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

September 21, 2014 22 h 39

Abstract

We consider 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 [24]. Many combinatorial properties were described in the paper by Coven and Hedlund [11].

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. A Sturmian sets is the set of factors of one Sturmian word. For more details, see [19, 23].

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.

Interval exchange transformations were introduced by Oseledec [25] 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 set of factors of the natural codings of a regular interval exchange transformation is called an interval exchange set.

Even though 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 [9, 27]).

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.

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 S of words on an alphabet A, a finite bifix code is S-maximal if and only if it is the basis of a subgroup of finite index of the free group on A. 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 S is a uniformly recurrent tree set on the alphabet A, then for any $n \geq 1$, the set $S \cap A^n$ is a basis of the subgroup formed by the words of length multiple of n (see Theorem 5.9).

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 S is a uniformly recurrent tree set and f a coding morphism for a finite S-maximal bifix code, then $f^{-1}(S)$ is a uniformly recurrent tree set. The family of regular interval exchange sets is closed under maximal bifix decoding (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 5.12). This property, which is interesting in its own, generalizes the fact that the derived word of a Sturmian word is Sturmian [21].

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). We prove in [**?**] that the class of regular interval exchange sets is closed under maximal bifix decoding.

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 5.4). We prove that the family of uniformly recurrent tree sets is closed under derivation (Theorem 5.12). 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 5.18).

In Section 6, we turn to the notion of H-adic representation of sets, introduced in [17], using a terminology initiated by Vershik and coined out by B. Host (it is usually called S-adic). We deduce from the previous result that uniformly recurrent tree sets have a primitive H_e-adic representation (Theorem 6.5) where H_e is the finite set of positive elementary automorphisms of the free group.

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 let A^* denote the set of all finite words over A and A^+ the set of finite nonempty words over A. The empty word is denoted by 1 or by ε . We let $|w|$ denote 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 word v is a *factor* of a word x if $x = uvw$. A set of words is said to be *factorial* if it contains the factors of its elements. Let S be a set of words on the alphabet A. For $w \in S$, we denote

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

and further

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

These notions depend upon S but it is assumed from the context. 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 S is called *right-extendable* (resp. *left-extendable*, resp. *biextendable*) if every word in S is right-extendable (resp. left-extendable, resp. biextendable).

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

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

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

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 biextendable.

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

A *morphism* f from 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 *fixed point* of the morphism f.

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

exampleFibonacci

Example 2.1 Let $A = \{a, b\}$. The Fibonacci word is the fixed point $x = abaababa\ldots$ of the morphism $f : A^* \to A^*$ defined by $f(a) = ab$ and $f(b) = a$. It is a Sturmian word (see [23]). The set $\mathrm{Fac}(x)$ of factors of x is the *Fibonacci set.*

exampleTribonacci

Example 2.2 Let $A = \{a, b, c\}$. The Tribonacci word is the fixed point $x = f^\omega(a) = abacaba\cdots$ 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 S be a set of words. A prefix code $X \subset S$ is S-maximal if it is not properly contained in any prefix code $Y \subset S$[1]. Equivalently, a prefix code $X \subset S$ is S-maximal if any word in S 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 $X \subset S$ is *right S-complete* if every word of S is a prefix of a word in X^*. If S is factorial, a prefix code is S-maximal if and only if it is right S-complete [2, Proposition 3.3.2].

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

A *parse* of a word $w \in A^*$ 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 $d_X(w)$ 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 [3, Proposition 6.1.6].

By definition, the S-degree of a bifix code X, denoted $d_X(S)$, is the maximal number of parses of all words in S with respect to X. 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 S be a recurrent set and let X be a finite S-maximal bifix code of S-degree d. A word $w \in S$ is such that $d_X(w) < d$ if and only if it is an internal factor of X, that is

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

eqInternal

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

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

The *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 $K(X) = I(X) \cap X$.

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

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

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

2.3 Group codes

We let $\mathcal{A} = (Q, i, T)$ denote 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 (for a general introduction to automata theory, see [16] for example).

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 $u \in A^*$ (see Section 2.1 for the notation $u^{-1}L$). 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 $Q = \{1, 2, 3\}$,

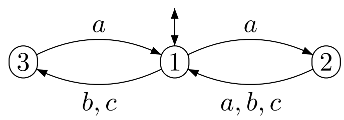

Figure 2.1: The minimal automaton of $\{aa, ab, ac, ba, ca\}^*$.

$i = 1$ 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].

propositionGroupAutomaton

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 propositionGroupAutomaton 2.6 is called a *group code* of degree d.

