Maximal bifix decoding

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

[1]CNRS, Université Paris 7, [2]Università degli Studi di Salerno, [3]Université Paris Est, LIGM, [4]Université du Luxembourg, [5]Université du Québec à Montréal

March 3, 2014 7 h 54

Abstract

We introduce a class of sets of words which is a natural common generalization of Sturmian sets and of interval exchange sets. This class of sets consists of the uniformly recurrent tree sets, where the tree sets are defined by a condition on the possible extensions of bispecial factors. We prove that this class is closed under maximal bifix decoding. The proof uses the fact that the class is also closed under decoding with respect to return words.

Contents

1 Introduction

This paper studies the properties of a common generalization of Sturmian sets and regular interval exchange sets. We first give some elements on the background of these two families of sets.

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. Their first in-depth study is by Morse and Hedlund [25]. Many combinatorial properties were described in the paper by Coven and Hedlund [12].

We understand here by Sturmian words the generalization to arbitrary alphabets, often called strict episturmian words or Arnoux-Rauzy words (see the survey [20]), of the classical Sturmian words on two letters. Sturmian sets are the sets of factors of Sturmian words. For more details, see [19, 24].

Sturmian words are closely related to the free group. This connection is one of the main points of the series of papers [2, 4, 5] and the present one. A striking feature of this connection is the fact that our results do not hold only for two-letter alphabets or for two generators but for any number of letters and generators.

In a paper with part of the present list of authors on bifix codes and Sturmian words [2] we proved that Sturmian sets satisfy the finite index basis property, in the sense that, given a set F of words on an alphabet A, a finite bifix code is F-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 [26] following an earlier idea of Arnold [1]. These transformations form a generalization of rotations of the circle. The class of regular interval exchange transformations was introduced by Keane [22] who showed that they are minimal in the sense of topological dynamics. The factors of natural codings of regular interval exchange transformations are called interval exchange sets. In [5], we show that regular interval exchange sets satisfy the finite index basis property.

Even if they have the same factor complexity (that is, the same number of factors of a given length), Sturmian 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 [10, 28]).

The class of tree sets, introduced in [4] contains both the Sturmian sets and the regular interval exchange sets. They are defined by a condition on the possible extensions of bispecial factors. One of the results of [4] is that, in a uniformly recurrent tree set F, the set of first return words to a given word in F is a basis of the free group on the alphabet of F. The main statement of [5] is that uniformly recurrent tree sets satisfy the finite index basis property. This generalizes the result concerning Sturmian words of [2] quoted above. As an example of a consequence of this result, if F is a uniformly recurrent tree set on the alphabet A, then for any $n \geq 1$, the set $F \cap A^n$ is a basis of the subgroup formed by the words of length multiple of n (see Theorem 5.7).

Our main result here is that the class of uniformly recurrent tree sets is closed under maximal bifix decoding (Theorem 7.1). This means that if F is a uniformly recurrent tree set and f a coding morphism for a finite F-maximal bifix code, then $f^{-1}(F)$ is a uniformly recurrent tree set. The family of regular interval exchange sets is closed under maximal bifix decoding (see [5] Corollary 5.22) but the family of Sturmian sets is not (see Example 7.2 below). Thus, this result shows that the family of uniformly recurrent tree sets is the natural closure of the family of Sturmian sets. The proof uses the finite index basis property of uniformly recurrent tree sets.

The proof of Theorem 7.1 uses the closure of uniformly recurrent tree sets under decoding with respect to return words (Theorem 5.10). This property, which is interesting in its own, generalizes the fact that the derived word of a Sturmian word is Sturmian [21].

We also focus on tree sets defined on a ternary alphabet. In this case, uniformly recurrent tree sets are uniquely ergodic (which means that they have a unique invariant probability measure). We give a characterization of the S-adic representation of ternary tree sets (Theorem 6.6) in terms of infinite paths in a Büchi automaton, where S is the set of positive elementary automorphisms of the free group on three letters..

The paper is organized as follows. In Section 2, we introduce the notation and recall some basic results. We define the composition of prefix codes.

In Section 3, we introduce one important subclass of tree sets, namely interval exchange sets. We recall the definitions concerning minimal and regular interval exchange transformations. We state the result of Keane expressing that regular interval exchange transformations are minimal (Theorem 3.4).

In Section 4, we define return words, derived words and derived sets and prove some elementary properties.

In Section 5, we recall the definition of tree sets. We also recall that a regular interval exchange set is a tree set (Proposition 5.3). We prove that the family of uniformly recurrent tree sets is invariant under derivation (Theorem 5.10). We further prove that all bases of the free group included in a uniformly recurrent tree set are tame, that is obtained from the alphabet by composition of elementary positive automorphisms (Theorem 5.16).

In Section 6, We deduce from the previous result that uniformly recurrent tree sets have a primitive S_e-adic representation (Theorem 6.5) where S_e is the finite set of positive elementary automorphisms of the free group. We give a more precise result in the case of a ternary alphabet. It characterizes tree sets by their S-adic representation (Theorem 6.6).

In Section 7, we state and prove our main result (Theorem 7.1), namely the closure under maximal bifix decoding of the family of uniformly recurrent tree sets.

Finally, in Section 7.3, we use Theorem 7.1 to prove a result concerning the composition of bifix codes (Theorem 7.12) showing that the degrees of the terms of a composition are multiplicative.

2 Preliminaries

In this section, we recall some notions and definitions concerning words, codes and automata. For a more detailed presentation, see [2]. We also introduce the notion of composition of codes.

2.1 Words

Let A be a finite nonempty alphabet. All words considered below, unless stated explicitly, are supposed to be on the alphabet A. We denote by A^* the set of all words on A. The empty word is denoted by 1 or by ε . We denote by $|w|$ the length of a word w. For a set X of words and a word x, we denote

$$x^{-1}X = \{y \in A^* \mid xy \in X\}, \quad Xx^{-1} = \{z \in A^* \mid zx \in X\}.$$

A set of words is said to be *factorial* if it contains the factors of its elements. Let F be a set of words on the alphabet A. For $w \in F$, we denote

$$\begin{aligned} L(w) &= \{a \in A \mid aw \in F\} \\ R(w) &= \{a \in A \mid wa \in F\} \\ E(w) &= \{(a,b) \in A \times A \mid awb \in F\} \end{aligned}$$

and further

$$\ell(w) = \mathrm{Card}(L(w)), \quad r(w) = \mathrm{Card}(R(w)), \quad e(w) = \mathrm{Card}(E(w)).$$

A word w is *right-extendable* if $r(w) > 0$, *left-extendable* if $\ell(w) > 0$ and *biextendable* if $e(w) > 0$. A factorial set F is called *right-essential* (resp. *left-essential*, resp. *biessential*) if every word in F 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 $\ell(w) \geq 2$. It is called *bispecial* if it is both right and left-special.

We denote by $\mathrm{Fac}(x)$ the set of factors of an infinite word $x \in A^{\mathbb{N}}$. The set $\mathrm{Fac}(x)$ is factorial and right essential. An infinite word $x \in A^\omega$ is *recurrent* if for any $u \in \mathrm{Fac}(x)$ there is a $v \in \mathrm{Fac}(x)$ such that $uvu \in \mathrm{Fac}(x)$.

A factorial set of words $F \neq \{1\}$ is *recurrent* if for every $u, w \in F$ there is a word $v \in F$ such that $uvw \in F$. For any recurrent set F there is an infinite word x such that $\mathrm{Fac}(x) = F$.

For any infinite word x, the set $\mathrm{Fac}(x)$ is recurrent if and only if x is recurrent (see [2]).

Note that any recurrent set not reduced to the empty word is biessential.

A set of words F is said to be *uniformly recurrent* if it is right-essential and if, for any word $u \in F$, there exists an integer $n \geq 1$ such that u is a factor of every word of F of length n. A uniformly recurrent set is recurrent.

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 $f^\omega(a)$ which has all words $f^n(a)$ as prefixes. It is called a *fixpoint* 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 fixpoint of f is uniformly recurrent (see [19] Proposition 1.2.3 for example).

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. 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 $r(u) = \mathrm{Card}(A)$.

A *Sturmian set* is a set of words which is the set of factors of a strict episturmian word. Any Sturmian set is uniformly recurrent (see [2], Proposition 2.3.3).

Example 2.1 Let $A = \{a, b\}$. The Fibonacci word is the fixpoint $x = abaababa\ldots$ of the morphism $f : A^* \to A^*$ defined by $f(a) = ab$ and $f(b) = a$. It is a Sturmian word (see [24]). The set $\mathrm{Fac}(x)$ of factors of x is the *Fibonacci set*.

Example 2.2 Let $A = \{a, b, c\}$. The Tribonacci word is the fixpoint $x = f^\omega(a) = abacaba\cdots$ 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 [21]). The set $\mathrm{Fac}(x)$ of factors of x is the *Tribonacci set*.

2.2 Bifix codes

Recall that a set $X \subset A^+$ of nonempty words over an alphabet A is 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 $i = 1, \ldots, n$. For the general theory of codes, see [3].

A *prefix code* is a set of nonempty words which does not contain any proper prefix of its elements. 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.

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

Let F be a set of words. A prefix code $X \subset F$ is F-maximal if it is not properly contained in any prefix code $Y \subset F$[1]. Equivalently, a prefix code $X \subset F$ is F-maximal if any word in F is comparable for the prefix order with some word of X.

A set of words M is called *right unitary* if $u, uv \in M$ imply $v \in M$. The submonoid M generated by a prefix code is right unitary. One can show that conversely, any right unitary submonoid of A^* is generated by a prefix code (see [3]). The symmetric notion of a *left unitary* set is defined by the condition $v, uv \in M$ implies $u \in M$.

We denote by X^* the submonoid generated by X. A set $X \subset F$ is *right F-complete* if any word of F is a prefix of a word in X^*. If F is factorial, a prefix code is F-maximal if and only if it is right F-complete (Proposition 3.3.2 in [2]).

Similarly a bifix code $X \subset F$ is F-maximal if it is not properly contained in a bifix code $Y \subset F$. For a recurrent set F, a finite bifix code is F-maximal as a bifix code if and only if it is an F-maximal prefix code (see [2], Theorem 4.2.2). For a uniformly recurrent set F, any finite bifix code $X \subset F$ is contained in a finite F-maximal bifix code (Theorem 4.4.3 in [2]).

A *parse* of a word w with respect to a set 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 $\delta_X(w)$ the number of parses of w.

Let X be a bifix code. The number of parses of a word w is also equal to the number of suffixes of w which have no prefix in X and the number of prefixes of w which have no suffix in X (see Proposition 6.1.6 in [3]).

By definition, the F-degree of a bifix code X, denoted $d_F(X)$, is the maximal number of parses of a word in F. It can be finite or infinite.

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 $uwv \in X$.

Let F be a recurrent set and let X be a finite F-maximal bifix code of F-degree d. A word $w \in F$ is such that $\delta_X(w) < d$ if and only if it is an internal factor of X, that is

$$I(X) = \{w \in F \mid \delta_X(w) < d\} \tag{2.1}$$

(Theorem 4.2.8 in [2]). Thus any word of X of maximal length has d parses. This implies that the F-degree d is finite.

Example 2.3 Let F be a recurrent set. For any integer $n \geq 1$, the set $F \cap A^n$ is an F-maximal bifix code of F-degree n.

The *kernel* of a set of words X is the set of words in X which are internal factors of words in X. We denote by $K(X)$ the kernel of X. Note that $K(X) = I(X) \cap X$.

[1] Note that in this paper we use $\subset$ to denote the inclusion allowing equality.

For any recurrent set F, a finite F-maximal bifix code is determined by its F-degree and its kernel (see [2], Theorem 4.3.11).

Example 2.4 Let F be a recurrent set containing the alphabet A. The only F-maximal bifix code of F-degree 1 is the alphabet A. This is clear since A is the unique F-maximal bifix code of F-degree 1 with empty kernel.

2.3 Group codes

We denote $\mathcal{A} = (Q, i, T)$ a deterministic automaton with Q as set of states, $i \in Q$ as initial state and $T \subset Q$ as set of terminal states. 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 set *recognized* by the automaton is the set of words $w \in A^*$ such that $i \cdot w \in T$. A set of words is *rational* if it is recognized by a finite automaton. Two automata are *equivalent* if they recognize the same set.

All automata considered in this paper 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 i 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.

An automaton $\mathcal{A} = (Q, i, T)$ 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 sets $u^{-1}L = \{v \in A^* \mid uv \in L\}$ for $u \in A^*$. 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 and the terminal states are the sets $u^{-1}L$ for $u \in L$.

