(Abstract) Numeration Systems

Michel Rigo

Department of Mathematics, University of Liège
http://www.discmath.ulg.ac.be/

Outline of the talk

What is a numeration system ?

Connection with formal languages theory

Base dependence

Characterizations of k-recognizable sets

Some puzzling properties of P.-T.-M. sequence

Let's come back to U-recognizability

Motivation for a generalization

Abstract numeration systems

First results

Representing real numbers

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.

(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 $\mathrm{rep}_U(n) = c_\ell \cdots c_0$.

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.

(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 $\mathrm{rep}_U(n) = c_\ell \cdots c_0$.

SOME CONDITIONS ON $U = (U_i)_{i\geq 0}$

- $U_{i+1} < U_i$, *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

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

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

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

- $U_{i+1} < U_i$, *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

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

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

SOME CONDITIONS ON $U = (U_i)_{i \geq 0}$

- $U_{i+1} < U_i$, *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

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

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, $A_U = \{0, 1\}$.

What use for numeration systems ?

- Number theory, Transcendence
- Combinatorics on words
- Automatic sequences, Digital sequences
- Sloane and Plouff's encyclopedia of integer sequences
- Theoretical computer science, Automata theory
- Formal languages theory
- Logic, Complexity theory
- Symbolic dynamics
- Ergodic theory
- Graph theory, Game theory
- Physics, Quasi-crystals, Fractal geometry

FACT

Any $n \geq 0$ is represented by a *word* $\mathrm{rep}_U(n)$ over A_U.
A set $X \subseteq \mathbb{N}$ corresponds to a set of words, i.e., a *language*.

The Chomsky's hierarchy :

- Recursively enumerable languages (Turing Machine)
- Context-sensitive languages (linear bounded T.M.)
- Context-free languages (pushdown automaton)
- Regular (or rational) languages (finite automaton)

FACT

Any $n \geq 0$ is represented by a *word* $\mathrm{rep}_U(n)$ over A_U.
A set $X \subseteq \mathbb{N}$ corresponds to a set of words, i.e., a *language*.

The Chomsky's hierarchy :

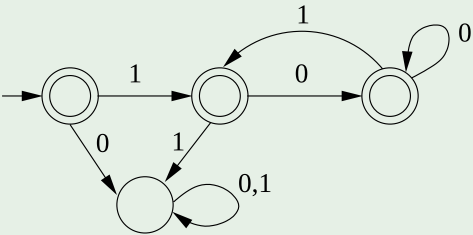
- Recursively enumerable languages (Turing Machine)
- Context-sensitive languages (linear bounded T.M.)
- Context-free languages (pushdown automaton)
- Regular (or rational) languages (finite automaton)

DETERMINISTIC FINITE AUTOMATON

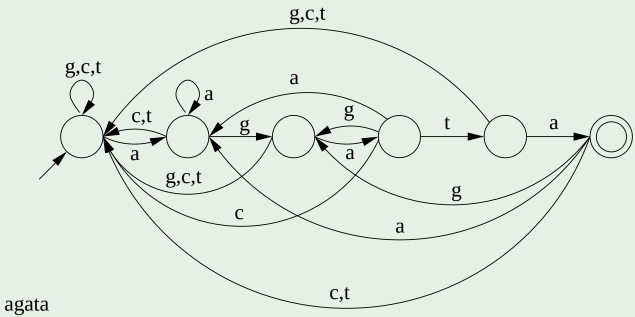
$\mathcal{A} = (Q, q_0, \Sigma, \delta, F)$

- Q finite set of states, $q_0 \in Q$ initial state
- $\delta : Q \times \Sigma \to Q$ transition function
- $F \subseteq Q$ set of final (or accepting) states

EXAMPLE (FIBONACCI)



EXAMPLE (USE IN BIO-INFORMATICS, DNA: `a,c,g,t`)



EXAMPLE (USE IN COMPUTER SCIENCE)

Complete algorithmic solution for model checking, program verification, . . .

What are we looking for ?

The “simplest” sets X of integers are the ones such that

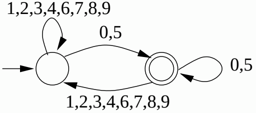
$$\mathrm{rep}_U(X) \text{ is regular,}$$

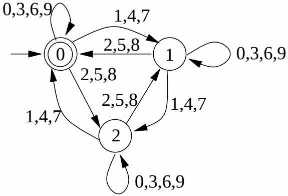
i.e., representations are accepted by some finite automaton. Such sets are said to be U-recognizable.

Divisibility criterion in base k

Let $k \geq 2$. The set $X = \{n \mid n \equiv r \pmod{s}\}$ is k-recognizable.

Example





What are we looking for ?

