Groups, bifix codes and minimal sets

Dominique Perrin

25 janvier 2016

Outline

- Uniformly recurrent sets
- Group of a bifix code
- Examples

Uniformly recurrent sets

A set F of finite words on A is called factorial if it contains the alphabet A and all the factors of its elements.
It is called extendable if for any word w in F, there are letters a, b such that $awb \in F$.
A factorial set $F \neq \{\varepsilon\}$ is recurrent (or irreducible) if for any $u, v \in F$, there is $w \in F$ such that $uwv \in F$.
An infinite factorial set is said to be uniformly recurrent (or minimal) if for any $w \in S$ there is an integer n such that w is a factor of any word of F of length n.

Pseudowords

We consider the free profinite monoid $\widehat{A^*}$.
Let F be a uniformly recurrent set of finite words on the alphabet A. The closure $\bar{F}$ of F in $\widehat{A^*}$ is also factorial.
As seen before, all the infinite pseudowords in the closure $\bar{F}$ of F are are $\mathcal{J}$-equivalent. We denote by $J(F)$ their $\mathcal{J}$-class.

Example

The Fibonacci morphism φ is primitive. The set F of factors of the words $\varphi^n(a)$ for $n \geq 1$ is called the Fibonacci set. It contains the infinitely recurrent pseudowords $\varphi^\omega(a)$ and $\varphi^\omega(b)$.

Return words

A return word to $x \in F$ is a nonempty word $w \in F$ which begins and ends by x but no internal factor of w has the same property. We denote by $\mathcal{R}_F(x)$ the set of return words to x.

Example

Let F be the Fibonacci set. The set of return words to a is $\mathcal{R}_F(a) = \{a, ba\}$. Similarly, $R_F(b) = \{ab, aab\}$.

Multiplicity

Let F be a factorial set on the alphabet A. For $w \in F$, we denote

$$\begin{aligned} L_F(w) &= \{a \in A \mid aw \in F\}, \\ R_F(w) &= \{a \in A \mid wa \in F\}, \\ E_F(w) &= \{(a,b) \in A \times A \mid awb \in F\} \end{aligned}$$

and further

$$\ell_F(w) = \mathrm{Card}(L_F(w)), \quad r_F(w) = \mathrm{Card}(R_F(w)), \quad e_F(w) = \mathrm{Card}(E_F(w)$$

For $w \in F$, we define the multiplicity of w as

$$m_F(w) = e_F(w) - \ell_F(w) - r_F(w) + 1.$$

Neutral sets

A word w is called neutral if $m_F(w) = 0$. A factorial set F is neutral if every word in F is neutral. By a result of (Balkova, Pelantova, Steiner, 2008), in a uniformly recurrent neutral set, one has

$$\mathrm{Card}(\mathcal{R}_F(x)) = \mathrm{Card}(A) \tag{1}$$

for every $x \in F$.

Return words

Let F be neutral set. For $x \in J(F)$, a return word to x is the limit of a sequence of return words to x_n for a sequence (x_n) of words of F converging to x.

Example

Let φ be the Fibonacci morphism and F be the Fibonacci set. It is a neutral set. The set of return words to a is $\mathcal{R}_F(a) = \{a, ba\}$. Accordingly, the set of return words to $\varphi^\omega(a)$ is $\{\varphi^\omega(a), \varphi^\omega(ba)\}$. Actually, any set of return words has two elements.

Schützenberger groups

Let x be an elementof a semigroup S and H be its $\mathcal{J}$-class. We denote by $G(x)$ the Schützenberger group of x. It is, by definition, the group of translations $\rho(z) : y \in H \mapsto yz \in H$ for all z such that $Hz = z$. When S is a topological semigroup, it is a topological group.

When H is a group, it is isomorphic to H.

Theorem (Almeida, Costa, 2013)

Let F be a uniformly recurrent neutral set. For any $x \in \mathcal{J}(F)$, the group $G(x)$ is the closure of the subgroup generated by any set of return words to x.

