** *Let $T, \xi(T)$ be two equivalent regular interval exchange transformations with $\xi \in \{\varphi, \psi\}^*$. There exists a primitive morphism $\eta$ and a point $z \in D(T)$ such that the natural coding of $T$ relative to $z$ is a fixed point of $\eta$.*

*Proof.* By Proposition 6.1.12, the set $\mathcal{L}(T)$ is uniformly recurrent. Thus, there exists a positive integer $N$ such that every letter of the alphabet appears in every word of length $N$ of $\mathcal{L}(T)$. Moreover, by Theorem 7.2.7, applying iteratively the Rauzy induction, the length of the domains tends to zero.

Consider $T' = \chi^m(T)$, for a positive integer $m$, such that $D(T') < \varepsilon$, where $\varepsilon$ is the positive real number for which, by Lemma 6.1.10, the first return map for every point of the domain is "longer" than $N$, i.e. $T'(z) = T^{n(z)}(z)$, with $n(z) \geq N$, for every $z \in D(T')$.

By Theorem 7.2.15 and the remark following it, there exists an automorphism $\eta$ of the free group and a point $z \in D(T') \subseteq D(T)$ such that the natural coding of $T$ relative to $z$ is a fixed point of $\eta$, that is $\Sigma_T(z) = \eta(\Sigma_T(z))$.

By the previous argument, the image of every letter by $\eta$ is longer than $N$, hence it contains every letter of the alphabet as a factor. Therefore, $\eta$ is a primitive morphism. ■

**Theorem 7.3.12** *Let $T$ be a regular interval exchange transformation defined over a quadratic field. The interval exchange set $\mathcal{L}(T)$ is primitive morphic.*

*Proof.* By Theorem 7.3.1 there exists a regular interval transformation $S$ such that we can find in the induction graph $\mathcal{IG}(T)$ a path from $[T]$ to $[S]$ followed by a cycle on $[S]$. Thus, by Theorem 7.2.15 there exists a point $z \in D(S)$ and two automorphisms $\eta, \zeta$ of the free group such that $\Sigma_T(z) = \eta(\Sigma_S(z))$, with $\Sigma_S(z)$ a fixed point of $\zeta$.

By Proposition 7.3.11 we can suppose, without loss of generality, that $\zeta$ is primitive. Therefore, $\mathcal{L}(T)$ is a primitive morphic set. ■

**Example 7.3.13** Let $T = T_{\lambda,\pi}$ be the transformation of Example 6.1.7 (see also 6.1.13). The set $\mathcal{L}(T)$ is primitive morphic. Indeed the transformation $T$ is regular and the length vector $\lambda = (1 - 2\alpha, \alpha, \alpha)$ belongs to $\mathbb{Q}\left[\sqrt{5}\right]^3$.

# Chapter 8

# Linear involutions

In this chapter, we define linear involutions, which are a generalization of interval exchange transformations seen in Chapters 6 and 7.

As for interval exchanges, we can associate to every linear involution $T$ a language, its natural coding $\mathcal{L}(T)$. The family of natural codings of linear involutions without connection is an important class of specular sets.

Section 8.1 is devoted to the dynamical properties of linear involutions. We study, in particular, some remarkable classes, such as coherent, orientable and minimal linear involutions.

In Section 8.2 we define the natural coding of a linear involution and we show that, under certain hypothesis, this set is a specular set (Theorem 8.2.11). We also give some results about orientability (Proposition 8.2.5) and mixed return words (Section 8.2.4) in this framework. We end the section with the notion of admissible interval for a linear involution (Section 8.2.5), that generalize the analougous notion seen in Chapter 7 for interval exchanges.

## 8.1 Linear involution

In this section we introduce linear involutions. This family of dynamical systems is closely related to the family of interval exchanges seen in Chapters 6 and 7.

After giving the basic definitions in Section 8.1.1, we discuss the similarity between linear involutions and interval exchanges in Section 8.1.2. Sections 8.1.3 and 8.1.4 are devoted to some remarkable classes of linear involutions: coherent, orientable and minimal ones.

### 8.1.1 Generalized permutations and linear involutions

Let us 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.

Let $A$ be an alphabet of cardinality $k$, with $k$ an even number. Let $\theta$ be an involution on $A$. We denote by $a^{-1}$ or $\bar{a}$ the inverse of a letter $a \in A$.

A *generalized permutation* on $A$ of type $(\ell, m)$, 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 $(\ell, m, \pi)$ 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 $\ell$ 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 $\Sigma$ 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 $\sigma_1$ 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 $\sigma_2$ 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 8.1.1** 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 8.1 (we represent $I \times \{0\}$ above $I \times \{1\}$) with the assumption that the restriction of $\sigma_1$ to $I_a$ and $I_d$ is a symmetry while its restriction to $I_b, I_c$ is a translation.

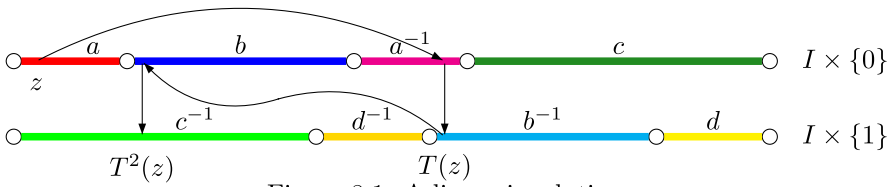


Figure 8.1: A linear involution.

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.

### 8.1.2 Linear involutions and interval exchanges

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 $\sigma_1$ 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 [57, 58]. Assume again conditions (i) and (ii), but now that the restriction of $\sigma_1$ 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 chapter interval exchange transformations defined by a partition of an open interval minus $\ell - 1$ points in $\ell$ open intervals, instead that using a partition of a semi-interval in a finite number of semi-intervals as in Chapter 6.

### 8.1.3 Coherent and orientable linear involutions

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

The set $\Sigma$ 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 $\pi$ of type $(\ell, m)$ is said to be *nonorientable* 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 $\sigma_1$ to $I_a$ is a translation if and only if $I_a$ and $I_{a^{-1}}$ belong to distinct components of $\hat{I}$.

**Example 8.1.2** The linear involution of Example 8.1.1 is coherent.

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

Orientable linear involutions correspond to orientable laminations (see [15]), whereas coherent linear involutions correspond to orientable surfaces. Thus coherent nonorientable involutions correspond to nonorientable laminations on orientable surfaces (see [15]).

### 8.1.4 Minimal involutions

As for interval exchanges, we define a *connection* of a linear involution $T$ as a triple $(x, y, n)$ such that $x$ is a singularity of $T^{-1}$, $y$ is a singularity of $T$, $n \geq 0$ and $T^n(x) = y$.

**Example 8.1.3** Let us consider the linear involution $T$ which is the same as in Example 8.1.1 but such that the restriction of $\sigma_1$ 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 8.1.4** Let $T$ be the linear involution on $I = ]0, 1[$ represented in Figure 8.2. We assume that the restriction of $\sigma_1$ 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.

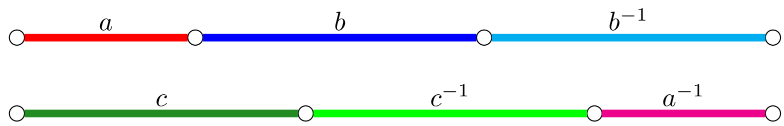


Figure 8.2: A linear involution without connections.

Let $T$ be a linear involution without connections. Let

$$\mathcal{O} = \bigcup_{n \geq 0} T^{-n}(\Sigma) \quad \text{and} \quad \hat{\mathcal{O}} = \mathcal{O} \cup \sigma_2(\mathcal{O}) \tag{8.1}$$

be respectively the negative orbit of the singular points and its closure under mirror image. Then $T$ is a bijection from $\hat{I} \setminus \hat{\mathcal{O}}$ onto itself. Indeed, assume that $T(z) \in \hat{\mathcal{O}}$. If $T(z) \in \mathcal{O}$ then $z \in \mathcal{O}$. Next if $T(z) \in \sigma_2(\mathcal{O})$, 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 $z \in T^{n-1}(\sigma_2(\Sigma)) = \sigma_2(T^{-n+1}(\Sigma)) \subset \sigma_2(\mathcal{O})$.

