Abstract Numeration Systems or Decimation of Languages

Emilie Charlier

University of Waterloo, School of Computer Science

Languages and Automata Theory Seminar
Waterloo, April 1 and 8, 2011

Positional numeration systems

A positional numeration system (PNS) is given by a sequence of integers $U = (U_i)_{i\geq 0}$ such that

- $U_0 = 1$
- $\forall i \; U_i < U_{i+1}$
- $(U_{i+1}/U_i)_{i\geq 0}$ is bounded $\quad\quad \rightarrow C_U = \sup_{i\geq 0}\lceil U_{i+1}/U_i \rceil$

The greedy U-representation of a positive integer n is the unique word $\mathrm{rep}_U(n) = c_{\ell-1}\cdots c_0$ over $\Sigma_U = \{0,\ldots,C_U - 1\}$ satisfying

$$n = \sum_{i=0}^{\ell-1} c_i\, U_i,\ c_{\ell-1} \neq 0 \ \text{ and } \ \forall t\ \sum_{i=0}^{t} c_i U_i < U_{t+1}.$$

A set $X \subseteq \mathbb{N}$ is U-recognizable or U-automatic if $\mathrm{rep}_U(X) \subseteq \Sigma_U^*$ is regular.

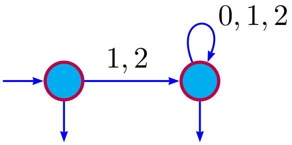
Integer base $b \geq 2$

$U = (b^i)_{i \geq 0}$

$\mathrm{rep}_U,\ \Sigma_U \ \rightarrow\ \mathrm{rep}_b,\ \Sigma_b$

$\Sigma_b = \{0, \cdots, b-1\}$

$\mathcal{L}_b = \mathrm{rep}_b(\mathbb{N}) = \Sigma_b^* \setminus 0\,\Sigma_b^*$



$\mathrm{rep}_3(\mathbb{N})$

$\mathbb{N}$ is 3-recognizable

27	9	3	1	
			ε	0
			1	1
			2	2
		1	0	3
		1	1	4
		1	2	5
		2	0	6
		2	1	7
		2	2	8
	1	0	0	9

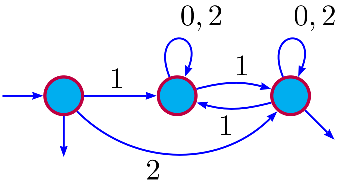
Integer base $b \geq 2$

$U = (b^i)_{i \geq 0}$

$\mathrm{rep}_U,\ \Sigma_U \ \rightarrow\ \mathrm{rep}_b,\ \Sigma_b$

$\Sigma_b = \{0, \cdots, b-1\}$

$\mathcal{L}_b = \mathrm{rep}_b(\mathbb{N}) = \Sigma_b^* \setminus 0\,\Sigma_b^*$



$\mathrm{rep}_{U_3}(2\,\mathbb{N})$

$2\,\mathbb{N}$ is 3-recognizable

27	9	3	1	
			ε	0
			1	1
			2	2
		1	0	3
		1	1	4
		1	2	5
		2	0	6
		2	1	7
		2	2	8
	1	0	0	9

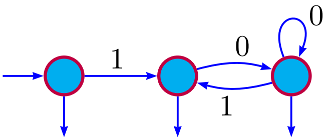
Fibonacci (or Zeckendorf) numeration system

Let $F = (F_i)_{i \geq 0} = (1, 2, 3, 5, 8, 13, 21, \ldots)$ be defined by

$$F_0 = 1,\ F_1 = 2 \text{ and } \forall i \in \mathbb{N},\ F_{i+2} = F_{i+1} + F_i.$$

$\Sigma_F = \{0, 1\}$

The factor 11 is forbidden :



$$\mathrm{rep}_F(\mathbb{N}) = 1\{0, 01\}^* \cup \{\varepsilon\}$$

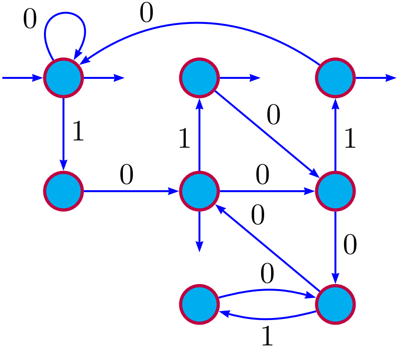
$\mathbb{N}$ is F-recognizable

13	8	5	3	2	1	
					ε	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

Fibonacci (or Zeckendorf) numeration system

Let $F = (F_i)_{i\geq 0} = (1, 2, 3, 5, 8, 13, 21, \dots)$ be defined by

$$F_0 = 1,\ F_1 = 2 \ \text{ and } \ \forall i \in \mathbb{N},\ F_{i+2} = F_{i+1} + F_i.$$



$\mathrm{rep}_F(2\,\mathbb{N})$

13	8	5	3	2	1	
					ε	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

$2\,\mathbb{N}$ is F-recognizable

U-recognizability of $\mathbb{N}$

Is the set $\mathbb{N}$ U-recognizable? Otherwise stated, is the numeration language regular? Not necessarily:

Theorem (Shallit 1994)

Let U be a PNS. If $\mathbb{N}$ is U-recognizable, then U is linear, i.e., it satisfies a linear recurrence relation over $\mathbb{Z}$.

Loraud (1995) and Hollander (1998) gave sufficient conditions for the numeration language to be regular : "The characteristic polynomial of the recurrence relation has a particular form".

Abstract numeration systems

An abstract numeration system (ANS) is a triple $S = (L, \Sigma, <)$ where L is an infinite regular language over a totally ordered alphabet $(\Sigma, <)$.
By enumerating the words of L w.r.t. the radix order $<_{rad}$ induced by $<$, we define a bijection :

$$\mathrm{rep}_S : \mathbb{N} \to L \qquad \mathrm{val}_S = \mathrm{rep}_S^{-1} : L \to \mathbb{N}.$$

A set $X \subseteq \mathbb{N}$ is S-recognizable if $\mathrm{rep}_S(X)$ is regular.

$L = \{a,b\}^* \qquad \Sigma = \{a,b\} \qquad a < b$

n	0	1	2	3	4	5	6	7	$\cdots$
$\mathrm{rep}(n)$	ε	a	b	aa	ab	ba	bb	aaa	$\cdots$

$L = a^*b^* \qquad \Sigma = \{a,b\} \qquad a < b$

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

A generalization

ANS generalize PNS having a regular numeration language:

Let U be a PNS and let $x, y \in \mathbb{N}$. We have

$$x < y \Leftrightarrow \mathrm{rep}_U(x) <_{rad} \mathrm{rep}_U(y).$$

Example (Fibonacci)

$\mathrm{rep}_F(\mathbb{N}) = 1\{0, 01\}^* \cup \{\varepsilon\}$

$6 < 7$ and $1001 <_{rad} 1010$ (same length)

$6 < 8$ and $1001 <_{rad} 10000$ (different lengths)

13	8	5	3	2	1	
					ε	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

A generalization

ANS generalize PNS having a regular numeration language:

Let U be a PNS and let $x, y \in \mathbb{N}$. We have

$$x < y \Leftrightarrow \text{rep}_U(x) <_{rad} \text{rep}_U(y).$$

Example (Fibonacci)

$\text{rep}_F(\mathbb{N}) = 1\{0, 01\}^* \cup \{\varepsilon\}$

$6 < 7$ and $1001 <_{rad} 1010$ (same length)

$6 < 8$ and $1001 <_{rad} 10000$ (different lengths)

13	8	5	3	2	1	
					ε	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

A generalization

ANS generalize PNS having a regular numeration language:

Let U be a PNS and let $x, y \in \mathbb{N}$. We have

$$x < y \Leftrightarrow \mathrm{rep}_U(x) <_{rad} \mathrm{rep}_U(y).$$

Example (Fibonacci)

$$\mathrm{rep}_F(\mathbb{N}) = 1\{0, 01\}^* \cup \{\varepsilon\}$$

$\mathrm{rep}_F(n)$	n
ε	0
1	1
10	2
100	3
101	4
1000	5
1001	6
1010	7
10000	8

Decimation of languages

Let L be a language ordered w.r.t. the radix order.

