The link between return words and extensions of factors

France Gheeraert

July 4, 2025

LAMFA

Some notations

We consider

- a two-sided infinite sequence x

Some notations

We consider

- a two-sided infinite sequence x
- over the alphabet $\mathcal{A}$ (minimal)

Some notations

We consider

- a two-sided infinite sequence x
- over the alphabet $\mathcal{A}$ (minimal)
- where its language is denoted $\mathcal{L}(x)$

Some notations

We consider

- a two-sided infinite sequence x
- over the alphabet $\mathcal{A}$ (minimal)
- where its language is denoted $\mathcal{L}(x)$
- and $\mathcal{L}_n(x)$ is the set of length-n words in x.

1. The protagonists

Extensions

For any $w \in \mathcal{L}(x)$,

- its *left extensions* are the letters in

$$\mathrm{E}^L_x(w) = \{a \in \mathcal{A} : aw \in \mathcal{L}(x)\},$$

Extensions

For any $w \in \mathcal{L}(x)$,

- its *left extensions* are the letters in

$$\mathrm{E}_x^L(w) = \{a \in \mathcal{A} : aw \in \mathcal{L}(x)\},$$

- its *right extensions* are the letters in

$$\mathrm{E}_x^R(w) = \{b \in \mathcal{A} : wb \in \mathcal{L}(x)\},$$

Extensions

For any $w \in \mathcal{L}(x)$,

- its *left extensions* are the letters in

$$\mathrm{E}_x^L(w) = \{a \in \mathcal{A} : aw \in \mathcal{L}(x)\},$$

- its *right extensions* are the letters in

$$\mathrm{E}_x^R(w) = \{b \in \mathcal{A} : wb \in \mathcal{L}(x)\},$$

- its *bi-extensions* are the pairs of letters in

$$\mathrm{E}_x(w) = \{(a, b) \in \mathcal{A}^2 : awb \in \mathcal{L}(x)\}.$$

Extensions and factor complexity

Definition

- $s_x\colon n \mapsto p_x(n+1) - p_x(n)$ is the *first difference of complexity*;

Extensions and factor complexity

Definition

- $s_x \colon n \mapsto p_x(n+1) - p_x(n)$ is the *first difference of complexity*;
- $m_x(w) = \#\mathrm{E}_x(w) - \#\mathrm{E}_x^L(w) - \#\mathrm{E}_x^R(w) + 1$ is the *multiplicity* of $w \in \mathcal{L}(x)$.

Extensions and factor complexity

Definition

- $s_x \colon n \mapsto p_x(n+1) - p_x(n)$ is the *first difference of complexity*;
- $m_x(w) = \#\mathrm{E}_x(w) - \#\mathrm{E}_x^L(w) - \#\mathrm{E}_x^R(w) + 1$ is the *multiplicity* of $w \in \mathcal{L}(x)$.

Proposition

For all n,

$$s_x(n+1) - s_x(n) = \sum_{w \in \mathcal{L}_n(x)} m_x(w).$$

Return words

$$x = \cdots 0011002001100220011000 11\cdots$$

$\mathcal{R}_x(00) =$

Return words

$$x = \cdots \mid 00110020011002200110001 1\cdots$$

$\mathcal{R}_x(00) =$

Return words

$$x = \cdots \mid 0011 \mid 0020011002200110011\cdots$$

$\mathcal{R}_x(00) = \{0011$

Return words

$$x = \cdots \mid 0011 \mid 002 \mid 0011002200110011\cdots$$

$\mathcal{R}_x(00) = \{0011, 002$

Return words

$$x = \cdots \mid 0011 \mid 002 \mid 0011 \mid 0022001100011\cdots$$

$\mathcal{R}_x(00) = \{0011, 002$

Return words

$$x = \cdots \mid 0011 \mid 002 \mid 0011 \mid 0022 \mid 001100011\cdots$$

$\mathcal{R}_x(00) = \{0011, 002, 0022$

Return words

$$x = \cdots \mid 0011 \mid 002 \mid 0011 \mid 0022 \mid 0011 \mid 00011 \cdots$$

$\mathcal{R}_x(00) = \{0011, 002, 0022$

Return words

$$x = \cdots \mid 0011 \mid 002 \mid 0011 \mid 0022 \mid 0011 \mid 0 \mid 0011 \cdots$$

$$\mathcal{R}_x(00) = \{0011, 002, 0022, 0\}$$

Return words

$$x = \cdots \mid 0011 \mid 002 \mid 0011 \mid 0022 \mid 0011 \mid 0 \mid 0011 \cdots$$

$\mathcal{R}_x(00) = \{0011, 002, 0022, 0\}$

Definition

A *return word* for w is a word u such that

$$uw \in \mathcal{L}(x) \cap w\mathcal{A}^+ \setminus \mathcal{A}^+ w \mathcal{A}^+.$$

The *set of return words* for w is denoted $\mathcal{R}_x(w)$.

Return words

$$x = \cdots \mid 0011 \mid 002 \mid 0011 \mid 0022 \mid 0011 \mid 0 \mid 0011 \cdots$$

$\mathcal{R}_x(00) = \{0011, 002, 0022, 0\}$

Definition

A *return word* for w is a word u such that

$$uw \in \mathcal{L}(x) \cap w\mathcal{A}^+ \setminus \mathcal{A}^+ w \mathcal{A}^+.$$

The *set of return words* for w is denoted $\mathcal{R}_x(w)$.

Remark: $\mathcal{R}_x(\varepsilon) = \mathcal{A}$

Decomposition and derivation

$$x = \quad \cdots \quad 0011 \quad 002 \quad 0011 \quad 0022 \quad 0011 \quad 0 \quad 0011 \cdots$$

Decomposition and derivation

$$
\begin{array}{rccccccccc}
x = & \cdots & 0011 & 002 & 0011 & 0022 & 0011 & 0 & 0011\cdots \\
 & & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\
D_{00}(x) = & \cdots & 0 & 1 & 0 & 2 & 0 & 3 & \cdots
\end{array}
$$

Decomposition and derivation

$$\begin{array}{rccccccccc}
x = & \cdots & 0011 & 002 & 0011 & 0022 & 0011 & 0 & 0011\cdots \\
 & & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\
D_{00}(x) = & \cdots & 0 & 1 & 0 & 2 & 0 & 3 & \cdots
\end{array}$$

Definition

The derived sequence of x with respect to w is the sequence $D_w(x) \in \mathcal{B}^{\mathbb{Z}}$ such that $x = \theta(D_w(x))$ for a morphism θ defining a bijection between $\mathcal{B}$ and $\mathcal{R}_x(w)$.

