Some numeration systems

Savinien Kreczman

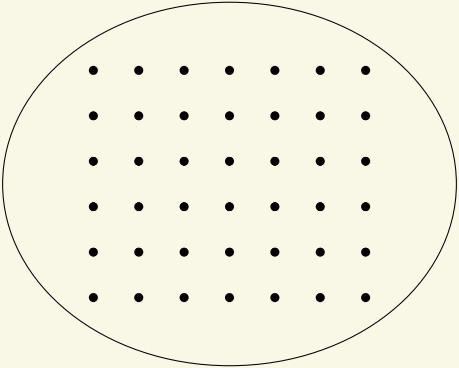
31 May 2025

A message from our sponsors

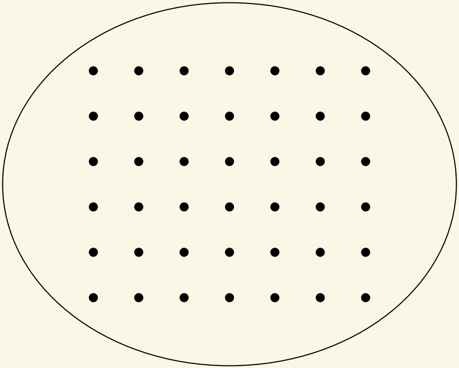
Avec le soutien de Wallonie-Bruxelles International

Why not try an Erasmus in Liège some time? We have an agreement with ČVUT and a research group in discrete mathematics too.

Integers

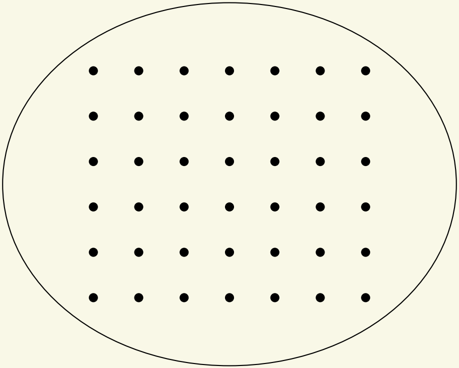


Integers



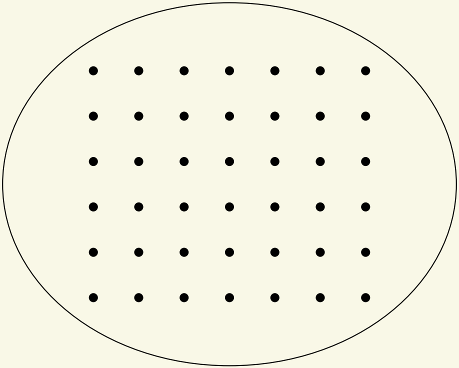
The number of points represented above is commonly written as 42 in decimal,

Integers

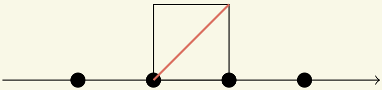


The number of points represented above is commonly written as 42 in decimal, but if it were in the memory of a computer it would be written as 101010.

Integers



The number of points represented above is commonly written as 42 in decimal, but if it were in the memory of a computer it would be written as 101010.
"Numbers exist independently of the way we represent them"



The red segment has length $\sqrt{2}$, which in decimal is written as $1.414213\cdots$. In binary, this number would be written as $1.0110101\cdots$.

Numeration systems

A *numeration system* is a way to associate a set of words with a set of numbers.

Definition

A *numeration system* over the *domain* $\mathbb{D} \in \{\mathbb{N}, \mathbb{R}^+\}$ is given by:

- *representation map* $\mathsf{rep}\colon \mathbb{D} \to A^*|A^+$
- *evaluation map* $\mathsf{val}\colon L \to \mathbb{D}$, where $\mathsf{rep}(\mathbb{D}) \subseteq L \subseteq A^*|A^+$

such that $\mathsf{val} \circ \mathsf{rep} = id_{\mathbb{D}}$.

