Factor complexity of infinite words associated with β-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
29th August, 2008

Rényi expansion of unity in base $\beta > 1$

$$d_\beta(1) = t_1 t_2 t_3 \cdots , \qquad t_i = \lfloor \beta T_\beta^{i-1}(1) \rfloor,$$

where

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

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

Simple Parry numbers

$d_\beta(1) = t_1 \cdots t_m$

Canonical substitution φ_β over the alphabet $\mathcal{A} = \{0, 1, \dots, m-1\}$

$$\begin{array}{rcl} \varphi_\beta(0) & = & 0^{t_1}1 \\ \varphi_\beta(1) & = & 0^{t_2}2 \\ & \vdots & \\ \varphi_\beta(m-2) & = & 0^{t_{m-1}}(m-1) \\ \varphi_\beta(m-1) & = & 0^{t_m} \end{array}$$

Non-simple Parry numbers

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

Canonical substitution φ_β over the alphabet $\mathcal{A} = \{0, 1, \ldots, m+p-1\}$

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

Fixed point $\mathbf{u}_\beta = \lim_{n\to\infty} \varphi_\beta^n(0) = 0^{t_1} 1 \cdots$.

Basic definitions – factor complexity

$\mathcal{A} = \{0, 1, \ldots, q-1\}$	an alphabet
$\mathbf{u} = (\mathbf{u}_i)_{i\in\mathbb{N}}$, $\mathbf{u}_i \in \mathcal{A}$	an infinite word over $\mathcal{A}$
$w = \mathbf{u}_j\mathbf{u}_{j+1}\cdots\mathbf{u}_{j+n-1}$	a factor of $\mathbf{u}$ of length n
$\mathcal{L}_n(\mathbf{u})$	the set of factors of $\mathbf{u}$ of length n
$\mathcal{L}(\mathbf{u}) = \bigcup_{n\in\mathbb{N}} \mathcal{L}_n(\mathbf{u})$	the language of $\mathbf{u}$

The *factor complexity* of **u** is the function $\mathcal{C} : \mathbb{N} \to \mathbb{N}$ given by

$$\mathcal{C}(n) := \#\mathcal{L}_n(\mathbf{u}).$$

Basic definitions – fixed point of substitution

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

A substitution φ 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 φ is *primitive and injective*.

In general, the factor complexity of a fixed point of any primitive substitution is a sublinear function $\mathcal{C}(n) \leq an + b$, $a, b \in \mathbb{N}$.

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 $\mathcal{C}(n)$ is known,
- in particular, $(m-1)n + 1 \leq \mathcal{C}(n) \leq mn$, for all $n \geq 1$,
- $\mathcal{C}(n)$ is affine $\Leftrightarrow$

  1) $t_m = 1$

  2) for all $i = 2, 3, \ldots, m-1$ we have

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

  Then $\mathcal{C}(n) = (m-1)n + 1$.

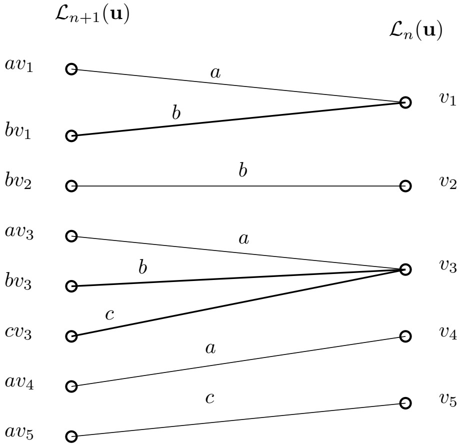
Special factors

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

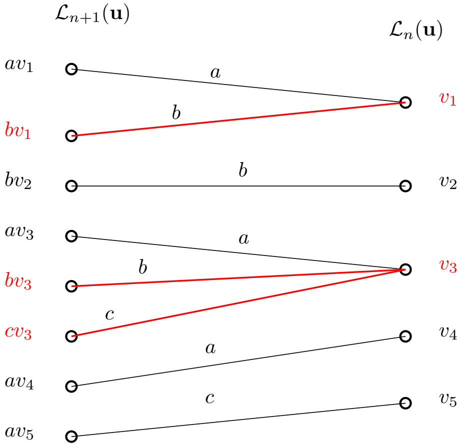
$$\mathrm{Lext}(v) := \{a \in \mathcal{A} \mid av \in \mathcal{L}(\mathbf{u})\}.$$

