Maximal bifix decoding of a tree set

Francesco Dolce

Journées Annuelles Systèmes Dynamiques, Automates et Algorithmes

Marne-la-Vallée, 9^{th} April 2015

Joint work with

V. Berthé, C. De Felice, J. Leroy, D. Perrin, C. Reutenauer and G. Rindone

Motivation

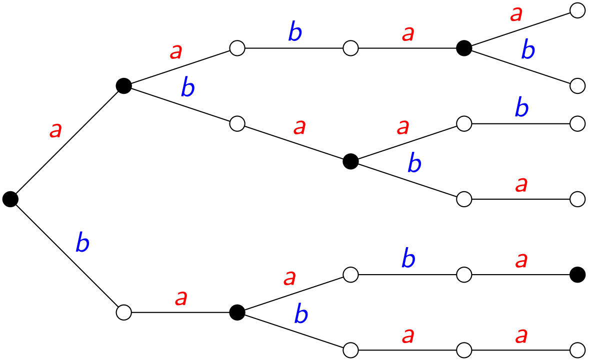
$$x = abaababaabaababa\cdots$$

$$x = \varphi^\omega(a)$$

$$\varphi : \begin{cases} a \mapsto ab \\ b \mapsto a \end{cases}$$

Motivation

$$x = abaababaabaababa\cdots$$



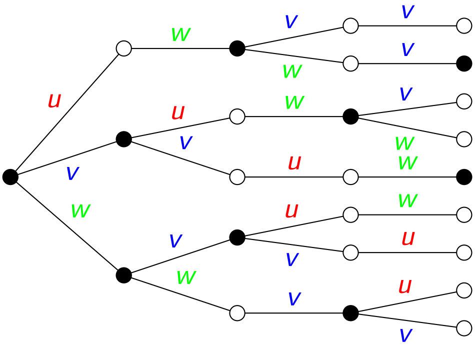
n	0	1	2	3	4	5	$\cdots$
(2-1)n+1	1	2	3	4	5	6	$\cdots$

Motivation

$$x = \underline{ab}\,\underline{aa}\,\underline{ba}\,\underline{ba}\,\underline{ab}\,\underline{aa}\,\underline{ba}\,\underline{ba}\cdots$$

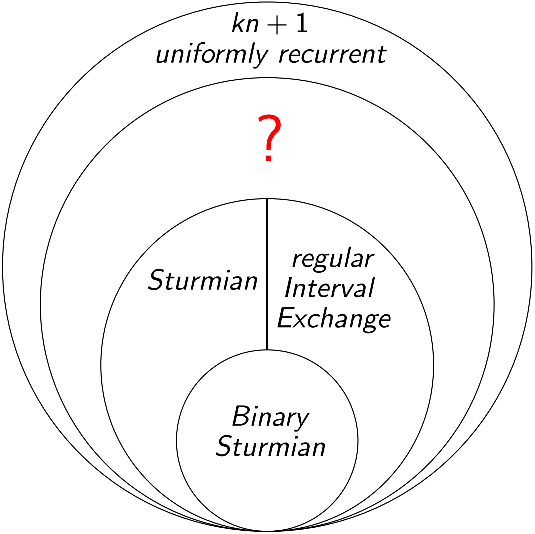
$$f(x) = v\;u\;w\;w\;v\;u\;w\;w\cdots$$

$$f : \begin{cases} u = aa \\ v = ab \\ w = ba \end{cases}$$



n	0	1	2	3	4	$\cdots$
(3-1)n+1	1	3	5	7	9	$\cdots$

Motivation


$kn+1$
uniformly recurrent
?
Sturmian
regular
Interval
Exchange
Binary
Sturmian

Outline

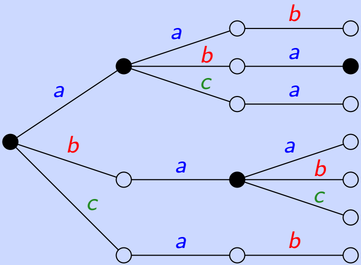
Outline

A *Sturmian* set is the set of factors of a *strict episturmian word* (i.e. of a word x whose set of factors $F(x)$ is closed under reversal and for each n contains exactly one right-special word w_n of length n with $w_n A \subset F(x)$).

