Factor complexity of Arnoux-Rauzy and Poincaré S-adic sequences

Sébastien Labbé

Laboratoire d'Informatique Algorithmique : Fondements et Applications
Université Paris Diderot Paris 7

Journées annuelles SDA2 du GDRIM
Amiens
June 10-12th 2013

Joint work with Valérie Berthé

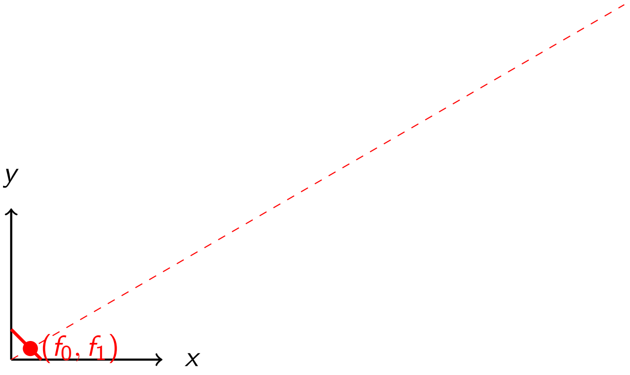
Outline

Plan

1. Introduction
2. Multidimensional Euclidean Algorithm
3. Result on factor complexity
4. Future work

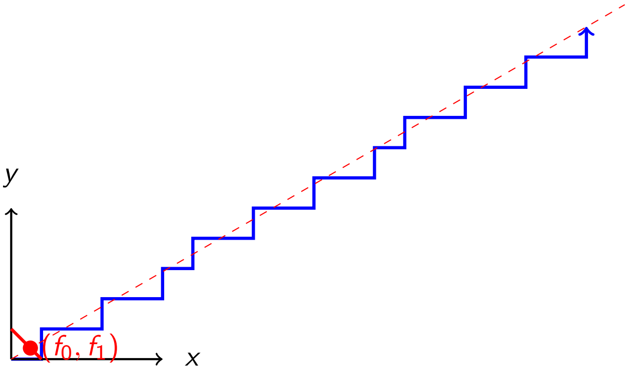
Question

Given $(f_0, f_1) \in \mathbb{R}^2$ *such that* $f_0 + f_1 = 1$, *can we construct an infinite sequence* $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$ *on the alphabet* $\mathcal{A} = \{0, 1\}$ *such that the frequency of digit* i *is* f_i *for all* $i \in \mathcal{A}$ *?*



Question

Given $(f_0, f_1) \in \mathbb{R}^2$ *such that* $f_0 + f_1 = 1$, *can we construct an infinite sequence* $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$ *on the alphabet* $\mathcal{A} = \{0, 1\}$ *such that the frequency of digit* i *is* f_i *for all* $i \in \mathcal{A}$ *?*



$0100100101001001001010010010 01\cdots$

Factors, frequencies, vectors

Factor : finite string of consecutive digits. Let

$$w = 0100100101001001001010010011001.$$

Then 00100 and 1001 are factors of w. 1001 is a suffix of w.

Factors, frequencies, vectors

Factor : finite string of consecutive digits. Let

$$w = 010010010100100100101001001001.$$

Then 00100 and 1001 are factors of w. 1001 is a suffix of w.

$|w|$: the length of the factor w. $|w| = 30$.

Factors, frequencies, vectors

Factor : finite string of consecutive digits. Let

$$w = 010010010100100100101001001001.$$

Then 00100 and 1001 are factors of w. 1001 is a suffix of w.

$|w|$: the length of the factor w. $|w| = 30$.

$|w|_u$: the number of occurences of the factor u in w.

$$|w|_0 = 19, \quad |w|_1 = 11$$

$$|w|_{00} = 8, \quad |w|_{01} = 11, \quad |w|_{10} = 10, \quad |w|_{11} = 0.$$

Factors, frequencies, vectors

Factor : finite string of consecutive digits. Let

$$w = 010010010100100100101001001001.$$

Then 00100 and 1001 are factors of w. 1001 is a suffix of w.

$|w|$: the length of the factor w. $|w| = 30$.

$|w|_u$: the number of occurences of the factor u in w.

$$|w|_0 = 19, \quad |w|_1 = 11$$

$$|w|_{00} = 8, \quad |w|_{01} = 11, \quad |w|_{10} = 10, \quad |w|_{11} = 0.$$

$\vec{u} = (|u|_0, |u|_1)$: the abelian vector of the factor u.

$$\overrightarrow{00100} = (4, 1), \quad \overrightarrow{1001} = (2, 2).$$

Normal numbers

Let **w** be the development in base b of a number $x \in \mathbb{R}$.
Let p be a finite prefix of **w**.

Definition

The number x is called normal in base b if

$$\lim_{|p|\to\infty} \frac{|p|_s}{|p|} = \frac{1}{b^k}$$

for every string s of length k.

The number

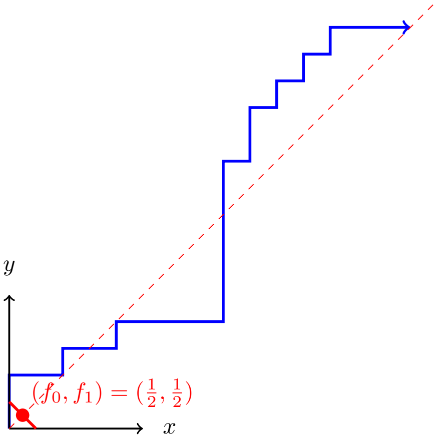
$$\pi = 11.00100100001111110110101010001\cdots$$

is conjectured normal in base 2.

Normal numbers : not a good answer

Normal number are such that :

- Really different factors appear in the sequence (ex : 0000 and 1111) ;
- All factors (b^k factors of length k) appear in the sequence.



π=11.0010010000111111011010101000$\cdots$

Balanced sequences

Definition

An infinite word $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$ is said to be finitely balanced or C-balanced or balanced if there exists a constant $C \in \mathbb{N}$ such that
for all pairs of factors u, v of $\mathbf{w}$ of the same length,

$$||\vec{u} - \vec{v}||_\infty \leq C.$$

Base 2 development of π is not balanced for $C = 1$ because the factors

$$0000 \qquad \text{and} \qquad 1111$$

appear in the development. If π was proved normal, then 0^k and 1^k would also appear for all k, thus it would not be balanced.

