Factor complexity of infinite words associated with $\beta$-expansions

Karel Klouda$^1,2$

karel@kloudak.eu

Joint work with E.Pelantová$^1$

$^1$FNSPE, CTU in Prague

$^2$LIAFA, Université Paris 7

Mons Theoretical Computer Science Days
29$^\text{th}$ August, 2008
Rényi expansion of unity in base $\beta > 1$

\[ d_\beta(1) = t_1 t_2 t_3 \cdots, \quad t_i = \lfloor \beta T^{-1}_\beta(1) \rfloor, \]

where

\[ T_\beta : [0, 1] \rightarrow (0, 1), \quad T_\beta(x) := \beta x - \lfloor \beta x \rfloor = \{ \beta x \}. \]

- Parry number: $d_\beta(1)$ is eventually periodic,
- simple Parry number: $d_\beta(1) = t_1 \cdots t_m$,
- non-simple Parry number: $d_\beta(1) = t_1 \cdots t_m t_{m+1} t_{m+2} \cdots t_{m+p} \omega$. 

Simple Parry numbers

d_β(1) = t_1 \cdots t_m

Canonical substitution \( \varphi_\beta \) over the alphabet \( \mathcal{A} = \{0, 1, \ldots, m - 1\} \)

\[
\begin{align*}
\varphi_\beta(0) & = 0^{t_1}1 \\
\varphi_\beta(1) & = 0^{t_2}2 \\
& \vdots \\
\varphi_\beta(m-2) & = 0^{t_{m-1}}(m-1) \\
\varphi_\beta(m-1) & = 0^{t_m}
\end{align*}
\]
Non-simple Parry numbers

\[ d_\beta(1) = t_1 \cdots t_m(t_{m+1}t_{m+2} \cdots t_{m+p})^\omega \]

Canonical substitution \( \varphi_\beta \) over the alphabet \( A = \{0, 1, \ldots, m + p - 1\} \)

\[
\begin{align*}
\varphi_\beta(0) &= 0^{t_1}1 \\
\varphi_\beta(1) &= 0^{t_2}2 \\
& \vdots \\
\varphi_\beta(m-1) &= 0^{t_m}m \\
\varphi_\beta(m) &= 0^{t_{m+1}}(m+1) \\
& \vdots \\
\varphi_\beta(m+p-2) &= 0^{t_{m+p-1}}(m+p-1) \\
\varphi_\beta(m+p-1) &= 0^{t_{m+p}}m
\end{align*}
\]

Fixed point \( u_\beta = \lim_{n \to \infty} \varphi_\beta^n(0) = 0^{t_1}1 \cdots \).
Basic definitions – factor complexity

\[ \mathcal{A} = \{0, 1, \ldots, q - 1\} \] an alphabet

\[ u = (u_i)_{i \in \mathbb{N}}, u_i \in \mathcal{A} \] an infinite word over \( \mathcal{A} \)

\[ w = u_j u_{j+1} \cdots u_{j+n-1} \] a factor of \( u \) of length \( n \)

\[ \mathcal{L}_n(u) \] the set of factors of \( u \) of length \( n \)

\[ \mathcal{L}(u) = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n(u) \] the language of \( u \)

The factor complexity of \( u \) is the function \( \mathcal{C} : \mathbb{N} \rightarrow \mathbb{N} \) given by

\[ \mathcal{C}(n) := \#\mathcal{L}_n(u). \]
Basic definitions – fixed point of substitution

If \( \varphi(0) = 0v \), \( v \in \mathcal{A}^+ \), then the fixed point of \( \varphi \) given by
\[
u := \lim_{n \to \infty} \varphi^n(0) = \varphi^\infty(0)
\]
is an infinite word which is uniformly recurrent.

A substitution \( \varphi \) is primitive if there exists \( k \in \mathbb{N} \) such that for all \( a, b \in \mathcal{A} \) the word \( \varphi^k(a) \) contains \( b \). In what follows, we assume that \( \varphi \) is primitive and injective.

