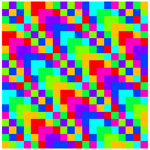
AUTOUR DES SYSTÈMES DE NUMÉRATION ABSTRAITS

Michel Rigo

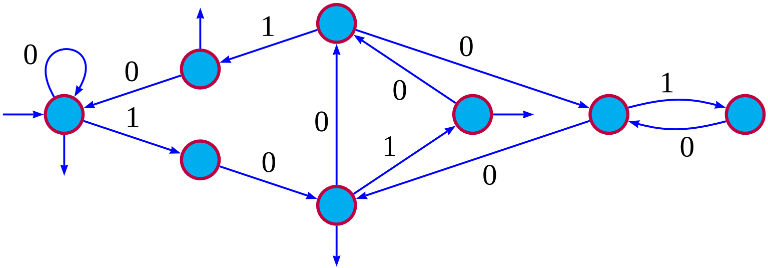
http://www.discmath.ulg.ac.be/
http://orbi.ulg.ac.be/handle/2268/124195

JMC 2012, journés SDA2-2012, LITIS Rouen, 11–13 juin 2012

UNIVERSITÉ de Liège



In the Chomsky's hierarchy, the simplest models of computation are finite automata accepting regular languages.



100100, 1000, 1000100, 0000001, . . .

With this model in mind, *what is a “simple” set of integers ?*

DEFINITION

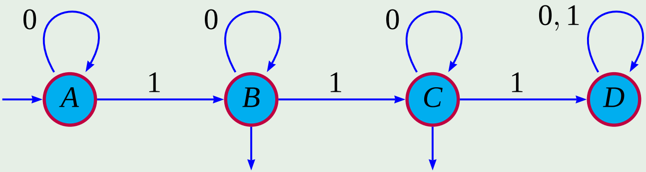
A set $X \subset \mathbb{N}$ is *k-recognizable*, if the set of base k expansions of the elements in X is accepted by some finite automaton, *i.e.*, $\mathrm{rep}_k(X)$ is a regular language.

Much “simpler” than a *recursive set* of integers for which there is an algorithm that decides whether or not a given number belongs to the set.

Some examples

A 2-recognizable set

$X = \{n \in \mathbb{N} \mid \exists i, j \geq 0 : n = 2^i + 2^j\} \cup \{1\}$

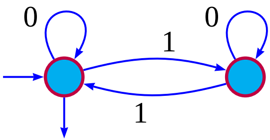


$$X = \{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 16, 17, 18, 20, 24, \ldots\}$$

$$\text{rep}_2(X) = \{1, 10, 11, 100, 101, 110, 1000, 1001, 1010, 1100, \ldots\}$$

Some examples

- The set of even integers is 2-recognizable.
- The Prouhet–Thue–Morse set is 2-recognizable,



$$X = \{n \in \mathbb{N} \mid s_2(n) \equiv 0 \bmod 2\}$$

$$X = \{0, 3, 5, 6, 9, 10, 12, 15, 17, 18, \ldots\}$$

$$\mathrm{rep}_2(X) = \{\varepsilon, 11, 101, 110, 1001, 1010, 1100, 1111, 10001, 10010, \ldots\}$$

- The set of powers of 2 is 2-recognizable.

More examples

Let $X = \{x_0 < x_1 < x_2 < \cdots\} \subseteq \mathbb{N}$. Define

$$\mathbf{R}_X := \limsup_{i\to\infty} \frac{x_{i+1}}{x_i} \text{ and } \mathbf{D}_X := \limsup_{i\to\infty} (x_{i+1} - x_i).$$

Gap theorem (Cobham'72)

Let $k \geq 2$. If $X \subseteq \mathbb{N}$ is a k-recognizable infinite subset of $\mathbb{N}$, then either $\mathbf{R}_X > 1$ or $\mathbf{D}_X < +\infty$.

A. Cobham, Uniform tag, Theory Comput. Syst. 6, (1972), 164–192.

Corollary

Let $k, t \geq 2$ be integers.
The set $\{n^t \mid n \geq 0\}$ is NOT k-recognizable.

S. Eilenberg, Automata, Languages, and Machines, 1974.

MORE EXAMPLES

MINSKY–PAPERT 1966

The set $\mathcal{P}$ of prime numbers is not k-recognizable.

A proof using the gap theorem :
Since $n! + 2, \ldots, n! + n$ are composite numbers, $\mathbf{D}_{\mathcal{P}} = +\infty$
Since $p_n \in (n \ln n, n \ln n + n \ln \ln n)$, $\mathbf{R}_{\mathcal{P}} = 1$

E. Bach, J. Shallit, Algorithmic number theory, MIT Press

M.-P. SCHÜTZENBERGER (1968)

No infinite subset of $\mathcal{P}$ can be recognized by a finite automaton.

BASE SENSITIVITY

Is this notion of recognizability base dependent ?

- Is the set of even integers 3-recognizable ? (exercise)
- Is the set of powers of 2 also 3-recognizable ?

$$2, 11, 22, 121, 1012, 2101, 11202, 100111, 200222, 1101221,$$

$$2210212, 12121201, 102020102, 211110211, 1122221122, 10022220021,$$

$$20122210112, 111022121001, 222122012002, 1222021101011, \ldots$$

Base sensitivity

Two integers $k, \ell \geq 2$ are *multiplicatively independent* if $k^m = \ell^n \Rightarrow m = n = 0$, i.e., if $\log k / \log \ell$ is irrational.

Cobham's theorem (1969)

Let $k, \ell \geq 2$ be two multiplicatively independent integers.
A set $X \subseteq \mathbb{N}$ is k-rec. AND ℓ-rec. IFF X is ultimately periodic, *i.e.*, X is a finite union of arithmetic progressions.

V. Bruyère, G. Hansel, C. Michaux, R. Villemaire, Logic and p-recognizable sets of integers, **BBMS**'94.

F. Durand, M. Rigo, On Cobham's theorem, to appear in Handbook of Automata.

Tool (Kronecker's theorem)

Let θ be an irrational number.
The sequence $(\{n\theta\})_{n\geq 0}$ is dense in $[0,1)$.

Base sensitivity

S. Eilenberg (p. 104): “*The proof is correct, long and hard. It is a challenge to find a more reasonable proof of this fine theorem*”

The easy part, *e.g.*, conversion between base 2 and base 4,

$$\begin{array}{c|c} 00 & 0 \\ 01 & 1 \\ 10 & 2 \\ 11 & 3 \end{array}$$

- such a transformation preserves regularity
- L is regular IFF 0^*L is regular

Base sensitivity

Some consequences of Cobham's theorem from 1969:

- k-recognizable sets are easy to describe but non-trivial,
- motivates characterizations of k-recognizability,
- motivates the study of "exotic" numeration systems,
- generalizations of Cobham's result to various contexts:

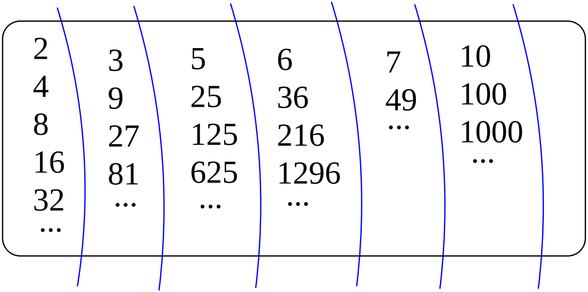
  multidimensional setting, logical framework, extension to Pisot systems, substitutive systems, fractals and tilings, simpler proofs, . . .

B. Adamczewski, J. Bell, G. Hansel, D. Perrin, F. Durand, V. Bruyère, F. Point, C. Michaux, R. Villemaire, A. Bès, J. Honkala, S. Fabre, C. Reutenauer, A.L. Semenov, L. Waxweiler, M.-I. Cortez, . . .

BASE SENSITIVITY

There are three kinds of sets:

- Ultimately periodic sets are recognizable in all bases,
- Sets that are k-recognizable for some k, and only k^m-recognizable, $m \geq 1$,
- Sets that are not k-recognizable.



multiplicative dependence is trivially an equivalence relation.

Logical characterization

Büchi–Bruyère Theorem

A set $X \subset \mathbb{N}^d$ is k-recognizable IFF it is definable by a first order formula in the extended Presburger arithmetic $\langle \mathbb{N}, +, V_k \rangle$.

$V_k(n)$ is the largest power of k dividing $n \geq 1$, $V_k(0) = 1$.

$$\varphi_1(x) \equiv V_2(x) = x$$

$$\varphi_2(x) \equiv (\exists y)(V_2(y) = y) \wedge (\exists z)(V_2(z) = z) \wedge x = y + z$$

$$\varphi_3(x) \equiv (\exists y)(x = y + y + y + y + 3)$$

from formula to automata from automata to formula...