Discrepancy

Definition

The discrepancy of an infinite word $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$ *is defined as*

$$\limsup_{i \in \mathcal{A}, p \text{ prefix of } \mathbf{w}} \left| f_i \cdot |p| - |p|_i \right|.$$

where f_i *is the frequency of the letter* $i \in \mathcal{A}$*, if it exists :*

$$f_i = \lim_{p \text{ prefix of } \mathbf{w}} \frac{|p|_i}{|p|}$$

$\mathbf{w}$ is balanced $\iff$ $\mathbf{w}$ has finite discrepancy $\iff$ $\mathbf{w}$ stays at bounded distance from the euclidean line of direction (f_0, f_1)

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 00010001000100100010001000100100010001001$$

$$\mathrm{L}_w(4) = \{ \quad , \quad , \quad , \quad , \quad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 0001000100010010001000100010010001001001$$

$$\mathrm{L}_w(4) = \{0001, \quad , \quad , \quad , \quad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 0001000100010010001000100010010001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, \quad , \quad , \quad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 000100010001001000100010001001000100010001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, \qquad , \qquad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 0001000100010010001000100010010001001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, 1000, \qquad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 0001000100010010001000100010010001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, 1000, \qquad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 00010001000100100010001000100100010001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, 1000, \quad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 0001000100010010001000100010010001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, 1000, \qquad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 00010001000100100010001000100100010001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, 1000, \qquad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 000100010001001000100010001001000100010001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, 1000, \qquad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 0001000100010010001000100010010001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, 1000, \qquad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 00010001000100100010001000100100010001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, 1000, \qquad \}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 0001000100010010001000100010010001001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, 1000, 1001\}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 00010001000100100010001000100100010001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, 1000, 1001\}$$

n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Factor complexity

Let $w \in \mathcal{A}^{\mathbb{N}}$. The factor complexity is a function $p_w(n) : \mathbb{N} \to \mathbb{N}$ counting the number of factors of length n, noted $\mathrm{L}_w(n)$, in the sequence w.

$$w = 00010001000100100010001000100100010001001$$

$$\mathrm{L}_w(4) = \{0001, 0010, 0100, 1000, 1001\}$$

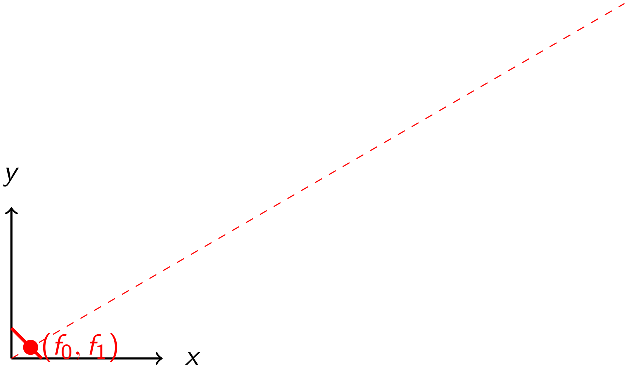
n	$p_w(n)$
0	1
1	2
2	3
3	4
4	5

Upper bound : $p_w(n) \leq |\mathcal{A}|^n$.

Let $\Delta_d = \{(f_1, f_2, \dots, f_d) \in \mathbb{R}_+^d : f_1 + f_2 + \cdots + f_d = 1\}$.

Question (updated)

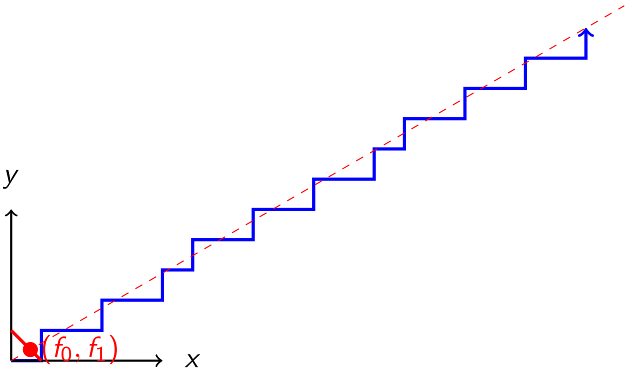
Given a vector $(f_1, f_2, \cdots, f_d) \in \Delta_d$, can we construct an infinite word $\mathbf{w}$ on the alphabet $\mathcal{A} = \{1, 2, \cdots, d\}$ such that the frequency of each letter $i \in \mathcal{A}$ exists and is equal to f_i, $\mathbf{w}$ is balanced and has a linear factor complexity?



Let $\Delta_d = \{(f_1, f_2, \dots, f_d) \in \mathbb{R}_+^d : f_1 + f_2 + \cdots + f_d = 1\}$.

Question (updated)

Given a vector $(f_1, f_2, \cdots, f_d) \in \Delta_d$, can we construct an infinite word **w** *on the alphabet $\mathcal{A} = \{1, 2, \cdots, d\}$ such that the frequency of each letter $i \in \mathcal{A}$ exists and is equal to f_i,* **w** *is balanced and has a linear factor complexity ?*



010010010100100100101001001001$\cdots$

Answer for $d = 2$: Sturmian words

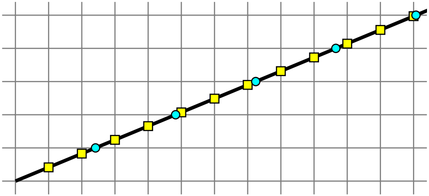
Proposition

Sturmian words are 1-balanced and satisfy $p(n) = n + 1$.