Comparison

Extensions

- letters or pairs of letters

Return words

- (long) words

Comparison

Extensions

- letters or pairs of letters
- very local

Return words

- (long) words
- mildly local

Comparison

Extensions

- letters or pairs of letters
- very local
- gives the complexity of x

Return words

- (long) words
- mildly local
- gives a decomposition of x

Comparison

Extensions

- letters or pairs of letters
- very local
- gives the complexity of x
- used to define or characterize famous families of sequences

Return words

- (long) words
- mildly local
- gives a decomposition of x
- used for S-adic representations and critical exponents

First observation

Knowing the return words for w,

- we know the left and right extensions of w,

First observation

Knowing the return words for w,

- we know the left and right extensions of w,
  example: if the return words for 011 are $01101, 0110, 011101$,
  then $\mathrm{E}_x^L(w) = \{0, 1\}$ and $\mathrm{E}_x^R(w) = \{0, 1\}$

First observation

Knowing the return words for w,

- we know the left and right extensions of w,
  example: if the return words for 011 are $01101, 0110, 011101$,
  then $\mathrm{E}_x^L(w) = \{0, 1\}$ and $\mathrm{E}_x^R(w) = \{0, 1\}$

First observation

Knowing the return words for w,

- we know the left and right extensions of w,
  example: if the return words for 011 are $01101, 0110, 011101$,
  then $\mathrm{E}_x^L(w) = \{0,1\}$ and $\mathrm{E}_x^R(w) = \{0,1\}$

First observation

Knowing the return words for w,

- we know the left and right extensions of w,
  example: if the return words for 011 are $01101, 0110, 011101$, then $\mathrm{E}_x^L(w) = \{0, 1\}$ and $\mathrm{E}_x^R(w) = \{0, 1\}$
  and in particular,

$$\#\mathcal{R}_x(w) \geq \max\{\#\mathrm{E}_x^L(w), \#\mathrm{E}_x^R(w)\};$$

First observation

Knowing the return words for w,

- we know the left and right extensions of w,
  example: if the return words for 011 are $01101, 0110, 011101$,
  then $\mathrm{E}_x^L(w) = \{0, 1\}$ and $\mathrm{E}_x^R(w) = \{0, 1\}$
  and in particular,

$$\#\mathcal{R}_x(w) \geq \max\{\#\mathrm{E}_x^L(w), \#\mathrm{E}_x^R(w)\};$$

- we can't know the bi-extensions of w.

First observation

Knowing the return words for w,

- we know the left and right extensions of w,
  example: if the return words for 011 are $01101, 0110, 011101$, then $\mathrm{E}_x^L(w) = \{0,1\}$ and $\mathrm{E}_x^R(w) = \{0,1\}$
  and in particular,
  $$\#\mathcal{R}_x(w) \geq \max\{\#\mathrm{E}_x^L(w), \#\mathrm{E}_x^R(w)\};$$
- we can't know the bi-extensions of w.

Knowing the extensions of w,

- we can't know the return words for w.

2. Number of extensions and return words

A first result

Theorem (Vuillon)

x is Sturmian if and only if $\#\mathcal{R}_x(w) = 2$ for all $w \in \mathcal{L}(x)$.

A first result

Theorem (Vuillon)

x is Sturmian if and only if $\#\mathcal{R}_x(w) = 2$ for all $w \in \mathcal{L}(x)$.

Tools for the proof:

- Rauzy graphs of Sturmian sequences

A first result

Theorem (Vuillon)

x is Sturmian if and only if $\#\mathcal{R}_x(w) = 2$ for all $w \in \mathcal{L}(x)$.

Tools for the proof:

- Rauzy graphs of Sturmian sequences
- $\#\mathcal{R}_x(w) = 2$ for all $w \in \mathcal{L}(x) \implies \#\mathcal{R}_{D_a(x)}(u) = 2$ for all $u \in \mathcal{L}(D_w(x))$

A first result

Theorem (Vuillon)

x is Sturmian if and only if $\#\mathcal{R}_x(w) = 2$ for all $w \in \mathcal{L}(x)$.

Tools for the proof:

- Rauzy graphs of Sturmian sequences
- $\#\mathcal{R}_x(w) = 2$ for all $w \in \mathcal{L}(x) \implies \#\mathcal{R}_{D_a(x)}(u) = 2$ for all $u \in \mathcal{L}(D_w(x))$
- S-adic characterization of Sturmian sequences

