An introduction to combinatorics

V. Berthé

IRIF-CNRS-Université Paris

An overview

- A brief overview
- Generating functions and a symbolic dictionary
- Combinatorics on words
- A detour to quasicrystals and tilings
- Analysis of Euclid's algorithm

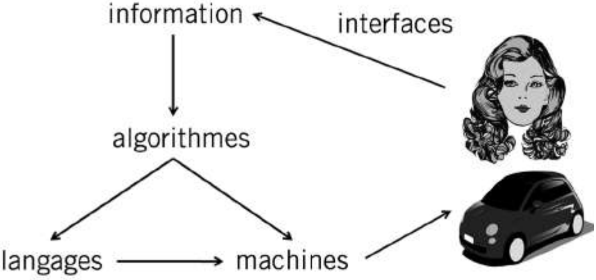


Image from G. Berry, Cours au collège de France

- Algorithmique, combinatoire, graphes, automates, systèmes dynamiques discrets.
- Calcul formel et calcul certifié, arithmétique des ordinateurs, codage et cryptographie.
- Logique, complexité algorithmique et structurelle, sémantique, modèles de calcul.
- Programmation, génie logiciel, vérification et preuves.
- Recherche opérationnelle, aide à la décision, optimisation discrète et continue, satisfaction de contraintes, SAT.
- Systèmes de production, logistique, ordonnancement.
- Intelligence artificielle, système multi-agent, ingénierie / représentation et traitement des connaissances, représen traitement de l'incertitude, formalisation des raisonnements, fusion information.
- Environnements informatiques pour l'apprentissage humain.
- Sûreté de fonctionnement, sécurité informatique, protection de la vie privée, réseaux sociaux.
- Réseaux, télécommunications, systèmes distribués, réseaux de capteurs.
- Internet du futur, intelligence ambiante.
- Calcul distribué, grilles, cloud, calcul à haute performance, parallélisme, architecture et compilation, infrastructure
- Cognition, modélisation pour la médecine, neurosciences computationnelles.
- Systèmes d'informations, web sémantique, masses de données, fouille de données, base de données, gestion de d d'informations, apprentissage.
- Bioinformatique.

Section 06, CNRS

What is combinatorics?

- Enumerative combinatorics
- Analytic combinatorics
- Algebraic combinatorics
- Probabilistic combinatorics
- Bijective combinatorics
- Extremal combinatorics
- Combinatorics on words
- Graph theory

What is combinatorics?

- Enumerative combinatorics (permutations, partitions, maps, etc.)
- Analytic combinatorics (complex analysis)
- Algebraic combinatorics
- Probabilistic combinatorics
- Bijective combinatorics
- Extremal combinatorics
- Combinatorics on words
- Graph theory

- Geometric combinatorics
- Topological combinatorics
- Arithmetic combinatorics

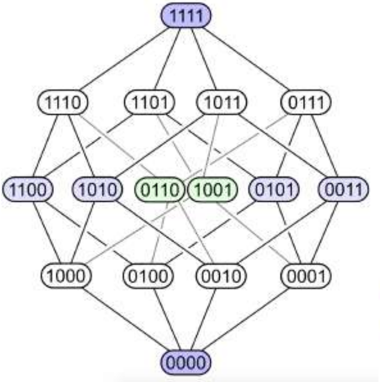
What is combinatorics?

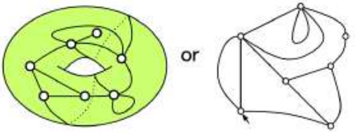
- Enumerative combinatorics
- Analytic combinatorics
- Algebraic combinatorics
- Probabilistic combinatorics
- Bijective combinatorics
- Extremal combinatorics
- Combinatorics on words
- Graph theory

- Partition theory, Design theory, Order theory, Matroid theory
- Combinatorial optimization, Coding theory, Discrete and computational geometry, Combinatorics and dynamical systems
- Combinatorics and physics

Some influential people in France

- M.-P. Schützenberger
- M. Nivat
- P. Flajolet
- X. Viennot
- M. Bousquet-Mélou





Images from Wikipedia and G.Chapuy

On enumerative combinatorics

Most of the questions that we study start like this: given a set of discrete objects, equipped with a notion of size (say permutations on n elements), how many objects of size n are there? Of course you do not want a number for particular values of n but a formula or, more realistically, a characterisation (e.g. a recurrence relation) valid for general n.

Sometimes, more important than getting a counting formula for a certain problem is the fact that to arrive at such a formula requires information about the combinatorial structure under study. Hence, counting is sometimes just a pretext and the important thing is to understand, or discover, a structure in some discrete objects.

M. Bousquet-Mélou, EMS Newsletter 2017.

On enumerative combinatorics

The objects that we (try to) count come from various branches of mathematics, including probability (of course the interaction with this area is particularly strong via discrete probability), algebra (e.g. in connection with representations of classical groups and algebras) and mathematical physics (via the study of discrete models, like the famous Ising model).

M. Bousquet-Mélou, EMS Newsletter 2017.

On enumerative combinatorics

Most French combinatorialists work in computer science departments. There are several reasons for that, partly historical but mostly scientific: there is no real boundary between some parts of theoretical computer science (e.g. the study of formal languages) and discrete mathematics. There is also a strong interaction between enumerative combinatorics and the study of the complexity of algorithms, as launched a long time ago by Don Knuth and pursued in France by Philippe Flajolet and his school. The rough idea is that in order to understand the complexity of an algorithm, one has to determine how many entries of a given length get processed in a given time – a well-posed bivariate counting problem.

M. Bousquet-Mélou, EMS Newsletter 2017.

Alea

- Study of discrete random structures coming from various disciplines: fundamental computer science and algorithmics, discrete mathematics and probability, statistical physics...
- Objects: trees, words, permutations, paths, cellular automata, etc.
- Methods: enumeration, asymptotic properties and analytic combinatorics, probabilistic properties, random generation...

Domaine un peu paradoxal, la combinatoire se présente comme

- simple et complexe
- pauvre et riche
- facile et difficile
- pure et appliquée

Elle occupe aujourd'hui une place quasi-centrale en mathématiques en particulier à cause de des interactions

algèbre, théorie des nombres, probabilités, topologie, géométrie algébrique

Informatique, mathématiques, physique (statistique)

Extrait de la description du cours au collège de France de Timothy Gowers, 2021.

Generating functions

Generating functions are used to describe families of combinatorial objects. Let $\mathcal{C}$ denote the family of objects to count.

A combinatorial class is a set $\mathcal{C}$, equipped with a size function $|.|\colon \mathcal{C} \to \mathbb{N}$, such that for any n the set $\mathcal{C}_n$ of objects of size n is finite. Let c_n stand for its cardinality. The generating function of $\mathcal{C}$ is the formal power series

$$C(x) = \sum_{n=0}^{\infty} c_n x^n.$$

There are various natural operations on generating functions such as addition, multiplication, differentiation, etc., which have a combinatorial meaning.

A symbolic dictionary

The generating function of $\mathcal{C}$ is the formal power series

$$C(x) = \sum_{n=0}^{\infty} c_n x^n,$$

where c_n is the number of elements of size n

A symbolic dictionary

The generating function of $\mathcal{C}$ is the formal power series

$$C(x) = \sum_{n=0}^{\infty} c_n x^n,$$

where c_n is the number of elements of size n

- Disjoint union $\longleftrightarrow$ addition
- Product $\longleftrightarrow$ pairs $\mathcal{C} = \mathcal{A} \times \mathcal{B}$ with $|(a,b)| = |a| + |b|$
- Sequence $\mathcal{C} = \cup_{k\geq 0}\mathcal{A}^k$ $c = a_1 \cdots a_k$

$$\longleftrightarrow C(x) = \frac{1}{1 - A(x)} = \sum_{k\geq 0} A(x)^k \quad (\mathcal{A}_0 = \emptyset)$$

- Differentiation $\longleftrightarrow$ expectation

A symbolic dictionary

The generating function of $\mathcal{C}$ is the formal power series

$$C(x) = \sum_{n=0}^{\infty} c_n x^n,$$

where c_n is the number of elements of size n

- Disjoint union $\longleftrightarrow$ addition
- Product $\longleftrightarrow$ pairs $\mathcal{C} = \mathcal{A} \times \mathcal{B}$ with $|(a, b)| = |a| + |b|$
- Sequence $\mathcal{C} = \cup_{k \geq 0} \mathcal{A}^k$ $c = a_1 \cdots a_k$

$$\longleftrightarrow C(x) = \frac{1}{1 - A(x)} = \sum_{k \geq 0} A(x)^k \quad (\mathcal{A}_0 = \emptyset)$$

- Differentiation $\longleftrightarrow$ expectation

Example: Let $\mathcal{C}$ be the set of all finite binary words, with size given by the length. Then

$$A(x) = 2x, \; C(x) = \frac{1}{1 - A(x)} = \frac{1}{1 - 2x}, \; c_n = 2^n.$$

Counting binary trees

Let $\mathcal{C}_n$ be the number of binary trees that have n binary branching nodes, and hence n +1 external nodes.
A tree is a connected graph without cycle. A (complete) binary rooted plane tree is such that:

- there is a distinguished vertex, called the root;
- the tree is drawn from the root, so there is a natural genealogical structure, and in particular, a notion of children of a vertex;
- the children of every vertex are ordered from left to right.
- A complete binary tree is such that all vertices have arity 0 or 2.

`https://www.irif.fr/~chapuy/`
`/chapuyCombinatoricsNotesMPRI.pdf`

Counting binary trees

Let $\mathcal{C}_n$ be the number of binary trees that have n binary branching nodes, and hence n +1 external nodes.

$$c_0 = 1,\ c_1 = 1,\ c_2 = 2,\ c_3 = 5,\ c_4 = 14,\ c_5 = 42$$

6 AN INVITATION TO ANALYTIC COMBINATORICS

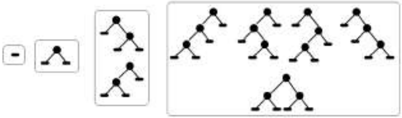


Figure 0.3. The collection of binary trees with $n = 0, 1, 2, 3$ binary nodes, with respective cardinalities 1, 1, 2, 5.

From THE book Analytic combinatorics [Flajolet-Sedgewick]

Counting binary trees

Let $\mathcal{C}_n$ be the number of binary trees that have n binary branching nodes, and hence n +1 external nodes.