In general, the factor complexity of a fixed point of any primitive substitution is a sublinear function \( C(n) \leq an + b \), \( a, b \in \mathbb{N} \).
Known results for simple Parry numbers

**Simple Parry numbers** (Bernat, Frougny, Masáková, Pelantová):

- $t_1 = t_2 = \cdots = t_{m-1}$ or $t_1 > \max\{t_2, \ldots, t_{m-1}\}$ the exact value of $C(n)$ is known,
- in particular, $(m - 1)n + 1 \leq C(n) \leq mn$, for all $n \geq 1$,
- $C(n)$ is affine $\iff$
  1) $t_m = 1$
  2) for all $i = 2, 3, \ldots, m-1$ we have

$$t_i t_{i+1} \cdots t_{m-1} t_1 \cdots t_{i-1} \preceq t_1 t_2 \cdots t_{m-1}.$$ 

Then $C(n) = (m - 1)n + 1$. 
Special factors

For $v \in \mathcal{L}(u)$ we define the set of *left extensions*

$$\text{Lext}(v) := \{ a \in A \mid av \in \mathcal{L}(u) \}.$$ 

If $\#\text{Lext}(v) > 1$, then $v$ is said to be *left special (LS) factor*. Analogously are defined right special (RS) factors.
LS factors and factor complexity

\[ \mathcal{L}_{n+1}(u) \quad \mathcal{L}_n(u) \]

\[
\begin{align*}
av_1 & \quad a \\
bv_1 & \quad b \\
bv_2 & \quad b \\
av_3 & \quad a \\
bv_3 & \quad a \\
cv_3 & \quad c \\
av_4 & \quad c \\
av_5 & \quad c
\end{align*}
\]

\[ C_u(n+1) - C_u(n) = 3 \]
LS factors and factor complexity

\[ \mathcal{L}_{n+1}(u) \]

\[ \mathcal{L}_n(u) \]

\[ a v_1 \]

\[ b v_1 \]

\[ b v_2 \]

\[ a v_3 \]

\[ b v_3 \]

\[ c v_3 \]

\[ a v_4 \]

\[ a v_5 \]

\[ C(n + 1) - C(n) = 3 \]
LS factors and factor complexity

For the first difference of the complexity function it holds that

\[ \Delta C(n) := C(n + 1) - C(n) = \sum_{\substack{v \in L_n(u) \\ n \text{ is LS}}} (\#\text{Lext}(v) - 1). \]

Complete knowledge of all LS factors along with the number of their left extensions allow us to evaluate \( C(n) \).

\[ \Delta C(n) \geq 1 \text{ for all } n \in \mathbb{N} \Leftrightarrow u \text{ is aperiodic.} \]
Structure of LS factors – infinite LS branches

Definition

An infinite word \( w \) is said to be an infinite LS branch of \( u \) if each prefix of \( w \) is a LS factor of \( u \).

\[
Lext(w) = \bigcap_{v \text{ prefix } w} Lext(v).
\]
Structure of LS factors – infinite LS branches

Definition
An infinite word $w$ is said to be an infinite LS branch of $u$ if each prefix of $w$ is a LS factor of $u$.

$Lext(w) = \bigcap_{v \text{ prefix } w} Lext(v)$.

- $u$ periodic $\Rightarrow$ no infinite LS branches,
Structure of LS factors – infinite LS branches

Definition
An infinite word $w$ is said to be an infinite LS branch of $u$ if each prefix of $w$ is a LS factor of $u$.

$$Lext(w) = \bigcap_{v \text{ prefix } w} Lext(v).$$

- $u$ periodic $\Rightarrow$ no infinite LS branches,
- $u$ aperiodic $\Rightarrow$ at least one infinite LS branch,
Structure of LS factors – infinite LS branches

Definition
An infinite word \( w \) is said to be an infinite LS branch of \( u \) if each prefix of \( w \) is a LS factor of \( u \).

\[
\text{Lext}(w) = \bigcap_{v \text{ prefix } w} \text{Lext}(v).
\]

