Symbolic discrepancy

V. Berthé

LIAFA-CNRS-Paris-France
berthe@liafa.univ-paris-diderot.fr
http ://www.liafa.univ-paris-diderot.fr/~berthe

DynA3S-Février 2014

Some combinatorial definitions

We are given an infinite word on $\{1, 2, \cdots, d\}^{\mathbb{N}}$

- Factor complexity
- Frequencies
- Discrepancy

Some combinatorial definitions

We are given an infinite word on $\{1, 2, \cdots, d\}^{\mathbb{N}}$
Symbolic dynamics

- Factor complexity
- Frequencies
- Discrepancy

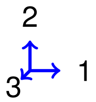
Discrete geometry We can associate with it a discrete line/path in $\mathbb{R}^d$ with vertices in $\mathbb{Z}^d$: replace letters by canonical vectors

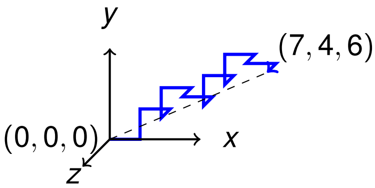
- Number of local configurations
- Frequencies of local configurations, slope of the line
- Distance to the line

A discrete segment associated with the word u

- Let $l : A^* \to \mathbb{N}^d,\ w \mapsto {}^t(|w|_1, \ldots, |w|_d)$ stand for the Parikh mapping
- One associates with any word a discrete line with set of vertices equal to $\{l(u_0 \cdots u_{n-1}) \mid n \in \mathbb{N}\}$

$u = 12132131321321313$





First combinatorial definitions

Factor complexity number of factors of a given length

The frequency f_i of a letter $i \in \mathcal{A}$ in $u = (u_n)_{n\in\mathbb{N}}$ is defined as the following limit, if it exists

$$f_i = \lim_{n\to\infty} \frac{|u_0 \cdots u_{n-1}|_i}{n}$$

$\rightsquigarrow$ Frequency of a word

The Fibonacci word

Fibonacci word $\sigma : a \mapsto ab,\ b \mapsto a$ σ is called a substitution

a
ab
aba
$abaab$
$abaababa$

$\sigma^\infty(a) = abaababaabaababaababaabaababaabaababaabab\ldots$

There are $n + 1$ factors of length n

The Fibonacci word

Fibonacci word $\sigma : a \mapsto ab,\ b \mapsto a$ σ is called a substitution

a
ab
aba
$abaab$
$abaababa$

$$\sigma^{\infty}(a) = abaababaabaababaababaabaababaabaababaabab\ldots$$

There are $n+1$ factors of length n

Frequencies exist

The Fibonacci word

Fibonacci word $\sigma : a \mapsto ab,\ b \mapsto a$ σ is called a substitution

$$a$$
$$ab$$
$$aba$$
$$abaab$$
$$abaababa$$

$$\sigma^\infty(a) = abaababaabaababaababaabaababaabaababaabab\ldots$$

There are $n+1$ factors of length n

Frequencies exist

Frequencies exist for all the infinite words with factor complexity $n+1$

$\limsup p_n/n < 3$ implies unique ergodicity [Boshernitzan]

Symbolic dynamical system

Let $u = (u_n)$ be an infinite word with values in the finite set $\mathcal{A}$

Let S be the shift

$$S((u_n)_n) = (u_{n+1})_n$$

The symbolic dynamical system generated by u is (X_u, S)

$$X_u := \overline{\{S^n(u);\ n \in \mathbb{N}\}} \subset \mathcal{A}^{\mathbb{N}}$$

This is the set of infinite words whose language is included in the language of u

set of factors $=$ language

Word combinatorics vs. symbolic dynamics

Let $u \in \mathcal{A}^{\mathbb{N}}$ be an infinite word

- Word combinatorics

  Study of the number of factors of a given length (factor complexity), frequencies, powers

- Symbolic dynamics Let

$$X_u := \overline{\{S^n u \mid n \in \mathbb{N}\}}$$

$$S((u_n)_n) = (u_{n+1})_n$$

(X_u, S) is a symbolic dynamical system

Compacity, study of invariant measures, recurrence properties, finding geometric representations

Discrepancy of a sequence

Let $(u_n)_n$ be a sequence with values in $[0, 1]$

$$\Delta_N = \limsup_{I \text{ interval}} |\{\mathrm{Card}\,\{0 \leq n \leq N; u_n \in I\} - N\mu(I)|$$

Symbolic discrepancy

Take a word $(u_n)_n$ with values in a finite alphabet $\mathcal{A}$

The frequency f_i of a letter i in $u = (u_n)_{n\in\mathbb{N}}$ is defined as the following limit, if it exists

$$f_i = \lim_{n\to\infty} \frac{|u_0 \cdots u_{N-1}|_i}{N}$$

where $|x|_j$ stands for the number of occurrences of the letter j in the factor x

Symbolic discrepancy

Take a word $(u_n)_n$ with values in a finite alphabet $\mathcal{A}$

The frequency f_i of a letter i in $u = (u_n)_{n\in\mathbb{N}}$ is defined as the following limit, if it exists

$$f_i = \lim_{n\to\infty} \frac{|u_0 \cdots u_{N-1}|_i}{N}$$

where $|x|_j$ stands for the number of occurrences of the letter j in the factor x

Assume that each letter i has frequency f_i in u

Symbolic discrepancy

$$\Delta_N = \max_{i\in\mathcal{A}} ||u_0 u_1 \ldots u_{N-1}|_i - N \cdot f_i|$$

Balancedness

An infinite word $u \in \mathcal{A}^{\mathbb{N}}$ is said to be finitely balanced if there exists a constant $C > 0$ such that for any pair of factors of the same length v, w of u, and for any letter $i \in \mathcal{A}$,

$$||v|_i - |w|_i| \leq C$$

Balancedness

An infinite word $u \in \mathcal{A}^{\mathbb{N}}$ is said to be finitely balanced if there exists a constant $C > 0$ such that for any pair of factors of the same length v, w of u, and for any letter $i \in \mathcal{A}$,

$$||v|_i - |w|_i| \leq C$$

Fibonacci word $\sigma : a \mapsto ab,\ b \mapsto a$ σ is called a substitution

a
ab
aba
abaab
abaababa

$$\sigma^{\infty}(a) = abaababaabaababaababaabaababaabaababaabab\ldots$$

The factors of length 5 contain 3 or 4 *a*'s

Symbolic discrepancies

$$X_u := \overline{\{S^n(u);\ n \in \mathbb{N}\}} \subset \mathcal{A}^{\mathbb{N}} \quad \text{minimal}$$

We assume X_u minimal : $\emptyset$ and X_u are the only closed shift-invariant subsets of X_u

$\rightsquigarrow$ Every infinite word $v \in X_u$ has the same language as u

Symbolic discrepancies

$$X_u := \overline{\{S^n(u);\ n \in \mathbb{N}\}} \subset \mathcal{A}^{\mathbb{N}} \quad \text{minimal}$$

$$\Delta_N = \max_{i \in \mathcal{A}} ||u_0 u_1 \ldots u_{N-1}|_i - N \cdot f_i|$$

$$\begin{aligned} \widetilde{\Delta}_N &= \max\nolimits_{i \in \mathcal{A},\ k} ||u_k \cdots u_{k+N-1}|_i - N \cdot f_i| \\ &= \max\nolimits_{i \in \mathcal{A},\ w \in L_N(u)} ||w|_i - N \cdot f_i| \\ &= \max\nolimits_{i \in \mathcal{A},\ v \in X_u} ||v_0 u_1 \ldots v_{N-1}|_i - N \cdot f_i| \end{aligned}$$