Restatement of Cobham's thm.

Let $k, \ell \geq 2$ be two multiplicatively independent integers.
A set $X \subseteq \mathbb{N}$ is k-rec. AND ℓ-rec. IFF X is definable in $\langle \mathbb{N}, + \rangle$.

Logical characterization

Applications to decision problems and, in computer science, to model-checking and formal verification.

Theorem (Bruyère 1985)

The theory $\langle \mathbb{N}, +, V_k \rangle$ is decidable.

Example

Let X be a k-recognizable set of integers.
Decide whether or not X is ultimately periodic ?

Let $\varphi(x)$ be a formula such that $a \in X$ IFF $\varphi(a)$ holds true.
Consider the sentence

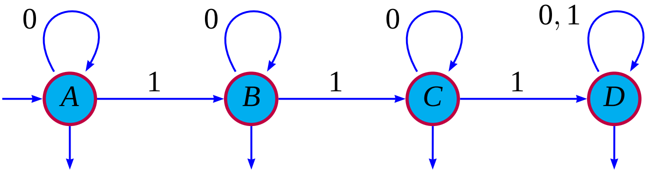
$$(\exists p)(\exists i)(\forall a \geq i)(\varphi(a) \Leftrightarrow \varphi(a+p)).$$

Theorem (Cobham 1972)

An infinite word $\mathbf{x}$ is morphic and generated by a k-uniform morphism + coding IFF $\mathbf{x}$ is k-automatic, *i.e.*, $\forall n \geq 0$, $\mathbf{x}_n$ is generated by an automaton reading $\mathrm{rep}_k(n)$.

$$f : A \mapsto AB, \quad B \mapsto BC, \quad C \mapsto CD, \quad D \mapsto DD$$

$$f^\omega(A) = ABBCBCCDBCCDCDDDBCCDCDDDCDDDDDDD \cdots$$



Morphic characterization

Corollary

A set $X \subseteq \mathbb{N}$ is k-recognizable IFF its characteristic sequence is k-automatic.

Link with combinatorics on words

$$f(0) = 01, \quad f(1) = 10$$

$$f^\omega(0) = 0110100110010110100101100110100 1\cdots$$

A. Thue (1912)

The Thue–Morse word is overlap free.

Morphic characterization

The *k-kernel* of $x = (x_n)_{n\geq 0}$ is defined by

$$N_k(x) = \{(x_{k^e n+d})_{n\geq 0} \mid e \geq 0,\ 0 \leq d < k^e\}$$

S. Eilenberg (1974)

A sequence $x = (x_n)_{n\geq 0}$ is k-automatic IFF $N_k(x)$ is finite.

Definition (J.-P. Allouche, J. Shallit 1992)

Let R be a (possibly infinite) commutative ring. Let $x = (x_n)_{n\geq 0} \in R^{\mathbb{N}}$. If the R-module generated by all sequences in $N_k(x)$ is finitely generated then x is said to be (R,k)-*regular*.

J.-P. Allouche, J. Shallit, The ring of k-regular sequences, TCS **98**, 163–197.

Morphic characterization

A sequence of C. Mallows

There is a unique monotone sequence $(a(n))_{n\geq 0}$ of non-negative integers such that $a(a(n)) = 2n$ for all $n \neq 1$,

n	0	1	2	3	4	5	6	7	8	9	10	11	12
$a(n)$	0	1	3	4	6	7	8	10	12	13	14	15	16

This sequence $(a(n))_{n\geq 0}$ is $(\mathbb{Z}, 2)$-regular.

J.-P. Allouche, J. Shallit, The ring of k-regular sequences II.

J. Bell (2005)

Let R be a commutative ring. Let k, ℓ be two multiplicatively independent integers. If a sequence $x \in R^{\mathbb{N}}$ is both (R, k)-regular and (R, ℓ)-regular, then it satisfies a linear recurrence over R.

Non-standard numeration systems

Definition

Consider an increasing sequence $(U_n)_{n\geq 0}$ of integers such that

- $U_0 = 1$
- $\sup U_{n+1}/U_n$ is bounded

Any integer n can be written as

$$n = \sum_{i=0}^{\ell} c_i\, U_i, \quad c_i \geq 0.$$

We choose the greedy representation: $\text{rep}_U(n) = c_\ell \cdots c_0$.

A. Fraenkel, Systems of numeration, Amer. Math. Monthly, 1985

M. Lothaire, Algebraic Combinatorics on Words, Cambridge Univ. Press 2002, Chap. by Ch. Frougny

Combinatorics, Automata and Number Theory, V. Berthé, M. Rigo (Eds.), Cambridge Univ. Press 2010, Chap. 2& 3

canonical alphabet $A_U = \{0, \ldots, \lceil \max U_{n+1}/U_n \rceil - 1\}$

$\text{rep}_U : \mathbb{N} \to A_U^*$

for any alphabet $B \subset \mathbb{Z}$, $\text{val}_U : B^* \to \mathbb{Z}$

$$\text{val}_U(d_\ell \cdots d_0) = \sum_{i=0}^{\ell} d_i \, U_i .$$

REMARK

We have *positional* numeration systems.

NON-STANDARD NUMERATION SYSTEMS

FIBONACCI (ZECKENDORF 1972)

$$\mathrm{rep}_F(11) = 10100 \text{ but } \mathrm{val}_F(10100) = \mathrm{val}_F(10011) = \mathrm{val}_F(1111)$$
$$U_0 = 1,\ U_1 = 2 \text{ and } U_{n+2} = U_{n+1} + U_n.$$

E. Zeckendorf, Bull. Soc. Roy. Sci. Liège **41**, 179–182.

$\ldots, 610, 377, 233, 144, 89, 55, 34, 21, 13, 8, 5, 3, 2, 1$

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

Can we extend Cobham's theorem on recognizability into two integer base systems to non-standard numeration systems ?

DEFINITION

A set $X \subset \mathbb{N}$ is *U-recognizable*, if the set of greedy expansions of the elements of X is accepted by some finite automaton, *i.e.*, $\mathrm{rep}_U(X)$ is a regular language.

If $X \subset \mathbb{N}$ is U-rec. and V-rec., U and V being "sufficiently independent", does it imply that X is ultimately periodic ?

We can also study the U-recognizable sets of integers for themselves !

It is natural to ask whether or not $\text{rep}_U(\mathbb{N})$ is regular...
check with a DFA if a word over A_U is a valid representation ?

OBSERVATION (G. HANSEL, J. SHALLIT, ...)

If $\mathbb{N}$ is U-recognizable, then $(U_n)_{n\geq 0}$ satisfies a linear recurrence relation with (constant) integer coefficients.

$\text{rep}_U(U_\ell) = 10^\ell$ for all $\ell \geq 0$. Amongst the words of length $\ell+1$ in $\text{rep}_U(\mathbb{N})$, the smallest one for the genealogical ordering is 10^ℓ.

Consequently, $U_{\ell+1} - U_\ell = \#(\text{rep}_U(\mathbb{N}) \cap A^{\ell+1})$.

Since $\text{rep}_U(\mathbb{N})$ is regular, it is accepted by a DFA and the number of words of length n in $\text{rep}_U(\mathbb{N})$ is equal to the number of paths of length n from the initial state to the final ones (then use Cayley-Hamilton theorem).

$\mathbb{N}$ being U-recognizable is somehow a minimal requirement,

PROPOSITION

Let $p, r \geq 0$. If $(U_n)_{n\geq 0}$ is a numeration system satisfying a linear recurrence relation with integer coefficients, then

$$\mathrm{val}_{A_U,U}^{-1}(p\,\mathbb{N}+r) = \left\{ c_\ell \cdots c_0 \in A_U^* \mid \sum_{k=0}^{\ell} c_k\, U_k \in p\,\mathbb{N}+r \right\}$$

is accepted by a DFA that can be effectively constructed.

COROLLARY

If $\mathbb{N}$ is U-recognizable, then any utimately periodic set is U-recognizable.

Satisfying a linear recurrence is not enough...

COUNTER-EXAMPLE (SHALLIT 1994)

Take $(U_n)_{n\geq 0}$ defined by $U_n = (n+1)^2$.
We have $U_0 = 1$, $U_1 = 4$, $U_2 = 9$ and
$U_{n+3} = 3\,U_{n+2} - 3\,U_{n+1} + U_n$. In that case,

$$\mathrm{rep}_U(\mathbb{N}) \cap 10^*10^* = \{10^a10^b \mid b^2 < 2a+4\}$$

showing with the pumping lemma that $\mathbb{N}$ is not U-recognizable.

