Around Cobham's theorem and some of its extensions

Véronique Bruyère
University of Mons

November 25, 2010

Dynamical Aspects of Automata and Semigroup Theories
Satellite Workshop of Highlights of AutoMathA

Two influential papers

Cobham wrote two very influential papers

- On the Base-Dependence of Sets of Numbers Recognizable by Finite Automata, *Mathematical Systems Theory* **3** (1969) 186-192
- Uniform Tag Sequences, *Mathematical Systems Theory* **6** (1972) 164-192

First Cobham's Theorem

Definition

Given a base $r \geq 2$, a set $X \subseteq \mathbb{N}$ is called *r-recognizable* if X written in base r is accepted by a finite automaton
(All possible leading 0 are considered)

Example

$X = \{2^n \mid n \geq 0\}$ is 2-recognizable and 4-recognizable

- base 2 : 0^*10^*
- base 4 : $0^*(1+2)0^*$
- 3-recognizable ? $0^*(1+2+11+22+121+1012+2101+\cdots)$

Theorem (Cobham 1969)

A set $X \subseteq \mathbb{N}$ is r-recognizable for every base $r \geq 2$ iff X is a finite union of constants and arithmetic progressions.

More precisely ...

Definition

Two bases $r, s \geq 2$ are multiplicatively dependent if $r^k = s^l$ for some $k, l \in \mathbb{N} \setminus \{0\}$.

Example

Bases $2, 4$ are multiplicatively dependent. Bases $2, 3$ are not.

Theorem (Cobham 1969)

Let $r, s \geq 2$ be two multiplicatively independent bases.
A set $X \subseteq \mathbb{N}$ is r- and s-recognizable iff X is a finite union of constants and arithmetic progressions.

Example

$X = \{2^n \mid n \geq 0\}$ is not 3-recognizable.
(It is exactly 2^k-recognizable for every $k \geq 1$).

Second Cobham's Theorem

Theorem (Cobham 1972)

A set $X \subseteq \mathbb{N}$ is r-recognizable iff its characteristic sequence is generated by the iteration of a r-uniform morphism, followed by a coding.

Example

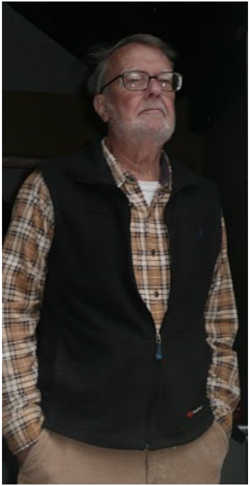
$X = \{2^n \mid n \geq 0\}$

g :	$a \to ab$,	$b \to bc$,	$c \to cc$	2-uniform morphism
f :	$a \to 0$,	$b \to 1$,	$c \to 0$	coding

a
ab
$abbc$
$abbcbccc$
$abbcbcccbccccccc$
$abbcbcccbcccccccbccccccccccccccc$

$01101000100000001000000000000000 \cdots$

Picture



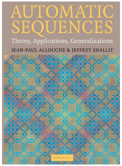
Alan Belmont Cobham
Born November 4, 1927, San Francisco
He lives in Middletown, Connecticut

Picture from Jeffrey O. Shallit 's blog
http ://www.cs.uwaterloo.ca/∼shallit/

Great impact of the two Cobham's theorems

Basis of hundreds of papers exploring the theory of automatic sequences and generalizing them.

- J.-P. Allouche, J. Shallit
  Automatic Sequences :
  Theory, Applications, Generalizations
  Cambridge University Press (2003)



First Cobham's theorem

- Simpler proofs and generalizations to various contexts : multidimensional setting, logical framework, non standard bases, substitutive systems, fractals and tilings, . . .
- B. Adamczewski, J. Bell, A. Bès, B. Boigelot, J. Brusten, V. Bruyère, F. Durand, S. Fabre, I. Fagnot, G. Hansel, C. Michaux, A. Muchnik, D. Perrin, F. Point, M. Rigo, A. Semenov, R. Villemaire, . . .

