The finite index basis property

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

June 6, 2014 17 h 4

Abstract

We describe in this paper a connection between bifix codes, symbolic dynamical systems and free groups. This is in the spirit of the connection established previously for the symbolic systems corresponding to Sturmian words. We introduce a class of sets of factors of an infinite word with linear factor complexity containing Sturmian sets and regular interval exchange sets, namely the class of tree sets. We prove as a main result that for a uniformly recurrent tree set S, a finite bifix code X on the alphabet A is S-maximal of S-degree d if and only if it is the basis of a subgroup of index d of the free group on A.

Contents

1 Introduction

In this paper we study a relation between symbolic dynamical systems and bifix codes. The paper is a continuation of the paper with part of the present list of authors on bifix codes and Sturmian words [3]. We understand here by Sturmian words the generalization to arbitrary alphabets, often called strict episturmian words or Arnoux-Rauzy words (see the survey [12]), of the classical Sturmian words on two letters.

As a main result, we prove that, under natural hypotheses satisfied by a Sturmian set S, a finite bifix code X on the alphabet A is S-maximal of S-degree d if and only if it is the basis of a subgroup of index d of the free group on A (Theorem 4.4 called below the Finite Index Basis Theorem).

The proof uses the property, proved in [5], that the sets of first return words in a uniformly recurrent tree set containing the alphabet A form a basis of the free group on A (this result is referred to below as the Return Words Theorem).

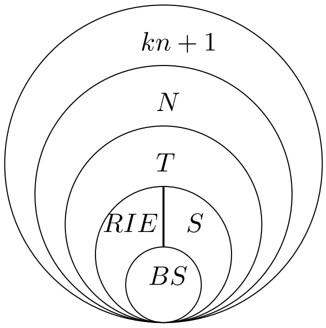
We actually introduce several classes of uniformly recurrent sets of words on $k+1$ letters having all $kn+1$ elements of length n for all $n \geq 0$.

The smallest class (BS) is formed of the Sturmian sets on a binary alphabet, that is, with $k = 1$ (see Figure 1.1). It is contained both in the class of regular interval exchange sets (denoted RIE) and of Sturmian sets (denoted S). Moreover, it can be shown that the intersection of RIE and S is reduced to BS. Indeed, Sturmian sets on more than two letters are not the set of factors of an interval exchange transformation with each interval labeled by a distinct letter (the construction in [2] allows one to obtain the Sturmian sets of 3 letters as an exchange of 7 intervals labeled by 3 letters).

The next one is the class of uniformly recurrent sets satisfying the tree condition (T), which contains the previous ones. The class of uniformly recurrent sets satisfying the neutrality condition (N) contains the class (T). All these classes are contained in the class of uniformly recurrent sets of complexity $kn+1$ on an alphabet with $k+1$ letters.

We have tried in all the paper to use the weakest possible conditions to prove our results. As an example, we prove that, under the neutrality condition, any finite S-maximal bifix code of S-degree d has $1 + d(\mathrm{Card}(A) - 1)$ elements (Theorem 3.6 called below the Cardinality Theorem).

The class RIE is closed under decoding by a maximal bifix code (Corollary 7.2 in [7] referred to as the Bifix Decoding Theorem) but it is not the case for Sturmian sets. In contrast, the uniformly recurrent tree sets form a class of sets containing the Sturmian sets and the regular interval exchange sets which is closed under decoding by a maximal bifix code (see [6]) and for which the Finite Index Basis Theorem is true.



	CT	RT	BT	BD
S	yes	yes	yes	no^4
RIE	yes	yes	yes	yes
T	yes	yes	yes	yes
N	yes	no^2	no^3	?^{5}
$kn+1$	no^1	no	no	no^6

Figure 1.1: The classes of uniformly recurrent sets on $k+1$ letters: Binary Sturmian (BS), Regular interval exchange (RIE), Sturmian (S), Tree (T), Neutral (N), and finally of complexity $kn+1$ (1: see Example 3.10 below, 2: see Example 5.9 in [5], 3: see Example 4.9 below, 4: see Example 4.4 in [7], 5: it can be shown that the neutrality is preserved but it is not known whether the uniform recurrence is, 6: see Example 3.11 below).

For each class, the array on the right of Figure 1.1 indicates whether it satisfies the Cardinality Theorem (CT), the Return Words Theorem (RT), the Finite Index Basis Theorem (BT) or the Bifix Decoding Theorem (BD). All these classes are distinct.

The paper is organized as follows.

In Section 3, we introduce strong, weak and neutral sets. We prove the Cardinality Theorem in neutral sets (Theorem 3.6). We also prove a converse in the sense that a uniformly recurrent set S containing the alphabet and such that the Cardinality Theorem holds for any finite S-maximal bifix code is neutral (Theorem 3.12).

In Section 4, we introduce acyclic and tree sets. The family of tree sets contains Sturmian sets and, as shown in [7], regular interval exchange sets. We prove, as a main result, that in uniformly recurrent tree sets the Finite Index Basis Theorem holds (Theorem 4.4), a result which is proved in [3] for a Sturmian set. The proof uses a result of [5] concerning bifix codes in acyclic sets (Theorem 4.2 referred to as the Saturation Theorem). It also uses the Return Words Theorem proved in [5].

Ackowledgement This work was supported by grants from Région Ile-de-France, the ANR projects Eqinocs and Dyna3S, the Labex Bezout, the FARB Project "Aspetti algebrici e computazionali nella teoria dei codici, degli automi e dei linguaggi formali" (University of Salerno, 2013) and the MIUR PRIN 2010-2011 grant "Automata and Formal Languages: Mathematical and Applicative Aspects". We warmly thank the referee for his useful remarks on the first version of the paper.

2 Preliminaries

In this section, we first recall some definitions concerning words, prefix codes and bifix codes. We give the definitions of recurrent and uniformly recurrent sets of words. We also give the definitions and basic properties of bifix codes (see [3] for a more detailed presentation).