2.4 Composition of codes

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

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

eq1.6.1

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 [3, Proposition 2.6.1] and if Y and Z are prefix (suffix) codes, then X is a prefix (suffix) code. Conversely, if X is a prefix (suffix) code, then Y is a prefix (suffix) code.

We extend the notation alph as follows. For two codes $X, Z \subset A^*$ we denote

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

The following is Proposition 2.6.6 in [3].

prop266 **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.

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

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

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

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

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

sectionIntervalExchange

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 [10] or [7]).

A *semi-interval* is a nonempty subset of the real line of the form $[\alpha, \beta) = \{z \in \mathbb{R} \mid \alpha \le z < \beta\}$. 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 $\mu_a = \sum_{b \le_1 a} \lambda_b$ and $\nu_a = \sum_{b \le_2 a} \lambda_b$. 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.

exampleRotation

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* with T.

Let $s \geq 2$ be an integer. 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 $\pi = (abc)$.

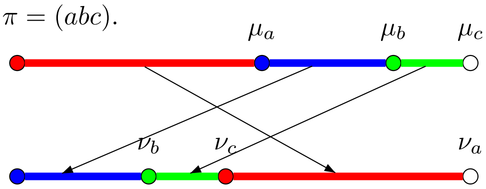

Figure 3.1: A 3-interval exchange transformation

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 $1 \leq i \leq s$.

There are several equivalent terms used instead of regular. 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 *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 $\Sigma_T(z) = a_0 a_1 \cdots$ 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 [23, Chapter 2] for example).

For a word $w = b_0 b_1 \cdots b_{m-1}$, let I_w be the set

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

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 $n \geq 0$

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

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 [23]). 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 $x = a_0 a_1 \cdots, y = b_0 b_1 \cdots \in A^\omega$ 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 σ be the shift on X. The pair (X, σ) is a *symbolic dynamical system*, formed of a topological space X and a continuous transformation σ. Such a system is said to be minimal if the only closed subsets invariant by σ are $\emptyset$ or X. It is well-known that (X, σ) is minimal if and only if the set $\mathrm{Fac}(X)$ of factors of the $x \in X$ is uniformly recurrent (see for example [23] Theorem 1.5.9).

We have the commutative diagram of Figure 3.2.

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

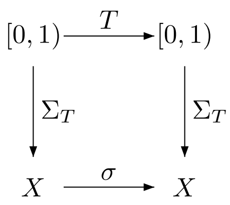

Figure 3.2: A commutative diagram. commutativeDiagram

(see [22] or [7] page 349). However, if the transformation T is minimal, the symbolic dynamical system (X, σ) is minimal (see [7] page 392). Thus, we obtain the following statement.

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

exampleDivision **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 figure3interval2 3.3). The transformation T is regular since α is irrational. The words

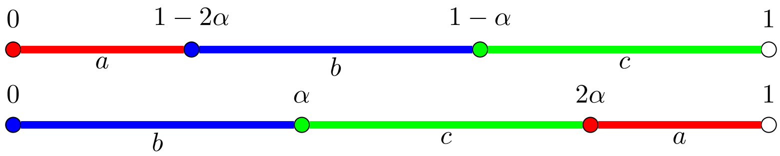

Figure 3.3: A regular 3-interval exchange transformation. figure3interval2

of length at most 5 of the set $S = \mathrm{Fac}(T)$ are represented in Figure figureSetF 3.4. Since

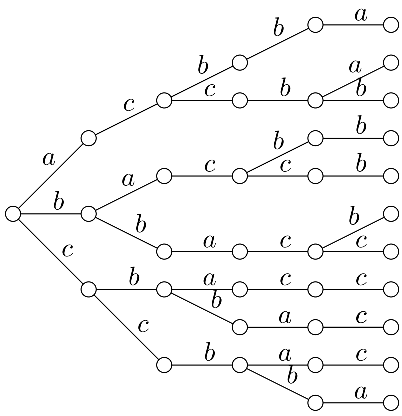

Figure 3.4: The words of length ≤ 5 of the set S. figureSetF

$T = R^2$, where R is the transformation of Example exampleFiboNatCoding 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

$$y = baccbaccbbacbbacbbacc\cdots \tag{3.3}$$

Eqy

Actually, the word y is the fixed-point $g^\omega(b)$ of the morphism $g : a \mapsto baccb, b \mapsto bacc, c \mapsto bacb$. This follows from the fact that the cube of the Fibonacci morphism $f : a \mapsto ab, b \mapsto a$ sends each letter on a word of odd length and thus preserves the set of words of even length.

4 Return words

sectionReturn

In this section, we introduce the notion of return and first return words. We prove elementary results about return words which extendablely already appear in [12].