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

LS factors and factor complexity



LS factors and factor complexity



$\mathcal{C}(n+1) - \mathcal{C}(n) = 3$

LS factors and factor complexity

For the first difference of the complexity function it holds that

$$\triangle \mathcal{C}(n) := \mathcal{C}(n+1) - \mathcal{C}(n) = \sum_{\substack{v \in \mathcal{L}_n(\mathbf{u}) \\ v \text{ is LS}}} (\#\mathrm{Lext}(v) - 1).$$

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

$$\triangle \mathcal{C}(n) \geq 1 \text{ for all } n \in \mathbb{N} \Leftrightarrow \mathbf{u} \text{ is aperiodic.}$$

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**.

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

Structure of LS factors – infinite LS branches

Definition

An infinite word $\mathbf{w}$ *is said to be an infinite LS branch of* $\mathbf{u}$ *if each prefix of* $\mathbf{w}$ *is a LS factor of* $\mathbf{u}$*.*

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

- $\mathbf{u}$ 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***.*

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

- **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 $\mathbf{w}$ *is said to be an infinite LS branch of* $\mathbf{u}$ *if each prefix of* $\mathbf{w}$ *is a LS factor of* $\mathbf{u}$*.*

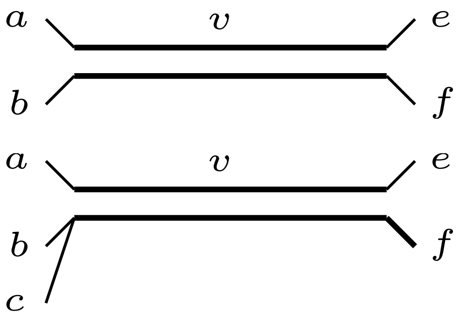
$$Lext(\mathbf{w}) = \bigcap_{v \text{ prefix } \mathbf{w}} Lext(v).$$

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

Structure of LS factors – maximal LS factors

Definition

A LS factor v having left extensions $a, b \in Lext(v)$ 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: $\varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534$

$\mathbf{u} = \varphi^{\infty}(1)$

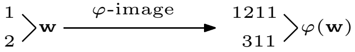
w is a LS factor or an infinite LS branch of $\mathbf{u}$ with left extensions 1, 2 and 3:

Images of LS factors

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

$\mathbf{u} = \varphi^{\infty}(1)$

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

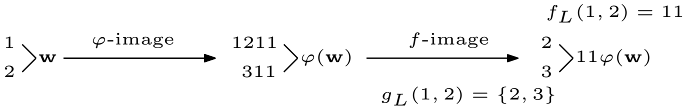


Images of LS factors

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

$\mathbf{u} = \varphi^{\infty}(1)$

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

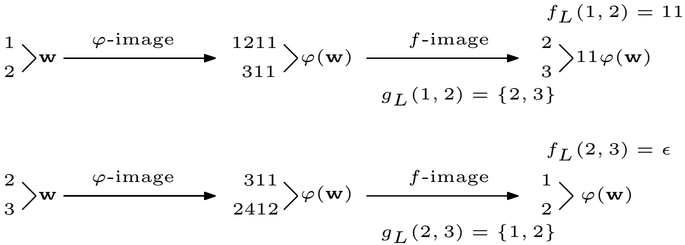


Images of LS factors

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

$\mathbf{u} = \varphi^{\infty}(1)$

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

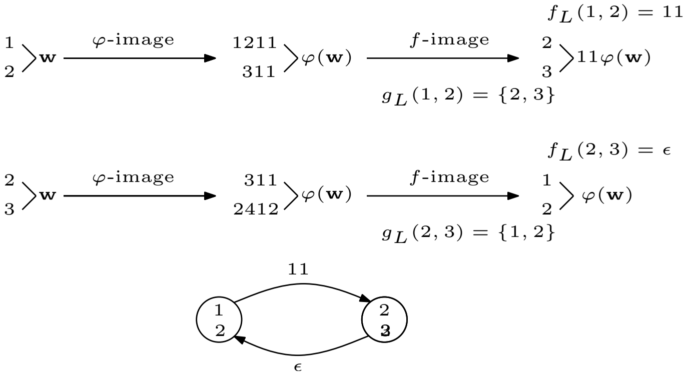


Images of LS factors

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

$\mathbf{u} = \varphi^{\infty}(1)$

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



Graph GL_φ

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

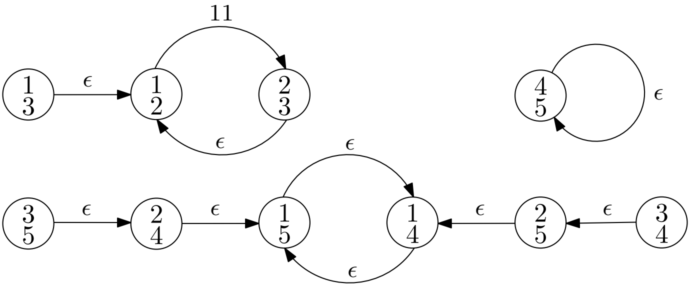
Edges: if $g_L(a, b) = (c, d)$, then there is an edge between (a, b) and (c, d) with label $f_L(a, b)$.

Graph GL_φ

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

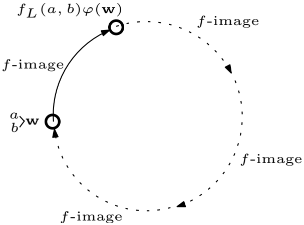


Structure of infinite branches

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**′.



Structure of infinite branches

Assumption: For each infinite LS branch $\mathbf{w}$ it holds that

a) f-image of $\mathbf{w}$ is uniquely given,

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

Theorem

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

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

Structure of infinite branches

Assumption: For each infinite LS branch $\mathbf{w}$ it holds that

a) f-image of $\mathbf{w}$ is uniquely given,

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

Theorem

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

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

- $f_L = \epsilon \Rightarrow \mathbf{w} = \varphi^l(\mathbf{w})$ and (a, b) is a vertex of a cycle in GL_φ labelled by ϵ only,

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 $\mathbf{w}'$ such that **w** is f-image of $\mathbf{w}'$.

Theorem

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

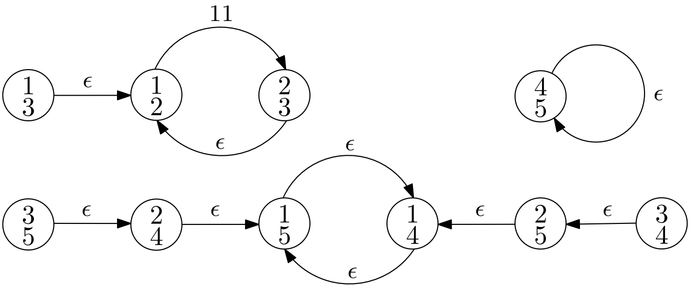
$$\mathbf{w} = f_L(g_L^{l-1}(a,b)) \cdots \varphi^{l-2}(f_L(g_L(a,b)))\varphi^{l-1}(f_L(a,b))\varphi^l(\mathbf{w}).$$

- $f_L = \epsilon \Rightarrow \mathbf{w} = \varphi^l(\mathbf{w})$ and (a, b) is a vertex of a cycle in GL_φ labelled by ϵ only,
- otherwise, (a, b) is a vertex of a cycle in GL_φ labelled not only by ϵ.