Sturmian word are also obtained from coding of rotations :

$$w_n = \begin{cases} 1 & \text{if } (\alpha + n\beta) \mod 1 \in [0, \beta[\\ 0 & \text{else} \end{cases} .$$

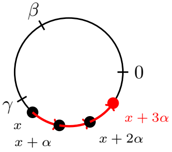
for some $\alpha, \beta \in [0, 1[$
or from the cutting sequence of a line :

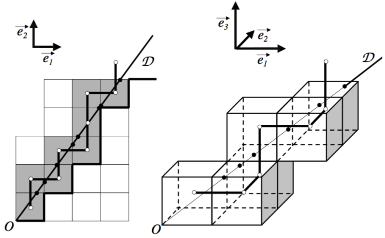


Possible answers for $d > 2$

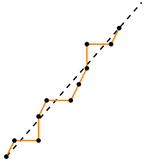
Typical answers to this question are :

	$\forall v \in \mathbb{R}^d_+$	$p(n)$ is linear	Balanced
Coding of Rotations	Yes	Yes	No
Coding of Interval Exchange Tr.	Yes	Yes	No
Billiard words	Yes	No	Yes
E. Andres discrete lines	Yes	No	Yes





Borel (2006)



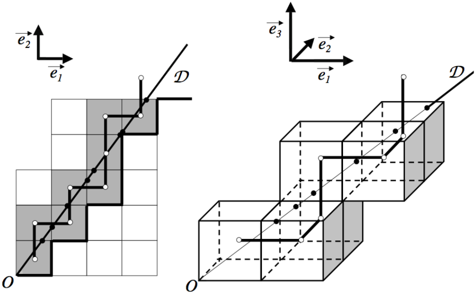
$$-7/2 \leq 2x - 5z < 7/2$$
$$-8/2 \leq 3x - 5y < 8/2$$
$$-5/2 \leq 2y - 3z < 5/2$$

Andres (2003)

Factor Complexity of the Billiard word

Theorem (Baryshnikov, 1995 ; Bédaride, 2003)

If both the direction $(\alpha_1, \alpha_2, \alpha_3)$ *and* $(\alpha_1^{-1}, \alpha_2^{-1}, \alpha_3^{-1})$ *are* $\mathbb{Q}$ *independent, the number of factors appearing in the Billiard word in a cube is exactly* $p(n) = n^2 + n + 1$.



Source of image : J.-P. Borel, Complexity of Degenerated Three Dimensional Billiard Words, Developments in Language Theory 4036 (2006) 386-396.

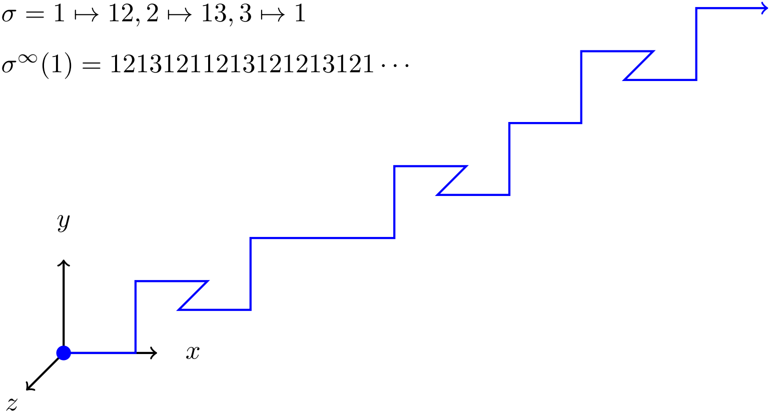
Hope : Tribonacci Example from Rauzy (1982)

The Tribonacci word is an infinite word over $\{1, 2, 3\}$ containing $2n + 1$ factors of length n and is balanced.

Can we generalize this to any 3D directions ?

$\sigma = 1 \mapsto 12, 2 \mapsto 13, 3 \mapsto 1$

$\sigma^\infty(1) = 1213121121312121 3121 \cdots$

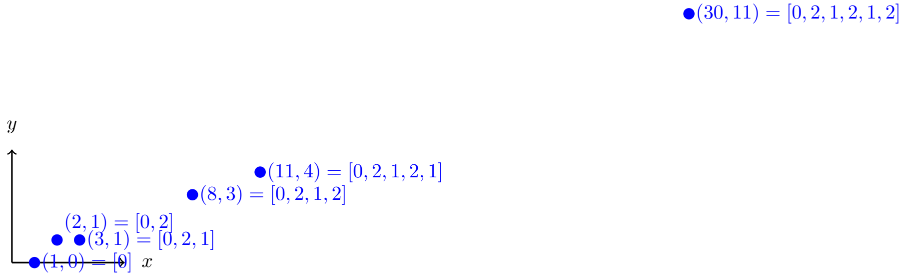


Let $\alpha = \frac{\sqrt{3}-1}{2} = 0.36602540\cdots$. We have

$$\alpha = 0 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cdots}}}} = [0; 2, 1, 2, 1, 2, 1, \cdots]$$

The convergents p_n/q_n are

$$0, \frac{1}{2}, \frac{1}{3}, \frac{3}{8}, \frac{4}{11}, \frac{11}{30}, \frac{15}{41}, \frac{41}{112}, \frac{56}{153}, \frac{153}{418}, \frac{209}{571}, \frac{571}{1560}, \frac{780}{2131}, \cdots$$



Continued fractions : matrices from convergents

With

$$L = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad R = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

the convergents can be obtained as

$$\begin{aligned}
\begin{pmatrix} q_0 \\ p_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} &= R^0 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
\begin{pmatrix} q_1 \\ p_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} &= R^0 L^2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix} &= R^0 L^2 R^1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
\begin{pmatrix} q_3 \\ p_3 \end{pmatrix} = \begin{pmatrix} 8 \\ 3 \end{pmatrix} &= R^0 L^2 R^1 L^2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
\begin{pmatrix} q_4 \\ p_4 \end{pmatrix} = \begin{pmatrix} 11 \\ 4 \end{pmatrix} &= R^0 L^2 R^1 L^2 R^1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
\begin{pmatrix} q_5 \\ p_5 \end{pmatrix} = \begin{pmatrix} 30 \\ 11 \end{pmatrix} &= R^0 L^2 R^1 L^2 R^1 L^2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}
\end{aligned}$$

Morphisms and incidence matrices

A morphism is a function $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ such that $\sigma(uv) = \sigma(u)\sigma(v)$ for all $u, v \in \mathcal{A}^*$.
The incidence matrix M_σ of a morphism σ is the unique matrix over the non negative integers such that, for all $u \in \mathcal{A}^*$:

$$M_\sigma(\overrightarrow{u}) = \overrightarrow{\sigma(u)}.$$

$$\begin{array}{ccc} \mathcal{A}^* & \xrightarrow{\sigma} & \mathcal{A}^* \\ \downarrow \vec{\cdot} & & \downarrow \vec{\cdot} \\ \mathbb{N}^d & \xrightarrow{M_\sigma} & \mathbb{N}^d \end{array}$$

