Abstract numeration systems

Emilie Charlier

Department of Mathematics
University of Liège

Herbertov 2007

Integer base numeration system, $k \geq 2$

$$n = \sum_{i=0}^{\ell} c_i \, k^i, \quad \text{with} \quad c_i \in \Sigma_k = \{0, \ldots, k-1\}, \; c_\ell \neq 0$$

Any integer n corresponds to a word $\mathrm{rep}_k(n) = c_\ell \cdots c_0$ over Σ_k.

Definition

A set $X \subseteq \mathbb{N}$ is *k-recognizable* if $\mathrm{rep}_k(X) \subseteq \Sigma_k^*$ is a regular language (accepted by a DFA).

Divisibility criteria

If $X \subseteq \mathbb{N}$ is ultimately periodic,
then X is k-recognizable for any $k \geq 2$.

(Non-standard) system built upon a sequence $U = (U_i)_{i\geq 0}$ of integers

$$n = \sum_{i=0}^{\ell} c_i\, U_i, \quad \text{with} \quad c_\ell \neq 0 \quad \text{greedy expansion}$$

Any integer n corresponds to a word $\mathrm{rep}_U(n) = c_\ell \cdots c_0$.

Definition

A set $X \subseteq \mathbb{N}$ is *U-recognizable* if $\mathrm{rep}_U(X) \subseteq \Sigma_k^*$ is a regular language (accepted by a DFA).

Some conditions on $U = (U_i)_{i \geq 0}$

- $U_i < U_{i+1}$, *non-ambiguity*
- $U_0 = 1$, *any integer can be represented*
- $\frac{U_{i+1}}{U_i}$ is bounded, *finite alphabet of digits* A_U

Some conditions on $U = (U_i)_{i \geq 0}$

- $U_i < U_{i+1}$, *non-ambiguity*
- $U_0 = 1$, *any integer can be represented*
- $\frac{U_{i+1}}{U_i}$ is bounded, *finite alphabet of digits* A_U

Example $(U_i = 2^{i+1} : 2, 4, 8, 16, 32, \dots)$

you cannot represent odd integers !

Some conditions on $U = (U_i)_{i\geq 0}$

- $U_i < U_{i+1}$, *non-ambiguity*
- $U_0 = 1$, *any integer can be represented*
- $\frac{U_{i+1}}{U_i}$ is bounded, *finite alphabet of digits* A_U

Example ($U_i = 2^{i+1}$: $2, 4, 8, 16, 32, \ldots$)

you cannot represent odd integers !

Example ($U_i = (i + 1)!$: $1, 2, 6, 24, \ldots$)

Any integer n can be uniquely written as

$$n = \sum_{i=1}^{\ell} c_i \, i! \quad \text{with} \quad 0 \leq c_i \leq i$$

Fraenkel'85, Lenstra'06 (EMS Newsletter, profinite numbers)

A nice setting

Take $(U_i)_{i\geq 0}$ satisfying a linear recurrence equation,

$$U_{i+k} = a_{k-1}U_{i+k-1} + \cdots + a_0 U_i, \quad a_j \in \mathbb{Z},\ a_0 \neq 0.$$

Example ($U_{i+2} = U_{i+1} + U_i,\ U_0 = 1,\ U_1 = 2$)

Use greedy expansion, $\ldots, 21, 13, 8, 5, 3, 2, 1$

1	1	8	10000	15	100010
2	10	9	10001	16	100100
3	100	10	10010	17	100101
4	101	11	10100	18	101000
5	1000	12	10101	19	101001
6	1001	13	100000	20	101010
7	1010	14	100001	21	1000000

The "pattern" 11 is forbidden, $A_U = \{0, 1\}$.

Question

Let $U = (U_i)_{i\geq 0}$ be a strictly increasing sequence of integers,

is the whole set $\mathbb{N}$ U-recognizable ?
i.e., is $\mathcal{L}_U = \mathrm{rep}_U(\mathbb{N})$ regular ?