Great impact of the two Cobham's theorems

First Cobham's theorem - surveys

- D. Perrin, Finite automata, In *Handbook of TCS*, Vol B, Elsevier - MIT Press (1990) 1-57
- V. Bruyère, G. Hansel, C. Michaux and R. Villemaire, Logic and p-recognizable sets of integers, *Bull. Belg. Math. Soc.* **1** (1994) 191-238
- M. Rigo, Numeration systems : a link between number theory and formal language theory, *Proc. DLT'10*, LNCS **6224** Springer (2010) 33-53
- F. Durand, M. Rigo, On Cobham's theorem, In Handbook of Automata Theory (AutoMathA project), in preparation, 39 p

Outline of this talk

First Cobham's theorem

- Some known extensions
  - Logical characterizations
  - Extension to $\mathbb{Z}$
  - Extension to $\mathbb{N}^m$
- Recent extensions to $\mathbb{R}$ and $\mathbb{R}^m$
  - Recognizability
  - Logical characterizations
  - Cobham's theorem extended to $\mathbb{R}$
  - Weak automata
  - Main steps of the proof
  - Cobham's theorem extended to $\mathbb{R}^m$
- Conclusion and other related works

Some known extensions

Logical characterizations

Theorem (Büchi 1960)

X is r-recognizable iff X is first-order definable in $\langle \mathbb{N}, +, V_r \rangle$.

- $V_r(x) = y$ means that y is the largest power of r dividing x.
  $V_r(0) = 1$
- Formulae - (first-order) variables $x, y, z, \ldots$ over $\mathbb{N}$
  equality $=$, addition $+$, function V_r
  connectives $\wedge, \vee, \neg, \rightarrow, \leftrightarrow$
  quantifiers $\exists, \forall$ over variables

Example

- $V_2(20) = 4,\ V_3(20) = 1$
- $\leq$ is first-order definable ; any constant is first-order definable
- $X = 2\mathbb{N} + 1$ is definable by the formula $\varphi(x)$:
  $(\exists y)\ (x = y + y + 1)$
- $X = \{2^n \mid n \geq 0\}$ is definable by the formula $\varphi(x)$: $V_2(x) = x$

Logical characterizations

Theorem

X is a finite union of constants and arithmetic progressions
iff X is ultimately periodic
iff X is first-order definable in Presburger arithmetic $\langle \mathbb{N}, + \rangle$.

- X is ultimately periodic iff
  $(\exists l \geq 0)(\exists p \geq 1)(\forall n \geq l)\ (n \in X \Leftrightarrow n + p \in X)$

Theorem (Cobham's theorem restated)

Let $r, s \geq 2$ be two multiplicatively independent bases.

A set $X \subseteq \mathbb{N}$ is r- and s-recognizable,
(X is first-order definable in $\langle \mathbb{N}, +, V_r \rangle$ and in $\langle \mathbb{N}, +, V_s \rangle$)

iff X is a finite union of constants and arithmetic progressions
(X is first-order definable in $\langle \mathbb{N}, + \rangle$)

Extension to $\mathbb{Z}$

Automata

- In base r, a positive (resp. negative) number always begins with 0 (resp. $r-1$).

Example

In base 2, $-6 = -8 + 2$ is written as 1010 (2's complement), and 10 as 01010

$X = \{2^n \mid n \geq 0\} \cup \{-2^n \mid n \geq 0\}$ is 2-recognizable.
Base 2 : $0^+10^* + 1^+10^*$

Logical structures

- Structures $\langle \mathbb{Z}, +, \leq \rangle$ and $\langle \mathbb{Z}, +, \leq, V_r \rangle$
- X is first-order definable in $\langle \mathbb{Z}, +, \leq \rangle$ iff X is a finite union of constants, arithmetic progressions, and opposite of arithmetic progressions

Extension to $\mathbb{N}^m$

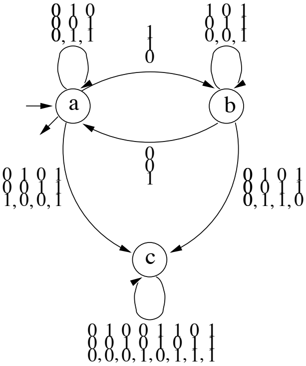
Automata

$\binom{3}{9}$ is written as $\binom{0011}{1001} = \binom{0}{1}\binom{0}{0}\binom{1}{0}\binom{1}{1}$ in base 2.

