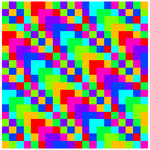
RECOGNIZABLE SETS OF INTEGERS

Michel Rigo

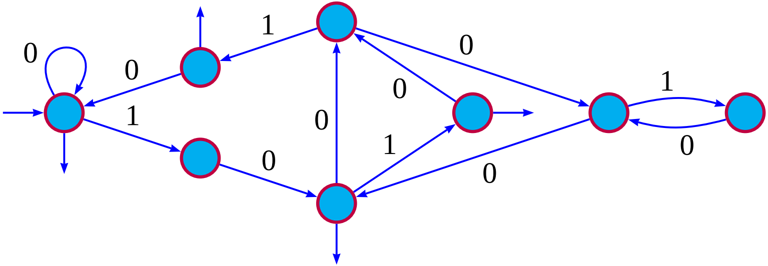
http://www.discmath.ulg.ac.be/

1st Joint Conference of the Belgian, Royal Spanish and Luxembourg Mathematical Societies, June 2012, Liège

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

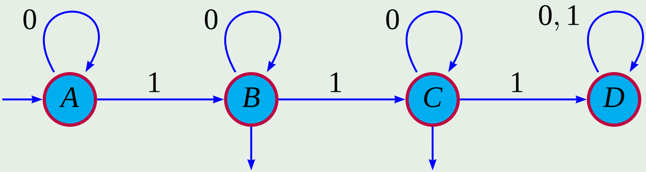
A set $X \subset \mathbb{N}$ is *k-recognizable*, if the set of base k expansions of the elements of X is accepted by some finite automaton, *i.e.*, $\mathrm{rep}_k(X)$ is a regular language.

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

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

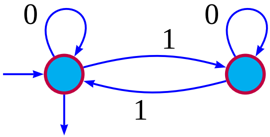


$$X = \{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 16, 17, 18, 20, 24, \ldots\}$$

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

Some examples

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



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

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

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

- 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

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 ? (exercice)
- 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 irational number.
The sequence $(\{n\theta\})_{n\geq 0}$ is dense in $[0,1)$.

Base sensitivity

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

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

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, . . .

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

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

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

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

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

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

DEFINTION

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

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

We choose the greedy representation: $\mathrm{rep}_U(n) = c_\ell \cdots c_0$.
Usually, we ask that $\mathrm{rep}_U(\mathbb{N}$ is a regular language.

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

FIBONACCI (ZECKENDORF 1972)

$\ldots, 34, 21, 13, 8, 5, 3, 2, 1 = (F_n)_{n\geq 0}$ and $\text{rep}_F(11) = 10100$
but $\text{val}_F(10100) = \text{val}_F(10011) = \text{val}_F(1111)$ $\quad U_{n+2} = U_{n+1} + U_n$.

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

NON-STANDARD NUMERATION SYSTEMS

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

DEFINITION

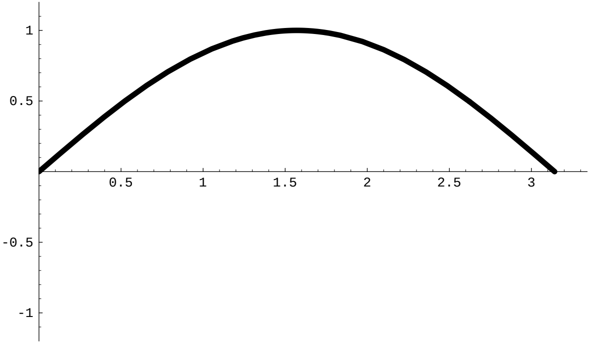
A set $X \subset \mathbb{N}$ is *U-recognizable*, if the set of greedy expansions of the elements of X is accepted by some finite automaton, *i.e.*, $\mathrm{rep}_U(X)$ is a regular language.

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 ?