Let $X \subset A^*$ be a prefix code. Then there is a simple automaton $\mathcal{A} = (Q, 1, 1)$ that recognizes X^*. Moreover, the minimal automaton of X^* is simple.

Example 2.5 The automaton $\mathcal{A} = (Q, 1, 1)$ represented in Figure 2.1 is the minimal automaton of X^* with $X = \{aa, ab, ac, ba, ca\}$. We have $Q = \{1, 2, 3\}$,

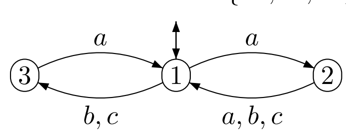

Figure 2.1: The minimal automaton of $\{aa, ab, ac, ba, ca\}^*$.

$i = 1$ and $T = 1$. The initial state is indicated by an incoming arrow and the terminal one by an outgoing arrow.

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 $\mathcal{A} = (P, 1, 1)$ with transitions defined for $p \in P$ and $a \in A$ by

$$p \cdot a = \begin{cases} pa & \text{if } pa \in P\,, \\ 1 & \text{if } pa \in X\,, \\ \emptyset & \text{otherwise.} \end{cases}$$

One verifies that this automaton recognizes X^*.

An 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.

The following result is proved in [2] (Proposition 6.1.5).

Proposition 2.6 *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 φ 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 2.6 is called a *group code* of degree d.

2.4 Composition

We introduce the notion of composition of codes (see [3] 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 f from B^* into A^*. If f is such a morphism, then Y and Z are called composable *through* f. The set

$$X = f(Y) \subset Z^* \subset A^* \tag{2.2}$$

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 2.7 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 (Proposition 2.6.1 of [3]) 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

$$\mathrm{alph}_Z(X) = \{z \in Z \mid \exists\ u, v \in Z^*,\ uzv \in X\}.$$

The following is Proposition 2.6.6 in [3].

Proposition 2.8 *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 Propositions 2.6.4 and 2.6.12 of [3] for prefix codes.

Proposition 2.9 *Let Y, Z be finite prefix codes composable through f and let $X = Y \circ_f Z$.*

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

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

Proof. (i) Let $w \in f(G)$ and set $w = f(v)$ with $v \in G$. Since Y is G-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 $f(G)$-maximal.
(ii) Since X is an F-maximal prefix code, any word in F is prefix comparable with some element of X and thus with some element of Z. Therefore, Z is F-maximal. Next if $u \in f^{-1}(F)$, $v = f(u)$ is in F 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 F is recurrent. Let w be a word in F of length strictly larger than the sum of the maximal length of the words of X and Z. Since F is recurrent, the set Z is right F-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 $u \in F$, we have $v \in f^{-1}(F)$. 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 F-maximal. ∎

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

Note also that when F is recurrent (or even uniformly recurrent), $G = f^{-1}(F)$ need not be recurrent. Indeed, let F 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 $G = u^* \cup v^*$ which is not recurrent.

3 Interval exchange sets

In this section, we recall the definition and the basic properties of interval exchange transformations.

3.1 Interval exchange transformations

Let us recall the definition of an interval exchange transformation (see [11] or [7]).

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 Δ, Γ, we denote $\Delta < \Gamma$ if $x < y$ for any $x \in \Delta$ and $y \in \Gamma$.

Let $(A, <)$ be an ordered set. A partition $(I_a)_{a\in A}$ of $[0, 1[$ in semi-intervals is *ordered* if $a < b$ implies $I_a < I_b$.

Let A be a finite set ordered by two total orders $<_1$ and $<_2$. Let $(I_a)_{a\in A}$ be a partition of $[0, 1[$ in semi-intervals ordered for $<_1$. Let λ_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 $T : [0, 1[\to [0, 1[$ 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 μ_a is the right boundary of I_a and that ν_a is the right boundary of J_a. We additionally denote by γ_a the left boundary of I_a and by δ_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 $[0, 1[$ ordered for $<_2$. In particular, the transformation T defines a bijection from $[0, 1[$ 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 3.1 Let R be the interval exchange transformation corresponding to $A = \{a, b\}$, $a <_1 b$, $b <_2 a$, $I_a = [0, 1-\alpha[$, $I_b = [1-\alpha, 1[$. The transformation R is the rotation of angle α 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 π of A such that $a <_1 b$ if and only if $\pi(a) <_2 \pi(b)$. Conversely, $<_2$ is determined by $<_1$ and π and $<_1$ is determined by $<_2$ and π. The permutation π is said to be *associated* to T.

If we set $A = \{a_1, a_2, \ldots, a_s\}$ with $a_1 <_1 a_2 <_1 \cdots <_1 a_s$, the pair (λ, π) formed by the family $\lambda = (\lambda_a)_{a \in A}$ and the permutation π 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 composition and by taking inverses.

Example 3.2 A 3-interval exchange transformation is represented in Figure 3.1. One has $A = \{a, b, c\}$ with $a <_1 b <_1 c$ and $b <_2 c <_2 a$. The associated permutation is the cycle $\pi = (abc)$.

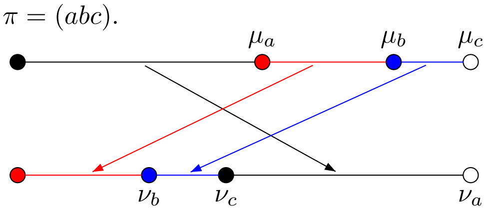

Figure 3.1: A 3-interval exchange transformation

3.2 Regular interval exchange transformations

The *orbit* of a point $z \in [0, 1[$ is the set $\{T^n(z) \mid n \in \mathbb{Z}\}$. The transformation T is said to be *minimal* if for any $z \in [0, 1[$, the orbit of z is dense in $[0, 1[$.

Set $A = \{a_1, a_2, \ldots, a_s\}$ with $a_1 <_1 a_2 <_1 \ldots <_1 a_s$, $\mu_i = \mu_{a_i}$ and $\delta_i = \delta_{a_i}$. The points $0, \mu_1, \ldots, \mu_{s-1}$ form the set of *separation points* of T, denoted $\mathrm{Sep}(T)$.

An interval exchange transformation $T_{\lambda,\pi}$ is called *regular* if the orbits of the nonzero separation points $\mu_1, \ldots, \mu_{s-1}$ are infinite and disjoint. Note that the orbit of 0 cannot be disjoint of the others since one has $T(\mu_i) = 0$ for some i with $1 \le i \le s$.

A regular interval exchange transformation is also said to satisfy the *idoc* condition (where idoc stands for “infinite disjoint orbit condition”). It is also said to have the Keane property or to be without *connection* (see [8]).

Example 3.3 The 2-interval exchange transformation R of Example 3.1 which is the rotation of angle α is regular if and only if α is irrational.

The following result is due to Keane [22].

Theorem 3.4 (Keane) *A regular interval exchange transformation is minimal.*

The converse is not true. Indeed, consider the rotation of angle α with α irrational, as a 3-interval exchange transformation with $\lambda = (1 - 2\alpha, \alpha, \alpha)$and $\pi = (132)$. 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 $\mu_2 = T(\mu_1)$.

3.3 Natural coding

Let T be an interval exchange transformation relative to $(I_a)_{a\in A}$. For a given real number $z \in [0,1[$, the *natural coding* of T relative to z is the infinite word $\Sigma_T(z) = a_0a_1\cdots$ on the alphabet A defined by

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

Example 3.5 Let $\alpha = (3-\sqrt{5})/2$ and let R be the rotation of angle α on $[0,1[$ as in Example 3.1. The natural coding of R with respect to α is the Fibonacci word (see [24, Chapter 2] for example).

For a word $w = b_0b_1\cdots b_{m-1}$, let I_w be the set

$$I_w = I_{b_0} \cap T^{-1}(I_{b_1}) \cap \ldots \cap T^{-m+1}(I_{b_{m-1}}). \tag{3.1}$$

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.

Then one has for any $n \geq 0$

$$a_na_{n+1}\cdots a_{n+m-1} = w \Longleftrightarrow T^n(z) \in I_w \tag{3.2}$$

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 (as for Sturmian words, see [24]). Since it depends only on T, we denote it by $\mathrm{Fac}(T)$. When T is regular (resp. minimal), such a set is called a *regular interval exchange set* (resp. a minimal interval exchange set).

Let T be an interval exchange transformation. The natural codings $\Sigma_T(z)$ of T with $z \in [0,1[$ are infinite words on A. The set A^ω of infinite words on A is a topological space for the topology induced by the metric defined by the following distance. For $x = a_0a_1\cdots, y = b_0b_1\cdots \in A^\omega$ with $x \neq y$, one sets $d(x,y) = 2^{-n(x,y)}$ if $n(x,y)$ is the least n such that $a_n \neq b_n$. Let X be the closure in the space A^ω of the set of all $\Sigma_T(z)$ for $z \in [0,1[$ and let S be the shift on X. 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 $F(S)$ is uniformly recurrent (see for example [24] Theorem 1.5.9).

We have the following commutative diagram.

The map Σ_T is neither continuous nor surjective. This can be corrected by embedding the interval $[0,1[$ into a larger space on which T is a homeomophism (see [22] or [7] page 349). However, if the transformation T is minimal, the symbolic dynamical system (X,S) is minimal (see [7] page 392). Thus, we obtain the following statement.

$$\begin{tikzcd}
{[0,1[} \arrow[r, "T"] \arrow[d, "\Sigma_T"] & {[0,1[} \arrow[d, "\Sigma_T"] \\
X \arrow[r, "S"] & X
\end{tikzcd}$$

Proposition 3.6 *For any minimal interval exchange transformation T, the set* $\mathrm{Fac}(T)$ *is uniformly recurrent.*

Example 3.7 Set $\alpha = (3-\sqrt{5})/2$ and $A = \{a,b,c\}$. Let T be the interval exchange transformation on $[0,1[$ which is the rotation of angle $2\alpha \bmod 1$ on the three intervals $I_a = [0, 1-2\alpha[$, $I_b = [1-2\alpha, 1-\alpha[$, $I_c = [1-\alpha, 1[$ (see Figure 3.2). The transformation T is regular since α is irrational. The words

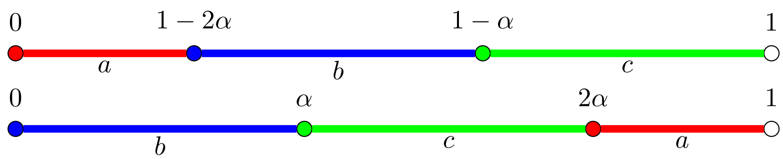

Figure 3.2: A 3-interval exchange transformation.

of length at most 5 of the set $F = \mathrm{Fac}(T)$ are represented in Figure 3.3. Since

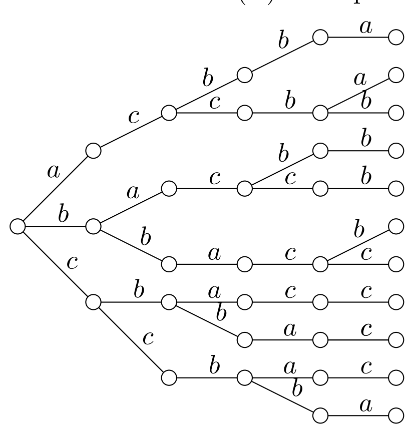

Figure 3.3: The words of length ≤ 5 of the set F.

$T = R^2$, where R is the transformation of Example 3.5, the natural coding of T relative to α is the infinite word $y = \gamma^{-1}(x)$ where x is the Fibonacci word and γ is the morphism defined by $\gamma(a) = aa$, $\gamma(b) = ab$, $\gamma(c) = ba$. One has

$$y = baccbaccbbacbbacc\cdots$$

4 Return words

In this section, we introduce the notion of return and first return words. We prove elementary results about return words which essentially already appear in [14].

Let F be a set of words. For $w \in F$, let $\Gamma_F(w) = \{x \in F \mid wx \in F \cap A^+w\}$ be the set of *right return words* to w and let $\mathcal{R}_F(w) = \Gamma_F(w) \setminus \Gamma_F(w)A^+$ be the set of *first right return words* to w. By definition, the set $\mathcal{R}_F(w)$ is, for any $w \in F$, a prefix code. If F is recurrent, it is a $w^{-1}F$-maximal prefix code.

