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

Emilie Charlier and Michel Rigo

Institute of Mathematics, University of Liège, Grande Traverse 12 (B 37), B-4000 Liège, Belgium
{echarlier,M.Rigo}@ulg.ac.be

Abstract. Consider a non-standard numeration system like the one built over the Fibonacci sequence where nonnegative integers are represented by words over $\{0,1\}$ without two consecutive 1. Given a set X of integers such that the language of their greedy representations in this system is accepted by a finite automaton, we consider the problem of deciding whether or not X is a finite union of arithmetic progressions. We obtain a decision procedure under some hypothesis about the considered numeration system. In a second part, we obtain an analogous decision result for a particular class of abstract numeration systems built on an infinite regular language.

1 Introduction

Definition 1. *A* positional 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. Let $A_U=\{0,\ldots,C_U-1\}$. 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 satisfying*

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

We set $\mathrm{rep}_U(0)$ to be the empty word ε. 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 deterministic finite automaton, DFA). If $x=x_\ell\cdots x_0$ is a word over a finite alphabet of integers, then the U*-numerical value *of x is*

$$\mathrm{val}_U(x)=\sum_{i=0}^{\ell} x_i\,U_i.$$

Remark 1. Let x,y be two words over A_U. As a consequence of the greediness of the representation, if xy is a greedy U-representation and if the first letter of y is not 0, then y is also a greedy U-representation. Notice that for $m,n\in\mathbb{N}$, we have $m<n$ if and only if $\mathrm{rep}_U(m)<_{gen}\mathrm{rep}_U(n)$ where $<_{gen}$ is the genealogical ordering over A_U^*: words are ordered by increasing length and for words of same

E. Ochmański and J. Tyszkiewicz (Eds.): MFCS 2008, LNCS 5162, pp. 241–252, 2008.

length, one uses the lexicographical ordering induced by the natural ordering of the digits in the alphabet A_U. Recall that for two words $x, y \in A_U^*$ of same length, x is lexicographically smaller than y if there exist $w, x', y' \in A_U^*$ and $a, b \in A_U$ such that $x = wax'$, $y = wby'$ and $a < b$.

For a positional numeration system U, it is natural to expect that $\mathbb{N}$ is U-recognizable. A necessary condition is that the sequence U satisfies a linear recurrence relation [12].

Definition 2. *A positional numeration system $U = (U_i)_{i \geq 0}$ is said to be* linear*, if the sequence U satisfies a homogenous linear recurrence relation. For all $i \geq 0$, we have*

$$U_{i+k} = a_1 U_{i+k-1} + \cdots + a_k U_i \tag{1}$$

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

Example 1. Consider the sequence defined by $F_0 = 1$, $F_1 = 2$ and for all $n \geq 0$, $F_{n+2} = F_{n+1} + F_n$. The *Fibonacci (linear numeration) system* is given by $F = (F_i)_{i \geq 0} = (1, 2, 3, 5, 8, 13, \ldots)$. For instance, $\mathrm{rep}_F(15) = 100010$ and $\mathrm{val}_F(101001) = 13 + 5 + 1 = 19$.

In this paper, we address the following decidability question.

Problem 1. Given a linear numeration system U and a set $X \subseteq \mathbb{N}$ such that $\mathrm{rep}_U(X)$ is recognized by a (deterministic) finite automaton. Is it decidable whether or not X is ultimately periodic, i.e., whether or not X is a finite union of arithmetic progressions ?

Ultimately periodic sets of integers play a special role. On the one hand such infinite sets are coded thanks to a finite amount of information. On the other hand the celebrated Cobham's theorem asserts that these sets are the only sets that are recognizable in all integer base systems [3]. It is the reason why they are also referred in the literature as *recognizable* sets of integers (the recognizability being in that case independent of the base). Moreover, Cobham's theorem has been extended to various situations and in particular, to numeration systems given by substitutions [4].

J. Honkala showed in [8] that Problem 1 turns out to be decidable for the usual integer base $b \geq 2$ numeration system defined by $U_n = b\, U_{n-1}$ for $n \geq 1$. Let us also mention [1] where the number of states of the minimal automaton accepting numbers written in base b and divisible by d is given explicitly.

The question under inspection in this paper was raised by J. Sakarovitch during the "*Journées de Numération*" in Graz, May 2007. The question was initially asked for a larger class of systems that the one treated here, namely for any abstract numeration systems defined on an infinite regular language [9].