$L_N(u)$ is the set of factors of u of length N

Symbolic discrepancies

$$X_u := \overline{\{S^n(u);\ n \in \mathbb{N}\}} \subset \mathcal{A}^{\mathbb{N}} \quad \text{minimal}$$

$$\Delta_N = \max_{i \in \mathcal{A}} \left| |u_0 u_1 \ldots u_{N-1}|_i - N \cdot f_i \right|$$

$$\begin{aligned} \widetilde{\Delta}_N = & \quad \max_{i \in \mathcal{A},\ k} \left| |u_k \cdots u_{k+N-1}|_i - N \cdot f_i \right| \\ = & \quad \max_{i \in \mathcal{A},\ w \in L_N(u)} \left| |w|_i - N \cdot f_i \right| \\ = & \quad \max_{i \in \mathcal{A},\ v \in X_u} \left| |v_0 u_1 \ldots v_{N-1}|_i - N \cdot f_i \right| \end{aligned}$$

$L_N(u)$ is the set of factors of u of length N

We can also consider factors w and not only letters

Remark [B. Adamczewski] There exists an infinite word $u \in \{0,1\}^{\mathbb{N}}$ such that

- u has has a frequency vector
- $\Delta_N = O(f(N))$ with $f(N) = o(N)$
- for every integer N, $\widetilde{\Delta}_N = O(N)$

Remark [B. Adamczewski] There exists an infinite word $u \in \{0,1\}^{\mathbb{N}}$ such that

- u has has a frequency vector
- $\Delta_N = O(f(N))$ with $f(N) = o(N)$
- for every integer N, $\widetilde{\Delta}_N = O(N)$

Take

$$u = 01\,0^{[f(1)]}1^{[f(1)]}\,0101\,0^{[f(2)]}1^{[f(2)]}\,\cdots(01)^n\,0^{[f(n)]}1^{[f(n)]}$$

$$||u_0\cdots u_{N-1}|_i - N/2| \leq 1/2f(N)$$

Equidistribution vs. well-equidistribution

Let u be an infinite word with values in the finite alphabet $\mathcal{A}$

$$\widetilde{\Delta}_N = \limsup_{i \in \mathcal{A},\ k} ||u_k \cdots u_{k+N-1}|_i - N \cdot f_i|$$

u is well-distributed with respect to letters if $\widetilde{\Delta}_N = o(N)$
$\rightsquigarrow$ uniformly in k

The frequency of a factor w in u is defined as the limit when n tends towards infinity, if it exists, of the number of occurrences of w in $u_0 u_1 \cdots u_{n-1}$ divided by n

Equidistribution vs. well-equidistribution

Let u be an infinite word with values in the finite alphabet $\mathcal{A}$

$$\widetilde{\Delta}_N = \limsup_{i \in \mathcal{A},\ k} \left| |u_k \cdots u_{k+N-1}|_i - N \cdot f_i \right|$$

u is well-distributed with respect to letters if $\widetilde{\Delta}_N = o(N)$
$\rightsquigarrow$ uniformly in k

The frequency of a factor w in u is defined as the limit when n tends towards infinity, if it exists, of the number of occurrences of w in $u_0 u_1 \cdots u_{n-1}$ divided by n

The infinite word u has uniform factor frequencies if, for every factor w of u, the number of occurrences of w in $u_k \cdots u_{k+n-1}$ divided by n has a limit when n tends to infinity, uniformly in k

Balance and equidistribution

An infinite word $u \in \mathcal{A}^{\mathbb{N}}$ is finitely balanced if and only if

- it has uniform letter frequencies
- there exists a constant B such that for any factor w of u, we have $||w|_i - f_i|w|| \leq B$ for all letter i in $\mathcal{A}$

where f_i is the frequency of i

Balance and equidistribution

An infinite word $u \in \mathcal{A}^{\mathbb{N}}$ is finitely balanced if and only if

- it has uniform letter frequencies
- there exists a constant B such that for any factor w of u, we have $||w|_i - f_i|w|| \leq B$ for all letter i in $\mathcal{A}$

where f_i is the frequency of i

Proof

Let u be an infinite word with letter frequency vector f and such that $||w|_i - f_i|w|| \leq B$ for every factor w and all letters i in $\mathcal{A}$

Balance and equidistribution

An infinite word $u \in \mathcal{A}^{\mathbb{N}}$ is finitely balanced if and only if

- it has uniform letter frequencies
- there exists a constant B such that for any factor w of u, we have $||w|_i - f_i|w|| \leq B$ for all letter i in $\mathcal{A}$

where f_i is the frequency of i

Proof

Let u be an infinite word with letter frequency vector f and such that $||w|_i - f_i|w|| \leq B$ for every factor w and all letters i in $\mathcal{A}$

For every pair of factors w_1 and w_2 with the same length n, we have

$$||w_1|_i - |w_2|_i| \leq ||w_1|_i - nf_i| + ||w_2|_i - nf_i| \leq 2B$$

Hence u is $2B$-balanced

Finite balancedness implies the existence of uniform letter frequencies

Proof Assume that u is C-balanced and fix a letter i

Let N_p be such that for every word of length p of u, the number of occurrences of the letter i belongs to the set

$$\{N_p, N_p + 1, \cdots, N_p + C\}$$

The sequence $(N_p/p)_{p\in\mathbb{N}}$ is a Cauchy sequence. Indeed consider a factor w of length pq

$$pN_q \leq |w|_i \leq pN_q + pC, \quad qN_p \leq |w|_i \leq qN_p + qC.$$

$$-C/p \leq N_p/p - N_q/q \leq C/q$$

Finite balancedness implies the existence of uniform letter frequencies

Proof Assume that u is C-balanced and fix a letter i

Let N_p be such that for every word of length p of u, the number of occurrences of the letter i belongs to the set

$$\{N_p, N_p + 1, \cdots, N_p + C\}$$

The sequence $(N_p/p)_{p\in\mathbb{N}}$ is a Cauchy sequence. Indeed consider a factor w of length pq

$$pN_q \leq |w|_i \leq pN_q + pC, \qquad qN_p \leq |w|_i \leq qN_p + qC.$$

$$-C/p \leq N_p/p - N_q/q \leq C/q$$

Let $f_i = \lim N_q/q$

$$-C \leq N_p - pf_i \leq 0 \quad (q \to \infty)$$

Then, for any factor w

$$\left| \frac{|w|_i}{|w|} - f_i \right| \leq \frac{C}{|w|} \qquad \rightsquigarrow \text{uniform frequencies}$$

Frequencies and measures

$$X_u := \overline{\{S^n(u);\ n \in \mathbb{N}\}} \subset \mathcal{A}^{\mathbb{N}}$$

- Having frequencies is a property of the infinite word u while having uniform frequencies is a property of the associated language or shift X_u