Therefore in both cases $z \in \hat{\mathcal{O}}$. The converse implication is proved in the same way.

A linear involution $T$ on $I$ without connections is *minimal* if for any point $z \in \hat{I} \setminus \hat{\mathcal{O}}$ 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 [29] that noncoherent linear involutions are almost surely not minimal.

**Example 8.1.5** Let us consider the noncoherent linear involution $T$ which is the same as in Example 8.1.1 but such that the restriction of $\sigma_1$ to $I_c$ is a symmetry, as in Example 8.1.3. We assume that $I = ]0,1[$, that $\lambda_a = \lambda_d$, 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 8.3). 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 $T^3(z) = z$.

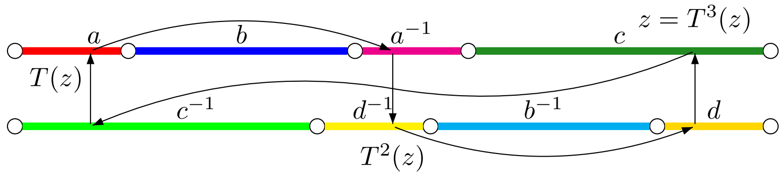

Figure 8.3: A noncoherent linear involution.

Let $X \subset I \times \{0,1\}$. The *return time* $\rho_X$ to $X$ is the function from $I \times \{0,1\}$ to $\mathbb{N} \cup \{\infty\}$ defined on $X$ by

$$\rho_X(x) = \inf\{n \geq 1 \mid T^n(x) \in X\}.$$

The following result was already proved in [18, Proposition 4.2] for the class of coherent involutions. The proof uses Keane's theorem proving that an interval exchange transformation without connections is minimal (Theorem 6.1.6). The proof of Keane's theorem also implies that for each interval of positive length, the return time to this interval is bounded.

**Proposition 8.1.6** *Let $T$ be a linear involution without connections on $I$. If $T$ is nonorientable, it is minimal. Otherwise, its restriction to each component of $\hat{I}$ is minimal. Moreover, for each interval of positive length included in $\hat{I}$, the return time to this interval takes a finite number of values.*

*Proof.* Consider the set $\widetilde{I} = \hat{I} \times \{0,1\} = I \times \{0,1\}^2$ and the transformation $\widetilde{T}$

on $\widetilde{I}$ defined for $(x, \delta) \in \widetilde{I}$ by

$$\widetilde{T}(x,\delta) = \begin{cases} (T(x),\delta) & \text{if } T \text{ is a translation on a neighborhood of } x \\ (T(x), 1-\delta) & \text{otherwise.} \end{cases}$$

Let $T'$ be the transformation induced by $\widetilde{T}$ on $I' = I \times \{0,0\}$. Note that if $x \in I'$ is recurrent, that is, $\widetilde{T}^n(x) \in I'$ for some $n > 0$, then the restriction of $T'$ to some neighborhood of $x$ is a translation. Indeed, there is an even number of indices $i$ with $0 \le i < n$ such that $T$ is a symmetry on a neighborhood of $T^i(x)$.

Let us show that $T'$ is an interval exchange transformation. Let $\Sigma$ be the set of singularities of $T$. For each $z \in \Sigma$, let $s(z)$ be the minimal integer $s > 0$ (or $\infty$) such that $\widetilde{T}^{-s}(z) \in I'$. Let $N = \{\widetilde{T}^{-s(z)}(z) \mid z \in \Sigma \text{ with } s(z) < \infty\}$. The set $N$ divides $I'$ into a finite number of disjoint open intervals. If $J$ is such an open interval, it contains, by the Poincaré Recurrence Theorem, at least one recurrent point $x \in I'$ for $\widetilde{T}$, that is such that $\widetilde{T}^n(x) \in I'$ for some $n > 0$. By definition of $N$, all the points of $J$ are recurrent. Moreover, as we have seen above, the restriction of $T'$ to $J$ is a translation. This shows that $T'$ is an interval exchange transformation.

We can now conclude the proof. Since $T$ has no connection, $T'$ has no connection. Thus, by Keane's theorem, it is minimal. This shows that the intersection with $I \times \{0\}$ of the nonnegative orbit of any point in $I \times \{0\}$ is dense in $I \times \{0\}$. A similar proof shows that the same is true for $I \times \{1\}$. If $T$ is nonorientable, the nonnegative orbit of any $x \in I \times \{0\}$ contains a point in $I \times \{1\}$. Thus its nonnegative orbit is dense in $\hat{I}$. The same holds symmetrically for $x \in I \times \{1\}$.

Let $J$ be an interval of positive length included in $I$. By Keane's theorem, the return time to $J \times \{0,0\}$ relative to $T'$ takes a finite number of values. Thus the return time to $J \times \{0\}$ with respect to $T$ takes also a finite number of values. A similar argument holds for an interval included in $I \times \{1\}$. ■

## 8.2 Natural coding

In this section we use the linear involutions defined in the previous part of the chapter to construct a laminary set of words: the natural coding $\mathcal{L}(T)$ of a linear involution $T$.

The definition of natural coding is given in Section 8.2.1. We show that a word $u$ is in $\mathcal{L}(T)$ if and only if the correspective open interval $I_u$ is nonempty (Lemma 8.2.2). In Proposition 8.2.3 we show that $\mathcal{L}(T)$ is actually a laminary set.

Section 8.2.2 is devoted to the study of orientable sets and orientable linear involutions. In particular we show in Proposition 8.2.5 how the dynamical notion is related to the one on laminary sets.

In Section 8.2.3 we show that the natural coding of a linear involution without connection is a specular set (Theorem 8.2.11).

In Section 8.2.4 we return to de notion of mixed return words defined in Chapter 5. We also give a geometrical characterization of the set of mixed return words (Lemma 8.2.14).

Finally, in Section 8.2.5 we introduce the notion of admissible interval for a linear involution and we show that if $T$ is without connection, every interval $I_w$ is admissible with respect to $T$ (Proposition 8.2.19).

### 8.2.1 Infinite natural coding

Let $T$ be a linear involution on $I$, let $\hat{I} = I \times \{0,1\}$ and let $\hat{\mathcal{O}}$ be the set defined by Equation (8.1).

Given $z \in \hat{I} \setminus \hat{\mathcal{O}}$, the *infinite natural coding* of $T$ relative to $z$ is the infinite word $\Sigma_T(z) = a_0 a_1 \ldots$ 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.

As for the interval exchange sets (see Chapter 6), 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 8.2.1** Let $T$ be the linear involution of Example 8.1.4. The words of length at most 3 of $S = \mathcal{L}(T)$ are represented in Figure 5.3.

Similarly that with interval exchanges (see Section 6.1.2), given a word $w = b_0 b_1 \cdots b_{m-1}$, we define the interval $I_w$ as

$$I_w = I_{b_0} \cap T^{-1}(I_{b_1}) \cap \ldots \cap T^{-m+1}(I_{b_{m-1}}). \tag{8.2}$$

By convention, $I_\varepsilon = \hat{I}$.

For any $z \in \hat{I} \setminus \hat{\mathcal{O}}$, one has $z \in I_u$ if and only if $u$ is a prefix of $\Sigma_T(z)$.

