Codes and automata in minimal sets

Dominique Perrin

Codes and automata in minimal sets

Dominique Perrin

Université Paris Est, LIGM

Abstract. We explore several notions concerning codes and automata in a restricted set of words S. We define a notion of S-degree of an automaton and prove an inequality relating the cardinality of a prefix code included in a minimal set S and its S-degree.

1 Introduction

We have introduced in [5] the notion of tree set as a common generalization of Sturmian sets and of interval exchange sets. In this paper, we investigate several new directions concerning codes and automata in minimal sets.

Codes and automata in restricted sets of words have already been investigated several times. In particular, Restivo has investigated codes in sets of finite type [13] and Reutenauer has studied the more general notion of codes of paths in a graph [14]. We have initiated in [3] with several other authors, a systematic study of bifix codes in Sturmian sets, a subject already considered before in [8]. The overall conclusion of this study is that very surprising phenomena appear in this context in relation with subgroups of finite index of the free group, allowing one to obtain positive bases of the subgroups contained in a given minimal set.

In this paper, we investigate several notions concerning codes and automata in relation with a factorial set S. This includes a definition of minimal S-rank of an automaton, which is equal to 1 if and only if the automaton is synchronized. We prove a result which allows to compute the minimal S-rank when S is minimal (Theorem 3.1). We also show that for a recurrent set S and a strongly connected automaton $\mathcal{A}$, the set of elements of the transition monoid M of minimal S-rank is included in a $\mathcal{D}$-class of M called its S-minimal $\mathcal{D}$-class (Proposition 3.2). This regular $\mathcal{D}$-class is unique when S is minimal and it is related with the results of [1] and [2] on the regular $\mathcal{J}$-classes of free profinite semigroups.

We define the S-degree of a prefix code X included in S as the minimal S-rank of the minimal automaton of X^*. We show that the cardinality of a prefix code is bounded below by a linear function of its S-degree (Theorem 4.4).

Let X be a prefix code and let M be the transition monoid of the minimal automaton of X^*. We associate to X a permutation group denoted $G_X(S)$ which is the structure group of the S-minimal $\mathcal{D}$-class of M. We show that for any uniformly recurrent tree set S and any finite S-maximal bifix code X, the group $G_X(S)$ is equivalent to the representation of the free group on the cosets of the subgroup generated by X (Theorem 4.5).

2 Neutral and tree sets

Let A be a finite alphabet. We denote by A^* the set of all words on A. We denote by ε or 1 the empty word. A set of words on the alphabet A and containing A is said to be *factorial* if it contains the factors of its elements. An *internal factor* of a word x is a word v such that $x = uvw$ with u, w nonempty.

2.1 Neutral sets

Let S be a factorial set on the alphabet A. For $w \in S$, we denote $L_S(w) = \{a \in A \mid aw \in S\}$, $R_S(w) = \{a \in A \mid wa \in S\}, E_S(w) = \{(a, b) \in A \times A \mid awb \in S\}$, and further $\ell_S(w) = \mathrm{Card}(L_S(w))$, $r_S(w) = \mathrm{Card}(R_S(w))$, $e_S(w) = \mathrm{Card}(E_S(w))$.

We omit the subscript S when it is clear from the context. A word w is *right-extendable* if $r(w) > 0$, *left-extendable* if $\ell(w) > 0$ and *biextendable* if $e(w) > 0$. 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 left-special and right-special. For $w \in S$, we denote

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

A word w is called *neutral* if $m_S(w) = 0$. We say that a set S is *neutral* if it is factorial and every nonempty word $w \in S$ is neutral. The *characteristic* of S is the integer $\chi(S) = 1 - m_S(\varepsilon)$.

A neutral set of characteristic 1, simply called a neutral set, is such that all words (including the empty word) are neutral.

The following is a trivial example of a neutral set of characteristic 2.

Example 2.1. Let $A = \{a, b\}$ and let S be the set of factors of $(ab)^*$. Then S is neutral of characteristic 2.

As a more interesting example, any Sturmian set is a neutral set [5] (by a Sturmian set, we mean the set of factors of a strict episturmian word, see [11]).

The following example is the classical example of a Sturmian set.

Example 2.2. Let $A = \{a, b\}$ and let $f : A^* \to A^*$ be the *Fibonacci morphism* defined by $f(a) = ab$ and $f(b) = a$. The infinite word $x = \lim_{n\to\infty} f^n(a)$ is the *Fibonacci word*. One has $x = abaababa\cdots$. The *Fibonacci set* is the set of factors of the Fibonacci word. It is a Sturmian set, and thus a neutral set.

The *factor complexity* of a factorial set S of words on an alphabet A is the sequence $p_n = \mathrm{Card}(S \cap A^n)$. The complexity of a Sturmian set is $p_n = n(\mathrm{Card}(A) - 1) + 1$. The following result (see [10]) shows that a neutral set has linear complexity.

