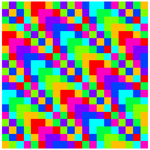
NUMERATION SYSTEMS: A LINK BETWEEN NUMBER THEORY AND FORMAL LANGUAGE THEORY

Michel Rigo

http://www.discmath.ulg.ac.be/

DLT 2010 – UWO, London, Ontario, 19th August 2010

UNIVERSITÉ de Liège



Sets of numbers	Numeration system	finite/infinite words or sequences
$\mathbb{N}$	integer base	
$\mathbb{Z}$	linear recurrence	
$\mathbb{Q}$	numeration basis	
$\mathbb{R}$	substitutive	
Gaussian int.	abstract	
$\mathbb{C}$	Ostrowski system	
$\mathbb{F}_q[X]$	factorial system	
	β-expansions	
vectors	continued fractions	
of these	canonical number sys.	
$\vdots$	$\vdots$	

Sets of numbers	Numeration system	finite/infinite words or sequences
$\mathbb{N}$	integer base	
$\mathbb{Z}$	linear recurrence	$A_2 = \{0, 1\}$
$\mathbb{Q}$	numeration basis	
$\mathbb{R}$	substitutive	$\mathrm{rep}_2(n)$, $n \in \mathbb{N}$, is a
Gaussian int.	abstract	finite word
$\mathbb{C}$	Ostrowski system	
$\mathbb{F}_q[X]$	factorial system	with $X \subseteq \mathbb{N}$,
	β-expansions	$\mathrm{rep}_2(X)$ is a
vectors	continued fractions	language over A_2
of these	canonical number sys.	
$\vdots$	$\vdots$	

Integer base, e.g., $k = 2$

$$\mathrm{rep}_2 : \mathbb{N} \to \{0,1\}^*,\ n = \sum_{i=0}^{\ell} d_i\, 2^i,\ \mathrm{rep}_2(n) = d_\ell \cdots d_0$$

$$\mathrm{rep}_2(37) = 100101 \quad \text{and} \quad \mathrm{val}_2(100101) = 37$$

Sets of numbers	Numeration system	finite/infinite words or sequences
$\mathbb{N}$	integer base	
$\mathbb{Z}$	linear recurrence	$A_2 = \{0, 1\}$
$\mathbb{Q}$	numeration basis	
$\mathbb{R}$	substitutive	$\text{rep}_2(r)$, $r \in \mathbb{R}$, is an
Gaussian int.	abstract	infinite word
$\mathbb{C}$	Ostrowski system	
$\mathbb{F}_q[X]$	factorial system	with $X \subseteq \mathbb{R}$,
	β-expansions	$\text{rep}_2(X)$ is an
vectors	continued fractions	ω-language over A_k
of these	canonical number sys.	
⋮	⋮	maybe several rep.

Integer base, e.g., $k = 2$ (base-complement for neg. numbers)

$$\text{rep}_2 : \mathbb{R} \to \{0, 1\}^* \star \{0, 1\}^\omega,\ \{r\} = \sum_{i=1}^{+\infty} d_i\, 2^{-i}.$$

The set of representations of $3/2$ is $0^+1 \star 10^\omega \cup 0^+1 \star 01^\omega$.

Sets of numbers	Numeration system	finite/infinite words or sequences
$\mathbb{N}$	integer base	$A_F = \{0, 1\}$
$\mathbb{Z}$	linear recurrence	
$\mathbb{Q}$	numeration basis	greedy choice
$\mathbb{R}$	substitutive	$\mathrm{rep}_F(n)$, $n \in \mathbb{N}$, is a
Gaussian int.	abstract	finite word
$\mathbb{C}$	Ostrowski system	
$\mathbb{F}_q[X]$	factorial system	with $X \subseteq \mathbb{N}$,
	β-expansions	$\mathrm{rep}_2(X)$ is a
vectors	continued fractions	language over A_F
of these	canonical number sys.	
$\vdots$	$\vdots$	maybe several rep.

Fibonacci numeration system (Zeckendorf 1972)

$\ldots, 34, 21, 13, 8, 5, 3, 2, 1 = (F_n)_{n\geq 0}$ and $\mathrm{rep}_F(11) = 10100$

but $\mathrm{val}_F(10100) = \mathrm{val}_F(10011) = \mathrm{val}_F(1111)$ $\quad U_{n+2} = U_{n+1} + U_n$.