EXERCISE FOR 1ST YEAR STUDENTS (J.-P. ALLOUCHE'99)

Study the sign over the interval $[0, \pi]$ of the function

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

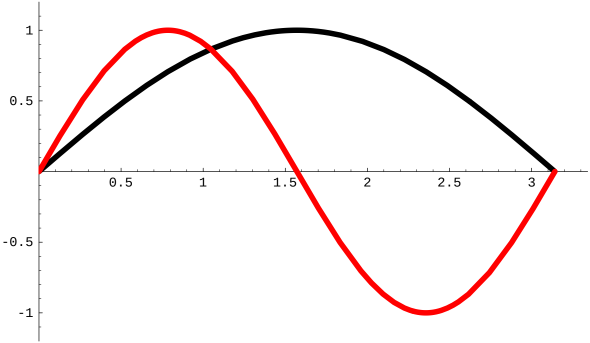


An interlude

Exercise for 1st year students (J.-P. Allouche'99)

Study the sign over the interval $[0, \pi]$ of the function

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

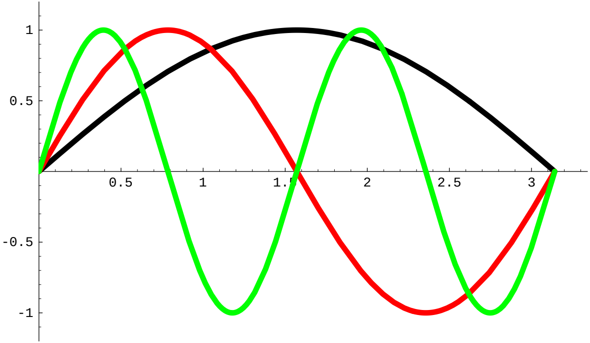


AN INTERLUDE

EXERCISE FOR 1ST YEAR STUDENTS (J.-P. ALLOUCHE'99)

Study the sign over the interval $[0, \pi]$ of the function

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

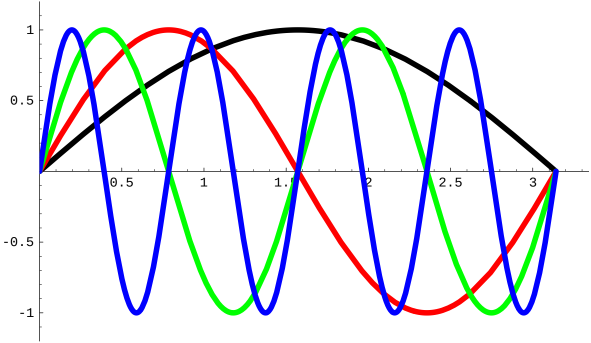


An interlude

Exercise for 1st year students (J.-P. Allouche'99)

Study the sign over the interval $[0, \pi]$ of the function

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



AN INTERLUDE

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

$$\mathbf{t} = + - - + - + + - - + + - + - - + \cdots$$

$$f : \{+, -\} \to \{+, -\}^* : + \mapsto +-, \quad - \mapsto -+$$

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

THEOREM (A. THUE 1912, M. MORSE 1921)

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

M. Lothaire, Combinatorics on words, reprinted in 1997 by Cambridge University Press

Morphic words

Definition (on an example)

Let $A = \{a, b, c\}$ and $B = \{0, 1\}$ be finite alphabets. Consider the morphisms $f : A^* \to A^*$ and $g : A^* \to B^*$ given by

$$f : \begin{cases} a \mapsto abc \\ b \mapsto ca \\ c \mapsto b \end{cases} \quad (\textit{prolongable}) \qquad g : \begin{cases} a \mapsto 0 \\ b \mapsto 1 \\ c \mapsto 0 \end{cases} \quad (\textit{coding})$$

purely morphic

$$f^\omega(a) = \lim_{n\to\infty} f^n(a) = abccabbabccacaabccabbabcbabc\cdots$$

morphic $\quad g(f^\omega(a)) = 0100011010000001000110101010\cdots$

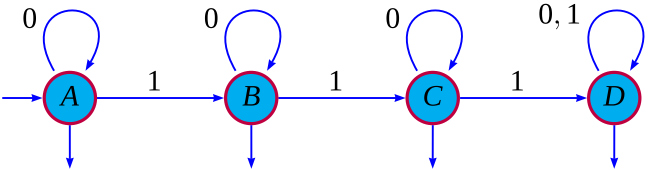
Special case of a uniform morphism f, like Thue–Morse