Each set $I_u$ is a (possibly empty) open interval. Indeed, this is true if $u$ is a letter. Next, assume that $I_u$ is an open interval. Note that

$$I_{au} = I_a \cap T^{-1}(I_u). \tag{8.3}$$

Then, by (8.3), for $a \in A$, we have $T(I_{au}) = T(I_a) \cap I_u$ and thus $T(I_{au})$ is an open interval. Since $I_{au} \subset I_a$, $T(I_{au})$ is the image of $I_{au}$ by a continuous map and thus $I_{au}$ is also an open interval.

We have the following result.

**Lemma 8.2.2** *Let $T$ be a linear involution. Let $u$ be a nonempty word. We have*

$$u \in \mathcal{L}(T) \Longleftrightarrow I_u \neq \emptyset.$$

*Proof.* If $u$ is a factor of $\Sigma_T(z)$ for some $z \in \hat{I} \setminus \hat{\mathcal{O}}$, then $T^n(z) \in I_u$ for some $n \geq 0$ and thus $I_u \neq \emptyset$. Conversely, if $I_u \neq \emptyset$, since $I_u$ is an open interval, it contains some $z \in \hat{I} \setminus \hat{\mathcal{O}}$. Then $u$ is a prefix of $\Sigma_T(z)$ and thus $u \in \mathcal{L}(T)$. ■

Observe that if $T$ is nonorientable and without connection, then by Proposition 8.1.6, $\mathcal{L}(T)$ is the set of factors of $\Sigma_T(z)$ for any $z \in \hat{I} \setminus \hat{\mathcal{O}}$, that is, the set of factors of $\Sigma_T(z)$ does not depend on $z$. Indeed, if $I_u \neq \emptyset$, since the orbit of $z$ is dense in $\hat{I}$, there is an $n \geq 0$ such that $T^n(z) \in I_u$ and thus $u$ is a factor of $\Sigma_T(z)$.

**Proposition 8.2.3** *Let $T = \sigma_2 \circ \sigma_1$ be a linear involution. For any nonempty word $u \in \mathcal{L}(T)$, one has $I_{u^{-1}} = \sigma_1 T^{|u|-1}(I_u)$. Consequently the set $\mathcal{L}(T)$ is closed under taking inverses. It is thus a laminary set.*

*Proof.* To prove the assertion, we use an induction on the length of $u$. The property holds for $|u| = 1$ by definition of $\sigma_1$. Next, consider $u \in \mathcal{L}(T)$ and $a \in A$ such that $ua \in \mathcal{L}(T)$. We assume by induction hypothesis that $I_{u^{-1}} = \sigma_1 T^{|u|-1}(I_u)$.

Since $T^{-1} = \sigma_1 \circ \sigma_2$,

$$\begin{aligned}
\sigma_1 T^{|u|}(I_{ua}) &= \sigma_1 T^{|u|}(I_u \cap T^{-|u|}(I_a)) = \sigma_1 T^{|u|}(I_u) \cap \sigma_1(I_a) \\
&= \sigma_1 \sigma_2 \sigma_1 T^{|u|-1}(I_u) \cap \sigma_1(I_a) = \sigma_1 \sigma_2 (I_{u^{-1}}) \cap I_{a^{-1}} \\
&= I_{a^{-1}u^{-1}}
\end{aligned}$$

where the last equality results from the application of Equation (8.3) to the word $a^{-1}u^{-1}$.

We easily deduce that the set $\mathcal{L}(T)$ is closed under taking inverses. Furthermore it is a factorial subset of the free group $F_A$. It is thus a laminary set. ■

**Example 8.2.4** Let $T$ be the linear involution of Example 8.1.4. As seen in Example 5.2.3, the set $S = \mathcal{L}(T)$ can actually be defined directly as the set of factors of the substitution

$$f : a \mapsto c\bar{b}, \quad b \mapsto c, \quad c \mapsto a\bar{b}$$

(where we use the notation $\bar{\cdot}$ instead of $\cdot^{-1}$ ) which extends to an automorphism of the free group $F_A$. The verification uses the (one-side) Rauzy induction seen in Chapter 7 extended to linear involutions (see [18] for more details). The Rauzy induction applied to $T$ gives the linear involution $T'$ represented in Figure 8.4 on the left. It is the transformation induced by $T$ on the interval obtained by erasing the smallest interval on the right, namely $I_{\bar{a}}$.

The Rauzy induction applied on $T'$ is obtained by erasing the smallest interval on the right, namely $I_{\bar{b}}$. It gives a transformation $T''$ represented in Figure 8.4 on the right.

The transformation $T''$ is the same as $T$ up to normalization of the length of the interval, exchange of the two components and the permutation (written in cycle form) $\pi = (a\,c\,b\,\bar{a}\,\bar{c}\,\bar{b})$ (see Figure 8.4) which sends $a$ to $c$, $c$ to $b$ and so on.

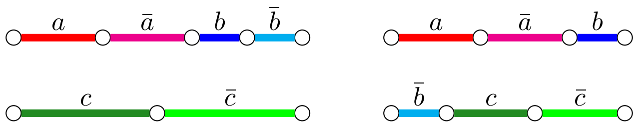


Figure 8.4: The transforms $T'$ and $T''$ of $T$ by Rauzy induction.

Set $S = \mathcal{L}(T)$, $S' = \mathcal{L}(T')$ and $S'' = \mathcal{L}(T'')$. Since $T'$ is obtained from $T$ by a Rauzy induction, there is an associated automorphism $\tau'$ of the free group such that $S = \mathrm{Fac}(\tau'(S'))$ (see Section 7.2.4). One has actually

$$\tau : a \mapsto a\bar{b}, \quad b \mapsto b, \quad c \mapsto c.$$

Similarly, one has $S' = \mathrm{Fac}(\tau''(S''))$ with

$$\tau'' : a \mapsto a, \quad b \mapsto b\bar{c}, \quad c \mapsto c.$$

Set $\tau = \tau' \circ \tau''$. It is easy to verify that $f = \tau \circ \pi^{-1}$. Since $S = \mathrm{Fac}(\tau(S'')) = \mathrm{Fac}(\tau\pi^{-1}(S)) = \mathrm{Fac}(f(S))$, we obtain that $S$ is the set of factors of the fixpoint of $f$ as claimed above.

### 8.2.2 Orientability and uniform recurrence

We gather here basic properties of the language $\mathcal{L}(T)$ of a linear involution. Recall from Section 1.1.1 that a laminary set $S$ is called orientable if $S = S_+ \cup S_-$ with $S_+, S_-$ two factorial sets such that $S_+ \cap S_- = \{\varepsilon\}$ and for any $x \in S$, one has $x \in S_-$ if and only if $x^{-1} \in S_+$.

**Proposition 8.2.5** *Let $T$ be a linear involution. If $T$ is orientable, then $\mathcal{L}(T)$ is orientable. The converse is true if $T$ has no connection.*

*Proof.* Let $T$ be a linear involution and let $S = \mathcal{L}(T)$. Assume that $T$ is orientable. Set $S_+ = \{u \in S \mid I_u \subset I \times \{0\}\} \cup \{\varepsilon\}$ and $S_- = \{u \in S \mid I_u \subset I \times \{1\}\} \cup \{\varepsilon\}$. Then $S = S_+ \cup S_-$. Since $T$ is orientable, we have $u \in S_+$ (resp. $u \in S_-$) if and only if all letters of $u$ are in $S_+$ (resp. in $S_-$). This shows that $S_+ \cap S_- = \{\varepsilon\}$, that $S_+, S_-$ are factorial, and that $u \in S_+$ if and only if $u^{-1} \in S_-$. Thus $S$ is orientable.

Conversely, assume that $T$ is nonorientable and has no connection. Let $a \in A$ be such that $I_a, I_{a^{-1}} \subset I \times \{0\}$. Since $T$ is minimal by Proposition 8.1.6, there is some $z \in I_a$ and $n > 0$ such that $T^n(z) \in I_{a^{-1}}$. Thus $S$ contains a word of the form $aua^{-1}$. This implies that $S$ is nonorientable. ∎

The following statement can be easily deduced from the similar statement for interval exchange transformations (see [16, p. 392]).