Example – how to identify infinite LS branche

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

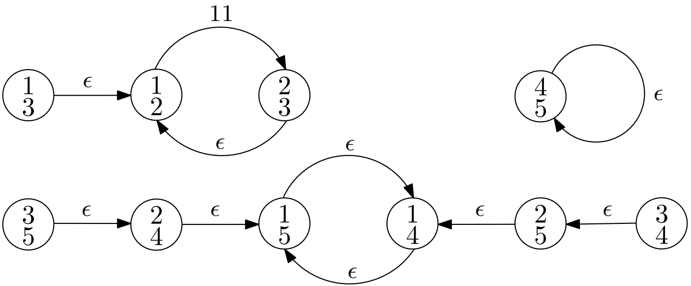
$\mathrm{Lext}(1) = \{1, 2, 3, 4, 5\}, \mathrm{Lext}(2) = \{1, 4, 5\}, \mathrm{Lext}(3) = \{1, 4, 5\}, \mathrm{Lext}(4) = \{1, 2, 3\}, \mathrm{Lext}(5) = \{1, 2, 3\}$



Example – how to identify infinite LS branche

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

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

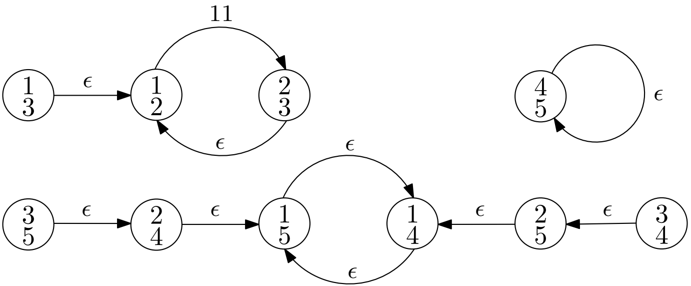


- $\mathbf{w} = 11\varphi^2(\mathbf{w}) \quad \rightarrow \quad 11\varphi^2(11)\varphi^4(11)\cdots$

Example – how to identify infinite LS branche

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

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

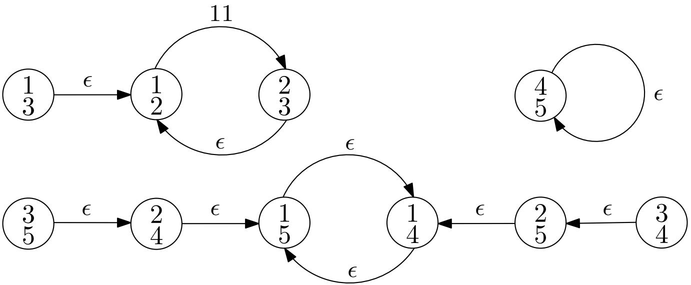


- $\varphi(11)\varphi^3(11)\cdots, \quad 11\varphi^2(11)\varphi^4(11)\cdots$

Example – how to identify infinite LS branche

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

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

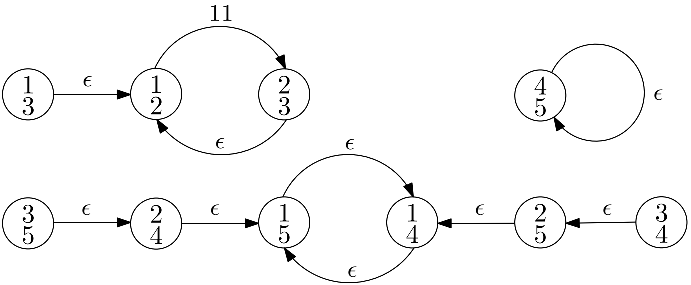


- $\varphi(11)\varphi^3(11)\cdots, \quad 11\varphi^2(11)\varphi^4(11)\cdots$
- $\varphi^\omega(1), \varphi^\omega(4), \varphi^\omega(5), (\varphi^2)^\omega(2), (\varphi^2)^\omega(3)$