N. Loraud, β-shift, systèmes de numération et automates, JTNB **7** (1995), 473-–498.

M. Hollander, Greedy numeration systems and regularity, *Theory Comput. Systems* **31** (1998), 111–133.

Pisot numeration systems

Definition

Consider a linear numeration system such that the characteristic polynomial of $(U_n)_{n\geq 0}$ is the minimal polynomial of a Pisot number (i.e., an algebraic integer $\alpha > 1$ whose Galois conjugates have modulus less than 1).

V. Bruyère, G. Hansel, Bertrand numeration systems and recognizability, TCS **181** (1997).

$$\lim_{n\to\infty} \frac{U_n}{c\alpha^n} = 1.$$

For these systems, all the “nice” properties hold true

- $\mathrm{rep}_U(\mathbb{N})$ is regular (for any reasonable initial conditions),
- for a precise choice of intial conditions, we have a Bertrand system (i.e., $v \in \mathrm{rep}_U(\mathbb{N}) \Leftrightarrow v0 \in \mathrm{rep}_U(\mathbb{N})$),
- normalization is computable by some finite automaton,
- the logical characterization can be extended,
- the morphic characterization too.

PISOT NUMERATION SYSTEMS

A link with the expansions of real numbers
$L(\beta)$ is the set of factors in some sequences $d_\beta(x)$, $x \in [0,1]$

$$\text{greedy } \beta\text{-expansion } d_\beta(x) = x_1 x_2 \cdots , \; x = \sum_{i=1}^{+\infty} x_i \beta^{-i}$$

A. BERTRAND (1989)

Let U be a numeration system. It is a Bertrand system if and only if there exists a real number $\beta > 1$ such that

$$\mathrm{rep}_U(\mathbb{N}) = L(\beta).$$

In this case, if U is linear, then β is a root of the characteristic polynomial of U.

D_β is the set of greedy β-expansions of numbers of $[0,1)$.

W. PARRY (1960)

Let $\beta > 1$ and let s be an infinite sequence of non-negative integers. The sequence s belongs to D_β IFF

$$\forall k \geq 0, \quad \sigma^k(s) <_{lex} d^*_\beta(1)$$

and s belongs to S_β, i.e., closure of D_β, IFF

$$\forall k \geq 0, \quad \sigma^k(s) \leq_{lex} d^*_\beta(1).$$

Parry, W. On the β-expansions of real numbers, Acta Math. Acad. Sci. Hung. **11**, (1960) 401-–416.

Pisot numeration systems

A. Bertrand (1986)

Let $\beta > 1$ be a real number. The language $L(\beta)$ is regular if and only if β is a *Parry number*.

Corollary

The DFA accepting $\mathrm{rep}_U(\mathbb{N})$ has a very special form.

The β-shift S_β is a dynamical system which is

- sofic IFF $d_\beta(1)$ is ultimately periodic,
- of finite type IFF $d_\beta(1)$ is finite.

Ito and Takahashi (1974), Bertrand-Mathis (1986), Blanchard (1989)

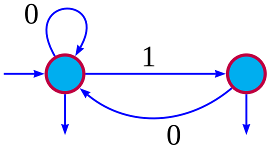
Pisot numeration systems

Integer base systems are special case of Pisot systems.

Fibonacci

$U_{n+2} = U_{n+1} + U_n$ with $U_0 = 1$ and $U_1 = 2$
$P(X) = X^2 - X - 1$ has roots $\frac{1+\sqrt{5}}{2}, \quad \frac{1-\sqrt{5}}{2}$

- $d_\beta(1) = 11$, $\mathrm{rep}_U(\mathbb{N})$ is regular (no block 11)



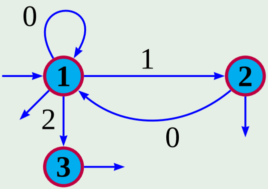
- we have a Bertrand system
  (i.e., $v \in \mathrm{rep}_U(\mathbb{N}) \Leftrightarrow v0 \in \mathrm{rep}_U(\mathbb{N})$),

Pisot numeration systems

(modified) Fibonacci

$U_{n+2} = U_{n+1} + U_n$ with the inital conditions $U_0 = 1$, $U_1 = 3$

$$(U_n)_{n \geq 0} = 1, 3, 4, 7, 11, 18, 29, 47, \ldots$$



PISOT NUMERATION SYSTEMS

Normalization $\nu_U : B^* \to A_U^*$ seems to be an essential tool, $B \subset \mathbb{Z}$, if $\mathrm{val}_U(w) \geq 0$, then $\nu_U(w) = \mathrm{rep}_U(\mathrm{val}_U(w))$.
Example for Fibonacci

$$\nu_F : 11011 \mapsto 100100, \quad 11100 \mapsto 100100, \ldots, \quad 22 \mapsto 1001$$

THEOREM (CH. FROUGNY 1992)

For any given alphabet B, for a Pisot system U, ν_U is realisable by a finite letter-to-letter transducer

COROLLARY

Addition is a U-recognizable ternary relation.

Ch. Frougny, Representations of numbers and finite automata, Math. Systems Theory **25**, (1992) 37—60.

Ch. Frougny, J. Sakarovitch, Number representation and finite automata, CANT Ch. 2, Cambridge Univ. Press (2010).

Logical characterization

BÜCHI–BRUYÈRE–HANSEL THEOREM

A set $X \subset \mathbb{N}$ is U-recognizable IFF it is definable by a first order formula in the extended Presburger arithmetic $\langle \mathbb{N}, +, V_U \rangle$.

$V_U(n)$ is the smallest U_i occurring in $\mathrm{rep}_U(n)$ with a non-zero coefficient.

What about a multidimensional context ?

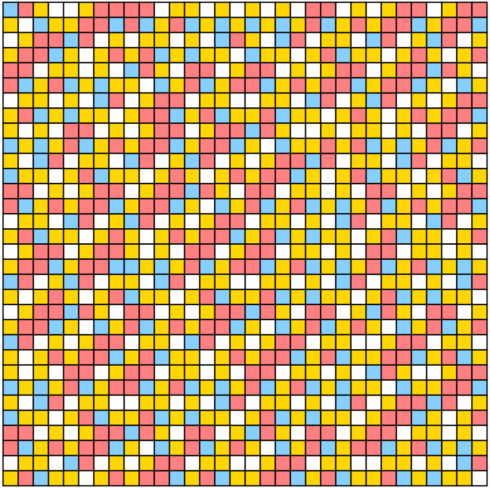
Everything works fine !

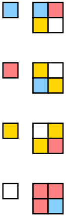
- automata reading n-tuples (with leading zeroes),
- morphisms with images being n-cubes of size k,
- logical characterization in $\langle \mathbb{N}, +, V_k \rangle$,
- extension to Cobham–Semenov' theorem

Cobham–Semenov' theorem

Let $k, \ell \geq 2$ be two multiplicatively independent integers.
A set $X^n \subseteq \mathbb{N}$ is k-rec. AND ℓ-rec. IFF X is definable in $\langle \mathbb{N}, + \rangle$.

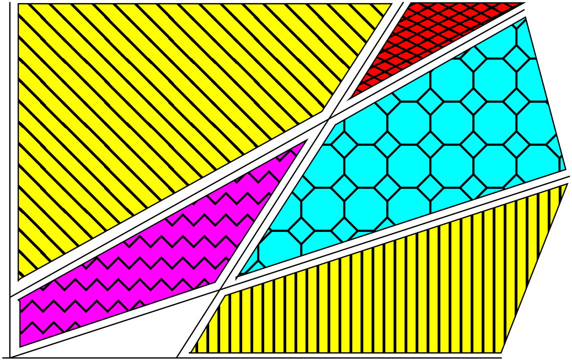
What about a multidimensional context ?





What about a multidimensional context ?

Looking at Cobham–Semenov' theorem, the *right* extension of ultimate periodicity over $\mathbb{N}$ could be definability in $\langle \mathbb{N}, + \rangle$



cf. local periodicity and Muchnik criterion

A. A. Muchnik, The definable criterion for definability in Presburger arithmetic and its applications, *Theoret. Comput. Sci* **290** (2003) 1433–1444.

SUMMARY

So far, we have seen

- integer base systems and recognizable sets
- Pisot numeration systems and recognizable sets

P. Lecomte (1997): "*everyone is taking an increasing sequence of integers then look at the regularity of* $\mathrm{rep}_U(\mathbb{N})$*.*
We could proceed the other way round and start directly by taking a regular language!"

REMARK

For positional numeration systems, rep_U is an increasing map:

$$x < y \Leftrightarrow \mathrm{rep}_U(x) < \mathrm{rep}_U(y).$$

DEFINITION ANS (P. LECOMTE, M.R. 2001)

An abstract numeration system $\mathcal{S} = (L, A, <)$ is a regular language L over a totally ordered finite alphabet $(A, <)$.