Similarly, for $w \in F$, we denote $\Gamma'_F(w) = \{x \in F \mid xw \in F \cap wA^+\}$ the set of *left return words* to w and $\mathcal{R}'_F(w) = \Gamma'_F(w) \setminus A^+\Gamma'_F(w)$ the set of *first left return words* to w. By definition, the set $\mathcal{R}'_F(w)$ is, for any $w \in F$, a suffix code. If F is recurrent, it is an Fw^{-1}-maximal suffix code. The relation between $\mathcal{R}_F(w)$ and $\mathcal{R}'_F(w)$ is simply

$$w\mathcal{R}_F(w) = \mathcal{R}'_F(w)w\,. \tag{4.1}$$

Let $f : B^* \to A^*$ is a coding morphism for $\mathcal{R}_F(w)$. The morphism $f' : B^* \to A^*$ defined for $b \in B$ by $f'(b)w = wf(b)$ is a coding morphism for $\mathcal{R}'_F(w)$ called the coding morphism *associated* to f.

Example 4.1 Let F be the uniformly recurrent set of Example 3.7. We have

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

Note that $\Gamma_F(w) \cup \{1\}$ is right unitary and that

$$\Gamma_F(w) \cup \{1\} = \mathcal{R}_F(w)^* \cap w^{-1}F. \tag{4.2}$$

Indeed, if $x \in \Gamma_F(w)$ is not in $\mathcal{R}_F(w)$, we have $x = zu$ with $z \in \Gamma_F(w)$ and u nonempty. Since $\Gamma_F(w)$ is right unitary, we have $u \in \Gamma_F(w)$, whence the conclusion by induction on the length of x. The converse inclusion is obvious.

Proposition 4.2 *A recurrent set F is uniformly recurrent if and only if the set $\mathcal{R}_F(w)$ is finite for any $w \in F$.*

Proof. Assume that all sets $\mathcal{R}_F(w)$ for $w \in F$ are finite. Let $n \geq 1$. Let N be the maximal length of the words in $\mathcal{R}_F(w)$ for a word w of length n, then any word of length $N + 2n - 1$ contains an occurrence of w. Conversely, for $w \in F$, let N be such that w is a factor of any word in F of length N. Then the words of $\mathcal{R}_F(w)$ have length at most $|w| + N - 1$. ■

Let F be a recurrent set and let $w \in F$. Let f be a coding morphism for $\mathcal{R}_F(w)$. The set $f^{-1}(w^{-1}F)$, denoted $D_f(F)$, is called the *derived set* of F with respect to f. Note that if f' is the coding morphism for $\mathcal{R}'_F(w)$ associated to f, then $D_f(F) = f'^{-1}(Fw^{-1})$.

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

Proposition 4.3 *Let F be a recurrent set. For $w \in F$, let f be a coding morphism for the set $\mathcal{R}_F(w)$. Then*

$$D_f(F) = f^{-1}(\Gamma_F(w)) \cup \{1\}. \tag{4.3}$$

Proof. Let $z \in D_f(F)$. Then $f(z) \in w^{-1}F \cap R_F(w)^*$ and thus $f(z) \in \Gamma_F(w) \cup \{1\}$. Conversely, if $x \in \Gamma_F(w)$, then $x \in \mathcal{R}_F(w)^*$ by Equation (4.2) and thus $x = f(z)$ for some $z \in D_f(F)$. ∎

Let F be a recurrent set and x be an infinite word such that $F = \mathrm{Fac}(x)$. Let $w \in F$ and let f be a coding morphism for the set $\mathcal{R}_F(w)$. Since w appears infinitely often in x, there is a unique factorization $x = vwz$ with $z \in \mathcal{R}_F(w)^\omega$ and v such that vw has no proper prefix ending with w. The infinite word $f^{-1}(z)$ is called the *derived word* of x relative to f. If f' is the coding morphism for $\mathcal{R}'_F(w)$ associated to f, we have $f^{-1}(z) = f'^{-1}(wz)$ and thus f, f' define the same derived word.

The following well-known result (for a proof, see [6] for example), shows in particular that the derived set of a recurrent set is recurrent.

Proposition 4.4 *Let F be a recurrent set and let x be a recurrent infinite word such that $F = \mathrm{Fac}(x)$. Let $w \in F$ and let f be a coding morphism for the set $\mathcal{R}_F(w)$. The derived set of F with respect to f is the set of factors of the derived word of x with respect to f, that is $D_f(F) = \mathrm{Fac}(D_f(x))$.*

Example 4.5 Let F be the uniformly recurrent set of Example 3.7. Let f be the coding morphism for the set $\mathcal{R}_F(c)$ given by $f(a) = bac$, $f(b) = bbac$, $f(c) = c$. Then the derived set of F with respect to f is represented in Figure 4.1.

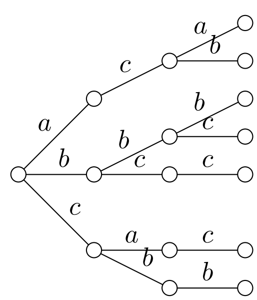

Figure 4.1: The words of length ≤ 3 of the derived set of F.

5 Uniformly recurrent tree sets

In this section, we recall the notion of tree set introduced in [4]. We recall that the factor complexity of a tree set on $k+1$ letters is $p_n = kn + 1$. Observe that uniformly recurrent ternary tree sets, which will be considered in Section 6, are uniquely ergodic as a consequence of the fact that a minimal symbolic system such that $\limsup p_n/n < 3$ is uniquely ergodic [9].

We recall a result concerning the decoding of tree sets (Theorem 5.5). We also recall the finite index basis property of uniformly recurrent tree sets (Theorems 5.6 and 5.7) that we will use in Section 7. We prove that the family of uniformly recurrent tree sets is invariant under derivation (Theorem 5.10). We further prove that all bases of the free group included in a uniformly recurrent tree set are tame (Theorem 5.16).

5.1 Tree sets

For a biextendable word w, we consider the undirected graph $G(w)$ on the set of vertices which is the disjoint union of $L(w)$ and $R(w)$ with edges the pairs $(a,b) \in E(w)$. The graph $G(w)$ is called the *extension graph* of w in F.

Recall that an undirected graph is a tree if it is connected and acyclic.

We say that F is a *tree set* (resp. an acyclic set) if it is biessential and if for every word $w \in F$, the graph $G(w)$ is a tree (resp. is acyclic).

It is not difficult to verify the following statement (see [4], Proposition 4.3) which shows that the factor complexity of a tree set is linear.

Proposition 5.1 *Let F be a tree set on the alphabet A and let $k = \mathrm{Card}(A \cap F) - 1$. Then $\mathrm{Card}(F \cap A^n) = kn + 1$ for all $n \geq 0$*

The following result is also easy to prove.

Proposition 5.2 *A Sturmian set F is a uniformly recurrent tree set.*

Proof. We have already seen that a Sturmian set is uniformly recurrent. Let us show that it is a tree set. Consider $w \in F$. If w is not left-special there is a unique $a \in A$ such that $aw \in F$. Then $E(w) \subset a \times A$ and thus $G(w)$ 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 $E(w) = a \times A \cup A \times b$ and thus $G(w)$ is a tree. ■

Putting together Propositions 3.6 and Proposition 5.8 in [5], we have the similar statement.

Proposition 5.3 *A regular interval exchange set is a uniformly recurrent tree set.*

Proposition 5.3 is actually a particular case of a result of [18] which characterizes the regular interval exchange sets.

Let F be a set. For $w \in F$, and $U, V \subset F$, let $U(w) = \{\ell \in U \mid \ell w \in F\}$ and let $V(w) = \{r \in V \mid wr \in F\}$. The *generalized extension graph* of w relative to U, V is the following undirected graph $G_{U,V}(w)$. The set of vertices is made of two disjoint copies of $U(w)$ and $V(w)$. The edges are the pairs (ℓ, r) for

$\ell \in U(w)$ and $r \in V(w)$ such that $\ell w r \in F$. The extension graph $G(w)$ defined previously corresponds to the case where $U, V = A$.

The following result is proved in [4] (Proposition 4.9).

Proposition 5.4 *Let F be a tree set. For any $w \in F$, any finite F-maximal suffix code $U \subset F$ and any finite F-maximal prefix code $V \subset F$, the generalized extension graph $G_{U,V}(w)$ is a tree.*

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

The following result is Theorem 4.13 in [4].

Theorem 5.5 *Any maximal bifix decoding of a recurrent tree set is a tree set.*

We have no example of a bifix decoding of a recurrent tree set which is not recurrent (in view of Theorem 7.1 to be proved hereafter, such a set would be the decoding of a recurrent tree set which is not uniformly recurrent).

5.2 The finite index basis property

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

We recall the main result of [5] (Theorem 6.1).

Theorem 5.6 *A uniformly recurrent tree set containing the alphabet A has the finite index basis property.*

Recall from Section 2.3 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.

We will use the following result. It is stated for a Sturmian set F in [2] (Theorem 7.2.5) but the proof only uses the fact that F is uniformly recurrent and satisfies the finite index basis property. We reproduce the proof for the sake of clarity.

For a set of words X, we denote by $\langle X \rangle$ the subgroup of the free group on A generated by X. The free group on A itself is consistently denoted $\langle A \rangle$.

Theorem 5.7 *Let $Z \subset A^+$ be a group code of degree d. For every uniformly recurrent tree set F containing the alphabet A, the set $X = Z \cap F$ is a basis of a subgroup of index d of $\langle A \rangle$.*

Proof. By Theorem 4.2.11 in [2], the code X is an F-maximal bifix code of F-degree $e \leq d$. Since F is a uniformly recurrent, by Theorem 4.4.3 of [2], X is finite. By Theorem 5.6, 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 F is a uniformly recurrent tree set, then $F \cap A^n$ is a basis of the subgroup formed by the words of length multiple of n (where the length is not the length of the reduced word but the sum of values 1 for the letters in A and -1 for the letters in A^{-1}).

We will use the following results from [4]. The first one is Corollary 5.8 in [4].

Theorem 5.8 *Let F be a uniformly recurrent tree set containing the alphabet A. For any word $w \in F$, the set $\mathcal{R}_F(w)$ is a basis of the free group on A.*

The next result is Theorem 6.2 in [4]. A submonoid M of A^* is *saturated* in a set F if $M \cap F = \langle M \rangle \cap F$.

Theorem 5.9 *Let F be an acyclic set. The submonoid generated by any bifix code $X \subset F$ is saturated in F.*

5.3 Derived sets of tree sets

We will use the following closure property of the family of uniformly recurrent tree sets. It generalizes the fact that the derived word of a Sturmian word is Sturmian (see [21]).

Theorem 5.10 *Any derived set of a uniformly recurrent tree set is a uniformly recurrent tree set.*

Proof. Let F be a uniformly recurrent tree set containing A, let $v \in F$ and let f be a coding morphism for $X = \mathcal{R}_F(v)$. By Theorem 5.8, X is a basis of the free group on A. Thus $f : B^* \to A^*$ extends to an isomorphism from $\langle B \rangle$ onto $\langle A \rangle$.

Set $H = f^{-1}(v^{-1}F)$. By Proposition 4.3, the set H is recurrent and $H = f^{-1}(\Gamma_F(v)) \cup \{1\}$.

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

$$(a, b) \in G(x) \Leftrightarrow (f'(a), f(b)) \in G_{X',X}(vy)$$

Indeed,

$$axb \in H \Leftrightarrow f(a)yf(b) \in \Gamma_F(v) \Leftrightarrow vf(a)yf(b) \in F \Leftrightarrow f'(a)vyf(b) \in F.$$

The set X' is an Fv^{-1}-maximal suffix code and the set X is a $v^{-1}F$-maximal prefix code. By Proposition 5.4 the generalized extension graph $G_{X',X}(vy)$ is a tree. Thus the graph $G(x)$ is a tree. This shows that H is a tree set.

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

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

Example 5.11 Let F be the Tribonacci set (see Example 2.2). It is the set of factors of the infinite word $x = abacaba\cdots$ which is the fixpoint of the morphism f defined by $f(a) = ab$, $f(b) = ac$, $f(c) = a$. We have $\mathcal{R}_F(a) = \{a, ba, ca\}$. Let g be the coding morphism for $\mathcal{R}_F(a)$ defined by $g(a) = a$, $g(b) = ba$, $g(c) = ca$ and let g' be the associated coding morphism for $\mathcal{R}'_F(a)$. We have $f = g'\pi$ where π 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 F with respect to a is the set $\pi(F)$.

5.4 Tame bases

An automorphism α of the free group on A is *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* 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

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

Thus $\alpha_{a,b}$ places a b after each a and $\tilde{\alpha}_{a,b}$ places a 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 [27]).

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 α such that $X = \alpha(A)$.

Example 5.12 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.

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

The following result will play a key role in the proof of the main result of this section (Theorem 5.16).