Proposition 2.1 *The factor complexity of a neutral set on k letters is given by $p_0 = 1$ and $p_n = n(k - \chi(S)) + \chi(S)$ for every $n \geq 1$.*

Example 2.3. The complexity of the set of Example 2.1 is $p_n = 2$ for any $n \geq 1$.

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

2.2 Tree sets

Let S be a biextendable set of words. For $w \in S$, we consider the set $E(w)$ as an undirected graph on the set of vertices which is the disjoint union of $L(w)$ and $R(w)$ with edges the pairs $(a, b) \in E(w)$. This graph is called the *extension graph* of w. We sometimes denote $1 \otimes L(w)$ and $R(w) \otimes 1$ the copies of $L(w)$ and $R(w)$ used to define the set of vertices of $E(w)$. We note that since $E(w)$ has $\ell(w) + r(w)$ vertices and $e(w)$ edges, the number $1 - m_S(w)$ is the Euler characteristic of the graph $E(w)$.

A biextendable set S is called a *tree set* of characteristic c if for any nonempty $w \in S$, the graph $E(w)$ is a tree and if $E(\varepsilon)$ is a union of c trees. Note that a tree set of characteristic c is a neutral set of characteristic c.

Example 2.4. The set S of Example 2.1 is a tree set of characteristic 2.

A tree set of characteristic 1, simply called a tree set as in [5], is such that $E(w)$ is a tree for any $w \in S$.

As an example, a Sturmian set is a tree set [5] .

Example 2.5. Let $A = \{a, b\}$ and let $f : A^* \to A^*$ be the morphism defined by $f(a) = ab$ and $f(b) = ba$. The infinite word $x = \lim_{n\to\infty} f^n(a)$ is the *Thue-Morse word*. The *Thue-Morse set* is the set of factors of the Thue-Morse word. It is uniformly recurrent but it is not a tree set since $E(\varepsilon) = A \times A$.

Let S be a set of words. For $w \in S$, let $\Gamma_S(w) = \{x \in S \mid wx \in S \cap A^+w\}$. If S is recurrent, the set $\Gamma_S(w)$ is nonempty. Let

$$\mathrm{Ret}_S(w) = \Gamma_S(w) \setminus \Gamma_S(w)A^+$$

be the set of *return words* to w.

Note that a recurrent set S is uniformly recurrent if and only if the set $\mathrm{Ret}_S(w)$ is finite for any $w \in S$. Indeed, if N is the maximal length of the words in $\mathrm{Ret}_S(w)$ for a word w of length n, any word in S of length $N + n$ contains an occurrence of w. The converse is obvious.

We will use the following result [5, Theorem 4.5] . We denote by F_A the free group on A.

Theorem 2.2 (Return Theorem). *Let S be a uniformly recurrent tree set. For any $w \in S$, the set $\mathrm{Ret}_S(w)$ is a basis of the free group F_A.*

Note that this result implies in particular that for any $w \in S$, the set $\mathrm{Ret}_S(w)$ has $\mathrm{Card}(A)$ elements.

Example 2.6. Let S be the Tribonacci set. It is the set of factors of the infinite word $x = abacaba\cdots$ which is the fixed point of the morphism f defined by $f(a) = ab$, $f(b) = ac$, $f(c) = a$. It is a Sturmian set (see [11]). We have $\mathrm{Ret}_S(a) = \{a, ba, ca\}$.

3 Automata

All automata considered in this paper are deterministic and strongly connected and we simply call them automata. An automaton on a finite set Q of states is given by a partial map from $Q \times A$ into Q denoted $p \mapsto p \cdot a$, and extended to words with the same notation. For a word w, we denote by $\varphi_{\mathcal{A}}$ the map $p \in Q \mapsto p \cdot w \in Q$.

The *transition monoid* of the automaton $\mathcal{A}$ is the monoid M of partial maps from Q to itself of the form $\varphi_{\mathcal{A}}(w)$ for $w \in A^*$. The rank of an element m of M is the cardinality of its image, denoted $\mathrm{Im}(m)$.

Let $\mathcal{A}$ be an automaton and let S be a set of words. Denote by $\mathrm{rank}_{\mathcal{A}}(w)$ the rank of the map $\varphi_{\mathcal{A}}(w)$, also called the *rank* of w with respect to the automaton $\mathcal{A}$. The *S-minimal rank* of $\mathcal{A}$ is the minimal value of $\mathrm{rank}_{\mathcal{A}}(w)$ for $w \in S$. It is denoted $\mathrm{rank}_{\mathcal{A}}(S)$. A word of rank 1 is called *synchronizing*.

The following result gives a method to compute $\mathrm{rank}_{\mathcal{A}}(S)$ and thus gives a method to decide if $\mathcal{A}$ admits synchronizing words.

