Enumeration formulæ in neutral sets

Francesco Dolce

Réunion EQINOCS
Grenoble, 7 octobre 2015

Joint work with :

Valérie Berthé (Université Paris-Diderot)
Clelia De Felice (Universitá degli Studi di Salerno)
Vincent Delecroix (Université de Bordeaux)
Julien Leroy (Université de Liège)
Dominique Perrin and Giuseppina Rindone (Université Paris-Est)
Christophe Reutenauer (Université du Québec á Montréal)

Outline

1. Neutral sets
2. Interval exchange sets
3. Bifix codes in neutral sets

Outline

1. Neutral sets
   - Multiplicity and characteristic
   - Factor complexity
   - Tree sets
2. Interval exchange sets
3. Bifix codes in neutral sets

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

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

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 $p_n = \text{Card}\,(S \cap A^n)$.

Its *entropy* is defined as $\lim_{n \to \infty} \frac{1}{n} \log(p_n)$.

Proposition [J. Cassaigne (1997)]

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

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

Its entropy is then 0.

Example

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

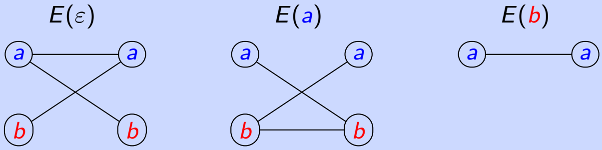
The *extension graph* of a word $w \in S$ is the undirected bipartite graph $G(w)$ with vertices the disjoint union of

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

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

Example

Here are the extensions graphs of the words of length at most 1 inside the Fibonacci set.



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

A biextendable set S is called a *tree set* of *characteristic* c if for any nonempty $w \in S$, the graph $E(w)$ is a tree (acyclic and connected) and if $E(\varepsilon)$ is a union of c trees.

Example

The Fibonacci set is a tree set of characteristic 1.

A tree set of characteristic c is clearly a neutral set of characteristic c.

A biextendable set S is called a *tree set* of *characteristic* c if for any nonempty $w \in S$, the graph $E(w)$ is a tree (acyclic and connected) and if $E(\varepsilon)$ is a union of c trees.

Example

The Fibonacci set is a tree set of characteristic 1.

A tree set of characteristic c is clearly a neutral set of characteristic c.

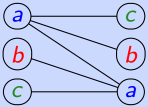
Proposition [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2014, *Monatsh. Math.*)]

A Sturmian set is a uniformly recurrent tree set of characteristic 1.

Example

The *Tribonacci* set is a tree set of characteristic 1.

$E(\varepsilon)$

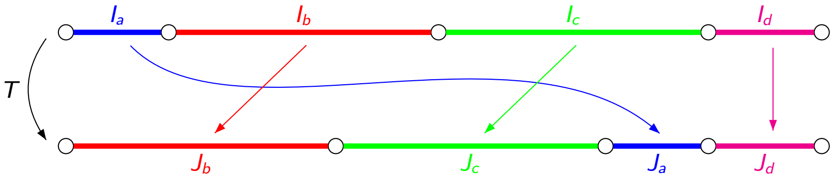


Outline

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

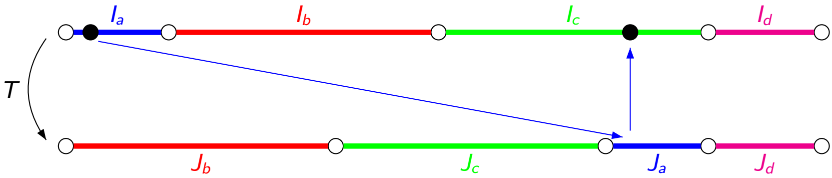
$$T(z) = z + \alpha_z \quad \text{if } z \in I_a.$$



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

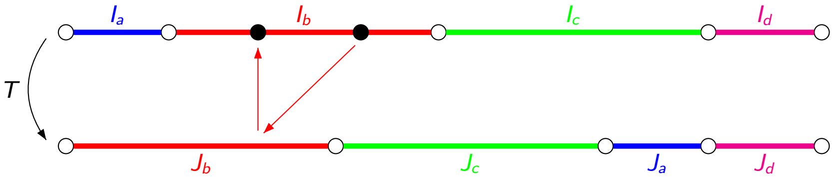
$$T(z) = z + \alpha_z \quad \text{if } z \in I_a.$$



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

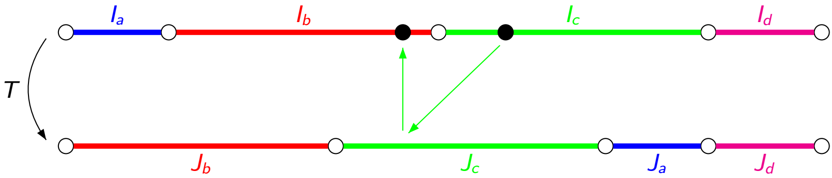
$$T(z) = z + \alpha_z \quad \text{if } z \in I_a.$$



