Enumeration formulæ in neutral sets

Francesco Dolce and Dominique Perrin

UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Gaspard - Monge Laboratoire d'informatique

RDMath IdF Domaine d'Intérêt Majeur (DIM) en Mathématiques

îledeFrance

DLT 2015
Liverpool, 28^{th} July 2015

Overview

- Study of symbolic dynamical systems (essentially factors of infinite words) of linear complexity called “neutral”, containing the Sturmian dynamical systems.
- Proof of enumeration formulæ in these sets for bifix codes (and return words).
- Link with interval exchange transformations.

Outline

Outline

Let A a finite alphabet and S be a *factorial* set on A.

For a word $w \in S$, we denote

$$\begin{array}{lclccc} \ell(w) & = & \text{the number of letters} & a & \text{such that} & aw \in S, \\ r(w) & = & \text{the number of letters} & a & \text{such that} & wa \in S, \\ e(w) & = & \text{the number of pairs} & (a,b) & \text{such that} & awb \in S. \end{array}$$

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

Let A a finite alphabet and S be a *factorial* set on A.

For a word $w \in S$, we denote

$$\begin{array}{lclccc} \ell(w) & = & \text{the number of letters} & a & \text{such that} & aw \in S, \\ r(w) & = & \text{the number of letters} & a & \text{such that} & wa \in S, \\ e(w) & = & \text{the number of pairs} & (a,b) & \text{such that} & awb \in S. \end{array}$$

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

The *multiplicity* of a word w is the quantity

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

A word is called *neutral* if $m(w) = 0$.

A set S is *neutral* if it is factorial and every nonempty word $w \in S$ is neutral.

The integer $\chi(S) = 1 - m(\varepsilon) = \ell(\varepsilon) + r(\varepsilon) - e(\varepsilon)$ is called the *characteristic* of S.

A set S is *neutral* if it is factorial and every nonempty word $w \in S$ is neutral.

The integer $\chi(S) = 1 - m(\varepsilon) = \ell(\varepsilon) + r(\varepsilon) - e(\varepsilon)$ is called the *characteristic* of S.

Proposition

The following are neutral sets of characteristic 1 :

- *Sturmian sets* (sets of factors of an Arnoux-Rauzy word) and
- *Regular Interval Exchange sets* (see later).

Example

The *Fibonacci set* is the set of factors of the Fibonacci word, that is the fixed point $\varphi^\omega(a) = abaababaaba\cdots$ of the morphism

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

It is a neutral set of characteristic 1.
Indeed, $m(w) = 0$ for every w in the set (including the empty word).

The *factor complexity* of a factorial set $S \subset A^*$ is the sequence $p_n = \text{Card}\,(S \cap A^n)$.

Proposition (J. Cassaigne)

The factor complexity of a neutral set is given by $p_0 = 1$ and

$$p_n = n\big(\text{Card}\,(A) - \chi(S)\big) + \chi(S).$$

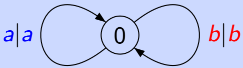
Example

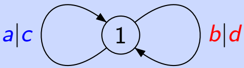
The Fibonacci set has factor complexity $p_n = n + 1$.

Example

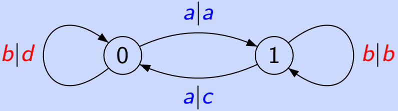
Let us consider two *doublings* of the Fibonacci set :

- the set of factors of the two infinite sequences $abaababa\cdots$ and $cdccdcdc\cdots$,





- the set of factors of the two infinite sequences $abcabcda\cdots$ and $cdacdabc\cdots$.



Both are neutral set of characteristic 2. Their factor complexity is $2n+2$.

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 $Y \subset S$.

Example

Let S be the Fibonacci set. The set $X = \{aa, ab, ba\}$ is an S-maximal bifix code. It is not an A^*-maximal bifix code, indeed $X \subset Y = X \cup \{bb\}$.

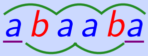
A *parse* of a word w with respect to a bifix code X is a triple (q, x, p) such that

- $w = qxp$,
- q has no suffix in X,
- $x \in X^*$ and
- p has no prefix in X.

Example

Let $X = \{aa, ab, ba\}$ and $w = abaaba$. The two possible parses of w are

- $(\varepsilon, ab\,aa\,ba, \varepsilon)$,
- $(a, ba\,ab, a)$.



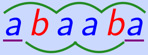
A *parse* of a word w with respect to a bifix code X is a triple (q, x, p) such that

- $w = qxp$,
- q has no suffix in X,
- $x \in X^*$ and
- p has no prefix in X.

