Maximal bifix decoding of a tree set

Francesco Dolce

Automatic Sequences
Liège, 25th May 2015

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

\[ x = abaababaabaababa \ldots \]

\[ x = \varphi^\omega(a) \]

\[ \varphi : \begin{cases} 
  a &\mapsto ab \\
  b &\mapsto a 
\end{cases} \]
Motivation

\[ x = abaababaababa \cdots \]

\begin{align*}
\begin{array}{cccccc}
 n & 0 & 1 & 2 & 3 & 4 & 5 & 6 & \cdots \\
(2-1)n+1 & 1 & 2 & 3 & 4 & 5 & 6 & \cdots \\
\end{array}
\end{align*}
Motivation

\[ x = \alpha \beta \alpha \beta \alpha \beta \cdot \cdot \cdot \]
\[ f(x) = v u w w v u w w \cdot \cdot \cdot \]

\[
\begin{align*}
\quad u & = \alpha \beta \\
\quad v & = \alpha \beta \\
\quad w & = \beta \alpha \\
\end{align*}
\]
Outline

Motivation
1. Two important classes
2. Acyclic, connected and tree sets
3. Maximal bifix decoding
Outline

Motivation

1. Two important classes
   - Sturmian sets
   - Interval Exchange sets
2. Acyclic, connected and tree sets
3. Bifix decoding
A *Sturmian* set is the set of factors of a *strict episturmian word* (i.e. of a word $x$ whose set of factors $F(x)$ is closed under reversal and for each $n$ contains exactly one right-special word $w_n$ of length $n$ with $w_nA \subset F(x)$).

**Example**

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

$$\psi : a \mapsto ab, \quad b \mapsto ac, \quad c \mapsto a.$$
Let \( A \) be a finite set ordered by \( <_1 \) and \( <_2 \).

An **interval exchange transformation** (IET) is a map \( T : [0, 1[ \to [0, 1[ \) defined by

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

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

**Theorem [Keane (1975)]**

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

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

**Theorem [Keane (1975)]**

A regular interval exchange transformation is minimal.

The converse is not true.
Let $T$ be an IET relative to $(l_a)_{a \in A}$.

The *natural coding* of $T$ relative to $z \in [0, 1]$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

**Example**

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

---

**Diagram:**

A diagram illustrating the action of the transformation $T$ on a line segment divided into intervals, with the transformation mapping points within each interval to the next one.

---

**Notes:**

- Two Important Classes
- Interval Exchange Sets

**References:**

Let $T$ be an IET relative to $(I_a)_{a \in A}$.

The **natural coding** of $T$ relative to $z \in [0,1]$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

**Example**

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

$$\Sigma_T(\alpha) = a$$
Let $T$ be an IET relative to $(I_a)_{a \in A}$.

The *natural coding* of $T$ relative to $z \in [0, 1]$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

**Example**

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

$$\Sigma_T(\alpha) = a b$$
Let $T$ be an IET relative to $(I_a)_{a \in A}$.

The *natural coding* of $T$ relative to $z \in [0, 1]$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

**Example**

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

$$\Sigma_T(\alpha) = a b a$$
Let $T$ be an IET relative to $(I_a)_{a \in A}$.

The *natural coding* of $T$ relative to $z \in [0, 1]$ is the infinite word $\Sigma_T(z) = a_0a_1\cdots \in A^\omega$ defined by

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

**Example**

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

$$\Sigma_T(\alpha) = a \ b \ a \ a$$
Let $T$ be an IET relative to $(I_a)_{a \in A}$.

The natural coding of $T$ relative to $z \in [0, 1]$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

**Example**

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

$$\Sigma_T(\alpha) = a \ b \ a \ a \ b$$
Let $T$ be an IET relative to $(I_a)_{a \in A}$.

The **natural coding** of $T$ relative to $z \in [0, 1]$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots \in A^\omega$ defined by

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

**Example**

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

\[
\Sigma_T(\alpha) = a \ b \ a \ a \ b \ a \ \cdots
\]
The set $F(T) = \bigcup_{z \in [0,1]} (\Sigma_T(z))$ is said a (minimal, regular) interval exchange set.

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

Example

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

$$F(T) = \{\varepsilon, a, b, aa, ab, ba, aab, aba, baa, \ldots\}$$
Sturmian sets and regular interval exchange sets have both complexity function $p(n) = kn + 1$, with $k = \text{Card}(A) - 1$. 
Sturmian sets and regular interval exchange sets have both complexity function $p(n) = kn + 1$, with $k = \text{Card}(A) - 1$.

