

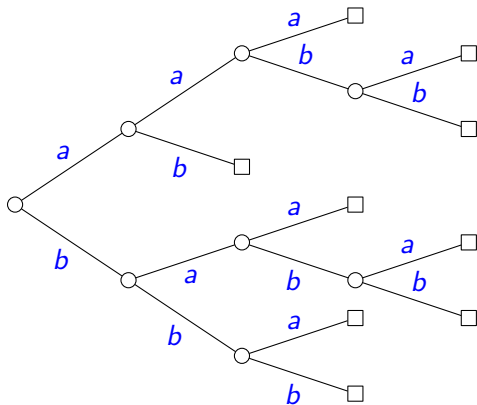
Bifix codes and Sturmian words

Dominique Perrin

2 mai 2011

A miracle

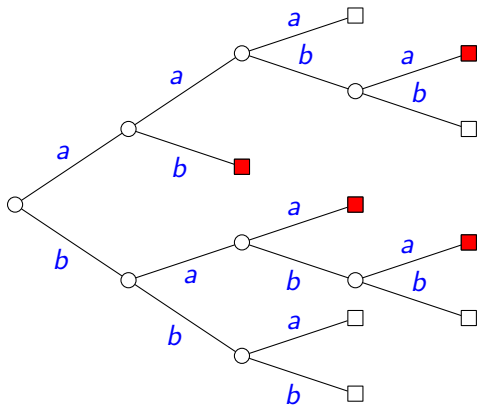
Consider the maximal bifix code of degree 3 below.



Let F be the set of factors of the Fibonacci word. The set $X \cap F$ (red nodes) has 4 elements.

A miracle

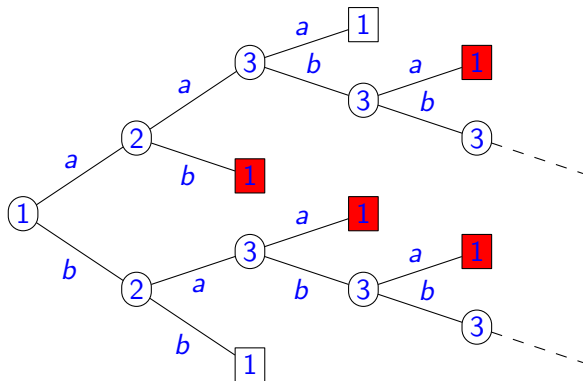
Consider the maximal bifix code of degree 3 below.



Let F be the set of factors of the Fibonacci word. The set $X \cap F$ (red nodes) has 4 elements.

Second miracle

Consider the group code of degree 3 below ($a \mapsto (123)$, $b \mapsto (12)$).



The intersection with F is the same as the previous one.

Outline

We show that

- in a Sturmian set F , any finite F -maximal bifix code of degree d on k letters has $(k - 1)d + 1$ elements (Cardinality Theorem).
- if an infinite word x is such that $\text{Card}(F(x) \cap X) \leq d$ for some finite maximal bifix code X of degree d , then x is ultimately periodic (Periodicity Theorem).
- in a Sturmian set, any finite F -maximal bifix code of F -degree d is a basis of a subgroup of index d of the free group on A and conversely (Sturmian Basis Theorem).

Based on **Bifix codes and Sturmian words**, by Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, Giuseppina Rindone (BDPRR, 2010).

- 1 Factorial sets
 - Recurrent sets
 - Sturmian sets
 - Bifix codes
- 2 Bifix codes in Sturmian sets
 - Cardinality
 - Periodicity
 - Sturmian Basis Theorem
- 3 Syntactic groups
 - Holonomy groups
 - The Syntactic Group Theorem

Factorial sets

A nonempty set $F \subset A^*$ of words is said to be **factorial** if it contains the factors of all its elements.

A set F is said to be **recurrent** if it is factorial and if for every $u, w \in F$ there is a $v \in F$ such that $uvw \in F$.

Example

Let $A = \{a, b\}$. Let F be the set of words on A without factor bb . Thus $F = A^* \setminus A^*bbA^*$. The set F is recurrent. Indeed, if $u, w \in F$, then $uaw \in F$.