Example

Let $X = \{aa, ab, ba\}$ and $w = abaaba$. The two possible parses of w are

- $(\varepsilon, ab\,aa\,ba, \varepsilon)$,
- $(a, ba\,ab, a)$.



The *S-degree* of X is the maximal number of parses with respect to X of a word of S.

Example

- For the Fibonacci set S, the set $X = \{aa, ab, ba\}$ has S-degree 2
- The set $X = S \cap A^n$ has S-degree n.

Theorem

Let S be a neutral set. For any finite S-maximal bifix code X of S-degree n, one has

$$\mathrm{Card}(X) = n\big(\mathrm{Card}(A) - \chi(S)\big) + \chi(S).$$

Theorem

Let S be a neutral set. For any finite S-maximal bifix code X of S-degree n, one has

$$\mathrm{Card}\,(X) = n\big(\mathrm{Card}\,(A) - \chi(S)\big) + \chi(S).$$

Example

The set S-maximal bifix code $X = \{aa, ab, ba\}$ of S-degree 2 verifies

$$\mathrm{Card}\,(X) = 2(2-1) + 1.$$

Let S be a factorial set and X be a finite S-maximal bifix code.
A *coding morphism* for X is a morphism $f : B^* \to A^*$ which maps bijectively an alphabet B onto X.

The set $f^{-1}(S)$ is called a *maximal bifix decoding* of S.

Theorem

Any maximal bifix decoding of a recurrent neutral set is a neutral set with the same characteristic.

Let S be a factorial set and X be a finite S-maximal bifix code.
A *coding morphism* for X is a morphism $f : B^* \to A^*$ which maps bijectively an alphabet B onto X.

The set $f^{-1}(S)$ is called a *maximal bifix decoding* of S.

Theorem

Any maximal bifix decoding of a recurrent neutral set is a neutral set with the same characteristic.

Example

Let us consider the Fibonacci set S, the S-maximal bifix code $X = \{aa, ab, ba\}$, the alphabet $B = \{u, v, w\}$, and the coding morphism

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

Both S and $f^{-1}(S)$ are neutral sets of characteristic 1.

Outline

Overview

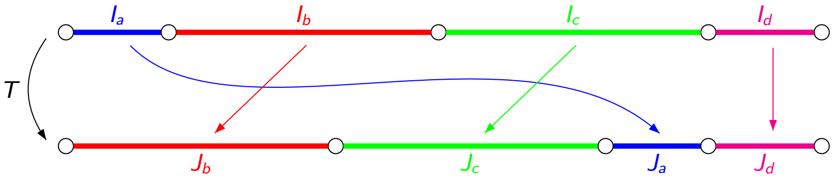
1. Neutral sets
2. Bifix codes in neutral sets
3. Interval exchange sets
   - Interval exchange transformations
   - Natural coding
   - Connections

Conclusions

Let $(I_a)_{a \in A}$ and $(J_a)_{a \in A}$ be two open partitions of the open set I (minus $\mathrm{Card}\,(A) - 1$ points), such that $|I_a| = |J_a|$ for every $a \in A$.

An *interval exchange transformation* is a map $T : I \to I$ defined by

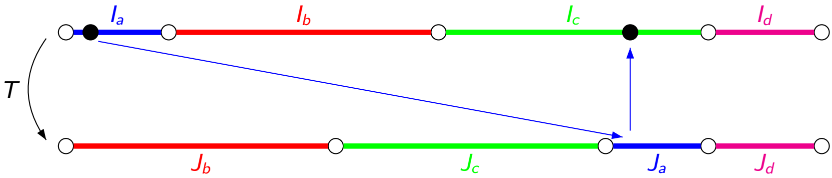
$$T(z) = z + \alpha_z \quad \text{if } z \in I_a.$$



Let $(I_a)_{a\in A}$ and $(J_a)_{a\in A}$ be two open partitions of the open set I (minus $\mathrm{Card}\,(A) - 1$ points), such that $|I_a| = |J_a|$ for every $a \in A$.

An *interval exchange transformation* is a map $T : I \to I$ defined by

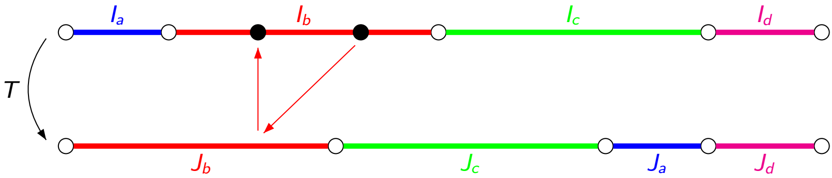
$$T(z) = z + \alpha_z \quad \text{if } z \in I_a.$$



Let $(I_a)_{a\in A}$ and $(J_a)_{a\in A}$ be two open partitions of the open set I (minus $\mathrm{Card}\,(A)-1$ points), such that $|I_a| = |J_a|$ for every $a \in A$.

