A Decision Problem for ultimately periodic Sets in non-standard Numeration Systems

Emilie Charlier   Michel Rigo

Department of Mathematics
University of Liège

Journées de Numération
Prague May 26-30 2008

Outline of the talk

- Preliminaries and Motivation
- A Decision Problem
- An upper Bound for the Period
- An upper Bound for the Preperiod
- A Decision Procedure
- Abstract Numeration Systems

Definition

A *numeration system* is given by a (strictly) increasing sequence $U = (U_i)_{i\geq 0}$ of integers such that $U_0 = 1$ 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 $\mathrm{rep}_U(n) = w_\ell \cdots w_0$ over $A_U := \{0, \ldots, C_U - 1\}$ satisfying $n = \sum_{i=0}^{\ell} w_i\, U_i,\ w_\ell \neq 0$ and $\sum_{i=0}^{t} w_i U_i < U_{t+1}$, $\forall t = 0, \ldots, \ell$. We set $\mathrm{rep}_U(0) = \varepsilon$.

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

Definition

A *numeration system* is given by a (strictly) increasing sequence $U = (U_i)_{i\geq 0}$ of integers such that $U_0 = 1$ 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 $\mathrm{rep}_U(n) = w_\ell \cdots w_0$ over $A_U := \{0, \dots, C_U - 1\}$ satisfying $n = \sum_{i=0}^{\ell} w_i\, U_i,\ \ w_\ell \neq 0$ and $\sum_{i=0}^{t} w_i U_i < U_{t+1}$, $\forall t = 0, \dots, \ell$. We set $\mathrm{rep}_U(0) = \varepsilon$.

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

Definition

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

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 $a_k \neq 0$.

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 $F = (F_i)_{i \geq 0} = (1, 2, 3, 5, 8, 13, \ldots)$. For instance, $\text{rep}_F(15) = 100010$ and $\text{val}_F(101001) = 13 + 5 + 1 = 19$.

Definition

Two integers p, $q \geq 2$ are *multiplicatively independant* if $p^k = p^\ell$ and $k, \ell \in \mathbb{N} \Rightarrow k = \ell = 0$.

Notation

If $p \geq 2$ and $U = (p^i)_{i \geq 0}$, a set $X \subseteq \mathbb{N}$ of integers is said p-recognizable if the language $\mathrm{rep}_U(X)$ over $A_U = \{0, \ldots, p-1\}$ is regular.

Theorem (Cobham, 1969)

Let $X \subseteq \mathbb{N}$ be a set of integers. If p and q are two multiplicatively independant integers, X is p-recognizable and q-recognizable if and only if X is ultimately periodic.

Theorem (J. Honkala, 1985)

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

Proposition

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

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 ?

Definition

Let $X \subseteq \mathbb{N}$ be a set of integers.
The *characteristic word of* X is an infinite word $x_0 x_1 x_2 \cdots$ over $\{0, 1\}$ defined by $x_i = 1$ if and only if $i \in X$.

Definition

Let $X \subseteq \mathbb{N}$ be a set of integers.
The *characteristic word of* X is an infinite word $x_0x_1x_2\cdots$ over $\{0,1\}$ defined by $x_i = 1$ if and only if $i \in X$.

If now $X \subseteq \mathbb{N}$ is ultimately periodic, its characteristic word is an infinite word over $\{0,1\}$ of the form

$$x_0x_1x_2\cdots = uv^\omega$$

where u and v are chosen of minimal length. We say that $|u|$ (resp. $|v|$) is the *preperiod* (resp. *period*) of X.

Idea of Honkala's Decision Procedure

The input is a finite automaton accepting $\text{rep}_U(X)$.

First, he gives an upper bound for the possible periods of X, by showing that, if Y is a ultimately periodic set of integers, then the number of states of any deterministic automaton accepting $\text{rep}_U(Y)$ grows with the period of Y.

Then, once the period of X is bounded, he gives an upper bound for the possible preperiods of X, in a similar way.

Notation

For a sequence $U = (U_i)_{i \geq 0}$ *of integers and an integer* $m \geq 2$, $N_U(m) \in \{1, \ldots, m\}$ *denotes the number of values that are taken infinitely often by the sequence* $(U_i \mod m)_{i \geq 0}$.

Notation

For a sequence $U = (U_i)_{i \geq 0}$ *of integers and an integer* $m \geq 2$*,* $N_U(m) \in \{1, \ldots, m\}$ *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 $N_F(11) = 7$.

