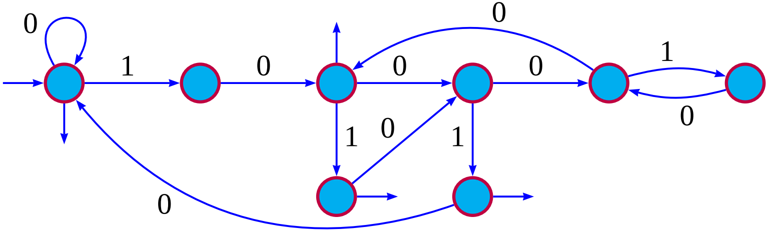
Structure of the minimal automaton of a numeration language & State complexity of testing divisibility

É. Charlier N. Rampersad M. Rigo L. Waxweiler

Département de mathématiques
Université de Liège

Journées montoises d'informatique théorique 2010
Amiens, September 6

An example first



13	8	5	3	2	1	
				1	0	2
			1	0	1	4
		1	0	0	1	6
	1	0	0	0	0	8
	1	0	0	1	0	10
	1	0	1	0	1	12
						⋮

The set $2\mathbb{N}$ of even integers is *F-recognizable* or *F-automatic*, i.e., the language $\mathrm{rep}_F(2\mathbb{N}) = \{\varepsilon, 10, 101, 1001, 10000, \ldots\}$ is accepted by some finite automaton.

Remark (in terms of the Chomsky hierarchy)

With respect to the Fibonacci system, *any* F-recognizable set can be considered as a "*particularly simple*" set of integers.

We get a similar definition for other numeration systems.

Numeration systems

- A *numeration system* is an increasing sequence of integers $U = (U_n)_{n \geq 0}$ such that
  - $U_0 = 1$ and
  - $C_U := \sup_{n \geq 0} \lceil U_{n+1}/U_n \rceil < +\infty$.
- U is *linear* if it satisfies a linear recurrence relation over $\mathbb{Z}$.

Example

Let $(F_n)_{n \geq 0}$ be the Fibonacci sequence with $F_0 = 1$ and $F_1 = 2$.

- Let $n \in \mathbb{N}$. A word $w = w_{\ell-1} \cdots w_0$ over $\mathbb{N}$ represents n if

$$\sum_{i=0}^{\ell-1} w_i U_i = n.$$

Greedy representations

- A representation $w = w_{\ell-1} \cdots w_0$ of an integer is *greedy* if
$$\forall j, \; \sum_{i=0}^{j-1} w_i \, U_i < U_j.$$
- In that case, $w \in \{0, 1, \ldots, C_U - 1\}^*$.
- $\mathrm{rep}_U(n)$ is the greedy representation of n with $w_{\ell-1} \neq 0$.
- $X \subseteq \mathbb{N}$ U-recognizable $\overset{\Delta}{\Leftrightarrow}$ $\mathrm{rep}_U(X)$ is accepted by a finite automaton.
- $\mathrm{rep}_U(\mathbb{N})$ is the numeration language.

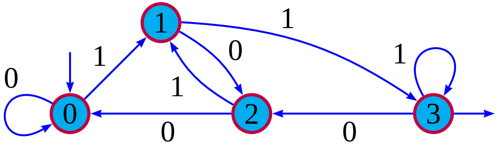
Motivations

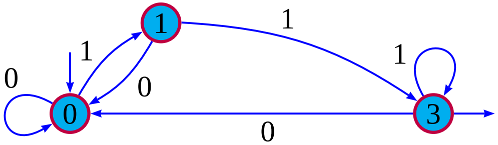
- Cobham's theorem for integer base systems (1969) shows that *recognizability depends on the choice of the base*. Only ultimately periodic sets are recognizable in all bases.
- Introduction of non-standard numeration systems and study U-recognizable sets.
- If $\mathbb{N}$ is U-recognizable, then U is linear and any ultimately periodic set is U-recognizable.

- V. Bruyère, G. Hansel, C. Michaux, R. Villemaire, Logic and p-recognizable sets of integers, *BBMS* **1** (1994).
- V. Bruyère, G. Hansel, Bertrand numeration systems and recognizability, *TCS* **181** (1997).

Motivations

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 $0^* \operatorname{rep}_U(p\mathbb{N}+r)$?

Background (I)

Theorem

If L accepted by an n-state DFA, then the minimal automaton accepting the language of words of L indexed by the multiples of m (w.r.t. the radix order) has at most nm^n states.