2.1 Words

In this section, we give definitions concerning extensions of words. We define recurrent sets and sets of first return words. For all undefined notions, we refer to [4].

2.1.1 Recurrent sets

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. We denote by 1 or by ε the empty word. We refer to [4] for the notions of prefix, suffix, factor of a word.

A set of words is said to be *prefix-closed* (resp. *factorial*) if it contains the prefixes (resp. 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

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

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.

A set of words $S \neq \{1\}$ is *recurrent* if it is factorial and if for every $u, w \in S$ there is a $v \in S$ such that $uvw \in S$. A recurrent set 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, and thus biextendable.

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 [11], Proposition 1.2.3 for example).

A morphism $f : A^* \to B^*$ is *trivial* if $f(a) = 1$ for all $a \in A$. The image of a uniformly recurrent set by a nontrivial morphism is uniformly recurrent (see [1], Theorem 10.8.6 and Exercise 10.11.38).

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 [3]).

Example 2.1 Let $A = \{a, b\}$. The Fibonacci word is the fixpoint $x = f^\omega(a) = abaababa\ldots$ of the morphism $f : A^* \to A^*$ defined by $f(a) = ab$ and $f(b) = a$. It is a Sturmian word (see [14]). The set $F(x)$ 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 [13]). The set $F(x)$ of factors of x is the *Tribonacci set*.

2.2 Bifix codes

In this section, we present basic definitions concerning prefix codes and bifix codes. For a more detailed presentation, see [4]. We also describe an operation on bifix codes called internal transformation and prove a property of this transformation (Proposition 2.9). It will be used in Section 3.3.

2.2.1 Prefix codes

A *prefix code* is a set of nonempty words which does not contain any proper prefix of its elements. A suffix code is defined symmetrically. A *bifix code* is a set which is both a prefix code and a suffix code.

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

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

We denote by X^* the submonoid generated by X. A set $X \subset S$ is *right S-complete* if any word of S is a prefix of a word in X^*. Given a factorial set S, a prefix code is S-maximal if and only if it is right S-complete (Proposition 3.3.2 in [3]).

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 with respect to X. Let X be a prefix code. By Proposition 4.1.6 in [3], for any $u \in A^*$ and $a \in A$, one has

$$\delta_X(ua) = \begin{cases} \delta_X(u) & \text{if } ua \in A^*X, \\ \delta_X(u) + 1 & \text{otherwise.} \end{cases} \tag{2.1}$$

2.2.2 Maximal bifix codes

Let S be a set of words. A bifix code $X \subset S$ is S-maximal if it is not properly contained in a bifix code $Y \subset S$. For a recurrent set S, a finite bifix code is S-maximal as a bifix code if and only if it is an S-maximal prefix code (see [3], Theorem 4.2.2).

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

For $S = A^*$, we use the term 'maximal bifix code' instead of A^*-maximal bifix code and 'degree' instead of A^*-degree. This is consistent with the terminology of [4].

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 [4]).

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

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

$$I(X) = \{w \in S \mid \delta_X(w) < d\}$$

(Theorem 4.2.8 in [3]). Thus any word of S which is not a factor of X 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 bifix code X is the set $K(X) = I(X) \cap X$. Thus it is the set of words of X which are also internal factors of X. By Theorem 4.3.11 of [3], an S-thin and S-maximal bifix code is determined by its S-degree and its kernel. Moreover, by Theorem 4.3.12 of [3], we have the following result.

Theorem 2.4 *Let S be a recurrent set. A bifix code $Y \subset S$ is the kernel of some S-thin S-maximal bifix code of S-degree d if and only if Y is not S-maximal and $\delta_Y(y) \le d-1$ for all $y \in Y$.*

Example 2.5 Let S be the Fibonacci set. The set $Y = \{a\}$ is a bifix code which is not S-maximal and $\delta_Y(a) = 1$. The set $X = \{a, baab, bab\}$ is the unique S-maximal bifix code of S-degree 2 with kernel $\{a\}$. Indeed, the word bab is not an internal factor and has two parses, namely $(1, bab, 1)$ and (b, a, b).

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

Proposition 2.6 *Let S be a recurrent set. For any S-thin and S-maximal bifix code X of S-degree d, there is a thin maximal bifix code X' of degree d such that $X = X' \cap S$.*

Proof. Let K be the kernel of X and let d be the S-degree of X. By Theorem 2.4, the set K is not S-maximal and $\delta_K(y) \le d-1$ for any $y \in K$. Thus, applying again Theorem 2.4 with $S = A^*$, there is a maximal bifix code X' with kernel K and degree d. Then, by Theorem 4.2.11 of [3], the set $X' \cap S$ is an S-maximal bifix code.

Let us show that $X \cup X'$ is prefix. Suppose that $x \in X$ and $x' \in X'$ are comparable for the prefix order. We may assume that x is a prefix of x' (the other case works symmetrically). If $x \in K$, then $x \in X'$ and thus $x = x'$. Otherwise, $\delta_X(x) = d$. Set $x = pa$ with $a \in A$. Then, by equation (2.1), $\delta_X(x) = \delta_X(p)$ and thus $\delta_X(p) = d$. But since all the factors of p which are in X are in K, we have $\delta_X(p) = \delta_K(p)$. Analogously, since all factors of p which are in X' are in K, we have $\delta_K(p) = \delta_{X'}(p)$. Therefore $\delta_{X'}(p) = d$. But, since X' has degree d, $\delta_{X'}(x) \le d$. Then, by Equation (2.1) again, we have $\delta_{X'}(x) = d$ and $x \in A^*X'$. Let z be the suffix of x which is in X'. If $x \ne x'$, then $z = x$ or $z \in K$ and in both cases $z \in X$. Since X' is prefix and X is suffix, this implies $z = x = x'$.