Example.

$$\sigma = \begin{array}{l} 0 \mapsto 0100111 \\ 1 \mapsto 101 \end{array} \qquad M_\sigma = \begin{pmatrix} 3 & 1 \\ 4 & 2 \end{pmatrix}$$

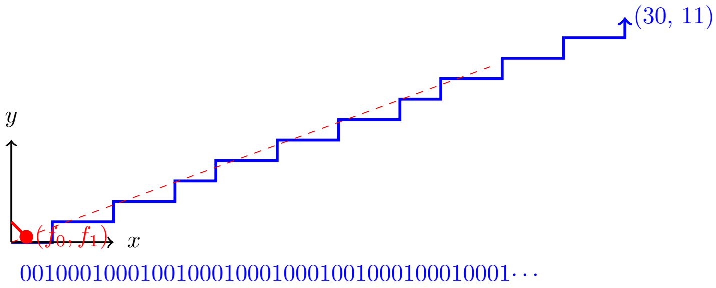
Continued fractions : substitutions from matrices

With

$$L = \begin{array}{l} 0 \mapsto 0 \\ 1 \mapsto 01 \end{array} \qquad \text{and} \qquad R = \begin{array}{l} 0 \mapsto 10 \\ 1 \mapsto 1 \end{array}$$

the convergents can be transformed into finite sequences over $\mathcal{A}$:

$$\begin{array}{lcl} w_0 &=& R^0(0) &=& 0 \\ w_1 &=& R^0L^2(1) &=& 001 \\ w_2 &=& R^0L^2R^1(0) &=& 0010 \\ w_3 &=& R^0L^2R^1L^2(1) &=& 00100010001 \\ w_4 &=& R^0L^2R^1L^2R^1(0) &=& 001000100010010 \\ w_5 &=& R^0L^2R^1L^2R^1L^2(1) &=& 00100010001001000100010001001000100010001 \end{array}$$



The previous command in Sage

Available in Sage since December 2009 :

```
sage: R = WordMorphism('0->01,1->1')
sage: L = WordMorphism('0->0,1->01')
sage: w5 = words.s_adic([L,L,R,L,L,R,L,L], '1')
sage: print w5
0001000100010010001000100010010001001
```

The proposed approach

Generalize the Tribonacci word to any $v \in \mathbb{R}^3_+$ using multidimensional continued fractions algorithms and S-adic sequences.

There are other motivations for this approach :

- Study S-adic sequences, in the perspective of the S-adic conjecture concerning factor complexity.
- Study multidimensional continued fractions algorithms from substitutions and combinatorics on words point of view.
- Extend Pisot conjecture and Rauzy fractals (usually defined for fixed point of morphisms) to $\mathcal{S}$-adic sequences.

Plan

Continued fractions : from Euclid Algorithm

With

$$L = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

the execution of Euclid Algorithm appears as

$$\begin{aligned}
\begin{pmatrix} 30 \\ 11 \end{pmatrix} &= L^2R^1L^2R^1L^2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
\begin{pmatrix} 8 \\ 11 \end{pmatrix} &= R^1L^2R^1L^2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
\begin{pmatrix} 8 \\ 3 \end{pmatrix} &= L^2R^1L^2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
\begin{pmatrix} 2 \\ 3 \end{pmatrix} &= R^1L^2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
\begin{pmatrix} 2 \\ 1 \end{pmatrix} &= L^2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}
\end{aligned}$$

3D Continued fraction algorithms

Brun's Algorithm : Subtract the second largest to the largest.

$(7,4,6) \to (1,4,6) \to (1,4,2) \to (1,2,2) \to (1,0,2) \to (1,0,1) \to (0,0,1).$

3D Continued fraction algorithms

Brun's Algorithm : Subtract the second largest to the largest.

$(7,4,6) \to (1,4,6) \to (1,4,2) \to (1,2,2) \to (1,0,2) \to (1,0,1) \to (0,0,1)$.

Selmer's Algorithm : Subtract the smallest to the largest.

$(7,4,6) \to (3,4,6) \to (3,4,3) \to (3,1,3) \to (2,1,3) \to (2,1,2) \to (1,1,2)$
$\to (1,1,1) \to (0,1,1) \to (0,0,1)$

3D Continued fraction algorithms

Brun's Algorithm : Subtract the second largest to the largest.

$(7,4,6) \to (1,4,6) \to (1,4,2) \to (1,2,2) \to (1,0,2) \to (1,0,1) \to (0,0,1).$

Selmer's Algorithm : Subtract the smallest to the largest.

$(7,4,6) \to (3,4,6) \to (3,4,3) \to (3,1,3) \to (2,1,3) \to (2,1,2) \to (1,1,2)$
$\to (1,1,1) \to (0,1,1) \to (0,0,1)$

Poincaré's Algorithm : Subtract the smallest to the mid and the mid to the largest.

$$(7,4,6) \to (1,4,2) \to (1,2,1) \to (1,1,0) \to (1,0,0)$$

3D Continued fraction algorithms

Brun's Algorithm : Subtract the second largest to the largest.

$$(7,4,6) \to (1,4,6) \to (1,4,2) \to (1,2,2) \to (1,0,2) \to (1,0,1) \to (0,0,1).$$

Selmer's Algorithm : Subtract the smallest to the largest.

$$(7,4,6) \to (3,4,6) \to (3,4,3) \to (3,1,3) \to (2,1,3) \to (2,1,2) \to (1,1,2)$$
$$\to (1,1,1) \to (0,1,1) \to (0,0,1)$$

Poincaré's Algorithm : Subtract the smallest to the mid and the mid to the largest.

$$(7,4,6) \to (1,4,2) \to (1,2,1) \to (1,1,0) \to (1,0,0)$$

Arnoux-Rauzy's Algorithm : Subtract the sum of the two smallest to the largest (not always possible).

$$(7,4,6) \to \text{Impossible}$$

3D Continued fraction algorithms

Brun's Algorithm : Subtract the second largest to the largest.