Proposition 5.13 *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 $X = (Y \setminus u) \cup uv$.*

Proof. Assume first that X is a tame basis of the free group on A. Then $X = \alpha(A)$ where α is a tame automorphism of $\langle A\rangle$. Then $\alpha = \alpha_1\alpha_2\cdots\alpha_n$ where the α_i are elementary positive automorphisms. We use an induction on n. If $n = 0$, then $X = A$. If α_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 β 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 5.14 *A tame basis which is a bifix code is the alphabet.*

Proof. Assume that X is a tame basis which is not the alphabet. By Proposition 5.13 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 [27].

Example 5.15 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 5.14.

The following result is a remarquable consequence of Theorem 5.6.

Theorem 5.16 *Any basis of the free group included in a uniformly recurrent tree set is tame.*

Proof. Let F be a uniformly recurrent tree set. Let $X \subset F$ be a basis of the free group on A. We use an induction on the sum $\ell(X)$ of the lengths of the words of X. If X is bifix, by Theorem 5.6, it is an F-maximal bifix code of F-degree 1. Thus $X = A$ (see Example 2.4). 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 $\ell(Y) < \ell(X)$. By induction hypothesis, Y is tame. Since $X = (Y \setminus v) \cup uv$, X is tame by Proposition 5.13. ∎

Example 5.17 The set $X = \{ab, acb, acc\}$ is a basis of the free group which is not tame (see Example 5.15). Accordingly, the extension graph $G(\varepsilon)$ relative to the set of factors of X is not a tree (see Figure 5.1).

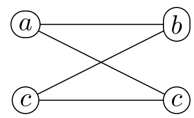

Figure 5.1: The graph $G(\varepsilon)$

6 S-adic representations

In this section we study S-adic representations of tree sets. This notion was introduced in [17], using a terminology initiated by Vershik and coined out by B. Host. We first recall a general construction allowing to build S-adic representations of any uniformly recurrent aperiodic set (Proposition 6.1) which is based on return words. Using Theorem 5.16, we show that this construction actually provides $\mathcal{S}_e$-representations of uniformly recurrent tree sets (Theorem 6.5), where $\mathcal{S}_e$ is the set of elementary positive automorphisms of the free group on A.

We then investigate the case of a ternary alphabet where a careful study of Rauzy graphs allows us to provide an $\mathcal{S}_3$-adic characterization of uniformly recurrent ternary tree sets with $\mathcal{S}_3$ being the set of elementary positive automorphisms of $\langle\{0,1,2\}\rangle$ (Theorem 6.6). In particular, this characterization can be expressed using some (non-deterministic) Büchi automaton (Theorem 6.9).

6.1 S-adic representation of tree sets

Let S be a set of morphisms and $\mathbf{s} = (\sigma_n)_{n\in\mathbb{N}}$ be a sequence in $S^{\mathbb{N}}$ with $\sigma_n : A_{n+1}^* \to A_n^*$. We let $F_{\mathbf{s}}$ 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 F an *S-adic set* if there exists $\mathbf{s} \in S^{\mathbb{N}}$ such that $F = F_{\mathbf{s}}$. In this case, the sequence $\mathbf{s}$ is called an *S-adic representation* of F.

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 F 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 S-adic representation of any aperiodic uniformly recurrent set F.

Proposition 6.1 *An aperiodic factorial set $F \subset A^*$ is uniformly recurrent if and only if it has a primitive S-adic representation for some (possibly infinite) set S of morphisms.*

Proof. Let S be a set of morphisms and $\mathbf{s} = (\sigma_n : A_{n+1}^* \to A_n^*)_{n\in\mathbb{N}} \in S^{\mathbb{N}}$ a primitive sequence of morphisms such that $F = \bigcap_{n\in\mathbb{N}} \mathrm{Fac}(\sigma_0 \cdots \sigma_n(A_{n+1}^*))$. Consider a word $u \in F$ and let us prove that $u \in \mathrm{Fac}(v)$ for all long enough $v \in F$. The sequence $\mathbf{s}$ 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 $F = \bigcap_{n\in\mathbb{N}} \mathrm{Fac}(\sigma_0 \cdots \sigma_n(A_{n+1}^*))$, 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 $\mathbf{s}$ 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 $v \in F$ such that $|v| \geq \max_{d\in A_t} |\sigma_0 \cdots \sigma_{t-1}(d)|$ and F is uniformly recurrent.

Let us prove the only if part. Let $(u_n)_{n\in\mathbb{N}} \in F^{\mathbb{N}}$ be a non-ultimately periodic sequence such that u_n is suffix of u_{n+1}. By assumption, F is uniformly recurrent so $\mathcal{R}_F(u_{n+1})$ is finite for all n. The set F being aperiodic, $\mathcal{R}_F(u_{n+1})$ also has cardinality at least 2 for all n. For all n, let $A_n = \{0, \ldots, \mathrm{Card}(\mathcal{R}_F(u_n))-1\}$ and let $\alpha_n : A_n^* \to A^*$ be a coding morphism for $\mathcal{R}_F(u_n)$. The word u_n being suffix of u_{n+1}, we have $\alpha_{n+1}(A_{n+1}) \subset \alpha_n(A_n^+)$. Since $\alpha_n(A_n) = \mathcal{R}_F(u_n)$ 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 $\mathcal{R}_F(u_n) = \alpha_0\sigma_0\sigma_1 \cdots \sigma_{n-1}(A_n)$ and $F = \bigcap_{n\in\mathbb{N}} \mathrm{Fac}(\alpha_0\sigma_0 \cdots \sigma_n(A_{n+1}^*))$. 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 F thus has an S-adic representation with $S = \{\sigma_n \mid n \in \mathbb{N}\}$.

Let us show that $\mathbf{s} = (\sigma_n)_{n\in\mathbb{N}}$ 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 $n \geq N$. This means that the word $r = \sigma_0 \cdots \sigma_n(a_n) \in F$ is a first return word to u_n for all $n \geq N$. The sequence $(|u_n|)_{n\in\mathbb{N}}$ being unbounded, the word r^k belongs to F for all positive integers k, which contradicts the uniform recurrence of F.

Let us show that $\mathbf{s}$ is primitive. The set F being uniformly recurrent, for all $n \in \mathbb{N}$ there exists N_n such that all words of $F \cap A^{\leq n}$ occur in all words of $F \cap A^{\geq N_n}$. 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 $\sigma_0 \cdots \sigma_{r-1}(A_r) = \mathcal{R}_F(u_r)$ is a prefix code, the letter $a \in A_r$ occurs in $\sigma_r \cdots \sigma_{s-1}(b)$ for all $b \in A_r$. ■

Remark 6.2 In the continuation of the proof of the above proposition, we could also consider a sequence $(a_n \in A_n)_{n\in\mathbb{N}}$ of letters such that $\sigma_n(a_{n+1}) \in a_n A_n^*$ (such a sequence exists by application of König's lemma). By doing so, we would build a uniformly recurrent infinite word $\mathbf{w} = \lim_{n\to+\infty} \sigma_0 \cdots \sigma_n(a_{n+1})$ with F for set of factors. According to Durand [14], $\mathbf{w}$ is substitutive if and only if there is a sequence of words $(u_n)_{n\in\mathbb{N}}$ that makes the sequence $(\sigma_n)_{n\in\mathbb{N}}$ be ultimately periodic.

Remark 6.3 In the proof of the previous proposition, the same construction works if we define the sequence $(u_n)_{n\in\mathbb{N}}$ such that u_n is prefix of u_{n+1} and if we consider $\mathcal{R}'_F(u_n)$ instead of $\mathcal{R}_F(u_n)$.

Remark 6.4 Still in the continuation of the proof, we can also slightly modify the construction in such a way that the sequence $(\sigma_n)_{n\in\mathbb{N}}$ is *proper*, *i.e.*, for all n, there is an integer $m > n$ and two letters $a, b \in A_n$ such that $\sigma_n \cdots \sigma_{m-1}(A_m) \subset aA_n^* \cap A_n^*b$. According to Durand [15, 16], if S is finite, then F is linearly recurrent if and only if there is an integer $k \geq 0$ such that for all $n \in \mathbb{N}$, all letters of A_n occur in $\sigma_n \cdots \sigma_{n+k}(a)$ for all $a \in A_{n+k+1}$ (this property is called *strong primitiveness*) and there are two letters $a, b \in A_n$ such that $\sigma_n \cdots \sigma_{n+k}(A_{n+k+1}) \subset aA_n^* \cap A_n^* b$.

Even for uniformly recurrent sets with linear factor complexity, the set of morphisms $S = \{\sigma_n \mid n \in \mathbb{N}\}$ considered in Proposition 6.1 usually is infinite as well as the sequence of alphabets $(A_n)_{n\in\mathbb{N}}$ usually is unbounded (see [13]). For tree sets F, the next theorem significantly improves the only if part of Proposition 6.1: For such sets, the set 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.

Theorem 6.5 *If F is a uniformly recurrent tree set over an alphabet A, then it has a primitive $\mathcal{S}_e$-adic representation.*

Proof. For any non-ultimately periodic sequence $(u_n)_{n\in\mathbb{N}} \in F^{\mathbb{N}}$ 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 6.1 is a primitive S-adic representation of F with $S = \{\sigma_n \mid n \in \mathbb{N}\}$. Therefore, all we need to do is to consider such a sequence $(u_n)_{n\in\mathbb{N}}$ such that σ_n is tame for all n.

Let $u_1 = a^{(0)}$ be a letter in A. By Theorem 5.8, the set $\mathcal{R}_F(u_1)$ is a basis of the free group on A. Therefore, by Theorem 5.16, 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 $u_2 \in \mathcal{R}_F(u_1)$ and u_1 is a suffix of u_2. By Theorem 5.10, the derived set $F^{(1)} = \sigma_0^{-1}(F)$ 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 $n \geq 2$. ■

6.2 The case of a ternary alphabet

In the case of a ternary alphabet A, Theorem 6.5 can again be significantly improved into an if and only if result. Let us first recall the notion of Rauzy graph. For $n \in \mathbb{N}$, the *Rauzy graph of order n* of F is the directed graph $G_F(n) = (V(n), E(n))$ where the set of vertices is $V(n) = F \cap A^n$ and there is an edge from u to v if there is a word in $F \cap A^{n+1}$ with prefix u and suffix v. We extend the notions of left special, right special and bispecial word to vertices of $G_F(n)$. Clearly, a vertex will be left special (resp. right special) if it has at least two incoming edges (resp. outgoing edges) and bispecial if it is both left and right special.

Theorem 6.6 *Let $\mathcal{S}_3$ be the set of elementary positive automorphisms over $\{0,1,2\}$. A set F is a uniformly recurrent tree set over $\{0,1,2\}$ if and only if it has a primitive $\mathcal{S}_3$-adic representation that labels an infinite path starting at vertex 2 in the directed graph represented in Figure 6.1 where edges are labeled by subsets of $\mathcal{S}_3^*$ given in Appendix A.*

The proof almost immediately follows from the proof of Theorem 5.24 in [23] that we recall below (with a slightly different statement). Observe that we do not recall hereafter the detailed formulation of [23] because it would force us to give a lot of developments that are useless for Theorem 6.6 as the vertex 1 of Figure 6.2 disappears.

Theorem 6.7 (Theorem 5.24 in [23]) *Let $M : \{0,1,2\}^* \to \{0,1,2\}^*$ be the morphism defined by $M(0) = M(2) = 0$ and $M(1) = 1$ and let $\mathcal{S} = \mathcal{S}_3 \cup \{M\}$. An aperiodic uniformly recurrent set F satisfies $p(n+1) - p(n) \leq 2$ if and only if it has a primitive $\mathcal{S}$-adic representation $(\sigma_n)_{n\in\mathbb{N}}$ that labels a path starting in vertex 1 or in vertex 2 in Figure* 6.2 *where edges are labeled by subsets of $\mathcal{S}^*$ given in Appendix* A *and Appendix* B *and that satisfies some (computable) condition on the lengths of $\sigma_0 \cdots \sigma_n$ when σ_n labels the edge from* 5/6 *or* 7/8 *to* 1*.*

The idea of the proof of Theorem 6.7 is the following. For the considered class of factor complexity, the Rauzy graphs G_n can have 10 different shapes. When computing the morphisms σ_n of Proposition 6.1, we notice that they only depend on the shapes of the Rauzy graphs of order $|u_n|$ and $|u_{n+1}|$. The result is thus obtained by choosing for each n a word of length n for u_n and computing all possible morphisms σ_n. The sequence $(\sigma_n)_{n\in\mathbb{N}}$ thus describes a path in what we called the *graph of graphs* that consists in the directed graph with one vertex for each shape of Rauzy graph and with an edge from a vertex i to a vertex j when a Rauzy graph G_n of shape i can evolve into a Rauzy graph G_{n+1} of shape j. A given Rauzy graph of shape i usually has several possibilities to evolve to a Rauzy graph of shape j. The edges are thus labeled by several morphisms. Theorem 6.7 is obtained by a careful description of the infinite path in the graph of graphs that really correspond to $\mathcal{S}$-adic representations of these sets. A detailed computation can be found in [23].