Even if U is linear, the answer is not completely known. . .

Question

Let $U = (U_i)_{i\geq 0}$ be a strictly increasing sequence of integers,

is the whole set $\mathbb{N}$ U-recognizable ?
i.e., is $\mathcal{L}_U = \mathrm{rep}_U(\mathbb{N})$ regular ?

Even if U is linear, the answer is not completely known...

Theorem (Shallit '94)

If $\mathcal{L}_U$ is regular,
then $(U_i)_{i\geq 0}$ satisfies a linear recurrent equation.

Question

Let $U = (U_i)_{i\geq 0}$ be a strictly increasing sequence of integers,

is the whole set $\mathbb{N}$ U-recognizable ?
i.e., is $\mathcal{L}_U = \mathrm{rep}_U(\mathbb{N})$ regular ?

Even if U is linear, the answer is not completely known. . .

Theorem (Shallit '94)

If $\mathcal{L}_U$ is regular,
then $(U_i)_{i\geq 0}$ satisfies a linear recurrent equation.

Theorem (N. Loraud '95, M. Hollander '98)

They give (technical) sufficient conditions for $\mathcal{L}_U$ to be regular:
"the characteristic polynomial of the recurrence has a special form".

Best known case : linear "Pisot systems"

If the characteristic polynomial of $(U_i)_{i\geq 0}$ is the minimal polynomial of a Pisot number θ then "everything" is fine:

Best known case : linear "Pisot systems"

If the characteristic polynomial of $(U_i)_{i\geq 0}$ is the minimal polynomial of a Pisot number θ then "everything" is fine:
$\mathcal{L}_U$ *is regular, addition preserves recognizability, logical first order characterization of recognizable sets*, ...
"*Just*" like in the integer case : $U_i \simeq \theta^i$.

A. Bertrand '89, C. Frougny, B. Solomyak, D. Berend, J. Sakarovitch, V. Bruyère and G. Hansel '97, ...

Definition

A Pisot (resp. Salem, Perron) number is an algebraic integer $\alpha > 1$ such that its Galois conjugates have modulus < 1 (resp. ≤ 1, $< \alpha$).

A question by P. Lecomte

A question by P. Lecomte

- Everybody takes first a sequence $(U_k)_{k\geq 0}$

A question by P. Lecomte

- Everybody takes first a sequence $(U_k)_{k\geq 0}$
- then ask for the language $\mathcal{L}_U$ of the numeration to be regular and play with recognizable sets

A question by P. Lecomte

- Everybody takes first a sequence $(U_k)_{k\geq 0}$
- then ask for the language $\mathcal{L}_U$ of the numeration to be regular and play with recognizable sets
- Why not proceed backwards ?

A question by P. Lecomte

- Everybody takes first a sequence $(U_k)_{k\geq 0}$
- then ask for the language $\mathcal{L}_U$ of the numeration to be regular and play with recognizable sets
- Why not proceed backwards ?

Remark

Let $x, y \in \mathbb{N}$, $x < y \Leftrightarrow \mathrm{rep}_U(x) <_{gen} \mathrm{rep}_U(y)$.

A question by P. Lecomte

- Everybody takes first a sequence $(U_k)_{k\geq 0}$
- then ask for the language $\mathcal{L}_U$ of the numeration to be regular and play with recognizable sets
- Why not proceed backwards ?

Remark

Let $x, y \in \mathbb{N}$, $x < y \Leftrightarrow \mathrm{rep}_U(x) <_{gen} \mathrm{rep}_U(y)$.

Example (Fibonacci)

$6 < 7$ and $1001 <_{gen} 1010$ (same length)
$6 < 8$ and $1001 <_{gen} 10000$ (different lengths).