The structure of this paper is the same as [8]. First we give an upper bound on the admissible periods of a U-recognizable set X when it is assumed to be ultimately periodic, then an upper bound on the admissible preperiods is obtained. These bounds depend essentially on the number of states of the (minimal) automaton recognizing $\mathrm{rep}_U(X)$. Finally, finitely many such periods and preperiods have to be checked. Even if the structure is the same, our arguments and techniques are quite different from [8]. Actually they cannot be applied to integer base systems (see Remark 5).

In the next section, Theorem 1 gives a decision procedure for Problem 1 whenever U is a linear numeration system such that $\mathbb{N}$ is U-recognizable and satisfying a relation like (1) with $a_k = \pm 1$ (the main reason for this assumption is that 1 and -1 are the only two integers invertible modulo n for all $n \geq 2$). In the last section, we consider the same decision problem but restated in the framework of abstract numeration systems [9]. We apply successfully the same kind of techniques to a large class of abstract numeration systems (for instance, an example consisting of two copies of the Fibonacci system is considered). The corresponding decision procedure is given by Theorem 2. All along the paper, we try whenever it is possible to state results in their most general form, even if later on we have to restrict ourselves to particular cases. For instance, results about the admissible preperiods do not require any extra assumption.

2 Decision Procedure for Linear Systems with $a_k = \pm 1$

We will often consider positional numeration systems $U = (U_i)_{i \geq 0}$ satisfying the following condition:

$$\lim_{i \to +\infty} U_{i+1} - U_i = +\infty. \tag{2}$$

Lemma 1. *Let $U = (U_i)_{i \geq 0}$ be a positional numeration system satisfying* (2)*. Then 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.

Proof. Notice that $\operatorname{rep}_U(U_j - 1)$ is the greatest word of length j in $\operatorname{rep}_U(\mathbb{N})$, since $\operatorname{rep}_U(U_j) = 10^j$. By hypothesis, there exists L such that for all $\ell \geq L$, $U_{\ell+1} - U_\ell > U_j - 1$. Therefore, for all $\ell \geq L$,

$$10^{\ell - j} \operatorname{rep}_U(U_j - 1)$$

is the greedy U-representation of $U_\ell + U_j - 1 < U_{\ell+1}$ and the conclusion follows.

Remark 2. Bertrand numeration systems associated with a real number $\beta > 1$ are defined as follows. Let $A_\beta = \{0, \ldots, \lceil \beta \rceil - 1\}$. Any $x \in [0, 1]$ can be written as

$$x = \sum_{i=1}^{+\infty} c_i \, \beta^{-i}, \text{ with } c_i \in A_\beta$$

and the sequence $(c_i)_{i \geq 1}$ is said to be a *β-representation* of x. The maximal β-representation of x for the lexicographical order is denoted $d_\beta(x)$ and is called the

β-*development* of x (for details see [10, Chap. 8]). We say that a β-development $(c_i)_{i\geq 1}$ is *finite* if there exists N such that $c_i = 0$ for all $i \geq N$. If there exists $m \geq 1$ such that $d_\beta(1) = t_1 \cdots t_m$ with $t_m \neq 0$, we set $d^*_\beta(1) := (t_1 \cdots t_{m-1}(t_m - 1))^\omega$, otherwise $d_\beta(1)$ is infinite and we set $d^*_\beta(1) := d_\beta(1)$.

We can now define a positional numeration system $U_\beta = (U_i)_{i\geq 0}$ associated with β (see [2]). If $d^*_\beta(1) = (t_i)_{i\geq 1}$, then

$$U_0 = 1 \text{ and } \forall i \geq 1,\ U_i = t_1 U_{i-1} + \cdots + t_i U_0 + 1. \quad (3)$$

If β is a Parry number (i.e., $d_\beta(1)$ is finite or ultimately periodic) then $d^*_\beta(1)$ is ultimately periodic and one can derive from (3) that the sequence U_β satisfies a linear recurrence relation and as a consequence of Bertrand's theorem [2] linking greedy U_β-representations and finite factors occurring in β-developments, the language $\mathrm{rep}_{U_\beta}(\mathbb{N})$ of the greedy U_β-representations is regular. The automaton accepting these representations is well-known [6] and has a special form (all states — except for a sink — are final and from all these states, an edge of label 0 goes back to the initial state). We therefore have the following property being much stronger than the previous lemma. If x and y are greedy U_β-representations then $x0y$ is also a greedy U_β-representation.

Example 2. The Fibonacci system is the Bertrand system associated with the golden ratio $(1+\sqrt{5})/2$. Since greedy representations in the Fibonacci system are the words not containing two consecutive ones [13], then for $x, y \in \mathrm{rep}_F(\mathbb{N})$, we have $x0y \in \mathrm{rep}_F(\mathbb{N})$.