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

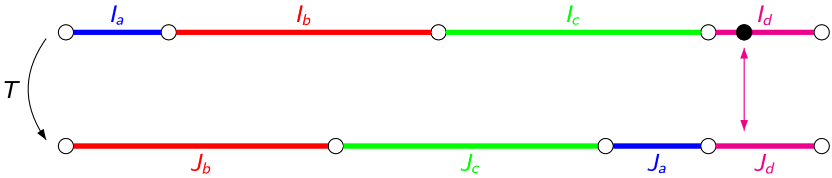
$$T(z) = z + \alpha_z \quad \text{if } z \in I_a.$$



Let $(I_a)_{a \in A}$ and $(J_a)_{a \in A}$ be two open partitions of the open set I (minus $\text{Card}(A) - 1$ 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

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

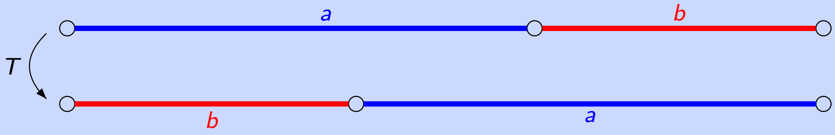


The *natural coding* of T relative to $z \in I$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

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. $T(z) = z + \alpha \bmod 1$.

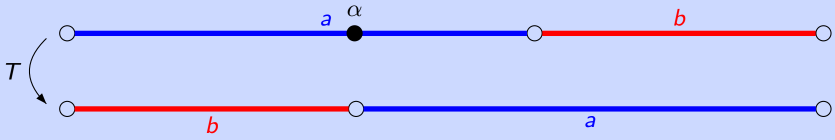


The *natural coding* of T relative to $z \in I$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

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. $T(z) = z + \alpha \bmod 1$.



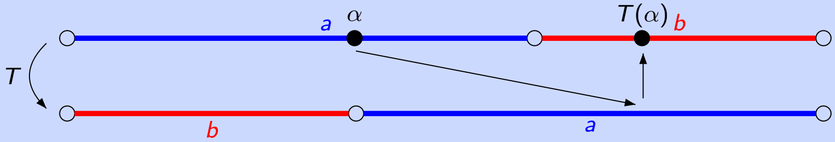
$\Sigma_T(\alpha) = a$

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. $T(z) = z + \alpha \mod 1$.



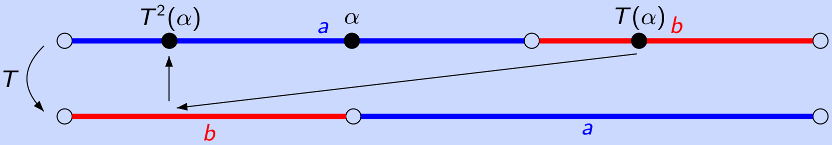
$\Sigma_T(\alpha) = a\, b$

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. $T(z) = z + \alpha \bmod 1$.



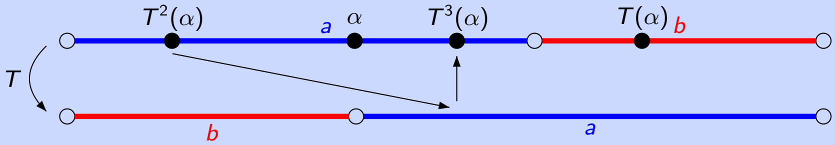
$\Sigma_T(\alpha) = a\,b\,a$

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. $T(z) = z + \alpha \bmod 1$.



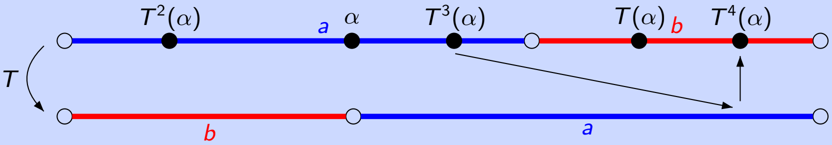
$$\Sigma_T(\alpha) = a\,b\,a\,a$$

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. $T(z) = z + \alpha \bmod 1$.



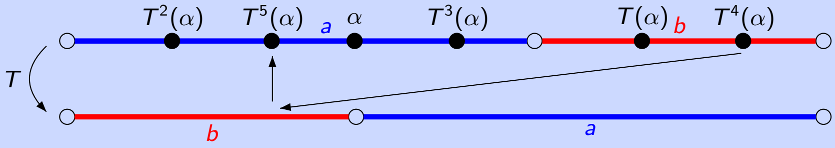
$\Sigma_T(\alpha) = a\,b\,a\,a\,b$

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. $T(z) = z + \alpha \bmod 1$.

