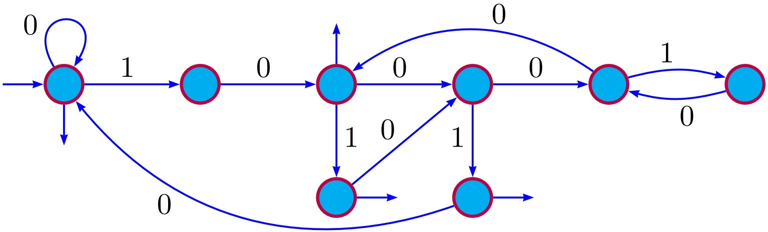
Syntactic complexity of recognizable sets

Émilie Charlier

Université libre de Bruxelles

1st Joint Conference of the Belgian, Royal Spanish and Luxembourg Mathematical Societies, June 2012, Liège

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
						⋮

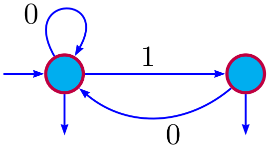
The set $2\mathbb{N}$ of even integers is *F-recognizable* or *F-automatic*, i.e., the language $\mathrm{rep}_F(2\mathbb{N}) = \{\varepsilon, 10, 101, 1001, 10000, \ldots\}$ is accepted by some finite automaton.

Remark (in terms of the Chomsky hierarchy)

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

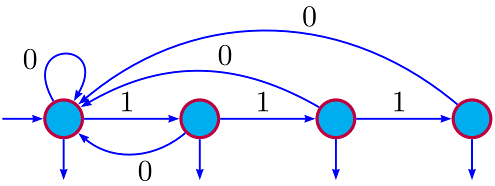
We get a similar definition for other numeration systems.

Zeckendorf (or Fibonacci) numeration system



- $F_{n+2} = F_{n+1} + F_n$
- $F_0 = 1,\ F_1 = 2$
- $\mathcal{A}_F$ 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.

U-recognizability of arithmetic progressions

Proposition

Let $U = (U_i)_{i\geq 0}$ *be a numeration system and let* $m, r \in \mathbb{N}$*.*

If $\mathbb{N}$ *is* U*-recognizable, then* $m\,\mathbb{N} + r$ *is* U*-recognizable and, given a DFA accepting* $\mathrm{rep}_U(\mathbb{N})$*, a DFA accepting* $\mathrm{rep}_U(m\,\mathbb{N} + r)$ *can be obtained effectively.*

Consequently, any ultimately periodic set is U*-recognizable.*

U-recognizability of $\mathbb{N}$

Is the set $\mathbb{N}$ U-recognizable? Otherwise stated, is the numeration language $\mathrm{rep}_U(\mathbb{N})$ 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 $\mathbb{Z}$.

The condition is *not* sufficient:

Example ($U_i = (i+1)^2$ for all $i \in \mathbb{N}$)

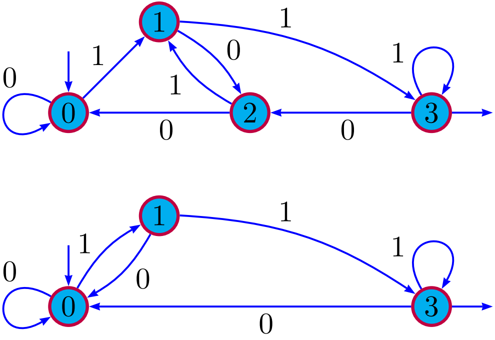
It is linear: $U_{i+3} = 3U_{i+2} - 3U_{i+1} + U_i$ for all $i \in \mathbb{N}$, but:

$$\begin{aligned}\mathrm{rep}_U(\mathbb{N}) \cap 10^*10^* &= \{10^a10^b \colon U_{a+b+1} + U_b < U_{a+b+2}\}\\ &= \{10^a10^b \colon b^2 < 2a+4\}\end{aligned}$$

Thus, $\mathrm{rep}_U(\mathbb{N})$ is not regular.

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 $\mathrm{rep}_U(m\mathbb{N}+r)$?

A general upper bound

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

Let $m, r \in \mathbb{N}$ *with* $m \geq 2$ *and* $r < m$.
If $\mathrm{rep}_U(\mathbb{N})$ *is accepted by a* n*-state DFA, then the minimal automaton of* $\mathrm{rep}_U(m\mathbb{N} + r)$ *has at most* $n\, m^n$ *states.*

NB: This result remains true for the larger class of *abstract numeration systems*.

