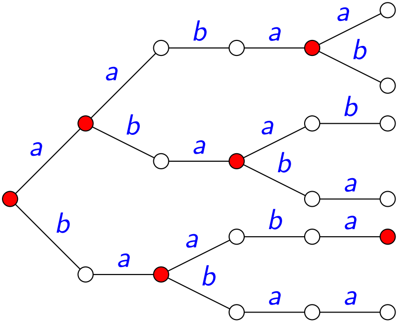
Combinatorics on Sturmian words

Dominique Perrin

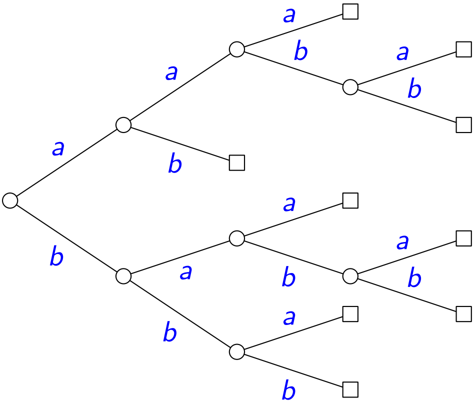
22 novembre 2011



The factors of length ≤ 5 of the Fibonacci word $x = abaababa\ldots$ fixpoint of $a \mapsto ab, \ b \mapsto a$.

A miracle

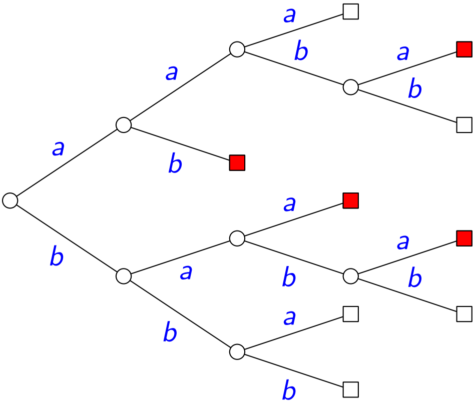
Consider the maximal bifix code of degree 3 below.



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

A miracle

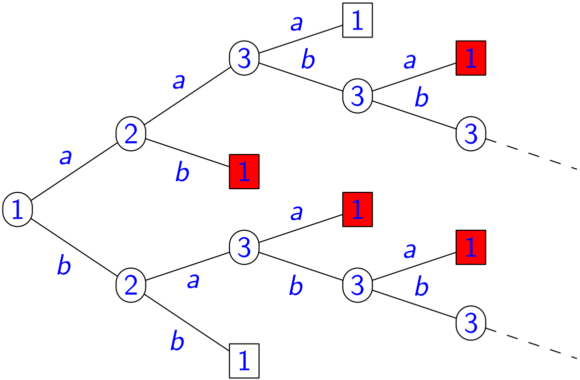
Consider the maximal bifix code of degree 3 below.



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 $\mathrm{Card}(F(x) \cap X) \leq d$ 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. Sturmian sets
2. Bifix codes
3. Sturmian sets and bifix codes
    - Cardinality Theorem
    - Periodicity Theorem
    - Sturmian Basis Theorem

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 $\mathrm{Card}(A)$.

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 inifinite word

$$x = abaababaabaababaababaabaababaabaab\cdots$$

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

Example

Set $A = \{a, b, c\}$. The morphism $f : A^* \to A^*$ 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

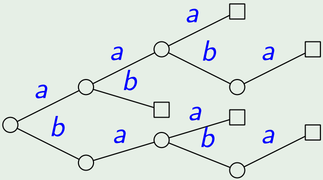
The set $X = \{a, bab\}$ is a bifix code.

Maximal bifix codes

A 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) $Y \subset F$.

Example

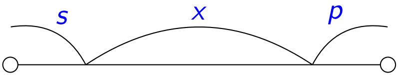
Let $A = \{a, b\}$ and let F be the set of words without factor bb . The set $X = \{aaa, aaba, ab, baa, baba\}$ is a finite F-maximal bifix code.