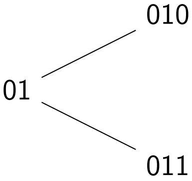
A tree to count return words

For the Thue-Morse sequence,

01

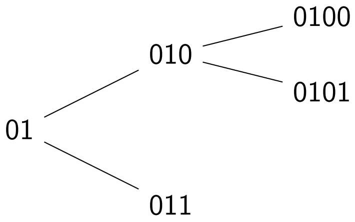
A tree to count return words

For the Thue-Morse sequence,



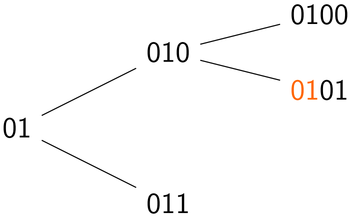
A tree to count return words

For the Thue-Morse sequence,



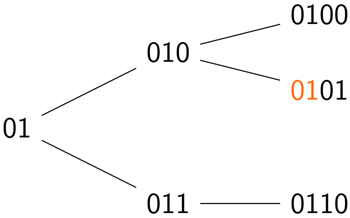
A tree to count return words

For the Thue-Morse sequence,



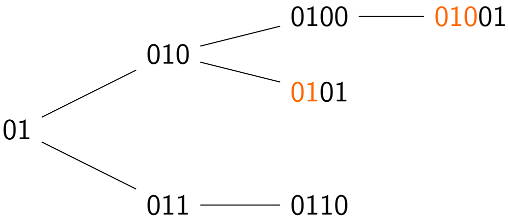
A tree to count return words

For the Thue-Morse sequence,



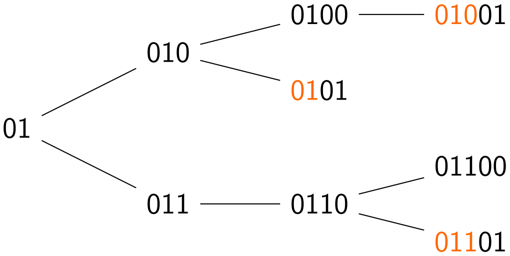
A tree to count return words

For the Thue-Morse sequence,



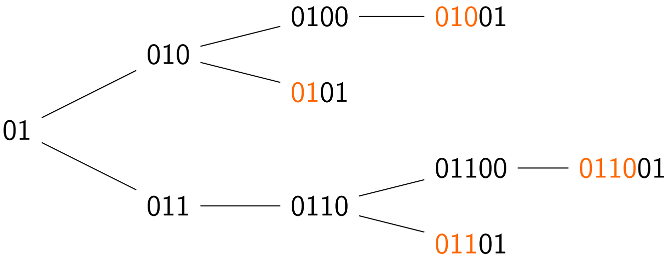
A tree to count return words

For the Thue-Morse sequence,



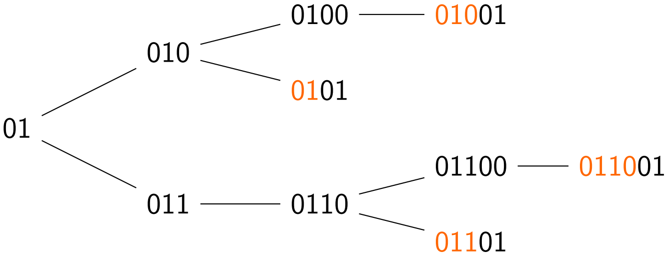
A tree to count return words

For the Thue-Morse sequence,



A tree to count return words

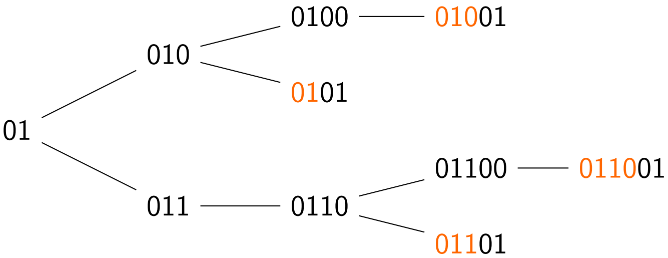
For the Thue-Morse sequence,



return words

A tree to count return words

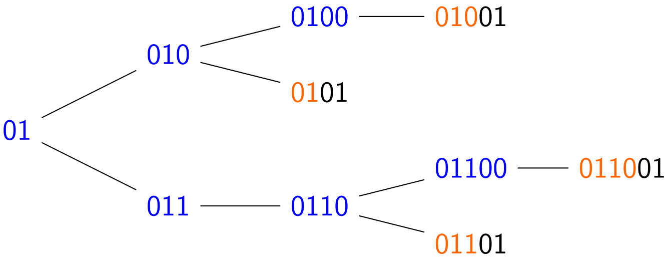
For the Thue-Morse sequence,



\# return words = # branches

A tree to count return words

For the Thue-Morse sequence,



$$\begin{aligned}\#\text{ return words} &= \#\text{ branches}\\ &= 1 + \sum_{u \in S} \left(\# \mathrm{E}_x^R(u) - 1 \right)\end{aligned}$$

General case

Proposition (Balková, Pelantová, Steiner)

If x is uniformly recurrent, then for every $w \in \mathcal{L}(x)$,

$$\#\mathcal{R}_x(w) = 1 + \sum_{u \in S_w} \left(\#\mathrm{E}_x^R(u) - 1\right)$$

where $S_w = \{u \in \mathcal{L}(x) \cap w\mathcal{A}^* : |u|_w = 1\}$

General case

Proposition (Balková, Pelantová, Steiner)

If x is uniformly recurrent, then for every $w \in \mathcal{L}(x)$,

$$\#\mathcal{R}_x(w) = 1 + \sum_{u \in S_w} \left(\#\mathrm{E}_x^R(u) - 1 \right)$$

where $S_w = \{u \in \mathcal{L}(x) \cap w\mathcal{A}^* : |u|_w = 1\}$ is an x-maximal suffix code.

Definition

A set $S \subseteq \mathcal{L}(x)$ is an *x-maximal suffix code* if

- for all $u, v \in S$, if $u \in \mathrm{Suff}(v)$, then $u = v$ (suffix code)
- for each $v \in \mathcal{L}(x)$, there exists $u \in S$ such that $v \in \mathrm{Suff}(u)$ or $u \in \mathrm{Suff}(v)$ (x-maximal).

Working on sums

Lemma

If S is a finite x-maximal suffix code, then

$$\sum_{u \in S} \Big(\#\mathrm{E}_x^R(u) - 1 \Big) = \#\mathcal{A} - 1 + \sum_{\substack{u \in \mathcal{L}(x) \\ \mathrm{Suff}(u) \cap S = \emptyset}} m_x(u).$$

Working on sums

Lemma

If S is a finite x-maximal suffix code, then

$$\sum_{u \in S} \left(\#\mathrm{E}_x^R(u) - 1 \right) = \#\mathcal{A} - 1 + \sum_{\substack{u \in \mathcal{L}(x) \\ \mathrm{Suff}(u) \cap S = \emptyset}} m_x(u).$$

Proposition (G.)

If x is uniformly recurrent, then for every $w \in \mathcal{L}(x)$,

$$\#\mathcal{R}_x(w) = \#\mathcal{A} + \sum_{\substack{u \in \mathcal{L}(x) \\ |u|_w = 0}} m_x(u).$$

Working on sums

Lemma

If S is a finite x-maximal suffix code, then

$$\sum_{u \in S} \left(\#\mathrm{E}_x^R(u) - 1 \right) = \#\mathcal{A} - 1 + \sum_{\substack{u \in \mathcal{L}(x) \\ \mathrm{Suff}(u) \cap S = \emptyset}} m_x(u).$$

Proposition (G.)

If x is uniformly recurrent, then for every $w \in \mathcal{L}(x)$,

$$\#\mathcal{R}_x(w) = \#\mathcal{A} + \sum_{\substack{u \in \mathcal{L}(x) \\ |u|_w = 0}} m_x(u).$$

Similar sums to study the evolution of factor complexity when applying a morphism.

Return words for a set

Condition:

$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots 0011002001100220011000110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) =$$
$$\mathcal{CR}_x(W) =$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0011002001100220011000110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) =$$
$$\mathcal{CR}_x(W) =$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0 \mid 011002001100220011000110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) = \{0$$
$$\mathcal{CR}_x(W) = \{0011$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0 \mid 011 \mid 002001100220011000110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) = \{0, 011$$
$$\mathcal{CR}_x(W) = \{0011, 01100$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0 \mid 011 \mid 002 \mid 001100220011000110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) = \{0, 011, 002$$
$$\mathcal{CR}_x(W) = \{0011, 01100, 00200$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0 \mid 011 \mid 002 \mid 0 \mid 01100220011000110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) = \{0, 011, 002$$
$$\mathcal{CR}_x(W) = \{0011, 01100, 00200$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0 \mid 011 \mid 002 \mid 0 \mid 011 \mid 0022001100011 0 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) = \{0, 011, 002$$
$$\mathcal{CR}_x(W) = \{0011, 01100, 00200$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0 \mid 011 \mid 002 \mid 0 \mid 011 \mid 0022 \mid 0011000110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) = \{0, 011, 002, 0022$$
$$\mathcal{CR}_x(W) = \{0011, 01100, 00200, 002200$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0 \mid 011 \mid 002 \mid 0 \mid 011 \mid 0022 \mid 0 \mid 011000110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) = \{0, 011, 002, 0022$$
$$\mathcal{CR}_x(W) = \{0011, 01100, 00200, 002200$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0 \mid 011 \mid 002 \mid 0 \mid 011 \mid 0022 \mid 0 \mid 011 \mid 000110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) = \{0, 011, 002, 0022$$
$$\mathcal{CR}_x(W) = \{0011, 01100, 00200, 002200$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0 \mid 011 \mid 002 \mid 0 \mid 011 \mid 0022 \mid 0 \mid 011 \mid 0 \mid 00110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) = \{0, 011, 002, 0022$$
$$\mathcal{CR}_x(W) = \{0011, 01100, 00200, 002200, 000$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0 \mid 011 \mid 002 \mid 0 \mid 011 \mid 0022 \mid 0 \mid 011 \mid 0 \mid 0 \mid 0110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) = \{0, 011, 002, 0022\}$$
$$\mathcal{CR}_x(W) = \{0011, 01100, 00200, 002200, 000\}$$

Return words for a set

Condition:
$W \subseteq \mathcal{L}(x)$ is a factor code, i.e. no element is factor of another

Example:

$$x = \cdots \mid 0 \mid 011 \mid 002 \mid 0 \mid 011 \mid 0022 \mid 0 \mid 011 \mid 0 \mid 0 \mid 0110 \cdots$$

and $W = \{00, 011\}$

$$\mathcal{R}_x(W) = \{0, 011, 002, 0022\}$$
$$\mathcal{CR}_x(W) = \{0011, 01100, 00200, 002200, 000\}$$

Definition

The *complete return words* for W are the elements of

$$\mathcal{CR}_x(W) = \mathcal{L}(x) \cap W\mathcal{A}^+ \cap \mathcal{A}^+W \setminus \mathcal{A}^+W\mathcal{A}^+.$$

Number of return words for a set

Proposition (G.)

If x is uniformly recurrent, then for every factor code $W \subseteq \mathcal{L}(X)$,

$$\#\mathcal{CR}_x(W) = \#W + \sum_{u \in S_W} \left(\#\mathrm{E}_x^R(u) - 1\right)$$

where $S_W = \{u \in \mathcal{L}(x) \cap W\mathcal{A}^* : |u|_W = 1\}$.

Number of return words for a set

Proposition (G.)

If x is uniformly recurrent, then for every factor code $W \subseteq \mathcal{L}(X)$,

$$\begin{aligned}\#\mathcal{CR}_x(W) &= \#W + \sum_{u \in S_W} \left(\#\mathrm{E}_x^R(u) - 1\right) \\ &= \#W - 1 + \#\mathcal{A} + \sum_{\substack{u \in \mathcal{L}(x) \\ |u|_W = 0}} m_x(u)\end{aligned}$$

where $S_W = \{u \in \mathcal{L}(x) \cap W\mathcal{A}^* : |u|_W = 1\}$.

Neutrality

Definition

A word $w \in \mathcal{L}(x)$ is

- *neutral* if $m_x(w) = 0$,
- *weak* if $m_x(w) < 0$,
- *strong* if $m_x(w) > 0$.

Neutrality

Definition

A word $w \in \mathcal{L}(x)$ is

- *neutral* if $m_x(w) = 0$,
- *weak* if $m_x(w) < 0$,
- *strong* if $m_x(w) > 0$.

Definition

A sequence x is *eventually neutral* if any long enough $w \in \mathcal{L}(x)$ is neutral.

Theorem (Balková, Pelantová, Steiner ; Dolce, Perrin)

Let x be uniformly recurrent with no weak $w \in \mathcal{L}_{\geq N}(x)$.
The following are equivalent:

Theorem (Balková, Pelantová, Steiner ; Dolce, Perrin)

Let x be uniformly recurrent with no weak $w \in \mathcal{L}_{\geq N}(x)$.
The following are equivalent:

1. *every $w \in \mathcal{L}_{\geq N}(x)$ is neutral;*
2. *$\exists\, K$ st. $\#\mathcal{CR}_x(W) = \#W + K$ for every $W \subseteq \mathcal{L}_{\geq N}(x)$;*
3. *$\exists\, K$ st. $\#\mathcal{R}_x(w) = 1 + K$ for every $w \in \mathcal{L}_{\geq N}(x)$.*

Moreover, $K = s_x(N)$.

Theorem (Balková, Pelantová, Steiner ; Dolce, Perrin)

Let x be uniformly recurrent with no weak $w \in \mathcal{L}_{\geq N}(x)$.
The following are equivalent:

1. *every $w \in \mathcal{L}_{\geq N}(x)$ is neutral;*
2. *$\exists\, K$ st. $\#\mathcal{CR}_x(W) = \#W + K$ for every $W \subseteq \mathcal{L}_{\geq N}(x)$;*
3. *$\exists\, K$ st. $\#\mathcal{R}_x(w) = 1 + K$ for every $w \in \mathcal{L}_{\geq N}(x)$.*

Moreover, $K = s_x(N)$.

$\Downarrow$:

$$\#\mathcal{CR}_x(W) = \#W + \#\mathcal{A} - 1 + \sum_{\substack{u \in \mathcal{L}(x) \\ |u|_W = 0}} m_x(u)$$

Theorem (Balková, Pelantová, Steiner ; Dolce, Perrin)

Let x be uniformly recurrent with no weak $w \in \mathcal{L}_{\geq N}(x)$.
The following are equivalent:

1. *every $w \in \mathcal{L}_{\geq N}(x)$ is neutral;*
2. *$\exists\, K$ st. $\#\mathcal{CR}_x(W) = \#W + K$ for every $W \subseteq \mathcal{L}_{\geq N}(x)$;*
3. *$\exists\, K$ st. $\#\mathcal{R}_x(w) = 1 + K$ for every $w \in \mathcal{L}_{\geq N}(x)$.*

Moreover, $K = s_x(N)$.

$\Downarrow$:

$$\#\mathcal{CR}_x(W) = \#W + \#\mathcal{A} - 1 + \sum_{\substack{u \in \mathcal{L}(x) \\ |u|_W = 0}} m_x(u)$$

$\Uparrow$: if $m_x(v) > 0$, then $\#\mathcal{R}_x(v) < \#\mathcal{R}_x(va)$ for $a \in \mathrm{E}_x^R(v)$

Low number of return words

Proposition (adaptation of Balková, Pelantová, Steiner)

Let x be uniformly recurrent.

