

Positional Numeration Systems : Ultimate Periodicity, Complexity and Automatic Sequences

Adeline Massuir

Liège - April 30, 2021

$$456 = 4 \times 10^{2} + 5 \times 10 + 6 \times 1$$

= 1 × 3⁵ + 2 × 3⁴ + 1 × 3³ + 2 × 3² + 2 × 3 + 0 × 1

$\begin{array}{l} 456 = 4 \times 10^2 + 5 \times 10 + 6 \times 1 \\ = 1 \times 3^5 + 2 \times 3^4 + 1 \times 3^3 + 2 \times 3^2 + 2 \times 3 + 0 \times 1 \end{array}$

$$8 + 2 = 10$$



recognizability

 \swarrow

automatic sequences

3/39

Part 1

Ultimate periodicity problem for linear numeration systems

Problem

Given a linear numeration system U and a deterministic finite automaton \mathscr{A} whose accepted language is contained in the numeration language rep_U(\mathbb{N}), decide whether the subset X of \mathbb{N} that is recognized by \mathscr{A} is ulti-

mately periodic, i.e. whether or not X is a finite union of arithmetic progressions (along a finite set).

Problem

Given a linear numeration system U and a deterministic finite automaton \mathscr{A} whose accepted language is contained in the numeration language rep_U(\mathbb{N}), decide whether the subset X of \mathbb{N} that is recognized by \mathscr{A} is ultimately periodic, i.e. whether or not X is a finite union of arithmetic progressions (along a finite set).

1986 : Honkala 1994 : Bruyère, Hansel, Michaux, Villemaire 2009 : Bell, Charlier, Fraenkel, Rigo

Problem

Given a linear numeration system U and a deterministic finite automaton \mathscr{A} whose accepted language is contained in the numeration language rep_U(\mathbb{N}), decide whether the subset X of \mathbb{N} that is recognized by \mathscr{A} is ultimately periodic, i.e. whether or not X is a finite union of arithmetic progressions (along a finite set).

1986 : Honkala 1994 : Bruyère, Hansel, Michaux, Villemaire 2009 : Bell, Charlier, Fraenkel, Rigo

Syntactic complexity : Lacroix, Rampersad, Rigo, Vandomme (2012) Logic : Muchnik (2003) – Leroux (2005) Automata : Marsault, Sakarovitch (2013) – Boigelot, Mainz, Marsault, Rigo (2017) – Marsault (2019) Morphic : Durand (2013) – Mitrofanov (2013) (H1) \mathbb{N} is *U*-recognizable,

(H2)
$$\limsup_{i\to+\infty} (U_{i+1} - U_i) = +\infty$$
,

(H3)
$$\exists N \geq 0, \forall i \geq 0, U_{i+1} - U_i \leq U_{i+2} - U_{i+1}.$$

- We are able to check with an automaton whether a representation is greedy,
- the numeration system is linear,
- ultimately periodic sets are recognizable.

- We are able to check with an automaton whether a representation is greedy,
- the numeration system is linear,
- ultimately periodic sets are recognizable.

Let U be a numeration system satisfying (H1), (H2) and (H3). There exists a constant Z such that if w is a greedy U-representation, then for all $z \ge Z$, $10^z w$ is also a greedy U-representation.

Example 1

Consider the numeration system $U_{i+4} = 2U_{i+3} + 2U_{i+2} + 2U_i$ with initial conditions $U_0 = 1$, $U_1 = 3$, $U_2 = 9$, $U_3 = 25$. The largest root is $\beta \approx 2.804$ and it has also a root $\gamma \approx -1.134$.

Example 2

Consider the numeration system $U_{i+3} = 12U_{i+2} + 6U_{i+1} + 12U_i$ with initial conditions $U_0 = 1$, $U_1 = 13$, $U_2 = 163$.