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

Example 2.7 Let S be the Fibonacci set. Let $X = \{a, baab, bab\}$ be the S-maximal bifix code of S-degree 2 with kernel $\{a\}$. Then $X' = a \cup ba^*b$ is the maximal bifix code with kernel $\{a\}$ of degree 2 such that $X' \cap S = X$.

2.2.3 Internal transformation

We will use the following transformation which operates on bifix codes (see [4, Chapter 6] for a more detailed presentation). For a set of words X and a word u, we denote $u^{-1}X = \{v \in A^* \mid uv \in X\}$ and $Xu^{-1} = \{v \in A^* \mid vu \in X\}$ the *residuals* of X with respect to u (one should not confuse this notation with that of the inverse in the free group). Let $X \subset S$ be a set of words and $w \in S$ a word. Let

$$G = Xw^{-1}, \qquad D = w^{-1}X, \tag{2.2}$$

$$G_0 = (wD)w^{-1} \quad D_0 = w^{-1}(Gw), \tag{2.3}$$

$$G_1 = G \setminus G_0, \quad D_1 = D \setminus D_0. \tag{2.4}$$

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

$$Y = (X \cup w \cup (G_1wD_0^*D_1 \cap S)) \setminus (Gw \cup wD) \tag{2.5}$$

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

$$Y = (X \cup w \cup (GwD \cap S)) \setminus (Gw \cup wD). \tag{2.6}$$

It is this form which is used in [3] to define the internal transformation.

Example 2.8 Let S be the Fibonacci set. Let $X = S \cap A^2$. The internal transformation applied to X with respect to b gives $Y = \{aa, aba, b\}$. The internal transformation applied to X with respect to a gives $Y' = \{a, baab, bab\}$.

The following result is proved in [3] in the case $G_0 = \emptyset$ (Proposition 4.4.5).

Proposition 2.9 *Let S be a uniformly recurrent set and let $X \subset S$ be a finite S-maximal bifix code of S-degree d. Let $w \in S$ be a nonempty word such that the sets G_1, D_1 defined by Equation* (2.4) *are nonempty. Then the set Y obtained as in Equation* (2.5) *is a finite S-maximal bifix code with S-degree at most d.*

Proof. By Proposition 2.6 there is a thin maximal bifix code X' of degree d such that $X = X' \cap S$. Let Y' be the code obtained from X' by internal transformation with respect to w. Then

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

with $G' = X'w^{-1}$, $D' = w^{-1}X'$, and $G_0' = (wD')w^{-1}$, $D_0' = w^{-1}(G'w)$, $G_1' = G' \setminus G_0'$, $D_1' = D' \setminus D_0'$. We have $G = G' \cap Sw^{-1}$, $D = D' \cap w^{-1}S$, and $D_i = D_i' \cap w^{-1}S$, $G_i = G_i' \cap Sw^{-1}$ for $i = 0, 1$. In particular $G_1 \subset G_1'$, $D_1 \subset D_1'$. Thus $G_1', D_1' \neq \emptyset$. This implies that Y' is a thin maximal bifix code of degree d (see Proposition 6.2.8 and its complement page 242 in [4]).

Since $w \in S$, we have $Y = Y' \cap S$. By Theorem 4.2.11 of [3], Y is an S-maximal bifix code of S-degree at most d. Since S is uniformly recurrent, this implies that Y is finite. ∎

When $G_0 = \emptyset$, the bifix code Y has S-degree d (see [3], Proposition 4.4.5). We will see in the proof of Theorem 3.12 another case where it is true. We have no example where it is not true.

Example 2.10 Let S be the Fibonacci set, as in Example 2.8. Let $X = S \cap A^2$ and let $w = a$. Then $Y = \{a, baab, bab\}$ is the S-maximal bifix code of S-degree 2 already considered in Example 2.8.

3 Strong, weak and neutral sets

In this section, we introduce strong, weak and neutral sets. We prove a theorem concerning the cardinality of an S-maximal bifix code in a neutral set S (Theorem 3.6).

3.1 Strong, weak and neutral words

Let S be a factorial set. For a word $w \in S$, let

$$m(w) = e(w) - \ell(w) - r(w) + 1.$$

We say that, with respect to S, w is *strong* if $m(w) > 0$, *weak* if $m(w) < 0$ and *neutral* if $m(w) = 0$.

A biextendable word w is called *ordinary* if $E(w) \subset a \times A \cup A \times b$ for some $(a, b) \in E(w)$ (see [8], Chapter 4). If S is biextendable, any ordinary word is neutral. Indeed, one has $E(w) = (a \times (R(w) \setminus b)) \cup ((L(w) \setminus a) \times b) \cup (a, b)$ and thus $e(w) = \ell(w) + r(w) - 1$.

Example 3.1 In a Sturmian set, any word is ordinary. Indeed, for any bispecial word w, there is a unique letter a such that aw is right-special and a unique letter b such that wb is left-special. Then $awb \in S$ and $E(w) = a \times A \cup A \times b$.

We say that a set of words S is *strong* (resp. *weak*, resp. *neutral*) if it is factorial and every word $w \in S$ is strong or neutral (resp. weak or neutral, resp. neutral).

The sequence $(p_n)_{n \geq 0}$ with $p_n = Card(S \cap A^n)$ is called the *complexity* of S. Set $k = \mathrm{Card}(S \cap A) - 1$.

Proposition 3.2 *The complexity of a strong (resp. weak, resp. neutral) set S is at least (resp. at most, resp. exactly) equal to $kn + 1$.*

Given a factorial set S with complexity p_n, we denote $s_n = p_{n+1} - p_n$ the first difference of the sequence p_n and $b_n = s_{n+1} - s_n$ its second difference. The following is from [9] (it is also part of Theorem 4.5.4 in [8, Chapter 4] and also Lemma 3.3 in [5]).