They are factorial and \textit{uniformly recurrent} (right-extendable and s.t. for any element $u \in S$ there exists an $n = n(u)$ with $u$ a factor of all words of $S \cap A^n$).
Sturmian sets and regular interval exchange sets have both complexity function $p(n) = kn + 1$, with $k = \text{Card}(A) - 1$.

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

However, the two families are distinct for $k \geq 2$. 
Sturmian sets and regular interval exchange sets have both complexity function $p(n) = kn + 1$, with $k = \text{Card}(A) - 1$.

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

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

Do they have other properties in common?
Outline

Motivation

1. Two important classes
2. Acyclic, connected and tree sets
   - Tree sets
   - Planar tree sets
3. Bifix decoding
Let $S$ be a factorial set over an alphabet $A$.

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

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

and edges the pairs in

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

**Example**

Let $S$ be the Fibonacci set.

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

![Diagram of extension graphs for \(\varepsilon, a, b\)](image-url)
A set $S$ is *acyclic* (resp. *connected*) if it is biextendable and if for every word $w \in S$, the graph $G(w)$ is acyclic (resp. connected).

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

**Example**

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

1. of characteristic $\chi(S) = 1$. 

Francesco Dolce  (Paris-Est)  
Maximal Bifix Decoding of a Tree Set  
Liège, 25th May 2015  
15 / 26
Proposition [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2014, 2015)]
Both Sturmian sets and regular interval exchange sets are uniformly recurrent tree sets.
Let $<_1$ and $<_2$ be two orders on $A$.
For a set $S$ and a word $w \in S$, the graph $G(w)$ is *compatible* with $<_1$ and $<_2$ if for any $(a, b), (c, d) \in E(w)$, one has

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

**Example**

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

![Graphs](image)

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

---

*Francesco Dolce (Paris-Est)  Maximal Bifix Decoding of a Tree Set  Liège, 25\textsuperscript{th} May 2015  17 / 26*
Example

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

$$G(\varepsilon)$$

$$G(a)$$

$$G(aba)$$

It is not possible to find two orders on $A$ making the three graphs planar.
Theorem [Ferenczi, Zamboni (2008)]

A set $S$ is a regular interval exchange set on $A$ if and only if it is a uniformly recurrent planar tree set containing $A$. 

$$kn + 1 \quad \text{uniformly recurrent}$$

Tree

uniformly recurrent

Planar Tree

regular interval exchange

Sturmian

BS

Francesco Dolce (Paris-Est) 
Maximal Bifix Decoding of a Tree Set 
Liège, 25th May 2015 
19 / 26
Outline

Motivation

1. Two important classes
2. Acyclic, connected and tree sets
3. Bifix decoding
   - Bifix codes
   - Maximal 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 set $X \subset A^+$ of nonempty words over an alphabet $A$ is a \textit{bifix code} if it does not contain any proper prefix or suffix of its elements.

\begin{itemize}
\item $\{aa, ab, ba\}$
\item $\{aa, ab, bba, bbb\}$
\item $\{ac, bcc, bcba\}$
\end{itemize}

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

\begin{itemize}
\item Let $S$ be the Fibonacci set. The set $X = \{aa, ab, ba\}$ is an \textit{$S$-maximal} bifix code. It is not an \textit{$A^*$-maximal} bifix code, indeed $X \subset Y = X \cup \{bb\}$. 
\end{itemize}
A **coding morphism** for a bifix code $X \subset A^+$ is a morphism $f : B^* \rightarrow A^*$ which maps bijectively $B$ onto $X$.

**Example**

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

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

is a coding morphism for $X$.

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

Let $S$ be the Fibonacci set.
Let us consider the $S$-maximal bifix code $X = \{aa, ab, ba\}$ and the coding morphism $f : u \mapsto aa, \ v \mapsto ab, \ w \mapsto ba$.

$f^{-1}(S)$ is not a Sturmian set. But it is a regular interval exchange sets (as $S$).
Example

Let $T$ be the Tribonacci set. Let us consider the $T$-maximal bifix code $X = \{aa, ab, ac, ba, ca\}$ and the coding morphism

$$g : u \mapsto aa, \quad v \mapsto ab, \quad w \mapsto ac, \quad t \mapsto ba, \quad z \mapsto ca.$$ 

$g^{-1}(T)$ is not a Sturmian set. But it is a tree set (as $T$).
Theorem [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2014, 2015)]

The family of uniformly recurrent tree sets is closed under maximal bifix decoding (and so is the family of u.r. planar tree sets).
Theorem [Berthé, De Felice, D., Leroy, Perrin, Reutenauer, Rindone (2014, 2015)]

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