Example

$X = \{(x, y, z) \mid x + y = z\}$ is 2-recognizable

state a : no carry
state b : carry
state c : error



Extension to $\mathbb{N}^m$

Theorem (Büchi 1960)

Let $m \geq 1$. A set $X \subseteq \mathbb{N}^m$ is r-recognizable iff it is first-order definable in $\langle \mathbb{N}, +, V_r \rangle$.

Theorem (Semenov 1977)

Let $m \geq 1$. Let $r, s \geq 2$ be two multiplicatively independent bases. A set $X \subseteq \mathbb{N}^m$ is r- and s-recognizable iff X is first-order definable in $\langle \mathbb{N}, + \rangle$.

- Elegant proof by (Muchnik 1991), by induction on m

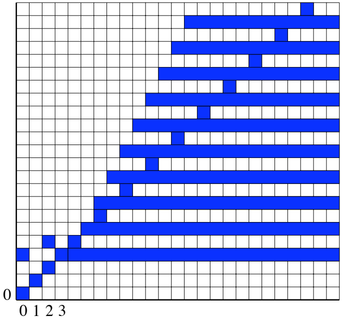
Extension of ultimate periodicity

- definability in $\langle \mathbb{N}, + \rangle$
- finite union of points and cones (semi-linear sets)
- Muchnik's definability criterion

Extension to $\mathbb{N}^m$

$\varphi(x, y)$

$$
\begin{array}{ll}
 & (x = 0 \land y = 3) \\
\lor & (x = 2 \land y = 4) \\
\lor & (x = y) \\
\lor & (\exists z)(\exists t)(x = z + t + 4) \land (y = t + t + 3)
\end{array}
$$



Two points and two cones :

- cone $\{(x, y) \mid (\exists z)(\exists t)\ (x, y) = z(1, 0) + t(1, 2) + (4, 3)\}$
- diagonal

Recent extensions to $\mathbb{R}$ and $\mathbb{R}^m$

Recognizability in $\mathbb{R}^m$

Definition

Given a base r, real numbers are positionally encoded as infinite words over $\{0, 1, \ldots, r-1, \star\}$

- a positive (resp. negative) number begins with 0 (resp. $r-1$)
- all possible encodings ; tuples
- integer numbers : infinite words $u \star 0^\omega$ and $u \star (r-1)^\omega$
- rational numbers : infinite words $u \star vw^\omega$

Example

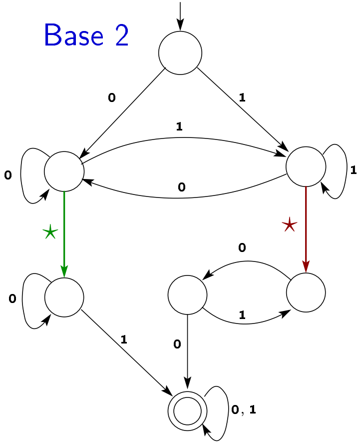
3.5 in base 10 : $0^+3 \star 50^\omega \ \cup \ 0^+3 \star 4\ 9^\omega$.