Frequencies and measures

$$X_u := \overline{\{S^n(u);\ n \in \mathbb{N}\}} \subset \mathcal{A}^{\mathbb{N}}$$

- A probability measure μ on X_u is said invariant if $\mu(S^{-1}A) = \mu(A)$ for all measurable subset $A \subset X$
- An invariant probability measure on a shift X is said ergodic if any shift-invariant measurable set has either measure 0 or 1
- The property of uniform frequency of factors for a shift X is equivalent to unique ergodicity : there exists a unique shift-invariant probability measure on X

Frequencies and measures

$$X_u := \overline{\{S^n(u);\ n \in \mathbb{N}\}} \subset \mathcal{A}^{\mathbb{N}}$$

- Having frequencies is a property of the infinite word u while having uniform frequencies is a property of the associated language or shift X_u
- Balancedness is a property of the associated shift and may be thought as a strong form of unique ergodicity

Birkhoff sums

Let μ is an ergodic measure on X_u. The Birkhoff Ergodic theorem says that for μ-a.e. x and for $f \in L_1(X_u, \mathbb{R})$

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \int f d\mu$$

The mean behaviour along an orbit=
the mean value of f with respect to μ

μ-almost every infinite word in X_u has frequency $\mu[w]$

$$[w] = \{u \in X; u_0 \dots u_{n-1} = w\}$$

but this frequency is not necessarily uniform

If X_u is uniquely ergodic, the unique invariant measure on X_u is ergodic and the convergence is uniform for all words in X_u

Motivations

- Discrete lines
- Bounded remainder sets and symbolic codings of Kronecker sequences
- The chairman assignment problem

Discrete lines in discrete geometry

The discrepancy of the word $u = (u_n)$ is defined as

$$\max_{i \in \mathcal{A},\ n} ||u_0 \cdots u_{n-1}|_i - f_i \cdot n|$$

We measure the distance to the vector directed by the frequencies

Discrete lines in discrete geometry

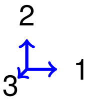
The discrepancy of the word $u = (u_n)$ is defined as

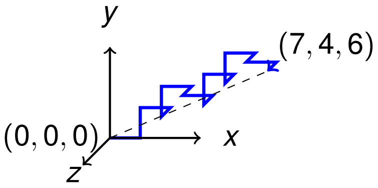
$$\max_{i \in \mathcal{A},\ n} \left| |u_0 \cdots u_{n-1}|_i - f_i \cdot n \right|$$

We measure the distance to the vector directed by the frequencies

$\rightsquigarrow$ If the discrepancy is bounded, the word u can be considered as a discretization of the vector line directed by the letter frequency vector $(f_1, \cdots, f_d)$

$u = 121321313213213 13$



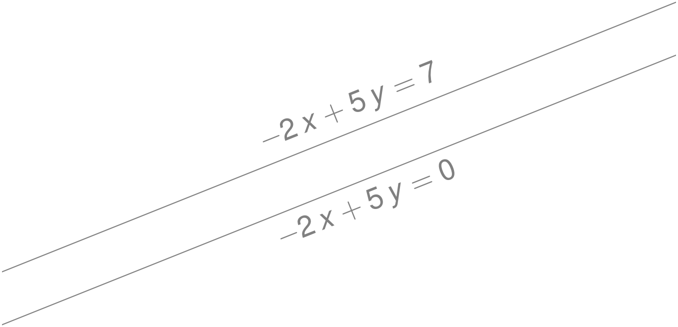


2D Discrete Lines and Sturmian words

$$0 < -2x + 5y \leq 7$$

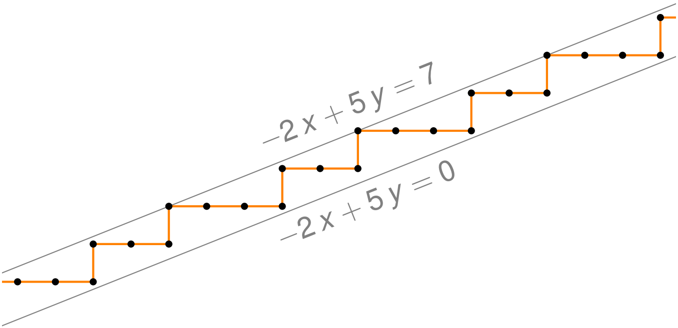
2D Discrete Lines and Sturmian words

$$0 < -2\,x + 5\,y \leq 7$$



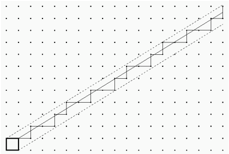
2D Discrete Lines and Sturmian words

$$0 < -2\,x + 5\,y \leq 7$$



Sturmian words

$$\forall n \in \mathbb{N},\ u_n = 0 \Longleftrightarrow R_\alpha^n(x) = n\alpha + x \in [0, 1-\alpha) \pmod 1$$



$\rightsquigarrow$ Diophantine approximation

- Sturmian words are known to be 1-balanced
- They thus have a bounded discrepancy Δ_N and $\widetilde{\Delta}_N$
- They even are exactly the 1-balanced infinite words that are not eventually periodic

This gives a combinatorial characterization of natural codings of Kronecker sequences (rotations on the unit circle) [Morse-Hedlund'42]

Bounded remainder sets and Kronecker sequences

Let $\alpha = (\alpha_1, \dots, \alpha_d) \in [0,1]^d$

with $1, \alpha_1, \cdots, \alpha_d$ $\mathbb{Q}$-linearly independent

We consider the Kronecker sequence

$$(\{n\alpha_1\}, \dots, \{n\alpha_d\})_n$$

Bounded remainder sets and Kronecker sequences

We consider the Kronecker sequence

$$(\{n\alpha_1\}, \ldots, \{n\alpha_d\})_n$$

associated with the translation over $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$

$$R_\alpha \colon \mathbb{T}^d \mapsto \mathbb{T}^d, \; x \mapsto x + \alpha$$

Bounded remainder sets and Kronecker sequences

We consider the Kronecker sequence

$$(\{n\alpha_1\}, \ldots, \{n\alpha_d\})_n$$

associated with the translation over $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$

$$R_\alpha \colon \mathbb{T}^d \mapsto \mathbb{T}^d, \ x \mapsto x + \alpha$$

Discrepancy

$$\Delta_N = sup_{B \text{ box}} \left| \text{Card} \left\{ 0 \leq n \leq N; R_\alpha^n(0) \in B \right\} - N \cdot \mu(B) \right|$$

Bounded remainder sets and Kronecker sequences

We consider the Kronecker sequence

$$(\{n\alpha_1\}, \dots, \{n\alpha_d\})_n$$

associated with the translation over $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$

$$R_\alpha \colon \mathbb{T}^d \mapsto \mathbb{T}^d,\ x \mapsto x + \alpha$$

Discrepancy

$$\Delta_N = sup_{B \text{ box}} \left|\text{Card}\,\{0 \leq n \leq N; R_\alpha^n(0) \in B\} - N \cdot \mu(B)\right|$$

Bounded remainder set X for which there exists $C > 0$ s.t. for all N

$$|\text{Card}\{0 \leq n \leq N; R_\alpha^n(0) \in X\} - N\mu(X)| \leq C$$

Bounded remainder sets

Case $d = 1$

Theorem [Kesten'66] Intervals that are bounded remainder sets are the intervals with length in $\mathbb{Z} + \alpha\mathbb{Z}$

Sturmian words are finitely balanced

Bounded remainder sets

Case $d = 1$

Theorem [Kesten'66] Intervals that are bounded remainder sets are the intervals with length in $\mathbb{Z} + \alpha\mathbb{Z}$

Sturmian words are finitely balanced

General dimension d

Theorem [Liardet'87] There are no nontrivial boxes that are bounded remainder sets

Bounded remainder sets

Case $d = 1$

Theorem [Kesten'66] Intervals that are bounded remainder sets are the intervals with length in $\mathbb{Z} + \alpha\mathbb{Z}$

Sturmian words are finitely balanced

General dimension d

Theorem [Liardet'87] There are no nontrivial boxes that are bounded remainder sets

Boxes are not bounded remainder sets

How well can one approximate a box by bounded remainder sets ?