Example

Let φ be the Fibonacci morphism and let F be the Fibonacci set. Let $x = \varphi^\omega(a)$ and $y = \varphi^\omega(b)$. The group $G(x)$ is the closure of the group generated by x and yx, that is, isomorphic to the free profinite group $\widehat{FG(A)}$.

Tree sets

Let F be a factorial set of words. For $w \in F$, we consider the set $E_F(w)$ as an undirected graph on the set of vertices which is the disjoint union of $L_F(w)$ and $R_F(w)$ with edges the pairs $(a, b) \in E_F(w)$. This graph is called the extension graph of w.
A factorial set is a tree set is for every $x \in F$, the graph $E_F(x)$ is a tree. A tree set is neutral.

Example 1

The Fibonacci set is a tree set. This follows from the fact that it is a Sturmian set and that every Sturmian set is a tree set.

Example 2

The Tribonacci set is the set of factors of the fixed point of the morphism $\varphi : a \mapsto ab, b \mapsto ac, c \mapsto a$. It is a also a tree set. The graph $E(\varepsilon)$ is represented below.

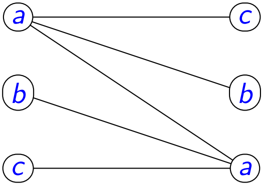


FIGURE: The extension graph of ε in the Tribonacci set.

The Return Theorem

Theorem (Berthe, De Felice, Dolce, Leroy, Perrin, Reutenauer, Rindone, 2013)

Let F be a uniformly recurrent tree set. For any $x \in F$, the set $R_F(x)$ is a basis of the free group $FG(A)$.

Example

Let F be the Tribonacci set on $A = \{a, b, c\}$. Then $R_F(a) = \{a, ba, ca\}$, which is easily seen to be a basis of $FG(A)$.

We denote by $G(F)$ the Schützenberger group of $J(F)$ (that is the Schützenberger group of any element of $J(F)$).

Theorem (Ameida, Costa, 2015)

Let F be a nonperiodic uniformly recurrent tree set on the alphabet A. The group $G(F)$ is the free profinite group on A.

Example

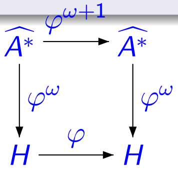
Let F be the Fibonacci set. We have seen that $G(F)$ is the free profinite group on A (Example 3).

Let $\varphi : A^* \to A^*$ be a primitive substitution and let $F(\varphi)$ be the set of factors of a fixed point of φ. We denote by $J(\varphi)$ the closure of $F(\varphi)$ and by $G(\varphi)$ the Schützenberger group of $J(\varphi)$.
A connexion for φ is a word ba with $b, a \in A$ such that $ba \in F(\varphi)$, the first letter of $\varphi^\omega(a)$ is a and the last letter of $\varphi^\omega(b)$ is b. Every primitive substitution has a connexion. A connective power of φ is a finite power $\tilde{\varphi}$ of φ such that the first letter of $\tilde{\varphi}(a)$ is a, the last letter of $\tilde{\varphi}(b)$ is b. We denote $X_\varphi(a, b) = aR_F(ba)a^{-1}$. The set $X_\varphi(a, b)$ is a code.
A substitution φ over A is proper if there are letters $a, b \in A$ such that for every $d \in A$, $\varphi(d)$ starts with a and ends with b.

Theorem (Almeida, Costa, 2013)

Let φ be a non-periodic proper primitive substitution over a finite alphabet A. Then $G(\varphi)$ admits the presentation

$$\langle A \mid \varphi_G^\omega(a) = a, a \in A\rangle_G.$$



Example

Let $A = \{a, b\}$ and let $\varphi : a \mapsto ab, b \mapsto a^3 b$. The Schützenberger group of $J(\varphi)$ has the presentation $\langle a, b \mid \varphi_G^\omega(a) = a, \varphi_G^\omega(b) = b \rangle$.

Bifix codes

Let F be a recurrent set. Given a bifix code $X \subset F$, a parse of a word w is a triple (s, x, p) such that

(i) $w = sxp$,
ii) s has no suffix in X,
(iii) $x \in X^*$,
(iv) p has no prefix in X.