Parses

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

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



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 $a \in A$
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.

Theorem (Schützenberger, 1965)

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

Example

Let F be the Fibonacci set. The set $X = \{a, bab, baab\}$ is an F-maximal bifix code of degree 2. The parses of bab are $(1, bab, 1)$ and (b, a, b).

Example

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

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

Let $x = a_0 a_1 \cdots$, 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

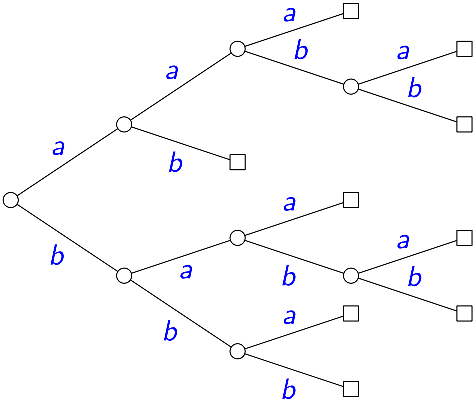
The following statement implies the Coven-Hedlund Theorem since A^d is a maximal bifix code of degree d.

Theorem (BDPRR, 2010)

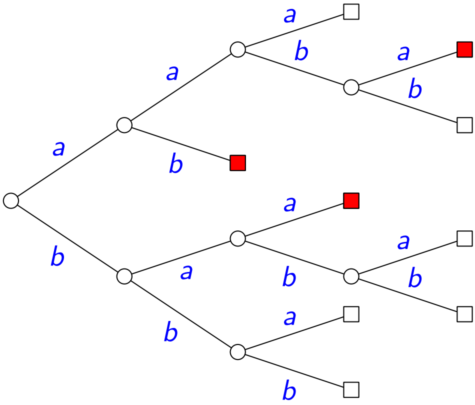
Let $x \in A^{\mathbb{N}}$ be an infinite word. If there exists a finite maximal bifix code X of degree d such that $\mathrm{Card}(X \cap F(x)) \leq d$, then x is ultimately periodic.

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 $\mathrm{Card}(X) = (k-1)d + 1$ 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 $i = d, \ldots, 0$ 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 satifies $N_{1,k} = 1$ 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

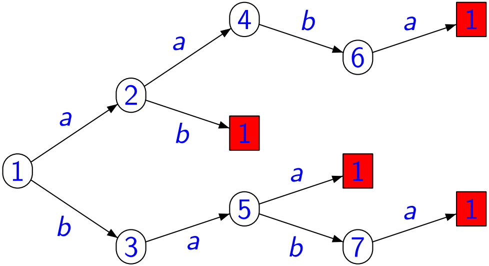
$$N_{d,2} = d\ d! - \sum_{i=1}^{d-1} (d-i)! N_{i,2}.$$

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

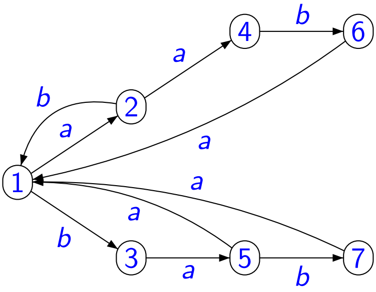
The formula is known to enumerate also the indecomposable permutations on $d+1$ elements (see Dress, Franz 1985, Ossona, Rosenstiehl 2004 and Cori 2009).

Stallings foldings



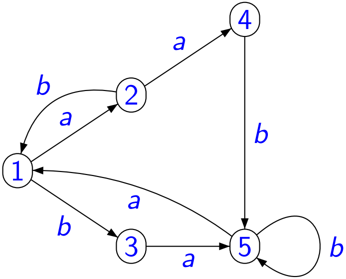
An F-maximal bifix code of F-degree 3.