Theorem (Cobham 1972)

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

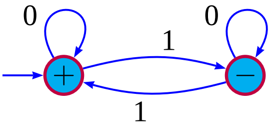
$$f : A \mapsto AB, \quad B \mapsto BC, \quad C \mapsto CD, \quad D \mapsto DD$$

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



$$\begin{cases} + \mapsto +- \\ - \mapsto -+ \end{cases}$$

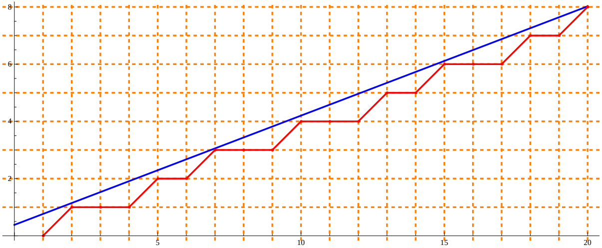
$$+ - - + - + + - - + + - + - - + - + + - + - - + \cdots$$



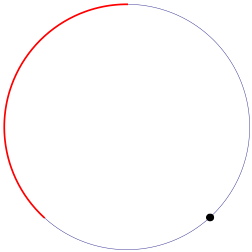
COROLLARY

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

Non-uniform case also of interested, e.g., Fibonacci word:
a morphic sturmian/mechanical word $0100101001001010 01\cdots$

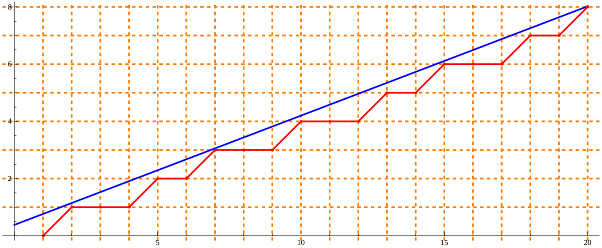


ρ: intercept, α: slope, $T : x \mapsto \{x + \alpha\}$, $(T^n(\rho))_{n\geq 0}$,

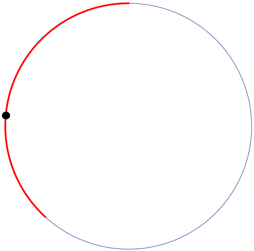


$\alpha = \rho = 1/\phi^2 \simeq .382$

Non-uniform case also of interested, e.g., Fibonacci word:
a morphic sturmian/mechanical word $0100101001001010 01\cdots$

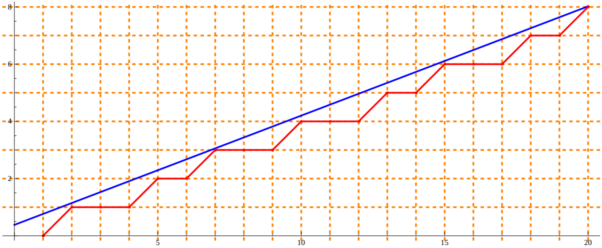


ρ: intercept, α: slope, $T : x \mapsto \{x + \alpha\}$, $(T^n(\rho))_{n \geq 0}$,

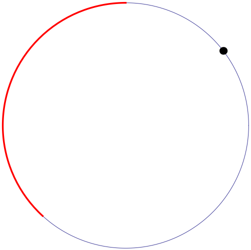


$\alpha = \rho = 1/\phi^2 \simeq .382$

Non-uniform case also of interested, e.g., Fibonacci word:
a morphic sturmian/mechanical word $0100101001001010 01\cdots$