The F-degree of X is the maximal number of parses of a word in F. It is finite for any finite F-maximal bifix code X.

Example

For any $n \geq 1$, the set $F \cap A^n$ is a finite F-maximal bifix code. It has F-degree n.

Cardinality Theorem

The following result shows that in a recurrent tree set, the cardinality of an F-maximal bifix code depends only of its F-degree.

Theorem (Berthé, De Felice, Dolce, Leroy, P., Reutenauer, Rindone, *J. Pure Appl. Algebra*, 2015)

Let F be a recurrent neutral set. For any finite F-maximal bifix code X, one has $\mathrm{Card}(X) = d_X(F)(\mathrm{Card}(A) - 1) + 1$.

The Finite Index Basis Theorem

Theorem (Berthé, De Felice, Dolce, Leroy, P., Reutenauer, Rindone, *J. Pure Appl. Algebra*, 2015)

Let F be a minimal tree set. A finite bifix code $X \subset F$ is F-maximal of F-degree d if and only if it is a basis of subgroup of finite index d of the free group on A.

This implies for tree sets the Cardinality Theorem because, by Schreier's Formula, a subgroup of index d of a free group of rank k has rank $d(k-1)+1$.

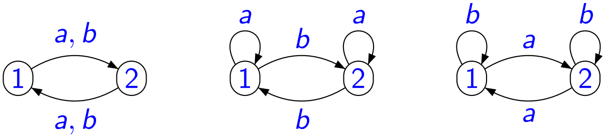
The proof uses the Return Theorem.

Examples

Let F be the Fibonacci set. The 3 F-maximal bifix codes of F-degree 2 are

$$aa, ab, ba,$$
$$a, bab, baab,$$
$$aa, aba, b$$

corresponding to the three subgroups of index 2



The F-degree of a prefix code

Let F be a factorial set and let $\mathcal{A}$ be a finite deterministic automaton. The rank of a word w is the number of states reached by w. The F-minimal rank of $\mathcal{A}$ is the minimal value of the rank of the words in F.
An automaton of F-minimal rank 1 is called F-synchronized.
Let $X \subset F$ be a finite F-maximal prefix code. The F-degree of X, denoted $d_X(F)$, is the F-minimal rank of the minimal automaton of X^*.

Example

Let F be the Fibonacci set. The X be the F-maximal prefix code represented below has F-degree 3. It is not bifix and it generates the free group.

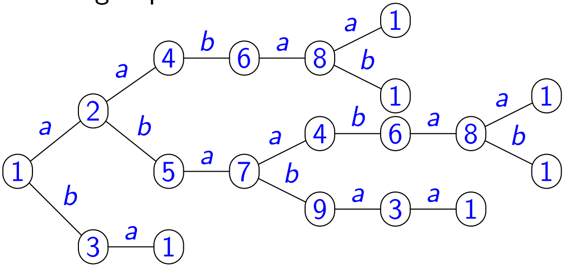
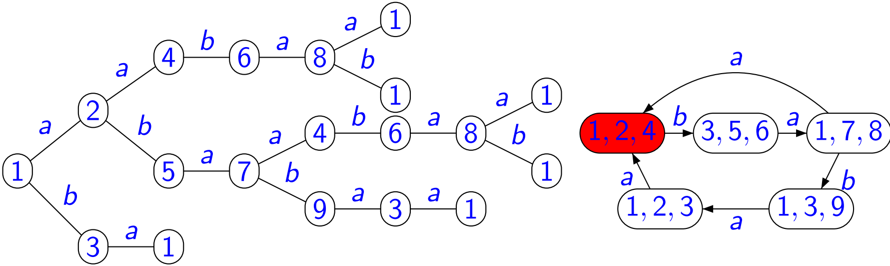


FIGURE: A prefix code of F-degree 3

The set $\mathrm{Im}(a^2)$ of states reachable by a^2 is $\mathrm{Im}(a^2) = \{1, 2, 4\}$. The action on the 3-element sets of states of the automaton is shown below (right).



Thus $d_X(F) = 3$.