A symbolic approach

We consider a partition $\{X_1, \cdots, X_k\}$ of $\mathbb{T}^d$

$$\mathbb{T}^d = \bigcup_{1 \le i \le k} X_i, \quad \mu(X_i \cap X_j) = 0, \text{ for all } i \neq j$$

We code the trajectory of x under the action of $R_\alpha \colon x \mapsto x + \alpha$ as follows

$$x \rightsquigarrow (u_n)_n \in \{1, 2, \ldots, k\}^{\mathbb{N}}$$

$$u_n = i \quad \text{if and only if} \quad R_\alpha^n(x) = x + n\alpha \in X_i$$

A symbolic approach

We consider a partition $\{X_1, \cdots, X_k\}$ of $\mathbb{T}^d$

$$\mathbb{T}^d = \bigcup_{1 \leq i \leq k} X_i, \quad \mu(X_i \cap X_j) = 0, \text{ for all } i \neq j$$

We code the trajectory of x under the action of $R_\alpha : x \mapsto x + \alpha$ as follows

$$x \rightsquigarrow (u_n)_n \in \{1, 2, \ldots, k\}^{\mathbb{N}}$$

$$u_n = i \quad \text{if and only if} \quad R_\alpha^n(x) = x + n\alpha \in X_i$$

Questions Which information on R_α can we get from the combinatorial properties of the sequence (u_n) ? What is a good coding ?

Example

Fibonacci substitution $\sigma : 1 \mapsto 12,\ 2 \mapsto 1$

$$u = \sigma^{\infty}(1) = 121121211211212\cdots$$

(X_u, S) is isomorphic to $(\mathbb{R}/\mathbb{Z}, R_{\frac{1+\sqrt{5}}{2}})$ where

$$R_{\frac{1+\sqrt{5}}{2}} : x \mapsto x + \frac{1+\sqrt{5}}{2}$$

Natural coding of a two-interval exchange

From letters to words

One wants to find good partitions for toral translations which provide natural codings that have bounded discrepancy for every length of factor

A translation on $\mathbb{T}^2$ is a map $R_\alpha : \mathbb{R}^2/L \to \mathbb{R}^2/L$, $x \mapsto x + \alpha \pmod{L}$, where $\alpha \in \mathbb{R}^2$ and where L is a lattice in $\mathbb{R}^2$

A coding u of $(R_\alpha, \mathbb{T}^2)$ is a natural coding if there exists a fundamental domain for a lattice L in $\mathbb{R}^2$ together with a finite partition of this domain such that on each element of the partition the map R_α is a translation by a vector

A symbolic measure-theoretical dynamical system (Ω, S) is a natural symbolic coding of (X, T) if every element of Ω is a coding of the orbit of some point of X, and if (Ω, S) and (X, T) are semi-topologically conjugate

The chairman assignment problem

The chairman assignment problem [R. Tijdeman] “Suppose k states form a union and every year a union chairman has to be selected in such a way that at any time the accumulated number of chairmen from each state is proportional to its weight.”

How to get in an effective way an assignment with small discrepancy ?

The chairman assignment problem

The chairman assignment problem [R. Tijdeman] “Suppose k states form a union and every year a union chairman has to be selected in such a way that at any time the accumulated number of chairmen from each state is proportional to its weight.”

How to get in an effective way an assignment with small discrepancy ?

The discrepancy of the word $u = (u_n)$ is defined as

$$D(u) = \max_{i \in \mathcal{A},\ n} ||u_0 \cdots u_{n-1}|_i - f_i \cdot n|$$

The chairman assignment problem

The chairman assignment problem [R. Tijdeman] “Suppose k states form a union and every year a union chairman has to be selected in such a way that at any time the accumulated number of chairmen from each state is proportional to its weight.”

How to get in an effective way an assignment with small discrepancy ?

Theorem [Meijer-Tijdeman]

$$\sup_{\mathbf{f}} \inf_{u} D(u) = 1 - \frac{1}{2d-2}$$

Remark $1 - 1/(2d-2) = 3/4$ for $d = 3$

R. Tijdeman has given an algorithmic way, given **f**, to construct a sequence u with $D(u) \leq 1 - \frac{1}{2d-2}$

See also [M.L. Balinski and H.P. Young]

A word formulation

Problem

Let $(f_1, \cdots, f_d) \in [0,1]^d$ such that $\sum_{i=1}^{d} f_i = 1$

How to construct a word u over the alphabet $\{1, 2, \cdots, d\}$ satisfying the following conditions

- u has linear complexity function
- u is uniformly balanced
- the letter frequencies in u are given by $(f_1, \cdots, f_d)$

Problem

Let $(f_1, \cdots, f_d) \in [0,1]^d$ such that $\sum_{i=1}^d f_i = 1$

How to construct a word u over the alphabet $\{1, 2, \cdots, d\}$ satisfying the following conditions

- u has linear complexity function
- u is uniformly balanced on factors
- the letter frequencies in u are given by $(f_1, \cdots, f_d)$

On fabrique une trajectoire qui reste à distance bornée de la droite

- Mots de billard (dans le cube)
- Mots de poursuite (Chevallier)

- Equilibre sur les lettres mais sur les mots ?
- Complexité quadratique

The algebraic case : substitutive dynamics

The Dumont-Thomas numeration system

- It is based on the greedy algorithm and acts on words
- Let $u = (u_n)$ such that $\sigma(u) = u$
- We decompose prefixes of $u_0 \cdots u_{N-1}$ into images by powers of σ of a finite number of words $\rightsquigarrow$ base σ
- Since $\sigma(u) = u$, there exists L such that

$$\sigma(u_0 \cdots u_{L-1}) \leq u_0 \cdots u_{N-1} < \sigma(u_0 \cdots u_L)$$

and thus a proper prefix p of $\sigma(u_L)$ s.t.

$$u_0 \cdots u_{N-1} = \sigma(u_0 \cdots u_{L-1})\, p \text{ with } \sigma(u_L) = p\, u_N\, s$$

- Hence, for every N, one has

$$u_0 \cdots u_{N-1} = \sigma^K(p_K)\sigma^{K-1}(p_{K-1}) \cdots \sigma(p_1)p_0,$$

the p_i belong to a finite set of words that only depends on σ $\rightsquigarrow$ digits

$\rightsquigarrow$ a numeration system on words... but also for integers and real numbers

The Dumont-Thomas numeration system

Every prefix w of the Tribonacci word $u = \sigma^\infty(1)$

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

can be uniquely expanded as

$$w = \sigma^n(p_n)\sigma^{n-1}(p_{n-1}) \cdots p_0,$$

where the words p_i are equal to the empy word or to the letter 1, and 111 $\nexists$.