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

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

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

eqAutomo

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

Example 4.1 Let S be the uniformly recurrent set of Example 3.7 (exampleDivision). We have

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

These sets can be read from the word y given in Equation (3.3) (Eqy). A coding morphism $f : B^* \to A^*$ with $B = A$ for the set $\mathcal{R}_S(c)$ is given by $f(a) = bac$, $f(b) = bbac$, $f(c) = c$.

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

$$\Gamma_S(w) \cup \{1\} = \mathcal{R}_S(w)^* \cap w^{-1}S. \tag{4.2}$$

eqGamma1

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

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

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

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

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

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

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

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

Let S be a recurrent set and x be an infinite word such that $S = \mathrm{Fac}(x)$. Let $w \in S$ and let f be a coding morphism for the set $\mathcal{R}_S(w)$. Since w appears infinitely often in x, there is a unique factorization $x = vwz$ with $z \in \mathcal{R}_S(w)^\omega$ 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 $\mathcal{R}'_S(w)$ associated with 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 S be a recurrent set and let x be a recurrent infinite word such that $S = \mathrm{Fac}(x)$. Let $w \in S$ and let f be a coding morphism for the set $\mathcal{R}_S(w)$. The derived set of S with respect to f is the set of factors of the derived word of x with respect to f, that is $D_f(S) = \mathrm{Fac}(D_f(x))$.*

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

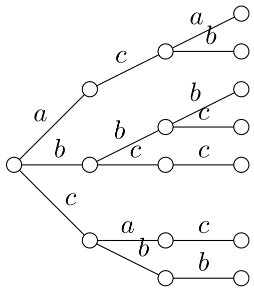

Figure 4.1: The words of length ≤ 3 of the derived set of S. figureDerived

5 Uniformly recurrent tree sets

sectionTreeNormal

In this section, we recall the notion of tree set introduced in [**?**]. We recall that the factor complexity of a tree set on $k+1$ letters is $p_n = kn + 1$.

We recall a result concerning the decoding of tree sets (Theorem 5.7). We also recall the finite index basis property of uniformly recurrent tree sets (Theorems **??** and 5.9) that we will use in Section 7. We prove that the family of uniformly recurrent tree sets is invariant under derivation (Theorem 5.12). We further prove that all bases of the free group included in a uniformly recurrent tree set are tame (Theorem 5.18).

5.1 Tree sets

Let S be a fixed factorial set. 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 S.

Example 5.1 Let S be the Fibonacci set. The extension graphs of ε, a, b, ab respectively are shown in Figure 5.1.

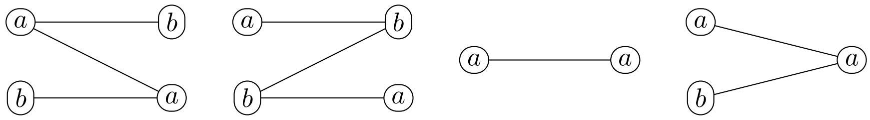

Figure 5.1: The extension graphs of ε, a, b, ab in the Fibonacci set. FigureExtensionGraph

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

We say that S is a *tree set* (resp. an acyclic set) if it is biextendable and if for every word $w \in S$, 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 5.2 *Let S be a tree set on the alphabet A and let $k = \mathrm{Card}(A \cap S) - 1$. Then $\mathrm{Card}(S \cap A^n) = kn + 1$ for all $n \geq 0$*

The following result is also easy to prove.

Proposition 5.3 *A Sturmian set S 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 $w \in S$. If w is not left-special there is a unique $a \in A$ such that $aw \in S$. 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 5.4 *A regular interval exchange set is a uniformly recurrent tree set.*

Proposition 5.4 is actually a particular case of a result of [18] which characterizes the regular interval exchange sets.

We give an example of a uniformly recurrent tree set which is neither a Sturmian set nor an interval exchange set.

Example 5.5 Let S be the Tribonacci set on the alphabet $A = \{a, b, c\}$ (see Example 2.2). Let $X = A^2 \cap S$. Then $X = \{aa, ab, ac, ba, ca\}$ is an S-maximal bifix code of S-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. We will see that the set $G = f^{-1}(S)$ is a uniformly recurrent tree set (this follows from Theorem 7.1 below). 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.

Let S be a set of words. For $w \in S$, and $U, V \subset S$, let $U(w) = \{\ell \in U \mid \ell w \in S\}$ and let $V(w) = \{r \in V \mid wr \in S\}$. 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 $\ell wr \in S$. 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 5.6 *Let S be a tree set. For any $w \in S$, any finite S-maximal suffix code $U \subset S$ and any finite S-maximal prefix code $V \subset S$, the generalized extension graph $G_{U,V}(w)$ is a tree.*

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

The following result is Theorem 4.13 in [4].