The F-group of a prefix code

Let F be a recurrent set and let $\mathcal{A}$ be a deterministic automaton. The F-group of the automaton is obtained by choosing a word $w \in F$ of minimal rank. It is generated by the permutations on $I = \mathrm{Im}(w)$ realized by the return words to w (in terms of semigroup theory, it is the group of the $\mathcal{D}$-class of the elements of F of minimal rank).

Let X be a finite prefix code. The F-group of X, denoted $G_X(F)$ is the F-group of the minimal automaton of X^*.

The following is proved for a Sturmian set in (Berstel,De Felice, P., Reutenauer, Rindone, *J. Algebra*, 2012).

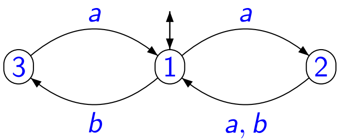
Theorem

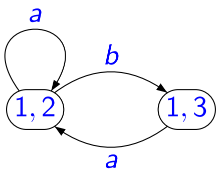
Let F be a uniformly recurrent tree set and let X be a finite F-maximal bifix code. The group $G_X(F)$ is equivalent to the representation of the free group F_A on the cosets of the group generated by X.

We conjecture that for a finite F-maximal prefix code X, the group $G_X(F)$ is transitive.

Example 0

Let F be the Fibonacci set and let $X = \{aa, ab, ba\}$. The minimal automaton of X^* is represented on the left.



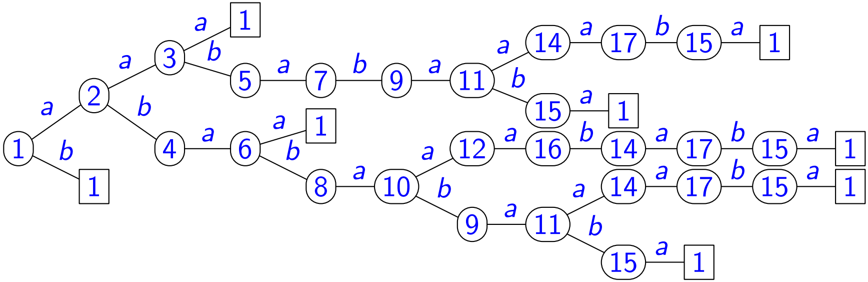


The word a has 2 parses and its image is the set $\{1, 2\}$. The action of the return words to a on the minimal images is indicated on the right. The word a defines the permutation (12) and the word ba the identity.

Example 1 : the alternating group A_5

Let F be the set of factors of a fixed point of $\varphi : a \mapsto ab, b \mapsto a^3 b$. We consider the morphim $h : A^* \to A_5$ from A^* onto the alternating group of degree 5 defined by $h : a \mapsto (123), b \mapsto (345)$. We denote by φ_G the map from G^A into itself defined for $f \in G^A$ and $a \in A$ by $\varphi_G(f)(a) = f(\varphi(a))$. We say that φ has finite f-order if there is an integer $n \geq 1$ such that $\varphi_G^n(f) = f$. The least such integer is called the *h-order* of φ.
The morphism φ has h-order 12 and thus, by a result of (Almeida,Costa, 2013), $\hat{h}$ induces a surjective map from any maximal subgroup of $J(\varphi)$ onto A_5.

Let Z be the bifix code generating the submonoid stabilizing 1 and let $X = Z \cap F$. The F-maximal bifix code X has 8 elements. It is represented below with the states of the minimal automaton indicated on its prefixes.



The $\mathcal{F}$-minimal $\mathcal{D}$-class

	1, 2, 3, 16, 17	1, 4, 5, 14, 15	1, 2, 6, 7, 17	1, 4, 8, 9, 15	1, 2, 6, 10
1/2, 4/3, 6, 15/ 8/9	a^3	a^3b	a^3ba	* a^3bab	
1/2/11, 17/6 7, 9, 10	ba^3		*		
1/3, 6, 15/9, 14 5, 8/4	aba^3	*			
1/11, 17/7, 16 3, 6/2	* $baba^3$			bab	
1/2, 4/9, 14 3, 6, 15/5, 12		*			
1/2, 4/3, 6, 15 10/11					*