Example

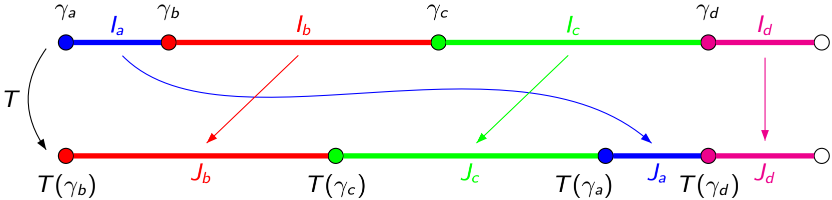
Let $A = \{a, b, c\}$. The *Tribonacci set* is the set of factors of the Tribonacci word, i.e. the fixpoint $x = f^\omega(a) = abacaba\ldots$ of the morphism

$$f : a \mapsto ab, \quad b \mapsto ac, \quad c \mapsto a.$$



Let A be a finite set ordered by $<_1$ and $<_2$.
An *interval exchange transformation* (IET) is a map $T : [0,1[\to [0,1[$ defined by

$$T(z) = z + \alpha_z \quad \text{if } z \in I_a.$$



$a <_1 b <_1 c <_1 d$ $\qquad$ $b <_2 c <_2 a <_2 d$

An interval exchange transformation T is said to be *minimal* if for any $z \in [0,1[$ the orbit $\mathcal{O}(z) = \{T^n(z) \mid n \in \mathbb{Z}\}$ is dense in $[0,1[$.

The transformation T is said *regular* if the orbits of the nonzero separation points are infinite and disjoint.

Theorem [Keane (1975)]

A regular interval exchange transformation is minimal.

An interval exchange transformation T is said to be *minimal* if for any $z \in [0,1[$ the orbit $\mathcal{O}(z) = \{T^n(z) \mid n \in \mathbb{Z}\}$ is dense in $[0,1[$.

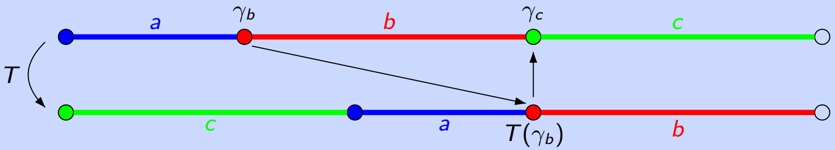
The transformation T is said *regular* if the orbits of the nonzero separation points are infinite and disjoint.

Theorem [Keane (1975)]

A regular interval exchange transformation is minimal.

The converse is not true.

Example



Let T be an IET relative to $(I_a)_{a \in A}$.
The *natural coding* of T relative to $z \in [0, 1[$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

$$a_n = a \quad \text{si } T^n(z) \in I_a.$$

Example

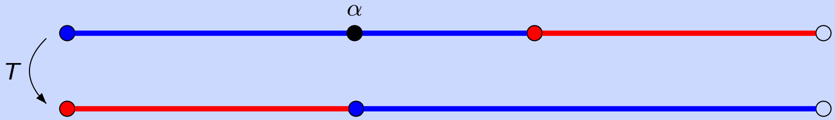
The *Fibonacci word* is the natural coding of the rotation of angle $\alpha = (3 - \sqrt{5})/2$ relative to the point α, i.e. $T(z) = z + \alpha \bmod 1$.

Let T be an IET relative to $(I_a)_{a \in A}$.
The *natural coding* of T relative to $z \in [0,1[$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

$$a_n = a \quad \text{si} \quad T^n(z) \in I_a.$$

Example

The *Fibonacci word* is the natural coding of the rotation of angle $\alpha = (3 - \sqrt{5})/2$ relative to the point α, i.e. $T(z) = z + \alpha \bmod 1$.



