An introduction to numeration systems: Cobham-like theorems, first-order logic and regular sequences

Émilie Charlier

Département de mathématiques, ULiège

Trends in Arithmetic Theories
2022, July 4

From numbers to words

Usually integers are represented by finite words while real numbers are represented by infinite words.

- In base 10: $148 \to 148$, $\frac{1}{3} \to 0.3333\cdots$, $\pi \to 3.141592\cdots$
- In base 2: $148 \to 10010100$, $\frac{1}{3} \to 0.01010101\cdots$, $\pi \to 11.001001000011\cdots$

The basic consideration is as follows: properties of numbers are translated into combinatorial properties of their representations.

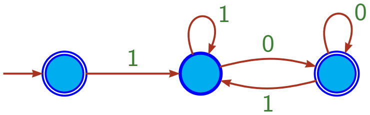
Recognizable sets of integers

A subset X of $\mathbb{N}$ is recognizable with respect to a given numeration system S, or *S-recognizable*, if the language

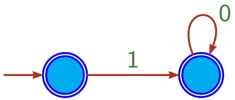
$$\{\mathrm{rep}_S(n) : n \in X\}$$

is accepted by a finite automaton.

- The set $2\mathbb{N}$ of even non-negative integers is 2-recognizable.

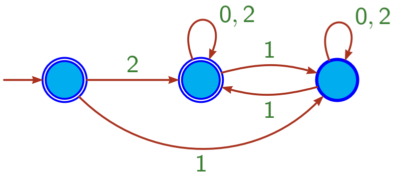


- The set $\{2^n : n \in \mathbb{N}\}$ of powers of 2 is 2-recognizable.



Changing the system

- The set $2\mathbb{N}$ of even non-negative integers is 3-recognizable.



In fact, the set $2\mathbb{N}$ is b-recognizable for all integer bases b.

- The set $\{2^n : n \in \mathbb{N}\}$ of powers of 2 is not 3-recognizable.

  This is a consequence of Cobham's theorem.

Cobham's theorem

Two integers k and ℓ are multiplicatively independent if $k^m = \ell^n$ and $m, n \in \mathbb{N}$ implies $m = n = 0$.

Theorem (Cobham 1969)

Let b and b' be multiplicatively independent integer bases. If a subset of $\mathbb{N}$ is simultaneously b-recognizable and b'-recognizable, then it is a finite union of arithmetic progressions (possibly finite).

$2\mathbb{N} \cup (3\mathbb{N} + 2) \cup \{3\}$

•		•	•	•	•	•		•		•	•	•		•		•	•	•		•		•	•	•		•		•	•	•		•		
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	...

Multidimensional version of Cobham's theorem

Theorem (Semenov 1977)

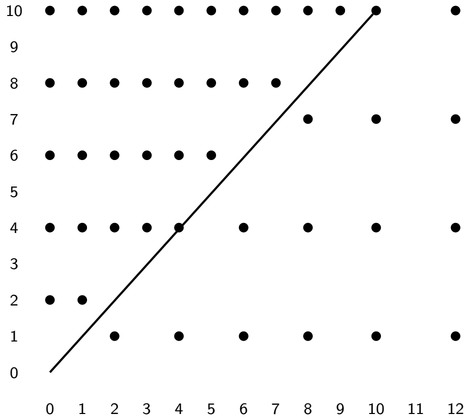
Let b and b' be multiplicatively independent integer bases. If a subset of $\mathbb{N}^d$ is simultaneously b-recognizable and b'-recognizable, then it is semi-linear.

A set $X \subseteq \mathbb{N}^d$ is linear if there exists $v_0, v_1, \cdots, v_t \in \mathbb{N}^d$ such that

$$X = \{v_0 + n_1 v_1 + n_2 v_2 + \cdots + n_t v_t : n_1, \ldots, n_t \in \mathbb{N}\}.$$

A subset of $\mathbb{N}^d$ is semi-linear if it is a finite union of linear sets.

$$\{(2m, 3n+1) : m, n \in \mathbb{N} \text{ and } 2m \geq 3n+1\} \cup \{(m, 2n) : m, n \in \mathbb{N} \text{ and } m < 2n\}$$



From words to numbers

On the other hand, infinite words may also represent sets of numbers: the characteristic sequence of a subset of $\mathbb{N}$ is a binary infinite word.

- The set $2\mathbb{N}$ gives the periodic infinite word $10101010\cdots$
- The set $\{2^n : n \in \mathbb{N}\}$ gives the aperiodic infinite word $01101000100000001000 0\cdots$

Exercise: Show that the characteristic sequence of a subset of $\mathbb{N}$ is ultimately periodic, that is, of the form $uvvv\cdots$, if and only if it is a finite union of arithmetic progressions (possibly finite).

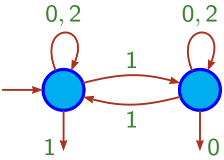
$2\mathbb{N} \cup (3\mathbb{N} + 2) \cup \{3\}$

For this reason, we also talk about ultimately periodic sets of integers.