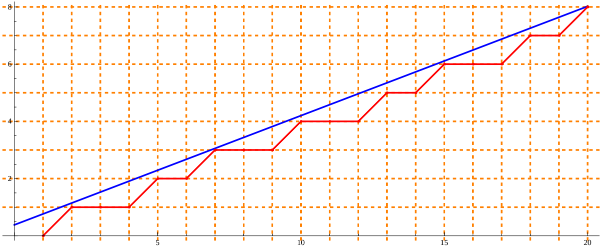


ρ: intercept, α: slope, $T : x \mapsto \{x + \alpha\}$, $(T^n(\rho))_{n\geq 0}$,

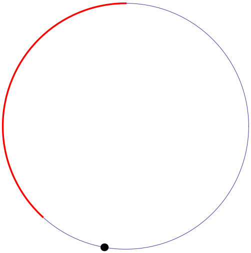


$\alpha = \rho = 1/\phi^2 \simeq .382$

Non-uniform case also of interested, e.g., Fibonacci word:
a morphic sturmian/mechanical word $01001010010101001\cdots$

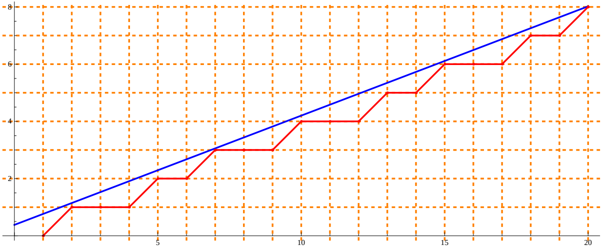


ρ: intercept, α: slope, $T : x \mapsto \{x+\alpha\}$, $(T^n(\rho))_{n\geq 0}$,

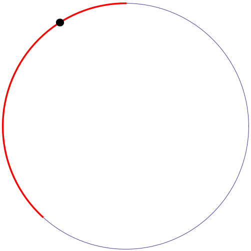


$\alpha = \rho = 1/\phi^2 \simeq .382$

Non-uniform case also of interested, e.g., Fibonacci word:
a morphic sturmian/mechanical word $0100101001001010 01\cdots$



ρ: intercept, α: slope, $T : x \mapsto \{x + \alpha\}$, $(T^n(\rho))_{n\geq 0}$,



$\alpha = \rho = 1/\phi^2 \simeq .382$

In fact, the Fibonacci word is purely morphic

$$f : 0 \mapsto 01, \quad 1 \mapsto 0$$

The transition matrix associated with f is

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

n	$f^n(0)$	$f^n(1)$
1	01	0
2	010	01
3	01001	010
4	01001010	01001

$$M_f^2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \; M_f^3 = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$$

Two formalisms: numeration systems / morphic words

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

An infinite word is morphic IFF it is S-automatic for some *abstract numeration system* S (whatever it is precisely)

We can try to reformulate Cobham's theorem from 1969...

KIND OF A RESULT

Let $\mathbf{x}$ be an infinite word such that

$$g(f^\omega(a)) = \mathbf{x} = s(r^\omega(b))$$

for "sufficiently independent" morphisms f and r.
Then $\mathbf{x}$ is ultimately periodic.

For the sakeof simplicity,
Assume f is a primitive morphism, *i.e.*, the matrix M_f is primitive.

From Perron's theorem,
M_f has a unique dominating real eigenvalue $\alpha_f > 1$.

THEOREM (F. DURAND 1998)

Let $\mathbf{x}$ be an infinite word such that $g(f^\omega(a)) = \mathbf{x} = s(r^\omega(b))$ for primitive morphisms f and r such that α_f and α_r are multiplicative independent real numbers.
Then $\mathbf{x}$ is ultimately periodic.

F. Durand, A Theorem of Cobham for non primitive substitutions, Acta Arithmetica (2002)
F. Durand, M. Rigo, Syndeticity and independent substitutions, Adv. in Applied Math. (2009)

F. Durand, Cobham's theorem for substitutions, J. Eur. Math. Soc. (2011)

Example

For the Fibonacci word, the matrix M_f is primitive having $(1+\sqrt{5})/2$ has Perron value.
Consequently, the Fibonacci word is not k-automatic.