1. $\#\mathcal{R}_x(w) = 1$ *for every long enough* w $\iff$ x *is ev. neutral with* $\lim_n s_x(n) = 0$ *(x is periodic).*

Low number of return words

Proposition (adaptation of Balková, Pelantová, Steiner)

Let x be uniformly recurrent.

1. $\#\mathcal{R}_x(w) = 1$ *for every long enough* w
   $\iff$ *x is ev. neutral with* $\lim_n s_x(n) = 0$ *(x is periodic).*
2. $\#\mathcal{R}_x(w) = 2$ *for every long enough* w
   $\iff$ *x is ev. neutral with* $\lim_n s_x(n) = 1$ *(x is quasi-Sturmian).*

Low number of return words

Proposition (adaptation of Balková, Pelantová, Steiner)

Let x be uniformly recurrent.

1. $\#\mathcal{R}_x(w) = 1$ *for every long enough* w
   $\iff$ x *is ev. neutral with* $\lim_n s_x(n) = 0$ *(x is periodic).*
2. $\#\mathcal{R}_x(w) = 2$ *for every long enough* w
   $\iff$ x *is ev. neutral with* $\lim_n s_x(n) = 1$ *(x is quasi-Sturmian).*
3. $\#\mathcal{R}_x(w) = 3$ *for every long enough* w
   $\iff$ x *is ev. neutral with* $\lim_n s_x(n) = 2$*.*

Low number of return words

Proposition (adaptation of Balková, Pelantová, Steiner)

Let x be uniformly recurrent.

1. $\#\mathcal{R}_x(w) = 1$ *for every long enough* w
   $\iff$ *x is ev. neutral with* $\lim_n s_x(n) = 0$ *(x is periodic).*
2. $\#\mathcal{R}_x(w) = 2$ *for every long enough* w
   $\iff$ *x is ev. neutral with* $\lim_n s_x(n) = 1$ *(x is quasi-Sturmian).*
3. $\#\mathcal{R}_x(w) = 3$ *for every long enough* w
   $\iff$ *x is ev. neutral with* $\lim_n s_x(n) = 2$.

Counter-example for $\#\mathcal{R}_x(w) = 4$: Thue-Morse

3. Structure of extensions and return words

Structure of return words: embedding in the free group

The free group $F_{\mathcal{A}}$ is the natural algebraic extension of $\mathcal{A}^*$ with the operations:

- inversion : $(w_1 w_2 \cdots w_n)^{-1} = w_n^{-1} \cdots w_2^{-1} w_1^{-1}$

Structure of return words: embedding in the free group

The free group $F_{\mathcal{A}}$ is the natural algebraic extension of $\mathcal{A}^*$ with the operations:

- inversion : $(w_1 w_2 \cdots w_n)^{-1} = w_n^{-1} \cdots w_2^{-1} w_1^{-1}$
- concatenation-cancellation : when concatenating, we erase factors aa^{-1} and $a^{-1}a$.

Structure of return words: embedding in the free group

The free group $F_{\mathcal{A}}$ is the natural algebraic extension of $\mathcal{A}^*$ with the operations:

- inversion : $(w_1 w_2 \cdots w_n)^{-1} = w_n^{-1} \cdots w_2^{-1} w_1^{-1}$
- concatenation-cancellation : when concatenating, we erase factors aa^{-1} and $a^{-1}a$.

Examples:

$$(ab^{-1}ac^{-1})^{-1} = ca^{-1}ba^{-1}$$

Structure of return words: embedding in the free group

The free group $F_{\mathcal{A}}$ is the natural algebraic extension of $\mathcal{A}^*$ with the operations:

- inversion : $(w_1 w_2 \cdots w_n)^{-1} = w_n^{-1} \cdots w_2^{-1} w_1^{-1}$
- concatenation-cancellation : when concatenating, we erase factors aa^{-1} and $a^{-1}a$.

Examples:

$$(ab^{-1}ac^{-1})^{-1} = ca^{-1}ba^{-1}$$
$$(abba^{-1}) \cdot (ab^{-1}aa) = abba^{-1}ab^{-1}aa$$

Structure of return words: embedding in the free group

The free group $F_{\mathcal{A}}$ is the natural algebraic extension of $\mathcal{A}^*$ with the operations:

- inversion : $(w_1 w_2 \cdots w_n)^{-1} = w_n^{-1} \cdots w_2^{-1} w_1^{-1}$
- concatenation-cancellation : when concatenating, we erase factors aa^{-1} and $a^{-1}a$.

Examples:

$$(ab^{-1}ac^{-1})^{-1} = ca^{-1}ba^{-1}$$
$$(abba^{-1}) \cdot (ab^{-1}aa) = abb\color{orange}{a^{-1}a}\color{black}{b^{-1}aa}$$

Structure of return words: embedding in the free group

The free group $F_{\mathcal{A}}$ is the natural algebraic extension of $\mathcal{A}^*$ with the operations:

- inversion : $(w_1 w_2 \cdots w_n)^{-1} = w_n^{-1} \cdots w_2^{-1} w_1^{-1}$
- concatenation-cancellation : when concatenating, we erase factors aa^{-1} and $a^{-1}a$.

Examples:

$$(ab^{-1}ac^{-1})^{-1} = ca^{-1}ba^{-1}$$
$$(abba^{-1}) \cdot (ab^{-1}aa) = ab\,bb^{-1}aa$$

Structure of return words: embedding in the free group

The free group $F_{\mathcal{A}}$ is the natural algebraic extension of $\mathcal{A}^*$ with the operations:

- inversion : $(w_1 w_2 \cdots w_n)^{-1} = w_n^{-1} \cdots w_2^{-1} w_1^{-1}$
- concatenation-cancellation : when concatenating, we erase factors aa^{-1} and $a^{-1}a$.