Theorem 3.1. *Let S be a recurrent set and let $\mathcal{A}$ be an automaton. Let w be in S and let $I = \mathrm{Im}(w)$. Then w has rank equal to $\mathrm{rank}_{\mathcal{A}}(S)$ if and only if $\mathrm{rank}_{\mathcal{A}}(wz) = \mathrm{rank}_{\mathcal{A}}(w)$ for any $z \in \mathrm{Ret}_S(w)$.*

Proof. Assume first that $\mathrm{rank}_{\mathcal{A}}(w) = \mathrm{rank}_{\mathcal{A}}(S)$. If z is in $\mathrm{Ret}_S(w)$, then wz is in S. Since $\mathrm{rank}_{\mathcal{A}}(wz) \leq \mathrm{rank}_{\mathcal{A}}(w)$ and since $\mathrm{rank}_{\mathcal{A}}(w)$ is minimal, this forces $\mathrm{rank}_{\mathcal{A}}(wz) = \mathrm{rank}_{\mathcal{A}}(w)$.

Conversely, assume that w satisfies the condition. For any $r \in \mathrm{Ret}_S(w)$, we have $I \cdot r = \mathrm{Im}(wr) \subset \mathrm{Im}(w) = I$. Since $\mathrm{rank}_{\mathcal{A}}(wr) = \mathrm{rank}_{\mathcal{A}}(w)$, this forces $I \cdot r = I$. Since $\Gamma_S(w) \subset \mathrm{Ret}_S(w)^*$, this proves that

$$\Gamma_S(w) \subset \{z \in S \mid I \cdot z = I\}. \tag{3.1}$$

Let u be a word of S of minimal rank. Since S is recurrent, there exists words v, v' such that $wvuv'w \in S$. Then $vuv'w$ is in $\Gamma_S(w)$ and thus $I \cdot vuv'w = I$ by (3.1). This implies that $\mathrm{rank}_{\mathcal{A}}(u) \geq \mathrm{rank}_{\mathcal{A}}(vuv'w) = \mathrm{rank}_{\mathcal{A}}(w)$. Thus w has minimal rank in S.

Theorem 3.1 can be used to compute the S-minimal rank of an automaton in an effective way for a uniformly recurrent set S provided one can compute effectively the finite sets $\mathrm{Ret}_S(w)$ for $w \in S$.

Example 3.1. Let S be the Fibonacci set and let $\mathcal{A}$ be the automaton given by its transitions in Figure 3.1 on the left. One has $\mathrm{Im}(a^2) = \{1, 2, 4\}$. The action on the 3-element sets of states of the automaton is shown on the right. By Theorem 3.1, we obtain $\mathrm{rank}_{\mathcal{A}}(S) = 3$.

	1	2	3	4	5	6	7	8	9
a	2	4	1	–	7	8	4	1	3
b	3	5	–	6	–	–	9	1	–

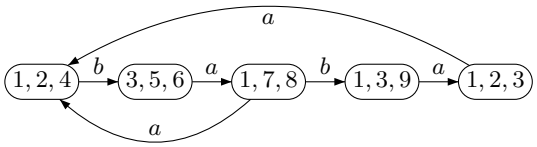


Fig. 3.1. An automaton of S-degree 3.

We denote by $\mathcal{L}, \mathcal{R}, \mathcal{D}, \mathcal{H}$ the usual Green relations on a monoid M (see [4]). Recall that $\mathcal{R}$ is the equivalence on M defined by $m\mathcal{R}n$ if $mM = nM$. The $\mathcal{R}$-class of m is denoted $R(m)$. Symmetrically, one denotes by $\mathcal{L}$ the equivalence defined by $m\mathcal{L}n$ if $Mm = Mn$. It is well-known that the equivalences $\mathcal{R}$ and $\mathcal{L}$ commute. The equivalence $\mathcal{R}\mathcal{L} = \mathcal{L}\mathcal{R}$ is denoted $\mathcal{D}$. Finally, one denotes by $\mathcal{H}$ the equivalence $\mathcal{R} \cap \mathcal{L}$.

The following result is proved in [3] in a particular case (that is, for an automaton recognizing the submomoid generated by a bifix code).

Proposition 3.2 *Let S be a recurrent set and $\mathcal{A}$ be a strongly connected automaton. Set $\varphi = \varphi_{\mathcal{A}}$ and $M = \varphi(A^*)$. The set of elements of $\varphi(S)$ of rank $\mathrm{rank}_{\mathcal{A}}(S)$ is included in a regular $\mathcal{D}$-class of M.*

Proof. Set $d = \mathrm{rank}_{\mathcal{A}}(S)$. Let $u, v \in S$ be two words of rank d. Set $m = \varphi(u)$ and $n = \varphi(v)$. Let w be such that $uwv \in S$. We show first that $m\mathcal{R}\varphi(uwv)$ and $n\mathcal{L}\varphi(uwv)$.