Conversely every finite word that can be decomposed under this form is a prefix of the Tribonacci word.

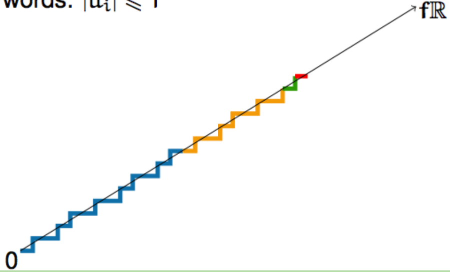
$$|w| = \sum_{i=0}^{n} \varepsilon_i T_i.$$

Such a numeration exists for every primitive substitution

- **Dumont-Thomas representation:** For every prefix of ω:

$$p = \phi_1\phi_2\ldots\phi_n(u_n)\cdot\phi_1\phi_2\ldots\phi_{n-1}(u_{n-1})\cdot\ \cdots\ \cdot\phi_1(u_1)\cdot u_0$$

- Brun/AR words: $|u_i| \leqslant 1$



Slide by T. Hejda

A substitution on words : the Tribonacci substitution

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

$$\sigma^{\infty}(1) = 1213121121312121 3 \cdots$$

The incidence matrix M_σ of σ is defined by

$$M_\sigma = (|\sigma(j)|_i)_{(i,j)\in\mathcal{A}^2},$$

where $|\sigma(j)|_i$ counts the number of occurrences of the letter i in $\sigma(j)$

The matrix M_σ has nonnegative entries $\rightsquigarrow$ Perron-Frobenius theory

Its incidence matrix is $M_\sigma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

Its characteristic polynomial is $X^3 - X^2 - X - 1$. Its Perron-Frobenius eigenvalue $\beta > 1$ is a Pisot number

It is primitive : there exists a power of M_σ which contains only positive entries

Pisot substitution

Pisot-Vijayaraghavan number An algebraic integer is a Pisot number if its algebraic conjugates λ (except itself) satisfy

$$|\lambda| < 1$$

Let σ be a substitution over the alphabet $\mathcal{A}$

Pisot irreduciblesubstitution σ is primitive, its Perron–Frobenius eigenvalue is a Pisot number and the characteristic polynomial of its incidence matrix is irreducible

Pisot substitutions have bounded discrepancy

Let σ be a Pisot irreducible substitution and $u = \sigma^\infty(1)$

$$l : A^* \to \mathbb{N}^d,\ w \mapsto {}^t(|w|_1, \ldots, |w|_d)$$

Fact The vectors $l(u_0 u_1 \ldots u_n)$ stay within bounded distance of the expanding (=the direction given by the vector of frequencies)

Pisot substitutions have bounded discrepancy

Let σ be a Pisot irreducible substitution and $u = \sigma^\infty(1)$
Fact The vectors $l(u_0 u_1 \ldots u_n)$ stay within bounded distance of the expanding (=the direction given by the vector of frequencies)
Proof Let (v_i) be a basis of eigenvectors associated with the eigenvalues λ_i, with $\lambda_1 > 1 \geq |\lambda_i|$
By Perron-Frobenius Theorem, v_1 has nonzero entries
Write

$$e_1 = \sum a_i v_i$$

We have

$$f(\sigma^k(1)) = M_\sigma^k e_1 = a_1 \lambda_1^k v_1 + \sum a_j \lambda_j^k v_j$$

The vectors $f(\sigma^k(1))$ converge exponentially fast to the expanding line, and their projections on the contracting plane converge to 0

Pisot substitutions have bounded discrepancy

Let σ be a Pisot irreducible substitution and $u = \sigma^\infty(1)$

Fact The vectors $l(u_0 u_1 \ldots u_n)$ stay within bounded distance of the expanding (=the direction given by the vector of frequencies)

Proof We have

$$f(\sigma^k(1)) = M_\sigma^k e_1 = a_1 \lambda_1^k v_1 + \sum a_j \lambda_j^k v_j$$

The vectors $f(\sigma^k(1))$ converge exponentially fast to the expanding line, and their projections on the contracting plane converge to 0

Dumont-Thomas numeration Any prefix w of u can be expanded as

$$w = \sigma^k(p_k)\sigma^{k-1}(p_{k-1}) \ldots p_0,$$

where the p_i belong to a finite set of words

Substitutive words [B. Adamczewski]

Let σ be a primitive substitution
According to the Perron-Frobenius Theorem, M_σ admits a dominant eigenvalue θ_1

Let α_2 be the multiplicity of the second eigenvalue θ_2

Substitutive words [B. Adamczewski]

Theorem

- If $\theta_2 < 1$, then the discrepancy is bounded
- If $\theta_2 > 1$, then $D_N = (O \cap \Omega)((\log N)^{\alpha_2 - 1} N^{(\log_{\theta_1} |\theta_2])})$
- If $|\theta_2| = 1$, then $D_N = (O \cap \Omega)((\log N)^{\alpha_2})$ or $D_N = (O \cap \Omega)((\log N)^{\alpha_2 - 1})$
- In particular there exist balanced fixed points of substitutions for which $|\theta_2| = 1$. All eigenvalues of modulus one of the incidence matrix have to be roots of unity

Tribonacci's substitution [Rauzy '82]

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

$$1213121121312121 3 \cdots$$

Question Is it possible to give a geometric representation of the associated substitutive dynamical system X_σ as a translation on an abelian compact group ?

Yes ! (X_σ, S) is isomorphic to a translation on the two-dimensional torus

Question How to produce explicitly a fundamental domain for this translation ?

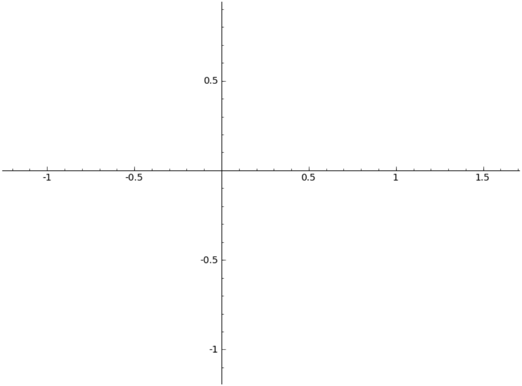
Rauzy fractal G. Rauzy introduced in the 80's a compact set with fractal boundary that tiles the plane which provides a geometric representation of (X_σ, S)
$\rightsquigarrow$Thurston for beta-numeration

The Rauzy fractal as a geometric representation

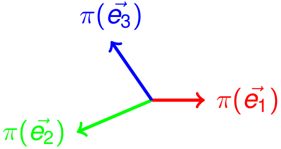
Consider the Tribonacci substitution

$$\sigma : 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131212131211213\cdots$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