- Input : DFA \mathscr{A}
- Upper bound on the admissible preperiods and periods
- For each pair (*N*, *p*) of possible preperiods and periods, there are at most 2^{*N*}2^{*p*} corresponding ultimately periodic sets *X*
- Equality test : \mathscr{A}_X and \mathscr{A}

Let $U = (U_i)_{i \in \mathbb{N}}$ satisfying (H1), (H2) and (H3).

$$U_{i+k} = a_{k-1}U_{i+k-1} + \cdots + a_0U_i, \ a_0 \neq 0$$

Suppose that the minimal automaton \mathscr{A}_X of $\operatorname{rep}_U(X)$ is given. Let π_X be a potential period for X and consider its prime decomposition.

- Factors that do not divide a_0 ,
- **(a)** factors that divide a_0 but not all the a_i ,
- factors that divide all the a_i .

Proposition

Let $X \subseteq \mathbb{N}$ an ultimately periodic *U*-recognizable set and let q be a divisor of π_X such that $(q, a_0) = 1$. Then the minimal automaton of rep_{*U*}(X) has at least q states.

Proposition

Let p be a prime not dividing all the coefficients of the recurrence relation and let $\lambda \ge 1$ be the least integer such that $(U_i \mod p^{\lambda})_{i \in \mathbb{N}}$ has a period containing a non-zero element. If $X \subseteq \mathbb{N}$ is an ultimately periodic U-recognizable set with period $\pi_X = p^{\mu} \cdot r$ where $\mu \ge \lambda$ and r is not divisible by p, then the minimal automaton of rep $_U(X)$ has at least $p^{\mu-\lambda+1}$ states.

Theorem

Let U be a numeration system satisfying (H1), (H2) and (H3) and such that the gcd of the coefficients of the recurrence relation of U is 1.

Given a DFA \mathscr{A} accepting a language contained in the numeration language rep_U(\mathbb{N}), it is decidable whether this DFA recognizes an ultimately periodic set.

Theorem

Let U be a numeration system satisfying (H1), (H2) and (H3) and such that the gcd of the coefficients of the recurrence relation of U is 1.

Given a DFA \mathscr{A} accepting a language contained in the numeration language rep_U(\mathbb{N}), it is decidable whether this DFA recognizes an ultimately periodic set.

$$U_{i+5} = 6U_{i+4} + 3U_{i+3} - U_{i+2} + 6U_{i+1} + 3U_i, \forall i \ge 0$$

with $U_0 = 1, U_1 = 7, U_2 = 45, U_3 = 291, U_5 = 1881$

Prime factors that divide all the coefficients

$$\pi_x = m_X \cdot p_1^{\mu_1} \cdots p_t^{\mu_t}$$

Prime factors that divide all the coefficients

$$\pi_x = m_X \cdot p_1^{\mu_1} \cdots p_t^{\mu_t}$$

Let F_X be the length of the preperiod of $(U_i \pmod{\frac{\pi_X}{m_X}})_{i \in \mathbb{N}}$.

Theorem

Let $X \subseteq \mathbb{N}$ be an ultimately periodic *U*-recognizable set with period $\pi_x = m_X \cdot p_1^{\mu_1} \cdots p_t^{\mu_t}$. Assume that $F_X - 1 - \left| \operatorname{rep}_U \left(\frac{\pi_X}{m_X} - 1 \right) \right| \ge Z$. Then there is a positive constant *C* such that the minimal automaton of $0^* \operatorname{rep}_U(X)$ has at least $\frac{C}{\gamma_{m_X}} \log_2 \left(\left| \operatorname{rep}_U \left(\frac{\pi_X}{m_X} - 1 \right) \right| + 1 \right)$ states.

Lemma

Let U be a numeration system satisfying (H1), (H2) and (H3). There exists a constant Z such that if w is a greedy U-representation, then for all $z \ge Z$, $10^z w$ is also a greedy U-representation.

$$F_X - 1 - \left| \mathsf{rep}_U \left(\frac{\pi_X}{m_X} - 1 \right) \right|$$

$$F_X - 1 - \left| \mathsf{rep}_U \left(\frac{\pi_X}{m_X} - 1 \right) \right|$$

Assume that for all $j \in [\![1, t]\!]$, there are $\alpha_j, \epsilon_j \in \mathbb{R}_{>0}$ and a nondecreasing function g_j such that

$$u_{p_j}(U_i) < \lfloor \alpha_j i \rfloor + g_j(i)$$

for all $i \in \mathbb{N}$ and there exists M_j such that $g_j(i) < \epsilon_j i$ for all $i > M_j$. Then for large enough μ_1, \ldots, μ_t ,

$$F_X > \frac{\max_{1 \le j \le t} \mu_j}{\max_{1 \le j \le t} (\alpha_j + \epsilon_j)}$$

$$F_X - 1 - \left| \operatorname{rep}_U \left(\frac{\pi_X}{m_X} - 1 \right) \right|$$