Numeration systems on a regular language, Theory Comput. Syst. **34** (2001), 27-–44.

- Enumerating the words in L using genealogical ordering provides a one-to-one correspondance between $\mathbb{N}$ and L :

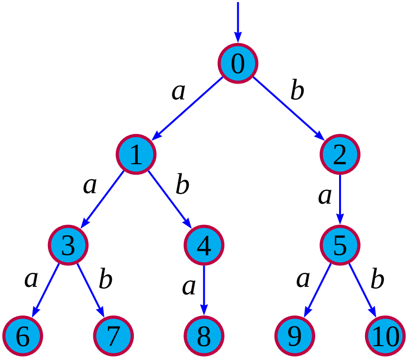
$$\mathrm{rep}_{\mathcal{S}} : \mathbb{N} \to L, \qquad \mathrm{val}_{\mathcal{S}} : L \to \mathbb{N}.$$

- This generalizes any positional system U for which $\mathbb{N}$ is U-recognizable.

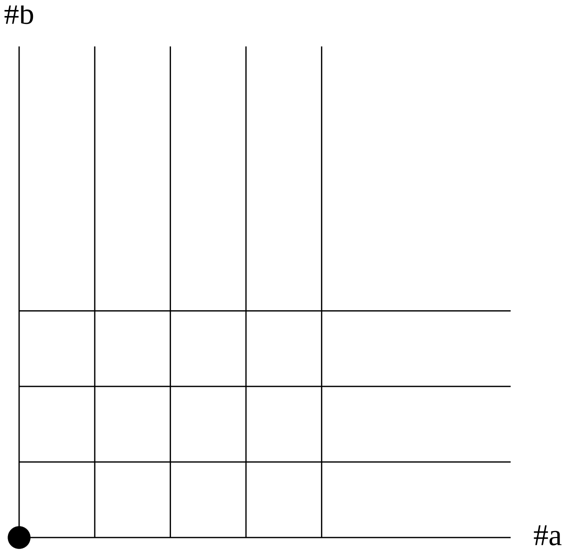
DEFINITION

A set $X \subseteq \mathbb{N}$ is *$\mathcal{S}$-recognizable*, if $\mathrm{rep}_{\mathcal{S}}(X)$ is regular.

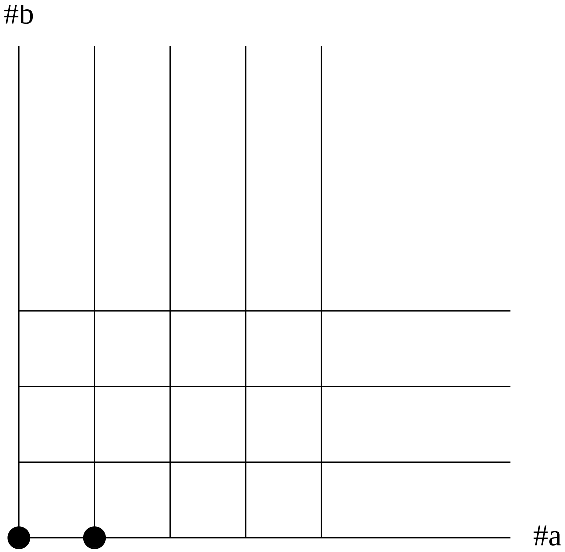
Example : consider a prefix-closed language $L = \{b, \varepsilon\}\{a, ab\}^*$