**Proposition 8.2.6** *Let $T$ be a linear involution without connection. If $T$ is nonorientable, then $\mathcal{L}(T)$ is uniformly recurrent. Otherwise, $\mathcal{L}(T)$ is uniformly semi-recurrent.*

*Proof.* Set $S = \mathcal{L}(T)$. Let $u \in S$ and let $N$ be the maximal return time to $I_u$ (this exists by Proposition 8.1.6). Thus for any $z \in \hat{I}$ such that $\rho_{I_u}(z)$ is finite, we have $\rho_{I_u}(z) \leq N$. Let $w$ be a word of $S$ of length $N + |u|$ and let $z \in \hat{I} \setminus \hat{O}$ be such that $\Sigma_T(z)$ begins with $w$.

If $T$ is nonorientable, by Proposition 8.1.6, it is minimal. Thus there exists $n > 0$ such that $T^n(z) \in I_u$. This implies that $\rho_{I_u}(z)$ is finite and thus that $\rho_{I_u}(z) \leq N$. This implies in turn that $u$ is a factor of $w$. We conclude that $S$ is uniformly recurrent.

If $T$ is orientable, then the restriction of $T$ to each component of $\hat{I}$ is minimal. By Proposition 8.2.5, $S$ is orientable. Thus $I_u$ and $I_{u^{-1}}$ cannot be included in the same component of $\hat{I}$, since otherwise $S$ would contain a word of the form $uvu^{-1}$, and $S$ would be nonorientable. Thus $I_w$ is in the same component as $I_u$ or $I_{u^{-1}}$, and we conclude as above that $u$ or $u^{-1}$ is a factor of $w$. This shows that $S$ is uniformly semi-recurrent. ∎

### 8.2.3 Linear involutions and specular sets

The following theorem is proved in a similar way as Theorem 6.1.16.

**Theorem 8.2.7** *The natural coding of a linear involution without connection is a planar tree set set of characteristic 2.*

In order to prove Theorem 8.2.7 we need some preliminary results. The following one is proved in the same way as Lemma 6.1.17 (see also Figure 8.5).

**Lemma 8.2.8** *Let $T$ be a linear involution. For every nonempty word $w$ and letter $a \in A$, one has*

*(i)* $a \in L(w) \Leftrightarrow \sigma_2(I_{a^{-1}}) \cap I_w \neq \emptyset$,

*(ii)* $a \in R(w) \Leftrightarrow \sigma_2(I_a) \cap I_{w^{-1}} \neq \emptyset$.

*Proof.* By Lemma 8.2.2, we have $a \in L(w)$ if and only if $I_{aw} \neq \emptyset$ which is also equivalent to $T(I_{aw}) \neq \emptyset$. As for interval exchanges, one has $T(I_{aw}) = T(I_a) \cap I_w$. Since $T = \sigma_2 \circ \sigma_1$ and since $\sigma_1(I_a) = I_{a^{-1}}$, $a \in L(w)$ if and only if $\sigma_2(I_{a^{-1}}) \cap I_w \neq \emptyset$. Next, since $\mathcal{L}(T)$ is closed under taking inverses (see 8.2.3), $aw \in S$ if and only if $w^{-1}a^{-1} \in S$. Thus $a \in R(w)$ if and only if $a^{-1} \in L(w^{-1})$, whence the second equivalence. ∎

Recall from Section 6.1 that, given two subsets $I, J$ of the real line, we write $I < J$ if $x < y$ for any $x \in I$ and $y \in J$.

Given a linear involution $T$ on $I$, we introduce two orders on $\mathcal{L}(T)$ as follows. For any $u, v \in \mathcal{L}(T)$, one has

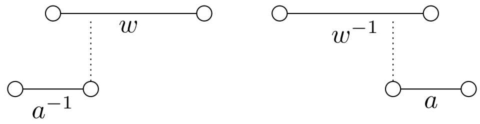


Figure 8.5: An illustration of $a \in L(w)$ and $a \in R(w)$.

(i) $u <_R v$ if and only if $I_u < I_v$,

(ii) $u <_L v$ if and only if $I_{u^{-1}} < I_{v^{-1}}$.

The following lemma is proved in the same way as Lemma 6.1.18.

**Lemma 8.2.9** *Let $T$ be a linear involutions on $I$ without connection. Let $w \in \mathcal{L}(T)$ and $a, a' \in L(w)$ (resp. $b, b' \in R(w)$). Then $1 \otimes a, 1 \otimes a'$ (resp. $b \otimes 1, b' \otimes 1$) are in the same connected component of $\mathcal{E}(w)$ if and only if $I_{a^{-1}}, I_{a'^{-1}}$ (resp. $I_b, I_{b'}$) are in the same component of $I$.*

*Proof.* If $(1 \otimes a, b \otimes 1) \in B(w)$, then $\sigma_2(I_{a^{-1}}) \cap I_{wb} \neq \emptyset$. Thus $I_{a^{-1}}$ and $I_{wb}$ belong to distinct components of $\hat{I}$. Consequently, if $a, a' \in L(w)$ (resp. $R(w)$) belong to the same connected component of $\mathcal{E}(w)$, then $I_{a^{-1}}, I_{a'^{-1}}$ (resp. $I_{wa}, I_{wa'}$) belong to the same component of $\hat{I}$.

Conversely, let $a, a' \in L(w)$ be such that $a, a'$ belong to the same component of $\hat{I}$. We may assume that $a <_L a'$. There is a reduced path (i.e., it does not use twice consecutively the same edge) in $\mathcal{E}(w)$ from $a$ to $a'$ which is the sequence $a_1, b_1, \ldots, b_{n-1}, a_n$ with $a_1 = a$ and $a_n = a'$ with $a_1 <_L a_2 <_L \cdots <_L a_n$, $wb_1 <_R wb_2 <_R \cdots <_R wb_{n-1}$ and $\sigma_2(I_{a_i^{-1}}) \cap I_{wb_i} \neq \emptyset$, $\sigma_2(I_{a_{i+1}^{-1}}) \cap I_{wb_i} \neq \emptyset$ for $1 \leq i \leq n-1$ (see Figure 8.6 for an illustration).

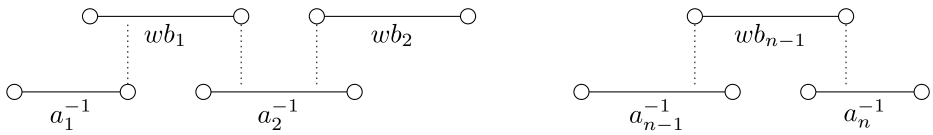


Figure 8.6: A path from $a_1$ to $a_n$ in $\mathcal{E}(w)$.

Note that the hypothesis that $T$ is without connection is needed since otherwise the right boundary of $\sigma_2(I_{a_i^{-1}})$ could be the left boundary of $I_{wb_i}$.

The assertion concerning $b, b' \in R(w)$ is a consequence of the first one since $b, b' \in R(w)$ if and only if $b^{-1}, b'^{-1} \in L(w^{-1})$ (see 8.2.3). ∎

We can now prove the main result of this section.

*Proof.*[ of Theorem 8.2.7] Let $T$ be a linear involution on $I$ without connection and let $S = \mathcal{L}(T)$. Let us first prove that for any $w \in \mathcal{L}(T)$, the graph $\mathcal{E}(w)$ is acyclic. Assume that $(1 \otimes a_1, b_1 \otimes 1, \ldots, 1 \otimes a_n, b_n \otimes 1)$ is a path in $\mathcal{E}(w)$ with $a_1, \ldots, a_n \in L(w)$ and $b_1, \ldots, b_n \in R(w)$. We may assume that the path is reduced, that $n \geq 2$ and also that $a_1 <_L a_2$. It follows that $a_1 <_L \ldots <_L a_n$ and $wb_1 <_R \ldots <_R wb_n$ (see Figure 8.6). Thus it is not possible to have an edge $(a_1, b_n)$, which shows that $\mathcal{E}(w)$ is acyclic.