If $\beta > 1$, there is a non-negative constant K such that

$$\operatorname{rep}_U(n)| < u \log_\beta(n) + K$$

for all $n \in \mathbb{N}$. In particular,

$$\left|\operatorname{rep}_{U}\left(\frac{\pi_{X}}{m_{X}}-1\right)\right| \leq u\left(\max_{1\leq j\leq t}\mu_{j}\right)\sum_{j=1}^{t}\log_{\beta}(p_{j})+K$$

$$F_X - 1 - \left| \mathsf{rep}_U \left(\frac{\pi_X}{m_X} - 1 \right) \right|$$

Assume that for all $j \in [\![1, t]\!]$, there are $\alpha_j, \epsilon_j \in \mathbb{R}_{>0}$ and a nondecreasing function g_j such that

$$u_{p_j}(U_i) < \lfloor \alpha_j i \rfloor + g_j(i)$$

for all $i \in \mathbb{N}$ and there exists M_j such that $g_j(i) < \epsilon_j i$ for all $i > M_j$. Assume also that $\beta > 1$. Then

$$F_X - 1 - \left| \operatorname{rep}_U \left(\frac{\pi_X}{m_X} - 1 \right) \right|$$

$$\geq \max_{1 \le j \le t} \mu_j \left(\frac{1}{\max(\alpha_j + \epsilon_j)} - u \sum_{j=1}^t \log_\beta(p_j) \right) - K - 1$$

Search for the α_j

Example 2

$$U_{i+3} = 12U_{i+2} + 6U_{i+1} + 12U_i, U_0 = 1, U_1 = 13, U_2 = 163.$$

Proposition

For all $i \in \mathbb{N}$, one has

$$u_3(U_i) < \left\lfloor \frac{i}{3} \right\rfloor + 2.$$

Search for the α_j

Example 2

$$U_{i+3} = 12U_{i+2} + 6U_{i+1} + 12U_i$$
, $U_0 = 1$, $U_1 = 13$, $U_2 = 163$.

Proposition

For all $i \in \mathbb{N}$, one has

$$u_3(U_i) < \left\lfloor \frac{i}{3} \right\rfloor + 2.$$

Proposition

Under Conjecture (\star) , one has

$$u_2(U_i) \leq rac{i}{2} + rac{536}{95}\log_2(i)$$

for all $i \ge 10$.

Part 2

Minimal automaton for multiplying and translating the Thue-Morse set

Theorem (Alexeev, 2004)

The state complexity of the language $0^* \operatorname{rep}_b(m\mathbb{N})$ is

$$\min_{N\geq 0}\left\{\frac{m}{\gcd\left(m,b^{N}\right)}+\sum_{n=0}^{N-1}\frac{b^{n}}{\gcd\left(b^{n},m\right)}\right\}$$

2011 : Charlier, Rampersad, Rigo, Waxweiler

Definition

The Thue-Morse set is the set

$$\mathscr{T} = \left\{ n \in \mathbb{N} : |\operatorname{\mathsf{rep}}_2(n)|_1 \in 2\,\mathbb{N}
ight\}.$$

Definition

The Thue-Morse set is the set

$$\mathscr{T} = \left\{ n \in \mathbb{N} : |\operatorname{\mathsf{rep}}_2(n)|_1 \in 2\,\mathbb{N}
ight\}.$$

 $1001011001101001 \cdots$

Definition

The Thue-Morse set is the set

$$\mathscr{T} = \left\{ n \in \mathbb{N} : |\operatorname{\mathsf{rep}}_2(n)|_1 \in 2\,\mathbb{N}
ight\}.$$