InverseImageTree **Theorem 5.7** *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

sectionNormal

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

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

theoremFIB **Theorem 5.8** *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 S in [2] (Theorem 7.2.5) but the proof only uses the fact that S is uniformly recurrent and satisfies the finite index basis property. We reproduce the proof for the sake of clarity.

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 $\langle A\rangle$.

theoremGroupCode **Theorem 5.9** *Let $Z \subset A^+$ be a group code of degree d. For every uniformly recurrent tree set S containing the alphabet A, the set $X = Z \cap S$ 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 S-maximal bifix code of S-degree $e \le d$. Since S is a uniformly recurrent, by Theorem 4.4.3 of [2], X is finite. By Theorem 5.8, 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 $e \le d$, this implies that $e = d$. ■

As an example of this result, if S is a uniformly recurrent tree set, then $S \cap A^n$ 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].

theoremJulien

Theorem 5.10 *Let S be a uniformly recurrent tree set containing the alphabet A. For any word $w \in S$, the set $\mathcal{R}_S(w)$ 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 S if $M \cap S = \langle M \rangle \cap S$.

propositionHcapE

Theorem 5.11 *Let S be an acyclic set. The submonoid generated by any bifix code $X \subset S$ is saturated in S.*

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]).

propositionReturns

Theorem 5.12 *Any derived set of a uniformly recurrent tree set is a uniformly recurrent tree set.*

Proof. Let S be a uniformly recurrent tree set containing A, let $v \in S$ and let f be a coding morphism for $X = \mathcal{R}_S(v)$. By Theorem 5.10, 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 $H = f^{-1}(v^{-1}S)$. By Proposition 4.3, the set H is recurrent and $H = f^{-1}(\Gamma_S(v)) \cup \{1\}$.

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

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

where $G_{X',X}(vy)$ denotes the generalized extension graph of vy relative to X', X. Indeed,

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

The set X' is an Sv^{-1}-maximal suffix code and the set X is a $v^{-1}S$-maximal prefix code. By Proposition 5.6 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 $\Gamma_H(x) = f^{-1}(\Gamma_S(vy))$ or equivalently $f(\Gamma_H(x)) = \Gamma_S(vy)$. Consider first $r \in \Gamma_H(x)$. Set $s = f(r)$. Then $xr = ux$ with $u, ux \in H$. Thus $ys = wy$ with $w = f(u)$.

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

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

Example 5.13 Let S be the Tribonacci set (see Example 2.2). It is the set of factors of the infinite word $x = abacaba\cdots$ which is the fixed point of the morphism f defined by $f(a) = ab$, $f(b) = ac$, $f(c) = a$. We have $\mathcal{R}_S(a) = \{a, ba, ca\}$. Let g be the coding morphism for $\mathcal{R}_S(a)$ defined by $g(a) = a$, $g(b) = ba$, $g(c) = ca$ and let g' be the associated coding morphism for $\mathcal{R}'_S(a)$. We have $f = g'\pi$ where 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 S with respect to a is the set $\pi(S)$.

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*[2] 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 [26]).

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 5.14 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 fact that X is a basis can be check directly by the fact that $c = (cba)(ba)^{-1}$, $c^{-2}(cca) = a$ and finally $(ba)a^{-1} = b$.

The following result will play a key role in the proof of the main result of this section (Theorem 5.18).

Proposition 5.15 *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 $X = \alpha_1\alpha_2\cdots\alpha_{n-1}(A)$

[2] The word *tame* (as opposed to *wild*) is used here on analogy with its use in ring theory.

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 5.16 *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 5.15 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 [26].

Example 5.17 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 5.16.

The following result is a remarkable consequence of Theorem 5.8.

Theorem 5.18 *Any basis of the free group included in a uniformly recurrent tree set is tame.*

Proof. Let S be a uniformly recurrent tree set. Let $X \subset S$ be a basis of the free group on A. Since A is finite, X is finite (and of the same cardinality as A). We use an induction on the sum $\lambda(X)$ of the lengths of the words of X. If X is bifix, by Theorem 5.8, it is an S-maximal bifix code of S-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 $\lambda(Y) < \lambda(X)$. By induction hypothesis, Y is tame. Since $X = (Y \setminus v) \cup uv$, X is tame by Proposition 5.15. ∎

Example 5.19 The set $X = \{ab, acb, acc\}$ is a basis of the free group which is not tame (see Example 5.17). Accordingly, the extension graph $G(\varepsilon)$ relative to the set of factors of X is not a tree (see Figure 5.2).