Linking recognizable sets and automatic sequences

For an integer base $b \geq 2$, a subset X of $\mathbb{N}$ is b-recognizable if and only if its characteristic sequence is *b-automatic*: there exists a DFAO that on input $\mathrm{rep}_b(n)$ ouputs 1 if $n \in X$, and outputs 0 otherwise.

For example, the DFAO



generates the periodic sequence

$$1010101010 \cdots$$

when reading 3-representations of integers, which corresponds to the subset of even non-negative integers

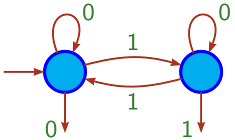
$$\{0, 2, 4, 6, 8, \ldots\}.$$

Automatic sequences

A sequence $f : \mathbb{N} \to B$ is called automatic with respect to a numeration system S, or S-automatic, if there exists a DFA0 $\mathcal{A} = (Q, q_0, \delta, A, \tau, B)$ such that

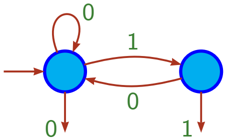
$$\forall n \in \mathbb{N}, \quad f(n) = \tau(\delta(q_0, \mathrm{rep}_S(n)))$$

- The Thue-Morse sequence $01101001100101 \cdots$ is generated by the DFAO



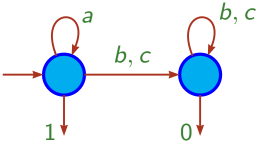
when reading integers in base 2.

- The Fibonacci sequence $0100101001001 \cdots$ is generated by the DFAO



when reading the Zeckendorf representations of the integers.

- The characteristic sequence $110010000100000010\cdots$ of the set of squares $\{0, 1, 4, 9, 16, 25, \dots\}$ is generated by the DFAO



when reading integers in the abstract numeration system $(a^*b^* \cup a^*c^*, a < b < c)$.

However, this sequence isn't b-automatic for any integer base b (Eilenberg 1974).

A range of numeration systems

- Unary representations

  A natural number n is represented by the finite word $\mathrm{rep}_1(n) = a^n$ where a is any fixed symbol.

  Exercise: Show that the 1-recognizable subsets of $\mathbb{N}$ are exactly the ultimately periodic sets.

- Binary representations

$\cdots$	16	8	4	2	1	
$\cdots$	c_4	c_3	c_2	c_1	c_0	n
						0
					1	1
				1	0	2
				1	1	3
			1	0	0	4
			1	0	1	5
			1	1	0	6
			1	1	1	7
		1	0	0	0	8

We have $n = \sum_{i=0}^{\ell-1} c_i 2^i$ with $c_{\ell-1} = 1$, and we write $\mathrm{rep}_2(n) = c_{\ell-1} \cdots c_0$.

- Integer base representations

  Let $b \geq 2$ be an integer. A natural number n is represented by the finite word $\mathrm{rep}_b(n) = c_{\ell-1} \cdots c_0$ obtained from the greedy algorithm:

$$n = \sum_{i=0}^{\ell-1} c_i b^i.$$

  The greedy algorithm only imposes to have a nonzero leading digit $c_{\ell-1}$.

  Thus, the set of all greedy representations is

$$\{1, \ldots, b-1\}\{0, \cdots, b-1\}^* \cup \{\varepsilon\}.$$

- Zeckendorf representations

Let $F = (F_i)_{i\geq 0} = (1, 2, 3, 5, 8, \ldots)$ be the sequence obtained from the rules:

$$F_0 = 1,\ \ F_1 = 2 \text{ and } F_{i+2} = F_{i+1} + F_i \text{ for } i \geq 0.$$

Again, we can use the greedy algorithm in order to get digits $c_{\ell-1} \cdots c_0$ such that $n = \sum_{i=0}^{\ell-1} c_i F_i$:

$\cdots$	8	5	3	2	1	
$\cdots$	c_4	c_3	c_2	c_1	c_0	n
						0
					1	1
				1	0	2
			1	0	0	3
			1	0	1	4
		1	0	0	0	5
		1	0	0	1	6
		1	0	1	0	7
	1	0	0	0	0	8

In addition to having a nonzero leading digit $c_{\ell-1}$, the greedy algorithm imposes that the valid representations do not contain two consecutive 1's.

The set of all greedy representations is

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

- Positional representations

  Let $U = (U_i)_{i\geq 0}$ be a base sequence, that is, an increasing sequence of positive integers such that $U_0 = 1$ and the quotients $\frac{U_{i+1}}{U_i}$ are bounded.

  A natural number n is represented by the finite word $\mathrm{rep}_U(n) = c_{\ell-1}\cdots c_0$ obtained from the greedy algorithm:

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

  A description of the numeration language $\{\mathrm{rep}_U(n) : n \in \mathbb{N}\}$ strongly depends on the base sequence U.

  Given such a system U, other choices of representations could be made, such as the lazy algorithm for instance.

Abstract numeration systems

In all the previous settings, the representations of the integers are ordered thanks to the radix order.

An ANS is a triple $S = (L, A, <)$ where L is an infinite regular language over a totally ordered alphabet $(A, <)$.

The S-representation function $\mathrm{rep}_S \colon \mathbb{N} \to L$ maps n onto the nth word of L in the radix order.