Uniformly recurrent sets

A set F is said to be **uniformly recurrent** if

- it is factorial and for any $w \in F$ there is at least one letter $a \in A$ such that $wa \in F$,
- for any word $u \in F$, there exists an integer $n \geq 1$ such that u is a factor of every word in $F \cap A^n$.

Proposition

A uniformly recurrent set is recurrent.

The converse is not true.

Fixpoints of morphisms

Let $f : A^* \rightarrow A^*$ be a morphism (of monoids) and assume there is a letter $a \in A$ such that $f(a) \in aA^+$. The words $f^n(a)$ for $n \geq 1$ are prefixes of one another. If $|f^n(a)| \rightarrow \infty$ with n , then we denote by $f^\omega(a)$ the infinite word which has all $f^n(a)$ as prefixes. It is called a **fixpoint** of f .

Example

Set $A = \{a, b\}$. The **Thue–Morse morphism** is the morphism $f : A^* \rightarrow A^*$ defined by $f(a) = ab$ and $f(b) = ba$. The **Thue–Morse word** $x = abbabaab \cdots$ is the fixpoint $f^\omega(a)$ of f . It is uniformly recurrent.

Sturmian sets

Given a set F of words over an alphabet A , the right order of a word u in F is the number of letters a such that $ua \in F$. A word u is **right-special** if its right order is at least 2. A right-special word is **strict** if its right order is equal to $\text{Card}(A)$.

A set of words F is **Sturmian** if it is the set of factors of an infinite word and if

- it is closed under reversal
- it contains, for each $n \geq 1$, exactly one right-special word u of length n which is moreover strict.

It is easy to see that for a Sturmian set F on an alphabet A with k letters, the set $F \cap A^n$ has $(k - 1)n + 1$ elements for each n .

Example

Set $A = \{a, b\}$. The **Fibonacci set** is the set of factors of the infinite word

$$x = abaababaabaababaababaababaabaab \cdots$$

called the Fibonacci word. It is the fixpoint $f^\omega(a)$ of the morphism $f : A^* \rightarrow A^*$ defined by $f(a) = ab$ and $f(b) = a$.

Example

Set $A = \{a, b, c\}$. The morphism $f : A^* \rightarrow A^*$ defined by $f(a) = ab$, $f(b) = ac$ and $f(c) = a$ has the fixpoint

$$x = abacabaabacababacabaabacabacabaabacab \cdots$$

called the **Tribonacci word**. The set $F(x)$ is Sturmian.

Bifix codes

A set X of nonempty words is a **prefix code** if any two distinct elements of X are incomparable for the prefix order.

Example

The set $X = \{a, ba\}$ is a prefix code.

A set X of nonempty words is a **bifix code** if any two distinct elements of X are incomparable for the prefix order and for the suffix order.

Example

The set $X = \{a, bab\}$ is a bifix code.

Maximal bifix codes

A set $X \subset F$ is said to be F -thin, if there exists a word of F which is not a factor of a word in X .

A prefix code (resp. a bifix code) $X \subset F$ is F -maximal if it is not properly contained in any other prefix code (resp. bifix code) $Y \subset F$.

The following result is due to Schützenberger (1961) for $F = A^*$.

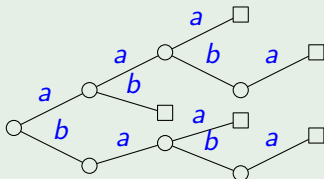
Theorem

Let F be a recurrent set and let $X \subset F$ be an F -thin set. The following conditions are equivalent.

- (i) X is an F -maximal bifix code.
- (ii) X is an F -maximal prefix code and a suffix code.

Example

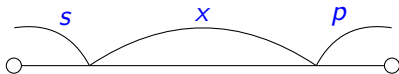
Let $A = \{a, b\}$ and let F be the set of words without factor bb . The set $X = \{aaa, aaba, ab, baa, baba\}$ is a finite F -maximal bifix code.



Parses

A **parse** of a word w with respect to a set X is a triple (s, x, p) such that

- s has no suffix in X ,
- $x \in X^*$
- p has no prefix in X
- $w = sxp$



For a word w , $\pi_X(w)$ denotes the number of parses of w .