$$(7,4,6) \to (1,4,6) \to (1,4,2) \to (1,2,2) \to (1,0,2) \to (1,0,1) \to (0,0,1).$$

Selmer's Algorithm : Subtract the smallest to the largest.

$$(7,4,6) \to (3,4,6) \to (3,4,3) \to (3,1,3) \to (2,1,3) \to (2,1,2) \to (1,1,2) \to (1,1,1) \to (0,1,1) \to (0,0,1)$$

Poincaré's Algorithm : Subtract the smallest to the mid and the mid to the largest.

$$(7,4,6) \to (1,4,2) \to (1,2,1) \to (1,1,0) \to (1,0,0)$$

Arnoux-Rauzy's Algorithm : Subtract the sum of the two smallest to the largest (not always possible).

$$(7,4,6) \to \text{Impossible}$$

Fully subtractive's Algorithm : Subtract the smallest to the other two.

$$(7,4,6) \to (3,4,2) \to (1,2,2) \to (1,1,1) \to (1,0,0)$$

Experimentations

	Min	Mean	Max	Std
Arnoux-Rauzy (AR)	0.6000	0.8922	1.200	0.09953
Fully subtractive	0.5000	6.047	14.21	4.385
Selmer	0.5000	2.151	12.75	2.076
Brun	0.5000	1.100	2.000	0.2625
Poincaré	0.5000	2.476	11.13	2.245
AR-Fully subtractive	0.5000	1.154	4.000	0.3759
AR-Selmer	0.5000	0.9991	1.600	0.1429
AR-Brun	0.5000	0.9169	1.520	0.1170
AR-Poincaré	0.5000	0.9066	1.320	0.1079

TABLE: Statistics for the discrepancy for strictly positive integer vectors (a_1, a_2, a_3) such that $a_1 + a_2 + a_3 = N$ and $N = 100$.

Arnoux-Rauzy words : $\mathcal{S}$-adic representation

Let $\mathcal{A} = \{1, 2, 3\}$, $\mathcal{S} = \{\sigma_i : i \in \mathcal{A}\}$ where

$$\sigma_i : i \mapsto i, j \mapsto ji \text{ for } j \notin \mathcal{A}.$$

An Arnoux-Rauzy word can be represented as an $\mathcal{S}$-adic sequence

$$\mathbf{w} = \lim_{n\to\infty} (\sigma_{i_0} \circ \sigma_{i_1} \circ \cdots \circ \sigma_{i_n})(1).$$

where every letter in $\mathcal{A}$ occurs infinitely often in $(i_m)_{m\geq 0}$. The sequence $(i_m)_{m\geq 0} \in \mathcal{A}^{\mathbb{N}}$ is called the $\mathcal{S}$-directive sequence.

Pierre Arnoux and Gérard Rauzy. Représentation géométrique de suites de complexité 2n+1. *Bull. Soc. Math. France*, 119(2) :199–215, 1991.

J. Cassaigne, S. Ferenczi, and L. Q. Zamboni. Imbalances in Arnoux-Rauzy sequences. *Ann. Inst. Fourier (Grenoble)*, 50(4) :1265–1276, 2000.

Balance properties of Arnoux-Rauzy words

Let $\mathbf{w}$ be an Arnoux-Rauzy word with $\mathcal{S}$-directive sequence $(i_m)_{m\geq 0}$.

Theorem (Berthé, Cassaigne, Steiner, 2012)

If the weak partial quotients are bounded by h, i.e., if we do not have $i_m = i_{m+1} = \cdots = i_{m+h}$ for any $m \geq 0$, then $\mathbf{w}$ is $(2h+1)$-balanced.

Theorem (Berthé, Cassaigne, Steiner, 2012)

Let X be the set of words $\{1121, 1122, 12121, 12122\}$ together with all the words that are obtained from one of these four words by a permutation of the letters 1, 2, and 3. If $(i_m)_{m\geq 0}$ contains no factor in X, then $\mathbf{w}$ is 2-balanced.

Valérie Berthé, Julien Cassaigne, and Wolfgang Steiner. Balance properties of arnoux-rauzy words. *arXiv :1212.5106*, December 2012.

Plan

Arnoux-Rauzy and Poincaré substitutions

For all $\{i, j, k\} = \{1, 2, 3\}$, we consider

$$\begin{array}{ll} \pi_{jk} : i \mapsto ijk, j \mapsto jk, k \mapsto k & \text{(Poincaré substitutions)} \\ \alpha_k : i \mapsto ik, j \mapsto jk, k \mapsto k & \text{(Arnoux-Rauzy substitutions)} \end{array}$$

Namely,

$$\pi_{23} = \begin{cases} 1 \mapsto 123 \\ 2 \mapsto 23 \\ 3 \mapsto 3 \end{cases}, \quad \pi_{13} = \begin{cases} 1 \mapsto 13 \\ 2 \mapsto 213 \\ 3 \mapsto 3 \end{cases}, \quad \alpha_3 = \begin{cases} 1 \mapsto 13 \\ 2 \mapsto 23 \\ 3 \mapsto 3 \end{cases},$$

$$\pi_{12} = \begin{cases} 1 \mapsto 12 \\ 2 \mapsto 2 \\ 3 \mapsto 312 \end{cases}, \quad \pi_{32} = \begin{cases} 1 \mapsto 132 \\ 2 \mapsto 2 \\ 3 \mapsto 32 \end{cases}, \quad \alpha_2 = \begin{cases} 1 \mapsto 12 \\ 2 \mapsto 2 \\ 3 \mapsto 32 \end{cases},$$

$$\pi_{31} = \begin{cases} 1 \mapsto 1 \\ 2 \mapsto 231 \\ 3 \mapsto 31 \end{cases}, \quad \pi_{21} = \begin{cases} 1 \mapsto 1 \\ 2 \mapsto 21 \\ 3 \mapsto 321 \end{cases}, \quad \alpha_1 = \begin{cases} 1 \mapsto 1 \\ 2 \mapsto 21 \\ 3 \mapsto 31 \end{cases}.$$