Definition 3. *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 $i \in X$.*

Consider for now $X \subseteq \mathbb{N}$ to be an ultimately periodic set. The characteristic word of X is therefore 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 the length $|u|$ of u (resp. the length $|v|$ of v) is the preperiod *(resp.* period*) of X. Hence, for all $n \geq |u|$, $n \in X$ if and only if $n + |v| \in X$.*

The following lemma is a simple consequence of the minimality of the period chosen to represent an ultimately periodic set.

Lemma 2. *Let $X \subseteq \mathbb{N}$ be an ultimately periodic set of period p_X and preperiod a_X. Let $i, j \geq a_X$. If $i \not\equiv j \mod p_X$ then there exists $t < p_X$ such that either $i + t \in X$ and $j + t \notin X$ or $i + t \notin X$ and $j + t \in X$.*

We assume that the reader is familiar with automata theory (see for instance [11]) but let us recall some classical results. Let $L \subseteq \Sigma^*$ be a language over a finite alphabet Σ and x be a finite word over Σ. We set

$$x^{-1}L = \{z \in \Sigma^* \mid xz \in L\}.$$

We can now define the Myhill-Nerode congruence. Let $x, y \in \Sigma^*$. We have $x \sim_L y$ if and only if $x^{-1}L = y^{-1}L$. Moreover L is regular if and only if $\sim_L$ has a finite index being the number of states of the minimal automaton of L.

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

Proposition 1. *Let $U = (U_i)_{i\geq 0}$ be a positional numeration system satisfying* (2)*. If $X \subseteq \mathbb{N}$ is an ultimately periodic U-recognizable set of period p_X, then any deterministic finite automaton accepting $\operatorname{rep}_U(X)$ has at least $N_U(p_X)$ states.*

Proof. Let a_X be the preperiod of X. By Lemma 1, there exists L such that for any $h \geq L$, the words

$$10^{h-|\operatorname{rep}_U(t)|}\operatorname{rep}_U(t),\ t = 0, \ldots, p_X - 1$$

are greedy U-representations. The sequence $(U_i \mod p_X)_{i\geq 0}$ takes infinitely often $N_U(p_X) =: N$ different values. Let $h_1, \ldots, h_N \geq L$ be such that

$$i \neq j \Rightarrow U_{h_i} \not\equiv U_{h_j} \mod p_X$$

and $h_1, \ldots, h_N$ can be chosen such that $U_{h_i} > a_X$ for all $i \in \{1, \ldots, N\}$.

By Lemma 2, for all $i, j \in \{1, \ldots, N\}$ such that $i \neq j$, there exists $t_{i,j} < p_X$ such that either $U_{h_i} + t_{i,j} \in X$ and $U_{h_j} + t_{i,j} \notin X$, or $U_{h_i} + t_{i,j} \notin X$ and $U_{h_j} + t_{i,j} \in X$. Therefore,

$$w_{i,j} = 0^{|\operatorname{rep}_U(p_X-1)|-|\operatorname{rep}_U(t_{i,j})|}\operatorname{rep}_U(t_{i,j})$$

is a word such that either

$$10^{h_i-|\operatorname{rep}_U(p_X-1)|}w_{i,j} \in \operatorname{rep}_U(X) \text{ and } 10^{h_j-|\operatorname{rep}_U(p_X-1)|}w_{i,j} \notin \operatorname{rep}_U(X),$$

or

$$10^{h_i-|\operatorname{rep}_U(p_X-1)|}w_{i,j} \notin \operatorname{rep}_U(X) \text{ and } 10^{h_j-|\operatorname{rep}_U(p_X-1)|}w_{i,j} \in \operatorname{rep}_U(X).$$

Therefore the words $10^{h_1-|\operatorname{rep}_U(p_X-1)|}, \ldots, 10^{h_N-|\operatorname{rep}_U(p_X-1)|}$ are pairwise non-equivalent for the relation $\sim_{\operatorname{rep}_U(X)}$ and the minimal automaton of $\operatorname{rep}_U(X)$ has at least $N = N_U(p_X)$ states.

The previous proposition has an immediate consequence.

Corollary 1. *Let $U = (U_i)_{i\geq 0}$ be a positional numeration system satisfying* (2)*. Assume that*

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

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

For a sequence $(U_i)_{i\ge 0}$ of integers, if $(U_i \mod m)_{i\ge 0}$ is ultimately periodic, we denote its (minimal) preperiod by $\iota_U(m)$ (we choose notation ι to remind the word index which is equally used as preperiod) and its (minimal) period by $\pi_U(m)$. The next lemma provides a special case where assumption about $N_U(m)$ in Corollary 1 is satisfied.