Examples:

$$(ab^{-1}ac^{-1})^{-1} = ca^{-1}ba^{-1}$$
$$(abba^{-1}) \cdot (ab^{-1}aa) = abaa$$

Free and generating sets

Definition

A set $S \subseteq F_{\mathcal{A}}$ is

- *free* if $s_1^{\eta_1} \cdots s_n^{\eta_n} \neq \varepsilon$ for any choice of $n \geq 1$, $s_1, \ldots, s_n \in S$, and of $\eta_1, \ldots, \eta_n \in \{1, -1\}$ such that $\eta_i = \eta_{i+1}$ if $s_i = s_{i+1}$;

Free and generating sets

Definition

A set $S \subseteq F_{\mathcal{A}}$ is

- *free* if $s_1^{\eta_1} \cdots s_n^{\eta_n} \neq \varepsilon$ for any choice of $n \geq 1$, $s_1, \ldots, s_n \in S$, and of $\eta_1, \ldots, \eta_n \in \{1, -1\}$ such that $\eta_i = \eta_{i+1}$ if $s_i = s_{i+1}$;
- *generating* if any element of $F_{\mathcal{A}}$ is obtained as a product of elements of S and their inverses;

Free and generating sets

Definition

A set $S \subseteq F_{\mathcal{A}}$ is

- *free* if $s_1^{\eta_1} \cdots s_n^{\eta_n} \neq \varepsilon$ for any choice of $n \geq 1$, $s_1, \ldots, s_n \in S$, and of $\eta_1, \ldots, \eta_n \in \{1, -1\}$ such that $\eta_i = \eta_{i+1}$ if $s_i = s_{i+1}$;
- *generating* if any element of $F_{\mathcal{A}}$ is obtained as a product of elements of S and their inverses;
- *a basis of $F_{\mathcal{A}}$* if it is free and generating.

Free and generating sets

Definition

A set $S \subseteq F_{\mathcal{A}}$ is

- *free* if $s_1^{\eta_1} \cdots s_n^{\eta_n} \neq \varepsilon$ for any choice of $n \geq 1$, $s_1, \ldots, s_n \in S$, and of $\eta_1, \ldots, \eta_n \in \{1, -1\}$ such that $\eta_i = \eta_{i+1}$ if $s_i = s_{i+1}$;
- *generating* if any element of $F_{\mathcal{A}}$ is obtained as a product of elements of S and their inverses;
- *a basis of $F_{\mathcal{A}}$* if it is free and generating.

Examples: $S = \{0, 01, 011\}$ is generating:

$$0 \in S \quad \textit{and} \quad 1 = (0)^{-1}(01)$$

Free and generating sets

Definition

A set $S \subseteq F_{\mathcal{A}}$ is

- *free* if $s_1^{\eta_1} \cdots s_n^{\eta_n} \neq \varepsilon$ for any choice of $n \geq 1$, $s_1, \ldots, s_n \in S$, and of $\eta_1, \ldots, \eta_n \in \{1, -1\}$ such that $\eta_i = \eta_{i+1}$ if $s_i = s_{i+1}$;
- *generating* if any element of $F_{\mathcal{A}}$ is obtained as a product of elements of S and their inverses;
- *a basis of $F_{\mathcal{A}}$* if it is free and generating.

Examples: $S = \{0, 01, 011\}$ is generating:

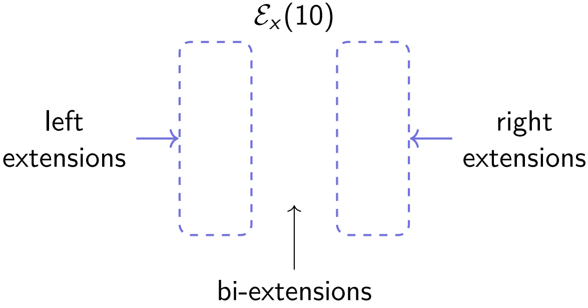
$$0 \in S \quad \textit{and} \quad 1 = (0)^{-1}(01)$$

but not free:

$$(01)(0)^{-1}(01)(011)^{-1} = \varepsilon.$$

Structure of extensions: extension graph

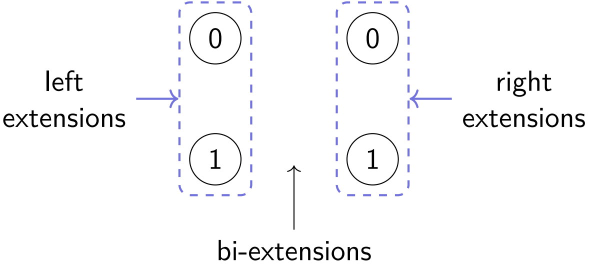
$$x = \cdots 10010011001001001101100 \cdots$$



Structure of extensions: extension graph

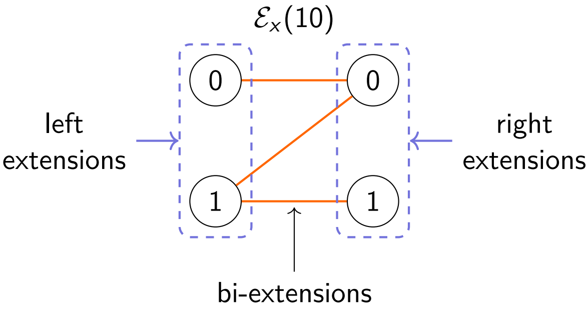
$$x = \cdots 100100110010010011 01100 \cdots$$

$$\mathcal{E}_x(10)$$



Structure of extensions: extension graph

$$x = \cdots 10010011001001001101100\cdots$$



Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 010020010020 \cdots$$

10 01

00

20 02

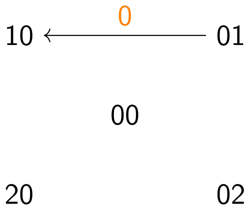
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 010020010020 \cdots$$



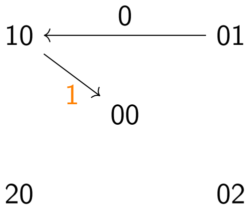
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 0100200 10020 \cdots$$



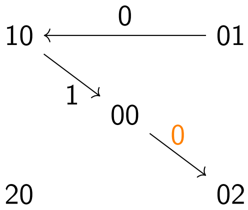
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 01002001 0020 \cdots$$



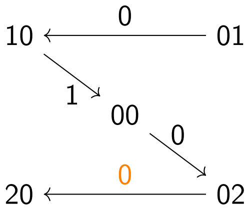
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 010020010020 \cdots$$



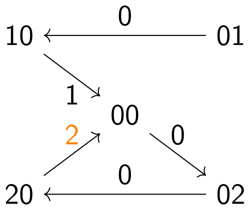
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 0100\textcolor{orange}{2}\textcolor{blue}{00}10020 \cdots$$



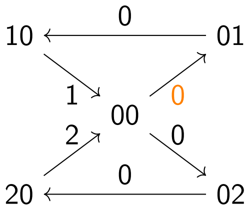
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 0100200010020 \cdots$$



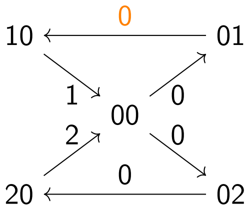
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 010020010020 \cdots$$



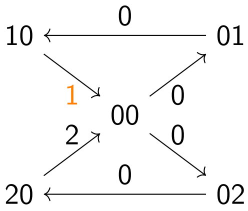
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 0100200\textcolor{orange}{1}\textcolor{blue}{00}20 \cdots$$