Sets of numbers	Numeration system	finite/infinite words or sequences
$\mathbb{N}$	integer base	
$\mathbb{Z}$	linear recurrence	$A_\beta = \{0, 1\}$
$\mathbb{Q}$	numeration basis	
$\mathbb{R}$	substitutive	β-expansions are
Gaussian int.	abstract	infinite words
$\mathbb{C}$	Ostrowski system	
$\mathbb{F}_q[X]$	factorial system	maybe several rep.
	β-expansions	
vectors	continued fractions	β-development is
of these	canonical number sys.	the lexico. largest
$\vdots$	$\vdots$	

β-expansions (Rényi 1957, Parry 1960), e.g., $\beta = (1+\sqrt{5})/2$

$$r \in (0,1),\ r = \sum_{i=1}^{+\infty} c_i\, \beta^{-i} \qquad \beta^2 = \beta + 1$$

$$d_\beta(\pi - 3) = 00001010100100010101010\cdots.$$

Sets of numbers	Numeration system	finite/infinite words or sequences
$\mathbb{N}$	integer base	
$\mathbb{Z}$	linear recurrence	$A = \mathbb{N}$
$\mathbb{Q}$	numeration basis	
$\mathbb{R}$	substitutive	$\mathrm{rep}(n)$, $n \in \mathbb{N}$, is a
Gaussian int.	abstract	finite word
$\mathbb{C}$	Ostrowski system	over an
$\mathbb{F}_q[X]$	factorial system	infinite alphabet
	β-expansions	
vectors	continued fractions	
of these	canonical number sys.	
$\vdots$	$\vdots$	

Factorial numeration system

$$\ldots,\ 720,\ 120,\ 24,\ 6,\ 2,\ 1 = (j!)_{j\geq 0}, \qquad n = \sum_{i=0}^{\ell} d_i\, i!,$$
$$\mathrm{rep}(719) = 54321.$$

H. Lenstra, Profinite Fibonacci numbers, **EMS Newsletter**'06

Sets of numbers	Numeration system	finite/infinite words or sequences
$\mathbb{N}$	integer base	
$\mathbb{Z}$	linear recurrence	$A = \{0, 1, X, X+1\}$
$\mathbb{Q}$	numeration basis	finite alphabet
$\mathbb{R}$	substitutive	
Gaussian int.	abstract	$\mathrm{rep}_B(P)$, $P \in \mathbb{F}_2[X]$ is
$\mathbb{C}$	Ostrowski system	a finite word
$\mathbb{F}_q[X]$	factorial system	
	β-expansions	with $\mathcal{T} \subseteq \mathbb{F}_2[X]$
vectors	continued fractions	$\mathrm{rep}_B(\mathcal{T})$ is a
of these	canonical number sys.	language over A
$\vdots$	$\vdots$	

"Polynomial base", e.g., $B = X^2 + 1$, $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$

$$P = \sum_{i=0}^{\ell} C_i\, B^i \text{ with } \deg C_i < \deg B,$$

$$X^6 + X^5 + 1 = 1.B^3 + (X+1).B^2 + 1.B + X.B^0$$

Sets of numbers	Numeration system	finite/infinite words or sequences
$\mathbb{N}$	integer base	
$\mathbb{Z}$	linear recurrence	
$\mathbb{Q}$	numeration basis	
$\mathbb{R}$	substitutive	
Gaussian int.	abstract	
$\mathbb{C}$	Ostrowski system	
$\mathbb{F}_q[X]$	factorial system	
	β-expansions	
vectors	continued fractions	
of these	canonical number sys.	
$\vdots$	$\vdots$	
numbers		formal languages
arithmetic/	$\Leftrightarrow$	theory
algebraic		syntactical
properties		properties

Verification

Cryptography

Dynamical systems

Number theory

Symbolic dynacmics

Automatic structures

Ergodic theory

Numeration systems

Combinatorics on words

Fractal

Formal language theory

Combinatorial games

Logic

Design of algorithms

Computability

Discrete geometry

Analysis of algorithms

Outline of this talk

- Sets of integers with an integer base
- Multidimensional setting
- Sets of reals with an integer base
- Moving to non-standard systems