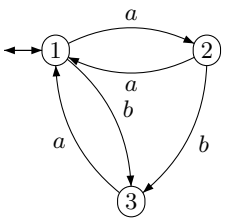
For this, let t be such that $uwvtu \in S$. Set $z = wvtu$. Since $uz \in S$, the rank of uz is d. Since $\mathrm{Im}(uz) \subset \mathrm{Im}(z) \subset \mathrm{Im}(u)$, this implies that the images are equal. Consequently, the restriction of $\varphi(z)$ to $\mathrm{Im}(u)$ is a permutation. Since $\mathrm{Im}(u)$ is finite, there is an integer $\ell \geq 1$ such that $\varphi(z)^\ell$ is the identity on $\mathrm{Im}(u)$. Set $e = \varphi(z)^\ell$ and $s = tuz^{\ell-1}$. Then, since e is the identity on $\mathrm{Im}(u)$, one has $m = me$. Thus $m = \varphi(uwv)\varphi(s)$, and since $\varphi(uwv) = m\varphi(wv)$, it follows that m and $\varphi(uwv)$ are $\mathcal{R}$-equivalent.

Similarly n and $\varphi(uwv)$ are $\mathcal{L}$-equivalent. Indeed, let t' be such that $vt'uwv \in S$. Set $z' = t'uwv$. Then $\mathrm{Im}(vz') \subset \mathrm{Im}(z') \subset \mathrm{Im}(v)$. Since vz' is a factor of z^2 and z has rank d, it follows that $d = \mathrm{rank}(z^2) \leq \mathrm{rank}(vz') \leq \mathrm{rank}(v) = d$. Therefore, vz' has rank d and consequently the images $\mathrm{Im}(vz')$, $\mathrm{Im}(z')$ and $\mathrm{Im}(v)$ are equal. There is an integer $\ell' \geq 1$ such that $\varphi(z')^{\ell'}$ is the identity on $\mathrm{Im}(v)$. Set $e' = \varphi(z')^{\ell'}$. Then $n = ne' = n\varphi(z')^{\ell'-1}\varphi(tuwv) = nq\varphi(uwv)$, with $q = \varphi(z')^{\ell'-1}\varphi(t)$. Since $\varphi(uwv) = \varphi(uw)n$, one has $n\mathcal{L}\varphi(uwv)$. Thus m, n are $\mathcal{D}$-equivalent, and $\varphi(uwv) \in R(m) \cap L(n)$.

Set $p = \varphi(wv)$. Then $p = \varphi(w)n$ and, with the previous notation, $n = ne' = nq\varphi(u)p$, so $L(n) = L(p)$. Thus $mp = \varphi(uwv) \in R(m) \cap L(p)$, and by Clifford and Miller's Lemma, $R(p) \cap L(m)$ contains an idempotent. Thus the $\mathcal{D}$-class of m, p and n is regular.

The $\mathcal{D}$-class containing the elements of $\varphi(S)$ of rank $\mathrm{rank}_{\mathcal{A}}(S)$ is called the *S-minimal* $\mathcal{D}$-class of M. This $\mathcal{D}$-class appears in a different context in [12] (for a survey concerning the use of Green's relations in automata theory, see [9]).

Example 3.2. Let S be the Fibonacci set and let $\mathcal{A}$ be the automaton represented in Figure 3.2 on the left. The S-minimal $\mathcal{D}$-class of the transition monoid of $\mathcal{A}$ is represented in Figure 3.2 on the right.



	3	1	2
1, 2	b	$^*ba$	$^*ba^2$
1, 2, 3	$^*ab$	*	*

Fig. 3.2. The automaton $\mathcal{A}$ and the S-minimal $\mathcal{D}$-class

Thus $\mathrm{rank}_{\mathcal{A}}(S) = 1$. We indicate with a $*$ the $\mathcal{H}$-classes containing an idempotent.

Let us recall some notions concerning groups in transformation monoids (see [3] for a more detailed presentation). Let M be a transformation monoid on a set Q. For $I \subset Q$, we denote

$$\mathrm{Stab}_M(I) = \{x \in M \mid Ix = I\}$$

or $\mathrm{Stab}(I)$ if the monoid M is understood. The *holonomy group* of M relative to I is the restriction of the elements of $\mathrm{Stab}_M(I)$ to the set I. It is denoted $\mathrm{Group}(I)$.

Let D be a regular $\mathcal{D}$-class in a transformation monoid M on a set Q. The holonomy groups of M relative to the sets Qm for $m \in D$ are all equivalent. The *structure group* of D is any of them.

Let $\mathcal{A}$ be an automaton with Q as set of states and let $I \subset Q$. Let w be a word such that $\varphi_{\mathcal{A}}(w) \in \mathrm{Stab}(I)$. The restriction of $\varphi_{\mathcal{A}}(w)$ to I is a permutation which belongs to $\mathrm{Group}(I)$. It is called the permutation *defined* by the word w on the set I.

