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

\textbf{Remark (in terms of the Chomsky hierarchy)}

With respect to the Zeckendorf system, \textit{any} $F$-recognizable set can be considered as a “\textit{particularly simple}” set of integers.

We get a similar definition for \textit{other} numeration systems.
Zeckendorf (or Fibonacci) numeration system

- \( F_{n+2} = F_{n+1} + F_n \)
- \( F_0 = 1, \ F_1 = 2 \)
- \( A_F \) accepts all words that do not contain 11.
The $\ell$-bonacci numeration system

- $U_{n+\ell} = U_{n+\ell-1} + U_{n+\ell-2} + \cdots + U_n$
- $U_i = 2^i$, $i \in \{0, \ldots, \ell - 1\}$
- $A_U$ accepts all words that do not contain $1^\ell$. 
$U$-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 $\text{rep}_U(\mathbb{N})$, a DFA accepting $\text{rep}_U(m \mathbb{N} + r)$ can be obtained effectively.

Consequently, any ultimately periodic set is $U$-recognizable.
\textit{U}-recognizability of $\mathbb{N}$

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

\textbf{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 \textit{not} sufficient:

\textbf{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:

$$\text{rep}_U(\mathbb{N}) \cap 10^*10^* = \{10^a10^b : U_{a+b+1} + U_b < U_{a+b+2}\}$$

$$= \{10^a10^b : b^2 < 2a + 4\}$$

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

What is the “best automaton” we can get?

DFAs accepting the binary representations of $4N + 3$.

Question

The general algorithm doesn’t provide a minimal automaton. What is the state complexity of $\text{rep}_U(mN + r)$?
A general upper bound


Let $m, r \in \mathbb{N}$ with $m \geq 2$ and $r < m$.

If $\text{rep}_U(\mathbb{N})$ is accepted by a $n$-state DFA, then the minimal automaton of $\text{rep}_U(m\mathbb{N} + r)$ 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

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

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.

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 $A_{U,m}$ 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 $A_U$ 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 $A_{U,m}$ is at least $|\text{rep}_U(m)|$. 
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 $(U_i \mod m)_{i \geq 0}$.
- 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 $b$ in $\{0, \ldots, m - 1\}^k$ such that the system

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

has at least one solution $x = (x_1, \ldots, x_k)$.

$S_{U,m} \leq m^k$. 
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
  \[
  \begin{align*}
  1x_1 + 3x_2 & \equiv b_1 \pmod{4} \\
  3x_1 + 7x_2 & \equiv b_2 \pmod{4}
  \end{align*}
  \]
- $2x_1 \equiv b_2 - b_1 \pmod{4}$
- 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 $A_U$ satisfies (H.1) and (H.2),
(b) $(U_n \mod m)_{n \geq 0}$ is purely periodic.

The number of states of $A_{U,m}$ from which infinitely many words are accepted is

$$|C_U| S_{U,m}.$$ 

(H.1) $A_U$ has a single strongly connected component $C_U$.

(H.2) For all states $p, q$ in $C_U$ with $p \neq q$, there exists a word $x_{pq}$ such that $\delta_U(p, x_{pq}) \in C_U$ and $\delta_U(q, x_{pq}) \notin C_U$, 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 $\ell$-bonacci system

Corollary
For $U$ the $\ell$-bonacci system, the number of states of $A_{U,m}$ is $\ell m^\ell$. 
Further work for state complexity

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

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

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 $\Sigma$.
- Myhill-Nerode equivalence relation for $L$: $u \sim_L v$ means that for all $y \in \Sigma^*$, $uy \in L \iff vy \in L$.
- Leads to the minimal automaton of $L$: $|A_L| = |\Sigma^*/\sim_L|$ is the state complexity of $L$.
- Syntactic congruence for $L$: $u \equiv_L v$ means that for all $x, y \in \Sigma^*$, $xuy \in L \iff xvy \in L$.
- 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 \( 0^* \text{rep}_U(X) \).

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

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 \( m \text{ord}_m(b) \).
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 $(n + 1) q \text{ord}_q(b) + n$. 
A general lower bound for the integer base case

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

Let \( b \geq 2 \) and \( m = db^nq \) 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

\[
\geq \max \left( q \text{ord}_q(b), \frac{\gamma + 1}{q \text{ord}_q(b)} \right),
\]

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* \((F_i \mod m)_{i \geq 0}\).*

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.