Let $a, a' \in A$. If $I_{a^{-1}}$ and $I_{a'^{-1}}$ are in the same component of $\hat{I}$, then $1 \otimes a, 1 \otimes a'$ are in the same connected component of $\mathcal{E}(\varepsilon)$. Thus $\mathcal{E}(\varepsilon)$ is a union of two trees with $2\,\mathrm{Card}(A)$ vertices.

If $w \in S$ is nonempty and $1 \otimes a, 1 \otimes a' \in L(w)$, then $I_{a^{-1}}$ and $I_{a'^{-1}}$ are in the same component of $\hat{I}$ (by Lemma 8.2.8), and thus $1 \otimes a, 1 \otimes a'$ are in the same connected component of $\mathcal{E}(w)$. Thus $\mathcal{E}(w)$ is a tree. ■

**Example 8.2.10** Let $T$ be the linear involution of Example 8.1.4. $\mathcal{L}(T)$ is a tree set of characteristic 2 over the alphabet $\{a, b, c, a^{-1}, b^{-1}, c^{-1}\}$. In Figure 8.7 are represented the extension graphs of the empty word (left) and of letters $a$ (center) and $c^{-1}$ (right) (where we note $\bar{a}$ instead of $a^{-1}$).

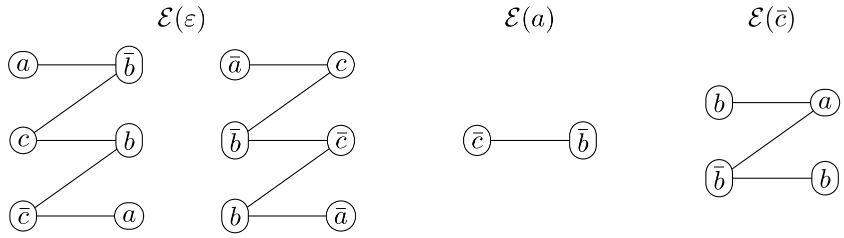


Figure 8.7: Some extension graphs.

We prove the following result.

**Theorem 8.2.11** *The natural coding of a linear involution without connections is a specular set.*

*Proof.* Let $T$ be a linear involution without connections. By Proposition 8.2.3, the set $\mathcal{L}(T)$ is a laminary set.

By 8.2.7, $\mathcal{L}(T)$ 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 $\theta$ has fixed points.

**Example 8.2.12** Let $A = \{a, b, c, d\}$ be as in Example 5.5.5 (in particular, $d = b^{-1}$, $a = a^{-1}$, $c = c^{-1}$).

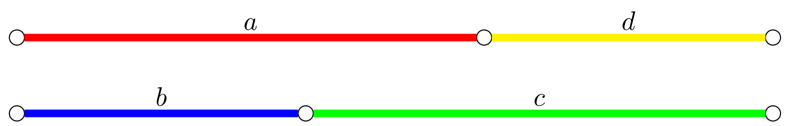


Figure 8.8: A linear involution on $A = \{a, b, c, d\}$.

Let $T$ be the linear involution represented in Figure 8.8 with $\sigma_1$ 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 5.2.22. Indeed, consider the interval exchange $V$ on the interval $Y = ]0, 2[$ represented in Figure 8.9 on the right, which is obtained by using two copies of the interval exchange $U$ defining the Fibonacci set (represented in Figure 8.9 on the left).

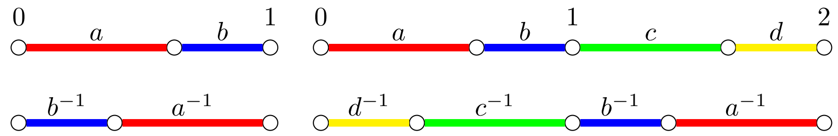


Figure 8.9: Interval exchanges $U$ and $V$ for the Fibonacci set and its doubling.

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 [15]).

### 8.2.4 Mixed return words

In this section we recall the definition of mixed return words given in Section 5.3.

First, note that in the natural coding of a linear involution, every word $w$ does not overlap with its inverse $w^{-1}$. Indeed, in the free group, a reduced word $w$ and its inverse do not overlap.

Recall that the mixed return words to $w$ are the words $N(u)$ associated with complete return words $u$ to $\{w, w^{-1}\}$ obtained erasing $w$ if it appears as a prefix and $w^{-1}$ if it appears as a suffix. The convention choses for the transformation $N$ corresponds to the induction on $I_{w^{-1}} \cup \sigma_2 (I_{w^{-1}})$ (see Lemma 8.2.14 below).

We denote by $\mathcal{MR}(w)$ the set of mixed return words to $w$ in $S$. If $T$ is an orientable linear involution we have, then $\mathcal{MR}(w)$ is the union of the set of right return words to $w$ with the set of left return words to $w^{-1}$.

**Example 8.2.13** Let $T$ be the linear involution of Example 8.1.4. The set of complete return words to the set of a letter and its inverse are:

$$\begin{aligned} \mathcal{CR}_S(\{a, \bar{a}\}) &= \{a\bar{b}cb\bar{a}, a\bar{b}cb\bar{c}a, \bar{a}c\bar{b}\bar{c}a, a\bar{b}\bar{c}b\bar{a}, \bar{a}cb\bar{c}a, \bar{a}c\bar{b}\bar{c}b\bar{a}\} \\ \mathcal{CR}_S(\{b, \bar{b}\}) &= \{b\bar{a}cb, b\bar{a}c\bar{b}, b\bar{c}a\bar{b}, \bar{b}cb, \bar{b}\bar{c}a\bar{b}, \bar{b}\bar{c}b\}, \\ \mathcal{CR}_S(\{c, \bar{c}\}) &= \{cb\bar{a}c, cb\bar{c}, c\bar{b}\bar{c}, \bar{c}a\bar{b}c, \bar{c}a\bar{b}\bar{c}, \bar{c}b\bar{a}c\}. \end{aligned}$$

Thus we have the following sets of mixed return words:

$$\begin{aligned} \mathcal{MR}_S(a) &= \{\bar{b}cb, \bar{b}cb\bar{c}a, \bar{a}c\bar{b}\bar{c}a, \bar{b}\bar{c}b, \bar{a}cb\bar{c}a, \bar{a}c\bar{b}\bar{c}b\} \\ \mathcal{MR}_S(b) &= \{\bar{a}cb, \bar{a}c, \bar{c}a, \bar{b}cb, \bar{b}\bar{c}a, \bar{b}\bar{c}b\}, \\ \mathcal{MR}_S(c) &= \{b\bar{a}c, b, \bar{b}, \bar{c}a\bar{b}c, \bar{c}a\bar{b}, \bar{c}b\bar{a}c\}. \end{aligned}$$

The reason for introducing the notion of mixed return words comes from the fact that, when $S$ is the natural coding of a linear involution, we are interested in the transformation induced on $I_w \cup \sigma_2(I_w)$. The natural coding of a point in $I_w$ begins with $w$ while the natural coding of a point $z$ in $\sigma_2(I_w)$ is preceded by $w^{-1}$ in the sense that the natural coding of $T^{-|w|}(z)$ begins with $w^{-1}$. To be more precise, the convention chosen for the transformation $N$ corresponds to the induction on $I_{w^{-1}} \cup \sigma_2(I_{w^{-1}})$, such as shown with the following lemma. Recall that the notation $\rho_X$ stands for the return time to $X$.