An *interval exchange transformation* is a map $T : I \to I$ defined by

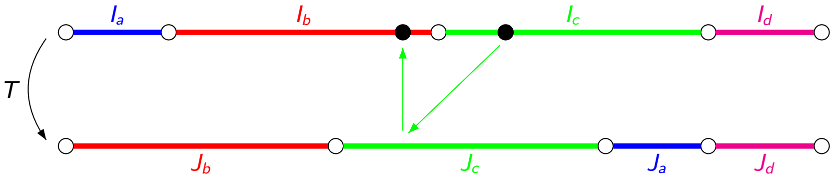
$$T(z) = z + \alpha_z \quad \text{if } z \in I_a.$$



Let $(I_a)_{a \in A}$ and $(J_a)_{a \in A}$ be two open partitions of the open set I (minus $\mathrm{Card}\,(A) - 1$ points), such that $|I_a| = |J_a|$ for every $a \in A$.

An *interval exchange transformation* is a map $T : I \to I$ defined by

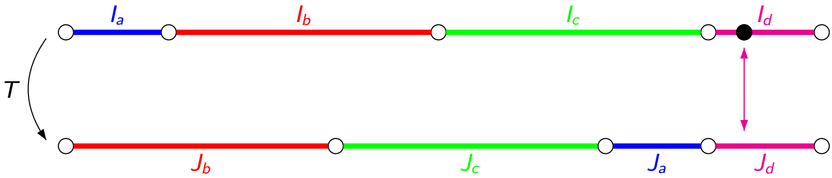
$$T(z) = z + \alpha_z \quad \text{if } z \in I_a.$$



Let $(I_a)_{a\in A}$ and $(J_a)_{a\in A}$ be two open partitions of the open set I (minus $\mathrm{Card}(A)-1$ points), such that $|I_a| = |J_a|$ for every $a \in A$.

An *interval exchange transformation* is a map $T : I \to I$ defined by

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

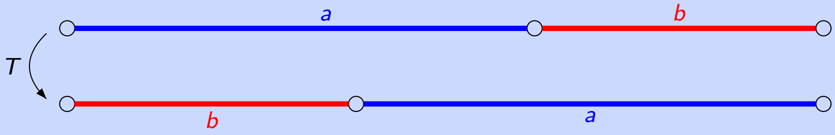


The *natural coding* of T relative to $z \in I$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

Example

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

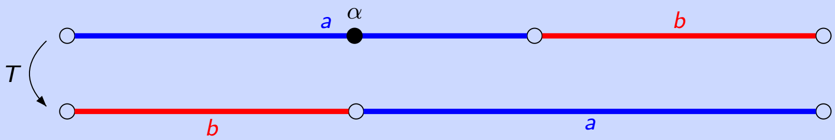


The *natural coding* of T relative to $z \in I$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

Example

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



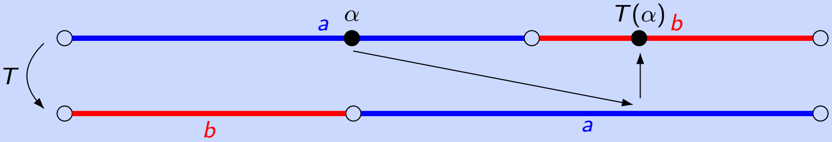
$\Sigma_T(\alpha) = a$

The *natural coding* of T relative to $z \in I$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

Example

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



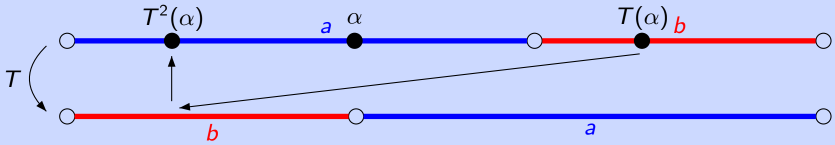
$\Sigma_T(\alpha) = a\, b$

The *natural coding* of T relative to $z \in I$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