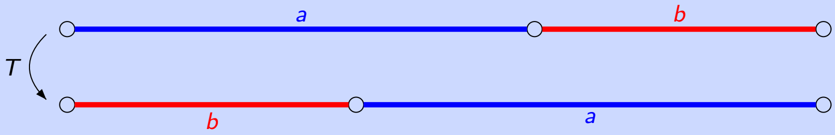


$$\Sigma_T(\alpha) = a\,b\,a\,a\,b\,a \cdots$$

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 $\alpha = (3 - \sqrt{5})/2$.



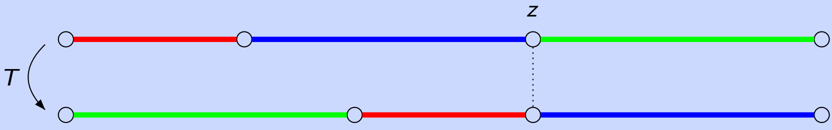
$$F(T) = \left\{\varepsilon, a, b, aa, ab, ba, aab, aba, baa, \ldots\right\}$$

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*.

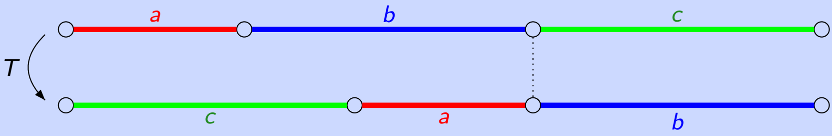
Proposition [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2014, *J.P.P.A.*)]

A regular interval exchange set is a tree set of characteristic 1.

Theorem [D., Perrin (2015, *DLT*)]

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

Example



The set $\mathcal{L}(T)$ is a tree set of characteristic 2.

Outline

1. Neutral sets
2. Interval exchange sets
3. Bifix codes in neutral sets
    - Bifix codes and S-degree
    - Cardinality Theorem for bifix codes
    - Bifix decoding

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 $X \subset Y = X \cup \{bb\}$.

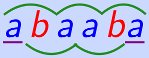
A *parse* of a word w with respect to a bifix code X is a triple (q, x, p) such that :

- $w = qxp$,
- 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

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



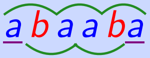
A *parse* of a word w with respect to a bifix code X is a triple (q, x, p) such that :

- $w = qxp$,
- 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 [D., Perrin (2015, *DLT*)]

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

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

$$\mathrm{Card}\,(X) = 2(2-1) + 1 = 3.$$

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

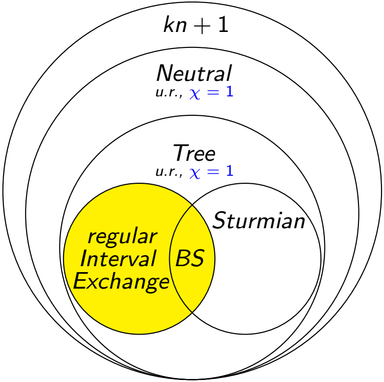
$$f : \begin{cases} u \mapsto aa \\ v \mapsto ab \\ w \mapsto ba \end{cases}$$

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.

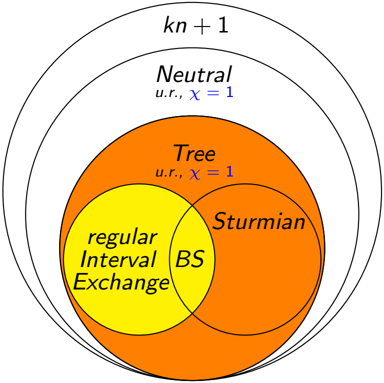
Theorem [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2014, *J.P.A.A.*)]

The family of regular interval exchange sets is closed by maximal bifix decoding (the cardinality of the alphabet might change).



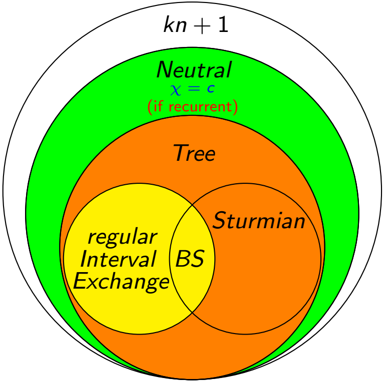
Theorem [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2015, *Discrete Math.*)]

The family of uniformly recurrent tree sets of characteristic 1 is closed by maximal bifix decoding.



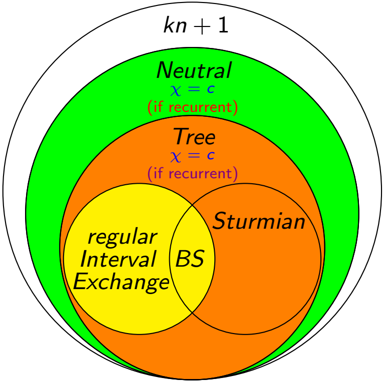
Theorem [D., Perrin (2015, *DLT*)]

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



Conjecture [D., Perrin]

Any maximal bifix decoding of a (uniformly) recurrent tree set is a tree set with the same characteristic.




Grant marci !