The map rep_S is a bijection and its reciprocal map is the S-value function $\mathrm{val}_S \colon L \to \mathbb{N}$.

Enumerate the words in $a^*b^* \cup a^*c^*$ thanks to the radix order induced by $a < b < c$:

n	$\mathrm{rep}_S(n)$	n	$\mathrm{rep}_S(n)$	n	$\mathrm{rep}_S(n)$
0	ε	9	aaa	18	$aaac$
1	a	10	aab	19	$aabb$
2	b	11	aac	20	$aacc$
3	c	12	abb	21	$abbb$
4	aa	13	acc	22	$accc$
5	ab	14	bbb	23	$bbbb$
6	ac	15	ccc	24	$cccc$
7	bb	16	$aaaa$	25	$aaaaa$
8	cc	17	$aaab$	26	$aaaab$

Enumerate the words in $a^*b^* \cup a^*c^*$ thanks to the radix order induced by $a < b < c$:

n	$\mathrm{rep}_S(n)$	n	$\mathrm{rep}_S(n)$	n	$\mathrm{rep}_S(n)$
0	ε	**9**	**aaa**	18	*aaac*
1	**a**	10	*aab*	19	*aabb*
2	*b*	11	*aac*	20	*aacc*
3	*c*	12	*abb*	21	*abbb*
4	**aa**	13	*acc*	22	*accc*
5	*ab*	14	*bbb*	23	*bbbb*
6	*ac*	15	*ccc*	24	*cccc*
7	*bb*	**16**	**aaaa**	**25**	**aaaaa**
8	*cc*	17	*aaab*	26	*aaaab*

For this ANS, it can be checked that $\mathrm{rep}_S(n^2) = a^n$ for all $n \in \mathbb{N}$.

Enumerate the words in $a^*b^* \cup a^*c^*$ thanks to the radix order induced by $a < b < c$:

n	$\mathrm{rep}_S(n)$	n	$\mathrm{rep}_S(n)$	n	$\mathrm{rep}_S(n)$
0	ε	**9**	**aaa**	18	*aaac*
1	**a**	10	*aab*	19	*aabb*
2	*b*	11	*aac*	20	*aacc*
3	*c*	12	*abb*	21	*abbb*
4	**aa**	13	*acc*	22	*accc*
5	*ab*	14	*bbb*	23	*bbbb*
6	*ac*	15	*ccc*	24	*cccc*
7	*bb*	**16**	**aaaa**	**25**	**aaaaa**
8	*cc*	17	*aaab*	26	*aaaab*

For this ANS, it can be checked that $\mathrm{rep}_S(n^2) = a^n$ for all $n \in \mathbb{N}$.

Theorem (Rigo 2002)

For all $k \in \mathbb{N}$, the set $\{n^k : n \in \mathbb{N}\}$ is S-recognizable for ANS S.

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $\underline{0}1$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0\underline{1}10$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $01\underline{1}010$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $011\underline{0}1001$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110\underline{1}00110$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $011010\underline{0}11001$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{0}1100101$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110\cdots$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110\cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110\cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 0$ from 0:

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110 \cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 0$ from 0:

  $\underline{0}1$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110\cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 0$ from 0:

  $0\underline{1}0$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110\cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 0$ from 0:

  $01\underline{0}01$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110 \cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 0$ from 0:

  $010\underline{0}101$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110\cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 0$ from 0:

  $0100\underline{1}010$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110\cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 0$ from 0:

  $01001\underline{0}1001$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110\cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 0$ from 0:

  $010010\underline{1}0010$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110\cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 0$ from 0:

  $0100101\underline{0}01001\cdots$

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $01101001\underline{1}0010110\cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 0$ from 0:

  $0100101\underline{0}01001\cdots$

  The so-obtained limit infinite word is the called Fibonacci sequence.

Morphic sequences

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 10$ iteratively from 0:

  $0110100\underline{1}10010110\cdots$

  The so-obtained limit infinite word is the called Thue-Morse sequence.

- Apply the rules $0 \mapsto 01$ and $1 \mapsto 0$ from 0:

  $0100101\underline{0}01001\cdots$

  The so-obtained limit infinite word is the called Fibonacci sequence.

A morphism $\sigma\colon A^* \to A^*$ is said to be prolongable on a letter $a \in A$ if $\sigma(a) = au$ for some nonempty word u such that $\sigma^n(u)$ is nonempty for all $n \geq 0$.

In this case, when iterating σ on a, we get longer and longer words and for each $n \in \mathbb{N}$, the word $\sigma^n(a)$ is a prefix of $\sigma^{n+1}(a)$.

An infinite sequence obtained as the limit $a\sigma(u)\sigma^2(u)\sigma^3(u)\cdots$ of such a process is said to be pure morphic or the fixed point of the morphism σ.

A morphic sequence is the image under a letter-to-letter morphism of a pure morphic sequence.

Automatic versus morphic

Theorem (Cobham 1972)

Let b be an integer base. A sequence if b-automatic if and only if it is the image under a letter-to-letter morphism of a fixed point of a b-uniform morphism.