Lemma 3.3 *We have*

$$b_n = \sum_{w \in A^n \cap S} m(w) \quad \text{and} \quad s_n = \sum_{w \in A^n \cap S} (r(w) - 1)$$

for all $n \geq 0$.

Proposition 3.2 follows easily from the following lemma.

Lemma 3.4 *If S is strong (resp. weak, resp. neutral), then $s_n \geq k$ (resp. $s_n \leq k$, resp. $s_n = k$) for all $n \geq 0$.*

Proof. Assume that S is strong. Then $m(w) \geq 0$ for all $w \in S$ and thus, by Lemma 3.3, the sequence (s_n) is nondecreasing. Since $s_0 = k$, this implies $s_n \geq k$ for all n. The proof of the other cases is similar. ∎

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

Example 3.5 Let $A = \{a, b, c\}$. The *Chacon word* on three letters is the fixpoint $x = f^\omega(a)$ of the morphism f from A^* into itself defined by $f(a) = aabc$, $f(b) = bc$ and $f(c) = abc$. Thus $x = aabcaabcbcabc \cdots$. The *Chacon set* is the set S of factors of x. It is of complexity $2n+1$ (see [11] Section 5.5.2).

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

3.2 The Cardinality Theorem

The following result, referred to as the Cardinality Theorem, is a generalization of a result proved in [3] in the less general case of a Sturmian set. Since $S \cap A^n$ is an S-maximal bifix code of S-degree n (see Example 2.3), it is also a generalization of Proposition 3.2.

Theorem 3.6 *Let S be a recurrent set containing the alphabet A and let $X \subset S$ be a finite S-maximal bifix code. Set $k = \mathrm{Card}(A) - 1$ and $d = d_X(S)$. If S is strong (resp. weak), then $\mathrm{Card}(X) - 1 \geq dk$ (resp. $\mathrm{Card}(X) - 1 \leq dk$). If S is neutral, then $\mathrm{Card}(X) - 1 = dk$.*

Note that, for a recurrent neutral set S, a bifix code $X \subset S$ may be infinite since this may happen for a Sturmian set S (see [3], Example 5.1.4).

We consider rooted trees with the usual notions of root, node, child and parent. The following lemma is an application of a well-known lemma on trees relating the number of its leaves to the sum of the degrees of its internal nodes.

Lemma 3.7 *Let S be a prefix-closed set. Let X be a finite S-maximal prefix code and let P be the set of its proper prefixes. Then* $\mathrm{Card}(X) = 1 + \sum_{p \in P}(r(p) - 1)$.

We order the nodes of a tree from the parent to the child and thus we have $m \leq n$ if m is a descendant of n. We denote $m < n$ if $m \leq n$ with $m \neq n$.

Lemma 3.8 *Let T be a finite tree with root r on a set N of nodes, let $d \geq 1$, and let π, α be functions assigning to each node an integer such that*

(i) *for each internal node n, $\pi(n) \leq \sum \pi(m)$ where the sum runs over the children of n,*

(ii) *for each leaf m of T, one has $\sum_{m \leq n} \alpha(n) = d$.*

Then $\sum_{n\in N}\alpha(n)\pi(n) \geq d\pi(r)$.

Proof. We use an induction on the number of nodes of T. If T is reduced to its root, then $d = \alpha(r)$ implies $\alpha(r)\pi(r) = d\pi(r)$ and the result is true. Assume that it holds for trees with less nodes than T. Since T is finite and not reduced to its root, there is an internal node such that all its children are leaves of T. Let m be such a node. Since $\sum_{x\leq n}\alpha(n) = \alpha(x) + \sum_{m\leq n}\alpha(n)$ has value d for each child x of m, the value $v = \alpha(x)$ is the same for all children of m. Let T' be the tree obtained from T by deleting all children of m. Let N' be the set of nodes of T'. Let π' be the restriction of π to N' and let α' be defined by

$$\alpha'(n) = \begin{cases} \alpha(n) & \text{if } n \neq m \\ \alpha(m) + v & \text{otherwise.} \end{cases}$$

It is easy to verify that T', π' and α' satisfy the same hypotheses as T, π and α. Then

$$\begin{aligned} \sum_{n\in N}\alpha(n)\pi(n) &= \sum_{n\in N'\setminus m}\alpha(n)\pi(n) + \alpha(m)\pi(m) + \sum_{x<m} v\pi(x) \\ &= \sum_{n\in N'\setminus m}\alpha'(n)\pi'(n) + \alpha(m)\pi(m) + v\sum_{x<m}\pi(x) \\ &\geq \sum_{n\in N'\setminus m}\alpha'(n)\pi'(n) + (\alpha(m)+v)\pi(m) \\ &= \sum_{n\in N'\setminus m}\alpha'(n)\pi'(n) + \alpha'(m)\pi'(m) = \sum_{n\in N'}\alpha'(n)\pi'(n) \end{aligned}$$

whence the result by the induction hypothesis. ∎

A symmetric statement holds replacing the inequality in condition (i) by $\pi(n) \geq \sum \pi(m)$ and the conclusion by $\sum_{n\in N}\alpha(n)\pi(n) \leq d\pi(r)$.

Proof of Theorem 3.6. Assume first that S is strong. Let N be larger than the lengths of the words of X.

Let U be the set of words of S of length at most N. By considering each word w as the father of aw for $a \in A$, the set U can be considered as a tree T with root the empty word ε. The leaves of T are the elements of S of length N.

For $w \in U$, set $\pi(w) = r(w) - 1$ and let

$$\alpha(n) = \begin{cases} 1 & \text{if } n \text{ is a proper prefix of } X \\ 0 & \text{otherwise.} \end{cases}$$

Let us verify that the conditions of Lemma 3.8 are satisfied. Let u be in U with $|u| < N$. Then since u is strong or neutral, $\sum_{a\in L(u)}(r(au) - 1) = e(u) - \ell(u) \geq r(u) - 1$. This implies that $\sum_{au\in S}\pi(au) \geq \pi(u)$ showing that condition (i) is satisfied.