Counting binary trees

Let $\mathcal{C}_n$ be the number of binary trees that have n binary branching nodes, and hence n +1 external nodes.

$$\mathcal{C} = \square \cup (\mathcal{C}, \bullet, \mathcal{C}) \qquad C(z) = \sum_{n \geq 0} c_n z^n$$

$$C(z) = 1 + zC(z)^2 \qquad C(z) = \frac{1 - \sqrt{1 - 4z}}{2z}$$

$$c_n = \frac{1}{n+1}\binom{2n}{n} \qquad c_n \sim \frac{1}{\sqrt{\pi}}(1/4)^n n^{-3/2}$$

These numbers are known as the Catalan numbers.

From singularities to asymptotic combinatorics

Let $Q(x)$ be a polynomial with complex coefficients.
Write

$$Q(x) = \prod_{i=1}^{k} (1 - \gamma_i x)^{d_i}$$

with distinct γ_i's.
Let

$$A(x) = \frac{P(x)}{Q(x)} = \sum_n a_n x^n$$

be a formal power series, with P polynomial with $\deg(P) < \deg(Q)$. Then, for all n

$$a_n = R_1(n)\gamma_1^n + \cdots + R_k(n)\gamma_k^n$$

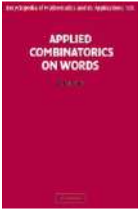
where $R_1, \cdots, R_k$ are polynomials with $\deg R_i < d_i$.

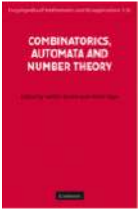
Combinatorics on words

Combinatorics on words

A wide field of applications: automata theory, bio-informatics, computational biology, algorithms on strings, text compression, number theory, Schrödinger operators.

Among the main questions: existence of patterns (e.g., squarefree words), repetitions and regularities, counting configurations, statistical properties.





[Lothaire, Algebraic combinatorics on words,
N. Pytheas Fogg, Substitutions in dynamics, arithmetics and combinatorics
CANT Combinatorics, Automata and Number theory]

Unavoidable regularities and patterns

The story starts with the work of A. Thue (1863–1922) with the existence of square-free infinite words.

Thue was interested in finding long sequences with few repetitions.

A word is square-free if it avoids the pattern xx.

Squares cannot be avoided on infinite binary words.

$$aa, \; ab, \; ba, \; bb$$

$$aba, bab$$

$$abaa, \; abab, \; baba, \; babb$$

The Thue-Morse substitution

Overlaps can be avoided on a binary alphabet. Consider the Thue-Morse substitution

$$\sigma : a \to ab,\ \ b \to ba$$

$$\begin{aligned} \sigma(a) &= ab \\ \sigma^2(a) &= abba \\ \sigma^3(a) &= abbabaab \end{aligned}$$

The infinite word $\sigma^\infty(a)$ is overlap-free: it has no factor of the form

$$uvuvu$$

for some words u, v with u nonempty

$$\sigma^{\infty}(a) = abbabaabbaababbabaababba\cdots$$

The word t derived from the Thue–Morse by the inverse morphism $A \to abb, B \to ab, C \to a$ is square-free

$$t = ABCACBABCBA\cdots$$

$$\sigma^\infty(a) = abbabaabbaababbabaababba\cdots$$

$$\sigma^\infty a = abb \mid ab \mid a \mid abb \mid a \mid ab \mid abb \mid ab \mid a \mid ab \mid abb \mid a\cdots$$

The word t derived from the Thue–Morse by the inverse morphism $A \to abb, B \to ab, C \to a$ is square-free

$$t = ABCACBABCBA\cdots$$

On Dejean's conjecture

- A repetition in a word w is a pair of words (p, q) such that pq is a factor of w, p is nonempty, and q is a prefix of pq.
- The exponent of a repetition (p, q) is $\frac{|pq|}{|q|}$.
- Squares are repetitions of exponent 2.
- A word is x-free if it does not contain a repetition of exponent y with $y \geq x$.

On Dejean's conjecture

- A repetition in a word w is a pair of words (p, q) such that pq is a factor of w, p is nonempty, and q is a prefix of pq.
- The exponent of a repetition (p, q) is $\frac{|pq|}{|q|}$.
- Squares are repetitions of exponent 2.
- A word is x-free if it does not contain a repetition of exponent y with $y \geq x$.
- For an integer $k \geq 2$, the repetition threshold $R(k)$ for k letters is the infimum over the set of x such that there exists an infinite x-free word over a k-letter alphabet.

On Dejean's conjecture

- A repetition in a word w is a pair of words (p, q) such that pq is a factor of w, p is nonempty, and q is a prefix of pq.
- The exponent of a repetition (p, q) is $\frac{|pq|}{|q|}$.
- Squares are repetitions of exponent 2.
- A word is x-free if it does not contain a repetition of exponent y with $y \geq x$.
- For an integer $k \geq 2$, the repetition threshold $R(k)$ for k letters is the infimum over the set of x such that there exists an infinite x-free word over a k-letter alphabet.
- Dejean's conjecture has been proven in 2011 [Rao, 2011) and (Currie and Rampersad, 2011]: the repetition treshold, i.e., the largest avoidable fractional power in an infinite word on k letters is $k/(k-1)$.

$$R(2) = 2,\ R(3) = 7/4,\ R(4) = 7/5,\ R(k) = \frac{k}{k-1},\ k \geq 5$$

Word combinatorics

Let $\mathcal{A}$ be a finite alphabet and let $u \in \mathcal{A}^{\mathbb{N}}$ be an infinite word

$$u = abaababaabaababaababaab\cdots$$

$$u = abaababaab\underbrace{aa}babaababaab\cdots$$

aa is a factor, bb is not a factor

Toward symbolic dynamics

Let $\mathcal{A}$ be a finite alphabet and let $u \in \mathcal{A}^{\mathbb{N}}$ be an infinite word

$$u = abaababaabaababaababaab \cdots$$

$$u = abaababaab \underbrace{aa} babaababaab \cdots$$

aa is a factor, bb is not a factor

The shift maps $u = (u_n)_{n \in \mathbb{N}}$ to $(u_{n+1})_{n \in \mathbb{N}}$

$$u = abaababaabaababaababaab \cdots$$

$$S(u) = baababaabaababaababaab \cdots$$

Discrete dynamical system

A discrete dynamical system is given by a map T acting on a set X

$$T\colon X \to X$$

Discrete stands for discrete time
The map T is the law of time evolution

We consider orbits/trajectories of points of X under the action of the map T

$$\{T^n x \mid n \in \mathbb{N}\}$$

Discrete dynamical system

A discrete dynamical system is given by a map T acting on a set X

$$T\colon X \to X$$

Discrete stands for discrete time
The map T is the law of time evolution

We consider orbits/trajectories of points of X under the action of the map T

$$\{T^n x \mid n \in \mathbb{N}\}$$

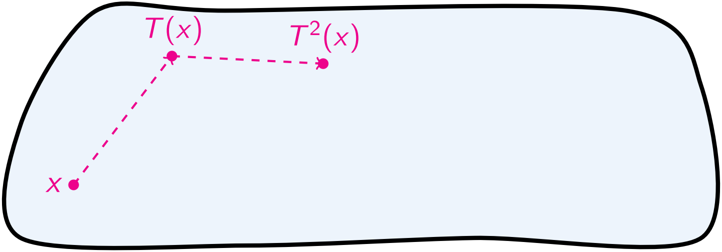
How well are the orbits distributed?

A trajectory for $T : X \to X$

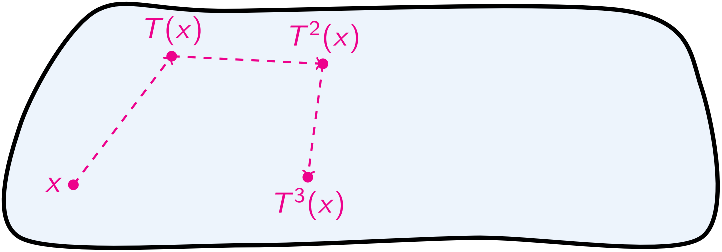
A trajectory for $T : X \to X$

A trajectory for $T : X \to X$

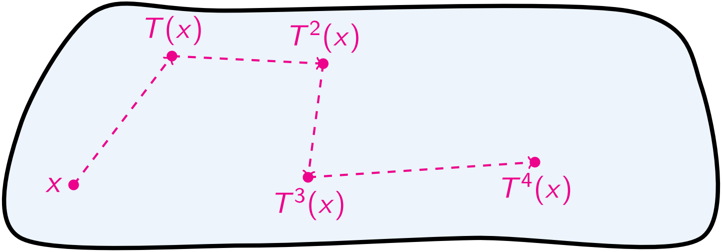
A trajectory for $T : X \to X$



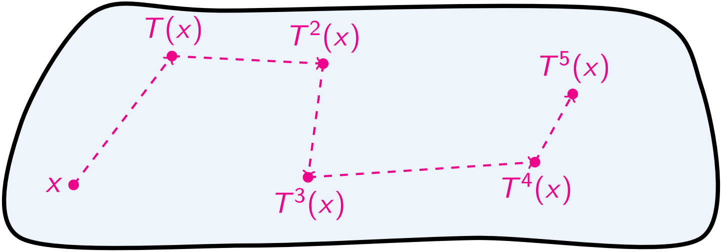
A trajectory for $T : X \to X$



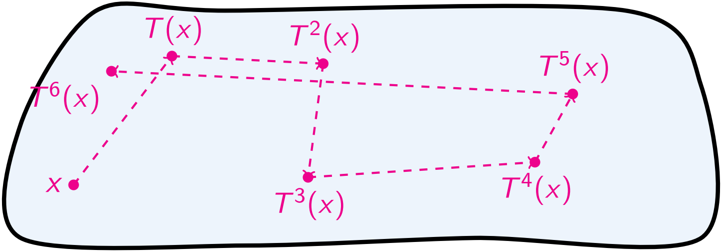
A trajectory for $T : X \to X$



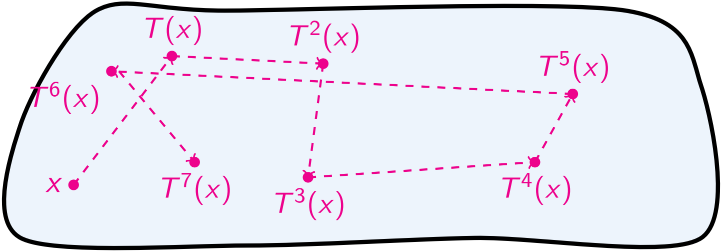
A trajectory for $T : X \to X$



A trajectory for $T : X \to X$



A trajectory for $T : X \to X$



What's the point of this formalization?