Notation

For a sequence $U = (U_i)_{i\geq 0}$ *of integers and an integer* $m \geq 2$, $N_U(m) \in \{1, \dots, m\}$ *denotes the number of values that are taken infinitely often by the sequence* $(U_i \mod m)_{i\geq 0}$.

Example (Fibonacci System, continued)

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

Proposition

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* $|v|$, *then any deterministic finite automaton accepting* $\text{rep}_U(X)$ *has at least* $N_U(|v|)$ *states.*

Corollary

Let $U = (U_i)_{i \geq 0}$ *be a numeration system satisfying* $\lim_{i \to +\infty} U_{i+1} - U_i = +\infty$. *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.*

Corollary

Let $U = (U_i)_{i \geq 0}$ be a numeration system satisfying $\lim_{i \to +\infty} U_{i+1} - U_i = +\infty$. 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

Let $U = (U_i)_{i\geq 0}$ be a numeration system satisfying condition $\lim_{i\to+\infty} U_{i+1} - U_i = +\infty$ and $X \subseteq \mathbb{N}$ be an ultimately periodic U-recognizable set of period $|v|$. If 1 occurs infinitely many times in $(U_i \mod |v|)_{i\geq 0}$ then any deterministic finite automaton accepting $\mathrm{rep}_U(X)$ has at least $|v|$ states.

Definition

Let $L \subseteq \Sigma^*$ be a language over a finite alphabet Σ and x be a finite word over Σ. We set $x^{-1}.L = \{z \in \Sigma^* \mid xz \in L\}$.
The Myhill-Nerode congruence $\sim_L$ is defined as follows. Let $x, y \in \Sigma^*$. We write $x \sim_L y$ if $x^{-1}.L = y^{-1}.L$.

Proposition

A language L over a finite alphabet Σ is regular if and only if $\sim_L$ has a finite index, being the number of states of the minimal automaton of L.

Definition

Let $L \subseteq \Sigma^*$ be a language over a finite alphabet Σ and x be a finite word over Σ. We set $x^{-1}.L = \{z \in \Sigma^* \mid xz \in L\}$.
The Myhill-Nerode congruence $\sim_L$ is defined as follows. Let $x, y \in \Sigma^*$. We write $x \sim_L y$ if $x^{-1}.L = y^{-1}.L$.

Proposition

A language L over a finite alphabet Σ is regular if and only if $\sim_L$ has a finite index, being the number of states of the minimal automaton of L.

Example (Fibonacci System, continued)

For all $m \geq 2$, the sequences $(F_i \mod m)_{i\geq 0}$ is purely periodic. So $F_0 = 1$ appears infinitely often in $(F_i \mod m)_{i\geq 0}$.
Let $X \subseteq \mathbb{N}$ be an ultimately periodic F-recognizable set of period $|v|$ and preperiod $|u|$.

Example (Fibonacci System, continued)

There exist $n_1, \ldots, n_{|v|}$ such that for all $t = 0, \ldots, |v| - 1$,

$$10^{n_{|v|}} 10^{n_{|v|-1}} \cdots 10^{n_1} 0^{|\operatorname{rep}_U(|v|-1)| - |\operatorname{rep}_U(t)|} \operatorname{rep}_U(t)$$

is a greedy F-representation.

Example (Fibonacci System, continued)

There exist $n_1, \dots, n_{|v|}$ such that for all $t = 0, \dots, |v| - 1$,

$$10^{n_{|v|}} 10^{n_{|v|-1}} \cdots 10^{n_1} 0^{|\operatorname{rep}_U(|v|-1)| - |\operatorname{rep}_U(t)|} \operatorname{rep}_U(t)$$

is a greedy F-representation. Moreover $n_1, \dots, n_{|v|}$ can be chosen such that, for all $j = 1, \dots, |v|$,

$$\operatorname{val}_U(10^{n_j} \cdots 10^{n_1 + |\operatorname{rep}_U(|v|-1)|}) \equiv j \mod |v|$$

and $\operatorname{val}_U(10^{n_1 + |\operatorname{rep}_U(|v|-1)|}) > |u|$.