Let w be a leaf of T, that is, a word of S of length N. Since N is larger than the maximal length of the words of X, the word w is not an internal factor of X and thus it has d parses with respect to X. It implies that it has d suffixes which are proper prefixes of X (since X is right S-complete, this is the same as to have no prefix in X). Thus $\sum_{w \leq u} \alpha(u) = d$. Thus condition (ii) is also satisfied.

By Lemma 3.8, we have $\sum_{n \in U} \alpha(n)\pi(n) \geq d\pi(\varepsilon)$. Let P be the set of proper prefixes of X. By definition of α, we have $\sum_{n \in U} \alpha(n)\pi(n) = \sum_{p \in P} \pi(p)$ and thus by definition of π, $d\pi(\varepsilon) = dk \leq \sum_{p \in P}(r(p) - 1)$. Since S is recurrent, X is an S-maximal prefix code. Thus, by Lemma 3.7, we have $\mathrm{Card}(X) = 1 + \sum_{p \in P}(r(p) - 1)$ and thus we obtain $\mathrm{Card}(X) \geq 1 + dk$ which is the desired conclusion.

The proof that $\mathrm{Card}(X) - 1 \leq dk$ if S is weak is symmetric, using the symmetric version of Lemma 3.8. The case where S is neutral follows then directly. ■

We illustrate Theorem 3.6 in the following example.

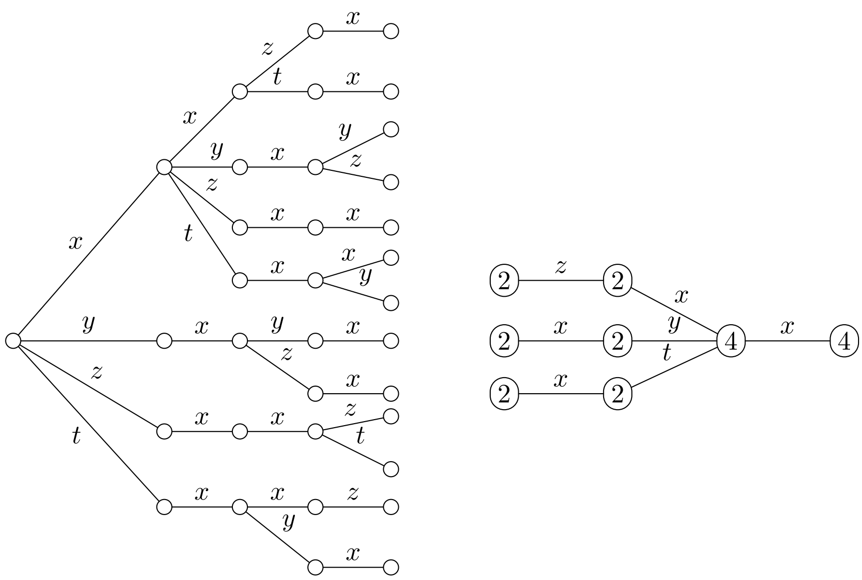


Figure 3.1: The words of length at most 4 of a neutral set G and the tree of right-special words.

Example 3.9 Consider the set G of words on the alphabet $B = \{x, y, z, t\}$ obtained as follows. Let S be the Fibonacci set and let $X \subset S$ be the S-maximal bifix code of S-degree 3 defined by $X = \{a, baabaab, baabab, babaab\}$. We consider the morphism $f : B^* \to A^*$ defined by $f(x) = a$, $f(y) = baabaab$, $f(z) = baabab$, $f(t) = babaab$. We set $G = f^{-1}(S)$.

The words of G of length at most 4 are represented in Figure 3.1 on the left. By the main result of [6], the set G is a uniformly recurrent neutral set. Indeed, since S is Sturmian, it is a tree set (see the definition in Section 4) and thus G is a tree set, which implies that it is neutral.

The tree of right-special words is represented on the right in Figure 3.1 with the value of r indicated at each node. The bifix codes

$$Y = \{xx, xyx, xz, xt, y, zx, tx\}, \quad Z = \{x, yxy, yxz, zxxz, zxxt, txxz, txy\}$$

are G-maximal and have both G-degree 2. In agreement with Theorem 3.6, we have $\text{Card}(Y) = \text{Card}(Z) = 1 + 2(\text{Card}(B) - 1) = 7$. The codes Y and Z are represented in Figure 3.2. The right-special proper prefixes p of Y and Z are

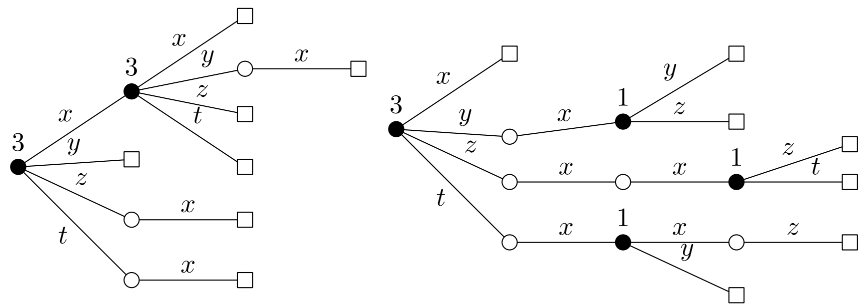


Figure 3.2: Two G-maximal bifix codes of G-degree 2.

indicated in black in Figure 3.2 with the value of $r(p)-1$ indicated for each one. In agreement with Lemma 3.7, the sum of the values of $r(p) - 1$ is 6 in both cases.

The following example illustrates the necessity of the hypotheses in Theorem 3.6.