- \( u \) periodic \( \Rightarrow \) no infinite LS branches,
- \( u \) aperiodic \( \Rightarrow \) at least one infinite LS branch,
- \( u \) is a fixed point of a primitive substitution \( \Rightarrow \) finite number of infinite LS branches
  (a consequence of the fact that \( \triangle C(n) \) is bounded (Mossé, Cassaigne))
Structure of LS factors – maximal LS factors

Definition
A LS factor $v$ having left extensions $a, b \in \text{Lext}(v)$ is called an $(a, b)$-maximal LS factor if for each letter $e \in \mathcal{A}$ we is not a LS factor with the left extensions $a$ and $b$. 
Images of LS factors

Example: \( \varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534 \)

\( u = \varphi^\infty(1) \)

\( w \) is a LS factor or an infinite LS branch of \( u \) with left extensions 1, 2 and 3:

\[
\begin{align*}
&1 \quad 2 \\
&\quad \quad \quad w
\end{align*}
\]
Images of LS factors

Example: \( \varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534 \)

\( u = \varphi^\infty(1) \)

\( w \) is a LS factor or an infinite LS branch of \( u \) with left extensions 1, 2 and 3:
Images of LS factors

Example: $\varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534$

$u = \varphi^\infty(1)$

$w$ is a LS factor or an infinite LS branch of $u$ with left extensions 1, 2 and 3:

\[
\begin{array}{c}
1 \\
2 \\
\end{array} \xrightarrow{\varphi\text{-image}}
\begin{array}{c}
1211 \\
311 \\
\end{array} \xrightarrow{f\text{-image}}
\begin{array}{c}
2 \\
3 \\
\end{array} \xrightarrow{11\varphi(w)}
\end{array}
\]

$\varphi(w) \\
\varphi(w)

\begin{align*}
f_L(1, 2) &= 11 \\
g_L(1, 2) &= \{2, 3\}
\end{align*}
Images of LS factors

Example: $\varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534$

$u = \varphi^\infty(1)$

$w$ is a LS factor or an infinite LS branch of $u$ with left extensions 1, 2 and 3:

$f_L(1, 2) = 11$

$g_L(1, 2) = \{2, 3\}$

$f_L(2, 3) = \epsilon$

$g_L(2, 3) = \{1, 2\}$
Images of LS factors

Example: \( \varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534 \)

\( u = \varphi^\infty(1) \)

\( w \) is a LS factor or an infinite LS branch of \( u \) with left extensions 1, 2 and 3:

\[
\begin{align*}
1 \xrightarrow{\varphi\text{-image}} 1211 \xrightarrow{\varphi(w)} 2 \xrightarrow{f\text{-image}} 11 \varphi(w) \\
3 \xrightarrow{\varphi\text{-image}} 311 \xrightarrow{\varphi(w)} 2 \xrightarrow{f\text{-image}} 11 \varphi(w)
\end{align*}
\]

\( f_L(1, 2) = 11 \)

\( g_L(1, 2) = \{2, 3\} \)

\[
\begin{align*}
2 \xrightarrow{\varphi\text{-image}} 311 \xrightarrow{\varphi(w)} 1 \xrightarrow{f\text{-image}} \varphi(w) \\
3 \xrightarrow{\varphi\text{-image}} 2412 \xrightarrow{\varphi(w)} 1 \xrightarrow{f\text{-image}} \varphi(w)
\end{align*}
\]

\( f_L(2, 3) = \epsilon \)

\( g_L(2, 3) = \{1, 2\} \)
Graph $GL_\varphi$

Vertices: unordered couples of distinct letters $(a, b)$.

Edges: if $g_L(a, b) = (c, d)$, then there is an edge between $(a, b)$ and $(c, d)$ with label $f_L(a, b)$. 

Example: $\varphi: 1 \mapsto \rightarrow 1211, 2 \mapsto \rightarrow 311, 3 \mapsto \rightarrow 2412, 4 \mapsto \rightarrow 435, 5 \mapsto \rightarrow 534$.
Graph $GL_\varphi$

Vertices: unordered couples of distinct letters $(a, b)$.