Definition (P. Lecomte, M.Rigo '01)

An *abstract numeration system* is a triple $S = (L, \Sigma, <)$ where L is a regular language over a totally ordered alphabet $(\Sigma, <)$. Enumerating the words of L with respect to the genealogical ordering induced by $<$ gives a one-to-one correspondence

$$\mathrm{rep}_S : \mathbb{N} \to L \qquad \mathrm{val}_S = \mathrm{rep}_S^{-1} : L \to \mathbb{N}.$$

remark

This generalizes “classical” Pisot systems like integer base systems or Fibonacci system.

Example (Positional)

$L = \{\varepsilon\} \cup \{1, \ldots, k-1\}\{0, \ldots, k-1\}^*$ or $L = \{\varepsilon\} \cup 1\{0, 01\}^*$

remark

This generalizes “classical” Pisot systems like integer base systems or Fibonacci system.

Example (Positional)

$L = \{\varepsilon\} \cup \{1, \dots, k-1\}\{0, \dots, k-1\}^*$ or $L = \{\varepsilon\} \cup 1\{0, 01\}^*$

Example (Non positional)

$L = a^*,\ \Sigma = \{a\}$

n	0	1	2	3	4	$\cdots$
$\mathrm{rep}(n)$	ε	a	aa	aaa	$aaaa$	$\cdots$

$L = \{a, b\}^*,\ \Sigma = \{a, b\},\ a < b$

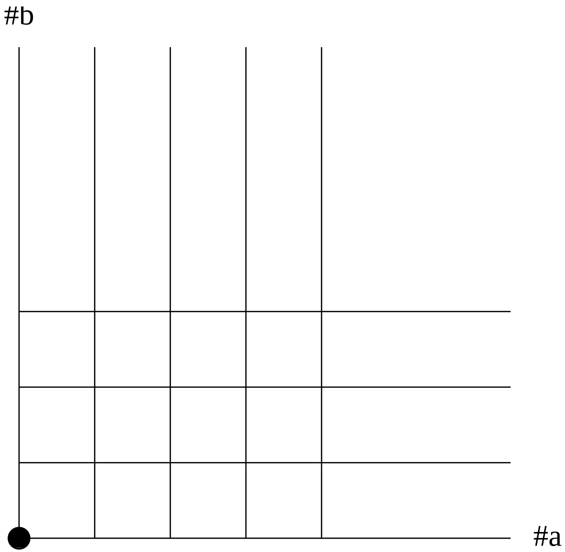
n	0	1	2	3	4	5	6	7	$\cdots$
$\mathrm{rep}(n)$	ε	a	b	aa	ab	ba	bb	aaa	$\cdots$

$L = a^* b^*,\ \Sigma = \{a < b\}$

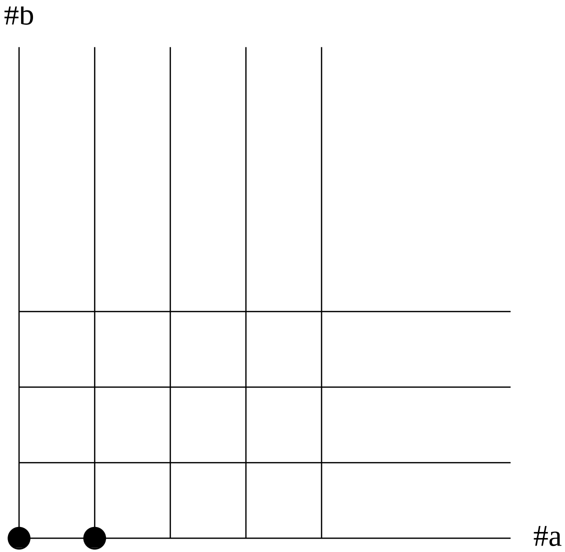
n	0	1	2	3	4	5	6	$\cdots$
$\mathrm{rep}(n)$	ε	a	b	aa	ab	bb	aaa	$\cdots$