The “simplest” sets X of integers are the ones such that

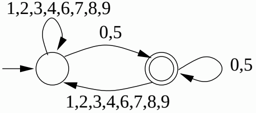
$$\mathrm{rep}_U(X) \text{ is regular,}$$

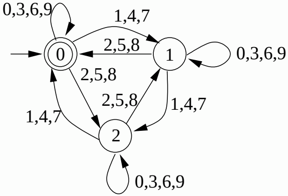
i.e., representations are accepted by some finite automaton. Such sets are said to be U-recognizable.

Divisibility criterion in base k

Let $k \geq 2$. The set $X = \{n \mid n \equiv r \pmod{s}\}$ is k-recognizable.

Example





What are we looking for ?

The "simplest" sets X of integers are the ones such that

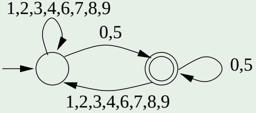
$$\mathrm{rep}_U(X) \text{ is regular,}$$

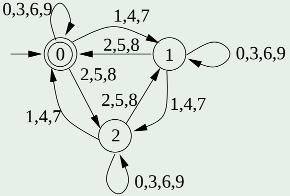
i.e., representations are accepted by some finite automaton. Such sets are said to be U-recognizable.

Divisibility criterion in base k

Let $k \geq 2$. The set $X = \{n \mid n \equiv r \pmod{s}\}$ is k-recognizable.

Example





A "NATURAL" QUESTION

If $X \subseteq \mathbb{N}$ is p-recognizable, is it also q-recognizable ?

p, q are *multiplicatively independent* if $p^k = q^\ell \Rightarrow k = \ell = 0$, i.e., $\log p / \log q$ is irrational.

"being multiplicatively dependent" is an equivalence relation,

2	3	5	6	7	10	11	$\cdots$
4	9	25	36	49	100	121	$\cdots$
8	27	125	216	343	1000	1331	$\cdots$

PROPOSITION

If $p, q \geq 2$ are multiplicatively dependent, then $X \subset \mathbb{N}$ is p-recognizable iff it is q-recognizable.

A "natural" question

If $X \subseteq \mathbb{N}$ is p-recognizable, is it also q-recognizable ?

p, q are *multiplicatively independent* if $p^k = q^\ell \Rightarrow k = \ell = 0$, i.e., $\log p / \log q$ is irrational.

"being multiplicatively dependent" is an equivalence relation,

2	3	5	6	7	10	11	$\cdots$
4	9	25	36	49	100	121	$\cdots$
8	27	125	216	343	1000	1331	$\cdots$

Proposition

If $p, q \geq 2$ are multiplicatively dependent, then $X \subset \mathbb{N}$ is p-recognizable iff it is q-recognizable.

A "NATURAL" QUESTION

If $X \subseteq \mathbb{N}$ is p-recognizable, is it also q-recognizable ?

p, q are *multiplicatively independent* if $p^k = q^\ell \Rightarrow k = \ell = 0$, i.e., $\log p / \log q$ is irrational.

"being multiplicatively dependent" is an equivalence relation,

2	3	5	6	7	10	11	$\cdots$
4	9	25	36	49	100	121	$\cdots$
8	27	125	216	343	1000	1331	$\cdots$

PROPOSITION

If $p, q \geq 2$ are multiplicatively dependent, then $X \subset \mathbb{N}$ is p-recognizable iff it is q-recognizable.

A "NATURAL" QUESTION

If $X \subseteq \mathbb{N}$ is p-recognizable, is it also q-recognizable ?

p, q are *multiplicatively independent* if $p^k = q^\ell \Rightarrow k = \ell = 0$, i.e., $\log p / \log q$ is irrational.

"being multiplicatively dependent" is an equivalence relation,

2	3	5	6	7	10	11	$\cdots$
4	9	25	36	49	100	121	$\cdots$
8	27	125	216	343	1000	1331	$\cdots$

PROPOSITION

If $p, q \geq 2$ are multiplicatively dependent, then $X \subset \mathbb{N}$ is p-recognizable iff it is q-recognizable.

Theorem (Cobham '69)

Let $p, q \geq 2$ be two multiplicatively independent integers.
If $X \subseteq \mathbb{N}$ is both p- and q-recognizable,
then X is ultimately periodic (finite union of A. P.).

Corollary

There exists sets which are

- *k-recognizable for any k (ultimately periodic sets),*
- *k-recognizable for some (minimal) k and exactly all the k^m,*
- *k-recognizable for no k.*

Example

The set of even integers is k-recognizable for any k.
The set $\{2^n \mid n \geq 0\}$ is 2-recognizable but not 3-recognizable.
The set of primes is never k-recognizable.

Theorem (Cobham '69)

Let $p, q \geq 2$ be two multiplicatively independent integers.
If $X \subseteq \mathbb{N}$ is both p- and q-recognizable,
then X is ultimately periodic (finite union of A. P.).

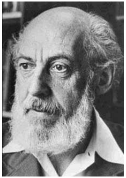
Corollary

There exists sets which are

- *k-recognizable for any k (ultimately periodic sets),*
- *k-recognizable for some (minimal) k and exactly all the k^m,*
- *k-recognizable for no k.*

Example

The set of even integers is k-recognizable for any k.
The set $\{2^n \mid n \geq 0\}$ is 2-recognizable but not 3-recognizable.
The set of primes is never k-recognizable.