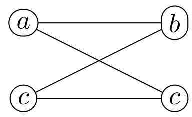

Figure 5.2: The graph $G(\varepsilon)$

6 H-adic representations

In this section we study H-adic representations of tree sets. This notion was introduced in [17], using a terminology initiated by Vershik and coined out by B. Host (it is usually called S-adic but we already use here the letter S for sets of words). We first recall a general construction allowing to build H-adic representations of any uniformly recurrent aperiodic set (Proposition 6.1) which is based on return words. Using Theorem 5.18, we show that this construction actually provides $\mathcal{H}_e$-representations of uniformly recurrent tree sets (Theorem 6.5), where $\mathcal{H}_e$ is the set of elementary positive automorphisms of the free group on A.

6.1 H-adic representations

Let H be a set of morphisms and $\mathbf{h} = (\sigma_n)_{n\in\mathbb{N}}$ be a sequence in $H^{\mathbb{N}}$ with $\sigma_n : A_{n+1}^* \to A_n^*$. We let $S_{\mathbf{h}}$ denote the set of words $\bigcap_{n\in\mathbb{N}} \mathrm{Fac}(\sigma_0 \cdots \sigma_n(A_{n+1}^*))$. We call a factorial set S an *H-adic set* if there exists $\mathbf{h} \in S^{\mathbb{N}}$ such that $S = S_{\mathbf{h}}$. In this case, the sequence $\mathbf{h}$ is called an *H-adic representation* of S.

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 S 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 S.

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

Proof. Let H be a set of morphisms and $\mathbf{h} = (\sigma_n : A_{n+1}^* \to A_n^*)_{n\in\mathbb{N}} \in H^{\mathbb{N}}$ be a primitive sequence of morphisms such that $S = \bigcap_{n\in\mathbb{N}} \mathrm{Fac}(\sigma_0 \cdots \sigma_n(A_{n+1}^*))$. Consider a word $u \in S$ and let us prove that $u \in \mathrm{Fac}(v)$ for all long enough $v \in S$. The sequence $\mathbf{h}$ being everywhere growing, there is an integer $r > 0$ such that $\min_{a\in A_r} |\sigma_0 \cdots \sigma_{r-1}(a)| > |u|$. As $S = \bigcap_{n\in\mathbb{N}} \mathrm{Fac}(\sigma_0 \cdots \sigma_n(A_{n+1}^*))$, there is an integer $s > r$, two letters $a, b \in A_r$ and a letter $c \in A_s$ such that $u \in \mathrm{Fac}(\sigma_0 \cdots \sigma_{r-1}(ab))$ and $ab \in \mathrm{Fac}(\sigma_r \cdots \sigma_{s-1}(c))$. The sequence $\mathbf{h}$ being primitive, there is an integer $t > s$ such that c occurs in $\sigma_s \cdots \sigma_{t-1}(d)$ for all $d \in$

A_t. Thus u is a factor of all words $v \in S$ such that $|v| \geq \max_{d \in A_t} |\sigma_0 \cdots \sigma_{t-1}(d)|$ and S is uniformly recurrent.

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

Let us show that $\mathbf{h} = (\sigma_n)_{n\in\mathbb{N}}$ is everywhere growing. If not, there is a sequence of letters $(a_n \in A_n)_{n\geq N}$ such that $\sigma_n(a_{n+1}) = a_n$ for all $n \geq N$. This means that the word $r = \sigma_0 \cdots \sigma_n(a_n) \in S$ 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 S for all positive integers k, which contradicts the uniform recurrence of S.

Let us show that $\mathbf{h}$ is primitive. The set S being uniformly recurrent, for all $n \in \mathbb{N}$ there exists N_n such that all words of $S \cap A^{\leq n}$ occur in all words of $S \cap A^{\geq N_n}$. 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 $\sigma_0 \cdots \sigma_{r-1}(A_r) = \mathcal{R}_S(u_r)$ 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 $\sigma_n(a_{n+1}) \in a_nA_n^*$ (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 S for set of factors. According to Durand [12], $\mathbf{w}$ 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 $\mathcal{R}'_S(u_n)$ instead of $\mathcal{R}_S(u_n)$.

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*, that is, 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 [13, 14], if H is finite, then S 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 $\sigma_n \cdots \sigma_{n+k}(A_{n+k+1}) \subset aA_n^* \cap A_n^*b$.

6.2 H-adic representation of tree sets

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 is usually infinite as well as the sequence of alphabets $(A_n)_{n\in\mathbb{N}}$ is usually unbounded (see [15]). For tree sets S, the next theorem significantly improves the only if part of Proposition 6.1: For such sets, the set H can be replaced by the set $\mathcal{H}_e$ of elementary positive automorphisms. In particular, A_n is equal to A for all n.