Example

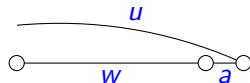
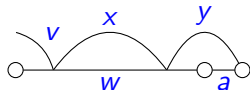
The set $X = \{a, bab\}$ is a finite bifix code. The parses of the word bab are $(1, bab, 1)$ and (b, a, b) . Thus $\pi_X(bab) = 2$.

Degree of a bifix code

Let X be a bifix code. For any word w and any letter $a \in A$

$$\pi_X(wa) = \begin{cases} \pi_X(w) & \text{if } wa \in A^*X \\ \pi_X(w) + 1 & \text{otherwise.} \end{cases}$$

As a consequence, the function π_X determines X .



The F -degree, denoted $d_F(X)$, of a bifix code X is the maximum of the number of parses of the words of F , that is

$$d_F(X) = \max_{w \in F} \pi_X(w).$$

Theorem

Let F be a recurrent set and let $X \subset F$ be a bifix code. Then X is an F -thin and F -maximal bifix code if and only if its F -degree is finite. In this case,

$$I(X) = \{w \in F \mid \pi_X(w) < d_F(X)\}$$

where $I(X)$ is the set of **internal** factors of X , that is the words v such that $uvw \in X$ for some nonempty words u, w .

Example

Let F be the Fibonacci set. The set $X = \{a, bab, baab\}$ is an F -maximal bifix code. The parses of the word bab are $(1, bab, 1)$ and (b, a, b) . Since bab is not an internal factor, one has $d_F(X) = 2$.

Example

Let F be the Fibonacci set. The set $X = \{aaba, ab, baa, baba\}$ is an F -maximal bifix code. It has F -degree 3. Indeed, the word $aaba$ has three parses $(1, aaba, 1)$, (a, ab, a) and $(aa, 1, ba)$ and it is in $F \setminus I(X)$.

The following was proved by Schützenberger (1961) for $F = A^*$.

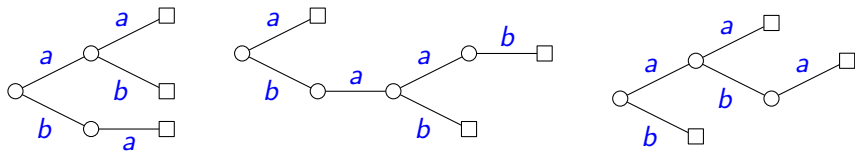
Theorem

For any recurrent set F and any integer $d \geq 1$ there is a finite number of finite F -maximal bifix codes $X \subset F$ of F -degree d .

For the Fibonacci set F , the numbers are :

d	1	2	3	4	5	6	7	8
	1	3	13	71	461	3447	29093	273343

This is Sloane's sequence A003319 : number of indecomposable permutations of $\{1, \dots, d\}$.



Example

Let $A = \{a, b\}$ and let $F \subset A^*$ be the Fibonacci set. There are three F -maximal bifix codes of degree 2 represented on the figure.

The following result shows that the case of a uniformly recurrent set contrasts with the case $F = A^*$ since in A^* , as soon as $\text{Card}(A) \geq 2$, there exist infinite maximal bifix codes of F -degree 2 and thus of all F -degrees $d \geq 2$.

Proposition

Let F be a uniformly recurrent set. Any F -thin bifix code $X \subset F$ is finite.

The Cardinality Theorem

The following result generalizes the fact that a Sturmian word has $d + 1$ factors of length d .

Theorem (BDPRR, 2010)

Let F be a Sturmian set on an alphabet with k letters. For any finite F -maximal bifix code $X \subset F$, one has

$$\text{Card}(X) = (k - 1)d_F(X) + 1.$$

Let $x = a_0a_1 \cdots$, with $a_i \in A$, be an infinite word. It is **periodic** if there is an integer $n \geq 1$ such that $a_{i+n} = a_i$ for all $i \geq 0$. It is **ultimately periodic** if the equalities hold for all i large enough. Thus, x is ultimately periodic if there is a word u and a periodic infinite word y such that $x = uy$. The following result, due to Coven and Hedlund, is well-known.

Theorem (Coven and Hedlund, 1973)

Let $x \in A^{\mathbb{N}}$ be an infinite word. If there exists an integer $d \geq 1$ such that x has at most d factors of length d then x is ultimately periodic.