- The Thue-Morse sequence is 2-automatic and it is also the fixed point of the 2-uniform morphim $0 \mapsto 01$, $1 \mapsto 10$.

Theorem (Maes & Rigo 2002)

A sequence if S-automatic for some abstract numeration system S if and only if it is morphic.

- The set of primes is never S-recognizable, since its characteristic sequence is not morphic (Mauduit 1988).
- The Fibonacci sequence is Zeckendorf-automatic and it is the fixed point of the non-uniform morphim $0 \mapsto 01$, $1 \mapsto 0$.

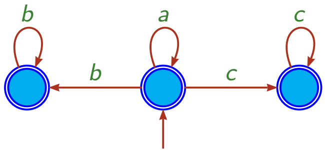
- For $S = (a^*b^* \cup a^*c^*, a < b < c)$, the set of squares is S-recognizable since $\{\mathrm{rep}_S(n^2) : n \in \mathbb{N}\} = a^*$. Hence its characteristic sequence $110010000100000010 \cdots$ is S-automatic.

- For $S = (a^*b^* \cup a^*c^*, a < b < c)$, the set of squares is S-recognizable since $\{\mathrm{rep}_S(n^2) : n \in \mathbb{N}\} = a^*$. Hence its characteristic sequence $110010000100000010\cdots$ is S-automatic.

  Let us see why it is also morphic.

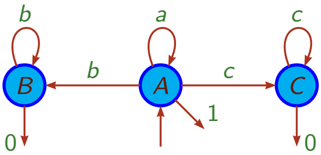
- For $S = (a^*b^* \cup a^*c^*, a < b < c)$, the set of squares is S-recognizable since $\{\mathrm{rep}_S(n^2) : n \in \mathbb{N}\} = a^*$. Hence its characteristic sequence $110010000100000010\cdots$ is S-automatic.

  Let us see why it is also morphic.



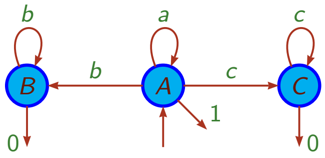
- For $S = (a^*b^* \cup a^*c^*, a < b < c)$, the set of squares is S-recognizable since $\{\mathrm{rep}_S(n^2) : n \in \mathbb{N}\} = a^*$. Hence its characteristic sequence $110010000100000010 \cdots$ is S-automatic.

  Let us see why it is also morphic.



- For $S = (a^*b^* \cup a^*c^*, a < b < c)$, the set of squares is S-recognizable since $\{\mathrm{rep}_S(n^2) : n \in \mathbb{N}\} = a^*$. Hence its characteristic sequence $110010000100000010\cdots$ is S-automatic.

  Let us see why it is also morphic.

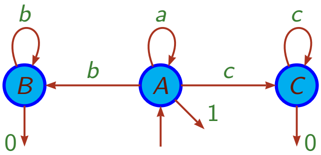


  We compute the (non uniform) morphism

$$\alpha \mapsto \alpha A, \quad A \mapsto ABC, \quad B \mapsto B, \quad C \mapsto C$$

- For $S = (a^*b^* \cup a^*c^*, a < b < c)$, the set of squares is S-recognizable since $\{\mathrm{rep}_S(n^2) : n \in \mathbb{N}\} = a^*$. Hence its characteristic sequence $110010000100000010 \cdots$ is S-automatic.

  Let us see why it is also morphic.



  We compute the (non uniform) morphism

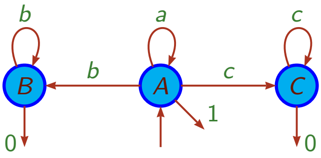
$$\alpha \mapsto \alpha A, \quad A \mapsto ABC, \quad B \mapsto B, \quad C \mapsto C$$

  Iterating this morphism from α, we get the sequence

$$\alpha AABCABCBCABCBCBCABC \cdots$$

- For $S = (a^*b^* \cup a^*c^*, a < b < c)$, the set of squares is S-recognizable since $\{\mathrm{rep}_S(n^2) : n \in \mathbb{N}\} = a^*$. Hence its characteristic sequence $110010000100000010 \cdots$ is S-automatic.

  Let us see why it is also morphic.



  We compute the (non uniform) morphism

$$\alpha \mapsto \alpha A, \quad A \mapsto ABC, \quad B \mapsto B, \quad C \mapsto C$$

  Iterating this morphism from α, we get the sequence

$$\alpha AABCABCBCABCBCBCABC \cdots$$

  Finally, applying the morphism

$$\alpha \mapsto \varepsilon, \; A \mapsto 1, \; B \mapsto 0, \; C \mapsto 0$$

  we obtain the desired sequence

$$1100100001000000100 \cdots$$

Alternative definitions of b-recognizable sets

There exist several equivalent definitions of b-recognizable sets of integers using

- automata
- uniform morphisms
- logic
- finiteness of the b-kernel
- formal series

There are also multidimensional versions of the previous definitions.

See the survey of Bruyère, Hansel, Michaux & Villemaire 1996.

