# *Acyclic, connected and tree sets*

Francesco Dolce

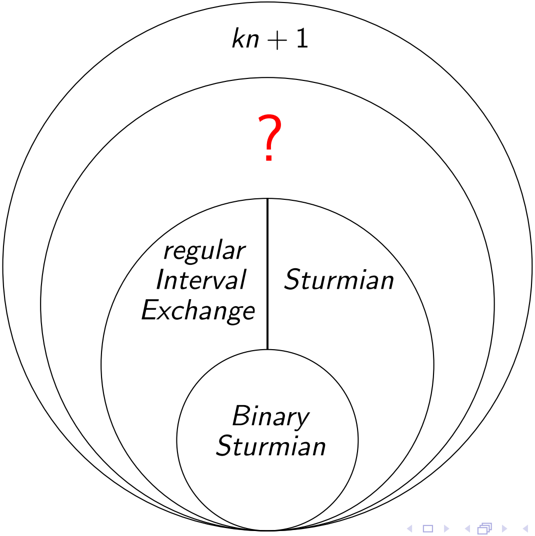
Créteil, $2^{nd}$ February 2015

Réunion EQINOCS

Joint work with :
V. Berthé, C. De Felice, J. Leroy, D. Perrin, C. Reutenauer and G. Rindone

# Motivation

## Uniformly Recurrent sets of linear complexity



# *Outline*

1. Acyclic, Connected and Tree Sets
2. Tree Sets and Bifix Codes
3. Return Words in Tree Sets

# *Outline*

1. Acyclic, Connected and Tree Sets
   - Extensions of Words
   - Recurrent Sets
   - Tree Sets
     - Sturmian Sets
     - Regular Interval Exchange Sets
2. Tree Sets and Bifix Codes
3. Return Words in Tree Sets

Let $A$ be a finite nonempty alphabet, and let $S \subset A^*$ be a *factorial* set.
For $w \in S$, we denote

$$\begin{aligned} L(w) &= \{a \in A \mid aw \in S\}, \\ R(w) &= \{a \in A \mid wa \in S\}, \\ E(w) &= \{(a,b) \in A \times A \mid awb \in S\}. \end{aligned}$$

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

Let $A$ be a finite nonempty alphabet, and let $S \subset A^*$ be a *factorial* set.
For $w \in S$, we denote

$$\begin{aligned} L(w) &= \{a \in A \mid aw \in S\}, \\ R(w) &= \{a \in A \mid wa \in S\}, \\ E(w) &= \{(a,b) \in A \times A \mid awb \in S\}. \end{aligned}$$

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

A word $w$ is *right-extendable* if $r(w) > 0$. Symmetrically for *left-extendable* and *biextendable*.
A factorial set $S$ is called right-extendable if every word in $S$ is right-extendable.

A word $w$ is called *right-special* (resp. *left-special*) if $r(w) \geq 2$ (resp. $\ell(w) \geq 2$. It is called *bispecial* if it is both right and left-special.

A set of words $S \neq \{\varepsilon\}$ is *recurrent* if it is factorial and for every $u, w \in S$ there is a $v \in S$ such thath $uvw \in S$. A recurrent set is biextendable.

A set of words $S$ is said to be *uniformly recurrent* if it is right-extendable and if, for any word $u \in S$ there exists an integer $n \geq 1$ such that $u$ is a factor of every word of $S$ of length $n$. A uniformly recurrent set is recurrent.

A set of words $S \neq \{\varepsilon\}$ is *recurrent* if it is factorial and for every $u, w \in S$ there is a $v \in S$ such thath $uvw \in S$. A recurrent set is biextendable.

A set of words $S$ is said to be *uniformly recurrent* if it is right-extendable and if, for any word $u \in S$ there exists an integer $n \geq 1$ such that $u$ is a factor of every word of $S$ of length $n$. A uniformly recurrent set is recurrent.

**Example**

Let $A = \{a, b\}$. The *Fibonacci set* $S$ is the set of factors of the *Fibonacci word*, that is the fixpoint $x = \varphi^\omega(a) = abaababaaba\cdots$ of the morphism

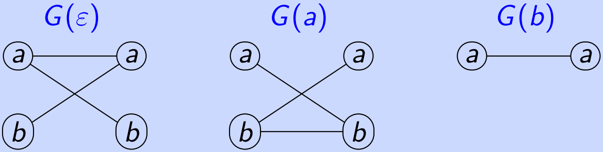
$$\varphi : a \mapsto ab, \qquad b \mapsto a.$$

$S$ is a uniformly recurrent set.