**Lemma 8.2.14** *Let $T$ be a linear involution with no connection and $w$ a nonempty word in its natural coding $\mathcal{L}(T)$. Let $K_w = I_{w^{-1}} \cup \sigma_2(I_{w^{-1}})$. Then the set of mixed first return words to $w$ are exactly the prefixes of length $\rho_{K_w}(z)$ of the infinite natural coding of points $z \in K_w$.*

*Proof.* Let $u$ be the prefix of length $\rho_{K_w}(z)$ of $\Sigma_T(z)$ for some $z \in K_w$. Let us first recall that $\sigma_2(I_{w^{-1}}) = T^{|w|}(I_w)$ (Proposition 8.2.3). Assume first that the length of $u$ is larger than or equal to the length of $w$. If $z \in I_{w^{-1}}$, then $u$ starts with $w^{-1}$ while if $z \in \sigma_2(I_{w^{-1}})$ then $wu$ is in $\mathcal{L}(T)$. Similarly, if $T^{|u|}(z) \in I_{w^{-1}}$ then $uw^{-1}$ is in $\mathcal{L}(T)$ while if $T^{|u|}(z) \in \sigma_2(I_{w^{-1}})$ then $u$ ends with $w$. In all four possible cases, $u$, $wu$, $uw^{-1}$ and $wuw^{-1}$ are in $\mathcal{L}(T)$.

Let

$$p = \begin{cases} \varepsilon & \text{if } z \in I_{w^{-1}}, \\ w & \text{if } z \in \sigma_2(I_{w^{-1}}), \end{cases} \qquad \text{and} \qquad s = \begin{cases} w^{-1} & \text{if } T^{|u|}(z) \in I_{w^{-1}}, \\ \varepsilon & \text{if } T^{|u|}(z) \in \sigma_2(I_{w^{-1}}). \end{cases}$$

Since $I_{w^{-1}}$ and $\sigma_2(I_{w^{-1}})$ are included into two distinct components, there is no cancellation in the product $pus$. Moreover, $|pus| \geq |u|$ and hence $pus$ starts and ends with an occurrence of $w$ or $w^{-1}$. It is thus a complete return word to $\{w, w^{-1}\}$. Furthermore one has $N(pus) = u$.

Let conversely $u$ be a mixed first return word to $w$ and let $u'$ be the complete first return word such that $u = N(u')$. Write $u' = pus$. Assume first that $u' = wu$. Then $wu$ ends with $w$. For any point $y \in I_{u'}$, set $x = T^{|w|}(y)$. Then $x \in T^{|w|}I_w = \sigma_2(I_{w^{-1}})$, $x \in I_u$, and thus $T^{|u|}x \in \sigma_2(I_{w^{-1}})$ and $\rho_{K_w}(x) = |w|$. Hence $u$ is the prefix of length $\rho_{J_w}(x)$ of $\Sigma_T(x)$. The proof in the three other cases is similar. ∎

As a corollary of Theorem 5.3.9 we obtain the following results.

**Corollary 8.2.15** *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.*

A geometric proof and interpretation of the next result is given in [15].

**Corollary 8.2.16** *Let $S$ be the natural coding of a linear involution without connections on the alphabet $A = A_+ \cup A_-$. For any $w \in S$, the set $\mathcal{MR}_S(w)$ is a monoidal basis of $F_B$.*

**Example 8.2.17** Let $T$ be the linear involution of Example 8.1.4. We have seen in Example 8.2.13 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 $\{a, b, c\}$.

**Example 8.2.18** Let $S$ be the specular set of Example 5.2.3. As seen in Example 5.5.5, the group of even words $G$ is a subgroup of index 2 and 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 8.2) is the interval exchange transformation represented in Figure 8.10. Its upper intervals are the $I_x$ for $x \in X$.

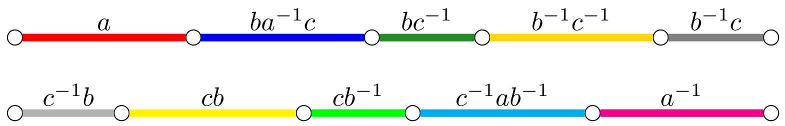


Figure 8.10: The transformation induced on the upper level.

This corresponds to 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 $I \times \{1\}$.

### 8.2.5 Admissible intervals

We have introduced in Chapter 7 the notion of an admissible semi-interval. We give an analogous definition for open intervals in the framework of linear involutions.

Let $T$ be a linear involution without connection defined on the interval $I$. The open interval $J = ]u, v[$ with $J \subset I$ is *admissible* with respect to $T$ if for each of its two endpoints $x = u, v$, there is

(i) either a singularity $z$ of $T^{-1}$ such that $x = T^n(z)$ and $T^k(z) \notin J$ for $0 \le k \le n$,

(ii) or a singularity $z$ of $T$ such that $z = T^n(x)$ and $T^k(x) \notin J$ for $0 \le k \le n$.

For any admissible interval of $I$ with respect to $T$, the transformation induced on $I$ is a $k$-linear involution without connection (see cite[Lemma 4.4]linearinvolutions).

The following is the counterpart for linear involutions of Theorem 7.2.3. The proof for linear involutions is the same. Recall that the intervals $I_w$, $w \in S$, are defined in Equation 8.2.

**Proposition 8.2.19** *Let $T$ be a linear involution without connection on $I$. The interval $I_w$, seen as a subinterval of $I$, is admissible with respect to $T$.*

*Proof.* Let $T$ be a $k$-linear involution. Recall that $\Sigma$ is the set of $2k-2$ division points separating the intervals $I_a$ for $a \in A$.

Let $n \geq 1$. Since $T$ is without connections, $T^{-i}(z)$ is well defined for any $z \in \Sigma$ and for any $i$ such that $0 \leq i \leq n-1$. Let $P_n = \{T^{-i}(z) \mid z \in \Sigma \mid z \in \Sigma, 0 \leq i \leq n-1\} \cup \{(\{\lambda\} \times \{0,1\}), $ where $\lambda$ stands for the left endpoint of the interval $I$. One has $\mathrm{Card}(P_n) = (2k-2)n + 2$. Consider two points $z$ and $z'$ in $\hat{I} \setminus \hat{\mathcal{O}}$ that belong to two different intervals of the partition by open intervals of $I \times \{0,1\}$ made by the points of $P_n$. Then the prefixes of size $n$ of theri respective infinite natural codings differ. On the other hand, the left boundary of each $I_w$, $|w| = n$, is the left boundary of some $T^{-i}(I_a)$ for some $0 \leq i \leq n-1$ and some $a \in A$. This proves that $P_n$ is the set of $2(k-1)n+2$ left boundaries of the intervals $I_w$ for all words $w$ with $|w| = n$, and that the family $(I_w)_{|w|=n}$ forms a partition of $I \times \{0,1\}$ (up to the points of $P_n$).

Let $I_w = ]u, v[$ nad $w = a_0 a_1 \cdots a_{n-1}$. We assume that $u \neq \lambda$. By construction, there exist a point $z \in \Sigma$ and an integer $i$ with $0 \leq i \leq n-1$ such that $u = T^{-i}(z)$, where $I_{a_i} = ]z, t[$ for some $t$ in $I$ or equal to the right boundary of $I$. For any $k$ with $0 \leq k \leq n-1$, the point $T^{-k}(z)$ is the left boundary of some interval $I_y$, with $|y| = n$. Thus, in particular, on gets $T^k(u) \notin I_w$, for $0 \leq k \leq i$.

The same reasoning applies to the right boundary $v$ of $I_w$. ■

# Conclusions

## Where do we come from? What are we? Where are we going?

In this manuscript are contained results from different papers signed by me and several other people: Jean Berstel, Valérie Berthé, Clelia De Felice, Vincent Delecroix, Julien Leroy, Dominique Perrin, Christophe Reutenauer, Giuseppina Rindone.