If $w_0 < w_1 < \cdots$ are the elements of L and $X \subseteq \mathbb{N}$, then

$$L[X] = \{w_n : n \in X\}.$$

If $S = (L, \Sigma, <)$, then $L[X] = \mathrm{rep}_S(X)$.

If $L[X]$ is accepted by a finite automaton, what does it imply on X? What conditions on X insures that $L[X]$ is regular?

Motivation for ANS

ANS are a generalization of all usual PNS like integer base numeration systems or linear numeration systems, and even rational numeration systems.

Thanks to this general point of view on numeration systems, we try to distinguish results that deeply depend on the algorithm used to represent the integers from results that only depend on the set of representations.

Due to the general setting of ANS, some new questions concerning languages arise naturally from this numeration point of view.

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- ...

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- ...

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- ...

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- ...

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- ...

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- ...

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- ...

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- ...

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- . . .

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- ...

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- ...

Some questions around ANS

- Rec. sets in a given ANS?
- Rec. sets in all ANS?
- Are there subsets of $\mathbb{N}$ that are never recognizable?
- Given a subset of $\mathbb{N}$ can we build an ANS for which it is rec.?
- How do rec. depend on the choice of the numeration?
- For which ANS do arithmetic operations preserve rec.?
- Operations preserving rec. in a given ANS?
- How to represent real numbers?
- Can we define automatic sequences in that context?
- Logical characterization of rec. sets?
- Extensions to the multidimensional setting?
- ...

S-automatic words

b-automatic words

An infinite word $x = (x_n)_{n\geq 0}$ is b-automatic if there exists a DFAO $\mathcal{A} = (Q, q_0, \Sigma_b, \delta, \Gamma, \tau)$ s.t. for all $n \geq 0$,

$$x_n = \tau(\delta(q_0, \mathrm{rep}_b(n))).$$

Theorem (Cobham 1972)

Let $b \geq 2$. An infinite word is b-automatic iff it is the image under a coding of an infinite fixed point of a b-uniform morphism.

S-automatic words

Let $S = (L, \Sigma, <)$ be an ANS.
An infinite word $x = (x_n)_{n \geq 0}$ is S-automatic if there exists a DFAO $\mathcal{A} = (Q, q_0, \Sigma, \delta, \Gamma, \tau)$ s.t. for all $n \geq 0$,

$$x_n = \tau(\delta(q_0, \mathrm{rep}_S(n))).$$

Theorem (Rigo-Maes 2002)

An infinite word is S-automatic for some ANS S iff it is the image under a coding of an infinite fixed point of a morphism, i.e. a morphic word.

Corollary

The factor complexity of an S-automatic word is $O(n^2)$.

Corollary

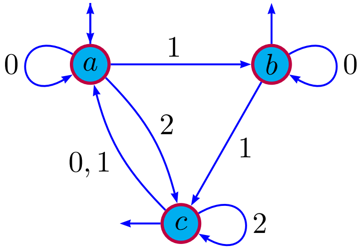
The set of primes is never S-recognizable.

Its characteristic sequence is not morphic (Mauduit 1988).

Idea of the proof

Example (Morphic $\to$ S-Automatic)

Consider the morphism μ defined by $a \mapsto abc$; $b \mapsto bc$; $c \mapsto aac$.
We have $\mu^\omega(a) = abcbcaacbcaacabcabcaacbcaacabcabc\cdots$.
One canonically associates the DFA $\mathcal{A}_{\mu,a}$



$$L_{\mu,a} = \{\varepsilon, 1, 2, 10, 11, 20, 21, 22, 100, 101, 110, 111, 112, 200, \ldots\}$$

If $S = (L_{\mu,a}, \{0,1,2\}, 0 < 1 < 2)$, then

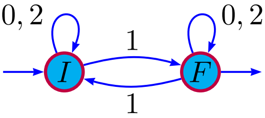
$$(\mu^\omega(a))_n = \delta_\mu(a, \mathrm{rep}_S(n)) \text{ for all } n \geq 0.$$

Idea of the proof

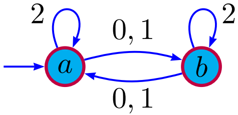
Example (S-Automatic $\to$ Morphic)

$S = (L, \{0,1,2\}, 0 < 1 < 2)$ where $L = \{w \in \Sigma^* : |w|_1 \text{ is odd}\}$

minimal automaton of L

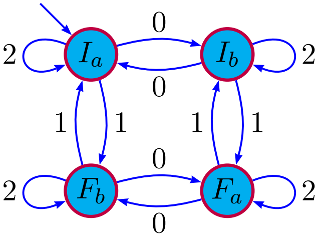


DFAO generating x



n	0	1	2	3	4	5	6	7	8	$\cdots$
$\mathrm{rep}_S(n)$	1	01	10	12	21	001	010	012	021	$\cdots$
x	b	a	a	b	b	b	b	a	a	$\cdots$

Example (Continued)



$$
\begin{aligned}
f\colon \alpha &\mapsto \alpha I_a & F_a &\mapsto F_b I_b F_a & g\colon \alpha, I_a, I_b &\mapsto \varepsilon \\
I_a &\mapsto I_b F_b I_a & F_b &\mapsto F_a I_a F_b & F_a &\mapsto a \\
I_b &\mapsto I_a F_a I_b & & & F_b &\mapsto b
\end{aligned}
$$

$L \subseteq \Sigma^*$		ε	0	1	2	00	01	02	10	11	12
$f^\omega(\alpha)$	α	I_a	I_b	F_b	I_a	I_a	F_a	I_b	F_a	I_a	F_b
x				b			a		a		b

$$g(f^\omega(\alpha)) = x$$

Multidimensional Case

A d-dimensional infinite word over an alphabet Σ is a map $x : \mathbb{N}^d \to \Sigma$. We use notation like $x_{n_1,\ldots,n_d}$ or $x(n_1, \ldots, n_d)$ to denote the value of x at $(n_1, \ldots, n_d)$.

If $w_1, \ldots, w_d$ are finite words over the alphabet Σ,

$$(w_1, \ldots, w_d)^{\#} := (\#^{m-|w_1|}w_1, \ldots, \#^{m-|w_d|}w_d)$$

where $m = \max\{|w_1|, \ldots, |w_d|\}$.

Example

$(ab, bbaa)^{\#} = (\#\#ab, bbaa) = (\#, b)(\#, b)(a, a)(b, a)$

A d-dimensional infinite word over an alphabet Γ is b-automatic if there exists a DFAO

$$\mathcal{A} = (Q, q_0, (\Sigma_b)^d, \delta, \Gamma, \tau)$$

s.t. for all $n_1, \ldots, n_d \geq 0$,

$$\tau\left(\delta\left(q_0, (\mathrm{rep}_b(n_1), \ldots, \mathrm{rep}_b(n_d))^0\right)\right) = x_{n_1,\ldots,n_d}.$$

Theorem (Salon 1987)

Let $b \geq 2$ and $d \geq 1$. A d-dimensional infinite word is b-automatic iff it is the image under a coding of a fixed point of a b-uniform d-dimensional morphism.

Theorem (C-Kärki-Rigo 2010)

Let $d \geq 1$. The d-dimensional infinite word is S-automatic for some ANS $S = (L, \Sigma, <)$ where $\varepsilon \in L$ iff it is the image under a coding of a shape-symmetric infinite d-dimensional word.