The Periodicity Theorem

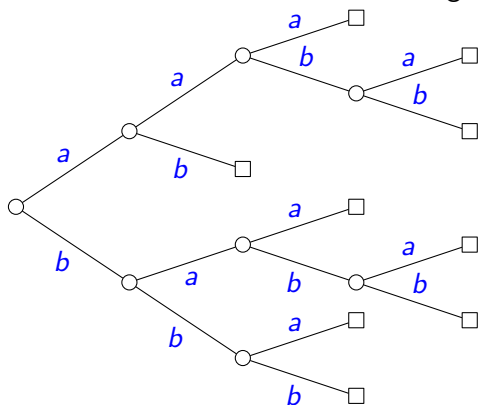
The following statement implies the Coven-Hedlund Theorem since A^d is a maximal bifix code of degree d .

Theorem (BDPRR, 2010)

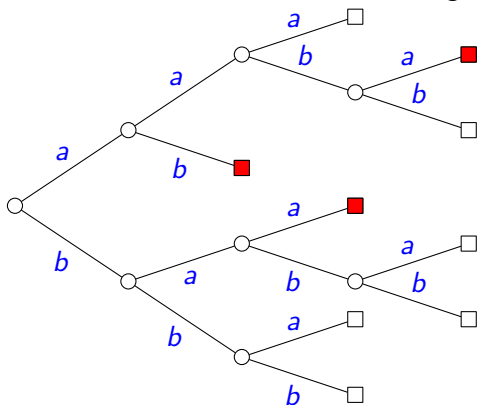
Let $x \in A^{\mathbb{N}}$ be an infinite word. If there exists a finite maximal bifix code X of degree d such that $\text{Card}(X \cap F(x)) \leq d$, then x is ultimately periodic.

The proof uses the Critical Factorization Theorem.

Consider the maximal bifix code of degree 3 below.



Consider the maximal bifix code of degree 3 below.



Assume that $X \cap F(x)$ is the set of red nodes. Then a factor aab can only be followed by a second aab . Thus $x = u(aab)^\omega$.

Sturmian Basis Theorem

Theorem (BDPRR, 2010)

Let F be a Sturmian set and let $d \geq 1$ be an integer. A bifix code $X \subset F$ is a basis of a subgroup of index d of A° if and only if it is a finite F -maximal bifix code of F -degree d .

Note that this contains the Cardinality Theorem. Indeed, by Schreier's formula, if H is a subgroup of rank n and index d of a free group of rank k , then

$$n - 1 = d(k - 1)$$

Let X be a F -maximal bifix code of F -degree d . By the above theorem, it is a basis of a subgroup of index d of the free group A° which has rank k . Thus $\text{Card}(X) = (k - 1)d + 1$ by Schreier's formula.

Before proving the Sturmian Basis Theorem, we list some corollaries.

Corollary

Let F be a Sturmian set. For any $n \geq 1$, the set $F \cap A^n$ is a basis of the subgroup of A° generated by A^n .

Direct proof : show by descending induction on $i = d, \dots, 0$ that for any $u \in F \cap A^i$, one has $uA^{d-i} \subset \langle X \rangle$. It is true for $i = d$. Next consider a right-special word $u \in F \cap A^i$. By induction hypothesis, we have $uaA^{d-i-1} \subset \langle X \rangle$ for any $a \in A$. Thus $uA^{d-i} \subset \langle X \rangle$. For another $v \in A^i$, let w be such that $vw \in F \cap A^d$. Then $vt = vw(uw)^{-1}ut$ for any $t \in A^{d-i}$.

Example

Let F be the Fibonacci set. We have $F \cap A^2 = \{aa, ab, ba\}$ and $bb = ba(aa)^{-1}ab$.

The following corollary contains the well-known fact that a subgroup of finite index of a free group has a positive basis.

Corollary

Let F be a Sturmian set. Any subgroup of finite index of the free group on A has a basis contained in F .

Let indeed H be a subgroup of index d of A° . Let Z be the bifix codes which generates the submonoid $H \cap A^*$. Then Z is a maximal bifix code of degree d . The set $X = Z \cap F$ is an F -maximal bifix code of degree $e \leq d$. By the Sturmian Basis Theorem, it is the basis of a subgroup K of index e . But then $K \subset H$ implies that d divides e . Thus $d = e$ and $H = K$.