Lemma 3. *If $U=(U_i)_{i\ge 0}$ is a linear numeration system satisfying a recurrence relation of order k of the kind* (1) *with $a_k=\pm 1$, then $\lim_{m\to+\infty} N_U(m)=+\infty$.*

Proof. For all $m\ge 2$, since U is a linear numeration system, the sequence $(U_i \mod m)_{i\ge 0}$ is ultimately periodic but here it is even purely periodic. Indeed, for all $i\ge 0$, U_{i+k} is determined by the k previous terms $U_{i+k-1},\ldots,U_i$. But since $a_k=\pm 1$, for all $i\ge 0$, U_i is also determined by the k following terms $U_{i+1},\ldots,U_{i+k}$. So, by definition of $N_U(m)$, the sequence $(U_i \mod m)_{i\ge 0}$ takes exactly $N_U(m)$ different values because any term appears infinitely often.

Since U is increasing, the function α mapping m onto the smallest index $\alpha(m)$ such that $U_{\alpha(m)}\ge m$ is nondecreasing and $\lim_{m\to+\infty}\alpha(m)=+\infty$. The conclusion follows, as $N_U(m)\ge\alpha(m)$. Indeed, $U_0,\ldots,U_{\alpha(m)-1}$ are distinct. So $(U_i \mod m)_{i\ge 0}$ takes infinitely often at least $\alpha(m)$ values.

Remark 3. Let $U=(U_i)_{i\ge 0}$ be a positional numeration system satisfying hypothesis of Lemma 3 and let X be a U-recognizable set of integers. If $\mathrm{rep}_U(X)$ is accepted by a DFA with d states, then the constant s_0 (depending on d) given in the statement of Corollary 1 can be estimated as follows.

By Lemma 3, $\lim_{m\to+\infty} N_U(m)=+\infty$. Define t_0 to be the smallest integer such that $\alpha(t_0)>d$, where α is defined as in the proof of Lemma 3. This integer can be effectively computed by considering the first terms of the linear sequence $(U_i)_{i\ge 0}$. Notice that $N_U(t_0)\ge\alpha(t_0)>d$. Consequently $s_0\le t_0$.

Moreover, if U satisfies (2) and if X is an ultimately periodic set, then, by Corollary 1, the period of X is bounded by t_0. So t_0 can be used as an upper bound for the period and it can be effectively computed.

A result similar to the previous corollary (in the sense that it permits to give an upper bound on the period) can be stated as follows. One has to notice that $a_k=\pm 1$ implies that 1 occurs infinitely often in $(U_i \mod m)_{i\ge 0}$ for all $m\ge 2$.

Proposition 2. *Let $U=(U_i)_{i\ge 0}$ be a positional numeration system satisfying* (2) *and $X\subseteq\mathbb{N}$ be an ultimately periodic U-recognizable set of period p_X. If* 1 *occurs infinitely many times in $(U_i \mod p_X)_{i\ge 0}$ then any deterministic finite automaton accepting $\mathrm{rep}_U(X)$ has at least p_X states.*

Proof. Let a_X be the preperiod of X. Applying several times Lemma 1, there exist $n_1,\ldots,n_{p_X}$ such that

$$10^{n_{p_X}}10^{n_{p_X-1}}\cdots 10^{n_1}0^{|\mathrm{rep}_U(p_X-1)|-|\mathrm{rep}_U(t)|}\mathrm{rep}_U(t),\ t=0,\ldots,p_X-1$$

are greedy U-representations. Moreover, since 1 occurs infinitely many times in the sequence $(U_i \mod p_X)_{i\ge 0}$, $n_1,\ldots,n_{p_X}$ can be chosen such that, for all $j=1,\ldots,p_X$,

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

and

$$\mathrm{val}_U(10^{n_1+|\,\mathrm{rep}_U(p_X-1)|}) > a_X.$$

For $i,j \in \{1,\ldots,p_X\}$, $i \neq j$, by Lemma 2 the words

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

are nonequivalent for $\sim_{\mathrm{rep}_U(X)}$. This can be shown by concatenating some word of the kind $0^{|\,\mathrm{rep}_U(p_X-1)|-|\,\mathrm{rep}_U(t)|}\,\mathrm{rep}_U(t)$ with $t < p_X$, as in the proof of Proposition 1. This concludes the proof.

Now we want to obtain an upper bound on the preperiod of any ultimately periodic U-recognizable set.

Proposition 3. *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 preperiod 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.*

The arguments of the following proof are similar to the one found in [8].