Stallings foldings



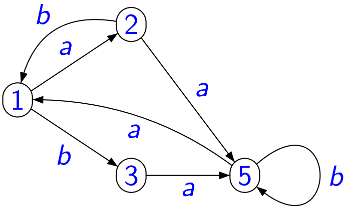
Fusion of $5, 6, 7$.

Stallings foldings



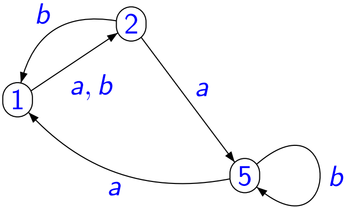
Fusion of $4, 5$.

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 $X^* \cap F$.

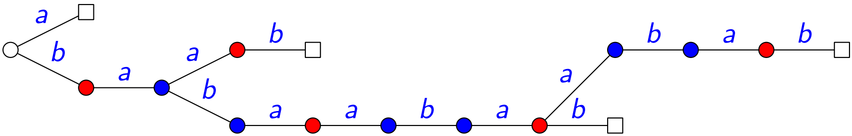
Proposition

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.



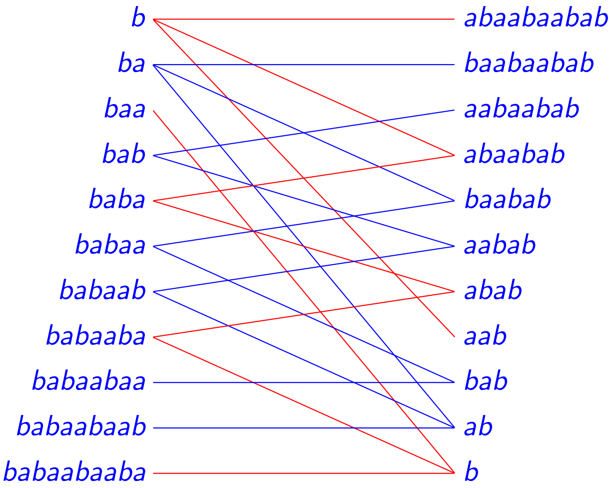


FIG.: The incidence graph of X.

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.

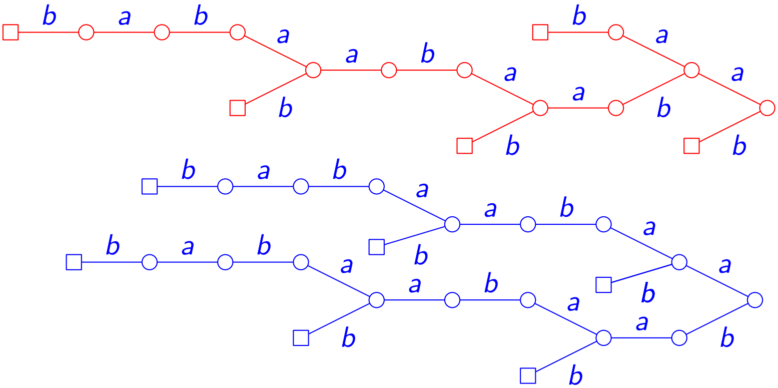
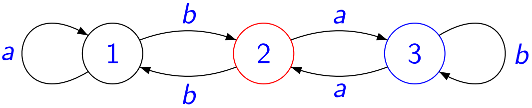


FIG.: The two suffix codes which are classes of the equivalence θ_X.

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

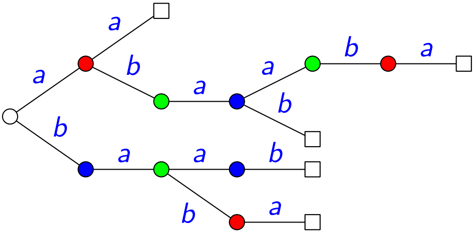
The following result is used in the proof.

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.

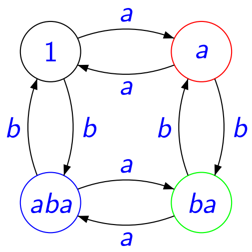


FIG.: The group $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.