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

Proof. For any non-ultimately periodic sequence $(u_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}$ 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 H-adic representation of S with $H = \{\sigma_n \mid n \in \mathbb{N}\}$. 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 5.10, the set $\mathcal{R}_S(u_1)$ is a basis of the free group on A. Therefore, by Theorem 5.18, 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 $u_2 \in \mathcal{R}_S(u_1)$ and u_1 is a suffix of u_2. By Theorem 5.12, the derived set $S^{(1)} = \sigma_0^{-1}(S)$ 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 $n \geq 2$. ■

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 5.7, 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 G be as in Example 5.5. The set G is a uniformly recurrent tree set by Theorem 7.1.

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 S 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 $\varphi(S) = G$.*

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 $X = Z \cap S$. By Theorem 5.9, X is a basis of a subgroup of index $\mathrm{Card}(G)$. Let x be a word of X of maximal length (since X is a basis, it is finite and has $\mathrm{Card}(A)$ elements). Then x is not an internal factor of X and thus it has $\mathrm{Card}(G)$ parses. Let $S(x)$ be the set of suffixes of x which are prefixes of X. If $s, t \in S(x)$, then they are comparable for the suffix order. Assume for example that $s = ut$. If $\varphi(s) = \varphi(t)$, then $u \in X^*$ which implies $u = 1$ since s is a prefix of X. Thus all elements of $S(x)$ have distinct images by φ. Since $S(x)$ has $\mathrm{Card}(G)$ elements, this forces $\varphi(S(x)) = G$ and thus $\varphi(S) = G$ since $S(x) \subset S$. ■

We illustrate the proof on the following example.

Example 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 S be the Fibonacci set. The code $X = Z \cap S$ is represented in Figure 7.2. The word

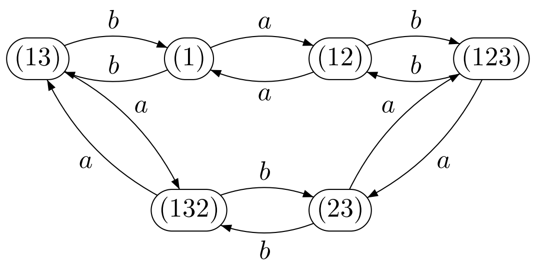

Figure 7.1: The group automaton corresponding to the regular representation of G.

$w = ababa$ 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 S 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 $w \in S$, one has $\varphi(\Gamma_S(w) \cup \{1\}) = G$.*

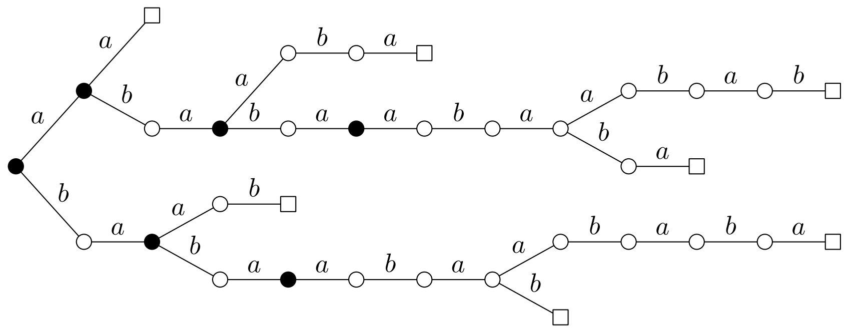

Figure 7.2: The code $X = Z \cap S$

Proof. Let $\alpha : B^* \to A^*$ be a coding morphism for $\mathcal{R}_S(w)$. Then $\beta = \varphi \circ \alpha : B^* \to G$ is a morphism from B^* into G. By Theorem 5.10, the set $\mathcal{R}_S(w)$ is a basis of the free group on A. Thus $\langle \alpha(B) \rangle = \langle A \rangle$. This implies that $\beta(\langle B \rangle) = G$. 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 5.12, the set $H = \alpha^{-1}(w^{-1}S)$ is a uniformly recurrent tree set. Thus $\beta(H) = G$ by Proposition 7.3. This implies that $\varphi(\Gamma_S(w) \cup \{1\}) = G$. ∎

7.2 Proof of the main result

Let S be a uniformly recurrent tree set containing A and let $f : B^* \to A^*$ be a coding morphism for a finite S-maximal bifix code Z. By Theorem 5.8, Z is a basis of a subgroup of index $d_S(Z)$ and, by Theorem 5.11, the submonoid Z^* is saturated in S.

We first prove the following lemma.