Example 3.10 Consider again the Chacon set S of Example 3.5. Let $X = S \cap A^4$ and let Y, Z be the S-maximal bifix codes of S-degree 4 represented in Figure 3.3. The first one is obtained from X by internal transformation with respect to *abc* . The second one with respect to *bca*. We have $\text{Card}(Y) = 10$ and $\text{Card}(Z) = 8$ showing that $\text{Card}(Y) - 1 > 8$ and $\text{Card}(Z) - 1 < 8$, illustrating the fact that S is neither strong nor weak.

The following example shows that the class of sets of factor complexity $kn+1$ is not closed by maximal bifix decoding.

Example 3.11 Let S be the Chacon set and let $f : B^* \to A^*$ be a coding morphism for the S-maximal bifix code Z of S-degree 4 with 8 elements of Example 3.10. One may verify that $\text{Card}(B^2 \cap f^{-1}(S)) = \text{Card}(Z^2 \cap S) = 17$. This shows that the set $f^{-1}(S)$ does not have factor complexity $7n + 1$.

3.3 A converse of the Cardinality Theorem

We end this section with a statement proving a converse of the Cardinality Theorem.

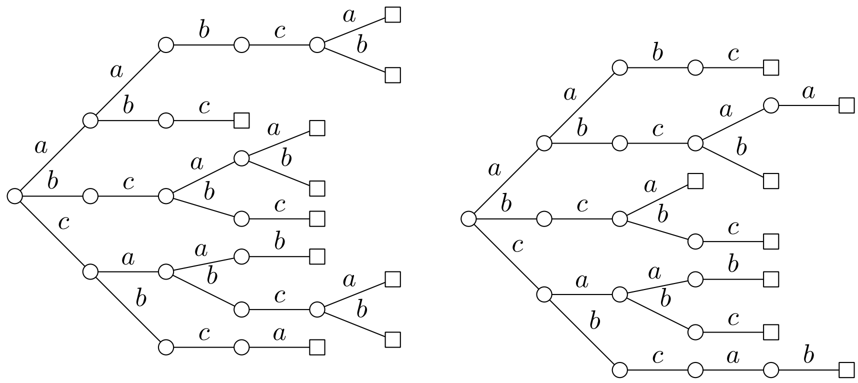


Figure 3.3: Two S-maximal bifix codes of S-degree 4.

Theorem 3.12 *Let S be a uniformly recurrent set containing the alphabet A. If any finite S-maximal bifix code of S-degree d has $d(\mathrm{Card}(A)-1)+1$ elements, then S is neutral.*

Proof. We may assume that A has more than one element. We argue by contradiction. Let $w \in S$ be a word which is not neutral. We cannot have $w = \varepsilon$ since otherwise the S-maximal bifix code $X = S \cap A^2$ has not the good cardinality.

Set $n = |w|$ and $X = S \cap A^{n+1}$. The set X is an S-maximal bifix code of S-degree $n+1$. Let Y be the code obtained by internal transformation from X with respect to w and defined by Equation (2.5). Note that $G = L(w)$ and $D = R(w)$.

We distinguish two cases.

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

The code Y is defined by Equation (2.6) and we have $\mathrm{Card}(GwD \cap S) = e(w)$. Since $D_0 = G_0 = \emptyset$, the hypotheses of Proposition 2.9 are satisfied and Y has S-degree $n+1$ (by Proposition 4.4.5 in [3]). This implies $\mathrm{Card}(X) = \mathrm{Card}(Y)$. On the other hand

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

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

Case 2. Assume next that $Gw \cap wD \neq \emptyset$. Then $w = a^n$ with $n > 0$ for some letter a and the sets G_0, D_0 defined by Equation 2.3 are $G_0 = D_0 = \{a\}$. Moreover $a^{n+1} \in X$.

Since w is not neutral, it is bispecial. Thus the sets G_1, D_1 are nonempty and the hypotheses of Proposition 2.9 are satisfied. Since S is uniformly recurrent and since $S \neq a^*$, the set $a^* \cap S$ is finite. Set $a^* \cap S = \{1, a, \ldots, a^m\}$. Thus $m \geq n+1$.

Let $b \neq a$ be a letter such that $a^m b \in S$. Then, $\delta_Y(a^m) = n$ since a^m has n suffixes which are proper prefixes of Y. Moreover, $a^m b$ has no suffix in Y. Indeed, if $a^t b \in Y$, we cannot have $t \geq n$ since $a^n \in Y$. And since all words in Y except a^n have length greater than n, $t < n$ is also impossible. Thus by Equation (2.1), we have $\delta_Y(a^m b) = \delta_Y(a^m) + 1$ and thus $\delta_Y(a^m b) = n+1$. This shows that the S-degree of Y is $n+1$ and thus that $\mathrm{Card}(Y) = \mathrm{Card}(X)$ as in Case 1.

We may assume that n is chosen maximal such that a^n is not neutral. This is always possible if a^m is neutral. Otherwise, Case 1 applies to $X = S \cap A^{m+1}$ and $w = a^m$.

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

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

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

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

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

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

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

4 Tree sets

We introduce in this section the notions of acyclic and tree sets. We state and prove the main result of this paper (Theorem 4.4). The proof uses results from [5].

4.1 Acyclic and tree sets

Let S be a set of words. For $w \in S$, the *extension graph* of w is the undirected bipartite graph $G(w)$ on the set of vertices which is the disjoint union of $L(w)$ and $R(w)$ with edges the pairs $(a, b) \in E(w)$. An edge $(a, b) \in E(w)$ goes from $a \in L(w)$ to $b \in R(w)$.

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

Let S be a biextendable set. We say that S is *acyclic* if for every word $w \in S$, the graph $G(w)$ is acyclic. We say that S is a *tree set* if $G(w)$ is a tree for all $w \in S$.

Clearly an acyclic set is weak and a tree set is neutral.