Edges: if $g_L(a, b) = (c, d)$, then there is an edge between $(a, b)$ and $(c, d)$ with label $f_L(a, b)$.

Example: $\varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534$
Structure of infinite branches

**Assumption:** For each infinite LS branch \( w \) it holds that

\( a) \) \( f \)-image of \( w \) is uniquely given,

\( b) \) there exists exactly one infinite LS branch \( w' \) such that \( w \) is \( f \)-image of \( w' \).

\[ f_L(a, b)\varphi(w) \]

\[ f \text{-image} \]

\[ f \text{-image} \]

\[ f \text{-image} \]
Structure of infinite branches

**Assumption:** For each infinite LS branch $w$ it holds that

a) $f$-image of $w$ is uniquely given,

b) there exists exactly one infinite LS branch $w'$ such that $w$ is $f$-image of $w'$.

**Theorem**

Let $w$ be an infinite LS branch, $a, b \in \text{Lex}(w)$. Then there exists $l > 0$ such that

$$w = f_L(g_L^{-1}(a, b)) \cdots \varphi^{l-2}(f_L(g_L(a, b)))\varphi^{l-1}(f_L(a, b))\varphi^l(w).$$
Structure of infinite branches

**Assumption:** For each infinite LS branch $w$ it holds that

- $a)$ $f$-image of $w$ is uniquely given,
- $b)$ there exists exactly one infinite LS branch $w'$ such that $w$ is $f$-image of $w'$.

**Theorem**

*Let $w$ be an infinite LS branch, $a, b \in \text{Lext}(w)$. Then there exists $l > 0$ such that*

$$w = f_L(g_L^{-1}(a, b)) \cdots \varphi^{l-2}(f_L(g_L(a, b)))\varphi^{l-1}(f_L(a, b))\varphi^l(w).$$

- $f_L = \epsilon \Rightarrow w = \varphi^l(w)$ and $(a, b)$ is a vertex of a cycle in $GL\varphi$ labelled by $\epsilon$ only,
Structure of infinite branches

**Assumption:** For each infinite LS branch $w$ it holds that

- $f$-image of $w$ is uniquely given,
- there exists exactly one infinite LS branch $w'$ such that $w$ is $f$-image of $w'$.

**Theorem**

Let $w$ be an infinite LS branch, $a, b \in \text{Lext}(w)$. Then there exists $l > 0$ such that

$$w = f_L(g_L^{-1}(a, b)) \cdots \varphi^{l-2}(f_L(g_L(a, b)))\varphi^{l-1}(f_L(a, b))\varphi^l(w).$$

- $f_L = \epsilon \Rightarrow w = \varphi^l(w)$ and $(a, b)$ is a vertex of a cycle in $GL\varphi$ labelled by $\epsilon$ only,
- otherwise, $(a, b)$ is a vertex of a cycle in $GL\varphi$ labelled not only by $\epsilon$. 
Example – how to identify infinite LS branch

\[ \varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534 \]

\[ \text{Lext}(1) = \{1, 2, 3, 4, 5\}, \text{Lext}(2) = \{1, 4, 5\}, \text{Lext}(3) = \{1, 4, 5\}, \text{Lext}(4) = \{1, 2, 3\}, \text{Lext}(5) = \{1, 2, 3\} \]
Example – how to identify infinite LS branchs

\( \varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534 \)

Lext(1) = \{1, 2, 3, 4, 5\}, Lext(2) = \{1, 4, 5\}, Lext(3) = \{1, 4, 5\}, Lext(4) = \{1, 2, 3\}, Lext(5) = \{1, 2, 3\}

• \( w = 11 \varphi^2(w) \rightarrow 11 \varphi^2(11) \varphi^4(11) \ldots \)
Example – how to identify infinite LS branch
e
φ : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534