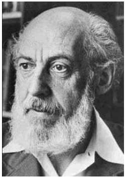
 S. Eilenberg'74 (p.118)

"*The proof is correct, long and hard. It is a challenge to find a more reasonable proof of this fine theorem*"

$$\{p^m/q^n \mid m,n \geq 0\} \text{ is dense in } [0,+\infty)$$

VARIOUS PROOF SIMPLIFICATIONS AND GENERALIZATIONS

G. Hansel'82, D. Perrin'90, F. Durand'05, V. Bruyère'97,
F. Point, C. Michaux'96, R. Villemaire, A. Bès'00, J. Bell'05,
J. Honkala, S. Fabre, C. Reutenauer, A.L. Semenov'77,
L. Waxweiler'06, M.R.'06…

 S. Eilenberg'74 (p.118)

"*The proof is correct, long and hard. It is a challenge to find a more reasonable proof of this fine theorem*"

$$\{p^m/q^n \mid m,n \geq 0\} \text{ is dense in } [0,+\infty)$$

VARIOUS PROOF SIMPLIFICATIONS AND GENERALIZATIONS

G. Hansel'82, D. Perrin'90, F. Durand'05, V. Bruyère'97,
F. Point, C. Michaux'96, R. Villemaire, A. Bès'00, J. Bell'05,
J. Honkala, S. Fabre, C. Reutenauer, A.L. Semenov'77,
L. Waxweiler'06, M.R.'06...

THEOREM (J.R. BÜCHI'60)

k-recognizable sets are exactly the sets definable by first order formula in the "extended" Presburger arithmetic $\langle \mathbb{N}, +, V_k \rangle$.

EXAMPLE

$\phi(x) \equiv (\exists y)(x = y + y)$, $X = \{x \in \mathbb{N} \mid \langle \mathbb{N}, +, V_k \rangle \models \phi(x)\}$.

THEOREM (G. CHRISTOL, T. KAMAE, M. MENDÈS FRANCE, G. RAUZY'80)

Let p prime. A set S is p-recognizable iff the formal power series

$$S(X) = \sum_{n \geq 0} \chi_S(n) X^n \text{ where } \chi_S(n) = 1 \Leftrightarrow n \in S,$$

is algebraic over $\mathbb{F}_p(X)$ (i.e., root of a polynomial over $\mathbb{F}_p[X]$).

THEOREM (J.R. BÜCHI'60)

k-recognizable sets are exactly the sets definable by first order formula in the "extended" Presburger arithmetic $\langle \mathbb{N}, +, V_k \rangle$.

EXAMPLE

$\phi(x) \equiv (\exists y)(x = y + y)$, $X = \{x \in \mathbb{N} \mid \langle \mathbb{N}, +, V_k \rangle \models \phi(x)\}$.

THEOREM (G. CHRISTOL, T. KAMAE, M. MENDÈS FRANCE, G. RAUZY'80)

Let p prime. A set S is p-recognizable iff the formal power series

$$S(X) = \sum_{n \geq 0} \chi_S(n) X^n \text{ where } \chi_S(n) = 1 \Leftrightarrow n \in S,$$

is algebraic over $\mathbb{F}_p(X)$ (i.e., root of a polynomial over $\mathbb{F}_p[X]$).

THEOREM (COBHAM'72)

"k-recognizable sets = k-automatic sets"

EXAMPLE ((PROUHET)-THUE-MORSE SEQUENCE)

$f : a \mapsto ab,\ b \mapsto ba$, $g : a \mapsto 0,\ b \mapsto 1$.

$f^0(a)$	a
$f^1(a)$	ab
$f^2(a)$	$abba$
$f^3(a)$	$abbabaab$
$f^4(a)$	$abbabaabbaababba$
$\vdots$	$\vdots$

$$\mathbf{t} = 01101001100101101001011001101001\cdots$$

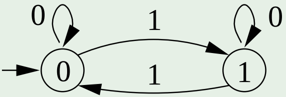
is 2-automatic : "generated by an iterated morphism of constant length 2".

EXAMPLE (CONT.)