 $1001011001101001 \cdots$



Definition

The Thue-Morse set is the set

$$\mathscr{T} = \left\{ n \in \mathbb{N} : |\operatorname{\mathsf{rep}}_2(n)|_1 \in 2\,\mathbb{N}
ight\}.$$

$1001011001101001 \cdots$



$$\mathscr{T} = \{n \in \mathbb{N} : |\operatorname{rep}_2(n)|_1 \in 2\mathbb{N}\}$$

Theorem

Let $p, m \in \mathbb{N}_{\geq 1}$ and $r \in [[0, m-1]]$. The state complexity of the language $0^* \operatorname{rep}_{2^p}(m\mathscr{T} + r)$ is

$$2k + \left\lceil \frac{z}{p} \right\rceil$$

if $m = k2^z$ with k odd.

Automaton	Accepted language
$\mathscr{A}_{\mathscr{T},2^p}$	$(0,0)^* \operatorname{rep}_{2^p} (\mathscr{T} imes \mathbb{N})$

Automaton	Accepted language
А _{Г,2} р А _{т,r,2} р	$(0,0)^* \operatorname{rep}_{2^p} (\mathscr{T} imes \mathbb{N}) \ (0,0)^* \operatorname{rep}_{2^p} (\{(n,mn+r): n \in \mathbb{N}\})$

Automaton	Accepted language
$egin{array}{llllllllllllllllllllllllllllllllllll$	$(0,0)^* \operatorname{rep}_{2^p} (\mathscr{T} \times \mathbb{N})$ $(0,0)^* \operatorname{rep}_{2^p} (\{(n,mn+r) : n \in \mathbb{N}\})$ $(0,0)^* \operatorname{rep}_{2^p} (\{(t,mt+r) : t \in \mathscr{T}\})$

Automaton	Accepted language
A 7 2P	$(0,0)^*$ rep $_{\mathcal{D}}(\mathscr{T} imes \mathbb{N})$
$\mathscr{A}_{m,r,2^p}$	$(0,0)^* \operatorname{rep}_{2^p} (\{(n,mn+r) : n \in \mathbb{N}\})$
$\mathscr{A}_{\mathscr{T},2^p} \times \mathscr{A}_{m,r,2^p}$	$(0,0)^* \operatorname{rep}_{2^p} (\{(t,mt+r) : t \in \mathscr{T}\})$
$\Pi\left(\mathscr{A}_{\mathscr{T},2^{p}}\times\mathscr{A}_{m,r,2^{p}}\right)$	$0^* \operatorname{rep}_{2^p}(m\mathscr{T}+r)$
The automaton $\mathscr{A}_{\mathcal{T},2^{p}}$

 $(0,0)^* \{ \operatorname{rep}_{2^p}(t,n) : t \in \mathscr{T}, n \in \mathbb{N} \}$



The automaton $\mathscr{A}_{m,r,b}$

 $(0,0)^* \{ \operatorname{rep}_b(n,mn+r) : n \in \mathbb{N} \}$

The automaton $\mathscr{A}_{m,r,b}$

 $(0,0)^* \{ \operatorname{rep}_b(n,mn+r) : n \in \mathbb{N} \}$



 $\delta_{m,b}(i, (d, e)) = j \Leftrightarrow bi + e = md + j$

The product automaton $\mathscr{A}_{m,r,2^p} \times \mathscr{A}_{\mathscr{T},2^p}$

 $(0,0)^* \{ \operatorname{rep}_{2^p}(t, mt+r) : t \in \mathscr{T} \}$



The projected automaton $\Pi\left(\mathscr{A}_{m,r,2^{p}}\times\mathscr{A}_{\mathscr{T},2^{p}}\right)$

$$0^* \operatorname{rep}_{2^p}(m\mathscr{T} + r) = 0^* \{\operatorname{rep}_{2^p}(mt + r) : t \in \mathscr{T}\}$$



Proposition

The automaton $\Pi\left(\mathscr{A}_{m,r,2^{p}}\times\mathscr{A}_{\mathscr{T},2^{p}}\right)$ is

- complete,
- deterministic,
- accessible,
- coaccessible.