\text{Lext}(1) = \{1, 2, 3, 4, 5\}, \text{Lext}(2) = \{1, 4, 5\}, \text{Lext}(3) = \{1, 4, 5\}, \text{Lext}(4) = \{1, 2, 3\}, \text{Lext}(5) = \{1, 2, 3\}

- \varphi(11)\varphi^3(11)\cdots, \ 11\varphi^2(11)\varphi^4(11)\cdots
Example – how to identify infinite LS branch

\( \varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534 \)

\( \text{Lext}(1) = \{1, 2, 3, 4, 5\}, \text{Lext}(2) = \{1, 4, 5\}, \text{Lext}(3) = \{1, 4, 5\}, \text{Lext}(4) = \{1, 2, 3\}, \text{Lext}(5) = \{1, 2, 3\} \)

- \( \varphi^{(11)} \varphi^{3(11)} \cdots \)
- \( 11 \varphi^{2(11)} \varphi^{4(11)} \cdots \)
- \( \varphi^{\omega}(1), \varphi^{\omega}(4), \varphi^{\omega}(5), (\varphi^2)^{\omega}(2), (\varphi^2)^{\omega}(3) \)
Example – how to identify infinite LS branch

\[ \varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534 \]

\[ \text{Lext}(1) = \{1, 2, 3, 4, 5\}, \text{Lext}(2) = \{1, 4, 5\}, \text{Lext}(3) = \{1, 4, 5\}, \text{Lext}(4) = \{1, 2, 3\}, \text{Lext}(5) = \{1, 2, 3\} \]

- \( \varphi(11)\varphi^3(11) \ldots, 11\varphi^2(11)\varphi^4(11) \ldots \)
- \( \varphi^\omega(1), (\varphi^2)^\omega(2), (\varphi^2)^\omega(3) \)
GL_{\varphi_\beta} for simple Parry numbers

\[ f_L(a, b) = \epsilon \text{ for all } a, b \in \{0, 1, \ldots, m - 1\} \text{ and } u_\beta = \varphi_\beta^\omega(0) \text{ is the only fixed point} \]

\[ k = 1, \ldots, m - 1 \]

⇒ \( u_\beta \) is the only infinite LS branch
$GL_{\varphi_\beta}$ for non-simple Parry numbers

$m - 1 \mapsto 0^t m, \ m + p - 1 \mapsto 0^{t+p} m, f_L(m - 1, m + p - 1) = 0^t m, t = \min\{t_m, t_{m+p}\}, \text{Lext}(0^t m) = \{0, z\}, s \geq 1$

$k \neq sp$

\[
\begin{array}{c}
0_k \\
\epsilon \\
\downarrow \\
1_k \oplus 1 \\
\epsilon \\
\downarrow \\
m_k \oplus m \\
\epsilon \\
\downarrow \\
m + p - 1_k \oplus m + p - 1 \\
\end{array}
\]

\[
\begin{array}{c}
0_{sp} \\
\epsilon \\
\downarrow \\
1_{sp} \oplus 1 \\
\epsilon \\
\downarrow \\
m - 1_{sp} \oplus m - 1 \\
\end{array}
\]

$z = sp$

$0^t m$

$\text{yes}$

$\text{no}$
Infinite LS factors

\[ t = \min\{t_m, t_{m+p}\}, \text{Lext}(0^t m) = \{0, z\}, \ s \geq 1 \]

**Definition**

\[ \beta \in S \iff z = sp \iff \]

a) \[ d_\beta(1) = t_1 \cdots t_m(0 \cdots 0t_{m+p})^\omega \quad \text{and} \quad t_m > t_{m+p} \]

b) \[ d_\beta(1) = t_1 \cdots t_{m-qp}0 \cdots 0t_m(t_m + 1 \cdots t_{m+p})^\omega, \quad q \geq 1, \ t_m < t_{m+p}, \]
\[ \neq 0 \quad q \neq 1 \]

\[ \beta \in S_0 \iff d_\beta(1) = t_1(0 \cdots 0(t_1 - 1))^\omega. \]
Infinite LS factors