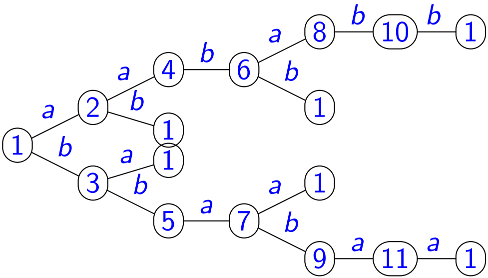
The word a^3 has rank 5 and $\mathcal{R}_F(a^3) = \{babaaa, babababaaa\}$. The corresponding permutations defined on the image $\{1, 2, 3, 16, 17\}$ of a^3 are respectively

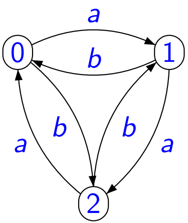
$$(1, 2, 16, 3, 17) \quad (1, 17, 16, 2, 3)$$

which generate A_5.

Example 2 : a trivial group

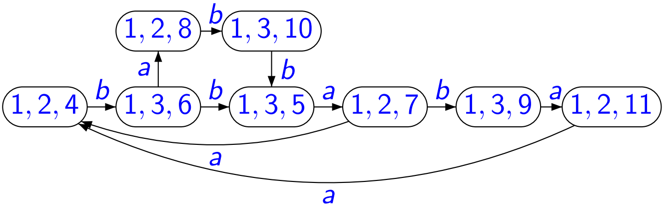
Let F be the Thue-Morse set and let $\mathcal{A}$ be the automaton represented below on the left.





The action on the minimal images

The word aa has rank 3 and image $I = \{1, 2, 4\}$. The action on the images accessible from I is given below.



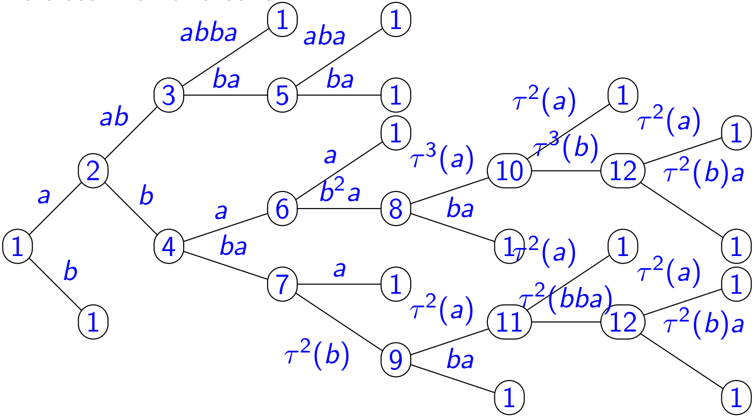
All words with image $\{1, 2, 4\}$ end with aa. The paths returning for the first time to $\{1, 2, 4\}$ are labeled by the set $\mathcal{R}_F(aa) = \{b^2a^2, bab^2aba^2, bab^2a^2, b^2aba^2\}$. Thus $\text{rank}_{\mathcal{A}}(F) = 3$. Moreover each of the words of $\mathcal{R}_F(a^2)$ defines the trivial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

Example 3

Consider again the Thue-Morse substitution τ and the Thue-Morse set F. Let h be the morphism $h : a \mapsto (123), b \mapsto (345)$ from A^* onto the alternating group A_5 (already used in Example 1). One may verify that τ has h-order 6 and thus, h extends to a surjective continuous morphim from any maximal subgroup of $J(\varphi)$ onto A_5. Let Z be the group code generating the submonoid stabilizing 1 and let $X = Z \cap F$.

The F-maximal bifix code X

We represent only the nodes corresponding to right special words, that is, vertices with two sons.



The image of $\tau^4(b)$ is $\{1,3,4,9,10\}$ and thus it is minimal. The action on its image is shown below. The return words to $\tau^4(b)$ are $\tau^4(b), \tau^3(a)$ and $\tau^5(ab)$. The permutations on the image of $\tau^4(b)$ are the 3 cycles of length 5 indicated below. Since they generate the group A_5, we have $G_X(F) = A_5$.