Positional numeration systems

Decimal system to represent integers: the evaluation is

$$a_\ell \cdots a_0 \mapsto \sum_{j=0}^{\ell} a_j 10^j.$$

Binary system: the evaluation is

$$a_\ell \cdots a_0 \mapsto \sum_{j=0}^{\ell} a_j 2^j.$$

Positional numeration systems

Decimal system to represent integers: the evaluation is

$$a_\ell \cdots a_0 \mapsto \sum_{j=0}^{\ell} a_j 10^j.$$

Binary system: the evaluation is

$$a_\ell \cdots a_0 \mapsto \sum_{j=0}^{\ell} a_j 2^j.$$

This may be generalized: select a sequence $(U_n)_{n\in\mathbb{N}}$, then the evaluation map is

$$\mathrm{val}_U \colon a_\ell \cdots a_0 \mapsto \sum_{j=0}^{\ell} a_j U_j.$$

Positional numeration systems

Decimal system to represent integers: the evaluation is

$$a_\ell \cdots a_0 \mapsto \sum_{j=0}^{\ell} a_j 10^j.$$

Binary system: the evaluation is

$$a_\ell \cdots a_0 \mapsto \sum_{j=0}^{\ell} a_j 2^j.$$

This may be generalized: select a sequence $(U_n)_{n\in\mathbb{N}}$, then the evaluation map is

$$\mathsf{val}_U \colon a_\ell \cdots a_0 \mapsto \sum_{j=0}^{\ell} a_j U_j.$$

We ask that $U_0 = 1$, U is increasing and $\frac{U_{n+1}}{U_n}$ is bounded. The resulting numeration system is called a *positional numeration system*.

Greed as a quality

The evaluation map is not injective. There are several equivalent ways to choose one of the many words that evaluate to n:

- Select the greatest representation in radix order.
- Select the unique representation that verifies $a_\ell \neq 0$ and

$$\sum_{j=0}^{i-1} a_j U_j < U_i \text{ for all } i$$

- Use the greedy algorithm to define the representation of n: set ℓ such that $U_\ell \leq n < U_{\ell+1}$ and $r_\ell = n$, then set $a_i = \left\lfloor \frac{r_i}{U_i} \right\rfloor$ and $r_{i-1} = r_i - a_i U_i$.

A running example: Tribonacci

Consider the Tribonacci sequence given by

$$T_0 = 1,\ T_1 = 2,\ T_2 = 4 \text{ and } T_{n+3} = T_{n+2} + T_{n+1} + T_n \quad \forall n.$$

The representations of the first few natural numbers are

n	0	1	2	3	4	5	6	7	8
$\mathrm{rep}_T(n)$	ε	1	10	11	100	101	110	1000	1001

The word 111 also has value 7, but 1000 is the greedy representation.

Abstract numeration systems

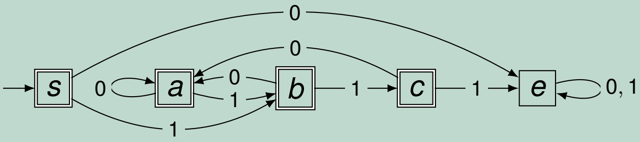
Rather than defining the weights of various positions, we define a numeration system by specifying its language, which is the set $L = \{\mathrm{rep}(n) | n \in \mathbb{N}\}$.
To define an abstract numeration system, we specify a regular language L. Once L is ordered by the radix order, there is a unique increasing bijection between it and $(\mathbb{N}, \leq)$. This bijection is the evaluation map, and its inverse is the representation map.

Return to Tribonacci

Consider the case of the language

$$L = \{0,1\}^* \setminus (0\{0,1\}^* \cup \{0,1\}^*111\{0,1\}^*)\,,$$

which is recognized by the following automaton:



The first few words in L in the radix order are

$$\varepsilon, 1, 10, 11, 100, 101, 110, 1000, \ldots,$$

so we have for instance $\mathsf{val}(110) = 6$ and $\mathsf{rep}(7) = 1000$. We find again the Tribonacci numeration system.