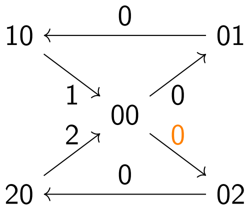
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 010020010020 \cdots$$



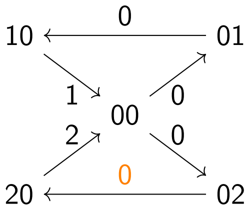
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 010020010020 \cdots$$



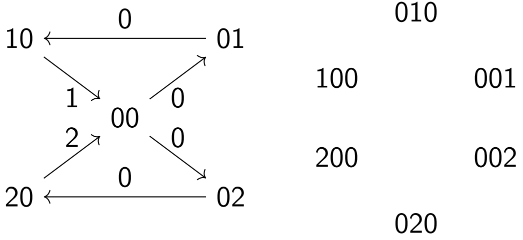
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 010020010020 \cdots$$



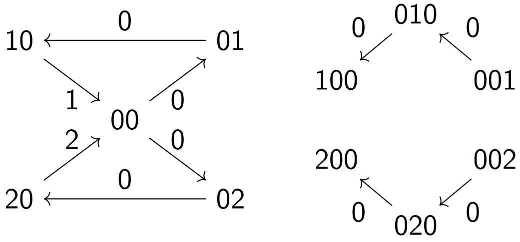
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 010020010020 \cdots$$



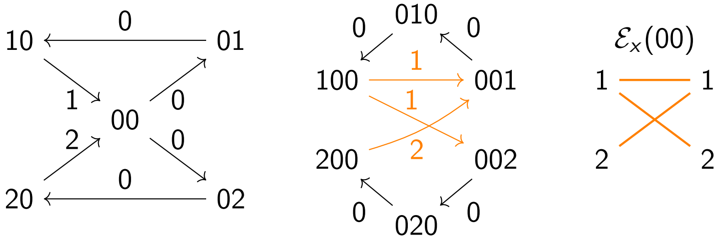
Link with Rauzy graphs

Definition

The *Rauzy graph of order n* is the graph $\Gamma_x(n)$ such that

- the vertices are the elements of $\mathcal{L}_n(x)$;
- there is an edge from u to v with label $a \in \mathcal{A}$ if $av \in u\mathcal{A} \cap \mathcal{L}(x)$.

$$\cdots 010020010020 \cdots$$



Dendricity and cie.

Definition

A word $w \in \mathcal{L}(x)$ is

- *acyclic* if $\mathcal{E}_x(w)$ is acyclic;
- *connected* if $\mathcal{E}_x(w)$ is connected;
- *dendric* if $\mathcal{E}_x(w)$ is a tree.

Dendricity and cie.

Definition

A word $w \in \mathcal{L}(x)$ is

- *acyclic* if $\mathcal{E}_x(w)$ is acyclic;
- *connected* if $\mathcal{E}_x(w)$ is connected;
- *dendric* if $\mathcal{E}_x(w)$ is a tree.

Definition

A sequence x is

- *connected* if every $w \in \mathcal{L}(x)$ is connected;
- *dendric* if every $w \in \mathcal{L}(x)$ is dendric.

Connected extensions and generating return words

Theorem (Berthé, De Felice, Dolce, Leroy, Perrin, Reutenauer, Rindone)

If x is uniformly recurrent and connected, then $\mathcal{R}_x(w)$ generates $F_{\mathcal{A}}$ for every $w \in \mathcal{L}(x)$.

Connected extensions and generating return words

Theorem (Berthé, De Felice, Dolce, Leroy, Perrin, Reutenauer, Rindone)

If x is uniformly recurrent and connected, then $\mathcal{R}_x(w)$ generates $F_{\mathcal{A}}$ for every $w \in \mathcal{L}(x)$.

Idea of proof:

- for each w, there exists u such that $\langle \mathcal{R}_x(w) \rangle$ contains the group G generated by the paths based on u in $\Gamma_x(|u|)$;

Connected extensions and generating return words

Theorem (Berthé, De Felice, Dolce, Leroy, Perrin, Reutenauer, Rindone)

If x is uniformly recurrent and connected, then $\mathcal{R}_x(w)$ generates $F_{\mathcal{A}}$ for every $w \in \mathcal{L}(x)$.

Idea of proof:

- for each w, there exists u such that $\langle \mathcal{R}_x(w) \rangle$ contains the group G generated by the paths based on u in $\Gamma_x(|u|)$;
- since x is connected, G is also the group generated by the paths based on $u_{[1,n]}$ in $\Gamma_x(n)$;

Connected extensions and generating return words

Theorem (Berthé, De Felice, Dolce, Leroy, Perrin, Reutenauer, Rindone)

If x is uniformly recurrent and connected, then $\mathcal{R}_x(w)$ generates $F_{\mathcal{A}}$ for every $w \in \mathcal{L}(x)$.

Idea of proof:

- for each w, there exists u such that $\langle \mathcal{R}_x(w) \rangle$ contains the group G generated by the paths based on u in $\Gamma_x(|u|)$;
- since x is connected, G is also the group generated by the paths based on $u_{[1,n]}$ in $\Gamma_x(n)$;
- the paths based on ε in $\Gamma_x(0)$ generate $F_{\mathcal{A}}$.

Connected extensions and generating return words

Theorem (Berthé, De Felice, Dolce, Leroy, Perrin, Reutenauer, Rindone)

If x is uniformly recurrent and connected, then $\mathcal{R}_x(w)$ generates $F_{\mathcal{A}}$ for every $w \in \mathcal{L}(x)$.

Idea of proof:

- for each w, there exists u such that $\langle \mathcal{R}_x(w) \rangle$ contains the group G generated by the paths based on u in $\Gamma_x(|u|)$;
- since x is connected, G is also the group generated by the paths based on $u_{[1,n]}$ in $\Gamma_x(n)$;
- the paths based on ε in $\Gamma_x(0)$ generate $F_{\mathcal{A}}$.

Theorem (Goulet-Ouellet)

If x is uniformly recurrent and suffix-connected, then $\mathcal{R}_x(w)$ generates $F_{\mathcal{A}}$ for every $w \in \mathcal{L}(x)$.