Definable sets

Let $\mathcal{S}$ be a logical structure whose domain is H and let $d \geq 1$. A set $X \subseteq H^d$ is definable in $\mathcal{S}$ if there exists a first-order formula $\varphi(x_1, \ldots, x_d)$ of $\mathcal{S}$ such that

$$X = \{(h_1, \ldots, h_d) \in H^d : \varphi(h_1, \ldots, h_d) \text{ is true}\}.$$

Let $V_b \colon \mathbb{N} \to \mathbb{N}$ be the function mapping $n \geq 1$ to the largest power of b dividing n, and mapping 0 to 1.

- The set $\{b^n : n \in \mathbb{N}\}$ is definable in $\langle \mathbb{N}, +, V_b \rangle$ by the formula $V_b(x) = x$.

Theorem (Büchi 1960, Bruyère 1985)

Let b be an integer base. A subset X of $\mathbb{N}^d$ is b-recognizable if and only if it is definable in $\langle \mathbb{N}, +, V_b \rangle$. Moreover, both directions are effective.

We may now reformulate the Cobham-Semenov theorem in logical terms:

Theorem (Cobham-Semenov)

Let b and b' be multiplicatively independent integer bases. If a subset of $\mathbb{N}^d$ is simultinaeously definable in $\langle \mathbb{N}, +, V_b \rangle$ and in $\langle \mathbb{N}, +, V_{b'} \rangle$, then it is definable in $\langle \mathbb{N}, + \rangle$.

Sets that are S-recognizable for all S

As linear sets are b-recognizable for all $b \geq 2$, we obtain:

Corollary

A subset of $\mathbb{N}^d$ is b-recognizable for all $b \geq 2$ if and only if it is semi-linear.

Sets that are S-recognizable for all S

As linear sets are b-recognizable for all $b \geq 2$, we obtain:

Corollary

A subset of $\mathbb{N}^d$ is b-recognizable for all $b \geq 2$ if and only if it is semi-linear.

However, semi-linear sets are not always S-recognizable!

Sets that are S-recognizable for all S

As linear sets are b-recognizable for all $b \geq 2$, we obtain:

Corollary

A subset of $\mathbb{N}^d$ is b-recognizable for all $b \geq 2$ if and only if it is semi-linear.

However, semi-linear sets are not always S-recognizable!

For example, the linear set $X = \{(n, 2n) : n \in \mathbb{N}\} = (1, 2)\mathbb{N}$ is not 1-recognizable since the language

$$\mathrm{rep}_1(X) = \{(\#^n a^n, a^{2n}) : n \in \mathbb{N}\} = \{(\#, a)^n (a, a)^n : n \in \mathbb{N}\}$$

is not regular (apply the pumping lemma).

Sets that are S-recognizable for all S

As linear sets are b-recognizable for all $b \geq 2$, we obtain:

Corollary

A subset of $\mathbb{N}^d$ is b-recognizable for all $b \geq 2$ if and only if it is semi-linear.

However, semi-linear sets are not always S-recognizable!

For example, the linear set $X = \{(n, 2n) : n \in \mathbb{N}\} = (1, 2)\mathbb{N}$ is not 1-recognizable since the language

$$\mathrm{rep}_1(X) = \{(\#^n a^n, a^{2n}) : n \in \mathbb{N}\} = \{(\#, a)^n (a, a)^n : n \in \mathbb{N}\}$$

is not regular (apply the pumping lemma).

Theorem (Charlier, Lacroix & Rampersad 2010)

A subset X of $\mathbb{N}^d$ is S-recognizable for all S if and only if it is 1-recognizable.

Applications to decidability questions for automatic sequences

By using the following corollary of the Büchi-Bruyère theorem, we can automatically prove many properties of automatic sequences.

Corollary

The first order theory of $\langle \mathbb{N}, +, V_b \rangle$ *is decidable.*

For example, the fact that the Thue-Morse sequence $T \colon \mathbb{N} \to \{0, 1\}$ is aperiodic translates as

$$\exists N,\ \exists p \geq 1,\ \forall n \geq N,\ T(n) = T(n + p).$$

Since T is a 2-automatic sequence, the previous formula is a closed first-order formula of $\langle \mathbb{N}, +, V_2 \rangle$.

This means that we can decide whether this formula is true or is false.

In practice

This method for deciding first-order expressible properties of b-automatic sequences has worst case complexity

$$2^{2^{\cdot^{\cdot^{\cdot^{2^n}}}}}$$

where n is the number of states of the given DFAO and the height of the tower is the number of alternating quantifiers in the first-order formula.

Nevertheless, this procedure was implemented by Mousavi, and Shallit and his coauthors were able to run their programs in order to prove (and reprove) many results about b-automatic sequences, in a purely mechanical way.

From automatic sequences to regular sequences

In what follows, $\mathbb{K}$ designates an arbitrary commutative semiring and $S = (L, A, <)$ is an arbitrary ANS.

A sequence $f : \mathbb{N} \to \mathbb{K}$ is called $(S, \mathbb{K})$-regular if there exist a morphism of monoids $\mu \colon A^* \to \mathbb{K}^{r \times r}$, and vectors $\lambda \in \mathbb{K}^{1 \times r}$ and $\gamma \in \mathbb{K}^{r \times 1}$ such that

$$\forall w \in L, \quad \lambda \mu(w) \gamma = f(\mathrm{val}_S(w)).$$

In this case, the triple (λ, μ, γ) is called a linear representation of the sequence f.