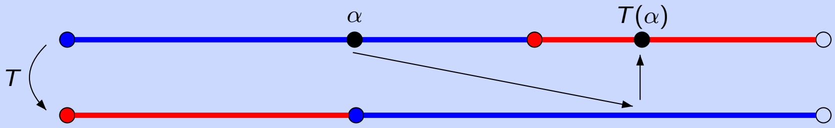
$$\Sigma_T(\alpha) = a$$

Let T be an IET relative to $(I_a)_{a \in A}$.
The *natural coding* of T relative to $z \in [0, 1[$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

$$a_n = a \quad \text{si} \quad T^n(z) \in I_a.$$

Example

The *Fibonacci word* is the natural coding of the rotation of angle $\alpha = (3 - \sqrt{5})/2$ relative to the point α, i.e. $T(z) = z + \alpha \bmod 1$.



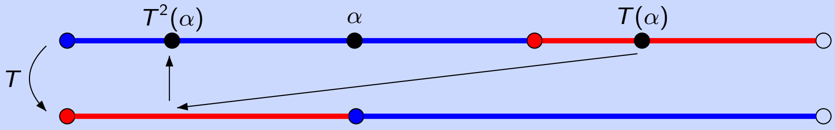
$\Sigma_T(\alpha) = a\ b$

Let T be an IET relative to $(I_a)_{a \in A}$.
The *natural coding* of T relative to $z \in [0,1[$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

$$a_n = a \quad \text{si } T^n(z) \in I_a.$$

Example

The *Fibonacci word* is the natural coding of the rotation of angle $\alpha = (3-\sqrt{5})/2$ relative to the point α, i.e. $T(z) = z + \alpha \bmod 1$.



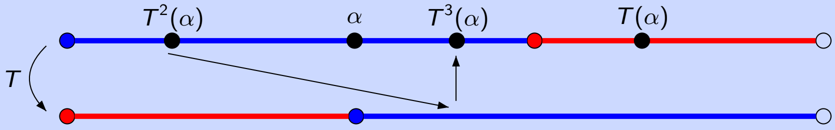
$\Sigma_T(\alpha) = a\ b\ a$

Let T be an IET relative to $(I_a)_{a\in A}$.
The *natural coding* of T relative to $z \in [0,1[$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

$$a_n = a \quad \text{si } T^n(z) \in I_a.$$

Example

The *Fibonacci word* is the natural coding of the rotation of angle $\alpha = (3-\sqrt{5})/2$ relative to the point α, i.e. $T(z) = z + \alpha \bmod 1$.



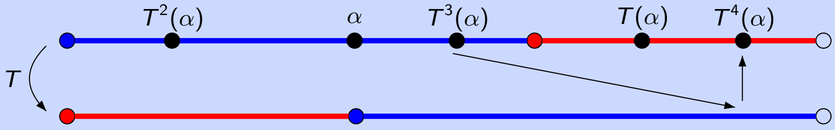
$$\Sigma_T(\alpha) = a\ b\ a\ a$$

Let T be an IET relative to $(I_a)_{a \in A}$.
The *natural coding* of T relative to $z \in [0,1[$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

$$a_n = a \quad \text{si} \quad T^n(z) \in I_a.$$

Example

The *Fibonacci word* is the natural coding of the rotation of angle $\alpha = (3 - \sqrt{5})/2$ relative to the point α, i.e. $T(z) = z + \alpha \bmod 1$.



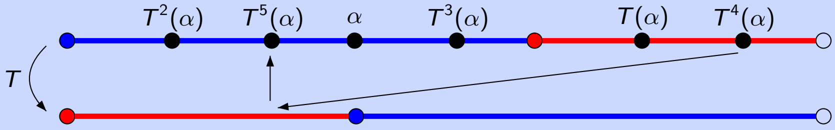
$$\Sigma_T(\alpha) = a\,b\,a\,a\,b$$

Let T be an IET relative to $(I_a)_{a\in A}$.
The *natural coding* of T relative to $z \in [0,1[$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

$$a_n = a \quad \text{si} \quad T^n(z) \in I_a.$$

Example

The *Fibonacci word* is the natural coding of the rotation of angle $\alpha = (3-\sqrt{5})/2$ relative to the point α, i.e. $T(z) = z + \alpha \bmod 1$.



