A Decision Problem for Ultimately Periodic Sets in Non-Standard Numeration Systems

Jason Bell[1] Emilie Charlier[2] Avierzri Fraenkel[3] Michel Rigo[2]

[1]Department of Mathematics
Simon Fraser University

[2]Department of Mathematics
University of Liège

[3]Department of Computer Science and Applied Mathematics
Weizmann Institute of Science

Séminaire Louvain-la-Neuve, December 12th 2008

Let's start with classical k-ary numeration system, $k \geq 2$:

$$n = \sum_{i=0}^{\ell} d_i\, k^i,\ d_\ell \neq 0, \quad \mathrm{rep}_k(n) = d_\ell \cdots d_0 \in \{0, \ldots, k-1\}^*$$

Definition

A set $X \subseteq \mathbb{N}$ is *k-recognizable*, if the language

$$\mathrm{rep}_k(X) = \{\mathrm{rep}_k(x) \mid x \in X\}$$

is regular, i.e., accepted by a finite automaton.

Examples of k-recognizable sets

- In base 2, the set of even integers : $\mathrm{rep}_2(2\mathbb{N}) = 1\{0,1\}^*0 + e$.
- In base 2, the set of powers of 2 : $\mathrm{rep}_2(\{2^i : i \in \mathbb{N}\}) = 1\,0^*$.
- In base 2, the *Thue-Morse set* :

$$\{n \in \mathbb{N} : |\mathrm{rep}_2(n)|_1 \equiv 0 \pmod 2\}.$$

- Given a k-automatic sequence $(x_n)_{n\geq 0}$ over an alphabet Σ, then for all $a \in \Sigma$, the following set is k-recognizable :

$$\{n \in \mathbb{N} \mid x_n = a\}.$$

Divisibility criteria

If $X \subseteq \mathbb{N}$ is ultimately periodic, then X is k-recognizable $\forall k \geq 2$.

$$X = (3\mathbb{N}+1) \cup (2\mathbb{N}+2) \cup \{3\}, \; \mathit{Index} = 4, \; \mathit{Period} = 6$$

$\chi_X =$  $\cdots$

Definition

Two integers k, $\ell \geq 2$ are *multiplicatively independant* if $k^m = \ell^n \Rightarrow m = n = 0$.

Theorem (Cobham, 1969)

Let $k, \ell \geq 2$ be two multiplicatively independant integers. If $X \subseteq \mathbb{N}$ is both k- and ℓ-recognizable, then X is ultimately periodic, i.e. a finite union of arithmetic progressions.

Theorem (J. Honkala, 1985)

Let $k \geq 2$. It is decidable whether or not a k-recognizable set is ultimately periodic.

Sketch of Honkala's Decision Procedure

- The input is a finite automaton $\mathcal{A}_X$ accepting $\mathrm{rep}_k(X)$.
- The number of states of $\mathcal{A}_X$ produces upper bounds on the possible (minimal) index and period for X.
- Consequently, there are finitely many candidates to check.
- For each pair (i, p) of candidates, produce a DFA for all possible corresponding ultimately periodic sets and compare it with $\mathcal{A}_X$.

Definition

A *positional numeration system* is an increasing sequence $U = (U_i)_{i\geq 0}$ of integers s.t. $U_0 = 1$ and $C_U := \sup_{i\geq 0}\lceil U_{i+1}/U_i \rceil$ is finite.

The *greedy U-representation* of a positive integer n is the unique finite word $\text{rep}_U(n) = d_\ell \cdots d_0$ over $A_U := \{0, \ldots, C_U - 1\}$ satisfying

$$n = \sum_{i=0}^{\ell} d_i\, U_i,\ \ d_\ell \neq 0 \text{ and } \sum_{i=0}^{t} d_i U_i < U_{t+1},\ \forall t = 0, \ldots, \ell.$$

If $x = x_\ell \cdots x_0$ is a word over a finite alphabet of integers, then the *U-numerical value* of x is $\text{val}_U(x) = \sum_{i=0}^{\ell} x_i U_i$.

A set $X \subseteq \mathbb{N}$ is *U-recognizable* if the language $\text{rep}_U(X)$ over A_U is regular.

Definition

A positional numeration system $U = (U_i)_{i \geq 0}$ is said to be *linear* if there exist $k \geq 1$ and constant coefficients $a_1, \ldots, a_k$ such that for all $i \geq 0$, we have

$$U_{i+k} = a_1 U_{i+k-1} + \cdots + a_k U_i, \quad \text{with } a_1, \ldots, a_k \in \mathbb{Z},\ a_k \neq 0.$$

We say that k is the *order* of the recurrence relation.