- The mathematical formalization of discrete dynamical system offers the framework of ergodic theory

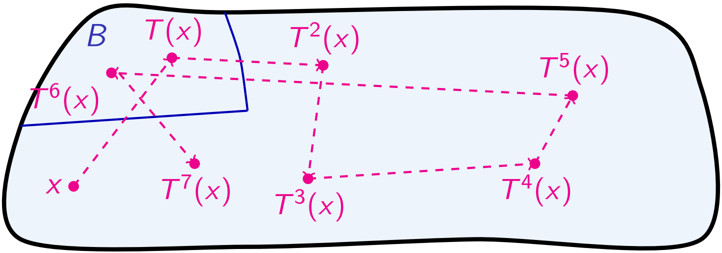
What's the point of this formalization?

- The mathematical formalization of discrete dynamical system offers the framework of ergodic theory

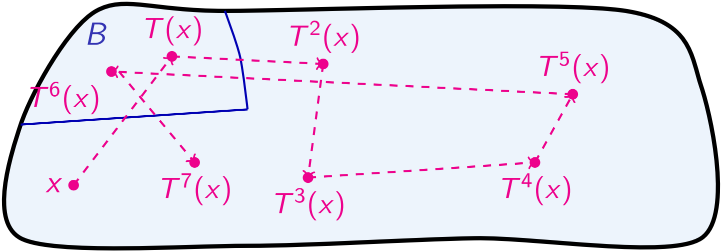
- Topological dynamics describes the qualitative/topological behaviour of trajectories
  The map T is continuous and the space X is compact

- Ergodicity describes the long term statistical behaviour of orbits
  The space X is endowed with a probability measure and T is measurable $(X, T, \mathcal{B}, \mu)$

Ergodic theorem



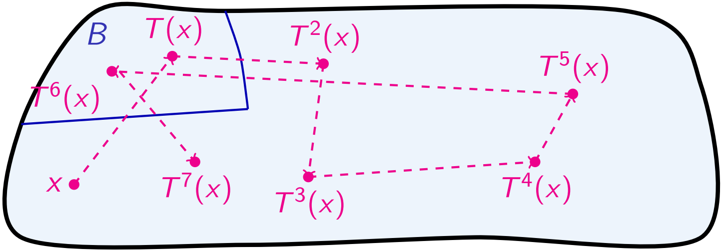
Ergodic theorem



Among the first N points of the orbit of x, how many of them enter B?

How often do they visit B?

Ergodic theorem



Let 1_B be the characteristic function of B

Among the first N points of the orbit of x, how many of them enter B? $\sum_{0\leq n<N} 1_B(T^n x)$

How often do they visit B? $\lim_{N\to\infty} \frac{1}{N} \sum_{0\leq n<N} 1_B(T^n x)$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{0\leq n<N} 1_B(T^n x) = \mu(P) \quad \text{a.e. } x$$

Ergodic theorem

We are given a dynamical system $(X, T, \mathcal{B}, \mu)$ with $T \colon X \to X$

- Average time values: one particle over the long term
- Average space values: all particles at a particular instant

Ergodicity

$$\mu(B) = \mu(T^{-1}B) \quad T\text{-invariance}$$

$$T^{-1}B = B \implies \mu(B) = 0 \text{ or } 1 \quad \text{ergodicity}$$

Ergodic theorem

space average= time average

$$f \in L_1(\mu) \qquad \lim_{N\to\infty} \frac{1}{N} \sum_{0 \le n < N} f(T^n x) = \int f \, d\mu \qquad \text{a.e. } x$$

Numeration dynamics

Numeration dynamical systems are simple algorithms that produce digits in classical representation systems

- Decimal expansions

$$T\colon [0,1] \to [0,1],\ x \mapsto 10x - [10x] = \{10x\}$$

Numeration dynamics

Numeration dynamical systems are simple algorithms that produce digits in classical representation systems

- Decimal expansions

$$T\colon [0,1] \to [0,1],\ x \mapsto 10x - [10x] = \{10x\}$$

$$x_1 = T(x) = 10x - [10x] = 10x - a_1$$

$$x = \frac{a_1}{10} + \frac{x_1}{10}$$

$$x_2 = T(x_1) = T^2(x) \qquad a_2 = \lfloor 10T(x) \rfloor$$

$$x = \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{x_2}{10^2} = \sum_{i=1}^{\infty} a_i 10^{-i}$$

Numeration dynamics

Numeration dynamical systems are simple algorithms that produce digits in classical representation systems

- Decimal expansions

$$T\colon [0,1] \to [0,1],\ x \mapsto 10x - [10x] = \{10x\}$$

The map T produces the digits

$$a_n = \lfloor 10T^{n-1}(x) \rfloor$$

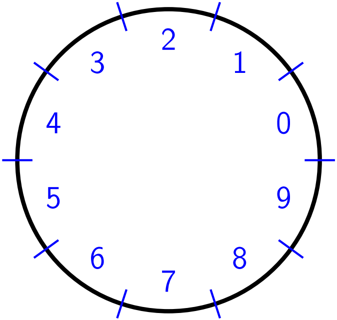
The action of T can be seen as a shift on the sequence of digits

$$x \sim a_1 a_2 a_3 a_4 \cdots \qquad T(x) \sim a_2 a_3 a_4 \cdots$$

Multiplication by 10 on $[0,1]$

$X = [0,1] \qquad T : x \mapsto 10\,x \pmod 1$

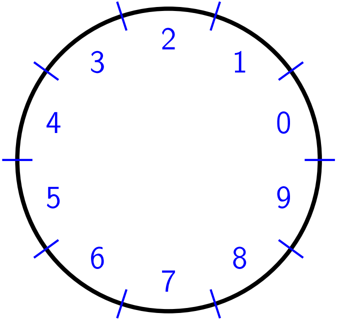
$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \leq i \leq 9 \right\}$



Multiplication by 10 on $[0, 1]$

$X = [0, 1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \leq i \leq 9 \right\}$

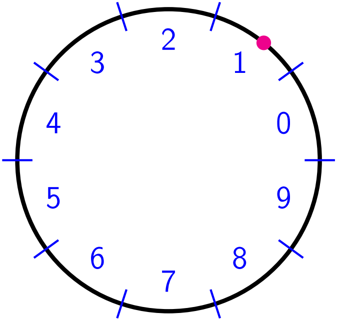


Orbit of $\pi - 3$

Multiplication by 10 on $[0,1]$

$X = [0,1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \leq i \leq 9 \right\}$



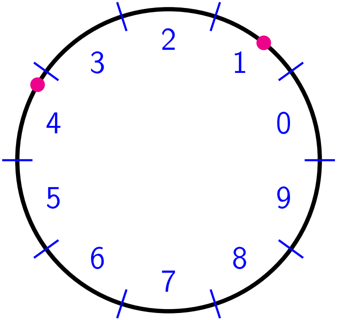
Orbit of $\pi - 3$

0.1

Multiplication by 10 on $[0,1]$

$X = [0,1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[\; : 0 \leq i \leq 9 \right\}$



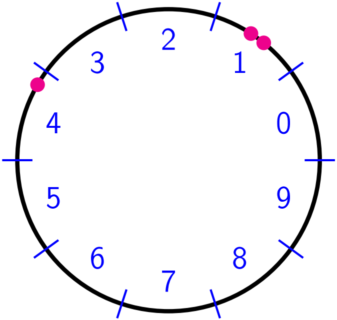
Orbit of $\pi - 3$

0.14

Multiplication by 10 on $[0,1]$

$X = [0,1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \leq i \leq 9 \right\}$



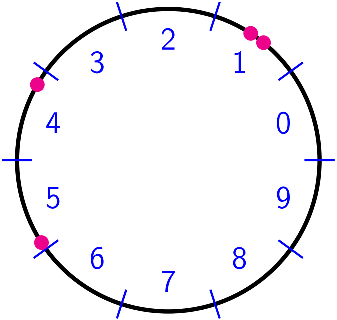
Orbit of $\pi - 3$

0.141

Multiplication by 10 on $[0, 1]$

$$X = [0, 1] \qquad T : x \mapsto 10\,x \pmod 1$$

$$\mathcal{P} = \left\{ \left[\tfrac{i}{10}, \tfrac{i+1}{10} \right[\, : 0 \le i \le 9 \right\}$$



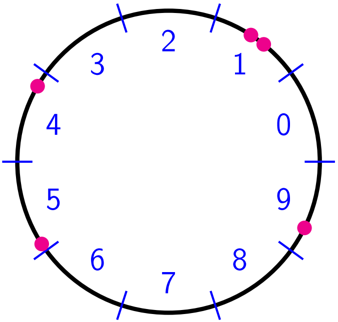
Orbit of $\pi - 3$

0.1415

Multiplication by 10 on $[0,1]$

$X = [0,1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \le i \le 9 \right\}$



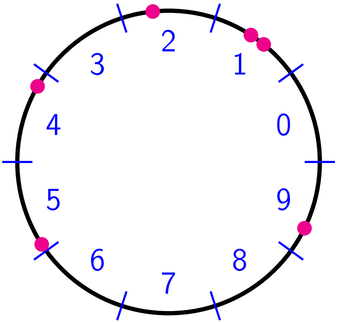
Orbit of $\pi - 3$

0.14159

Multiplication by 10 on $[0, 1]$

$X = [0, 1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \le i \le 9 \right\}$

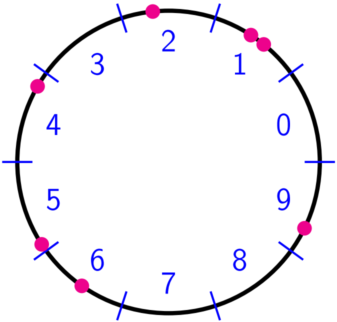


Orbit of $\pi - 3$
0.141592

Multiplication by 10 on $[0, 1]$

$$X = [0,1] \qquad T : x \mapsto 10\,x \pmod 1$$

$$\mathcal{P} = \left\{ \left[\tfrac{i}{10}, \tfrac{i+1}{10} \right[\; : 0 \le i \le 9 \right\}$$

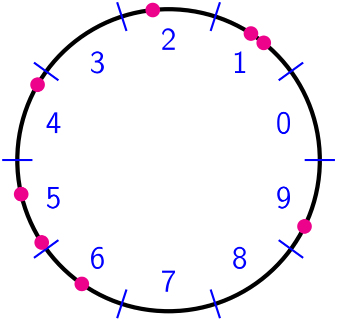


Orbit of $\pi - 3$
0.1415926

Multiplication by 10 on $[0,1]$

$X = [0,1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \leq i \leq 9 \right\}$