Proof of Theorem 6.6. If F is a uniformly recurrent tree set over $\{0,1,2\}$, it satisfies $p(n+1) - p(n) = 2$ for all n. Uniformly recurrent tree sets on three letters are thus particular cases of aperiodic uniformly recurrent sets with $p(n+1) - p(n) \leq 2$. To get the $\mathcal{S}_3$-adic characterization of Theorem 6.6, it thus suffices to remove from Theorem 6.7 all cases not corresponding to tree sets.

In Figure 6.2, the vertex 1 corresponds to a Rauzy graph with exactly one right special vertex with exactly two outgoing edges and exactly one left special vertex with exactly two incoming edges (these two vertices being possibly the same bispecial vertex). If the Rauzy graph G_n has such a shape, then $p(n+1) - p(n) = 1$, which is never the case for ternary tree sets (see Proposition 5.1). The vertex 1 can thus be removed from the graph of graphs of tree sets over $\{0,1,2\}$.

When computing all morphisms labeling the edges not related to the vertex 1, we observe that some of them correspond to evolutions of Rauzy graphs involving bispecial factors whose extension graph is not a tree. This is the case exactly for the morphisms given in Appendix B labeling the edge from the vertex

2 to the vertex 7/8. We thus remove these morphisms from the set of labels of this edge. Observe that this agrees with Theorem 6.5: these morphisms are exactly those that belong to $\mathcal{S}^+ \setminus \mathcal{S}_3^+$. ■

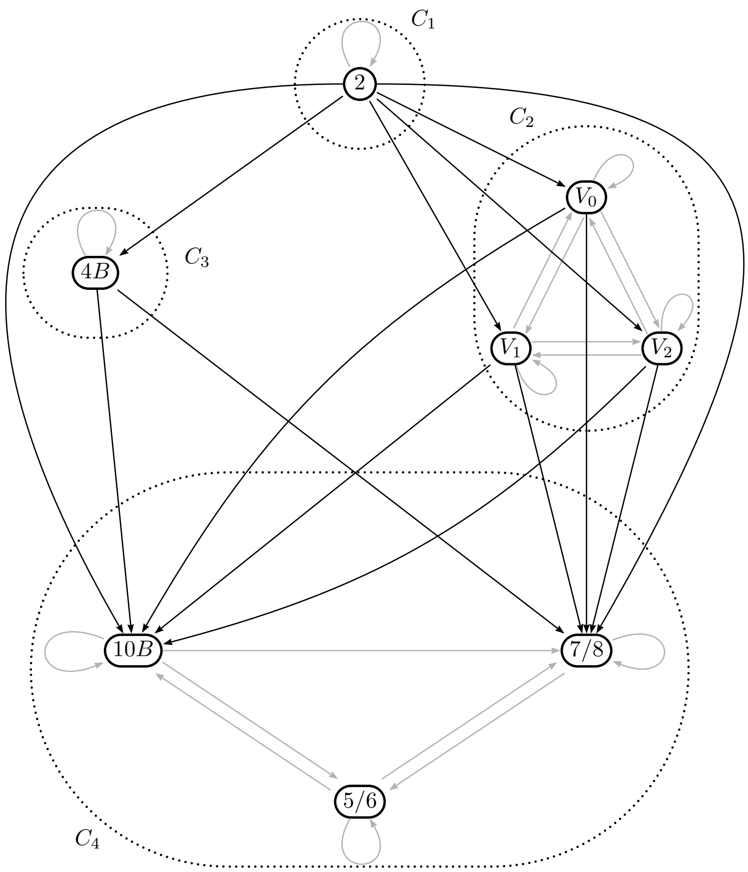

Figure 6.1: Graph of graphs for ternary tree sets.

Remark 6.8 There are four strongly connected components in Figure 6.1 that are denoted C_1, C_2, C_3 and C_4. The component C_1 corresponds to Arnoux-Rauzy words (also called strict episturmian words). The component C_2 corresponds to words such that for all large enough n, there is exactly one right special factor w with $r(w) = 3$ and two left special factors u, v, each with $\ell(u) = \ell(v) = 2$. The component C_3 is the opposite case of C_2: for all large enough n, there is exactly one left special factor w with $\ell(w) = 3$ and two right special factors u, v with $r(u) = r(v) = 2$. The component C_4 corresponds to

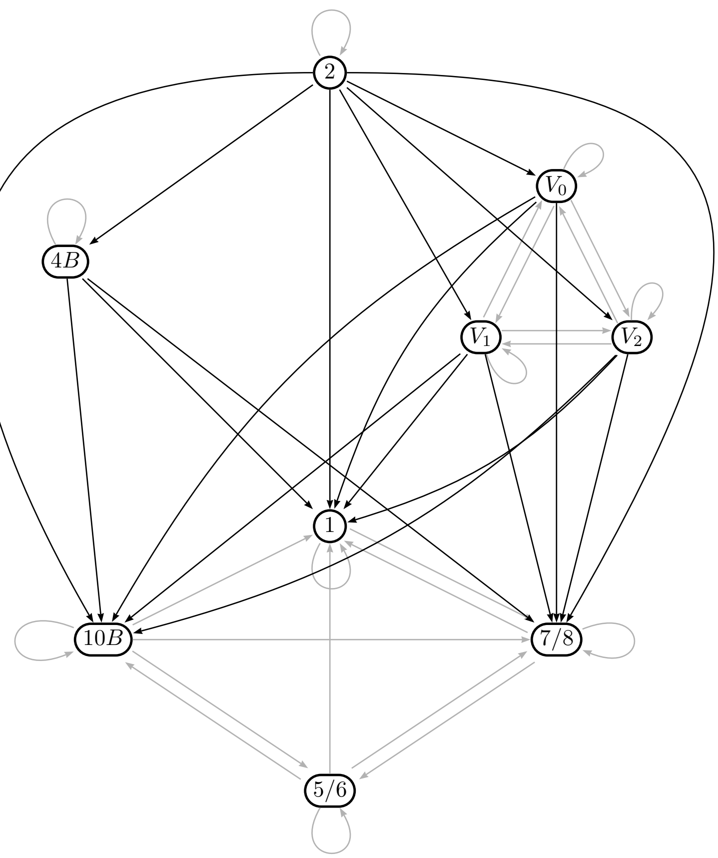

Figure 6.2: Modified graph of graphs.

words such that for all large enough n, there are exactly two right (resp. left) special factors u, v (resp. u', v') with $r(u) = r(v) = 2$ (resp. $\ell(u') = \ell(v') = 2$). All 3-interval exchange words eventually end up in C_4.

In Theorem 6.7 and Theorem 6.6, the condition of primitiveness of $(\sigma_n)_{n\in\mathbb{N}}$ can be hard to describe in the graph of graphs. Theorem 5.24 in [23] gives evidence of this fact: the description of the primitiveness in Figure 6.2 needs a 2 page-long statement. The next result shows that this condition can be verified by a (non-deterministic) automaton.

Recall that a *Büchi automaton* is an automaton with a condition of acceptance adapted to infinite words. An infinite word is accepted by such an automaton if it labels an infinite path starting in an initial state and visiting infinitely often terminal states.

Theorem 6.9 *There exists a Büchi automaton $\mathcal{A}$ over the alphabet $\mathcal{S}_3$ such that F is a uniformly recurrent tree set if and only if it has an $\mathcal{S}_3$-adic representation accepted by $\mathcal{A}$.*

We will use the following lemma.

Lemma 6.10 *The set of primitive sequences of morphisms in $\mathcal{S}_3$ is accepted by some Büchi automaton $\mathcal{P}$.*

Proof. The automaton will read infinite words $(\sigma_n)_{n\in\mathbb{N}} \in \mathcal{S}_3^{\mathbb{N}}$. The idea of the proof is to keep track of the letters occurring in $\sigma_k \cdots \sigma_{k+n}(a)$ for $a \in \{0,1,2\}$ and any k, n. This information will be registered in the vertices and the edges will be labeled by the morphisms σ_n.

Consider the set T of 3-tuples (u, v, w) of non-empty words in $0^*1^*2^*$ of length at most 3. These words are devoted to register the letters occurring in $\sigma_k \cdots \sigma_{k+n}(a)$ for $a \in \{0,1,2\}$ and we will put an edge from a state (u, v, w) to a state (u', v', w') with label σ if the letters occurring in u (resp. v, w) are exactly those occurring in $\sigma(u')$ (resp. $\sigma(v')$, $\sigma(w')$).

Let us formalize this. Given a 3-tuple (u', v', w') of non-empty words over $\{0,1,2\}$, we let $\lambda(u', v', w')$ denote the element $t = (u, v, w) \in T$ where u (resp. v, w) is the lexicographically smallest word in $0^*1^*2^*$ that contains an occurrence of $0, 1$ or 2 if and only if u' does (resp. v', w'). For instance, $\lambda((2011210, 1, 220002)) = (012, 1, 02)$. Clearly, a sequence of morphisms $(\sigma_n)_{n\in\mathbb{N}} \in \mathcal{S}_3^{\mathbb{N}}$ is primitive if and only if for all $n \geq 0$, there exists $m \geq n$ such that $\lambda(\sigma_n \cdots \sigma_m(0), \sigma_n \cdots \sigma_m(1), \sigma_n \cdots \sigma_m(2)) = (012, 012, 012)$.

Let us build the automaton. Let $P \subset T$ be the set of permutations of $(0, 1, 2)$. The set of states of the automaton is $Q = T \setminus P$. The initial state $t_0 = (012, 012, 012)$ is also the unique terminal state.

Let us define the transitions. A sequence of morphisms $(\sigma_n)_{n\in\mathbb{N}}$ is primitive if it can be cut into pieces $\Sigma_i = \sigma_{k_i} \cdots \sigma_{k_{i+1}-1}$ for an increasing sequence $(k_i)_{i\in\mathbb{N}}$ with $k_0 = 0$ such that $\lambda(\Sigma_i(t)) = t_0$ for all $t \in P$. Our aim is to define the transitions in such a way that all Σ_i label a path from t_0 to t_0.

For all $t, t' \in Q$, $t' \neq t_0$, for all $\sigma \in \mathcal{S}_3$, there is a transition with label σ from t to t' if and only if $t = \lambda(\sigma(t'))$. With these transitions, we can start from t_0 and, reading morphisms as labels, reach some states with smaller words as components. If the sequence of morphisms is primitive, we should be able to reach a triple $p \in P$. As $P \nsubseteq Q$, we add the following transitions that allow us to get back to t_0 when such a triple should be reached. For all $\sigma \in \mathcal{S}_3$ and all $t \in Q$, there is a transition with label σ from t to t_0 if and only if $t \in \sigma(P)$.

By construction, this automaton accepts exactly the primitive sequences of morphisms in $\mathcal{S}_3^{\mathbb{N}}$. ∎

Proof of Theorem 6.9. The Büchi automaton is obtained from Figure 6.1 and from the automaton $\mathcal{P}$ built in Lemma 6.10. Indeed, by Theorem 6.6 a set F is a tree set if and only if it has a primitive $\mathcal{S}_3$-adic representation labeling an infinite path starting in vertex 2 in Figure 6.1. Our goal thus is to turn Figure 6.1 into a Büchi automaton with initial state 2 and all states being terminal and then, to consider the intersection of it with $\mathcal{P}$. This can be achieved by showing that the edges in Figure 6.1 are labeled by rational subsets of $\mathcal{S}_3^+$.

We can observe in Appendix A that the edges in Figure 6.1 are labeled by subsets of $\mathcal{S}_3^{\leq K}\Sigma$ for some constant K, where Σ can be one of the sets given in Equation (A.1).