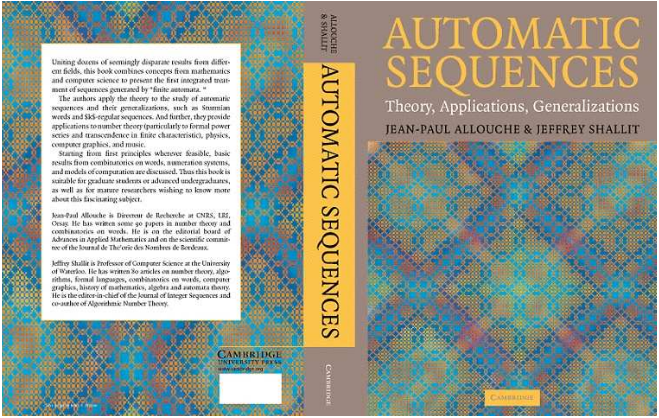
$\mathbf{t}_n = 1$ iff $\mathrm{rep}_2(n)$ contains an odd number of 1's,

0	1	2	3	4	5	6	7	8	$\cdots$
ε	1	10	11	100	101	110	111	100	$\cdots$
0	1	1	0	1	0	0	1	1	$\cdots$

This set is 2-recognizable,



Jean-Paul Allouche, Jeffrey Shallit ’04



PROUHET'S PROBLEM'1851

Find a partition of $A_n = \{0, 1, 2, 3, \ldots, 2^n - 1\}$ into two sets

$$I = \{i_1, \ldots, i_{2^{n-1}}\} \quad \text{and} \quad J = \{j_1, \ldots, j_{2^{n-1}}\}$$

such that (Multi-grades) : $\begin{cases} \sum_{i \in I} i & = & \sum_{j \in J} j \\ \sum_{i \in I} i^2 & = & \sum_{j \in J} j^2 \\ & \vdots & \\ \sum_{i \in I} i^{n-1} & = & \sum_{j \in J} j^{n-1} \end{cases}$

$n = 3$: 01101001, $I = \{0, 3, 5, 6\}$ and $J = \{1, 2, 4, 7\}$

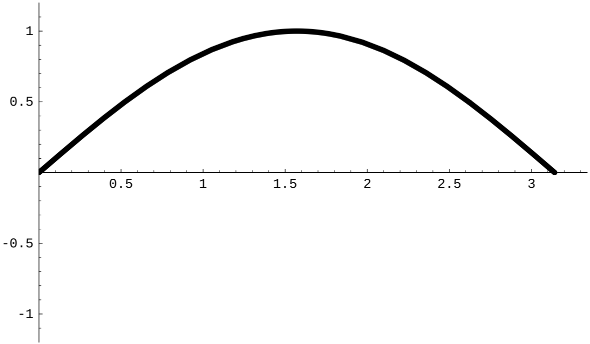
$$0 + 3 + 5 + 6 = 14 = 1 + 2 + 4 + 7 \text{ and}$$
$$0 + 9 + 25 + 36 = 70 = 1 + 4 + 16 + 49.$$

Question

What is the sign of

$$f_n(x) = \sin x \ \sin 2x \ \sin 4x \ \cdots \sin 2^n x$$

over $[0, \pi]$?

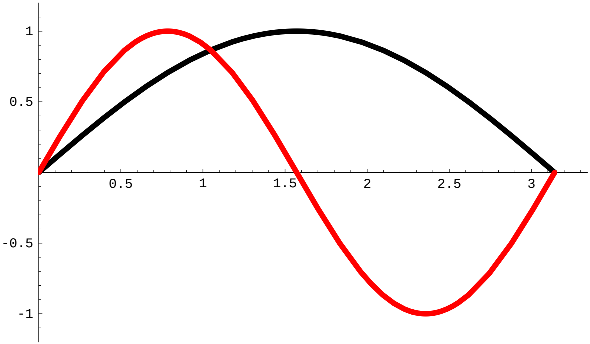


QUESTION

What is the sign of

$$f_n(x) = \sin x \ \sin 2x \ \sin 4x \ \cdots \sin 2^n x$$

over $[0, \pi]$?

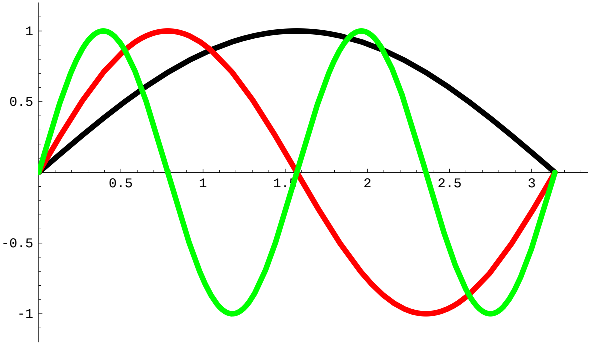


QUESTION

What is the sign of

$$f_n(x) = \sin x \ \sin 2x \ \sin 4x \ \cdots \sin 2^n x$$

over $[0, \pi]$?

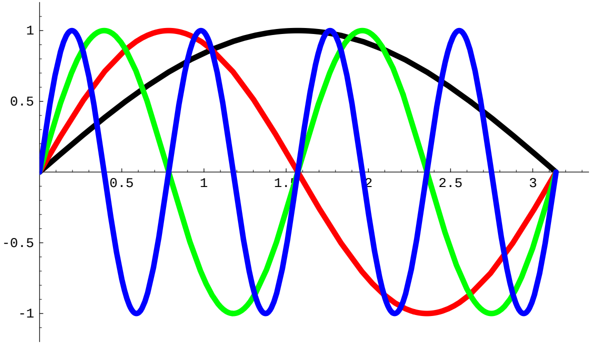


Question

What is the sign of

$$f_n(x) = \sin x \ \sin 2x \ \sin 4x \ \cdots \sin 2^n x$$

over $[0, \pi]$?