Example

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



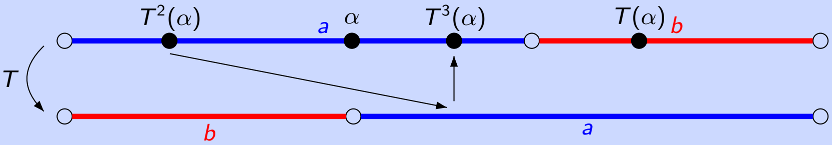
$$\Sigma_T(\alpha) = a\, b\, a$$

The *natural coding* of T relative to $z \in I$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

Example

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



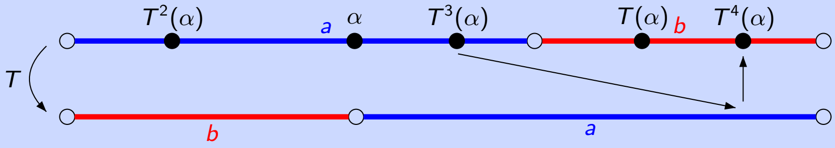
$$\Sigma_T(\alpha) = a\,b\,a\,a$$

The *natural coding* of T relative to $z \in I$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

Example

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



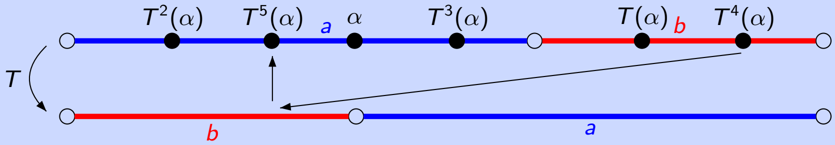
$$\Sigma_T(\alpha) = a\, b\, a\, a\, b$$

The *natural coding* of T relative to $z \in I$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

Example

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

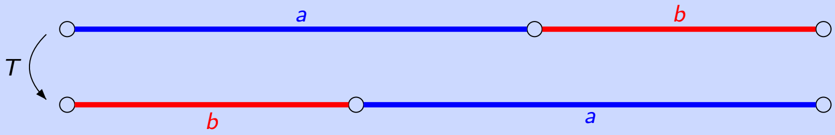


$$\Sigma_T(\alpha) = a\,b\,a\,a\,b\,a \cdots$$

The *interval exchange set* $\mathcal{L}(T)$ is the set of factors of all natural codings of T.

Example

The *Fibonacci set* is the set of factors of all natural codings of the rotation on the cirle (minus 2 points) by angle $\alpha = (3 - \sqrt{5})/2$.



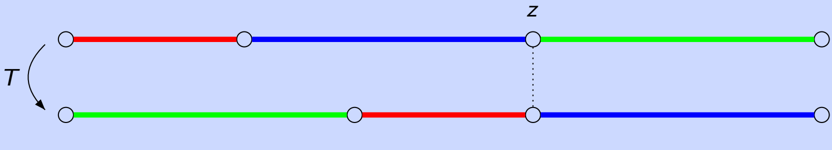
$$F(T) = \left\{\varepsilon, a, b, aa, ab, ba, aab, aba, baa, \ldots\right\}$$

A *connection* of *length* $n \geq 0$ of an interval exchange T is a triple (x, y, n) with

- x is a singularity of T^{-1},
- y is a singularity of T, and
- $T^n(x) = y$.

When $n = 0$, we say that $x = y$ is a connection.

Example



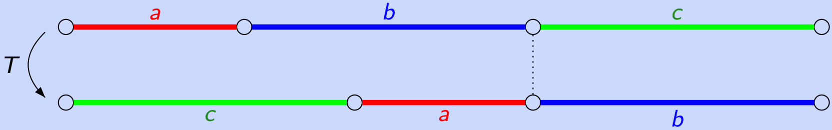
The point z is a connection of length 0.

An interval exchange without connections is said to be *regular*.

Theorem

Let T be an interval exchange with exactly c connections, all of length 0. Then, $\mathcal{L}(T)$ is a neutral set of characteristic $c+1$.

Example



$$\begin{aligned}\chi(\varepsilon) &= 1 - m(\varepsilon)\\ &= \ell(\varepsilon) + r(\varepsilon) - e(\varepsilon)\\ &= \mathrm{Card}\,(A) + \mathrm{Card}\,(A) - \mathrm{Card}\,(\{ab, bc, ca, cb\})\\ &= 2\end{aligned}$$

while $m(w) = 0$ for every $w \in A^+$.

Further research directions

- *Specular sets*, i.e. neutral sets of characteristic 2 satisfying additional "symmetric" properties.
- *Tree sets* of arbitrary characteristic, i.e. neutral sets with extra constraints of the extensions.
- Sets with a finite number of elements satisfying $m(w) \neq 0$.

THANKS
FOR YOUR
ATTENTION