Integer base case

Theorem (Alexeev 2004)

Let $b, m \geq 2$. *Let* N, M *be such that* $b^N < m \leq b^{N+1}$ *and*
$(m,1) < (m,b) < \cdots < (m,b^M) = (m,b^{M+1})$.
The minimal automaton recognizing $m\mathbb{N}$ *in base* b *has exactly*

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

In particular, if m and b are coprime, then this number is just m.
Further, if $m = b^n$, then this number is $n+1$.

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).
- A. Muchnick, The definable criterion for definability in Presburger arithmetic and its applications, *TCS* **290** (2003).
- J. Leroux, A Polynomial Time Presburger Criterion and Synthesis for Number Decision Diagrams, *LICS 2005* (2005).
- J.-P. Allouche, N. Rampersad, J. Shallit, Periodicity, repetitions, and orbits of an automatic sequence, *TCS* **410** (2009).
- J. Bell, ÉC, A. Fraenkel, M. Rigo, A decision problem for ultimately periodic sets in non-standard numeration systems, *IJAC* **19** (2009).
- F. Durand, Decidability of the HD0L ultimate periodicity problem, *arXiv* (2011).
- I. Mitrofanov, A proof for the decidability of HD0L ultimate periodicity, *arXiv* (2011).

Information we are looking for

Consider a linear numeration system U such that $\mathbb{N}$ is U-recognizable.

How many states does the trim minimal automaton $\mathcal{A}_{U,m}$ recognizing $m\,\mathbb{N}$ contain?

1. Give upper/lower bounds?
2. Study special cases, e.g., Zeckendorf numeration system.
3. Get information on the trim minimal automaton $\mathcal{A}_U$ recognizing $\mathbb{N}$.

A lower bound

Theorem (C-Rampersad-Rigo-Waxweiler 2011)

Let U be any numeration system (not necessarily linear). The number of states of $\mathcal{A}_{U,m}$ is at least $|\operatorname{rep}_U(m)|$.

The Hankel matrix

- Let $U = (U_n)_{n\geq 0}$ be a linear numeration system.
- Let $k = k_{U,m}$ be the length of the shortest linear recurrence relation satisfied by $(U_i \bmod m)_{i\geq 0}$.
- 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 $m \geq 2$, $k_{U,m}$ is also the largest t such that $\det H_t \not\equiv 0 \pmod{m}$.

A system of linear congruences

- 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 $\mathbf{x} = (x_1, \ldots, x_k)$.

- $S_{U,m} \leq m^k$.

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 $S_{U,4} = 8$.

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 $\mathcal{A}_{U,m}$ from which infinitely many words are accepted is

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

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

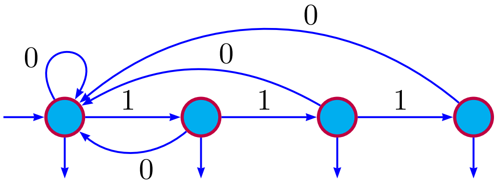
(H.2) For all states p, q in $\mathcal{C}_U$ with $p \neq q$, there exists a word x_{pq} such that $\delta_U(p, x_{pq}) \in \mathcal{C}_U$ and $\delta_U(q, x_{pq}) \notin \mathcal{C}_U$, or vice-versa.

Result for strongly connected automata

Corollary

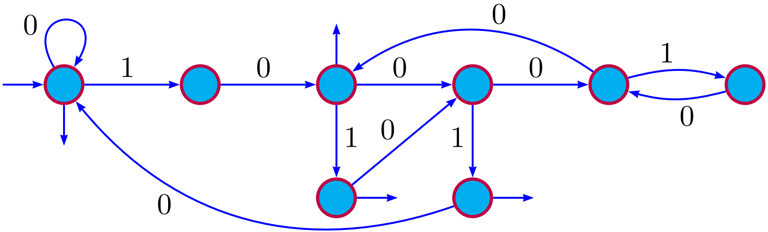
If U satisfies the conditions of the previous theorem and $\mathcal{A}_U$ is strongly connected, then the number of states of $\mathcal{A}_{U,m}$ is $|\mathcal{A}_U|\, S_{U,m}$.

Result for the ℓ-bonacci system



Corollary

For U the ℓ-bonacci system, the number of states of $\mathcal{A}_{U,m}$ 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
						⋮

Further work for state complexity

- Analyze the structure of $\mathcal{A}_U$ for systems with no dominant root.
- Remove the assumption that $(U_n \bmod m)_{n\geq 0}$ is purely periodic in the state complexity result.
- Look at any arithmetic progressions $X = m\,\mathbb{N}+r$.

Transition to syntactic complexity

Let $N_U(m) \in \{1, \ldots, m\}$ denote the number of values that are taken infinitely often by the sequence $(U_i \mod m)_{i\geq 0}$.