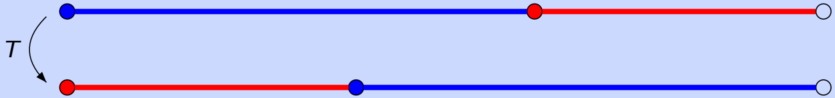
$$\Sigma_T(\alpha) = a\,b\,a\,a\,b\,a\cdots$$

Given an interval exchange transformation T (resp. minimal interval exchange transformation, resp. regular interval exchange transformation), the set $F(T) = \cup_z(\Sigma_T(z))$ is said an *interval exchange set* (resp. *minimal* regular interval exchange set, resp. *regular* interval exchange set).

Remark. If T is minimal, $F(\Sigma_T(z))$ does not depend on the point z.

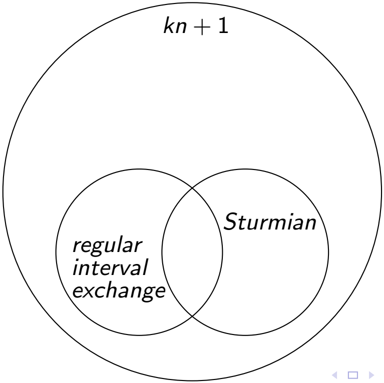
Example

The *Fibonacci set* is the set of factors of a natural coding of the rotation of angle $\alpha = (3-\sqrt{5})/2$.



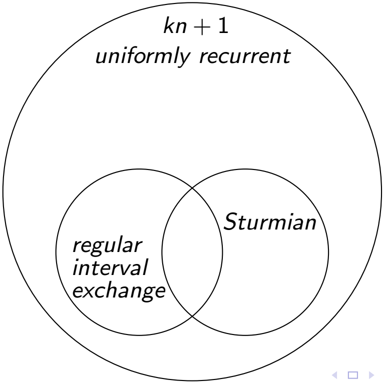
$$F(T) = \{\varepsilon, a, b, aa, ab, ba, aab, aba, baa, \dots\}$$

Sturmian sets and regular interval exchange sets have both complexity function $p(n) = kn + 1$, with $k = \text{Card}\,(A) - 1$.



Sturmian sets and regular interval exchange sets have both complexity function $p(n) = kn + 1$, with $k = \text{Card}(A) - 1$.

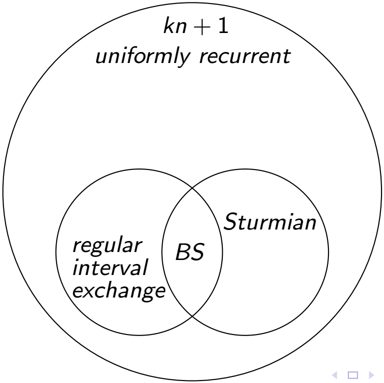
They are factorial and *uniformly recurrent* (right-extendable and s.t. for any element $u \in S$ there exists an $n = n(u)$ with u a factor of all words of $S \cap A^n$).



Sturmian sets and regular interval exchange sets have both complexity function $p(n) = kn + 1$, with $k = \text{Card}\,(A) - 1$.

They are factorial and *uniformly recurrent* (right-extendable and s.t. for any element $u \in S$ there exists an $n = n(u)$ with u a factor of all words of $S \cap A^n$).

However, the two families are distinct for $k \geq 2$.

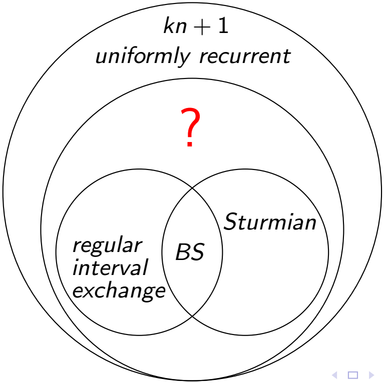


Sturmian sets and regular interval exchange sets have both complexity function $p(n) = kn + 1$, with $k = \text{Card}(A) - 1$.

They are factorial and *uniformly recurrent* (right-extendable and s.t. for any element $u \in S$ there exists an $n = n(u)$ with u a factor of all words of $S \cap A^n$).

However, the two families are distinct for $k \geq 2$.