- Transcendence of real numbers
- Some results about primes
- Adamczewski's positive view on k-recognizable sets

Sets of integers with an integer base 1/10

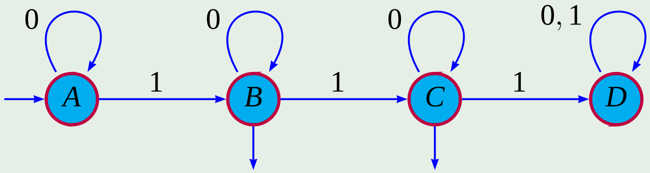
Sets of integers having a somehow simple description

Definition

A set $X \subset \mathbb{N}$ is *k-recognizable*, if $\mathrm{rep}_k(X)$ is a regular language.

A 2-recognizable set

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

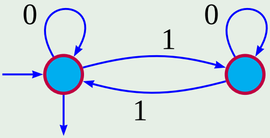


$1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 16, 17, 18, 20, 24, \ldots$

$1, 10, 11, 100, 101, 110, 1000, 1001, 1010, 1100, 10000, 10001, \ldots$

Prouhet–Thue–Morse set

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



$$0, 3, 5, 6, 9, 10, 12, 15, 17, 18, \ldots$$

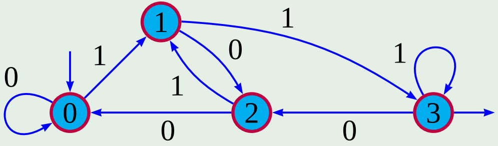
$$\varepsilon, 11, 101, 110, 1001, 1010, 1100, 1111, 10001, 10010, \ldots$$

The set of powers of 2

$$\mathrm{rep}_2(\{2^i \mid i \geq 0\}) = 10^*$$

$$1, 2, 4, 8, 16, 32, 64, \ldots$$

AN ULTIMATELY PERIODIC SET, E.G., $4\mathbb{N}+3$



$$3, 7, 11, 15, 19, 23, 27, 31, \dots$$

EXERCISE

Let $k \geq 2$. Show that any arithmetic progression $p\mathbb{N}+q$ is k-recognizable (and consequently any ultimately periodic set).

B. Alexeev, Minimal dfas for testing divisibility, **JCSS**'04

QUESTION

Does recognizability depends on the choice of the base ?
Is a 2-recognizable set also 3-recognizable or 4-recognizable ?

EXERCISE

Let $k, t \geq 2$. Show that $X \subset \mathbb{N}$ is k-recognizable
IFF it is k^t-recognizable. $0 \mapsto 00, 1 \mapsto 01, 2 \mapsto 10, 3 \mapsto 11$

Powers of 2 in base 3 :

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

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

$$20122210112, 111022121001, 222122012002, 1222021101011,$$

$$10221112202022, 21220002111121, 120210012000012, \ldots$$

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.

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

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, . . .

A possible application

The set of powers of 2 or the Thue–Morse set are 2-recognizable but NOT 3-recognizable.

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

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

Following G. Hansel, first part of the proof of Cobham's theorem is to show that X is *syndetic*, i.e., $\mathbf{D}_X < +\infty$.

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 $\mathbf{D}_X < +\infty$.

For instance, $\{n^t \mid n \geq 0\}$ is k-recognizable for no $k \geq 2$.

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

- Logical characterization of k-recognizable sets

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 $n \geq 1$, $V_k(0) = 1$.

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

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

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

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 $\langle \mathbb{N}, + \rangle$.

- Automatic characterization of k-recognizable sets

Theorem (Cobham 1972) – Uniform tag sequences

A set X is k-recognizable / k-automatic IFF its characteristic sequence is generated through a k-uniform morphism + a coding.

$$g : \begin{cases} A & \mapsto & AB \\ B & \mapsto & BC \\ C & \mapsto & CD \\ D & \mapsto & DD \end{cases} \qquad f : \begin{cases} A & \mapsto & 0 \\ B & \mapsto & 1 \\ C & \mapsto & 1 \\ D & \mapsto & 0 \end{cases}$$

$$g(A) = AB,\ g^2(A) = ABBC,\ g^3(A) = ABBCBCCD, \ldots$$

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

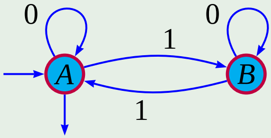
$$w = f(g^\omega(A)) = 01111101110100011101000100000000\cdots$$

feed a DFAO with k-ary rep. $, \forall n \geq 0,\ w_n = \tau(q_0 \cdot \mathrm{rep}_k(n))$