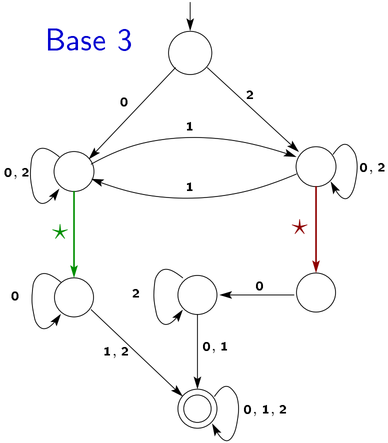
Definition

Let $m \geq 1$. Let $r \geq 2$ be a base.
A set $X \subseteq \mathbb{R}^m$ is r-recognizable if X written in base r is accepted by a finite (non deterministic) Büchi automaton

Example

$$\left\{2n + \left]0, \tfrac{4}{3}\right[\;\middle|\; n \in \mathbb{Z}\right\}$$





Logical characterization

Theorem (Boigelot-Rassart-Wolper 1998)

A set $X \subseteq \mathbb{R}^m$ is r-recognizable iff X is first-order definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z}, V_r \rangle$.

- Variables $x, y, z, \ldots$ over $\mathbb{R}$
- Predicate $\mathbb{Z}(x)$ means that x is an integer variable
- $V_r(x) = y$ means y is the largest power of r dividing x as follows : $x = ky$ with $k \in \mathbb{Z}$ (if such a power exists)

Example

- $V_{10}(3.5) = \frac{1}{10}$, $V_{10}(3.55) = \frac{1}{10^2}$
- $X = \{2^n \mid n \in \mathbb{Z}\}$ is definable by : $V_2(x) = x$
- any rational constant is first-order definable
- $X = \left\{2n +]0, \frac{4}{3}[\;\mid\; n \in \mathbb{Z}\right\}$ is definable by : $(\exists y)(\exists z) \;\; \mathbb{Z}(y) \wedge (x = y + y + z) \wedge (0 < z < \frac{4}{3})$

Ultimately periodically simple sets

Theorem (Weispfenning 1999)

$X \subseteq \mathbb{R}$ is first-order definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z} \rangle$ iff X is ultimately periodically simple.

- Characterization for dimension 1. Higher dimensions : see later.

Definition

X is ultimately periodically simple iff

- X is a finite union of sets of the form $Y_i + Z_i$ where
- each $Y_i \subseteq \mathbb{Z}$ is either an integer constant, either an arithmetic progression, or its opposite
- each $Z_i \subseteq [0,1]$ is an interval with rational endpoints

Example

$X = \left\{ 2n + \left]0, \frac{4}{3}\right[\;\middle|\; n \in \mathbb{Z} \right\}$ is ultimately periodically simple

Cobham's theorem extended to $\mathbb{R}$

Theorem (Boigelot-Brusten-Bruyère 2008)

Let r, s be two bases that do not have the same set of prime factors. A set $X \subseteq \mathbb{R}$ is r- and s-recognizable iff X is ultimately periodically simple

- If r, s do not have the same set of prime factors, then they are multiplicatively independent
- The converse is false (ex. $r = 6$ and $s = 12$)
- This theorem is false for two multiplicatively independent bases (see next slides)

Weak automata

Definition

A deterministic Büchi automaton is weak if each of its strongly connected components has either only accepting or only non accepting states.

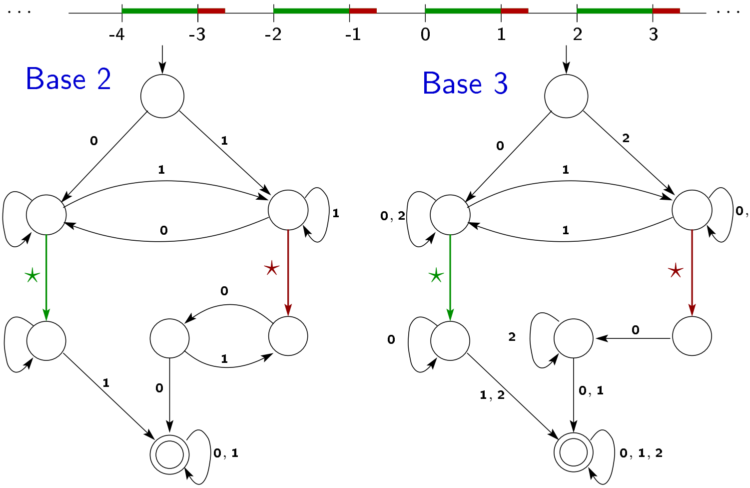
- Practically as easy to handle as finite-word automata
- Canonical minimal form (Löding 2001)