Do they have other properties in common ?



Outline

Motivation

1. Two important classes
2. Acyclic, connected and tree sets
    - Tree sets
    - Planar tree sets
    - Neutral sets
    - Return words in tree sets
3. Bifix decoding

Conclusions

Let S be a factorial over an alphabet A.
The *extension graph* of a word $w \in S$ is the undirected bipartite graph $G(w)$ with vertices the disjoint union of

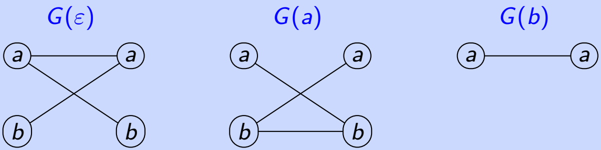
$$L(w) = \{a \in A \mid aw \in S\} \quad \text{and} \quad R(w) = \{a \in A \mid wa \in S\},$$

and edges the pairs in

$$E(w) = \{(a, b) \in A \times A \mid awb \in S\}.$$

Example

Let S be the Fibonacci set.



Indeed one has $S = \{\varepsilon, a, b, aa, ab, ba, aab, aba, baa, bab, \ldots\}$.

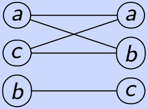
A set S is *acyclic* (resp. *connected*) if it is biextendable and if for every word $w \in S$, the graph $G(w)$ is acyclic (resp. connected).

A set S is a *tree set* (*of characteristic* 1) if $G(w)$ is acyclic and connected for every word $w \in S$.

Example

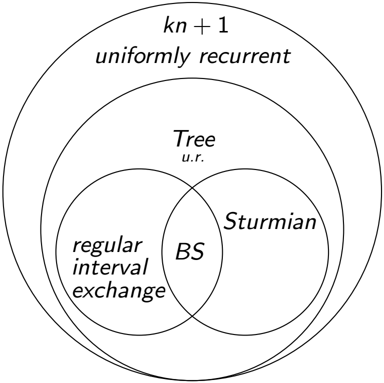
Let $A = \{a, b, c\}$. The set S of factors of $a^*\{bc, bcbc\}a^*$ is not a tree set. Actually it is neither acyclic nor connected.

$G(\varepsilon)$



Proposition [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2014, 2015)]

Both Sturmian sets and regular interval exchange sets are uniformly recurrent tree sets (*of characteristic* 1).

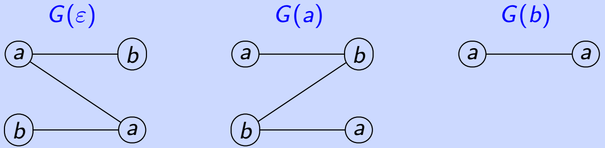


Let $<_1$ and $<_2$ be two orders on A.
For a set S and a word $w \in S$, the graph $G(w)$ is *compatible* with $<_1$ and $<_2$ if for any $(a, b), (c, d) \in E(w)$, one has

$$a <_1 c \implies b \leq_2 d.$$

Example

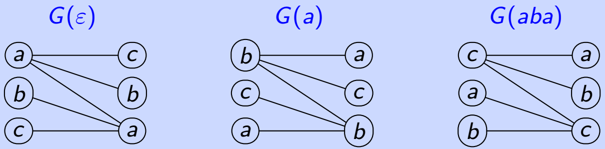
Let S be the Fibonacci set. Set $a <_1 b$ and $b <_2 a$.



We say that a biextendable set S is a *planar tree set* (*of characteristic* 1) w.r.t. $<_1$ and $<_2$ on A if for any $w \in S$, the graph $G(w)$ is a tree compatible with $<_1$ and $<_2$.

Example

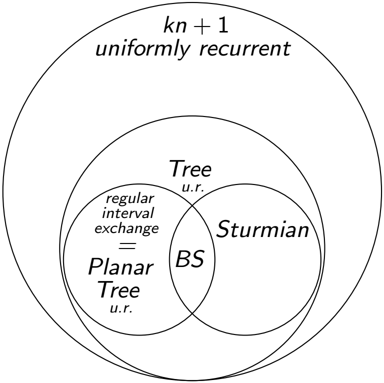
The *Tribonacci set* is not a planar tree set.
Indeed, let us consider the extension graphs of the bispecial words ε, a and aba.