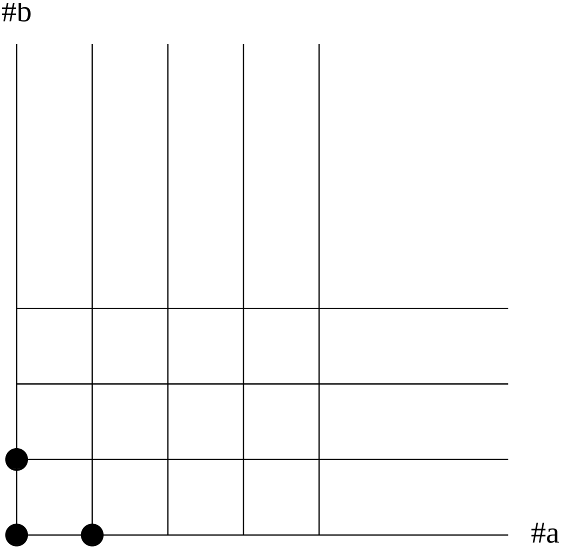
A non-positional ANS $L = a^* b^*$



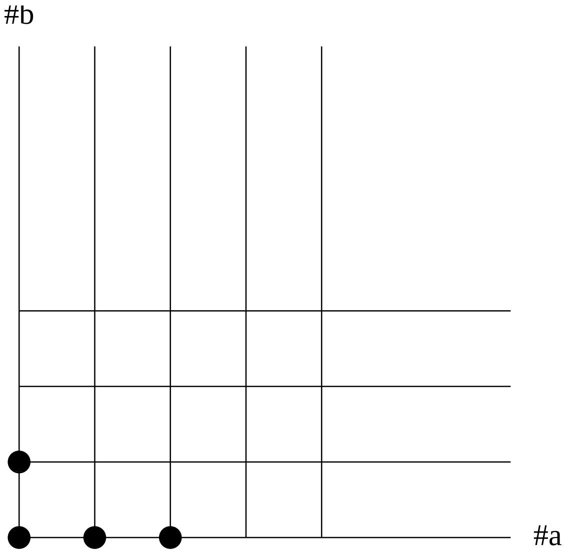
A non-positional ANS $L = a^*b^*$



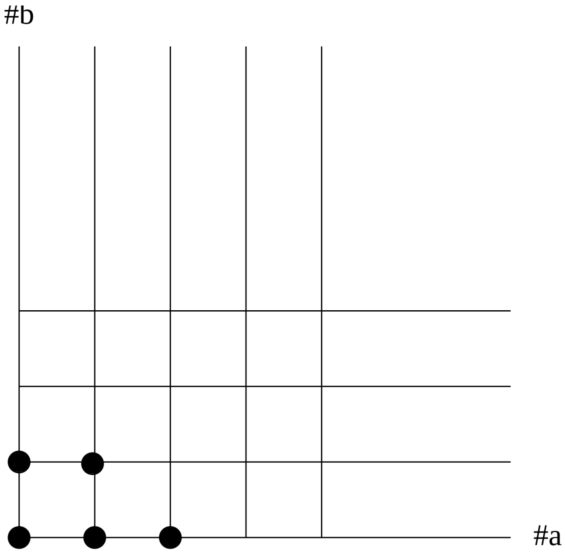
A non-positional ANS $L = a^*b^*$



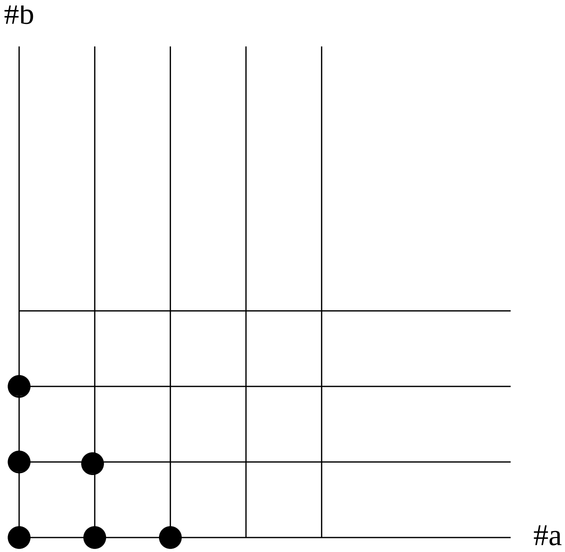
A non-positional ANS $L = a^*b^*$



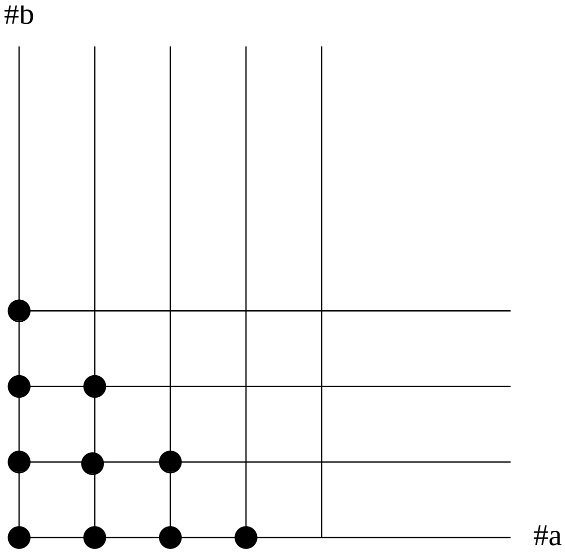
A non-positional ANS $L = a^*b^*$



A non-positional ANS $L = a^*b^*$

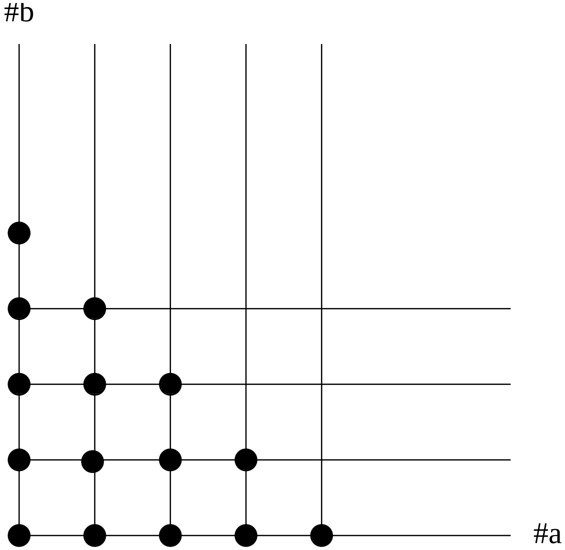


A non-positional ANS $L = a^*b^*$



Abstract numeration systems

A non-positional ANS $L = a^*b^*$



Abstract numeration systems

A non-positional ANS $L = a^*b^*$

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

ε	a	b	aa	ab	bb	aaa	$\cdots$
0	1	2	3	4	5	6	$\cdots$

$$U_0 = 1,\ U_1 = 2,\ p(a) = 1, p(b) = 2$$

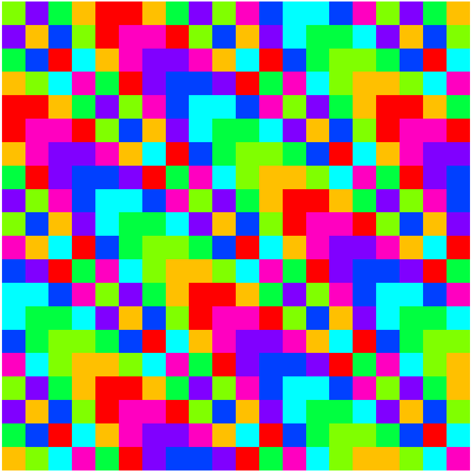
Generalization : $\mathrm{val}_\ell(a_1^{n_1} \cdots a_\ell^{n_\ell}) = \sum_{i=1}^{\ell} \binom{n_i + \cdots + n_\ell + \ell - i}{\ell - i + 1}.$

$$\forall n \in \mathbb{N}, \exists z_1, \ldots, z_\ell : n = \binom{z_\ell}{\ell} + \binom{z_{\ell-1}}{\ell-1} + \cdots + \binom{z_1}{1}$$

with the condition $z_\ell > z_{\ell-1} > \cdots > z_1 \geq 0$

[Katona, Gel'fand, Lehmer, Fraenkel, Lew, Morales, . . .]

ABSTRACT NUMERATION SYSTEMS



val$(a^p b^q)$ modulo 8

ABSTRACT NUMERATION SYSTEMS

THEOREM [P. LECOMTE, M.R.]

Let $\mathcal{S}$ be an ANS. Any ultimately periodic set of intergers is $\mathcal{S}$-recognizable.

EQUIVALENT FORMULATION [D. KRIEGER *et al.* TCS'09]

Let L be a regular language. Any “periodic decimation” of L is a regular language.

In general, this result does not hold for context-free languages.

Another example, an unambiguous positional system
$L = \{1, 2\}^*$

0	ε
1	1
2	2
3	11
4	12
5	21
6	22
7	111
8	112
9	121
10	122
$\vdots$	$\vdots$

Lemma

$$\mathrm{val}_{\mathcal{S}}(w) = \sum_{q\in Q} \sum_{i=1}^{|w|} b_{q,i}(w)\, \mathbf{u}_q(|w|-i)$$

with

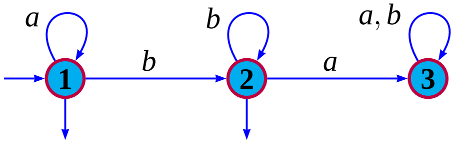
$$b_{q,i}(w) := \#\{a < w_i \mid q_0 \cdot w_1 \cdots w_{i-1} a = q\} + \mathbf{1}_{q_0,q}$$

and

$$\mathbf{u}_q(n) = \#\{v \in A^n \mid q \cdot v \in F\}.$$

$$\mathbf{v}_q(n) = \#\{v \in A^{\leq n} \mid q \cdot v \in F\} = \sum_{i=0}^{n} \mathbf{u}_q(i).$$

Let $\mathcal{S} = (L, A, <)$ an ANS.



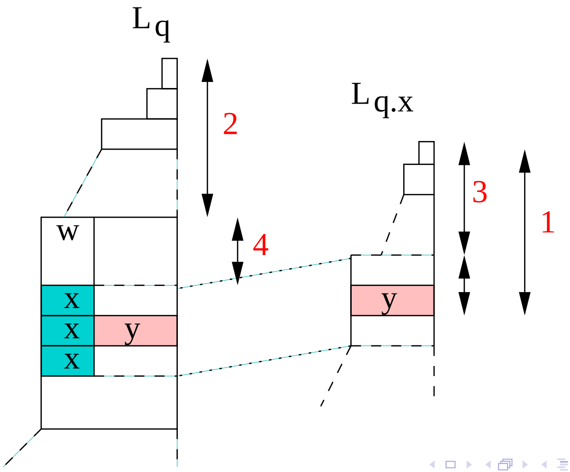
	L_1	L_2	L_3
0	ε	ε	$--$
1	a	b	$--$
2	b	bb	$--$
3	aa	bbb	$--$
4	ab	$bbbb$	$--$
5	bb	$bbbbb$	$--$

If xy belongs to L_q, $y \neq \varepsilon$, then

$$\mathrm{val}_q(xy) = \mathrm{val}_{q\cdot x}(y) + \mathbf{v}_q(|xy|-1) - \mathbf{v}_{q\cdot x}(|y|-1) + \sum_{\substack{w<x \\ |w|=|x|}} \mathbf{u}_{q\cdot w}(|y|).$$

If xy belongs to L_q, $y \neq \varepsilon$, then

$$\mathrm{val}_q(xy) = \mathrm{val}_{q\cdot x}(y) + \mathbf{v}_q(|xy|-1) - \mathbf{v}_{q\cdot x}(|y|-1) + \sum_{\substack{w<x \\ |w|=|x|}} \mathbf{u}_{q\cdot w}(|y|).$$



Many natural questions...

- What about $\mathcal{S}$-recognizable sets ?
  - Are ultimately periodic sets $\mathcal{S}$-recognizable for any $\mathcal{S}$?
  - For a given $X \subseteq \mathbb{N}$, can we find $\mathcal{S}$ s.t. X is $\mathcal{S}$-recognizable ?
  - For a given $\mathcal{S}$, what are the $\mathcal{S}$-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...

Recall that 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	$\cdots$

Folklore

II L is a regular language, then the set $\min(L)$ of minimal words of each length is again regular.

Theorem (M.R. 2002)

Let P_i be polynomials belonging to $\mathbb{Q}[x]$ such that $P_i(\mathbb{N}) \subset \mathbb{N}$ and α_i be non-negative integers, $i = 1, \ldots, k$, $k \geq 1$. Set

$$f(n) = \sum_{i=1}^{k} P_i(n)\, \alpha_i^n .$$

There exists an ANS $\mathcal{S}$ such that $f(\mathbb{N})$ is $\mathcal{S}$-recognizable.

Definition of growth rate

Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$

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

See also, E. Charlier, N. Rampersad, The growth function of S-recognizable sets, *Theoret. Comput. Sci.* **412** (2011), 5400–5408.

THEOREM (P. LECOMTE, M.R. 2001)

Let $\mathcal{S} = (a^*b^*, a < b)$. Multiplication by $\lambda \in \mathbb{N}_{>0}$ preserves $\mathcal{S}$-recognizability, i.e., for all $\mathcal{S}$-recognizable set $X \subseteq \mathbb{N}$, λX is $\mathcal{S}$-recognizable, IFF 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 ?

M. R., Numeration systems on a regular language : Arithmetic operations, Recognizability and Formal power series,

Theoret. Comp. Sci. 269 (2001), 469–498.