- D. Krieger, A. Miller, N. Rampersad, B. Ravikumar, J. Shallit, Decimations of languages and state complexity, *TCS* **410** (2009).

For $x, y \in \mathbb{N}$, we have $x < y \Leftrightarrow \mathrm{rep}_U(x) <_{\mathrm{rad}} \mathrm{rep}_U(y)$.

In particular, if $\mathrm{rep}_U(\mathbb{N})$ is accepted by an n-state DFA, then the minimal automaton accepting $\mathrm{rep}_U(m\mathbb{N})$ has at most nm^n states.

Background (II)

Alexeev's result

Let $b, m \geq 2$. Let N, M be such that $b^N < m \leq b^{N+1}$ and

$$(m, 1) < (m, b) < \cdots < (m, b^M) = (m, b^{M+1}) = (m, b^{M+2}) = \cdots .$$

The minimal automaton accepting the base b representations of the multiples of m has exactly

$$\frac{m}{(m, b^{N+1})} + \sum_{t=0}^{\inf\{N, M-1\}} \frac{b^t}{(m, b^t)} \text{ states.}$$

- B. Alexeev, Minimal DFA for testing divisibility, *JCSS* **69** (2004).

Background (III)

Honkala's decision procedure

Given any finite automaton recognizing a set X of integers written in base b, it is decidable whether X is ultimately periodic.

- J. Honkala, A decision method for the recognizability of sets defined by number systems, *Theor. Inform. Appl.* **20** (1986).
- J.-P. Allouche, N. Rampersad, J. Shallit, Periodicity, repetitions, and orbits of an automatic sequence, *TCS* **410** (2009).
- J. Bell, É. C., A. S. Fraenkel, M. Rigo, A decision problem for ultimately periodic sets in non-standard numeration systems, *IJAC* **19** (2009).

Information we are looking for

Consider a linear numeration system 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 $0^* \operatorname{rep}_U(\mathbb{N})$?

The Hankel matrix

- Let $U = (U_n)_{n\geq 0}$ be a numeration system.
- For $t \geq 1$ define

$$H_t := \begin{pmatrix} U_0 & U_1 & \cdots & U_{t-1} \\ U_1 & U_2 & \cdots & U_t \\ \vdots & \vdots & \ddots & \vdots \\ U_{t-1} & U_t & \cdots & U_{2t-2} \end{pmatrix}.$$

- For $m \geq 2$, define $k_{U,m}$ to be the largest t such that $\det H_t \not\equiv 0 \pmod{m}$.

Calculating $k_{U,m}$

- $U_{n+2} = 2U_{n+1} + U_n$, $(U_0, U_1) = (1, 3)$
- $(U_n)_{n\geq 0} = 1, 3, 7, 17, 41, 99, 239, \ldots$
- $(U_n \bmod 2)_{n\geq 0}$ is constant and trivially satisfies the recurrence relation $U_{n+1} = U_n$ with $U_0 = 1$.
- Hence $k_{U,2} = 1$.
- Modulo 4 we find $k_{U,4} = 2$.

A system of linear congruences

- Let $k = k_{U,m}$.
- Let $\mathbf{x} = (x_1, \ldots, x_k)$.
- Let $S_{U,m}$ denote the number of k-tuples $\mathbf{b}$ in $\{0, \ldots, m-1\}^k$ such that the system

$$H_k \, \mathbf{x} \equiv \mathbf{b} \pmod{m}$$

has at least one solution.

Calculating $S_{U,m}$

- $U_{n+2} = 2U_{n+1} + U_n$, $(U_0, U_1) = (1, 3)$
- $(U_n)_{n \geq 0} = 1, 3, 7, 17, 41, 99, 239, \ldots$
- Consider the system

$$\begin{cases} 1\,x_1 + 3\,x_2 & \equiv & b_1 & \pmod{4} \\ 3\,x_1 + 7\,x_2 & \equiv & b_2 & \pmod{4} \end{cases}$$

- $2x_1 \equiv b_2 - b_1 \pmod{4}$
- For each value of b_1 there are at most 2 values for b_2.
- Hence $S_{U,4} = 8$.

Properties of the automata we consider

(H.1) $\mathcal{A}_U$ has a single strongly connected component $\mathcal{C}_U$.

(H.2) For all states p, q in $\mathcal{C}_U$ with $p \neq q$, there exists a word x_{pq} such that $\delta_U(p, x_{pq}) \in \mathcal{C}_U$ and $\delta_U(q, x_{pq}) \notin \mathcal{C}_U$, or vice-versa.