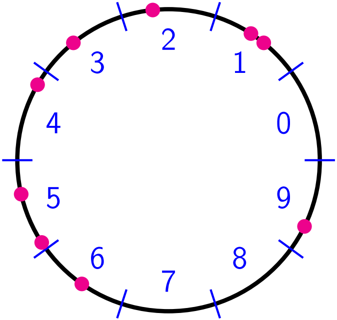


Orbit of $\pi - 3$
0.14159265

Multiplication by 10 on $[0, 1]$

$X = [0, 1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \leq i \leq 9 \right\}$

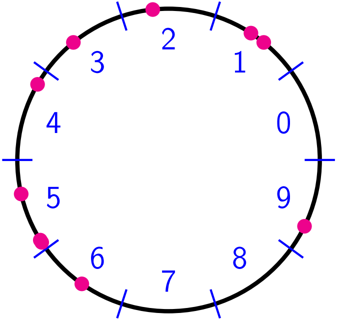


Orbit of $\pi - 3$
0.141592653

Multiplication by 10 on $[0,1]$

$X = [0,1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \leq i \leq 9 \right\}$

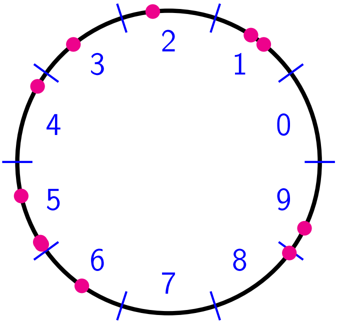


Orbit of $\pi - 3$
0.1415926535

Multiplication by 10 on $[0, 1]$

$X = [0, 1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \le i \le 9 \right\}$

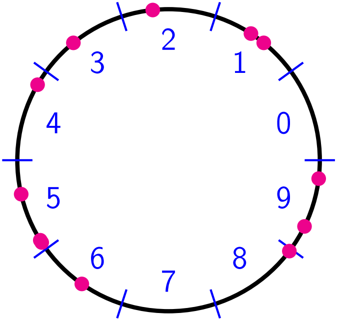


Orbit of $\pi - 3$
0.14159265358

Multiplication by 10 on $[0,1]$

$X = [0,1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \leq i \leq 9 \right\}$



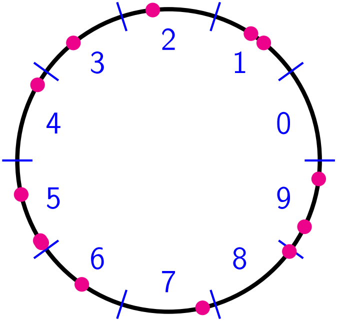
Orbit of $\pi - 3$

0.141592653589

Multiplication by 10 on $[0, 1]$

$X = [0, 1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[\, : 0 \leq i \leq 9 \right\}$

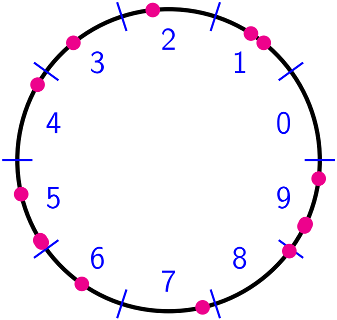


Orbit of $\pi - 3$
0.1415926535897

Multiplication by 10 on $[0, 1]$

$X = [0, 1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[: 0 \leq i \leq 9 \right\}$



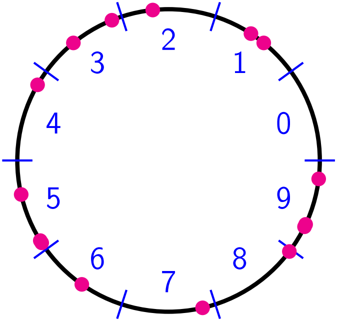
Orbit of $\pi - 3$

$0.14159265358979312\cdots$

Multiplication by 10 on $[0,1]$

$X = [0,1] \qquad T : x \mapsto 10\,x \pmod 1$

$\mathcal{P} = \left\{ \left[\frac{i}{10}, \frac{i+1}{10} \right[\,:\, 0 \leq i \leq 9 \right\}$



Orbit of $\pi - 3$

$0.14159265358979312\cdots$

Codings $\Longleftrightarrow$ decimal expansions

From numeration dynamics to symbolic dynamics

- Decimal expansion $T\colon [0,1] \to [0,1],\ x \mapsto \{10x\}$
- Beta-transformation $T\colon [0,1] \to [0,1],\ x \mapsto \{\beta x\}$
- Continued fractions $T\colon [0,1] \to [0,1],\ x \mapsto \{1/x\}$

From numeration dynamics to symbolic dynamics

- Decimal expansion $T\colon [0,1] \to [0,1],\ x \mapsto \{10x\}$

- Beta-transformation $T\colon [0,1] \to [0,1],\ x \mapsto \{\beta x\}$

$$\beta > 1 \qquad x = \sum_{i=1}^{\infty} a_i \beta^{-i}$$

- Continued fractions $T\colon [0,1] \to [0,1],\ x \mapsto \{1/x\}$

$$x = \frac{1}{a_1 + x_1} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cdots}}}}$$

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

$$X_u := \overline{\{S^n u \mid n \in \mathbb{N}\}} \text{ with the shift } S((u_n)_n) = (u_{n+1})_n$$

  (X_u, S) is a symbolic dynamical system

  Study of invariant measures, recurrence properties, finding geometric representations, spectral properties

From word combinatorics to symbolic dynamics

Let $\mathcal{A}$ be a finite alphabet and let $u \in \mathcal{A}^{\mathbb{N}}$ be an infinite word

Let S stand for the shift map

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

(X_u, S) is a symbolic dynamical system

$$X_u = \{v \,;\, \mathcal{L}_v \subset \mathcal{L}_u\}$$

This is the set of infinite words whose factors belong to the language $\mathcal{L}_u$ of u, i.e., the set of factors of u

Symbolic dynamics

- 1898, Hadamard: Geodesic flows on surfaces of negative curvature
- 1912, Thue: Prouhet-Thue-Morse substitution

$$\sigma : a \mapsto ab,\ \ b \mapsto ba$$

- 1921, Morse: Symbolic representation of geodesics on a surface with negative curvature. Recurrent geodesics

Symbolic dynamics

- 1898, Hadamard: Geodesic flows on surfaces of negative curvature
- 1912, Thue: Prouhet-Thue-Morse substitution

$$\sigma : a \mapsto ab, \;\; b \mapsto ba$$

- 1921, Morse: Symbolic representation of geodesics on a surface with negative curvature. Recurrent geodesics

From geometric dynamical systems to
symbolic dynamical systems and backwards

- Given a geometric system, can one find a good partition?
- And vice-versa?

A substitution on words: the Fibonacci substitution

Definition A substitution σ is a morphism of the free monoid $\sigma(uv) = \sigma(u)\sigma(v)$

Positive morphism of the free group, no cancellations

Example

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

1
12
121
12112
12112121

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

A substitution on words: the Fibonacci substitution

Definition A substitution σ is a morphism of the free monoid $\quad \sigma(uv) = \sigma(u)\sigma(v)$

Positive morphism of the free group, no cancellations

Example

$$\sigma : 1 \mapsto 12,\ 2 \mapsto 1 \quad \sigma^{\infty}(1) = 121121211211212\cdots$$

J. Berstel, D. Perrin / European Journal of Combinatorics 28 (2007) 996–1022

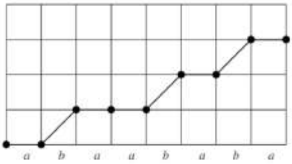


Fig. 3. The graphical representation of the Fibonacci word.

A substitution on words: the Fibonacci substitution

Definition A substitution σ is a morphism of the free monoid $\quad \sigma(uv) = \sigma(u)\sigma(v)$

Positive morphism of the free group, no cancellations

Example

$$\sigma : 1 \mapsto 12,\ 2 \mapsto 1 \quad \sigma^\infty(1) = 121121211211212\cdots$$

Why the terminology Fibonacci word?

$$\sigma^{n+1}(1) = \sigma^n(12) = \sigma^n(1)\sigma^n(2)$$

$$\sigma^n(2) = \sigma^{n-1}(1)$$

$$\sigma^{n+1}(1) = \sigma^n(1)\sigma^{n-1}(1)$$

The length of the word $\sigma^n(1)$ satisfies the Fibonacci recurrence

How to define a notion of order for an infinite word?

Consider the Fibonacci word

$u = \sigma^{\infty}(a) = abaababaabaababaababaabaababaabaababaababaa \cdots$

- There is a simple algorithmic way to construct it

  (cf. Kolmogorov complexity)

How to define a notion of order for an infinite word?

Consider the Fibonacci word

$u = \sigma^{\infty}(a) = abaababaabaababaababaabaababaabaababaababaa \cdots$

- There are few local configurations $=$ factors

  A factor is a word made of consecutive occurrences of letters

  ab is a factor, bb is not a factor of the Fibonacci word

  But

  $$\cdots aaaaaaaaaaabaaaaaaaaaa \cdots$$

  has as many factors of length n as

  $\cdots abaababaabaababaababaabaababaabaababaababaa \cdots$

  The Fibonacci word has $n+1$ factors of length n

How to define a notion of order for an infinite word?