$$\mathrm{val}(a^p b^q) = \frac{1}{2}(p+q)(p+q+1) + q$$

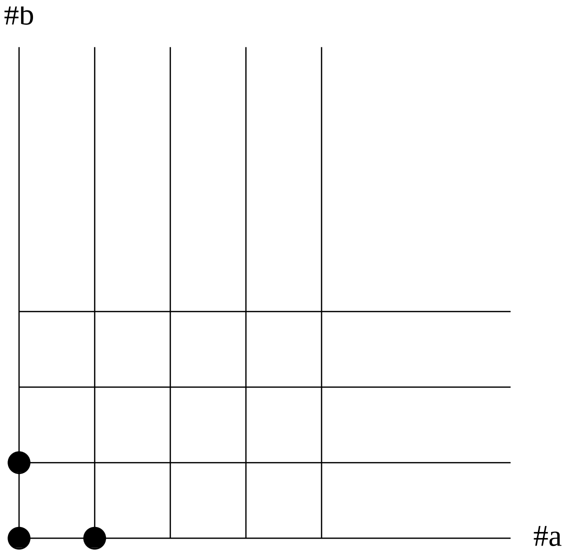
Example (continued...)



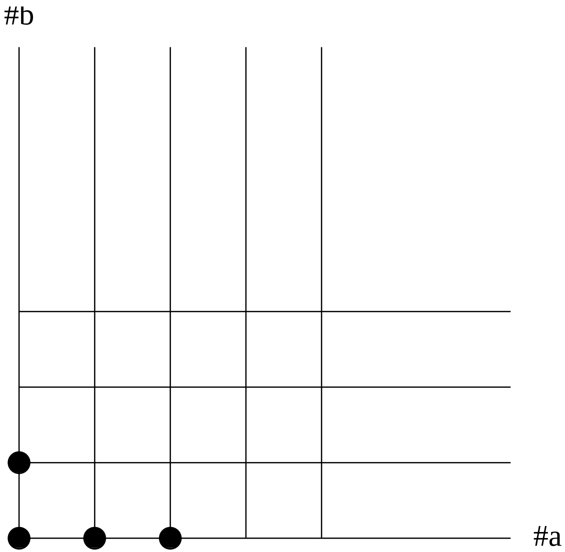
Example (continued...)



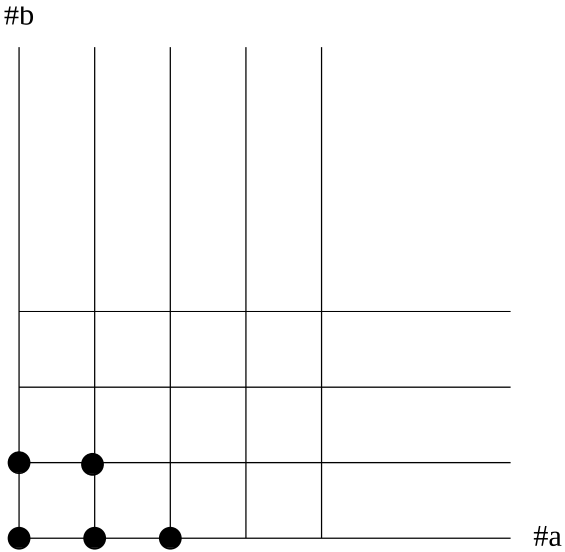
Example (continued...)



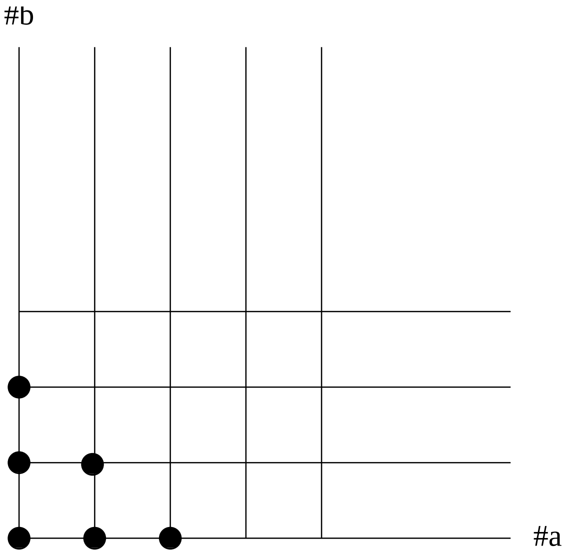
Example (continued...)