Example (Fibonacci System)

Consider the sequence defined by $F_0 = 1$, $F_1 = 2$ and $F_{i+2} = F_{i+1} + F_i$, $i \geq 0$. The *Fibonacci (linear numeration) system* is given by $F = (F_i)_{i \geq 0} = (1, 2, 3, 5, 8, 13, \ldots)$.

1	1	8	10000	15	100010
2	10	9	10001	16	100100
3	100	10	10010	17	100101
4	101	11	10100	18	101000
5	1000	12	10101	19	101001
6	1001	13	100000	20	101010
7	1010	14	100001	21	1000000

$\mathrm{rep}_F(\mathbb{N}) = 1(0 + 01)^*$, $A_F = \{0, 1\}$.

Lemma

Let $U = (U_i)_{i \geq 0}$ be a (linear) numeration system such that $\mathbb{N}$ is U-recognizable. Any ultimately periodic $X \subseteq \mathbb{N}$ is U-recognizable and a DFA accepting $\mathrm{rep}_U(X)$ can be effectively obtained.

Remark (J. Shallit, 1994)

If $\mathbb{N}$ is U-recognizable, then U is linear.

Problem

Given a linear numeration system U and a U-recognizable set $X \subseteq \mathbb{N}$. Is it decidable whether or not X is ultimately periodic, i.e., whether or not X is a finite union of arithmetic progressions ?

First part (Upper Bound on the Period)

"Pseudo-Result"

Let X be ultimately periodic with period p_X (X is U-recognizable).

Any DFA accepting $\mathrm{rep}_U(X)$ has at least $f(p_X)$ states, where f is increasing.

"Pseudo-Corollary"

Let $X \subseteq \mathbb{N}$ be a U-recognizable set of integers s.t. $\mathrm{rep}_U(X)$ is accepted by $\mathcal{A}_X$ with k states.

If X is ultimately periodic with period p, then

$$\boxed{f(p) \le k} \quad \text{with} \begin{cases} k & \text{fixed} \\ f & \text{increasing.} \end{cases}$$

$\Rightarrow$ The number of candidates for p is bounded from above.

A technical hypothesis :

$$\lim_{i \to +\infty} U_{i+1} - U_i = +\infty. \qquad (1)$$

Most systems are built on an exponential sequence $(U_i)_{i \geq 0}$.

Lemma

Let $U = (U_i)_{i \geq 0}$ be a numeration system satisfying (1). For all j, there exists L such that for all $\ell \geq L$,

$$10^{\ell - |\operatorname{rep}_U(t)|} \operatorname{rep}_U(t), \; t = 0, \ldots, U_j - 1$$

are greedy U-representations. Otherwise stated,
if w is a greedy U-representation, then for r large enough, $10^r w$ is also a greedy U-representation.

Proposition (Fibonacci)

Let $X \subseteq \mathbb{N}$ be ultimately periodic with period p_X (and index a_X). Any DFA accepting $\mathrm{rep}_F(X)$ has at least p_X states.

- $w^{-1}L = \{u \mid wu \in L\} \leftrightarrow$ states of minimal automaton of L
- $(F_i \mod p_X)_{i \geq 0}$ is *purely periodic*.
  Indeed, $F_{n+2} = F_{n+1} + F_n$ and $F_n = F_{n+2} - F_{n+1}$.
- If $i, j \geq a_X$, $i \not\equiv j \mod p_X$ then there exists $t < p_X$ s.t. either $i + t \in X$ and $j + t \notin X$, or $i + t \notin X$ and $j + t \in X$.

Idea of the Proof with the Fibonacci System

- $\exists n_1, \ldots, n_{p_X},\ \forall j = 1, \ldots, p_X,$

$$10^{n_{p_X}} \cdots 10^{n_1} \operatorname{rep}_F(p_X - 1) \in \operatorname{rep}_F(\mathbb{N}) \tag{2}$$

$$\operatorname{val}_F(10^{n_1 + |\operatorname{rep}_F(p_X - 1)|}) \geq a_X \tag{3}$$

$$\operatorname{val}_F(10^{n_j} \cdots 10^{n_1 + |\operatorname{rep}_F(p_X - 1)|}) \equiv j \mod p_X \tag{4}$$

- For $i, j \in \{1, \ldots, p_X\}$, $i \neq j$, the words

$$10^{n_i} \cdots 10^{n_1} \text{ and } 10^{n_j} \cdots 10^{n_1}$$

generate different states in the minimal automaton of $\operatorname{rep}_F(X)$. This can be shown by concatenating some word of length $|\operatorname{rep}_F(p_X - 1)|$.