Let us first show that the following sets are rational subsets of $\mathcal{S}_3^*$.

$$\begin{aligned}
\Sigma_1(i) &= \{[0, 1^k2, 1^{k-1}2] \mid k \geq i\}, \quad i \geq 1 \\
&= \tilde{\alpha}_{2,1}^* \tilde{\alpha}_{2,1}^{i-1} \alpha_{1,2}, \quad i \geq 1 \\
\Sigma_{2,-}(i) &= \{[1^{k-1}0, 21^k0, 21^{k-1}0] \mid k \geq i\}, \quad i \geq 1 \\
&= \tilde{\alpha}_{0,1}^* \tilde{\alpha}_{0,1}^{i-1} \tilde{\alpha}_{1,2} \alpha_{1,0} \alpha_{2,0}, \quad i \geq 1 \\
\Sigma_{2,<}(i) &= \{[1^\ell 0, 21^k0, 21^{k-1}0] \mid k-1 > \ell \geq i\}, \quad i \geq 0 \\
&= \alpha_{2,1}^+ \tilde{\alpha}_{0,1}^* \tilde{\alpha}_{0,1}^{i} \tilde{\alpha}_{1,2} \alpha_{1,0} \alpha_{2,0}, \quad i \geq 0 \\
\Sigma_{3,=}(i)[2,1,0] &= \{[1^{k-1}0, 1^k0, 21^k0] \mid k \geq i\}, \quad i \geq 1 \\
&= \tilde{\alpha}_{0,1}^* \tilde{\alpha}_{0,1}^{i-1} \alpha_{1,0} \alpha_{2,1}, \quad i \geq 1 \\
\Sigma_{3,-}(i)[2,1,0] &= \{[1^{k-1}0, 1^k0, 21^{k-1}0] \mid k \geq i\}, \quad i \geq 1 \\
&= \tilde{\alpha}_{0,1}^* \tilde{\alpha}_{0,1}^{i-1} \alpha_{1,0} \alpha_{2,0}, \quad i \geq 1 \\
\Sigma_{3,0}(i)[2,1,0] &= \{[1^{k-1}0, 1^k0, 20] \mid k \geq i\}, \quad i \geq 1 \\
&= \alpha_{2,0} \tilde{\alpha}_{0,1}^* \tilde{\alpha}_{0,1}^{i-1} \alpha_{1,0}, \quad i \geq 1
\end{aligned}$$

Obviously the sets $\Sigma_{2,\leq}(i) = \Sigma_{2,-}(i) \cup \Sigma_{2,<}(i)$, $i \geq 0$, are also rational as well as the set $\Sigma_{2,0}(i)$ as it can easily be deduced from $\Sigma_{2,<}(i)$

At the opposite, the following sets are not rational subsets of $\mathcal{S}_3^*$.

$$\begin{aligned}
\Sigma_{2,=}(i) &= \{[1^k0, 21^k0, 21^{k-1}0] \mid k \geq i\}, \quad i \geq 1 \\
\Sigma_{3,<}(i) &= \{[21^\ell 0, 1^k0, 1^{k-1}0] \mid k-1 > \ell \geq i\}, \quad i \geq 0 \\
\Sigma_{3,\leq}(i) &= \{[21^\ell 0, 1^k0, 1^{k-1}0] \mid k-1 \geq \ell \geq i\}, \quad i \geq 0
\end{aligned}$$

Our goal is therefore to modify these sets into rational ones while keeping $\mathcal{S}_3$-adic representations. We can observe that for all morphisms in these sets, the last letter of the image is always 0. Let β_0 be the inner automorphism of the free group $\langle\{0,1,2\}\rangle$ defined by $\beta_0 : u \mapsto 0u0^{-1}$. Let $(\sigma_n)_{n\in\mathbb{N}}$ be an $\mathcal{S}_3$-adic representation of a tree set F and suppose that $\sigma_r \cdots \sigma_s \in \Sigma_{2,=}(i) \cup \Sigma_{3,<}(i) \cup \Sigma_{3,\leq}(i)$ for some integers $r < s$ and i. Then the sequence of morphisms $(\sigma'_n)_{n\in\mathbb{N}}$ obtained from $(\sigma_n)_{n\in\mathbb{N}}$ by replacing $\sigma_r \cdots \sigma_s$ by a composition of morphisms in $\mathcal{S}_3$ which is equal to $\beta_0\sigma_r \cdots \sigma_s$ is also an $\mathcal{S}_3$-adic representation of F.

Thus we only need to show that $\beta_0(\Sigma_{2,=}(i))$, $\beta_0(\Sigma_{3,<}(i))$ and $\beta_0(\Sigma_{3,\leq}(i))$ are rational subsets of $\mathcal{S}_3^*$.

$$\begin{aligned}
\beta_0(\Sigma_{2,=}(i)) &= \{[01^k, 021^k, 021^{k-1}] \mid k \geq i\}, \quad i \geq 1\\
&= \tilde{\alpha}_{2,0}(\alpha_{0,1}\alpha_{2,1})^*(\alpha_{0,1}\alpha_{2,1})^{i-1}\tilde{\alpha}_{1,2}, \quad i \geq 1\\
\beta_0(\Sigma_{3,<}(i))[2,1,0] &= \{[01^{k-1}, 01^k, 021^\ell] \mid k-1 > \ell \geq i\}, \quad i \geq 0\\
&= \tilde{\alpha}_{2,0}\alpha_{0,1}^+(\alpha_{0,1}\alpha_{2,1})^*(\alpha_{0,1}\alpha_{2,1})^i\tilde{\alpha}_{1,0}, \quad i \geq 0\\
\beta_0(\Sigma_{3,\leq}(i))[2,1,0] &= \{[01^{k-1}, 01^k, 021^\ell] \mid k-1 \geq \ell \geq i\}, \quad i \geq 0\\
&= \tilde{\alpha}_{2,0}\alpha_{0,1}^*(\alpha_{0,1}\alpha_{2,1})^*(\alpha_{0,1}\alpha_{2,1})^i\tilde{\alpha}_{1,0}, \quad i \geq 0
\end{aligned}$$

The above equations conclude the proof. ∎

7 Maximal bifix decoding

In this section, we state and prove the main result of this paper (Theorem 7.1). In the first part, we prove two results concerning morphisms onto a finite group. In the second one we prove a sequence of lemmas leading to a proof of the main result.

7.1 Main result

The family of uniformly recurrent tree sets contains both the Sturmian sets and the regular interval exchange sets. The second family is closed under maximal bifix decoding (see [5], Corollary 5.22) but the first family is not (see Example 7.2 below). The following result shows that the family of uniformly recurrent tree sets is a natural closure of the family of Sturmian sets.

Theorem 7.1 *The family of uniformly recurrent tree sets is closed under maximal bifix decoding.*

Note that, in contrast with Theorem 5.5, assuming the uniform recurrence, instead of simply the recurrence, implies the same property for the decoding.

We illustrate Theorem 7.1 by the following example.

Example 7.2 Let F be the Tribonacci set on the alphabet $A = \{a, b, c\}$ (see Example 2.2). Let $X = A^2 \cap F$. Then $X = \{aa, ab, ac, ba, ca\}$ is an F-maximal bifix code of F-degree 2. Let $B = \{x, y, z, t, u\}$ and let $f : B^* \to A^*$ be the morphism defined by $f(x) = aa$, $f(y) = ab$, $f(z) = ac$, $f(t) = ba$, $f(u) = ca$. Then f is a coding morphism for X. The set $G = f^{-1}(F)$ is a uniformly recurrent tree set by Theorem 7.1. It is not Sturmian since y and t are two right-special words of length 1. It is not either an interval exchange set. Indeed, for any right-special word w of G, one has $r(w) = 3$. This is not possible in a regular interval exchange set T since, Σ_T being injective, the length of the interval J_w tends to 0 as $|w|$ tends to infinity.

We prove two preliminary results concerning the restriction to a uniformly recurrent tree set of a morphism onto a finite group (Propositions 7.3 and 7.5).

Proposition 7.3 *Let F be a uniformly recurrent tree set containing the alphabet A and let $\varphi : A^* \to G$ be a morphism from A^* onto a finite group G. Then $\varphi(F) = G$.*

Proof. Since the submonoid $\varphi^{-1}(1)$ is right and left unitary, there is a bifix code Z such that $Z^* = \varphi^{-1}(1)$. Let $X = Z \cap F$. By Theorem 5.7, X is a basis of a subgroup of index $\mathrm{Card}(G)$. Let x be a word of X of maximal length. Then x is not an internal factor of X and thus it has $\mathrm{Card}(G)$ parses. Let S be the set of suffixes of x which are prefixes of X. If $s, t \in S$, 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 $u = 1$ since s is a prefix of X. Thus all elements of S have distinct images by φ. Since S has $\mathrm{Card}(G)$ elements, this forces $\varphi(S) = G$ and thus $\varphi(F) = G$ since $S \subset F$. ■

We illustrate the proof on the following example.

Example 7.4 Let $A = \{a, b\}$ and let φ be the morphism from A^* onto the symmetric group G on 3 elements defined by $\varphi(a) = (12)$ and $\varphi(b) = (13)$. Let Z be the group code such that $Z^* = \varphi^{-1}(1)$. The group automaton corresponding to the regular representation of G is represented in Figure 7.1. Let F be the Fibonacci set. The code $X = Z \cap F$ is represented in Figure 7.2. The word

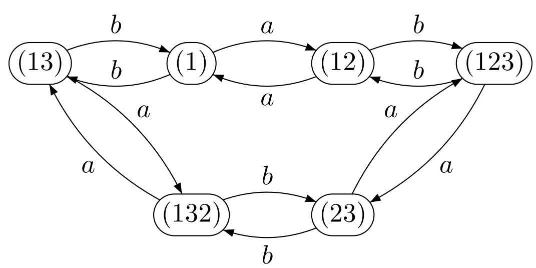

Figure 7.1: The group automaton corresponding to the regular representation of G.

$w = ababa$ is not an internal factor of X. All its 6 suffixes (indicated in black in Figure 7.2) are proper prefixes of X and their images by φ are the 6 elements of the group G.

Proposition 7.5 *Let F be a uniformly recurrent tree set containing the alphabet A and let $\varphi : A^* \to G$ be a morphism from A^* onto a finite group G. For any $w \in F$, one has $\varphi(\Gamma_F(w) \cup \{1\}) = G$.*

Proof. Let $\alpha : B^* \to A^*$ be a coding morphism for $\mathcal{R}_F(w)$. Then $\beta = \varphi \circ \alpha : B^* \to G$ is a morphism from B^* into G. By Theorem 5.8, the set $\mathcal{R}_F(w)$ is a basis of the free group on A. Thus $\langle \alpha(B) \rangle = \langle A \rangle$. This implies that $\beta(\langle B \rangle) = G$.

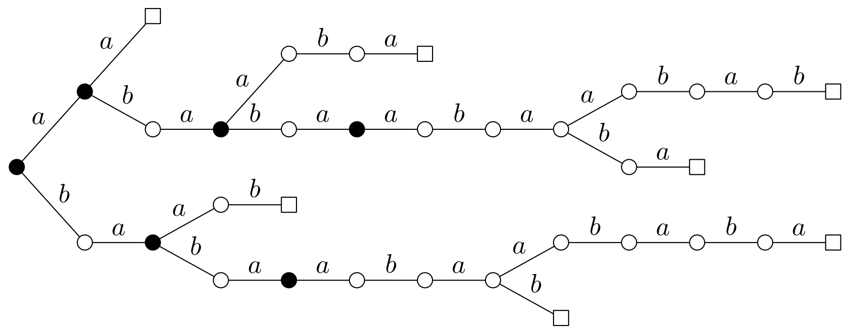

Figure 7.2: The code $X = Z \cap F$

This implies that $\beta(B)$ generates G. Since G is a finite group, $\beta(B^*)$ is a subgroup of G and thus $\beta(B^*) = G$. By Theorem 5.10, the set $H = \alpha^{-1}(w^{-1}F)$ is a uniformly recurrent tree set. Thus $\beta(H) = G$ by Proposition 7.3. This implies that $\varphi(\Gamma_F(w) \cup \{1\}) = G$. ■

7.2 Proof of the main result

Let F be a uniformly recurrent tree set containing A and let $f : B^* \to A^*$ be a coding morphism for a finite F-maximal bifix code Z. By Theorem 5.6, Z is a basis of a subgroup of index $d_F(Z)$ and, by Theorem 5.9, the submonoid Z^* is saturated in F.

We first prove the following lemma.

Lemma 7.6 *Let F be a uniformly recurrent tree set containing A and let $f : B^* \to A^*$ be a coding morphism for an F-maximal bifix code Z. The set $K = f^{-1}(F)$ is recurrent.*

Proof. Since F is factorial, the set K is factorial. Let $r, s \in K$. Since F is recurrent, there exists $u \in F$ such that $f(r)uf(s) \in F$. Set $t = f(r)uf(s)$. Let G be the representation of $\langle A \rangle$ on the right cosets of $\langle Z \rangle$. Let $\varphi : A^* \to G$ be the natural morphism from A^* onto G. By Proposition 7.5, we have $\varphi(\Gamma_F(t) \cup \{1\}) = G$. Let $v \in \Gamma_F(t)$ be such that $\varphi(v)$ is the inverse of $\varphi(t)$. Then $\varphi(tv)$ is the identity of G and thus $tv \in \langle Z \rangle$.