Theorem (Boigelot-Brusten 2007)

Let r, s be two independent bases. A set $X \subseteq \mathbb{R}$ is r- and s-recognizable by weak deterministic Büchi automata iff X is ultimately periodically simple

Weak automata

Example



Weak automata

The expressiveness of weak deterministic Büchi automata is limited

- level $\Sigma_2^0 \cap \Pi_2^0$ in Borel hierarchy
- instead of level $\Sigma_3^0 \cap \Pi_3^0$ for Büchi automata

Theorem (Boigelot-Jodogne-Wolper 2001)

Let $m \geq 1$. Let $r \geq 2$ be a base.
If a set $X \subseteq \mathbb{R}^m$ is definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z} \rangle$, then X written in base r is recognizable by a weak determinstic Büchi automaton.

- Same result with $\langle \mathbb{R}, +, \leq, \mathbb{Z}, P_r \rangle$, where predicate $P_r(x)$ means that x is a power of r (Brusten 2006)
- False for $\langle \mathbb{R}, +, \leq, \mathbb{Z}, V_r \rangle$

Cobham's theorem extended to $\mathbb{R}$

Theorem (Boigelot-Brusten-Bruyère 2008)

Let r, s be two bases that do not have the same set of prime factors. A set $X \subseteq \mathbb{R}$ is r- and s-recognizable iff X is ultimately periodically simple

- In other words, if X is definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z}, V_r \rangle$ and $\langle \mathbb{R}, +, \leq, \mathbb{Z}, V_s \rangle$, then it is definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z} \rangle$
- In other words, if a set X is recognizable by a Büchi automaton independently of the base, then it is recognizable by a weak deterministic Büchi automaton
- Theoretical justification to the use of weak deterministic automata as an effective symbolic representation of sets in the context of computer-aided verification (LASH tool)

Cobham's theorem extended to $\mathbb{R}$

Counterexample

$r = 6$ and $s = 12$

$X = \{x \in \mathbb{R} \mid \text{x can be encoded in base 6 as } u \star v0^\omega\}$

X is both 6- and 12-recognizable

X is not ultimately periodically simple

Intuition

- To have an infinite queue of zeros in base 6 is equivalent to have an infinite queue of zeros in base 12.
- X written in base 6 cannot be accepted by a weak Büchi automaton.

Theorem (Boigelot-Brusten-Bruyère 2009)

For any pair of bases r, s that have the same set of prime factors, the set $X = \{x \in \mathbb{R} \mid x \text{ can be encoded in base } r \text{ as } u \star v0^\omega\}$

- *is both r- and s-recognizable,*
- *but is not ultimately periodically simple.*

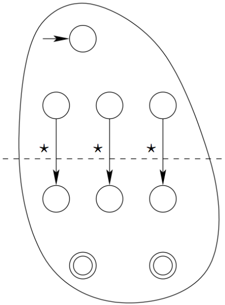
Main steps of the proof

Let r, s that do not have the same set of prime factors
Let $X \subseteq \mathbb{R}$ be a r- and s-recognizable set

1. Separate integer parts and fractional parts

We have $X = \bigcup_{i=1}^{n} Y_i + Z_i$ with

- $Y_i \subseteq \mathbb{Z}$ and $Z_i \subseteq [0, 1]$
- Y_i is r- and s-recognizable (finite-word automata)
- Z_i is r- and s-recognizable (Büchi automata)



By Cobham's theorem, each Y_i is first-order definable in $\langle \mathbb{Z}, +, \leq \rangle$
Thus, it is sufficient to prove each Z_i is a finite union of intervals with rational endpoints

Main steps of the proof

Let $X \subseteq [0, 1]$ be a r- and s-recognizable set.