Consider the Fibonacci word

$u = \sigma^\infty(a) = abaababaabaababaababaabaababaabababaa \cdots$

- Consider frequencies of occurrences of factors

  Symbolic discrepancy

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

if each letter i has frequency f_i in u

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

The Fibonacci word has bounded symbolic discrepancy

Complexity and periodicity

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

The factor complexity $p_u(n)$ counts the number of factors of length n

Theorem [Morse-Hedlund 1940]

If there exists n such that $p_u(n) \leq n$, then u is ultimately periodic

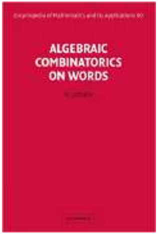
There exists T such that $u_n = u_{n+T}$ for all n large enough

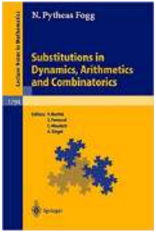
Proof

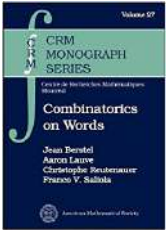
- We can assume $p_u(1) \geq 2$
- There exists $1 \leq k \leq n-1$ such that $p_u(k) = p_u(k+1)$
- Every factor w of length k admits a unique letter $a \in \mathcal{A}$ such that wa is also a factor of u
- Take a factor of length k that occurs at least twice in u

Sturmian words

Sturmian words [Morse-Hedlund, 1940] $p_u(n) = n + 1$ for all n





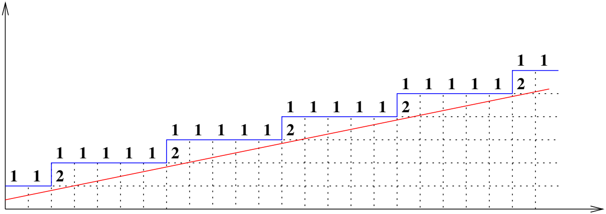


Sturmian words

A word $u \in \{0,1\}^{\mathbb{N}}$ is Sturmian if $p_u(n) = n + 1$ for all n

The Fibonacci word has $n + 1$ factors of length n

- Sturmian words are the words having the lowest factor complexity among non-periodic words
- They are codings of discrete lines



Sturmian words

Sturmian words are defined as the infinite words with factor complexity $n+1$ for all n

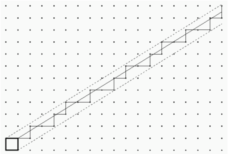
0110110101101101

Sturmian words

Sturmian words are defined as the infinite words with factor complexity $n+1$ for all n

011011010110110 1

11 and 00 cannot occur simultaneously



Sturmian words

Sturmian words are defined as the infinite words with factor complexity $n+1$ for all n

$$0110110101101101$$

One considers the substitutions

$$\sigma_0\colon 0 \mapsto 0,\ \ \sigma_0\colon 1 \mapsto 10$$

$$\sigma_1\colon 0 \mapsto 01,\ \ \sigma_1\colon 1 \mapsto 1$$

One has

$$01\,1\,01\,1\,01\,01\,1\,01\,1\,01 = \sigma_1(0101001010)$$

$$0\,10\,10\,0\,10\,10 = \sigma_0(011011)$$

$$01\,1\,01\,1 = \sigma_1(0101)$$

$$01\,01 = \sigma_1(00)$$

Sturmian words

Sturmian words are defined as the infinite words with factor complexity $n+1$ for all n

$$0110110101101101$$

One considers the substitutions

$$\sigma_0 \colon 0 \mapsto 0,\ \sigma_0 \colon 1 \mapsto 10$$

$$\sigma_1 \colon 0 \mapsto 01,\ \sigma_1 \colon 1 \mapsto 1$$

The Sturmian words of slope α are provided by an infinite composition of substitutions

$$\lim_{n\to+\infty} \sigma_0^{a_1}\sigma_1^{a_2}\cdots\sigma_{2n}^{a_{2n}}\sigma_{2n+1}^{a_{2n+1}}(0)$$

where the a_i are produced by the continued fraction expansion of α

Continued fractions

We represent real numbers in $(0,1)$ as

$$\cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cdots}}}}$$

with partial quotients (digits) $a_i \in \mathbb{N}^*$

Continued fractions

One represents α as

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}$$

in order to find good rational approximations of α

Continued fractions

One represents α as

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}$$

in order to find good rational approximations of α

$$\frac{p_n}{q_n} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots + \cfrac{1}{a_n}}}}$$

Continued fractions

One represents α as

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}$$

in order to find good rational approximations of α

$$\frac{p_n}{q_n} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots + \cfrac{1}{a_n}}}}$$

$$|\alpha - p_n/q_n] \leq 1/q_n^2$$

[http://images.math.cnrs.fr/Nombres-et-representations.html]

Complexity and periodicity

Theorem [Morse-Hedlund 1940]

If there exists n such that u has at most n factors of length n, then u is ultimately periodic

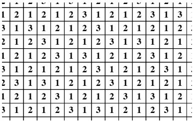
Nivat's conjecture

We now consider two-dimensional words $u \in \mathcal{A}^{\mathbb{Z}^2}$ and rectangular factors

1	2	1	2	1	2	3	1	2	1	2	3	1	3
3	1	3	1	2	1	2	3	1	2	1	2	1	2
2	1	2	3	1	2	1	2	3	1	3	1	2	1
1	2	1	2	3	1	3	1	2	1	2	3	1	2
3	1	2	1	2	1	2	3	1	2	1	2	3	1
2	3	1	3	1	2	1	2	3	1	2	1	2	1
1	2	1	2	3	1	2	1	2	3	1	3	1	2
3	1	2	1	2	3	1	3	1	2	1	2	3	1

Nivat's conjecture

We now consider two-dimensional words $u \in \mathcal{A}^{\mathbb{Z}^2}$ and rectangular factors



Nivat's conjecture [ICALP-1997]

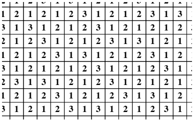
If there exists m, n such that u admits at most mn rectangular factors of size (m, n), i.e.,

$$p_u(m, n) \leq mn,$$

then u is periodic.

Nivat's conjecture

We now consider two-dimensional words $u \in \mathcal{A}^{\mathbb{Z}^2}$ and rectangular factors



Nivat's conjecture [ICALP-1997]

If there exists m, n such that u admits at most mn rectangular factors of size (m, n), i.e.,

$$p_u(m, n) \leq mn,$$

then u is periodic.

Periodic means periodic along one direction. There exists a non-zero vector (s, t) such that $u(m, n) = u(m + s, n + t)\ \forall (m, n)$.

Nivat's conjecture is not about full periodicity

A word in $\mathcal{A}^{\mathbb{Z}^d}$ is fully periodic if and only if its rectangular factor complexity function is bounded.

Proof

- One has $p_u(1, \cdots, 1, n, 1, \cdots, 1) \leq C$ for all n.
- Apply Morse-Hedlund's theorem.

Nivat's conjecture is not an equivalence

There exist periodic words with high factor complexity

There exists u periodic such that $p_u(m,n) = 2^{m+n-1}$ for all (m,n).

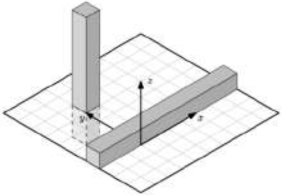
- Take a 1D word x with factor complexity $p_x(n) = 2^n$ for all n (e.g., Champernowne construction).
- Define $u \in \{0,1\}^{\mathbb{Z}^2}$ by $u(m,n) := x(0, m+n)$ for all (m,n).
- It has period $(-1,1)$.

Nivat's conjecture is a two-dimensional conjecture

Take $d = 3$

Define $u \in \{0,1\}^{\mathbb{Z}^3}$ as

- $u_{m,0,0} = 1$ for all m
- $u_{0,n_0,p} = 1$ for all p with $n_0 \neq 0$
- $u_{m,n,p} = 0$ otherwise



One has for $2 \leq n \leq n_0$

$$p_u(n, \cdots, n) = 2n^2 + 1 < n^3$$

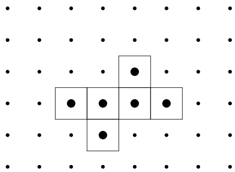
Note that u is a sum of two periodic words

Nivat's conjecture is about rectangular factors

What about general patterns? [Cassaigne]

If $p_u(D) \leq |D|$ for some D, is u periodic?

Not necessarily, even if the pattern D is an *hv*-convex polyomino (if two points in the same row or column are in the pattern, then all integer points on the segment between them should be included too)



What about convex patterns (the trace in $\mathbb{Z}^2$ of convex sets in $\mathbb{R}^2$)?

Some results toward Nivat's conjecture

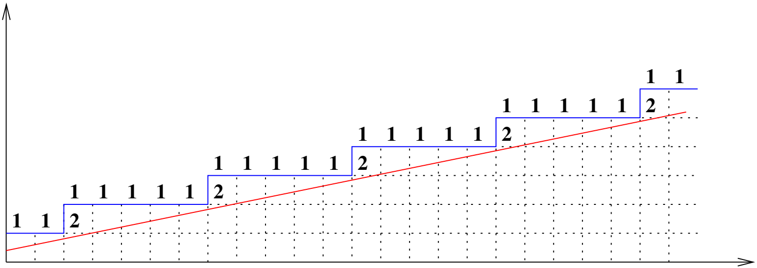
The following conditions imply periodicity

- $p_u(2, n) \leq 2n$ or $p_u(n, 2) \leq 2n$ for some n [Sander-Tijdeman 2002]
- $p_u(m, n) \leq \frac{1}{144}mn$ for some (m, n) [Epifanio-Koskas-Mignosi 2003]
- $p_u(m, n) \leq \frac{1}{16}mn$ for some (m, n) [Quas-Zamboni 2004] Combinatorial approach
- $p_u(m, n) \leq \frac{1}{2}mn$ for some (m, n) [Cyr-Kra 2015] Dynamical approach
- $p_u(m, 3) \leq 3mn$ for some (m, n) [Cyr-Kra 2016]
- $p_u(m, n) \leq mn$ for infinitely many pairs (m, n) [Kari-Szabados 2015] Algebraic approach

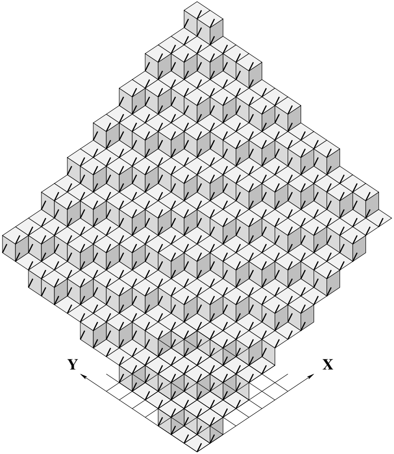
Sturmian words

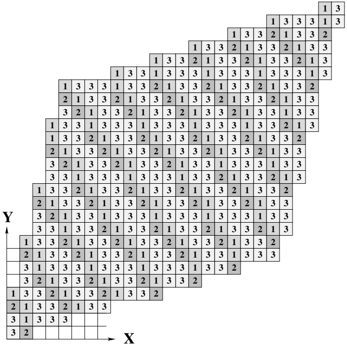
Sturmian words are the words that have $n+1$ factors of length n for all n

They are codings of discrete lines



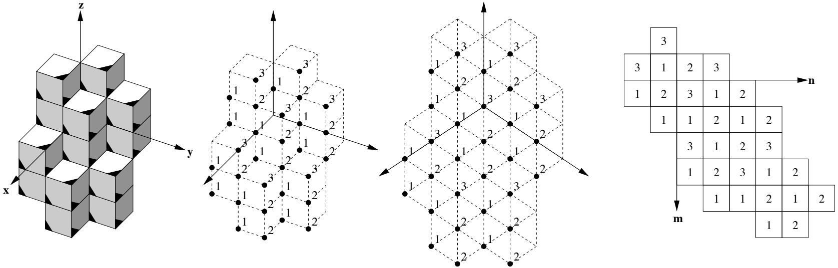
Discrete planes and 2D Sturmian words





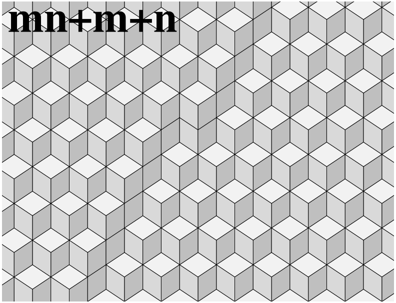
©Th. Fernique

Discrete planes and 2D Sturmian words



2D Sturmian words [B.-Vuillon]