Shape-symmetric

$$\mu(a) = \mu(f) = \begin{array}{|c|c|}\hline a & b \\ \hline c & d \\ \hline \end{array}\,;\ \mu(b) = \begin{array}{|c|}\hline e \\ \hline c \\ \hline \end{array}\,;\ \mu(c) = \begin{array}{|c|c|}\hline e & b \\ \hline \end{array}\,;\ \mu(d) = \begin{array}{|c|}\hline f \\ \hline \end{array}$$

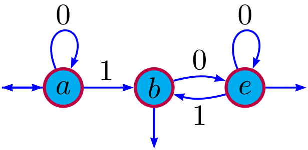
$$\mu(e) = \begin{array}{|c|c|}\hline e & b \\ \hline g & d \\ \hline \end{array}\,;\ \mu(g) = \begin{array}{|c|c|}\hline h & b \\ \hline \end{array}\,;\ \mu(h) = \begin{array}{|c|c|}\hline h & b \\ \hline c & d \\ \hline \end{array}\,.$$

$$\mu^{\omega}(a) = \begin{array}{|cc|c|cc|cc|c|c}\hline a & b & e & e & b & e & b & e & \cdots \\ c & d & c & g & d & g & d & c & \\ \hline e & b & f & e & b & h & b & f & \\ \hline e & b & e & a & b & e & b & e & \\ g & d & c & c & d & g & d & c & \\ \hline e & b & e & e & b & a & b & e & \\ g & d & c & g & d & c & d & c & \\ \hline h & b & f & e & b & e & b & f & \\ \hline \vdots & & & & & & & & \ddots \end{array}$$

Consider the morphism μ_1 defined by

$$a \mapsto ab\,;b \mapsto e\,;e \mapsto eb.$$

We have $\mu_1^\omega(a) = abeebebeebeebebeebebeebeeb\cdots$.
One canonically associates the DFA $\mathcal{A}_{\mu_1,a}$



$$L_{\mu_1,a} = \{\varepsilon, 1, 10, 100, 101, 1000, 1001, 1010, 10000, \dots\}$$

Open question

- If S and T are two ANS, (S,T)-automatic words are bidimensional infinite words $(x_{m,n})_{m,n\geq 0}$ for which there exists a DFAO $\mathcal{A} = (Q, (\Sigma \cup \{\#\})^d, \delta, q_0, \Gamma, \tau)$ s.t. $\forall m, n \in \mathbb{N}$,

$$x_{m,n} = \tau(\delta(q_0, (\mathrm{rep}_S(m), \mathrm{rep}_T(n))^{\#})).$$

  Can these (S,T)-automatic words be characterized by iterating morphisms?

b-kernel

An infinite word $(x_n)_{n\ge 0}$ is b-automatic iff its b-kernel

$$\{(x_{b^e n+r})_{n\ge 0} \colon e,r\in\mathbb{N},\ r<b^e\}$$

is finite. The b-kernel can be rewritten

$$\{(x_{b^{|w|}n+\mathrm{val}_b(w)})_{n\ge 0} \colon w\in\Sigma_b^*\}.$$

8	4	2	1		8	4	2	1			8	4	2	1	
			ε	0		1	1	0	6		1	1	0	0	12
			1	1		1	1	1	7		1	1	0	1	13
		1	0	2	1	0	0	0	8		1	1	1	0	14
		1	1	3	1	0	0	1	9		1	1	1	1	15
	1	0	0	4	1	0	1	0	10	1	0	0	0	0	16
	1	0	1	5	1	0	1	1	11	1	0	0	0	1	17

NB: $b^{|w|}n+\mathrm{val}_b(w)$ is the base-b value of the $(n+1)$-th word in $\mathcal{L}_b$ having w as a suffix.

Open question

The S-kernel of $(x_n)_{n\geq 0}$ is

$$\{(x_{f_w(n)})_{n\geq 0} \colon w \in \Sigma^*\}$$

where $f_w(n)$ is the S-value of the $(n+1)$-th word in L having w as a suffix.

Theorem (Rigo-Maes 2002)

An infinite word is S-automatic iff its S-kernel is finite.

- Does a similar characterization hold in the multidimensional setting?

Sets S-recognizable for all S

Ultimately periodic sets

It is an exercise to show that all ultimately periodic set are b-recognizable for all $b \geq 2$.

Theorem (Cobham 1969)

Let $k, \ell \geq 2$ be two multiplicatively independent integers. A subset of $\mathbb{N}$ is both k-recognizable and ℓ-recognizable iff it is ultimately periodic.

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

Corollary

A subset of $\mathbb{N}$ is b-recognizable for all $b \geq 2$ iff it is ultimately periodic.

Generalization to ANS

Theorem (Lecomte-Rigo 2001, Krieger et al. 2009)

Ultimately periodic sets are S-recognizable for all ANS S.

Corollary

A subset of $\mathbb{N}$ is S-recognizable for all ANS S iff it is ultimately periodic.

Theorem (Krieger et al. 2009, Angrand-Sakarovitch 2010)

Let $m, r \in \mathbb{N}$ with $m \geq 2$ and $0 \leq r \leq m-1$ and let $S = (L, \Sigma, <)$ be an ANS. If L is accepted by a n-state DFA, then the minimal DFA of $\mathrm{rep}_S(m\mathbb{N} + r)$ has at most nm^n states.

Semi-linear sets

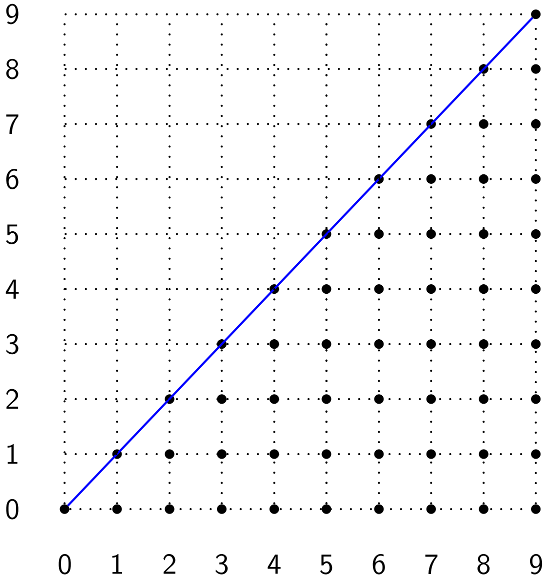
A subset X of $\mathbb{N}^d$ is b-recognizable if the language $(\mathrm{rep}_b(X))^\#$ over $(\{0,1,\ldots,b-1\} \cup \{\#\})^d$ is regular, where

$$\mathrm{rep}_b(X) = \{(\mathrm{rep}_b(n_1),\ldots,\mathrm{rep}_b(n_d)) \colon (n_1,\ldots,n_d) \in X\}.$$

Theorem (Cobham–Semenov, Semenov 1977)

Let $k, \ell \geq 2$ be two multiplicatively independent integers. A subset of $\mathbb{N}^d$ is both k-recognizable and ℓ-recognizable iff it is semi-linear.

A set $X \subseteq \mathbb{N}^d$ is linear if there exist $\mathbf{v}_0, \mathbf{v}_1, \cdots, \mathbf{v}_t \in \mathbb{N}^d$ such that $X = \mathbf{v}_0 + \mathbb{N}\,\mathbf{v}_1 + \mathbb{N}\,\mathbf{v}_2 + \cdots + \mathbb{N}\,\mathbf{v}_t$. A set $X \subseteq \mathbb{N}^d$ is semi-linear if it is a finite union of linear sets.



$$\{(n, m) : n, m \in \mathbb{N} \text{ and } n \geq m\} = \mathbb{N}(1,0) + \mathbb{N}(1,1)$$

Semi-linear sets: a good generalization?

Corollary

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

In the one-dimensional case, we have the following equivalences:

$$\text{semi-linear} \Leftrightarrow \text{ultimately periodic} \Leftrightarrow 1\text{-recognizable}.$$

Multidimensional case for ANS

One might therefore expect that the semi-linear sets are recognizable in all ANS. However, this fails to be the case, as the following example shows.

Example

The semi-linear set $X = \{n(1,2) : n \in \mathbb{N}\} = \{(n, 2n) \mid n \in \mathbb{N}\}$ is not 1-recognizable. Consider the language $\{(a^n\#^n, a^{2n}) \mid n \in \mathbb{N}\}$, consisting of the unary representations of the elements of X. Use the pumping lemma to show that this is not accepted by a finite automaton.

Let $S = (L, \Sigma, <)$ be an ANS.