It is not possible to find two orders on A making the three graphs planar.

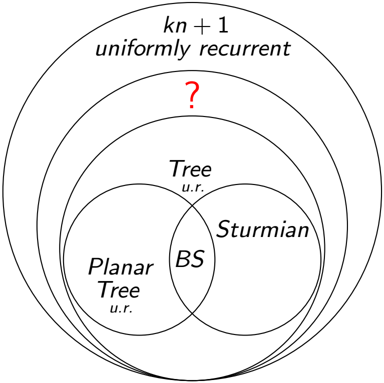
Theorem [Ferenczi, Zamboni (2008)]

A set S is a regular interval exchange set on A if and only if it is a uniformly recurrent planar tree set (*of characteristic* 1) containing A.



Theorem [Ferenczi, Zamboni (2008)]

A set S is a regular interval exchange set on A if and only if it is a uniformly recurrent planar tree set (*of characteristic* 1) containing A.



Given a factorial set S and a word $w \in S$, we define

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

We say that S is *neutral* (*of characteristic* 1) if for every $w \in S$ one has

$$m(w) = e(w) - \ell(w) - r(w) + 1 = 0.$$

Given a factorial set S and a word $w \in S$, we define

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

We say that S is *neutral* (*of characteristic* 1) if for every $w \in S$ one has

$$m(w) = e(w) - \ell(w) - r(w) + 1 = 0.$$

Example

The *Chacon set* is the set of factors of the Chacon word, i.e. the fixed point of the morphism

$$f : a \mapsto aabc, \quad b \mapsto bc, \quad c \mapsto abc.$$

It is not neutral. Indeed, one has $m(\varepsilon) = 0$, but $m(abc) = 1$ and $m(bca) = -1$.

Given a factorial set S and a word $w \in S$, we define

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

We say that S is *neutral* (*of characteristic* 1) if for every $w \in S$ one has

$$m(w) = e(w) - \ell(w) - r(w) + 1 = 0.$$

Example

The *Chacon set* is the set of factors of the Chacon word, i.e. the fixed point of the morphism

$$f : a \mapsto aabc, \quad b \mapsto bc, \quad c \mapsto abc.$$

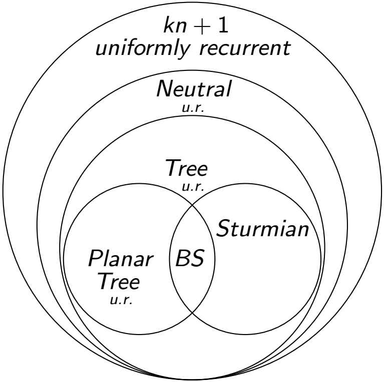
It is not neutral. Indeed, one has $m(\varepsilon) = 0$, but $m(abc) = 1$ and $m(bca) = -1$.

Proposition

A neutral set (*of characteristic* 1) S over a finite alphabet A has complexity function $p(n) = (\text{Card}\,(A) - 1)n + 1$.

Proposition

Tree sets (*of characteristic* 1) are neutral (*of characteristic* 1).



Let S be a set of words. For $w \in S$, let

$$\Gamma_S(w) = \{x \in S \mid wx \in S \cap A^+ w\} \quad \text{and} \quad \mathcal{R}_S(w) = \Gamma_S(w) \setminus \Gamma_S(w)A^+$$

be the set of (*right*) *return words* and *first* (*right*) *return words* to w.

Example

Let S be the Fibonacci set. One has $\mathcal{R}_S(aa) = \{baa, babaa\}$.

$$x = abaababaabaababaababaabaababaabaab\cdots$$

Let S be a set of words. For $w \in S$, let

$$\Gamma_S(w) = \{x \in S \mid wx \in S \cap A^+ w\} \quad \text{and} \quad \mathcal{R}_S(w) = \Gamma_S(w) \setminus \Gamma_S(w)A^+$$

be the set of (*right*) *return words* and *first* (*right*) *return words* to w.