As a further consequence of the Sturmian Basis Theorem, we have the following result.

Corollary

Let F be a Sturmian set on an alphabet with k letters. The number $N_{d,k}$ of finite F -maximal bifix codes $X \subset F$ of F -degree d satisfies $N_{1,k} = 1$ and

$$N_{d,k} = d(d!)^{k-1} - \sum_{i=1}^{d-1} [(d-i)!]^{k-1} N_{i,k}.$$

The formula results directly from the formula, due to Hall (1949), for the number of subgroups of index d in a free group of rank k . The values for $k = 2$ are given by the recurrence

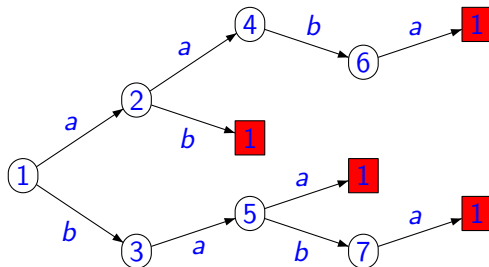
$$N_{d,2} = d d! - \sum_{i=1}^{d-1} (d-i)! N_{i,2}.$$

The first values are

d	1	2	3	4	5	6	7	8	9	10
$N_{d,2}$	1	3	13	71	461	3447	29093	273343	2829325	31998903

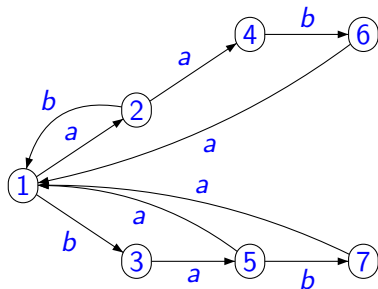
The formula is known to enumerate also the indecomposable permutations on $d + 1$ elements (see Dress, Franz 1985, Ossona, Rosenstiehl 2004 and Cori 2009).

Stallings foldings



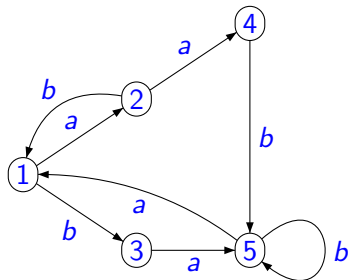
An F -maximal bifix code of F -degree 3.

Stallings foldings



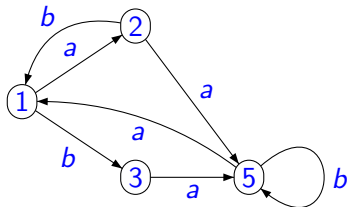
Fusion of 5, 6, 7.

Stallings foldings



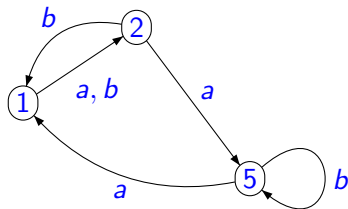
Fusion of 4, 5.

Stallings foldings



Fusion of 2, 3.

Stallings foldings



$a \mapsto (125)$, $b \mapsto (12)$.

Some preliminary results are used in the proof of the Sturmian Basis Theorem.

The first one is a **closure property** of the set $X^* \cap F$.

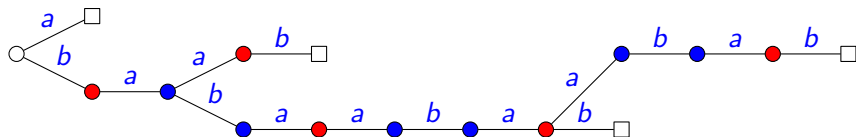
Proposition

Let F be a Sturmian set and let $X \subset F$ be a finite F -maximal bifix code. Then $\langle X \rangle \cap F = X^* \cap F$.

The incidence graph

Let X be a bifix code, let P be the set of its proper prefixes and S be the set of its proper suffixes. Set $P' = P \setminus 1$ and $S' = S \setminus 1$. The **incidence graph** of X is the undirected graph G defined as follows. The set of vertices is $V = 1 \otimes P' \cup S' \otimes 1$. The edges of G are the pairs $(1 \otimes p, s \otimes 1)$ and $(s \otimes 1, 1 \otimes p)$, for $p \in P'$ and $s \in S'$, such that $ps \in X$.