Quadratic complexity for ARP sequences

In general, it is possible that $p(n+1) - p(n) > 3$ for some values of n. Let

$$s = \pi_{23}\pi_{23}\pi_{13}\pi_{23}\pi_{23}\alpha_1\alpha_3\alpha_2(1).$$

Indeed,

$$p_s(n) = (1, 3, 5, 8, 11, 15, 19, 23, 27, 31, 35, 38, \cdots)$$

Quadratic complexity for ARP sequences

In general, it is possible that $p(n+1) - p(n) > 3$ for some values of n. Let

$$s = \pi_{23}\pi_{23}\pi_{13}\pi_{23}\pi_{23}\alpha_1\alpha_3\alpha_2(1).$$

Indeed,

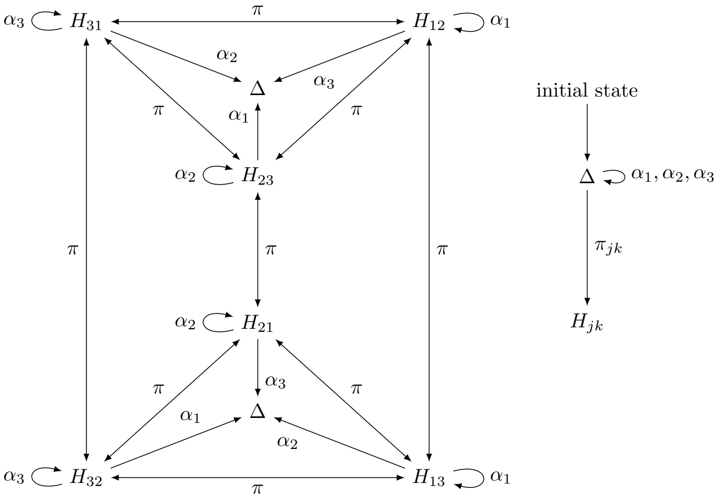
$$p_s(n) = (1, 3, 5, 8, 11, 15, 19, 23, 27, 31, 35, 38, \cdots)$$

Even worse, the fixed point of

$$\pi_{13}\pi_{23} : \begin{cases} 1 \mapsto 132133 \\ 2 \mapsto 2133 \\ 3 \mapsto 3 \end{cases}$$

starting with letter 1 has a quadratic factor complexity.

Deterministic and minimized automaton recognizing the language $\mathcal{L} \subset \mathcal{S}^{\mathbb{N}}$ of ARP algorithm :



Result on complexity

Theorem (Berthé, L., 2013)

For (Lebesgue) almost all $(f_1, f_2, f_3) \in \Delta_3$, *there exists*

$$\mathbf{w} = \lim_{n\to\infty} \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_n(a),$$

such that the frequency of letter i *in* $\mathbf{w}$ *is* f_i *and where* $a \in \{1, 2, 3\}$ *and* $(\sigma_i)_{i\geq 0} \in \mathcal{L} \subset \mathcal{S}^{\mathbb{N}}$.

Theorem (Berthé, L., 2013)

The factor complexity of $\mathbf{w}$ *is such that*

- $p(n+1) - p(n) \in \{2, 3\}$ *and*
- $\limsup_{n\to\infty} \frac{p(n)}{n} \leq \frac{5}{2} < 3$.

Arnoux-Rauzy Poincaré algorithm

Theorem (Boshernitzan, 1984)

A minimal symbolic system (X, S) *such that* $\limsup_{n\to\infty} \frac{p(n)}{n} < 3$ *is uniquely ergodic.*

Theorem (see CANT, Prop. 7.2.10)

A symbolic system (X_x, S) *is uniquely ergodic if, and only if,* x *has uniform frequencies.*

Sébastien Ferenczi and Thierry Monteil. Infinite words with uniform frequencies, and invariant measures. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 373–409. Cambridge Univ. Press, Cambridge, 2010.

Corollary

The frequencies of factors and letters in $\mathbf{w}$ *exist.*

Idea of the proof on complexity

Let $p(n)$ be the factor complexity function of $\mathbf{w}$. Let $s(n)$ and $b(n)$ be its sequences of finite differences of order 1 and 2 :

$$\begin{aligned} p(n) &= 1,3,5,7,9,11,14,17,20,22,24,26,28, \\ s(n) = p(n+1) - p(n) &= 2,2,2,2,2,3,3,3,2,2,2,2, \\ b(n) = s(n+1) - s(n) &= 0,0,0,0,+1,0,0,-1,0,0,0, \end{aligned}$$

Idea of the proof on complexity

Let $p(n)$ be the factor complexity function of **w**. Let $s(n)$ and $b(n)$ be its sequences of finite differences of order 1 and 2 :

$$\begin{aligned} p(n) &= 1, 3, 5, 7, 9, 11, 14, 17, 20, 22, 24, 26, 28, \\ s(n) = p(n+1) - p(n) &= 2, 2, 2, 2, 2, 3, 3, 3, 2, 2, 2, 2, \\ b(n) = s(n+1) - s(n) &= 0, 0, 0, 0, +1, 0, 0, -1, 0, 0, 0, \end{aligned}$$

Functions s and b are related to special and bispecial factors of **w**.

Theorem (Cassaigne, 1997 ; Cassaigne, Nicolas, 2010)

Let $\mathbf{u} \in A^{\mathbb{N}}$ be a infinite [recurrent] word. Then, for all $n \in \mathbb{N}$:

$$s(n) = \sum_{w \in RS_n(\mathbf{u})} (d^+(w) - 1) \quad \text{and} \quad b(n) = \sum_{w \in BS_n(\mathbf{u})} m(w)$$

Special and bispecial words

Let **u** be an infinite word $L(\mathbf{u})$ be its language.
Right extensions and right valence :

$$E^+(w) = \{x \in \mathcal{A} | wx \in L(\mathbf{u})\} \qquad d^+(w) = \mathrm{Card} E^+(w).$$

Left extensions and left valence :

$$E^-(w) = \{x \in \mathcal{A} | xw \in L(\mathbf{u})\} \qquad d^-(w) = \mathrm{Card} E^-(w).$$

A factor w is

- right special if $d^+(w) \geq 2$,
- left special if $d^-(w) \geq 2$,
- bispecial if it is left and right special.