The successor function can be computed by means of finite automata: It is realized by a (left or right) letter-to-letter finite transducer

Theorem (P.-Y. Angrand, J. Sakarovitch 2010)

The radix enumeration of a rational language is a finite union of co-sequential functions.

A cascade of (at most 2) sequential (right) transducers, that is, a first transducer reads the input and produces an output which is then taken as the input of second transducer which depends on the final state in the computation of the first one.

P.-Y. Angrand, J. Sakarovitch, Radix enumeration of rational languages, *RAIRO - Theoret. Informatics and Appl.* **44** (2010) 19–36.

Morphic words

Definition

Let $f : \Sigma \to \Sigma^*$ and $g : \Sigma \to \Gamma^*$ be to morphisms such that $f(a) \in a\Sigma^+$. We define a *morphic word* (a.k.a. substitutive) over Γ,

$$w = g(\lim_{n\to\infty} f^n(a)) = g(f^\omega(a)).$$

We can assume f non-erasing and g is a coding.

Example (Characteristic sequence of squares)

$f : a \mapsto abcd,\ b \mapsto b,\ c \mapsto cdd,\ d \mapsto d,\ g : a, b \mapsto 1,\ c, d \mapsto 0.$

$$f^\omega(a) = abcdbcdddbcdddddbcdddddddbc\cdots$$

$$g(f^\omega(a)) = 110010000100000010000000010\cdots$$

Morphic words

What is the link between morphic words and ANS ?

Recall this result (Cobham 1972)

An infinite word $\mathbf{x}$ is morphic and generated by a k-uniform morphism + coding IFF $\mathbf{x}$ is k-automatic, *i.e.*, $\forall n \geq 0$, $\mathbf{x}_n$ is generated by an automaton reading $\mathrm{rep}_k(n)$.

We can introduce S-automatic sequences...

Morphic words

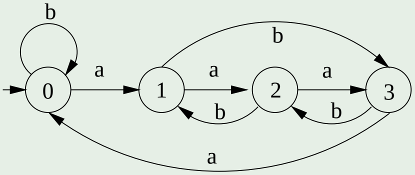
Definition

Let $S = (L, \Sigma, <)$ be an ANS and $\mathcal{M} = (Q, q_0, \Sigma, \delta, \Gamma, \tau)$ be a DFAO. Consider the *S-automatic sequence*

$$x_n = \tau(\delta(q_0, (\text{rep}_S(n))))$$

Example

$S = (a^*b^*, \{a, b\}, a < b)$



$$0102303120023101012302303120312023100231012 3 \cdots$$

MORPHIC WORDS

Extension of Cobham's result

THEOREM (A. MAES, M.R. 2002)

An infinite word x is morphic IFF there exists some ANS $\mathcal{S}$ such that x is a $\mathcal{S}$-automatic.

The set of $\mathcal{S}$-automatic sequences (for all $\mathcal{S}$) coincides with the set of morphic words.

REMARK

A set $X \subseteq \mathbb{N}$ is $\mathcal{S}$-recognizable IFF its characteristic sequence is $\mathcal{S}$-automatic.

MORPHIC WORDS

k-automatic sequence $\Updownarrow$ k-uniform morphism + coding [A. Cobham'72]	$\mathcal{S}$-automatic sequence $\Updownarrow$ non-erasing morphism + coding [A. Maes, M.R.'02]
multidimensional setup $x : \mathbb{N}^d \to A$ k-automatic sequence $\Updownarrow$ morphism $g : A \to (A^q)^d$ + coding [O. Salon'87]	 $\mathcal{S}$-automatic sequence $\Updownarrow$ "shape-symmetric" morphism + coding [É. Charlier, T. Kärki, M.R.'09]

$$\varphi : a \mapsto \begin{array}{|c|c|}\hline a & b \\ \hline c & d \\ \hline\end{array} \quad b \mapsto \begin{array}{|c|}\hline i \\ \hline e \\ \hline\end{array} \quad c \mapsto \begin{array}{|c|c|}\hline i & j \\ \hline\end{array} \quad d \mapsto \begin{array}{|c|}\hline i \\ \hline\end{array} \quad e \mapsto \begin{array}{|c|c|}\hline f & b \\ \hline\end{array}$$

$$f \mapsto \begin{array}{|c|c|}\hline g & b \\ \hline h & d \\ \hline\end{array} \quad g \mapsto \begin{array}{|c|c|}\hline f & b \\ \hline h & d \\ \hline\end{array} \quad h \mapsto \begin{array}{|c|c|}\hline i & m \\ \hline\end{array} \quad i \mapsto \begin{array}{|c|c|}\hline i & m \\ \hline h & d \\ \hline\end{array}$$

$$j \mapsto \begin{array}{|c|}\hline k \\ \hline c \\ \hline\end{array} \quad k \mapsto \begin{array}{|c|c|}\hline l & m \\ \hline c & d \\ \hline\end{array} \quad l \mapsto \begin{array}{|c|c|}\hline k & m \\ \hline c & d \\ \hline\end{array} \quad m \mapsto \begin{array}{|c|}\hline i \\ \hline h \\ \hline\end{array}$$

coding

$$\mu : e, g, j, l \mapsto 1, \quad a, b, c, d, f, h, i, k, m \mapsto 0$$

Morphic words

$a \mapsto$

a	b
c	d

$\mapsto$

a	b	i
c	d	e
i	j	i

$\mapsto$

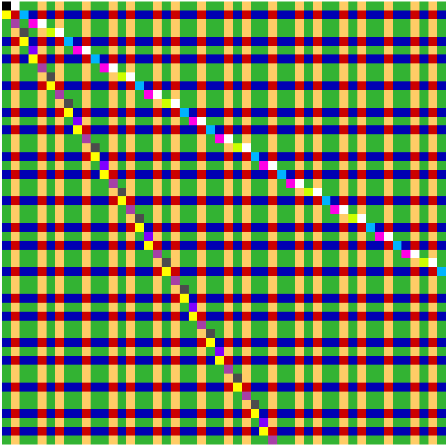
a	b	i	i	m
c	d	e	h	d
i	j	i	f	b
i	m	k	i	m
h	d	c	h	d

$\mapsto$

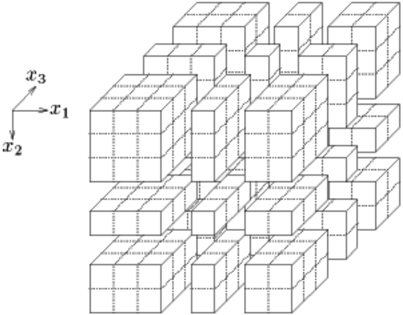
a	b	i	i	m	i	m	i
c	d	e	h	d	h	d	h
i	j	i	f	b	i	m	i
i	m	k	i	m	g	b	i
h	d	c	h	d	h	d	e
i	m	i	l	m	i	m	i
h	d	h	c	d	h	d	h
i	m	i	i	j	i	m	i

$\mapsto \cdots$

Morphic words



From E. Duchêne, A. S. Fraenkel, R. Nowakowski, M.R., Extensions and restrictions of wythoff's game preserving wythoff's sequence as set of $\mathcal{P}$-positions, **JCTA** (2010).



From A. Maes Ph.D. thesis, *Prédicats morphiques et applications à la décidabilité de théories arithmétiques*

MORPHIC WORDS

THEOREM (F. DURAND 1998)

Let (f, g, a) (resp. (f', g', a')) be a primitive substitution with a dominating eigenvalue $\alpha > 1$ (resp. $\beta > 1$). Let α and β be multiplicatively independent. If $x = g(f^\omega(a)) = g'(f'^\omega(a'))$, then x is ultimately periodic.

- F. Durand, A generalization of Cobham's theorem, *Theory of Computing Systems* **31** (1998), 169–185.
- F. Durand, A Theorem of Cobham for non primitive substitutions, *Acta Arithmetica* **104** (2002), 225–241.
- F. Durand, M. R., Syndeticity and independent substitutions, *Adv. in Applied Math.* **42** (2009), 1–22.
- F. Durand, Cobham's theorem for substitutions, *J. Eur. Math. Soc.* **13** (2011), 1797-–1812.

Morphic words

An “application”

Example

The Fibonacci word $0100101001\cdots$ generated by $f: 0 \mapsto 01, 1 \mapsto 0$ is not k-automatic.

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

Indeed, this (Sturmian) word is not ultimately periodic and for all integers k, k and $(1+\sqrt{5})/2$ are multiplicatively independent.

Of course, one can also use this result about frequency

Proposition

In any k-automatic sequence, if the frequency of a symbol exists, then it is rational.