The *extension graph* of $w$ is the undirected bipartite graph $G(w)$ with vertices $L(w) \sqcup R(w)$ and edges $E(w)$.

### Example

Let $S$ be the Fibonacci set.



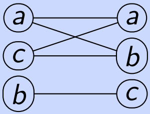

Indeed one has $S = \{\varepsilon,\ a, b,\ aa, ab, ba,\ aab, aba, baa, bab, \ldots\}$.

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

## 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**

The factor complexity of a tree set is $kn + 1$.

In particular, as for any set of polynomial complexity, the entropy is zero.

**Proposition**

The factor complexity of a tree set is $kn + 1$.

In particular, as for any set of polynomial complexity, the entropy is zero.

Two important classes of tree sets are :

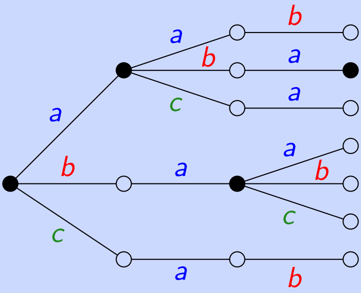
- Sturmian sets ;
- Regular interval exchange sets.

A *Sturmian* set is the set of factors of a *strict episturmian word* (i.e. of a word whose set of factors is closed under reversal and for each $n$ contains exactly one right-special word $w_n$ of length $n$ with $r(w_n) = \mathrm{Card}\,(A)$).

## Example

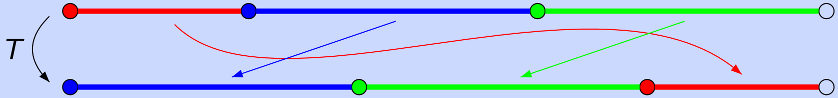
Let $A = \{a, b, c\}$. The *Tribonacci set* is the set of factors of the Tribonacci word, i.e. the fixpoint $x = f^\omega(a) = abacaba\dots$ of the morphism

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



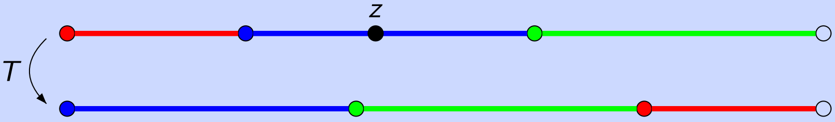
A *regular interval exchange set* is the set of factors of a natural coding of a *regular interval exchange transformation*.

## Example



A *regular interval exchange set* is the set of factors of a natural coding of a *regular interval exchange transformation*.

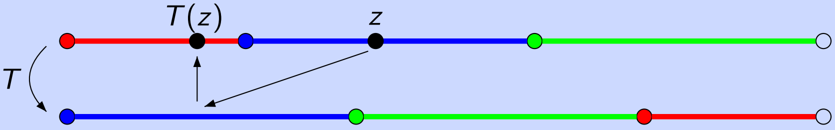
## Example




$$\Sigma_T(z) = b$$

A *regular interval exchange set* is the set of factors of a natural coding of a *regular interval exchange transformation*.

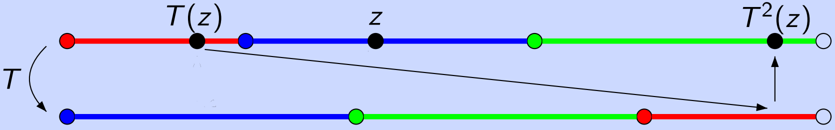
## Example




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

A *regular interval exchange set* is the set of factors of a natural coding of a *regular interval exchange transformation*.

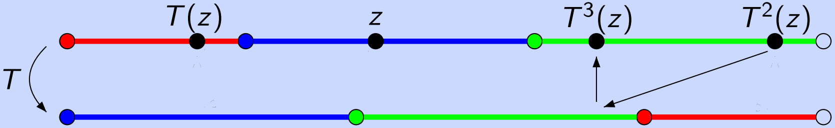
## Example




$$\Sigma_T(z) = b\,a\,c$$

A *regular interval exchange set* is the set of factors of a natural coding of a *regular interval exchange transformation*.

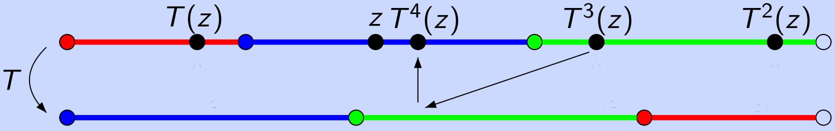
## Example




$$\Sigma_T(z) = b\,a\,c\,c$$

A *regular interval exchange set* is the set of factors of a natural coding of a *regular interval exchange transformation*.

