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

Emilie Charlier   Michel Rigo

Department of Mathematics
University of Liège

MFCS
Torun, August 25-29 2008

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

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

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: $\text{rep}_2(2\mathbb{N}) = 1\{0,1\}^*0 + \varepsilon$.
- In base 2, the set of powers of 2: $\text{rep}_2(\{2^i | i \in \mathbb{N}\}) = 1\,0^*$.
- In base 2, the "*Thue-Morse set*": $\{n \in \mathbb{N} \mid S(\text{rep}_2(n)) \equiv 0 \pmod 2\}$.
- Given a k-automatic sequence $(x_n)_{n\geq 0}$ over an alphabet Σ, then for all $\sigma \in \Sigma$, the set $\{i \in \mathbb{N} \mid x_i = \sigma\}$ is k-recognizable.

DIVISIBILITY CRITERIA

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

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

$\chi_X =$  $\cdots$

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 $\text{rep}_k(X)$.
- The number of states of $\mathcal{A}_X$ produces an upper bound 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 *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 numeration system $U = (U_i)_{i \geq 0}$ is said to be *linear (of order k)* , if the sequence U satisfies a homogenous linear recurrence relation like

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

for some $k \geq 1$, $a_1, \ldots, a_k \in \mathbb{Z}$ and $a_k \neq 0$.

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

The "pattern" 11 is forbidden, $A_F = \{0, 1\}$.

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 ?

"Pseudo-Result"

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

Any DFA accepting $\text{rep}_U(X)$ 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. $\text{rep}_U(X)$ is accepted by $\mathcal{A}_X$ with k states.

If X is ultimately periodic with period p, then

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

$\Rightarrow$ 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 $\ell \geq L$,

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

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.

$N_U(m) \in \{1, \ldots, m\}$ 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).
If $X \subseteq \mathbb{N}$ is an ultimately periodic U-recognizable 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.*

Lemma

If $U = (U_i)_{i\geq 0}$ *is a linear numeration system satisfying a recurrence relation of order* $k \geq 1$ *of the kind*

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

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

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 is *purely periodic* mod p_X.
  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 $j + t \in X$.
- $\exists n_1, \ldots, n_{p_X}$, $\forall t = 0, \ldots, p_X - 1$,

$$10^{n_{p_X}} 10^{n_{p_X - 1}} \cdots 10^{n_1} 0^{|\mathrm{rep}_F(p_X - 1)| - |\mathrm{rep}_F(t)|} \mathrm{rep}_F(t)$$

is a greedy F-representation.

- Moreover $n_1, \ldots, n_{p_X}$ can be chosen s.t. $\forall j = 1, \ldots, p_X$,

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

  and $\mathrm{val}_F(10^{n_1 + |\mathrm{rep}_F(p_X - 1)|}) \geq a_X$.

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

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

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

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

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

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	$2+2 \notin X$

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

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 U-recognizable set of period p_X and index a_X.

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

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

Theorem (E. C., M. Rigo)

Let $U = (U_i)_{i \geq 0}$ be a linear numeration system such that $\mathbb{N}$ is U-recognizable and satisfying a recurrence relation of order k of the kind

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

with $a_k = \pm 1$ and such that $\lim_{i \to +\infty} U_{i+1} - U_i = +\infty$.
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: $U_{n+1} = k\, U_n$
- $U_{n+2} = 2U_{n+1} + 2U_n$

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

Question

What happen whenever $\gcd(a_1, \ldots, a_k) = 1$ and $a_k \neq \pm 1$?

Learn more about linear recurrent sequences mod m ...

- H.T. Engstrom, On sequences defined by linear recurrence relations, *Trans. Amer. Math. Soc.* **33** (1931).
- M. Ward, The characteristic number of a sequence of integers satisfying a linear recursion relation, *Trans. Amer. Math. Soc.* **35** (1933).
- M. Hall, An isomorphism between linear recurring sequences and algebraic rings, *Trans. Amer. Math. Soc.* **44** (1938).
- G. Rauzy, Relations de récurrence modulo m, Séminaire Delange-Pisot, 1963/1964.

To solve the case where $\gcd(a_1, \ldots, a_k) = 1$.