Proposition

The automaton $\Pi\left(\mathscr{A}_{m,r,2^{p}}\times\mathscr{A}_{\mathscr{T},2^{p}}\right)$ is

- complete,
- deterministic,
- accessible,
- coaccessible.

Proposition

In the automaton $\Pi \left(\mathscr{A}_{m,r,2^p} \times \mathscr{A}_{\mathscr{T},2^p} \right)$, the states (i, T) and (i, B) are disjoined for all $i \in [0, m-1]$.

The automaton $\Pi\left(\mathscr{A}_{6,2,4}\times\mathscr{A}_{\mathscr{T},4}\right)$



The automaton $\Pi\left(\mathscr{A}_{6,2,4}\times\mathscr{A}_{\mathcal{T},4}\right)$



$$rep_4(23) = 113$$

$$rep_4(23) = 113$$

$$rep_4(23) = 113$$



$$rep_4(23) = 113$$



$$rep_4(23) = 113$$



$$rep_4(23) = 113$$

$$rep_4(23) = 113$$

$$rep_4(23) = 113$$

●●○○○○○○●○○○●○○○●○○○●○○○

$$rep_4(23) = 113$$

Let
$$N = \max\left\{ \begin{bmatrix} \frac{z}{p} \end{bmatrix}, |\operatorname{rep}_{2^{p}}(r)| \right\}$$
. For $\alpha \in \llbracket 0, N \rrbracket$,
 $C'_{\alpha} = \left\{ \begin{array}{l} \left\{ \left(\lfloor \frac{r}{2^{\alpha p}} \rfloor + \ell \frac{m}{2^{\alpha p}}, T_{\ell} \right) : 0 \le \ell \le 2^{\alpha p} - 1 \right\} \text{ if } \alpha \le \frac{z}{p} \\ \left\{ \left(\lfloor \frac{r}{2^{\alpha p}} \rfloor + \ell k, T_{\ell} \right) : 0 \le \ell \le 2^{z} - 1 \right\} \text{ else.} \end{array} \right\}$
and $C_{\alpha} = C'_{\alpha} \setminus \bigcup_{\beta=0}^{\alpha-1} C'_{\beta}$.

Let
$$N = \max\left\{ \begin{bmatrix} \frac{z}{p} \end{bmatrix}, |\operatorname{rep}_{2^p}(r)| \right\}$$
. For $\alpha \in \llbracket 0, N \rrbracket$,
 $C'_{\alpha} = \left\{ \begin{array}{l} \left\{ \left(\lfloor \frac{r}{2^{\alpha p}} \rfloor + \ell \frac{m}{2^{\alpha p}}, T_{\ell} \right) : 0 \le \ell \le 2^{\alpha p} - 1 \right\} \text{ if } \alpha \le \frac{z}{p} \\ \left\{ \left(\lfloor \frac{r}{2^{\alpha p}} \rfloor + \ell k, T_{\ell} \right) : 0 \le \ell \le 2^{z} - 1 \right\} \text{ else.} \end{array} \right\}$
and $C_{\alpha} = C'_{\alpha} \setminus \bigcup_{\beta=0}^{\alpha-1} C'_{\beta}$.

For $(j, X) \in (\llbracket 0, k-1 \rrbracket \times \{T, B\}) \setminus \{(0, T)\}$, we define

$$D'_{(j,X)} = \{(j + k\ell, X_\ell) : 0 \le \ell \le 2^z - 1\}$$

 $\quad \text{and} \quad$

$$D_{(j,X)} = D'_{(j,X)} \setminus \bigcup_{\alpha=0}^{N} C_{\alpha}.$$

Theorem

Let $p, m \in \mathbb{N}_{\geq} 1$ and $r \in [0, m-1]$. The state complexity of the language $0^* \operatorname{rep}_{2^p}(m\mathscr{T}+r)$ is equal to

$$2k + \left[\frac{z}{p}\right]$$

if $m = k2^z$ with k odd.





 $2\times 3+\lceil \tfrac{1}{2}\rceil=7$

Part 3

Automatic sequences based on Parry or Bertrand numeration systems

Integer base systems \subsetneq

 $U_{i+1} = 2U_i, U_0 = 1$

- Pisot numeration systems $F_{i+2}=F_{i+1}+F_i, F_0=1, F_1=2$
- \subseteq Bertrand numeration systems $B_{i+1}=3B_i+1, B_0=1$

The factor complexity function $p_x(n)$ of an infinite word x counts the number of factors of length n in x.