$p_u(m,n) = mn + m + n$ for all (m,n)



2D Sturmian words are

- codings of discrete planes
- they have low complexity function
- quasicrystals

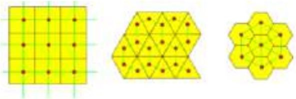
The geometry of discrete objects

can be

- algorithmic/computational ex: convex hull, Delaunay triangulation
- discrete/digital ex: discretization, segmentation, discrete convexity
- discrete differential ex: topological combinatorics, geometric estimators
- combinatorial ex: packings, hyperplane arrangements

Discrete geometry Digital geometry

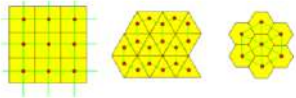
Analysis of geometric problems on objects defined on regular lattices



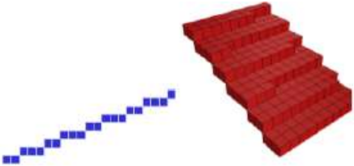
[D. Coeurjoly, Digital geometry in a Nutshell]

Discrete geometry Digital geometry

Analysis of geometric problems on objects defined on regular lattices



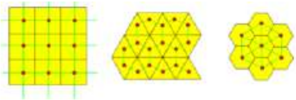
Among the most basic primitives one finds discrete lines and planes



[D. Coeurjoly, Digital geometry in a Nutshell]

Discrete geometry Digital geometry

Analysis of geometric problems on objects defined on regular lattices



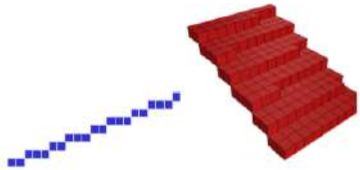
Example of application: segmentation into maximal discrete segments

[D. Coeurjoly, Digital geometry in a Nutshell]

Digital geometry

How to discretize a line in the space?

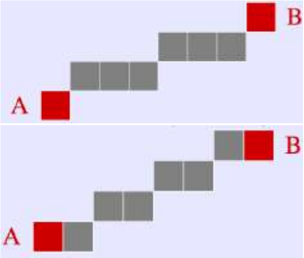
- There are the usual difficulties related to discrete geometry
- There are further difficulties due to the codimension > 1 for discrete lines



[D. Coeurjoly, Digital geometry in a Nutshell
http://liris.cnrs.fr/david.coeurjolly/doku/doku.php]

Euclid first axiom

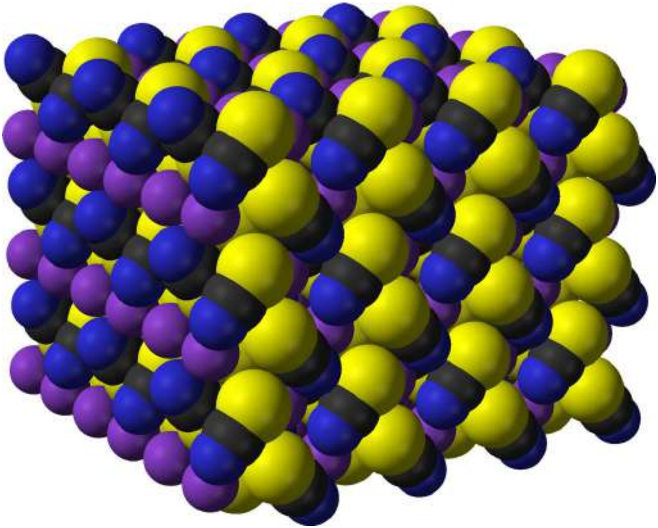
Given two points A and B, there exists a unique line that contains them
This is no more true in the discrete case



[D. Coeurjoly, Digital geometry in a Nutshell]

Words, tilings and quasicrystals

A crystal



A periodic arrangement of atoms

Quasiperiodicity and quasicrystals

Quasicrystals are solids discovered in 84 with an atomic structure that is both ordered and aperiodic [Shechtman-Blech-Gratias-Cahn]

An aperiodic system may have long-range order

(cf. Aperiodic tilings [Wang'61, Berger'66, Robinson'71,...])

Which mathematical models for quasicrystals?

There are mainly two methods for producing quasicrystals

- Substitutions
- Cut and project schemes

[WHAT IS.. a Quasicrystal? M. Senechal]

Which models for quasicrystals?

"His discovery was extremely controversial. In the course of defending his findings, he was asked to leave his research group. However, his battle eventually forced scientists to reconsider their conception of the very nature of matter."

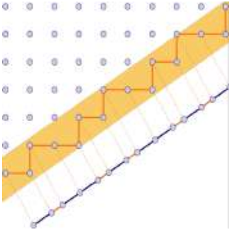
Aperiodic mosaics, such as those found in the medieval Islamic mosaics of the Alhambra Palace in Spain and the Darb-i Imam Shrine in Iran, have helped scientists understand what quasicrystals look like at the atomic level. In those mosaics, as in quasicrystals, the patterns are regular - they follow mathematical rules - but they never repeat themselves.

When scientists describe Shechtman's quasicrystals, they use a concept that comes from mathematics and art : the golden ratio.

Cut and project schemes

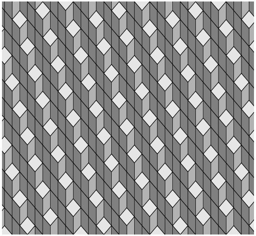
Projection of a “plane” slicing through a higher dimensional lattice

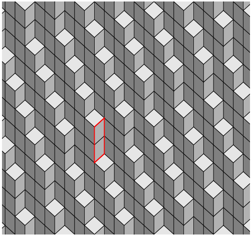
- The order comes from the lattice structure
- The nonperiodicity comes from the irrationality of the normal vector of the “plane”

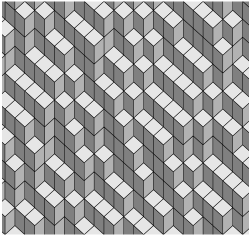


Sturmian words are 1D quasicrystals

Toward long-range aperiodic order







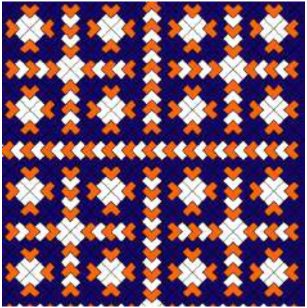
What is meant by quasiperiodicity?

The objects under consideration

- Infinite words (sequences with values in a finite alphabet)

$$abaababaabaababaababaabaababaabaababaababaa\cdots$$

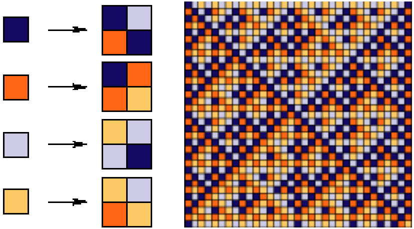
- Tilings



A tiling of the plane is a collection of tiles that covers the plane with no overlaps

Substitutions

- Substitutions on words and symbolic dynamical systems
- Substitutions on tiles : inflation/subdivision rules, tilings and point sets

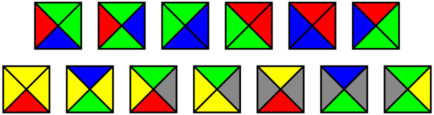


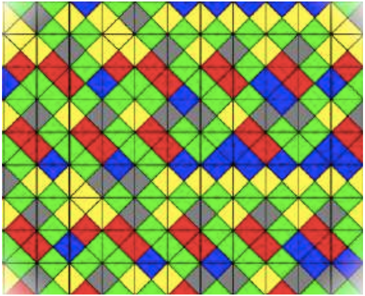
- Tilings Encyclopedia http://tilings.math.uni-bielefeld.de/

[E. Harriss, D. Frettlöh]

Wang tiles

These are square tiles with colors on each side and colors have to match.





A decision problem (1961)

Can one tile the plane with a given set of Wang tiles?

The Eternity game

A price of 2 millions of dollars!
256 Wang tiles to place on a 16×16 grid
The number of solutions is estimated to 20 000

https://fr.wikipedia.org/wiki/Eternity_(jeu)

A conjecture

If a set of Wang tiles can pave the plane, it can pave it in a periodic way

We then can decide the domino problem

which turned to be false

There exist aperiodic sets of tiles!

https://www.lri.fr/~aubrun/exposes/SML_Aubrun.pdf
http://images.math.cnrs.fr/Dominos-aperiodiques.html

Aperiodic sets of tiles

They only allow the production of aperiodic tilings

- Berger, 1964 20426 tiles (computability)
- Berger, 1964 104 tiles
- Robinson, 1967 52 tiles (computability and substitutions)
- Penrose, 1976 34 tiles (substitutions)

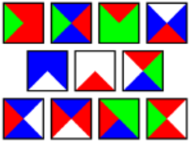
Aperiodic sets of tiles

They only allow the production of aperiodic tilings

- Berger, 1964 20426 tiles (computability)
- Berger, 1964 104 tiles
- Robinson, 1967 52 tiles (computability and substitutions)
- Penrose, 1976 34 tiles (substitutions)

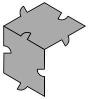
And the actual record is

- E. Jeandel and M. Rao, 2015 11 tiles and 4 colors

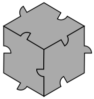


A periodic tiling

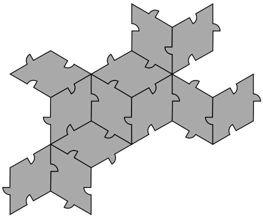
A periodic tiling



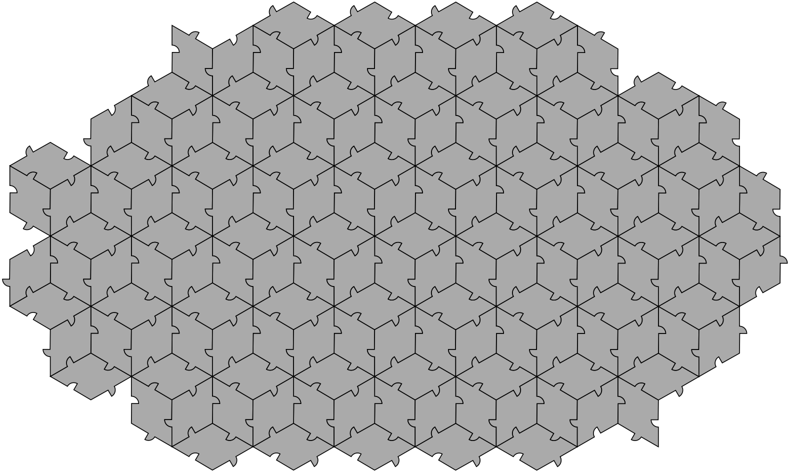
A periodic tiling



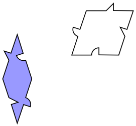
A periodic tiling



A periodic tiling

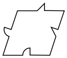


Penrose tiling



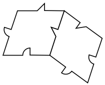
This aperiodic tiling is also generated by cut and projection and by substitution