Sets of integers with an integer base 10/10

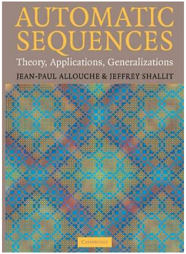
Another example (Thue–Morse)

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



$$g : A \mapsto AB,\ B \mapsto BA, \quad f : A \mapsto 1,\ B \mapsto 0$$

$$f(g^{\omega}(A)) = 10010110011010010110100110010110\cdots$$



J.-P. Allouche, J. Shallit, Cambridge Univ. Press, 2003.

Multidimensional setting 1/2

$k=2, d=2$

$$\mathrm{rep}_2\begin{pmatrix}5\\35\end{pmatrix}=\begin{pmatrix}000101\\100011\end{pmatrix},\quad \text{Alphabet } \left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix},\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}1\\1\end{pmatrix}\right\}$$

One can easily define k-recognizable subsets of $\mathbb{N}^d$.

Cobham–Semenov' Theorem (1977)

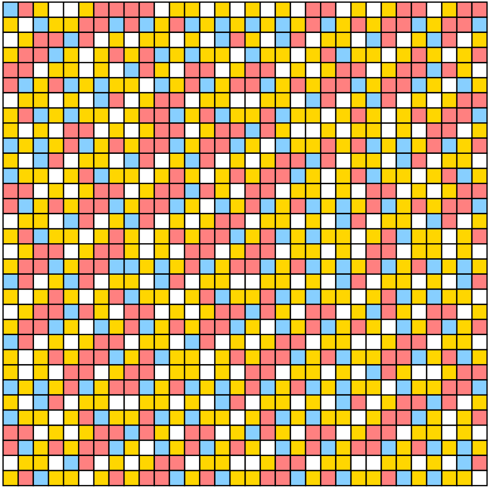
Let $k, \ell \geq 2$ be two multiplicatively independent integers.
A set $X \subseteq \mathbb{N}^d$ is k-rec. AND ℓ-rec. IFF X is definable in $\langle \mathbb{N}, + \rangle$

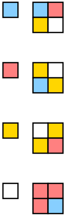
Natural extension of ultimate periodicity :

- definability in $\langle \mathbb{N}, + \rangle$,
- semi-linear sets,
- Muchnik's local periodicity (TCS'03)

Multidimensional setting 2/2

A 2-recognizable/2-automatic set in $\mathbb{N}^2$





O. Salon, Suites automatiques à multi-indices, **Sém TN Bord.**, 1986–1987.

Sets of reals with an integer base 1/2

$\text{rep}_2(\pi) = 11 \star 0010010000111111011010101000100010000\cdots$

Theorem (Boigelot–Jodogne–Wolper'05)

If $X \subseteq \mathbb{R}^d$ is first-order definable in $\langle \mathbb{R}, \mathbb{Z}, +, 0, < \rangle$, then X written in base $k \geq 2$ is recognizable by a weak deterministic RVA (Büchi automaton accepting *all* the encodings).

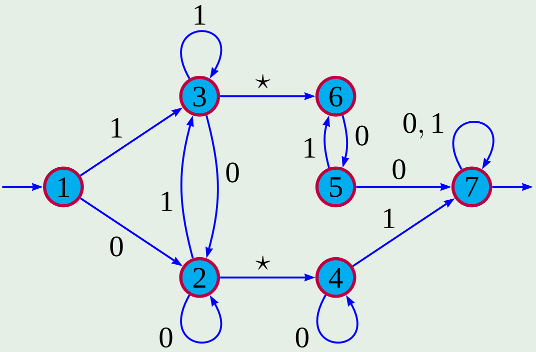
Theorem (Boigelot–Brusten'09)

Let $k, \ell \geq 2$ be multiplicatively independent integers. If $X \subseteq \mathbb{R}$ is both k- and ℓ-recognizable by two weak deterministic RVA, then it is definable in $\langle \mathbb{R}, \mathbb{Z}, +, 0, < \rangle$

Extension to $\mathbb{R}^d$: B. Boigelot, J. Brusten, J. Leroux, CADE'09, **LNCS** 5663.