Proof. W.l.o.g. we can assume that $|\,\mathrm{rep}_U(a_X - 1)| - \iota_U(p_X) > 0$. The sequence $(U_i \mod p_X)_{i\geq 0}$ is ultimately periodic with preperiod $\iota_U(p_X)$ and period $\pi_U(p_X)$. Proceed by contradiction and assume that $\mathcal{A}$ is a deterministic finite automaton with less than $|\,\mathrm{rep}_U(a_X - 1)| - \iota_U(p_X)$ states accepting $\mathrm{rep}_U(X)$. There exist words w, w_4 such that the greedy U-representation of $a_X - 1$ can be factorized as

$$\mathrm{rep}_U(a_X - 1) = ww_4$$

with $|w| = |\,\mathrm{rep}_U(a_X - 1)| - \iota_U(p_X)$. By the pumping lemma, w can be written $w_1w_2w_3$ with $w_2 \neq \varepsilon$ and for all $i \geq 0$,

$$w_1w_2^iw_3w_4 \in \mathrm{rep}_U(X) \Leftrightarrow w_1w_2w_3w_4 \in \mathrm{rep}_U(X).$$

By minimality of a_X and p_X, either $a_X - 1 \in X$ and for all $n \geq 1$, $a_X + np_X - 1 \notin X$, or $a_X - 1 \notin X$ and for all $n \geq 1$, $a_X + np_X - 1 \in X$. Using the ultimate periodicity of $(U_i \mod p_X)_{i\geq 0}$, we observe that repeating a factor of length multiple of $\pi_U(p_X)$ exactly p_X times does not change the value mod p_X and we get

$$\mathrm{val}_U(w_1w_2^{p_X\pi_U(p_X)}w_2w_3w_4) \equiv \mathrm{val}_U(w_1w_2w_3w_4) \mod p_X,$$

leading to a contradiction.

For the sake of completeness, we restate some well-known property of ultimately periodic sets (see for instance [11] for a prologue on the Pascal's machine for integer base systems).

Lemma 4. *Let a, b be nonnegative integers and $U = (U_i)_{i\geq 0}$ be a linear numeration system. The language*

$$\mathrm{val}_U^{-1}(a\mathbb{N}+b) = \{w \in A_U^* \mid \mathrm{val}_U(w) \in a\mathbb{N}+b\} \subset A_U^*$$

is regular. In particular, if $\mathbb{N}$ is U-recognizable then a DFA accepting $\mathrm{rep}_U(a\mathbb{N}+b)$ can be obtained efficiently and any ultimately periodic set is U-recognizable.

Remark 4. In the previous statement, the assumption about the U-recognizability of $\mathbb{N}$ is of particular interest. Indeed, it is well-known that for an arbitrary linear numeration system, $\mathbb{N}$ is in general *not* U-recognizable. If $\mathbb{N}$ is U-recognizable, then U satisfies a linear recurrence relation [12], but the converse does not hold. Sufficient conditions on the recurrence relation that U satisfies for $\mathbb{N}$ to be U-recognizable are given in [7].

Theorem 1. *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* (1) *with $a_k = \pm 1$ and condition* (2)*. It is decidable whether or not a U-recognizable set is ultimately periodic.*

Proof. Let X be a U-recognizable set and d be the number of states of the minimal automaton of $\mathrm{rep}_U(X)$.

As discussed in Remark 3, if X is ultimately periodic, then the admissible periods are bounded by the constant t_0, which is effectively computable (an alternative and easier argument is provided by Proposition 2). Then, using Proposition 3, the admissible preperiods are also bounded by a constant. Indeed, assume that X is ultimately periodic with period $p_X \leq t_0$ and preperiod a_X. We have $\iota_U(p_X) = 0$ and any DFA accepting $\mathrm{rep}_U(X)$ must have at least $|\,\mathrm{rep}_U(a_X - 1)|$ states. Therefore, the only values that a_X can take satisfy $|\,\mathrm{rep}_U(a_X - 1)| \leq d$.

Consequently the sets of admissible preperiods and periods that we have to check are finite. For each pair (a, p) of admissible preperiods and periods, there are at most $2^a 2^p$ distinct ultimately periodic sets. Thanks to Lemma 4, one can build an automaton for each of them and then compare the language L accepted by this automaton with $\mathrm{rep}_U(X)$. (Recall that testing whether $L \setminus \mathrm{rep}_U(X) = \emptyset$ and $\mathrm{rep}_U(X) \setminus L = \emptyset$ is decidable algorithmically).