Theorem

• If \( \beta \) is a non-simple Parry and \( p > 1 \), then \( u_\beta \) is an infinite LS branch with left extensions \( \{m, m + 1, \ldots, m + p - 1\} \).
Theorem

- If $\beta$ is a non-simple Parry and $p > 1$, then $u_\beta$ is an infinite LS branch with left extensions $\{m, m + 1, \ldots, m + p - 1\}$.
- If $\beta \notin S$, then $u_\beta$ is the only one infinite LS branch.
Infinite LS factors

Theorem

- If $\beta$ is a non-simple Parry and $p > 1$, then $u_\beta$ is an infinite LS branch with left extensions $\{m, m + 1, \ldots, m + p - 1\}$.
- If $\beta \notin S$, then $u_\beta$ is the only one infinite LS branch.
- If $\beta \in S$, then there are $m$ infinite LS branches

\[0^t m \varphi^m(0^t m) \varphi^{2m}(0^t m) \cdots \]

\[\vdots\]

\[\varphi^{m-1}(0^t m) \varphi^{2m-1}(0^t m) \varphi^{3m-1}(0^t m) \cdots .\]
Affine complexity

Theorem

- The factor complexity of $u_\beta$ is affine $\iff u_\beta$ does not contain any $(a, b)$-maximal factor $\iff \beta \in S_0 \iff d_\beta(1) = t_1(0 \cdots 0(t_1 - 1))^\omega$. Then $C(n) = (m + p - 1)n + 1$.

- The first equivalence is not valid in general (Chacon),
- $\beta \in S_0 \Rightarrow \beta$ is an unitary Pisot number (Frougny).
Affine complexity

Theorem

• The factor complexity of $u_\beta$ is affine $\iff u_\beta$ does not contain any $(a, b)$-maximal factor $\iff \beta \in S_0 \iff d_\beta(1) = t_1(0 \cdots 0(t_1 - 1))^\omega$. Then $C(n) = (m + p - 1)n + 1$.

• The first equivalence is not valid in general (Chacon),
• $\beta \in S_0 \Rightarrow \beta$ is an unitary Pisot number (Frougny).

• If $p > 1$ and $\beta \in S_0$, then $u_\beta$ and $0^{-1}u_\beta$ are the only infinite LS branches.
Affine complexity

Theorem

• The factor complexity of $u_\beta$ is affine $\iff$ $u_\beta$ does not contain any $(a, b)$-maximal factor $\iff$ $\beta \in S_0 \iff d_\beta(1) = t_1(0 \cdots 0(t_1 - 1))^\omega$. Then $C(n) = (m + p - 1)n + 1$.

• The first equivalence is not valid in general (Chacon),

• $\beta \in S_0 \Rightarrow \beta$ is an unitary Pisot number (Frougny).

• If $p > 1$ and $\beta \in S_0$, then $u_\beta$ and $0^{-1}u_\beta$ are the only infinite LS branches.

• known result: $u_\beta$ is Sturmian $\iff$ $p = 1$ and $\beta \in S_0$, i.e. $d_\beta(1) = t_1(t_1 - 1)^\omega$. 
Affine complexity

Theorem

- The factor complexity of $u_\beta$ is affine $\iff$ $u_\beta$ does not contain any $(a, b)$-maximal factor $\iff$ $\beta \in S_0 \iff d_\beta(1) = t_1(0 \cdots 0(t_1 - 1))\omega$. Then $C(n) = (m + p - 1)n + 1$.
  
  - The first equivalence is not valid in general (Chacon),
  - $\beta \in S_0 \Rightarrow \beta$ is an unitary Pisot number (Frougny).
  - If $p > 1$ and $\beta \in S_0$, then $u_\beta$ and $0^{-1}u_\beta$ are the only infinite LS branches.
  - known result: $u_\beta$ is Sturmian $\iff$ $p = 1$ and $\beta \in S_0$, i.e. $d_\beta(1) = t_1(t_1 - 1)\omega$.

THE END