Also see B. Adamczewski, J. P. Bell, An analogue of Cobham's thm. for fractals, to appear **TAMS**.

A BÜCHI AUTOMATON ACCEPTING $\{2n + (0, 4/3) \mid n \in \mathbb{Z}\}$



For instance $3/2$ is encoded by $0^+1 \star 10^\omega \cup 0^+1 \star 01^\omega$.

$3 \xrightarrow{\star} 6$: odd integer part $\quad 2 \xrightarrow{\star} 4$: even integer part

$\sum_{i=1}^{+\infty} 4^{-i} = 1/3$ corresponds to the cycle $\{5, 6\}$.

Recap: a set X is k-recognizable IFF its characteristic word is generated using a k-uniform morphism.

From k-automatic words to … morphic/substitutive words

$$\{\text{automatic words}\} \subsetneq \{\text{morphic words}\}$$

Tribonacci word

$$g : \begin{cases} A & \mapsto & AB \\ B & \mapsto & AC \\ C & \mapsto & A \end{cases} \qquad f : \begin{cases} A & \mapsto & 0 \\ B & \mapsto & 1 \\ C & \mapsto & 0 \end{cases}$$

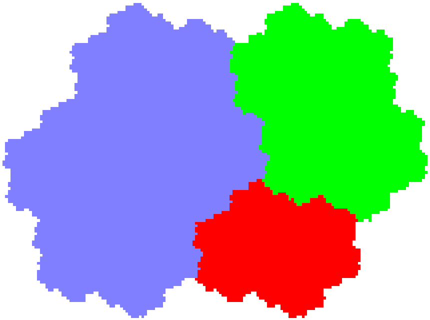
$$g(A) = AB,\ g^2(A) = ABAC,\ g^3(A) = ABACABA, \ldots$$

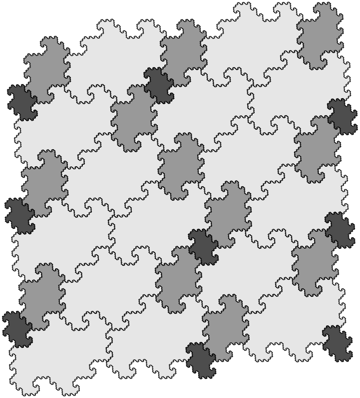
$$g^\omega(A) = ABACABAABACABABACABAABAC\cdots$$

$$f(g^\omega(A)) = 010001001000101000100100\cdots$$

B. Tan, Z.-Y. Wen, Some properties of the Tribonacci sequence, **Europ. J. Combin.**'07

Rauzy fractal





G. Rauzy, Nombres algébriques et substitutions, **BSMF**'82
V. Berthé, A. Siegel, Tilings associated with beta-numeration and substitutions, **INTEGERS**'05
V. Berthé, A. Siegel, J. Thuswaldner, Substitutions, Rauzy fractals and tilings, Chap. 4 in Combinatorics, Automata and Number Theory, CUP 2010.

SOMETHING MORE NASTY ?

$$g : \begin{cases} A \mapsto ABCC \\ B \mapsto \varepsilon \\ C \mapsto BA \end{cases} \qquad f : \begin{cases} A \mapsto 010 \\ B \mapsto 1 \\ C \mapsto \varepsilon \end{cases}$$

REMARK

We can always assume that f is a coding (letter-to-letter) and g is a non-erasing morphism

A. Cobham, On the Hartmanis-Stearns problem for a class of tag machines, '68
J.-P. Allouche, J. Shallit, CUP'03
J. Honkala, On the simplification of infinite morphic words, TCS'09

From k-automatic words to ... morphic/substitutive words

From k-recognizable subsets of $\mathbb{N}$ to ... substitutive sets

$$f(g^\omega(A)) = 0100010010001010001001000\cdots$$

Easy to generate the characteristic sequence of the substitutive set $\{1, 5, 8, 12, 14, 18, 21, \ldots\}$

We still have a notion of "automaticity":

Maes–R. (JALC 2002)

An infinite word w is morphic IFF there exists an abstract numeration system S such that w is S-automatic.