$n = 0$	$+$
$n = 1$	$+ \ -$
$n = 2$	$+ \ - \ - \ +$
$n = 3$	$+ \ - \ - \ + \ - \ + \ + \ -$

$+$	$-$	$-$	$+$	$-$	$+$	$+$	$-$	$-$	$+$	$+$	$-$	$+$	$-$	$-$	$+$	$\cdots$
0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	0	$\cdots$

A word w is an *overlap* if $w = avava$, a is a letter,

$$w = avava = avava.$$

THEOREM (MORSE-HEDLUND)

The infinite word $\mathbf{t}$ is overlap-free.

Recurrent geodesics on a surface of negative curvature (Morse 1921, Morse-Hedlund 1938–1940).

$n=0$	$+$
$n=1$	$+\ -$
$n=2$	$+\ -\ -\ +$
$n=3$	$+\ -\ -\ +\ -\ +\ +\ -$

$+$	$-$	$-$	$+$	$-$	$+$	$+$	$-$	$-$	$+$	$+$	$-$	$+$	$-$	$-$	$+$	$\cdots$
0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	0	$\cdots$

A word w is an *overlap* if $w = avava$, a is a letter,

$$w = avava = avava.$$

Theorem (Morse-Hedlund)

The infinite word **t** *is overlap-free.*

Recurrent geodesics on a surface of negative curvature (Morse 1921, Morse-Hedlund 1938–1940).

Let's come back to U-recognizability

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".

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".

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:
$\mathcal{L}_U$ *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, $< \alpha$).

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

After G. Hansel's talk during JM'94 in Mons and knowing Shallit's results, P. Lecomte has the following question :

- 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).

A QUESTION

After G. Hansel's talk during JM'94 in Mons and knowing Shallit's results, P. Lecomte has the following question :

- 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).

A QUESTION

After G. Hansel's talk during JM'94 in Mons and knowing Shallit's results, P. Lecomte has the following question :

- 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).

A QUESTION

After G. Hansel's talk during JM'94 in Mons and knowing Shallit's results, P. Lecomte has the following question :

- 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).

A QUESTION

After G. Hansel's talk during JM'94 in Mons and knowing Shallit's results, P. Lecomte has the following question :

- 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).

A QUESTION

After G. Hansel's talk during JM'94 in Mons and knowing Shallit's results, P. Lecomte has the following question :

- 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.R. '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 $L = \{\varepsilon\} \cup 1\{0, 01\}^*$

EXAMPLE (NON POSITIONAL)

Non positional numeration system : $L = a^*b^*$ $\Sigma = \{a < b\}$

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

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

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 $L = \{\varepsilon\} \cup 1\{0, 01\}^*$

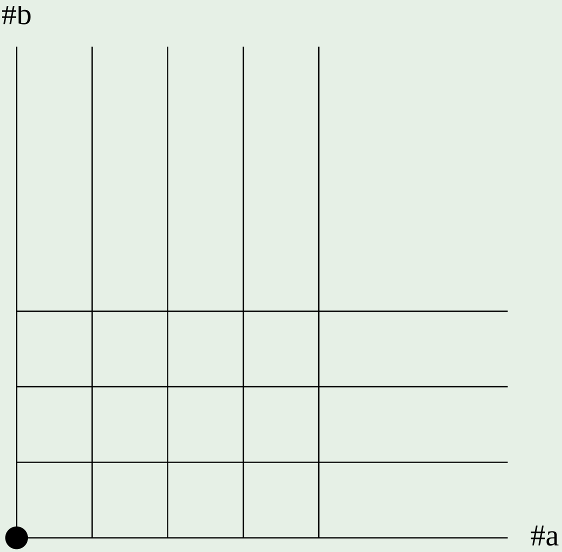
EXAMPLE (NON POSITIONAL)

Non positional numeration system : $L = a^* b^*$ $\Sigma = \{a < b\}$

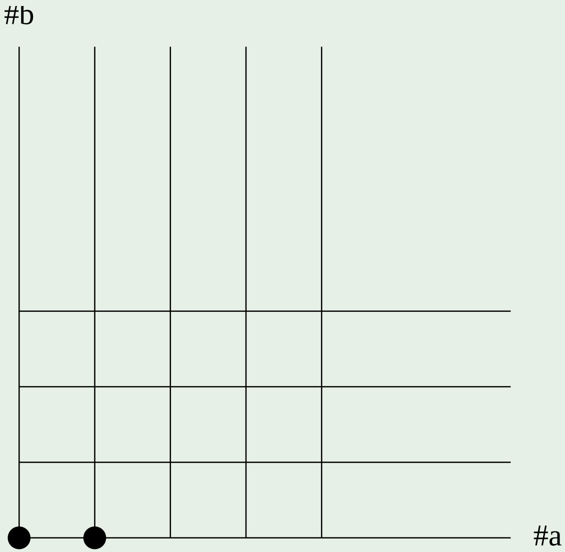
n	0	1	2	3	4	5	6	$\cdots$
$\mathrm{rep}(n)$	ε	a	b	aa	ab	bb	aaa	$\cdots$