Note that a biextendable set S is a tree set if and only if the graph $G(w)$ is a tree for every bispecial non-ordinary word w. Indeed, if w is not bispecial or if it is ordinary, then $G(w)$ is always a tree.

Proposition 4.1 *A Sturmian set S is a tree set.*

Indeed, S is biextendable and every bispecial word is ordinary (see Example 3.1).

The following example shows that there are neutral sets which are not tree sets.

Example 4.2 Let $A = \{a, b, c\}$ and let S be the set of factors of $a^*\{bc, bcbc\}a^*$. The set S is biextendable. One has $S \cap A^2 = \{aa, ab, bc, cb, ca\}$. It is neutral. Indeed the empty word is neutral since $e(\varepsilon) = \mathrm{Card}(S \cap A^2) = 5 = \ell(\varepsilon) + r(\varepsilon) - 1$. Next, the only nonempty bispecial words are bc and a^n for $n \geq 1$. They are neutral since $e(bc) = 3 = \ell(bc) + r(bc) - 1$ and $e(a^n) = 3 = \ell(a^n) + r(a^n) - 1$. However, S is not acyclic since the graph $G(\varepsilon)$ contains a cycle (and has two connected components, see Figure 4.1).

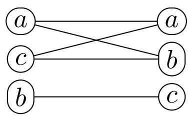


Figure 4.1: The graph $G(\varepsilon)$ for the set S.

In the last example, the set is not recurrent. We present now an example, due to Julien Cassaigne [10] of a uniformly recurrent set which is neutral but is not a tree set (it is actually not even acyclic).

Example 4.3 Let $A = \{a, b, c, d\}$ and let σ be the morphism from A^* into itself defined by

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

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

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

Let S be the set of factors of the infinite word $\tau(\sigma^\omega(a))$ (see Figure 4.2).

It is shown in [5] (Example 4.5) that S is a uniformly recurrent neutral set. It is not a tree set since $G(\varepsilon)$ is neither acyclic nor connected.

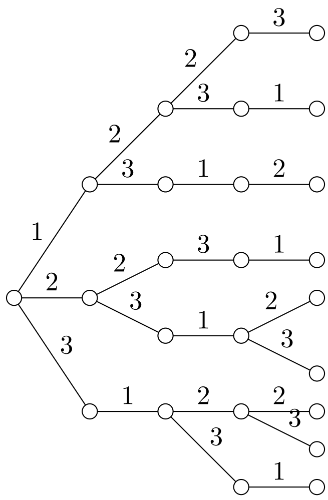


Figure 4.2: The words of length at most 4 of the set S.

4.2 Finite index basis property

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 will prove the following result, referred to as the Finite Index Basis Theorem.

Theorem 4.4 *Any uniformly recurrent tree set S containing the alphabet A has the finite index basis property.*

Note that the Cardinality Theorem (Theorem 3.6) holds for a set S satisfying the finite index basis property. Indeed, by Schreier's formula a basis of a subgroup of index d of a free group on s generators has $(s-1)d+1$ elements (actually we use Theorem 3.6 in the proof of Theorem 4.4).

We denote by by $FG(A)$ the free group on the set A and by $\langle X \rangle$ the subgroup generated by a set of words X. A submonoid M of A^* is called *saturated* in S if $M \cap S = \langle M \rangle \cap S$. We recall the following result from [5] (Theorem 6.2 referred to as the Saturation Theorem).

Theorem 4.5 *Let S be an acyclic set. The submonoid generated by a bifix code included in S is saturated in S.*

Actually, by a second result of [5] (Theorem 6.1 referred to as the Freeness Theorem), if S is acyclic, any bifix code $X \subset S$ is free, which means that it is a basis of the subgroup $\langle X \rangle$. We will not use this result here and thus we will prove directly that if S is a uniformly recurrent tree set, any finite S-maximal bifix code is free.

Before proving Theorem 4.4, we list some related results. The first one is the main result of [3].

Corollary 4.6 *A Sturmian set has the finite index basis property.*

Proof. This follows from Theorem 4.4 since a Sturmian set is a uniformly recurrent tree set (Proposition 4.1). ∎

The following examples shows that Theorem 4.4 may be false for a set S which does not satisfy some of the hypotheses.

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

Example 4.7 Let S be the Chacon set (see Example 3.5). We have seen that S is not neutral and thus not a tree set. The set $S \cap A^2 = \{aa, ab, bc, ca, cb\}$ is an S-maximal bifix code of S-degree 2. It is not a basis since $ca(aa)^{-1}ab = cb$. Thus S does not satisfy the finite index basis property.

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

Example 4.8 Let S be the set of Example 4.2. It is not a tree set (and it is not either uniformly recurrent). The set $S \cap A^2$ is the same as in the Chacon set. Thus S does not satisfy the finite index basis property.

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

Example 4.9 Let S be the set on the alphabet $B = \{1, 2, 3\}$ of Example 4.3. We have seen that S is neutral but not a tree set.

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

We close this section with a converse of Theorem 4.4.

Proposition 4.10 *A biextendable set S such that $S \cap A^n$ is a basis of the subgroup $\langle A^n \rangle$ for all $n \geq 1$ is a tree set.*

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

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

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

contradicting the fact that X is a basis.

Since $G(w)$ is an acyclic graph with $\ell(w)+r(w)$ vertices and $e(w)$ edges, we have $e(w) \leq \ell(w)+r(w)-1$. But then

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

Since $\mathrm{Card}(A^{m+2}\cap S) = k(m+2)+1$, we have $e(w) = \ell(w)+r(w)-1$ for all $w \in A^m$. This implies that $G(w)$ is a tree for all $w \in S$. Thus S is a tree set. ■

Corollary 4.11 *A uniformly recurrent set which has the finite index basis property is a tree set.*

Proof. Let S be a uniformly recurrent set having the finite index basis property. For any $n \geq 1$, the set $S \cap A^n$ is an S-maximal bifix code of S-degree n (Example 2.3). Thus it is a basis of a subgroup of index n. Since it is included in the subgroup generated by A^n, which has index n, it is a basis of this subgroup. This implies that S is a tree set by Proposition 4.10. ■

4.3 Proof of the Finite Index Basis Theorem

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. When S is recurrent, the set $\Gamma_S(w)$ is nonempty. Let

$$\mathcal{R}_S(w) = \Gamma_S(w) \setminus \Gamma_S(w)A^+$$

be the set of *first right return words*.