Since F is a tree set, it is acyclic and thus Z^* is saturated in F by Theorem 5.9. Thus $Z^* \cap F = \langle Z \rangle \cap F$. This implies that $tv \in Z^*$. 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 $q \in F$. Since Z^* is right unitary, $f(r), f(r)uf(s)v \in Z^*$ imply $uf(s)v = qf(s) \in Z^*$. In turn, since Z^* is left unitary, $qf(s), f(s) \in Z^*$ imply $q \in Z^*$ and thus $q \in Z^* \cap F$. Let $w \in K$ be such that $f(w) = q$. Then rws is in K. This shows that K is recurrent. ■

We prove a series of lemmas. In each of them, we consider a uniformly recurrent tree set F containing A and a coding morphism $f : B^* \to A^*$ for an F-maximal bifix code Z. We set $K = f^{-1}(F)$. We choose $w \in K$ and set $v = f(w)$. Let also $Y = R_K(w)$. Then Y is a $w^{-1}K$-maximal prefix code. Let $X = f(Y)$ or equivalently $X = Y \circ_f Z$. Then, since $f(w^{-1}K) = v^{-1}F$, by Proposition 2.9 (i), X is a $v^{-1}F$-maximal prefix code.

Finally we set $U = \mathcal{R}_F(v)$. Let $\alpha : C^* \to A^*$ be a coding morphism for U. Since $X \subset \Gamma_F(v)$, we have $X \subset U^*$. Since $uU^* \cap X \neq \emptyset$ for any $u \in U$, we have $\mathrm{alph}_U(X) = U$. Thus, by Proposition 2.8, we have $X = T \circ_\alpha U$ where T is the prefix code such that $\alpha(T) = X$.

Lemma 7.7 *We have* $X^* \cap v^{-1}F = U^* \cap Z^* \cap v^{-1}F$.

Proof. Indeed, the left handside is clearly included in the right one. Conversely, consider $x \in U^* \cap Z^* \cap v^{-1}F$. Since $x \in U^* \cap v^{-1}F$, $\alpha^{-1}(x)$ is in $\alpha^{-1}(v^{-1}F) = \alpha^{-1}(\Gamma_F(v)) \cup \{1\}$ by Proposition 4.3. Thus $x \in \Gamma_F(v) \cup \{1\}$. Since $x \in Z^*$, $f^{-1}(x) \in \Gamma_K(w) \cup \{1\} \subset Y^*$. Therefore x is in $f(Y^*) = X^*$. ∎

We set for simplicity $d = d_F(Z)$. Set $H = \alpha^{-1}(v^{-1}F)$. By Proposition 5.10, H is a uniformly recurrent tree set.

Lemma 7.8 *The set T is a finite H-maximal bifix code and $d_H(T) = d$.*

Proof. Since X is a prefix code, T is a prefix code. Since X is $v^{-1}F$-maximal, T is $\alpha^{-1}(v^{-1}F)$-maximal by Proposition 2.9 (ii) and thus H-maximal since $H = \alpha^{-1}(v^{-1}F)$.

Let $x, y \in C^*$ be such that $xy, y \in T$. Then $\alpha(xy), \alpha(y) \in X$ imply $\alpha(x) \in Z^*$. Since on the other hand, $\alpha(x) \in U^* \cap v^{-1}F$, we obtain by Lemma 7.7 that $\alpha(x) \in X^*$. This implies $x \in T^*$ and thus $x = 1$ since T is a prefix code. This shows that T is a suffix code.

To show that $d_H(T) = d$, we consider the morphism φ from A^* onto the group G which is the representation of $\langle A \rangle$ on the right cosets of $\langle Z \rangle$. Set $J = \varphi(Z^*)$. Thus J is a subgroup of index d of G. By Theorem 5.8, the set U is a basis of the free group on A. Therefore, since G is a finite group, the restriction of φ to U^* is surjective. Set $\psi = \varphi \circ \alpha$. Then $\psi : C^* \to G$ is a morphism which is onto since $U = \alpha(C)$ generates the free group on A. Let V be the group code of degree d such that $V^* = \psi^{-1}(J)$. Then $T = V \cap H$, as we will show now.

Indeed, set $W = V \cap H$. If $t \in T$, then $\alpha(t) \in X$ and thus $\alpha(t) \in Z^*$. Therefore $\psi(t) \in J$ and $t \in V^*$. This shows that $T \subset W^*$. Conversely, if $t \in W$, then $\psi(t) \in J$ and thus $\alpha(t) \in Z^*$. Since on the other hand $\alpha(t) \in U^* \cap F$, we obtain $\alpha(t) \in X^*$ by Lemma 7.7. This implies $t \in T^*$ and shows that $W \subset T^*$.

Thus, since H is a uniformly recurrent tree set, by Theorem 5.7, T is a basis of a subgroup of index d. Thus $d_H(T) = d$ by Theorem 5.6. ∎

Lemma 7.9 *The set Y is finite.*

Proof. Since T and U are finite, the set $X = T \circ U$ is finite. Thus $Y = f^{-1}(X)$ is finite. ∎

Proof of Theorem 7.1. Let F be a uniformly recurrent tree set containing A and let $f : B^* \to A^*$ be a coding morphism for a finite F-maximal bifix code Z. Set $K = f^{-1}(F)$.

By Lemma 7.6, K is recurrent. By Lemma 7.9 any set of first return words $Y = R_K(w)$ is finite. Thus K is uniformly recurrent. By Theorem 5.5, K is a tree set.

Thus we conclude that K is a uniformly recurrent tree set. ∎

Note that since K is a uniformly recurrent tree set, the set Y is not only finite as asserted in Lemma 7.9 but in fact a basis of the free group on B, by Theorem 5.8.

We illustrate the proof with the following example.

Example 7.10 Let F be the Fibonacci set on $A = \{a, b\}$ and let $Z = F \cap A^2 = \{aa, ab, ba\}$. Thus Z is an F-maximal bifix code of F-degree 2. Let $B = \{c, d, e\}$ and let $f : B^* \to A^*$ be the coding morphism defined by $f(c) = aa$, $f(d) = ab$ and $f(e) = ba$. Part of the set $K = f^{-1}(F)$ is represented in Figure 7.3 on the left.

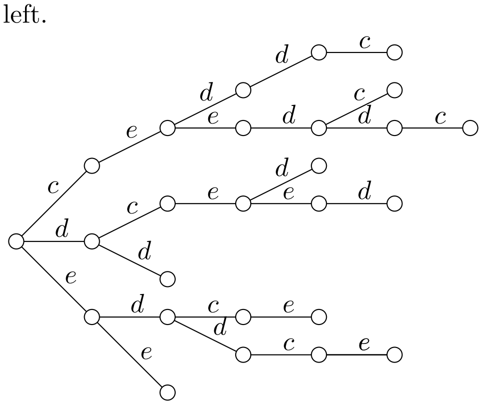

Figure 7.3: The sets K and H.

The set $Y = R_K(c)$ and $X = f(Y)$ are

$$Y = \{eddc, eedc, eeddc\}, \quad X = \{baababaa, babaabaa, babaababaa\}.$$

On the other hand, the set $U = \mathcal{R}_F(aa)$ is $U = \{baa, babaa\}$. Let $C = \{r, s\}$ and let $\alpha : C^* \to A^*$ be the coding morphism for U defined by $\alpha(r) = baa$, $\alpha(s) = babaa$. Part of the set $H = \alpha^{-1}((aa)^{-1}F)$ is represented in Figure 7.3 on the right. Then we have $T = \{rs, sr, ss\}$ which is an H-maximal bifix code of H-degree 2 in agreement with Lemma 7.8.

The following example shows that the condition that F is a tree set is necessary.

Example 7.11 Let F be the set of factors of $(ab)^*$. The set F does not satisfy the tree condition since $G(\epsilon)$ is not connected. Let $X = \{ab, ba\}$. The set X is a finite F-maximal bifix code. Let $f : \{u, v\}^* \to A^*$ be the coding morphism for X defined by $f(u) = ab$, $f(v) = ba$. Then $f^{-1}(F) = u^* \cup v^*$ is not recurrent.

7.3 Composition of bifix codes

In this section, we use Theorem 7.1 to prove a result showing that in a uniformly recurrent tree set, the degrees of the terms of a composition of maximal bifix codes are multiplicative (Theorem 7.12).

The following result is proved in [3] for a more general class of codes (including all finite codes and not only finite bifix codes), but in the case of $F = A^*$ (Proposition 11.1.2).

Theorem 7.12 *Let F be a uniformly recurrent tree set and let $X, Z \subset F$ be finite bifix codes such that X decomposes into $X = Y \circ_f Z$ where f is a coding morphism for Z. Set $G = f^{-1}(F)$. Then X is an F-maximal bifix code if and only if Y is a G-maximal bifix code and Z is an F-maximal bifix code. Moreover, in this case*

$$d_F(X) = d_G(Y)d_F(Z). \tag{7.1}$$

Proof. Assume first that X is an F-maximal bifix code. By Proposition 2.9 (ii), Y is a G-maximal prefix code and Z is an F-maximal prefix code. This implies that Y is a G-maximal bifix code and that Z is an F-maximal bifix code.

The converse also holds by Proposition 2.9.

To show Formula (7.1), let us first observe that there exist words $w \in F$ such that for any 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 $w \in F$ 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 7.1, the set G is a uniformly recurrent tree set and thus in particular, it is recurrent.

Let $w \in F$ 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 G is recurrent, there are $d_G(Y)$ such triples. This shows Formula (7.1). ∎

Example 7.13 Let F 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 $G = f^{-1}(F)$. The words of length at most 3 of G are represented on Figure 7.4.

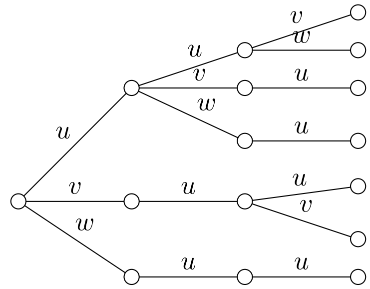

Figure 7.4: The words of length at most 3 in G.

The set $Z = f(B)$ is an F-maximal bifix code of F-degree 2 (it is the unique F-maximal bifix code of F-degree 2 with kernel $\{a\}$). Let $Y = \{uu, uvu, uw, v, wu\}$, which is a G-maximal bifix code of G-degree 2 (it is the unique G-maximal bifix code of G-degree 2 with kernel $\{v\}$).

The code $X = f(Y)$ is the F-maximal bifix code of F-degree 4 shown on Figure 7.5.

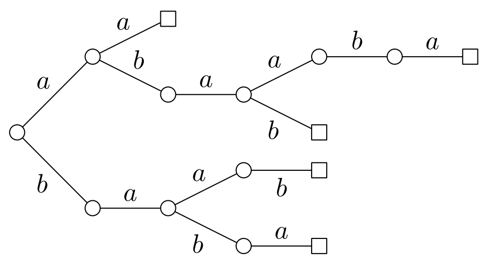

Figure 7.5: An F-maximal bifix code of F-degree 4.

Example 7.14 shows that Formula (7.1) does not hold if F is not a tree set.

Example 7.14 Let $F = F(ab)^*$ (see Example 7.11). 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 F-maximal bifix codes and $d_F(Z) = 2$. We have $d_F(X) = 3$ since $abab$ has three parses. Thus $d_F(Z)$ does not divide $d_F(X)$.