Penrose tiling



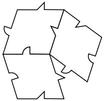
This aperiodic tiling is also generated by cut and projection and by substitution

Penrose tiling



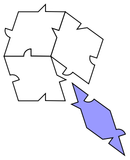
This aperiodic tiling is also generated by cut and projection and by substitution

Penrose tiling



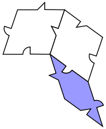
This aperiodic tiling is also generated by cut and projection and by substitution

Penrose tiling



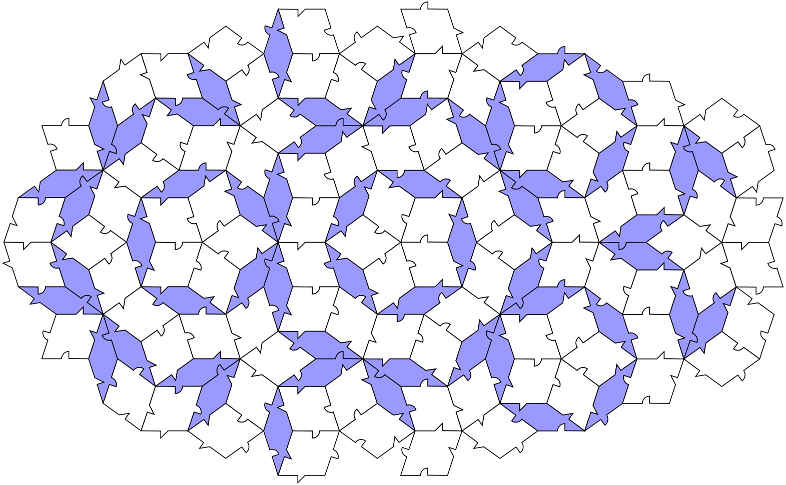
This aperiodic tiling is also generated by cut and projection and by substitution

Penrose tiling



This aperiodic tiling is also generated by cut and projection and by substitution

Penrose tiling



This aperiodic tiling is also generated by cut and projection and by substitution

Combinatorics and analysis of algorithms

Analysis of algorithms

- Analysis of algorithms [Knuth'63]

probabilistic, combinatorial, and analytic methods

- Analytic combinatorics [Flajolet-Sedgewick]

generating functions and complex analysis,
analysis of the singularities

- Dynamical analysis of algorithms [Vallée]

Transfer operators $\rightsquigarrow$ Generating functions of Dirichlet type

Average analysis of algorithms

Remark: Worst case vs. average analysis of algorithms

Average analysis of algorithms

Remark: Worst case vs. average analysis of algorithms

Elements for an average analysis

- An algorithm $\mathcal{A}$ whose inputs belong to some set Ω
- A cost function $X : \Omega \to \mathbb{R}^+$ that describes the algorithm (bit complexity, size of the output, memory/space complexity, ...)
- A size function: $\Omega = \bigcup_n \Omega_n$
- Each set Ω_n is endowed with a probability distribution (usually the uniform distribution)

Average analysis of algorithms

We consider a cost function $X : \Omega \to \mathbb{R}^+$

- [mean value] Compute the asymptotic mean value of X

$$\mathrm{E}_n[X] \underset{n\to\infty}{\sim}$$

ex: what is the average bit complexity of the algorithm when the input size n is large? Is it linear in n? Quadratic in n?

Average analysis of algorithms

We consider a cost function $X : \Omega \to \mathbb{R}^+$

- [mean value] Compute the asymptotic mean value of X

$$\mathrm{E}_n[X] \underset{n\to\infty}{\sim}$$

ex: what is the average bit complexity of the algorithm when the input size n is large? Is it linear in n? Quadratic in n?

- [variance] Compute the asymptotic of the variance

$$\mathrm{V}_n[X] \underset{n\to\infty}{\sim}$$

ex: is the probability to be far from the mean value asymptotically close to 0?

Average analysis of algorithms

We consider a cost function $X : \Omega \to \mathbb{R}^+$

- [mean value] Compute the asymptotic mean value of X

$$\mathrm{E}_n[X] \underset{n\to\infty}{\sim}$$

ex: what is the average bit complexity of the algorithm when the input size n is large? Is it linear in n? Quadratic in n?

- [variance] Compute the asymptotic of the variance

$$\mathrm{V}_n[X] \underset{n\to\infty}{\sim}$$

ex: is the probability to be far from the mean value asymptotically close to 0?

- [limit law] what is the limit law of X

$$\frac{X - \mathrm{E}_n[X]}{\sigma_n(X)} \underset{n\to\infty}{\to}$$

ex: what is asymptotically the probability that X is in the interval $[a, b]$?

On the Euclidean algorithm

We start from two positive integers u_0 and u_1

$$u_0 = u_1 \left[\frac{u_0}{u_1}\right] + u_2$$

$$u_1 = u_2 \left[\frac{u_1}{u_2}\right] + u_3$$

$$\vdots$$

$$u_{m-1} = u_m \left[\frac{u_{m-1}}{u_m}\right] + u_{m+1}$$

$$u_{m+1} = \gcd(u_0, u_1)$$

$$u_{m+2} = 0$$

Euclid algorithm and continued fractions

We start with two coprime integers u_0 and u_1

$$u_0 = u_1 a_1 + u_2$$

Euclid algorithm and continued fractions

We start with two coprime integers u_0 and u_1

$$u_0 = u_1 a_1 + u_2$$

$$\vdots$$

$$u_{m-1} = u_m a_m + u_{m+1}$$

$$u_m = u_{m+1} a_{m+1} + 0$$

$$u_{m+1} = 1 = \gcd(u_0, u_1)$$

Euclid algorithm and continued fractions

We start with two coprime integers u_0 and u_1

$$u_0 = u_1 a_1 + u_2$$

$$\frac{u_1}{u_0} = \frac{1}{a_1 + \frac{u_2}{u_1}}$$

$$u_1/u_0 = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots + \frac{1}{a_m + \frac{1}{a_{m+1}}}}}$$

On the number of Euclidean divisions for Euclid's algorithm

- Lamé (1850): *the worst case is linear w.r.t. the input binary size*
- Heilbron (69) and Dixon (70): *the mean number of divisions is linear w.r.t. the input binary size*
- Hensley (1994): *the number of divisions follows a gaussian limit law*

Number of steps for the Euclid algorithm

Let $L(u, v)$ stand for number of steps with $0 < v < u$

- Worst case

$$L(u, v) = O(\log v) \qquad (\leq 5 \log_{10} v, \text{ Lamé } 1844)$$

Number of steps for the Euclid algorithm

Let $L(u, v)$ stand for number of steps with $0 < v < u$

- Worst case

$$L(u, v) = O(\log v) \qquad (\leq 5 \log_{10} v,\ \text{Lamé } 1844)$$

- Mean case $\qquad 0 < v < u \leq N \qquad \gcd(u, v) = 1$

  Consider

$$\Omega_m := \{(u_1, u_2) \in \mathbb{N}^2,\ 0 \leq u_1, u_2 \leq m\}$$

endowed with the uniform distribution

$$\mathbb{E}_N[L] \sim \frac{12 \log 2}{\pi^2} \cdot \log N$$

[see Knuth, Vol. 2]

Number of steps for the Euclid algorithm

Let $L(u,v)$ stand for number of steps with $0 < v < u$

- Worst case

$$L(u,v) = O(\log v) \qquad (\leq 5 \log_{10} v, \text{ Lamé } 1844)$$

- Mean case $0 < v < u \leq N$ $\qquad \gcd(u,v) = 1$

$$\mathbb{E}_N[L] \sim \frac{12 \log 2}{\pi^2} \cdot \log N + \eta + O(N^{-\gamma})$$

η Porter's constant

asymptotically normal distribution

[Heilbronn'69,Dixon'70,Porter'75,Hensley'94,Baladi-Vallée'05...]

Formal power series with coefficients in $\mathbb{F}_q$

Formal power series

Let q be a power of a prime number p

We have the correspondence

- $\mathbb{Z} \sim \mathbb{F}_q[X]$
- $\mathbb{Q} \sim \mathbb{F}_q(X)$
- $\mathbb{R} \sim \mathbb{F}_q((X^{-1}))$

$$f = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_0 + a_{-1} X^{-1} + \cdots$$

Laurent formal power series

Formal power series

Let $f \in \mathbb{F}_q((X^{-1}))$ $\qquad f \neq 0$

$$f = a_n X^n + a_{n-1} X^{n-1} + \cdots \qquad a_n \neq 0$$

- Degree $\qquad \deg f = n$
- Distance $\qquad |f| = q^{\deg f}$

Ultrametric space

$$|f + g| \leq \max(|f|, |gl)$$

No carry propagation!

Generating functions

$$\Omega = \{(P, Q) \in \mathbb{F}_q[X]^2 \ : \ Q \text{ monic}, \ P = 0 \text{ or } \deg P < \deg Q\}$$

$$\Omega_m = \{(P, Q) \in \Omega \ : \ \deg Q = m\}$$

- Size of $(P, Q) := \deg Q$
- Generating function

$$T_\Omega(z) := \sum_{m \geq 0} |\Omega_m| z^m = \sum_{(P,Q) \in \Omega} z^{\deg (Q)}$$

- Fact

$$|\Omega_m| = q^{2m}$$

$$T_\Omega(z) := \sum_{m \geq 0} |\Omega_m| z^m = \frac{1}{1 - q^2 z}$$

A fundamental bijection

Euclid algorithm $\rightsquigarrow$ (P, Q) is uniquely determined by

partial quotients $(A_1, \cdots, A_L)$+ gcd (monic)

A fundamental bijection

Euclid algorithm $\rightsquigarrow$ (P, Q) is uniquely determined by

partial quotients $(A_1, \cdots, A_L)$+ gcd (monic)

$$\deg(Q) = \sum \deg(A_i) + \deg(\mathrm{gcd})$$

Ultrametricity!