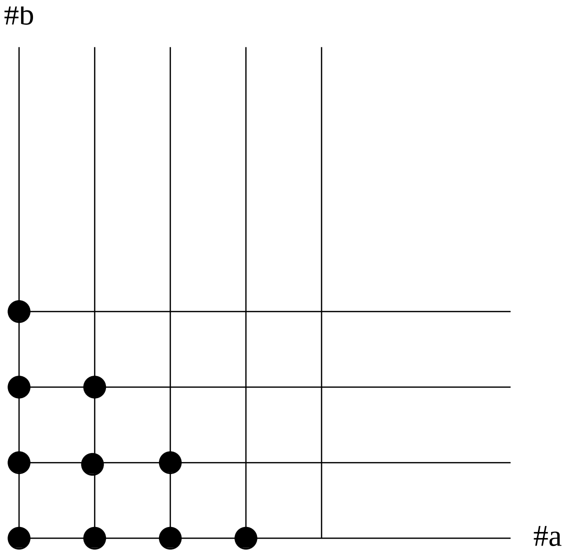
Example (continued...)



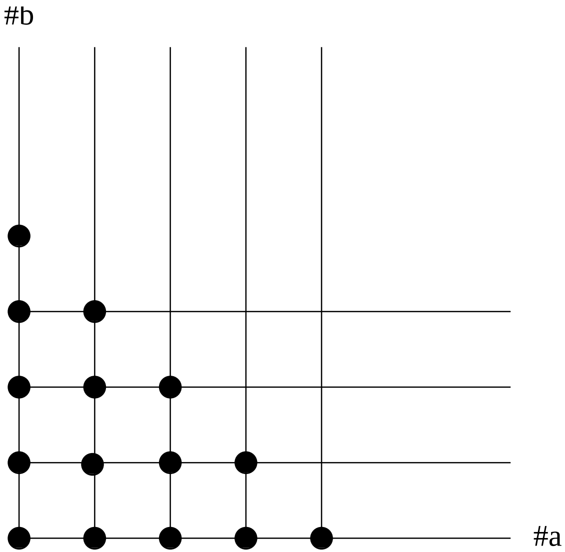
Example (continued...)



Example (continued...)



Example (continued...)



Definition of complexity

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a DFA accepting L.
For all $q \in Q$, $L_q = \{w \in \Sigma^* \mid \delta(q, w) \in F\}$.

$$\mathbf{u}_q(n) = \#(L_q \cap \Sigma^n) \quad \text{and} \quad \mathbf{v}_q(n) = \#(L_q \cap \Sigma^{\leq n}).$$

In particular, $\mathbf{u}_{q_0}(n) = \#(L \cap \Sigma^n)$.

Computing $\mathrm{val}_S : L \to \mathbb{N}$

If $\sigma w \in L_q$, $\sigma \in \Sigma$, $w \in \Sigma^+$, then

$$\mathrm{val}_{L_q}(\sigma w) = \mathrm{val}_{L_{q.\sigma}}(w) + v_q(|w|) - v_{q.\sigma}(|w| - 1) + \sum_{\sigma' < \sigma} u_{q.\sigma'}(|w|).$$

If $\sigma \in L_q \cap \Sigma$, then

$$\mathrm{val}_{L_q}(\sigma) = u_{L_q}(0) + \sum_{\sigma' < \sigma} u_{q.\sigma'}(0).$$

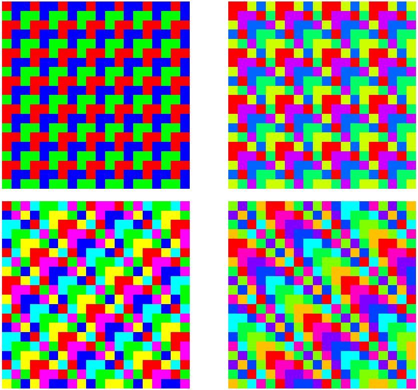
Many natural questions...