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

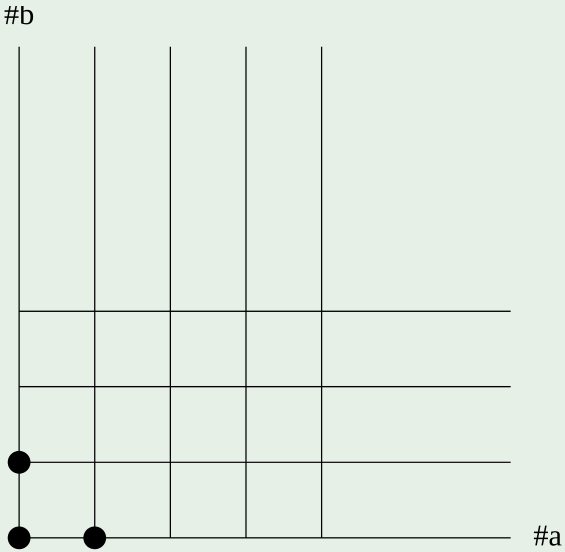
Example (continued...)



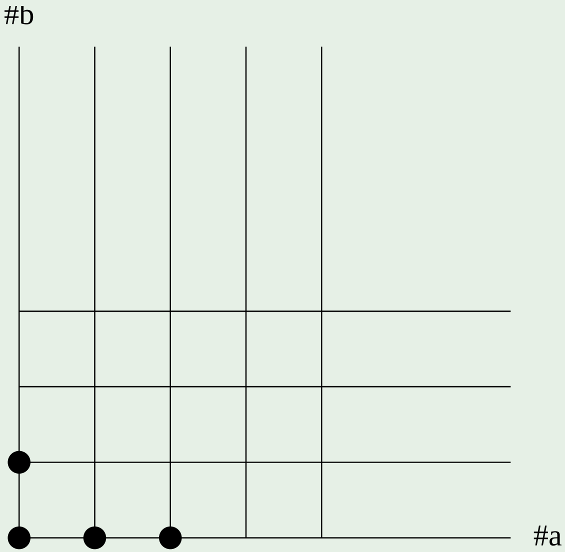
Example (continued...)



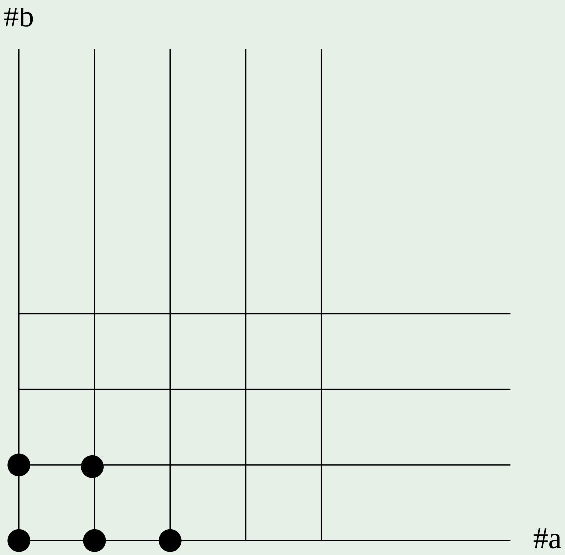
Example (continued...)



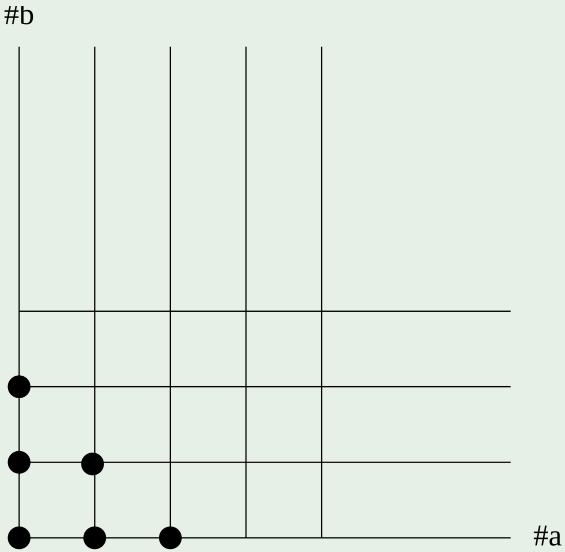
Example (continued...)



