

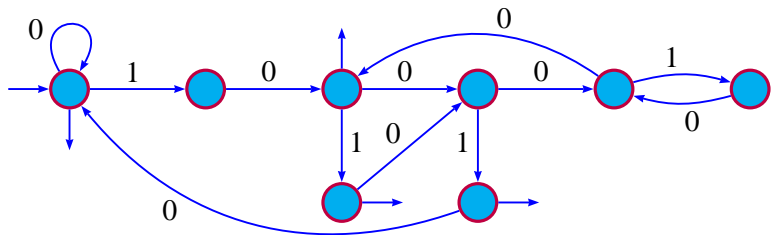
Structure of the minimal automaton of a numeration language

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

Département de mathématiques
Université de Liège

LaCIM 2010, Montréal, August 31st

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	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 $\text{rep}_F(2\mathbb{N}) = \{\varepsilon, 10, 101, 1001, 10000, \dots\}$ 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
 - ▶ $C_U := \sup_{n \geq 0} [U_{n+1}/U_n] < +\infty$.
- ▶ 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

$$\forall j, \sum_{i=0}^{j-1} w_i U_i < U_j.$$

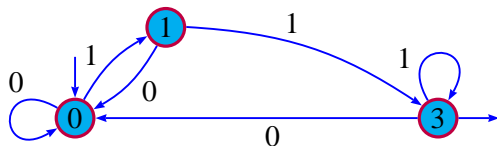
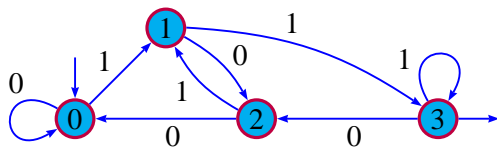
- ▶ In that case, $w \in \{0, 1, \dots, C_U - 1\}^*$.
- ▶ $\text{rep}_U(n)$ is the greedy representation of n with $w_{\ell-1} \neq 0$.
- ▶ $X \subseteq \mathbb{N}$ is *U-recognizable* if $\text{rep}_U(X)$ is accepted by a finite automaton.
- ▶ $\text{rep}_U(\mathbb{N})$ is the *numeration language*.

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?



DFA's 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 $0^* \text{rep}_U(p\mathbb{N} + r)$?

Background

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 $0^* \text{rep}_U(m\mathbb{N})$ 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 $0^* \text{rep}_U(\mathbb{N})$?

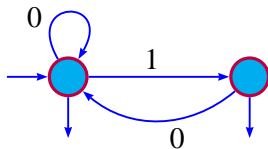
Information we are looking for

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

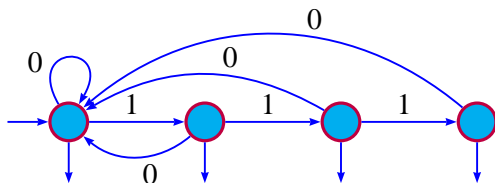
**Structure of the minimal automaton \mathcal{A}_U
recognizing $0^* \text{rep}_U(\mathbb{N})$**

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



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

First results

Theorem

Let U be a linear numeration system such that $\text{rep}_U(\mathbb{N})$ 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 $\lim_{n \rightarrow +\infty} U_{n+1} - U_n = +\infty$.
- (iv) If $\lim_{n \rightarrow +\infty} U_{n+1} - U_n = +\infty$, then $\delta_U(q_{U,0}, 1)$ is in \mathcal{C}_U .

Dominant root condition

- ▶ U satisfies the *dominant root condition* if $\lim_{n \rightarrow +\infty} U_{n+1}/U_n = \beta$ for some real $\beta > 1$.
- ▶ β is the *dominant root* of the recurrence.
- ▶ E.g., Fibonacci: dominant root $\beta = (1 + \sqrt{5})/2$

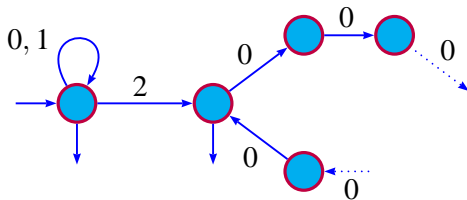
Theorem (cont'd.)

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

An example with two components

- ▶ Let $t \geq 1$.
- ▶ Let $U_0 = 1$, $U_{t+1} = 2U_t + 1$, and
- ▶ $U_{t+r} = 2U_{t+r-1}$, for $1 < r \leq t$.
- ▶ E.g., for $t = 2$ we have $U = (1, 3, 6, 13, 26, 53, \dots)$.
- ▶ Then $0^* \text{rep}_U(\mathbb{N}) = \{0, 1\}^* \cup \{0, 1\}^* 2(0^t)^*$.
- ▶ The second component is a cycle of t 0's.



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

With any Parry number β is associated a **canonical finite automaton** \mathcal{A}_β .

We will study the relationship between \mathcal{A}_U and \mathcal{A}_β .

