STATE COMPLEXITY OF TESTING
DIVISIBILITY...

É. Charlier, N. Rampersad, M. Rigo, L. Waxweiler
(University of Liège)

http://www.discmath.ulg.ac.be/

Numeration 2010 – Leiden, 18th June 2010
What is this talk about?

13 8 5 3 2 1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The set $2\mathbb{N}$ of even integers is $F$-recognizable or $F$-automatic, i.e., the language $\text{rep}_F(2\mathbb{N}) = \{10, 101, 1001, 10000, \ldots\}$ is accepted by some finite automaton.

**Remark (in terms of Chomsky’s 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.
A numeration system is an increasing sequence of integers $U = (U_n)_{n \geq 0}$ such that

- $U_0 = 1$ and
- $C_U := \sup_{n \geq 0} \left\lceil \frac{U_{n+1}}{U_n} \right\rceil < \infty$.

$U$ is linear if it satisfies a linear recurrence relation over $\mathbb{Z}$, for all large enough $n$.

**Example**

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

Let $A \subset \mathbb{Z}$ be an alphabet.
For any word $w = w_{\ell-1} \cdots w_0$ over $A$, we set

$$\text{val}_{A,U}(w) := \sum_{i=0}^{\ell-1} w_i U_i.$$
A greedy representation of a non-negative integer $n$ is a word $w = w_{\ell-1} \cdots w_0$ over $A_U = \{0, 1, \ldots, C_U - 1\}$ such that

$$\text{val}_U(w) = \sum_{i=0}^{\ell-1} w_i U_i = n,$$

and for all $j$

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

$\text{rep}_U(n)$ 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.
• 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.
• Ultimately periodic sets are still *$U$*-recognizable for any numeration system *$U$* such that $\mathbb{N}$ is *$U$*-recognizable.

**Motivations**

**Proposition**

Let $p, r \geq 0$. If $(U_n)_{n \geq 0}$ is a linear numeration system, then

$$\text{val}_{A_U, U}^{-1}(p \mathbb{N} + r) = \left\{ c_\ell \cdots c_0 \in A_U^* \mid \sum_{k=0}^{\ell} c_k U_k \in p \mathbb{N} + r \right\}$$

is accepted by a DFA that can be effectively constructed. In particular, if $\mathbb{N}$ is $U$-recognizable, then any ultimately periodic set is $U$-recognizable. The proof is effective.

A DFA accepting $\text{rep}(4\mathbb{N} + 3)$. 

![DFA Diagram](image-url)
Main question

What is the “best automaton” we can get?

In general, the derived algorithm does not provide a minimal automaton. What is the state complexity of the minimal automaton accepting $\text{rep}_U(p\mathbb{N} + r)$?
BACKGROUND (I)

**Honkala’s Decision Procedure**

Given any finite automaton recognizing a set $X$ of integers written in base $b$. It is (algorithmically) decidable whether or not $X$ is ultimately periodic.


ALEXEEV’ 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 of $0^* \text{rep}_b(m \mathbb{N})$ has exactly

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

states.

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

1. Give upper/lower bounds?
2. Study special cases, e.g., Fibonacci numeration system?
3. Get informations on the minimal automaton $A_U$ recognizing $0^* \text{rep}_U(\mathbb{N})$?
Structure of the minimal automaton $A_U$ recognizing $0^* \text{rep}_U(\mathbb{N})$
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 Fibonacci system is Bertrand.

**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 $\beta$-development of 1.

If $\beta$ is a Parry number, the system is linear and we have a minimal finite automaton $A_\beta$ accepting $\text{Fact}(D_{\beta})$. 
THEOREM

Let $U$ be a linear numeration system such that $\text{rep}_U(\mathbb{N})$ is regular.

(i) The automaton $A_U$ has a non-trivial strongly connected component $C_U$ containing the initial state.

(ii) If $p$ is a state in $C_U$, 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 $C_U$ by reading a 0.
The Fibonacci numeration system

- $U_{n+2} = U_{n+1} + U_n$ ($U_0 = 1$, $U_1 = 2$)
- $A_U$ 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 $1^\ell$. 
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.
Let $\beta$ be the largest root of $X^3 - 2X^2 - 1$.