P. Lecomte, R., Numeration systems on a regular language, **TOCS**'01.

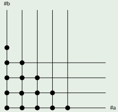
P. Lecomte, R., Abstract numeration systems, Chap. 3 in Combinatorics, Automata and Number Theory, CUP 2010.

MOVING TO NON-STANDARD SYSTEMS 5/9

An abstract numeration system is a regular language $L \subset A^*$ genealogically ordered where the alphabet A is totally ordered.

$L = a^*b^*$, $a < b$

ε	a	b	aa	ab	bb	aaa	aab	abb	$\cdots$
0	1	2	3	4	5	6	7	8	$\cdots$



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

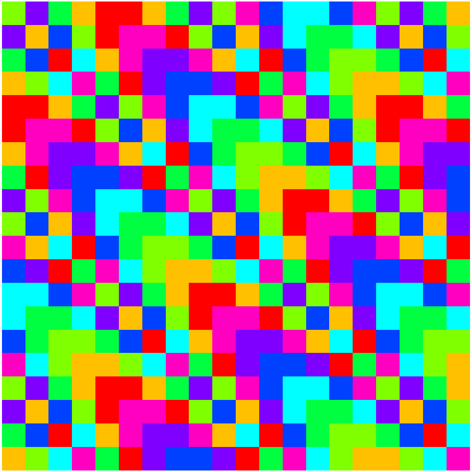
Katona, Lehmer, Fraenkel, Charlier, R., Steiner,...

feed a DFAO with k-ary rep. , $\forall n \geq 0,\ w_n = \tau(q_0 \cdot \mathrm{rep}_S(n))$

Two complementary formalisms:

morphisms and numeration systems

MOVING TO NON-STANDARD SYSTEMS 6/9



$\mathrm{val}(a^p b^q)$ modulo 8

THEOREM (P. LECOMTE, M.R.)

Let S be an abstract numeration system.
Any ultimately periodic set is S-recognizable.

THEOREM (D. KRIEGER *et al.* TCS'09)

Let L be a genealogically ordered regular language.
Any *periodic decimation* in L gives a regular language.
This result does not hold anymore if regular is replaced by context-free.

Matrix associated with a morphism
($\equiv$ adjacency matrix of the associated automaton)

Tribonacci morphism

$g : A \mapsto AB,\ B \mapsto AC,\ C \mapsto A$
$g^2 : A \mapsto ABAC,\ B \mapsto ABA,\ C \mapsto AB$
$g^3 : A \mapsto ABACABA,\ B \mapsto ABACAB,\ C \mapsto ABAC$

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad M^2 = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad M^3 = \begin{pmatrix} 4 & 3 & 2 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$\alpha_T \simeq 1.83929$$

Note: all letters have an occurrence in $g^\omega(A)$.

Primitive (or irreducible, i.e., strongly connected) components
$\rightarrow$ Perron–Frobenius theory $\rightarrow$ dominating eigenvalue

$$f(g^{\omega}(A)) = 0100010010001010001001000\cdots$$

the set $\{1,5,8,12,14,18,21,\ldots\}$ is α_T-substitutive ($\alpha_T \simeq 1.839$).

"META-THEOREM" F. DURAND

Let $\alpha,\beta>1$ be two multiplicatively independent Perron numbers. An infinite word is both α-substitutive and β-substitutive IFF it is ultimately periodic.

A *good substitution* has a primitive sub-substitution having the same dominating eigenvalue.

F. Durand, Sur les ensembles d'entiers reconnaissables, **JTNB**'98.
F. Durand, A generalization of Cobham's theorem, **TOCS**'98.
F. Durand, A thm. of Cobham for non primitive substitutions, **Acta Arith.**'02.
F. Durand, R., On Cobham's theorem, to appear Handbook (AutoMathA project).