A subset X of $\mathbb{N}^d$ is S-recognizable if the language $(\mathrm{rep}_S(X))^\#$ over $(\Sigma \cup \{\#\})^d$ is regular, where

$$\mathrm{rep}_S(X) = \{(\mathrm{rep}_S(n_1), \ldots, \mathrm{rep}_S(n_d)) \colon (n_1, \ldots, n_d) \in X\}.$$

It is 1-automatic if it is S-automatic for the ANS S built on a^*.

Multidimensional 1-recognizable sets

Theorem (C-Lacroix-Rampersad 2010)

A subset of $\mathbb{N}^d$ *is* S*-recognizable for all ANS* S *iff it is* 1*-recognizable.*

Theorem (C-Lacroix-Rampersad 2010)

The multidimensional 1*-recognizable sets are the finite unions of sets of the form*

$$\mathbb{N}\,\mathbf{v}_1 + \mathbf{a}_1 + \mathbb{N}\,\mathbf{v}_2 + \mathbf{a}_2 + \cdots + \mathbb{N}\,\mathbf{v}_t + \mathbf{a}_t,$$

where

- $\forall i \ \operatorname{Supp}(\mathbf{v}_i) = \operatorname{Supp}(\mathbf{a}_i)$
- $\operatorname{Supp}(\mathbf{v}_1) \supseteq \operatorname{Supp}(\mathbf{v}_2) \supseteq \cdots \supseteq \operatorname{Supp}(\mathbf{v}_t)$
- *All* $\mathbf{v}_i$ *and* $\mathbf{a}_i$ *are multiples of vectors all of whose components are* 0 *or* 1.

Recognizable sets

Another well-studied subclass of the class of semi-linear sets is the class of recognizable sets.

A subset X of $\mathbb{N}^d$ is recognizable if the right congruence $\sim_X$ has finite index ($x \sim_X y$ if $\forall z \in \mathbb{N}^d$ $(x + z \in X \Leftrightarrow y + z \in X)$).

When $d = 1$, we have again the following equivalences:

recognizable $\Leftrightarrow$ ultimately periodic $\Leftrightarrow$ 1-recognizable.

However, for $d > 1$ these equivalences no longer hold.

Multidimensional recognizable sets: a characterization

Theorem (Mezei)

The recognizable subsets of $\mathbb{N}^2$ are precisely finite unions of sets of the form $Y \times Z$, where Y and Z are ultimately periodic subsets of $\mathbb{N}$.

In particular, the diagonal set $D = \{(n, n) \mid n \in \mathbb{N}\}$ is not recognizable.

However, the set D is clearly a 1-recognizable subset of $\mathbb{N}^2$.

So we see that for $d > 1$, the class of 1-recognizable sets corresponds neither to the class of semi-linear sets, nor to the class of recognizable sets.

A description of S-recognizable sets

Which growth functions for recognizable sets?

Let L be a language over an alphabet Σ. Define

$$\mathbf{u}_L(n) = |L \cap \Sigma^n| \quad \text{and} \quad \mathbf{v}_L(n) = \sum_{i=0}^{n} \mathbf{u}_L(i) = |L \cap \Sigma^{\leq n}|$$

The maps $\mathbf{u}_L \colon \mathbb{N} \to \mathbb{N}$ and $\mathbf{v}_L \colon \mathbb{N} \to \mathbb{N}$ are the growth functions of L.

If $X \subseteq \mathbb{N}$, we let $t_X(n)$ denote the $(n+1)$-th term of X.
The map $t_X \colon \mathbb{N} \to \mathbb{N}$ is the growth function of X.

What do the growth functions of S-recognizable sets look like?

Theorem (C-Rampersad 2010)

Let $S = (L, \Sigma, <)$ *be an ANS built on a regular language and let* $X \subseteq \mathbb{N}$ *be an infinite* S*-recognizable set. Suppose*

$$\forall i \in \{0, \ldots, p-1\},\ \mathbf{v}_L(np+i) \sim a_i n^c \alpha^n \ (n \to +\infty),$$

for some $p, c \in \mathbb{N}$ *with* $p \geq 1$*, some* $\alpha \geq 1$ *and some positive constants* $a_0, \ldots, a_{p-1}$*, and*

$$\forall j \in \{0, \ldots, q-1\},\ \mathbf{v}_{\mathrm{rep}_S(X)}(nq+j) \sim b_j n^d \beta^n \ (n \to +\infty),$$

for some $q, d \in \mathbb{N}$ *with* $q \geq 1$*, some* $\beta \geq 1$ *and some positive constants* $b_0, \ldots, b_{q-1}$*. Then we have*

- $t_X(n) = \Theta\left((\log(n))^{c - d\frac{\log(\sqrt[p]{\alpha})}{\log(\sqrt[q]{\beta})}} n^{\frac{\log(\sqrt[p]{\alpha})}{\log(\sqrt[q]{\beta})}} \right)$ *if* $\beta > 1$*;*
- $t_X(n) = \Theta\left(n^{\frac{c}{d}} (\sqrt[p]{\alpha})^{\Theta(n^{\frac{1}{d}})} \right)$ *if* $\beta = 1$*.*

Exponential numeration language

Proposition (C-Rampersad 2010)

- *For all* $k, \ell \in \mathbb{N}$ *with* $\ell > 0$*, there exists an ANS* S *built on an exponential regular language and an infinite* S*-recognizable set* $X \subseteq \mathbb{N}$ *s.t.* $t_X(n) = \Theta((\log(n))^k n^\ell)$*.*
- *For all* $k, \ell \in \mathbb{N}$ *with* $\ell > 1$*, there exists an ANS* S *built on an exponential regular language and an infinite* S*-recognizable set* $X \subseteq \mathbb{N}$ *s.t.* $t_X(n) = \Theta\left(\frac{n^\ell}{(\log(n))^k}\right)$*.*
- *For all* $k \in \mathbb{N}$ *with* $k > 0$ *and for all ANS* S*, there is no* S*-recognizable set* $X \subseteq \mathbb{N}$ *s.t.* $t_X(n) = \Theta\left(\frac{n}{(\log(n))^k}\right)$*.*

Polynomial numeration language

Corollary

Let $S = (L, \Sigma, <)$ *be an ANS built on a polynomial regular language and let* $X \subseteq \mathbb{N}$ *be an infinite* S*-recognizable set. Then we have* $t_X(n) = \Theta(n^r)$ *for some rational number* $r \geq 1$*.*

Proposition (C-Rampersad 2010)

For every rational number $r \geq 1$*, there exists an ANS* S *built on a polynomial regular language and an infinite* S*-recognizable set* $X \subseteq \mathbb{N}$ *such that* $t_X(n) = \Theta(n^r)$*.*

Theorem (Eilenberg 1974)

A b-recognizable set $X = \{x_n : n \in \mathbb{N}\}$, where $x_0 < x_1 < \cdots$, satisfies either $\limsup_{n\to+\infty}(x_{n+1} - x_n) < +\infty$ *or* $\limsup_{n\to+\infty} \frac{x_{n+1}}{x_n} > 1$.

Corollary

The set of squares $\{n^2 \mid n \in \mathbb{N}\}$ is not b-automatic for all $b \geq 2$.

But it is S-automatic for the ANS built on $a^*b^* \cup a^*c^*$ with $a < b < c$ since we have $\mathrm{rep}_S(\{n^2 \mid n \in \mathbb{N}\}) = a^*$.

Proposition (Rigo 2002)

For all $k \in \mathbb{N}$, the set $\{n^k \mid n \in \mathbb{N}\}$ is S-automatic for some S.

In those constructions, the exhibited ANS are built on polynomial languages.

An example

Consider the base 4 numeration system, that is, the ANS built on $\mathcal{L}_4 = \{\varepsilon\} \cup \{1,2,3\}\{0,1,2,3\}^*$ with the natural order on the digits.

Let $X = \mathrm{val}_4(\{1,3\}^*) = \{1,3,5,7,13,15,21,23,29,31,\dots\}$. It is clearly 4-automatic.

We have $\mathbf{v}_{\mathcal{L}_4}(n) = 4^n$ and $\mathbf{v}_{\{1,3\}^*}(n) = 2^{n+1} - 1$.