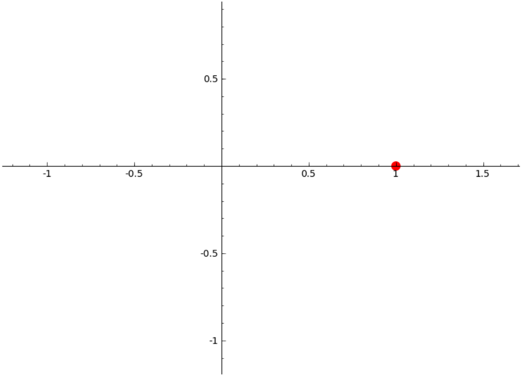
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

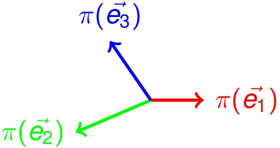
$$\sigma: 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131212131211213\cdots$

$\pi(\vec{e_1})$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



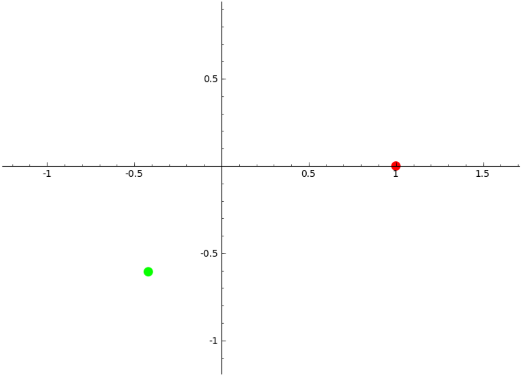
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

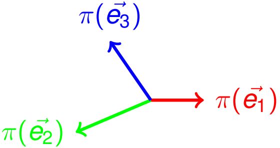
$$\sigma : 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131212131211213\cdots$

$\pi(\vec{e_1} + \vec{e_2})$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



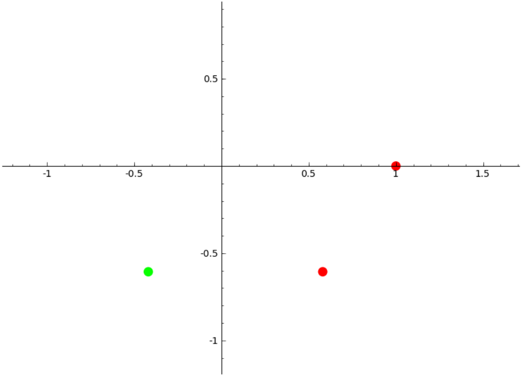
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

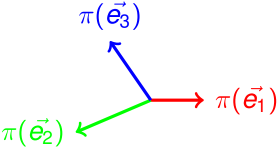
$$\sigma : 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131212131211213\cdots$

$\pi(\vec{e_1} + \vec{e_2} + \vec{e_1})$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



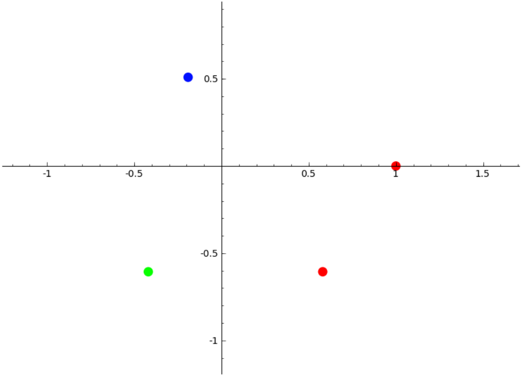
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

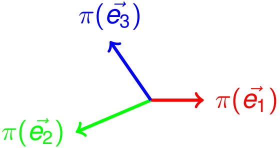
$$\sigma : 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131212131211213\cdots$

$\pi(\vec{e_1} + \vec{e_2} + \vec{e_1} + \vec{e_3})$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



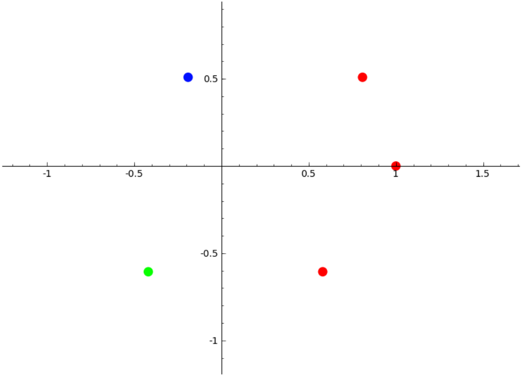
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

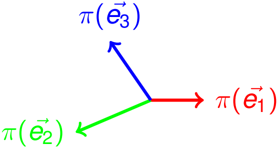
$$\sigma : 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131212131211213\cdots$

$\pi(\vec{e_1} + \vec{e_2} + \vec{e_1} + \vec{e_3} + \vec{e_1})$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



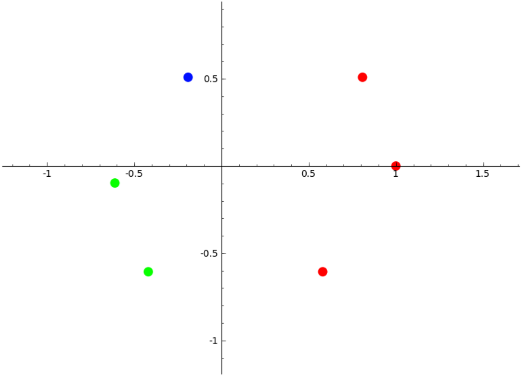
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

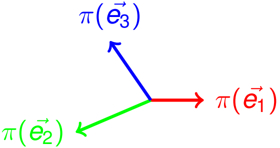
$$\sigma : 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131212131211213\cdots$

$\pi(\vec{e_1} + \vec{e_2} + \vec{e_1} + \vec{e_3} + \vec{e_1} + \vec{e_2})$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



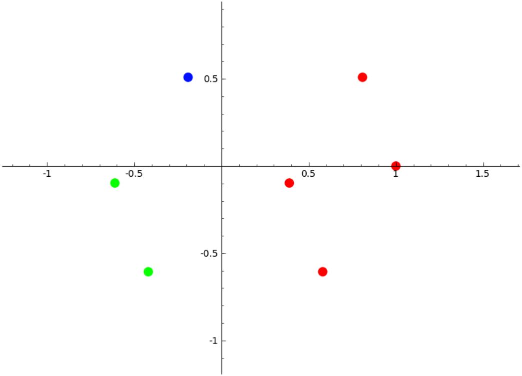
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

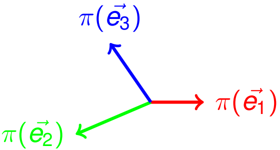
$$\sigma : 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131213121121 3 \cdots$

$\pi(\vec{e_1} + \vec{e_2} + \vec{e_1} + \vec{e_3} + \vec{e_1} + \vec{e_2} + \vec{e_1} + \cdots)$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



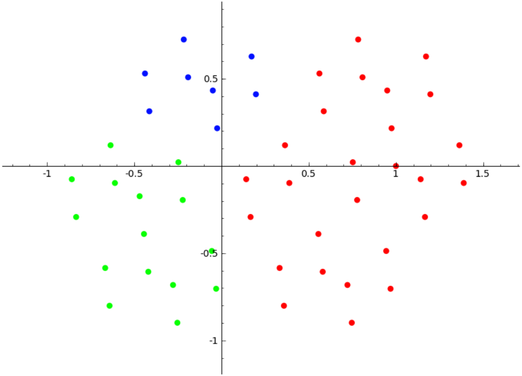
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