Theorem (Charlier, Cisternino & Stipulanti 2020)

Let $f : \mathbb{N} \to \mathbb{K}$.

- *If f is S-automatic then it is $(S, \mathbb{K})$-regular.*
- *If f is $(S, \mathbb{K})$-regular and takes only finitely many values, and if moreover $\mathbb{K}$ is finite or is a ring, then f is S-automatic.*

We also have a multidimensional version of this result.

Enumerating recognizable properties of automatic sequences gives rise to regular sequences

In this part, we focus on the semirings $\mathbb{N}$ and $\mathbb{N}_\infty = \mathbb{N} \cup \{\infty\}$.

Theorem (Charlier, Cisternino & Stipulanti 2020)

If X is an S-recognizable subset of $\mathbb{N}^{d+d'}$, then the sequence

$$f : \mathbb{N}^d \to \mathbb{N}_\infty,\ \boldsymbol{n} \mapsto \mathrm{Card}\{\boldsymbol{n}' \in \mathbb{N}^{d'} : \left(\begin{smallmatrix} \boldsymbol{n} \\ \boldsymbol{n}' \end{smallmatrix}\right) \in X\}$$

is $(S, \mathbb{N}_\infty)$-regular. If moreover $f(\mathbb{N}) \subseteq \mathbb{N}$ then f is $(S, \mathbb{N})$-regular.

S-recognizable predicates

A predicate P on $\mathbb{N}^d$ is S-recognizable if the set

$$\{(n_1, \dots, n_d) \in \mathbb{N}^d : P(n_1, \dots, n_d) \text{ is true}\}$$

is S-recognizable.

- The binary predicates $x = y$ and $x < y$ are always S-recognizable since the languages
  - $\mathrm{rep}_S\{(n, n) : n \in \mathbb{N}\} = \{(w, w) : w \in L\}$
  - $\mathrm{rep}_S\{(m, n) \in \mathbb{N}^2 : m < n\} = \{(u, v)^{\#} : u, v \in L,\ u <_{\mathrm{rad}} v\}$

  are both regular.
- Addition is not always S-recognizable since the subset

$$\Big\{(m, n, m + n) : m, n \in \mathbb{N}\Big\}$$

  of $\mathbb{N}^3$ is not S-recognizable in general.

  The most famous family of ANS for which addition is recognizable is that of Pisot numeration systems (Frougny & Solomyak 1996).

The following result generalizes ideas from Bruyère, Hansel, Michaux & Villemaire 1996 and Charlier, Rampersad & Shallit 2012 to ANS.

Proposition

Any predicate on $\mathbb{N}^d$ that is defined recursively from S-recognizable predicates by only using the logical connectives $\wedge, \vee, \neg, \implies, \iff$ and the quantifiers $\forall$ and $\exists$ on variables describing elements of $\mathbb{N}$, is S-recognizable.

Corollary

If P a such a predicate on $\mathbb{N}$ then the closed predicates $\forall x P(x)$, $\exists x P(x)$ and $\exists^\infty x P(x)$ are decidable.

Application to factor complexity

The factor complexity of $f\colon \mathbb{N} \to B$ is the function $\rho_f\colon \mathbb{N} \mapsto \mathbb{N}$ that maps each $s \in \mathbb{N}$ to the number of factors of size s occurring in f.

Corollary

Let S be an ANS such that addition is S-recognizable, i.e., the predicate $x + y = z$ is S-recognizable. Then the factor complexity of an S-automatic sequence is an $(S, \mathbb{N})$-regular sequence.

Proof.

Let f be an S-automatic sequence.

For all $s \in \mathbb{N}$, one has

$$\rho_f(s) = \mathrm{Card}\{p \in \mathbb{N} : \forall p' \in \mathbb{N} \left(p' < p \implies \exists i < s,\ f(p' + i) \neq f(p + i)\right)\}.$$

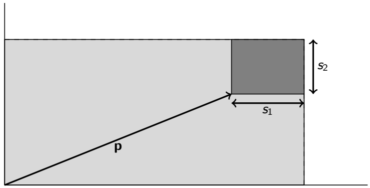
It now suffices to see that the set

$$\{(s, p) \in \mathbb{N}^2 : \forall p' \in \mathbb{N} \left(p' < p \implies \exists i < s,\ f(p' + i) \neq f(p + i)\right)\}$$

is S-recognizable. $\square$

Factor complexity of multidimensional sequence

The factor complexity of $f : \mathbb{N}^d \to B$ is the function $\rho_f : \mathbb{N}^d \mapsto \mathbb{N}$ that maps each $s \in \mathbb{N}^d$ to the number of rectangular d-dimensional factors of size s occurring in f.



Corollary

Let S be an ANS such that addition is S-recognizable, i.e., the predicate $x + y = z$ is S-recognizable. Then the factor complexity of a multidimensional S-automatic sequence is an $(S, \mathbb{N})$-regular sequence.