The proof of Theorem 4.4 uses several other results, among which Theorem 4.5 and the following result from [5] (Theorem 5.6).

Theorem 4.12 *Let S be a uniformly recurrent tree set containing the alphabet A. For any $w \in S$, the set $\mathcal{R}_S(w)$ is a basis of the free group on A.*

Proof of Theorem 4.4. Assume first that X is a finite S-maximal bifix code of S-degree d. Let P be the set of proper prefixes of X. Let H be the subgroup generated by X.

Let $u \in S$ be a word such that $\delta_X(u) = d$, or, equivalently, which is not an internal factor of X. Let Q be the set formed of the d suffixes of u which are in P.

Let us first show that the cosets Hq for $q \in Q$ are disjoint. Indeed, $Hp \cap Hq \neq \emptyset$ implies $Hp = Hq$. Any $p, q \in Q$ are comparable for the suffix order. Assuming that q is longer than p, we have $q = tp$ for some $t \in P$. Then $Hp = Hq$ implies

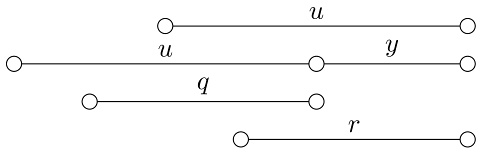


Figure 4.3: A word $y \in \mathcal{R}_S(u)$.

$Ht = H$ and thus $t \in H \cap S$. By Theorem 4.5, since S is acyclic, this implies $t \in X^*$ and thus $t = \varepsilon$. Thus $p = q$.

Let

$$V = \{v \in FG(A) \mid Qv \subset HQ\}.$$

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

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

We show that the set $\mathcal{R}_S(u)$ is contained in V. Indeed, let $q \in Q$ and $y \in \mathcal{R}_S(u)$. Since q is a suffix of u, qy is a suffix of uy, and since uy is in S (by definition of $\mathcal{R}_S(u)$), also qy is in S. Since X is an S-maximal bifix code, it is an S-maximal prefix code and thus it is right S-complete. This implies that qy is a prefix of a word in X^* and thus there is a word $r \in P$ such that $qy \in X^*r$. We verify that the word r is a suffix of u. Since $y \in \mathcal{R}_S(u)$, there is a word y' such that $uy = y'u$. Consequently, r is a suffix of $y'u$, and in fact the word r is a suffix of u. Indeed, one has $|r| \leq |u|$ since otherwise u is in the set $I(X)$ of internal factors of X, and this is not the case. Thus we have $r \in Q$ (see Figure 4.3). Since $X^* \subset H$ and $r \in Q$, we have $qy \in HQ$. Thus $y \in V$.

By Theorem 4.12, the group generated by $\mathcal{R}_S(u)$ is the free group on A. Since $\mathcal{R}_S(u) \subset V$, and since V is a subgroup of $FG(A)$, we have $V = FG(A)$. Thus $Qw \subset HQ$ for any $w \in FG(A)$. Since $1 \in Q$, we have in particular $w \in HQ$. Thus $FG(A) = HQ$. Since $\mathrm{Card}(Q) = d$, and since the right cosets Hq for $q \in Q$ are pairwise disjoint, this shows that H is a subgroup of index d. Since S is acyclic and recurrent, by Theorem 3.6, we have $\mathrm{Card}(X) \leq d(\mathrm{Card}(A) - 1) + 1$. But since X generates H, it contains a basis of H. In view of Schreier's Formula, this implies that X is a basis of H.

Assume conversely that the finite bifix code $X \subset S$ is a basis of the group $H = \langle X \rangle$ and that H has index d. Since X is a basis of H, by Schreier's Formula, we have $\mathrm{Card}(X) = (k-1)d + 1$, where $k = \mathrm{Card}(A)$. The case $k = 1$ is straightforward; thus we assume $k \geq 2$. By Theorem 4.4.3 in [3], if S is a uniformly recurrent set, any finite bifix code contained in S is contained in a finite S-maximal bifix code. Thus there is a finite S-maximal bifix code Y containing X. Let e be the S-degree of Y. By the first part of the proof, Y is a basis of a subgroup K of index e of the free group on A. In particular, it has $(k-1)e + 1$ elements. Since $X \subset Y$, we have $(k-1)d + 1 \leq (k-1)e + 1$ and thus $d \leq e$. On the other hand, since H is included in K, d is a multiple of e and thus $e \leq d$. We conclude that $d = e$ and thus that $X = Y$. ■

References

[1] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences*. Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations. 5

[2] Pierre Arnoux and Gérard Rauzy. Représentation géométrique de suites de complexité $2n+1$. *Bull. Soc. Math. France*, 119(2):199–215, 1991. 2

[3] 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, 10, 14, 18, 20

[4] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2009. 4, 5, 6, 7, 8

[5] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Acyclic, connected and tree sets. 2013. `http://arxiv.org/abs/1308.4260`. 2, 3, 9, 15, 16, 17, 19

[6] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. 2013. `http://arxiv.org/abs/1308.5396`. 2, 12

[7] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and interval exchange transformations. 2014. 2, 3

[8] Valérie Berthé and Michel Rigo, editors. *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010. 9

[9] Julien Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin*, 4(1):67–88, 1997. Journées Montoises (Mons, 1994). 9

[10] Julien Cassaigne. 2013. Personal communication. 16

[11] 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. 5, 10

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

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

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