Dendric extensions and basis of return words

Theorem (Berthé *et al.* & G., Goulet-Ouellet, Leroy, Stas)

Let x be uniformly recurrent. The following assertions are equivalent:

1. *x is dendric;*
2. *for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;*
3. *for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a tame basis of $F_{\mathcal{A}}$.*

Dendric extensions and basis of return words

Theorem (Berthé *et al.* & G., Goulet-Ouellet, Leroy, Stas)

Let x be uniformly recurrent. The following assertions are equivalent:

1. *x is dendric;*
2. *for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;*
3. *for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a tame basis of $F_{\mathcal{A}}$.*

Proof that $1 \implies 2$:

Dendric extensions and basis of return words

Theorem (Berthé *et al.* & G., Goulet-Ouellet, Leroy, Stas)

Let x be uniformly recurrent. The following assertions are equivalent:

1. *x is dendric;*
2. *for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;*
3. *for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a tame basis of $F_{\mathcal{A}}$.*

Proof that $1 \implies 2$:

- x connected $\implies \mathcal{R}_x(w)$ generates $F_{\mathcal{A}}$

Dendric extensions and basis of return words

Theorem (Berthé *et al.* & G., Goulet-Ouellet, Leroy, Stas)

Let x be uniformly recurrent. The following assertions are equivalent:

1. *x is dendric;*
2. *for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;*
3. *for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a tame basis of $F_{\mathcal{A}}$.*

Proof that $1 \implies 2$:

- x connected $\implies \mathcal{R}_x(w)$ generates $F_{\mathcal{A}}$
- x dendric $\implies x$ neutral $\implies \#\mathcal{R}_x(w) = \#\mathcal{A}$

Proof that $2 \implies 1$

1. x is dendric;
2. for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;

Proof that $2 \implies 1$

1. x is dendric;
2. for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;

- $2. \implies x$ is connected:

- x connected $+ \#\mathcal{R}_x(w) = \#\mathcal{A} \implies x$ dendric:

Proof that $2 \implies 1$

1. x is dendric;
2. for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;

- $2. \implies x$ is connected:
  - if a is a letter: $\mathcal{R}_x(a)$ generates $F_{\mathcal{A}} \implies \varepsilon$ is connected
- x connected $+ \#\mathcal{R}_x(w) = \#\mathcal{A} \implies x$ dendric:

Proof that $2 \implies 1$

1. x is dendric;
2. for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;

- $2. \implies x$ is connected:
  - if a is a letter: $\mathcal{R}_x(a)$ generates $F_{\mathcal{A}} \implies \varepsilon$ is connected
  - ε connected in $D_w(x) \implies w$ connected in x
- x connected $+ \#\mathcal{R}_x(w) = \#\mathcal{A} \implies x$ dendric:

Proof that 2 $\implies$ 1

1. x is dendric;
2. for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;

- 2. $\implies$ x is connected:
  - if a is a letter: $\mathcal{R}_x(a)$ generates $F_{\mathcal{A}} \implies \varepsilon$ is connected
  - ε connected in $D_w(x) \implies w$ connected in x
  - 2. $\implies \mathcal{R}_{D_w(x)}(a)$ generates $F_{\mathcal{L}_1(D_w(x))}$ for any letter a of $D_w(x)$
- x connected $+ \#\mathcal{R}_x(w) = \#\mathcal{A} \implies x$ dendric:

Proof that $2 \implies 1$

1. x is dendric;
2. for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;

- 2. $\implies$ x is connected:
  - if a is a letter: $\mathcal{R}_x(a)$ generates $F_{\mathcal{A}} \implies \varepsilon$ is connected
  - ε connected in $D_w(x) \implies w$ connected in x
  - 2. $\implies \mathcal{R}_{D_w(x)}(a)$ generates $F_{\mathcal{L}_1(D_w(x))}$ for any letter a of $D_w(x)$
- x connected $+ \#\mathcal{R}_x(w) = \#\mathcal{A} \implies x$ dendric:
  - x connected $\implies$ no weak factor

Proof that $2 \implies 1$

1. x is dendric;
2. for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;

- $2. \implies x$ is connected:
  - if a is a letter: $\mathcal{R}_x(a)$ generates $F_{\mathcal{A}} \implies \varepsilon$ is connected
  - ε connected in $D_w(x) \implies w$ connected in x
  - $2. \implies \mathcal{R}_{D_w(x)}(a)$ generates $F_{\mathcal{L}_1(D_w(x))}$ for any letter a of $D_w(x)$
- x connected $+ \#\mathcal{R}_x(w) = \#\mathcal{A} \implies x$ dendric:
  - x connected $\implies$ no weak factor
  - no weak factor $+ \#\mathcal{R}_x(w) = \#\mathcal{A} \implies x$ neutral

Proof that $2 \implies 1$

1. x is dendric;
2. for every $w \in \mathcal{L}(x)$, $\mathcal{R}_x(w)$ is a basis of $F_{\mathcal{A}}$;

- 2. $\implies$ x is connected:
  - if a is a letter: $\mathcal{R}_x(a)$ generates $F_{\mathcal{A}} \implies \varepsilon$ is connected
  - ε connected in $D_w(x) \implies w$ connected in x
  - 2. $\implies \mathcal{R}_{D_w(x)}(a)$ generates $F_{\mathcal{L}_1(D_w(x))}$ for any letter a of $D_w(x)$
- x connected $+ \#\mathcal{R}_x(w) = \#\mathcal{A} \implies x$ dendric:
  - x connected $\implies$ no weak factor
  - no weak factor $+ \#\mathcal{R}_x(w) = \#\mathcal{A} \implies x$ neutral
  - x connected $+$ x neutral $\implies x$ dendric

Preservation under derivation

Property	$x \implies D_w(x)$
$\#\mathcal{R}_x(u) = K$ for every u	✓

Preservation under derivation

Property	$x \implies D_w(x)$
$\#\mathcal{R}_x(u) = K$ for every long enough u	✓

Preservation under derivation

Property	$x \implies D_w(x)$
$\#\mathcal{R}_x(u) = K$ for every long enough u	✓
eventually neutral (resp., weak or neutral, strong or neutral)	✓

Preservation under derivation

Property	$x \implies D_w(x)$
$\#\mathcal{R}_x(u) = K$ for every long enough u	✓
eventually neutral (resp., weak or neutral, strong or neutral)	✓
eventually dendric (resp., acyclic, connected)	✓

Preservation under derivation

Property	$x \implies D_w(x)$
$\#\mathcal{R}_x(u) = K$ for every long enough u	✓
eventually neutral (resp., weak or neutral, strong or neutral)	✓
eventually dendric (resp., acyclic, connected)	✓
$\mathcal{R}_x(u)$ generates the free group over the alphabet for every u	X

Open questions

- Is there a formula for the number of return words for a set?

Open questions

- Is there a formula for the number of return words for a set?
- What can we say about freeness of return words?

Open questions

- Is there a formula for the number of return words for a set?
- What can we say about freeness of return words?
- Are there other link between return words and extensions?

Open questions

- Is there a formula for the number of return words for a set?
- What can we say about freeness of return words?
- Are there other link between return words and extensions?

Thank you for your attention!