Example

Let S be the Fibonacci set. One has $\mathcal{R}_S(aa) = \{baa, babaa\}$.

$$x = abaababaabaababaababaabaababaabaab\cdots$$

Remark. A recurrent set S is uniformly recurrent if and only if the set $\mathcal{R}_S(w)$ is finite for every $w \in S$.

Theorem [Balková, Palentová, Steiner (2008)]

Let S be a uniformly recurrent neutral set (*of characteristic* 1) containing the alphabet A. Then for every $w \in S$, the set $\mathcal{R}_s(w)$ has exactly $\mathrm{Card}\,(A)$ elements.

Theorem [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2014)]

Let S be a uniformly recurrent tree set (*of characteristic* 1) containing the alphabet A. Then, for any $w \in S$, the set $\mathcal{R}_S(w)$ is a basis of the free group on A.

Theorem [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2014)]

Let S be a uniformly recurrent tree set (*of characteristic* 1) containing the alphabet A. Then, for any $w \in S$, the set $\mathcal{R}_S(w)$ is a basis of the free group on A.

Example

Let S be the Fibonacci set. The set $\mathcal{R}_S(aa) = \{baa, babaa\}$ is a basis of the free group. Indeed,

$$\begin{aligned} a &= baa\,(babaa)^{-1}baa \\ b &= baa\,a^{-1}\,a^{-1} \end{aligned}$$

So, $\langle \mathcal{R}_S(aa) \rangle = \langle a, b \rangle = F_A$.

Outline

Motivation

1. Two important classes
2. Acyclic, connected and tree sets
3. Bifix decoding
   - Bifix codes
   - Maximal bifix decoding

Conclusions

A set $X \subset A^+$ of nonempty words over an alphabet A is a *bifix code* if it does not contain any proper prefix or suffix of its elements.

Example

- $\{aa, ab, ba\}$
- $\{aa, ab, bba, bbb\}$
- $\{ac, bcc, bcbca\}$

A set $X \subset A^+$ of nonempty words over an alphabet A is a *bifix code* if it does not contain any proper prefix or suffix of its elements.

Example

- $\{aa, ab, ba\}$
- $\{aa, ab, bba, bbb\}$
- $\{ac, bcc, bcbca\}$

A bifix code $X \subset S$ is *S-maximal* if it is not properly contained in a bifix code $Y \subset S$.

Example

Let S be the Fibonacci set. The set $X = \{aa, ab, ba\}$ is an S-maximal bifix code.

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

Example

Let's consider the bifix code $X = \{aa, ab, ba\}$ on $A = \{a, b\}$ and let $B = \{u, v, w\}$. The map

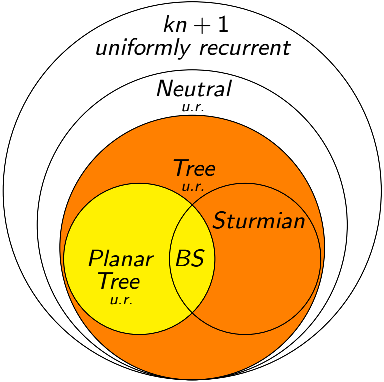
$$f : \begin{cases} u \mapsto aa \\ v \mapsto ab \\ w \mapsto ba \end{cases}$$

is a coding morphism for X.

If S is factorial and X is an S-maximal bifix code, we call the set $f^{-1}(S)$ a *maximal bifix decoding* of S.

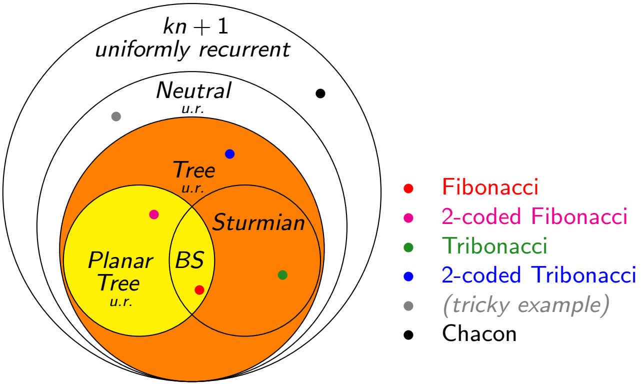
Theorem [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2014, 2015)]

The family of uniformly recurrent tree sets (*of characteristic* 1) is closed under maximal bifix decoding (and so is the family of u.r. planar tree sets (*of characteristic* 1)).