We obtain

$$t_X(n) = \Theta(n^2).$$

Theorem (Durand-Rigo 2009)

Let S be an ANS built on a polynomial regular language and let T be an ANS built on an exponential regular language. If a subset of $\mathbb{N}$ is both S-recognizable and T-recognizable, then it is ultimately periodic.

Corollary

Let S be an ANS built on an exponential regular language. If $f \in \mathbb{Q}[x]$ is a polynomial of degree greater than 1 such that $f(\mathbb{N}) \subseteq \mathbb{N}$, then the set $f(\mathbb{N})$ is not S-recognizable.

Open problem (hard)

- Find a Cobham-style theorem for ANS.

The periodicity problem

The periodicity problem

Open problem

- Given an ANS S and a DFA accepting an S-recognizable set, decide if this set it ultimately periodic.

Theorem (Honkala 1985)

The periodicity problem is decidable for integer bases.

Theorem (Muchnik 1991)

The periodicity problem is decidable for linear PNS U s.t. $\mathbb{N}$ is U-recognizable and addition is computable by a finite automaton.

Results for ANS

Theorem (Honkala-Rigo 2004)

The periodicity problem for all ANS is equivalent to the HD0L periodicity problem.

Theorem (C-Rigo 2008, Bell-C-Fraenkel-Rigo 2009)

The periodicity problem is decidable for a large class of linear PNS.

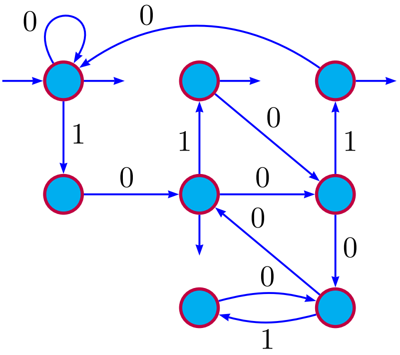
NB: Our class contains PNS for which addition is not computable by a finite automaton.

Intermediate open problem

- Given a PNS U s.t. $\mathbb{N}$ is U-recognizable and a DFA accepting an U-recognizable set, decide if this set it ultimately periodic.

Automata recognizing languages arising from linear PNS

Even numbers in Fibonacci system

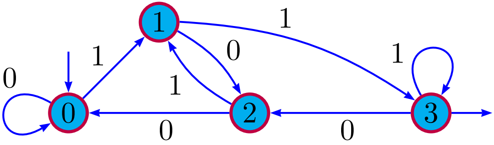


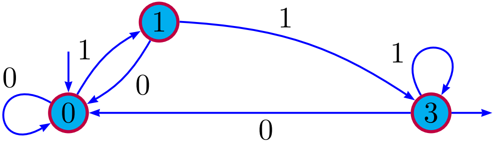
$\mathrm{rep}_F(2\,\mathbb{N})$

13	8	5	3	2	1	
					ε	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

Motivation

What is the "best automaton" we can get?





DFAs accepting the binary representations of $4\mathbb{N}+3$.

Question

The general algorithm doesn't provide a minimal automaton.
What is the state complexity of $0^* \operatorname{rep}_U(p\mathbb{N}+r)$?

Information we are looking for

Consider a linear PNS U such that $\mathbb{N}$ is U-recognizable. How many states does the minimal automaton recognizing $0^* \operatorname{rep}_U(m\mathbb{N})$ contain?

1. Give upper/lower bounds.
2. Study special cases, e.g., Fibonacci numeration system.
3. Get information on the minimal automaton $\mathcal{A}_U$ recognizing $0^* \operatorname{rep}_U(\mathbb{N})$.

Information we are looking for

Consider a linear PNS U such that $\mathbb{N}$ is U-recognizable. How many states does the minimal automaton recognizing $0^* \operatorname{rep}_U(m\mathbb{N})$ contain?

1. Give upper/lower bounds.
2. Study special cases, e.g., Fibonacci numeration system.
3. Get information on the minimal automaton $\mathcal{A}_U$ recognizing $0^* \operatorname{rep}_U(\mathbb{N})$.

First results

Theorem (C-Rampersad-Rigo-Waxweiler 2010)

Let U be a linear PNS such that $\mathrm{rep}_U(\mathbb{N})$ is regular.

(i) *The automaton $\mathcal{A}_U$ has a non-trivial strongly connected component $\mathcal{C}_U$ containing the initial state.*

(ii) *If p is a state in $\mathcal{C}_U$, then there exists $N \in \mathbb{N}$ such that $\delta_U(p, 0^n) = q_{U,0}$ for all $n \geq N$. In particular, one cannot leave $\mathcal{C}_U$ by reading a 0.*

Theorem (cont'd.)

(iii) *If* $\mathcal{C}_U$ *is the only non-trivial strongly connected component of* $\mathcal{A}_U$*, then* $\lim_{n\to+\infty} U_{n+1} - U_n = +\infty$.

(iv) *If* $\lim_{n\to+\infty} U_{n+1} - U_n = +\infty$*, then* $\delta_U(q_{U,0}, 1)$ *is in* $\mathcal{C}_U$*.*

Dominant root condition

U satisfies the dominant root condition if $\lim_{n\to+\infty} U_{n+1}/U_n = \beta$ for some real $\beta > 1$.

β is the dominant root of the recurrence.

E.g., Fibonacci: dominant root $\beta = (1+\sqrt{5})/2$

Theorem (cont'd.)

Suppose U has a dominant root $\beta > 1$.

(v) *If $\mathcal{A}_U$ has more than one non-trivial strongly connected component, then any such component other than $\mathcal{C}_U$ is a cycle all of whose edges are labeled 0.*

(vi) *If $\lim_{n\to+\infty} U_{n+1}/U_n = \beta^-$, then there is only one non-trivial strongly connected component.*

An example with two components

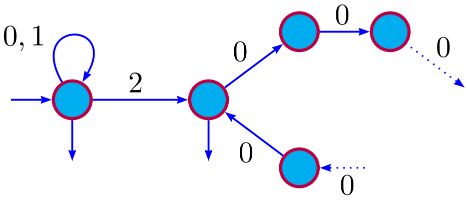
Let $t \geq 1$.

Let $U_0 = 1$, $U_{tn+1} = 2U_{tn} + 1$, and $U_{tn+r} = 2U_{tn+r-1}$, for $1 < r \leq t$.

E.g., for $t = 2$ we have $U = (1, 3, 6, 13, 26, 53, \ldots)$.

Then $0^* \operatorname{rep}_U(\mathbb{N}) = \{0,1\}^* \cup \{0,1\}^* 2(0^t)^*$.

The second component is a cycle of t 0's.



Theorem (Hollander 1998)

If U is a linear PNS has a dominant root β and if $\mathrm{rep}_U(\mathbb{N})$ is regular, then β is a Parry number.

With any Parry number β is associated a canonical finite automaton $\mathcal{A}_\beta$.

We will study the relationship between $\mathcal{A}_U$ and $\mathcal{A}_\beta$.

β-expansions

Let $\beta > 1$ be a real number.

The β-expansion of a real number $x \in [0,1]$ is the lexicographically greatest sequence $\mathsf{d}_\beta(x) := (x_i)_{i \geq 1}$ over $\{0, \ldots, \lfloor \beta \rfloor\}$ satisfying

$$x = \sum_{i=1}^{\infty} x_i \beta^{-i}.$$

Parry numbers

If $\mathsf{d}_\beta(1) = t_1 \cdots t_m 0^\omega$, with $t_m \neq 0$, then $\mathsf{d}_\beta(1)$ is finite.

In this case $\mathsf{d}^*_\beta(1) := (t_1 \cdots t_{m-1}(t_m - 1))^\omega$.

For instance, $\mathsf{d}_2(1) = 20^\omega$ and $\mathsf{d}^*_2(1) = 1^\omega$.

Otherwise $\mathsf{d}^*_\beta(1) := \mathsf{d}_\beta(1)$.

If $\mathsf{d}^*_\beta(1)$ is ultimately periodic, then β is a Parry number.