Substitutions

Substitution: map $\mu\colon A^* \to B^*$ such that $\mu(uv) = \mu(u)\mu(v)$ for all $u, v \in A^*$, that is non-erasing.
Fixed point of μ: $u \in A^{\mathbb{D}}$ such that $\mu(u) = u$. *Seed* of a fixed point: u_0. Knowing the seed is enough to know the whole fixed point, if it is a growing letter.

Substitutions

Substitution: map $\mu\colon A^* \to B^*$ such that $\mu(uv) = \mu(u)\mu(v)$ for all $u, v \in A^*$, that is non-erasing.
Fixed point of μ: $u \in A^{\mathbb{D}}$ such that $\mu(u) = u$. *Seed* of a fixed point: u_0. Knowing the seed is enough to know the whole fixed point, if it is a growing letter.
Tree representation: given a fixed point with seed a, build a tree $\mathcal{T}_{\mu,a}$ with root a and such that if $\mu(x) = y_0 \cdots y_{\ell-1}$, the node x has ℓ children labelled $y_0, \ldots, y_{\ell-1}$ with edges labelled $0, \ldots, \ell - 1$.

Return to Return to Tribonacci

Consider $\mu : \begin{cases} a & \mapsto & ab \\ b & \mapsto & ac \\ c & \mapsto & a \end{cases}$ and its fixed point starting with *a*:

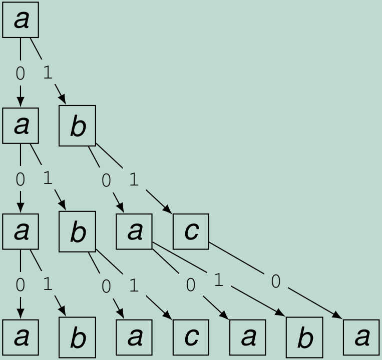
abacabaabacab....

Return to Return to Tribonacci

Consider μ: $\begin{cases} a & \mapsto & ab \\ b & \mapsto & ac \\ c & \mapsto & a \end{cases}$ and its fixed point starting with *a*:

abacabaabacab....

The initial segment of $\mathcal{T}_{\mu,a}$ is:

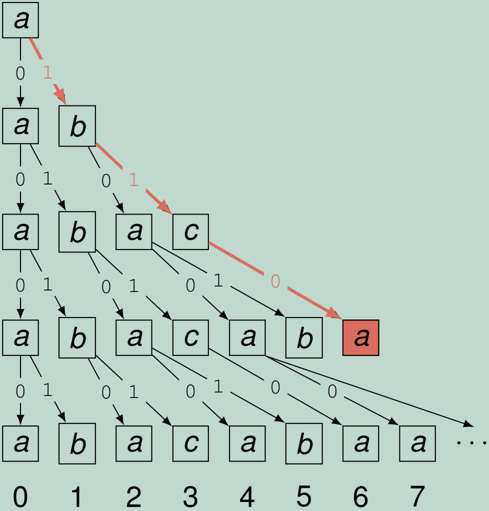


Dumont-Thomas numeration system

Definition

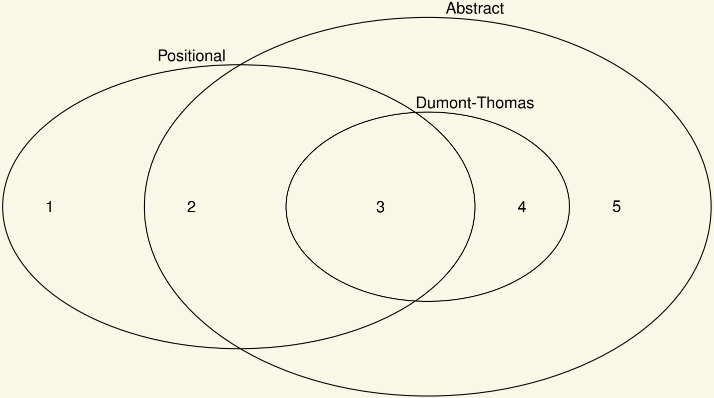
A Dumont-Thomas numeration system is obtained by organizing the tree in columns and representing n by the label of a shortest path from the root to a node in column n.