An "application"

Example

If $X \subseteq \mathbb{N}$ is both $\mathcal{S}$- and $\mathcal{T}$-recognizable where $\mathcal{S}$ (resp. $\mathcal{T}$) is built over an exponential (resp. a polynomial) language then X is ultimately periodic.

A few words on ω-HD0L ultimate periodicity

Question : *given f, g two morphisms, decide whether or not $g(f^\omega(a))$ is ultimately periodic.*

- Trivial for k-automatic sequences, thanks to first order logic.
- J. Honkala, A decision method for the recognizability of sets defined by number systems, *Theoret. Inform. Appl.* **20** (1986), 395–403.
- T. Harju, M. Linna, On the periodicity of morphisms on free monoids, *RAIRO Inform. Théor. Appl.* **20** (1986), 47–54.
- J.-J. Pansiot, Decidability of periodicity for infinite words, *RAIRO Inform. Théor. Appl.* **20** (1986), 43–46.
- J.-P. Allouche, N. Rampersad, J. Shallit, Periodicity, repetitions, and orbits of an automatic sequence, *Theoret. Comput. Sci* **5162**, pp. 241–252, Springer-Verlag (2008).

A FEW WORDS ON ω-HD0L ULTIMATE PERIODICITY

- J. Leroux, A Polynomial Time Presburger Criterion and Synthesis for Number Decision Diagrams, LICS 2005, IEEE Comp. Soc. (2005), 147–156.
- E. Charlier, M. Rigo, A decision problem for ultimately periodic sets in non-standard numeration systems, MFCS 2008, Lect. Notes in Comput. Sci. **5162** (2008), 241–252.

Equivalent question : *Let given a $\mathcal{S}$-recognizable set of integers, decide whether or not it is ultimately periodic.*

- F. Durand, Decidability of the HD0L ultimate periodicity problem, `arXiv:1111.3268v1`
- I. Mitrofanov, A proof for the decidability of HD0L ultimate periodicity, `arXiv:1110.4780`

Some open problems

- Give a proof based on ANS for the ω-HD0L ultimate periodicity (based on automata, we could have a better view/complexity).
- If $g(f^{\omega}(a))$ is infinite, one can always assume that f is non-erasing and g is a coding [Cobham'68, Allouche–Shallit'03, Honkala'09], again give a proof based only on automata.
- Given a ANS, decide whether or not this system is a positional numeration system.

How to represent real numbers

Example (base 10)

$$\pi - 3 = .14159265358979323846264338328\cdots$$

$$\frac{1}{10}, \quad \frac{14}{100}, \quad \frac{141}{1000}, \quad \ldots, \quad \frac{val(w_n)}{10^n}, \quad \ldots$$

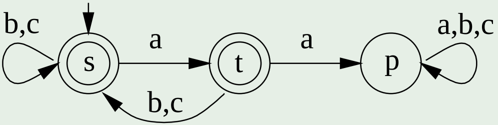
$$\frac{\mathrm{val}(w)}{\#\{\text{words of length } \leq |w|\}}$$

This deserves notation

$$\mathbf{v}_{q_0}(n) = \#(L \cap \Sigma^{\leq n}) = \sum_{i=0}^{n} \mathbf{u}_{q_0}(i).$$

How to represent real numbers

Example (Avoid *aa* on three letters)



w	$\mathrm{val}(w)$	$\mathbf{v}_{q_0}(\lvert w\rvert)$	$\mathrm{val}(w)/\mathbf{v}_{q_0}(\lvert w\rvert)$
bc	8	12	0.66666666666667
bac	19	34	0.55882352941176
babc	52	94	0.55319148936170
babac	139	258	0.53875968992248
bababc	380	706	0.53824362606232

$$\lim_{n\to\infty} \frac{\mathrm{val}((ba)^n c)}{\mathbf{v}_{q_0}(2n+1)} = \frac{1}{1+\sqrt{3}} + \frac{3}{9+5\sqrt{3}} \simeq 0.535898.$$

How to represent real numbers

$$\mathrm{val}_{\mathcal{S}}(w) = \sum_{q \in Q} \sum_{i=1}^{|w|} b_{q,i}(w)\, \mathbf{u}_q(|w| - i)$$

with

$$b_{q,i}(w) := \#\{a < w_i \mid q_0 \cdot w_1 \cdots w_{i-1} a = q\} + \mathbf{1}_{q_0,q}$$

Hypotheses: For all state q of $\mathcal{M}_L$, either

(i) $\exists N_q \in \mathbb{N} : \forall n > N_q,\ \mathbf{u}_q(n) = 0$, or

(ii) $\exists \beta_q \geq 1,\ P_q(x) \in \mathbb{R}[x]\ ,\ b_q > 0 : \lim_{n\to\infty} \frac{\mathbf{u}_q(n)}{P_q(n)\beta_q^n} = b_q.$

From automata theory, we have

$$\beta_{q_0} \geq \beta_q \text{ and } \beta_q = \beta_{q_0} \Rightarrow deg(P_q) \leq deg(P_{q_0})$$

How to represent real numbers

Let $\boxed{\beta = \beta_{q_0}}$ and for any state q, define

$$\lim_{n\to\infty} \frac{\mathbf{u}_q(n)}{P_{q_0}(n)\beta^n} = a_q \in \mathbb{Q}(\beta), \quad a_{q_0} > 0 \text{ and } a_q \text{ could be zero.}$$

If $(w_n)_{n\in\mathbb{N}}$ is converging to $W = W_1 W_2 \cdots$ then

$$\lim_{n\to\infty} \frac{\mathrm{val}(w_n)}{\mathbf{v}_{q_0}(|w_n|)} = \frac{\beta - 1}{\beta^2} \sum_{j=0}^{\infty} \sum_{q\in Q} \frac{a_q}{a_{q_0}} b_{q,j+1}(W)\, \beta^{-j} = x.$$

We say that W is a representation of x

A real number can have

- a unique expansion
- finitely many expansions
- countably many expansions

$x \in I_w$, if there exists an infinite word having w as prefix and representing x.

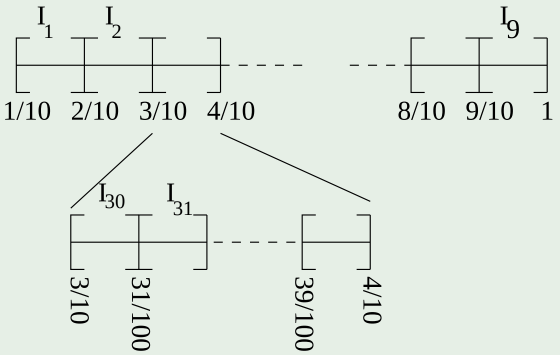
$\mathcal{W}_\ell$:=set of words of length ℓ that are prefixes of infinitely many words in L. Let $w \in \mathcal{W}_\ell$,

$$I_w = \left[\frac{1}{\beta} + \frac{\beta - 1}{\beta^{\ell+1}} \sum_{v<w,\ v\in\mathcal{W}_\ell} \frac{a_{q_0.v}}{a_{q_0}}, \frac{1}{\beta} + \frac{\beta - 1}{\beta^{\ell+1}} \sum_{v\leq w,\ v\in\mathcal{W}_\ell} \frac{a_{q_0.v}}{a_{q_0}} \right].$$

How to represent real numbers

In base 10

$L = \{\varepsilon\} \cup \{1, \ldots, 9\}\{0, \ldots, 9\}^*$, we represent $[1/10, 1]$.



NOTATION (FOR ALL STATES q)

Ratio of words starting with a, b or c...

$$\begin{aligned}
[0,1] \quad = & \quad [0, \lim \frac{u_{q.a}(n-1)}{u_q(n)}[& (A_{q,a}) \\
\cup & \quad [\lim \frac{u_{q.a}(n-1)}{u_q(n)}, \lim \frac{u_{q.a}(n-1)+u_{q.b}(n-1)}{u_q(n)}[& (A_{q,b}) \\
\cup & \quad [\lim \frac{u_{q.a}(n-1)+u_{q.b}(n-1)}{u_q(n)}, 1[& (A_{q,c})
\end{aligned}$$

If $I = [a,b] \ni x$, then $f_I : [a,b] \to [0,1] : x \mapsto (x-a)/(b-a)$

How to represent real numbers

Algorithm

Let $x \in [1/\beta, 1]$

Initialization

$q \leftarrow q_0$

$w \leftarrow \varepsilon$

$I \leftarrow [1/\beta, 1]$

$x \leftarrow f_I(x)$

repeat

Find the letter $\sigma \in \Sigma$ s.t. $x \in A_{q,\sigma}$.