The Parry automaton

Theorem (Parry 1960)

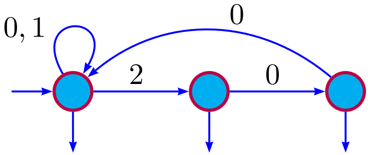
A sequence $(s_i)_{i\geq 1}$ over $\mathbb{N}$ is the β-expansion of a real number in $[0,1)$ iff $(s_{n+i})_{i\geq 1}$ is lexicographically less than $d^*_\beta(1)$ for all $n \in \mathbb{N}$.

Define D_β to be the set of all β-expansions of real numbers in $[0,1)$.

So, for β Parry, the language $\mathrm{Fact}(D_\beta)$ is regular.

For β Parry, let $\mathcal{A}_\beta$ be the minimal DFA accepting $\mathrm{Fact}(D_\beta)$.

An example of the automaton $\mathcal{A}_\beta$



Let β be the largest root of $X^3 - 2X^2 - 1$.

$\mathsf{d}_\beta(1) = 2010^\omega$ and $\mathsf{d}^*_\beta(1) = (200)^\omega$.

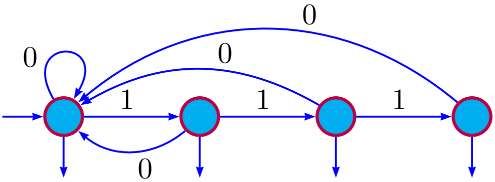
This automaton also accepts $0^* \operatorname{rep}_U(\mathbb{N})$ for U defined by $U_{n+3} = 2U_{n+2} + U_n, (U_0, U_1, U_2) = (1, 3, 7)$.

$\mathcal{A}_U = \mathcal{A}_\beta$

Bertrand numeration systems

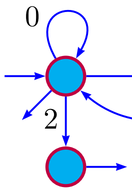
Bertrand numeration system: $w \in \mathrm{rep}_U(\mathbb{N}) \Leftrightarrow w0 \in \mathrm{rep}_U(\mathbb{N})$.

Example (The ℓ-bonacci system is Bertrand.)



$U_{n+\ell} = U_{n+\ell-1} + U_{n+\ell-2} + \cdots + U_n$
$U_i = 2^i$, $i \in \{0, \ldots, \ell - 1\}$
$\mathcal{A}_U$ accepts all words that do not contain 1^ℓ.

A non-Bertrand system



$$U_{n+2} = U_{n+1} + U_n, (U_0 = 1, U_1 = 3)$$

$$(U_n)_{n \geq 0} = 1, 3, 4, 7, 11, 18, 29, 47, \ldots$$

2 is a greedy representation but 20 is not.

Theorem (Bertrand 1989)

A PNS U is Bertrand iff there is a $\beta > 1$ such that

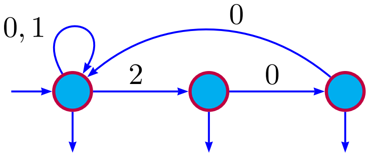
$$0^* \operatorname{rep}_U(\mathbb{N}) = \operatorname{Fact}(D_\beta).$$

Moreover, the system is derived from the β-development of 1.

If β is a Parry number, then U is linear and we have a minimal finite automaton $\mathcal{A}_\beta$ accepting $\operatorname{Fact}(D_\beta)$.

Consequently, $\operatorname{rep}_U(\mathbb{N})$ is regular and $\mathcal{A}_U = \mathcal{A}_\beta$.

Back to a previous example



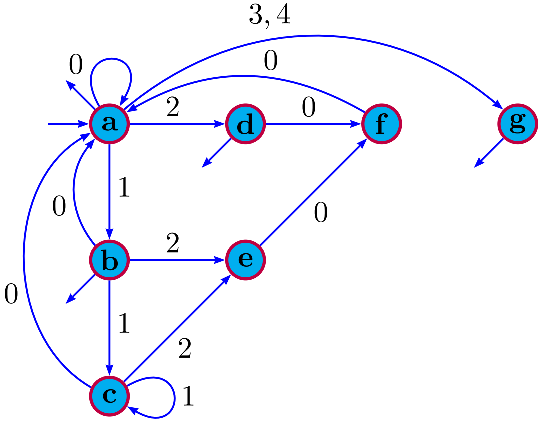
Let β be the largest root of $X^3 - 2X^2 - 1$.

$\mathsf{d}_\beta(1) = 2010^\omega$ and $\mathsf{d}^*_\beta(1) = (200)^\omega$.

This automaton accepts $0^* \operatorname{rep}_U(\mathbb{N})$ for U defined by $U_{n+3} = 2U_{n+2} + U_n, (U_0, U_1, U_2) = (1, 3, 7)$.

$\mathcal{A}_U = \mathcal{A}_\beta$

Changing the initial conditions



$U_{n+3} = 2U_{n+2} + U_n, (U_0, U_1, U_2) = (1, 3, 7)$

We change the initial values to $(U_0, U_1, U_2) = (1, 5, 6)$.

$\mathcal{A}_U \neq \mathcal{A}_\beta$

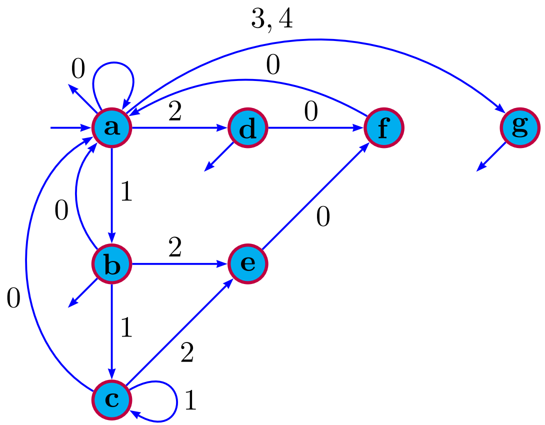
Relationship with $\mathcal{A}_\beta$

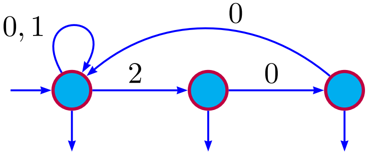
Theorem (cont'd.)

Suppose U has a dominant root $\beta > 1$. There is a morphism of automata Φ from $\mathcal{C}_U$ to $\mathcal{A}_\beta$.

Φ maps the states of $\mathcal{C}_U$ onto the states of $\mathcal{A}_\beta$ so that

- $\Phi(q_{U,0}) = q_{\beta,0}$,
- for all states q and all letters σ s.t. q and $\delta_U(q,\sigma)$ are in $\mathcal{C}_U$, we have $\Phi(\delta_U(q,\sigma)) = \delta_\beta(\Phi(q),\sigma)$.


3, 4
0
0
2
0
a
d
f
g
1
0
0
b
2
e
0
1
2
c
1


0, 1
0
2
0

Other results

When U has a dominant root $\beta > 1$, we can say more.

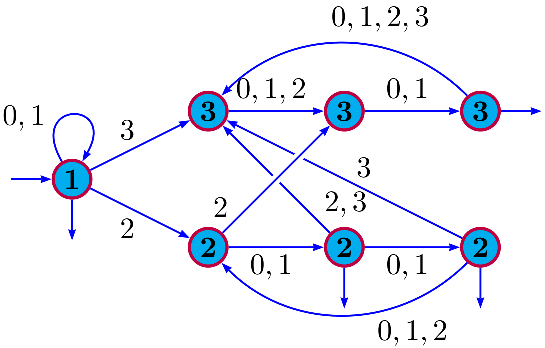
E.g., if $\mathcal{A}_U$ has more than one non-trivial strongly connected component, then $\mathrm{d}_\beta(1)$ is finite.

We can also give sufficient conditions for $\mathcal{A}_U$ to have more than one non-trivial strongly connected component.

In addition, we can give an upper bound on the number of non-trivial strongly connected components.

When U has no dominant root, the situation is more complicated.