Theorem [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2014, 2015)]

The family of uniformly recurrent tree sets (*of characteristic* 1) is closed under maximal bifix decoding (and so is the family of u.r. planar tree sets (*of characteristic* 1)).



We studied several classes of uniform recurrent sets of linear complexity (namely $p(n) = (\text{Card}\,(A) - 1)n + 1$), all of them satisfying certain properties for every word :

→ neutral condition,

$m(w) = 0$ for every word $w \in S$

→ tree condition,

$G(w)$ is a tree for every word $w \in S$

→ planar tree condition.

every $G(w)$ is a tree compatible with two orders $<_1$ and $<_2$

We studied several classes of uniform recurrent sets of linear complexity (namely $p(n) = (\text{Card}\,(A) - 1)n + 1$), all of them satisfying certain properties for every word :

→ neutral condition,

$m(w) = 0$ for every word $w \in S$

→ tree condition,

$G(w)$ is a tree for every word $w \in S$

→ planar tree condition.

every $G(w)$ is a tree compatible with two orders $<_1$ and $<_2$

The last two satisfy a closure property (maximal bifix decoding).

Neutrality (*of characteristic* 1) is preserved under maximal bifix decoding, but we do not known if the uniform recurrence is.

We can generalize the notions of tree sets and neutral sets considering the empty word as an “exception” (characteristic $1 - m(\varepsilon) \geq 1$).

We can generalize the notions of tree sets and neutral sets considering the empty word as an "exception" (characteristic $1 - m(\varepsilon) \geq 1$).

Theorem [D., Perrin (2015)]

The factor complexity of a neutral set of characteristic χ is given by $p(0) = 1$ and $p(n) = (\mathrm{Card}\,(A) - \chi)n + \chi$.
Moreover, any maximal bifix decoding of a recurrent neutral set is a neutral set with the same characteristic.

We can generalize the notions of tree sets and neutral sets considering the empty word as an "exception" (characteristic $1 - m(\varepsilon) \geq 1$).

Theorem [D., Perrin (2015)]

The factor complexity of a neutral set of characteristic χ is given by $p(0) = 1$ and $p(n) = (\mathrm{Card}\,(A) - \chi)n + \chi$.
Moreover, any maximal bifix decoding of a recurrent neutral set is a neutral set with the same characteristic.

$\rightarrow$ *What can we say about the recurrence?*

Example

Let S be the set of factors of the infinite word $(ab)^\omega$. Consider the S-maximal bifix code $X = \{ab, ba\}$ and the coding morphism $f : u \mapsto ab,\ v \mapsto ba$.
The set S is uniformly recurrent but $f^{-1}(S) = u^\omega \cup v^\omega$ is not even recurrent.

We can generalize the notions of tree sets and neutral sets considering the empty word as an "exception" (characteristic $1 - m(\varepsilon) \geq 1$).

Theorem [D., Perrin (2015)]

The factor complexity of a neutral set of characteristic χ is given by $p(0) = 1$ and $p(n) = (\text{Card}\,(A) - \chi)n + \chi$.
Moreover, any maximal bifix decoding of a recurrent neutral set is a neutral set with the same characteristic.

$\rightarrow$ *What can we say about the recurrence ?*

Example

Let S be the set of factors of the infinite word $(ab)^\omega$. Consider the S-maximal bifix code $X = \{ab, ba\}$ and the coding morphism $f : u \mapsto ab,\ v \mapsto ba$.
The set S is uniformly recurrent but $f^{-1}(S) = u^\omega \cup v^\omega$ is not even recurrent.

$\rightarrow$ *What about tree sets of characteristic $\chi \geq 1$ ($G(\varepsilon)$ acyclic) ?*

Proposition [D., Perrin (2015)]

A tree set of characteristic χ is a neutral set of characteristic χ.

Merci