- What about S-recognizable sets ?
  - Are ultimately periodic sets S-recognizable for any S ?
  - For a given $X \subseteq \mathbb{N}$, can we find S s.t. X is S-recognizable ?
  - For a given S, what are the S-recognizable sets ?
- Can we compute "easily" in these systems ?
  - Addition, multiplication by a constant, ...
- Are these systems equivalent to something else ?
- Any hope for a Cobham's theorem ?
- Can we also represent real numbers ?
- Number theoretic problems like additive functions ?
- Dynamics, odometer, tilings, logic...

Theorem

Let $S = (L, \Sigma, <)$ be an abstract numeration system. Any ultimately periodic set is S-recognizable.

Example (For a^*b^* mod $3, 5, 6$ and 8)



Well-known fact (see Eilenberg's book)

The set of squares is never recognizable in any integer base system.

Example

Let $L = a^* b^* \cup a^* c^*$, $a < b < c$.

Well-known fact (see Eilenberg's book)

The set of squares is never recognizable in any integer base system.

Example

Let $L = a^*b^* \cup a^*c^*$, $a < b < c$.

0	1	2	3	4	5	6	7	8	9	$\cdots$
ε	a	b	c	aa	ab	ac	bb	cc	aaa	$\cdots$

Well-known fact (see Eilenberg's book)

The set of squares is never recognizable in any integer base system.

Example

Let $L = a^*b^* \cup a^*c^*$, $a < b < c$.

0	1	2	3	4	5	6	7	8	9	$\cdots$
ε	a	b	c	aa	ab	ac	bb	cc	aaa	$\cdots$

Theorem

If $P \in \mathbb{Q}[X]$ is such that $P(\mathbb{N}) \subseteq \mathbb{N}$ then there exists an abstract system S such that $P(\mathbb{N})$ is S-recognizable.

Consider multiplication by a constant...

Theorem

Let $S = (a^*b^*, \{a < b\})$. Multiplication by $\lambda \in \mathbb{N}$ preserves S-recognizability iff λ is an odd square.

Example

There exists $X_3 \subseteq \mathbb{N}$ such that X_3 is S-recognizable but such that $3X_3$ is not S-recognizable. (3 is not a square)

There exists $X_4 \subseteq \mathbb{N}$ such that X_4 is S-recognizable but such that $4X_4$ is not S-recognizable. (4 is an even square)

For any S-recognizable set $X \subseteq \mathbb{N}$, $9X$ or $25X$ is also S-recognizable.

Theorem

Let ℓ be a positive integer. For the abstract numeration system

$$S = (a_1^* \dots a_\ell^*, \{a_1 < \dots < a_\ell\}),$$

multiplication by $\lambda > 1$ preserves S-recognizability if and only if one of the following condition is satisfied :

- $\ell = 1$
- $\ell = 2$ *and λ is an odd square.*

Theorem ("Multiplication by a constant")

slender language	$\mathbf{u}_{q_0}(n) \in \mathcal{O}(1)$	*OK*
polynomial language	$\mathbf{u}_{q_0}(n) \in \mathcal{O}(n^k)$	*NOT OK*
exponential language with polynomial complement	$\mathbf{u}_{q_0}(n) \in 2^{\Omega(n)}$	*NOT OK*
exponential language with exponential complement	$\mathbf{u}_{q_0}(n) \in 2^{\Omega(n)}$	*OK ?*

Example

"Pisot" systems belong to the last class.