Remark 5. We have thus obtained a decision procedure for our Problem 1 when the coefficient a_k occurring in (1) is equal to ± 1. On the other hand, 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 assumption about $N_U(m)$ in Corollary 1 does not hold [5]. Indeed, the only value taken infinitely often by the sequence $(U_i \mod g^n)_{i\geq 0}$ is 0, so $N_U(m)$ equals 1 for infinitely many values of m. Notice in particular, that the same observation can be made for the usual integer base $b \geq 2$ numeration system where the only value taken infinitely often by the sequence $(b^i \mod b^n)_{i\geq 0}$ is 0, for all $n \geq 1$.

3 A Decision Procedure for a Class of Abstract Numeration Systems

An *abstract numeration system* $S = (L, \Sigma, <)$ is given by an infinite regular language L over a totally ordered alphabet $(\Sigma, <)$ [9]. By enumerating the words of L in genealogical order, we get a one-to-one correspondence denoted rep_S between $\mathbb{N}$ and L. In particular, 0 is represented by the first word in L. The reciprocal map associating a word $w \in L$ to its index in the genealogically ordered language L is denoted val_S. 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).

Let $S = (L, \Sigma, <)$ be an abstract numeration system built over an infinite regular language L having $\mathcal{M}_L = (Q_L, q_{0,L}, \Sigma, \delta_L, F_L)$ as minimal automaton. The transition function $\delta_L : Q_L \times \Sigma \to Q_L$ is extended on $Q_L \times \Sigma^*$. 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 $\mathcal{M}_L$. By classical arguments, 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 $q \in Q_L$ (for details, see Remark 6).

In this section, we consider, with some extra hypothesis on the abstract numeration system, the following decidability question analogous to Problem 1.

Problem 2. Given an abstract numeration system S and a set $X \subseteq \mathbb{N}$ such that $\text{rep}_S(X)$ is recognized by a (deterministic) finite automaton, is it decidable whether or not X is ultimately periodic, i.e., whether or not X is a finite union of arithmetic progressions ?

Abstract numeration systems are a generalization of positional numeration systems $U = (U_i)_{i\geq 0}$ for which $\mathbb{N}$ is U-recognizable.

Example 3. Take the language $L = \{\varepsilon\} \cup 1\{0, 01\}^*$ and assume $0 < 1$. Ordering the words of L in genealogical order: $\varepsilon, 1, 10, 100, 101, 1000, 1001, \ldots$ gives back the Fibonacci system.

Example 4. Consider the language $L = \{\varepsilon\} \cup \{a, ab\}^* \cup \{c, cd\}^*$ and the ordering $a < b < c < d$ of the alphabet. If we order the first words in L we get

0	ε	5	cc	10	ccc	15	$aaba$	20	$ccdc$
1	a	6	cd	11	ccd	16	$abaa$	21	$cdcc$
2	c	7	aaa	12	cdc	17	$abab$	22	$cdcd$
3	aa	8	aab	13	$aaaa$	18	$cccc$	23	$aaaaa$
4	ab	9	aba	14	$aaab$	19	$cccd$	24	$aaaab$

Notice that there is no bijection between $\{a, b, c, d\}$ and a set of integers leading to a positional linear numeration system. Otherwise stated, a, b, c, d cannot be identified with usual "digits". For all $n \geq 1$, we have $\mathbf{u}_n(q_{0,L}) = 2F_n$ and $\mathbf{u}_0(q_{0,L}) = 1$. Consequently, for $n \geq 1$,

$$\mathbf{v}_n(q_{0,L}) = 1 + \sum_{i=1}^{n} \mathbf{u}_i(q_{0,L}) = 1 + 2\sum_{i=1}^{n} F_i.$$

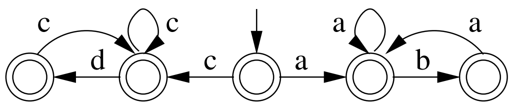

Fig. 1. A DFA accepting L

Notice that for $n \geq 1$, $\mathbf{v}_n(q_{0,L}) - \mathbf{v}_{n-1}(q_{0,L}) = \mathbf{u}_n(q_{0,L}) = 2F_n$. Consequently, by definition of the Fibonacci sequence, we get for all $n \geq 3$,

$$\mathbf{v}_n(q_{0,L}) - \mathbf{v}_{n-1}(q_{0,L}) = (\mathbf{v}_{n-1}(q_{0,L}) - \mathbf{v}_{n-2}(q_{0,L})) + (\mathbf{v}_{n-2}(q_{0,L}) - \mathbf{v}_{n-3}(q_{0,L}))$$

and $\mathbf{v}_n(q_{0,L}) = 2\mathbf{v}_{n-1}(q_{0,L}) - \mathbf{v}_{n-3}(q_{0,L})$, with $\mathbf{v}_0(q_{0,L}) = 1$, $\mathbf{v}_1(q_{0,L}) = 3$, $\mathbf{v}_2(q_{0,L}) = 7$.