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

Proof. Since S is factorial, the set K is factorial. Let $r, s \in K$. Since S is recurrent, there exists $u \in S$ such that $f(r)uf(s) \in S$. Set $t = f(r)uf(s)$. Let G be the representation of $\langle A \rangle$ 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 $\varphi(\Gamma_S(t) \cup \{1\}) = G$. Let $v \in \Gamma_S(t)$ be such that $\varphi(v)$ is the inverse of $\varphi(t)$. Then $\varphi(tv)$ is the identity of G and thus $tv \in \langle Z \rangle$.

Since S is a tree set, it is acyclic and thus Z^* is saturated in S by Theorem 5.11. Thus $Z^* \cap S = \langle Z \rangle \cap S$. 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 $q \in S$. 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 $q \in Z^* \cap S$.

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 S containing A and a coding morphism $f : B^* \to A^*$ for an S-maximal bifix code Z. We set $K = f^{-1}(S)$. 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 $f(w^{-1}K) = v^{-1}S$, by Proposition 2.9 (i), X is a $v^{-1}S$-maximal prefix code.

Finally we set $U = \mathcal{R}_S(v)$. Let $\alpha : C^* \to A^*$ be a coding morphism for U. Since $X \subset \Gamma_S(v)$, 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* $X^* \cap v^{-1}S = U^* \cap Z^* \cap v^{-1}S$.

Proof. Indeed, the left handside is clearly included in the right one. Conversely, consider $x \in U^* \cap Z^* \cap v^{-1}S$. Since $x \in U^* \cap v^{-1}S$, $\alpha^{-1}(x)$ is in $\alpha^{-1}(v^{-1}S) = \alpha^{-1}(\Gamma_S(v)) \cup \{1\}$ by Proposition 4.3. Thus $x \in \Gamma_S(v) \cup \{1\}$. 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 $d = d_S(Z)$. Set $H = \alpha^{-1}(v^{-1}S)$. By Proposition 5.12, 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 $v^{-1}S$-maximal, T is $\alpha^{-1}(v^{-1}S)$-maximal by Proposition 2.9 (ii) and thus H-maximal since $H = \alpha^{-1}(v^{-1}S)$.

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, $\alpha(x) \in U^* \cap v^{-1}S$, 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 5.10, 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 $\alpha(t) \in U^* \cap S$, 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 5.9, T is a basis of a subgroup of index d. Thus $d_H(T) = d$ by Theorem 5.8. ■

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

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 5.7, 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 5.10.

We illustrate the proof with the following example.

Example 7.10 Let S be the Fibonacci set on $A = \{a, b\}$ and let $Z = S \cap A^2 = \{aa, ab, ba\}$. Thus Z is an S-maximal bifix code of S-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 $K = f^{-1}(S)$ is represented in Figure 7.3 on the left.

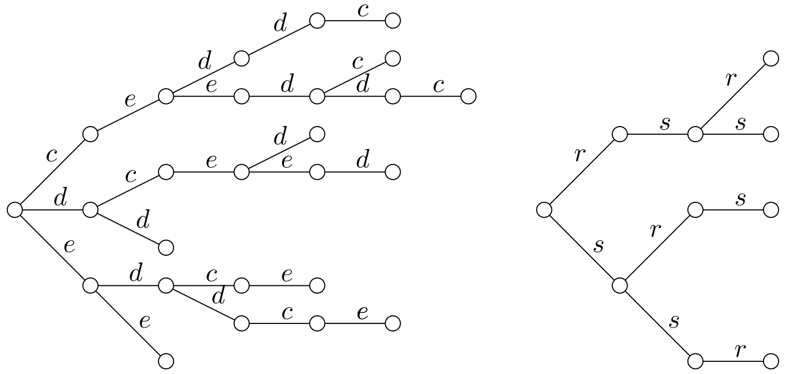

Figure 7.3: The sets K and H.

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 $U = \mathcal{R}_S(aa)$ 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 $H = \alpha^{-1}((aa)^{-1}S)$ 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 S is a tree set is necessary.

Example 7.11 Let S be the set of factors of $(ab)^*$. The set S does not satisfy the tree condition since $G(\epsilon)$ is not connected. Let $X = \{ab, ba\}$. The set X is a finite S-maximal bifix code. Let $f : \{u, v\}^* \to A^*$ be the coding morphism for X defined by $f(u) = ab$, $f(v) = ba$. Then $f^{-1}(S) = u^* \cup v^*$ 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 $S = A^*$ (Proposition 11.1.2).

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

$$d_X(S) = d_Y(G)d_Z(S). \tag{7.1}$$

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

The converse also holds by Proposition 2.9.

To show Formula (7.1), let us first observe that there exist words $w \in S$ 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 $w \in S$ is larger than $3n$. Then if (v, x, u) is a parse of w, we have $|u|, |v| < n$ and thus $|x| > n$. This implies that x is not a factor of X.