Let $\mathcal{A}$ be a strongly connected automaton and let S be a recurrent set of words. The *S-group* of $\mathcal{A}$ is the structure group of its S-minimal $\mathcal{D}$-class. It is denoted $G_{\mathcal{A}}(S)$.

For the set $S = A^*$ and a strongly connected automaton, the group $G_{\mathcal{A}}(S)$ is a transitive permutation group of degree $d_X(S)$ (see [4, Theorem 9.3.10]). We conjecture that it holds for a uniformly recurrent tree set. It is not true for any uniformly recurrent set S, as shown in the following examples.

Example 3.3. Let S be the set of factors of $(ab)^*$ and let $\mathcal{A}$ be the automaton of Figure 3.3. The minimal S-rank of $\mathcal{A}$ is 2 but the group $G_{\mathcal{A}}(S)$ is trivial.

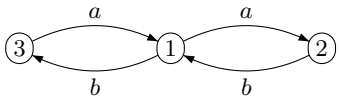


Fig. 3.3. An automaton of S-rank 2 with trivial S-group

Example 3.4. Let S be the Thue-Morse set and let $\mathcal{A}$ be the automaton represented in Figure 3.4 on the left. The word aa has rank 3 and image $I = \{1, 2, 4\}$.

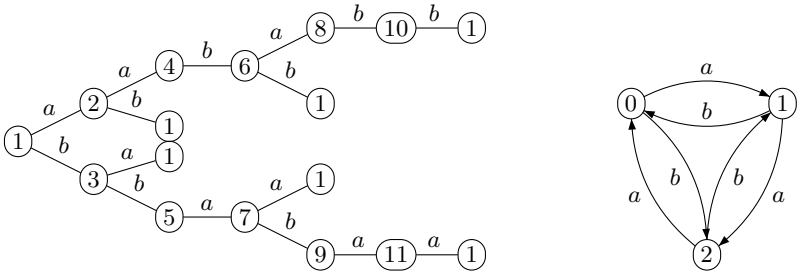


Fig. 3.4. An automaton of S-degree 3 with trivial S-group

The action on the images accessible from I is given in Figure 3.5. All words

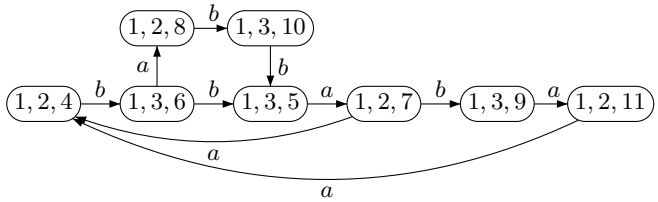


Fig. 3.5. The action on the minimal images

with image $\{1, 2, 4\}$ end with aa. The paths returning for the first time to $\{1, 2, 4\}$ are labeled by the set $\mathrm{Ret}_S(aa) = \{b^2a^2, bab^2aba^2, bab^2a^2, b^2aba^2\}$. Thus $\mathrm{rank}_{\mathcal{A}}(S) = 3$ by Theorem 3.1. Moreover each of the words of $\mathrm{Ret}_S(a^2)$ defines the trivial permutation on the set $\{1, 2, 4\}$. Thus $G_{\mathcal{A}}(S)$ is trivial.

The fact that $d_{\mathcal{A}}(S) = 3$ and that $G_{\mathcal{A}}(S)$ is trivial can be seen directly as follows. Consider the group automaton $\mathcal{B}$ represented in Figure 3.4 on the right and corresponding to the map sending each word to the difference modulo 3 of the number of occurrences of a and b. There is a reduction ρ from $\mathcal{A}$ onto $\mathcal{B}$ such that $1 \mapsto 0$, $2 \mapsto 1$, and $4 \mapsto 2$. This accounts for the fact that $d_{\mathcal{A}}(S) = 3$. Moreover, one may verify that any return word x to a^2 has equal number of a and b (if $x = uaa$ then $aauaa$ is in S, which implies that aua and thus uaa have

the same number of a and b). This implies that the permutation $\varphi_{\mathcal{B}}(x)$ is the identity, and therefore also the restriction of $\varphi_{\mathcal{A}}(x)$ to I. The same argument holds for Example 3.3 by considering the parity of the length.

4 Codes

A *code* is a set X such that for any $n, m \geq 0$ any $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ in X, one has $x_1 \cdots x_n = y_1 \cdots y_m$ only if $n = m$ and $x_1 = y_1$,..., $x_n = y_n$. A *prefix code* is a set X 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.

Let S be a set of words. A prefix code $X \subset S$ is said to be *S-maximal* if it is not properly contained in any prefix code $Y \subset S$. The notion of an S-maximal suffix or bifix code are symmetrical.