Consider the F -maximal bifix code of F -degree 3 in the Fibonacci set F given below.



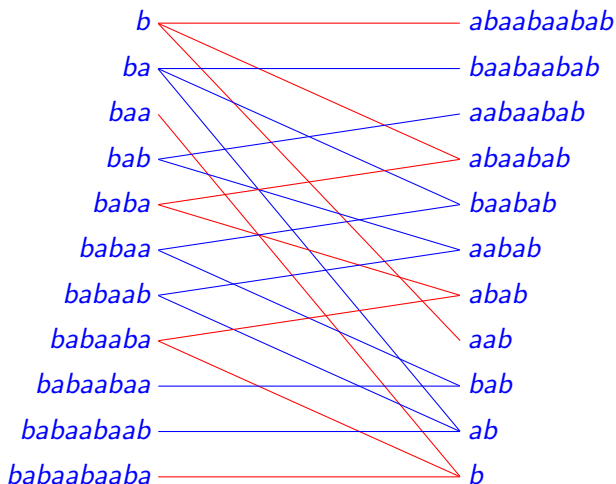


FIG.: The incidence graph of X .

Lemma

Let F be a Sturmian set and let $X \subset F$ be a bifix code. Let P' be the set of nonempty proper prefixes of X and let G be the incidence graph of X . The trace on P' of a connected component of G is a suffix code.

Consider the code X of the previous example. The two suffix-codes which are the traces of the incidence graph on the set of nonempty proper prefixes of X are shown below.

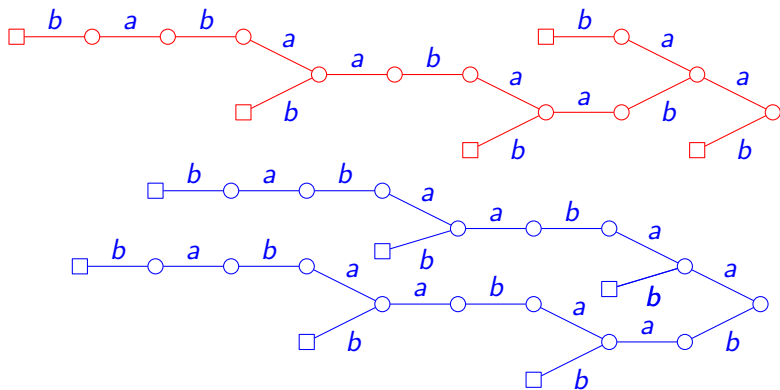
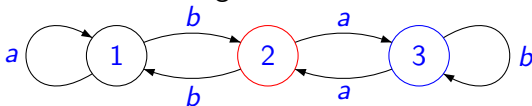


FIG.: The two suffix codes which are classes of the equivalence θ_X .

For the previous code X , the automaton induced on the classes has three states. State 2 is the class containing b , and state 3 represents the class containing ba .



Return words

Let F be a factorial set. For $u \in F$, define

$$\Gamma_F(u) = \{z \in F \mid uz \in A^+u \cap F\}$$

and

$$R_F(u) = \Gamma_F(u) \setminus \Gamma_F(u)A^+.$$

When $F = F(x)$ for an infinite word x , the sets $\Gamma_F(u)$ and $R_F(u)$ are respectively the set of **right return words** to u and **first right return words** to u in x .

Example

Let F be the Fibonacci set. The set $R_F(u)$ is given below for the first words of F .

u	1	a	b	aa	ab	ba
$R_F(u)$	a, b	a, ba	ab, aab	$baa, babaa$	ab, aab	ba, aba

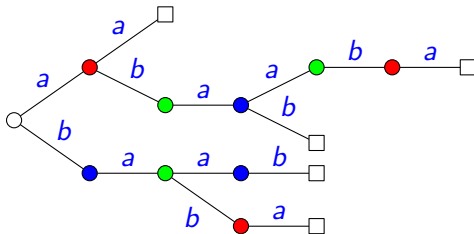
The following result is used in the proof.

Theorem (Justin and Vuillon, 2000)

Let F be a Sturmian set. For any word $u \in F$, the set $R_F(u)$ is a basis of the free group A° .

Example

Let F be the set of factors of the Fibonacci word. Let X be the F -maximal bifix code of F -degree 4 shown on the figure. The four classes are indicated in colors.