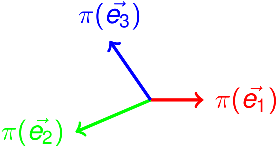
$$\sigma : 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131212131211213 \cdots$

$\pi(\vec{e_1} + \vec{e_2} + \vec{e_1} + \vec{e_3} + \vec{e_1} + \vec{e_2} + \vec{e_1} + \cdots)$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



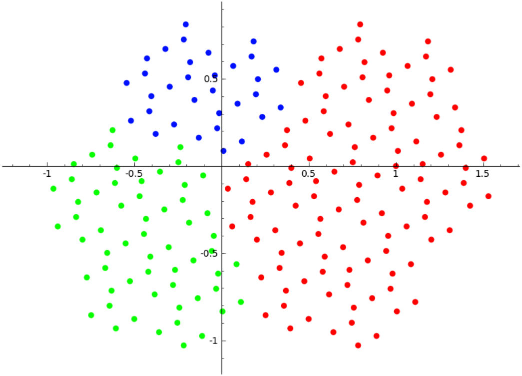
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

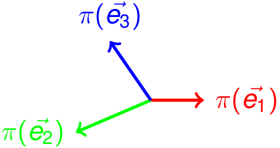
$$\sigma : 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$1213121121312121312112 13 \cdots$

$\pi(\vec{e_1} + \vec{e_2} + \vec{e_1} + \vec{e_3} + \vec{e_1} + \vec{e_2} + \vec{e_1} + \cdots)$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



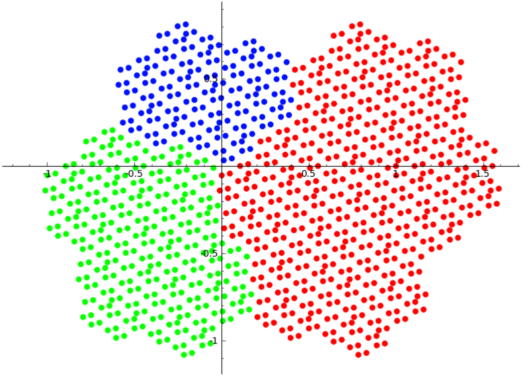
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

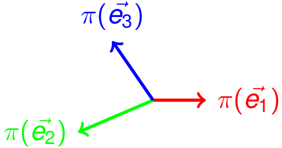
$$\sigma : 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131212131211213\cdots$

$\pi(\vec{e_1} + \vec{e_2} + \vec{e_1} + \vec{e_3} + \vec{e_1} + \vec{e_2} + \vec{e_1} + \cdots)$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



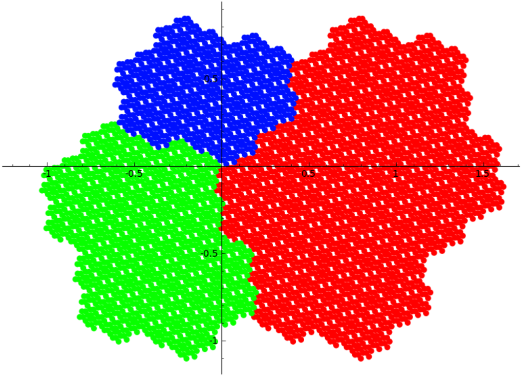
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

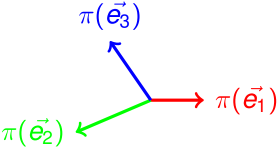
$$\sigma\colon 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131212131211213\cdots$

$\pi(\vec{e_1} + \vec{e_2} + \vec{e_1} + \vec{e_3} + \vec{e_1} + \vec{e_2} + \vec{e_1} + \cdots)$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



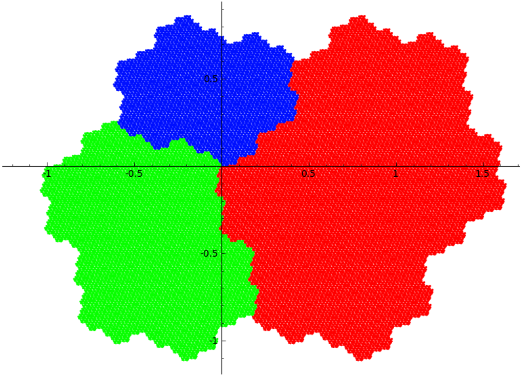
The Rauzy fractal as a geometric representation

Consider the Tribonacci substitution

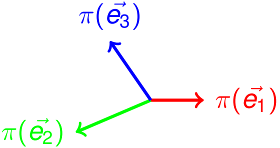
$$\sigma : 1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 1$$

$121312112131212131211213\cdots$

$\pi(\vec{e_1} + \vec{e_2} + \vec{e_1} + \vec{e_3} + \vec{e_1} + \vec{e_2} + \vec{e_1} + \cdots)$



π projection along the expanding eigenline onto the contracting plane of the incidence matrix of M_σ



S-adic expansions

S-adic expansions

Definition An infinite word ω is said *S-adic* if there exist

- a finite set of substitutions $\mathcal{S}$
- an infinite sequence of substitutions $(\sigma_n)_{n\geq 1}$ with values in $\mathcal{S}$

such that

$$\omega = \lim_{n\to+\infty} \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_n(0)$$

An S-adic representation defined by the directive sequence $(\sigma_n)_{n\in\mathbb{N}}$, where $\sigma_n : \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*$, is everywhere growing if for any sequence of letters $(a_n)_n$, one has

$$\lim_{n\to+\infty} |\sigma_0 \cdots \sigma_{n-1}(a_n)| = +\infty.$$

Unique ergodicity

Let X be an S-adic shift with directive sequence $\sigma = (\sigma_n)_n$ with everywhere growing sequence $(\sigma_n)_n$
Denote by $(M_n)_n$ the associated sequence of incidence matrices
If the cone $C^{(0)}$ is one-dimensional

$$C^{(0)} = \bigcap_{n\to\infty} M_{[0,n)}\mathbb{R}_+^d$$

then X has uniform letter frequencies

Unique ergodicity

Let X be an S-adic shift with directive sequence $\sigma = (\sigma_n)_n$ with everywhere growing sequence $(\sigma_n)_n$
Denote by $(M_n)_n$ the associated sequence of incidence matrices
If the cone $C^{(0)}$ is one-dimensional

$$C^{(0)} = \bigcap_{n \to \infty} M_{[0,n)} \mathbb{R}_+^d$$

then X has uniform letter frequencies
If

$$C^{(k)} = \bigcap_{n \to \infty} M_{[k,n)} \mathbb{R}_+^d$$

is one-dimensional, then the S-adic dynamical system (X, S) is uniquely ergodic

Back to the initial problem

For a.e. vector $f = (f_1, \cdots, f_d) \in [0, 1]^d$, one can construct a word u over the alphabet $\{1, 2, \cdots, d\}$ satisfying the following conditions

- u has linear complexity function
- u is uniformly balanced
- the letter frequencies in u are given by $(f_1, \cdots, f_d)$