Remark 6. The computation given in the previous example to obtain a homogenous linear recurrence relation for the sequence $(\mathbf{v}_j(q_{0,L}))_{j\geq 0}$ can be carried on in general. Let $q \in Q_L$. The sequence $(\mathbf{u}_j(q))_{j\geq 0}$ satisfies a homogenous linear recurrence relation of order t whose characteristic polynomial is the characteristic polynomial of the adjacency matrix of $\mathcal{M}_L$. There exist $a_1, \ldots, a_t \in \mathbb{Z}$ such that for all $j \geq 0$, $\mathbf{u}_{j+t}(q) = a_1\mathbf{u}_{j+t-1}(q) + \cdots + a_t\mathbf{u}_j(q)$. Consequently, we have for all $j \geq 0$, $\mathbf{v}_{j+t+1}(q) - \mathbf{v}_{j+t}(q) = \mathbf{u}_{j+t+1}(q) = a_1(\mathbf{v}_{j+t}(q) - \mathbf{v}_{j+t-1}(q)) + \cdots + a_t(\mathbf{v}_{j+1}(q) - \mathbf{v}_j(q))$. Therefore the sequence $(\mathbf{v}_j(q))_{j\geq 0}$ satisfies a homogenous linear recurrence relation of order $t+1$.

As shown by the following lemma, in an abstract numeration system, the different sequences $(\mathbf{u}_j(q))_{j\geq 0}$, for $q \in Q_L$, are replacing the single sequence $(U_j)_{j\geq 0}$ defining a positional numeration system as in Definition 1.

Lemma 5. *[9] Let $w = \sigma_1 \cdots \sigma_n \in L$. We have*

$$\mathrm{val}_S(w) = \sum_{q\in Q_L} \sum_{i=1}^{|w|} \beta_{q,i}(w)\, \mathbf{u}_{|w|-i}(q) \tag{4}$$

where

$$\beta_{q,i}(w) := \#\{\sigma < \sigma_i \mid \delta_L(q_{0,L}, \sigma_1 \cdots \sigma_{i-1}\sigma) = q\} + \mathbf{1}_{q,q_{0,L}} \tag{5}$$

for $i = 1, \ldots, |w|$.

Recall that $\mathbf{1}_{q,q'}$ is equal to 1 if $q = q'$ and it is equal to 0 otherwise.

Proposition 4. *[9] Let $S = (L, \Sigma, <)$ be an abstract numeration system built over an infinite regular language L over Σ. Any ultimately periodic set X is S-recognizable and a DFA accepting $\mathrm{rep}_S(X)$ can be effectively obtained.*

Recall that an automaton is *trim* if it is accessible and coaccessible (each state can be reached from the initial state and from each state, one can reach a final state).

Proposition 5. *Let $S = (L, \Sigma, <)$ be an abstract numeration system such that for all states q of the trim minimal automaton $\mathcal{M}_L = (Q_L, q_{0,L}, \Sigma, \delta_L, F_L)$ of L,*

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

and $\mathbf{u}_j(q_{0,L}) > 0$ for all $j \geq 0$. If $X \subseteq \mathbb{N}$ is an ultimately periodic set of period p_X, then any deterministic finite automaton accepting $\mathrm{rep}_S(X)$ has at least $\lceil N_{\mathbf{v}}(p_X)/\#Q_L \rceil$ states where $\mathbf{v} = (\mathbf{v}_j(q_{0,L}))_{j \geq 0}$.

To prove this result one has to adapt the arguments given in the proof of Proposition 1 to the framework of abstract numeration systems.

Corollary 2. *Let $S = (L, \Sigma, <)$ be an abstract numeration system having the same properties as in Proposition 5. Assume that the sequence $\mathbf{v} = (\mathbf{v}_j(q_{0,L}))_{j \geq 0}$ is such that*

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

Then the period of an ultimately periodic set $X \subseteq \mathbb{N}$ such that $\mathrm{rep}_S(X)$ is accepted by a DFA with d states is bounded by the smallest integer s_0 such that for all $m \geq s_0$, $N_{\mathbf{v}}(m) > d\,\#Q_L$, where Q_L is the set of states of the (trim) minimal automaton of L.

