The state complexity of testing divisibility

Narad Rampersad

Department of Mathematics

University of Liège

Joint work with: É. Charlier, M. Rigo, L. Waxweiler

Representing numbers in base 2

- In base 2, we expand using powers of 2:

$$13 = 1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0.$$

- The representation of 13 in base 2 is 1101.

Representing numbers using other sequences

- Suppose we expand using the terms of the Fibonacci sequence:

$$13 = 1 \cdot 13 + 0 \cdot 8 + 0 \cdot 5 + 0 \cdot 3 + 0 \cdot 2 + 0 \cdot 1.$$

- The representation of 13 in the Fibonacci system is 100000.
- 13 also has the representation 11000.

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 $\mathbb{Z}$.

Greedy representations

- A greedy representation of a non-negative integer n is a word $w = w_{\ell-1} \cdots w_0$ over $\{0, 1, \ldots, C_U - 1\}$ such that

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

and for all j

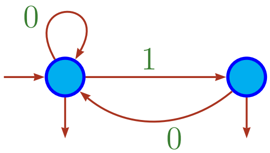
$$\sum_{i=0}^{j-1} w_i U_i < U_j.$$

- $\mathrm{rep}_U(n)$ is the greedy representation of n with $w_{\ell-1} \neq 0$.

Numeration languages recognized by automata

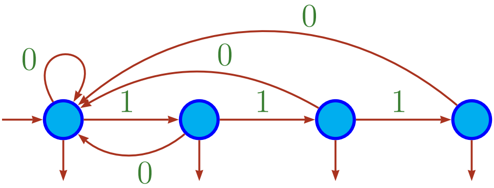
- Suppose that the language $\mathrm{rep}_U(\mathbb{N})$ of greedy representations is a regular language.
- Let $\mathscr{A}_U$ be the minimal automaton accepting $0^* \mathrm{rep}_U(\mathbb{N})$.
- $\mathscr{A}_U = (Q_U, \{0, \ldots, C_U - 1\}, \delta_U, q_{U,0}, F_U)$

The Fibonacci numeration system



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

The ℓ-bonacci numeration system



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

Recognizable sets

- A set X of integers is U-recognizable if $\mathrm{rep}_U(X)$ is accepted by a finite automaton.
- If $\mathbb{N}$ is U-recognizable, then U is linear.
- The converse is not true in general.

Recognizing arithmetic progressions

Theorem (Lecomte and Rigo 2001)

Let L be a regular language ordered first by length and then lexicographically. The language obtained by extracting from L those words whose indices belong to an ultimately periodic set is regular.

In particular, if $\mathbb{N}$ is U-recognizable then so is $m\,\mathbb{N}$.

State complexity

Theorem (Krieger, Miller, R., Ravikumar, Shallit 2009)

If L accepted by an n-state DFA, then the minimal DFA accepting the language of words of L indexed by the multiples of m has at most nm^n states.

An exact result for the integer bases

Theorem (Alexeev 2004)

Let $\lambda(x, y) = \frac{x}{\gcd(x,y)}$. The number of states of the minimal automaton accepting the base b representations of the multiples of m is

$$\lambda(m, b^A) + \sum_{i=0}^{A-1} \lambda(b^i, m),$$

where A is the least non-negative integer i for which $\lambda(m, b^i) - \lambda(m, b^{i+1}) < \lambda(b^i, m)$.

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 $k_{U,2} = 1$.
- Mod 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 $S_{U,m}$

- $U_{n+2} = 2U_{n+1} + U_n$, $(U_0, U_1) = (1, 3)$
- 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) $\mathscr{A}_U$ has a single strongly connected component $\mathscr{C}_U$.

(H.2) For all states p, q in $\mathscr{C}_U$ with $p \neq q$, there exists a word x_{pq} such that $\delta_U(p, x_{pq}) \in \mathscr{C}_U$ and $\delta_U(q, x_{pq}) \notin \mathscr{C}_U$, 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 $\mathscr{A}_U$ satisfies (H.1) and (H.2),

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

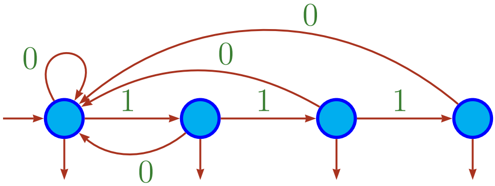
The number of states of the trim minimal automaton accepting $0^* \operatorname{rep}_U(m\mathbb{N})$ from which infinitely many words are accepted is $|\mathscr{C}_U| S_{U,m}$.

Result for strongly connected automata

Corollary

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

Result for the ℓ-bonacci system



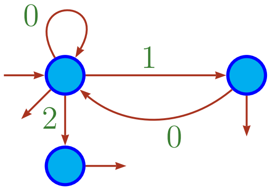
Corollary

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.

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.

β-expansions

- Bertrand systems are associated with 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 $\mathsf{d}_\beta(x) := (t_i)_{i \geq 1}$ over $\{0, \ldots, \lceil \beta \rceil - 1\}$ satisfying

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

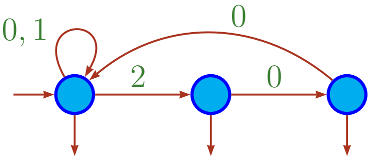
Parry numbers

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

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

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



- 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 $\mathrm{rep}_U(\mathbb{N})$ for U defined by $U_{n+3} = 2U_{n+2} + U_n, (U_0, U_1, U_2) = (1, 3, 7)$.

Characterization of Bertrand systems

Theorem (Bertrand)

A system U is Bertrand if and only if there is a $\beta > 1$ such that $0^* \operatorname{rep}_U(\mathbb{N}) = \operatorname{Fact}(D_\beta)$ (that is, $\mathscr{A}_U = \mathscr{A}_\beta$).

Applying our 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 $\mathscr{A}_U$ 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.

Further work

- 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?

The End