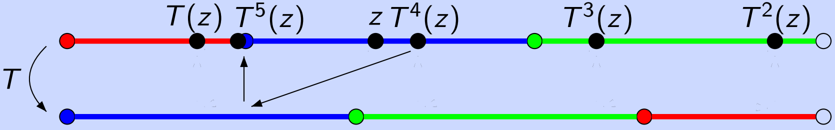
**Example**




$$\Sigma_T(z) = b\ a\ c\ c\ b$$

A *regular interval exchange set* is the set of factors of a natural coding of a *regular interval exchange transformation*.

**Example**




$$\Sigma_T(z) = b\,a\,c\,c\,b\,a\cdots$$

# *Outline*

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 $Y \subset S$.

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 $X$.

Let $S$ be a factorial set and $f$ be a coding morphism for a finite ($S$-maximal) bifix code $X \subset S$.
The set $f^{-1}(S)$ is called a (*maximal*) *bifix decoding* of $S$.

**Theorem (2014, *Monatsh. Math.*)**

Any biextendable set which is the bifix decoding of an acyclic set is acyclic.

Let $S$ be a factorial set and $f$ be a coding morphism for a finite ($S$-maximal) bifix code $X \subset S$.
The set $f^{-1}(S)$ is called a (*maximal*) *bifix decoding* of $S$.

**Theorem (2014, *Monatsh. Math.*)**

Any biextendable set which is the bifix decoding of an acyclic set is acyclic.

**Example**

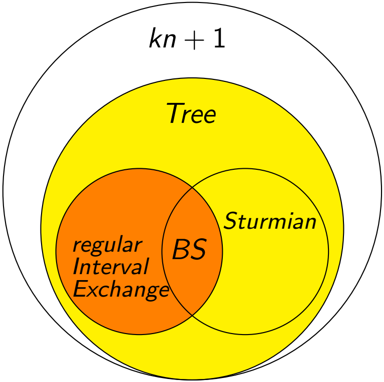
Let $S$ be the Fibonacci set, $X = S \cap A^2 = \{aa, ab, ba\}$, $B = \{u, v, w\}$ and

$$f : u \mapsto aa, \quad v \mapsto ab, \quad w \mapsto ba.$$

The set $f^{-1}(S)$ is an acyclic set. Actually, it is a tree set.

**Theorem (2014, *Monatsh. Math.* - 2015, *Discrete Math.*)**

The family of uniformly recurrent tree sets is closed under maximal bifix decoding (and so it is the family of interval exchange sets).



Denote by $F_A$ the free group on the alphabet $A$. A subset $X$ of the free group is called *free* if it is a basis of the subgroup $\langle X \rangle$.

**Theorem (2014, *Monatsh. Math.*)**

A set $S$ is acyclic if and only if any bifix code $X \subset S$ is a free subset of the free group $F_A$.

Denote by $F_A$ the free group on the alphabet $A$. A subset $X$ of the free group is called *free* if it is a basis of the subgroup $\langle X \rangle$.

**Theorem (2014, *Monatsh. Math.*)**

A set $S$ is acyclic if and only if any bifix code $X \subset S$ is a free subset of the free group $F_A$.

Let $M$ be a submonoid of $A^*$ and $\langle M \rangle$ the subgroup of $F_A$ generated by $M$. The submonoid $M$ is *saturated* in a set of word $S$ if $M \cap S = \langle M \rangle \cap S$.

**Theorem (2014, *Monatsh. Math.*)**

Let $S$ be an acyclic set. The submonoid generated by a bifix code included in $S$ is saturated in $S$.

# *Outline*

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

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

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

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

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

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

## Example

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

$$x = ab\,aa\,bab\,\underline{aa}\,baababaabab\,aa\,b\,\underline{aa}\,babaabaab\cdots$$

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

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

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

## Example

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

$$x = abaababaabaababaababaabaababaabaab\cdots$$

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

## Theorem (Balková, Palentová, Steiner, 2008)

Let $S$ be a uniformly recurrent tree [a] set containing the alphabet $A$. Then for every $w \in S$, the set $\mathcal{R}_s(w)$ has $\mathrm{Card}\,(A)$ elements.

a. They proved it for the more general class of *neutral* sets.

**Theorem** (2014, *J. Pure Appl. Algebra*)

Let $S$ be a uniformly recurrent tree set 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 (2014, *J. Pure Appl. Algebra*)**

Let $S$ be a uniformly recurrent tree set 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,

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

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


Merci