Proposition 6. *Let $S = (L, \Sigma, <)$ be an abstract numeration system. If $X \subseteq \mathbb{N}$ is an ultimately periodic set of period p_X such that $\mathrm{rep}_S(X)$ is accepted by a DFA with d states, then the preperiod a_X of X is bounded by a constant C depending only on d and p_X.*

To prove this result one has to adapt the arguments given in the proof of Proposition 3 to the framework of abstract numeration systems.

Remark 7. The constant C of the previous result can be effectively computed. Using notation of the previous proof, one has to choose a constant C such that $a_X > C$ implies $|\,\mathrm{rep}_S(a_X - 1)| - d\,\#Q_L > I(p_X)$. Since the abstract numeration system S, the period p_X and the number d of states are given, $I(p_X)$ and $\mathrm{rep}_S(n)$ for all $n \geq 0$ can be effectively computed.

Theorem 2. *Let $S = (L, \Sigma, <)$ be an abstract numeration system such that for all states q of the trim minimal automaton $\mathcal{M}_L = (Q_L, q_{0,L}, \Sigma, \delta_L, F_L)$ of L*

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

and $\mathbf{u}_j(q_{0,L}) > 0$ for all $j \geq 0$. Assume moreover that $\mathbf{v} = (\mathbf{v}_i(q_{0,L}))_{i \geq 0}$ satisfies a linear recurrence relation of the form (1) with $a_k = \pm 1$. It is decidable whether or not a S-recognizable set is ultimately periodic.

Proof. The proof is essentially the same as the one of Theorem 1. Let X be a S-recognizable set and d be the number of states of the minimal automaton of $\mathrm{rep}_S(X)$. With the same reasoning as in the proof of Lemma 3, $\lim_{m \to +\infty}$

$N_{\mathbf{v}}(m) = +\infty$. If X is ultimately periodic, then its period is bounded by a constant t_0 that can be effectively estimated.

If X is ultimately periodic with period $p_X \le t_0$, then using Proposition 6, its preperiod is bounded by a constant (which can also be computed effectively thanks to Remark 7).

Consequently, the sets of admissible periods and preperiods we have to check are finite. Thanks to Proposition 4, one has to build an automaton for each ultimately periodic set corresponding to a pair of admissible preperiods and periods and then compare the accepted language with $\mathrm{rep}_S(X)$.

Example 5. The abstract numeration system given in Example 4 satisfies all the assumptions of the previous theorem.

Acknowledgments

We warmly thank Jacques Sakarovitch for casting a new light on Juha Honkala's paper and also Aviezri Fraenkel for fruitful electronic discussions about linear numeration systems that allowed us to reconsider reasonable hypothesis to tackle this problem.

References

1. Alexeev, B.: Minimal DFAS for testing divisibility. J. Comput. Syst. Sci. 69, 235–243 (2004)
2. Bertrand-Mathis, A.: Comment écrire les nombres entiers dans une base qui n'est pas entière. Acta Math. Acad. Sci. Hungar. 54, 237–241 (1989)
3. Cobham, A.: On the base-dependence of sets of numbers recognizable by finite automata. Math. Systems Theory 3, 186–192 (1969)
4. Durand, F.: A theorem of Cobham for non primitive substitution. Acta Arithmetica 104, 225–241 (2002)
5. Fraenkel, A.: Personal communication (February 2008)
6. Frougny, C., Solomyak, B.: On the representation of integers in linear numeration systems. In: Ergodic theory of Z_d actions (Warwick, 1993–1994). London Math. Soc. Lecture Note Ser, vol. 228, pp. 345–368. Cambridge Univ. Press, Cambridge (1996)
7. Hollande, M.: Greedy numeration systems and regularity. Theory Comput. Syst. 31, 111–133 (1998)
8. Honkala, J.: A decision method for the recognizability of sets defined by number systems. Theoret. Inform. Appl. 20, 395–403 (1986)
9. Lecomte, P.B.A., Rigo, M.: Numeration systems on a regular language. Theory Comput. Syst. 34, 27–44 (2001)
10. Lothaire, M.: Algebraic Combinatorics on Words. Encyclopedia of Mathematics and its Applications, vol. 90. Cambridge University Press, Cambridge (2002)
11. Sakarovitch, J.: Élements de théorie des automates, Vuibert. English translation: Elements of Automata Theory. Cambridge University Press, Cambridge (2003)
12. Shallit, J.: Numeration systems, linear recurrences and regular sets. Inform. and Comput. 113, 331–347 (1994)
13. Zeckendorf, E.: Representation des nombres naturels par une somme des nombres de Fibonacci ou de nombres de Lucas. Bull. Soc. Roy. Sci. Liege 41, 179–182 (1972)