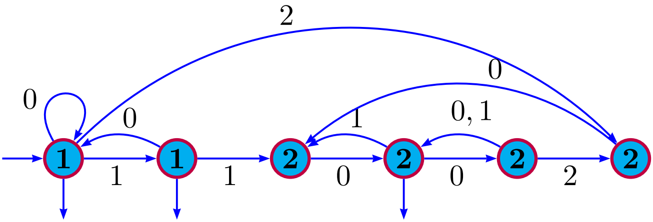
A system with no dominant root



$U_{n+3} = 24U_n, (U_0, U_1, U_2) = (1, 2, 6)$

3 non-trivial strongly connected components

A system with no dominant root



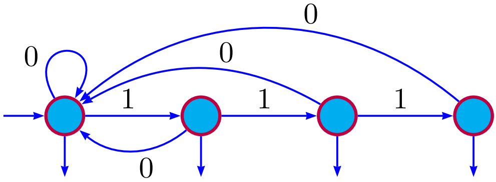
$$U_{n+4} = 3U_{n+2} + U_n, (U_0, U_1, U_2, U_3) = (1, 2, 3, 7)$$

U_{n+1}/U_n doesn't converge, but

$$\lim_{n\to+\infty} U_{2n+2}/U_{2n} = \lim_{n\to+\infty} U_{2n+3}/U_{2n+1} = (3+\sqrt{13})/2$$

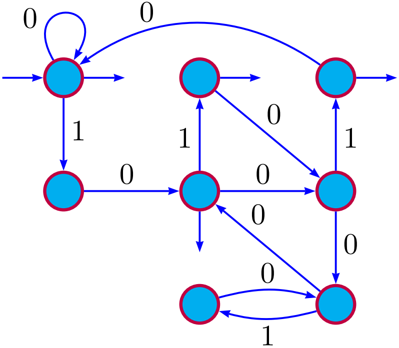
M. Hollander, Greedy numeration systems and regularity, *Theory Comput. Systems* **31** (1998).

Application to the ℓ-bonacci system



Corollary

For U the ℓ-bonacci numeration system, the number of states of the trim minimal automaton accepting $0^* \operatorname{rep}_U(m\mathbb{N})$ is ℓm^ℓ.



$\mathrm{rep}_F(2\,\mathbb{N})$

13	8	5	3	2	1	
					ε	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

Further work in this area

- Analyze the structure of $\mathcal{A}_U$ for systems with no dominant root.
- Remove the assumption that $(U_i \mod m)_{i \geq 0}$ is purely periodic in the state complexity result that we have.
- Get analogous syntactic complexity or radius complexity results.

Real numbers in ANS

The decimal representation of $\frac{11}{13}$ is $0.(846153)^\omega$:

$$\frac{8}{10}, \frac{84}{100}, \frac{846}{1000}, \frac{8461}{10000}, \frac{84615}{100000}, \ldots$$

$$n\text{-th fraction} = \frac{\mathrm{val}_{10}(\text{prefix of length } n \text{ of } (846153)^\omega)}{10^n}$$

$$\forall L \subseteq \Sigma^*,\ \mathbf{u}_L(n) = \mathrm{Card}(L \cap \Sigma^n);$$

$$\mathbf{v}_L(n) = \mathrm{Card}(L \cap \Sigma^{\leq n}) = \sum_{i=0}^{n} \mathbf{u}_L(i).$$

For the integer base $b \geq 2$:

$$\mathcal{L}_b = \{\varepsilon\} \cup \{1, \ldots, b-1\}\{0, \ldots, b-1\}^*$$

$$\mathbf{v}_{\mathcal{L}_b}(n) = \sum_{i=0}^{n} \mathbf{u}_{\mathcal{L}_b}(i) = b^n.$$

The decimal representation of $\frac{11}{13}$ is $0.(846153)^\omega$:

$$\frac{8}{10}, \frac{84}{100}, \frac{846}{1000}, \frac{8461}{10000}, \frac{84615}{100000}, \ldots$$

$$n\text{-th fraction} = \frac{\mathrm{val}_{10}(\text{prefix of length } n \text{ of } (846153)^\omega)}{\mathbf{v}_{\mathcal{L}_{10}}(n)}$$

The binary representation of $\frac{11}{13}$ is $0.(110110001001)^\omega$:

$$\frac{1}{2}, \frac{3}{4} = \frac{6}{8}, \frac{13}{16}, \frac{27}{32} = \frac{54}{64} = \frac{108}{128} = \frac{216}{256}, \frac{433}{512} = 0.845703125, \ldots$$

$$n\text{-th fraction} = \frac{\mathrm{val}_2(\text{prefix of length } n \text{ of } (110110001001)^\omega)}{\mathbf{v}_{\mathcal{L}_2}(n)}$$

7-th fraction: $108 = 64 + 32 + 8 + 4 = \mathrm{val}_2(1101100)$

$128 = 2^7 = \mathbf{v}_{\mathcal{L}_2}(7)$.

Lecomte and Rigo

- $S = (L, \Sigma, <)$
- $w \in \Sigma^\omega$
- $(w^{(n)})_{n \geq 0} \in L^{\mathbb{N}}$, $\qquad w^{(n)} \to w$ as $n \to +\infty$

POINT: To show that, under certain hypotheses, the limit

$$\lim_{n \to +\infty} \frac{\mathrm{val}_S(w^{(n)})}{\mathbf{v}_L(|w^{(n)}|)}$$ exists and only depends on w.

In that case, w is an S-representation of the corresponding real.

QUESTION: And when L is not regular?

Example

The $\frac{3}{2}$-number system introduced by Akiyama, Frougny and Sakarovitch (2008) has a numeration language which is not context-free.

AIM: To provide a unified approach for representing real numbers

Generalization to non-regular languages

- Arbitrary infinite language L (not necessarily regular)
- Minimal automaton of L: $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$
- "Generalized" ANS: $S = (L, \Sigma, <)$

For all $x \in L$, the numerical value $\mathrm{val}_S(x)$ of x is given by

$$\mathbf{v}_L(|x|-1) + \sum_{i=0}^{|x|-1} \sum_{a<x[i]} \mathbf{u}_{L_{\delta(q_0, x[0,i-1]a)}}(|x|-i-1),$$

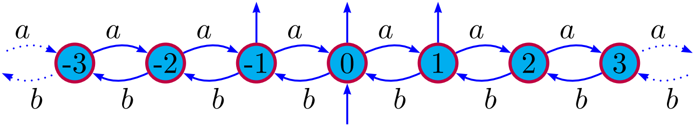
where $x[0, i-1]$ = prefix of length i of x
and L_q = language accepted from q in $\mathcal{A}$.

- $w =$ limit of words in $L \Leftrightarrow \mathrm{Pref}(w) \subseteq \mathrm{Pref}(L)$
  $\Leftrightarrow w \in \mathrm{Adh}(L)$

Since $\mathrm{Adh}(L) = \mathrm{Adh}(\mathrm{Pref}(L))$, there is no new representation if we assume that L is prefix-closed.

Example: $L = \{w \in \{a,b\}^* \mid ||w|_a - |w|_b| \leq 1\}$

$= \{\varepsilon, a, b, ab, ba, aab, aba, abb, baa, bab, bba, aabb, \ldots\}$



For $S = (L, \{a,b\}, a < b)$, we can compute

$$\lim_{n\to+\infty} \frac{\mathrm{val}_S((ab)^n)}{\mathbf{v}_L(2n)} = \frac{3}{4} \quad \text{and} \quad \lim_{n\to+\infty} \frac{\mathrm{val}_S((ab)^n a)}{\mathbf{v}_L(2n+1)} = \frac{3}{5}$$

which shows that $\displaystyle\lim_{n\to+\infty} \frac{\mathrm{val}_S((ab)^\omega[0,n-1])}{\mathbf{v}_L(n)}$ does not exist.

L not prefix-closed: $\mathrm{Pref}(L) = \{a,b\}^*$

Hypotheses needed?

- $(\mathrm{H1})$ L is prefix-closed
- $(\mathrm{H2})$ $\mathrm{Adh}(L)$ is uncountable

QUESTION: What conditions must L satisfy so that

the limits $\displaystyle\lim_{n\to+\infty} \frac{\mathrm{val}_S(w[0,n-1])}{\mathbf{v}_L(n)}$ exist for all $w \in \mathrm{Adh}(L)$?