$q \leftarrow q \cdot \sigma$

$w \leftarrow$ concat(w, σ)

$I \leftarrow A_{q,\sigma}$

$x \leftarrow f_I(x)$

until some halt condition.

NOTATION

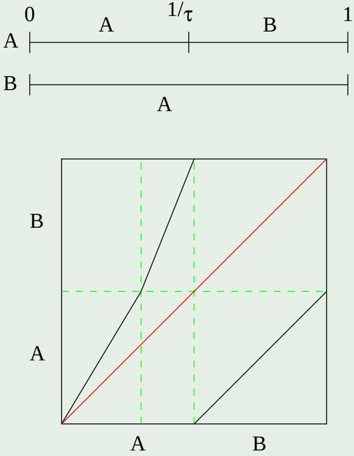
$h : Q \times [0,1] \to Q \times [0,1] : (q,x) \mapsto (q',x')$
There exists a unique letter σ s.t. $x \in A_{q,\sigma}$ hence

$$\begin{cases} q' = q \cdot \sigma \\ x' = f_{A_{q,\sigma}}(x) \end{cases}$$

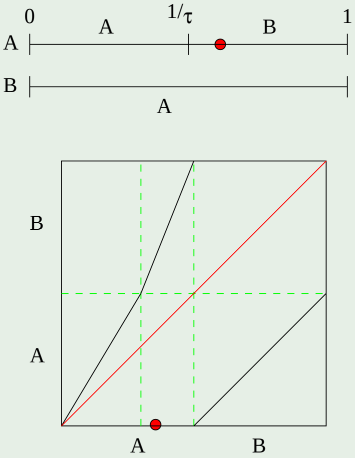
GENERAL QUESTION

are there $i < j$ such that $h^i(q_0,x) = h^j(q_0,x)$?

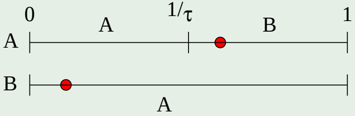
FIBONACCI

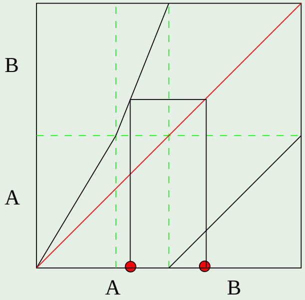


FIBONACCI

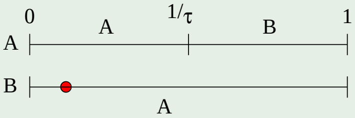


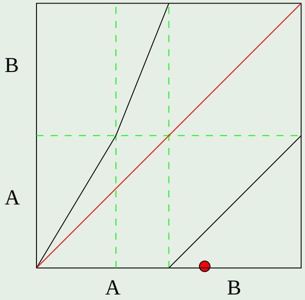
FIBONACCI



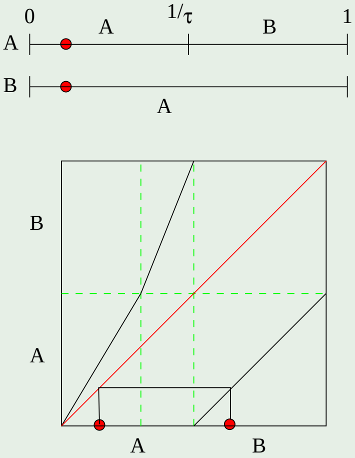


Fibonacci

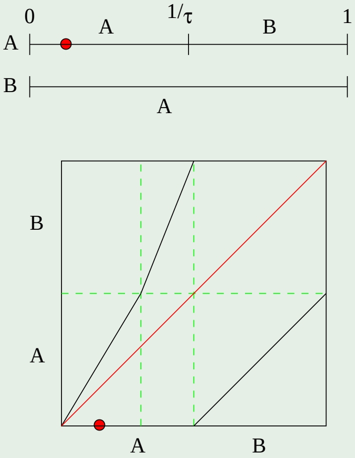




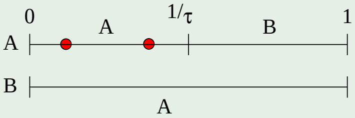
FIBONACCI

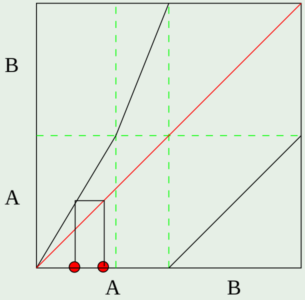


FIBONACCI

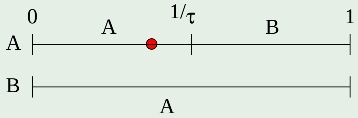


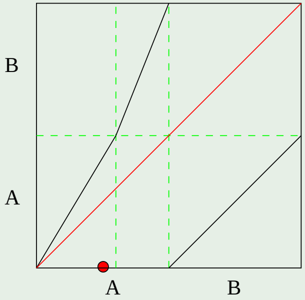
FIBONACCI



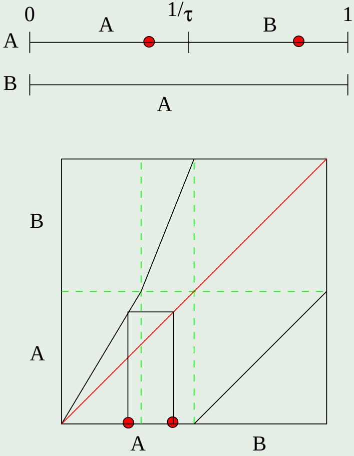


FIBONACCI

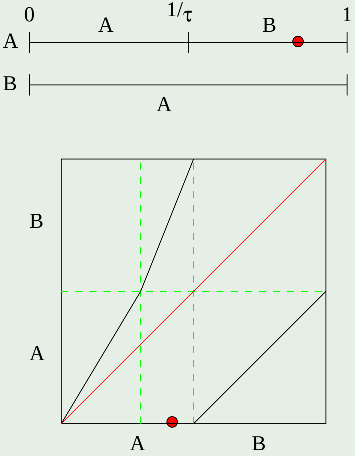




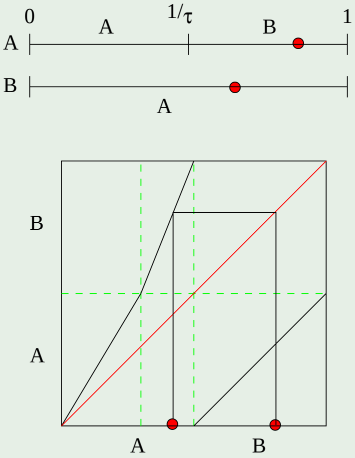
FIBONACCI



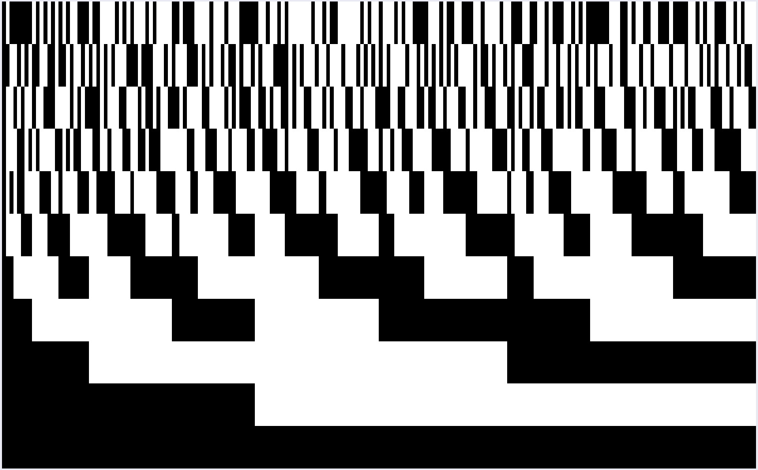
FIBONACCI



Fibonacci



In general, quite difficult



- S. Akiyama, Ch. Frougny, J. Sakarovitch, Powers of rationals modulo 1 and rational base number systems, *Israel J. Math.* **168** (2008) 53–91.
- E. Charlier, M. Le Gonidec, M. Rigo, Representing reals numbers in a generalized numeration system, *J. Comput. Syst. Sci.* **77** (2011), 743–759.
- P. Lecomte, M. Rigo, Real numbers having ultimately periodic representations in abstract numeration systems, *Inform. and Comput.* **192** (2004), 57–83.
- M. Rigo, W. Steiner, Abstract beta-expansion and ultimately periodic representations, *J. Theor. Nombres Bordeaux* **17** (2005), 288–299.
- P. Lecomte, M. Rigo, On the representation of real numbers using regular languages, *Theory Comput. Syst.* **35** (2002), 13–38.
- Fibred systems, see for instance, M. Madritsch