Return to Return to Return to Tribonacci



We have $\mathrm{rep}(6) = 110$, $\mathrm{rep}(7) = 1000$: we find again the same numeration system.

Overlap between different families of systems



Specific examples:

1: $U_n = n^2$

2: $U_0 = 1,\ U_{n+1} = 3.2^n$, language $\varepsilon \cup 0\{0,1\}^*\{0,1,2\}$.

3: Tribonacci

4: $\rho\colon\ a \mapsto abb,\ b \mapsto ab$

5: $L = 1^*2^*$

Base 10 for real numbers

Our usual decimal numeration system can represent not only natural numbers but also reals, using right-infinite words.

$$\mathsf{val}\colon 0.w_1 w_2 w_3 \cdots \mapsto \sum_{i=1}^{\infty} \frac{w_i}{10^i}.$$

The map rep can be defined using an algorithm: first set $r_0 = x$, then iteratively define

$$w_{i+1} = \lfloor 10 r_i \rfloor \text{ and } r_{i+1} = 10 r_i - \lfloor 10 r_i \rfloor .$$

β-numeration

This algorithm can be used not just for 10, but for any value of $\beta > 1$, and can also be used for to represent 1.

Example

Set β the positive root of $x^3 - x^2 - x - 1$, which is approximately 1.84. If we want to represent 1 in this base, we compute

i	βr_{i-1}	w_i	r_i
1	$\beta \simeq 1.84$	1	$\beta - 1$
2	$\beta^2 - \beta \simeq 1.54$	1	$\beta^2 - \beta - 1$
3	$\beta^3 - \beta^2 - \beta = 1$	1	0

As a result, $\mathrm{rep}_\beta(1) = 1110^\omega$. We also deduce $\mathrm{val}_\beta(110)^\omega = 1$.

Figuring out the language

Question

Different words may have the same value. How do we identify the *representation* of a number?

Greedy expansions are such that

$$\beta^j > \sum_{k<j} a_k \beta^k \quad \forall j,$$

thus the word $a_k a_{k+1} \cdots$ should have a value less than 1.

Proposition

An infinite word a is the expansion of some real number if, and only if, it contains no factor lexicographically greater than or equal to $d_\beta^*(1)$.

Return to Return to Return to Return to Tribonacci

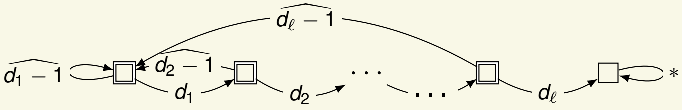
A word is the representation of a number in $[0, 1)$ if it contains no factor lexicographically greater than or equal to $(110)^\omega$. A finite word is the representation of a number in $[0, 1)$ if it contains no factor lexicographically greater than 111.
This language is 0^*L, where L is the language of the Tribonacci numeration system over the integers.

The general picture

Select a $\beta > 1$ such that the β-expansion of 1 is $d_1 \cdots d_\ell$ (simple Parry number). One can define a substitution

$$\mu_\beta \colon 0 \mapsto 0^{d_1}1,\ 1 \mapsto 0^{d_2}2,\ \ldots (\ell - 1) \mapsto 0^{d_l},$$

To this substitution is associated an automaton



One can define a Dumont-Thomas numeration system from the substitution and an abstract numeration system from the automaton. The incidence matrix of the automaton has characteristic polynomial $x^\ell - d_1 x^{\ell-1} - \ldots - d_\ell$. With the correct initial conditions, we obain a positional numeration system identical to the two above. We have $U_n/U_{n-1} \to \beta$.