Next, we observe that by Theorem 7.1, the set G is a uniformly recurrent tree set and thus in particular, it is recurrent.

Let $w \in S$ be a word with the above property. Let $\Pi_X(w)$ denote the set of parses of w with respect to X and $\Pi_Z(w)$ the set of its parses with respect to Z. We define a map $\varphi : \Pi_X(w) \to \Pi_Z(w)$ as follows. Let $\pi = (v, x, u) \in \Pi_X(w)$. Since Z is a bifix code, there is a unique way to write $v = sy$ and $u = zr$ with $s \in A^* \setminus A^*Z$, $y, z \in Z^*$ and $r \in A^* \setminus ZA^*$. We set $\varphi(\pi) = (s, yxz, r)$. The triples (y, x, z) are in bijection with the parses of $f^{-1}(yxz)$ with respect to Y. Since x is not a factor of X by the hypothesis made on w, and since G is recurrent, there are $d_Y(G)$ such triples. This shows Formula (7.1). ∎

Example 7.13 Let S be the Fibonacci set. Let $B = \{u, v, w\}$ and $A = \{a, b\}$. Let $f : B^* \to A^*$ be the morphism defined by $f(u) = a$, $f(v) = baab$ and

$f(w) = bab$. Set $G = f^{-1}(S)$. The words of length at most 3 of G are represented on Figure 7.4.

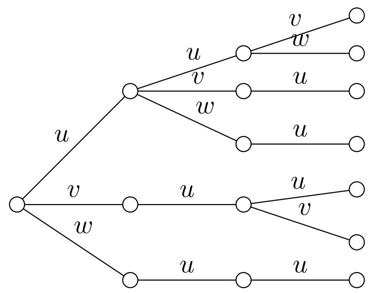

Figure 7.4: The words of length at most 3 in G. figureG

The set $Z = f(B)$ is an S-maximal bifix code of S-degree 2 (it is the unique S-maximal bifix code of S-degree 2 with kernel $\{a\}$). Let $Y = \{uu, uvu, uw, v, wu\}$, which is a 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 S-maximal bifix code of S-degree 4 shown on Figure 7.5.

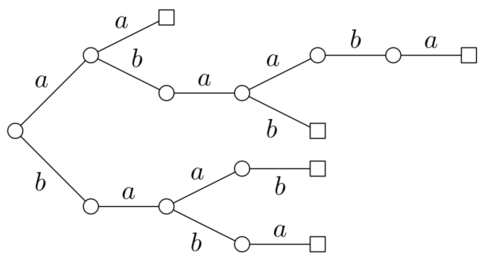

Figure 7.5: An S-maximal bifix code of S-degree 4. figureCodeGiuseppina

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

exampleNotMult **Example 7.14** Let $S = F(ab)^*$ (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 $d_X(F) = 3$ since $abab$ has three parses. Thus $d_F(Z)$ does not divide $d_X(F)$.

References

Arnold1963 [1] Vladimir 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] Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012.

[3] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*, volume 129 of *Encyclopedia Math. Appl.* Cambridge University Press, 2009.

[4] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Acyclic, connected and tree sets. 2014. to appear in Monatsh. Math.

[5] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. The finite index basis property. 2014. to appear in J. Pure Appl. Algebra.

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

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

[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.

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

[10] Issai P. Cornfeld, Serguei V. Fomin, and Yakov 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.

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

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

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

[14] 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.

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

[16] Samuel Eilenberg. *Automata, languages, and machines. Vol. A*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York, 1974. Pure and Applied Mathematics, Vol. 58.

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

FerencziZamboni2008 [18] Sébastien Ferenczi and Luca Q. Zamboni. Languages of k-interval exchange transformations. *Bull. Lond. Math. Soc.*, 40(4):705–714, 2008.

PytheasFogg2002 [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.

GlenJustin2009 [20] Amy Glen and Jacques Justin. Episturmian words: a survey. *Theor. Inform. Appl.*, 43:403–442, 2009.

JustinVuillon2000 [21] Jacques Justin and Laurent Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5):343–356, 2000.

Keane1975 [22] Michael Keane. Interval exchange transformations. *Math. Z.*, 141:25–31, 1975.

Lothaire2002 [23] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

MorseHedlund1940 [24] Marston Morse and Gustav A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.*, 62:1–42, 1940.

Oseledec1966 [25] V. I. Oseledec. The spectrum of ergodic automorphisms. *Dokl. Akad. Nauk SSSR*, 168:1009–1011, 1966.

TanWenZhang2004 [26] 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.

Zorich1997 [27] Anton Zorich. Deviation for interval exchange transformations. *Ergodic Theory Dynam. Systems*, 17(6):1477–1499, 1997.