It follows from results of [3] that for a recurrent set S, a finite bifix code $X \subset S$ is S-maximal as a bifix code if and only if it is S-maximal as a prefix code.

Given a set $X \subset S$, we denote $\lambda_S(X) = \sum_{x \in X} \lambda_S(x)$ where λ_S is the map defined by $\lambda_S(x) = e_S(x) - r_S(x)$. The following result is [10, Proposition 4].

Proposition 4.1 *Let S be a neutral set of characteristic c on the alphabet A, and let X be a finite S-maximal prefix code. Then $\lambda_S(X) = \mathrm{Card}(A) - c$.*

Symmetrically, one denotes $\rho_S(x) = e_S(x) - \ell_S(x)$. The dual of Proposition 4.1 holds for suffix codes instead of prefix codes with ρ_S instead of λ_S.

Note that when S is Sturmian, one has $\lambda_S(x) = \mathrm{Card}(A) - 1$ if x is left-special and $\lambda_S(x) = 0$ otherwise. Thus Proposition 4.1 expresses the fact that any finite S-maximal prefix code contains exactly one left-special word [3, Proposition 5.1.5].

Example 4.1. Let S be the Fibonacci set and let $X = \{aa, ab, b\}$. The set X is an S-maximal prefix code. It contains exactly one left-special word, namely ab. Accordingly, one has $\lambda_S(X) = 1$.

Let S be a factorial set and let $X \subset S$ be a finite prefix code. The *S-degree* of X is the S-minimal rank of the minimal automaton of X^*. It is denoted $d_X(S)$.

When X is a finite bifix code, the S-degree can be defined in a different way. A *parse* of a word w is a triple (s, x, p) such that $w = sxp$ with $s \in A^* \setminus A^*X$, $x \in X^*$ and $p \in A^* \setminus XA^*$. For a recurrent set S and an S-maximal bifix code X, $d_X(S)$ is the maximal number of parses of a word of S. A word $w \in S$ has $d_X(S)$ parses if and only if it is not an internal factor of a word of X (see [3]).

The following result is [6, Theorem 4.4].

Theorem 4.2 (Finite Index Basis Theorem). *Let S be a uniformly recurrent tree set and let $X \subset S$ be a finite bifix code. Then X 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 F_A.*

Note that the result implies that any S-maximal bifix code of S-degree n has $d(\mathrm{Card}(A)-1)+1$ elements. Indeed, by Schreier's Formula, a subgroup of index d of a free group of rank r has rank $d(r-1)+1$.

Example 4.2. Let S be a Sturmian set. For any $n \geq 1$, the set $X = S \cap A^n$ is an S-maximal bifix code of S-degree n. According to theorem 4.2, it is a basis of the subgroup which is the kernel of the group morphism from F_A onto the additive group $\mathbb{Z}/n\mathbb{Z}$ sending each letter to 1.

The following statement generalizes [3, Theorem 4.3.7] where it is proved for a bifix code (and in this case with a stronger conclusion).

Theorem 4.3. *Let S be a recurrent set and let X be a finite S-maximal prefix code of S-degree n. The set of nonempty proper prefixes of X contains a disjoint union of $n-1$ S-maximal suffix codes.*

Proof. Let P be the set of proper prefixes of X. Any word of S of rank n of length larger than the words of X has n suffixes which are in P.

We claim that this implies that any word in S is a suffix of a word with at least n suffixes in P. Indeed, let $x \in S$ be of minimal rank. For any $w \in S$, since S is recurrent, there is some u such that $xuw \in S$. Then xuw is of rank n and has n suffixes in P. This proves the claim.

Let Y_i for $1 \leq i \leq n$ be the set of $p \in P$ which have i suffixes in P. One has $Y_1 = \{\varepsilon\}$ and each Y_i for $2 \leq i \leq d$ is clearly a suffix code. It follows from the claim above that it is S-maximal. Since the Y_i are also disjoint, the result follows.

Corollary 1. *Let S be a recurrent neutral set of characteristic c, and let X be a finite S-maximal prefix code of S-degree n. The set P of proper prefixes of X satisfies $\rho_S(P) \geq n(\mathrm{Card}(A)-c)$.*

Proof. By Theorem 4.3, there exist $n-1$ pairwise disjoint S-maximal suffix codes Y_i $(2 \leq i \leq n)$ such that P contains all Y_i. By the dual of Proposition 4.1, we have $\rho_S(Y_i) = \mathrm{Card}(A) - c$ for $2 \leq i \leq n$. Since $\rho_S(\varepsilon) = e_S(\varepsilon) - \ell_S(\varepsilon) = m_S(\varepsilon)+r_S(\varepsilon)-1 = \mathrm{Card}(A)-c$, we obtain $\rho_S(P) \geq \rho_S(\varepsilon)+(n-1)(\mathrm{Card}(A)-c) = n(\mathrm{Card}(A)-c)$.