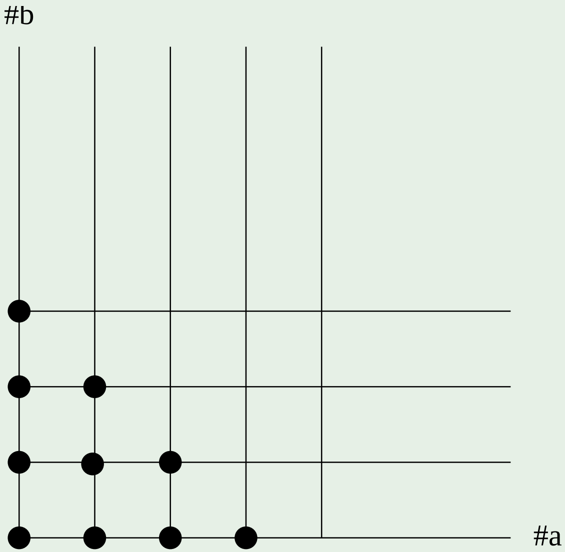
Example (continued...)



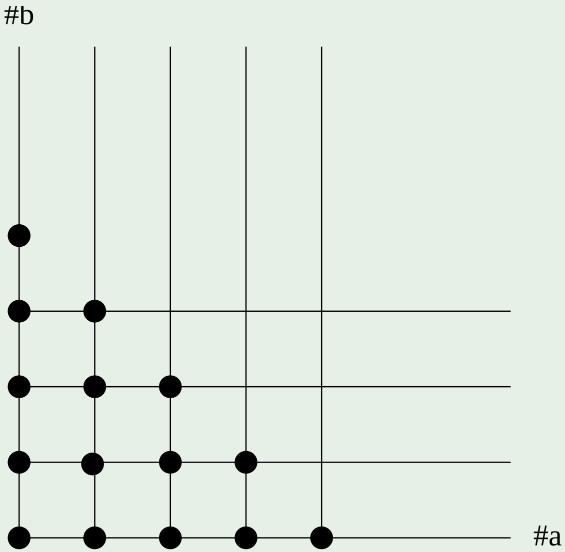
EXAMPLE (CONTINUED...)



Example (continued...)



Example (continued...)



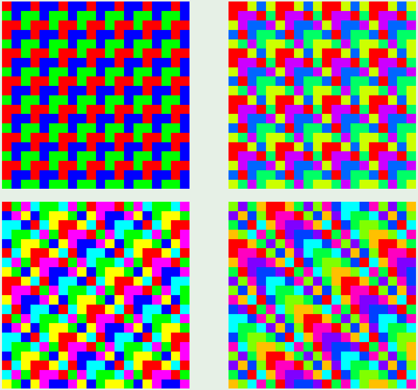
Many natural questions...

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

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.

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.

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.

Bounded languages $a_1^* \cdots a_\ell^*$

Let $S = (a_1^* \cdots a_\ell^*, \{a_1 < \cdots < a_\ell\})$. We have

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

Corollary (Katona's expansion'66)

Let $\ell \in \mathbb{N} \setminus \{0\}$. Any integer n can be uniquely written as

$$n = \binom{z_\ell}{\ell} + \binom{z_{\ell-1}}{\ell - 1} + \cdots + \binom{z_1}{1}$$

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

Theorem

For the abstract numeration system $S = (a^*b^*c^*, \{a < b < c\})$*, if* $\beta \in \mathbb{N} \setminus \{0, 1\}$ *is such that* $\beta \not\equiv \pm 1 \pmod 6$ *then the multiplication by* β^3 *does not preserve the* S*-recognizability.*

Conjecture

Multiplication by β^ℓ preserves S-recognizability for the abstract numeration system $S = (a_1^* \cdots a_\ell^*, \{a_1 < \cdots < a_\ell\})$ iff

$$\beta = \prod_{i=1}^{k} p_i^{\theta_i}$$

where $p_1, \ldots, p_k$ are prime numbers strictly greater than ℓ.

Why just looking at multiplication by β^{ℓ} ?

Definition of complexity

Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$, $\mathbf{u}_q(n) = \#\{w \in \Sigma^n \mid \delta(q, w) \in F\}$
i.e., number of words of length n accepted from q in $\mathcal{A}$.

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

Theorem

If $\mathbf{u}_{q_0}(n) = \#(L \cap \Sigma^n)$ is in $\Theta(n^k)$ and if $\lambda \neq \beta^{k+1}$ then there exists X which is S-recognizable and such that λX is not.

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.

EXAMPLE (CHARACTERISTIC SEQUENCE OF SQUARES)

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

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

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

Analogous to the Cobham's result from '72

THEOREM (A. MAES, M.R. '02)

A set is "morphic" iff it is S-recognizable for some abstract system S.

EXAMPLE (CHARACTERISTIC SEQUENCE OF SQUARES)

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

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

$$g(f^\omega(a)) = 1100100001000000100000000010\cdots$$

Analogous to the Cobham's result from '72

THEOREM (A. MAES, M.R. '02)

A set is "morphic" iff it is S-recognizable for some abstract system S.