Example (Zeckendorf system)

$(F_i \mod 4) = (1, 2, 3, 1, 0, 1, 1, 2, 3, \ldots)$ and $N_F(4) = 4$.
$(F_i \mod 11) = (1, 2, 3, 5, 8, 2, 10, 1, 0, 1, 1, 2, 3, \ldots)$ and $N_F(11) = 7$.

Theorem (C-Rigo 2008)

Let $U = (U_i)_{i\geq 0}$ be a numeration system satisfying
$$\lim_{i\to+\infty} U_{i+1} - U_i = +\infty.$$
If $X \subseteq \mathbb{N}$ is an ultimately periodic U-recognizable set of period p, then any DFA accepting $\mathrm{rep}_U(X)$ has at least $N_U(p)$ states.

- If $N_U(m) \to +\infty$ as $m \to +\infty$, then we obtain a decision procedure to the periodicity problem.
- If U is a LNS satisfying

$$U_{i+k} = a_1 U_{i+k-1} + \cdots + a_k U_i, \; i \geq 0, \quad \text{with} \quad a_k = \pm 1,$$

  then $\lim_{m \to +\infty} N_U(m) = +\infty$.
- Works for the Zeckendorf system.
- Not true for integer base b: $N(b^n) = 1$ for all $n \geq 0$.

- The formula for the state complexity of $m\mathbb{N}$ for the Zeckendorf system is much simpler than the formula for integer base b systems.
- In this point of view, state complexity is not completely satisfying.
- Hope: Find a complexity that would handle all these systems in a kind of uniform way.

Syntactic complexity

- Let L be a language over the finite alphabet Σ.
- Myhill-Nerode equivalence relation for L: $u \sim_L v$ means that for all $y \in \Sigma^*$, $uy \in L \Leftrightarrow vy \in L$.
- Leads to the minimal automaton of L: $|\mathcal{A}_L| = |\Sigma^*/{\sim_L}|$ is the state complexity of L.
- Syntactic congruence for L: $u \equiv_L v$ means that for all $x, y \in \Sigma^*$, $xuy \in L \Leftrightarrow xvy \in L$.
- Leads to the syntactic monoid of L: $|\mathcal{H}_L| = |\Sigma^*/{\equiv_L}|$ is the syntactic complexity of L.

Theorem

A language L is regular if and only if $\Sigma^*/{\equiv_L}$ is finite.

Syntactic complexity for integer bases

The syntactic complexity of $X \subseteq \mathbb{N}$ is the syntactic complexity of the language $0^* \operatorname{rep}_U(X)$.

Let $\operatorname{ord}_m(b) = \min\{j \in \mathbb{N}_0 \colon b^j \equiv 1 \pmod{m}\}$.

Theorem (Rigo-Vandomme 2011)

- *Let $m, b \geq 2$ be coprime integers.*
  If $X \subseteq \mathbb{N}$ is periodic of minimal period m, then the syntactic complexity of X is equal to $m \operatorname{ord}_m(b)$.

Theorem (continued)

- *Let* $b \geq 2$ *and* $m = b^n$ *with* $n \geq 1$.
  (a) *The syntactic complexity of* $m\mathbb{N}$ *is equal to* $2n + 1$.
  (b) *If* $X \subseteq \mathbb{N}$ *is periodic of minimal period* m, *then the syntactic complexity of* X *is* $\geq n + 1$.
- *Let* $b \geq 2$ *and* $m = b^n q$ *with* $n \geq 1$ *and* $(b, q) = 1$.
  Then the syntactic complexity of $m\mathbb{N}$ *is equal to*
  $(n + 1)\, q \operatorname{ord}_q(b) + n$.

A general lower bound for the integer base case

Theorem (Lacroix-Rampersad-Rigo-Vandomme, to appear)

Let $b \geq 2$ *and* $m = db^n q$ *with* $n \geq 1$ *and* $(b, q) = 1$ *and where* n *and* q *are chosen to be maximal.*
If $X \subseteq \mathbb{N}$ *is periodic of minimal period* m*, then the syntactic complexity of* X *is*

$$\geq \max\left(q \operatorname{ord}_q(b), \frac{\gamma + 1}{q \operatorname{ord}_q(b)}\right),$$

where $\gamma \to +\infty$ *as* n *or* $d \to +\infty$*.*

Zeckendorf numeration system and further work

Theorem

The syntactic complexity of $m\,\mathbb{N}$ *is*

$$4m^2 p_F(m) + 2$$

where $p_F(m)$ *is the minimal period of* $(F_i \bmod m)_{i \geq 0}$.

Further work for syntactic complexity:

- Try to estimate the syntactic complexity of periodic sets for a larger class of numeration systems.

Syntactic complexity seems to allow us to handle integer bases and the Zeckendorf system at once.