Representing real numbers

In general real numbers are represented by infinite words.

In this context, we consider Büchi automata. An infinite word is accepted when the corresponding path goes infinitely many times through an accepting state.

We talk about ω-languages and ω-regular languages.

Regular and ω-regular languages share some important properties: they both are stable under

- complementation
- finite union
- finite intersection
- morphic image
- inverse image under a morphism.

Nevertheless, they differ by some other aspects. One of them is determinism.

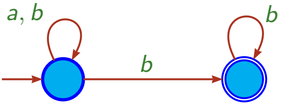
Deterministic Büchi automata

As for DFAs, we can define deterministic Büchi automata.

But one has to be careful as the family of ω-languages that are accepted by deterministic Büchi automata is strictly included in that of ω-regular languages.

Example

No deterministic Büchi automaton accepts the language accepted by



β-representation of real numbers

Let $\beta > 1$ be a real number and let $C \subset \mathbb{Z}$ be an alphabet. For a real number x, any infinite word $u = u_k \cdots u_1 u_0 \star u_{-1} u_{-2} \cdots$ over $C \cup \{\star\}$ such that

$$\sum_{-\infty < i \leq k} u_i \, \beta^i = x$$

is a β-representation of x.

In general, this is not unique.

- Consider $\beta = \frac{1+\sqrt{5}}{2}$ (golden ratio) and $x = \sum_{i \geq 1} \beta^{-2i}$.

  As we also have $x = \sum_{i \geq 3} \beta^{-i}$, the words

  $$u = 0 \star 001111 \cdots$$

  and

  $$v = 0 \star 0101010 \cdots$$

  are both β-representations of x.

β-expansions of real numbers

For $x \geq 0$, among all such β-representations of x, we distinguish the β-expansion

$$d_\beta(x) = x_k \cdots x_1 x_0 \star x_{-1} x_{-2} \cdots$$

which is the infinite word over $A_\beta = \{0, \ldots, \lceil \beta \rceil - 1\}$ containing exactly one symbol $\star$ and obtained by the greedy algorithm.

Then we may also define the β-expansions of negative real numbers as well as of real vectors.

A set $X \subseteq \mathbb{R}^d$ is β-recognizable if the set $d_\beta(X)$ is accepted by a Büchi automaton.

First order theory for mixed real and integer variables

Let X_β be the finite collection of binary predicates $\{X_{\beta,a} : a \in \tilde{A}_\beta\}$ defined by $X_{\beta,a}(x,y)$ is true whenever $y = \beta^i$ for some $i \in \mathbb{Z}$, and

- either $|x| < y$ and $a = 0$,
- or $|x| \geq y$, $i \leq k$ and $x_i = a$.

Theorem (Boigelot-Rassart-Wolper 1998)

Let b be an integer base. A subset of $\mathbb{R}^d$ is b-recognizable if and only if it is definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z}, X_b \rangle$.

As the emptiness of an ω-regular language is decidable, we obtain

Corollary

The first order theory of $\langle \mathbb{R}, +, \leq, \mathbb{Z}, X_b \rangle$ is decidable.

Deciding topological properties

The following properties of b-recognizable subsets X of $\mathbb{R}^d$ are decidable:

- X has a nonempty interior:

$$(\exists x \in X)\ (\exists \varepsilon > 0)\ (\forall y)\ (|x - y| < \varepsilon \implies y \in X).$$

- X is open:

$$(\forall x \in X)\ (\exists \varepsilon > 0)\ (\forall y)\ (|x - y| < \varepsilon \implies y \in X).$$

- X is closed: OK as $\mathbb{R}^d \setminus X$ is b-recognizable.
- ...

A Cobham theorem for real numbers

Theorem (Boigelot-Brusten-Bruyère-Jodogne-Leroux 2001, 2008, 2009)

Let b and b' be multiplicatively independent integer bases.
A subset $X \subseteq \mathbb{R}^d$ is simultaneously weakly b-recognizable and b'-recognizable if and only if it is definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z} \rangle$.

For $d = 1$, this result is equivalent to

Theorem (Adamczewski-Bell 2011)

Let $b, b' \geq 2$ be multiplicatively independent integers. A compact set $X \subseteq [0, 1]$ is simultaneously b-self-similar and b'-self-similar if and only if it is a finite union of closed intervals with rational endpoints.

b-self-similarity

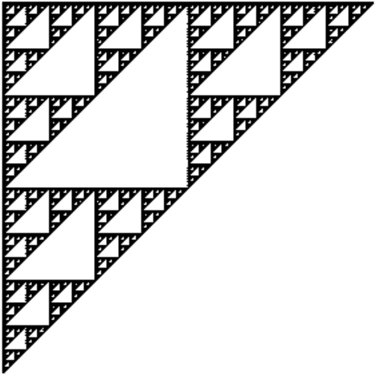
Let $b \geq 2$ be an integer.