4.1 A cardinality theorem for prefix codes

Theorem 4.4. *Let S be a uniformly recurrent neutral set of characteristic c. Any finite S-maximal prefix code has at least $d_X(S)(\mathrm{Card}(A)-c)+1$ elements.*

Proof. Let P be the set of proper prefixes of X. We may identify X with the set of leaves of a tree having P as set of internal nodes, each having $r_S(p)$ sons. By a well-known argument on trees, we have $\mathrm{Card}(X) = 1 + \sum_{p \in P}(r_S(p)-1)$. Thus $\mathrm{Card}(X) = 1 + \rho_S(P)$. By Corollary 1, we have $\rho_S(P) \geq n(\mathrm{Card}(A)-c)$.

The next example shows that the prefix code can have strictly more than $d_X(S)(\mathrm{Card}(A)-c)+1$ elements.

Example 4.3. Let S be the Fibonacci set. Let X be the S-maximal prefix code represented in Figure 4.1. The states of the minimal automaton of X^* are represented on the figure. The automaton coincides with that of Example 3.1. Thus

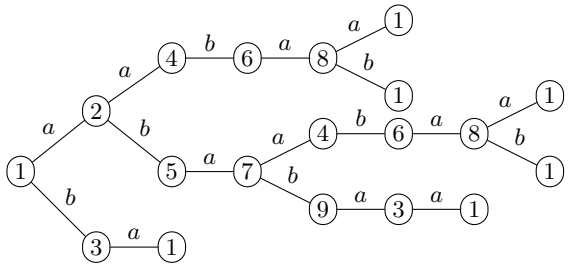


Fig. 4.1. A prefix code of S-degree 3

$d_X(S) = 3$ and $\mathrm{Card}(X) = 6$ while $d_X(S)(\mathrm{Card}(A) - 1) + 1 = 4$.

If X is bifix, then it has $d_X(S)(\mathrm{Card}(A) - c) + 1$ elements by a result of [10]. The following example shows that an S-maximal prefix code can have $d_X(S)(\mathrm{Card}(A) - c) + 1$ elements without being bifix.

Example 4.4. Let S be the Fibonacci set and let

$$X = \{aaba, ab, ba\}.$$

The literal automaton of X^* is represented in Figure 4.2 on the left. The prefix

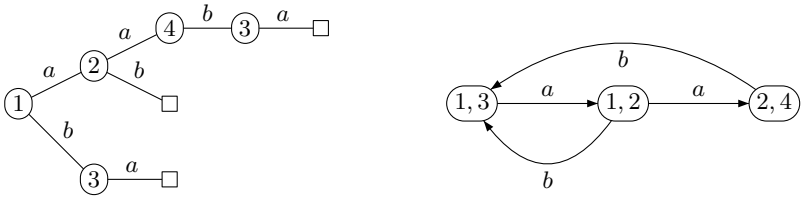


Fig. 4.2. The S-maximal prefix code X and the action on 2-subsets.

code X is S-maximal. The word ab has rank 2 in the literal automaton of X^*. Indeed, $\mathrm{Im}(ab) = \{1, 3\}$. Moreover $R_S(ab) = \{ab, aab\}$. The ranks of $abab$ and $abaab$ are also equal to 2, as shown in Figure 4.2 on the right. Thus the S-degree of X is 2 by Proposition 3.1. The code X is not bifix since ba is a suffix of $aaba$.

4.2 The group of a bifix code

The following result is proved in [3, Theorem 7.2.5] for a Sturmian set S. Recall that a *group code* of degree d is a bifix code Z such that $Z^* = \varphi^{-1}(K)$ for a

surjective morphism φ from A^* onto a finite group G and a subgroup K of index d in G. Equivalently, a bifix code Z is a group code if it generates the submonoid $H \cap A^*$ where H is a subgroup of index d of the free group F_A.

The S-group of a prefix code, denoted $G_X(S)$, is the group $G_{\mathcal{A}}(S)$ where $\mathcal{A}$ is the minimal automaton of X^*.

Theorem 4.5. *Let Z be a group code of degree d and let S be a uniformly recurrent tree set S. The set $X = Z \cap S$ is an S-maximal bifix code of S-degree d and $G_X(S)$ is equivalent to the representation of F_A on the cosets of the subgroup generated by X.*

Proof. The first part is [7, Theorem 5.10], obtained as a corollary of the Finite Index Basis Theorem. To see the second part, let H be the subgroup generated by X of the free group F_A. Consider a word $w \in S$ which is not an internal factor of X. Let P be the set of proper prefixes of X which are suffixes of w. Then P has d elements since for each $p \in P$, there is a parse of w of the form (s, x, p). Moreover P is a set of representatives of the right cosets of H. Indeed, let $p, q \in P$ and assume that $p = uq$ with $u \in S$. If $p \in Hq$, then $u \in X^* \cap S$. Since p cannot have a prefix in X, we conclude that $p = q$. Since H has index d, this implies the conclusion.