Transcendence of real numbers 1/6

$r \in (0,1)$, $k \in \mathbb{N} \setminus \{0,1\}$

$$r = \sum_{i=1}^{+\infty} c_i \, k^{-i} \qquad c_1 c_2 c_3 \cdots$$

Factor (or subword) complexity function : $p_w(n)$ is the number of distinct factors of length n occurring in w.

$$1 \leq p_w(n) \leq (\#A)^n \quad \text{and} \quad p_w(n) \leq p_w(n+1)$$

Morse–Hedlund theorem

The following conditions are equivalent:

- The word w is ultimately periodic, i.e., $w = xy^\omega$.
- The complexity function p_w is bounded by a constant,
- There exists $m \in \mathbb{N}$ such that $p_w(m) = p_w(m+1)$.

TRANSCENDENCE OF REAL NUMBERS 2/6

COBHAM 1972

If w is k-automatic, then p_w is $\mathcal{O}(n)$.

PANSIOT (LNCS 172, 1984)

If w is pure morphic (no coding) and not ultimately periodic, then there exist constants C_1, C_2 such that $C_1 f(n) \le p_w(n) \le C_2 f(n)$ where $f(n) \in \{n, n\log n, n\log\log n, n^2\}$.

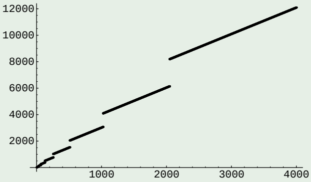
J.-P. Allouche, Sur la complexité des suites infinies, **BBMS**'94,
J. Cassaigne, F. Nicolas, Factor complexity, Chap. 4 in Combinatorics, Automata and Number Theory, CUP 2010.

Transcendence of real numbers 3/6

Thue–Morse word

$$t = 10010110011010010110100110010110 \cdots$$

$$p_t(n) = \begin{cases} 1 & \text{if } n = 0 \\ 2 & \text{if } n = 1 \\ 4 & \text{if } n = 2 \\ 4n - 2 \cdot 2^m - 4 & \text{if } 2 \cdot 2^m < n \leq 3 \cdot 2^m \\ 2n + 4 \cdot 2^m - 2 & \text{if } 3 \cdot 2^m < n \leq 4 \cdot 2^m \end{cases}$$



S. Brlek, Enumeration of factors in the Thue-Morse word, **DAM**'89
A. de Luca, S. Varricchio, On the factors of the Thue-Morse word on three symbols, **IPL**'88

Transcendence of real numbers 4/6

Cobham's conjecture

Let α be an algebraic irrational real number. Then the k-ary expansion of α cannot be generated by a finite automaton.

Following this question :

Hartmanis–Stearns (Trans. AMS'65)

Does it exist an algebraic irrational real number computable in linear time by a (multi-tape) Turing machine? i.e., the first n digits of the representation computable in $\mathcal{O}(n)$ operations.

J. P. Bell, B. Adamczewski, Automata in Number Theory, to appear Handbook (AutoMathA project).

ADAMCZEWSKI–BUGEAUD'07

Let $k \in \mathbb{N} \setminus \{0, 1\}$. The factor complexity of the k-ary expansion w of an algebraic irrational real number satisfies

$$\lim_{n \to +\infty} \frac{p_w(n)}{n} = +\infty.$$

Let $k \geq 2$ be an integer.
If α is an irrational real number whose k-ary expansion w has factor complexity in $\mathcal{O}(n)$, then α is transcendental.
So in particular, if w is k-automatic.

BUGEAUD–EVERTSE'08

Let $k \geq 2$ be an integer and ξ be an algebraic irrational real number with $0 < \xi < 1$. Then for any real number $\eta < 1/11$, the factor complexity $p(n)$ of the k-ary expansion of ξ satisfies

$$\lim_{n \to +\infty} \frac{p(n)}{n(\log n)^{\eta}} = +\infty.$$

The following slides are based on a talk given by B. Adamczewski in Leiden (Numeration, June 2010)

Minsky–Papert 1966

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

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

Schützenberger 1968

No infinite subset of the set of prime numbers can be recognized by a finite automaton.

Some results about primes 2/2

Fouvry–Mauduit 1996

Given a non-empty automatic set X associated with a strongly connected automaton, there exists $r > 0$ such that X contains infinitely many r-almost primes (product of at most r primes).