$w^{-1}L = \{u \mid wu \in L\} \leftrightarrow$ states of minimal automaton of L

$X = (11\mathbb{N}+3) \cup \{2\}$, $a_X = 3$, $p_X = 11$, $|\mathrm{rep}_F(10)| = 5$

Working in $(F_i \mod 11)_{i \geq 0}$:

$\cdots$ 2 1	1 0 1 10 2 8 5 3 2 1	1 0 1 10 2 8 5 3 2 1	
	1	0 0 0 0 0 0 0 0 0 0	1
1	0 0 0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 0 0 0	2
	1	0 0 0 0 0 0 0 0 1 0	$1+2 \in X$
1	0 0 0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 0 1 0	$2+2 \notin X$

$$\Rightarrow (10^5)^{-1}\,\mathrm{rep}_F(X) \neq (10^9 10^5)^{-1}\,\mathrm{rep}_F(X)$$

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

Example (Fibonacci system, continued)

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

Proposition

Let $U = (U_i)_{i\geq 0}$ be a numeration system satisfying (1).

Let $X \subseteq \mathbb{N}$ be an ultimately periodic set of period p_X.

Then any DFA accepting $\mathrm{rep}_U(X)$ has at least $N_U(p_X)$ states.

Corollary

Let $U = (U_i)_{i \geq 0}$ *be a numeration system satisfying* (1). *Assume that*

$$\lim_{m \to +\infty} N_U(m) = +\infty.$$

Then the period of an ultimately periodic set $X \subseteq \mathbb{N}$ *such that* $\mathrm{rep}_U(X)$ *is accepted by a DFA with* d *states is bounded by the smallest integer* s_0 *such that for all* $m \geq s_0$, $N_U(m) > d$, *which is effectively computable.*

For a sequence $U = (U_i)_{i\geq 0}$ of integers, if $(U_i \mod m)_{i\geq 0}$ is ultimately periodic, we denote its (minimal) index by $\iota_U(m)$.

Proposition

Let $U = (U_i)_{i\geq 0}$ be a linear numeration system.

Let $X \subseteq \mathbb{N}$ be an ultimately periodic set of period p_X and index a_X.

Then any deterministic finite automaton accepting $\text{rep}_U(X)$ has at least $|\text{rep}_U(a_X - 1)| - \iota_U(p_X)$ states.

If p_x is bounded and a_x is increasing, then the number of states is increasing.

Theorem

Let $U = (U_i)_{i\geq 0}$ *be a linear numeration system such that* $\mathbb{N}$ *is* U*-recognizable, satisfying condition* (1). *Assume that*

$$\lim_{m\to+\infty} N_U(m) = +\infty.$$

Then it is decidable whether or not a U*-recognizable set is ultimately periodic.*

Remark

Whenever $\gcd(a_1, \ldots, a_k) = g \geq 2$, for all $n \geq 1$ and for all i large enough, we have $U_i \equiv 0 \mod g^n$ and $N_U(m)$ does not tend to infinity.

Examples

- Honkala's integer bases: $U_{n+1} = k\, U_n$
- $U_{n+2} = 2U_{n+1} + 2U_n$

$$a, b, 2(a+b), 2(2a+3b), 4(3a+4b), 4(8a+11b) \ldots$$

Lemma

Let $U = (U_i)_{i \geq 0}$ *be an increasing sequence satisfying*

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

with $a_1, \ldots, a_k \in \mathbb{Z}$, $a_k \neq 0$. *The following assertions are equivalent:*

(I) $\lim_{m \to +\infty} N_U(m) = +\infty$

(II) *for all prime divisors* p *of* a_k, $\lim_{v \to +\infty} N_U(p^v) = +\infty$.

In particular, if $a_k = \pm 1$, *then* $\lim_{m \to +\infty} N_U(m) = +\infty$.