Special and bispecial words

The extension type of a factor w of $\mathbf{u}$ is

$$E(w) = \{(a, b) \in \mathcal{A} \times \mathcal{A} | awb \in L(\mathbf{u})\}.$$

The bilateral multiplicity of a factor w is

$$m(w) = \mathrm{Card}\, E(w) - d^-(w) - d^+(w) + 1.$$

A bispecial factor is said

weak if $m(w) < 0$,   neutral if $m(w) = 0$,   strong if $m(w) > 0$.

Examples of extension types $E(w)$ of bispecial factors w

$$m(w) = \mathrm{Card}\, E(w) - d^{-}(w) - d^{+}(w) + 1.$$

	1	2
1		×
2	×	

$m(w) = -1$
weak

	1	2
1	×	×
2	×	

$m(w) = 0$
neutral and ordinary

	1	2
1	×	×
2	×	×

$m(w) = 1$
strong

Examples of extension types $E(w)$ of bispecial factors w

$$m(w) = \mathrm{Card}\, E(w) - d^-(w) - d^+(w) + 1.$$

	1	2
1		$\times$
2	$\times$	

$m(w) = -1$
weak

	1	2
1	$\times$	$\times$
2	$\times$	

$m(w) = 0$
neutral and ordinary

	1	2
1	$\times$	$\times$
2	$\times$	$\times$

$m(w) = 1$
strong

A bispecial factor w is ordinary if

$$E(w) \subseteq (\{a\} \times A) \cup (A \times \{b\}) \quad \text{for a pair of letters} \quad (a, b) \in E(w).$$

Lemma

If a bispecial factor is ordinary, then it is neutral.

On a binary alphabet, the reciprocal also holds.

Examples of extension types $E(w)$ of bispecial factors w

	1	2	3
1		×	
2		×	
3	×	×	×

$m(w) = 0$
neutral and ordinary

	1	2	3
1		×	
2			
3	×	×	×

$m(w) = 0$
neutral and ordinary

	1	2	3
1		×	
2			×
3	×	×	×

$m(w) = 0$
neutral but not ordinary

	1	2	3
1		×	
2		×	
3	×		×

$m(w) = -1$
weak

	1	2	3
1		×	
2			
3			×

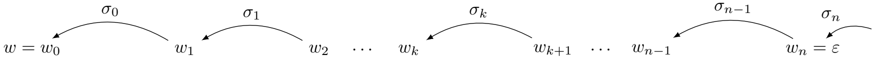
$m(w) = -1$
weak

	1	2	3
1			
2		×	×
3	×	×	×

$m(w) = 1$
strong

Life and history of a bispecial factor

- Some synchronization lemmas allows to define uniquely the antecedent w_{k+1} of a (bispecial) factor $w_k = s\sigma_k(w_{k+1})p$.
- w_k is an extended image of w_{k+1}
- We always have $|w_{k+1}| < |w_k|$.
- The history of w is $\sigma_0\sigma_1\cdots\sigma_n$.
- The life of w is $(w_k)_{0\leq k\leq n}$.



- w_{k+1} has only one extended image w_k under an Arnoux-Rauzy substitution.
- w_{k+1} has one or two extended images under Poincaré substitution.

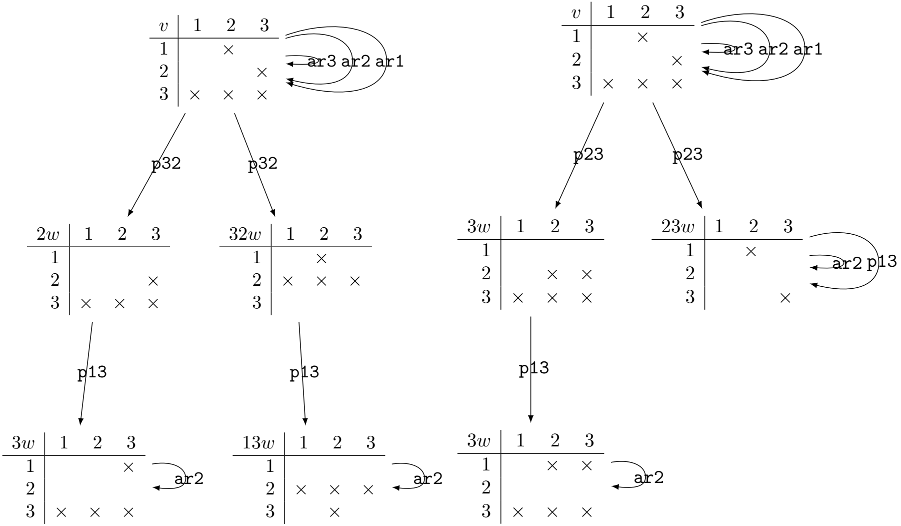
Extension type of the empty word

Lemma

Let $\mathbf{u} \in \mathcal{A}^* \cup \mathcal{A}^{\mathbb{N}}$ *be such that all letters of* $\mathcal{A}$ *appear as proper factors of* $\mathbf{u}$. *Considered as a bispecial factor of the language of the word* $\alpha_k(\mathbf{u})$, *the empty word* ε *is ordinary. Considered as a bispecial factor of the language of the word* $\pi_{jk}(\mathbf{u})$, *the empty word* ε *is neutral but not ordinary :*

$$E_{\alpha_k(\mathbf{u})}(\varepsilon) = \begin{array}{c|ccc} & i & j & k \\ \hline i & & & \times \\ j & & & \times \\ k & \times & \times & \times \end{array} \quad \text{and} \quad E_{\pi_{jk}(\mathbf{u})}(\varepsilon) = \begin{array}{c|ccc} & i & j & k \\ \hline i & & \times & \\ j & & & \times \\ k & \times & \times & \times \end{array}.$$

Giving birth to pair of bispecials words



Special and bispecial words

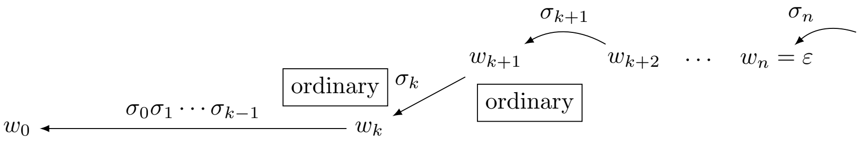
Lemma

Let $\mathbf{u} = \lim_{n\to\infty} \sigma_0\sigma_1\cdots\sigma_n(a_n)$*. Let* w *be a bispecial factor of* $\mathbf{u}$ *such that* $n = \mathrm{age}(w)$ *and* $\lim_{m\to\infty} \sigma_{n+1}\sigma_{n+2}\cdots\sigma_m(a_m)$ *contains all letters of* $\mathcal{A}$ *as proper factors. Let* z *be the other bispecial factor of the same age as* w *if it exists. Then the history* $\sigma_0\sigma_1\cdots\sigma_n$ *of* w *determines the left valence, multiplicity and extension type of* w *and* z *according to the following table.*