EXAMPLE (BASE 10)

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

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

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

THIS DESERVES NOTATION

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

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

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

EXAMPLE (BASE 10)

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

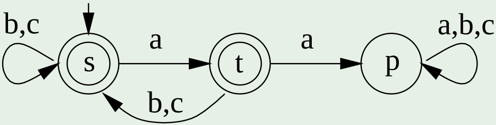
$$\frac{1}{10}, \quad \frac{14}{100}, \quad \frac{141}{1000}, \quad \ldots, \quad \frac{val(w_n)}{10^n}, \quad \cdots$$

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

THIS DESERVES NOTATION

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

Example (Avoid *aa* on three letters)



w	$\mathrm{val}(w)$	$\mathbf{v}_{q_0}(\|w\|)$	$\mathrm{val}(w)/\mathbf{v}_{q_0}(\|w\|)$
bc	8	12	0.666666666666667
bac	19	34	0.558823529411765
babc	52	94	0.553191489361702
babac	139	258	0.538759689922481
bababc	380	706	0.538243626062323
bababac	1035	1930	0.536269430051813
babababc	2828	5274	0.536215396283656

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

Numerical value of a word $w = w_1 \cdots w_\ell \in L$

$$\mathrm{val}(w) = \sum_{i=1}^{\ell} \sum_{q \in Q} (\theta_{q,i}(w) + \delta_{q,s})\, \mathbf{u}_q(|w| - i)$$

where $\theta_{q,i}(w) = \#\{\sigma < w_i \mid s.w_1 \cdots w_{i-1}\sigma = q\}$

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

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

Numerical value of a word $w = w_1 \cdots w_\ell \in L$

$$\mathrm{val}(w) = \sum_{i=1}^{\ell} \sum_{q \in Q} (\theta_{q,i}(w) + \delta_{q,s})\, \mathbf{u}_q(|w| - i)$$

where $\theta_{q,i}(w) = \#\{\sigma < w_i \mid s.w_1 \cdots w_{i-1}\sigma = q\}$

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

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

Let $\beta = \beta_{q_0}$ 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

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

Remark [W. Steiner, M.R. '05]

By "normalizing" we can specify the value of a_{q_0}, why not take $a_{q_0} = 1 - \frac{1}{\beta}$?

Doing so, we obtain something close to β-expansion...

Let $\beta = \beta_{q_0}$ 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

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

Remark [W. Steiner, M.R. '05]

By "normalizing" we can specify the value of a_{q_0}, why not take $a_{q_0} = 1 - \frac{1}{\beta}$?

Doing so, we obtain something close to β-expansion...

Let $\beta = \beta_{q_0}$ 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

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

Remark [W. Steiner, M.R. '05]

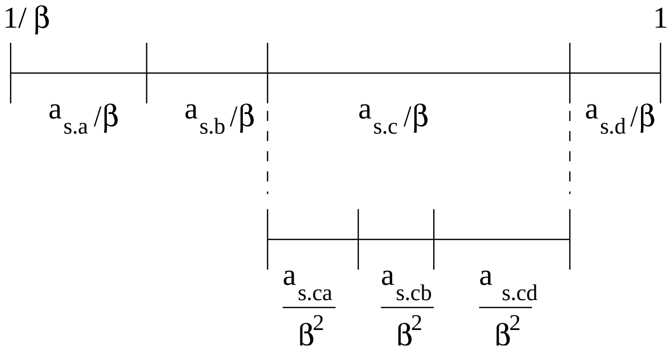
By "normalizing" we can specify the value of a_{q_0}, why not take $a_{q_0} = 1 - \frac{1}{\beta}$?

Doing so, we obtain something close to β-expansion...

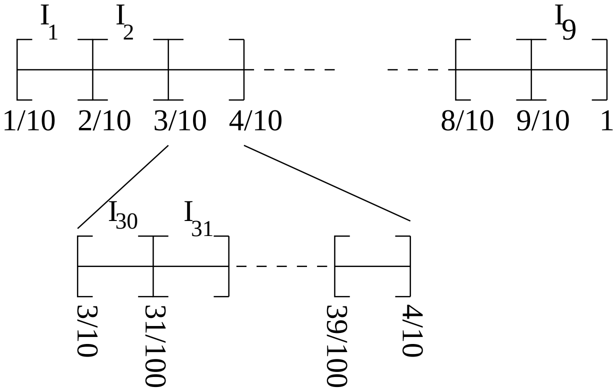
$W = W_1 W_2 \cdots$

$$x = \frac{1}{\beta} + \sum_{j=1}^{\infty} \alpha_{q_0.W_1 \cdots W_{j-1}}(W_j)\ \beta^{-j}$$

where $\alpha_q(\sigma) = \sum_{\tau < \sigma} a_{q.\tau}$.



This generalizes classical base 10 system :



This gives rises to several question…

Which real number have an ultimately periodic representation ?