$d_\beta(1) = 2010^\omega$ and $d^*_\beta(1) = (200)^\omega$.

This automaton accepts $\text{rep}_U(\mathbb{N})$ for $U$ defined by

$U_{n+3} = 2U_{n+2} + U_n, (U_0, U_1, U_2) = (1, 3, 7)$. 
$U_{n+3} = 2U_{n+2} + U_n$, $(U_0, U_1, U_2) = (1, 3, 7)$
we change the initial values to $(U_0, U_1, U_2) = (1, 5, 6)$. 
(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 $\mathcal{C}_U$. 
- $U$ satisfies the dominant root condition if $\lim_{n \to \infty} \frac{U_{n+1}}{U_n} = \beta$ for some real $\beta > 1$.
- $\beta$ 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 $A_U$ has more than one non-trivial strongly connected component, then any such component other than $C_U$ is a cycle all of whose edges are labeled $0$. 
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 \( 0^* \text{rep}_U(\mathbb{N}) = \{0, 1\}^* \cup \{0, 1\}^*2(0^t)^* \).
- The second component is a cycle of \( t \) 0's.
Suppose $U$ has a dominant root $\beta > 1$. There is a morphism of automata $\Phi$ from $C_U$ to $A_\beta$.

$\Phi$ maps the states of $C_U$ onto the states of $A_\beta$ so that

- $\Phi(q_{U,0}) = q_{\beta,0}$,
- for all states $q$ and all letters $\sigma$ such that $q$ and $\delta_U(q, \sigma)$ are in $C_U$, we have $\Phi(\delta_U(q, \sigma)) = \delta_\beta(\Phi(q), \sigma)$. 
When $U$ has a dominant root $\beta > 1$, we can say more.

E.g., if $\mathcal{A}_U$ has more than one strongly connected component, then $d_\beta(1)$ is finite.

We can also give sufficient conditions for $\mathcal{A}_U$ to have only one strongly connected component and sufficient conditions for $\mathcal{A}_U$ to have more than one strongly connected component.

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 strongly connected components
A SYSTEM WITH NO DOMINANT ROOT

\( U_{n+4} = 3U_{n+2} + U_n, (U_0, U_1, U_2, U_3) = (1, 2, 3, 7) \)

\( U_{n+1}/U_n \) does not converge, but

\[
\lim_{n \to \infty} \frac{U_{2n+2}}{U_{2n}} = \lim_{n \to \infty} \frac{U_{2n+3}}{U_{2n+1}} = (3 + \sqrt{13})/2
\]

Back to state complexity issues
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 $k_{U,m}$

- $U_{n+2} = 2U_{n+1} + U_n$, $(U_0, U_1) = (1, 3)$
- $(U_n)_{n \geq 0} = 1, 3, 7, 17, 41, 99, 239, \ldots$
- $(U_n \mod 2)_{n \geq 0}$ 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 $k_{U,4} = 2$. 
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

\[
\begin{align*}
1x_1 + 3x_2 & \equiv b_1 \pmod{4} \\
3x_1 + 7x_2 & \equiv b_2 \pmod{4}
\end{align*}
\]

- $2x_1 \equiv b_2 - b_1 \pmod{4}$
- For each value of $b_1$ there are at most 2 values for $b_2$.
- Hence $S_{U,4} = 8$. 
(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.
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 the trim minimal automaton accepting $0^* \text{rep}_U(m\mathbb{N})$ from which infinitely many words are accepted is

$$|C_U|S_{U,m}.$$
RESULT FOR STRONGLY CONNECTED AUTOMATA

**Corollary**

If $U$ satisfies the conditions of the previous theorem and $A_U$ is strongly connected, then the number of states of the trim minimal automaton accepting $0^* \text{rep}_U(m\mathbb{N})$ is $|C_U|S_{U,m}$. 
**Result for the \( \ell \)-bonacci system**

**Corollary**

For \( U \) the \( \ell \)-bonacci numeration system, the number of states of the trim minimal automaton accepting \( 0^* \text{rep}_U(m\mathbb{N}) \) is \( \ell m^\ell \).
FURTHER WORK

- Analyze the structure of $A_U$ 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 $\text{rep}_U(X)$, is it decidable whether $X$ is an ultimately periodic set?
What I should not forget!