2. Product stability

Definition
X is f-product-stable if for all x : $x \in X \Leftrightarrow f \cdot x \in X$

- r^j-product stability and s^k-product stability, for some $j, k \geq 1$
- In relation with some cycles in the automata

3. Sum stability

Definition
X is d-sum-stable if for all x : $x \in X \Leftrightarrow x + d \in X$

- Second use of Cobham's Theorem
- Very technical proof. Simpler proof?

Cobham's theorem extended to $\mathbb{R}^m$

Theorem (Boigelot-Brusten-Leroux 2009)

Let $m \geq 1$. Let r, s that do not have the same set of prime factors. A set $X \subseteq \mathbb{R}^m$ is r- and s-recognizable iff X is first-order definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z} \rangle$.

$X \subseteq \mathbb{R}^m$ is first-order definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z} \rangle$ iff

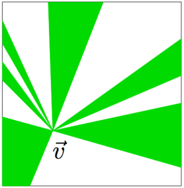
- X is a finite union of sets of the form $Y_i + Z_i$ where
- each $Y_i \subseteq \mathbb{Z}^m$ is first-order definable in $\langle \mathbb{Z}, +, \leq \rangle$
- each $Z_i \subseteq [0, 1]^m$ is first-order definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$, i.e., is a boolean combination of linear constraints with rational coefficients

Boigelot-Brusten-Leroux's definability criterion like in (Muchnik 1991)

Cobham's theorem extended to $\mathbb{R}^m$

Main steps of the proof

- Separation of integer and fractional parts : $X = \bigcup_{i=1}^{n} Y_i + Z_i$
- By Semenov's theorem, each Y_i is first-order definable in $\langle \mathbb{Z}, +, \leq \rangle$
- For each Z_i,
  - r^j-product stability and s^k-product stability, for some $j, k \geq 1$
  - Conical structure of Z_i
  - Each face of $Z_i \cap [0,1]^m$ has dimension $m-1$ and is r-, s-recognizable



- By induction on the dimension, each Z_i is first-order definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$

Conclusion and other related works

Conclusion

In this talk

- Extension of first Cobham's theorem to $\mathbb{Z}$, $\mathbb{N}^m$, $\mathbb{R}$ and $\mathbb{R}^m$
- Logical approach to find the right statements
- Precise description of the structure of automata, when the recognizability is independent of the base

Morphic approach

- Another extension of first's Cobham theorem for sets $X \subseteq \mathbb{N}$
- Orthogonal and beautiful extension
- See next slides

Morphic approach

Theorem (Cobham 1972)

A set $X \subseteq \mathbb{N}$ is r-recognizable iff its characteristic sequence is generated by the iteration of a r-uniform morphism, followed by a coding.

Example

$X = \{2^n \mid n \geq 0\}$

g :	$a \rightarrow ab$,	$b \rightarrow bc$,	$c \rightarrow cc$	2-uniform morphism
f :	$a \rightarrow 0$,	$b \rightarrow 1$,	$c \rightarrow 0$	coding

a
ab
$abbc$
$abbcbccc$
$abbcbcccbccccccc$
$abbcbcccbcccccccbccccccccccccccc$

$0110100010000000100000000000000 \cdots$

Morphic approach

Definition

A set $X \subseteq \mathbb{N}$ is α-recognizable if its characteristic sequence is generated by the iteration of a morphism g, followed by a coding f, such that $\alpha > 1$ is the dominating eigenvalue of the incidence matrix of g.

- Example Fibonacci morphism $g: \quad a \to ab, \quad b \to a$ with incidence matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ and dominating eigenvalue $\frac{1+\sqrt{5}}{2}$

Theorem (Durand 2010)

Let α and β two mutiplicatively independent Perron numbers. A set $X \subseteq \mathbb{N}$ is α- and β-recognizable iff X is a finite union of constants and arithmetic progressions.

- See reference "F. Durand, M. Rigo, On Cobham's theorem, In Handbook of Automata Theory, in preparation, 39 pages"

Thank you . . .