General state complexity result

Theorem

Let $m \geq 2$ be an integer. Let $U = (U_n)_{n\geq 0}$ be a linear numeration system such that

(a) $\mathbb{N}$ is U-recognizable and $\mathcal{A}_U$ satisfies (H.1) and (H.2),

(b) $(U_n \bmod m)_{n\geq 0}$ is purely periodic.

The number of states of the trim minimal automaton accepting $0^* \operatorname{rep}_U(m\mathbb{N})$ from which infinitely many words are accepted is

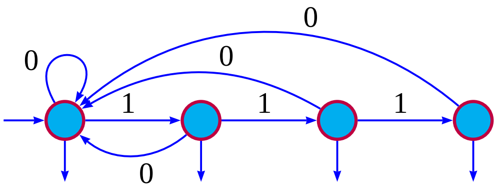
$$|\mathcal{C}_U| S_{U,m}.$$

Result for strongly connected automata

Corollary

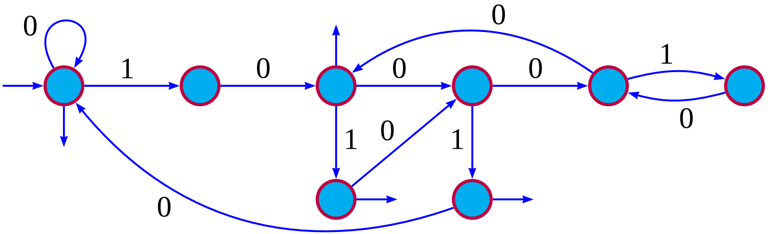
If U satisfies the conditions of the previous theorem and $\mathcal{A}_U$ is strongly connected, then the number of states of the trim minimal automaton accepting $0^* \operatorname{rep}_U(m\mathbb{N})$ is $|\mathcal{C}_U| S_{U,m}$.

Result for the ℓ-bonacci system



Corollary

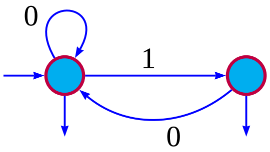
For U the ℓ-bonacci numeration system, the number of states of the trim minimal automaton accepting $0^* \operatorname{rep}_U(m\mathbb{N})$ is ℓm^ℓ.



13	8	5	3	2	1	
				1	0	2
			1	0	1	4
		1	0	0	1	6
	1	0	0	0	0	8
	1	0	0	1	0	10
	1	0	1	0	1	12
						⋮

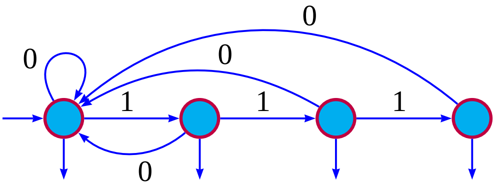
Structure of the minimal automaton $\mathcal{A}_U$ recognizing $0^* \operatorname{rep}_U(\mathbb{N})$

The Fibonacci numeration system



- $U_{n+2} = U_{n+1} + U_n$ $(U_0 = 1,\, U_1 = 2)$
- $\mathcal{A}_U$ accepts all words that do not contain 11.

The ℓ-bonacci numeration system



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

First results

Theorem

Let U be a linear numeration system 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 $\beta = (1 + \sqrt{5})/2$

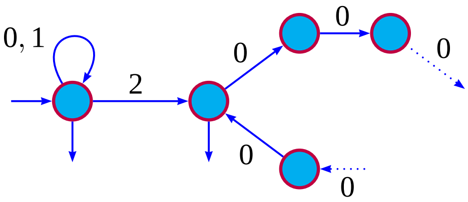
Theorem (cont'd.)

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

- 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.
- 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.



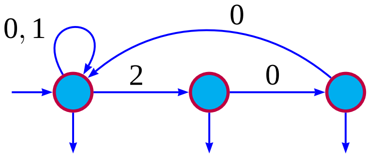
If U is a linear numeration system has a dominant root β and if $\mathrm{rep}_U(\mathbb{N})$ is regular, then β is a Parry number.