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

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 $\mathbf{d}_\beta(x) := (x_i)_{i \geq 1}$ over $\{0, \dots, \lfloor \beta \rfloor\}$ satisfying

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

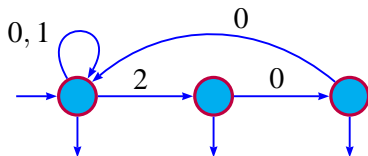
Parry numbers

- ▶ If $d_\beta(1) = t_1 \cdots t_m 0^\omega$, with $t_m \neq 0$, then $d_\beta(1)$ is finite.
- ▶ In this case $d_\beta^*(1) := (t_1 \cdots t_{m-1}(t_m - 1))^\omega$.
- ▶ For instance, $d_2(1) = 20^\omega$ and $d_2^*(1) = 1^\omega$.
- ▶ Otherwise $d_\beta^*(1) := d_\beta(1)$.
- ▶ If $d_\beta^*(1)$ is ultimately periodic, then β is a *Parry number*.

The Parry automaton

- ▶ Let $\text{Fact}(D_\beta)$ be the set of all words w lexicographically less than or equal to the prefix of $d_\beta^*(1)$ of length $|w|$.
- ▶ For β Parry, let \mathcal{A}_β be the minimal finite automaton accepting $\text{Fact}(D_\beta)$.

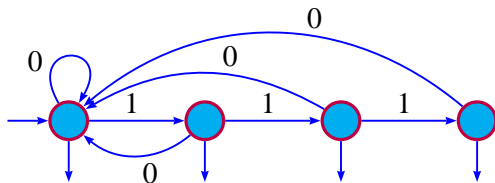
An example of the automaton \mathcal{A}_β



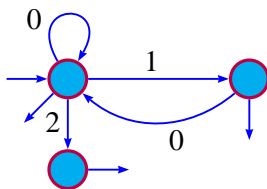
- ▶ Let β be the largest root of $X^3 - 2X^2 - 1$.
- ▶ $d_\beta(1) = 2010^\omega$ and $d_\beta^*(1) = (200)^\omega$.
- ▶ This automaton also accepts $0^* \text{rep}_U(\mathbb{N})$ for U defined by $U_{n+3} = 2U_{n+2} + U_n$, $(U_0, U_1, U_2) = (1, 3, 7)$.
- ▶ $\mathcal{A}_U = \mathcal{A}_\beta$

Bertrand numeration systems

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



A non-Bertrand system



- ▶ $U_{n+2} = U_{n+1} + U_n, (U_0 = 1, U_1 = 3)$
- ▶ $(U_n)_{n \geq 0} = 1, 3, 4, 7, 11, 18, 29, 47, \dots$
- ▶ 2 is a greedy representation but 20 is not.

Theorem (Bertrand)

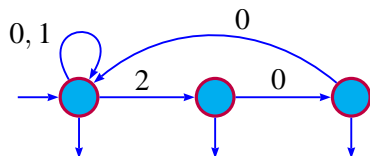
A system U is Bertrand if and only if there is a $\beta > 1$ such that

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

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

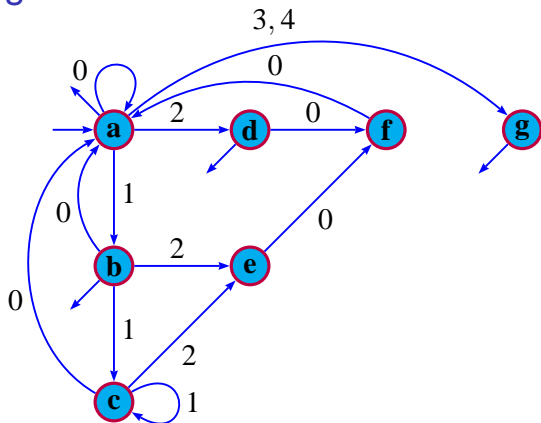
- ▶ If β is a Parry number, the system is linear and we have a minimal finite automaton \mathcal{A}_β accepting $\text{Fact}(D_\beta)$.
- ▶ Consequently, $\text{rep}_U(\mathbb{N})$ is regular and $\mathcal{A}_U = \mathcal{A}_\beta$.