The story of this thesis starts even before the beginning of my PhD. Back in 2012, I was a master student at the Università degli Studi di Palermo. My master thesis supervisor, Antonio Restivo, to "keep me busy" let me study a paper ([7], a preprint at the time) of more than fifty pages in which the five authors (Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer and Giuseppina Rindone) discussed bifix codes, episturmian words and subgroups of the free group. Most of the concept in this paper were completely new to me. Nevertheless, I became interested in the subject and started to read related papers. In my master thesis I also manage to give a (minimal) contribution to the theory with a counterexample disproving a conjecture about a converse of the Cardinality Theorem for bifix codes in Sturmian (Arnoux-Rauzy) sets, namely that there exist an infinite word $x$ and a family $(X_d)_{d>0}$ of maximal bifix codes satisfying the formula $\mathrm{Card}(X_d \cap F(x)) = d(X_d) + 1$ with $x$ not a Sturmian word.

After my master thesis in Palermo I arrived in Marne-la-Vallée, first for a year as a winner of the scholarship "International Master in Mathematics and Computer Science - the Bezout Excellence Track", then as a "natural" continuation of my master internship, as a PhD student with Dominique Perrin as my supervisor. This opportunity allowed me to join the working group continuing the work started in [7] (in the same period Valérie Berthé and Julien Leroy joined the group as well, while Jean Berstel retired).

"Bifix codes and Sturmian words" ([7]) was a seminal paper that already contained several results extended afterwards to larger classes than Arnoux-Rauzy sets: the Cardinality Theorem for bifix codes, the Finite Index Basis Theorem, the Return Theorem, etc.

One of the open question arised from this paper concerned the closure of an Arnoux-Rauzy set under maximal bifix decoding. Indeed, the maximal bifix decoding of the set of factors of a Sturmian (or strict episturmian) word is not an Arnoux-Rauzy set anymore. But an Arnoux-Rauzy set on a binary alphabet is also an interval exchange set. This led to the results of "Bifix codes and interval exchanges"([12]) where we show that the class of natural codings of regular interval exchange transformations is the natural closure under maximal bifix decoding of the class of Sturmian (binary Arnoux-Rauzy) sets.

And what about Arnoux-Rauzy sets over an arbitrary alphabet? This question led us to define a common generalization of Arnoux-Rauzy sets and regular interval exchange sets: the tree sets. The first definition of these sets as well as some generalization of [7] are provided in "Acyclic, connected and tree sets"([11]). The study of the maximal bifix decoding of a tree sets is treated in "Maximal bifix decoding" ([14]), while the paper "The finite index basis property" ([13]) shows that uniformly recurrent tree sets satisfy the finite index basis property.

In the wake of the study of tree sets, we tried to generalize some of the results to a larger class of sets: neutral sets. In the conference paper "Enumeration formulæ in neutral sets" ([33]) and its longer version "Neutral and tree sets of arbitrary characteristic" ([34]), we managed to generalize several results as well as the definition itself of neutral and tree sets, which led to the definition of characteristic of a neutral set. The results and the new tools introduced in these two papers allowed us to simplify several proofs from the previous papers. As a consequence, in this manuscript one can find shorter versions of the main results of [11, 12, 13, 14]. Further simplifications in the statement and proofs of several results are obtained using the surprising and unexpected fact, proved in [34], that in a neutral set (and thus, in particular, in a tree set) the notions of recurrence and uniformly recurrence coincide. This allows us also to answer an open question of [11] and [14]: we now know that the maximal bifix decoding of a recurrent tree set preserve both the recurrence and the tree property.

With the same group of authors of [11, 12, 13, 14] with the addition of Vincent Delecroix, we studied a family of dynamical systems closed to interval exchanges: linear involutions. We soon realized that the natural coding of a linear involution without connections satisfies the tree condition. Moreover, other peculiar symmetric properties are satisfied by this family of sets. This led us to the definition of specular sets and specular groups and to the publication of two papers: "Return words of linear involutions and fundamental groups" ([15]) with a more topological and geometric flavor, devoted to the specific case of linear involutions, and "Specular sets" ([9], a long version of the conference paper [10]), whose results are developed in a wider combinatorial context.

In parallel with the previous works, we focused on interval exchanges, generalizing some results of Rauzy and Boshernitzan concerning the Rauzy induction and the case of interval exchanges defined over a quadratic field. From this study comes the paper written by me and Dominique Perrin "Interval exchanges, admissibility and branching Rauzy induction" ([32], a longer version of the conference contribution "A note on regular interval exchange sets over a quadratic field").

In summary, this manuscript could be seen as a revised and unifying versions of the material contained in [9, 11, 12, 13, 14, 15, 32, 34]. However, note that there is not a one-to-one correspondence between this thesis and the series of papers. Indeed, several results published (or submitted) before are here refined or presented with a shorter and simplified proof.

On the other hand, some notions of the papers are not treated here. Firstly, in order to give consistency to the manuscript I tried, as far as possible, to put all the results in a uniform framework. As a second reason, I chose to include in this thesis the results to which my contribution has been particularly substantial. This is why, for instance, we do not talk in this manuscript about the Rauzy fractal (as done in [12]) or foliations and surfaces (as done in [15]). On the contrary, my own contribution ot the series if papers presented here is especially strong in what concerns the maximal bifix decoding results, the theory of specular sets as well as the branching Rauzy induction and the study of interval exchanges over a quadratic field. Therefore I decided to develop here these topics as much in detail as possible.

And what's next? Some of the notions we introduced have been used by other authors. This is the example of the branching induction and the admissibility we defined in this thesis for interval exchanges, used by Fickenscher in [40], or the use of the Return Theorem in profinite semigroups in [1].

The main topic of this thesis, the tree condition and the study of the extension grahs, also seems to be a promising topic in different fields: words and palindromes ([49]), Schützenberger groups ([1]), $\mathcal{S}$-adic representations ([50]). Recently, Julien Leroy and Revekka Kyriakoglou and myself submitted a paper, “Decidable properties of extension graphs for substitutive languages” ([31]), where we study the case of minimal dynamical systems arising from a substitutive language and we show that the tree properties is decidable.

It is very likely that in the future we, as well as other people, will continue the study of these properties and these families of sets.

Paris,<br>June 2016

# Bibliography


[1] Jorge Almeida and Alfredo Costa. A geometric interpretation of the Schützenberger group of a minimal subshift. *Arkiv för Matematik*, pages 1–33, 2016. (http://dx.doi.org/10.1007/s11512-016-0233-7).

[2] V. I. Arnol′d. Small denominators and problems of stability of motion in classical and celestial mechanics. *Uspehi Mat. Nauk*, 18(6 (114)):91–192, 1963.

[3] Pierre Arnoux and Gérard Rauzy. Représentation géométrique de suites de complexité $2n + 1$. *Bull. Soc. Math. France*, 119(2):199–215, 1991.

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

[5] L'ubomíra Balková, Edita Pelantová, and Wolfgang Steiner. Sequences with constant number of return words. *Monatsh. Math.*, 155(3-4):251–263, 2008. (http://dx.doi.org/10.1007/s00605-008-0001-2).

[6] Laurent Bartholdi. Growth of groups and wreath products. 2014. Preprint.

[7] Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012. (http://dx.doi.org/10.1016/j.jalgebra.2012.07.013).

[8] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and automata*, volume 129 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010.

[9] Valérie Berthé, Clelia De Felice, Vincent Delecroix, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Specular sets. submitted.

[10] Valérie Berthé, Clelia De Felice, Vincent Delecroix, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Specular sets. In *Combinatorics on Words*, Spinger LNCS. 2015.

[11] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Acyclic, connected and tree sets. *Monatsh. Math.*, 176(4):521–550, 2015. (http://dx.doi.org/10.1007/s00605-014-0721-4).

[12] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and interval exchanges. *J. Pure Appl. Algebra*, 219(7):2781–2798, 2015. (http://dx.doi.org/10.1016/j.jpaa.2014.09.028).

[13] 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.