References

[1] V. I. Arnold. Small denominators and problems of stability of motion in classical and celestial mechanics. *Uspehi Mat. Nauk*, 18(6 (114)):91–192, 1963. 2

[2] Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012. 2, 3, 4, 5, 6, 7, 8, 17

[3] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2009. 5, 6, 8, 9, 34

[4] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Acyclic, connected and tree sets. 2013. 2, 3, 15, 16, 17, 18

[5] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. The finite index basis property. 2013. 2, 3, 16, 17, 29

[6] Valérie Berthé, Clelia De Felice, Francesco Dolce, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Two-sided Rauzy induction. 2013. 15

[7] Valérie Berthé and Michel Rigo. *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.* Cambridge Univ. Press, Cambridge, 2010. 10, 12

[8] 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. 11

[9] Michael Boshernitzan. A unique ergodicity of minimal symbolic flows with linear block growth. *J. Analyse Math.*, 44:77–96, 1984/85. 16

[10] Julien Cassaigne, Sébastien Ferenczi, and Ali Messaoudi. Weak mixing and eigenvalues for Arnoux-Rauzy sequences. *Ann. Inst. Fourier (Grenoble)*, 58(6):1983–2005, 2008. 3

[11] 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. 10

[12] Ethan M. Coven and G. A. Hedlund. Sequences with minimal block growth. *Math. Systems Theory*, 7:138–153, 1973. 2

[13] F. Durand, J. Leroy, and G. Richomme. Do the properties of an S-adic representation determine factor complexity? *J. Integer Seq.*, 16(2):Article 13.2.6, 30, 2013. 23

[14] Fabien Durand. A characterization of substitutive sequences using return words. *Discrete Math.*, 179(1-3):89–101, 1998. 14, 22

[15] Fabien Durand. Linearly recurrent subshifts have a finite number of non-periodic subshift factors. *Ergodic Theory Dynam. Systems*, 20(4):1061–1078, 2000. 22

[16] Fabien Durand. Corrigendum and addendum to: "Linearly recurrent subshifts have a finite number of non-periodic subshift factors" [Ergodic Theory Dynam. Systems **20** (2000), no. 4, 1061–1078; MR1779393 (2001m:37022)]. *Ergodic Theory Dynam. Systems*, 23(2):663–669, 2003. 22

[17] Sébastien Ferenczi. Rank and symbolic complexity. *Ergodic Theory Dynam. Systems*, 16(4):663–682, 1996. 21

[18] Sébastien Ferenczi and Luca Q. Zamboni. Languages of k-interval exchange transformations. *Bull. Lond. Math. Soc.*, 40(4):705–714, 2008. 16

[19] 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. 2, 5

[20] Amy Glen and Jacques Justin. Episturmian words: a survey. *Theor. Inform. Appl.*, 43:403–442, 2009. 2

[21] Jacques Justin and Laurent Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5):343–356, 2000. 3, 5, 18

[22] Michael Keane. Interval exchange transformations. *Math. Z.*, 141:25–31, 1975. 2, 11, 12

[23] J. Leroy. An S-adic characterization of minimal subshifts with first difference of complexity $1 \leq p(n+1) - p(n) \leq 2$. *ArXiv e-prints*, May 2013. 24, 26

[24] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002. 2, 5, 12

[25] Marston Morse and Gustav A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.*, 62:1–42, 1940. 2

[26] V. I. Oseledec. The spectrum of ergodic automorphisms. *Dokl. Akad. Nauk SSSR*, 168:1009–1011, 1966. 2

[27] 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. 19, 20

[28] Anton Zorich. Deviation for interval exchange transformations. *Ergodic Theory Dynam. Systems*, 17(6):1477–1499, 1997. 3

A Labels in Figure 6.1

In this section we provide the full list of labels of the edges in Figure 6.1. To shorten the presentation, for all words u, v, w, we let $[u, v, w]$ denote the morphism defined by $\sigma(0) = u$, $\sigma(0) = v$ and $\sigma(2) = w$. We also define the following notations. Finally, when we use the letters x, y and z to denote the morphism $[x, y, z]$, it is understood that $\{x, y, z\} = \{0, 1, 2\}$. In the same way, when one of the letters x, y, z is fixed to some $i \in \{0, 1, 2\}$, the others are understood such that one still has $\{x, y, z\} = \{0, 1, 2\}$. For instance, in Table A.3, the morphism $[0, x, y]$ can be $[0, 1, 2]$ or $[0, 2, 1]$.

$$\begin{aligned}
\Sigma_1(i) &= \{[0, 1^k2, 1^{k-1}2] \mid k \geq i\} \\
\Sigma_{2,=}(i) &= \{[1^k0, 21^k0, 21^{k-1}0] \mid k \geq i\} \\
\Sigma_{2,-}(i) &= \{[1^{k-1}0, 21^k0, 21^{k-1}0] \mid k \geq i\} \\
\Sigma_{2,<}(i) &= \{[1^\ell 0, 21^k0, 21^{k-1}0] \mid k-1 > \ell \geq i\} \\
\Sigma_{2,\leq}(i) &= \{[1^\ell 0, 21^k0, 21^{k-1}0] \mid k-1 \geq \ell \geq i\} = \Sigma_{2,-}(i) \cup \Sigma_{2,<}(i) \\
\Sigma_{2,0}(i) &= \{[0, 21^k0, 21^{k-1}0] \mid k \geq i\} \\
\Sigma_{3,=}(i) &= \{[21^k0, 1^k0, 1^{k-1}0] \mid k \geq i\} \\
\Sigma_{3,-}(i) &= \{[21^{k-1}0, 1^k0, 1^{k-1}0] \mid k \geq i\} \\
\Sigma_{3,<}(i) &= \{[21^\ell 0, 1^k0, 1^{k-1}0] \mid k-1 > \ell \geq i\} \\
\Sigma_{3,\leq}(i) &= \{[21^\ell 0, 1^k0, 1^{k-1}0] \mid k-1 \geq \ell \geq i\} = \Sigma_{3,-}(i) \cup \Sigma_{3,<}(i) \\
\Sigma_{3,0}(i) &= \{[20, 1^k0, 1^{k-1}0] \mid k \geq i\}
\end{aligned} \tag{A.1}$$

$2 \to 2$	$\{\alpha_{1,0}\alpha_{2,0},\ \alpha_{0,1}\alpha_{2,1},\ \alpha_{0,2}\alpha_{1,2}\}$
$2 \to V_0$	$\{\alpha_{2,0}\alpha_{1,2},\ \alpha_{1,0}\alpha_{2,1}\}$
$2 \to V_1$	$\{\alpha_{0,1}\alpha_{2,0},\ \alpha_{2,1}\alpha_{0,2}\}$
$2 \to V_2$	$\{\alpha_{0,2}\alpha_{1,0},\ \alpha_{1,2}\alpha_{0,1}\}$
$2 \to 4B$	$[x, y, z]\,(\{\alpha_{1,0}\alpha_{2,1},\ \alpha_{2,0}\alpha_{1,2}\} \cup \Sigma_{2,-}(2) \cup \Sigma_{3,-}(2))$
$2 \to 7/8$	$[x, y, z]\,(\Sigma_{2,0}(2) \cup \Sigma_{2,<}(1) \cup \Sigma_{3,<}(1) \cup \alpha_{1,2}(\Sigma_{3,<}(0) \cup \Sigma_{2,\leq}(1))$ $\cup \tilde{\alpha}_{2,1}\alpha_{1,0}\Sigma_1(1) \cup \{\alpha_{2,0},\ \alpha_{0,2}\}\Sigma_1(2))$
$2 \to 10B$	$[x, y, z]\,(\Sigma_{2,=}(1) \cup \Sigma_{3,=}(2) \cup \alpha_{1,2}(\Sigma_{2,=}(2) \cup \Sigma_{3,-}(2)))$

Table A.1: Labels of outgoing edges of vertex 2.

$V_x \to V_x$	$\alpha_{y,x}\alpha_{z,x}$
$V_x \to V_y$	$\alpha_{x,z}$
$V_x \to 7/8$	$[x,y,z](\Sigma_{2,0}(1) \cup \Sigma_{2,-}(1) \cup \Sigma_{3,<}(0) \cup \Sigma_{2,<}(1)) \cup [y,x,z]\Sigma_1(2)$
$V_x \to 10B$	$\tilde{\alpha}_{z,x}\alpha_{x,y}[y,x,z] \cup [x,y,z]\,(\Sigma_{2,=}(1) \cup \Sigma_{3,-}(2))$

Table A.2: Labels of outgoing edges of vertices V_x.

$4B \to 4B$	$\{\alpha_{1,0}\alpha_{2,0}\} \cup [x,y,0]\,(\Sigma_{2,-}(1) \cup \Sigma_{3,-}(1))$
$4B \to 7/8$	$[0,x,y]\alpha_{2,0}\Sigma_1(1) \cup [x,y,0]\,(\Sigma_{2,<}(0) \cup \Sigma_{3,<}(0) \cup \alpha_{1,0}\Sigma_{2,\leq}(0))$
$4B \to 10B$	$[x,y,0]\,(\Sigma_{2,=}(1) \cup \Sigma_{3,=}(1) \cup \alpha_{1,0}(\Sigma_{2,=}(1) \cup \Sigma_{3,-}(1)))$

Table A.3: Labels of outgoing edges of vertex $4B$.

$5/6 \to 5/6$	$[2,0,1]\,(((\Sigma_{2,-}(1) \cup \Sigma_{3,-}(1))[1,0,2]) \cup ((\Sigma_{2,=}(1) \cup \Sigma_{3,=}(1))[2,0,1]))$
$5/6 \to 7/8$	$[1,0,2]\Sigma_1(1) \cup [0,2,1](\Sigma_{2,\leq}(0) \cup \Sigma_{3,\leq}(1) \cup \Sigma_{3,=}(1))$
$5/6 \to 10B$	$\alpha_{0,1}[1,0,2] \cup [0,2,1](\Sigma_{2,=}(1) \cup \Sigma_{3,-}(1))$

Table A.4: Labels of outgoing edges of vertex 5/6.

$7/8 \to 5/6$	$\tilde{\alpha}_{x,0}\alpha_{0,y}[x,y,0] \cup \alpha_{0,y}[x,0,y]$
$7/8 \to 7/8$	$\alpha_{1,0}\alpha_{2,0}$

Table A.5: Labels of outgoing edges of vertex 7/8.

$10B \to 5/6$	$(\Sigma_{2,-}(1) \cup \Sigma_{3,-}(1))[1,0,2] \cup (\Sigma_{2,=}(1) \cup \Sigma_{3,=}(1))[2,0,1]$
$10B \to 7/8$	$[0,2,1]\Sigma_1(1) \cup [2,1,0](\Sigma_{2,\leq}(0) \cup \Sigma_{3,<}(0))$
$10B \to 10B$	$\alpha_{2,0}[0,2,1] \cup [2,1,0](\Sigma_{2,=}(1) \cup \Sigma_{3,-}(1))$

Table A.6: Labels of outgoing edges of vertex $10B$.

B Labels in Figure 6.2

We only given the labels of Figure 6.2 that are not labels of Figure 6.1. We use the same notation as in the previous appendix. Recall that M is the morphism defined by $M(0) = M(2) = 0$ and $M(1) = 1$.

$1 \to 1$	$\{[01,1],[1,01],[0,10],[10,0]\}$
$1 \to 7/8$	$[x,y,2]M\Sigma_1(2)$

Table B.1: Labels of outgoing edges of vertex 1.

$2 \to 1$	$[x,y,z]\{\alpha_{2,0}\alpha_{1,2},\ \alpha_{2,1}\alpha_{0,2},\ \alpha_{0,2}\alpha_{1,2},\ \alpha_{0,2}\alpha_{1,0},\ \alpha_{1,2}\alpha_{0,1}\}$ $\cup[x,y,z](\Sigma_{2,=}(2) \cup \Sigma_{2,-}(2) \cup \Sigma_{3,=}(2) \cup \Sigma_{3,-}(2))$ $\cup[x,y,z]\alpha_{1,2}(\Sigma_{2,=}(1) \cup \Sigma_{3,=}(1) \cup \Sigma_{3,-}(2))$
$2 \to 7/8$	$[x,y,z]\{\alpha_{1,2},\ \alpha_{0,2}\}M\Sigma_1(2)$
$V_x \to 1$	$\{\alpha_{y,x}\alpha_{z,i}[y,z,x],\ \alpha_{x,z}[x,y,z],\ \alpha_{x,z}[y,x,z]\}$ $\cup[x,y,z](\Sigma_{2,=}(1) \cup \Sigma_{2,-}(2) \cup \Sigma_{3,=}(1))$
$4B \to 1$	$\{\mathrm{id}\}$ $\cup[x,y,0](\Sigma_{2,=}(1) \cup \Sigma_{2,-}(1) \cup \Sigma_{3,=}(1) \cup \Sigma_{3,-}(1))$ $\cup[x,y,0]\alpha_{1,2}(\Sigma_{2,=}(1) \cup \Sigma_{3,=}(1) \cup \Sigma_{3,-}(1))$
$5/6 \to 1$	$[x,y,z]\{\mathrm{id},\ \alpha_{0,1},\ \alpha_{1,0}\}$ $\cup[0,2,1](\Sigma_{2,=}(1) \cup \Sigma_{3,=}(1) \cup \Sigma_{3,-}(1))$
$7/8 \to 1$	$[x,y,z]\{\mathrm{id},\ \alpha_{0,1},\ \alpha_{1,0}\}$
$10B \to 1$	$[2,1,0](\Sigma_{2,=}(1) \cup \Sigma_{3,=}(1) \cup \Sigma_{3,-}(1))$

Table B.2: Labels of incoming edges of vertex 1.