A fundamental bijection

Euclid algorithm $\rightsquigarrow$ (P, Q) is uniquely determined by

partial quotients $(A_1, \cdots, A_L)$+ gcd (monic)

$$\deg(Q) = \sum \deg(A_i) + \deg(\mathrm{gcd})$$

Ultrametricity!

$\mathcal{G} = \{P \in \mathbb{F}_q[X] \,:\, \deg P \geq 1\}$ partial quotients
$\mathcal{U} = \{P \in \mathbb{F}_q[X] \,:\, P \text{ is monic}\}$ gcd

Fact $\quad \Omega = \mathtt{Seq}(\mathcal{G}) \times \mathcal{U}$

$\mathtt{Seq}(\mathcal{G})$:= finite sequences of elements of $\mathcal{G}$

$$\Omega = \{(P, Q) \in \mathbb{F}_q[X]^2 \,:\, Q \text{ monic},\ P = 0 \text{ or } \deg P < \deg Q\}$$

Generating functions

$$\Omega = \{(P, Q) \in \mathbb{F}_q[X]^2 \,:\, Q \text{ monic},\ P = 0 \text{ or } \deg P < \deg Q\}$$
$$\mathcal{G}_m = \{P \in \mathbb{F}_q[X] \,:\, \deg P \geq 1,\ \deg(P) = m\} \qquad \text{quotients}$$
$$\mathcal{U}_m = \{P \in \mathbb{F}_q[X] \,:\, P \text{ is monic},\ \deg(P) = m\} \qquad \text{gcd}$$

- Generating function

$$U(z) := \sum_{m \geq 0} |\mathcal{U}_m| z^m = \frac{1}{1 - qz}$$

$$G(z) := \sum_{m \geq 0} |\mathcal{G}_m| z^m = (q-1)\left(\frac{1}{1-qz} - 1\right) = \frac{(q-1)qz}{1-qz}$$

- Fact $\qquad \Omega = \mathtt{Seq}(\mathcal{G}) \times \mathcal{U}$

$$\rightsquigarrow T_\Omega(z) = \left(\sum_{k \geq 0} G^k(z)\right) \times U(z) = \frac{1}{1 - G(z)} \times U(z)$$

$$T_\Omega(z) = \sum \frac{1}{1 - q^2 z}, \qquad |\Omega_m| = q^{2m}$$

Additive costs

Let c be a cost defined on the set $\mathcal{G}$ of quotients

Additive costs

Let c be a cost defined on the set $\mathcal{G}$ of quotients

Additive cost C on Ω

$$C(P, Q) := \sum_{i=1}^{L(P,Q)} c(A_i)$$

The A_i are the quotients

Example: $C = 1$ Number of steps

We introduce a further variable for the cost u

Additive costs

Let c be a cost defined on the set $\mathcal{G}$ of quotients

$\rightsquigarrow$ generating functions with two variables

Additive costs

Let c be a cost defined on the set $\mathcal{G}$ of quotients

$$S_c(z,u) = \sum_{P \in \mathcal{G}} z^{\deg P} \cdot u^{c(P)} \qquad \mathcal{G} = \{P \in \mathbb{F}_q[X] \, : \, \deg P \geq 1\}$$

$$S_c(z,u) = \sum_{m,k} |\{P \in \mathcal{G}, \deg P = m, \ c(P) = k\}| \, z^m u^k$$

$$T_c(z,u) = \sum_{(P,Q) \in \Omega} z^{\deg Q} \cdot u^{C(P,Q)}$$

$$T_C(z,u) = \sum_{m,k} |\{(P,Q) \in \Omega, \deg Q = m, \ C(P,Q) = k\}| \, z^m u^k$$

Additive costs

Let c be a cost defined on the set $\mathcal{G}$ of quotients

$$S_c(z,u) = \sum_{P \in \mathcal{G}} z^{\deg P} \cdot u^{c(P)}$$

$$T_c(z,u) = \sum_{(P,Q) \in \Omega} z^{\deg Q} \cdot u^{C(P,Q)}$$

Fact

$$T_\Omega(z) = \frac{1}{1 - G(z)} \cdot U(z)$$

$$T_C(z,u) = \frac{1}{1 - S_c(z,u)} \cdot U(z)$$

An example of an additive cost

$$\Omega = \{(P,Q) \in \mathbb{F}_q[X]^2 \,:\, Q \text{ monic},\ P = 0,\ \text{or } \deg P < \deg Q\}$$

$$\mathcal{G} = \{P \in \mathbb{F}_q[X] \,:\, \deg P \geq 1\} \quad \mathcal{U} = \{P \in \mathbb{F}_q[X] \,:\, P \text{ is monic}\}$$

An example of an additive cost

- $c = 1$, $C =$ number of steps for Euclid algorithm

$$S_c(z, u) = u \cdot G(z)$$

$$G(z) = \frac{(q-1)qz}{1-qz} \qquad \mathcal{G} = \{P \in \mathbb{F}_q[X] \,:\, \deg P \geq 1\}$$

$$T_C(z, u) = \frac{1}{1-uG} \cdot U(z)$$

How to get expectations?

$$\Omega = \{(P,Q) \in \mathbb{F}_q[X]^2 \;:\; Q \text{ monic},\; P = 0,\; \text{or } \deg P < \deg Q\}$$

$$\mathcal{G} = \{P \in \mathbb{F}_q[X] \;:\; \deg P \geq 1\} \qquad \text{partial quotients}$$

$$\mathcal{U} = \{P \in \mathbb{F}_q[X] \;:\; P \text{ is monic}\} \qquad \text{gcd}$$

$$T_\Omega(z) = \frac{1}{1 - G(z)} \cdot U(z), \qquad T_C(z,u) = \frac{1}{1 - S_c(z,u)} \cdot U(z)$$

$$S_c(z,u) = \sum_{m,k} |\{P \in \mathcal{G},\, \deg P = m,\; c(P) = k\}|\, z^m u^k$$

$$T_C(z,u) = \sum_{m,k} |\{(P,Q) \in \Omega,\, \deg Q = m,\; C(P,Q) = k\}|\, z^m u^k$$

How to get expectations?

$$T_\Omega(z) = \frac{1}{1 - G(z)} \cdot U(z), \quad T_C(z,u) = \frac{1}{1 - S_c(z,u)} \cdot U(z)$$

$$S_c(z,u) = \sum_{m,k} |\{P \in \mathcal{G},\ \deg\ P = m,\ c(P) = k\}|\, z^m u^k$$

$$T_C(z,u) = \sum_{m,k} |\{(P,Q) \in \Omega,\ \deg\ Q = m,\ C(P,Q) = k\}|\, z^m u^k$$

By taking derivatives w.r.t. u

$$\frac{\partial}{\partial u} T_C|_{u=1} = \sum_m k|\{(P,Q) \in \Omega, \deg\ Q = m, C(P,Q) = k\}|z^m$$

$$\mathbb{E}_m[C] = \frac{[z^m]\frac{\partial}{\partial u} T_C(z,u)|_{u=1}}{q^{2m}}$$

$\rightsquigarrow$Expectation, variance, asymptotic Gaussian law

Number of steps

$$T_L(z,u) = \frac{1}{1-uG} \cdot U(z)$$

$$\frac{\partial}{\partial u} T_L\big|_{u=1} = G\left(\frac{1}{1-G}\right)^2 \cdot U(z)$$

Expectation $\quad \mathbb{E}_m[L] = \dfrac{[z^m]\frac{\partial}{\partial u} T_L(z,u)\big|_{u=1}}{q^{2m}}$

Number of steps

$$T_L(z,u) = \frac{1}{1-uG} \cdot U(z)$$

$$\frac{\partial}{\partial u} T_L|_{u=1} = G \left(\frac{1}{1-G} \right)^2 \cdot U(z)$$

Expectation $\quad \mathbb{E}_m[L] = \frac{[z^m]\frac{\partial}{\partial u} T_L(z,u)|_{u=1}}{q^{2m}}$

Looking for singularities

$$G(z) = \frac{(q-1)qz}{1-qz} \quad \text{singularity } 1/q$$

$$\left(\frac{1}{1-G} \right)^2 = \frac{1-qz}{1-q^2 z} \text{ singularity } 1/q^2$$

The smallest pole is $1/q^2$

Number of steps

$$T_L(z,u) = \frac{1}{1-uG} \cdot U(z)$$

$$\frac{\partial}{\partial u} T_L|_{u=1} = G\left(\frac{1}{1-G}\right)^2 \cdot U(z)$$

Expectation $\quad \mathbb{E}_m[L] = \frac{[z^m]\frac{\partial}{\partial u} T_L(z,u)|_{u=1}}{q^{2m}}$ $\quad$ linear in m

$$T_\Omega(z) = \frac{1}{1-G} \cdot U(z) = \frac{1}{1-q^2 z} \text{ singularity } 1/q^2 \rightsquigarrow q^{2m}$$

$$\left(\frac{1}{1-G}\right)^2 = \frac{1-qz}{1-q^2 z} \text{ singularity } 1/q^2 \text{ of order } 2 \rightsquigarrow mq^{2m}$$

Costs for Euclid algorithm

- Theorem [Vallée-Lhote]
  $L :=$ number of steps

$$\mathbb{E}_m[L] = \frac{q-1}{q} \cdot m \qquad \mathbb{V}_m[L] = \cdot \frac{q-1}{q^2} \cdot m$$

  Gaussian law

- Theorem [B.-Nakada-Natsui-Vallée]
  $N :=$ number of non-zero monomials

$$\mathbb{E}_m[N] = 2 \cdot \frac{q-1}{q} \cdot m + O(1) \qquad \mathbb{V}_m[N] = 2 \cdot \frac{q-1}{q^2} \cdot m + O(1)$$

  Gaussian law

Asymptotic Gaussian law

Let R be a cost defined on Ω

$$\mathbb{P}_m\left[(P,Q)\in\Omega_m, \frac{R(P,Q)-a_m}{\sqrt{b_m}}\le y\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-t^2/2}dt + r_m(y)$$

$(r_m)_m$ is sequence of functions $r_m\colon \mathbb{R}\to\mathbb{R}$, with

$$\lim_{m\to\infty}\sup\{r_m(y)\ :\ y\in\mathbb{R}\}=0$$

$$\mathbb{E}_m[R]\sim a_m, \quad \mathbb{V}_m[R]\sim b_m.$$

- Combinatorics is ubiquitous
- An interplay between discrete and continuous structures

  $\rightsquigarrow$ Concrete mathematics: A Foundation for Computer Science [Graham-Knuth-Patashnik]
- A domain of inter/transdisciplinarity