In 1968, Gelfond asked about the collection of prime numbers that belong to the Thue–Morse set

Mauduit–Rivat (Annals of Math. 2010)

$$\lim_{N \to +\infty} \frac{\#\{n \in \mathcal{P} \mid n \leq N \text{ and } s_2(n) \equiv 0 \bmod 2\}}{\#\{n \in \mathcal{P} \mid n \leq N\}} = \frac{1}{2}.$$

Negative answers :-(

- expansions of algebraic irrational real numbers are *not* automatic,
- the set $\mathcal{P}$ is *not* k-recognizable.

Positive view on k-recognizable sets 1/5

Let $\mathbb{K}$ be a field, $a(n) \in \mathbb{K}^{\mathbb{N}}$ be a $\mathbb{K}$-valued sequence and $k_1, \ldots, k_d \in \mathbb{K}$. The sequence $a(n)$ satisfies a *linear recurrence* over $\mathbb{K}$ if

$$a(n) = k_1 a(n-1) + \cdots + k_d\, a(n-d), \quad \forall n >>$$

Skolem–Mahler–Lech theorem

Let $a(n)$ be a linear recurrence over a field of characteristic 0. Then the zero set

$$\mathcal{Z}(a) = \{n \in \mathbb{N} \mid a(n) = 0\} \text{ is ultimately periodic.}$$

Remark

If $\mathbb{K}$ is a finite field, $a(n)$ (and so $\mathcal{Z}(a)$) is trivially ultimately periodic.

T. Tao, Effective Skolem–Mahler–Lech theorem in Structure and Randomness, **AMS**'08.

Positive view on k-recognizable sets 2/5

If $\mathbb{K}$ is an infinite field of positive characteristic...

Lech's example

$$a(n) := (1+t)^n - t^n - 1 \in \mathbb{F}_p(t).$$

The sequence a satisfies a linear recurrence, for $n > 3$

$$a(n) = (2+2t)\,a(n-1) + (1+3t+t^2)\,a(n-2) - (t+t^2)\,a(n-3).$$

We have

$$a(p^j) = (1+t)^{p^j} - t^{p^j} - 1 = 0$$

while $a(n) \neq 0$ if n is not a power of p, and so we obtain that

$$\mathcal{Z}(a) = \{1, p, p^2, p^3, \ldots\}.$$

POSITIVE VIEW ON k-RECOGNIZABLE SETS 3/5

DERKSEN'S EXAMPLE

Consider the sequence $a(n)$ in $\mathbb{F}_p(x, y, z)$ defined by

$$a(n) := (x+y+z)^n - (x+y)^n - (x+z)^n - (y+z)^n + x^n + y^n + z^n.$$

It can be proved that :

- The sequence $a(n)$ satisfies a linear recurrence.
- The zero set is given by

$$\mathcal{Z}(a) = \{p^n \mid n \in \mathbb{N}\} \cup \{p^n + p^m \mid n, m \in \mathbb{N}\}.$$

$\mathcal{Z}(a)$ can be *more pathological* than in characteristic zero
but. . . think about p-recognizable sets !

THEOREM (H. DERKSEN'07)

Let $a(n)$ be a linear recurrence over a field of characteristic p. Then the set $\mathcal{Z}(a)$ is a p-recognizable set.

Derksen gave a further refinement of this result:
not all p-recognizable sets are zero sets of linear recurrences defined over fields of characteristic p.

THEOREM (ADAMCZEWSKI–BELL'2010)

Let $\mathbb{K}$ be a field and Γ be a finitely generated subgroup of $\mathbb{K}^*$. Consider the linear equations

$$a_1X_1 + \cdots + a_dX_d = 1$$

where $a_1, \ldots, a_d \in \mathbb{K}$ and look for solutions in Γ^d. The set of solutions is a "p-automatic subset of Γ^d" (not defined here).

If $\mathbb{K}$ is a field of characteristic 0, many contributions due to Beukers, Evertse, Lang, Mahler, van der Poorten, Schlickewei and Schmidt.

J.-H. Evertse, H.P. Schlickewei, W.M. Schmidt, Linear equations in variables which lie in a multiplicative group, Annals of Math. 2002.

In the conference proceedings :

- Connection with combinatorial game theory
- Abridged bibliographic notes
- A list of open problems

Combinatorics, Automata and Number Theory, CUP 2010, Encycl. of Math. and its Appl., V. Berthé, M. R. Eds.

How many times did the name Cobham appear in this talk ?

2^4