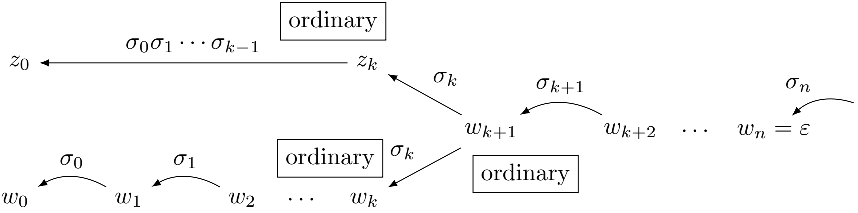
$\sigma_0\sigma_1\cdots\sigma_n \in$	$d^-(w)$	$m(w)$	*ordinary*	$d^-(z)$	$m(z)$	*ordinary*
$\mathcal{S}_\alpha^*\mathcal{S}_\alpha$	3	0	yes			
$\mathcal{S}_\alpha^*\mathcal{S}_\pi$	3	0	no			
$\mathcal{S}^*\,\pi_{jk}\,\mathcal{S}_\alpha^*\{\alpha_k\}$	2	0	yes			
$\mathcal{S}^*\,\pi_{jk}\,\mathcal{S}_\alpha^*\{\alpha_i,\alpha_j\}$	2	0	yes	2	0	yes
$\mathcal{S}^*\,\pi_{jk}\,\mathcal{S}_\alpha^*\{\pi_{ji},\pi_{ki},\pi_{ij},\pi_{kj}\}$	2	0	yes	2	0	yes
$\mathcal{S}^*\,\pi_{jk}\,\mathcal{S}_\alpha^*\{\pi_{ik},\pi_{jk}\}$	2	$+1$	no	2	-1	no

Typical life of ordinary bispecial factors

If $\sigma_0\sigma_1\cdots\sigma_n \in \mathcal{S}_\alpha^*\{\alpha_k\} \cup \mathcal{S}^*\{\pi_{ij}, \pi_{kj}, \pi_{ji}, \pi_{ki}\}\mathcal{S}_\alpha^*\{\alpha_k\}$:

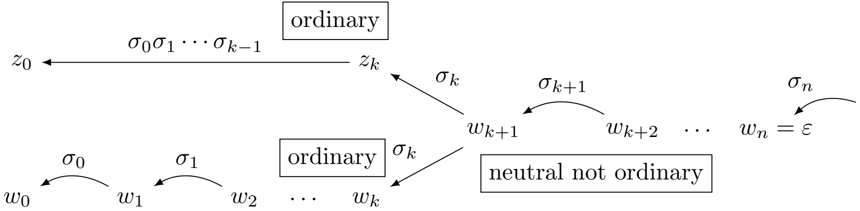


If $\sigma_0\sigma_1\cdots\sigma_n \in \mathcal{S}^*\{\pi_{ik}, \pi_{jk}\}\mathcal{S}_\alpha^*\{\alpha_k\}$:

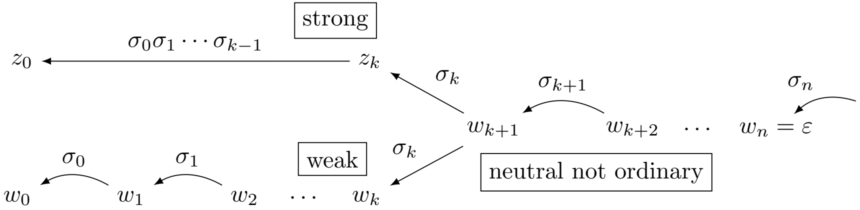


Typical life of neutral not ordinary bispecial factors

If $\sigma_0\sigma_1\cdots\sigma_n \in \mathcal{S}^*\{\pi_{ij},\pi_{kj},\pi_{ji},\pi_{ki}\}\mathcal{S}_\alpha^*\{\pi_{jk}\}$:

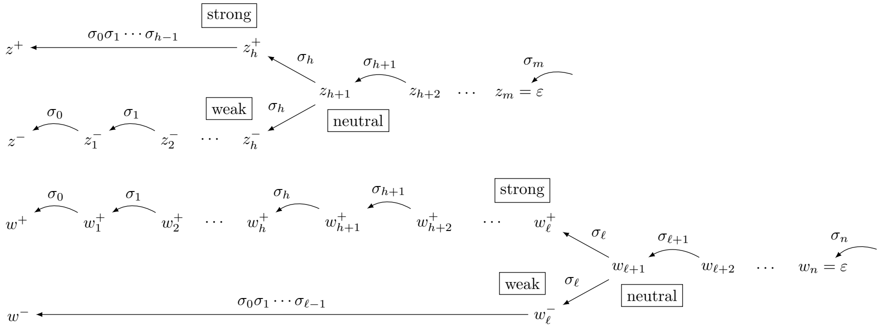


If $\sigma_0\sigma_1\cdots\sigma_n \in \mathcal{S}^*\{\pi_{ik},\pi_{jk}\}\mathcal{S}_\alpha^*\{\pi_{jk}\}$:



Idea of the proof on complexity

The proof consists in a complete characterization of the life of bispecial factors (strong, neutral and weak), not all of them being ordinary.
Lifes of two pairs of strong and weak bispecial factors : z^+, z^- and w^+, w^-.



Plan

Questions

- Compute, if it exists, the invariant measure associated to the **Arnoux-Rauzy-Poincaré** algorithm.
- Find under which conditions the **Brun** and **Arnoux-Rauzy-Poincaré** algorithms lead to bounded balance sequences.
- Compute the factor complexity of S-adic sequences obtain using **Brun** algorithm.
- Study ergodic properties of those previous fusion algorithms.