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

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

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

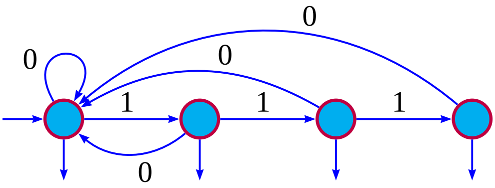
An example of the automaton $\mathcal{A}_\beta$



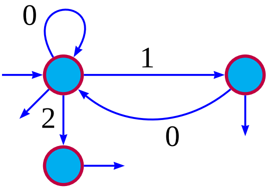
- Let β be the largest root of $X^3 - 2X^2 - 1$.
- $\mathrm{d}_\beta(1) = 2010^\omega$ and $\mathrm{d}^*_\beta(1) = (200)^\omega$.
- This automaton also accepts $\mathrm{rep}_U(\mathbb{N})$ for U defined by $U_{n+3} = 2U_{n+2} + U_n, (U_0, U_1, U_2) = (1, 3, 7)$.
- $\mathcal{A}_U = \mathcal{A}_\beta$

Bertrand numeration systems

- *Bertrand numeration system*: w is in $\mathrm{rep}_U(\mathbb{N})$ if and only if $w0$ is in $\mathrm{rep}_U(\mathbb{N})$.
- E.g., the ℓ-bonacci system is Bertrand.



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)

A system U is Bertrand if and only if 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, the system 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 $\mathcal{A}_U = \mathcal{A}_\beta$.

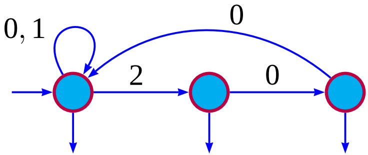
Applying our state complexity result to the Bertrand systems

Proposition

Let U be the Bertrand numeration system associated with a non-integer Parry number $\beta > 1$. The set $\mathbb{N}$ is U-recognizable and the trim minimal automaton $\mathcal{A}_U$ of $0^* \operatorname{rep}_U(\mathbb{N})$ fulfills properties (H.1) and (H.2).

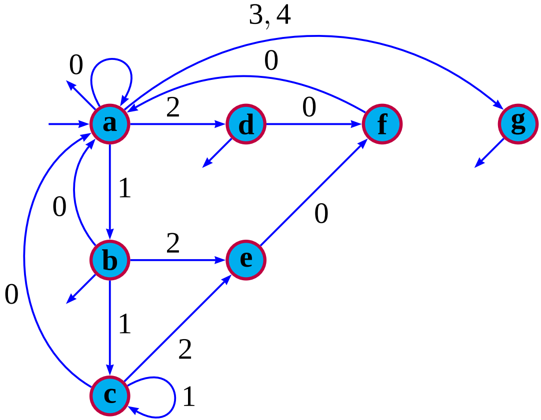
Our state complexity result thus applies to the class of Bertrand numeration systems.

Back to a previous example



- Let β be the largest root of $X^3 - 2X^2 - 1$.
- $\mathrm{d}_\beta(1) = 2010^\omega$ and $\mathrm{d}^*_\beta(1) = (200)^\omega$.
- This automaton accepts $\mathrm{rep}_U(\mathbb{N})$ 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 $(U_0, U_1, U_2) = (1, 5, 6)$.
- $\mathcal{A}_U \neq \mathcal{A}_\beta$

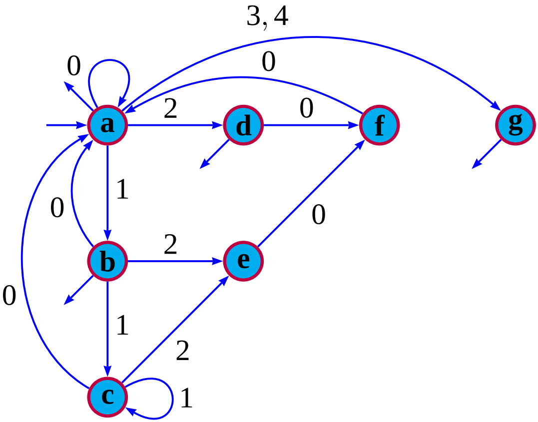
Relationship with $\mathcal{A}_\beta$

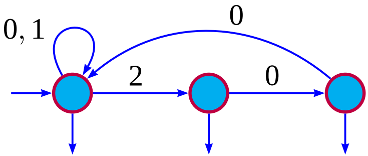
Theorem (cont'd.)

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 such that 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.

Further work

- Analyze the structure of $\mathcal{A}_U$ for systems with no dominant root.
- Remove the assumption that U is purely periodic in the state complexity result.
- Big open problem: Given an automaton accepting $\mathrm{rep}_U(X)$, is it decidable whether X is an ultimately periodic set?