Example (Fibonacci System, continued)

There exist $n_1, \ldots, n_{|v|}$ such that for all $t = 0, \ldots, |v| - 1$,

$$10^{n_{|v|}} 10^{n_{|v|-1}} \cdots 10^{n_1} 0^{|\operatorname{rep}_U(|v|-1)| - |\operatorname{rep}_U(t)|} \operatorname{rep}_U(t)$$

is a greedy F-representation. Moreover $n_1, \ldots, n_{|v|}$ can be chosen such that, for all $j = 1, \ldots, |v|$,

$$\operatorname{val}_U(10^{n_j} \cdots 10^{n_1 + |\operatorname{rep}_U(|v|-1)|}) \equiv j \mod |v|$$

and $\operatorname{val}_U(10^{n_1 + |\operatorname{rep}_U(|v|-1)|}) > |u|$. For $i, j \in \{1, \ldots, |v|\}$, $i \neq j$, the words

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

are nonequivalent for $\sim_{\operatorname{rep}_U(X)}$. This can be shown by concatenating some word of the kind $0^{|\operatorname{rep}_U(|v|-1)| - |\operatorname{rep}_U(t)|} \operatorname{rep}_U(t)$ with $t < |v|$.

Notation

For a sequence $U = (U_i)_{i \geq 0}$ *of integers, if* $(U_i \mod m)_{i \geq 0}$, $m \geq 2$, *is ultimately periodic, we denote its (minimal) preperiod by* $\iota_U(m)$ *and its (minimal) period by* $\pi_U(m)$.

Notation

For a sequence $U = (U_i)_{i\geq 0}$ *of integers, if* $(U_i \mod m)_{i\geq 0}$, $m \geq 2$, *is ultimately periodic, we denote its (minimal) preperiod by* $\iota_U(m)$ *and its (minimal) period by* $\pi_U(m)$.

Example (Fibonacci System, continued)

$(F_i \mod 4) = (1, 2, 3, 1, 0, 1, 1, 2, 3, \ldots)$ and $\pi_F(4) = 6$.
$(F_i \mod 11) = (1, 2, 3, 5, 8, 2, 10, 1, 0, 1, 1, 2, 3, \ldots)$ and $\pi_F(11) = 10$.
We have $\iota_F(m) = 0$, for all $m \geq 2$.

Notation

For a sequence $U = (U_i)_{i \geq 0}$ *of integers, if* $(U_i \mod m)_{i \geq 0}$, $m \geq 2$, *is ultimately periodic, we denote its (minimal) preperiod by* $\iota_U(m)$ *and its (minimal) period by* $\pi_U(m)$.

Example (Fibonacci System, continued)

$(F_i \mod 4) = (1, 2, 3, 1, 0, 1, 1, 2, 3, \ldots)$ and $\pi_F(4) = 6$.
$(F_i \mod 11) = (1, 2, 3, 5, 8, 2, 10, 1, 0, 1, 1, 2, 3, \ldots)$ and $\pi_F(11) = 10$.
We have $\iota_F(m) = 0$, for all $m \geq 2$.

Remark

If $U = (U_i)_{i \geq 0}$ is a linear numeration system of order k, then for all $m \geq 2$, we have $N_U(m) \geq \sqrt[k]{\pi_U(m)}$.

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* $|v|$ *and preperiod* $|u|$ *such that* $|\operatorname{rep}_U(|u|-1)| - \iota_U(|v|) > 0$*.*
Then any deterministic finite automaton accepting $\operatorname{rep}_U(X)$ *has at least* $|\operatorname{rep}_U(|u|-1)| - \iota_U(|v|)$ *states.*

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.*

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.

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.

Question

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

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	$\cdots$
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* $\mathcal{M}_L$*.*

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 $q \in Q_L$.

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

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

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).*

Definition

A set $X \subseteq \mathbb{N}$ of integers is *S-recognizable if the language* $\text{rep}_S(X)$ *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 $\text{rep}_S(X)$ *can be effectively obtained.*

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 and let* $\mathcal{M}_L = (Q_L, q_{0,L}, \Sigma, \delta_L, F_L)$ *the trim minimal automaton of* L*. Assume that*

$$\forall q \in Q_L \ \lim_{j\to\infty} \mathbf{u}_j(q) = +\infty;$$

$$\forall j \geq 0 \ \mathbf{u}_j(q_{0,L}) > 0.$$

Assume moreover that $\mathbf{v} = (\mathbf{v}_i(q_{0,L}))_{i\geq 0}$ *satisfies a linear recurrence relation of the form*

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

with $k \geq 1$, $a_1, \ldots, a_k \in \mathbb{Z}$ *and* $a_k = \pm 1$*.*
It is decidable whether or not a S*-recognizable set is ultimately periodic.*