Back to a previous example



- ▶ Let β be the largest root of $X^3 - 2X^2 - 1$.
- ▶ $d_\beta(1) = 2010^\omega$ and $d_\beta^*(1) = (200)^\omega$.
- ▶ This automaton accepts $0^* \text{rep}_U(\mathbb{N})$ 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 $(U_0, U_1, U_2) = (1, 5, 6)$.
- ▶ $\mathcal{A}_U \neq \mathcal{A}_\beta$

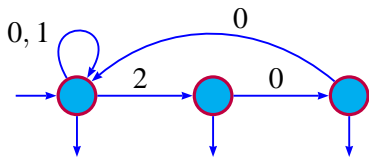
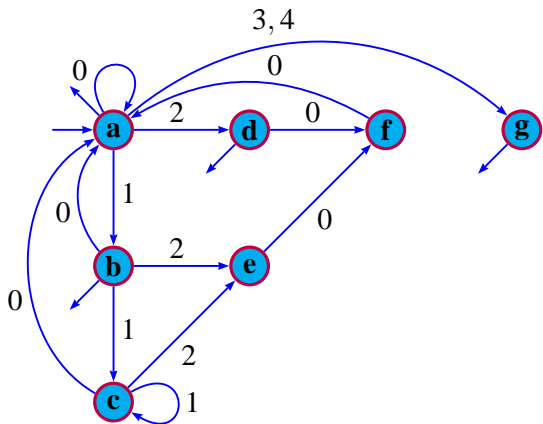
Relationship with \mathcal{A}_β

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}_β .

Φ maps the states of \mathcal{C}_U onto the states of \mathcal{A}_β so that

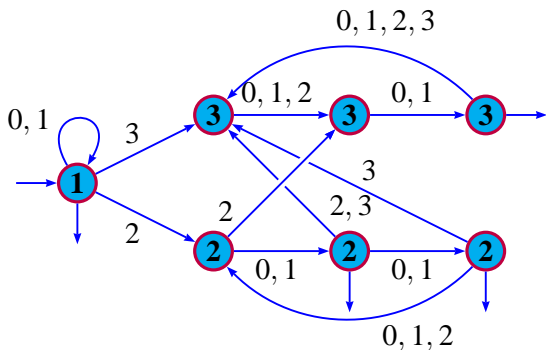
- ▶ $\Phi(q_{U,0}) = q_{\beta,0}$,
- ▶ for all states q and all letters σ such that q and $\delta_U(q, \sigma)$ are in \mathcal{C}_U , we have $\Phi(\delta_U(q, \sigma)) = \delta_\beta(\Phi(q), \sigma)$.



Other results

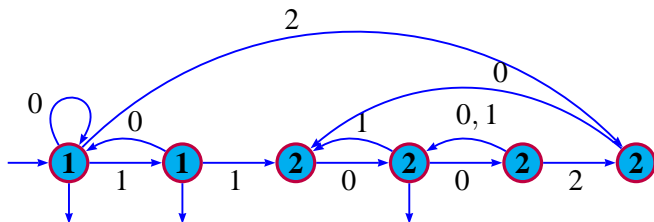
- ▶ When U has a dominant root $\beta > 1$, we can say more.
- ▶ E.g., if \mathcal{A}_U has more than one non-trivial strongly connected component, then $d_\beta(1)$ is finite.
- ▶ We can also give sufficient conditions for \mathcal{A}_U to have more than one non-trivial strongly connected component.
- ▶ 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

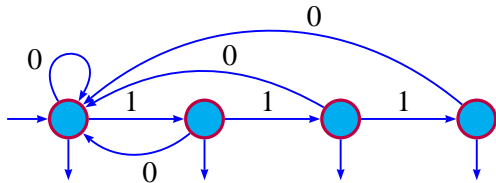
A system with no dominant root



- ▶ $U_{n+4} = 3U_{n+2} + U_n, (U_0, U_1, U_2, U_3) = (1, 2, 3, 7)$
- ▶ U_{n+1}/U_n doesn't converge, but
- ▶ $\lim_{n \rightarrow +\infty} U_{2n+2}/U_{2n} = \lim_{n \rightarrow +\infty} U_{2n+3}/U_{2n+1} = (3 + \sqrt{13})/2$

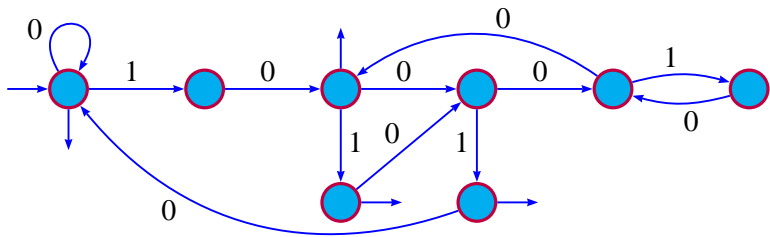
M. Hollander, Greedy numeration systems and regularity, *Theory Comput. Systems* **31** (1998).

Application to the state complexity of the ℓ -bonacci system



Corollary

For U the ℓ -bonacci numeration system, the number of states of the trim minimal automaton accepting $0^* \text{rep}_U(m\mathbb{N})$ 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	0	8
1	0	0	1	0	0	10
1	0	1	0	1	0	12
						⋮

Further work

- ▶ Analyze the structure of \mathcal{A}_U for systems with no dominant root.
- ▶ Remove the assumption that $(U_i \bmod m)_{i \geq 0}$ is purely periodic in the state complexity result that we have.
- ▶ Big open problem: Given an automaton accepting $\text{rep}_U(X)$, is it decidable whether X is an ultimately periodic set?