Example – how to identify infinite LS branche

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

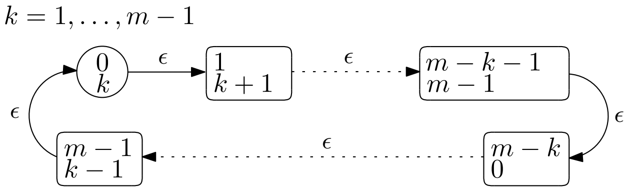
$\mathrm{Lext}(1) = \{1,2,3,4,5\}, \mathrm{Lext}(2) = \{1,4,5\}, \mathrm{Lext}(3) = \{1,4,5\}, \mathrm{Lext}(4) = \{1,2,3\}, \mathrm{Lext}(5) = \{1,2,3\}$



- $\varphi(11)\varphi^3(11)\cdots, \quad 11\varphi^2(11)\varphi^4(11)\cdots$
- $\varphi^\omega(1), (\varphi^2)^\omega(2), (\varphi^2)^\omega(3)$

GL_{φ_β} for simple Parry numbers

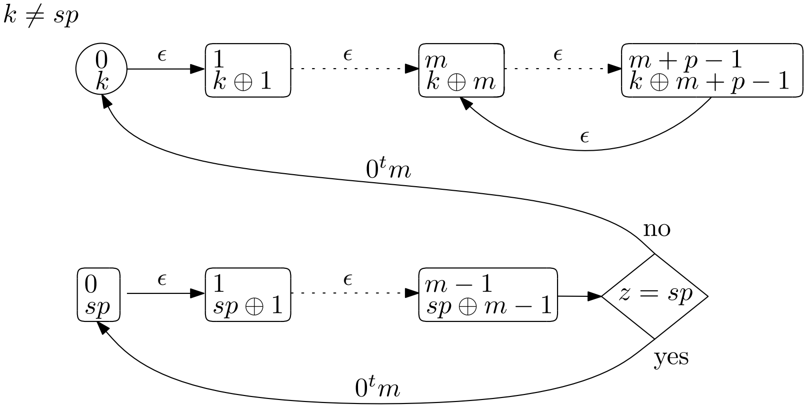
$f_L(a,b) = \epsilon$ for all $a, b \in \{0, 1, \ldots, m-1\}$ and $\mathbf{u}_\beta = \varphi_\beta^\omega(0)$ is the only fixed point



$\Rightarrow$ $\mathbf{u}_\beta$ is the only infinite LS branch

GL_{φ_β} for non-simple Parry numbers

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



Infinite LS factors

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

Definition

$\beta \in \mathcal{S} \Leftrightarrow z = sp \Leftrightarrow$

a) $d_\beta(1) = t_1 \cdots t_m (0 \cdots 0 t_{m+p})^\omega$ *and* $t_m > t_{m+p}$

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

$\beta \in \mathcal{S}_0 \Leftrightarrow d_\beta(1) = t_1 (0 \cdots 0 (t_1 - 1))^\omega.$

Infinite LS factors

Theorem

- *If* β *is a non-simple Parry and* $p > 1$*, then* $\mathbf{u}_\beta$ *is an infinite LS branch with left extensions* $\{m, m+1, \ldots, m+p-1\}$*.*

Infinite LS factors

Theorem

- *If β is a non-simple Parry and $p > 1$, then $\mathbf{u}_\beta$ is an infinite LS branch with left extensions $\{m, m+1, \dots, m+p-1\}$.*
- *If $\beta \notin \mathcal{S}$, then $\mathbf{u}_\beta$ is the only one infinite LS branch.*