Let $\mathcal{A} = (Q, i, i)$ be the minimal automaton of X^*. Set $I = Q \cdot w$. Let $\mathrm{Stab}(I)$ be the set of words $x \in A^*$ such that $I \cdot x = I$. Note that $\mathrm{Stab}(I)$ contains the set $\mathrm{Ret}_S(w)$ of right return words to w. For $x \in \mathrm{Stab}(I)$, let $\pi(x)$ be the permutation defined by x on I. By definition, the group $G_X(S)$ is generated by $\pi(\mathrm{Stab}(I))$. Since $\mathrm{Stab}(I)$ contains $\mathrm{Ret}_S(w)$ and since $\mathrm{Ret}_S(w)$ generates the free group F_A, the set $\mathrm{Stab}(I)$ generates F_A.

Let $x \in \mathrm{Stab}(I)$. For $p, q \in I$, let $u, v \in P$ be such that $i \cdot u = p$, $i \cdot v = q$. Let us verify that

$$p \cdot x = q \Leftrightarrow ux \in Hv. \tag{4.1}$$

Indeed, let $t \in S$ be such that $vt \in X$. Then, one has $p \cdot x = q$ if and only if $uxt \in X^*$ which is equivalent to $ux \in Hv$. Since $\mathrm{Stab}(I)$ generates F_A, Equation (4.1) shows that the bijection $u \mapsto i \cdot u$ from P onto I defines an equivalence from $G_X(S)$ onto the representation of F_A on the cosets of H.

Example 4.5. Let S be the Fibonacci set and let $Z = A^2$ which is a group code of degree 2 corresponding to the morphism from A^* onto the additive $\mathbb{Z}/2\mathbb{Z}$ sending each letter to 1. Then $X = \{aa, ab, ba\}$. The minimal automaton of X^* is represented in Figure 4.3 on the left. The word a has 2 parses and its image

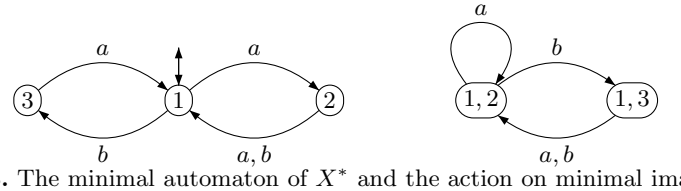


Fig. 4.3. The minimal automaton of X^* and the action on minimal images.

is the set $\{1, 2\}$. We have $\mathrm{Ret}_S(a) = \{a, ba\}$ and the action of $\mathrm{Ret}_S(a)$ on the

minimal images is indicated in Figure 4.3 on the right. The word a defines the permutation (12) and the word ba the identity.

Theorem 4.5 is not true for an arbitrary minimal set instead of a minimal tree set (see Example 3.4). The second part is true for an arbitrary finite S-maximal bifix code by the Finite Index Basis Theorem. We have no example where the second part is not true when X is S-maximal prefix instead of S-maximal bifix.

References

1. Jorge Almeida and Alfredo Costa. On the transition semigroups of centrally labeled Rauzy graphs. *Internat. J. Algebra Comput.*, 22(2):1250018, 25, 2012.
2. Jorge Almeida and Alfredo Costa. Presentations of Schützenberger groups of minimal subshifts. *Israel J. Math.*, 196(1):1–31, 2013.
3. Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012.
4. Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2009.
5. Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Acyclic, connected and tree sets. *Monatsh. Math.*, 176:521–550, 2015.
6. Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. The finite index basis property. *J. Pure Appl. Algebra*, 219:2521–2537, 2015.
7. Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. *Discrete Math.*, 338:725–742, 2015.
8. Arturo Carpi and Aldo de Luca. Codes of central Sturmian words. *Theoret. Comput. Sci.*, 340(2):220–239, 2005.
9. Thomas Colcombet. Green's relations and their use in automata theory. In *Language and Automata Theory and Applications - 5th International Conference, LATA 2011, Tarragona, Spain, May 26-31, 2011. Proceedings*, pages 1–21, 2011.
10. Francesco Dolce and Dominique Perrin. Enumeration formulæ in neutral sets. In *DLT 2015*, Springer LNCS. 2015. `http://arxiv.org/abs/1503.06081`.
11. Xavier Droubay, Jacques Justin, and Giuseppe Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255(1-2):539–553, 2001.
12. Dominique Perrin and Paul Schupp. Automata on the integers, recurrence, distinguishability, and the equivalence of monadic theories. In *LICS 1986*, pages 301–304, 1986.
13. Antonio Restivo. Codes and local constraints. *Theoret. Comput. Sci.*, 72(1):55–64, 1990.
14. Christophe Reutenauer. Ensembles libres de chemins dans un graphe. *Bull. Soc. Math. France*, 114(2):135–152, 1986.