Proposition (Cobham, 1972)

The factor complexity function of a b-automatic sequence is sublinear.

The factor complexity function $p_x(n)$ of an infinite word x counts the number of factors of length n in x.

Proposition (Cobham, 1972)

The factor complexity function of a b-automatic sequence is sublinear.

Theorem

The factor complexity function of a Parry-automatic sequence is sublinear.

The factor complexity function $p_x(n)$ of an infinite word x counts the number of factors of length n in x.

Proposition (Cobham, 1972)

The factor complexity function of a b-automatic sequence is sublinear.

Theorem

The factor complexity function of a Parry-automatic sequence is sublinear.

Theorem

There exists a Bertrand-automatic sequence with superlinear factor complexity.

Closure properties

Integer base systems \subsetneq Pisot n. s. \subsetneq Parry n. s. \subsetneq Bertrand n. s.

Proposition (Bruyère, Hansel, 1997)

The image of a Pisot-automatic sequence under a substitution of constant length is a Pisot-automatic sequence.

Closure properties

Integer base systems \subsetneq Pisot n. s. \subsetneq Parry n. s. \subsetneq Bertrand n. s.

Proposition (Bruyère, Hansel, 1997)

The image of a Pisot-automatic sequence under a substitution of constant length is a Pisot-automatic sequence.

Theorem

There is a Parry numeration system U such that the class of U-automatic sequences is not closed under taking image by a uniform morphism.

Closure properties

Integer base systems \subsetneq Pisot n. s. \subsetneq Parry n. s. \subsetneq Bertrand n. s.

Proposition (Bruyère, Hansel, 1997)

The image of a Pisot-automatic sequence under a substitution of constant length is a Pisot-automatic sequence.

Theorem

There is a Parry numeration system U such that the class of U-automatic sequences is not closed under taking image by a uniform morphism.

 $U_{i+4} = 3U_{i+3} + 2U_{i+2} + 3U_i, U_0 = 1, U_1 = 4, U_2 = 15, U_3 = 54$ Consider the characteristic sequence x of the set $\{U_i : i \in \mathbb{N}\}$:

$$\mu: \mathbf{0} \mapsto \mathbf{0}^t, \mathbf{1} \mapsto \mathbf{10}^{t-1}, t \geq 4$$

Proposition (Bruyère, Hansel, 1997)

Pisot-automatic sequences are closed under periodic deletion.

Theorem

There exists a Parry numeration system U such that the class of U-automatic sequences is not closed under periodic deletion.

Proposition (Bruyère, Hansel, 1997)

Pisot-automatic sequences are closed under periodic deletion.

Theorem

There exists a Parry numeration system U such that the class of U-automatic sequences is not closed under periodic deletion.

 $\{U_i/2: i \in \mathbb{N}, U_i \in 2\mathbb{N}\}$

Perspectives

Ultimate periodicity

- Can we weaken our hypotheses?
- Who are the numeration systems not satisfying

$$F_X - 1 - \left| \operatorname{rep}_U \left(\frac{\pi_X}{m_X} - 1 \right) \right| \ge Z?$$

$$U_{i+2} = 4U_i : 1, 3, 4, 6, 8, 12, \cdots$$

- Is there a strategy working for both integer bases and other positional numeration systems?
- What about the time complexity?
- Replacing \mathscr{T} by a *b*-recognizable set X : mX + r?
- What about positional numeration system which are not an integer base ?
- LSDF

Theorem (Pansiot, 1984)

Let x be a purely morphic word. Then one of the following holds :

- $p_{\mathbf{x}}(n) = \Theta(1)$,
- $p_{\mathbf{x}}(n) = \Theta(n)$,
- $p_{\mathbf{x}}(n) = \Theta(n \log \log n)$,
- $p_{\mathbf{x}}(n) = \Theta(n \log n)$,
- $p_{\mathbf{x}}(n) = \Theta(n^2)$.