Hypotheses needed?

AIM: Define some approximation intervals of reals.

Their length should decrease as the prefix that is read becomes larger and larger.

$$\forall x \in L \cap \Sigma^n,\ \underbrace{\frac{\mathbf{v}_L(n-1)}{\mathbf{v}_L(n)}}_{=1-\frac{\mathbf{u}_L(n)}{\mathbf{v}_L(n)}} \leq \frac{\mathrm{val}_S(x)}{\mathbf{v}_L(n)} \leq \underbrace{\frac{\mathbf{v}_L(n)}{\mathbf{v}_L(n)}}_{=1}$$

If $\lim_{n\to+\infty} \frac{\mathbf{u}_L(n)}{\mathbf{v}_L(n)}$ exists , then it is denoted by r_ε and the represented interval is $I_\varepsilon = [1 - r_\varepsilon, 1]$.

Hypotheses needed?

Recall that, for all $x \in L$,

$$\mathrm{val}_S(x) = \mathbf{v}_L(|x|-1) + \sum_{i=0}^{|x|-1} \sum_{a<x[i]} \mathbf{u}_{L_{\delta(q_0,x[0,i-1]a)}}(|x|-i-1)$$

- (H3) $\forall x \in \Sigma^*,\ \exists r_x \geq 0,\ \lim_{n\to+\infty} \frac{\mathbf{u}_{L_{\delta(q_0,x)}}(n-|x|)}{\mathbf{v}_L(n)} = r_x$

Hypotheses needed?

In general, $|I_x| = r_x$

- (H4) $\forall w \in \mathrm{Adh}(L)$, $\lim_{\ell \to +\infty} r_{w[0,\ell-1]} = 0$

Let $\mathrm{Center}(L) = \mathrm{Pref}(\mathrm{Adh}(L))$. Then $x \notin \mathrm{Center}(L) \Leftrightarrow r_x = 0$.

4 hypotheses

Theorem (C.-Le Gonidec-Rigo 2010)

The limits $\lim_{n\to+\infty} \dfrac{\mathrm{val}_S(w[0,n-1])}{\mathbf{v}_L(n)}$ *exist when* L *satisfies the following conditions:*

- (H1) L *prefix-closed*
- (H2) $\mathrm{Adh}(L)$ *uncountable*
- (H3) $\forall x \in \Sigma^*,\ \exists r_x \geq 0,\ \lim_{n\to+\infty} \dfrac{\mathbf{u}_{L_{\delta(q_0,x)}}(n-|x|)}{\mathbf{v}_L(n)} = r_x$
- (H4) $\forall w \in \mathrm{Adh}(L),\ \lim_{\ell\to+\infty} r_{w[0,\ell-1]} = 0$

For all $w \in \mathrm{Adh}(L)$, $\mathrm{val}_S(w) = \lim_{n \to +\infty} \frac{\mathrm{val}_S(w[0, n-1])}{\mathbf{v}_L(n)}$

is the numerical value of w.

The infinite word w is an S-representation of $\mathrm{val}_S(w)$.

Proposition (C.-Le Gonidec-Rigo 2010)

Let $L \subseteq \Sigma^*$, $S = (\mathrm{Pref}(L), \Sigma, <)$ be a (generalized) ANS. If $\mathrm{Pref}(L)$ satisfies (H1), (H2) and (H3), then for all sequences $(w^{(n)})_{n\geq 0} \in L^{\mathbb{N}}$ converging to a word $w \in \mathrm{Adh}(L)$, we have

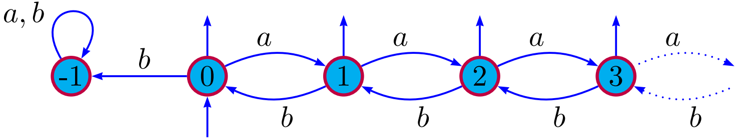
$$\lim_{n\to+\infty} \frac{\mathrm{val}_S(w^{(n)})}{\mathbf{v}_{\mathrm{Pref}(L)}(|w^{(n)}|)} = \mathrm{val}_S(w).$$

Example: Prefixes of Dyck words

$$D = \{w \in \{a,b\}^* \mid |w|_a = |w|_b \text{ and } \forall u \in \mathrm{Pref}(w),\ |u|_a \geq |u|_b\}$$

not prefix-closed $\longrightarrow$ we consider $S = (\mathrm{Pref}(D), \{a,b\}, a < b)$

$$\begin{aligned}\mathrm{Pref}(D) &= \{w \in \{a,b\}^* \mid \forall u \in \mathrm{Pref}(w),\ |u|_a \geq |u|_b\} \\ &= \{\varepsilon, a, aa, ab, aaa, aab, aba, aaaa, aaab, aaba, aabb, \ldots\}.\end{aligned}$$



Example: Prefixes of Dyck words (continued)

$$\lim_{n\to+\infty} \frac{\mathrm{val}_S((aab)^\omega[0,n-1])}{\mathbf{v}_L(n)} = \frac{39}{49} = 0.795918\ldots$$

x	$\mathrm{val}_S(x)$	$\mathbf{v}_L(\lvert x\rvert)$	$\frac{\mathrm{val}_S(x)}{\mathbf{v}_L(\lvert x\rvert)}$
a	1	2	0.50000
aa	2	4	0.50000
aab	5	7	0.71429
$aaba$	9	13	0.69231
$aabaa$	17	23	0.73913
$aabaab$	32	43	0.74419
$aabaaba$	60	78	0.76923
$aabaabaa$	112	148	0.75676
$aabaabaab$	213	274	0.77737
$aabaabaaba$	404	526	0.76806
$aabaabaabaa$	771	988	0.78036

Example: Prefixes of Dyck words (continued)

Since $\lim_{n \to +\infty} \frac{\mathbf{v}_L(n-1)}{\mathbf{v}_L(n)} = \frac{1}{2}$, we represent the interval $I_\varepsilon = [\frac{1}{2}, 1]$

$\mathrm{Center}(\mathrm{Pref}(D)) = \mathrm{Pref}(D)$:

- $I_a = [1/2, 1]$
- $I_{aa} = [1/2, 7/8]$ $I_{ab} = [7/8, 1]$
- $I_{aaa} = [1/2, 3/4]$ $I_{aab} = [3/4, 7/8]$ $I_{aba} = [7/8, 1]$
- $\ldots$

Example: Prefixes of Dyck words (continued)

$\forall x \in [\frac{1}{2}, 1]$, Q_x designates the set of representations of x.

We have $Q_{1/2} = \{a^\omega\}$ and $Q_1 = \{(ab)^\omega\}$.

If $x \in]1/2, 1[$ and $x = \sup I_w = \inf I_z$ then $Q_x = \{\bar{w}(ab)^\omega, za^\omega\}$, where $\bar{w} =$ the least Dyck word having w as a prefix.

Proposition (C.-Le Gonidec-Rigo 2010)

If L is context-free, then the representations of the endpoints of the intervals are ultimately periodic.

Open problems

- Characterize the automata recognizing a language L such that the corresponding ω-language $\mathrm{Adh}(L)$ is uncountable.

Open problems

Theorem (Boasson-Nivat 1980)

For every context-free language L, there exists a sequential mapping f such that $f(\mathrm{Adh}(D_2)) = \mathrm{Adh}(L)$, where D_2 is the Dyck language over two kinds of parentheses.

- Let S and T be abstract numeration systems built respectively on $\mathrm{Pref}(D_2)$ and $\mathrm{Pref}(L)$. Give a mapping g such that the following diagram commutes.

$$\begin{tikzcd}
\mathrm{Adh}(D_2) \arrow[r, "f"] \arrow[d, "\mathrm{val}_S"'] & \mathrm{Adh}(L) \arrow[d, "\mathrm{val}_T"] \\
{[s_0, 1]} \arrow[r, "g"] & {[t_0, 1]}
\end{tikzcd}$$