The representation of A° on the cosets of the subgroup generated by X is shown below.

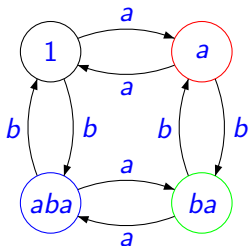


FIG.: The group $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

Holonomy groups

Let M be a monoid of transformations on a set Q . For $P \subset Q$, let

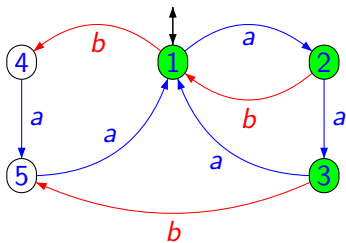
$$\text{Stab}(P) = \{m \in M \mid Pm = P\}$$

be the **stabilizer** of P . The restriction of $\text{Stab}(P)$ to P is a permutation group called the **holonomy group** of M relative to P (Eilenberg).

The holonomy groups of an automaton \mathcal{A} are those of its monoid of transitions $M(\mathcal{A})$. A **syntactic group** of a prefix code X is a holonomy group in the monoid of transitions of $\mathcal{A}(X^*)$.

Example

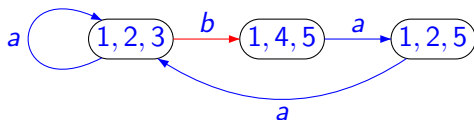
Let \mathcal{A} be the automaton represented below.



The holonomy group relative to $\{1, 2, 3\}$ is the symmetric group S_3 generated by $a = (123)$ and $baa = (23)$.

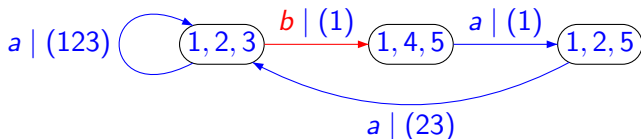
Computation of the holonomy group

Action on the sets with 3 elements.



The Schützenberger representation

The previous computation can be represented by a transducer with output in the holonomy group.



The Rank Conjecture

The results of Schützenberger around the Critical Factorization Theorem lead to the following conjecture.

Conjecture (P. 1981)

Let X be a finite bifix code and let G be a transitive permutation group of degree d which can be generated by k elements. If G is a syntactic group of X , then $\text{Card}(X) \geq (k - 1)d + 1$.

The inequality is related to Schreier's Formula.

The degree of nonspecial groups

Let X be a prefix code and let $\mathcal{A} = \mathcal{A}(X^*)$. A syntactic group G of X is called **special** if $\varphi_{\mathcal{A}}^{-1}(G)$ is a cyclic submonoid. In particular a special syntactic group is cyclic.

The **degree** of a permutation group G on a set R is the cardinality of R . The group G is **transitive** if for any $r, s \in R$ there is some $g \in G$ such that $rg = s$.

Theorem (P., Rindone 2003)

Let G be a permutation group of degree d . If G is a nonspecial syntactic group of a prefix code X , then $\text{Card}(X) \geq d + 1$.

This shows that the conjecture is true for groups of minimal rank 2.

This theorem was proved before by Schützenberger (1979)

- with a weaker bound ($\text{Card}(X) \geq d$)
- but with a more general hypothesis (with an arbitrary set X of words instead of a prefix code).

The general idea is that some parameters in the transition monoid of the minimal automaton of X^* can be bounded in terms of $\text{Card}(X)$ only, instead of the sum of the lengths of the words of X . The proof of Schützenberger uses the Critical Factorization Theorem (see Lothaire, 1983)

The theorem is clearly not true for special syntactic groups since $\mathbb{Z}/n\mathbb{Z}$ is a syntactic group of $X = a^n$ for any $n \geq 1$.

The Syntactic Group Theorem

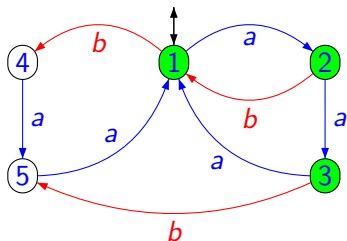
The following result shows that the bound can be reached for any transitive permutation group.