A compact set $X \subset [0,1]^d$ is b-self-similar if its b-kernel

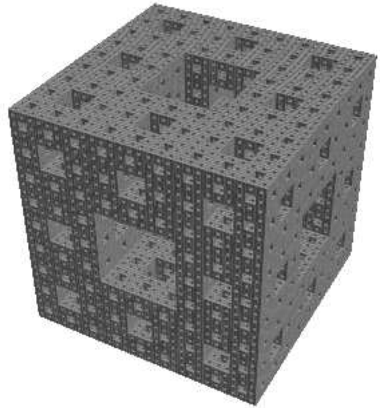
$$\left\{ (b^k X - \mathbf{a}) \cap [0,1]^d \ : \ k \geq 0,\ \mathbf{a} = (a_1, \ldots, a_d) \in \mathbb{Z}^d,\ (\forall i)\, 0 \leq a_i < b^k \right\}$$

is finite.

Pascal's triangle modulo 2 is 2-self-similar.



Menger sponge is 3-self-similar.



Sets definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z} \rangle$

A rational polyhedron is a region of $\mathbb{R}^d$ delimited by a finite number of hyperplanes whose equations have integer coefficients.

Any finite union of rational polyhedra is *b*-self-similar.

A bounded subset $X \subseteq \mathbb{R}^d$ definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z} \rangle$ is a finite union of rational polyhedra.

In particular, for $d = 1$, a subset $X \subseteq [0, 1]$ is definable in $\langle \mathbb{R}, +, \leq, \mathbb{Z} \rangle$ if and only if it is a finite union of closed intervals with rational endpoints.

Linking b-self-similarity and b-recognizability

Theorem (Charlier-Leroy-Rigo 2015)

A subset of $[0,1]^d$ is b-self-similar if and only if it is weakly b-recognizable.

Corollary (simultaneously obtained by Chan-Hare 2014)

Let $b, b' \geq 2$ be two multiplicatively independent integers.
A compact set $X \subset [0,1]^d$ is simultaneously b-self-similar and b'-self-similar if and only if it is a finite union of rational polyhedra.

In fact, we proved the above link in the more general case of a real Pisot base β.

Characterizing β-recognizable sets using logic

Theorem (Charlier-Leroy-Rigo 2015)

- *If β is Parry then every β-recognizable $X \subseteq \mathbb{R}^d$ is β-definable.*
- *If β is Pisot then every β-definable $X \subseteq \mathbb{R}^d$ is β-recognizable.*

As a consequence of this and the fact that emptiness of an ω-language is decidable, we obtain

Corollary

If β is a Pisot number, then the first order theory of $\langle \mathbb{R}, +, \leq, \mathbb{Z}_\beta, X_\beta \rangle$ is decidable.

References

- Berstel and Reutenauer, Noncommutative rational series with applications, *Encyclopedia of Mathematics and its Applications* **137**, Cambridge Univ. Press, 2011.
- Bruyère, Hansel, Michaux and Villemaire, Logic and p-Recognizable Sets of Integers, Bull. Belg. Math. Soc. Simon Stevin **1**, 1994, 191–238.
- Boigelot, Brusten and Leroux, A generalization of Semenov's theorem to automata over real numbers, Automated deduction—CADE-22, LNCS **5663**, 469–484.
- Boigelot, Jodogne and Wolper, On the use of weak automata for deciding linear arithmetic with integer and real variables, Automated reasoning, LNCS **2083**, 611–625.
- Charlier, First-order logic and numeration systems, Sequences, groups, and number theory, Trends Math., 89–141, Birkhäuser/Springer, Cham, 2018.
- Charlier, Cisternino and Stipulanti, Regular sequences and synchronized sequences in abstract numeration systems, European J. Combin. **101**, 2022, Paper No. 103475.
- Charlier, Lacroix and Rampersad, Multi-dimensional sets recognizable in all abstract numeration systems, RAIRO Theor. Inform. Appl. **46**, 2012, 51–65.
- Charlier, Leroy and Rigo, An analogue of Cobham's theorem for graph directed iterated function systems, Adv. Math. **280**, 2015, 86–120.
- Charlier, Rampersad and Shallit, Enumeration and decidable properties of automatic sequences, Internat. J. Found. Comput. Sci. **23**, 2012, 1035–1066.

- Cobham, Uniform tag sequences, Math. Systems Theory **6**, 1972, 164–192.
- Cobham, On the base-dependence of sets of numbers recognizable by finite automata, Math. Systems Theory **3**, 1969, 186–192.
- Frougny and Solomyak, On representation of integers in linear numeration systems, Ergodic theory of $\mathbb{Z}^d$ actions, London Math. Soc. Lecture Note Ser. **228**, 345–368, Cambridge Univ. Press, Cambridge, 1996.
- Lecomte and Rigo, Numeration systems on a regular language, TCS **34**, 2001, 27–44.
- Mauduit, Sur l'ensemble normal des substitutions de longueur quelconque, J. Number Theory **29**,1988, 235–250.
- Mousavi, Walnut software, available at `https://github.com/hamousavi/Walnut`.
- Rigo and Maes, More on generalized automatic sequences, J. Autom. Lang. Comb. **7**, 2002, 351–376.
- Shallit, A logical approach to automatic sequences (Exploring combinatorics on word with Walnut), London Mathematical Society Lecture Note Series, Series Number 482, 2022.