Infinite LS factors

Theorem

- *If β is a non-simple Parry and $p > 1$, then $\mathbf{u}_\beta$ is an infinite LS branch with left extensions $\{m, m+1, \ldots, m+p-1\}$.*
- *If $\beta \notin \mathcal{S}$, then $\mathbf{u}_\beta$ is the only one infinite LS branch.*
- *If $\beta \in \mathcal{S}$, 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* $\mathbf{u}_\beta$ *is affine* $\Leftrightarrow$ $\mathbf{u}_\beta$ *does not contain any* (a,b)*-maximal factor* $\Leftrightarrow \beta \in \mathcal{S}_0 \Leftrightarrow d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$. *Then* $\mathcal{C}(n) = (m+p-1)n+1$.
    - *The first equivalence is not valid in general (Chacon),*
    - $\beta \in \mathcal{S}_0 \Rightarrow \beta$ *is an unitary Pisot number (Frougny).*

Affine complexity

Theorem

- *The factor complexity of $\mathbf{u}_\beta$ is affine* $\Leftrightarrow$ $\mathbf{u}_\beta$ *does not contain any* (a,b)*-maximal factor* $\Leftrightarrow \beta \in \mathcal{S}_0 \Leftrightarrow d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$. *Then* $\mathcal{C}(n) = (m+p-1)n+1$.
  - *The first equivalence is not valid in general (Chacon),*
  - $\beta \in \mathcal{S}_0 \Rightarrow \beta$ *is an unitary Pisot number (Frougny).*
- *If* $p > 1$ *and* $\beta \in \mathcal{S}_0$, *then* $\mathbf{u}_\beta$ *and* $0^{-1}\mathbf{u}_\beta$ *are the only infinite LS branches.*

Affine complexity

Theorem

- *The factor complexity of* $\mathbf{u}_\beta$ *is affine* $\Leftrightarrow$ $\mathbf{u}_\beta$ *does not contain any* (a,b)*-maximal factor* $\Leftrightarrow \beta \in \mathcal{S}_0 \Leftrightarrow d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$. *Then* $\mathcal{C}(n) = (m+p-1)n+1$.
  - *The first equivalence is not valid in general (Chacon),*
  - $\beta \in \mathcal{S}_0 \Rightarrow \beta$ *is an unitary Pisot number (Frougny).*
- *If* $p > 1$ *and* $\beta \in \mathcal{S}_0$, *then* $\mathbf{u}_\beta$ *and* $0^{-1}\mathbf{u}_\beta$ *are the only infinite LS branches.*
- *known result:* $\mathbf{u}_\beta$ *is Sturmian* $\Leftrightarrow p = 1$ *and* $\beta \in \mathcal{S}_0$, *i.e.* $d_\beta(1) = t_1(t_1-1)^\omega$.

Affine complexity

Theorem

- *The factor complexity of* $\mathbf{u}_\beta$ *is affine* $\Leftrightarrow$ $\mathbf{u}_\beta$ *does not contain any* (a,b)*-maximal factor* $\Leftrightarrow \beta \in \mathcal{S}_0 \Leftrightarrow d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$. *Then* $\mathcal{C}(n) = (m+p-1)n+1$.
  - *The first equivalence is not valid in general (Chacon),*
  - $\beta \in \mathcal{S}_0 \Rightarrow \beta$ *is an unitary Pisot number (Frougny).*
- *If* $p > 1$ *and* $\beta \in \mathcal{S}_0$*, then* $\mathbf{u}_\beta$ *and* $0^{-1}\mathbf{u}_\beta$ *are the only infinite LS branches.*
- *known result:* $\mathbf{u}_\beta$ *is Sturmian* $\Leftrightarrow p = 1$ *and* $\beta \in \mathcal{S}_0$*, i.e.* $d_\beta(1) = t_1(t_1-1)^\omega$.

THE END