Theorem

Let $U = (U_i)_{i\geq 0}$ *be a linear recurrence sequence satisfying*

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

with $a_1, \ldots, a_k \in \mathbb{Z}$, $a_k \neq 0$, *and no recurrence relation of smaller order than* k. *One has* $N_U(p^v) \not\to +\infty$ *as* $v \to +\infty$ *if and only if* $P_U(x) = A(x)B(x)$ *with* $A(x), B(x) \in \mathbb{Z}[x]$ *such that:*

(I) $A(0) = B(0) = 1$;

(II) $B(x) \equiv 1 \pmod{p\mathbb{Z}[x]}$;

(III) $A(x)$ *has no repeated roots and all its roots are roots of unity.*

Definition

An *abstract numeration system* is a triple $S = (L, \Sigma, <)$ where L is a regular language over a totally ordered alphabet $(\Sigma, <)$.

Enumerating the words of L with respect to the genealogical ordering induced by $<$ gives a one-to-one correspondence

$$\mathrm{rep}_S : \mathbb{N} \to L \qquad \mathrm{val}_S = \mathrm{rep}_S^{-1} : L \to \mathbb{N}.$$

Example

$L = a^*,\ \Sigma = \{a\}$

n	0	1	2	3	4	$\cdots$
$\mathrm{rep}\,(n)$	ε	a	aa	aaa	$aaaa$	$\cdots$

Example

$L = \{a, b\}^*,\ \Sigma = \{a, b\},\ a < b$

n	0	1	2	3	4	5	6	7	$\cdots$
rep (n)	ε	a	b	aa	ab	ba	bb	aaa	$\cdots$

Example

$L = a^*b^*,\ \Sigma = \{a, b\},\ a < b$

n	0	1	2	3	4	5	6	$\cdots$
rep (n)	ε	a	b	aa	ab	bb	aaa	$\cdots$

Remark

This generalizes non-standard numeration systems $U = (U_i)_{i \geq 0}$ for which $\mathbb{N}$ is U-recognizable, like integer base p systems or Fibonacci system.

$$L = \{\varepsilon\} \cup \{1, \ldots, p-1\}\{0, \ldots, p-1\}^* \text{ or } L = \{\varepsilon\} \cup 1\{0, 01\}^*$$

Notation

If $S = (L, \Sigma, <)$ *is an abstract numeration system and if* $\mathcal{M}_L = (Q_L, q_{0,L}, \Sigma, \delta_L, F_L)$ *is the minimal automaton of* L*, we denote by* $\mathbf{u}_j(q)$ *(resp.* $\mathbf{v}_j(q)$*) the number of words of length* j *(resp.* $\leq j$*) accepted from* $q \in Q_L$ *in* $\mathcal{M}_L$*.*

Remark

The sequences $(\mathbf{u}_j(q))_{j\geq 0}$ (resp. $(\mathbf{v}_j(q))_{j\geq 0}$) satisfy the same homogenous linear recurrence relation for all $q \in Q_L$.

Lemma

Let $w = \sigma_1 \cdots \sigma_n \in L$*. We have*

$$\mathrm{val}_S(w) = \sum_{q \in Q_L} \sum_{i=1}^{|w|} \beta_{q,i}(w)\, \mathbf{u}_{|w|-i}(q) \tag{5}$$

where $\beta_{q,i}(w) := \#\{\sigma < \sigma_i \mid \delta_L(q_{0,L}, \sigma_1 \cdots \sigma_{i-1}\sigma) = q\} + \mathbf{1}_{q,q_{0,L}}$*,*
for $i = 1, \ldots, |w|$*.*

Definition

A set $X \subseteq \mathbb{N}$ of integers is *S-recognizable if the language* $\mathrm{rep}_S(X)$ *over Σ is regular (i.e., accepted by a finite automaton).*

Proposition

Let $S = (L, \Sigma, <)$ be an abstract numeration system built over an infinite regular language L. Any ultimately periodic set X is S-recognizable and a DFA accepting $\mathrm{rep}_S(X)$ *can be effectively obtained.*

Problem

Given an abstract numeration system S and a S-recognizable set $X \subseteq \mathbb{N}$. Is it decidable whether or not X is ultimately periodic ?

THEOREM

Let $S = (L, \Sigma, <)$ be an abstract numeration system such that for all states q of the trim minimal automaton of L $\mathcal{M}_L = (Q_L, q_{0,L}, \Sigma, \delta_L, F_L)$, we have $\lim_{i\to\infty} \mathbf{u}_i(q) = +\infty$ and $\mathbf{u}_i(q_{0,L}) > 0$ for all $i \geq 0$. Assume moreover that $\mathbf{v} = (\mathbf{v}_i(q_{0,L}))_{i\geq 0}$ is such that

$$\lim_{m\to+\infty} N_{\mathbf{v}}(m) = +\infty.$$

It is decidable whether or not an S-recognizable set is ultimately periodic.