Theorem (BDPRR, 2010)

Any transitive permutation group of degree d which can be generated by k elements is a syntactic group of a bifix code with $(k - 1)d + 1$ elements.

Example 1 : The symmetric group S_3

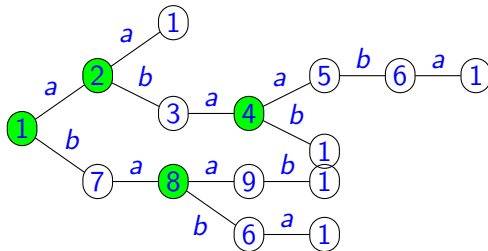
Let $X = \{aaa, aaba, ab, baa\}$. The minimal automaton of X^* is represented below.



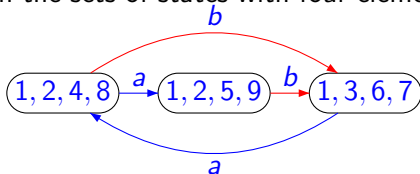
The holonomy group relative to $\{1, 2, 3\}$ is the symmetric group S_3 generated by $a = (123)$ and $baa = (23)$. Such a construction can be used to realize any group generated by a d -cycle α and another permutation β using $X = a^d \cup \{a^i ba^{d-(i+1)}\beta \mid 0 \leq i \leq d-1\}$.

Example 2 : The abelian group $(\mathbb{Z}/2\mathbb{Z})^2$

Let X be the bifix code with 5 elements represented below.



The action on the sets of states with four elements is shown below.



The word ba defines the permutation $(18)(24)$ and the word aba the permutation $(14)(28)$. Thus $(\mathbb{Z}/2\mathbb{Z})^2$ is a syntactic group of this code.

The Syntactic Group Theorem follows from the following result.

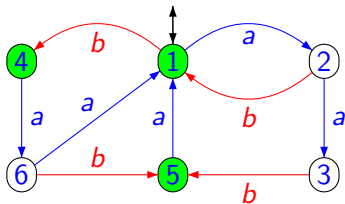
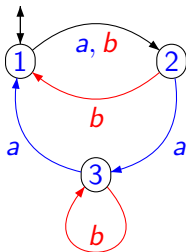
Theorem

Let $Z \subset A^$ be a group code of degree d . Let F be a Sturmian set. The set $X = Z \cap F$ is an F -maximal bifix code of degree d and $G(Z)$ is a syntactic group of X .*

Indeed, by the Cardinality Theorem, the code X has $(k - 1)d + 1$ elements.

Example 1 : The symmetric group S_3

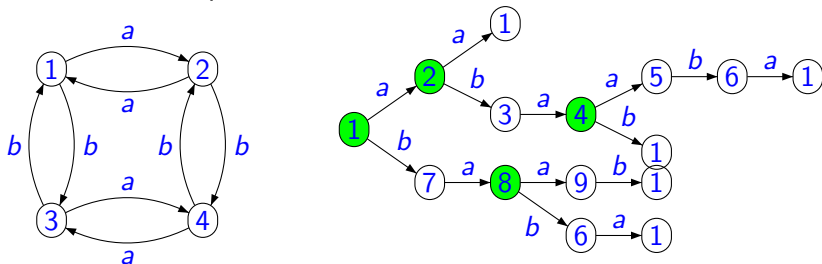
Let Z be the group code corresponding to the automaton on the left. Let F be the Fibonacci set and let $X = Z \cap F$. The minimal automaton of X^* is represented on the right.



The holonomy group relative to $\{1, 4, 5\}$ is the symmetric group S_3 generated by $ab = (45)$ and $aab = (15)$.

Example 2 : the Abelian group $(\mathbb{Z}/2\mathbb{Z})^2$

Let Z be group code defined by $Z^* = \varphi^{-1}(0,0)$ where $\varphi : \{a,b\}^* \rightarrow (\mathbb{Z}/2\mathbb{Z})^2$ is defined by $\varphi(a) = (1,0)$ and $\varphi(b) = (0,1)$. Let F be the Fibonacci set. The bifix code $X = Z \cap F$ is represented below.



We have already seen that the holonomy group relative to $\{1, 2, 4, 8\}$ is $(\mathbb{Z}/2\mathbb{Z})^2$.