Back to the initial problem

For a.e. vector $f = (f_1, \cdots, f_d) \in [0, 1]^d$, one can construct a word u over the alphabet $\{1, 2, \cdots, d\}$ satisfying the following conditions

- u has linear complexity function
- u is uniformly balanced
- the letter frequencies in u are given by $(f_1, \cdots, f_d)$

- Linear complexity for S-adic Brun words [Labbé-Leroy]
- Convergence issues : strong convergence ($< -$ second Lyapunov exponent negative)
- Characterization of uniformly balanced sequences

Our strategy

- We apply a multidimensional continued fraction algorithm to the line in $\mathbb{R}^3$ directed by a given vector $\mathbf{u} = (u_1, u_2, u_3)$
- We then associate with the matrices produced by the algorithm substitutions, with these substitutions having the matrices produced by the continued fraction algorithm as incidence matrices

$$\mathbf{u} = \mathbf{u}_0 \xleftarrow{M_1} \mathbf{u}_1 \xleftarrow{M_2} \mathbf{u}_2 \xleftarrow{M_3} \cdots \xleftarrow{M_k} \mathbf{u}_k$$

$$w = w_0 \xleftarrow{\sigma_1} w_1 \xleftarrow{\sigma_2} w_2 \xleftarrow{\sigma_3} \cdots \xleftarrow{\sigma_k} w_k \in \{1, 2, 3\}$$

$$\mathbf{u} = M_1 \cdots M_k \mathbf{u}_k$$

Applying Brun algorithm on $(7, 4, 6)$

$$
(7,4,6) \xleftarrow{\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}} (1,4,6) \xleftarrow{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}} (1,4,2) \xleftarrow{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}} (1,0,2) \xleftarrow{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}} (1,0,0)
$$

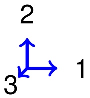
$$
\mathbf{w}_0 \xleftarrow{\begin{array}{ccc} 1 & \mapsto & 1 \\ 2 & \mapsto & 2 \\ 3 & \mapsto & 13 \end{array}} \mathbf{w}_1 \xleftarrow{\begin{array}{ccc} 1 & \mapsto & 1 \\ 2 & \mapsto & 23 \\ 3 & \mapsto & 3 \end{array}} \mathbf{w}_2 \xleftarrow{\begin{array}{ccc} 1 & \mapsto & 1 \\ 2 & \mapsto & 2 \\ 3 & \mapsto & 223 \end{array}} \mathbf{w}_3 \xleftarrow{\begin{array}{ccc} 1 & \mapsto & 133 \\ 2 & \mapsto & 2 \\ 3 & \mapsto & 3 \end{array}} \mathbf{w}_4
$$

Applying Brun algorithm on $(7, 4, 6)$

$$(7,4,6) \xleftarrow{\begin{pmatrix}1&0&1\\0&1&0\\0&0&1\end{pmatrix}} (1,4,6) \xleftarrow{\begin{pmatrix}1&0&0\\0&1&0\\0&1&1\end{pmatrix}} (1,4,2) \xleftarrow{\begin{pmatrix}1&0&0\\0&1&2\\0&0&1\end{pmatrix}} (1,0,2) \xleftarrow{\begin{pmatrix}1&0&0\\0&1&0\\2&0&1\end{pmatrix}} (1,0,0)$$

$$\mathbf{w}_0 \xleftarrow{\begin{array}{rcl}1&\mapsto&1\\2&\mapsto&2\\3&\mapsto&13\end{array}} \mathbf{w}_1 \xleftarrow{\begin{array}{rcl}1&\mapsto&1\\2&\mapsto&23\\3&\mapsto&3\end{array}} \mathbf{w}_2 \xleftarrow{\begin{array}{rcl}1&\mapsto&1\\2&\mapsto&2\\3&\mapsto&223\end{array}} \mathbf{w}_3 \xleftarrow{\begin{array}{rcl}1&\mapsto&133\\2&\mapsto&2\\3&\mapsto&3\end{array}} \mathbf{w}_4$$

$$\mathbf{w} = \mathbf{w}_0 = 1213213132132131 3$$

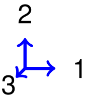


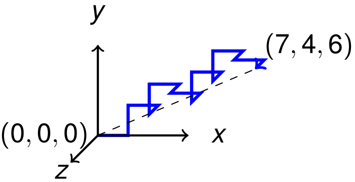
Applying Brun algorithm on $(7,4,6)$

$$
(7,4,6) \xleftarrow{\begin{pmatrix}1&0&1\\0&1&0\\0&0&1\end{pmatrix}} (1,4,6) \xleftarrow{\begin{pmatrix}1&0&0\\0&1&0\\0&1&1\end{pmatrix}} (1,4,2) \xleftarrow{\begin{pmatrix}1&0&0\\0&1&2\\0&0&1\end{pmatrix}} (1,0,2) \xleftarrow{\begin{pmatrix}1&0&0\\0&1&0\\2&0&1\end{pmatrix}} (1,0,0)
$$

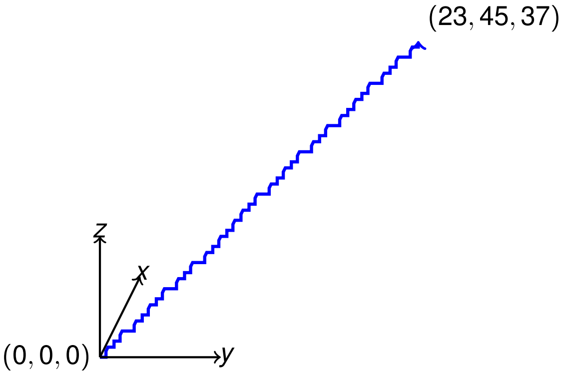
$$
\mathbf{w}_0 \xleftarrow{\begin{array}{ccc}1&\mapsto&1\\2&\mapsto&2\\3&\mapsto&13\end{array}} \mathbf{w}_1 \xleftarrow{\begin{array}{ccc}1&\mapsto&1\\2&\mapsto&23\\3&\mapsto&3\end{array}} \mathbf{w}_2 \xleftarrow{\begin{array}{ccc}1&\mapsto&1\\2&\mapsto&2\\3&\mapsto&223\end{array}} \mathbf{w}_3 \xleftarrow{\begin{array}{ccc}1&\mapsto&133\\2&\mapsto&2\\3&\mapsto&3\end{array}} \mathbf{w}_4
$$

$$\mathbf{w} = \mathbf{w}_0 = 12132131321321313$$





Applying Brun algorithm on $(23, 45, 37)$



Brun

En cours Linear complexity for S-adic Brun words
[Labbé-Leroy]

Theorem [Delecroix, Hejda, Steiner]

- For almost every $f \in \mathbb{R}_3^+$ the Brun word with frequency vector f is finitely balanced
- There exist (uncountably many) Brun words that are not finitely balanced

Theorem [B.-Steiner-Thuswaldner]

- For almost every $(\alpha, \beta) \in [0, 1]^2$, the S-adic system associated with the Brun multidimensional continued fraction algorithm of (α, β) is measurably conjugate to the translation by (α, β) on the torus $\mathbb{T}^2$

Proof Based on

- "adic IFS"
- and finiteness results. Finite products of Brun substitutions have discrete spectrum [B.-Bourdon-Jolivet-Siegel]