[14] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. *Discrete Math.*, 338(5):725–742, 2015. (http://dx.doi.org/10.1016/j.disc.2014.12.010).

[15] 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.

[16] Valérie Berthé and Michel Rigo. Preliminaries. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 1–33. Cambridge Univ. Press, Cambridge, 2010.

[17] A. Blondin Massé, S. Brlek, S. Labbé, and L. Vuillon. Palindromic complexity of codings of rotations. *Theoret. Comput. Sci.*, 412(46):6455–6463, 2011. (http://dx.doi.org/10.1016/j.tcs.2011.08.007).

[18] 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.

[19] M. D. Boshernitzan and C. R. Carroll. An extension of Lagrange's theorem to interval exchange transformations over quadratic fields. *J. Anal. Math.*, 72:21–44, 1997. (http://dx.doi.org/10.1007/BF02843152).

[20] Michael D. Boshernitzan. Rank two interval exchange transformations. *Ergodic Theory Dynam. Systems*, 8(3):379–394, 1988. (http://dx.doi.org/10.1017/S0143385700004521).

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

[22] Julien Cassaigne. Personal communication. 2013.

[23] Julien Cassaigne, Sébastien Ferenczi, and Ali Messaoudi. Weak mixing and eigenvalues for Arnoux-Rauzy sequences. *Ann. Inst. Fourier (Grenoble)*, 58(6):1983–2005, 2008.

[24] P. M. Cohn. *Free rings and their relations*, volume 19 of *London Mathematical Society Monographs*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London, second edition, 1985.

[25] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinaĭ. *Ergodic theory*, volume 245 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1982. Translated from the Russian by A. B. Sosinskiĭ.

[26] 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.

[27] Ethan M. Coven and G. A. Hedlund. Sequences with minimal block growth. *Math. Systems Theory*, 7:138–153, 1973.

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

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

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

[31] Francesco Dolce, Revekka Kyriakoglou, and Julien Leroy. Decidable properties of extension graphs for substitutive languages. 2016. submitted.

[32] Francesco Dolce and Dominique Perrin. Interval exchanges, admissibility and branching Rauzy induction. submitted.

[33] Francesco Dolce and Dominique Perrin. Enumeration formulæ in neutral sets. In *DLT 2015*, Spinger LNCS. 2015.

[34] Francesco Dolce and Dominique Perrin. Neutral and tree sets of arbitrary characteristic. *Theoret. Comput. Sci.*, 2016.

[35] 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.

[36] Fabien Durand. A characterization of substitutive sequences using return words. *Discrete Math.*, 179(1-3):89–101, 1998. (http://dx.doi.org/10.1016/S0012-365X(97)00029-0).

[37] Fabien Durand, Julien Leroy, and Gwenaël Richomme. Do the properties of an $S$-adic representation determine factor complexity? *J. Integer Seq.*, 16(2):Article 13.2.6, 30, 2013.

[38] Sébastien Ferenczi. Rank and symbolic complexity. *Ergodic Theory Dynam. Systems*, 16(4):663–682, 1996. (http://dx.doi.org/10.1017/S0143385700009032).

[39] Sébastien Ferenczi and Luca Q. Zamboni. Languages of $k$-interval exchange transformations. *Bull. Lond. Math. Soc.*, 40(4):705–714, 2008. (http://dx.doi.org/10.1112/blms/bdn051).

[40] Jon Fickenscher. Decoding Rauzy induction: An answer to bufetov's general question. 2016. (http://arxiv.org/abs/1507.07481).

[41] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[42] Amy Glen and Jacques Justin. Episturmian words: a survey. *Theor. Inform. Appl.*, 43(3):403–442, 2009. (http://dx.doi.org/10.1051/ita/2009003).

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

[44] Yann Jullian. An algorithm to identify automorphisms which arise from self-induced interval exchange transformations. *Math. Z.*, 274(1-2):33–55, 2013. (http://dx.doi.org/10.1007/s00209-012-1056-2).

[45] Jacques Justin and Laurent Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5):343–356, 2000. (http://dx.doi.org/10.1051/ita:2000121).

[46] Ilya Kapovich and Alexei Myasnikov. Stallings foldings and subgroups of free groups. *J. Algebra*, 248(2):608–668, 2002. (http://dx.doi.org/10.1006/jabr.2001.9033).

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

[48] P. Kůrka. *Topological and symbolic dynamics*, volume 11 of *Cours Spécialisés [Specialized Courses]*. Société Mathématique de France, Paris, 2003.

[49] Sébastian Labbé, Edita Pelantová, and Štěpán Starosta. On the zero defect conjecture. 2016. http://arxiv.org/abs/1606.05525.

[50] Julien Leroy. An s-adic characterization of ternary tree sets. 2014. In preparation.

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

[52] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002. A collective work by Jean Berstel, Dominique Perrin, Patrice Seebold, Julien Cassaigne, Aldo De Luca, Steffano Varricchio, Alain Lascoux, Bernard Leclerc, Jean-Yves Thibon, Veronique Bruyere, Christiane Frougny, Filippo Mignosi, Antonio Restivo, Christophe Reutenauer, Dominique Foata, Guo-Niu Han, Jacques Desarmenien, Volker Diekert, Tero Harju, Juhani Karhumaki and Wojciech Plandowski, With a preface by Berstel and Perrin.

[53] Roger C. Lyndon and Paul E. Schupp. *Combinatorial group theory*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1977 edition.

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

[55] Tomasz Miernowski and Arnaldo Nogueira. Exactness of the Euclidean algorithm and of the Rauzy induction on the space of interval exchange transformations. *Ergodic Theory Dynam. Systems*, 33(1):221–246, 2013. (http://dx.doi.org/10.1017/S014338571100085X).

[56] Marston Morse and Gustav A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.*, 62:1–42, 1940.

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

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

[59] V. I. Oseledec. The spectrum of ergodic automorphisms. *Dokl. Akad. Nauk SSSR*, 168:1009–1011, 1966.

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

[61] Gérard Rauzy. Échanges d'intervalles et transformations induites. *Acta Arith.*, 34(4):315–328, 1979.

[62] C. Reutenauer. Sur mon article: "Une topologie du monoïde libre" [Semigroup Forum **18** (1979), no. 1, 33–49; MR 80j:20075]. *Semigroup Forum*, 22(1):93–95, 1981. (http://dx.doi.org/10.1007/BF02572788).

[63] Alexandra Skripchenko. Symmetric interval identification systems of order three. *Discrete Contin. Dyn. Syst.*, 32(2):643–656, 2012. (http://dx.doi.org/10.3934/dcds.2012.32.643).

[64] Bo Tan, Zhi-Xiong Wen, and Yiping Zhang. The structure of invertible substitutions on a three-letter alphabet. *Adv. in Appl. Math.*, 32(4):736–753, 2004. (http://dx.doi.org/10.1016/S0196-8858(03)00102-7).

[65] William A. Veech. Finite group extensions of irrational rotations. *Israel J. Math.*, 21(2-3):240–259, 1975. Conference on Ergodic Theory and Topological Dynamics (Kibbutz Lavi, 1974).

[66] Marcelo Viana. Ergodic theory of interval exchange maps. *Rev. Mat. Complut.*, 19(1):7–100, 2006.

[67] Laurent Vuillon. On the number of return words in infinite words constructed by interval exchange transformations. *Pure Math. Appl. (PU.M.A.)*, 18(3-4):345–355, 2007.

[68] Jean-Christophe Yoccoz. Interval exchange maps and translation surfaces. In *Homogeneous flows, moduli spaces and arithmetic*, volume 10 of *Clay Math. Proc.*, pages 1–69. Amer. Math. Soc., Providence, RI, 2010.

[69] Anton Zorich. Deviation for interval exchange transformations. *Ergodic Theory Dynam. Systems*, 17(6):1477–1499, 1997. (http://dx.doi.org/10.1017/S0143385797086215).